Max Planck Graduate Center
mit der Johannes Gutenberg-Universität

Interactions of the ion channel TRPV4 with regulatory lipids and proteins

An (un)structural study

Dissertation zur Erlangung des Grades

„Doktor der Naturwissenschaften“

Der Fachbereiche:

08 – Physik, Mathematik und Informatik

09 – Chemie, Pharmazie, Geographie und Geowissenschaften

10 – Biologie

Universitätsmedizin

der Johannes-Gutenberg-Universität Mainz

JG|U
JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

Benedikt Goretzki

Geboren in Lich

Mainz, November 2020

Dekan: xxx

1. Gutachterin: xxx

2. Gutachter: xxx

Tag der mündlichen Prüfung: 18.12.2020

I Table of contents

II Abbreviations

2

2-APB ..... *2-aminoethoxydiphenyl borate*

4

4α-PDD *4a-phorbol 12,13-didecanoate*

5

5',6'-EET ..... *5',6'- epoxyeicosatrienoic acid*

A

AA ..... *Arachidonic acid*
ACC ..... *Autocorrelation curve*
ACN ..... *acetonitrile*
ADBE ..... *Activity-dependent bulk endocytosis*
AEA ..... *Arachidonoylethanolamide a.k.a. Anandamide*
AIP4 ..... *Atrophin-1-interacting protein 4*
ANTH ..... *AP180 N-terminal homology domain*
ATP ..... *Adenosine triphosphate*

B

BAA ..... *Bisandrographolide A*
BSA ..... *Bovine serum albumin*

C

CaM ..... *Calmodulin*
CD ..... *Circular dichroism*
cDNA ..... *Complementary DNA*
CFP ..... *Cyan fluorescent protein*
CME ..... *Clathrin-mediated endocytosis*
CRAC ..... *Calcium-release-activated calcium*
CSP ..... *Chemical shift perturbation*
CTD ..... *C-terminal domain*
cyt P450 ..... *Cytochrome P450*

D

ddH_2O ..... *Double-distilled water*
DDX3X ..... *DEAD box RNA helicase*
diC_8-PC *1,2-dioctanoyl-sn-glycero-3-phosphocholine*
diC_8-PG *1,2-dioctanoyl-sn-glycero-3-phosphoglycerol*
diC_8-PI(3,5)P_2 ..... *1,2-dioctanoyl-sn-glycero-3-phospho-(1'-myo-inositol-3',5'-bisphosphate)*
diC_8-PI(4,5)P_2 ..... *1,2-dioctanoyl-sn-glycero-3-phospho-(1'-myo-inositol-4',5'-bisphosphate)*
DkTx ..... *Double-knot toxin*
DMF ..... *Dimethylformamide*
DNA ..... *Deoxyribonucleic acid*
DSS ..... *Discuccinimidyl suberate*
DTT ..... *Dithiothreitol*

E

ECD ..... *Extracellular domain*
EDTA ..... *Ethylenediaminetetraacetic acid*
EHD ..... *Eps15 homology-domain*
ELD ..... *Extraluminal domain*
EM ..... *Electron microscopy*
ER ..... *Endoplasmic reticulum*

F

FA ..... *Formic acid*
F-BAR ..... *Fes/CIP4 homology Bin-Amphiphysin-Rvs*
FDAB ..... *Familial digital arthropathy-brachydactyly*
FRET ..... *Förster resonance energy transfer*

G

GA ..... *Genetic algorithm*
GE ..... *General Electrics™*
GPCR ..... *G-protein coupled receptor*

H

HECT ..... *Homologous to E6-AP C-terminus*
HEK293 ..... *Human embryonic kidney cells*

I

IMAC ..... *Immobilized metal ion chromatography*
IP_3 ..... *Inositol-1,4,5-trisphosphate*
IP_3R ..... *Inositol-1,4,5-trisphosphate receptor*
IPTG ..... *Isopropyl-β-D-thiogalactopyranosid*
ITCH ..... *E3 ubiquitin-protein ligase Itchy homolog*

L

LB ..... *Luria Broth*, *Luria broth*
LUV ..... *Large unilamellar vesicles*

M

MRE ..... *Mean residue ellipticity*
MRW ..... *Mean residue weight*
MWCO ..... *Molecular weight cutoff*

N

NHS ..... *N-hydroxysuccinimide*
Ni-NTA ..... *Nickel(II)-nitrilotriacetic acid*
NMA ..... *Normal mode analysis*
NMR ..... *Nuclear Magnetic Resonance*
NTD ..... *N-terminal domain*
N-WASP ..... *Neural Wiskott-Aldrich Syndrome protein*

O

OD_{600} ..... *Optical density at 600 nm*

P

Pacsin. *Protein kinase C and casein kinase substrate in neurons*
PAR_2 ..... *Protease Activated Receptor*
PBD ..... *Phosphoinositide binding domain*
PCR ..... *Polymer chain reaction*
PDB ..... *Protein data bank*
PET ..... *Photoinduced electron transfer*
PGE ..... *prostaglandin E_2*
PIP_2 ..... *Phosphatidylinositol 4,5-bisphosphate*
PKA ..... *Protein kinase A*
PKC ..... *Protein kinase C*
PLA_2 ..... *Phospholipase A_2*
PLC ..... *Phospholipase C*
PM ..... *Plasma membrane*
PMA ..... *phorbol 12-myristate 13-acetate*
PMSF ..... *Phenylmethylsulfonyl fluoride*
POPC *1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine*
POPG *1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol*
PRR ..... *Proline-rich region*

R

RDC ..... *Residual dipolar couplings*
RNA ..... *Ribonucleic acid*
RT ..... *Room temperature*
RTX ..... *Resiniferatoxin*

S

SA ..... *Simulated annealing*
SASA ..... *Solvent accessible surface area*
SDS-PAGE ..... *Sodium dodecyl sulfate–polyacrylamide gel electrophoresis*
SGK1 ..... *Glucocorticoid protein kinase-1*
SH3 ..... *Src homology domain*
STIM1 ..... *Stromal interaction module 1*
SUMO ..... *Small Ubiquitin Modifier*
SUV ..... *Small unilamellar vesicles*

T

TB ..... *Terrific Broth*
TCEP ..... *Tris(2-carboxyethyl)phosphine*
TFA ..... *Trifluoroacetic acid*
TMD ..... *transmembrane domain*
Tris ..... *Tris(hydroxymethyl)aminomethane*
TRP ..... *Transient Receptor Potential*
TRPA ..... *TRP Ankyrin*
TRPC ..... *TRP Canonical*
TRPM ..... *TRP Melastatin*
TRPML ..... *TRP Mucolipin*
TRPP ..... *TRP Polycystin*
TRPV ..... *TRP Vanilloid*
Tx100 ..... *Triton X-100*

U

Ulp-1 ..... *UBL-specific protease 1*
UV-Vis ..... *Ultraviolet–visible spectroscopy*

W

Wt ..... *Wildtype*

X

XL-MS ..... *Cross-linking mass spectrometry*

Y

YFP ..... *Yellow fluorescent protein*

III Summary

Transient receptor potential vanilloid 4 (TRPV4) is a polymodal eukaryotic cation channel that acts as a transducer of mechanical and osmotic stress, temperature, and pain in multiple human tissues. TRPV4 is regulated by a remarkably diverse spectrum of interaction partners to maintain cellular homeostasis. Perturbances in these regulatory pathways, e.g. through mutations in TRPV4, have been implicated with various sensory and motor neuropathies. For TRPV4 to react to temperature and osmotic stress stimuli, the lipid phosphatidylinositol 4,5-bisphosphate (PIP_2) needs to interact with the TRPV4 phosphoinositide binding domain (PBD). The channel's sensitivity for temperature and osmotic stress decreases when the F-BAR domain protein Pacsin3 binds with its SH3 domain to a proline-rich region (PRR) in the proximity of the PBD. PRR and the PBD are both located in a ~150 residue region preceding an ankyrin repeat domain (ARD) in the cytosolic TRPV4 N-terminus. This region is missing in the currently available high-resolution structure of a full-length TRPV4 channel. Sequence analyses predict it to be an intrinsically disordered region (IDR). Due to the lack of structural information, the molecular mechanism underlying the PIP_2- and Pacsin3-mediated regulation of TRPV4 remains enigmatic.

The results of this thesis provide detailed insights into the structural dynamics of a hitherto uncharacterized region of TRPV4 and how it interacts with channel modulators. A structural and dynamic coupling between ARD and IDR emerges as the mechanistic basis for sensing ligand-binding in the IDR in the remaining channel regions. The binding partners PIP_2 and Pacsin3 may induce counteracting IDR conformations with opposing effects on the ARD/IDR interaction. PIP_2 binding to the PBD leads to an IDR association with the plasma membrane, thereby sensitizing TRPV4. In contrast, Pacsin3 binding to the PRR stabilizes a conformation that tilts the IDR into the cytosol. An exclusive IDR-interaction of PIP_2 or Pacsin3 appears as an elegant explanation of the antagonistic relationship between both channel modulators. Beyond the PIP_2/Pacsin3 mediated regulation of TRPV4, the ARD/IDR communication may link other events in the IDR such as post-translational modifications or channelopathy-mutations to the functional state of the channel. The ARD/IDR communication, thus, forms the basis for a novel mechanism that may explain how the intrinsically unfolded N-terminus in a physiologically important TRP vanilloid member can regulate its channel activity.

Small-angle X-ray scattering (SAXS) and nuclear magnetic resonance (NMR) spectroscopy combined with hydrogen/deuterium-exchange (HDX-MS) and chemical cross-linking mass spectrometry (XL-MS) reveal that the entirely disordered IDR is structurally and dynamically coupled to the ARD. In contrast to the expectations based on high-resolution structures of the isolated ARD, this a-helical domain is a comparably labile protein in solution and seems to adopt transiently unfolded states. In the context of the NTD, the ARD becomes structurally and dynamically stabilized through an interaction with the IDR. Strikingly, the PBD seems to contribute to the IDR-ARD communication and may thus link the dynamics in the ARD to a PBD/PIP_2 interaction and potentially to Pacsin3 binding to the PRR. NMR and fluorescence spectroscopy revealed that the PBD is generally capable of binding anionic lipids and that regions beyond the PBD interact with PIP_2 and other anionic lipids, even though with a lower affinity. This observation indicates that the IDR will be associated with the plasma membrane in the sensitized state of TRPV4, therefore enabling the PBD to bind to a PIP_2 lipid. NMR spectroscopy was further used to characterize the Pacsin3 SH3 domain's interaction with the TRPV4 PRR and the putative cross-talk with the PBD/PIP_2 interaction. Competitive binding experiments show that the Pacsin3 SH3 domain and PIP_2 can bind to their binding sites in the IDR simultaneously. A solution NMR structure of the Pacsin3 SH3 domain in complex with the TRPV4 PRR reveals that Pacsin3 binding requires a proline *trans* to *cis* isomerization in the PRR. This isomerization reorients the PRR and presumably rearranges the preceding IDR. Pacsin3 may antagonize the effects of PIP_2 on TRPV4 by stabilizing an IDR conformation where the PBD loses contact with a PIP_2 molecule in the plasma membrane. Structural analyses of full-length Pacsin3 via SAXS, X-ray crystallography, NMR, and XL-MS suggest that an F-BAR/SH3 domain interaction maintains Pacsin3 in a compact, putatively auto-inhibited, conformation in the absence of a substrate. The binding of the TRPV4 PRR releases the SH3 domain from the F-BAR domain and presumably activates Pacsin3, thus leading to F-BAR domain-induced membrane curvate. Thus, the regulation of TRPV4 by PIP_2 and Pacsin3 appears to be governed by structural rearrangements in the participating protein interaction partners.

IV Zusammenfassung

Transient receptor potential vanilloid 4 (TRPV4) ist ein polymodaler eukaryotischer Kationenkanal, der als Rezeptor von mechanischem und osmotischem Stress, Temperatur und Schmerz in mehreren menschlichen Geweben fungiert. TRPV4 wird durch ein bemerkenswert vielfältiges Spektrum von Interaktionspartnern reguliert, um die zelluläre Homöostase aufrechtzuerhalten. Störungen dieser Regulationswege, z.B. durch Mutationen in TRPV4, sind mit verschiedenen sensorischen und motorischen Neuropathien in Verbindung gebracht worden. Damit TRPV4 auf Temperatur und osmotischen Stress reagieren kann, muss das Lipid Phosphatidylinositol-4,5-bisphosphat (PIP_2) mit der TRPV4-Phosphoinositid-Bindungedomäne (PBD) interagieren. Die Sensitivität für Temperatur und osmotischen Stress nimmt ab, wenn das F-BAR-Domänenprotein Pacsin3 mit seiner SH3-Domäne an eine prolinreiche Region (PRR) in der Nähe der PBD bindet. Die PRR und die PBD befinden sich beide in einer ~150 Aminosäuren umfassenden Region vor einer Ankyrin-Repeat-Domäne (ARD) im zytosolischen TRPV4 N-Terminus. Diese Region ist in der gegenwärtig verfügbaren hochauflösenden Struktur des TRPV4-Kanals nicht enthalten. Sequenzanalysen sagen voraus, dass es sich um eine intrinsisch ungeordnete Region (IDR) handelt. Aufgrund des Mangels an strukturellen Informationen bleibt der molekulare Mechanismus, der der PIP_2- und Pacsin3-vermittelten Regulation von TRPV4 zugrunde liegt, ungeklärt.

Die Ergebnisse dieser Arbeit geben detaillierte Einblicke in die Strukturdynamik einer bisher nicht charakterisierten Region des TRPV4 Kanals und wie sie mit Regulatoren interagiert. Eine strukturelle und dynamische Kopplung zwischen ARD und IDR erweist sich als die mechanistische Grundlage für die Erfassung von Ligandenbindung in der IDR in den übrigen Kanalregionen. Die Bindungspartner PIP_2 und Pacsin3 können womöglich gegenläufige IDR-Konformationen mit gegensätzlichen Auswirkungen auf die ARD/IDR-Interaktion bewirken. Die PIP_2-Bindung an die PBD führt zu einer IDR-Assoziation mit der Plasmamembran, wodurch TRPV4 sensibilisiert wird. Im Gegensatz dazu stabilisiert die Pacsin3-Bindung an die PRR eine Konformation, die die IDR in das Zytosol kippt. Eine exklusive IDR-Interaktion von PIP_2 oder Pacsin3 erscheint als eine elegante Erklärung für die antagonistische Beziehung zwischen den beiden Kanalmodulatoren. Über die PIP_2/Pacsin3-vermittelte Regulierung von TRPV4 hinaus kann die ARD/IDR-Kommunikation vermutlich auch andere Ereignisse in der IDR, wie posttranslationale Modifikationen oder krankheitsauslösende Mutationen, mit dem Funktionszustand des Kanals verknüpfen. Die ARD/IDR-Kommunikation bildet somit die Grundlage für einen neuartigen Mechanismus, der erklären könnte, wie der intrinsisch ungefaltete N-Terminus in einem physiologisch wichtigen TRP-Vanilloid-Mitglied dessen Kanalaktivität regulieren kann.

Röntgenkleinwinkelstreuung (SAXS) und Kernspinresonanzspektroskopie (NMR) kombiniert mit Wasserstoff/Deuterium-Austausch Massenspektrometrie (HDX-MS) sowie chemischer Quervernetzung gekoppelt mit Massenspektrometrie (XL-MS) zeigen, dass die völlig ungeordnete IDR strukturell und dynamisch an die ARD gekoppelt ist. Entgegen den Erwartungen von hochauflösenden Strukturen ist die isolierte ARD ein vergleichsweise labiles Protein in Lösung und scheint transient entfaltete Zustände anzunehmen. Im Kontext der NTD wird die ARD durch eine Interaktion mit der IDR strukturell und dynamisch stabilisiert. Auffälligerweise scheint die PBD zur IDR-ARD-Kommunikation beizutragen und könnte so die Dynamik in der ARD mit einer PBD/PIP_2-Interaktion und möglicherweise auch mit Pacsin3 verknüpfen. Mittels NMR und Fluoreszenzspektroskopie konnte gezeigt werden, dass die PBD generell in der Lage ist, anionische Lipide zu binden, und dass Regionen jenseits der PBD mit PIP_2 und anderen anionischen Lipiden interagieren, wenn auch mit einer geringeren Affinität. Diese Beobachtung lässt vermuten, dass die IDR im sensitivierten Zustand von TRPV4 mit der Plasmamembran assoziiert, so dass die PBD an ein PIP_2-Lipid binden kann. NMR-Spektroskopie wurde weiterhin verwendet, um die Wechselwirkung der Pacsin3-SH3-Domäne mit der TRPV4-PRR und eine mögliche Überschneidung mit der PBD/PIP_2-Interaktion zu charakterisieren. Kompetitive Bindungsexperimente zeigen, dass die Pacsin3-SH3-Domäne und PIP_2 gleichzeitig an ihre Bindungsstellen in der IDR binden können. Durch die Bestimmung einer NMR-Struktur der Pacsin3-SH3-Domäne im Komplex mit der TRPV4-PRR konnte gezeigt werden, dass die Pacsin3-Bindung eine *trans*-zu-*cis*-Prolinisomerisierung in der PRR erfordert. Diese Isomerisierung richtet die PRR neu aus und ordnet vermutlich die vorhergehende IDR um. Pacsin3 kann womöglich die Sensitivierung von TRPV4 durch PIP_2 antagonisieren, indem es eine IDR-Konformation stabilisiert, bei der die PBD den Kontakt mit einem PIP_2-Molekül in der Plasmamembran verliert. Die Strukturanalyse des Volllänge Pacsin3 Proteins mittels SAXS, Röntgenkristallographie, NMR und XL-MS lässt vermuten, dass eine F-BAR/SH3-Domänen-Wechselwirkung Pacsin3 in Abwesenheit eines Substrats in einer kompakten, womöglich autoinhibierten Konformation hält. Die Bindung der TRPV4-PRR setzt die SH3-Domäne aus der F-BAR-Domäne frei und aktiviert Pacsin3, sodass die F-BAR domäne Membrankrümmung induzieren kann. Somit scheint die Regulation von TRPV4 durch Pacsin3 durch strukturelle Umlagerungen in beiden Interaktionspartnern bestimmt zu werden.

V Publications and supervised student projects

Publications

The published studies and manuscripts in preparation listed below were implemented into this thesis.

Peer-reviewed publications (**equal contribution*)

[1] Benedikt Goretzki*, Nina A. Glogowski*, Erika Diehl, Elke Duchardt-Ferner, Carolin Hacker, Rachelle Gaudet, Ute A. Hellmich (2018) Structural Basis of TRPV4 N Terminus Interaction with Syndapin/PACSIN1-3 and PIP_2. ***Structure*** *26*, 1-11, doi: 10.1016/j.str.2018.08.002

[2] Dania Rose-Sperling, Mai A. Tran, Luca M. Lauth, Benedikt Goretzki, Ute A. Hellmich (2019). ^{19}F NMR as a versatile tool to study membrane protein structure and dynamics. ***Biological Chemistry***, 400(10), pp. 1277-1288, doi:10.1515/hsz-2018-0473

[3] Julia C. Heiby*, Benedikt Goretzki*, Christopher M. Johnson, Ute A. Hellmich & Hannes Neuweiler (2019), Methionine in a protein hydrophobic core drives tight interactions required for assembly of spider silk; ***Nature Communications 10***, 4378, doi:10.1038/s41467-019-12365-5

[4] Benedikt Goretzki, Julia C. Heiby, Hannes Neuweiler & Ute A. Hellmich (2019). NMR Backbone and Side Chain Assignment of a Mutant *E. australis* Major Ampullate Spidroin 1 N-terminal Domain. ***Biomolecular NMR Assignments***, doi.org/10.1007/s12104-019-09922-w

[5] Brett A. McCray, Erika Diehl, Jeremy M. Sullivan, William H. Aisenberg, Nicholas W. Zaccor, Alexander R. Lau, Dominick J. Rich, Benedikt Goretzki, Ute A. Hellmich, Thomas E. Lloyd, and Charlotte J. Sumner (2020). Neuropathy-causing TRPV4 mutations disrupt TRPV4-RhoA interactions and impair neurite extension. ***Nature Communications***, in revision

Manuscripts in preparation (**equal contribution*)

[6] Benedikt Goretzki, Charlotte Guhl, Frederike Tebbe & Ute A. Hellmich (in prep) Intrinsically disordered regions in TRP ion channels. ***J. Mol. Biol.***, invited contribution, Special Edition "Ion Channels", Eds.: Vera Moiseenkoval-Bell, Daniel L. Minor, Lucie Delemotte

[7] Benedikt Goretzki, Julia Nöth, Sarah-Ana Mitrovic & Ute A. Hellmich (in prep). NMR Backbone Assignment of the intrinsically disordered N-terminus of TRPV4.

[8] Benedikt Goretzki*, Frederike Tebbe*, Julia Nöth, Jasmin Jansen, Sascha Schmitt, Sarah-Ana Mitrovic, Cy Jeffries, Kaloian Koynov, Hans-Jürgen Butt, Florian Stengel & Ute A. Hellmich (in prep). Structural dynamics, ankyrin repeat domain and lipid interactions of the N-terminal intrinsically disordered region of the TRPV4 ion channel.

[9] Benedikt Goretzki*, Frederike Tebbe*, Julia Nöth, Jasmin Jansen, Cy Jeffries, Florian Stengel, Albert Guskov & Ute A. Hellmich (in prep). Intramolecular interaction between the SH3 and F-BAR domains of Syndapin/PACSIN3 is released by the TRPV4 ion channel N-terminus.

[10] Charlotte Guhl*, Tina Berger*, Benedikt Goretzki*, Daria Maltsevad, Yu-Kai Chao, Christian Grimm, Grazia Gonella & Ute A. Hellmich (in prep). Structural basis of PIP_2 binding in the disordered TRPML1 N-terminus.

Supervised student projects which contributed to this thesis

Erika Pfeiffer, MSc Biomedizinische Chemie (2016), *Mass spectroscopic elucidation of the TRPV4 N-terminus interactome*. Master thesis

Sarah-Ana Mitrovic (2017), *Structural studies on the TRPV4 intrinsically disordered N-terminus*. Bachelor thesis

Dorothea Winkelvoß (2017), *NMR spectroscopic studies of Pacsin SH3 domain interactions*. Internship

Carolyn Barnes (2018). *Fluorescence correlation spectroscopic studies of protein-lipid interactions*. **DAAD RISE Internship**

Axel Guthart (2018). *Structural and functional studies of PACSIN3*. Bachelor thesis

Simon Schneider (2018). *^{19}F- and fluorescence-labelling of recombinant proteins*. Internship.

Nina Jacobs (2019), *NMR spectroscopic studies of TRPV4-protein and TRPV4-lipid interactions*. Bachelor thesis (**Award of the Gutenberg Lehrkolleg for her outstanding Bachelor thesis**)

Julia Nöth (2019). *Structural and functional investigations of the TRPV4 intrinsically disordered N-terminus*. Master thesis

Ilja Gordijenko (2019). *Lipid-binding studies of PACSIN3*. Bachelor thesis

Sarah-Ana Mitrovic (2019). *Structural investigations of the TRPV4 intrinsically disordered N-terminus & structural studies of MIP-inhibitor interactions*. Master thesis (**Award by the German Biochemical Society (GBM) for her outstanding Master thesis**)

Nicolas Grinberg (2019). *NMR protein backbone assignments of intrinsically disordered proteins*. Internship**, Schull-Yang International Experience Award, McGill University**

Nina Jacobs (2020). *NMR spectroscopic studies on the lipid interaction of the intrinsically disordered TRPML1 N-terminus*. Internship

VI Interactions of the ion channel TRPV4 with regulatory lipids and proteins – an (un)structural study

1 Introduction

1.1 The transient receptor potential (TRP) ion channel family

In 1969, Cosens and Manning identified a spontaneous mutation in the fruit fly *Drosophila melanogaster*, causing irregularities in the retina's electric activity in response to a light stimulus (*1*). Instead of a steady electrical response evoked by a sustained light pulse, the mutation provoked a transient response that ultimately impaired the receptor potential of photoreceptor cells. Accordingly, the mutant was termed transient receptor potential, TRP, and the corresponding *trp* gene was later found to encode a calcium selective ion channel (*2*). Since then, many ion channels have been identified, mostly in mammals, which show similarities with the *D. melanogaster* TRP channel and gave rise to the TRP ion channel superfamily. The superfamily is divided into six subfamilies in mammals where it encompasses a total of 28 members: TRPC (Canonical, seven members), TRPM (Melastatin, eight members), TRPV (Vanilloid, six members), TRPA (Ankyrin, one member), TRPML (Mucolipin, three members), and TRPP (Polycystin, three members). TRP channels play vital roles in an impressively wide variety of physiological functions. The most salient feature of most TRP channels is their central role as molecular receptors of chemical and physical stimuli in primary sensory neurons (*3*). They participate in photo- (*4*), chemo- (*5*), thermo- (*6, 7*), mechano- and osmosensation (*8*), nociception (*9*), and other sensory modalities. Most TRP channels are Ca^{2+} selective ion channels, but many also gate monovalent and other divalent cations (*10*). Some TRPs appear to respond mainly to a single stimulus. However, many of them show a remarkable polymodality, i.e. the ability to respond to very diverse and seemingly unrelated stimuli such as natural and artificial compounds, voltage, temperature or cell swelling, to name just a few (*11*). The molecular mechanisms responsible for the polymodal gating of TRP channels is unclear to this date. Malfunctioning TRP channels, e.g. because of mutations, are increasingly linked to a variety of diseases, such as skeletal dysplasia, respiratory diseases, several types of cancer, and diseases of the nervous system (*12*).

The TRP channel superfamily exhibits an enormous structural and functional diversity which has made the classification of TRP channels a cumbersome process (*13, 14*). In fact, the exact functional roles of several TRP channels have been unknown for a long time (and partly remained so), and the molecular structures of most TRP channels have been enigmatic until only very recently. Therefore, the commonly accepted classification of TRP subfamilies is solely based on sequence similarity (*15*). Within the TRP channel superfamily, sequence similarity and functional roles do not seem to correlate as it is known for other protein families. For instance, members of the same TRP subfamily can have very different functions, whereas members of different TRP subfamilies can have overlapping functions. Nevertheless, the classification based on sequence similarity turned out to group the overall structurally diverse TRP superfamily according to very specific and characteristic structural features. For more than a decade, the structural characterization of TRP channels at the atomic level via X-ray crystallography (*16–28*) or NMR spectroscopy (*29–31*) was limited mainly to cytosolic domains or fragments in complex with a ligand. Early cryo-electron microscopy (EM) studies on full-length TRP channels could not provide detailed insights into the channel architecture and remained limited to low-resolution models (*32–37*). The structural understanding of TRP channels, however, has made an enormous leap with the onset of the "resolution revolution" (*38*), which was initiated by the high-resolution structure determination of TRPV1 via single-particle cryoEM in 2013 (*39, 40*). To date, the number of published cryo-EM structures of TRP channels in apo or ligand-bound states has culminated to 136*, and high-resolution structures of at least

* This number refers to high-resolution structures deposited in the PBD between December 2013 to October 2020. Note: TRPV6 and TRPV4 are the only TRP channels for which high-resolution structures could be determined via X-ray crystallography (*41, 42*).

one TRP channel from each mammalian subfamily have been determined (*43*). These structures provide unprecedented insights into the structural diversity of TRP channels and the molecular mechanisms underlying their function. However, all currently available TRP channel structures are incomplete, thus leaving many open questions regarding the molecular details of TRP channel function.

All TRP channels have in common that they assemble as tetramers to form a functional channel. They share relatively conserved transmembrane domains (TMDs) which are structurally related to the TMDs of voltage-gated potassium channels (K_V) (Figure 1 A and B) (*44*). The TMD of one TRP subunit comprises six transmembrane helices, termed S1-S6. S1-S4 form a helix bundle similar to the voltage sensor domain in K_V channels. This structural element is accordingly termed the voltage sensor-like domain (VSLD). The helices S5 and S6 together with one or in some subfamilies two short helices (P) between S5 and S6 form the pore domain (Figure 1 A). In a TRP channel tetramer, the pore domains of neighbouring subunits are intertwined to form a channel for cations (Figure 1 B). The VSLD and the pore region are connected by the S4-S5 linker (Figure 1 A). In most TRP channels, the S4-S5 linker forms a helix that is tilted almost perpendicular to the membrane. As in K_V channels, the movement of the S4-S5 linker is directly coupled to movements of the pore domain (*45, 46*). Subtle differences in the pore diameters between TRP channels determine their permeability for specific ions and fine-tune their ion gating mechanisms (*43*). Multiple high-resolution cryo-EM studies of full-length TRP channels have located lipid-binding sites in the TMDs which might either be required to maintain the protein structural integrity or serve as channel regulators (*47–51*).

The structural diversity between the TRP channel subfamilies predominantly manifests in the intra- and extracytosolic domains (Figure 1 C and D). These domains are crucial for regulating the opening and closing of the pore and are primarily responsible for the diverse functions of TRP channels by conferring peripheral signals to the central ion channel pore. They serve as binding sites for proteins, lipids, nucleotides, ions, small molecules or couple channels to the cytoskeleton (*52*). Based on structural criteria, the six mammalian TRP subfamilies can in fact be further divided into two groups (Figure 1 C and D). Group I comprises the subfamilies TRPM, TRPA, TRPV, and TRPC. All group I TRP channels contain large cytosolic domains in the N-terminus. A prominent structural element in TRPA, TRPC, and TRPV channels are ankyrin repeat domains (ARDs) (*48, 53*). These α-helical motifs have been described to serve as binding hubs for protein interaction partners and contribute to channel assembly (*8, 54*). A linker region connects the ARDs to a pre-S1 helix which is directly connected to the first helix of the TMD (*40, 51, 53, 55*). TRPM channels possess a unique N-terminal domain which was termed the Melastatin homology region (MHR) (*56–59*). It is connected to the pre-S1 helix in the C-terminus. TRPM and TRPC channels contain a helix between the pre-S1 and S1 helix, termed the pre-S1 elbow, which inserts into the membrane. C-terminal to the TMD, a group 1 TRP channel-specific helix is positioned below the VSLD and the pore domain (*40, 51, 53, 55*). This so-called TRP box is oriented perpendicular to the membrane and contacts the S4-S5 linker of the TMD and the linker regions of TRPA, TRPC, TRPV or the MHR of TRPM in the N-terminal domains (Figure 1 C and D). In this position, the TRP box is perfectly positioned to mediate structural changes in the large cytosolic domains to the pore by acting as a lever on the S4-S5 linker region (*60, 61*). Several additional domains have been reported to succeed the TRP box such as a helix termed the post-TRP elbow in TRPM and TRPC, which together with the pre-S1 elbow inserts into the membrane (*48, 51*). This post-TRP elbow is followed by a so-called rib helix (*48*) and a coiled coil domain succeeding which is essential for channel assembly (*25, 26*). A coiled coil domain is also present in the C-terminus of TRPA (*53*). In TRPV channels, the TRP box is succeeded by a C-terminal domain (CTD) which interacts with the N-terminal domain to form a β-sheet structure at the cytosolic subunit interfaces (Figure 1 C and D) (*40*).

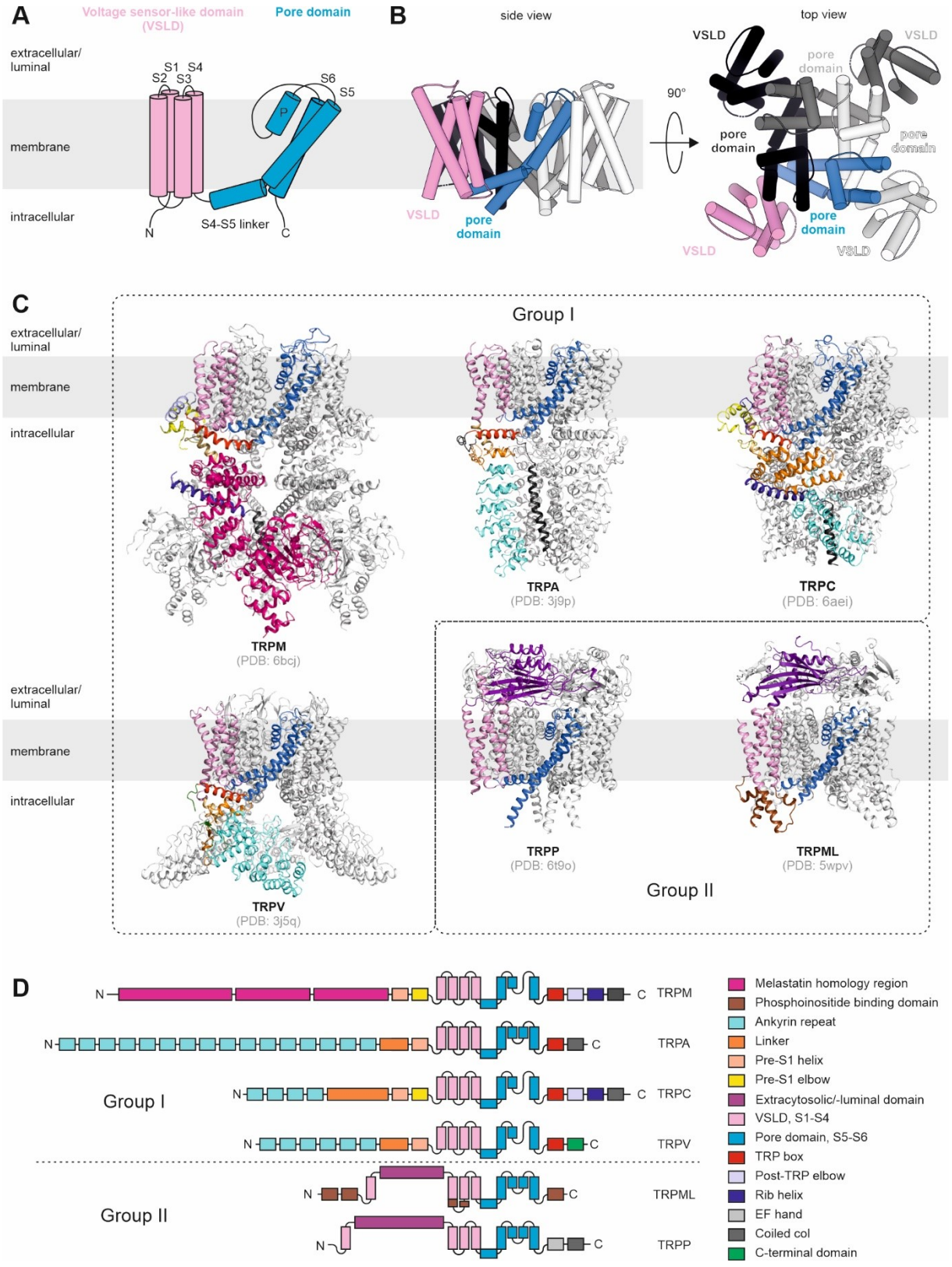


Figure 1: Structural characteristics of the mammalian TRP channel superfamily. Topology of the conserved TRP channel transmembrane domain (TMD) showing the voltage sensor-like domain, VSLD (S1-S4), in pink and the pore domain (S5-P-S6) in blue. Depending on the TRP channel family, one or two pore helices (P) can exist between S5 and S6. (**B**) Tetrameric assembly of the TMD viewed from the side and the top. (**C**) Representative high-resolution structures of group I and group II TRP channels from the six mammalian TRP subfamilies. Group I: TRPM4 (PDB:6bcj), TRPA1 (PDB: 3j9p), TRPC5 (PDB. 6aei), TRPV1 (PDB: 3j5q); group II: TRPP2 (PDB: 6t9o), TRPML1 (PDB: 5wpv). The representative TRP subfamily members were chosen based on the completeness of their high-resolution structures, i.e. how well subfamily-specific domains are resolved. (**D**) See next page.

Caption of **Figure 1 D**: Topology representation of the general domain architecture found in TRP channel subfamilies. See figure for colour coding and legend. (D) was modified from ref. (*43*). Importantly, the topology models only show regions resolved high-resolution structures or expected to be structured due to sequence homology with resolved regions (e.g. the N-terminal ankyrin repeats from TRPA).

Group II encompasses the TRPML and TRPP subfamilies (Figure 1 C and D). They contain large domains between S1 and S2 (*62–65*). TRPP channels primarily localize in the plasma membrane (PM) and in the endoplasmic reticulum (ER) (*66*), whereas TRPML channels are predominantly found in the lysosome (but also in the PM) (*67*). Therefore, the domain between S1 and S2 is called the extracytosolic domain (ECD) in TRPP channels, whereas it also termed the extraluminal domain (ELD) in TRPML channels. Besides the ECD/ELD, group II TRP channels differ from the group I members by the absence of large N-terminal domains or a TRP box in the C-terminus (Figure 1 C and D). TRPML channels contain an intracellular phosphoinositide binding domain (PBD) which is formed by extensions of TMD helices S1, S2, S3, and S6 (*68*). The PBD is crucial for regulating TRPML activity in the lysosomal pathway (*69*). For TRPP channels, no structural motifs in the N-terminus have been described yet. However, they possess an EF-hand in the C-terminus, which provides an element for calcium-dependent channel regulation (*70–72*). Following the EF-hand, a coiled coil domain is supposed to assist channel assembly and regulation (*16, 73*).

The extreme pace at which new TRP channel cryo-EM structures are being deposited in the PDB might suggest that the major obstacles to understanding TRP channels at the atomic level have been overcome. However, the inherent difficulties of cryo-EM to resolve flexible protein regions have immensely hampered the determination of 'truly' full-length TRP channels (*74*). Amino acid sequence analysis suggests that most TRP channels contain large intrinsically disordered N- and C-termini (Figure 2). These regions were either deleted in the constructs used for structure determination or not resolved in the cryo-EM density maps. The proclaimed full-length TRP channel structures in the PDB should therefore rather be considered as 'near-full length' structures. Various studies have demonstrated that the putatively disordered N- and C-termini of TRP channels contain binding sites for regulatory lipids and proteins (*75–82*). Due to the lack of structural information, little is known yet about how ligand binding in the distal N- and C-termini is allosterically coupled to conformational changes in the remainder of the channel. It has been suggested that intrinsically disordered regions may serve to expand and diversify a membrane protein's interactome (*83*). One might speculate that the structural and functional variability conferred by intrinsically disordered regions is crucial for the polymodality of TRP channels. Among the TRP superfamily, members of the vanilloid subfamily stand out due to their exceptional polymodality (*3*). Intriguingly, these channels exhibit large disordered N- and C-termini, which are emerging as crucial regulatory domains.

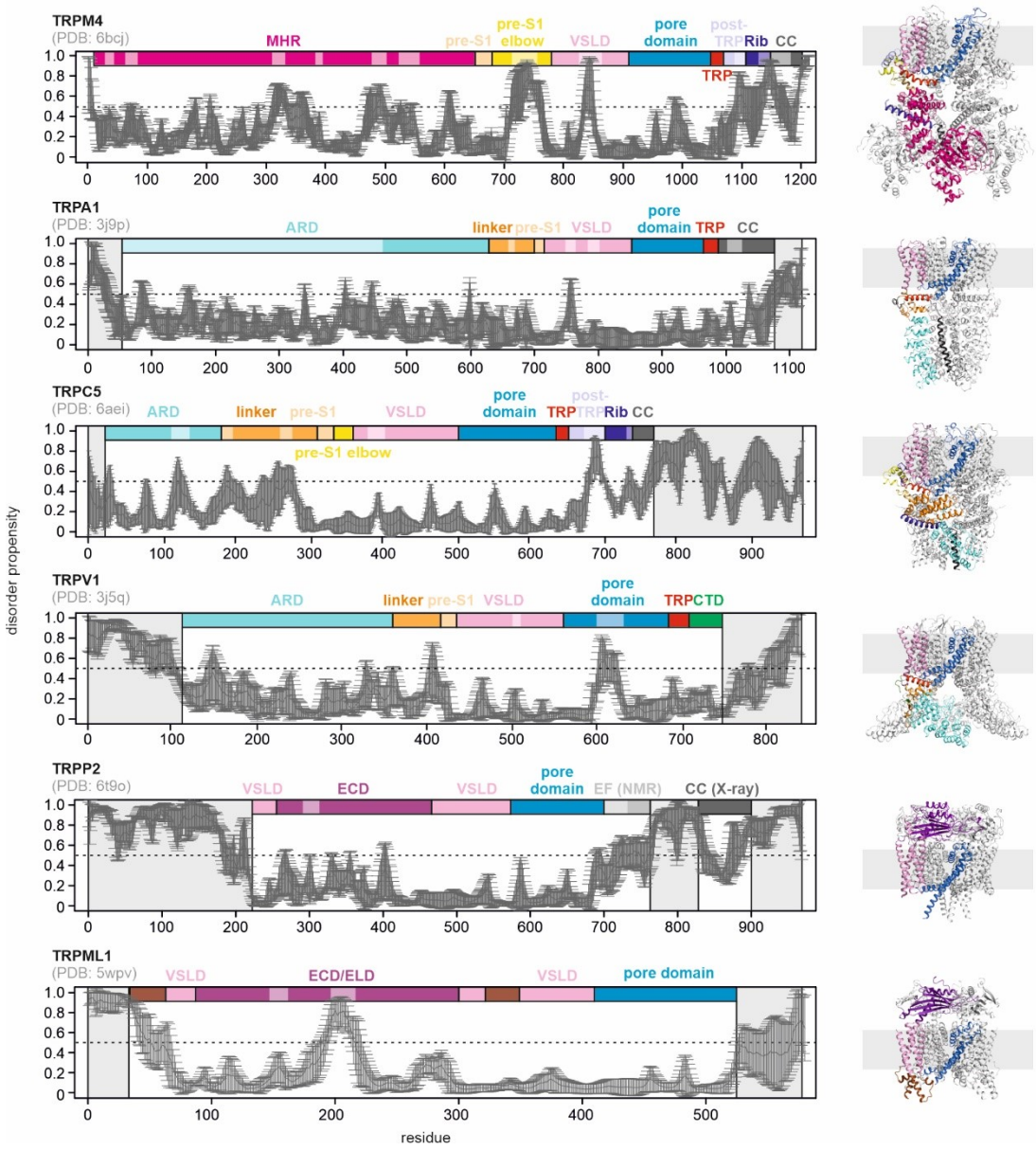


Figure 2: Protein disorder in the N- and C-terminal regions of TRP channels. Amino acid sequence-based disorder predictions of representative TRP subfamily members (same as shown in Figure 1) based on IUPRED *(84)*, PONDR *(85–90)*, PONDR-FIT *(91)*, and DISOPRED *(92)*. Plotted is the disorder propensity averaged over all prediction algorithms. The error bars represent the standard deviation from the mean value across all four prediction programs (?). Values larger than 0.5 indicate disordered protein regions. The representative members of each subfamily are TRPM4 (PDB:6bcj), TRPA1 (PDB: 3j9p), TRPC5 (PDB. 6aei), TRPV1 (PDB: 3j5q); group II: TRPP2 (PDB: 6t9o), TRPML1 (PDB: 5wpv). The topology of the channels, as is observed in the cryo-EM structures (on the right), is shown as coloured boxes on the top of each diagram. Regions that were not resolved in the cryo-EM structures are shown with 50% transparency in the topology diagrams. Where applicable, N- and C-terminal disordered regions are highlighted in light grey. These were either unresolved in the cryo-EM structures or not included in the constructs used for structure determination. In TRPP2, the EF-hand and the CC structures were determined in isolation via NMR or X-ray crystallography, respectively.

1.2 The TRP vanilloid subfamily

The transient receptor potential vanilloid subfamily has six members, termed TRPV1-6, and is further divided into two groups. Group I contains TRPV1-4 (Figure 3 A), and group II contains TRPV5 and 6 (Figure 3 B). All group I TRP vanilloids are thermoTRPVs, i.e. they respond to temperature stimuli (*93*) (Figure 3 A). They are widely expressed, e.g. in primary sensory neurons, dorsal root ganglia, trigeminal ganglia, testis, spleen,

intestine, bladder, brain, heart, liver, endothelia, and others (*94–99*). TRPV1-4 are mostly known for being involved in nociception and thermal reception. They are mildly calcium selective and exceptionally polymodal (*3*). Amongst the thermoTRPVs, TRP vanilloid 1, also known as the capsaicin receptor, is the most famous and best-characterized member (*100, 101*). It was not only the first discovered TRPV channel but also the first TRP channel in general for which a 'near full-length' high-resolution structure could be determined (*40*). TRPV1 is an ideal example of the polymodal gating behaviour of the thermoTRPVs.

TRPV1 is activated by temperatures above 43°C (*100, 101*). It responds to pungent and toxic chemical compounds such as capsaicin (from chili) (*100*), resiniferatoxin (RTX, cactus) (*102*), piperine (pepper) (*103*), allicin (garlic) (*104*), a double-knot toxin (DkTx) from a tarantula (*105*), as well as to compounds with analgesic effects such as the endocannabinoid anandamide or camphor (*106, 107*). TRPV1-mediated Ca^{2+} influx upon heat stimuli is enhanced by low pH (*108*). The shift in the TRPV1 heat-activation threshold and the potentiated responses towards capsaicin is also caused by ethanol (*109*), nicotine (*110*) and proinflammatory cytokines (*111*). Additionally, TRPV1-mediated Ca^{2+}-currents are potentiated by a decrease in PIP_2 levels in the plasma membrane (*112*) and via phosphorylation by protein kinase C (PKC) (*113*). Both mechanisms are preceded by the activation of phospholipase C (PLC) through proalgesic agents such as bradykinin and nerve growth factor (*114*). These examples impressively demonstrate how TRPV1 can integrate multiple signals invoked by rather different noxious stimuli. The other thermoTRPVs display similarly polymodal gating behaviours. TRPV2 is activated by noxious heat >52°C (*115*). It is functionally upregulated by phosphatidylinositol-3-kinase activation (*116*) and hypotonicity-induced cell swelling (*117*). TRPV3 and TRPV4 respond to warm temperatures in the ranges 33-39°C (*118*)and 27-34°C (*119*), respectively. TRPV3 activity is modulated by PLC (*120*), by polyunsaturated fatty acids such as arachidonic acids (*121*), by several monoterpenoids (*122*), and by compounds of spices such as oregano, cloves, and thymes (*123*). TRPV4 responds to cell swelling (*124, 125*), lipid metabolites (*126*), or phorbol esters (*127*), and its activity is sensitive to the membrane environment (*128, 129*).

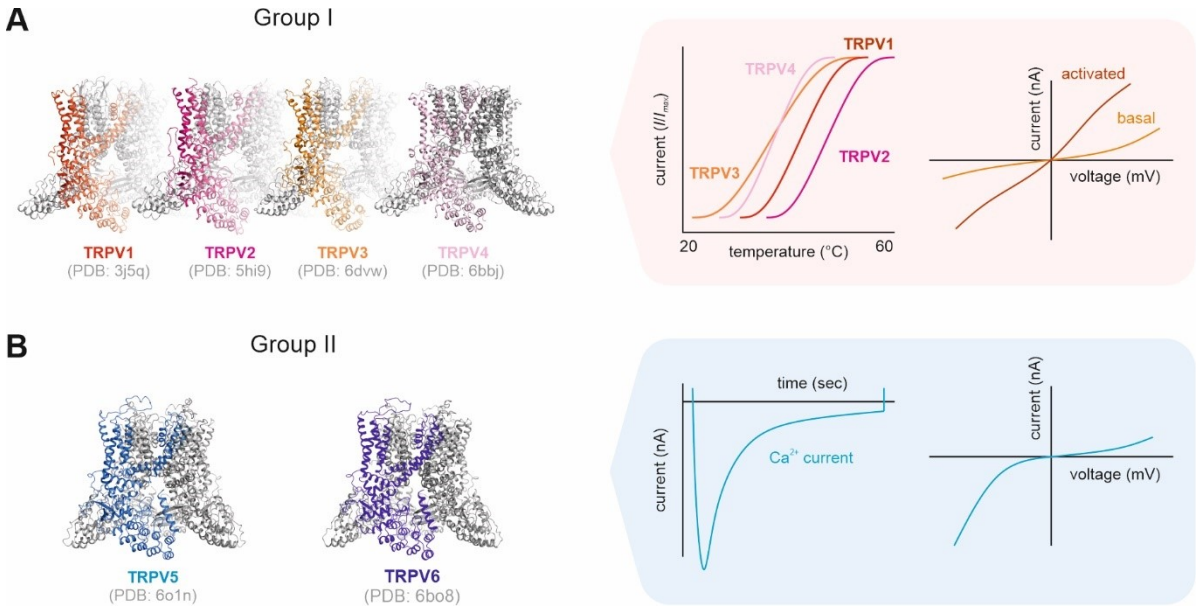


Figure 3: Functional characteristics of group I and group II TRP vanilloid channels (**A**) TRPV1, 2, 3, and 4 belong to the group of thermosensitive TRP channels (thermoTRPVs). ThermoTRPVs are non-selective cation channels and exhibit a low basal activity. They are activated in response to physical stimuli, e.g. by temperature or by the binding of pungent and toxic agents. Each thermoTRPV responds to an individual temperature range. However, the activation temperatures of all thermoTRPVs overlap and span a range between mild and high temperatures. (**B**) TRP vanilloid group II comprises TRPV5 and V6. These proteins are highly selective for Ca^{2+}, likely not ligand-gated, and supposed to be constitutively active. A negative feedback loop upon increased intracellular calcium concentrations followed by calmodulin-binding governs their regulation. TRPV5/6 show inward rectifying ion gating behaviour.

ThermoTRPVs exhibit a low basal ion channel activity (Figure 3 A) which is potentiated by external stimuli. This gating behaviour is in stark contrast to that of TRPV5 and V6, the second group of TRPV channels (Figure 3 B). TRPV5/6 are mainly expressed in the kidney, where they regulate calcium reabsorption (*130*). They are likely not ligand-gated and are constitutively active under normal physiological conditions. In contrast to the thermoTRPVs, TRPV5/6 are calcium-selective and display an inward rectifying ion gating behaviour (*130*) (Figure 3 B). They are regulated by a negative feedback loop with calmodulin (CaM). When intracellular calcium-levels increase, Ca^{2+}/CaM binds and blocks the pore on the cytosolic side of the channel, thus ablating inward Ca^{2+}-currents (Figure 3 B) (*131–133*).

The structure of TRPV channels is highly conserved

To date, high-resolution structures have been determined for all vanilloid members (Figure 4 A) (*40–42, 49, 50, 131, 134–149*). Even though the TRPV channels display quite diverse gating properties, they share a conserved channel architecture. The most characteristic structural feature is the ankyrin repeat domain with six ankyrin repeats (AR) in the cytosolic N-terminus (Figure 4 B). One AR is composed of an inner and an outer helix. In the case of TRPV channels, the inner helix points towards the membrane, the outer helix points into the cytosol. A long loop connects the outer helix of one AR with the inner helix of a subsequent AR. The loops stick out from the ARD as 'fingers'. As a key functional element, the ARD is essential for channel assembly and has been implicated in channel regulation by ATP (*18, 56*), lipids (*150*), and CaM (*78, 151*). Moreover, it is a hot spot for splice variants (*152*) and disease-causing mutations (*12*). The ARD connects to the transmembrane domain (TMD) by a linker region containing three helices (LH1-3) and a β-sheet with two strands (β1 and β2) followed by a pre-S1 helix (Figure 4 B). The linker region and the pre-S1 helix have recently been implicated in mediating regulatory interactions to lipids and proteins (*153*).

The TRP vanilloid TMDs display the TRP channel-characteristic architecture consisting of the VSLD and the pore domain. The VSLD has been implicated in coupling the binding of ligands to the pore domain during channel gating. More recently, it was identified as a crucial structural feature for thermosensation in TRPV1 (*154*). The VSLD is connected to the pore domain via the S4-S5 linker, which in most TRPV channels, forms an α-helix almost perpendicular to the membrane (Figure 4 B). The pore domain has two main restriction points that are supposed to gate TRPV channels (Figure 4 C). The first restriction point, termed the selectivity filter, is at the upper part of the pore domain in the pore helix (P) and the loop between P and S6. It typically comprises a stretch of four amino acids with at least one negatively charged side chain that recruits and coordinates cations. The diameter of the selectivity filter determines the ion selectivity. The selectivity filter in group I TRPV channels has a G(M/L)G(D/E) motif and a TIID motif in group II TRPVs(Figure 4 C). The second restriction point composed of isoleucine and methionine is located towards the end of S6 and forms the lower gate of the ion-conducting pore. The residues in the lower gate are largely conserved between TRPV1-6. However, TRPV5/6 contain a tryptophan residue at the end of S6 lining the pore exit in the cytosolic side of the channel (Figure 4 C). This tryptophan residue is absent in TRPV1-4. It has been implicated in the CaM-dependent inactivation mechanism of TRPV5 and TRPV6 (*155*).

At the C-terminus of the TMD, the group I TRP channel characteristic TRP box is positioned perpendicular to the membrane (Figure 4 C). The TRP box was proposed to act as a lever on the S4-S5 linker, which connects the pore domain to the VSLD (*60, 61*). Accordingly, movement of the TRP box directly controls the movement of the pore helices and thereby opens or closes the ion-conducting pore. Importantly, the TRP box makes contact with the pre-S1 helix and the linker domain in the N-terminus (*156*). This interaction network provides for a conformational coupling of the ARD with the pore region. Conformational changes in the ARD upon ligand binding likely propagate via the linker region and pre-S1 helix to the TRP box which controls the pore domain via the S4-S5 linker. C-terminal to the TRP box, the peptide chain first forms a disordered loop and then a β-strand (β3) which assembles as a β-sheet together with β1 and β2 from the linker region (Figure 4 C). This β-sheet makes contacts to an ARD of a neighbouring subunit and is therefore ideal to couple conformational changes in the neighbouring subunit to the TRP box and the pore region (Figure 4 D). In this sense, the

TRP box seems to play a dual role by coupling conformational changes in the ARD of the same subunit and the ARD of the neighbouring subunit to the pore region (*155*).

Despite their different functions, only subtle differences exist between the thermoTRPV and TRPV5/6 structures. The most obvious characteristic of the TRPV5/6 structures is an N-terminal helix which directly precedes and folds onto the ARD (Figure 4 D) (*41, 148*). The TRPV5/6 structures also display a longer linker between S1-S2 compared to TRPV1-4 (Figure 4 D). The extended S1-S2 linker makes contacts with the pore domain of an adjacent subunit (*149*). The thermoTRPVs TRPV1, V2 and V4, in contrast to TRPV3, V5, and V6, possess a 'pore turret', an extended loop between S5 and the pore helix, in the pore domain (Figure 4 D). The pore turret has been implicated as a part of the temperature sensing apparatus; however, its exact role is still under debate (*147, 157, 158*). A recent functional and computational study of TRPV1 orthologs suggests that the exact composition of the pore turret fine-tunes the heat-activation thresholds of TRPV1 channels (*159*). The pore turret also seems to dictate channel gating of TRPV1 in response to stimuli such as the binding of the agonists DkTx and capsaicin (*160*)

A large number of high-resolution structures of TRPV channels in apo and liganded states, showing closed or open pore conformations, provides detailed insights into their gating mechanisms (*20–24, 39–42, 49, 50, 78, 131, 134–150, 161–163*). A comparison of the available structures of TRPV1-4 suggests that the pore diameter of the upper gate, i.e. the selectivity filter, is quite flexible in TRPV1-4. In contrast, the selectivity filter is rather rigid in TRPV5 and V6. The reduced TRPV5/6 selectivity filter flexibility has been linked to the extended S1-S2 linker (Figure 4 D) and bulky phenylalanine residues in S6, which are not present in TRPV1-4. The S1-S2 linker in TRPV5/6 makes excessive contacts with the pore domain of an adjacent subunit, thus potentially stabilizing the conformation of the selectivity filter. This might suggest that ion selectivity in TRPV channels is determined by the pore flexibility (*155*). Ion permeation through the pores of TRPV channels not only depends on the selectivity filter but also requires the movement of S6 to open the lower gate (Figure 4 D). The exact S6 movement seems to differ between the different TRPV subtypes and on the agonist used to activate them. In TRPV2, V3, and V6, an α-to-π-helix transition upon movement of S6 was observed to expose several polar and acidic side chains in the lower gate and open the pore (*135, 136, 143*). The residues involved in the α-to-π-helix transition are conserved in the TRPV family. A similar transition was, however, not observed in the structures of TRPV1 and TRPV5 (*49, 149*).

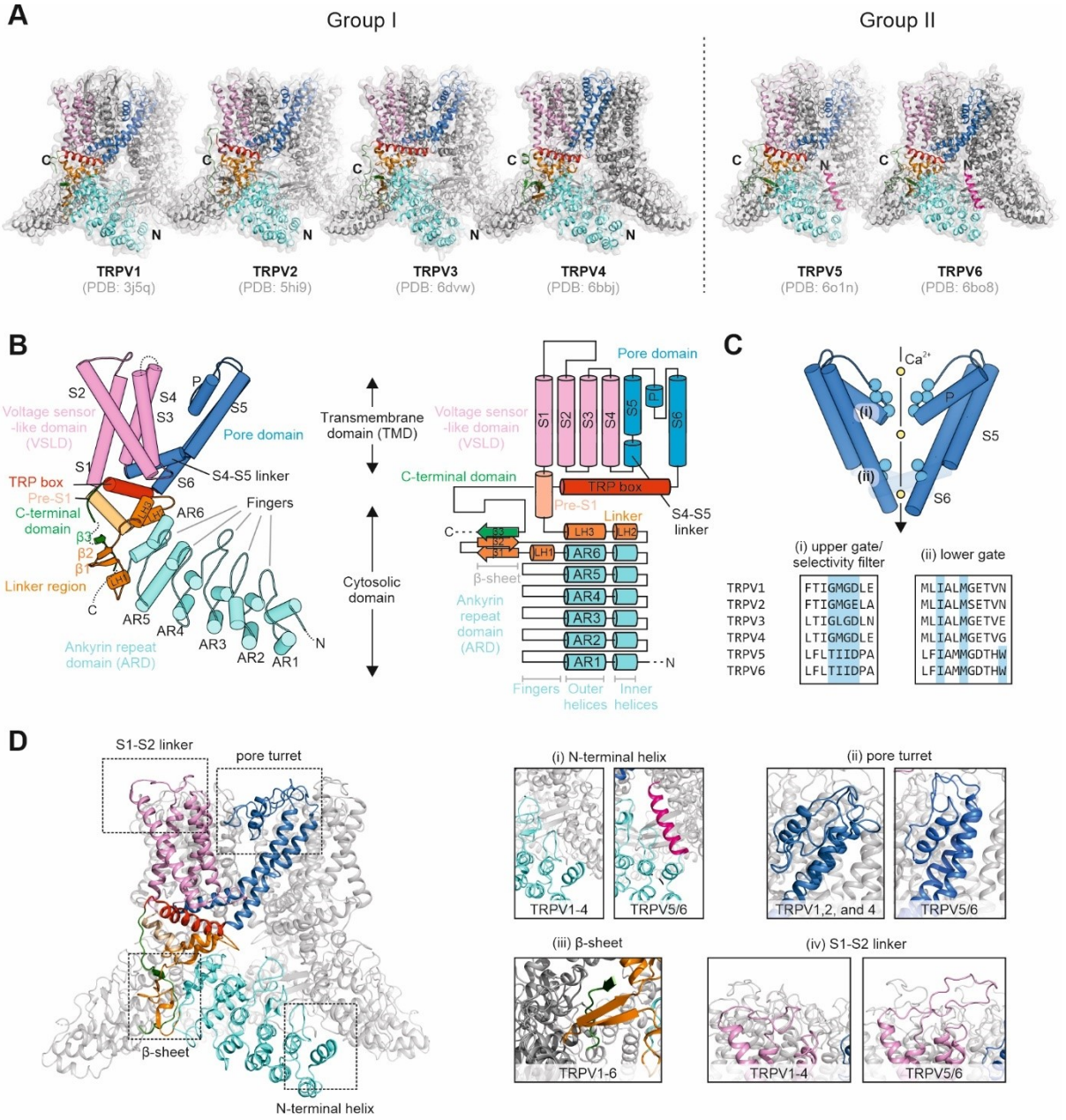


Figure 4: Structure of TRP vanilloid channels. (**A**) Cryo-EM structures of group I and group II TRP vanilloid channels. (**B**) Cartoon and topology representation of a typical TRP vanilloid subunit with structural elements highlighted in different colours: ARD with individual Ankyrin repeats (AR) and loops (fingers)between them (cyan), linker region with linker helices (LH) and β-strands β1 and β2 (orange), pre-S1 helix (light orange), VSLD consisting of S1-S4 (pink), pore domain (S5-S6) with S4-S5 linker (blue), TRP box helix (red), C-terminal domain with β-strand β3 (green). (**C**) The architecture of a TRP vanilloid pore. The upper gate, which forms the selectivity filter, has a conserved G(M/L)G(D/E) motif in group I TRPV channels. In group II TRPV channels, this motif consists of TIID. The lower gate contains conserved isoleucine and methionine residues. TRPV5/6 channels have an additional tryptophan residue at the end of S6. This tryptophan is essential for a Ca^{2+}/CaM-dependent negative feed-back mechanism (see Figure 5 B). (**D**) Structural features of TRPV group I and II channels. The TRPV2 structure (PDB: 6bo4) is shown as a representative of group I TRPV channels, whereas sections of the TRPV6 structure (PDB. 6bo8) are shown to highlight structural features in group II TRPV channels. (i) In contrast to TRPV1-4, TRPV5/6 have an α-helix in the N-terminus directly preceding the ARD. This N-terminal helix folds onto the ARD where it interacts with the ARD's fingers. (ii) In TRPV1, 2, and 4, an elongated loop between the pore loop and S5, the so-called the pore turret (*157*), folds onto the pore domain. The pore turret was implicated in modulating the heat sensitivity of TRPV channels (*147, 157–159*). The thermoresponsive channel TRPV3 and the non-thermoresponsive TRPV5/6 channels do not have a pore turret. (iii) In all TRPV channels, a β-sheet formed by the linker region and the CTD connects the cytosolic domains of neighbouring subunits. (iv) TRPV5/6 have an elongated S1-S2 linker in the VSLD, which interacts with the pore domain of an adjacent subunit.

Regulation of TRPV channels by lipids

High-resolution structures of TRPV channels in apo or ligand-bound states have not only revealed details of ion conduction through the pore. They have also expanded our molecular understanding of how exogenous and endogenous molecules modulate channel activity (Figure 5 A, B, C, and D). Structures of TRPV1 in complex with capsaicin, capsazepine, or RTX revealed a binding pocket for (ant)agonists between the S4-S5 linker and S3 helix of the VSLD, termed the vanilloid pocket (Figure 5 A, box (i)) (*40, 49*). A similar binding pocket exists in other TRPV channels, even though they are not sensitive to TRPV1-specific ligands. The ligand specificity depends on the exact residues in the binding pocket, which are required to hold the ligands in their proper orientations. Mutating a few residues in the vanilloid pockets of TRPV2 and TRPV3 confers sensitivity for vanilloid compounds (*164–167*). In the absence of exogenous ligands, the vanilloid pocket seems to be occupied by a resident lipid (Figure 5 A, box (ii)). In the cryo-EM structure of TRPV1 reconstituted in lipid nanodiscs, the resident lipid could be assigned confidently as phosphatidylinositol (PI) (*49*). It presumably stabilizes a closed channel conformation under basal conditions. When agonists such as vanilloid compounds activate thermoTRPV channels, they may replace the resident lipid in the vanilloid pocket and induce a conformational change to an open conformation (Figure 5 B) (*49, 135, 136, 140*). A mutation in TRPV1 which abrogated binding of the resident lipid rendered the channel insensitive to stimulation by capsaicin (*168*).In contrast to the thermoTRPVs, the resident lipid was proposed to stabilize the open conformation of TRPV5 and TRPV6 (*143, 148*). Mutation of the pocket leading to the displacement of the lipid rendered TRPV6 constitutively closed (Figure 5 D) (*143*). Thus, the vanilloid pocket seems to act as a general binding site for regulatory lipids in TRPV channels. The exact regulatory mechanism of lipid binding, however, differs between the thermoTRPVs and TRPV5/6.

A second lipid binding pocket was observed in several TRPV structures between the TRP box and the VSLD (*49, 135, 140, 143, 146, 149*). Similar to the vanilloid pocket, the VSLD/TRP box pocket seems to act as a site for competitive binding between a resident lipid and a ligand, e.g. 2-aminoethoxydiphenyl borate (2-APB) in TRPV3 and TRPV6 (*139, 140*). As observed for ligand binding to the vanilloid pocket, the directionality of channel regulation upon ligand binding in the second lipid binding pocket differs between group I and group II TRP vanilloid channels (Figure 5 C and D). In TRPV3 (and in TRPV1 and 2), 2-APB acts as an agonist, whereas it inhibits TRPV6 channel activity (*169*). The binding of an exogenous ligand in the second binding site may also displace the resident lipid in the vanilloid pocket (Figure 5 D). Lipid displacement in the TRPV6 vanilloid pocket was proposed to trigger the formation of a hydrophobic cluster which induces a π-to-α-helix transition in S6 to close the lower gate (*139, 143*).

High-resolution cryo-EM structures supported by functional and mutagenesis studies highlight the regulatory roles of lipids for TRPV function. However, all cryo-EM studies were carried out on purified TRPV channels reconstituted in lipid environments which may not necessarily reflect their physiological lipid environment (*121, 142, 145, 146*). Thus, the exact identity of the resident lipids in the vanilloid pocket or the VSLD/TRP box pocket remains to be conclusively determined. Nevertheless, the importance of lipids for the regulation of TRPV channels has been recognized long before the first lipid molecules were identified in TRPV cryo-EM structures. Functional studies identified various lipids and lipid metabolites as TRPV agonists, antagonists, sensitizers, and desensitizers (*170*). A particularly important lipid for TRP channel regulation, including TRP vanilloids, is phosphatidylinositol 4,5-bisphosphate ($PI(4,5)P_2$, hereafter solely referred to as PIP_2 unless further specified). PIP_2 is a lipid second messenger which modulates the activity of a diversity of ion channels (*171, 172*). The PIP_2 levels in the plasma membrane are regulated by several receptor signaling pathways via activation of phospholipase C (PLC) and subsequent hydrolysis of PIP_2 into inositol 4,5-bisphosphate (IP_3) and diacylglycerol (DAG). Thus, PIP_2 can act as a link between receptor signalling and ion channel activity. Regulatory effects of PIP_2 have been described for every member of the TRPV family. Considering the complex mechanisms which regulate the PIP_2 levels in a cell, it is not surprising that the regulatory effects of PIP_2 on TRPV activity often vary with the experimental setups used in functional studies, such as the used cell-types, the patch-clamp technique, or the way how PIP_2 levels are modulated (PIP_2 hydrolysis or inhibition of PIP_2 synthesis) (*173*). Accordingly, sometimes the roles of PIP_2 for TRPV gating appear contradictory. Most studies describe PIP_2 as an activator of TRPV1 (*174–181*). However, several publications also indicate inhibitory effects of PIP_2

on TRPV1 activity (*112, 114, 182*). Similar discrepancies have been observed for the effect of PIP_2 on TRPV4 activity. Two studies described PIP_2 as an inhibitor of TRPV4 (*150, 183*), whereas another study described it as a sensitizer of TRPV4 for heat and osmotic stimuli (*128*). So far, only few studies have investigated the effect of PIP_2 on TRPV2 and TRPV3. They suggest that TRPV2 is activated by PIP_2 (*184*), whereas TRPV3 is inhibited by PIP_2 (*120*). TRPV5 and TRPV6 were both found to be activated by PIP_2 (*185–187*).

The PIP_2 binding site in TRPV1 was initially identified to encompass basic and hydrophobic residues in the C-terminal part of the TRP box (*188*). With the advent of the first high-resolution structures of TRPV1, more detailed computational, functional, and mutagenesis studies narrowed down the PIP_2 binding site to a binding pocket between the TRP box and the S4-S5 linker in the TMD (*179*). In this binding pocket, several lysine and arginine side chains interact with the negatively charged PIP_2 headgroup. Molecular dynamics (MD) studies showed that PIP_2 binding to the TRP box/S4-S5 linker induces a conformational change in the TRPV1 TMD to open the channel gate (*179*). Intriguingly, PIP_2 binding residues in the S4-S5 linker and the TRP box are conserved between TRPV1-4 and partly conserved in TRPV5/6. In a recent cryo-EM study of TRPV5, EM densities at the TRP box/S4-S5 linker region could be modelled confidently with a PIP_2 molecule (Figure 5 A) (*146*). The position and orientation of PIP_2 in the TRPV5 cryo-EM structure agreed very well with the predicted PIP_2 binding mode in MD studies of TRPV1 (*179, 189*). The TRP box/S4-S5 linker region may thus represent a general PIP_2 binding site in TRPV channels. In TRPV1 MD studies, an arginine side chain in the S4-S5 linker makes contacts to a phosphate group of PIP_2. TRPV5/6 have a glycine instead of the arginine side chain in the S4-S5 linker. Mutation of the glycine in TRPV5/6 to arginine markedly increased the PIP_2 affinity and ablated TRPV5/6 inhibition upon PIP_2 depletion via PLC activation (*189*). A low-affinity PIP_2 binding site in TRPV5/6 is presumed to provide an additional layer to the Ca^{2+}-dependent negative feedback mechanism. At low intracellular Ca^{2+} concentrations, PIP_2 binding is supposed to help maintaining TRPV5/6 constitutively active. Increasing intracellular Ca^{2+} levels may activate PLC in a Ca^{2+}/CaM-dependent manner. The subsequent hydrolysis of PIP_2, causes a collapse of the open pore conformation (Figure 5 D) (*189*).

Importantly, the TRP box/S4-S5 linker may not be the only binding site for PIP_2 in TRPV channels. As outlined in more detail in section VI1.3, two different PIP_2 binding sites with opposing regulatory effects have been described in TRPV4 (*128, 150*). With one proposed binding site in the region preceding the ARD (*128*) and the other within the ARD (*150*), they are both somewhat distant from the TRP box/S4-S5 linker pocket. Thus, even though all TRPV channels share a conserved TRP box/S4-S5 linker amino acid composition as a putative PIP_2 binding site, additional PIP_2 binding sites may substantially contribute to the regulation of a TRPV channel. Moreover, multiple PIP_2 binding sites with opposite effects on channel gating may communicate allosterically and thereby fine-tune channel activity.

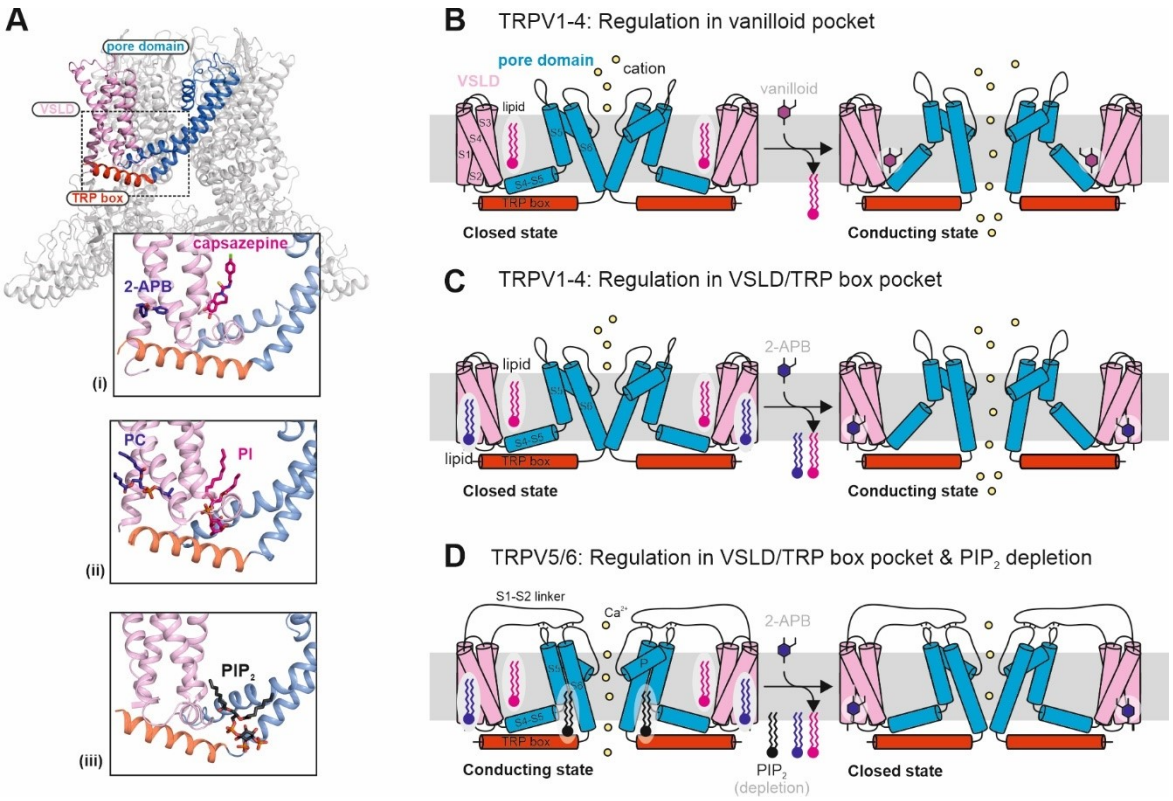


Figure 5: Proposed lipid-dependent gating mechanisms of group I and group II TRPV channels. (**A**) Ligand binding sites observed in high-resolution structures of TRPV channels. (i) Identified binding sites of the exogenous ligands capsazepine (PDB: 5is0) (*49*) and 2-aminoethoxydiphenyl borate (2-APB) (PDB: 6dvy) (*140*). The TRPV1 agonist capsazepine was found to occupy a pocket between the VSLD and on top of the S4-S5 linker, termed the vanilloid pocket. The same binding pocket is occupied by other vanilloid compounds such as capsaicin (*190*). The TRPV1-3 agonist and TRPV5/6 antagonist 2-APB binds to a pocket between the VSLD and the TRP box. (ii) In the absence of an agonist, the vanilloid pocket and the VSLD/TRP box pocket are occupied by lipids. In the cryo-EM studies, the resident lipid in the TRPV1 vanilloid pocket was identified as phosphatidylinositol (PI). In contrast, the lipid identified in the VSLD/TRP box pocket was modelled as phosphatidylcholine (PC) (both in PDB: 5irz) (*49*). (iii) A $PI(4,5)P_2$ binding pocket between the TRP box and the S4-S5 linker was identified in the cryo-EM structure of TRPV5 (PDB: 6dmu) (*146*). Functional studies paired with mutagenesis and computational modelling confirmed that the same binding pocket exists in TRPV1 and TRPV6 (*189*). The residues in this region are highly conserved, thus indicating that it may represent a general PIP_2 binding site in TRPV channels. (**B**) Vanilloid-induced activation mechanism proposed for TRPV1-4 channels. In the absence of a stimulus, lipids bind to the vanilloid pocket of TRPV1-4 channels and maintain the TRPV channels in a closed state. Exogenous compounds like capsazepine can compete with the lipid in the vanilloid binding pocket and trigger conformational changes of the S4-S5-linker and the pore domain to open the lower gate (conducting state). (**C**) A second regulatory binding site for exogenous ligands and lipids resides between the VSLD and the TRP box. As the vanilloid pocket, the VSLD/TRP box pocket is occupied by a lipid in the absence of a ligand. The lipid helps to maintain a closed state conformation. Binding of the agonist 2-APB displaces the resident lipid which presumably also displaces the resident lipid in the vanilloid pocket and thus triggers a conformational change of the pore helix S6 to a conducting state. (**D**) Unlike in the thermoTRPVs, the resident lipids in the vanilloid and the VSLD/TRP box pocket of TRPV5/6 stabilize a conducting state conformation. Displacement of the resident lipid in the vanilloid pocket was shown via mutagenesis to render TRPV6 constitutively closed. As opposed to agonizing thermoTRPVs, 2-APB antagonizes TRPV5/6. 2-APB binding to the VSLD/TRP box pocket of TRPV6 was proposed to displace the resident lipids in both the VSLD/TRP box and the vanilloid pocket and stabilize a closed state. Notably, a $PI(4,5)P_2$ molecule was identified in a TRPV5 cryo-EM study to bind between the S4-S5 linker and the TRP box (PDB: 6dmu) (*146*). PIP_2 binding to TRPV5/6 was proposed to maintain the channels constitutively active. Upon increased intracellular Ca^{2+} levels, Ca^{2+}/CaM induced activation of PLC causes PIP_2 depletion and leads to channel inhibition (*189*).

Regulation of TRPV channels by cytosolic interaction partners

TRPV channels participate in complex signalling pathways where they interact with different cytosolic proteins or small molecule second messengers. The large cytosolic domains, mainly composed of the ARD, are considered to be vital elements in mediating the binding of cytosolic interaction partners to TRPV channels. Binding sites for regulatory TRPV binding partners have also been identified in the N-terminal regions preceding the ARD (*76, 77*) and in the C-terminal regions succeeding the TRP box (*168, 191*). It is, however, still largely

unclear how interactions in the cytosolic domains elicit conformational changes which propagate to the pore and there regulate the gates of TRPV channels. High-resolution structures of TRPV channels with ligands bound in their cytosolic domains are still scarce. Until recently, they were largely restricted to X-ray crystal structures of isolated ARDs in complex with small molecule ligands, e.g. ATP (*18, 21*) or IP_3 (*150*), or of TRPV peptides in complex with a protein interaction partner, e.g. a TRPV1 C-terminal peptide with calmodulin (CaM) (*78*). These structures alone cannot explain how binding of an interaction partner in the cytosolic domains induces conformational changes in the pore regions.

The recently determined cryo-EM structures of TRPV5 and TRPV6 tetramers each in complex with a single calmodulin protein present the first examples of how TRPV channels interact with cytosolic interaction partners (*131, 146*). In combination with functional data (*133, 185, 191–195*) and NMR structural data of CaM binding to the C-terminal region of TRPV5 (*196*), which was not resolved in cryo-EM structures, they have enabled to deduce a conserved CaM-dependent TRPV5/6 inhibition mechanism (Figure 6) (*131, 149*). TRPV5/6 are constitutively open at basal calcium levels, and the CaM C-lobe is effectively always bound to Ca^{2+}. The CaM C-lobe presumably always binds to the distal end of the TRPV5/6 C-terminus at basal calcium levels. At increased intracellular calcium concentrations, the CaM N-lobe binds Ca^{2+} and interacts with the more proximal end of the TRPV5/6 C-terminus. This pulls the C-lobe towards the pore regions, where a lysine residue in the CaM C-lobe inserts between the TRPV5/6 tryptophan side chains at the lower end of S6 and thereby blocks the channel. The Ca^{2+}/CaM-depentend negative feedback loop of TRPV5/6 may be paralleled by the hydrolysis of PIP_2 upon PLC activation (Figure 6, see also Figure 5 D).

Interestingly, CaM was also described to downregulate TRPV1, 3, and 4 after activation (*151*). The identified or proposed binding sites in the C-terminus of TRPV1, 3, and 4, however, do not align well with those described in TRPV5/6 (*168, 191*). Moreover, the ARD was described to contribute to the CaM-TRPV interaction of thermoTRPVs (*151*). It is therefore possible that the CaM-dependent inactivation mechanism differs between the thermoTRPVs and TRPV5/6. The exact details of channel regulation via CaM in thermoTRPVs has remained largely unknown. This is partly because the putative CaM binding sites are not resolved in the cryo-EM structures of TRPV1-4, presumably because they are disordered, i.e. they are very flexible and do not adopt a defined conformation.

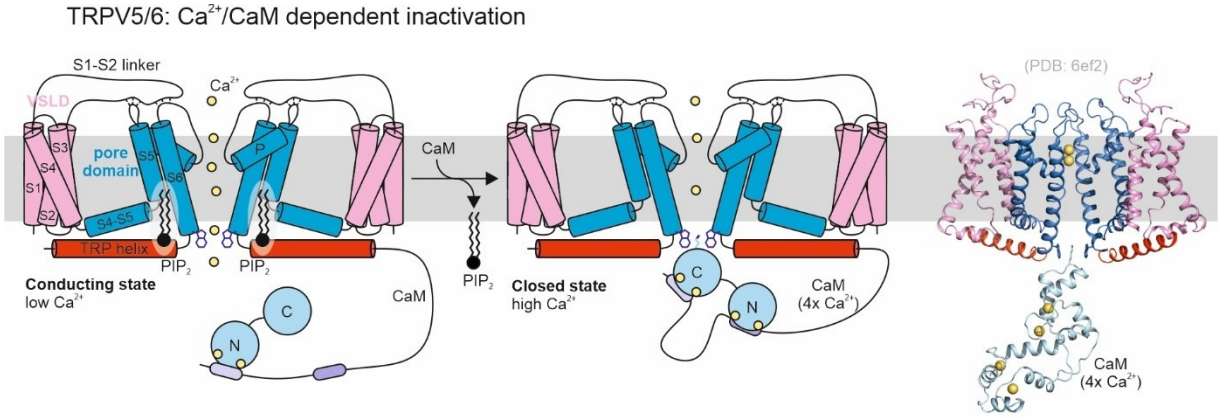


Figure 6: Calmodulin-dependent inhibition mechanism proposed for TRPV5/6. At low intracellular Ca^{2+} concentrations, TRPV5/6 are constitutively open (conducting). Calmodulin in complex with two Ca^{2+} ions in the C-lobe binds to a C-terminal region of TRPV5/6. With increasing intracellular Ca^{2+} levels, the CaM N-lobe binds Ca^{2+} and subsequently to a second binding site in the TRPV5/6 C-terminus. This triggers a conformational change in CaM upon which it inserts a lysine side chain into the TRPV5/6 pore, where it is coordinated by the S6 tryptophan side chains of the lower gate (see Figure 3 C). Channel inhibition by Ca^{2+}/CaM-binding to the lower gate may be paralleled by the Ca^{2+}/CaM-induced activation of PLC and the subsequent depletion of PIP_2 levels. PIP_2 binding to a pocket between the S4-S5 linker and the TRP boxes was proposed to favor a ion conducting pore conformation in TRPV5/6 under basal conditions. PIP_2 hydrolysis may thus lead to a collapse of the open pore conformation and block ion conduction.

The distal TRPV N- and C-termini engage in channel regulation

Sequence-based predictions indicate that the N- and C-termini of TRPV channels are disordered throughout almost the entire subfamily (Figure 7). The N- and C-termini contain binding sites for regulatory interaction partners such as lipids and proteins (*75–78, 191, 195*) and they are the targets of post-translational modifications (PTMs) (*197–202*) or the splicing machinery (*203*). Mutations in the disordered N- and C-termini of TRPV4 were shown to disturb channel trafficking or function and thus are responsible for the development of several pathological conditions (*204*). Given the lack of structural data, it is mostly unclear how these regions are structurally coupled to the rest of the channel and how they can propagate conformational changes induced upon ligand binding or PTMs to the pore region.

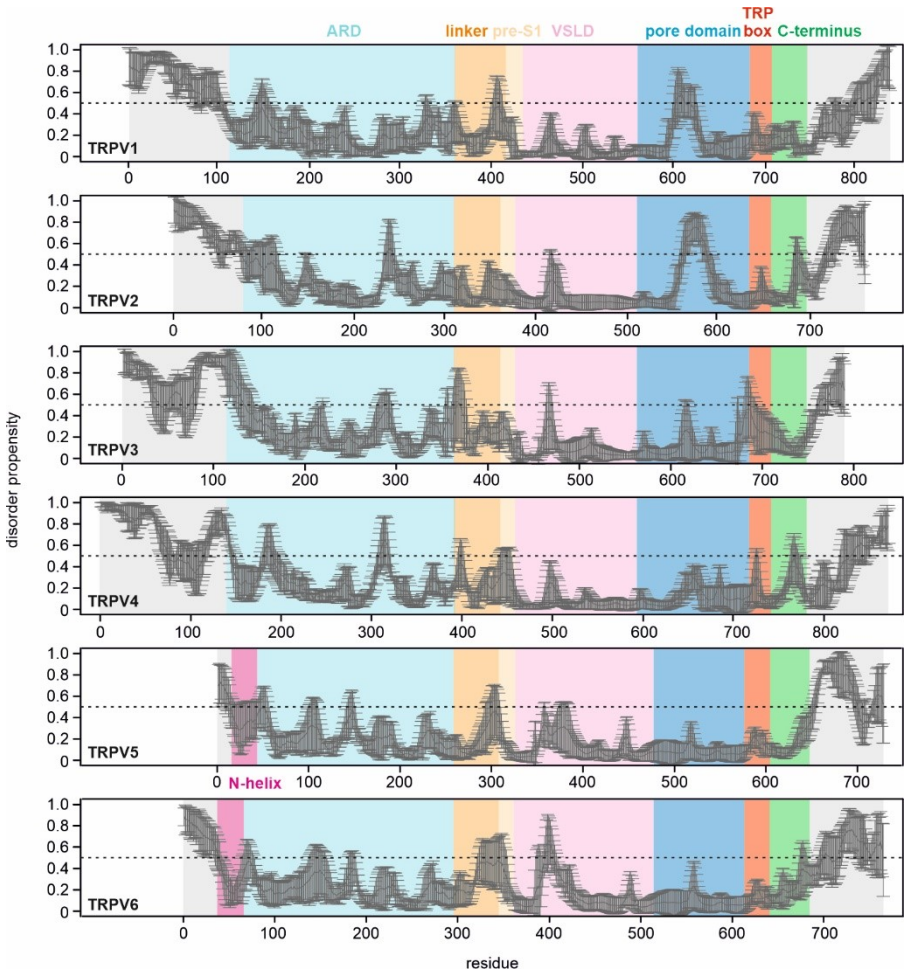


Figure 7: Disorder analysis in the TRP vanilloid ion channel subfamily. Sequence-based disorder predictions of *H. sapiens* TRPV1-6 channels via IUPRED (*84*), PONDR (*85–90*), PONDR-FIT (*91*), and DISOPRED (*92*). Plotted is the average disorder propensity with the standard error between all predictions. Colour code adheres to the domain colouring scheme shown in Figure 4. The disordered N- and C-terminal regions are highlighted in grey.

Even though disordered regions commonly evade structural analysis with single-particle cryo-EM, they can become accessible to cryo-EM by being trapped in a specific state, as observed in cryo-EM studies on the activation mechanism of TRPV3 (*134, 141*). Two groups independently observed a structural switch involving disordered regions of the C-terminal domain (CTD) and the N-terminal domain (NTD, the region preceding the ARD) between TRPV3 in a closed and a sensitized state (Figure 8) (*134, 141*). In the closed state of TRPV3 ('switch off'), the region in the CTD downstream of the β-sheet forms a disordered coil structure. A salt-bridge holds the disordered coil in a cavity between the β-sheet and the fingers of an adjacent ARD (hereafter termed

the ARD/β-sheet cavity). In the activated state, the CTD forms an α-helix and moves underneath the ARD. In the activated state, the CTD in the ARD/β-sheet cavity is replaced by a region of the disordered NTD. The CTD-NTD switch is coupled to the rest of the channel by a displacement of ankyrin repeat 5 in the ARDs and the subsequent movement of the pore helices via the TRP box/S4-S5 linker coupling (*134*). Importantly, the resolutions obtained in regions modelled as the CTD or the NTD are very low compared to other regions in the channels, thus indicating that they are very flexible. The EM density accounting for the NTD in the 'switch on' state could not be assigned to specific residues and was modelled as a polyalanine chain. Moreover, its assignment as NTD residues was solely based on the observation that the EM density had no connectivity to the CTD helix. Thus, it was concluded that the density has to belong to the NTD of an adjacent subunit. The assignment of the EM density in the ARD/β-sheet cavity in the 'switch on' state as a region of the NTD is therefore still ambiguous (*134*).

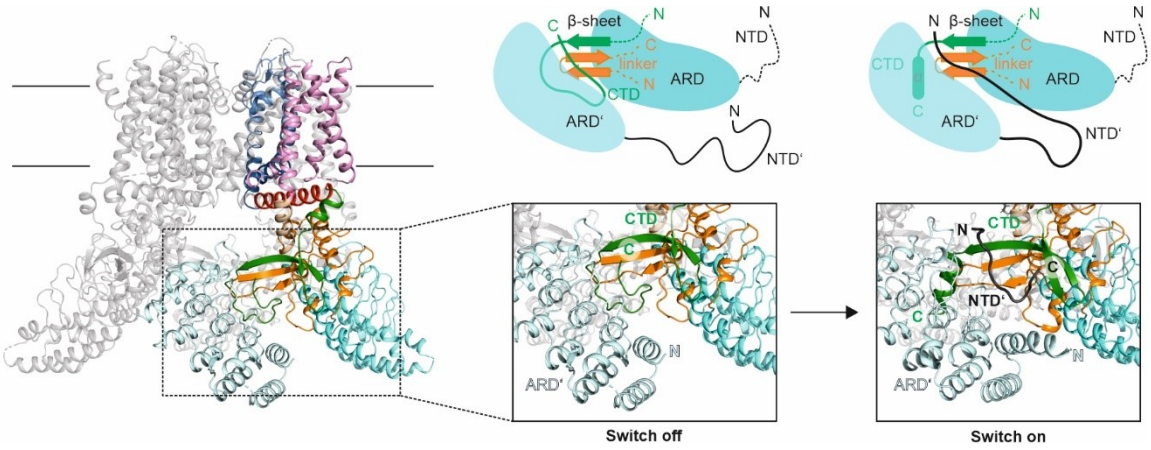


Figure 8: The CTD-NTD switch mechanism proposed for TRPV3. In the closed state, the distal CTD is wrapped around the β-sheet at the intersubunit interface via a salt bridge interaction with the ARD of an adjacent subunit (switch off). The binding site between the ARD and the β-sheet of two neighbouring subunits is termed the ARD/β-sheet cavity. Upon displacement from the ARD/β-sheet cavity (switch on), the CTD undergoes a coil-to-helix transition, which leads to a conformational change in the ARD. This conformational change is coupled to conformational changes in the pore region and leads to the opening of the channel. In 'switch off' state, the CTD in the ARD/β-sheet cavity is replaced by the NTD, i.e. the region preceding the ARD (*134*). Left: Cryo-EM structure of TRPV3 with 'switch off' (PDB: 6mho). Right: TRPV3 cryo-EM structure with 'switch on' (PDB: 6ot5).

The CTD-NTD switch mechanism proposed for TRPV3 provides the only structural insights into how conformational changes in the disordered N- and C-termini of a TRPV channel are coupled to changes in the rest of the channel (*134*). Importantly, all other TRPV channels have disordered C-termini longer than that of TRPV3 (Figure 7). Thus, the C-terminal domain can theoretically bind to the ARD/β-sheet cavity in all TRPV channels. However, the disordered nature of the ARD-preceding region, which appears necessary for the CTD-NTD switch, is only conserved in the thermoTRPVs but not in TRPV5/6 (Figure 7). This structural feature is exciting regarding the different gating behaviours observed between thermoTRPVs and TRPV5/6. In TRPV5/6, the ARDs are preceded by an α-helix with only a short disordered fragment (Figure 4 and Figure 7) (*143, 148*). In all structures of TRPV5/6, the N-terminal α-helix occupies virtually the same space in the ARD/β-sheet cavity as the disordered NTD in the cryo-EM structure of the TRPV3 'switch on' state (*41, 131, 134, 139, 143, 145, 146, 148, 149*). The helix appears as a relatively rigid structure that probably always stays in the ARD/β-sheet cavity. The N-terminal helix folded on the ARD might mimic a 'switch on' state and thus maintain TRPV5/6 channels constitutively active. One might speculate that the CTD-NTD switch mechanism proposed for TRPV3 requires a certain length and flexibility of the region preceding the ARD to reversibly bind in the ARD/β-sheet cavity (Figure 8). The CTD-NTD switch might also explain why ligand binding in the disordered N- and C-termini of thermoTRPVs is capable of modulating channel activity, i.e. by interfering with a CTD-NTD switch.

Overall, the disordered N- and C-termini in TRPV channels are emerging as critical regulators of channel function. Thus, there is an urgent need to elucidate the mechanistic details of how structural changes in the disordered regions are coupled to the rest of the channels. Considering the rare examples in which the disordered regions of TRPV channels were partially resolved in cryo-EM structures, it is unlikely that cryo-EM alone can accomplish this (*74*). Instead, integrated structural biology approaches based on complementary techniques, e.g. cryo-EM, NMR, SAXS or smFRET, will likely be at the heart of future approaches targeted at the (un)structural biology of TRP channels (*205*). One member of the TRPV subfamily for which multiple ligands, PTMs and disease-causing mutations have been identified in the disordered N- and C-termini is TRPV4 (*206*). However, the mechanistic features of TRPV4 regulation in the disordered regions have remained enigmatic.

1.3 TRP vanilloid 4 – a polymodally regulated cation channel

TRPV4 is crucial to human physiology

TRP vanilloid 4 was initially identified as an osmosensitive calcium channel mediating responses to changing osmotic conditions in mammals. Over time, it has been recognized that TRPV4 function is much broader than mere osmoregulation (Figure 9). TRPV4 stands out from other TRP channels due to its exceptionally pronounced polymodality. It is activated by temperature (*207–210*), hypotonic stimuli (*125*), mechanical force (*211*), and several noxious stimuli (*212, 213*). Moreover, a broad diversity of exogenous and endogenous interaction partners has been reported to modulate how TRPV4 responds to the environment and gates ions (*126, 214–218*). Its extraordinary polymodality and broad expression in numerous tissues make TRPV4 a critical signal transducer in human physiology (*98, 99, 125, 219, 220*). It plays crucial roles in osteoclast and chondrocyte development (*221–223*), neuritogenesis (*19, 224*), cell volume regulation (*225, 226*), viral infectivity (*227*), as well as in temperature (*207–210*) and pain sensation (*212, 213*). Therefore, it comes as no surprise that defective TRPV4 signalling is the cause of a variety of pathological conditions, so-called channelopathies (*204*).

Cell-based assays have shown that TRPV4 is activated by osmotic stress, and TRPV4$^{-/-}$ mice show impaired osmoregulation in the bladder (*124, 228, 229*). Osmoregulation describes the active control of the osmotic pressure of intra- and extracellular fluids to maintain homeostasis. Changes in osmolarity can lead to cell swelling or shrinkage. TRPV4 channels form complexes with aquaporins (AQ), a family of water channels, to control regulatory volume decrease under changing osmotic conditions (*230, 231*). A consequence of cell swelling under hypotonic conditions is an increase in membrane tension, thus leading to different forms of mechanical stress. TRPV4 can be stimulated by mechanical stress, e.g. flow-induced shear stress, which plays a role in controlling Ca^{2+} influx in endothelial cells during vasodilation (*232*). The response of TRPV4 to shear stress seems to be coupled to intracellular signalling and causes TRPV4 to traffic from cytoplasmic vesicles to the PM (*233*). Another form of mechanical stress is the deformation of the cell membrane through direct pressure on the cell membranes, thus deforming the cytoskeleton. TRPV4 was proposed to sense such deformations by being directly coupled to the cytoskeleton (*234*). Besides its role as a mechano- and osmotransducer, TRPV4 was implicated in sensing non-noxious temperatures in the range between 27-34°C.

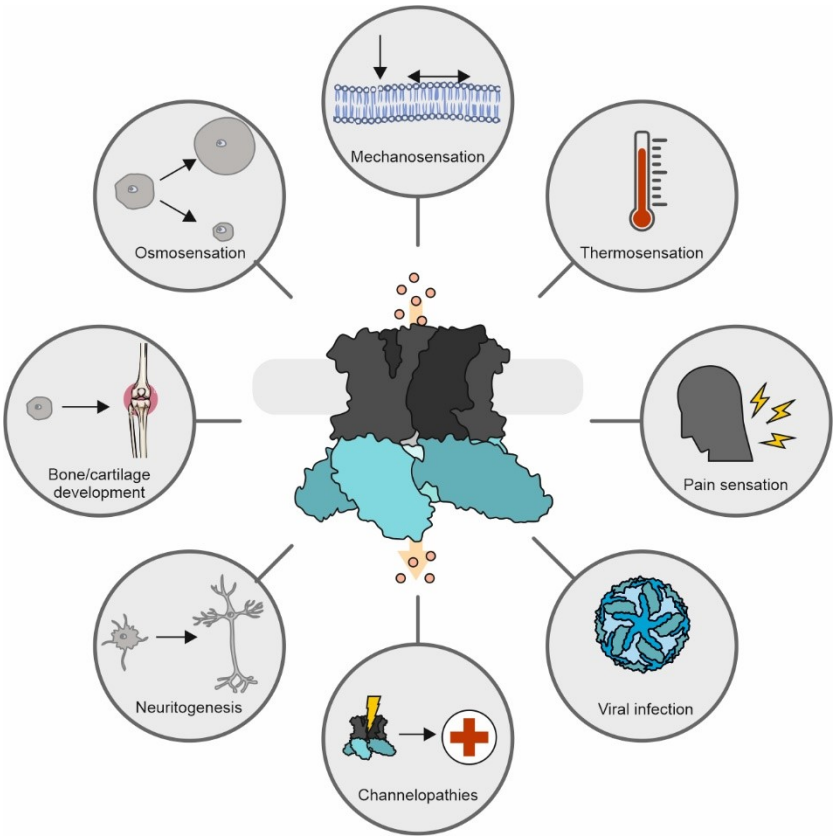


Figure 9: Physiological functions of TRPV4. TRPV4 has been implicated to act as a sensor of osmotic (cell swelling/shrinking), thermal, mechanic, and painful stimuli. Its polymodal gating behaviour and broad tissue distribution make TRPV4 a critical signal transducer in diverse physiological processes. It is involved in osteoclast and chondrocyte development (bone and cartilage), neuronal development (neuritogenesis), and has been linked to viral infectivity. Malfunctioning (TRP) ion channels are the cause of a variety of diseases, so-called channelopathies.

One of the most prominent roles of TRPV4 is the transduction of painful stimuli (*235*). The mechanism by which TRPV4 senses painful stimuli are multifaceted. Mostly, they are coupled to inflammatory signalling pathways which generate compounds that directly activate or sensitize TRPV4. Alternatively, they activate kinases, which phosphorylate TRPV4 and thereby sensitize it to stimuli which are otherwise not painful. As such, TRPV4 seems to play a role in thermal hyperalgesia which is the sensation of moderately hot temperatures (35-50°C) under pathological conditions (*236*). TRPV4 plays a role in inflammation-induced mechanical hyperalgesia and in neuropathy-dependent mechanical hyperalgesia (*213, 237*). Therefore, it attracts increasing interest as a novel analgesic (*235*). The potential of TRPV4 as a drug target is not only relevant to pain therapy. TRPV4 has also been in implicated in several other pathological conditions, such as cancer (*238*) or brain injuries and diseases (*239*). More recently, TRPV4-mediated Ca^{2+}-influx has been identified as a requirement for viral infectivity in a CaM and CaM-kinase-dependent pathway involving the RNA helicase DDX3X (*227*). Inhibiting TRPV4 blocked infection of cells by Zika, Dengue and hepatitis C virus. Moreover, a vast number of mutations in the *trpv4* gene have been linked to the development of severe channelopathies (*204*). Most diseases are gain-of-function diseases, and the severity of the diseases seems to correlate with the extent to which TRPV4 gating is enhanced in mutants compared to wildtype TRPV4 channels. The observation that TRPV4 point mutations are the cause of severe diseases is astounding, considering that $TRPV4^{-/-}$ mice are viable and show relatively mild phenotypes (*228, 240*). This suggests that other channels can compensate for the loss of TRPV4, but uncontrolled TRPV4-mediated Ca^{2+} influx is detrimental for downstream signalling pathways.

Diverse ligands and pathways control TRPV4 function

The severe consequences of TRPV4 dysregulation imply that TRPV4 activity needs to be tightly regulated in the cell. Various interaction partners found to activate TRPV4 or sensitize it for distinct stimuli directly include lipids (*128*) and lipid metabolites (*126*), proteins† (*76, 77, 221, 242*), nucleotides (*21*), pH (*240*), ions (*243*), and small organic compounds (*127*) (Figure 10). The definitions of activation and sensitization are often blurred in the polymodal gating behaviour of TRPV4. Activation describes the opening of TRPV4 by an agonist, whereas sensitization is the augmentation of the response elicited by a specific stimulus. Whether an interaction partner acts as an agonist or a sensitizer may sometimes be unclear because the effects of several other stimuli on channel response need to be taken into account to distinguish between both mechanisms clearly.

Several lipid-like molecules and exogenous compounds were identified as TRPV4 agonists. The phorbol ester 4α-phorbol 12,13-didecanoate (4α-PDD) was described to elicit a similar TRPV4 response as the application of heat (*127, 215*). Other phorbol esters such as the protein kinase C activator phorbol 12-myristate 13-acetate (PMA) were also found induce TRPV4-mediated Ca^{2+} currents in mouse, rat, or human cells, though the currents were lower than observed for 4α-PDD (*127*). The lipid metabolite 5',6'- epoxyeicosatrienoic acid (5',6'-EET) acts as an endogenous TRPV4 agonist (*126*). 5',6'-EET is a downstream product of cell swelling which activates phospholipase A_2 (PLA_2) to produce arachidonic acid (AA). AA is converted via cytochrome (cyt) P450 epoxygenase into 5',6'-EET (*214*). AA is also generated from the endocannabinoid anandamide, also known as arachidonoylethanolamide (AEA), via the enzyme fatty acid amidohydrolase (FAAH) and subsequently converted to 5',6'-EET via cyt P450. Consequently, treatment of cells expressing TRPV4 with AA, AEA, and 5',6'-EET evokes almost identical responses (*214*). A broad palette of plant-derived compounds was also observed to agonize TRPV4, such as bisandrographolide A (BAA), apigenin (*244*), eugenol (*245*), and several plant cannabinoids like cannabidivarin and tetrahydrocannabivarin (*246*). Of these, only BAA was shown to be selective for TRPV4 (*247*). It is a highly potent TRPV4 agonist similar to capsaicin for TRPV1. Currently, no endogenous TRPV4 antagonists are known. Antagonists are extremely important to study the physiological roles of TRPV4 and as potential drugs for novel pain therapies (*248*). The currently most frequently used TRPV4 antagonists are synthetic compounds. They are, however, often used at concentrations which might be non-selective.

Important groups of intracellular TRPV4 regulators are proteins such as the calcium-binding protein calmodulin (CaM). Almost all TRPV channels have been described to be regulated by the binding of CaM in response to increased intracellular Ca^{2+} concentrations (*242, 249, 250*). In TRPV4, the regulation by CaM is tightly connected to the sensitization of TRPV4 by ATP. The nucleotide was observed to potentiate basal TRPV4 currents in HEK293 cells and TRPV4 currents induced by 4α-PDD (*151*). This effect was competed by CaM binding to TRPV4. Crosstalk between both regulators for TRPV4 binding is supposed to present a way to fine-tune TRPV4 sensitivity to fluctuations in calcium concentrations. Proteins not only regulate TRPV4-mediated Ca^{2+} currents through direct interaction but also by altering its subcellular localization. Binding of the protein Pacsin3 (protein kinase C and casein kinase substrate in neurons 3) was observed to reduce basal TRPV4 currents (*77*). Simultaneously, the number of TRPV4 channels in the plasma membrane was increased (*76*). The HECT (homologous to E6-AP C-terminus) family ubiquitin-ligase ITCH (also known as AIP4) was found to ubiquitinate TRPV4 in its N-terminus (*251*). Instead of degradation, TRPV4 is endocytosed upon ubiquitination, thus decreasing the amount of TRPV4 in the plasma membrane. The ER-localized membrane protein OS-9 directly binds TRPV4 monomers to prevent underdeveloped TRPV4 species from trafficking to the plasma membrane (*252*). Trafficking of TRPV4 to the plasma membrane upon Ca^{2+} store depletion was demonstrated to require interaction between TRPV4 and the auxiliary protein STIM1 (stromal interaction module 1) (*253*). STIM1 is a calcium-sensing transmembrane protein in the ER which controls intracellular Ca^{2+} influx after intracellular Ca^{2+} stores become depleted. It was proposed that the TRPV4/STIM1 complex guides TRPV4 from the ER to the plasma membrane.

† A total of 38 proteins have been identified as TRPV4 interaction partners according to the TRIP data base (*241*).

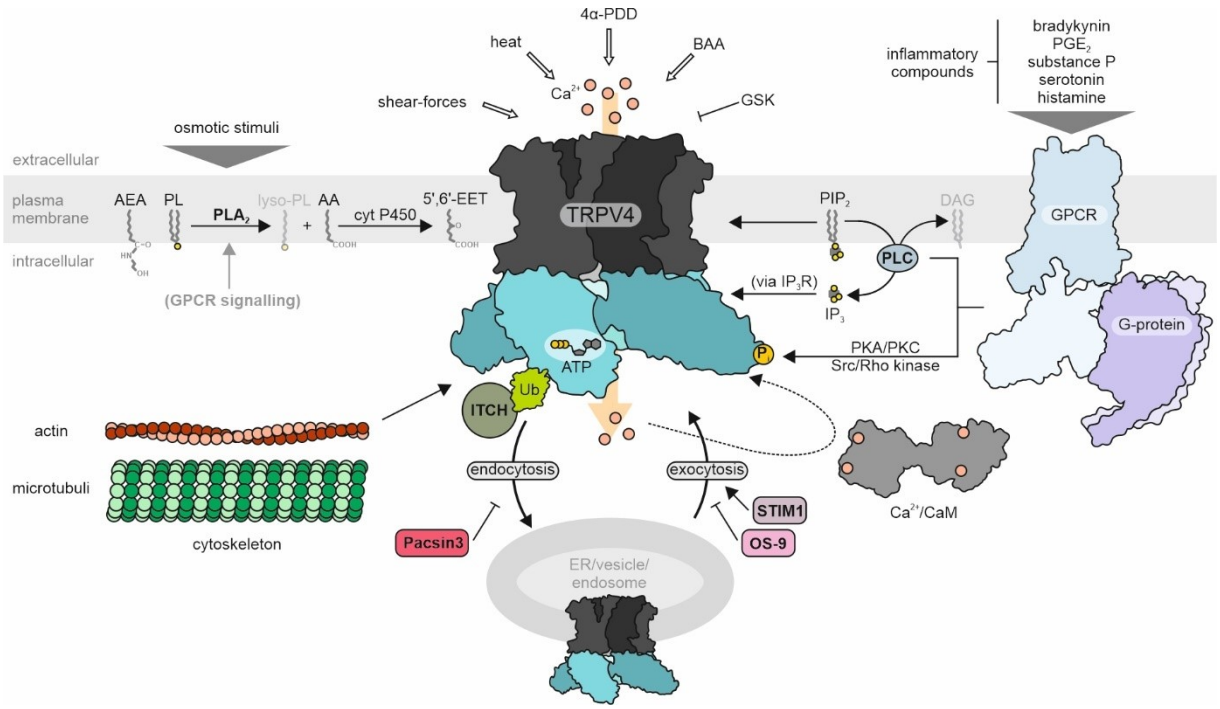


Figure 10: Diverse signalling pathways and interaction partners regulate TRPV4. Multiple stimuli have been identified to activate or modulate TRPV4 either through direct interaction with the channel or indirectly through receptor signalling pathways or by altering the subcellular localization of TRPV4. External stimuli known to activate TRPV4 are osmotic pressure, mechanical force, heat, and several organic molecules like phorbol esters (4α-PDD) or plant compounds (BAA). Antagonists of TRPV4 (GSK) are currently limited to synthetic compounds. Osmotic stimuli activate TRPV4 indirectly through the activation of phospholipase A_2 (PLA_2), which cleaves phospholipids (PL) into arachidonic acid (AA) and lysophospholipids (lyso-PL). AA is converted into 5',6'-epoxyeicosatrienoic acid (5',6'-EET), a TRPV4 agonist via cytochrome P450 (cyt p450). The endogenous cannabinoid anandamide (AEA) activates TRPV4 in a similar pathway via PLA_2, AA, and 5',6'-EET. PLA_2 may also be activated by exogenous ligands in GPCR-dependent signalling pathways. GPCR signalling, e.g. through inflammatory compounds, also controls TRPV4 function by activating kinases (PKA, PKC, Src/Rho-kinase) or phospholipase C (PLC). Kinases were shown to phosphorylate (P_i) the cytosolic domains of TRPV4 and thus modulate its activity. PLC regulates the levels of phosphatidylinositol 4,5-bisphosphate (PIP_2) in the plasma membrane by cleaving it into diacylglycerol (DAG) and inositol 1,4,5,-trisphosphate (IP_3). PIP_2 and IP_3 (via its receptor, IP_3R) both were shown to modulate TRPV4 activity. Cytosolic molecules such as ATP or Ca^{2+} activated calmodulin (Ca^{2+}/CaM) were observed to regulate TRPV4 through interaction with the ARD (cyan). The cytosolic domains of TRPV4 were also described to couple the channel to the cytoskeleton (actin and microtubules) and thereby confer mechanosensitivity. An important regulatory mechanism for TRPV4 is the modulation of its plasma membrane presentation. The cytosolic domains of TRPV4 are targeted for ubiquitination (Ub) by ITCH, thus leading to TRPV4 endocytosis. Endocytosis of TRPV4 seems to be inhibited by the protein Pacsin3. Other proteins regulate TRPV4 presentation in the plasma membrane such as STIM1 and OS-9 by affecting the exocytosis of the channel (see text).

Besides ligand binding, TRPV4 sensitization can depend on post-translational modifications which are commonly coupled to receptor signalling (*253–258*). Several phosphorylation sites targeted by protein kinase C (PKC) and protein kinase A (PKA) were identified in TRPV4 (*254, 255, 257, 258*). Mutation of these sites ablated TRPV4 responses for hypotonic stress. Sensitization to the agonist 4α-PDD and heat can occur through phosphorylation via serum glucocorticoid protein kinase-1 (SGK1) (*259*). Phosphorylation by Src family kinases was observed to sensitize TRPV4 to heat, shear-stress, hypotonic swelling, and PMA (*258*). Phosphorylation was recently established as a component of TRPV4-dependent inflammatory mechanical hyperalgesia (*213*). The mechanism involves activation of the Protease Activated Receptor (PAR_2), a GPCR, during inflammation through neutrophil elastase. The subsequent activation of TRPV4 was dependent on Rho-kinase and PKA. Other GPCR signalling pathways triggered by the inflammatory mediators bradykinin, prostaglandin E_2 (PGE), substance P, serotonin and histamine were identified to hypersensitize TRPV4 for osmotic stimuli (*260*). Serotonin was also observed to activate TRPV4 via the production of agonist 5',6'-EET in pulmonary hypertension in a PLA_2 dependent pathway. (*261*). As outlined in section VI1.2, TRPV channel function is tightly coupled to the phosphoinositide metabolism, which is under the control of GPCRs via phospholipase C (PLC). PIP_2 has been reported as both a sensitizer and desensitizer of TRPV4 for various stimuli (*128, 150*).

The PLC-derived PIP_2 metabolite IP_3 was also observed to desensitize TRPV4 indirectly to 5',6'-epoxyeicosatrienoic acid (EET) via the IP_3 receptor (IP_3R) in the endoplasmic reticulum (ER) (*211, 262*).

The multitude of proposed TRPV4 functions and the multitude of TRPV4 modulators make this protein a very complex system for functional studies. Many of the mechanisms proposed to govern TRPV4 function are still poorly understood. The complexity of *in vivo* or cell-based experiments substantially complicates to identify which cellular components are responsible for TRPV4 modulation when exposed to specific stimuli. Functional analyses would be facilitated by *in vitro* experiments, such as liposome-based functional assays with purified TRPV4 proteins. However, the difficulty in purifying recombinant TRPV4 proteins at sufficient amounts has not only hampered *in vitro* functional studies but also structural studies for a long time. These are essential to understand the polymodal gating behaviour of TRPV4 on the molecular level. Moreover, extracting TRPV4 from its natural environment in the plasma membrane and losing contact with cytoskeletal proteins or other auxiliary proteins might destabilize the TRPV4 channel and cause a physiologically non-relevant gating behaviour. Nevertheless, the recent breakthroughs in TRP channel structural biology through advancements in cryo-EM technologies have significantly expanded our molecular understanding of TRPV4.

Insights into TRPV4 function from high-resolution structures

A human TRPV4 monomer is composed of 871 aa. It displays the typical TRP vanilloid topology with a cytosolic ARD in the N-terminus followed by a linker region and pre-S1 in the N-terminus, a TMD containing the VSLD and the pore domain, and a TRP box and CTD in the C-terminus (Figure 11 A). For several years, high-resolution structural insights into TRPV4 had been limited to the isolated ARDs in the apo (absence of ligands) or ATP- and IP_3-bound states (Figure 11 B) (*21, 150*). These structures can explain how ligands bind to their binding sites in the ARD. However, the ARD makes up for only one-third of the TRPV4 sequence, and structures of isolated ARDs cannot explain how ligand binding in the ARD alters the channel's pore conformation. Recently, the first high-resolution structures of a near 'full-length' TRPV4 channel from *Xenopus tropicalis* (frog) were determined via cryo-EM and X-ray crystallography (Figure 11 C) (*42*). They show in an unprecedented manner how the cytosolic domains crucial for channel function and regulation are connected to the pore domain in the TMD and how ions permeate along the pore of the channel. In combination with functional and biochemical studies, the high-resolution structures of near-full-length TRPV4 and of isolated ARDs in complex with ligands give insights into how the individual TRPV4 domains contribute to channel function and how this might be regulated by ligand binding.

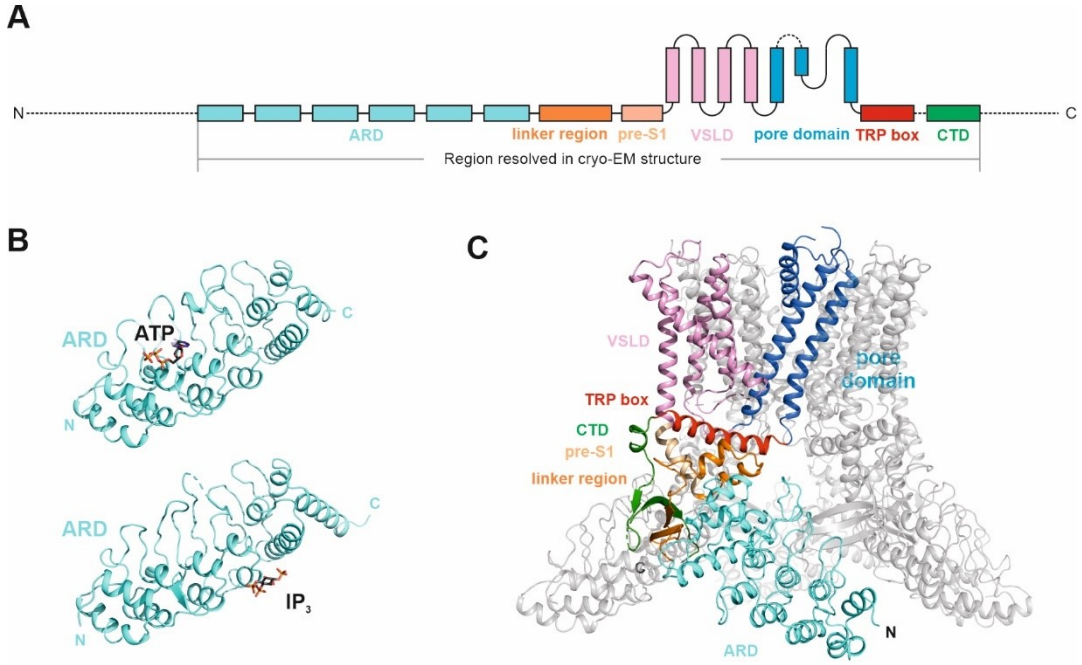


Figure 11: High-resolution structure of TRPV4 and TRPV4 domains. (**A**) Topology model of a TRPV4 monomer. (**B**) High-resolution X-ray crystal structures of the isolated TRPV4 ARD in complex with ATP (PDB: 4dx2) or IP_3 (PDB: 3w9f) yield insights into how ligands may interact with the cytosolic domains (*21, 150*). (**C**) The cryo-EM structure of TRPV4 from *Xenopus tropicalis* (PDB: 6bbj) encompassing residues 146-744 *(42)*. Residues preceding the ARD or succeeding the CTD were either not included in the TRPV4 construct used for structure determination or were unresolved in the cryo-EM maps (shown as dotted lines in A). No full-length structures of TRPV4 in complex with channel regulators are currently available.

As in all TRP channels, the TRPV4 TMD forms the core of the channel as it contains the VSLD and the pore domain. It forms a non-selective cation conducting pore through the membrane along helices S6 and the pore helix (P). Movement of this region controls the opening and closing of the channel and is regulated by a variety of exogenous and endogenous stimuli. The TRPV4 TMD shows a different organization of the VSLD helices compared to those observed in other TRPV channel structures (Figure 12 A). It rotates ~90° counterclockwise around S4, as viewed from the extracellular side, compared to the VSLD in TRPV1 (Figure 12 B). This results in a novel packing of the S4-S5 linker that connects the VSLD to the pore domain. Whereas in all other TRPV channels, the S4-S5 linker forms an α-helix which is tilted more perpendicular to a membrane, the S4-S5 linker in TRPV4 forms a loop which merges into an extended S5 helix (Figure 12 B). In this unique packing, bulky hydrophobic side chains of S3 and of S6 from an adjacent subunit mediate a tight zipper-like interaction (Figure 12 C). Gating stimuli acting on the VSLD might elicit conformational changes in the pore domain via S3-S6 coupling (*42*). The unique TMD packing mode of TRPV4 might explain the unique set of ligands identified to bind the TMD of TRPV4, and the different profile of responses to exogenous stimuli compared the other TRP vanilloids. The lack of an α-helical S4-S5 linker and the altered conformation of the pore domain with respect to the VSLD has two immediate consequences for ligand binding in the TMD. First, TRPV4 does not have a cavity between the VSLD, the pore domain, and the S4-S5 linker that might serve ligand binding as the vanilloid pocket in TRPV1 (see Figure 5 A). This might explain why TRPV4 is not sensitive to vanilloid compounds (*166, 167*). Second, the S4-S5 linker is not well-positioned to form a PIP_2 binding pocket with the TRP box and the VSLD as it is observed in all other TRPV channels (see Figure 5 A). Thus, even though the residues implicated in PIP_2 binding in the TRP box/S4-S5 linker pocket are relatively conserved between the TRPV channels, it may not bind PIP_2 in the TRPV4 channel. As outlined further below, PIP_2 binding sites in TRPV4 have been identified in other regions of the channel (*128, 150*).

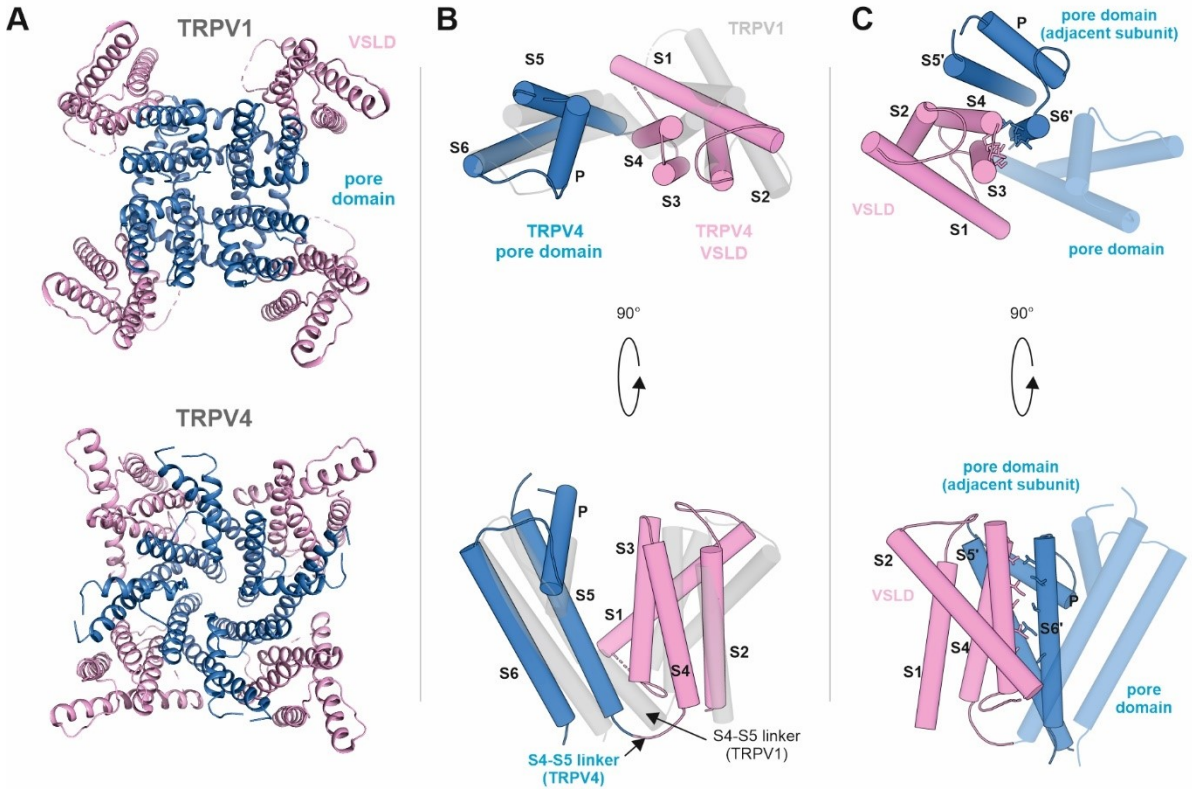


Figure 12: Structural characteristics of the TRPV4 transmembrane domain. (**A**) A top view of the tetrameric transmembrane domains (TMD) from TRPV1 (PDB: 3j5p) and TRPV4 (PDB: 6bbj) shows striking differences in the packing of the voltage sensor-like domain (VSLD) relative to the pore domain. (**B**) An overlay of the TMD of a TRPV1 subunit (grey, transparent) with the TMD of a TRPV4 subunit (coloured). The top view shows that the VSLDs tilt by ~90°. The side view shows that the S4-S5 linker in TRPV1 forms an α-helix, whereas it is a long loop in TRPV4. (**C**) Packing of the S3 helix with the S6 helix of another subunit (termed S6'). Bulky hydrophobic side chains of S3 and S6' mediate a tight zipper-like interaction and may couple conformational changes in the VSLD with the pore domain.

Currently, no high-resolution structural data about ligand binding in the TMD are available, but several mutagenesis-based functional studies identified residues essential for the binding of agonists in the TMD (*263*). The agonist 4α-PDD and the plant compound BAA were described to bind between S3 and S4. However, the residues observed to be critical for TRPV4 activation are slightly different for both agonists, which suggests that the compounds bind differently between S3 and S4 (*215*). A functional study paired with computational modelling suggests that the osmotransducer 5',6'-EET binding site is formed by residues from the S2-S3 linker, S4 and the S4-S5 linker (*264*). This position is very similar to the competitive binding site between 2-APB and lipids in other TRPV channels (see Figure 5) (*49*). Similar as proposed for the competition between 2-APB and lipids in the binding pocket of the VSLD, the identified 5',6'-EET binding site in TRPV4 might also be occupied by a lipid under basal conditions. Computational studies and functional studies paired with mutagenesis strongly suggest that 4α-PDD, BAA, and 5',6'-EET occupy different binding sites in the TRPV4 TMD (*264*). Importantly, residues involved in activation 4α-PDD only partly overlap with those required for heat sensation (*215*). It appears that activation of TRPV4 by 4α-PDD, heat or cell swelling proceed via distinct pathways (*263*).

The most salient structural feature of TRPV channels is the N-terminal ARD. It plays a significant role in the binding of cytosolic regulatory interaction partners and channel assembly. In TRPV4, the ARD (res. 149-398) was shown to comprise binding sites for CaM and adenosine triphosphate (ATP) (*151*). ATP acts as a sensitizer of TRPV4 whereas CaM-binding inhibits TRPV4 after activation, presumably by competing with ATP binding. An X-ray crystal structure of the TRPV4 ARD in complex with ATP (PDB:4dx2) showed that the nucleotide binds along AR1-3 on the ARD surface facing the membrane (*21*). Aligning the ARD-ATP complex structure with the full-length TRPV4 structure shows that the ATP binding pocket is at the intersection between two

subunits (Figure 13 A). In this position, ATP not only binds the ARD of one subunit but presumably also makes contacts to the β-sheet structure of an adjacent subunit, thus being perfectly suited to modulate inter-subunit communication. Strikingly, the ATP binding site in the ARD overlaps with the ARD/β-sheet cavity, which mediates sensitization via a CTD-NTD switch in TRPV3 (see section VI1.2, Figure 8). Therefore, ATP binding seems to target a binding site which is also exploited by in other regulatory pathways (*151*). Several residues important for the TRPV4-CaM interaction have been identified in the ARD, the exact binding mode is, however, unclear (*151*). Notably, also a C-terminal residue (W822) was shown to be essential for CaM binding (*242*).

The ARD was also described to contain a binding pocket for PIP_2, which supposedly desensitizes TRPV4 to activation by 4α-PDD (*150*). In an X-ray crystal structure of an isolated ARD (PDB:3w9f), the binding site of the PIP_2 headgroup, i.e. IP_3, localizes at AR4. This binding site was confirmed in functional and biochemical assays by the same authors (*150*). In the full-length TRPV4 channel, however, the putative PIP_2 binding site in the ARD faces away from the membrane into the cytosol (Figure 13 A). In a physiological context, it is unlikely that a PIP_2 lipid molecule is pulled out of the membrane to bind in this position. The observed binding site might thus be an artefact from the experimental setup used to study the TRPV4-PIP_2 interactions. Lipid-binding experiments were performed with the isolated ARD which can bind in physiologically non-relevant orientations to membranes. For ion-flux assays, cells expressing TRPV4 were perfused with soluble PIP_2 derivatives. This might allow for lipid binding to regions which are not accessible for lipid binding in a physiological context (*150*). It is thus questionable whether the ARD is an actual site for PIP_2 binding of whether PIP_2 binding occurs at the described site.

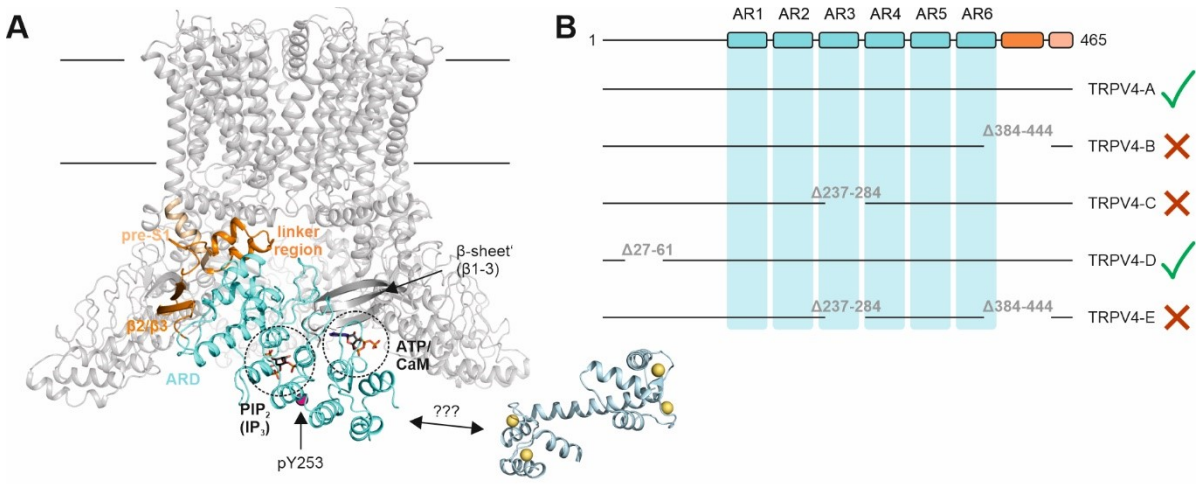


Figure 13: Structural and functional characteristics of the cytosolic TRPV4 N-terminus. (**A**) Structural view of ligand binding in the TRPV4 N-terminus (ARD, linker region, pre-S1). X-ray crystal structures of the TRPV4 ARD in complex with ATP (PDB: 4dx2) or IP_3 (PDB: 3w9f) were aligned with the ARD in the TRPV4 full-length structure (PDB: 6bbj). (**B**) TRPV4 spliced variants TRPV4-A to E. 'Δ' indicates the residues missing in a splice variant. TRPV4-A is commonly referred to as the wildtype protein. Whereas TRPV4-A and TRPV4-D are functionally intact, TRPV4-B, -C, and -E are retained in the ER of cells, presumably due to misfolding (*152*).

The ARD is not only important as a binding site for regulatory interaction partners but seems to be crucial for channel assembly and trafficking (*152*). Several TRPV4 splice variants have been identified to target the ARD for deletion. The splice variant TRPV4-A, which contains all TRPV4 exons, forms functional channels which are targeted to the plasma membrane. The spliced variants B (Δ384-444), C (Δ237-284), and E (Δ237-284 and Δ384-444) are retained in the ER and cannot oligomerize. TRPV4-C and E miss most of AR3, and TRPV4-B has a partly truncated AR6. Truncating ARs might impair channel assembly or lead to ARD misfolding and TRPV4 retention in the ER (*152*). In agreement with this hypothesis, the full-length TRPV4 structure reveals

that AR3 and AR4 make contacts with the β-sheet of an adjacent subunit. This interaction is likely to be disturbed when AR3 is deleted. Deleting AR3 also removes the ATP binding site and might therefore abolish the binding of interaction partners important for channel trafficking or regulation. Furthermore, AR3 contains Y253, which is required to be phosphorylated by Src family kinase for TRPV4 to be sensitive towards stimulation by hypotonic conditions (*258*). Thus, several regulatory mechanisms may be disturbed in TRPV4 splice variants lacking AR3.

The C-terminus of TRPV4 contains the TRP box, which in all TRP channels, forms a helix perpendicular to the membrane (Figure 14 A) (*42*). Positioned underneath the S4-S5 linker, it can propagate conformational changes from the cytosolic domains to the pore. Functional studies paired with MD simulations have identified a hydrogen bond between the oxygen atom of L596 in the S4-S5 linker (TMD) and the indole amide hydrogen of W733 in the TRP box. The hydrogen bond was proposed to act as a latch to stabilize a closed state of the TRPV4 channel (Figure 14 A) (*61*). Based on mutagenesis studies and MD simulations, the side chain of L596 is suggested to bind to lipids of the inner leaflet, thus counteracting the L596-W733 interaction and allowing S6 to move to a more open conformation (Figure 14 B). Lipid binding at the S4-S5 linker, therefore, seems to modulate TRPV4 channel opening and closing (*265*). The TRP box also makes contacts with the pre-S1 and linker region in the N-terminus (Figure 14 A). This assembly seems to be held together by an alternating hydrogen bond network between K462 (pre-S1), E745 (TRP box), and D425 (linker region). The interaction was described to be essential for a structural conformation which enables inter-subunit interactions for correct channel assembly and trafficking (*156*). Following the TRP box, the C-terminal domain (CTD) of TRPV4 forms a loop that connects to β-strand 3, which forms the β-sheet with the linker region (Figure 14 A). The β-sheet resides at the intersection between ARDs from neighbouring subunits and seems to be necessary for subunit assembly and communication. Residues after β3 are not resolved in the cryo-EM structure (*42*).

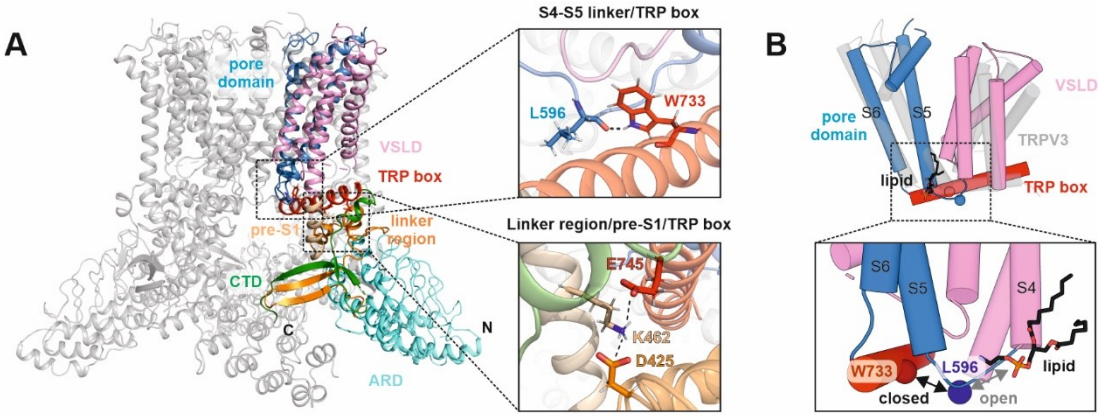


Figure 14: Interdomain communication between the N-terminus, the TMD, and the C-terminus modulates TRPV4 function. (**A**) The cryo-EM structure of TRPV4 reveals how the individual domains in the TRPV4 N-terminus, the TMD, and the C-terminus interact. A hydrogen bond between W733 in the TRP box and L596 in the S4-S5 linker was identified in MD simulations to maintain TRPV4 in a closed conformation (*61*). This provides a structural explanation of how the displacement of the TRP box controls the pore. The TRP box interacts with the linker region and the pre-S1 helix in the N-terminus via an alternating hydrogen bond network between residues D425, K462, and E745. (**B**) A hydrophobic tug exerted by a lipid was identified via mutagenesis and MD simulations to mediate the opening of TRPV4 (*265*). A lipid can interfere with the L596-W733 interaction by binding to the L596 side chain and thus pulling it away from W733 in the TRP box. The loss of the L596-W733 interaction subsequently opens the pore. Notably, a lipid at this position was observed in the cryo-EM structure of TRPV3 reconstituted in lipid nanodiscs (PDB: 6uw4). Importantly, the lipid-binding site in the hydrophobic tug mechanism is different from the putative PIP_2 binding site proposed for TRPV channels (see Figure 5).

Full-length TRPV4 cryo-EM and X-ray crystal structures provide detailed insights into the TRPV4 channel architecture and its gating mechanism. They can partly explain observations from functional and biochemical studies on the molecular level and open the path to elucidate the mechanistic details of TRPV4 signalling. The

complex subunit interaction network identified from its high-resolution structure provides a basis to explore how TRPV4 can sense its broad spectrum of stimuli and convert them into ion currents. More structural studies of TRPV4 in complex with agonists, antagonists or other functional regulatory are, however, needed to unravel the polymodal gating behaviour in molecular detail. In this context, the available structure of the unliganded TRPV4 channel provides a template for computer-guided screenings of putative ligand binding sites or investigations via molecular dynamics (MD) simulations.

Putatively disordered N- and C-termini regulate TRPV4

The full-length structure of TRPV4 is crucial to complement functional and biochemical data such as ion-flux assays, mutagenesis, pull-downs, or proteomics approaches in order to understand TRPV4 function on a molecular level. However, around 25% of the TRPV4 sequence is not contained in the TRPV4 cryo-EM model. The *X. tropicalis* TRPV4 construct used for cryo-EM and X-ray crystallographic studies lacked the N-terminal residues 1-132 as well as the C-terminal residues 798-868. Moreover, several residues which were included in the construct, i.e. a proline-rich region (PRR) spanning residues 133-143 in the N-terminus and residues 785-798 in the C-terminus, were still not resolved in the final structure (see Figure 11) (*42, 266*). This is in line with cryo-EM studies on other TRPV channels, in which the distal N- and C-termini were either deleted in the full-length TRPV constructs or were not resolved in the final EM density (*20–24, 39–41, 49, 50, 78, 131, 134–150, 161–163, 196*).

The TRPV4 amino acid sequence is highly conserved across species, particularly in the transmembrane domain and in the cytosolic ankyrin repeat domain (ARD). *X. tropicalis* TRPV4 shares about 78% sequence identity with the human protein (Appendix Figure 124) (*42*). However, the N-terminal region preceding the ARD and the ~90 C-terminal residues downstream of the TRP box share only 49% and 60% sequence identity between *X. tropicalis* and *H. sapiens* TRPV4, respectively. Both regions are substantially enriched in disorder promoting charged and polar residues, as well as glycine and proline in TRPV4 proteins across species (Figure 15 A, Appendix Figure 124). Accordingly, bioinformatic analyses predict high disorder probabilities in the TRPV4 N-terminus preceding the ARD across species (Figure 15 A), similar to what is predicted for N- and C-termini of other TRPV channels (Figure 7). Likewise, the C-terminal region which was deleted or unresolved in the *X. tropicalis* TRPV4 cryo-EM structure is predicted to be disordered, but notably, with a lower average propensity compared to the N-terminus (Figure 15 A). In agreement with X-ray crystallographic and cryo-electron microscopic data (*21, 42, 150*), the rest of the channel is predicted to be ordered in bioinformatic analyses.

TRPV4 contains ~150 residues in the distal N-terminus preceding the ARD for which no structural information is yet available (*267*). This region contains a proline-rich region (PRR) which is unique within the TRPV family (Figure 15 B). The PRR (res. 138-144) was found to interact with Pacsin (protein kinase C and casein kinase substrate in neurons) proteins, of which Pacsin3 regulates the subcellular localization of TRPV4 (*76, 77*). Simultaneously, it reduces the basal TRPV4 activity and desensitizes TRPV4 for specific physiological stimuli such as heat and hypotonicity. Several residues upstream of the PRR, a phosphoinositide binding domain was identified as a binding site for PIP_2 (res. 121-125) (Figure 15 B). In contrast to its putative interaction with the ARD (*150*), PIP_2 binding to the PBD in the distal N-terminus was observed to sensitize TRPV4 to stimulation by temperature and hypotonicity (*128*). Other binding sites for regulatory interaction partners in regions preceding the PBD, i.e. res. 1-120, have not been described yet. However, several phosphorylation sites (S88, Y110, and S134) in the distal N-terminus of TRPV4 have been identified (Figure 15 B). They are in close proximity to the PIP_2 and Pacsin binding sites. Phosphorylation at Y110 by an Src-type tyrosine kinase was proposed to modulate the response of TRPV4 to shear-stress and PMA, but not to 4α-PDD and heat (*257*). Residues S88 and S134 were proposed as putative sites targets of PKC, but no functional implications have been reported yet (*268*). Phosphorylation at S134 might influence how proteins bind to the PRR, whereas phosphorylation at Y110 might modulate lipid interactions in the PBD. Phosphorylation at S88 might point to yet unknown functional roles of the regions preceding the PBD. In a fifth splice variant, termed TRPV4-D, residues 27-61 are deleted, but the channel is functionally intact and correctly targeted to the PM (Figure 13 B) (*152*). This might

suggest that these residues have no essential functions. However, they might play functional roles in certain tissues or under specific conditions which were not screened in functional studies, yet.

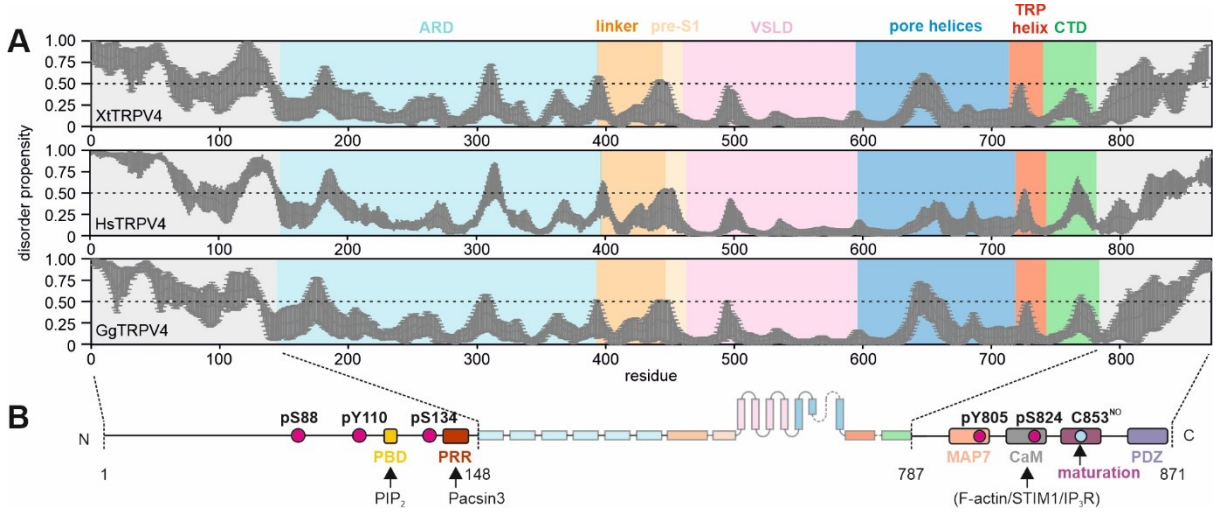


Figure 15: Predicted disorder in the full-length TRPV4 channel. (**A**) Sequence-based disorder prediction of the *X. tropicalis*, *H. sapiens*, and *G. gallus* TRPV4 channel based on IUPRED (*84*), PONDR (*85–90*), PONDR-FIT (*91*), and DISOPRED (*92*). Plotted is the average disorder propensity of all predictions. The error bars represent the standard deviation from the mean value. (**B**) Functional domains or residues identified in the disordered N- and C-terminal regions of TRPV4. Phosphorylation sites are indicated with pink dots. The corresponding residue is indicated with a 'p'. The blue dot at C853 indicates a putative nitrosylation site ($C853^{NO}$). PBD: Phosphoinositide binding domain – binding site for PIP_2 (yellow); PRR: Proline-rich region – Pacsin3 binding site (red); MAP7: Identified binding site of microtubuli-associated protein 7; CaM: Calmodulin binding site – overlaps with the binding sites of F-actin, STIM1, and the IP_3 receptor; Maturation: Residues important for proper folding, maturation, and trafficking of TRPV4; PDZ: Putative binding site for PDZ domain-containing proteins.

Similar as the distal N-terminus, the C-terminal region after the CTD contains several protein binding sites, such as CaM and MAP7 (microtubule-associated protein 7) and a sequence which suggests the binding of PDZ domain containing proteins, termed the PDZ binding like domain (Figure 15 B). MAP7 binding at residues 789-809 mediates the coupling of TRPV4 with mictrotubules, which redistributes TRPV4 towards the membrane and increases its functional expression (*37*). The MAP7 binding site contains a pohosphorylation site at Y805 with yet unknown functional roles (Figure 15 B) (*257*). The region comprising residues 838-857 was identified as a binding site of CaM (*242*). Binding of CaM was demonstrated to potentiate TRPV4-mediated Ca^{2+} currents in response to hypotonic conditions or 4α-PDD. Potentiation of TRPV4-mediated Ca^{2+}-currents was followed by inhibition of TRPV4, presumably by a Ca^{2+}-dependent negative feedback mechanism (*242*). Interestingly, another study showed that CaM binding is not restricted to the C-terminus but also seems to interact with the ARD, thereby modulating ATP-dependent sensitization of TRPV4 for various stimuli (*151*). The role of CaM for TRPV4 activity appears complicated and the mechanism of CaM-binding not entirely understood (*242, 249, 250*).

Notably, binding sites for several other proteins seem to overlap with the CaM binding site, e.g. F-actin (Figure 15 B) (*234*). The interaction with F-actin requires phosphorylation at S824 by serum glucocorticoid-induced protein kinase 1 (SGK1) (*256*). PKA-dependent phosphorylation at S824 was shown to be required for propper subcellular localization of TRPV4 and increases TRPV4 mediated Ca^{2+}-currents in resting cells or cells treated with AA (*254, 255*). The stromal interaction molecule 1 (STIM1) is another C-terminal interaction partner of TRPV4 (*253*). STIM1 is a membrane protein which migrates from the ER to the PM after the depletion of Ca^{2+}-stores to activiate calcium-release-activated calcium (CRAC) channels (*269*). In epithelial cells, it acts as an auxilary protein of TRPV4 by binding to its C-terminus. STIM1 seems to be required for TRPV4 trafficking from the ER to the PM. The interaction is regulated by phosphorylation of TRPV4 residue S824 which suggests that the STIM1 binding site overlaps with that of CaM (Figure 15 B) (*253*). Similarly, the CaM binding seems

to overlap with the binding of the IP_3 receptor (IP_3R) which mediates the IP_3-dependent sentization of TRPV4 to the agonist 5',6'-EET (Figure 15 B) (*211, 262*). Residues 838-857 downstream of the CaM binding site were found to form a small domain required for channel folding, maturation, and trafficking. Within this region, nitric oxide (NO) was shown to cause *S*-nitrosylation of TRPV4 at C853, thereby inhibiting the activation of TRPV4 (Figure 15 B) (*270*).

The examples described above demonstrate that TRPV4 regions which were not accessible via cryo-EM or X-ray crystallography contain many regulatory domains or residues. These regions are crucial for proper channel function and therefore require more detailed analyses, both structurally and functionally.

TRPV4 channelopathies are caused by mutations in regulatory domains

TRPV4 has been linked to several pathological conditions and is attracting increasing interest as a potential drug target (*204, 271*). Calcium signalling by TRPV4 has been implicated in inflammatory- or neuropathy-induced hyperalgesia (*197, 198, 213, 236, 260, 272*), cancer (*238, 273–275*), or brain injuries (*226, 239, 276*). More so, over 60 point-mutations in the *trpv4* gene have become infamous for causing several hereditary diseases with diverse phenotypes and symptoms ranging from mild to lethal, so-called TRPV4 channelopathies (*204*). Most diseases linked to TRPV4 mutations can be classified as skeletal dysplasias, i.e. abnormalities or malformations of bone structures, cartilage, and tissue (*277–281*), as well as peripheral axonal neuropathies, i.e. degenerative diseases of the peripheral nerves (*224, 279, 282, 283*). The channelopathy mutations predominantly show gain-of-function characteristics, and the severity of the diseases seems to correlate with the extent to which TRPV4 signalling is enhanced. Three loss-of-function mutations in the *trpv4* gene were found to cause the osteoarthropathy familial digital arthropathy-brachydactyly (FDAB), a disease where patients have shortened fingers (*284*). A single mutation is known to cause hyponatremia, a condition describing low sodium concentrations in the blood serum. TRPV4 channels carrying the hyponatremia mutation suffer from sensitivity to hypotonic stress or 5',6'-EET (*285*).

The phenotypic diversity of TRPV4 mutations is unique among ion channels and has mostly remained enigmatic (*204*). Several mechanisms by which mutations can dysregulate TRPV4 and may cause channelopathies have been proposed from functional studies, among them TRPV4-induced Ca^{2+} overload and toxicity (Figure 16 A) (*19, 286–288*). Many TRPV4 mutations are gain-of-function mutants that ultimately lead to an increased Ca^{2+} influx via TRPV4. Cells expressing mutant TRPV4 channel might therefore face problems with maintaining non-toxic intra-cellular Ca^{2+} levels. Besides increasing intracellular Ca^{2+} concentrations to toxic levels, TRPV4 gain-of-function mutations can also affect Ca^{2+}-dependent downstream signalling pathways, as proposed in peripheral neuropathies. The outgrowth of neurites of peripheral neurons in response to human growth factor (NGF) is mediated by a PLA_2-dependent TRPV4 signalling pathway (*224*). TRPV4 neuropathy mutants might directly disturb this pathway by altering TRPV4-mediated Ca^{2+} signalling and thereby lead to a defective neuronal development (Figure 16 B). Disturbed TRPV4 function might alter the expression of genes downstream of TRPV4 activation (Figure 16 C), as suggested for the development of skeletal dysplasia. Bone formation in the growth plate is initiated by the transcription factor SOX-9, which itself is presumably expressed downstream of a Ca^{2+}/CaM-dependent TRPV4 signalling pathway (*223*). Dysregulation of SOX-9 expression and activity from the defective signalling of mutant TRPV4 channels might therefore directly disturb the controlled initiation of bone formation, thus resulting in bone deformities. Mutations in TRPV4 might also dysregulate cellular trafficking and maturation of TRPV4 (Figure 16 D), either through structural perturbations and the subsequent retainment of TRPV4 in the ER (*156*), or through disturbing interactions with proteins involved in TRPV4 trafficking, e.g. Pacsin3, OS-9, STIM1, or ITCH (*76, 77, 251, 252*). Dysregulated protein-protein interactions might also lead to uncontrolled TRPV4 signalling and defective downstream pathways (Figure 16 E) (*21, 250*).

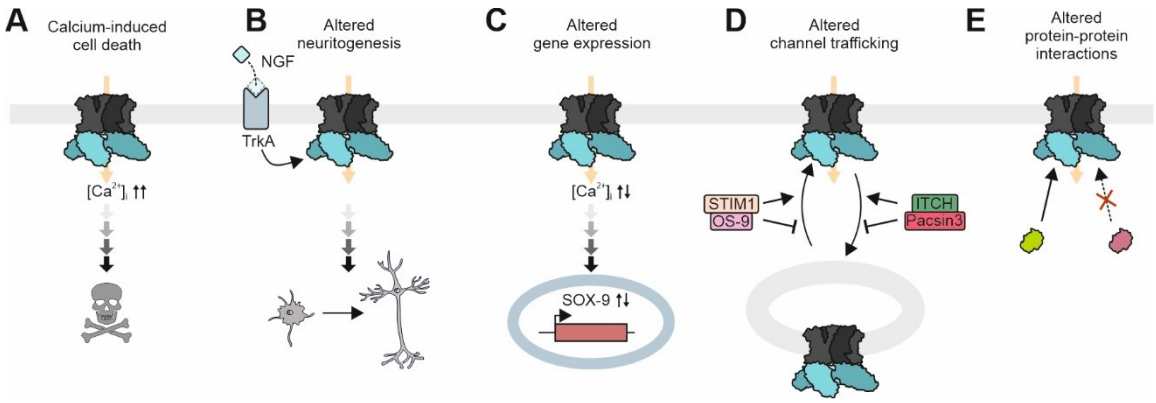


Figure 16: Proposed TRPV4 channelopathy mechanisms. (**A**) Gain-of-function mutations in TRPV4 channels lead to Ca^{2+} overload of cells followed by apoptosis (*19, 286–288*). (**B**) TRPV4 plays a crucial role in neuron development. The signalling pathway involves the binding of nerve growth factor (NGF) to tropomyosin receptor kinase A (TrkA) and subsequent activation of PLA_2 (*224*). Defective TRPV4 channels may disturb the signalling cascade and impair neurite outgrowth. (**C**) Dysregulated Ca^{2+} signalling may generally affect downstream signalling events. In this sense, TRPV4 mutations were proposed to alter gene expression of SOX-9, a transcription factor involved in bone and cartilage development (*223*). (**D**) Mutations might alter the trafficking of TRPV4 between the plasma membrane, the ER, or intracellular vesicles, e.g. by disturbing the binding of proteins such as Pacsin3 (*76*), OS-9 (*252*), ITCH (*251*), or STIM1 (*253*). (**E**) Disturbed TRPV4-protein interactions are believed to be a general mechanism underlying the aetiology of TRPV4 channelopathies (*21, 250*). Two possible scenarios for how disturbed TRPV4-protein interactions may cause enhanced TRPV4 signalling: Enhanced interaction with positive TRPV4 regulators, or reduced interaction with negative TRPV4 regulators. Notably, the binding of ligands other than proteins, e.g. lipids and small molecules, may also be affected by channelopathy mutations.

The consequences of channelopathy mutations on TRPV4 gating are commonly assessed in cell-based experiments or mouse models. For this, TRPV4 mutant channels are expressed heterologously, and the TRPV4-mediated ion currents are assessed under basal conditions and in response to different stimuli via patch-clamp recordings or calcium imaging. The recent determination of a high-resolution cryo-EM structure in combination with X-ray crystal structures of the isolated TRPV4 ARD in complex with ligands and has opened the path to relate the functional consequences of channelopathy mutations to structural details of the TRPV4 channel (*42*). Mapping the known disease mutants onto the full-length TRPV4 structure illustrates how the mutations are distributed throughout the protein (Figure 17). The TMD almost exclusively contains mutations linked to skeletal dysplasia, with T701I as a neuropathy mutation being the only exception (Figure 17 A). A hot spot for disease mutations in the TMD is the S4-S5 linker and surrounding regions (F417del, S542Y, F592L, F594H, L596P, G600W, Y602C, I604M, and T740I) (*289–293*). The S4-S5 linker connects the VSLD to the pore domain and is itself connected to the TRP box through hydrogen bonds. The TRP box is connected to the linker region and pre-S1 helix in the N-terminus through a hydrogen bond network (see Figure 14) (*60, 61, 265*). All these domains are targeted by at least one channelopathy mutation, mostly affecting hydrophobic or aromatic residues involved in tertiary interactions (Figure 17 B). The interaction network between VSLD, pore domain, S4-S5 linker, and TRP box has been recognized as crucial for gating of the pore domain (*61, 265*). Conformational perturbation in this region through mutations might therefore uncouple the pore domain from the rest of the channel leading to uncontrolled ion flux through the pore (*60*). Conformational changes of the pore domain may also be caused directly by channelopathy mutations in the S5 and S6 helices of the pore. Several skeletal dysplasia-causing mutations (R616Q, F617L, L618Q, V620I, L709M, and A716S) cluster in the centre of S5 and S6 (Figure 17 C) (*280, 289, 294, 295*). Some of the targeted residues mediate S5-S6 contacts through tertiary interactions between hydrophobic side chains. Mutations probably destabilize the conformation of the pore helices, thus allowing ion to pass the pore more easily. Skeletal dysplasia mutations are not restricted to the TMD. They also localize in the cytosolic domains of TRPV4, mostly in the ARD (E183K, K197R, L199F, K276E, E278K, T295A, I331F/T, D333G, V342F) (Figure 17 D) (*21, 277, 289, 293, 296*), and also in the C-terminal β-sheet (R775K, C777Y) (Figure 17 A) (*289*). Most of the mutations target residues in the interfaces which connect the cytosolic domain to the TMD or mediate contacts between subunits. They might therefore cause improper channel assembly or conformational uncoupling between the TMD and the cytosolic domains or between subunits. Mutations causing peripheral axonal neuropathies almost entirely line the outer perimeter

of the ARD (an exception is T701I in S6) (Figure 17 D). In this position, mutations not only have the potential to perturb the conformation of the ARD but also to interfere with the binding of regulatory interaction partners. In contrast to TRPV4 mutants involved in skeletal dysplasia and neuropathies, osteoarthropathy mutations (G270V, R271P, F272L) result in a loss-of-function (*284*). All affected residues localize to the loop between AR3 and AR4, which contacts the β-sheet of an adjacent subunit (Figure 17 E). In this position, residue F273 is surrounded by aromatic side chains of AR3, AR4 and the β-sheet of the adjacent subunit. The mutation F237L might weaken intersubunit interactions by destabilizing the interaction network of aromatic side chains. The mutations R271P and G270V might have similar effects. Destabilization of intersubunit interactions might interfere with proper channel assembly.

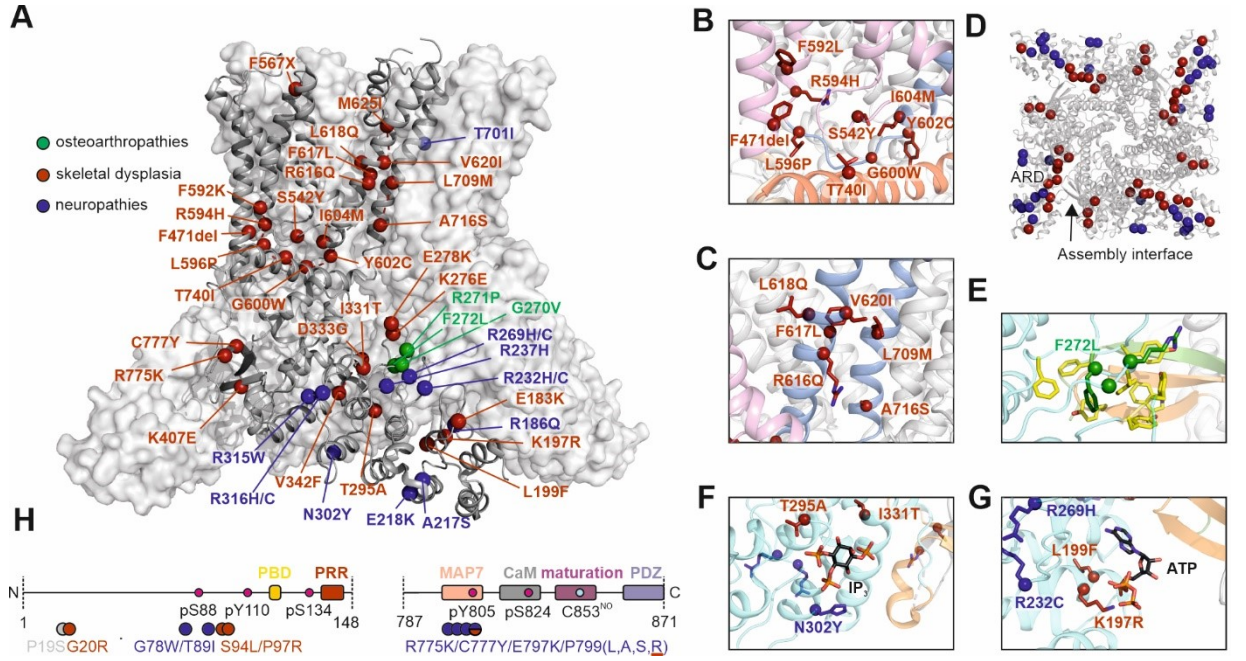


Figure 17: Structural basis of TRPV4 channelopathy mutations. (**A**) Residues involved in TRPV4 channelopathies mapped onto the cryo-EM structure of *X. tropicalis* TRPV4 (PDB: 6bbj). Peripheral axonal neuropathy-causing mutants are shown in blue, skeletal dysplasia mutations red, and osteoarthropathy mutants in green. The residue number correspond to the human TRPV4 sequence. (**B**) Magnified view of disease mutations in the intersection between VSLD, S4-S5 linker, and TRP box. Mutations in S5/S6. This region is solely targeted by skeletal dysplasia mutations. (**C**) Skeletal dysplasia mutants in the central region of the pore helices S5 and S6. (**D**) Skeletal dysplasia and neuropathy mutants in the cytosolic domains of a tetrameric assembly. The TMD is not shown for better visualization. Skeletal dysplasia mutations mainly localize in the intersections between the ARD and the TMD or between the subunits. The neuropathy mutations appear to line the periphery of the cytosolic domains. (**E**) The three osteoarthropathy-causing mutations cluster in the loop between AR3 and AR4. The residues are embedded in a tight network of aromatic side chains and make contacts to the β-sheet of a neighbouring subunit. (**F**) TRPV4 disease mutations in the vicinity of the IP_3/PIP_2 binding site, as observed in the X-ray crystal structure of the TRPV4 ARD in complex with IP_3 (PDB: 3w9f). (G) TRPV4 disease mutations in the vicinity of the ATP binding site observed in the X-ray crystal structure of the TRPV4 ARD/ATP complex (PDB: 4dx2). (**H**) The N-terminal regions which were not resolved in the cryo-EM structure are also targeted by channelopathy mutations (shown on the bottom). These partly overlap with identified ligand binding sites or are close to phosphorylation sites (same colour coding scheme and abbreviations as in Figure 15 B). Interestingly, the N-terminus displays a mutation at P19S, which is neither of the three major channelopathy classes but causes a condition called hyponatraemia (*285*).

Three residues (T295A, I331T, N302Y) targeted by channelopathy mutations are in the vicinity of an IP_3 binding site, which was identified in an X-ray crystallographic study (Figure 17 F, see also Figure 13) (*150*). However, no functional data are available whether these mutations affect the interaction of the ARD with IP_3 or PIP_2, respectively. Interestingly, several channelopathy mutations in the ARD which are distant from the IP_3/PIP_2 binding site on the opposite side of the ARD (R232C, R269H, R315W, R316H) were observed to abolish the interaction of PIP_2-containing liposomes with the isolated ARD *in vitro* and suppressed the inhibitory effect of PIP_2 binding to the ARD on TRPV4 activity in HEK293 cells (*150*). These residues are located in the fingers

between the ARs and make contacts to an adjacent subunit. An explanation for the suppressed IP_3/PIP_2 interaction of TRPV4 channelopathy mutants might be an altered conformation of the ARD with a disturbed IP_3/PIP_2 binding pocket. Residues implicated in channelopathies (K197, L199, R232, and R269) also line the ATP binding site and thus might affect the ARD-ATP interaction (Figure 17 G). Indeed, isolated ARDs with the channelopathy mutations K197R (skeletal dysplasia), R232C (neuropathy), and R269H (neuropathy) reduce binding of ATP (*21*). The same effect was observed for the V342F mutation (skeletal dysplasia), even though V342 is relatively distant from the ATP binding site. By contrast, E183K and E278K enhanced ATP binding, although both residues have no direct interaction with ATP (*21*). These examples demonstrate that not only residues in the direct vicinity of ligand binding sites in the TRPV4 ARD can lead to defective interactions. The mutation of distant residues can achieve similar effects. This suggests that ligand binding sites in the ARD are allosterically coupled to distant regions. Therefore, mutations along the allosteric pathway can not only disturb the binding of an interaction partner but also interfere with the propagation of conformational changes from the ligand binding site to the TMD.

Several of the hitherto identified disease mutations are located in regions for which no structural data are yet available, i.e. the putatively disordered N- and C-terminal regions (Figure 17 H). So far, six disease mutations have been described which target the N-terminal region preceding the ARD, i.e. P19S (hyponatraemia) (*285*), G20R (motor neuropathy) (*296*), G78W and T89I (both metabotropic dysplasia) (*277, 297*), S94L and P97R (both distal spinal muscular atrophy) (*297, 298*). The mutations G78W, T89I, S94L, and P97R are close to the PBD and the PRR, two regulatory sites for lipid and protein binding, and proximal to two putative phosphorylation sites at S88 and Y110. They might elicit their pathogenic potential by disturbing interactions of TRPV4 with PIP_2 or with Pacsin3 in the distal N-terminus or by interfering with TRPV4 phosphorylation. The mutations P19S and G20R are relatively distant from the other channelopathy mutations. Thus, even though no ligand binding or phosphorylation sites have been identified in this region yet, it seems to be essential for channel regulation. Several identified skeletal dysplasia-related mutations have localize to the distal TRPV4 C-terminus including R775K, C777Y, E797K, and P799L/A/S/R (Figure 17 H) (*277–280, 289, 291*). Interestingly, the mutation P799R causes not only a skeletal dysplasia but also a peripheral neuropathy (*278, 279*). The mutants E797K and P799R localize in the MAP7 binding site, which directly succeeds the C-terminal β-strand (*299*). The introduction of positively charged residues may thus interfere with and uncouple TRPV4 from the cytoskeleton. Interestingly, E797K and P799R mutations could be shown to impede the CaM-dependent regulation TRPV4 and lead to constitutively active channels. (*250*). It was proposed that the charge switch through E797K may interfere with electrostatic interactions and that P799R (or P799L/A/S) induces a conformational change of the C-terminus.

In summary, TRPV4 displays a diverse spectrum of TRPV4 channelopathies which are caused by mutations that spread throughout the entire protein sequence (*204*). The high-resolution structure of TRPV4 links the amino acid sequence positions of the mutants to their positions in the three-dimensional channel architecture. Notably, the mutations do not appear to be arbitrarily dispersed (Figure 17 A). Instead, they cluster in regions which were previously identified as regulatory ligand binding sites or seemed to control intersubunit contacts and pore opening or closing in the TMD. The combination of functional and biochemical studies and high-resolution structural studies of TRPV4 hold immense promise to explain how mutations dysregulate TRPV4 signalling. However, the disordered TRPV4 N- and C-termini have so far evaded structure determination, thus complicating attempts to explain the mechanistic details of how mutations in these regions interfere with TRPV4 function (Figure 17 H). More structural and functional investigations which take the distal N- and C-termini into account are therefore required to unravel the puzzle of TRPV4 channelopathies.

PIP_2 and Pacsin3 modulate TRPV4 by binding to its disordered N-terminus

TRPV4 is an essential transducer of hypotonic stimuli and mild to warm temperatures in various physiological processes (*206*). The ion gating of TRPV4 needs to be tightly regulated such that a stimulus does not cause a continuous activation of TRPV4. This can have dramatic consequences, as demonstrated by the fact that gain-of-function mutations in TRPV4 cause severe diseases (*204*). The disordered TRPV4 N-terminus appears to

be particularly important to fine-tune how TRPV4 responds to thermal and hypotonic stimuli. Close to the ARD in the disordered N-terminus, a phosphoinositide binding domain (PBD) with the sequence ^{121}KRWRK^{125} was identified as the binding site for the lipid PIP_2 (Figure 18) (*128*). Mutating the PBD to ^{121}AAWAA^{125} or depleting PIP_2 in HEK293 cells rendered TRPV4 insensitive to heat and osmotic stimuli. PIP_2 was accordingly classified as a TRPV4 sensitizer which is required to sense heat and hypotonic stimuli under physiological conditions (*128*). FRET studies of C-terminally CFP- and YFP-tagged TRPV4 channels expressed in HEK293 cells revealed that PIP_2 binding favours an expanded conformation of the cytosolic TRPV4 domains. Deleting all N-terminal residues up to the PRR (Δ1-130), mutating the PBD to ^{121}AAWAA^{125}, or depleting PIP_2 levels by activation of 5-phosphatase IV leads to a more compact conformation of the cytosolic domains (*128*). PIP_2 was also shown to bind to the ARD and thereby suppress TRPV4 channel activity in HEK293 cells (*150*). In the corresponding studies, however, the effect of PIP_2 on TRPV4 activity was solely assessed by perfusing cells expressing TRPV4 with soluble PIP_2 derivatives. In addition, the binding site of the PIP_2 head group in the ARD identified via X-ray crystallography is in a position which faces away from the membrane in a full-length TRPV4 channel. It is thus questionable whether the binding site identified via *in vitro* experiments is physiologically relevant. A more recent study reproduced the inhibitory effects of PIP_2 on TRPV4 activity in mouse brain capillary cells (*183*). In this study, the inhibition of PIP_2 synthesis by treating cells expressing TRPV4 with synthetic inhibitors and the perfusion of the cells with soluble PIP_2 derivatives had identical effects on TRPV4 activity. The exact role of PIP_2 for TRPV4 function thus seems complicated. Whether the PBD in the disordered N-terminus and the putative PIP_2 binding site in the ARD act together or compete for PIP_2 binding, or whether different conditions prefer occupation of one PIP_2 binding site or the other is unclear.

PIP_2 is an important lipid second messenger which makes up for approx. 1-2% of the lipids in the plasma membrane and localizes to the inner leaflet with its head group directed into the cytosol (*171*). In several GPCR- and RTK-dependent signalling pathways, PIP_2 is metabolized by PLC into diacylglycerol (DAG) and IP_3 (*171, 300*). Receptor-signalling thus has direct control over the PIP_2 levels in the PM and is an indirect regulator of PIP_2-dependent protein functions (see Figure 10). In mouse brain capillary endothelial cells, activation of a GPCR-signalling cascade involving activation of PLC and subsequent PIP_2 hydrolysis was shown to relieve the inhibition of TRPV4 by PIP_2 (*183*). Interestingly, the PIP_2 metabolite IP_3 was also shown to sensitize TRPV4 for hypotonic stimuli via the ER-located IP_3 receptor (*211, 262*). Thus, the PIP_2-dependent regulation of TRPV4 can be expected to be tightly connected to the cellular state, which determines PIP_2 levels in the PM or cytosolic IP_3 concentrations.

Between the PBD and the ARD, TRPV4 exhibits a proline-rich region (PRR) which is unique within the TRPV family (Figure 18). The PRR was found to act as the binding site for the protein Pacsin (*76*). Pacsin exists in three isoforms (Pacsin1, 2, and 3) of which each has been assigned different roles in activity-dependent bulk endocytosis (ADBE), clathrin-mediated endocytosis (CME), endosomal and vesicle trafficking, actin cytoskeleton dynamics, neuronal development, and cell migration (*301*). Some of these pathways seem to overlap with those of TRPV4, e.g. cytoskeleton dynamics and neuronal development (*224, 234, 302*). The three Pacsin isoforms share a conserved structure with an N-terminal F-BAR (Fes/CIP4 homology Bin-Amphiphysin-Rvs) domain. The F-BAR domain is a dimer with a concave shape which confers the ability to bind curved membranes or induce membrane curvature (*303*). Structural studies of the Pacsin F-BAR domains have revealed variations in the F-BAR domain curvature between the isoforms which also reflect in the generation of membrane tubules with different degrees of curvature. The Pacsin F-BAR domains are connected to C-terminal Src homology 3 (SH3) domains by disordered linkers. The linkers of Pacsin isoforms 1 and 2 exhibit multiple Asp-Pro-Phe (NPF) motifs which are not present in Pacsin3. NPF motifs are commonly targeted by Eps15 homology-domain (EHD)-containing proteins involved in vesicle recycling (*304*). The SH3 domain mediates interactions of Pacsin proteins with proline-rich substrates, such as the TRPV4 proline-rich region with the sequence ^{138}PAPQPPP^{144}. Other common interaction partners recognized by the Pacsin SH3 domain are dynamin, synaptojanin, synapsin (*305*) and neural Wiskott-Aldrich Syndrome protein (N-WASP) (*306*), i.e. proteins involved in vesicle formation or remodelling of the actin cytoskeleton, respectively. The SH3 domain also seems to regulate the tubulation activity of the Pacsin F-BAR domains (Figure 18). For Pacsin1, a regulatory autoinhibition mechanism was proposed in which the SH3 domain associates with the F-BAR domain in a clamped

conformation, and thus prevents it from binding to the membrane (*307*). The interaction with a proline-rich substrate, e.g. dynamin, breaks the SH3/F-BAR complex and releases Pacsin1 from the clamped conformation, thus allowing for membrane binding and tubulation. The TRPV4 PRR is also a potential candidate to release the Pacsin autoinhibition. A less well-understood mechanism for regulating the tubulation activity of Pacsin is the phosphorylation of conserved serine and threonine residues in the F-BAR domain (*308–310*).

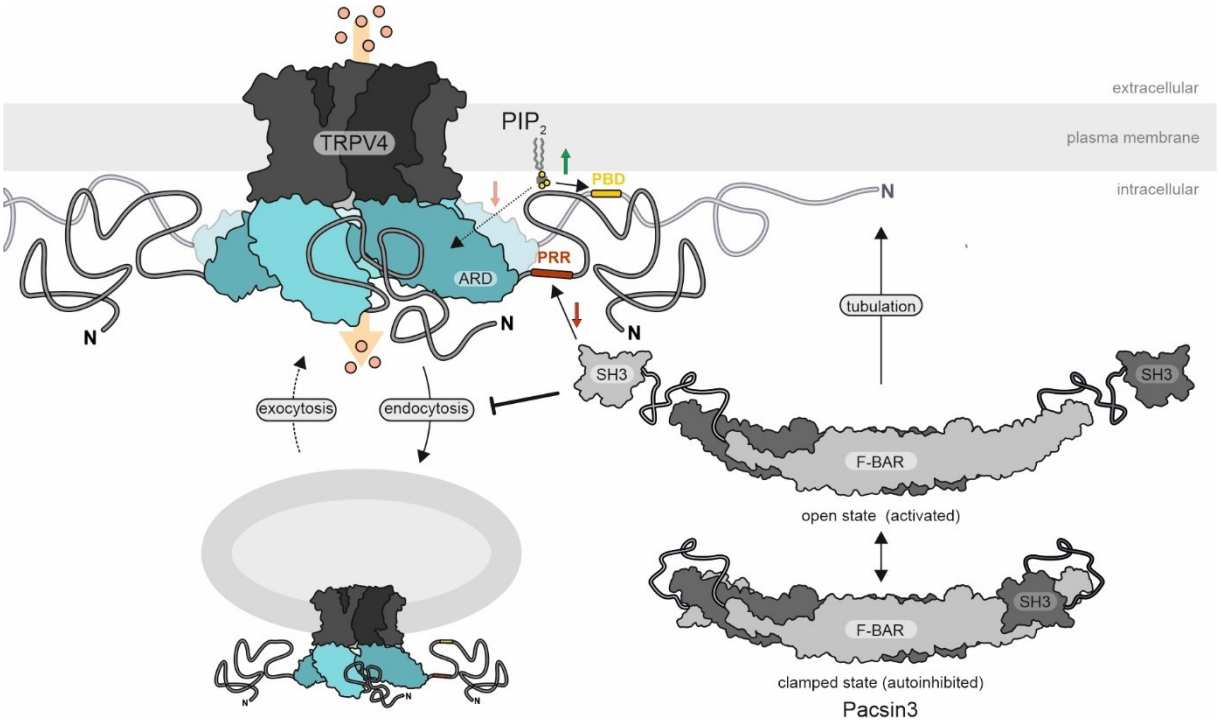


Figure 18: Regulation of TRPV4 by PIP_2 and Pacsin3. TRPV4 contains two binding sites for regulatory interaction partners in its putatively disordered N-terminus: a phosphoinositide binding domain (PBD, yellow) and a proline-rich region (PRR, red). The PBD acts as the binding site for the lipid second messenger phosphatidylinositol 4,5-bisphosphate (PIP_2). The interaction with PIP_2 is required for TRPV4 to be sensitive for heat and osmotic stimuli (*128*). PIP_2 binding to the PBD induces an expanded conformation of the TRPV4 cytosolic domains. The ARD was also identified to bind PIP_2, leading to an inhibition of TRPV4. However, the physiological relevance of the PIP_2 binding site in the ARD seems highly unlikely as it localizes to a region distant from the membrane (*150*). In close proximity to the PBD, the PRR serves as a binding site for the SH3 domain of Pacsin proteins (*76, 77*). The Pacsin SH3 domains are connected via a flexible linker to a dimeric F-BAR domain. F-BAR proteins are able to bind curved membranes or induce membrane curvature, thus leading to membrane tubulation (*303, 309, 311*). All three Pacsin isoforms were identified to bind TRPV4, but only Pacsin3 seems to regulate TRPV4. Pacsin3 increases the amount of TRPV4 in the plasma membrane by inhibition of its endocytosis, and it reduces the basal activity of TRPV4 and desensitizes it to heat and osmotic stimuli. Moreover, Pacsin3 binding induces a compact conformation of the TRPV4 cytosolic domains (*128*). Thus, Pacsin3 appears to antagonize the effects of PIP_2 on TRPV4 activity. Pacsin proteins are proposed to exists in a clamped state in the absence of a substrate. In this state, the SH3 domain binds to the F-BAR domain and inhibits the interaction with membranes. Binding of a substrate to the SH3 domain releases Pacsin to an open state and initiates membrane binding. TRPV4 is, thus, a potential candidate to release Pacsin isoforms from their autoinhibited state.

All Pacsin isoforms bind to the TRPV4 PRR with their C-terminal SH3 domain (*76*). The coexpression of Pacsin3 with TRPV4 in HEK293 cells increased the amount of TRPV4 channels presented to the plasma membrane compared to when a binding-incompetent Pacsin3 P415L mutant was coexpressed (*76, 77*). The enhanced surface expression of TRPV4, when coexpressed with Pacsin3, could be explained by Pacsin3 inhibiting the endocytosis of TRPV4 (Figure 18) (*76*). An effect on the subcellular localization was not observed when Pacsin1 and Pacsin2 were coexpressed (*76*). Strikingly, coexpression of Pacsin3 reduced the basal activity of TRPV4 and reduced the sensitivity of TRPV4 to heat and osmotic stimuli (*76, 77*). The Pacsin3-mediated desensitization of TRPV4 could be abolished by mutating residues P142 and P143 in the TRPV4-PRR to alanine and leucine (*77*). These observations suggest that Pacsin3 antagonizes the PIP_2-mediated sensitization of TRPV4. FRET studies revealed that a conformational change of the cytosolic domains to a more compact state

accompanies the binding of Pacsin3 to the PRR, similar to that observed upon mutating the PBD or depleting cellular PIP_2 levels (Figure 18) (*128*). Importantly, coexpression of TRPV4 with the isolated Pacsin3 SH3 domain did not reduce the sensitivity for hypotonic stimuli or heat (*128*). Whether Pacsin1 and Pacsin2 have a similar effect on the TRPV4 gating behaviour as Pacsin3 has not been tested yet.

The currently available information on how TRPV4 interacts with PIP_2 and Pacsin3 in the putative IDR are limited to mutagenesis and low-resolution FRET studies (*76, 77, 128*). These alone cannot explain how ligand binding in a disordered region of TRPV4 distant from the pore domain can modulate the ion gating behaviour. The recently determined TRPV4 cryo-EM structure together with computational and functional studies indicate that structural changes in the ARD are coupled to the pore domain through an allosteric pathway (*42, 61, 156*). Considering that the ARD connects the IDR to the rest of the channel, it is likely that an ARD/IDR communication is required to pass ligand-induced structural changes in the IDR to the pore. The lack of structural information on the region preceding the ARD, however, substantially hampers a greater understanding of the molecular details underlying the regulation of TRPV4 by PIP_2 and Pacsin3.

1.4 Scope of this thesis

TRPV4 is a physiologically important receptor for various stimuli and when deregulated causes a variety of diseases (*204*). It is therefore of high interest to understand its regulatory mechanisms in molecular detail. A yet structurally uncharacterized regulatory domain in TRPV4 is the distal N-terminus which provides the binding sites for the sensitizer PIP_2 and the desensitizer Pacsin3 (*76, 77, 128*). Due to the lack of structural data on the PIP_2 and Pacsin3 binding sites (*42*), it has remained unclear how both binding partners modulate the gating behaviour of TRPV4. The fact that TRPV N-termini is not accessible to the conventional methods for the structure determination of transmembrane proteins like cryo-EM or X-ray crystallography emphasizes the need for complementary structural techniques to investigate the TRPV4 N-terminus. Recent studies have demonstrated the power of integrative structural biology approaches to study disordered regions in membrane proteins (*205, 312*). They exploit the ability to express disordered regions isolated from the rest of the full-length proteins. In isolation, IDPs become amenable to a plethora of biophysical techniques which are extremely complicated or impossible when using membrane proteins (*83, 313–315*). The resulting information from isolated IDPs can then be pieced together with existing high-resolution structures from the rest of the full-length proteins and integrated with functional data. The central goal of this thesis was to elucidate the structure and dynamics of the TRPV4 N-terminal domain and how it interacts with Pacsin3, PIP_2, and other lipids via an integrative, divide and conquer structural biology approach (Figure 19). A prerequisite for structural studies was to establish purification protocols yielding high amounts of the isolated TRPV4 N-terminal domain (NTD) and several truncation constructs as soluble proteins. The NTD is composed of the ARD and the putatively intrinsically disordered region (IDR). The ARD has been characterized previously via X-ray crystallography and in cryo-EM studies but information on its dynamics and how they depend on the presence of the IDR are unavailable (*19, 21, 42, 150, 151*). A major focus was therefore on investigating the structural dynamics of the isolated ARD and IDR and compare them to the features observed in the NTD to deduce on a potential ARD/IDR communication. The primary tools to for investigating the TRPV4 NTD and its isolated domains were small-angle X-ray scattering (SAXS), nuclear magnetic resonance (NMR) spectroscopy, chemical crosslinking (XL-MS) and hydrogen/deuterium mass spectrometry (HDX-MS), and tryptophan fluorescence spectroscopy (Figure 19).

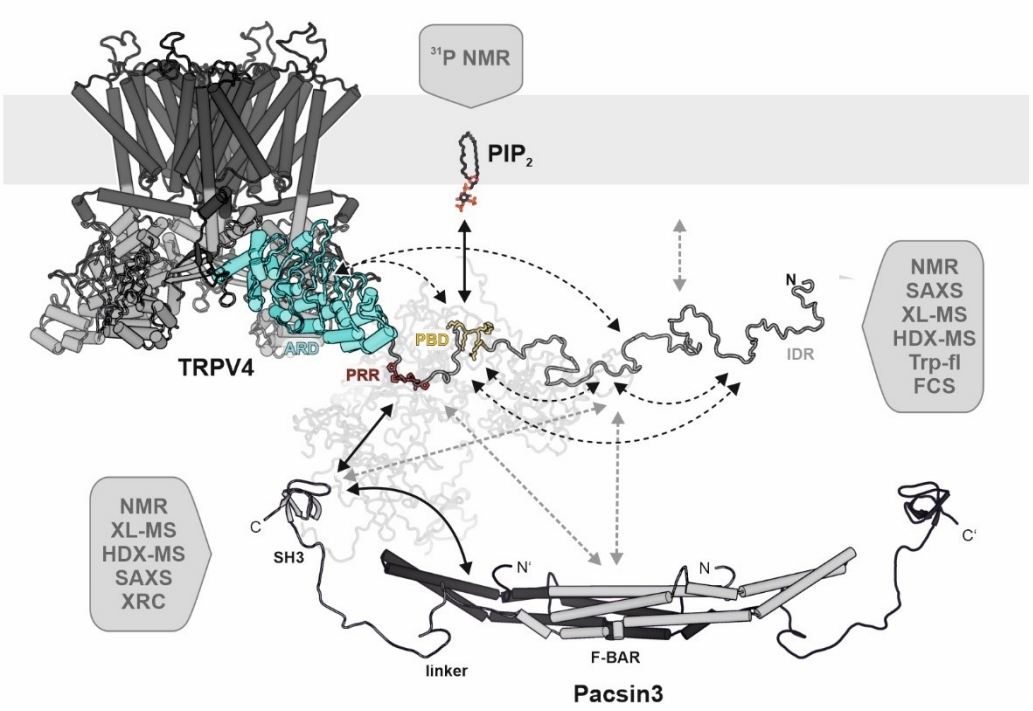


Figure 19: Overview of the integrative structural biology approach used to investigate the structural dynamics in the TRPV4 N-terminus and its interaction withPacsin3 and PIP_2. To elucidate a potential ARD/IDR communication and intrachain interactions in the IDR (black, dotted arrows), the structure and dynamics of the TRPV4 N-terminal domain (NTD, i.e. ARD & IDR) were investigated in a divide and conquer approach using small-angle X-ray scattering (SAXS), nuclear magnetic resonance (NMR) spectroscopy, chemical cross-linking mass spectrometry (XL-MS), hydrogen/deuterium mass spectrometry (HDX-MS), or tryptophan fluorescence spectroscopy (Trp-fl). Investigations on the lipid interaction of the TRPV4 NTD in and potentially beyond the PBD were based on NMR, Trp-fl, and FCS. These studies were directly related to the question of how the Pacsin3 SH3 domain binds to the PRR, and whether it cross-talks with a PBD/PIP_2 interaction. The major tool for this aspect of this thesis was NMR spectroscopy. Experiments on the isolated Pacsin3 SH3 domain were complemented by XL-MS to investigate whether the Pacsin3 interacts with regions beyond the PRR (grey, dashed arrows). The combination of NMR, XL-MS, HDX-MS, and X-ray crystallography (XRC) was used to obtain a structural model of full-length Pacsin3 and its conformational change upon binding to the TRPV4 NTD.

The functional implications of a regulatory ligand on its binding target are encoded in the structural details of both interaction partners. In this sense, the mechanism governing the regulation of TRPV4 by PIP_2 and Pacsin3 is presumably determined by local and global conformational changes in TRPV4 and in Pacsin3 or the plasma membrane (PIP_2). As a soluble protein, the TRPV4 NTD and its deletion mutants could be used to probe an interaction of the PBD and the remaining NTD regions with lipids such as PIP_2 using conventional liposome binding assays, tryptophan fluorescence spectroscopy, or fluorescence correlation spectroscopy (FCS). Residue-specific details of the interactions were accessible via NMR spectroscopy using the IDR or smaller peptide fragments of the TRPV4 NTD. Details on which phosphate moieties in PIP_2 contribute to the PBD and whether this is influence by TRPV4 NTD regions beyond the PBD could be inferred from ^{31}P NMR spectroscopy (Figure 19). The complementarity of the used methods enabled to investigate the interaction between PIP_2 and TRPV4 from the perspective of both interaction partners.

For investigating how Pacsin3 binds to the TRPV4 PRR and potentially cross-talks with the PBD/PIP_2 interaction, the divide and conquer approach was extended by partitioning Pacsin3 in its individual domains. Previous studies on the Pacsin1/dynamin interaction have demonstrated the feasibility to use Pacsin SH3 domains for NMR experiments (*316*). Therefore, the isolated SH3 domain was chosen as a reporter for chemical shift perturbation (CSP) experiments using different constructs of the TRPV4 NTD to identify the exact Pacsin3 binding site in the TRPV4 NTD. A comparison of the Pacsin3 SH3 domain with the Pacsin1 and Pacsin2 SH3 domains and how they interact with the TRPV4 N-terminus was carried out to understand whether they bind TRPV4 similarly and, thus, potentially have shared regulatory functions. The identified Pacsin3 SH3 binding site was then used for solution NMR structure determination to obtain atomistic insights into the binding mode

of Pacsin3 with TRPV4. Importantly, besides several X-ray crystal structures of F-BAR domains, SH3 domains, and an F-BAR/SH3 complex, there is currently no real full-length structure of a Pacsin isoform available (*303, 307, 311, 317*). Structure determination of Pacsin proteins seems to be substantially complicated by the structural heterogeneity of the disordered linker and the putatively dynamic association between SH3 and F-BAR domain (*307*). The exact domain organization is, however, an important piece of information to understand how Pacsin3 interacts with TRPV4. Divide and conquer-based integrative structural biology approaches have previously proven very useful to study the structure of BAR domain proteins or other multidomain proteins (*318–321*). A similar approach was therefore applied to investigate the structure and conformational dynamics of full-length Pacsin3 and how it changes upon interaction with TRPV4. NMR CSP experiments, SAXS, X-ray crystallography (XRC), XL-MS, HDX-MS, were at the heart of these investigations (Figure 19).

In an unrelated side project, a dynamically-perturbed spider silk protein was investigated to understand its dimerization mechanism. A combination of NMR spectroscopy with Trp-fl, SEC-MALS, and photo-electron transfer (PET)-FCS was used to elucidate structural and dynamic effects of mutations in the hydrophobic core of the protein on its ability to dimerize. The results are described in the Appendix (see section VIII).

2 Material & Methods

2.1 Text editing, graph plotting, and figure design

This thesis was written with Microsoft Word (Office 365). References were managed with Citavi 6 (Swiss Academic Software). All figures were generated with the CorelDRAW 2020 Graphics Suite (Corel Corporation). Graphs were plotted in Origin Pro 8 (OriginLab) and imported into CorelDRAW for further editing. Structures of biological macromolecules and 3D structures of lipids or other small molecules were visualized and rendered with PyMOL (*322, 323*). Coordinate files of biological macromolecules were obtained from the Protein data bank (PDB). Small-molecule structures were drawn with ChemDraw Ultra 8.0 (PerkinElmer).

2.2 Laboratory equipment

All laboratory equipment used in this thesis for standard procedures such as centrifugation, pH adjustment or SDS-PAGE is listed in Table 2. Specific equipment is mentioned in the respective sections.

Table 1: Standard laboratory equipment used in this thesis.

Name	Description	Manufacturer
Avanti J-26XP	Centrifuge used for cell harvesting and clearing of cell lysates	Beckman Coulter
Optima Max	Tabletop ultra-centrifuge	Beckman Coulter
JLA 16.250	Fixed-angle rotor for Avanti J-26XP centrigue fitting 250 mL volumes	Beckmann Coulter
JLA 8.1000	Fixed-angle rotor for Avanti J-26XP centrigue fitting 1 L volumes	Beckmann Coulter
JA 25.50	Fixed-angle rotor for Avanti J-26XP centrigue fitting 50 mL volumes	Beckmann Coulter
MLA-130	Fixed-angle rotor ultra-centrifuge rotor	Beckmann Coulter
TLA-100	Fixed-angle rotor ultra-centrifuge rotor	Beckmann Coulter
Centrifuge 5415 R	Tabletop centrifuge for 1-2 mL volumes	Eppendorf
Centrifuge 5810 R	Tabletop centrifuge for 10-50 ml volumes	Epspendorf
Concentrator plus	Vacuum Centrifuge	Eppendorf
pH electrode LE409	pH-meter	Mettler Toledo
Benchtop pH Meter Five Easy	pH-meter	Mettler Toledo
PeqSTAR 2x Gradient	PCR thermocycler	Peqlab
Mixing Block MB-102	Thermoblock for 1-2 mL tubes	Bioer
PowerPacBasic	Power supply/Gel electrophoresis	BioRad
Mini-PROTEAN Tetra cell	SDS-PAGE chamber and equipment	BioRad
NanoDrop 2000c	UV-Vis spectrometer for the determination of protein and DNA concentrations	Thermo Scientific
ViewPix 700	Scanner for imaging of SDS-PAGE gels	BioStep

2.3 Chemicals

Unless further specified in the respective sections, all chemicals were purchased from Roth (Karlsruhe), Sigma-Aldrich (Munich), New England Biolabs (Frankfurt), Avanti Polar Lipids (Alabama), and VWR (Darmstadt).

2.4 Consumables

All consumbles which are not further specified in the respective sections are listed in Table 2.

Table 2: Consumables used for this thesis.

Name	Description	Manufacturer
Vivaspin-6 & Vivaspin-20	Centrifugal concentration of proteins with MWCO of 2, 3.5, 5, 10, and 30 kDa	Sartorius
Amicon Ultra-6	Centrifugal concentration of proteins with MWCO 10 kDa	Merck Millipore
Spectra/Por Float-A-Lyzer G2	Peptide dialysis with MWCO 500 Da	Roth
ZelluTrans	Protein dialysis with MWCO 3.5 and 12 kDa	Roth
E.Z.N.A Plasmid Mini Kit	Isolation of plasmid DNA from *E. coli*	Omega Bio-Tek
GeneJet PCR Purification	Purification of PCR end products	Thermo Scientific
KAPA HIFI HotStart PCR Kit	PCR ênzymes and buffers	Roche
DpnI	Digestion of template DNA	New England BioLabs
Precision Plus Protein Dual Color Standard	Molecular marker for SDS-PAGE analysis of proteins	BioRad
1 kb DNA ladder	Molecular marker for agarose gel electrophoresis of linear DNA fragments	New England BioLabs

2.5 Cloning, expression, and purification of TRPV4 constructs

Cloning of TRPV4 N-terminal constructs

The DNA sequences encoding for the TRPV4 N-terminal domain from *H. sapiens* (human) and *G. gallus* (chicken), respectively, were cloned from cDNA into a pET11a vector with an N-terminal His_6SUMO-tag via Gibson Assembly (*324*). SUMO (small ubiquitin-related modifier) is a ubiquitin-like protein (UBL) commonly used as an expression tag to render proteins more soluble and as protection against proteolytic degradation (*325*). SUMO is best positioned in the N-terminus of a protein, as it can be cleaved after its C-terminal reside using the protease Ulp-1 (UBL-specific protease 1). Therefore, after SUMO-tag cleavage, proteins do not contain residual artificial amino acids. The His_6-tag N-terminally to the SUMO-tag enables protein purification via Ni-NTA immobilized ion chromatography (IMAC). The combination of His_6-tag and SUMO-tag in the N-terminus proved highly beneficial for the expression and purification of TRPV4 N-terminal constructs from *E. coli*. The TRPV4 N-terminal domain (NTD) contains a large intrinsically disordered region which is targeted for proteolytic degradation during expression in *E. coli*, thus yielding multiple degradation products with similar sizes as the intact tag-less NTD. Having an N-terminal His_6SUMO-tag in the TRPV4 NTD enables the separation of degradation products in a four-step purification procedure: 1) Ni-NTA IMAC, 2) Ulp-1 digest, 3) Ni-NTA IMA, and 4) SEC (Figure 20).

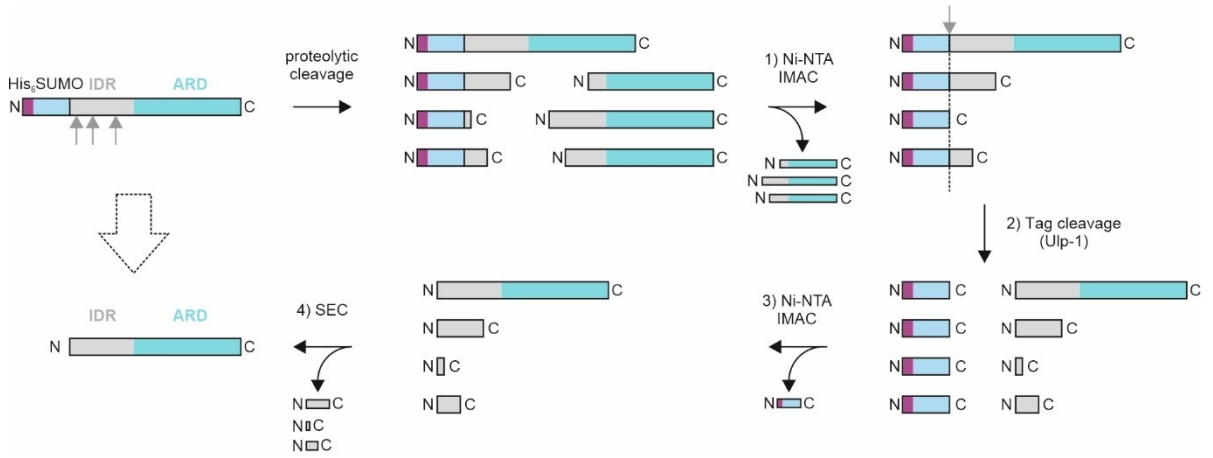


Figure 20: Purification procedure of N-terminally His_6SUMO-tagged TRPV4 N-terminal constructs. During expression in *E. coli*, TRPV4 N-terminal constructs are prone to proteolytic degradation at unspecific sites (grey arrows) in the intrinsically disordered regions (IDR). In the case of the TRPV4 N-terminal domain (NTD) which is composed of the IDR and the ankyrin repeat domain (ARD), proteolytic cleavage in the IDR yields His_6SUMO-tagged IDR fragments of different sizes, and N-terminally truncated IDR-ARD tandem constructs. In a first Ni-NTA IMAC step, the intact His_6SUMO-NTD and His_6SUMO-tagged IDR fragments are separated from cleaved fragments IDR-ARD tandem constructs, which do not have a His_6SUMO-tag. In a subsequent step, the His_6SUMO-tag is cleaved off (grey arrows) in all remaining proteins using the protease Ulp-1, thus yielding a mixture of isolated His_6SUMO-tag, the intact NTD, and IDR fragments of different sizes. A second Ni-NTA IMAC step enables the separation of the isolated His_6SUMO-tag. Due to a sufficient size difference between the intact NTD and the degraded IDR fragments, the NTD can be isolated using a SEC purification step. His_6-tag: purple; SUMO-tag: light blue; IDR: grey; ARD: cyan.

Expression plasmids encoding for the isolated intrinsically disordered region (IDR) and the isolated Ankyrin repeat domain (ARD) were obtained from the NTD encoding vectors using a Gibson Deletion protocol (*326*). Based on the expression plasmid of the *G. gallus* TRPV4 NTD, a set of N-terminal truncations, termed NTD ΔN54, NTD ΔN97, NTD ΔN104, and NTD ΔN120, were generated, again using the Gibson Deletion protocol. Here, NTD Δ*n* denotes the deletion of *n* N-terminal residues in the TRPV4 NTD. The Gibson Assembly protocol was further used to clone several TRPV4 N-terminal peptides, termed PH-PRR (res. 97-134), PBD-PRR (res. 105-134), PRR (121-134), and AH1-PRR (121-148). Mutation of the PIP_2 binding site (^{107}KRWRR^{111} to ^{107}AAWAA^{111}, forward primer: GTGAAAACGCAGCCTGGGCCGCGCGTGTGGTTGAAAAACCAGTGG, reverse primer: CACACGCGCGGCCCAGGCTGCGTTTTCACCACCAATCTGTTCACG) in the *G. gallus* TRPV4 NTD, IDR, PH-PRR, and PBD-PRR constructs and mutation of the proline-rich region in the PBD-PRR construct (P128A, forward primer: CCGAACGCGCCGCCAGTGCTGAAAGTG, reverse primer: ACTGGCGGCGCGTTCGGCGCCGGAC, P128A/P129L, forward primer: GCCGAACGCGCTGCCAGTGCTGAAAGTG, reverse primer: CACTTTCAGCACTGGCAGCGCGTTCGG) were obtained by site-directed mutagenesis using the quick change polymerase chain reaction (PCR) protocol (*327*).

Expression, and purification of TRPV4 N-terminal domains

The TRPV4 NTD, ARD, IDR and NTD Δ*n* deletion constructs were expressed in *E. coli* BL21-Gold(DE3) (Agilent Technologies) grown in terrific broth (TB) medium (Roth) supplemented with 0.04% (w/v) glucose and 1 mg/mL Ampicillin. Cells were grown at 37 °C to an OD_{600} of 0.4, moved to RT, grown to OD_{600} of 0.8 for induction (0.15 mM IPTG) and then grown overnight at 20 °C (typically 16 hrs). After harvest, cells were stored at -80 °C until further use. All purification steps were carried out at 4°C. Harvested cells were dissolved in lysis buffer (20 mM Tris pH 8, 20 mM imidazole, 300 mM NaCl, 0.1% (v/v) Tx100, 1 mM DTT, 1 mM benzamidine, 1 mM PMSF, lysozyme, DNAse, RNAse and protease inhibitor (Sigmafast)) and sonicated (Branson Sonifier 250) or homogenized under high pressure (Maximator HPL6). Debris was removed by centrifugation and the supernatant applied to a Ni-NTA gravity flow column (Qiagen). After washing (20 mM Tris pH 8, 20 mM imidazole, 300 mM NaCl), the protein was eluted with 500 mM imidazole. Protein containing

fractions were dialyzed overnight (20 mM Tris pH 7, 300 mM NaCl, 10% glycerol, 1 mM DTT, 0.5 mM PMSF) in the presence of Ulp-1 protease in a molar ratio of 20:1 protein vs protease (see below for Ulp-1 purification protocol). Processed proteins then only comprised the native residues of the TRPV4 N-terminus. After dialysis, cleaved proteins were separated from the His_6SUMO-tag and uncleaved proteins with a reverse Ni-NTA affinity chromatography step and subsequently purified via size exclusion chromatography (SEC) using a HiLoad prep grade 16/60 Superdex200 column (GE Healthcare) and 20 mM Tris pH 7, 300 mM NaCl, 1 mM DTT as buffer. Pure and tag-free TRPV4 N-terminal constructs. Proteins were afterwards concentrated with Vivaspin centrifugal concentrators (Merck), flash-frozen in liquid nitrogen and stored at -20 °C until further use.

Expression, and purification of TRPV4 N-terminal peptides

A slightly modified protocol was used for the expression and purification of the TRPV4 N-terminal peptides PH-PRR, PBD-PRR, PRR, and AH1-PRR. The TRPV4 peptides were expressed *E. coli* BL21-Gold(DE3) (Agilent Technologies) grown in LB medium and 1 mg/mL. Cells were grown to an OD_{600} of 0.8 for induction (0.5 mM IPTG) and then further grown at 37 °C for 3 hrs. Harvested cells were dissolved in lysis buffer (20 mM Tris pH 8, 20 mM imidazole, 300 mM NaCl, 0.1% (v/v) Tx100, 1 mM DTT, 1 mM benzamidine, 1 mM PMSF, lysozyme, DNAse, RNAse and protease inhibitor (Sigmafast)) and sonicated (Branson Sonifier 250) or homogenized under high pressure (Maximator HPL6). Debris was removed by centrifugation and the supernatant applied to a Ni-NTA gravity flow column (Qiagen). After washing (20 mM Tris pH 8, 20 mM imidazole, 300 mM NaCl), the protein was eluted with 500 mM imidazole and purified via SEC using a HiLoad prep grade 16/60 Superdex75 column (GE Healthcare) and 20 mM Tris pH 8, 300 mM NaCl, 1 mM DTT as buffer. The purified His_6SUMO-peptides were dialyzed overnight (20 mM Tris pH 8, 300 mM NaCl, 10% glycerol, 1 mM DTT, 0.5 mM PMSF) in the presence of Ulp-1 protease in a molar ratio of 20:1 protein vs protease. The cleaved peptides were subsequently separated from the His_6SUMO-tag and uncleaved peptides with a reverse Ni-NTA affinity chromatography step. The pure, tag-less peptides were extensively dialyzed against autoclaved bidest. H_2O, lyophilized, and stored in solid form at -20 °C. Peptides could be dissolved in desired amounts of buffer to concentrations up to 10 mM.

2.6 Cloning, expression, and purification of Pacsin3 constructs

Cloning of Pacsin constructs

The gene fragments encoding the *G. gallus* Pacsin1, 2 and 3 SH3 domains (termed Pac1/2/3 SH3) and the full-length Pacsin3 (termed fl-Pac3) DNA sequence were obtained commercially (Genescript). The Pacsin1-3 SH3 domain gene fragments were cloned into a pET11a vector with a C-terminal His_6 tag. The full-length Pacsin3 gene fragment was cloned into a pET11a vector with an N-terminally encoded His_6SUMO-tag. The isolated Pacsin3 F-BAR domain (termed Pac3 F-BAR) and the Pacsin3 SH3 domain including the N-terminal linker (termed Pac3 l-SH3) were obtained from the full-length Pacsin3 encoding plasmid using the Gibson Deletion protocol (*326*). The triple point mutation Q390R/E391R/E394R in the Pacsin3 SH3 domain was generated via site-directed mutagenesis (forward primer GCCGTCGTGCGGATCGTCTGAGCTTCAAGGCGGGCG AGG, reverse primer ACGATCCGCACGACGGCCGGTATAGTCGTACAGCGCAC) in a single PCR reaction using the quick-change protocol (*327*).

Expression and purification of Pacsin SH3 domains

Constructs were expressed in *E. coli* BL21-Gold(DE3) (Agilent Technologies) in LB medium with 1 mg/mL ampicillin at 37 °C. Cells were grown at 37 °C to an OD_{600} of 0.4, moved to RT, grown to OD_{600} of 0.8 for induction (75 µM IPTG) and then grown overnight at 20 °C (typically 16 hrs). After harvest, cells were stored at -80 °C until further use. For purification, cells were dissolved in lysis buffer (20 mM Tris pH s8, 20 mM imidazole, 300 mM NaCl, 0.1% (v/v) Tx100, 1 mM DTT, 1 mM benzamidine, 1 mM PMSF, lysozyme,

DNAse, RNAse and protease inhibitor (Sigmafast)) and sonicated (Branson sonifier 250). Debris was removed by centrifugation and the supernatant applied to a Ni-NTA gravity flow column (Qiagen). After washing (20 mM Tris pH 8, 20 mM imidazole, 300 mM NaCl), the protein was eluted with 250 mM imidazole. Protein containing fractions were concentrated and run on a size exclusion column (HiLoad prep grade 16/60 Superdex75, GE Healthcare) with 10 mM Tris pH 7, 100 mM NaCl. All steps were carried out at 4 °C. After SEC, samples were concentrated to required concentrations, flash-frozen in liquid nitrogen and stored at -80 °C until further use.

Expression and purification of full-length Pacsin3, F-BAR domain, and Linker-SH3 domain

The fl-Pac3 and Pac3 l-SH3 constructs were expressed in *E. coli* Rosetta gami cells and Pac3 F-BAR was expressed in BL21-Gold(DE3) (both Agilent Technologies) grown in Luria broth (LB) medium (Roth) supplemented with 1 mg/mL ampicillin. Cells were grown at 37 °C to an OD_{600} of 0.4, moved to 18 °C, grown to OD_{600} of 0.8-1.0 for induction (1.0 mM IPTG) and then grown overnight at 18 °C (typically 16 hrs). After harvest, cells were stored at -80 °C until further use. All purification steps were carried out at 4°C. Harvested cells were dissolved in lysis buffer (20 mM Tris pH 8, 20 mM imidazole, 300 mM NaCl, 0.1% (v/v) Tx100, 1 mM DTT, 1 mM benzamidine, 1 mM PMSF, lysozyme, DNAse, RNAse and protease inhibitor (Sigmafast)) and homogenized under high pressure (Maximator HPL6). Debris was removed by centrifugation and the supernatant applied to a Ni-NTA gravity flow column (Qiagen). After washing (20 mM Tris pH 8, 20 mM imidazole, 300 mM NaCl), the protein was eluted with 500 mM imidazole. Protein containing fractions were dialyzed overnight (20 mM Tris pH 8, 300 mM NaCl, 1 mM DTT) in the presence of Ulp-1 protease in a molar ratio of 20:1 protein vs protease. Processed proteins then only comprised the native Pacsin3 residues. After dialysis, cleaved proteins were separated from the His_6SUMO-tag and uncleaved proteins with a reverse Ni-NTA affinity chromatography step and subsequently purified via size exclusion chromatography (SEC) using a HiLoad prep grade 16/60 Superdex200 column (GE Healthcare) and 20 mM Tris pH 7, 300 mM NaCl, 1 mM DTT as buffer. Pure and tag-free TRPV4 N-terminal constructs. Proteins were afterwards concentrated with Vivaspin centrifugal concentrators (Merck), flash-frozen in liquid nitrogen and stored at -20 °C until further use.

2.7 Expression and purification of Ulp-1 protease

The Ulp-1 expression plasmid was a generous gift from the group of Prof. Dr. Jens Wöhnert (Goethe University, Frankfurt). For expression, *E. coli* BL21-Gold(DE3) cells (Agilent Technologies) were grown in LB medium containing 0.1 mg/mL kanamycin at 37 °C to an OD_{600} of 0.8-0.9. Protein expression was induced with IPTG at a final concentration of 1 mM and cells were grown for further 4 hrs at 37 °C, before being harvested by centrifugation. For purification (all steps at 4 °C), the frozen cell pellet was suspended in lysis buffer (20 mM Tris pH 8, 20 mM imidazole, 300 mM NaCl, 0.1% (v/v) Tx100, 1 mM DTT, lysozyme, DNAse, RNAse). No protease inhibitors were added to not compromise Ulp-1 activity. The cells were lysed under cooling on ice via sonication (Branson Sonifier 250) and cleared via centrifugation. The cell lysate was loaded onto a NiNTA gravity flow column (Qiagen). After washing (20 mM Tris pH 8, 20 mM imidazole, 300 mM NaCl), the protein was eluted with 100/150/200/250 mM imidazole. Fractions containing Ulp-1 were pooled and dialyzed against 50% glycerol, 75 mM Tris pH 8, 0.5 mM DTT, 1 mM EDTA overnight at 4°C. After dialysis, Ulp-1 samples were frozen in liquid nitrogen without further purification and stored at -80 °C.

2.8 Analytical size-exclusion chromatography

In size-exclusion chromatography (SEC), molecules are separated according to their hydrodynamic radius, also termed the Stokes radius (R_S) for proteins. The components to be separated are partitioned between a mobile

phase and a stationary phase. The stationary phase commonly is a packed column with a porous material with defined pore sizes. The mobile phase contains the sample and percolates through the stationary phase. Proteins with R_S smaller than the pore size of the stationary phase will fit inside the pores where they are retained for a short time. Proteins with R_S larger than the pore sizes will not diffuse into the pores and follow the solvent of the mobile phase. The different retention times between small and large proteins give rise to the so-called molecular sieving effect. Consequently, large proteins pass a stationary phase earlier than small molecules. The retention time, also expressed as the elution volume is directly related to the hydrodynamic properties of a protein. Thus, the elution volume obtained from a SEC experiment can be used to determine the Stokes radius R_S of a protein which itself is a versatile parameter to evaluate a protein's conformational state (*328*).

Analytical size exclusion chromatography (SEC) experiments were carried out using an NGC Quest (BioRad) chromatography system equipped with UV-Vis detection. A volume of 250 µL of protein at a concentration of 2-3 mg mL^{-1} was loaded onto a 1 mL loop and injected with 3 CV into the chromatography system. Depending on the protein, a GE Superdex75 or Superdex200 16/60 increase column was used for protein separation (see Table 3). Protein was detected at 230 and 280 nm. The flow rate was 0.5 mL min^{-1}. All analytical SEC runs were carried out at 4 °C.

Table 3: Experimental conditions of analytical size exclusion experiments. All experiments were carried out at 4 °C with a sample volume of 250 µL and a protein concentration of 2-3 mg mL^{-1}.

Protein	Column	Buffer
NTD, NTD^{AAWAA}, ARD, IDR, IDR^{AAWAA}, NTD ΔN54/ ΔN97/ΔN104/ΔN120, PH-PRR, PH^{AAWAA}-PRR	Superdex200 16/60 increase	20 mM Tris pH 7, 300 mM NaCl, 1 mM
IDR, PH-PRR, PBD-PRR, AH1-PRR	Superdex75 16/60 increase	10 mM Tris pH 7, 100 mM NaCl, 1 mM
Pac1 SH3, Pac2 SH3, Pac3 SH3	Superdex75 16/60 increase	20 mM Tris pH 7, 300 mM NaCl, 1 mM
Fl-Pac3, Pac3 F-BAR, Pac3 l-SH3, Pac3 SH3	Superdex200 16/60 increase	20 mM Tris pH 7, 300 mM NaCl, 1 mM

For Stokes radius (R_s) determination, SEC columns were calibrated with a protein standard kit (GE Healthcare) containing Ferritin (M_W = 440 kDa, R_s = 61.0 nm), Alcohol dehydrogenase (150 kDa, 45.0 nm), Conalbumin (75 kDa, 36.4 kDa), Ovalbumin (43 kDa, 30.5 nm), Carbonic anhydrase (29 kDa, 23.0 nm), Ribonuclease A (13.7 kDa, 16.4 nm), and Aprotinin (6.5 kDa, 13.5 nm). Stokes radii were obtained from ref. (*329*). The SEC elution volume, V_e (in mL), of the protein standards was plotted versus the log(R_s), with R_s in nm, and fitted with a linear regression (Equation 1).

$$V_e = m \cdot \log(R_S) + b \qquad 1$$

Here, m is the slope of the linear regression and b is the y-axis section. Equation 1 was then used to calculate the Stokes radius of the proteins in Table 3 from their respective SEC elution volumes.

R_S-M_W scaling laws

The Stokes radius R_S of a protein describes the radius of a sphere with the molecular weight of the respective protein. It is determined by the proteins molecular mass and by its shape. Large and expanded proteins have larger radii than small and compact proteins. At equal molecular weight, an elongated protein has a larger Stokes radius than a spherical protein. Moreover, the Stokes radius of a protein increases when it transitions from a folded, globular protein to an unfolded and thus disordered protein. Thus, the Stokes radius of a protein can provide valuable information on its conformational state, given that the molecular weight and the oligomeric

state of the protein is known. The analysis of the hydrodynamic properties of proteins in distinct conformational states has given rise to a set of empirical scaling relationships between R_S and M_W (equations 2-8) (*328*).

$$\log\left(R_s^{N}\right) = -(0.204 \pm 0.023) + (0.357 \pm 0.005) \cdot \log(M_W) \quad 2$$

$$\log\left(R_s^{MG}\right) = -(0.053 \pm 0.094) + (0.334 \pm 0.021) \cdot \log(M_W) \quad 3$$

$$\log\left(R_s^{PMG}\right) = -(0.210 \pm 0.180) + (0.392 \pm 0.041) \cdot \log(M_W) \quad 4$$

$$\log\left(R_s^{U(urea)}\right) = -(0.649 \pm 0.016) + (0.521 \pm 0.004) \cdot \log(M_W) \quad 5$$

$$\log\left(R_s^{U(GdmCl)}\right) = -(0.723 \pm 0.033) + (0.521 \pm 0.007) \cdot \log(M_W) \quad 6$$

$$\log\left(R_s^{NU(coil)}\right) = -(0.551 \pm 0.032) + (0.493 \pm 0.008) \cdot \log(M_W) \quad 7$$

$$\log\left(R_s^{NU(PMG)}\right) = -(0.239 \pm 0.055) + (0.403 \pm 0.012) \cdot \log(M_W) \quad 8$$

Here, N describes a native conformation, i.e. a globular almost spherical protein. MG and PMG correspond to the molten globule and pre-molten globule state. Both describe intermediate states between a folded and an unfolded protein. U(urea) and U(GdmCl) correspond to chemically denatured proteins. NU(coil) and NU(PMG) describe proteins with a random coil or pre-molten globule like behaviour under the physiological conditions. Notably, the relative errors of the scaling laws do not generally exceed 10%. Therefore, the Stokes radius of a protein in a variety of conformations can be predicted based on the chain length with an accuracy of 90% with the scaling laws (*328*).

2.9 Multi-angle light scattering

Multi-angle light scattering (MALS) is a versatile method to determine the molecular weight of a protein. During a MALS experiment, the light intensity scattered from a sample is detected as a function of the scattering angle relative to the incident beam. In parallel, a differential refractive index detector (dRI) determines the analyte concentration based on the change in the refractive index of the solution caused by the presence of the analyte. MALS is usually used in combination with SEC, termed SEC-MALS, to ensure the measurement of monodisperse particles (*330*). Using equation 9, the molecular weight, M_W, can be determined from the reduced Rayleigh ratio extrapolated to zero, $R(0)$, which is the light intensity scattered from the analyte relative to the intensity of the incident beam.

$$M_W = \frac{R(0)}{K\, c\, \left(\frac{dn}{dc}\right)^2} \quad 9$$

Here, c is the concentration of the analyte and (dn/dc) is the refractive index increment which is commonly set 0.185 mL g^{-1} (*331*). K is an optical constant which depends on the used wavelength and the solvent refractive index.

Multi-angle light scattering coupled with size exclusion chromatography (SEC-MALS) of the *G. gallus* TRPV4 NTD, ARD, and IDR were performed and analyzed by Dr. Rupert Abele (Goethe University, Frankfurt). SEC-MALS experiments were performed with a GE Superdex200 16/60 increase column run at 0.5 ml min^{-1} on a Jasco HPLC unit (Jasco Labor und Datentechnik) connected to a light scattering detector measuring at three angles (miniDAWN TREOS, Wyatt Technology). The column was equilibrated for at least 16 hrs with 20 mM

Tris pH 7, 300 mM NaCl, 1 mM DTT. The buffer was filtered through 0.1 μm pore size VVLP filters (Millipore) beforehand. For protein molecular weight determination, 200 μL of protein samples at a concentration of 2 mg mL^{-1} were separated on the column. The ASTRA software package (Wyatt Technology) was used for data analysis, assuming a Zimm model (*332*). The refractive index increment *dn/dc* value of the protein was defined as 0.185 mL g^{-1}, which is the standard value for proteins (*331*). The protein extinction coefficients at 280 nm were calculated from the respective amino acid sequences using the ProtParam tool (*333*).

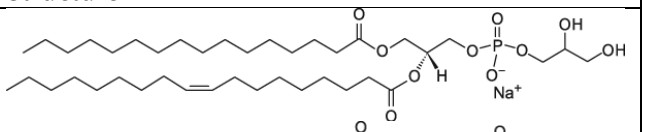
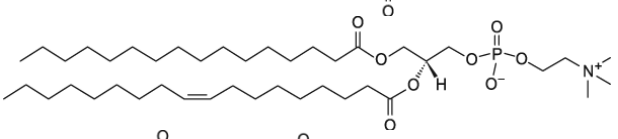
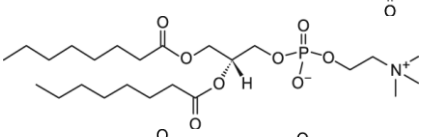
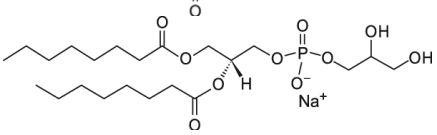
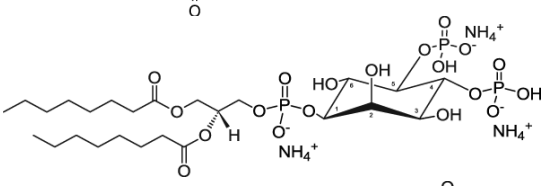
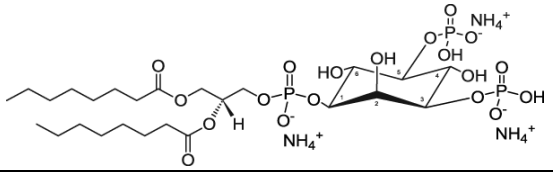
2.10 Lipid and liposome preparation

All lipids were purchased from Avanti Polar Lipids™ and Caymen Chemicals™ and stored at -20°C.

Liposomes were prepared from the two lipids 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol (POPG) (Table 4). POPG has a negatively charged headgroup. POPC is a zwitterion with a net charge of zero. Pure POPC or POPG/POPC mixed in a 1:1 (n/n) ratio were prepared in chloroform at the required amount. The organic solvent was then removed via nitrogen flux and under vacuum via desiccation overnight. The lipid cake was suspended in 1 mL buffer (10 mM Tris pH 7, 100 mM NaCl) to get 4 mM lipid and incubated for 20 min at 37°C and 1000 rpm. The lipid suspension was then subjected to five freeze and thaw cycles where the lipid sample was flash-frozen in liquid nitrogen and then thawed at 47.5 °C and 1000 rpm for 4 minutes. The resulting large unilamellar vesicles (LUVs) were then incubated for 20 min at 21 °C and 1000 rpm. To obtain a homogeneous solution of small unilamellar vesicles (SUVs), the mixture was extruded 15 times through a 100 nm membrane using the Mini Extruder (Avanti Polar Lipids). This yielded a liposome stock solution of 100 nm liposomes with 4.0 mM lipid in 10 mM Tris pH 7, 100 mM NaCl.

For NMR experiments, lipid stocks containing 1,2-dioctanoyl-*sn*-glycero-3-phosphocholine (diC_8-PC) and 1,2-dioctanoyl-*sn*-glycero-3-phosphoglycerol (diC_8-PG) (Table 4) were prepared by first removing the organic solvent (chloroform) via nitrogen flux and under vacuum via desiccation overnight. Afterwards, the lipids were dissolved in the appropriate amount of buffer to yield 10 mM lipid stocks. For 1,2-dioctanoyl-*sn*-glycero-3-phospho-(1'-*myo*-inositol-4',5'-bisphosphate) (diC_8-PI(4,5)P_2) and 1,2-dioctanoyl-*sn*-glycero-3-phospho-(1'-*myo*-inositol-3',5'-bisphosphate) (diC_8-PI(3,5)P_2) were purchased as powders and directly dissolved in the approprate amount of buffer.

Table 4: Long- and short-chain lipids used for protein-lipid interaction studies.

Lipid	IUPAC name	Structure
POPG	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoglycerol	
POPC	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine	
diC_8-PC	1,2-dioctanoyl-*sn*-glycero-3-phosphocholine	
diC_8-PG	1,2-dioctanoyl-*sn*-glycero-3-phosphoglycerol	
diC_8-PI(4,5)P_2	1,2-dioctanoyl-*sn*-glycero-3-phospho-(1'-*myo*-inositol-4',5'-bisphosphate)	
diC_8-PI(3,5)P_2	1,2-dioctanoyl-*sn*-glycero-3-phospho-(1'-*myo*-inositol-3',5'-bisphosphate)	

2.11 Liposome co-sedimentation assay

Liposome co-sedimentation provides a straightforward tool to analyze protein-lipid interactions. Proteins are mixed with liposomes in a reaction tube and incubated to allow for complex formation. Afterwards, the mixture is ultra-centrifuged which results in the sedimentation of the liposomes. If the proteins bind to lipids, they will co-sediment together with the liposomes in the centrifugation pellet. If not, the proteins will remain in the supernatant. The distribution of protein in the pellet or in the supernatant can be analyzed via SDS-PAGE followed by Western-Blot or Coomassie staining, or photometrically (Figure 21 A) (*334*). Importantly, aggregation due to protein instability also results in protein sedimentation. To exclude that protein detected in the pellet is an aggregation artefact, a liposome-free sample is always analyzed as a control (Figure 21 B).

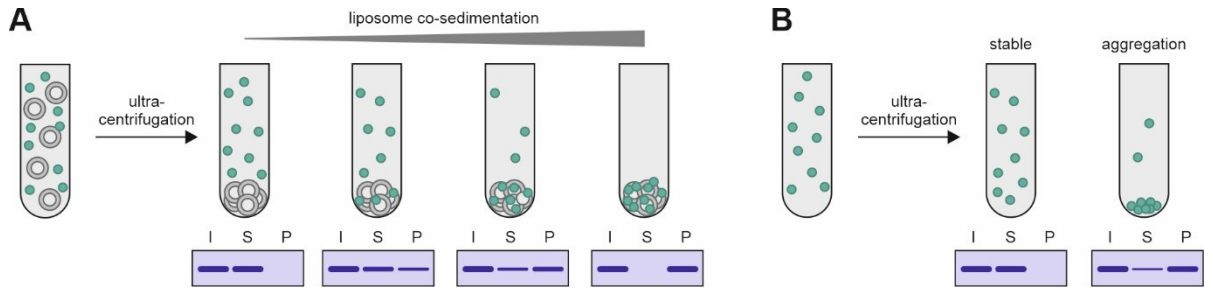


Figure 21: Schematic workflow of the liposome co-sedimentation assay. (**A**) Protein (green dots) and liposomes (grey circles) are mixed in a reaction tube and incubated before being applied to ultra-centrifugation. The distribution of the protein between the pellet and supernatant is afterwards analyzed via SDS-PAGE and Coomassie-staining, Here, 'I' is the input (sample before centrifugation), 'S' is the supernatant, and 'P' is the pellet fraction. The Coomassie-stained protein bands in the SDS-PAGE gels can be quantified densitometrically. (**B**) Liposome-free control sample. In the absence of liposomes, the protein remains in the supernatant during centrifugation if it is stable under the experimental conditions. Instability of the protein can lead to aggregation and sedimentation during centrifugation. Under these conditions, the liposome co-sedimentation assay is not suited to investigate protein-lipid interaction.

All liposome co-sedimentation experiments were carried out with Julia Nöth and Sarah-Ana Mitrovic.

For liposome co-sedimentation experiments with TRPV4 N-terminal constructs, liposomes were prepared as described in section VI2.10 with a diameter of 100 nm containing a mixture of POPG POPC in a 1:1 molar ratio. Proteins and liposomes were mixed in a 200 µL rection tube with 20 mM Tris pH 7, 100 mM NaCl to a final concentration of 2.5 µM protein and 2 mg mL^{-1} lipid. After incubation at 4°C for 1 hr under mild shaking, an SDS-PAGE sample of the input was taken (15 µL). The mixture was then ultra-centrifuged at 40 000 rpm for 1 hr at 4 °C (ML-130 rotor, Beckmann Coulter). Afterwards, the supernatant and pellet were carefully separated by pipetting the supernatant into a new reaction tube. The liposome pellet was then resuspended in an equal volume of assay buffer. SDS-PAGE samples (15 µL) were taken from both the supernatant and the suspended pellet. Control samples without liposomes were run in parallel to verify the protein stability under the experimental conditions. The samples were run on 12% SDS-PAGE gels, and protein bands were visualized via Coomassie staining. The protein distribution between the pellet and supernatant fractions was determined densitometrically using imageJ (*335*). Sedimentation assays were carried out three times for each protein liposome mixture and protein only control sample. Error bars were calculated as the standard deviation from the mean value of three replicates.

2.12 Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy is a versatile method to study protein secondary structure. Unlike NMR spectroscopy and X-ray crystallography, it provides low-resolution structural information but is much less demanding in terms of sample and time requirements. Additionally, it is a non-destructive technique that allows proteins to be studied under various conditions. The most frequent application of CD spectroscopy is the monitoring of secondary structure changes in proteins upon ligand binding. (*336*)

Plane polarized light can be described by two circular polarized components of identical magnitude, one rotating clockwise (L) and the other one counterclockwise (R). The CD effect describes the differential absorption of these components when a chiral chromophore is irradiated with plane-polarized light. Chiral chromophores or chromophores in chiral environments such as the peptide bonds in proteins tend to absorb one component more than the other. When a sample absorbs light with different L and R magnitudes, the result is an elliptically polarized wave. Practically, in a CD spectroscopic measurement, the incident radiation is modulated via switching between L and R components. The modulator, a quartz crystal, is subjected to an alternating electrical field which induces structural changes in the crystal. At the extremes of the field, the crystal transmits circularly polarized light. The transmitted radiation is detected by a photomultiplier. (*337*)

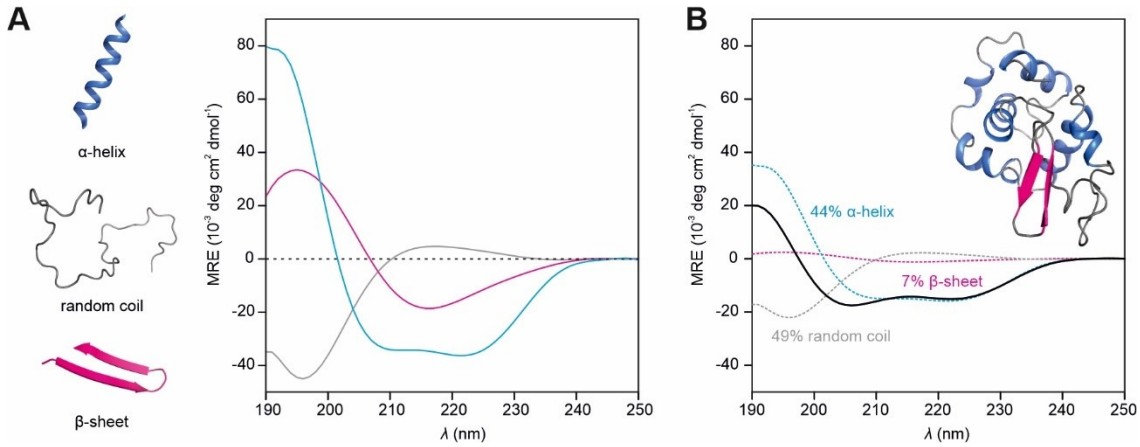


Figure 22: Circular dichroism spectra of protein secondary structure elements. (**A**) The secondary structure elements α-helix (blue), β-sheet (magenta), and random coil (grey) give rise to characteristic maxima and minima in the CD spectrum of a protein. The basis spectrum of random coil structures displays a minimum at ~197 nm, α-helix shows two minima at 208 and 222 nm, respectively, and β-sheet structures display a maximum at 197 nm and a minimum at 218 nm. The basis spectra were obtained from ref. (*338*). (**B**) The CD spectrum of a protein (here lysozyme, PDB: 6lyz) containing a mix of secondary structures is the population-weighted linear combination of the basis spectra of the individual secondary structure elements.

The CD signal is detected as the difference in L and R absorption and is given in degrees (deg). For proteins the CD effect is relatively weak and the typical values are around 10 mdeg. (*337*)

CD measurements

CD measurements were carried out with Julia Nöth and Sarah-Ana Mitrovic.

CD measurements were carried out on a Jasco-815 CD spectrometer (Jasco™, Gross-Umstadt, Germany) with 1 mm quartz cuvettes (Hellma Macro Cell). Spectra were recorded at 20 °C in a spectral range between 190 to 260 nm with 1 nm scanning intervals, 5 nm bandwidth and 50 nm min^{-1} scanning speed. All spectra were obtained from the automatic averaging of three measurements with automatic baseline correction. The measured ellipticity θ in degrees (deg) was converted to the mean residue ellipticity, MRE, via equation 10. (*337*)

$$\mathrm{MRE}_\lambda = \frac{\mathrm{MRW} \cdot \theta_\lambda}{10 \cdot d \cdot c} \qquad 10$$

Here, MRE_λ is the mean residue ellipticity, and θ_λ is the measured ellipticity at wavelength λ, d is the pathlength of the light through the sample (in cm), and c is the protein concentration (g mL^{-1}). MRW is the mean residue weight, $\mathrm{MRW} = M_W \cdot (N\text{-}1)^{-1}$, where M_W is the molecular weight of the protein (in Da), and N is the number of residues. For titration experiments, TRPV4 N-terminal peptides were used in a concentration of 30 µM in bidest. H_2O in the presence of TFE (2,2,2-trifluoroethanol, 0-90% (v/v)), SDS (0.5, 1.0, 2.5, 5.0 and 8.0 mM) and liposomes (0.5 and 1.0 mM). Liposomes were prepared from POPG and POPC at a molar ratio of 3:1.

BeStSel analysis

Secondary structure analysis of CD spectra was carried out using the Beta Structure Selection (BeStSel) tool (*339*). The BeStSel tool fits the experimental CD curve with a linear combination of fixed basis spectra of eight secondary structure classes: Regular α-helix, distorted α-helix, left-twisted β-sheet, relaxed antiparallel β-sheet, right-twisted antiparallel β-strand, parallel β-sheet, turn, and others. For most analyses, the eight classes were

merged into three classes: α-helix (regular and distorted), β-sheet (left-twisted, relaxed, right-twisted, antiparallel, parallel), and random coil (turns and others). The BeStSel algorithm converts the MRE value into an absorption units-based value, termed $\Delta\varepsilon$ equation. The MRE and $\Delta\varepsilon$ are related by MRE = 3298 · $\Delta\varepsilon$. All predictions discussed in this thesis fulfil a $\Delta\varepsilon$ RMSD < 0.15 between the experimental and fitted CD curves.

2.13 Tryptophan fluorescence spectroscopy

Of the three fluorescent aromatic amino acids tryptophan, tyrosine, and phenylalanine, the tryptophan indole side chain is the most dominant fluorophore. Tryptophan fluorescence is remarkably sensitive to the polarity of the environment. According to a process called solvent relaxation, polar solvent molecules like H_2O can decrease the energy of the excited state by aligning their dipoles with the transition dipole of the fluorophore. This results in a lower energy gap between the solvent relaxed, excited state and the ground state and thus in a redshift of the fluorescence emission (higher wavelength). An increasing solvent polarity enhances the redshift of the fluorescence emission. Conversely, apolar solvents result in a blue shift (Figure 23 A). Aside from shifting the fluorescence emission maximum, the local environment of a tryptophan commonly influences its fluorescence quantum yield, which reflects in changes of the fluorescence emission intensity (Figure 23 A). The currently proposed mechanism underlying a decrease in tryptophan fluorescence intensity is electron transfer quenching by local peptide carbonyl groups or by neighbouring amino acid side chains (*340*).

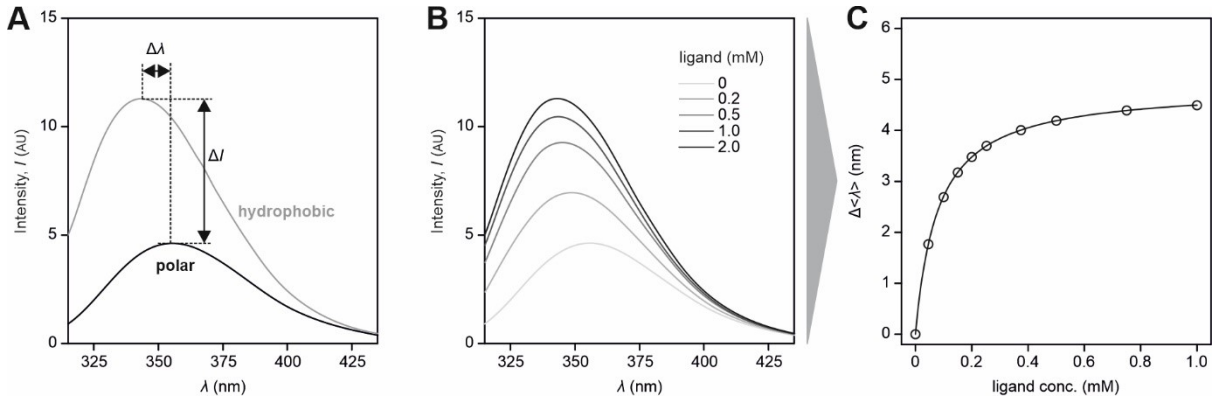


Figure 23: Dependence of tryptophan fluorescence on the polarity of the environment. (**A**) Fluorescence emission of a tryptophan residue in a hydrophobic (grey) and a polar (black) environment. The tryptophan fluorescence intensity is quenched, Δ*I*, and its wavelength shifts, Δ*λ*, when transitioning between hydrophobic and polar environments. (**B**) Ligand binding or conformational changes can transition a tryptophan residue between solvent-exposed and buried positions. This can be monitored as wavelength shifts and fluorescence quenching in the tryptophan fluorescence spectra. (**C**) Plotting the fluorescence shifts in the tryptophan fluorescence spectrum of a protein caused by the binding of a ligand can be fitted with a binding isotherm to derive the dissociation constants of the protein-ligand complex.

Its environment-sensitive spectral properties make tryptophan a versatile probe to monitor conformational changes or ligand binding which change the position of a tryptophan side chain from a hydrophobic to a solvent-exposed position or *vice versa* (Figure 23 B and C) (*341, 342*). In this thesis, tryptophan fluorescence was used to monitor the chemical denaturation of proteins or protein/peptide binding to phospholipid-containing vesicles. All tryptophan fluorescence measurements were carried out on a Fluro Max-4 fluorimeter with an excitation wavelength of 280 nm and a detection range between 300 nm to 550 nm. The fluorescence wavelength was determined as the intensity-weighted fluorescence wavelength between 320-380 nm (hereafter referred to as the average fluorescence wavelength) according to equation 11.

$$\langle\lambda\rangle = \frac{\sum_{i=320\text{ nm}}^{n=380\text{ nm}} I_i \cdot \lambda_i}{\sum_{i=320\text{ nm}}^{n=380\text{ nm}} I_i} \qquad 11$$

Here, $\langle\lambda\rangle$ is the average fluorescence wavelength and I_i is the fluorescence intensity at wavelength λ_i.

All tryptophan fluorescence studies were carried out with Julia Nöth and Sarah-Ana Mitrovic.

Liposome binding of the TRPV4 N-terminus

To monitor liposome binding of the TRPV4 N-terminal constructs via tryptophan fluorescence, fluorescence emission spectra were recorded in the presence of increasing lipid concentrations. Lipids were prepared as SUVs with 100 nm diameters as described in section VI2.10. The protein concentrations were kept constant at 5 µM. The reaction buffer contained 10 mM Tris pH 7, 100 mM NaCl. The protein-liposome mixtures were incubated for 10 min prior to recording the emission spectra. The tryptophan fluorescence wavelength at each lipid concentration was quantified by determining the average fluorescence wavelength using equation 11. The dissociation constants, K_D, of the protein liposome complexes were determined by plotting the changes in $\langle\lambda\rangle$ as $\Delta\langle\lambda\rangle$ against the lipid concentration c and fitting the data with a Langmuir binding isotherm (equation 12).

$$\Delta\langle\lambda\rangle = \frac{\Delta\langle\lambda\rangle_{max} \cdot c}{K_D + c} \qquad 12$$

Here, $\Delta\langle\lambda\rangle$ describes the wavelength shift between the start spectrum, that is in the absence of lipids, and a given titration step at lipid concentration c. $\Delta\langle\lambda\rangle_{max}$ indicates the maximum wavelength shift in the saturation regime of the binding curve. Under the assumption that proteins cannot diffuse through liposome membranes and therefore can only bind to the outer lipids, the lipid concentration c was set to half of the titrated lipid concentration and is termed the effective lipid concentration.

Chemical denaturation of the TRPV4 N-terminus

Chemical denaturation using urea as denaturant is a versatile method to evaluate the stability of proteins and to investigate the folding and unfolding of proteins (*343*). Urea denatures proteins by interacting with peptide groups through hydrogen bonding, thereby disturbing the protein's native hydrogen bonding network. It is also suggested that urea molecules replace some of the water molecules in the hydration shell formed around non-polar groups of the protein and thereby diminish the hydrophobic effect (*344*).

The unfolding of TRPV4 N-terminal constructs was investigated by measuring fluorescence emission spectra at urea concentrations from 0-8 M in 0.5 M steps using 10 mM Tris pH 7, 100 mM NaCl as the reaction buffer. The protein concentration was kept constant at 5 µM. All samples were incubated for 20 min prior to recording the fluorescence emission spectra. The tryptophan fluorescence wavelength at each urea concentration was quantified by determining the average fluorescence wavelength using equation 11.

2.14 Fluorescence correlation spectroscopy

Theoretical background

Fluorescence correlation spectroscopy (FCS) is an established tool to measure the diffusion behaviour of fluorescent particles in solution (*345*). In an FCS experiment, a confocal laser scanning microscope monitors the diffusion of single fluorescent particles in and out of a confocal volume as a time trace of fluorescence fluctu-

ations (Figure 24 A). The temporal self-similarities of the fluorescence fluctuations around an average fluorescence are quantified with a second-order normalized autocorrelation function $G(\tau)$, which is given by equation 13.

$$G(\tau) = 1 + \frac{\langle \delta I(t) \cdot \delta I(t+\tau) \rangle}{\langle I(t) \rangle^2} \qquad 13$$

Here, $\delta I(t) = I(t) - \langle I(t) \rangle$ is the fluorescence fluctuation from the temporally averaged intensity $\langle I(t) \rangle = 1/\int_0^T I(t)dt$ at a given time t (Figure 24 A). τ denotes the lag time, which is the time difference between two correlated intensity fluctuations in the fluorescence time trace. According to equation 13, the autocorrelation curve (ACC) will have the maximum value at $\tau = 0$. Increasing τ leads to a decrease of the self-similarity between fluorescence fluctuations. The typical $G(\tau)$ is a sigmoidal curve (Figure 24 A, bottom) which decays to 1 when the self-similarity between $I(t)$ and $I(t+\tau)$ is 0 (equation 13). For small particles, the self-similarity between $I(t)$ and $I(t+\tau)$ decreases rapidly due to the fast diffusion out of the confocal volume, and thus $G(\tau)$ decays at small τ values. Conversely, large particles stay longer in the confocal volume, and $G(\tau)$ decays at larger τ values. The slower a particle diffuses, the further $G(\tau)$ is shifted to larger τ values (Figure 24 A). If the monitored fluorescence fluctuations are solely caused by the diffusion of a single particle species through the confocal volume, $G(\tau)$ can be related to the diffusion time of the particle according to equation 14.

$$G(\tau) = 1 + \frac{1}{N} \cdot \frac{1 + \frac{\tau}{\tau_D}}{\sqrt{1 + \frac{\tau}{S^2 \cdot \tau_D}}} \qquad 14$$

Here, τ_D is the diffusion time of the particle, N is the average number of particles in the confocal volume, and S is the so-called structure factor $S = z_0 \cdot r_0^{-1}$, where z_0 and r_0 represent the axial and radial dimension of the confocal volume, respectively. Based on equation 14, FCS enables to determine the diffusion time of a particle which relates to the diffusion coefficient D according to equation 15.

$$D = \frac{{r_0}^2}{6 \cdot \tau_D} \qquad 15$$

The diffusion coefficient is useful to calculate the hydrodynamic radius of the diffusing particle, also termed Stokes radius R_S (see section VI2.8) using the Stokes-Einstein relation (equation 16).

$$R_S = \frac{k_B \cdot T}{6\pi \cdot \eta \cdot D} \qquad 16$$

Here, T is the temperature, k_B the Boltzmann constant and η the viscosity of the solvent.

The ability of FCS to determine the size of a fluorescent particle in solution makes it an attractive tool to investigate the interaction of fluorescently labelled proteins with liposomes (Figure 24 B) (*346*). Two different diffusing species will occur during the titration of a fluorescently labelled protein with liposomes, one corresponding to the free protein with τ_1 (fast diffusion) and the other to the liposome-bound protein with τ_2 (slow diffusion). Such a two-component system containing a fast and a slowly diffusing species gives rise to an ACC with two decays, one at τ_1 and the other at τ_2 (Figure 24 B). With increasing liposome concentrations, the population of the free protein will decrease in favour of the liposome-bound protein population. Consequently, the decay of the ACC at τ_1 will diminish while the second decay at τ_2 will emerge. $G(\tau)$ of a multi-component system of m different types of freely diffusion fluorescent species can be described according to equation 17 (*347*).

$$G(\tau) = 1 + \frac{1}{N}\sum_{i=1}^{m}\frac{f_i}{\left(1+\frac{\tau}{\tau_{D,i}}\right)\cdot\sqrt{1+\frac{\tau}{S^2\cdot\tau_{D,i}}}} \qquad 17$$

Here, $\tau_{D,i}$ is the diffusion time of the i-th species, and f_i is the fraction of the i-th component. In the context of a liposome titration experiment, equation 17 enables to determine the size of the free protein and the liposome/protein complex and their corresponding fractions from the ACC. In this way, the faction of liposome-bound protein can be monitored at each liposome concentration, thus enabling to determine the affinity of the protein for lipids.

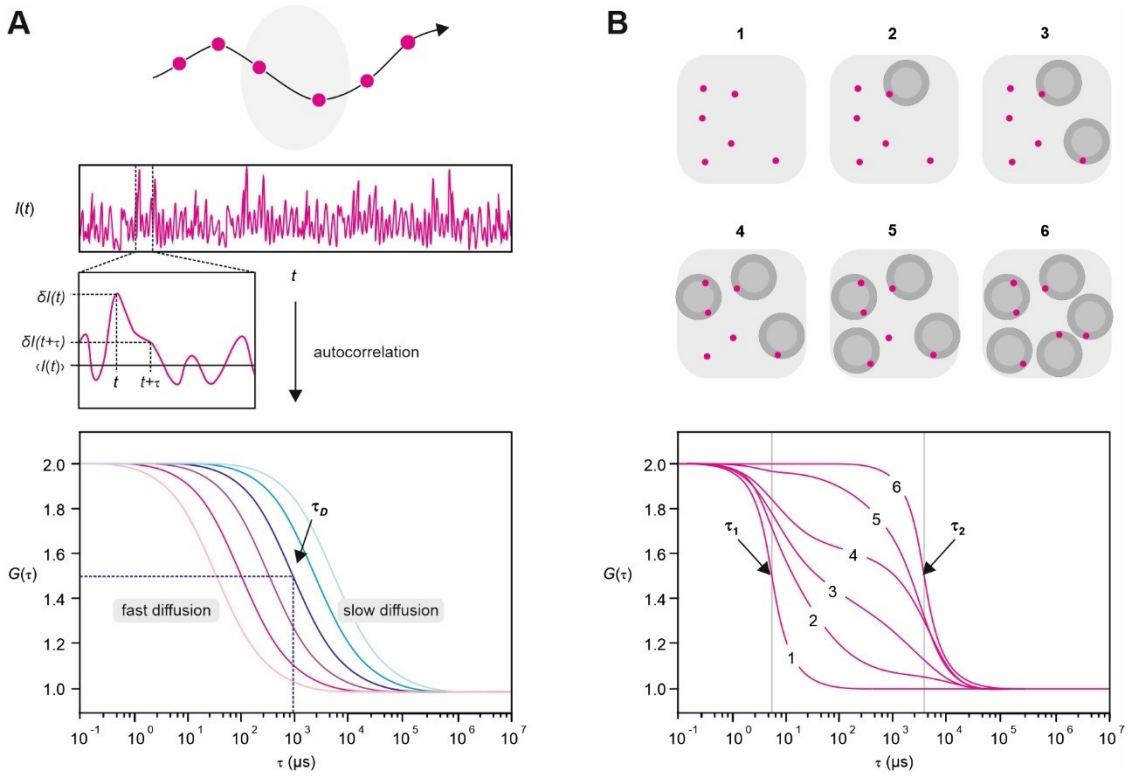


Figure 24: Principle of fluorescence correlation spectroscopy. (**A**) Depiction of a fluorescent particle (pink) diffusing through a confocal volume (grey). The diffusion in and out of the confocal volume gives rise to fluctuations in a fluorescence time trace, $I(t)$ vs t. Fluctuations $\delta I(t)$ from the average fluorescence $\langle I(t)\rangle$ are illustrated in the box below the fluorescence time trace. The fluorescence time trace can be analyzed by an autocorrelation function $G(\tau)$ (equation 13) which compares the self-similarity between $\delta I(t)$ and $\delta I(t+\tau)$ as a function of τ. The resulting autocorrelation curve can be fitting with equation 14 to calculate the diffusion time τ_D of the fluorescent particle. (**B**) Liposome binding to a fluorescent protein monitored via FCS. In the absence of liposomes (1), a fluorescent protein (pink dot) diffuses with τ_1 and gives rise to an ACC centred at τ_1. Upon titrating liposomes to the sample (2), a second species of diffusing particles arises, i.e. proteins bound to liposomes which diffuse with τ_2 ($\tau_1 < \tau_2$). This second species gives rise to a second decay in the ACC at τ_2. Increasing the liposome concentration (3-5) increases the amount of liposome-bound proteins at the cost of the free proteins. This shifts gradually diminishes the decay at τ_1 in favour of the decay at τ_2 in the ACC. When all proteins are bound to liposomes (6), the ACC solely decays with τ_2.

Protein labelling

A *H. sapiens* TRPV4 IDR V148C mutant was generated from the wild-type IDR through site-directed mutagenesis (forward primer CCCCCATCCTCAAATGCTAAGGATCCGGCTG, reverse primer AGCCGGATCCTTAGCATTTGAGGATGGGGGG) in a single PCR reaction using the quick-change protocol (*327*). The introduced cysteine residue at position 148 is the only cysteine in the entire IDR sequence and thus enables for site-specific labelling with a fluorescent probe via cysteine-maleimide conjugation. The IDR V148C mutant could be expressed and purified in analogy to the wild-type IDR (see section VI2.5). However, the mutant was

gel filtrated into 10 mM Tris pH 7, 100 mM NaCl, 1 mM TCEP in the last purification step. Fluorescence-labeling was achieved by incubating the V148C mutant with a 10-fold molar excess of Atto488-maleimide (Sigma Aldrich) in 10 mM Tris pH 7, 100 mM NaCl, 1 mM TCEP for 4 hrs at 4 °C. The labelled protein was separated from the free dye via SEC. The labelling degree of 12% was determined from the OD_{280} and OD_{501} in the SEC chromatogram according to equation 18 (*348*).

$$DOL = \frac{OD_{501} \cdot \varepsilon_{280}}{(OD_{280} - \varepsilon_{280} \cdot CF) \cdot \varepsilon_{501}} \qquad 18$$

Here, DOL is the degree of labelling, ε_{501} is the extinction coefficient of Atto488-maleimide at the emission maximum of 501 nm (90 000 M^{-1} cm^{-1}), and ε_{280} is the extinction coefficient of the protein at 280 nm (9960 M^{-1} cm^{-1}). The correction factor (CF) is the ration between OD_{280} and OD_{501} of the free Atto488 dye.

FCS measurements

FCS experiments were carried out in collaboration with Sascha Schmidt and Dr. Kaloian Koynov (Max Planck Institute for Polymer Research, MPI-P). All measurements were performed using a commercial setup LSM 880 microscope (Carl Zeiss, Jena, Germany). For excitation of Atto488 an argon-ion laser (488 nm) was used. The excitation light was focused into the sample by a C-Apochromat 40x/1.2 W (Carl, Zeiss, Jena, Germany) water immersion objective. The fluorescence light was collected with the same objective and after passing through a confocal pinhole, directed to a spectral detection unit (Quasar, Carl Zeiss). The detected emission range was in the spectral range of 508-562 nm. Calibration of the detection volumes was carried out with Alexa Fluor488®. FCS data were analysed using the software ZEN.

All measurements were performed in an eight-well polystyrene-chambered coverglass (Laboratory-Tek, Nalge Nunc Internation, Penfield, NY, USA). Protein adsorption during FCS measurements was suppressed by coating the eight-well plate chambers with bovine serum albumin (BSA). For this, the chambers were preincubated overnight at room temperature with a 0.1 mg mL^{-1} BSA in 10 mM Tris pH 7, 100 mM NaCl, 1 mM DTT, 0.01% Tween. Afterwards, the chambers were washed three times with measurement buffer (10 mM Tris pH 7, 100 mM NaCl, 1 mM DTT, 0.01% Tween). All samples were measured twenty times, with a total duration of 3 minutes. Titration experiments were performed with liposome concentrations of 1000 μM to 7 μM and a fixed IDR_V148C_Atto488 concentration. All solutions were incubated for 10 minutes before the measurement.

2.15 Cross-linking mass spectrometry

Theoretical background

Chemical cross-linking is a versatile method to obtain structural restraints within proteins or protein complexes. In a cross-linking experiment, a chemical agent is added to a protein or a mixture of proteins and connects two functional groups of amino acid side chains either within the same protein or between different proteins in a complex (Figure 25 A). The cross-linker contains two reactive groups which can form covalent bonds to specific side chains. Commonly targeted side chains by the reactive groups in cross-linking experiments are lysine and cysteine. A linker of a defined length separates the reactive groups in the cross-linker. It is referred to as a 'molecular ruler' (*349*). Altering the length of the molecular ruler enables to sample different distances in proteins and protein complexes. The identification of cross-linked residues in proteins is achieved by enzymatic digestion into peptides, which are subsequently analyzed via liquid chromatography electrospray-ionization tandem mass spectrometry (LC/ESI-MS/MS) (Figure 25 A).

The distance constraints of the linker serve as the basis to deduce the spatial proximity of cross-linked protein residues. They become particularly useful for the modelling of protein complex for which high-resolution structures of the individual components are known. The identified library of cross-linked peptides reveals contacts between residues within the same protein and between different proteins (Figure 25 B). Comparing intra-XLs with high-resolution structures may enable to assess whether intra-residue distances in a protein differ in solution, e.g. due to conformational flexibility of a protein. Cross-links between peptides of different proteins can guide the reconstruction of the spatial arrangement of proteins in a complex. In this sense, chemical cross-linking mass spectrometry (XL-MS) can be a versatile complement for structural studies when high-resolution methods are not applicable due to restrictions by protein flexibility or size (*350*).

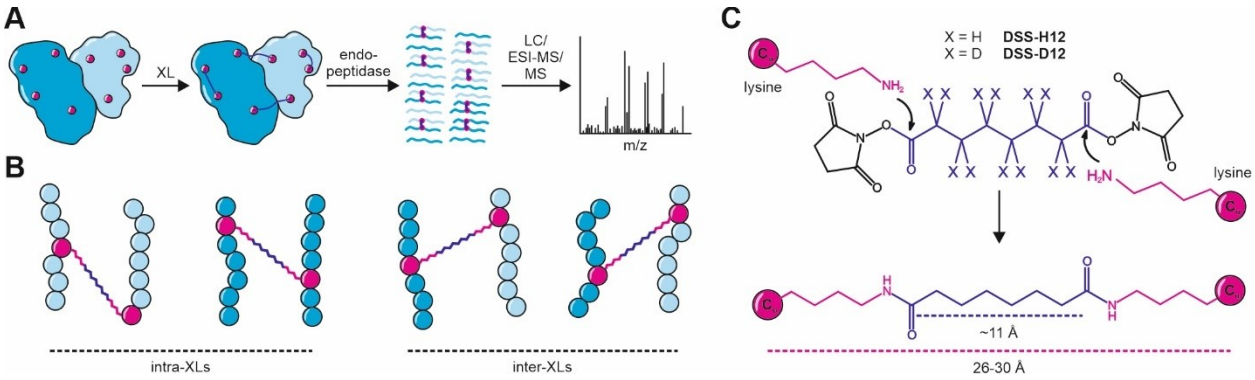


Figure 25: Basic principle of chemical cross-linking mass spectrometry for structural studies of proteins and protein complexes (**A**) In an XL-MS experiment, a sample containing a protein or, as shown in the example, a protein complex is incubated with a cross-linker (XL) which connects proximal pairs of lysine side chains. After the cross-linking reaction, the proteins are digested with an endopeptidase, which yields a mixture of cross-linked and non-cross-linked peptides. The mixture is applied to liquid chromatography electrospray-ionization tandem mass spectrometry (LC/ESI-MS/MS) to identify the cross-linked peptides. (**B**) The library of cross-linked peptides identified in the MS analysis reveals intra-XLs, i.e. XLs between lysine residues within the same protein, and inter-XLs, i.e. between different proteins. (**C**) A commonly used XL agent is the N-hydroxysuccinimide (NHS) ester disuccinimidyl suberate (DSS) with a linker length of ~11 Å. DSS can cross-link a pair of lysine residues with C_α-C_α distances of 26-30 Å. Deuteration of the suberate linker region facilitates the identification of cross-linked peptides during MS analysis.

The most popular cross-linkers used in XL-MS are homobifunctional *N*-hydroxysuccinimide (NHS) esters. At pH 7-8, NHS esters predominantly react with amine groups such as the side chains of lysine. Lysine residues are versatile cross-linking targets as they are abundant on protein surfaces and thus may enable to probe a multitude of intra-protein and inter-protein distances in a single measurement. The identification of cross-linked peptides in the mixture with unlinked peptides is facilitated with deuterated cross-linkers. In the MS analysis, isotope-specific m/z shifts in mass spectra indicate a cross-link (*349*). A standard NHS-based cross-linker is disuccinimidyl suberate (DSS) (Figure 25 C). With a linker length of ~11 Å, DSS can cross-link lysine residues with C_α-C_α distances of 26-30 Å (*351*).

Protein cross-linking

Purified *G. gallus* TRPV4 NTD and *G. gallus* fl-Pac3 in 20 mM HEPES pH 7, 300 mM NaCl were mixed in a 1:1 molar ratio to a final concentration of 0.1 mg mL^{-1} in a 100 µL reaction volume and incubated with 6 µL freshly prepared 25 mM H12/D12-disuccinimidyl suberate (H12/D12-DSS) dissolved in dimethylformamide (DMF). After 2 h incubation on ice, the cross-linking reaction was quenched by the addition of 5 µL of a 1 M ammonium bicarbonate and further incubation on ice for 2 h. The samples were subsequently dried in a SpeedVac centrifuge (Eppendorf) and stored at -20 °C until further use.

Sample processing

Processing of cross-linked samples for subsequent MS-analysis was carried out by Jasmin Jansen (University of Konstanz). The dried, crosslinked (H12/D12-DSS crosslinker from Creative Molecules) protein was denatured in 100 µL 8 M urea (BioReagent, Sigma Life Sciences) before 5 µL 50 mM TCEP (BioUltra, Sigma Life Sciences) were added, and the reduction was performed at 37 °C for 30 min. The sample was cooled down before cysteines were alkylated with 5 µL 100 mM iodoacetamide (freshly prepared; BioUltra, *Sigma Life Sciences*) for 30 min at room temperature in the dark. The solution was diluted to a maximal urea concentration of 1 M urea using 50 mM ammonium bicarbonate. Then, digestion was performed with 5 µL 0.5 µg/µL trypsin (Sequencing grade modified, Promega) at 37 °C overnight and stopped with 16 µL 98 % formic acid (FA, Sigma Aldrich). The sample was desalted on a 1 mL C18 Sep-Pak column (Sep-Pak tC18 1 cc Vac cartridge, 50 mg, Waters). The C18 column was wetted with 1 mL acetonitrile (ACN; UHPLC grade, Fisher Scientific UK) and equilibrated with 2 mL 5 % ACN, 0.1 % FA before the sample was loaded and washed twice with 2 mL 5 % ACN, 0.1 % FA. The peptides were eluted with 1 mL 50 % ACN, 0.1 % FA and dried on a SpeedVac. For enrichment of cross-linked peptides, the sample was thoroughly resuspended in 22 µL SEC sample buffer (70 % water, 30 % ACN, 0.1 % TFA) of which 18 µL were loaded onto the SEC column (column: Superdex peptide 3.2/300, GE Healthcare, running buffer: 30 % ACN, 0.1 % TFA, flow: 0.05 mg/mL). Afterwards, the crosslink fractions were collected and dried on a SpeedVac. The samples were frozen at -20°C until MS analysis.

Mass spectrometric analysis and evaluation

The dried samples were thoroughly resuspended in MS buffer (5 % ACN, 0.1 % FA) and transferred to MS vials. Mass-spectrometric analysis was done on an Orbitrap Fusion™ Tribrid™ mass spectrometer coupled to EASY-nLC™ 1200 system with an Acclaim™ PepMap™ 100 column (all Thermo Scientific). Cross-link data analysis of crosslinked samples was done with the xQuest software (*352*). Data were searched using ion-tag mode with a precursor mass tolerance of 10 pm. Crosslinks which were identified in at least two of three biological samples with at least one ld-score ≥ 25 were regarded as confident and were visualized using the software xiNET (*353*).

2.16 Hydrogen/deuterium-exchange mass spectrometry

Theoretical background

Hydrogen/deuterium-exchange mass spectrometry (HDX-MS) measures the exchange rates of amide hydrogen atoms with deuterated solvent. The exchange rates and kinetics of hydrogen atoms is related to the structure and dynamics of a protein. Therefore, HDX-MS is an exquisite tool to probe the conformational flexibility of a protein and its dynamic response to external stimuli such as ligand binding, pH, or temperature. In an HDX-MS experiment, a protein sample is incubated in a deuterated buffer, and the deuteration reaction is quenched after various time points with a low pH buffer. Similar as in XL-MS (see section VI2.15, Figure 25 A), the protein is digested into smaller peptides and the deuteration of the individual peptides at different time points assessed via mass spectrometric analysis. This provides sequence-specific information on the deuteration rate of the protein (*354*). Deuteration reflects in the mass spectra of the proteolytic peptides as a shift of their corresponding isotopic envelopes to higher m/z ratios. The deuteration kinetics reflect on local or large-scale conformational dynamics.

The exchange of hydrogen for deuterium in the amide groups of a protein is a base- and acid-catalyzed process with a global minimum of the exchange rate at around pH 2.2. The rate of exchange of a given amide group depends on the presence of secondary structure elements and its solvent accessibility. At pH 7 and 25 °C, the rate of exchange of any given amide hydrogen in an unfolded peptide typically occurs in the order of millisec-

onds to seconds (Figure 26 A). The exchange rate of a given amide group in an unfolded peptide can be regarded as the amide group's intrinsic hydrogen exchange rate because it has the maximum solvent exposure and is not stabilized by intramolecular hydrogen bonds. The exchange rate is reduced by factors up to 10^8 when a hydrogen atom is involved in the formation of secondary structures or when it is in a solvent-inaccessible position (Figure 26 A). However, even hydrogens in very stable secondary structures or buried positions can exchange through conformational fluctuations in the protein which transiently disrupt hydrogen bonds or transiently expose the hydrogen atoms to the solvent. The mechanism governing the exchange of amide hydrogen atoms can be described by equation 19 (*355*).

$$\text{N-H}_{cl} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{N-H}_{op} \xrightarrow[D_2O]{k_{in}} \text{N-D}_{op} \underset{k_{op}}{\overset{k_{cl}}{\rightleftharpoons}} \text{N-D}_{cl} \qquad 19$$

Here, N-H_{cl} represents an amide group in a protected (closed, cl) conformation which is in equilibrium with an exposed conformation N-H_{op} (open, op). The exchange rate from the closed to the open conformation is given by k_{op}, whereas the exchange rate for the reverse transition is given by k_{cl}. Upon addition of an excess of D_2O, the amide group in the N-H_{op} conformation is irreversibly deuterated to give N-D_{op}. After being deuterated, the open conformation re-equilibrates with the protected conformation, indicated as N-D_{cl} (*355*).

The exchange rate of amide hydrogen atoms in folded protein regions depends on the exact rates of interconversion between an exchange-incompetent (closed) and an exchange-competent (open) conformation as well as on its intrinsic deuteration rate. In most amide groups of folded proteins, so-called EX2 kinetics are observed when k_{in} is much smaller than k_{cl} and k_{op}. In this case, the interconversion between exchange-incompetent and an exchange-competent states and thus opening and closing of the amide group is much faster than the intrinsic hydrogen exchange rate. Therefore, each interconversion cycle only allows a subset of the hydrogens to exchange. EX2 hydrogen exchange is characterized by unimodal isotopic envelopes in the mass spectra (Figure 26 A). Multiple opening-closing cycles are needed to deuterate peptide segments over time completely. Therefore, the mass spectra envelopes shift gradually to higher m/z ratios over time. The rates and extents of exchange underlying EX2 kinetics reflect on the local conformational flexibility of a protein (*354*).

When an exchange-competent state is sufficiently long-lived such that k_{in} is much higher than k_{cl} and k_{op}, a complete deuteration of all amide hydrogen atoms within a single interconversion cycle can occur. This so-called EX1 mode results in bimodal isotopic envelopes in the mass spectra (Figure 26 B). One envelope corresponds to the fully hydrogenated peptide and the other envelope to the fully deuterated peptide. Over time, the envelope of the deuterated peptide increases at the cost of the envelope of the hydrogenated peptide. EX1 kinetics are considered to reflect large-scale rearrangements. They often occur in regulatory regions such as lids and hinges or in binding sites. Moreover, EX1 is an indicator of transiently unfolded states (*356*). In a folded state of a protein region, deuteration is slow due to the amide hydrogens being involved in hydrogen bonds. Upon transition an unfolded state, the deuteration rate is drastically increased such that all hydrogen atoms in the unfolded region can exchange rapidly. Because deuteration of unfolded regions is extremely fast, the unfolded state does not need to be extremely long-lived to observe EX1 behaviour. It is important to not that deuteration kinetics not always follow strict EX2 or EX1 kinetics but can display features of both extremes. This can depend on the intrinsic dynamics of the protein, but it can also occur when a peptide constitutes a sequence in which one part exchanges which EX2 and the other part with EX1 in the protein (*357*).

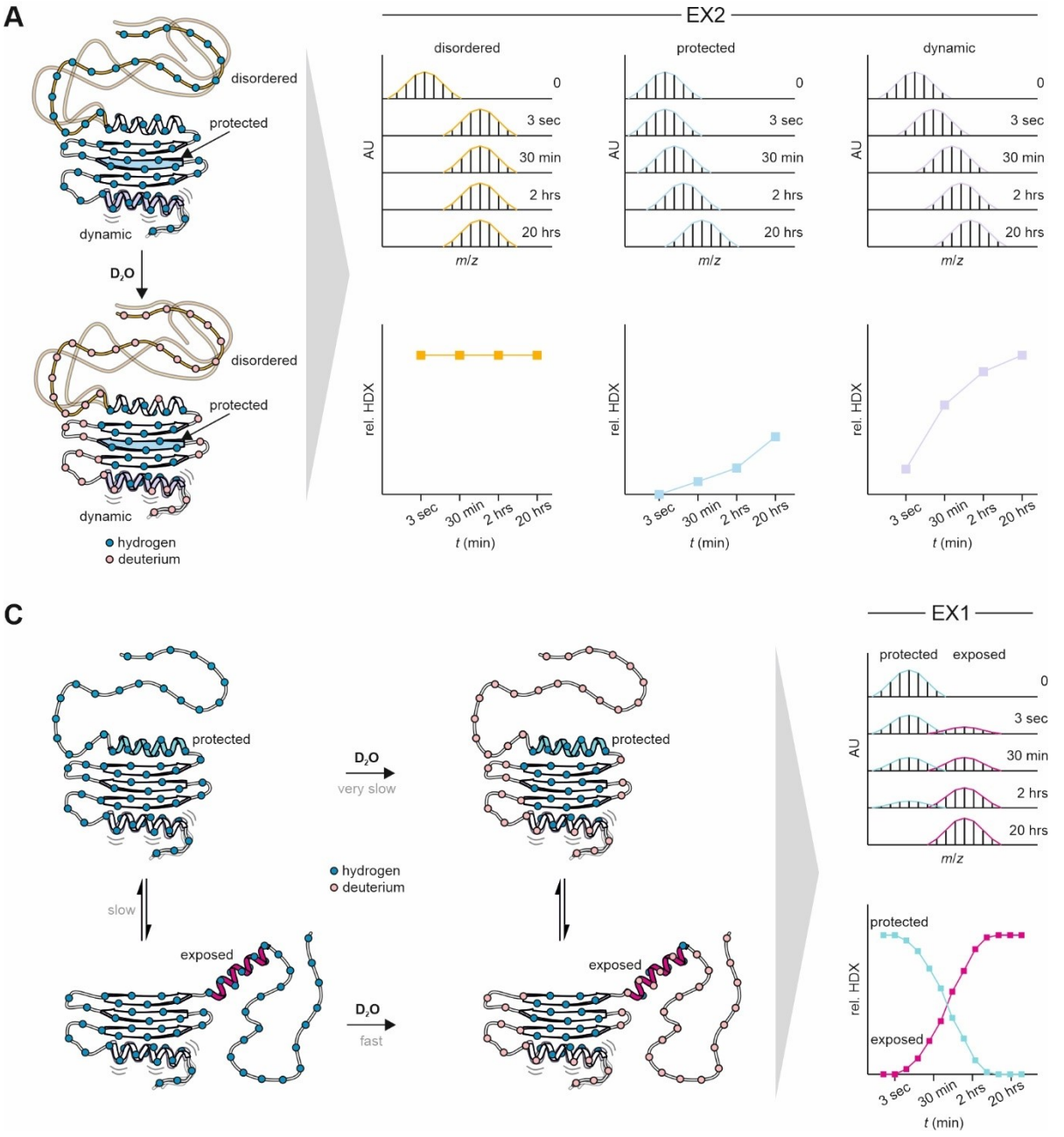


Figure 26: Protein structure and dynamics monitored via hydrogen/deuterium-exchange mass spectrometry. (**A**) Under equilibrium conditions, most folded proteins display EX2 deuteration kinetics, i.e. gradual incorporation of deuterium into their backbone. The rate and extent of deuteration can be directly related to the local structure and dynamics which cause fluctuations in amide hydrogen accessibility. Deuteration manifests as gradual shifts of the isotopic envelopes towards higher m/z ratios in the mass spectra of the proteolytic peptides. In disordered regions (shown in yellow), hydrogen exchange is extremely fast because all amide hydrogens are solvent-exposed and not stabilized by secondary structures. A full deuteration of a disordered protein or protein region usually occurs within milliseconds to seconds. Amide hydrogens involved in stable secondary structures and solvent excluded regions (light blue) are protected from deuteration. Hydrogen exchange still occurs slowly through local conformational flexibility. Deuterium uptake is facilitated in dynamic regions with increased local flexibility (light purple). (**B**) Large-scale conformational changes such as the displacement of helices (shown in this example) or transient unfolding can lead to so-called EX1 deuteration kinetics. This can be illustrated by a helix in a protein existing in an equilibrium between an exchange-incompetent state (protected) and an exchange-competent conformation (exposed) and a slow interconversion rate. When the amide hydrogens in the exposed helix exchange faster than the conversions to a protected helix conformation, a full deuteration of the helix will occur before the exposed state transitions to a protected state. This reflects as a bimodal isotopic envelope in the mass spectra of the peptides corresponding to the helix. One envelope reflects the non-deuterated helix and the other protein the fully deuterated helix. Whenever a non-deuterated helix becomes exposed, it is fully deuterated. Thus, over time, the mass spectrum envelope corresponding to the non-deuterated helix will decrease in favour of the envelope corresponding to the fully deuterated helix.

HDX-MS measurements

Hydrogen-deuterium exchange mass spectrometry (HDX-MS) measurements and data analyses were carried out by Dr. Wieland Steinchen (Philipps-University Marburg) as described previously (*358, 359*). Preparation of samples for HDX-MS was aided by a two-arm robotic autosampler (LEAP Technologies). Purified *G. gallus* TRPV4 IDR, ARD or NTD protein and *G. gallus* fl-Pac3 or F-BAR (50 µM) was diluted 10-fold in D_2O-containing buffer (20 mM Tris pH 7, 300 mM NaCl) and for 10, 95, 1,000 or 10,000 s at 25 °C. The reactions were stopped by mixing with an equal volume of pre-dispensed quench buffer (400 mM KH_2PO_4/H_3PO_4, 2 M guanidine-HCl; pH 2.2) kept at 1 °C, and 100 µl of the resulting mixture injected into an ACQUITY UPLC M-Class System with HDX Technology (*360*). Undeuterated samples were generated by similar procedure through 10-fold dilution in H_2O-containing buffer. The injected HDX samples were washed out of the injection loop (50 µl) with water + 0.1% (*v/v*) formic acid at a flow rate of 100 µl min^{-1} for three minutes and guided over a column containing immobilized porcine pepsin kept at 12 °C. The resulting peptic peptides were collected on a trap column (2 mm x 2 cm) kept at 0.5 °C and filled with POROS 20 R2 material (Thermo Scientific). After three minutes, the trap column was placed in line with an ACQUITY UPLC BEH C18 1.7 µm 1.0 x 100 mm column (Waters) and the peptides eluted with a gradient of water + 0.1% (v/v) formic acid (eluent A) and acetonitrile + 0.1% (*v/v*) formic acid (eluent B) at 30 µl min^{-1} flow rate as follows: 0-7 min/95-65% A, 7-8 min/65-15% A, 8-10 min/15% A. Eluting peptides were guided to a Synapt G2-Si mass spectrometer (Waters) and ionized with by electrospray ionization (capillary temperature and spray voltage of 250 °C and 3.0 kV, respectively). Mass spectra were acquired over a range of 50 to 2000 m/z in enhanced high definition MS ($HDMS^E$) (*361, 362*) or high definition MS (HDMS) mode for undeuterated and deuterated samples, respectively. Lock mass correction was implemented conducted with [Glu1]-Fibrinopeptide B standard (Waters). During separation of the peptides on the ACQUITY UPLC BEH C18 column, the pepsin column was washed three times by injecting 80 µl of 0.5 M guanidine hydrochloride in 4 % (*v/v*) acetonitrile. Blank runs (injection of double-distilled water instead of the sample) were performed between each sample. All measurements were carried out in triplicate.

Peptides were identified and evaluated for their deuterium incorporation with software ProteinLynx Global SERVER 3.0.1 (PLGS) and DynamX 3.0 (both Waters). Peptides were identified with PLGS from the non-deuterated samples acquired with $HDMS^E$ employing low energy, elevated energy and intensity thresholds of 300, 100 and 1 000 counts, respectively and matched using a database containing the amino acid sequences of IDR, ARD, NTD, fl-Pac3, and F-BAR, porcine pepsin and their reversed sequences with search parameters as follows: Peptide tolerance = automatic; fragment tolerance = automatic; min fragment ion matches per peptide = 1; min fragment ion matches per protein = 7; min peptide matches per protein = 3; maximum hits to return = 20; maximum protein mass = 250 000; primary digest reagent = non-specific; missed cleavages = 0; false discovery rate = 100. For quantification of deuterium incorporation with DynamX, peptides had to fulfil the following criteria: Identification in at least 2 of the 3 non-deuterated samples; the minimum intensity of 10 000 counts; maximum length of 30 amino acids; minimum number of products of two; maximum mass error of 25 ppm; retention time tolerance of 0.5 minutes. All spectra were manually inspected and omitted if necessary, e.g. in case of low signal-to-noise ratio or the presence of overlapping peptides disallowing the correct assignment of the isotopic clusters.

2.17 Nuclear magnetic resonance spectroscopy

NMR spectroscopy is widely used to investigate the structure and dynamics of proteins and their complexes. Its advantage over other high-resolution methods such as X-ray crystallography or cryo-EM is the ability to study very flexible systems in solution under ambient conditions. Therefore, it has emerged as one of the primary tools in the research on intrinsically disordered proteins (IDPs) and is currently the only experimental method to obtain a per-residue resolution of IDPs (*314*). A drawback of NMR is the limitation to globular proteins of typically <30 kDa or IDPs of < 150 residues.

Theoretical introduction

NMR spectroscopy makes use of the magnetic spin of atomic nuclei as a direct reporter of local structure and dynamics in a protein. Protons and neutrons exhibit a magnetic spin which gives rise to a spin angular momentum and thus a nuclear dipole moment. Nuclei with an odd number of neutrons and protons exhibit an overall spin larger than zero, $I > 0$, a fundamental requirement for NMR. The currently most powerful NMR experiments for the study of biological macromolecules were developed for $I = ½$ systems. Naturally occurring nuclei with $I = ½$ in proteins are ^{1}H, ^{13}C, ^{15}N, ^{31}P. The natural abundance of ^{13}C and ^{15}N is, however, extremely low such that they need to be enriched in proteins to enable the detection of carbon and nitrogen (see section p. 69) (*363*).

Nuclear spins with $I = ½$ behave like bar magnets and interact with magnetic fields. In the absence of an external magnetic field, all spin orientations are energetically equivalent. In the presence of an external magnetic field B_0 (along the z-axis), the degeneracy splits into two states, a phenomenon called Zeeman splitting. The so-called $I = +½$ spin state describes a parallel orientation with respect to the external field (spin up, denoted α), and the $I = -½$ spin state, which an anti-parallel orientation (spin down, denoted β). The parallel orientation α is energetically favoured over the anti-parallel orientation β. In the magnetic field, the nuclear spins are slightly tilted away from the z-axis and precess around z by a frequency ω_0, the so-called Larmor frequency. This frequency increases with the gyromagnetic ratio γ of a nucleus, and the strength of B_0 (equation 20) (*364*).

$$\omega_0 = -\frac{1}{2\pi}\gamma B_0 \qquad 20$$

The gyromagnetic ratio is a nucleus specific property and describes the ratio of the magnetic moment of a spin to its angular momentum. The α and β precess in opposite directions with respect to their angular momentum. The energy separation between α and β is determined by the Larmor frequency as $\Delta E = h\,\omega_0$ (h is the Planck constant). The Larmor frequency of ^{1}H in a 18.8 T magnet is 800 MHz which translates to an energy difference of $7.6\ 10^{-5}$ kcal/mole which results in an extremely low occupation ratio $N_\beta/N_\alpha = 0.99$ according to the Boltzmann equation (equation 22). The low occupation ratio explains the extremely low sensitivity of NMR.

$$\frac{N_\beta}{N_\alpha} = \exp\left(\frac{\Delta E}{kT}\right) = \exp\left(\frac{h\omega_0}{kT}\right) \qquad 21$$

Here, N_i is the number of spins in state i, T is the temperature, and k is the Boltzmann constant. When nuclear spins are exposed to a radiofrequency pulse that is resonant to the Larmor frequency, they transition from the α to the β state. During the α to the β transition, the spins are flipped into the xy-plane before flipping further to $-z$. By adjusting the length of the pulse, the spins are stopped in the xy plane (90° pulse) as a phase-coherent bulk magnetization which continues to precess around z. The precession generates oscillating magnetization in x, y, $-x$, $-z$ which can be detected as an oscillating current by a receiver coil. Over time, the oscillating current decays due to longitudinal relaxation, which is the transition of the bulk magnetization back to the z-direction. Furthermore, the loss of phase coherence between the spins decreases the bulk magnetization in the xy-plane. The decaying oscillating current is called the free induction decay (FID). It oscillates at the Larmor frequency of the observed nucleus, and its decay is determined primarily by translational relaxation in proteins. Fourier transformation of the FID gives a typical NMR spectrum with a peak at the position of the Larmor frequency and a line width of $1/\pi T_2$, where T_2 is the time constant of translational relaxation. T_2 decreases with increasing size of a molecule such that for proteins >30 kDa, the line width gets extremely broad (*364*).

In a molecule, the electrons of neighbouring atoms induce minuscule magnetic fields which slightly shift the Larmor frequency of a nucleus (in the sub-ppm range). Nuclei with identical magnetic environments

will have identical Larmor frequencies. Conversely, nuclei with different magnetic environments will resonate at different frequencies. The resulting FID of a molecule contains the frequencies of all atoms such that the Fourier transformation yields a spectrum with multiple peaks, each corresponding to a frequency of a nucleus in the molecule. In a simple onedimensional ^{1}H NMR experiment, a sample is allowed to equilibrate in a preparation phase and then irradiated with a 90° RF-pulse at the ^{1}H Larmor frequency. This pulse transitions all ^{1}H spins in the molecule to the *xy*-plane. The FID is recorded directly afterwards for a time t_1 which needs to be as long as it takes the spins to return to equilibrium. This sequence is repeated multiple times to achieve a high signal-to-noise ratio. A given spin *i* in the molecule will precess in the *xy*-plane at frequency ω_i and translationally relax with a time constant $T_{2,i}$. The recorded FID will be a superposition of the FID of each spin. After Fourier transformation, the NMR spectrum contains a peak for each spin at frequency ω_i with a line width $1/\pi T_{2,i}$. The frequencies in an NMR spectrum are normalized to a reference frequency such it becomes independent of the magnetic field strength. It is called the chemical shift δ (in ppm). Being directly dependent on the local environment of a nucleus, the chemical shift and the line with are essential parameters for the structural and dynamic analysis of molecules (*364*).

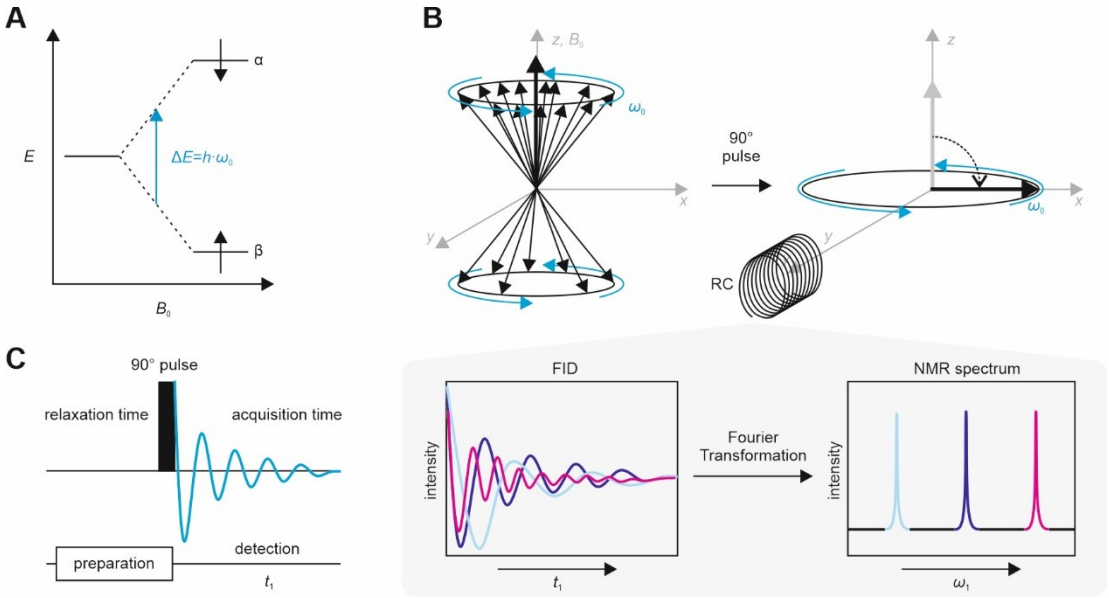


Figure 27: Basic principle of nuclear magnetic resonance. (**A**) Zeeman splitting for a nucleus with spin I = ½: In the absence of an external magnetic field, the orientation of nuclear spins is degenerate. With increasing strength of a magnetic field B_0, the energy levels split into the energetically favoured α state (I = ½) with spins oriented in the direction of B_0, and the β state (I = -½) with an antiparallel orientation. The energy difference between the α and β, ΔE, is proportional to the Larmor frequency ω_0. (**B**) Vektor model of a spin I = ½ nucleus. The spins in α and β rotate around the direction of B_0 at the Larmor frequency ω_0. At equilibrium, the α state is slightly more populated than the β state which results in a bulk magnetization (bold arrow) in the *z*-direction, i.e. into the direction of the magnetic field B_0. A 90° radio frequency (RF) pulse along the *x*-axis flips the bulk magnetization around the *y*-axis into the *xy*-plane, creating a phase-coherent bulk magnetization, which continues to precess around the *z*-axis. The oscillating magnetization along the *y*-axis is detected by a receiver coil (RC) as an oscillating current. Due to several translational and longitudinal relaxation, the magnetization in the *xy*-plane decays, thus leading to a decay in the current (free induction decay, FID). The FID (intensity vs time) is Fourier transformed into the NMR spectrum (intensity vs frequency) where each frequency in the FID gives rise to a peak with an intensity proportional to the number of spins and the line width $1/\pi T_2$. (**C**) Standard pulse sequence for one-dimensional NMR experiments. In a preparation phase, the system is allowed to relax to equilibrium with magnetization in the *z*-direction. A 90° RF pulse creates magnetization in the *xy*-plane which is detected as a free induction decay (FID) for the time t_1 by a receiver coil. After t_1, the pulse sequence can be repeated.

Protein NMR spectra

Unlike small organic compounds, proteins are built up of thousands of atoms, and their ^{1}H NMR spectra display a dramatic peak overlap. Multidimensional NMR experiments have been developed which enable to detect the nuclei of specific functional groups, thus limiting the number of detected peaks and increasing the spectral resolution by a second frequency domain. Multidimensional NMR experiments are based on the ability to transfer magnetization between nuclei which are either connect through a covalent bond (*J*-coupling) or are proximal in space (dipolar coupling). When magnetization is transferred from nucleus 1 to nucleus 2, nucleus 2 will resonate at a Larmor frequency modulated by the Larmor frequency of nucleus 1. Thus, the FID of nucleus 2 will carry the frequency of nucleus 1 as a label. The pulse sequence of a 2D experiment consists of a preparation time upon which an RF pulse transitions the spins of nucleus 1 to the *xy*-plane. In the evolution phase, the nuclei precess freely in the *xy*-plane for time t_1. In the subsequent mixing phase, the magnetization at t_1 is transferred to nucleus 2, and the FID is recorded for an acquisition time t_2. The FID as a function of t_2 is recorded at varying t_1, thus yielding a 2D FID, which is then transformed into a 2D spectrum with the frequencies ω_1 and ω_2 (*363*).

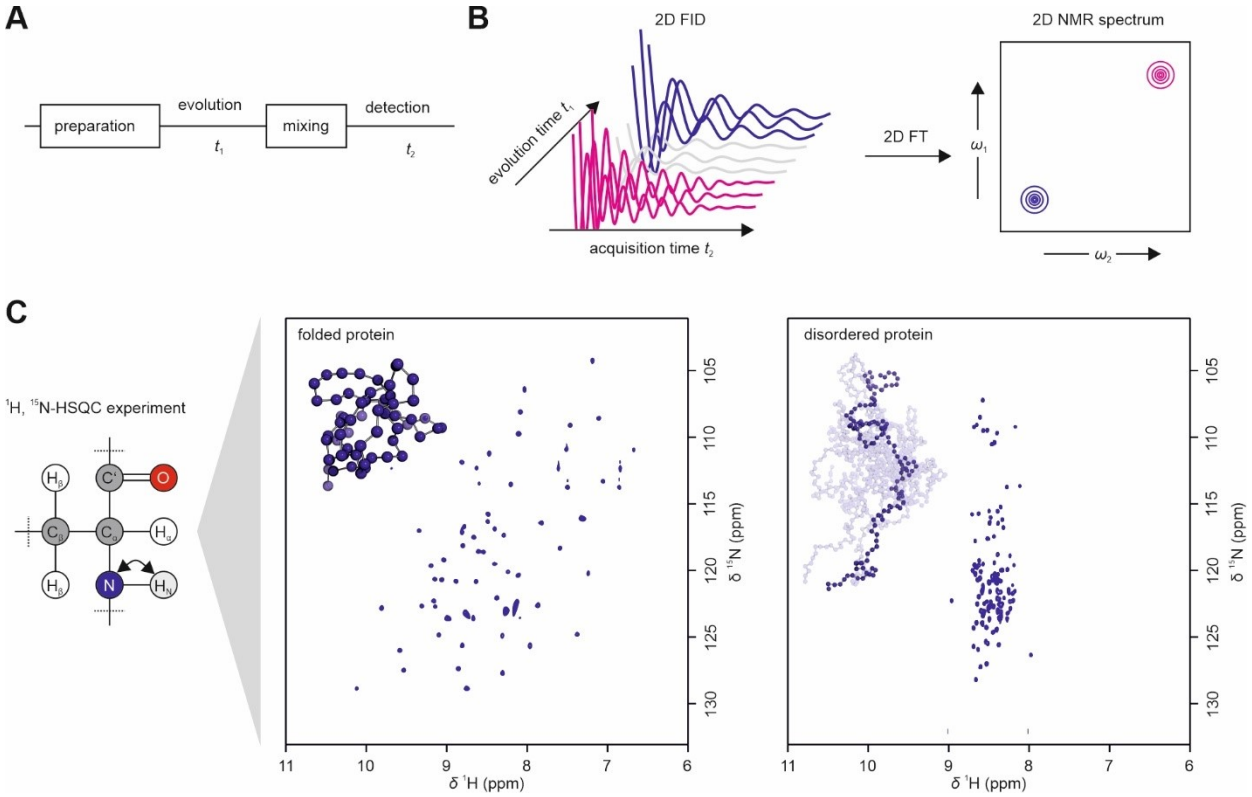


Figure 28: The NMR fingerprint of a protein. Basic scheme of a two-dimensional NMR pulse sequence. In the preparation step, spins are allowed to relax to equilibrium magnetization in z-direction before being flipped to the *xy*-plane. During evolution time, the spins precess freely for the time t_1. In the mixing phase, magnetization is transferred to another nucleus and then detected for the time t_2. The pulse sequence is repeated at specific time intervals of t_1. (**B**) The FID in a 2D experiment is recorded as a function of t_2 at different times t_1, thus yielding a 2D FID. The 2D Fourier transformation yields a 2D NMR spectrum with the chemical shifts of the directly observed nucleus in frequency domain ω_2 and the chemical shifts of the indirectly observed nucleus in frequency domain ω_1. (**C**) The ^{1}H, ^{15}N-HSQC spectrum can be regarded as a protein's fingerprint. It detects amide resonance through a magnetization transfer from ^{1}H to ^{15}N and back. Each peak in the ^{1}H, ^{15}N-HSQC spectrum will correspond to the amide group in a protein (mostly backbone amide groups). In folded proteins, amide groups participate in hydrogen-bonding, thus leading to structural diversity in their local environment. The results are well-dispersed spectra (left spectrum). Disordered proteins have little structural features, and all amide groups are similarly exposed to the solvent. The local environment is similar for each amide group, thus leading to a very narrow peak dispersion.

In proteins, one of the most useful functional groups for NMR studies is the amide group. It can be detected in the so-called ^{1}H, ^{15}N-heteronuclear single quantum coherence (HSQC) experiment. In a ^{1}H, ^{15}N-HSQC experiment magnetization is transferred from amide ^{1}H nuclei to their neighbouring amide ^{15}N nuclei and back. In the final FID, the ^{15}N-modulated frequencies of the ^{1}H nuclei are detected, and the 2D Fourier transformation yields a spectrum with the chemical shifts of ^{1}H in the direct frequency domain (ω_2) and the chemical shifts of ^{15}N in the indirect frequency domain (ω_1). Each cross-peak in the ^{1}H, ^{15}N-HSQC NMR spectrum corresponds to one amide group in the protein. Because every amino acid in a protein has one amide group in the peptide bond (some also in side chains), the ^{1}H, ^{15}N-HSQC spectrum will have at least as many peaks as amino acids in the protein. Only proline residues do not give a peak as they do not have a proton on the nitrogen atom, and the amide resonance of the first residue is not detected due to solvent exchange of the proton. Each amide group in a given protein has a unique structural surrounding and therefore is also magnetically unique. As a consequence, the amide groups in a protein resonate at different frequencies and their peaks disperse in the HSQC. Particularly in folded proteins, where different secondary structure elements lead to hydrogen-bonding, the ^{1}H, ^{15}N-HSQC spectra show broad and specific peak dispersion patterns. Therefore, the ^{1}H, ^{15}N-HSQC spectrum is also called the fingerprint spectrum of a protein (*363*).

The broad peak dispersion in ^{1}H, ^{15}N-HSQC spectra of folded proteins arises from the combination of a unique amino acid sequence with a unique three-dimensional fold. The less structural diversity in a protein chain, the less dispersed is its ^{1}H, ^{15}N-HSQC spectrum. IDPs exhibit very little structural features, and all of their amide groups are similarly exposed to the solvent. Therefore, the chemical shifts of their amide groups depend almost entirely on the type of residue and its direct neighbour residues, i.e. the amino acid sequence. This results in very narrow peak dispersions with substantial peak overlap in the ^{1}H dimension (*313*).

Proteins are dynamic molecules and display structural fluctuations on many time scales. The rate constants of interconversion between different conformations (k) and their chemical shift differences ($\Delta\delta$) determine whether a protein gives a single set of peaks or multiple peak sets in an NMR spectrum. Local fluctuations of the protein backbone or side chains usually occur at time scales faster than the temporal resolution of NMR. When $k >> \Delta\delta$ (fast exchange), the detected peak of a nucleus which fluctuates rapidly between different conformations will give a single peak that represents the population-averaged chemical shift of all amide conformations (Figure 29 A) (*365*). In many folded proteins, the displacement of amide groups caused by local fluctuations is comparably small, and amide groups mostly give rise to a single peak. The set of ^{1}H, ^{15}N-HSQC peaks of a folded protein is representative of a preferred conformation around which the protein fluctuates. In an entirely disordered protein where no secondary structures keep the protein backbone in a preferred conformation, the backbone fluctuations are substantially increased compared to folded proteins, both in terms of amplitude and time scale. Therefore, an IDP cannot be described by a single structure around which the protein fluctuates, but rather by an ensemble of structures. The large amplitude of the amide group fluctuations is compensated by the fast time scales such that IDPs mostly display a single set of peaks in the ^{1}H, ^{15}N-HSQC spectrum (*313, 314*).

When the rate of interconversion between two or more conformations in a protein is in a similar rate as the chemical shift difference between both conformations, i.e. $k \approx \Delta\delta$ (fast exchange), the peaks broaden (Figure 29 A). This phenomenon significantly decreases spectra quality such that proteins which exchange at intermediate rates can be difficult to study and may need to be stabilized, e.g. by changing the temperature or adding a ligand. At slow exchange rates ($k << \Delta\delta$), each conformation will give rise to an own set of peaks in the NMR spectrum. Importantly, proteins can display mixtures of intermediate and fast exchange or intermediate and low exchange. Exchange rates can also vary between different regions of a protein (*365*).

The time scales of interconversion between different states play a substantial role in chemical shift perturbation (CSP) experiments where spectra of a protein are recorded as a function of a ligand concentration. The binding of a ligand causes conformational changes or changes the local structure in the ligand binding site of the protein and thus perturbs the chemical shifts of the protein. This can manifest in the NMR spectra as peak shifts, line

broadening, or the splitting of one peak into two peaks. When the exchange rate of the protein between the free (apo) and the ligand-bound (complex) form is in a fast regime ($k << \Delta\delta$), a peak at the chemical shift of the apo state will shift gradually to the chemical shift of the complex. This behaviour is usually the case for dissociation constants in the mid-µM regime. At $k \approx \Delta\delta$ (fast exchange), titration of the ligand will cause the peak to shift from the apo to the complex chemical shift while undergoing line broadening. When a complex has a dissociation constant in the mid-nM regime or lower, the exchange will be slow ($k >> \Delta\delta$) and the apo peak will decrease whereas the complex peak increases with the ligand concentration. The CSP can be plotted against the ligand concentration, either as the build-up/decrease of the complex/apo peak in slow exchange or as the chemical shift displacement ΔCS from the apo chemical shift in fast exchange. The resulting binding curves can be used to determine the dissociation constant of a complex. CSP experiments are a powerful tool to map the binding sites of ligands in proteins because CSP will be most substantial for peaks corresponding the residues in the ligand binding site (*366*).

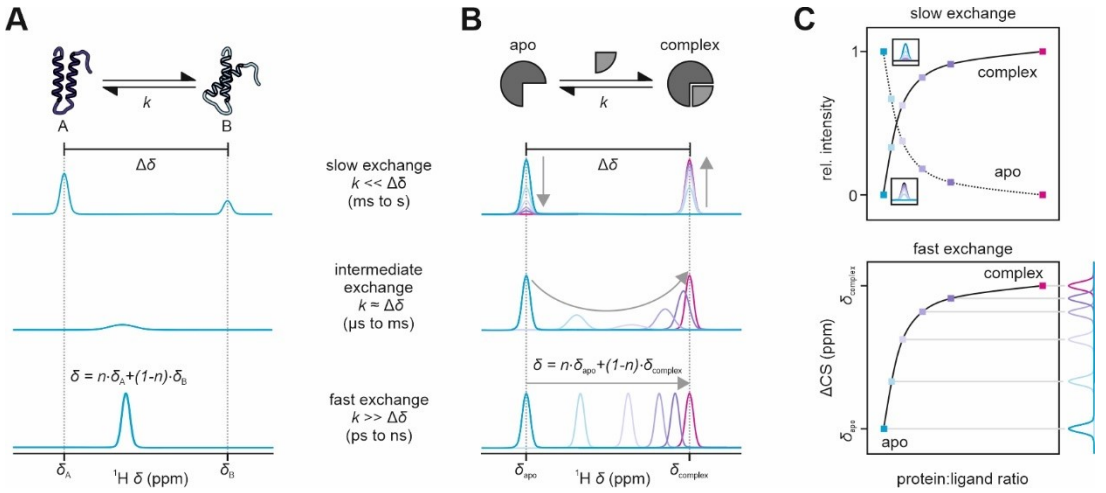


Figure 29: The effect of chemical exchange on NMR spectra. A protein which exchanges between two conformations, A and B, will produce different chemical shift and line width depending on the exchange rate k and the chemical shift difference between state A and B, $\Delta\delta$. At $k >> \Delta\delta$ (fast exchange), the peak position will be at the population-averaged of the A and B chemical shifts (δ_A and δ_B), given by $\delta = n \quad \delta_A + (1\text{-}n) \quad \delta_B$ (with n being the fraction of state A). At $k \approx \Delta\delta$ (intermediate exchange). The signal is also at a population-averaged chemical shift and additionally broadened. At $k << \Delta\delta$ (slow exchange), each state will give rise to a peak at δ_A and δ_B, respectively, whereas the peak intensity ratio between δ_A and δ_B reflects the population ratio. (**B**) In chemical shift perturbation (CSP) experiments, the interaction of a protein with a ligand gives rise to conformational exchange between an apo state (no ligand) and a complex (ligand-bound). During a ligand titration, the peak will shift from δ_{apo} to $\delta_{complex}$ when $k >> \Delta\delta$ (fast exchange). At $k \approx \Delta\delta$ (intermediate exchange), the signal will shift and broaden. At $k << \Delta\delta$ (slow exchange), ligand addition will decrease the peak intensity at δ_{apo} in favour of the peak intensity a $\delta_{complex}$. (**C**) Chemical shift perturbation can be mapped against the ligand concentration, either as rel. intensity decrease/increase or as chemical shift change (ΔCS). This can be used to calculate dissociation constants of protein-ligand complexes.

Backbone resonance assignment

Exploiting the full potential of a protein's ^{1}H, ^{15}N-HSQC spectrum requires a backbone resonance assignment, i.e. one needs to assign each amide resonance to its corresponding amino acid in the protein. It is the prerequisite to correlate structural and dynamic features encoded in ^{1}H, ^{15}N-HSQC spectra to their position in the protein sequence of threedimensional structure. Several so-called triple resonance experiments have been developed in which magnetization is transferred between ^{1}H, ^{15}N, and ^{13}C nuclei. They complement the HSQC experiment by a third frequency domain, typically the chemical shift of the carbonyl (C') or the C_α and C_β nuclei. Magnetization is transferred from ^{1}H to ^{15}N and then to ^{13}C, and the chemical shifts of all nuclei are detected. The 3D spectrum reflects the ^{1}H, ^{15}N-HSQC spectrum extended by a third ^{13}C dimension. The special feature of triple resonance experiments is that magnetization is not only transferred between nuclei within an amino acid but also between amino acid neighbours in the protein sequence (mostly to a predecessor). The 3D peaks

can be visualized as strips along the thrid the third dimension for each peak in the HSQC spectrum. A strip of a given amide resonance $^{1}H_{i}$-$^{15}N_{i}$ displays the chemical shift of the $^{13}C_{i}$ nucleus from the same amino acid *i* as the amide resonance, but also the $^{13}C_{i-1}$ chemical shift of the amino acid neighbour *i*-1. For a $^{13}C_{i-1}$ peak of amino acid *i*, an amide peak corresponding to amino acid *i-1* should exist in the HSQC spectrum which has a $^{13}C_{i}$ resonance at the exact same frequency. The $^{13}C_{i}$ /$^{13}C_{i-1}$ pairs are identified in an iterative approach by aligning the peak strips. The final goal is to identify a chain of amide resonances which are connected through $^{13}C_{i}$ /$^{13}C_{i-1}$ pairs. The typical experiments used for backbone assignments are HNCA, HNCACB, and HN(CA)CO. In HNCA, the C_{α} resonances are correlated, in HNCAB the C_{α} and C_{β} resonances, and in HN(CA)CO the C' resonance. In most cases, an unambiguous identification of amide resonance chains is obtained via $^{13}C_{\alpha,\,i}$ /$^{13}C_{\alpha,\,i-1}$, $^{13}C_{\beta,\,i}$ /$^{13}C_{\beta,\,i-1}$, and $^{13}C'_{i}$ /$^{13}C'_{i-1}$ pairs. 'Anchor' residues with specific chemical shift features, e.g. glycine, serine/threonine/alanine, enable to link the amide resonance chains to a sequence in the protein. Glycine can be identified due to a missing C_{β} resonance, serine/threonine have C_{β} resonances at high frequencies or alanine with C_{β} resonances at low frequencies. A proline in residue in the protein sequence interrupts a chain of linked amide resonances because it has no peak in the ^{1}H, ^{15}N-HSQC spectrum.

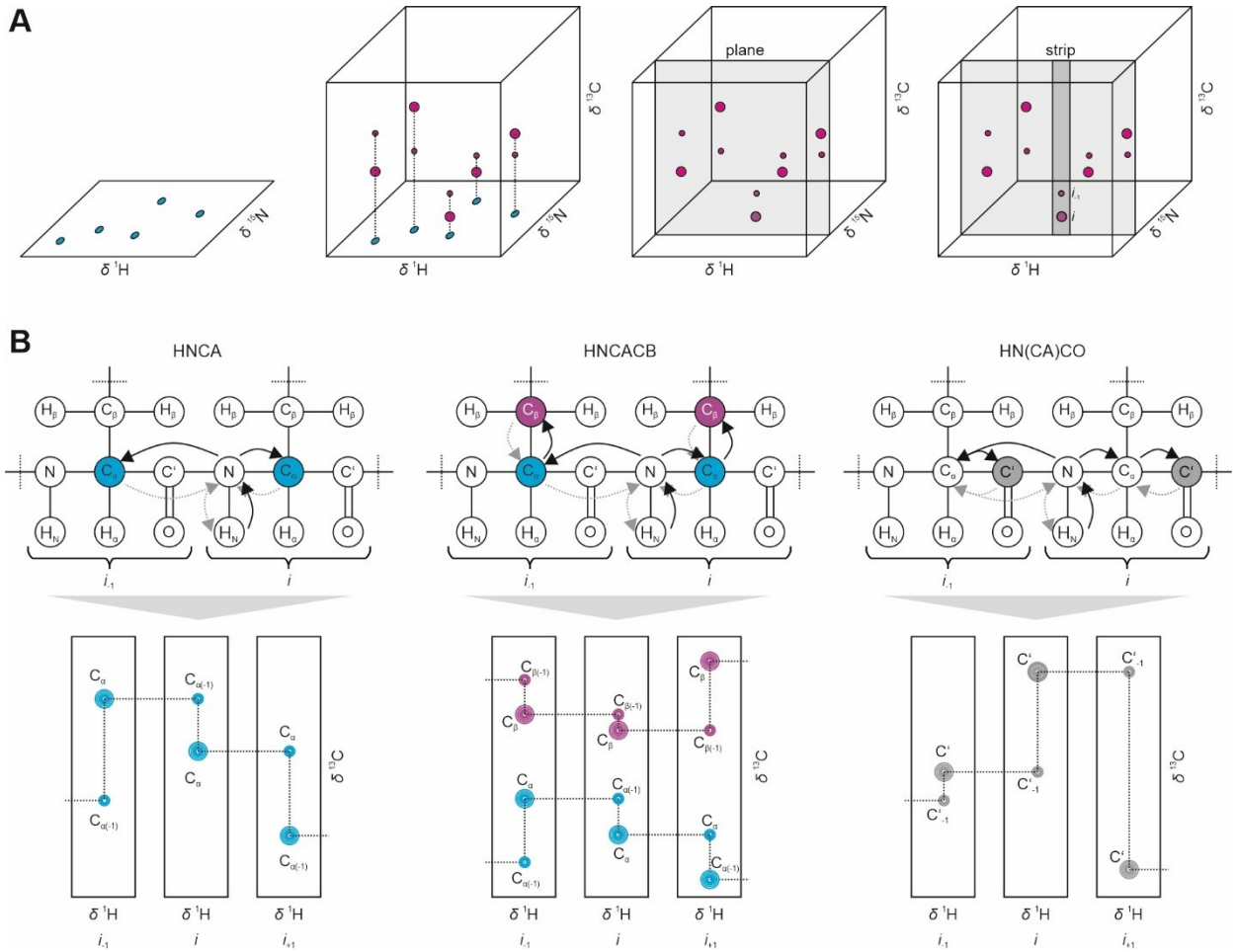


Figure 30: Backbone NMR assignment. (**A**) Basic principle of a triple-resonance experiment for NMR protein backbone assignments. Triple-resonance experiments are based on a ^{1}H, ^{15}N-HSQC experiment extended by a third dimension in which C', C_{α} and C_{β} and C'$_{-1}$, $C_{(\alpha-1)}$ and $C_{\beta(-1)}$ resonances are correlated to the amide resonances. The 3D spectrum is scanned along the ^{15}N frequency in ^{1}H, ^{13}C planes. The strip at each amide resonance contains its correlated ^{13}C resonances. The strips are searched $^{13}C_{i}$ /$^{13}C_{i-1}$ to find the connectivities between the amide resonances and correlated them to the protein sequence. (**B**) In the HNCA experiment, magnetization is transferred from the backbone amide group of a residue *i* to the $C_{\alpha,\,i}$ and $C_{\alpha,\,i-1}$ nuclei. In the peak strips, resonances of neighbouring residues have matching $C_{\alpha,\,i}$ and $C_{\alpha,\,i-1}$ resonances. (**C**) In the HNCACB experiment, magnetization is additionally transferred to the C_{β} and $C_{\beta,\,i-1}$ nuclei. The resonances of neighbouring residues can therefore be identified additionally by C_{β}/$C_{\beta,\,i-1}$ matches. (**D**) The HNCACO experiment transfers magnetization to the C'_{i} and C'_{i-1} nuclei. This enables assignments via C'_{i}/C'_{i-1} matches.

Several additional triple-resonance experiments exist and can be used when HNCA, HNCACB, and HN(CA)CO experiments are insufficient for an unambiguous resonance assignment. Moreover, the selective isotope-labelling of individual residues can significantly improve the assignment process. With a complete resonance assignment, the ^{1}H, ^{15}N-HSQC spectrum provides residue-specific structural and dynamic information on a protein an is the basis for more extensive investigations. In addition to backbone resonance assignments, the side chain resonances can also be assigned via multidimensional NMR spectra. This process becomes relevant for the NMR-based structure determination of a protein (*367*).

NMR structure determination

The magnetization transfer between nuclei of covalently linked atoms is based on a mechanism called *J*-coupling, which describes a hyperfine interaction between nuclei and local electrons. Nuclei can also be coupled through space via dipole-dipole interactions (dipolar coupling). The strength of a dipolar interaction depends on the distance between to nuclei and is observable between spins separated by up to 4-5 Å. Magnetization transfer can be achieved through the so-called nuclear Overhauser effect (NOE). This can be illustrated by a simple two spin system of nuclei A and X which are spatially close but not *J*-coupled. A and X can exist in the energetically favoured α state and in the unfavored β state, thus giving four states $\alpha_A\alpha_X$, $\alpha_A\beta_X$, $\beta_A\alpha_X$, and $\beta_A\beta_X$. When the spins of nucleus A are saturated by an RF pulse, the α_A state and the β_A state become equally populated ($N_{\beta,A}/N_{\alpha,A} = 1$). Two different relaxation processes can now occur and return spin A from state β_A to state α_A. The so-called double-quantum transition (W_2) describes the relaxation from $\beta_A\beta_X$ to $\alpha_A\alpha_X$ (two spins are flipped), whereas the so-called zero-quantum transition (W_0) describes the relaxation from $\beta_A\alpha_X$ to $\alpha_A\beta_X$. The W_2 transition depopulates the β_X state and increases the population difference, thus leading to an enhancement of the NMR signal of spin X. By contrast, the W_0 transition populates the β_X state and decreases the population ratio $N_{\beta,X}/N_{\alpha,X}$, thus weakening the NMR signal of spin X. The contribution of W_2 and W_0 depends on the correlation time τ_c (molecular tumbling). For small organic molecules, W_2 is the dominating relaxation mechanism, whereas for proteins, W_0 dominates. The NOE-induced signal change (V_{NOE}) is quantified as the deviation of the signal intensity with saturation (I_s) from the signal intensity without saturation (I_0) normalized to I_0, $V_{NOE} = (I_s\text{-}I_0)/I_0$.

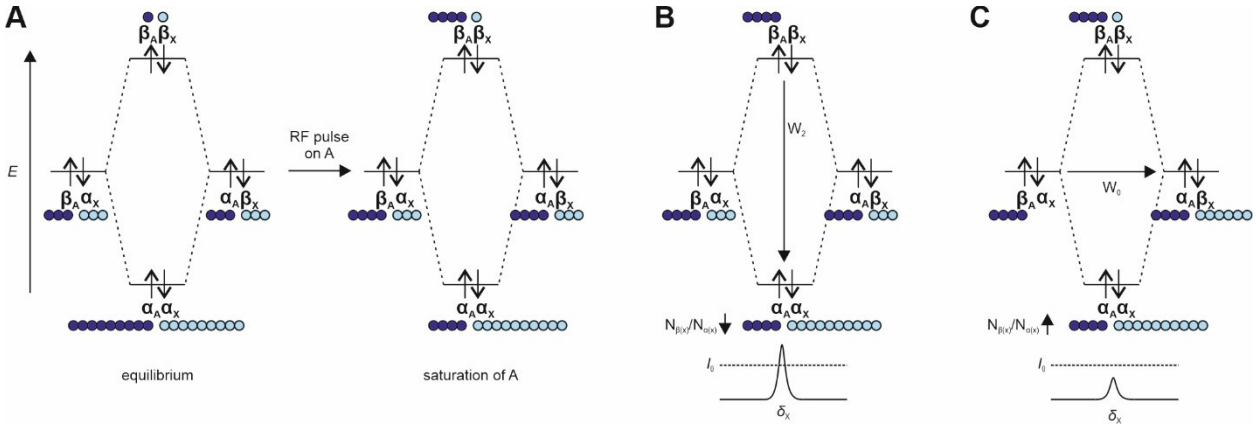


Figure 31: Principle of the nuclear Overhauser effect. Shown is a two-spin system of with nuclei A and X which are unaffected by *J*-coupling. The circles indicate the population of the four states $\alpha_A\alpha_X$, $\alpha_A\beta_X$, $\beta_A\alpha_X$, and $\beta_A\beta_X$. The α state is energetically favoured over the β state. (**A**) An RF pulse on spins A establishes equally populated α_A and β_X states (saturation). The return of spins A to equilibrium can proceed via two pathways which lead to different population ratios of the X spins. (**B**) The double-quantum transition (W_2) describes the return of spins A from the $\beta_A\beta_X$ state to the $\alpha_A\alpha_X$ state. This depopulates the β_X states and increases the population difference of the X spins (here 3β : 13α vs. 4β : 12α at equilibrium). The result is a signal enhancement of spin X with respect to the signal without saturation of A (I_0). This relaxation mechanism is dominant in small organic molecules. (**C**) The zero-quantum transition (W_2) describes the return of spins A from the $\beta_A\alpha_X$ state to the $\alpha_A\beta_X$ state. This populates the β_X states and decreases the population difference of the X spins (here 6β : 12α vs. 4β : 12α at equilibrium). The result is a signal loss of spin X relative to I_0. This relaxation mechanism is dominant in proteins.

Due to its dependence on the molecular tumbling of the investigated protein, the NOE-effect can be used to investigate the dynamics of a protein. In the so-called {^{1}H}, ^{15}N-hetNOE experiment, ^{1}H saturation is implemented into the ^{1}H, ^{15}N-HSQC experiment. The mobility of amide groups reflects in a decrease of the peak intensity in the ^{1}H, ^{15}N-HSQC spectrum. The more flexible an amide group, the weaker is its signal upon proton saturation relative to the signal without saturation (*368*).

The NOE-effect is also dependent on the distance between the dipolar coupled nuclei by $V_{NOE} \sim c/r^6$, where c is a calibration constant and r ist the distance. Based on this relationship, the NOE-effect provides the basis for NMR structure determination of proteins. The NOE effect can be implemented into 3D HSQC experiments called ^{15}N-NOESY-HSQC or ^{13}C-NOESY-HSQC. At the start of these experiments, magnetization is exchanged between all hydrogen nuclei via the NOE-effect. Subsequently, the magnetization is transferred to an amide ^{15}N and then ^{1}H nucleus for detection (^{15}N-NOESY-HSQC) or to a ^{13}C nucleus and back to ^{1}H for detection (^{13}C-NOESY-HSQC). This yields ^{1}H, ^{15}N- or ^{1}H, ^{13}C-HSQC spectra which are extended by a third ^{1}H frequency domain. Each peak in the HSQC spectra corresponds to an amide or a methyl group resonance, and their strips contain the peaks corresponding to spatially close ^{1}H nuclei. The signal intensity of the peaks in the strips encodes the distance information. With the collected NOE-signals and the extracted ^{1}H-^{1}H distance restraints, it is possible to calculate the 3D structure of the investigated protein. This, however, precludes high completeness of the backbone and resonance assignments to correlated the NOE-signals to the distance between the respective atoms (*363*).

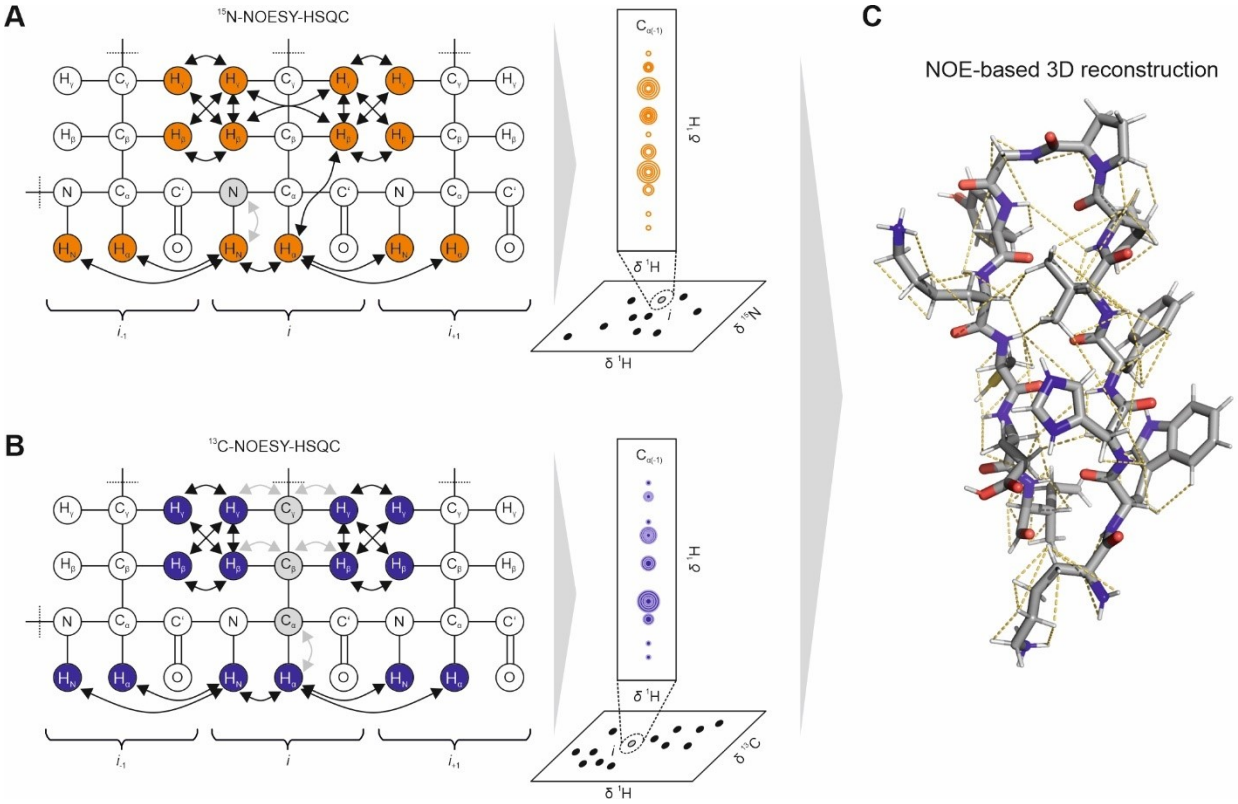


Figure 32: NOE-based structure determination. (**A**) ^{15}N-NOESY-HSQC experiment: Magnetization is exchanged between all hydrogen nuclei via the NOE-effect. The magnetization is then transferred to the amide ^{15}N and subsequently to the amide ^{1}H nucleus for detection. The 3D spectrum can be regarded as a ^{1}H, ^{15}N-HSQC spectrum extended by a third ^{1}H frequency domain. The strip of each amide resonance contains the peaks corresponding to spatially close ^{1}H nuclei, and the peak intensities encode the ^{1}H-^{1}H distances. (**B**) ^{13}C-NOESY-HSQC experiment: As in (A), magnetization is exchanged between all hydrogen nuclei via the NOE-effect. Magnetization is then transferred to a ^{13}C nucleus and back to ^{1}H for detection. The 3D spectrum is a ^{1}H, ^{13}C-HSQC spectrum in which each peak corresponds to a methyl group. This HSQC is extended by a third ^{1}H frequency domain such that each strip contains the peaks corresponding to spatially close ^{1}H nuclei and the peak intensities encode their distances. (**C**) The ^{1}H-^{1}H distance restraints obtained from the spectra in (A) and (B) are used to reconstruct the three-dimensional structure of the protein.

Isotope-labeling of proteins for NMR experiments

Most multi-dimensional NMR spectroscopic experiments for structural and dynamic investigations of proteins are based on magnetization transfer between hydrogen, nitrogen, and carbon nuclei with spin ½. The spin ½ nuclei in proteins are ^{1}H, ^{15}N, and ^{13}C. However, whereas ^{1}H has a natural abundance of 99.98%, the natural abundance of ^{13}C (1.1%) and ^{15}N (0.04%) is too low to for efficient magnetization transfers in multi-dimensional NMR experiments. Therefore, proteins need to be enriched in ^{15}N and/or ^{13}C. This can be achieved through protein expression in bacteria grown in a minimal medium supplemented with ^{15}N and ^{13}C sources. The growth medium must contain all essential trace elements and salts needed by the bacteria but must not contain ^{14}N or ^{12}C sources. In most standard NMR applications, it is desirable to have uniformly ^{15}N- or ^{15}N, ^{13}C-labeled proteins. By supplementing a minimal medium with ^{15}N-ammonium chloride and ^{13}C-glucose bacteria will synthesize uniformly ^{15}N- and ^{13}C, ^{15}N-labeled amino acids during growth and incorporate them into their synthesized proteins (*369*). The most commonly used medium for the generation of isotope-labelled proteins is M9 minimal medium (Table 5).

All isotope-labelled proteins used in this thesis were generated by expression in *E. coli* grown in isotope-supplemented M9 medium. For uniform ^{15}N-labeling, the medium was supplemented with ^{15}N-ammonium chloride and ^{12}C-glucose. For uniform ^{13}C-labeling, the ^{12}C-glucose was replaced by ^{13}C-glucose (see Table 5). A spin down protocol was used in which *E. coli* cells were first grown in LB medium to an OD_{600} ~0.4 and then centrifuged (room temperature, 5 000 g, 10 min). The LB medium was discarded, and the cells were resuspended in freshly prepared M9 medium supplemented with an appropriate antibiotic. The cells were further grown to the desired $OD_{600,}$ and protein expression was induced with IPTG. All subsequent steps follow the same protocols as for non-labelled proteins.

Table 5: Composition of M9 medium for ^{15}N- and ^{13}C-isotope labelling of proteins.

Component	Ingredients (per L medium)
Salts	15.0 g KH_2PO_4 33.9 g Na_2HPO_4 2.5 g NaCl 2 mM $MgCl_2$ 10 µM $FeCl_3$
Isotopes	0.75 g ^{15}N-NH_4Cl 2.00 g ^{13}C-glucose (or 4.00 g ^{12}C-glucose for only ^{15}N-labeling)
Magic Mix	1 mL of Centrum® vitamin tablet dissolved in 20 mL ddH_2O
Trace elements	2 mg $CaCl_2$-dihydrate 2 mg $ZnSO_4$-heptahydrate 2 mg $MnSO_4$-monohydrate 1 mg biotin 50 mg thiamine and niacin

Instrumentation and software

NMR measurements were carried out on Bruker 600, 700, 800, 900 and 950 MHz spectrometers equipped with cryogenic triple resonance probes (at the Centre for Biomolecular Magnetic Resonance, BMRZ, Goethe-University Frankfurt). The proton chemical shifts of ^{13}C, ^{15}N-labeled Pacsin3 SH3 domain were externally (apo) and internally (in complex with PRR) referenced to 2.2-dimethyl-2-silapentane-5-sulfonic acid (DSS) and the heteronuclear ^{13}C and ^{15}N chemical shifts were indirectly referenced with the appropriate conversion factors. All spectra were processed using TopSpin 2.1, 3.2, or 4.1 (Bruker) and analyzed using the programs CARA (*370*) and CcpNmr Analysis 2.2 (*371*).

Backbone NMR assignments

Backbone NMR assignments of Pacsin1 and 2 SH3 domains: Backbone resonance assignments of ^{13}C, ^{15}N-labelled Pacsin1 and 2 SH3 domains were obtained from HNCO, HN(CA)CO, HNCA, HN(CO)CA and HNCACB spectra recorded with standard Bruker pulse sequences. The buffers contained 10 mM Tris pH 7, 100 mM NaCl, 10% D_2O. All spectra were recorded at 298 K. The backbone chemical shifts were assigned to 66% for the Pacsin1 SH3 domain and to 68% for the Pacsin2 SH3 domain (no Ha assignments). The backbone amide resonances of the native protein sequence were completely assigned. Missing amide resonance assignments correspond to residues of the artificial His_6-tag in the C-terminus.

Backbone NMR assignments of apo Pacsin3 SH3 domain and in complex with PRR: Backbone resonance assignments of ^{13}C, ^{15}N-labelled Pacsin3 SH3 domains in the absence of the TRPV4 PRR peptide were carried out using HNCO, HN(CA)CO, HNCA, HN(CO)CA and HNCACB spectra recorded with standard Bruker pulse sequences. The buffers contained 10 mM Tris pH 7, 100 mM NaCl, 10% D_2O. All spectra were recorded at 298 K. The backbones assignments from the apo PACSIN3 SH3 domain could be transferred to the ^{13}C, ^{15}N-labelled chicken Pacsin3 SH3 domain in the presence of 5-fold unlabeled TRPV4-PRR peptide (V131I, see results section VI3.6) by using chemical shift perturbation experiments. The buffers contained 10 mM Tris pH 7, 100 mM NaCl, 10% D_2O. All spectra were recorded at 298 K. and were verified by (H)C(CO)NH (mixing time 25 ms) and H(CCO)NH (mixing time 25 ms) experiments used for the side chain assignment. Additional triple-resonance experiments to assign the side chains were ^{13}C-NOESY-HSQC (mixing time 150 ms, aliphatic and aromatic carbons) and ^{15}N-NOESY-HSQC (mixing time 150 ms) experiments (*367*). This approach yielded a nearly complete backbone assignment of the Pacsin3 SH3 domain bound to the TRPV4-PRR peptide. The assignment of the aliphatic side chain protons and carbons was completed to 97 and 75%, respectively. The aromatic protons and carbons were assigned with 64%. (Assignments were carried out by Dr. Ute A. Hellmich and Dr. Nina A. Glogowski)

Assignment of the ^1H resonances of the unlabeled TRPV4 PRR peptide in complex with the ^{13}C, ^{15}N-labeled Pacsin3 SH3 domain was achieved with ^{13}C-filtered TOCSY (mixing time 60 ms) (*372, 373*) and NOESY (mixing time 120 ms) (*374–376*) experiments. Additionally, ^{13}C- or ^{15}N-HSQC and 2D ^{13}C-NOESY-HSQC (mixing time 200 ms) spectra of selectively labelled ^{13}C, ^{15}N TRPV4 PRR peptides (see Table S2) in the presence of unlabeled Pacsin3 SH3 domain were used to support assignment and verify overlapping resonances. The assignment of the ^1H resonances of the bound TRPV4 PRR peptide was completed to 96%. (Assignments were carried out by Dr. Ute A. Hellmich and Dr. Nina A. Glogowski)

Backbone NMR assignments of TRPV4 IDR constructs: Backbone resonance assignments of ^{13}C, ^{15}N-labeled *G. gallus* and *H. sapiens* TRPV4 IDR constructs were carried out using HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HAHB(CO)NH and ^1H, ^{13}C-HSQC spectra recorded with standard Bruker pulse sequences. The spectra were recorded at 298 K in 20 mM NaP_i pH 4.5, 150 mM NaCl, 10% D_2O. All amide resonances besides the first residue A1 could be assigned for the *G. gallus* TRPV4 IDR. The overall backbone resonance shift assignments (N, H_N, Hα, Cα, C') were completed to 96%. For the *H. sapiens* TRPV4 IDR, no resonances corresponding to the amide group of His117 could be identified. All other backbone amide resonances could be assigned to their corresponding amino acids. The overall backbone resonance assignments could be completed to 92%.

Backbone NMR assignments of the TRPV4 PBD-PRR peptide: Backbone resonance assignments of ^{13}C, ^{15}N-labeled *G. gallus* TRPV4 PBD-PRR were carried out using HNCO, HN(CA)CO, HNCA, and HNCACB spectra recorded with standard Bruker pulse sequences. Side chain resonances were obtained with (H)C(CO)NH (mixing time 25 ms) and H(CCO)NH (mixing time 25 ms) experiments. The spectra were recorded at 293 K in 20 mM KP_i pH 6, 500 mM NaCl, 10% D_2O. The backbone assignments were completed to 82% and side chain assignments to 50%. All backbone amide resonances in could be assigned.

Backbone NMR assignments of TRPV4 PH-PRR/PH^{AAWAA}-PRR peptide: Backbone resonance assignments of ^{13}C, ^{15}N-labeled *G. gallus* TRPV4 PH-PRR were carried out using HNCO, HN(CA)CO, HNCA, and HNCACB, spectra recorded with standard Bruker pulse sequences. The spectra were recorded at 298 K in 20 mM NaP_i pH 4.5, 150 mM NaCl, 10% D_2O. The backbone amide resonances could be fully assigned, and the overall

backbone resonance assignments were completed to 90%. The resonance assignments of the wt-PH-PRR peptide could be largely transferred to the PH^{AAWAA}-PRR mutant. The remaining assignments could be completed with HN(CA)CO and HNCACB triple-resonance experiments yielding an overall backbone assignment completeness of 91%.

NMR chemical shift-based secondary structure prediction

In a disordered peptide or protein where no secondary structures or intra-chain interactions occur, the backbone NMR chemical shifts of are almost entirely dependent on the amino acid sequence (and pH and temperature) (*377*). Therefore, the backbone chemical shifts of a given amino acid in a protein sequence can be estimated based on empirically derived chemical shift values of pentapeptides (*378*). In secondary structures, the protein backbone adopts defined backbone torsion angles and amide, as well as carbonyl groups, form hydrogen bonds. Therefore, backbone chemical shifts in secondary structures differ significantly from the random coil chemical shifts (RCCS). By comparing the calculated random coil chemical shifts (RCCS) of a protein with its experimentally determined chemical shifts, one can therefore predict the presence of disordered regions or secondary structures. The difference between the experimentally determined chemical shift and the RCCS is called secondary chemical shift (SCC). Strikingly, different secondary structures have different backbone torsion angles and hydrogen bond patterns, thus producing different chemical shifts. Therefore, a given amino acid sequence will have a different SCCs in an α-helix conformation than in a β-sheet structure (*379*). The program TALOS+ automatically calculates the SCCs of a target protein based on its amino acid sequence and the determined H_N, H_α, N, C_α, and C' chemical shifts. The SCCs are then integrated into an artificial neural network analysis (ANN) of a database of 200 high-resolution protein structures with available SCCs to predict the backbone torsion angles phi (ϕ) and psi (ψ) and to provide a measure of certainties for these predictions (*380*).

Solution NMR structure calculation

The structure calculation of the Pacsin3 SH3/TRPV4-PRR complex was carried out by Dr. Ute A. Hellmich and Dr. Nina A. Glogowski.

First, peak picking and NOE assignment of the Pacsin3 SH3 domain was performed with the ATNOS/CANDID module in UNIO (*381*) in combination with CYANA (*382*) using the 3D NOESY spectra. Peak lists were reviewed manually and corrected in case of artefacts and possible NOEs to the TRPV4 PRR peptide. Distance restraints within the PACSIN3 SH3 domain were obtained using the automated NOE assignment and structure calculation protocol available in CYANA (*382*). Obvious NOEs to the peptide were assigned manually. In addition, 2D and 3D ^{13}C-filtered and ^{13}C-edited NOESY-HSQC experiments (*372, 383–385*) were used to obtain distance restraints of the ^{13}C, ^{15}N-labeled Pacsin3 SH3 domain to the unlabeled PRR peptide. The above described 2D ^{13}C-NOESY-HSQC experiments of selectively labelled ^{13}C, ^{15}N TRPV4-PRR peptides in the presence of unlabeled Pacsin3 SH3 domain were also the source for intermolecular distance restraints. In total, 52 intermolecular distance restraints for structure calculation of the complex were assigned manually. Furthermore, 12 intrapeptide distance restraints of the PRR peptide were manually obtained using the above described TOCSY and NOESY spectra. Backbone H, N, C_α, C_β chemical shifts were used to calculate torsion angle restraints using TALOS-N (*386*). Disordered residues in the SH3 domain were identified via $\{^1H\}$, ^{15}N-hetNOE measurements of ^{15}N-Pacsin3 SH3 domain with a 5-fold excess of PRR peptide (see p. 73).

For the structure calculation of the complex, the peptide sequence was connected to the C-terminus of the Pacsin3 SH3 domain sequence through a set of weightless noninteracting dummy atoms. Three sets of distance restraints were used to calculate 100 conformers with CYANA. Ten structures with the lowest target function were submitted to a restrained energy refinement with OPALp (*387*) and the AMBER94 force field (*388*). Structure validation was carried out with the Protein Structure Validation Software suite 1.5 (*389*) restricted to residues with hetNOE values >0.6. (see Table 9)

Chemical shift perturbation experiments

For peptide and lipid titration experiments, the standard Bruker HSQC pulse sequences were used. Solutions with 100-250 μM of ^{15}N-Pacsin1, 2 or 3 SH3 domain in 10 mM Tris pH 7, 100 mM NaCl were titrated with peptide and/or lipid from a concentrated stock solution. The chemical shifts were determined using TopSpin 3.2 (Bruker) The ^{1}H and ^{15}N weighted chemical shift differences observed in ^{1}H, ^{15}N-HSQC spectra were calculated according to equation 22 (*390*):

$$\Delta\delta = \sqrt{\Delta\delta_{\mathrm{H}}{}^{2} + \left(\frac{\Delta\delta_{\mathrm{N}}}{6.5}\right)^{2}} \qquad 22$$

Here, $\Delta\delta_{\mathrm{H}}$ is the ^{1}H chemical shift difference, $\Delta\delta_{\mathrm{N}}$ is the ^{15}N chemical shift difference, and $\Delta\delta$ is the ^{1}H and ^{15}N weighted chemical shift in ppm.

For ^{31}P NMR measurements, diC$_8$-PI(4,5)P$_2$ was used at 500 μM in 10 mM Tris pH 7, 100 mM NaCl, 10% D_2O and titrated with peptide or protein from a concentrated stock solution. Affinities were determined from chemical shift mapping using equation 23 (*391*).

$$\Delta\delta_{\mathrm{obs}} = \Delta\delta_{\mathrm{max}} \frac{\{([P]_{\mathrm{t}} + [L]_{\mathrm{t}} + K_d) - [([\mathrm{P}]_{\mathrm{t}} + [\mathrm{L}]_{\mathrm{t}} + K_d)^2 - 4v[P]_{\mathrm{t}}[L]_{\mathrm{t}}]^{0.5}\}}{2\,[P]_{\mathrm{t}}} \qquad 23$$

Here, $[P]_t$ and $[L]_t$ are the total protein and ligand concentrations, K_D is the dissociation constant, $\Delta\delta_{\mathrm{obs}}$ is the observed chemical shift difference, and $\Delta\delta_{\mathrm{max}}$ is the maximum chemical shift difference.

For NMR titrations of ^{15}N-labeled TRPV4 peptides or the ^{15}N-labeled TRPV4 IDR with liposomes (SUVs) where line broadening instead of peak shifts were observed, the interaction of the reporter with liposomes was quantified using the peak signal loss in response to liposome titration. The signal loss at a lipid concentration c_i was calculated as the relative peak integral decrease $\Delta V/V_0$ according to equation 24.

$$\frac{\Delta V}{V_0} = \frac{V_{\mathrm{i}} - V_0}{V_0} \qquad 24$$

Here, V_0 is the peak integral in the absence of SUVs, and V_i is the peak integral in the presence of a lipid concentration c_i.

Chemical shift perturbation data of ^{15}N-labeled TRPV4 peptides or the ^{15}N-labeled TRPV4 IDR titrated with dIC$_8$-lipids were not compatible with a fit described by equation 23. Instead, the sigmoidal line shape of the binding curves was fitted with a Hill-Langmuir model (equation 25)

$$\Delta\delta_{\mathrm{obs}} = \frac{\Delta\delta_{\mathrm{max}} \cdot c^{n}}{K_D + c^{n}} \qquad 25$$

Here, $\Delta\delta_{\mathrm{obs}}$ is the observed chemical shift difference, and $\Delta\delta_{\mathrm{max}}$ is the maximum chemical shift difference, K_D is the dissociation constant, c is the lipid concentration, and n is the Hill coefficient. The Hill coefficient describes the cooperativity underlying the binding of a ligand to a target. A coefficient of $n > 1$ indicates positive cooperativity, i.e. the binding of a ligand to a target facilitates the binding of additional ligands. A coefficient of $n < 1$ indicates negative cooperativity, i.e. the binding of a ligand to a target hampers the binding of additional ligands. The Hill coefficient does not report on the number of ligands bound to a target (*392*).

Heteronuclear Overhauser effect experiments

HetNOE values were determined as hetNOE = $(V_s\text{-}V_0)/V_0$, where Vs is the integral of an amide cross-peak in the ^{1}H, ^{15}N-HSQC spectrum with proton saturation, and V_0 is the integral in the spectrum without proton saturation. Peak integrals were obtained using Bruker TopSpin 3.2.

{^{1}H},^{15}N-hetNOE experiments of ^{15}N-labeled G. gallus and H. sapiens TRPV4 IDR: {^{1}H},^{15}N-heteronuclear nuclear Overhauser effect (hetNOE) experiments for ^{15}N-labeled IDR (100 μM) constructs were recorded using Bruker standard pulse sequences. Experiments were run in an interleaved fashion with and without proton saturation (7 s) during the recovery delay. The spectra were recorded at 298 K in 20 mM NaP_i pH 4.5, 150 mM NaCl, 10% D_2O.

{^{1}H},^{15}N-hetNOE experiments of 15N-labeled G. gallus Pacsin3 SH3 domain in complex with the TRPV4 PRR: The{^{1}H},^{15}N-heteronuclear nuclear Overhauser effect (hetNOE) experiments of ^{15}N-labeled Pacsin3 SH3 (250 μM) constructs were recorded using Bruker standard pulse sequences. The spectra were recorded in an interleaved manner with and without proton saturation (5 s) during the recovery delay. The spectra were recorded at 293 K in 10 mM Tris pH 7, 100 mM NaCl, 10% D_2O with 1.25 mM TRPV4 PRR peptide.

2.18 Small-angle X-ray scattering

Small-angle X-ray scattering (SAXS) is a versatile tool to investigate the structure of biological macromolecules which are unamenable to high-resolution structure determination methods such as NMR spectroscopy or cryo-EM (both limited by protein size), and X-ray crystallography (proteins need to crystallize) (*393*). It is principally applicable to any type of protein, including extremely dynamic and flexible systems (*394, 395*). It has therefore emerged as a popular tool for studying intrinsically disordered proteins or proteins with large disordered domains. (*396, 397*).

Theoretical introduction

When X-rays hit a molecule, they are coherently scattered by the electrons contained in the atoms of the molecule. The spatial distribution of the electrons in the molecule leads to path differences between the X-rays scattered from different electrons which result in phase shifts between the X-rays and thus to constructive and destructive interference. This can be described in a system of two electrons, e_1 and e_2, separated by the vector **r** (Figure 33 A). When two X-rays with a wave vector $\mathbf{k_i}$ hit the electrons, they are coherently scattered in a new direction described by wave vector $\mathbf{k_f}$. The path difference $(t + s)$ between both waves depends on the position of e_1 with respect to e_2 and the direction of the deflected beam $\mathbf{k_f}$. Because t is a projection of **r** on $\mathbf{k_i}$ and s is the projection of **r** on $\mathbf{k_f}$, $t = \lambda \cdot \mathbf{r} \cdot \mathbf{k_i}$ and $s = -\lambda \cdot \mathbf{r} \cdot \mathbf{k_f}$ (the λ originates from $\mathbf{k_i}$ and $\mathbf{k_f}$ being vectors of magnitude $1/\lambda$). Therefore, the path difference between both waves is $(t + s) = \lambda \cdot \mathbf{r} \cdot (\mathbf{k_f} - \mathbf{k_i})$. From this follows a phase difference between wave 1 and wave 2 of $2\pi \cdot \mathbf{r} \cdot \mathbf{q}$, where $\mathbf{q} = \mathbf{k_f} - \mathbf{k_i}$ is the scattering vector and is perpendicular to the reflecting plane. The waves can be regarded as being reflected against a plane with θ as the reflecting angle and $q = 2(\sin\theta)/\lambda$.

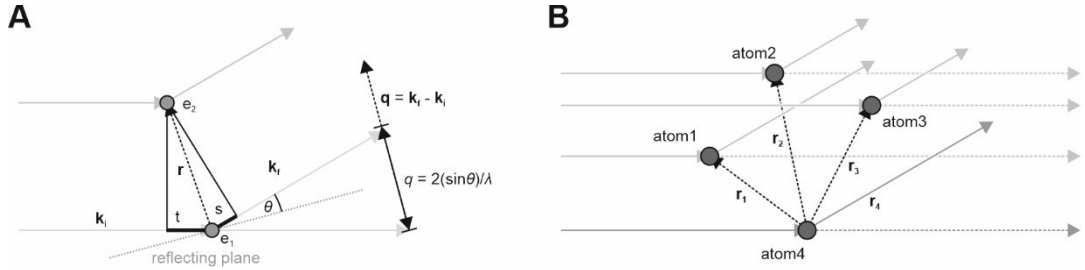


Figure 33: Basic principle of X-ray scattering from electrons in a molecule. (**A**) In a two-electron system (e_1 and e_2), the incident waves described by vector $\mathbf{k_i}$ are scattered into a new direction given by vector $\mathbf{k_f}$. The difference between $\mathbf{k_f}$ and $\mathbf{k_i}$ is described by $\mathbf{q}$, the scattering vector. The separation between e_1 and e2 is given by the vector $\mathbf{r}$. Depending on $\mathbf{q}$ and $\mathbf{r}$, the scattered waves will have a path difference (t + s) which results in a phase difference $2\pi\ \mathbf{r}\cdot\mathbf{q}$. The phase difference results in constructive or destructive interference and determines the amplitude of the scattered X-rays. (**B**) In a molecule, the position of each atom is given by a vector $\mathbf{r}_j$ (here, j = 1, 2, 3, 4 for a four atom molecule). All atoms scatter the X-rays coherently in the same direction. The phase shift between the X-rays depends on the atomic positions given by $\mathbf{r}_j$ and the scattering vector $\mathbf{q}$.

The scattering behaviour of an atom is determined by the phase difference between the waves scattered from its electrons which in turn depends on the position of the electrons, given by vector $\mathbf{r}$, and the direction of the scattered X-rays, given by the vector $\mathbf{q}$. The scattering from an atom is described by the atomic scattering factor $f(\mathbf{q})$ which is an integral over the scattering contribution from all electrons contained in the atom according to 30.

$$f(\mathbf{q}) = \int \rho(\mathbf{r})\ \exp(2\pi i \cdot \mathbf{r} \cdot \mathbf{q})\ d\mathbf{r} \qquad 26$$

Here, $\rho(\mathbf{r})$ is the electron density at position $\mathbf{r}$ with respect to an origin. In a molecule, all atoms contribute to the scattering behaviour (Figure 33 B) and the scattering of a molecule is described by a structure factor $\mathbf{F}(\mathbf{q})$. The structure factor depends on the spatial distribution of the atoms and is thus unique for each molecule. It is described by equation 33.

$$\mathbf{F}(\mathbf{q}) = \sum_{j=1}^{n} f_j \exp(2\pi i \cdot \mathbf{r}_j \cdot \mathbf{q}) \qquad 27$$

Here, $\mathbf{r}_j$ is the position of the j-th atom with scattering factor f_j, n is the total number of atoms contained in the molecule. Instead of a summation over all atoms in the molecule, the structure factor can also be expressed by integrating over all electrons in the molecule (equation 28).

$$\mathbf{F}(\mathbf{q}) = \int \rho(\mathbf{r})\ \exp(2\pi i \cdot \mathbf{r} \cdot \mathbf{q})\ dv \qquad 28$$

Here, $dv = V \cdot dx \cdot dy \cdot dz$ with V being the volume occupied by the molecule. The x, y, and z coordinates in Cartesian space are contained in $\mathbf{r}$. The relationship between the structure factor and the electron density distribution has the typical reciprocal form of the Fourier transformation $\mathcal{F}$ (equations 29 and 30).

$$(\mathcal{F}f)(x) = \int f(x)\ \exp(-2\pi i \cdot y \cdot x)\ dy \qquad 29$$

$$f(x) = \int (\mathcal{F}f)(x)\ \exp(2\pi i \cdot y \cdot x)\ dx \qquad 30$$

Thus, the scattering of X-rays from the electrons of a molecule can be described as a Fourier transformation of the spatial arrangement of the electrons in the molecule. Conversely, if the structure factor $F(\mathbf{q})$ of a molecule is known, the electronic structure can be derived via Fourier transformation. The structure factor is related to the measured scattering intensity by $I(\mathbf{q}) = |\mathbf{F}(\mathbf{q})|^2$. Thus, from measuring the X-ray scattering of a molecule, it is possible to derive its electronic structure.

The scattering of X-rays from a molecule is usually detected on two-dimensional detectors such that the detected intensities reflect a slice through the three-dimensional molecular transform. 3D-information can be obtained from measuring the X-ray scattering of the molecule in a minimal set of orientations. The orientation-dependent 2D molecular transforms can then be reassembled into a 3D transform from which the 3D structure of the molecule can be obtained via Fourier transformation (*398*). SAXS and X-ray crystallography both exploit the reciprocal relationship between the X-ray scattering pattern and the spatial distributions of the electrons in a molecule to determine the structure of molecules. In protein X-ray crystallography, the orientation of the proteins is fixed in a crystal and tilting the crystal enables to collect a series of orientation-dependent scattering patterns, termed diffraction patterns (see section 2.19). These can be subsequently merged, and Fourier transformed to give the proteins three-dimensional electronic structure (*399*). However, the isotropic molecular tumbling of the proteins in solution during a SAXS measurement produces a spherically averaged X-ray scattering pattern which is detected as a radially symmetric 2D scattering pattern (Figure 34 A). The information encoding the three-dimensional electronic structure of the molecules is thus irreversibly reduced to a 1D profile. The 1D SAXS profile encodes low-resolution information on the overall shape and size of the studied protein (*398*).

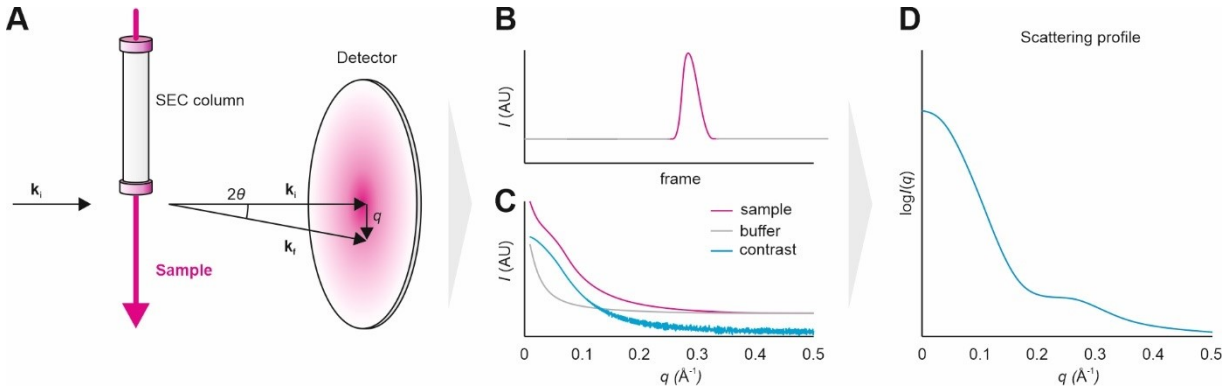


Figure 34: Basic principle of a small-angle X-ray scattering experiment. (**A**) An X-ray beam with vector k_i enters a protein-containing sample (blue), which isotropically scatters the X-rays in all directions at different angles (2θ). The scattered X-rays (k_f) are detected at different distances q from non-scattered (k_i) X-rays on a detector. To ensure sample monodispersity, SAXS can be coupled to size exclusion chromatography (SEC). (**B**) In a SEC-SAXS experiment, the scattering of the sample eluted from the SEC column (pink peak) and the scattering of the solvent (grey baseline) is detected in a series of frames yielding a chromatogram with I versus frame number. (**C**) The scattering profile of the biomolecule of interest (blue) plotted as I versus q is obtained by subtracting the X-ray scattering detected in the baseline regions (buffer) from the X-ray scattering detected in the protein elution peak. (**D**) The final SAXS profile is usually displayed as a log$I(q)$ vs q plot.

In a SAXS experiment, an incident wave of collimated X-rays (described by wave vector $\mathbf{k_i}$) irradiates a sample which scatters the intensities of the X-rays (wave vector $\mathbf{k_f}$) by an angle 2θ onto a detector (Figure 34 A). The radial scattering of the X-rays can be expressed by the scattering vector q which describes the distance between the incidence of the scattered (k_f) and the non-scatted beam (k_i) on the detector (equation 31). Note that in SAXS, it is sufficient to treat the scattering vector as a scalar due to the spherical averaging of the X-ray scattering profile.

$$q = |\mathbf{k_i} - \mathbf{k_f}| \quad (31)$$

Based on the relationship between the magnitude of wavevector k and wavelength λ (equation 32), q can be expressed by the scatting angle 2θ according to equation 33.

$$k = \frac{2\pi}{\lambda} \quad (32)$$

$$q = \frac{4\pi \sin(\theta)}{\lambda} \quad (33)$$

Unlike in X-ray diffraction, molecules in a SAXS experiment tumble in solution and scatter the incident X-ray intensities isotropically thus yielding a spherically averaged X-ray scattering pattern. In addition to the lost three-dimensional information content of X-ray scattering, the limitation to small scattering angles makes SAXS a low-resolution method. In an X-ray diffraction experiment, the lattice of a protein crystal acts as a signal amplifier that generates intense reflexes at wide scattering angles which encode structural information at resolutions below 2 Å. Unlike that, SAXS detects the contrast $\Delta\rho$ between the scattering of the biomolecule of interest, ρ, and the scattering of the bulk solvent, ρ_s (equation 34) (*398*).

$$\Delta\rho = \rho - \rho_s \quad (34)$$

The SAXS pattern of the biomolecule of interest is obtained via subtraction of the solvent SAXS pattern from the SAXS pattern of the sample. It is usually displayed as the logarithmic intensity $\log I(q)$ versus scattering vector q. The scattering of a protein in aqueous solution is only about 10% above the background. Accurate buffer subtraction and subsequent data processing, therefore, depends on precisely matched buffers between the protein sample and a blank sample. An elegant way to achieve accurate buffer matching conditions is the coupling of the SAXS detection to a SEC column (Figure *34* A, B and C). In a SEC-SAXS experiment, the X-ray scattering profile of the sample is obtained from the protein elution peak, and the solvent X-ray scattering profile is obtained from the buffer baseline (*400*). A significant side effect of using a SEC-SAXS system is the separation of potential aggregates which can dramatically change the outcome of a SAXS experiment. According to equation 36, the scattering intensity in SAXS is directly proportional to the square of the particles molecular mass.

$$I(0) = N\,(\Delta\rho V)^2 \quad (35)$$

Here, $I(0)$ is the X-ray scattering intensity at $q = 0$, N is the number of scattering, and V is the particle volume. Thus, a few aggregated particles with large molecular weights, i.e. $\Delta\rho V$ can mask the scattering signal of the biomolecule of interest. Moreover, aggregates due to attractive interparticle interactions can introduce long-range order in the sample, which influences the scattering profile. To understand this, one needs to consider that in SAXS the X-ray scattering intensity $I(q)$ of a particle is a convolution of a form factor $F(q)$ and a structure factor $S(q)$ (equation 36). The form factor describes intraparticle interactions, i.e. the distances between scattering centres within a molecule and thus its electronic structure. The structure factor accounts for interparticle interactions through repulsion and attraction, which leads to long-range order in the studied system. Long-range order can occur through aggregation or in highly concentrated samples where repulsion or attraction leads to constrained distances between the scattering particles.

$$I(q) = F(q) * S(q) \quad (36)$$

SAXS data describing the structure of a particle in solution require the structure factor to be equal to one such that $I(q)$ is solely described by the form factor. Whether or not a structure factor contributes to the detected X-ray scattering pattern can be validated in a Guinier plot (Figure 35 A). The Guinier method is based on the approximation that for sufficiently low q values, the logarithmic scattering intensity decreases linearly with the squared scattering vector q (equation 37) (*401*).

$$\log I(q) = \log I(0) - \frac{R_g^2}{3} q^2 \quad (37)$$

Here, R_g is the radius of gyration, which corresponds to the root mean square distance of the particles scattering centres, i.e. its electrons, from its centre of mass. R_g is an important structural parameter for the interpretation of SAXS data. If no structure factor contributes to log$I(q)$, the Guinier plot shows a linear decrease in the low q^2 regime. Otherwise, log$I(q)$ increases at low angles in the case of attractive interparticle interactions or it decreases in the case of repulsion. During data collection, it is impossible to measure at X-ray scattering at angles close to zero as $I(0)$ coincides with the beam of the non-scattered X-rays. $I(0)$ is usually obtained by extrapolating the Guinier plot to the intercept with the log$I(q)$. Nevertheless, it is important to detect the scattering at sufficiently small angles which correspond to intraparticle distances larger than the maximum particle dimension D_{max}. The lack of scattering information at low q values can else lead to an inaccurate underestimation of the particle size. The lowest q value should thus satisfy $q_{min} < \pi/D_{max}$ and the highest q value should be $q_{max} > 1.3/R_g$ (Figure 35 A) (*398*).

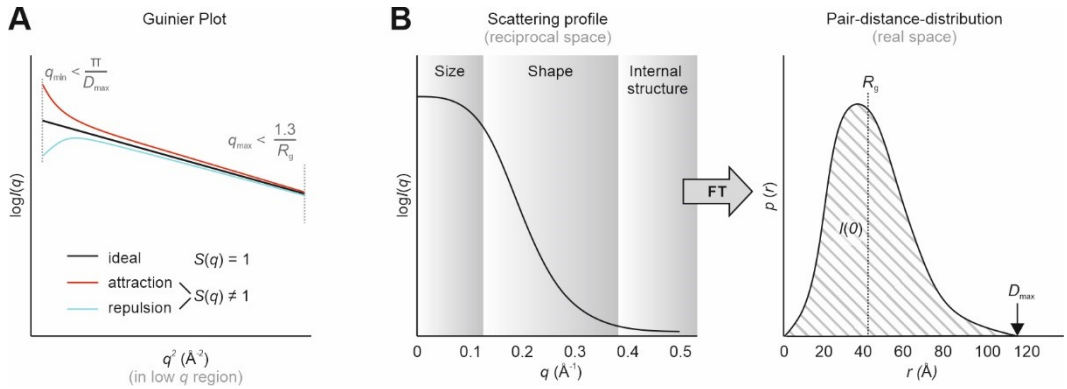


Figure 35: SAXS data validation and processing. (**A**) The Guinier plot is used to assess the potential contribution of a structure factor the scattering data. When no attractive (red) or repulsive (light blue) interparticle interactions contribute to the SAXS signal, log$I(q)$ should decrease linearly (black) with increasing q^2 values in the low q regime. In the Guinier plot, the maximum q value (q_{max}) depends on the radius of gyration (R_g), whereas the minimal q value (q_{min}) depends on the maximum particle dimension D_{max}. (**B**) Different q regimes in the SAXS profile encode information on the size, shape, and the internal structure of the scattering particle in reciprocal space. The scattering profile can be converted into a real space distance-distribution $p(r)$ via Fourier transformation. $P(r)$ provides a more accurate estimation of the radius of gyration R_g than the Guinier method. It also provides the maximum particle dimension D_{max} (maximum r-value) and $I(0)$ (integral under $p(r)$ curve, see equation 40).

The SAXS scattering profile $I(q)$ encodes the spherically averaged electronic structure of the scattering particle in reciprocal space. The reciprocal nature of $I(q)$ is, however, not very intuitive. Therefore, $I(q)$ can be converted into a real-space distance distribution, $p(r)$, via indirect Fourier transformation (equation 38, Figure 35 B) (*402*). $P(r)$ provides information on the particle shape, size, and flexibility.

$$I(q) = \int p(r) \frac{\sin(qr)}{qr} dr \qquad 38$$

The real space distance-distribution provides a more accurate way to calculate R_g and $I(0)$ than the Guinier method since $p(r)$ includes the scattering information from all measured angles (equations 39 and 40). Moreover, $p(r)$ is used to determine D_{max}. Notably, D_{max} is the most error-prone of all SAXS derived parameters, especially in flexible systems (*403*).

$$R_g = \frac{\int_0^{D_{max}} r^2 p(r)\, dr}{\int_0^{D_{max}} p(r)\, dr} \qquad 39$$

$$I(0) = 4\pi \int_0^{D_{max}} p(r)\, dr \qquad 40$$

The real space distance-distribution indicates the overall shape and flexibility of a protein (Figure 36 A). Compact globular proteins usually have bell-shaped $p(r)$ curves. In contrast, disordered proteins are characterized by $p(r)$ curves which gradually decay towards D_{max} values which are substantially larger than the D_{max} values of globular proteins with the same molecular mass. Multidomain proteins with mixtures of globular domains and disordered regions display the features of globular as well as unfolded proteins (Figure 36 B). Besides evaluating the protein fold, $p(r)$ is suitable to visualize conformational changes in proteins in response to pH, temperature, or ligand binding. A complimentary depiction of collected SAXS data to evaluate the flexibility of a studied protein is the Kratky plot which is a plot of $q^2 \cdot I(q)$ versus q (Figure 36 C). As in $p(r)$, a compact protein displays a bell-shaped curve in the Kratky plot, which decreases to zero at high q values. A disordered protein shows an asymptotic behaviour towards higher q-values. Increasing $q^2 \cdot I(q)$ at high q values are a strong indicator of flexible regions, whereas $q^2 \cdot I(q) \sim 0$ at high q values is indicative of very compact structures. Consequently, the Kratky plots of proteins with mixtures of globular and folded domains have bell-shapes in low q regimes and show asymptotic behaviours in high q regimes (*395, 404*).

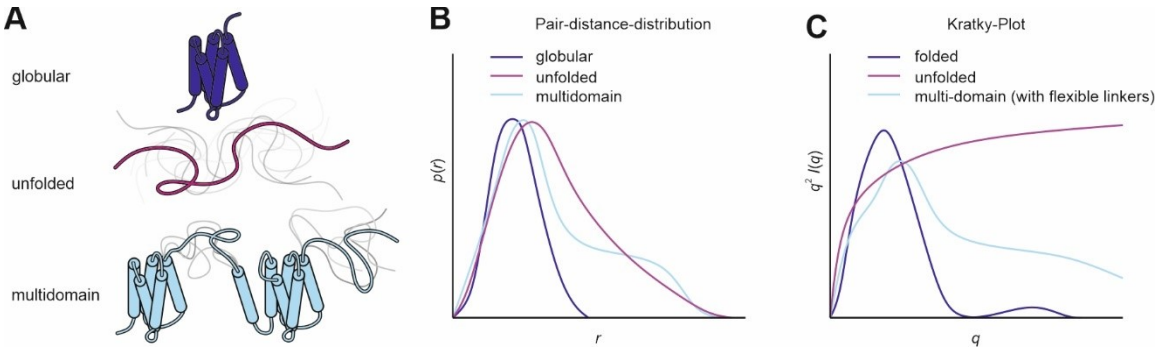


Figure 36: SAXS analysis of structural and dynamic. (**A**) Schematic representation of a globular, an unfolded, and a multidomain protein. In multidomain proteins, folded domains are connected by flexible linker regions. (**B**) Representative real space distance-distributions, and (**C**) Kratky plots of globular (dark blue), unfolded (magenta) and multidomain proteins (light blue).

Even though SAXS is a low-resolution method and limited by the spherical averaging of the X-ray scattering data, SAXS-guided *ab initio* structure modelling approaches can be alternative to high-resolution structure determination methods, e.g. when proteins refrain from crystallization, yield low quality NMR spectra, or are too small for cryo-EM (*405*). More importantly than being an alternative, SAXS-based rigid-body modelling is ideal to complement structural studies of proteins when high-resolution data are available but incomplete. Several

hybrid methods combining rigid-body and *ab initio* modelling have been developed which allow integrating available high-resolution structures of isolated protein domains into structural models of the corresponding full-length proteins (*406*). More recently, SAXS has emerged as a powerful tool for studying the structure and dynamics of intrinsically disordered proteins. Various tools to model isolated IDPs or IDRs in multidomain proteins have been implemented and were used to study the intrinsic disorder in the TRPV4 N-terminus in this thesis (*394, 403, 407*). Theoretical and practical details of the respective methods are discussed further below.

SAXS data collection and reduction

Small-angle X-ray scattering (SAXS) experiments were carried out at the EMBL-P12 bioSAXS beam line (*408*) in collaboration with Dr. Cy Jeffries (BioSAXS group, Prof. Dr. Dimtri Svergun, EMBL Hamburg). SEC-SAXS data collection, $I(q)$ vs q, where $q = 4\pi\sin\theta/\lambda$; 2θ is the scattering angle and λ the X-ray wavelength (0.124 nm; 10 keV) was performed at 20 °C using S75 and S200 Increase 5/150 analytical SEC columns (GE Healthcare) equilibrated in the appropriate buffers (see Table 6) at flow rates of 0.3 ml min^{-1}. Automated sample injection and data collection were controlled using the *BECQUEREL* beam line control software (*409*). The sample injection volume and load concentrations are listed in Table 6. The SAXS intensities were measured as a continuous series of 0.25 s individual X-ray exposures, from the continuously-flowing column eluent, using a Pilatus 6M 2D-area detector for a total of one column volume (ca. 600-3000 frames in total, see Table 6). The 2D-to-1D data reduction, i.e., radial averaging of the data to produce 1D $I(q)$ vs q profiles, were performed using the SASFLOW pipeline incorporating RADAVER from the ATSAS 2.8 suite of software tools (*410, 411*). The individual frames obtained for each SEC-SAXS run were processed using CHROMIXS (*400*). Briefly, individual SAXS data frames were selected across the respective sample SEC-elution peaks and an appropriate region of the elution profile, corresponding to SAXS data measured from the solute-free buffer, were identified, averaged and then subtracted to generate individual background-subtracted sample data frames. These data frames underwent further CHROMIXS analysis, including the assessment of the radius of gyration (R_g) of each individual sample frame, scaling of frames with equivalent R_g, and subsequent averaging to produce the final 1D-reduced and background-corrected scattering profiles. Only those scaled individual SAXS data frames with a consistent R_g through the SEC-elution peak that were also evaluated as statistically similar through the measured q-range were used to generate the final SAXS profiles. Corresponding UV traces were not measured; the column eluate was flowed directly to the P12 sample exposure unit after the small column, forging UV absorption measurements, in order to minimize unwanted band-broadening of the sample. All SAXS data-data comparisons and data-model fits were assessed using the reduced χ^2 test and the Correlation Map, or CORMAP, p-value (*412*). Fits within the χ^2 range of 0.9–1.1 or having a CORMAP p-values higher than the significance threshold cutoff of $\alpha = 0.01$ are considered excellent, i.e., no systematic differences are present between the data-data or data-model fits at the significance threshold.

Table 6: Experimental conditions of SAXS measurements. All measurements were carried out at a temperature of 20 °C, an X-ray wavelength of 0.124 nm, a sample to detector distance of 3 m, a flow rate of 0.3 mL min^{-1}. Used buffers for SEC-SAXS were as follows: Buffer A = 20 mM Tris pH 7 300 mM NaCl 10 mM DTT; Buffer B = 20 mM Tris pH 7 100 mM NaCl 10 mM DTT, Buffer C = buffer A + 10% (v/v) glycerol.

Sample		Column type	Injection volume (µL)	Load concentration (mg mL^{-1})	buffer	Exposure time (s) /number of frames
G. gallus TRPV4	NTD	S200	40	10.0	A	0.25/2400
	NTD^{AAWAA}	S200	40	8.2	A	0.25/2400
	NTD ΔN120	S200	45	11.4	A	0.25/600
	NTD ΔN104	S200	45	10.4	A	0.25/600
	NTD ΔN97	S200	45	5.8	A	0.25/600
	NTD ΔN54	S200	45	5.0	A	0.25/600
	ARD	S200	40	10.7	A	0.25/2400
	IDR	S200	25	11.4	A	0.25/2400
	IDR^{AAWAA}	S200	25	11.4	B	0.25/2400
	IDR	S75	22	10.1	B	0.25/2400
H. sapiens TRPV4	NTD	S200	65	4.44	C	0.25/2400
	IDR	S200	40	9.11	B	0.25/2400
G. gallus Pacsin3	Fl-Pac3	S200	35	10.4	B	0.25/2400
	F-BAR	S200	35	10.2	B	0.25/2400
	l-SH3	S75	40	10.5	B	0.25/2400
	SH3	S75	30	10.1	B	0.25/2400

Data processing and normalization

Real space distance-distributions, $p(r)$ vs. r, were obtained by inverse Fourier transformation of the processed SAXS scattering data, $I(q)$ vs. q, using the CHROMIXS software package. To enable the comparison of the real space distance distributions between different measurements, the $p(r)$ values were multiplied with a normalization constant B to reduce the $p(r)$ function to a probability density of one (equation 41).

$$B = \frac{1}{\int_0^{D_{max}} p(r)\ dr} \qquad 41$$

Kratky plots were generated by plotting $q^2I(q)$ versus q. To enable comparability between measurements and to allow for semi-quantitative analyses of protein flexibility and disorder, Kratky plots were normalized to the size and concentration of the scattering particle in a dimensionless plot with $(qR_g)^2\ I(q)/I(0)$ versus qR_g. The R_g and $I(0)$ values used for the normalization were obtained from the respective $p(r)$ curves.

Rigid-body and *ab initio* modelling

All SAXS-based modelling approaches were performed using the ATSAS software package (*411*).

CRYSOL

CRYSOL calculates the scattering profile of a protein based on atomic coordinates obtained from X-ray crystallography, cryo-EM, or NMR spectroscopy (*411*). Simultaneously, it can fit the calculated scattering profile to experimental SAXS data. Discrepancies between theoretical curves and experimental data may be indicators that the structure of a protein in solution is different from that in a crystal.

For the CRYSOL analysis of the *G. gallus* TRPV4 ARD (see section VI3.2, Figure 58), the *G. gallus* TRPV4 ARD X-ray crystal structure (PDB: 3w9g) was used as input. The coordinates file was curated in PyMOL by removing all crystallographic water molecules and residues M132, A384-H387 as these are not contained in the construct used for SAXS measurements. The $p(r)$ curve of the Pacsin3 F-BAR/SH3 domain complex (see section VI3.7, Figure 109) was obtained by first calculating its scattering profile with CRYSOL and subsequent Fourier transformation in CHROMIXS. The coordinate file used as input was generated as described in section VI2.20.

Ab initio shape reconstruction

The programs DAMMIN, GASBOR, and MONSA have been developed for the *ab initio* reconstruction of low-resolution envelopes from collected SAXS data (*413*). In the modelling procedure, a search volume with a radius R enclosing the measured protein is filled with N spheres with radius $r << R$. The spheres are called beads or dummy atoms. Each bead can belong to a phase X_j with $1 < j < N$ and X_j between 0 (solvent) and K ($K \geq 1$). In the easiest case (DAMMIN and GASBOR), $K = 1$, i.e. a bead is either a solvent molecule or an atom/residue in the modelled protein. The bead model is described by a vector which assigns the phase X with N components. For the *ab initio* modelling of proteins, the number of beads assigned to the protein phase typically equals the number of residues contained in the protein. In DAMMIN, the bead positions are restricted to a grid, whereas in GASBOR, the bead positions are free of choice with the only restriction being the absence of clashes. MONSA is an extension of the DAMMIN algorithm. It allows integrating multiple data sets with K > 1 to calculate the phases in a particle. A typical practical application of MONSA is the modelling of multidomain proteins when the scattering data of a full-length protein and its isolated domains are available. MONSA can integrate the datasets and assign the phases in the resulting bead model to the individual domains.

SAXS-based *ab initio* bead modelling approaches follow a common principle. In a set of $M \geq 1$ SAXS curves $I^{(i)}_{\exp}(q)$, $i = 1, \ldots, M$, bead models are assembled in the search for a configuration X which minimizes χ^2 between the experimental data and the theoretical scattering curve of the generated bead model according to equation 42.

$$\chi^2 = \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{N(i)} \left[\frac{I^{(i)}_{\exp}(q_j) - I^{(i)}(q_j)}{\sigma(q_j)} \right]^2 \qquad 42$$

Here, $N(i)$ is the number of points in the i-th curve, $\sigma(q)$ is the experimental error, and $I^{(i)}(q_j)$ is the calculated curve of configuration X. The assembly of beads to minimize χ^2 follows a simulated annealing (SA) approach to prevent running into local χ^2 minima. The basic principle behind simulated is to find a configuration X minimizing a goal function $f(X) = \chi^2 + aP(X)$ where $P(X)$ is a measure of the compactness of a given configuration X, also termed the average looseness and $a > 0$ is the weight of the looseness penalty (typically ~10 for compact bodies). Simulated annealing randomly modifies the configuration X and moves to the configurations that decrease $f(X)$. With the probability $\exp(-\Delta/T)$, the algorithm can also move to a configuration that increases f(X). Here, T is the temperature of the system. At the beginning of the annealing process, T is high, and changes in X are almost random. The probability of accepting configurations increasing $f(X)$ decreases in the course of the annealing process, i.e. the system cools down. At cold temperates when a configuration is reached, which does not improve $f(X)$ upon modification, the algorithm stops (*413*). The sequential steps can be described as follows:

1) Start from a random configuration X_0 at a high temperature T_0, e.g. $T_0 = f(X_0)$
2) Select a random atom and change its phase randomly, then recompute $\Delta = f(X') - f(X)$
3) If $\Delta < 0$ go to X'. If $\Delta > 0$ go to X' with a probability of $\exp(-\Delta/T)$. Then repeat step 2 either from X' or X
4) Hold T constant for $100N$ reconfigurations or $10N$ successful reconfigurations, then cool the system down ($T'=0.97T$)

5) Continue cooling until no improvement $\Delta \sim 0$, i.e. no improvement in $f(X)$

The simulated annealing procedure only changes a single dummy atom per move. This facilitates and accelerates the recalculation of a scattering curve of a configuration X compared to calculating a new configuration from scratch. It is important to note that *ab initio* modelling can result in multiple dummy models with different shapes but equally good fits. This is due to the fact that differently shaped particles can scatter X-rays in almost identical ways. Moreover, enantiomorphic particles have identical SAXS patterns (*414*).

In this thesis, DAMMIN was used for the ab initio modelling of the *G. gallus* TRPV4 ARD SAXS envelope (section VI3.2, Figure 58). GASBOR and MONSA were used for the shape reconstruction of the *G. gallus* Pacsin3 F-BAR domain and full-length Pacsin3, respectively (section VI3.7, Figure 109). For the MONSA model of the *G. gallus* full-length Pacsin3 model, the SAXS data of Pacsin3 F-BAR were used input for the first phase, and the SAXS data of fl-Pac3 were used as input for the second phase. For each model, ten bead models were generated by DAMMIN, GASBOR or MONSA, respectively, with P2 symmetry imposed, and the most probable models to fit the experimental data were selected with DAMSEL (9 out of 10), aligned with DAMSUP, and averaged with DAMAVER (*415*). The final bead models were aligned with available high-resolution structures with DAMSUP (*416*).

Rigid-body modelling

The conformation of proteins in high-resolution structures obtained from X-ray crystallography often does not represent the conformation of the protein in solution, e.g. due to crystal contacts. Moreover, high-resolution structures represent a protein in a defined conformation. However, in solution, proteins often interconvert between multiple conformations. The conformational flexibility of proteins in solution often manifest in discrepancies between the theoretical SAXS curves of proteins calculated from high-resolution structures and the experimental SAXS curves.

SREFLEX: The rigid-body modelling tool SREFLEX allows to estimate the flexibility of high-resolution models of proteins using a normal mode analysis (NMA) algorithm (*417*). In NMA, protein residues are modelled as point masses at their C_α positions which are connected by springs. The protein structure is then divided into pseudo-domains encompassing a set of C_α atoms, and their potential energy function is calculated according to equation 43.

$$V = \sum_{\substack{r_{ij}^0 < R_c \\ i<j}} k\left(r_{ij}-r_{ij}^0\right)^2 \qquad 43$$

Here, r_{ij}^0 is the distance between atoms i and j. The values R_c = 10 Å and k = 1.0 kcal mol^{-1} Å^{-2} are used as a standard by SREFLEX. The normal modes are ordered to their vibrational energies starting with lower-energy normal modes that correspond to global rearrangements (low frequencies). High-energy normal modes correspond to local movements (high frequencies). During a SREFLEX analysis, a protein structure is partitioned into pseudo-domains of C_α atoms which are then displaced in Cartesian space following either a given normal mode or a combination of different normal mode to generate a new conformation of the protein. The conformer is then scored against the experimental SAXS data according to equation 42. The algorithm proceeds hierarchically by first probing large rearrangements (low-frequency modes) and progressing to smaller, more localized motions (high-frequency modes). The output comprises an ensemble of possible conformers that improve the agreement with the experimental SAXS data (*417*).

SREFLEX was used to estimate the flexibility of the *G. gallus* TRPV4 ARD (section VI3.2, Figure 58) using the curated (see CRYSOL) *G. gallus* TRPV4 ARD X-ray crystal structure (PDB: 3w9g) as input. Pseudo-domains were automatically assigned by the algorithm, and 100 conformers were generated and scored against the experimental SAXS data yielding an ensemble of 10 conformers with an average χ^2 = 1.31, with the best conformer having χ^2 = 1.21.

SASREF: SAXS can also be used to model the quarternary structure of a complex formed by subunits with known atom structure. In the program called SASREF, a simulated annealing protocol constructs interconnected ensembles of subunits without clashes, while minimizing the discrepancy with the experimental data (equation 42) (*418*). The goal of the SA protocol is to perform random modifications of the interconnected subunits which decrease a target function E, but also occasionally also to modifications which increase E (depending on the temperature, compare the SA protocol of the DAMMIN algorithm). The temperature is high at the beginning of the annealing procedure and modifications are random. With decreasing temperatures, the annealing reaches a conformation with the best score against the experimental data. The minimization starts from an arbitrary assembly of K subunits which can be defined manually or automatically. The theoretical scattering curves are calculated with CRYSOL. A modification step includes the rotation of a randomly selected subunit by an arbitrary angle about a rotation axis with a subsequent shift along a random direction. The target function in SASREF to be minimized is described by equation 44.

$$E = \sum (\chi^2)_i + a_{\text{dis}}P_{\text{dis}} + a_{\text{cross}}P_{\text{cross}} + a_{\text{cont}}P_{\text{cont}} \qquad 44$$

Here, a are the weighting factors of the penalties P_{dis}, P_{cross}, and P_{cont}. The penalty $P_{\text{dis}} = -\ln(K/K_G)$ ensures the interconnectivity of the subdomains. K is the number of assembled subunits and K_G is the number of subunits in the longest interconnected subset found in their current arrangement. P_{cross} constraints the assembly to the absence of overlaps between the subunits and P_{cont} is a user-specified C_α-C_α distance between the intersections of the subunits.

SASREF was used as a complement to investigating the conformational flexibility of the *G. gallus* TRPV4 ARD (section VI3.2, Figure 58). The curated (see CRYSOL) *G. gallus* TRPV4 ARD X-ray crystal structure (PDB: 3w9g) was partitioned into four subunits: sub1 = res. 135-147, sub2 = res. 148-163, sub3 = res. 164-250, sub4 = res. 251-382. C_α-C_α distance restraints between the intersections of the subunits were set to the distances observed in the ARD X-ray structure (PDB: 3w9g). SASREF generate six ARD subunit assemblies which fit the experimental SAXS data with an χ^2 of 1.15-1.25.

Modelling of IDP conformational ensembles

The structures of intrinsically disordered proteins are inaccurately described by a single conformation. They are devoid of secondary structures and interconvert rapidly between a continuum of conformations. Consequently, SAXS-based *ab initio* modelling or rigid-body modelling usually fails for IDPs. Instead, the so-called ensemble optimization method (EOM) can be suitable to describe the conformational ensemble of an IDP or of a protein with a large disordered region (*419*). EOM assumes that the collected SAXS intensity of an IDR, $I(q)$, is composed of the averaged scattering intensities $I_k(q)$ from K components (equation 45).

$$I(q) = \sum_{k}^{K} v_k I_k(q) \qquad 45$$

Here, v_k is the volume fraction of the k-th component. For a given IDP with N residues, EOM calculates a pool of 10 000 conformations of random chains composed of N beads. Each bead represents the C_α atom of a residue in the IDP sequence. The dihedral angles and distances between C_α-C_α beads are selected randomly while adhering to the quasi-Ramachandran diagram and being free of clashes (*420*). Notably, EOM can also be combined with rigid-body modelling to describe the conformational ensemble of disordered regions connected to folded domains. For this, a pool of 10 000 conformations of the disordered region modelled in as a chain of beads to a high-resolution structure of the folded domain are generated. EOM calculates the scattering profiles of each generated conformer and selects subsets of conformations from the conformational ensemble using a genetic algorithm (GA). The theoretical intensity $I(q)$ of the conformational subset is calculated from the theo-

retical scattering curves of the conformers in the subset according to equation 45. The search for the conformational subset, which best describes the conformational ensemble of the measured IDP is based on minimizing the discrepancy between the experimental data and the calculated curves according to equation 46.

$$\chi^2 = \frac{1}{K\text{-}1} \sum_{j=1}^{K} \left[\frac{\mu I(q_j) - I_{\exp}(q_j)}{\sigma(q_j)} \right]^2 \qquad 46$$

Here, K is the number of experimental curves, $\sigma(q)$ are standard deviations, and μ is a scaling factor.

The output of an EOM analysis contains the R_g and D_{max} distributions in the conformational subsets selected by the GA, which best fits the data. The R_g and D_{max} distributions are then compared those of the random pool. EOM also provides coordinate files of class averages of the conformational subset, termed EOM refined conformers, as well as volume and end-to-end distance distributions.

EOM was employed for the analysis of the NTDs and IDRs from *G. gallus* and TRPV4 *H. sapiens* (see section VI3.1, Figure 48). For the TRPV4 IDRs, 10 000 random chain conformers with $N = 133$ (*G. gallus*) and $N = 147$ (*H. sapiens*) dummy residues, respectively, were generated and EOM was run withouth any symmetry imposed. Structures of the NTDs were generated using the published X-ray crystal structures of the *G. gallus* and the *H. sapiens* TRPV4 ARD (PDB codes 3w9g and 4dx1, respectively) as templates for the ARD. The structures were curated by removing all crystallographic water molecules, ligands, and residues which are not contained in the constructs used for SAX data collection. The sequence accounting for the IDR was modelled by a chain of dummy residues, in analogy to the EOM analysis of the isolated IDRs. Again, no symmetry was imposed in the EOM run. The genetic algorithm of EOM selected the final conformational subsets with χ^2 of 0.98-1.14.

EOM was also employed to model the flexible linker in the clamped (autoinhibited) state and in the open state of full-length *G. gallus* Pacsin3 (section VI3.7, Figure 110). The *G. gallus* Pacsin3 F-BAR domain X-ray structure (see section VI2.18, Figure 106, Table 10) and the coordinates of the SH3 domain in the NMR structure of the *G. gallus* Pacsin3 SH3 domain in complex with the TRPV4 PRR (PDB: 6f55) were used as templates to calculate pools of 10 000 full-length Pacsin3 structures in which the linker residues were modelled as random bead chains. For the generation of open conformers, the F-BAR domain was fixed in position, and the SH3 domains were allowed to move freely. The Pacsin3 F-BAR/SH3 assembly of the putative clamped conformation was obtained by aligning the Pacsin3 F-BAR and SH3 domains with the X-ray crystal structure of the *M. musculus* Pacsin1 F-BAR/SH3 complex (PDB: 2x3x). The residues of the linker residues were again modelled as random bead chains in the Pacsin3 F-BAR/SH3 complex. EOM was run by keeping the positions of both the F-BAR domain and the SH3 domains fixed. Notably, it was assumed that the SH3/F-BAR interaction in the clamped Pacsin3 conformation occurs in *trans*, i.e. the SH3 domain of chain A attaches to the F-BAR domain of chain B. A *cis* interaction in the putative F-BAR/SH3 complex (i.e. $SH3^A$/$F\text{-}BAR^A$ & $SH3^B$/$F\text{-}BAR^B$) is physically not possible due to a too short linker. Because the SH3 domains can be expected to move independently from each other in the Pacsin3 open conformer, the Pacsin3 dimer was treated as a single peptide chain, and P1 symmetry was imposed for the EOM runs. A similar approach was used for the EOM analysis of the closed Pacsin3 conformer to account for the independent movement of the linker.

2.19 X-ray crystallography

Macromolecular X-ray crystallography has been the workhorse for structural biology for decades. Currently, 88.5% of the high-resolution structures deposited in the PDB were solved via X-ray crystallography‡. As described in section VI2.1873, it makes use of the reciprocal relationship between the patterns produced by X-rays scattered from a molecule and its electronic structure. The major difference in X-ray crystallography compared to SAXS is the molecules being trapped in a protein crystal instead of tumbling isotropically in solution. This not only comes with the possibility to collect 3D X-ray diffraction data sets but also to collect high-resolution data. However, the limiting factor for the use of X-ray crystallography is finding appropriate conditions for crystallization. Especially proteins which switch between multiple conformations often refrain from assembling into ordered crystal lattices. The more flexible a protein is, the less likely it usually will be to crystallize. Thus, X-ray crystallography is mostly applicable to folded proteins but not to disordered proteins unless they can be trapped in a defined state through an interaction partner.

Theoretical background

An ideal crystal is composed of a large number of identical unit cells which are related by translational symmetry. The position of a specific unit cell in a crystal can be described by the translation vectors **a**, **b**, and **c** and the integers t, u, v as $t \cdot \mathbf{a} + u \cdot \mathbf{b} + v \cdot \mathbf{c}$ (**Figure *37*** A). The total wave $\mathbf{K}(\mathbf{q})$ scattered by crystal with n_1 unit cells in direction **a**, n_2 unit cells in **b**, and n_3 unit cells in **c** is obtained by summation over the scattering of all unit cells according to equation 47.

$$\mathbf{K}(\mathbf{q}) = \mathbf{F}(\mathbf{q}) \times \sum_{t=0}^{n_1} f_j \exp(2\pi i \cdot \mathbf{a} \cdot \mathbf{q}) \times \sum_{u=0}^{n_1} f_j \exp(2\pi i \cdot \mathbf{b} \cdot \mathbf{q}) \times \sum_{v=1}^{n_1} f_j \exp(2\pi i \cdot \mathbf{c} \cdot \mathbf{q}) \qquad 47$$

Here, $\mathbf{F}(\mathbf{q})$ is the scattering from a single unit cell. Equation 47 always becomes zero unless $\mathbf{a} \cdot \mathbf{q}$ is an integer h, $\mathbf{b} \cdot \mathbf{q}$ is an integer k, and $\mathbf{c} \cdot \mathbf{q}$ is an integer l (h, k, and l can be positive, negative, or zero). When this so-called Laue condition is fulfilled, all unit cells scatter in phase, and $\mathbf{K}(\mathbf{q})$ is proportional to $\mathbf{F}(\mathbf{q})$ and the number of unit cells in the crystal. The integers h, k, and l are called Miller indices and describe the planes in a crystal lattice from which X-rays are scattered in phase.

Similar to the reflection of X-rays by a molecule (section VI2.18, Figure 33), the reflection of waves from a plane is described by a vector **q** which is perpendicular to the reflecting plane. The position at which a wave is scattered with respect to an origin O is defined by **r** (Figure 37 B). The phase between two waves is given by $\mathbf{r} \cdot \mathbf{q}$ (see section VI2.18) and is the same for all points in the reflecting plane. This is true because for each point, the projection of **r** on **q** has the same length. In a unit cell with the origin O, waves from a reflecting plane through O have $\mathbf{r} \cdot \mathbf{q} = 0$, thus no phase shift. For a parallel plane with $\mathbf{r} \cdot \mathbf{q} = 1$, the phase angles are shifted by $1 \times 2\pi$, for $\mathbf{r} \cdot \mathbf{q} = 2$ by $2 \times 2\pi$ and so on. In other words, all parallel planes with $\mathbf{r} \cdot \mathbf{q}$ being an integer scatter in phase. Considering that the plane $\mathbf{r} \cdot \mathbf{q} = 1$ cuts the **a**-axis in a unit cell at a position defined by $\mathbf{r} = \mathbf{a}/\xi$, it follows that $\mathbf{a}/\xi \cdot \mathbf{q} = 1$ for in-phase scattering. Because the Laue condition requires $\mathbf{a}/h \cdot \mathbf{q} = 1$, ξ must be equal to h (Figure 37 B). The plane with $\mathbf{r} \cdot \mathbf{q} = 1$ consequently cuts **a** at **a**/h. In analogy, this means that the reflecting plane cuts **b** at $\mathbf{b}/k$ and **c** at $\mathbf{c}/l$. Consequently, the planes which reflect waves in phase are the lattice planes of the crystal and are described by the miller indices. The projection of $\mathbf{a}/h$ on **q** has a length of $1/q$, but it is also equal to the distance d between lattice planes. From $d = 1/q$ and $q = 2(\sin\theta)/\lambda$ (see section VI2.18) follows Bragg's law (equation 47).

$$\frac{2d \sin\theta}{\lambda} = 1 \qquad 48$$

‡ Statistics are from October, 2020.

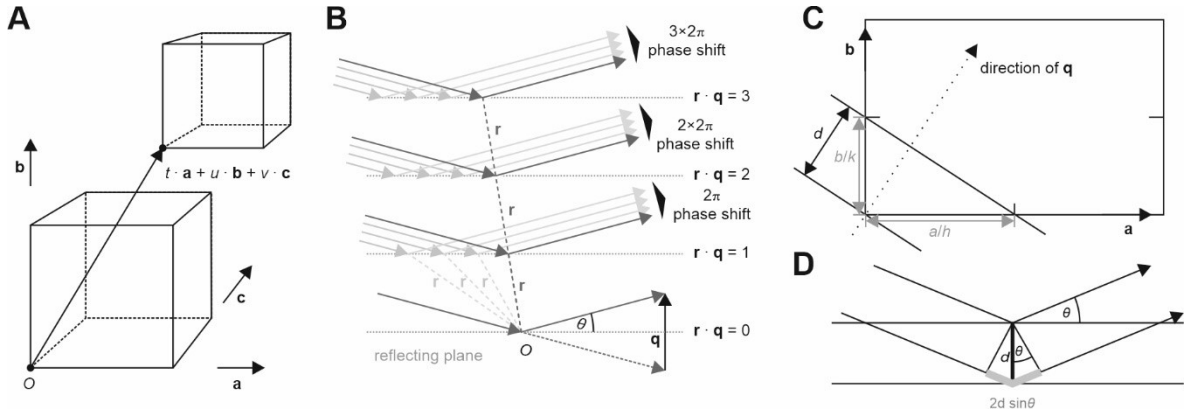


Figure 37: Physical basis of X-ray diffraction from crystals. (**A**) Translational symmetry between unit cells in a crystal. Every unit cell in a crystal can be transferred into any other unit cell by the translation vectors **a**, **b**, and **c** and the integers t, u, v using the operation $t \cdot \mathbf{a} + u \cdot \mathbf{b} + v \cdot \mathbf{c}$. (**B**) Reflection of X-ray waves from parallel planes by an angle θ. The scattering vector **q** is perpendicular to the planes. The phase of waves scattered against a plane is the same in all points of the plane, i.e. the projection of **r** in each point on **q** has the same length. Waves scattered by a reflecting plane going through the origin of the system O have no phase shift ($\mathbf{r} \cdot \mathbf{q} = 0$). A parallel plane with $\mathbf{r} \cdot \mathbf{q} = 1$ scatters waves with a phase shift of 2π, i.e. in phase. All other parallel planes with $\mathbf{r} \cdot \mathbf{q}$ being an integer scatter in phase. (**C**) Relationship between Miller indices h, k, and l with the crystal planes which scatter X-rays in phase. For simplicity, a 2D unit cell with vectors **a** and **b** is shown. The endpoints of the vectors a/h and b/k span a lattice plane perpendicular to vector **q**. All parallel planes separated by an integral multiple of the distance d scatter in phase. (**D**) Illustration of Bragg's law with two parallel reflection planes separated by the distance d. An incident beam is deflected by an angle θ when hitting a reflection plane. The path difference between two beams (grey lines) deflected by parallel planes is given by $2d \sin\theta$. Positive interference occurs when the path difference is an integer of the X-ray wavelength λ.

The physical basis of Bragg's law is illustrated in Figure 37 C. The incident and the reflected waves form an equal angle θ with the reflecting plane. The phase difference between the reflected waves (Figure 37 C, grey line) can be derived via trigonometric rules to be $2d \sin\theta$. Constructive interference occurs when the phase difference is 2π, i.e. a complete wavelength λ. The distance d is related to the Miller indices h, k, and l and the geometry of the unit cell. Importantly, all planes which do not satisfy Bragg's law will lead to destructive interference between the deflected X-rays. In this sense, a crystal acts as a diffraction grating that produces an X-ray diffraction pattern which is characterized by single spots of X-ray intensities, so-called Bragg reflexes. Each reflex can be assigned to a set of miller indices. The intensity of a Bragg reflex $I(h\,k\,l)$ is related to the structure factor $\mathbf{F}(h\,k\,l)$ of the unit cell. $\mathbf{F}(h\,k\,l)$ can be written in analogy to the structure factor of a molecule (equation 28). However, instead of integrating over all atoms in a molecule, $\mathbf{F}(h\,k\,l)$ integrates over all electrons in the unit cell. From inserting $\mathbf{r} \cdot \mathbf{q} = (\mathbf{a} \cdot x + \mathbf{b} \cdot y + \mathbf{c} \cdot z) \cdot \mathbf{q} = \mathbf{a} \cdot \mathbf{q} \cdot x + \mathbf{b} \cdot \mathbf{q} \cdot y + \mathbf{c} \cdot \mathbf{q} \cdot z = hx + ky + lz$, into equation 28 follows equation 49.

$$\mathbf{F}(h\,k\,l) = \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(x\,y\,z) \exp[2\pi i(h\,x + k\,l + l\,z)]dx\,dy\,dz \qquad 49$$

Here, $\rho(x\,y\,z)$ is the electron density at a position in Cartesian space with the coordinates x, y, z. The Fourier transformation of equation 49, with the relationship $\mathbf{F} = F \exp(i\alpha)$, gives the electron density at every position x, y, z as a function of the structure factor $\mathbf{F}(h\,k\,l)$. Because the Laue condition restricts in phase scattering to discrete directions, the transformation is written as a summation instead of an integral.

$$\rho(x\,y\,z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} F(h\,k\,l) \exp[2\pi i(h\,x + k\,l + l\,z) + \mathrm{i}\alpha(h\,k\,l)] \qquad 50$$

The $F(h\,k\,l)$'s can be derived from the intensities of the diffracted waves $I(h\,k\,l)$, i.e. the Bragg reflexes in the diffraction patterns. However, calculating $\rho(x\,y\,z)$ requires the phase angles $\alpha(h\,k\,l)$. The phases cannot be

derived from the Bragg reflexes. Aside from finding appropriate crystallization conditions, the solution of the phase problem can be the bottleneck for high-resolution structure determination with X-ray crystallography. The most common phasing methods are isomorphous replacement, anomalous scattering, and molecular replacement. Their physical theory is very complex and therefore, not discussed hereafter. Details can be found in ref. (*399*).

During X-ray diffraction data collection, a crystal is irradiated with X-rays from a defined orientation. The detected diffraction pattern represents a slice through the three-dimensional molecular transform of the unit cell. However, instead of a continuous pattern, the detected intensities are limited to the Bragg reflexes which account for in-phase scattered X-rays. The positions of the Bragg reflexes reflect the symmetry of the crystal lattice, whereas the intensities of the reflexes contain the information of the electronic structure of the unit cell. In an X-ray diffraction experiment, the crystal lattice works as a signal amplifier that generates intense reflexes even at wide scattering angles which provide structural information at resolutions of around 2 Å or less. How well a crystal diffracts depends on the number of unit cells contained in the crystal and on how 'perfect' the crystal is. Increasing the number of unit cells is like increasing the number of slits in a grating and theoretically comes with more intense reflexes at wide angles. Therefore, a protein crystal typically requires to be sufficiently large for diffraction to high-resolutions. However, with an increasing crystal size comes with an increased risk for crystal defects which counteract the desirable diffraction behaviour.

Diffraction data collected at a given orientation of the crystal with respect to the beam contain 2D structural information of the molecules in the unit cell. By tilting the crystal, a series of orientation-dependent diffraction patterns can be obtained and reassembled into a 3D diffraction pattern. The 3D data set is also termed the reciprocal lattice because the symmetry of the lattice is reciprocal to the symmetry of the crystal. The intensities of the Bragg reflexes in the reciprocal lattice encode the 3D molecular transform of the unit cell. After obtaining the phases of each Bragg reflex, e.g. by molecular replacement or anomalous scattering, the data set can be computed into a 3D electron density map via Fourier transformation. The resolution in the EM density map depends on the number of Bragg reflexes collected in the diffraction experiment. The electron density map is used for building the structural model of the studied protein (*399*).

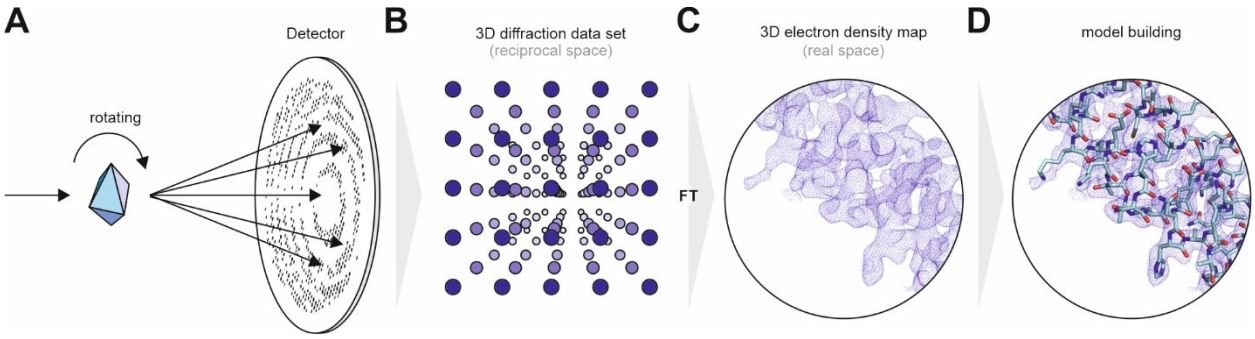


Figure 38: Schematic principle of protein structure determination via X-ray crystallography. (**A**) Schematic depiction of an X-ray diffraction experiment. A protein crystal is irradiated with X-rays, and the intensities of the diffracted intensities are detected on a 2D detector. 3D diffraction data set are obtained from rotating the crystal and collecting a series of orientation-dependent 2D diffraction patterns is recorded. (**B**) The 2D diffraction patterns can be assembled into a 3D data set which represents the 3D molecular transform of the molecules in the unit cell of the crystal in reciprocal space. The lattice of the 3D Bragg reflexes is reciprocal to the real space crystal lattice. (**C**) The 3D diffraction data set is Fourier transformed into a 3D electron density map of the unit cell. (**D**) The electron density map provides the template for model building.

Crystallization of full-length Pacsin3

The described steps were all carried out in a cold room at 4 °C. The purified *G. gallus* full-length Pacsin3 protein (fl-Pac3) was freshly gel filtrated into 20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT using a Superdex200 16/60 column (GE Healthcare). The protein was concentrated to 10 mg/mL using centrifugal concentrators (Merck Millipore) and used for sitting drop crystal screenings with the MemGold™ Screen (Molecular Dimensions). For screenings, 200 nL drops were prepared in Swissci 96-Well 2-Drop MRC Crystallization Plates (Molecular Dimensions) using a mosquito™ pipetting robot (SPT Labtech). Two drops were prepared per screening condition, one with 75 nL protein and 125 nL mother liquor, the other drop with 75 µL mother liquor and 125 µL protein. The reservoir contained 200 µL mother liquor. Crystal growth was observed after two weeks in drops prepared from 125 nL protein and 75 nL mother liquor containing 2.5 M ammonium sulfate, 50 mM HEPES pH 7.5 (MemGold A9 condition). The protein crystals were fished with a 0.4 mm loop; the crystals were cryo-protected with 30% PEG-3000 (1/3 mother liquor, 2/3 parts 60% PEG stock), and flash-frozen in liquid nitrogen.

Data collection and structure determination of full-length Pacsin3

X-ray diffraction data were collected at the DIAMOND beamline (Oxford, UK) from cooled (100 K) single crystals of fl-Pac3 obtained in MemGold condition A9. The best data set diffracted to a resolution of up to 2.7 Å. The data were processed with XDS (*421*) and scaled with Xscale (*422*). The structure of *G. gallus* full-length Pacsin3 was solved via molecular replacement with PHASER MR (*423*) using the *M. musculus* mouse Pacsin3 F-BAR domain X-ray crystal structure (PDB: 3m3w) as a search model (*317*). Structure refinement to a resolution of 2.1 Å was carried out with CCP4 (*424*) and PHENIX (*425*). Manual corrections were carried out in COOT (*426*). Notably, only the F-BAR domain was resolved in the electron density map. The final Ramachandran statistics are 96% for favoured regions, 4% for allowed regions, and no outliers (see Table 10).

2.20 Generation of full-length TRPV4 and Pacsin structural models

TRPV4 models

Full-length TRPV4: For generating a structural model of a full-length *G. gallus* TRPV4 tetramer (section VI3.1, Figure 49) with different IDR conformations, SAXS data collected in this thesis were combined with existing high-resolution cryo-EM data of *X. tropicalis* TRPV4 cryo-EM structure (PDB: 6bbj) (*42*). First, a homology model of TRPV4 was generated with the Swiss Model tool (*427*) using the *X. tropicalis* TRPV4 structure as a template. This yielded an atomic model of *G. gallus* TRPV4 (residues 134-774) in which regions which were not resolved in the cryo-EM structure (e.g. the loop between S5 and P) were modelled as loops by the Swiss Model algorithm. The region preceding the ARD (res. 1-133) and the region succeeding the C-terminal β-strand (res. 775-852) are missing in the homology model. The EOM refined ensembles of the TRPV4 NTD were aligned with their ARD to the ARD of the *G. gallus* TRPV4 homology model in PyMOL (*322, 323*). The RMSD for the alignment was 0.975 Å. Thus, the ARD is an adequate platform for docking the NTDs with the TRPV4 subunits.

TRPV4 NTD: Atomic models of the TRPV4 NTD (Figure 55, Figure 59, Figure 84, Figure 85) were generated from EOM refined NTDs models. The NTD EOM models contain the full set of atomic coordinates of the ARD from a curated *G. gallus* TRPV4 ARD crystal structure (PDB: 3w9g). The IDR is modelled as a chain of beads whereas each bead corresponds to the C_α atom of the respective residue. The backbone and side chains of the IDR was reconstructed from the EOM-derived C_α coordinates using the tool PD2 ca2main (*428*). The obtained NTD structure was afterwards refined manually in COOT to correct for Ramachandran outliers (*426*). Moreover, the PRR was modelled with the average backbone dihedral angles obtained from the NMR analysis of the PBD-PRR peptide (sectionVI3.6, Figure 99).

TRPV4 IDR and PH-PRR: Models of the PH-PRR peptide (see section VI3.3) were generated with the peptide builder in PyMOL (*322, 323*). The model of the Atto488-labeled IDR V148C mutant (section VI3.3, Figure 70) was obtained by reconstructing the backbone and sidechains of the EOM refined IDR conformers with PD2 ca2main (*428*). The Atto488-maleimide was fused to the C148 thiol group of the IDR V148C mutant with the builder tool in PyMOL (*322, 323*).

Pacsin3 models

Pacsin3 'open' conformation: Coordinates of the Pacsin3 linker were generated by running an EOM analysis of the SAXS measurements of the Pacsin3 construct l-SH3 (linker & SH3 domain) using the curated Pacsin3 SH3 domain NMR structure (PDB: 6F55, TRPV4 PRR was removed) as a rigid-body template for the SH3 domain. The C_α atom coordinates of the modelled linker were used to reconstruct the backbone and sidechains with PD2 ca2main (*428*). The N-terminal residue of the l-SH3 structure was then fused to the C-terminus of the *G. gallus* Pacsin3 F-BAR domain X-ray crystal structure (see section VI3.7, Figure 106) in PyMOL (*322, 323*), and the linker was corrected for clashes and Ramachandran outliers in COOT (*426*).

Pacsin3 F-BAR/SH3 complex: The *M. musculus* Pacsin1 X-ray crystal structure (PDB: 2x3x) was used as a template to generate a *G. gallus* Pacsin3 F-BAR/SH3 assembly. The Pacsin1 structure contains an F-BAR dimer with one SH3 domain associated in the hinge region of one F-BAR subunit. The SH3 domain was projected onto the other F-BAR subunit to obtain a symmetric Pacsin1 F-BAR/SH3 complex with a 2:2 stoichiometry. PyMOL (*322, 323*) was used to align the determined *G. gallus* Pacsin3 F-BAR domain X-ray crystal structure (see section VI3.7, Figure 106) the Pacsin1 F-BAR domain. The coordinates of the SH3 domain were extracted from the NMR structure of the *G. gallus* Pacsin3 SH3 domain in complex with the TRPV4 PRR. The Pacsin3 SH3 domains were aligned with the SH3 domains in the Pacsin1 F-BAR/SH3 complex. This yielded a Pacsin3 F-BAR/SH3 domain complex with a 2:2 stoichiometry which was used for all SAXS-based modelling approaches of the clamped Pacsin3 conformation.

3 Results & discussion

Parts of this chapter were published in:

Benedikt Goretzki*, Nina A. Glogowski*, Erika Diehl, Elke Duchardt-Ferner, Carolin Hacker, Rachelle Gaudet, Ute A. Hellmich (2018) Structural Basis of TRPV4 N Terminus Interaction with Syndapin/PACSIN1-3 and PIP_2. ***Structure*** *26*, 1-11, doi: 10.1016/j.str.2018.08.002

Brett A. McCray, Erika Diehl, Jeremy M. Sullivan, William H. Aisenberg, Nicholas W. Zaccor, Alexander R. Lau, Dominick J. Rich, Benedikt Goretzki, Ute A. Hellmich, Thomas E. Lloyd, and Charlotte J. Sumner (2020). Neuropathy-causing TRPV4 mutations disrupt TRPV4-RhoA interactions and impair neurite extension. ***Nature Communications***, in revision

3.1 The TRPV4 N-terminus contains a large, intrinsically disordered region

Recent advances in cryo-EM technology have revolutionized the structural biology of TRP channels (*13*). Within less than seven years, high-resolution structures have been determined via cryo-EM for each TRP subfamily (*155*). Among these, TRP vanilloids are the most intensively studied TRP channels and structures of all vanilloid members have been determined, including TRPV4. However, the distal N- and C-terminal regions of TRPV4 and other vanilloid channels have evaded all structure determination attempts. These regions are emerging as critical regulatory domains of TRP vanilloid channels and therefore demand for detailed structural investigations. The TRPV4 N-terminus preceding the ARD seems to be particularly important for channel regulation. Not only does it contain binding sites for regulatory lipids and proteins (*75–77, 128*). Moreover, up to six point mutations implicated in severe diseases target this region of TRPV4 (*281, 285, 293, 296–298*). Based on the predicted disorder in this region spanning ~130-150 amino acids across different TRPV4 channels (Figure 15), it is subsequently termed the **intrinsically disordered region (IDR)** for this thesis. As a consequence of the high disorder content and in agreement with previous structural studies of TRPV4 (*42*), structure determination of the complete TRPV4 N-terminus using conventional high-resolution methods such as X-ray crystallography or cryo-EM does not appear to be feasible.

Most of the currently best tools for structural and dynamic studies of IDPs or IDRs face inherent problems when it comes to studying large membrane proteins such as the ~440 kDa tetrameric TRPV4 channel. Previous studies on intrinsically disordered regions of membrane proteins have exploited the fact that membrane protein IDRs can be expressed as soluble proteins in isolation or in the context with other cytosolic domains (*205, 429*). As such, they become amenable to biophysical techniques such as NMR spectroscopy or small-angle X-ray scattering (SAXS) which are otherwise substantially complicated by the presence of detergent or lipid molecules or by the large size and oligomeric states of the full-length membrane proteins. A divide and conquer strategy also appears promising to investigate the overall domain architecture of the TRPV4 N-terminus. Structural information obtained from the isolated IDR or from the IDR together with the ARD, which corresponds to the complete TRPV4 N-terminal domain (NTD), can be combined with the existing high-resolution structural data of TRPV4 to generate a model of the IDR in the context the cytosolic N-terminal domain (ARD + IDR) or in the context of the full-length TRPV4 channel (Figure 39). An indispensable prerequisite for a divide and conquer-based structural and dynamic investigation of the TRPV4 N-terminus, however, is the preserved structural integrity of the isolated TRPV4 domains when they are separated from the rest of the channel (*430*).

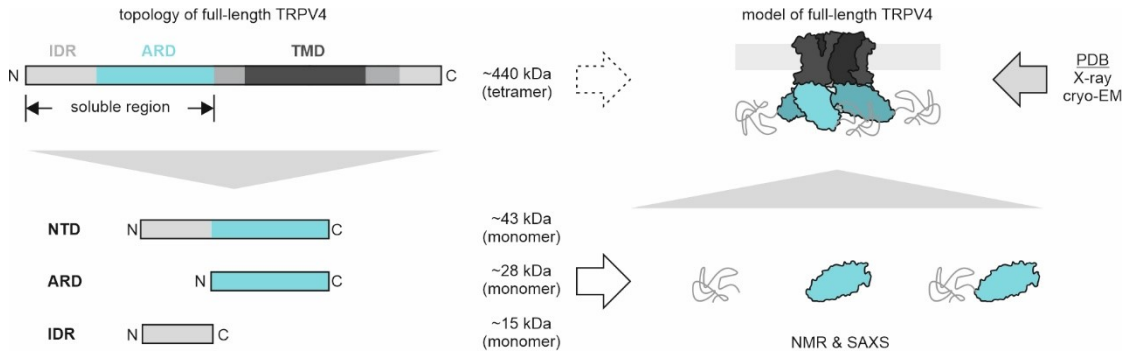


Figure 39: Divide and conquer strategy for studying the disordered TRPV4 N-terminus and generating a full-length TRPV4 model. Due to the large size (~440 kDa) and tetrameric state of a functional TRPV4 channel, structural studies on the TRPV4 IDR in the context of the full-length channel via NMR spectroscopy or SAXS are extremely complicated or sheerly impossible. By expressing the cytosolic TRPV4 NTD, ARD, and IDR as isolated monomers, the proteins become amenable to NMR and SAXS analysis. Structural and dynamic information gathered from the isolated TRPV4 N-terminal constructs can be combined with high-resolution structural information from X-ray crystallography or cryo-EM deposited in the protein data bank (PDB) to generate models of a full-length TRPV4 tetramer with a disordered N-terminus.

Several studies have demonstrated that the isolated ARDs of TRP vanilloid channels are expressed in *E. coli* as soluble and monomeric proteins suitable for structural studies, among them the TRPV4 ARDs from *Gallus gallus* (chicken) and *Homo sapiens* (human) (*18–21, 24, 150, 161*). Considering that membrane protein IDRs are commonly soluble proteins, a construct in which the IDR is attached to the TRPV4 ARD, that is the NTD, likely will remain a soluble and monomeric protein. Based on this assumption, a protocol for the expression and purification of soluble TRPV4 N-terminal constructs suitable for structural studies via NMR and SAXS was established. For this, the region comprising the IDR and the ARD was cloned from the full-length TRPV4 gene into a new vector for expression of the ~42 kDa TRPV4 NTD in *E. coli*. Using the NTD as a template, constructs comprising the isolated IDR (~15 kDa) and the isolated ARD (~28 kDa) were generated. (Figure 39) The expression of IDPs or proteins with large IDRs is often complicated by the fact that disordered protein regions are prone to degradation during expression in the host organism and during purification. Several strategies have been implemented to circumvent such problems, including the fusion of disordered proteins to cleavable, globular proteins (*431, 432*). For the expression of TRPV4 N-terminal constructs, an N-terminal His_6SUMO-tag (hexahistidine-tag fused to a Small Ubiquitin Modifier protein) was chosen as protection against degradation of the IDR during expression and purification. SUMO proteins not only protect against degradation. They also render proteins more soluble (*433*). SUMO-tags can be cleaved off with the protease Ulp1 without leaving non-native residues in the tagged protein. The position of the His_6SUMO-tag in the N-terminus of TRPV4 N-terminal constructs further allows separating degraded from non-degraded proteins via two consecutive Ni-NTA IMAC steps followed by size exclusion chromatography (SEC) (see VI2.5). The protocol established for the expression and purification of TRPV4 N-terminal constructs yields non-degraded proteins entirely composed of native residues. Notably, all constructs miss the first amino acid (M1) due to the insertion of the His_6SUMO-tag between residue 1 and 2.

Because expression screenings of full-length TRPV4 orthologs showed the highest stability for chicken TRPV4 (data not shown), structural characterization of the TRPV4 N-terminus was mainly based on constructs from *G. gallus* (Figure 40 A). The *G. gallus* TRPV4 N-terminus comprises residues 1-382, including both the IDR (residues 1-134) and the ARD (residues 135-382). The phosphoinositide binding domain (PBD) is formed by the sequence ^{107}KRWRR^{111}, the proline-rich region (PRR) comprises residues ^{124}PAPNPPP^{130}. TRPV4 N-terminal domain constructs from *Homo sapiens* (human) were used to verify that structural features observed in *G. gallus* constructs are conserved across species. In the *H. sapiens* TRPV4 protein, the IDR comprises residues 1-148, the ARD comprises residues 149-398. The *H. sapiens* TRPV4 PBD contains the sequence ^{121}KRWRK^{125}, the PRR is formed by residues ^{138}PAPQPPP^{145} (Figure 40 A). Using the expression and purification procedure outlined above and in section VI2.5, the *G. gallus* and *H. sapiens* TRPV4 ARD, NTD, and IDR could be obtained

at high purity and monodispersity as indicated by SDS-PAGE and size exclusion chromatography (SEC) analysis (Figure 40 B and C). Noticeably, the elution volumes (V_e) of ARD, IDR, and NTD from both *G. gallus* and *H. sapiens* observed in the SEC profiles do not follow the trend expected for globular proteins with the calculated molecular weight (*MW*) of ARD, IDR, and NTD. In SEC, the elution volume correlates inversely with the size and shape of a particle. Accordingly, large and elongated proteins elute at low volumes, whereas small and compact proteins elute at high volumes. The observation that the IDR elutes at lower V_e than the ARD, despite its lower *MW*, might indicate an oligomeric state of the IDR or an extended conformation. Thus, size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) was carried out and confirmed that the *G. gallus* TRPV4 NTD, ARD, and IDR are monomers in solution (Figure 40 D). The determined molecular weight values of ARD (26.6 ± 0.1 kDa), IDR (16.4 ± 0.2 kDa), and NTD (44.5 ± 0.2 kDa) are close to the calculated monomer molecular weights (Table 7). Therefore, the low elution volume of the IDR strongly indicates an extended conformation, which increases the hydrodynamic dimensions of the protein.

Several empirical scaling laws describe the relationships between the hydrodynamic radius, termed Stokes radius (R_s), and the molecular weight of native, unfolded, or partly unfolded proteins (section VI2.8, equations 2-8) (*328*). The scaling laws provide a versatile tool for a conformational state analysis of the ARD, IDR, and NTD. For this, SEC experiments were carried out with a mix of globular proteins with available R_s. The resulting linear relationship between V_e and $\log(R_s)$ (Figure 40 F) enabled to determine the Stokes radii of the NTD, ARD, and IDR from their corresponding SEC elution volumes which could then be compared to the molecular weight of the proteins, MW_{calc} (Table 7). The R_s values determined for the isolated ARD from both *G. gallus* and *H. sapiens* agree well with a globular protein state (Figure 40 G, Table 7). For IDR and NTD, the calculated R_s values are substantially increased compared to the theoretical R_s values expected for globular IDR and NTD monomers (Table 7). The IDR R_s values are comparable with a native coil-like protein, whereas the NTD R_s values indicate a premolten-globule state (Figure 40 G). The premolten-globule state describes an intermediate between unfolded and molten globule states (*328*). It contains a mixture of folded and unfolded regions.

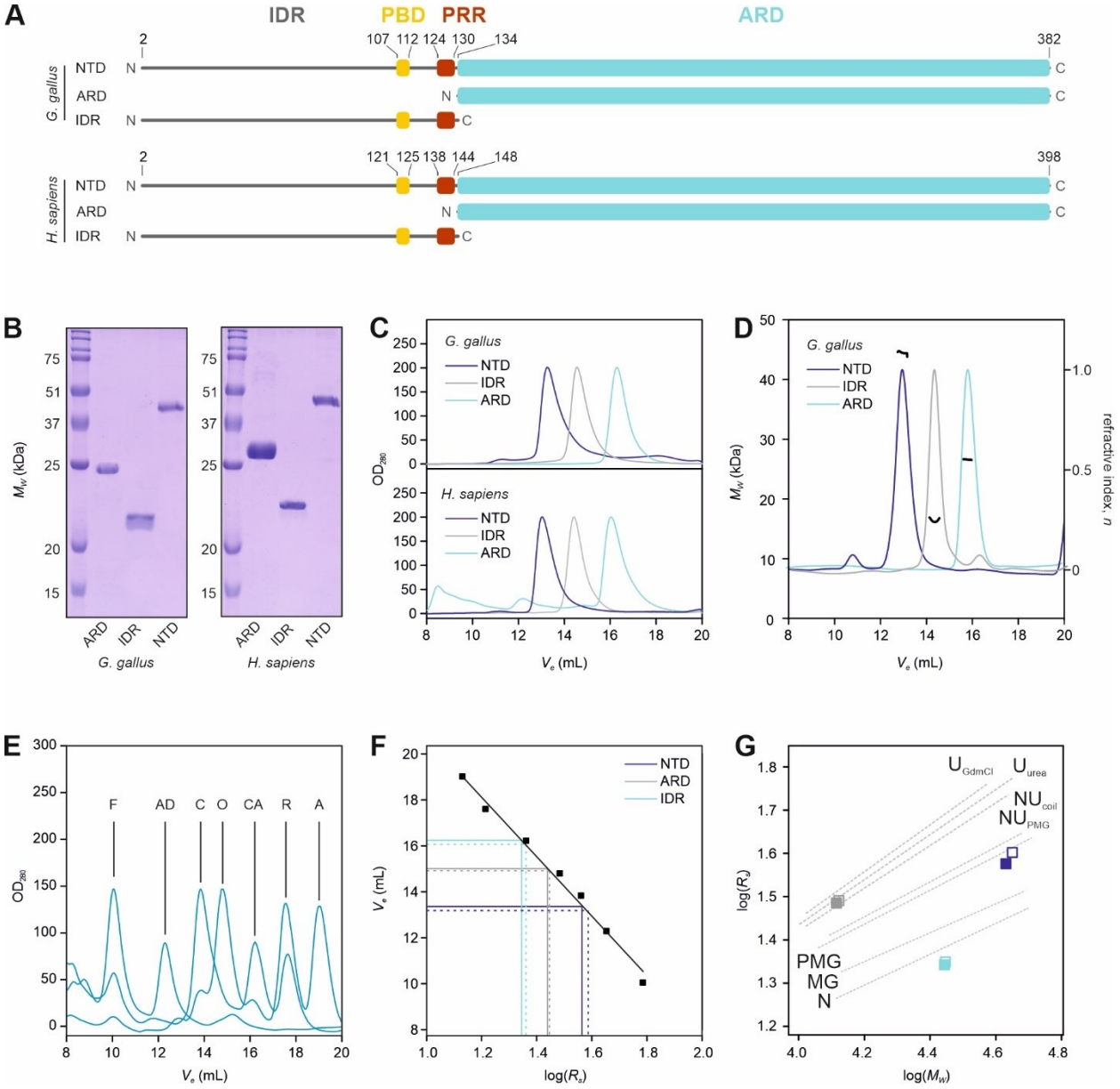


Figure 40 Size exclusion chromatography analysis of TRPV4 N-terminal proteins from *G. gallus* and *H. sapiens*. (**A**) Topology of the TRPV4 N-terminal constructs. The N-terminal domain (NTD) comprises residues 2-381 (*G. gallus*) and residues 2-398 (*H. sapiens*), the isolated Ankyrin repeat domain (ARD) residues 135-381 (*G. gallus*) and residues 149-398 (*H. sapiens*), and the intrinsically disordered region (IDR) comprises residues 2-134 (*G. gallus*) and residues 2-148 (*H. sapiens*). (**B**) SDS-PAGE and (**C**) SEC profiles of purified TRPV4 constructs from *G. gallus* and *H. sapiens*. The SEC profiles are normalized to an OD_{280} = 200 of the peak maxima. (**D**) SEC-MALS analysis of the *G. gallus* TRPV4 NTD, ARD, and IDR. Plotted are the molecular weight (*MW*) and the refractive index (*n*) versus the elution volume (V_e). The determined molecular weights classify the NTD, ARD, and IDR as monomers. The determined *MW*s deviate from the calculated monomer values by 5.4%, 4.5%, and 11.3%, respectively. NTD: 44.5 ± 0.2 kDa, ARD: 26.6 ± 0.1 kDa, IDR: 16.4 ± 0.2 kDa. (**E**) SEC profiles of protein standards used for Stokes radius (R_s) estimation. F: Ferritin, AD: Alcohol dehydrogenase, C: Conalbumin, O: Ovalbumin, CA: Carbonic anhydrase, R: Ribonuclease A, A: Aprotinin (see section VI2.8) for *MW* and R_s values). (**F**) Calibration curve (black) generated from the elution volumes of the protein standards in (E). Plotted is the elution volume (V_e) versus log(R_s) with the linear fit $V_e = 33.86 - 13.07 \cdot \log(R_s)$, where V_e is in mL and R_s in Å. R_s estimates of TRPV4 constructs were obtained from their SEC elution volumes (see Table 7). The solid lines indicate *G. gallus,* and the dashed lines indicate *H. sapiens* constructs. (**G**) Conformational analysis of NTD, ARD, and IDR from their estimated R_s values and calculated *MW*. Plotted is the log(R_s) versus log(*MW*). Empirical scaling relationships between log(R_s) and log(*MW*) for proteins in different conformational states are indicated as dotted lines. N: Native; MG: molten globule; PMG: pre-molten globule; NU_{PMG}: native pre-molten globule-like; NU_{coil}: native coil-like; U_{urea}: urea unfolded; U_{GdmCl}: guanidinium-hydrochloride unfolded (*328*). The ARD can be classified as a native protein, the NTD as a pre-molten globule-like protein, and the IDR as a native coil-like protein (solid boxes: *G. gallus* constructs, empty boxes: *H. sapiens* constructs). All SEC-experiments were carried out at 4°C with a Superdex 200 16/60 column.

Table 7: Estimated Stokes radii of *G. gallus* and *H. sapiens* TRPV4 NTD, ARD, and IDR. Stokes radii (R_s) were calculated from the SEC elution volumes (V_e) using the calibration curve in (Figure 40 F) Theoretical R_s values assuming monomeric and globular states under native conditions, termed R_s^N, were calculated according to ref. (*328*) based on the calculated molecular weights of ARD, IDR, and NTD monomer (MW_{calc}).

Construct	*G. gallus*				*H. sapiens*			
	V_e (mL)	R_s (Å)	R_s^N (Å)	MW_{calc} (kDa)	V_e (mL)	R_s (Å)	R_s^N (Å)	MW_{calc} (kDa)
ARD	16.3	22.1	24.2	28.0	16.0	23.1	24.6	29.3
IDR	14.6	30.0	19.0	14.2	14.4	30.8	19.6	15.6
NTD	13.4	36.7	28.1	42.5	13.0	39.3	28.4	43.8

The SEC-based conformational analysis of the ARD, IDR, and NTD (Figure 40 F) suggests a substantial degree of coil-like regions in the TRPV4 N-terminus, predominantly clustering in the IDR. To further investigate the degree and distribution of disorder in the NTD, the secondary structure of the *G. gallus* and *H. sapiens* ARD, IDR, and NTD were assessed via far-UV circular dichroism (CD) spectroscopy (Figure 41). The CD spectrum of the ARD shows two minima at ~208 and ~222 nm, which are characteristic for highly α-helical proteins (Figure 41 A and B). The minima are preserved in the NTD, however, with lower amplitude. The IDR CD spectrum displays a single minimum at ~200 nm, which is characteristic of the absence of secondary structure. The CD spectroscopy-based secondary structure analysis of the ARD via BeStSel (65% α-helix, no β-sheet, 35% random coil) (*434*) is in line with the high α-helical content (~62%) observed in the X-ray crystal structure of the isolated ARD (PBD: 3w9g) (*150*) (Figure 41 C). The α-helical content is reduced to ~40% in the NTD in favour of an increased fraction of disorder of ~25%. No BeStSel analysis based on the IDR CD spectrum was carried out due to its inaccuracy for predicting highly disordered proteins (*434*). Instead, the secondary structure content of the IDR was assessed quantitatively from the residue-weighted difference between the BeStSel predictions of NTD and ARD, termed 'Δ(NTD-ARD)BeStSel'. In line with the absence of secondary structure features in the CD spectrum of the IDR, this approach reveals that the IDR is almost entirely disordered. In agreement with the R_s-*MW* comparison of ARD, IDR, and NTD (Figure 40 F), the CD spectroscopic analysis classifies the NTD as a two-domain protein composed of the globular α-helical ARD and the highly disordered IDR.

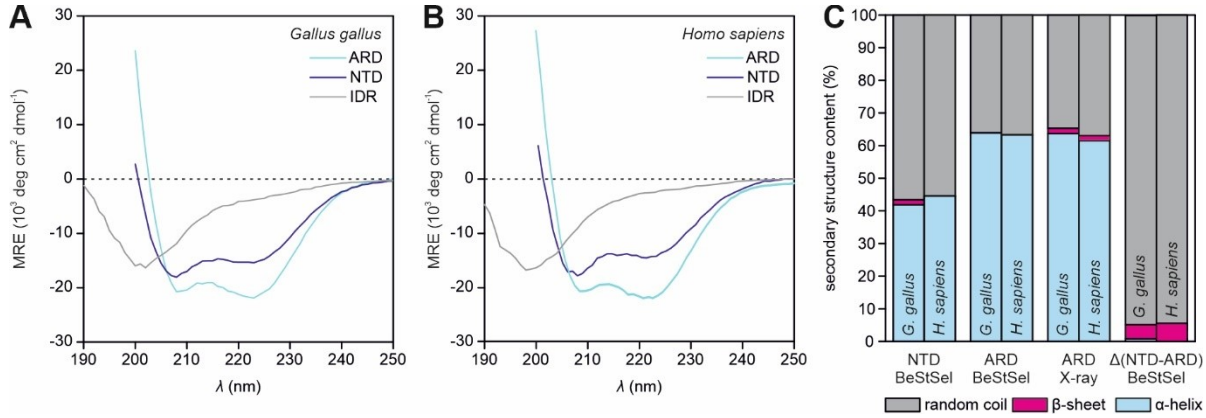


Figure 41: Secondary structure analysis of *G. gallus* and *H. sapiens* TRPV4 ARD, IDR, and NTD via far-UV CD spectroscopy. Far-UV CD spectra of (**A**) *G. gallus* and (**B**) *H. sapiens* ARD (cyan), IDR (grey), and NTD (blue). Spectra were recorded at 293 K, 0.3 mg/mL protein concentration, in 80 mM Tris pH 7, 20 mM NaCl. Plotted is the mean residue ellipticity MRE versus the wavelength λ. (**C**) Secondary structure of ARD, IDR, and NTD analysis via BeStSel (*434*). 'NTD BeStSel' and 'ARD BeStSel' indicate the secondary structure content of the NTD and ARD based on the spectra displayed in (A) and (B). 'ARD X-ray' indicates the secondary structure content derived from the available X-ray crystal structures of *G. gallus* TRPV4 ARD (PDB code 3w9g) and *H. sapiens* TRPV4 ARD (PDB code 4dx1). The secondary structure content of the IDR was estimated from the residue-weighted differences between the ARD and NTD BeStSel analyses, termed 'Δ(NTD-ARD) BeStSel'. Bars on the left represent *G. gallus,* and bars on the right represent *H. sapiens* constructs.

To further investigate the structure of the TRPV4 N-terminal domain, the ARD, NTD, and IDR from *G. gallus* TRPV4 were subjected to SEC-coupled small-angle X-ray scattering (SAXS) analysis (for experimental details see section VI2.18). The SEC-SAXS traces and the Guinier analyses (Figure 42 A and C) confirm that the recorded scattering profiles (Figure 42 B) are obtained from monodisperse proteins. The ARD generates a scattering profile typical for a globular protein. The narrow distance distribution with a radius of gyration R_g = 2.5 (Figure 42 D) and the bell-shaped Kratky-plot (Figure 42 E) are indicative of a compact structure. While the Kratky plot of a perfectly globular protein tends towards zero at high $q \cdot R_g$ values, the distances up to D_{max} = 11.5 nm in the real-space distance distribution and the ARD Kratky plot increasing to higher values at $q \cdot R_g > 6$ indicate a short and flexible extension protruding from a globular domain. The IDR generates a scattering profile that is commensurate with a random chain-like conformation. The asymptotic behaviour towards $(q \cdot R_g)^2 \cdot I \cdot I_0^{-1} \approx 2$ in the dimensionless Kratky plot (Figure 42 E) is also entirely consistent with an intrinsically disordered protein (*395*). The real-space distance distribution reveals a radius of gyration of R_g = 3.5 nm with a maximal particle dimension of D_{max} = 14.0 nm. The NTD scattering profile indicates a multidomain protein consisting of globular and disordered regions. The distance distribution has a similar shape as that of the IDR, suggesting that the NTD is a highly flexible protein. The NTD has a larger radius of gyration of R_g = 4.1 nm than the IDR and displays a maximum particle dimension of D_{max} = 19 nm. The presence of a flexible region in the NTD becomes evident from increasing values beyond $q \cdot R_g > 8$ in the Kratky-plot. The bell-like shape between $0 < q \cdot R_g < 8$, however, indicates the presence of a globular domain in the NTD. This strongly suggests that the NTD consists of the globular ARD attached to the highly flexible IDR (Figure 42 F).

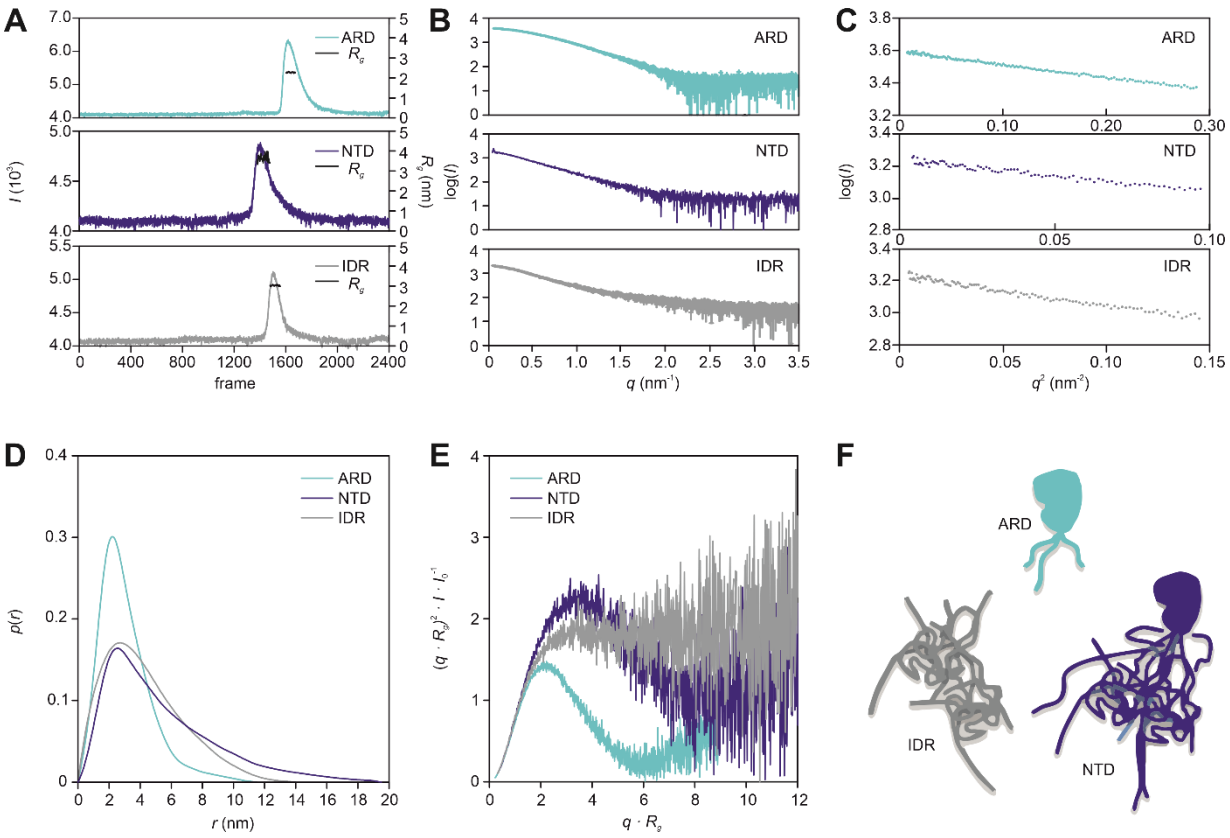


Figure 42: Small-angle X-ray scattering analysis of the *G. gallus* TRPV4 N-terminus. (**A**) SEC profiles of ARD (cyan), NTD (blue), and IDR (grey). Plotted is the scattering intensity *I* at the lowest detected *q* value (left axis) versus the frame number. The frames included in the final scattering profile and the corresponding radius of gyration (R_g) are indicated in black (right axis). (**B**) X-ray scattering profiles of ARD (cyan), NTD (blue), and IDR (grey). Plotted is the logarithmic scattering intensity log(*I*) versus the scattering vector *q*. (**C**) Guinier-plot of ARD (cyan), NTD (blue), and IDR (grey). Plotted is the logarithmic scattering intensity log(*I*) versus the squared scattering vector q^2. (**D**) Normalized real-space distance distribution, *p*(*r*), of ARD (cyan), NTD (blue), and IDR (grey). (**E**) Dimensionless Kratky-plot of ARD (cyan), NTD (blue), and IDR (grey) as $(q \cdot R_g)^2 \cdot I \cdot I_0^{-1}$ plotted versus $q \cdot R_g$ (see section VI2.18). (**F**) Schematic representation of the structure and domain organization of ARD (cyan), NTD (blue), and IDR (grey) as inferred from the SAXS and CD measurements (Figure 41).

To verify that the disordered and flexible state of the IDR is conserved across species, the *H. sapiens* TRPV4 NTD and IDR were also included in the SAXS analysis. The SEC-SAXS trace and Guinier analysis (Figure 43 A and C) confirm the high quality of the collected SAXS data (Figure 43 B). The real-space distance distribution and dimensionless Kratky-plots of the *H. sapiens* TRPV4 NTD and IDR (Figure 43 D and E) show very similar features as observed for the *G. gallus* TRPV4 constructs (Figure 42 D and E). The *H. sapiens* TRPV4 IDR (R_g = 3.4 nm, D_{max} = 13.5 nm) can thus also be classified as an intrinsically disordered protein, whereas the NTD (R_g = 4.6, D_{max} = 19 nm) behaves as a globular protein with a large flexible region (Figure 43 F). Differences in the determined R_g and D_{max} values between *G. gallus* and *H. sapiens* constructs might be attributed to the longer amino acid sequence of the *H. sapiens* IDR (148 residues) compared to the *G. gallus* IDR (133 residues). Together with the SEC and CD spectroscopic analysis, and SAXS data confirms that the region preceding the ARD in the TRPV4 N-terminus, the IDR, is a highly disordered protein domain. In good agreement with sequence-based secondary structure and disorder predictions, the overall domain organization of the NTD seems to be conserved between *G. gallus* and *H. sapiens* TRPV4. Structural and dynamic features inferred from *G. gallus* TRPV4 constructs can therefore be considered to exist across species.

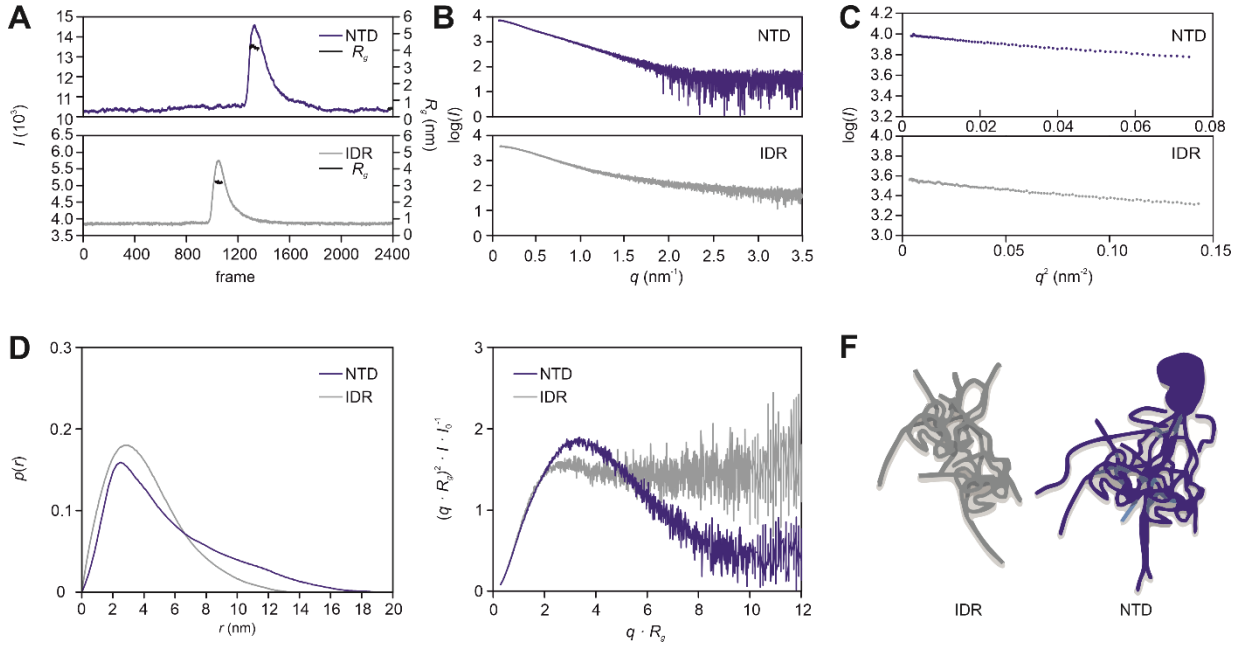


Figure 43: Small-angle X-ray scattering analysis of the *H. sapiens* TRPV4 N-terminus. (**A**) SEC profiles of NTD (blue), and IDR (grey). Plotted is the scattering intensity *I* at the lowest detected *q* value (left axis) versus the frame number. The frames included in the final scattering profile and the corresponding radius of gyration (R_g) are indicated in black (right axis). (**B**) X-ray scattering profiles of NTD (blue), and IDR (grey). Plotted is the logarithmic scattering intensity log(*I*) versus the scattering vector *q*. (**C**) Guinier-plot of NTD (blue), and IDR (grey). Plotted is the logarithmic scattering intensity log(*I*) versus the squared scattering vector q^2. (**D**) Normalized real-space distance distribution, *p*(*r*), of NTD (blue), and IDR (grey). (**E**) Dimensionless Kratky-plot of ARD (cyan), NTD (blue), and IDR (grey) showing $(q \cdot R_g)^2 \cdot I \cdot I_0^{-1}$ plotted versus $q \cdot R_g$ (see section VI2.18). (**F**) Schematic representation of the structure and domain organization of ARD (cyan), NTD (blue), and IDR (grey) as inferred from the SAXS and CD measurements (Figure 41). Importantly, findings for the *H. sapiens* TRPV4 N-terminus are in agreement with observations for the respective *G. gallus* constructs (Figure 42).

With approximately 20%, the IDR makes up for a substantial fraction of the entire TRPV4 sequence (Figure 15). Besides containing binding sites for channel regulators (*75–77, 128*), little is yet known about the structure and dynamics of the large disordered region in the cytosolic TRPV4 N-terminus. Intrinsically disordered protein regions such as the TRPV4 IDR are inadequately described by a single structure. Instead, their structures are better depicted by large ensembles of rapidly interconverting conformations. A thorough structural and dynamic description is a necessary step for understanding whether and how the conformational ensemble of the IDR contributes to the functional state of the channel.

The CD spectrum of the TRPV4 IDR (Figure 41) displays the IDP-characteristic minimum at ~200 nm and shows no detectable secondary structure-characteristic features. In combination with SAXS measurements, this suggests that the IDR is highly unstructured. However, the absence of notable secondary structure-characteristic features in the CD spectrum of the TRPV4 IDR in the native state does not rule out the presence of residual secondary structures in the IDR. Residual or transiently forming structures in IDPs are difficult or impossible to identify from CD spectra via visual inspection or BeStSel analysis (*434*). Urea is known to unfold proteins and disrupt secondary structures. Spectral changes and loss of secondary structure-characteristic features in IDPs upon treatment with denaturant can, therefore, report on residual structures in the native states of the proteins (*435*). The *G. gallus* and *H. sapiens* TRPV4 IDR both display very low CD-signals at the secondary structure characteristic wavelength of 222 nm in the native state (Figure 44 A and B). Upon urea-treatment up to a concentration of 10 M, the CD-signal diminishes completely. This strongly suggests the existence of residual or transient secondary structures in the native state of the IDR.

Many IDPs have been found to regulate protein function through disorder-order transitions in response to changes in the environment or ligand binding (*436*). To probe the ability of the IDR to form secondary structures, CD spectra of the *G. gallus* and *H. sapiens* TRPV4 IDR were recorded at increasing concentrations of trifluoroethanol (TFE) (Figure 44 C and D). TFE promotes the formation of secondary structures by increasing the hydrophobicity of the solution and thereby strengthening intramolecular hydrogen-bonding, which results in disorder-order transitions in unfolded peptides or IDPs (*437*). The *G. gallus* and *H. sapiens* TRPV4 IDR both respond to TFE with the formation of α-helical structures, as indicated by the minima formed at 208 and 222 nm. However, high concentrations of solvent were required to observe significant structural changes (>20% TFE) and even at the maximum concentration of 60% TFE, the CD spectra show only low α-helical content. This indicates a low probability of the IDR to undergo secondary structure formation.

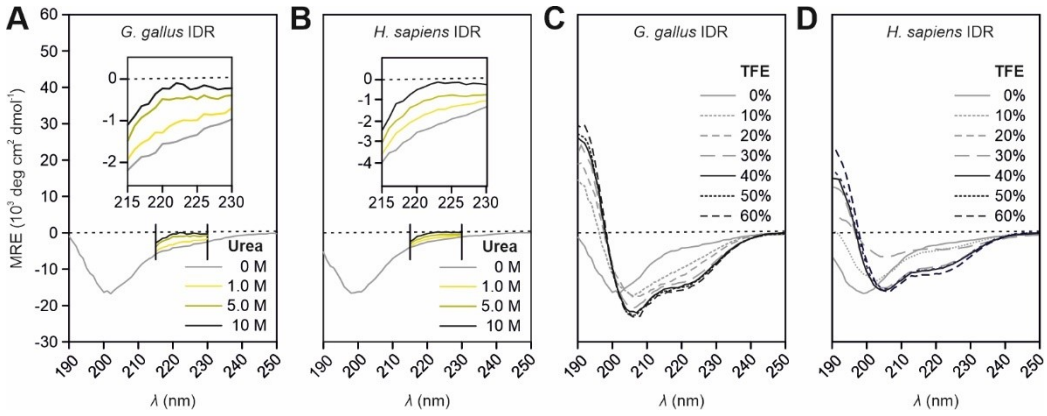


Figure 44: Secondary structure analysis of the *G. gallus* and *H. sapiens* TRPV4 IDR via CD spectroscopy. Far-UV CD spectra of *G. gallus* and *H. sapiens* TRPV4 IDR at increasing concentrations of (**A**, **B**) urea, and (**C**, **D**) TFE. Spectra were recorded at 293 K, 0.3 mg mL^{-1} protein concentration, in 5 mM Tris pH 7, 2 mM NaCl. Plotted is the mean residue ellipticity, *MRE,* versus the wavelength, *λ*. Due to high detector voltages in the presence of high urea concentrations, CD spectra of the IDR in the presence of urea were limited to wavelengths above 215 nm.

The presence of residual or transient secondary structures in the IDR was further investigated on the per-residue level using nuclear magnetic resonance (NMR) spectroscopy. For this, ^{15}N-labeled *G. gallus* and *H. sapiens* TRPV4 IDR constructs were expressed in *E. coli* and purified as outlined in section VI2.17. As commonly observed for disordered proteins, fast solvent exchange of amide hydrogen atoms caused severe line broadening of various resonances at pH 7. Solvent exchange can be suppressed either at low temperatures or under low pH conditions (*438*). Because CD spectroscopy revealed that the secondary structure of the IDR is sensitive to temperature but not to pH (Figure 45 A and B), all NMR measurements were carried out at pH 4.5 and 298 K. The ^{1}H, ^{15}N-TROSY NMR spectra of *G. gallus,* and *H. sapiens* ^{15}N-IDR display IDP-typical narrow ^{1}H peak dispersions (<1 ppm). Triple resonance NMR experiments and ^{13}C, ^{15}N-labeled constructs (section VI2.17)

enabled near-complete NMR backbone assignments of the IDRs (Figure 45 C and D). The backbone resonances (H^N, H, N, C', C_α, H_α) of the *H. sapiens* ^{13}C, ^{15}N- IDR were assigned to 93%, whereas the backbone resonances of the *G. gallus* ^{13}C, ^{15}N-IDR were assigned to 97% (see section VI2.17, p. 70).

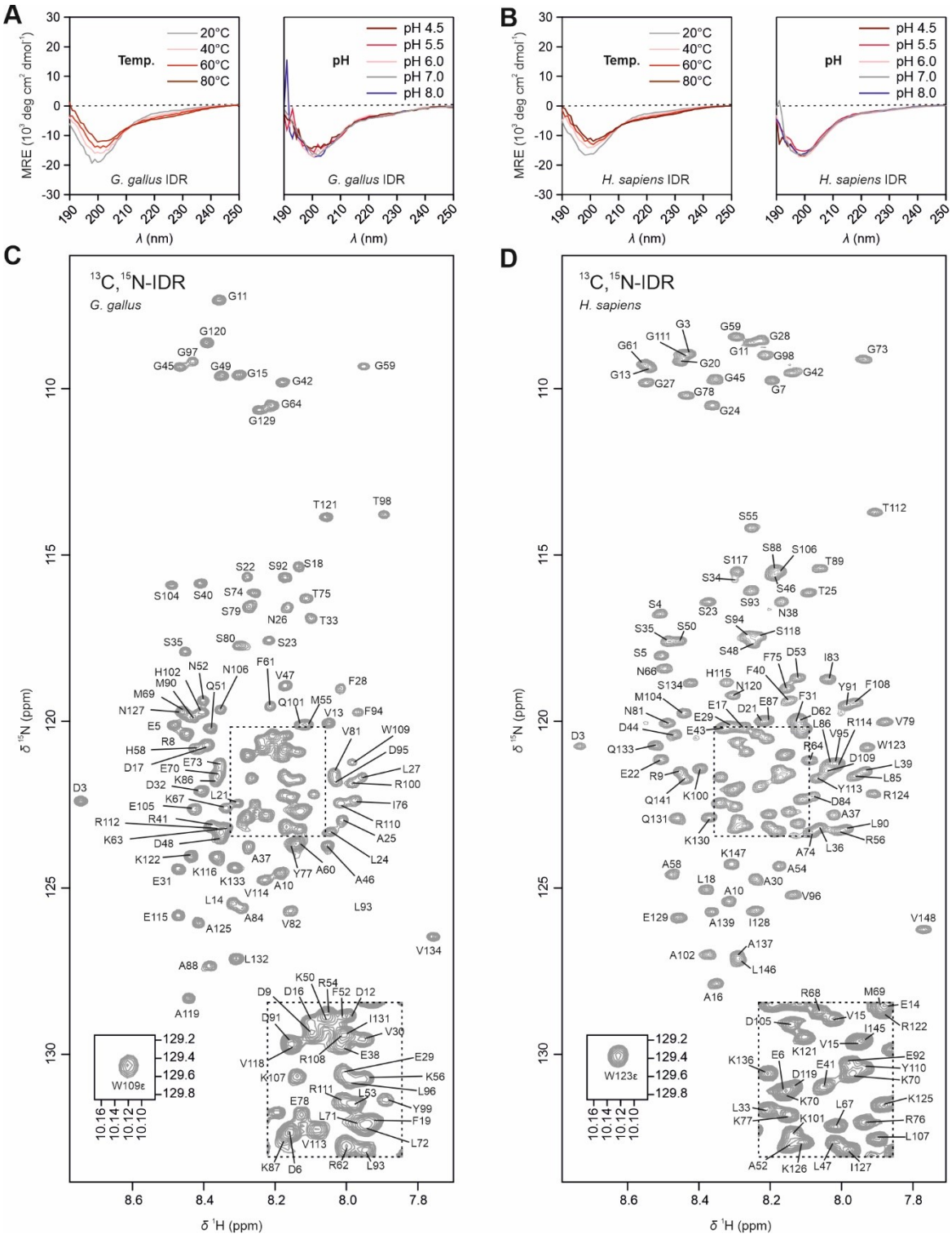


Figure 45: NMR backbone resonance assignment of the *G. gallus* and *H. sapiens* TRPV4 IDR. Far-UV CD spectra of (**A**) *G. gallus* and (**B**) *H. sapiens* TRPV4 IDR at varying pH and temperature. See section VI2.12 for exact experimental conditions. Plotted is the mean residue ellipticity, MRE, versus the wavelength, λ. ^1H, ^{15}N-TROSY NMR spectrum of (**C**) 130 µM ^{13}C, ^{15}N-IDR from *G. gallus* and (**D**) 125 µM ^{13}C, ^{15}N-IDR from *H. sapiens*. Spectra were recorded at 298 K in 20 mM NaP_i pH 4.5, 150 mM NaCl, 1 mM DTT, 10% D_2O. An expanded view of the crowded central region (dashed box) is shown in the bottom right corner of the spectra.

The assigned backbone NMR resonances were used for a chemical shift-based backbone torsion angle and secondary structure prediction for the *G. gallus* and the *H. sapiens* TRPV4 IDR via TALOS+ (section VI2.17, p. 71) (*380*). In line with CD spectroscopic analysis, TALOS+ predicts the absence of secondary structure motives almost throughout the entire IDR sequence. Several regions with an elevated propensity for α-helix or β-sheet structures, however, might point to transient secondary structure formation as also observed in CD-based urea titrations (Figure 44). Importantly, predicted β-sheet structures could also denote extended peptide chains conformations. As a complement to the secondary structure analysis, the residue-specific backbone dynamics of the IDR were probed on the ps-ns timescale via $\{^1H\}$, ^{15}N-hetNOE measurements (section VI2.17, p. 73). Disordered protein regions commonly display very low hetNOE-values below 0 to 0.3, whereas folded domains can exhibit hetNOE-values of 0.6-1.0 (*439*)s. Residual secondary structures in IDPs can reflect in elevated hetNOE-values between 0.4-0.6 (*440*). Consistent with the secondary structure prediction, low hetNOE-values are detected throughout the sequence of both the *G. gallus* and the *H. sapiens* TRPV4 IDR. Several regions with hetNOE-values above 0.5 can be observed and might designate residual or transiently formed secondary structures. However, to confirm whether the elevated hetNOE-values reflect on transient folding requires the determination of additional parameters such as R_1 and R_2 relaxation rates or residual dipolar couplings (RDCs) (*440*).

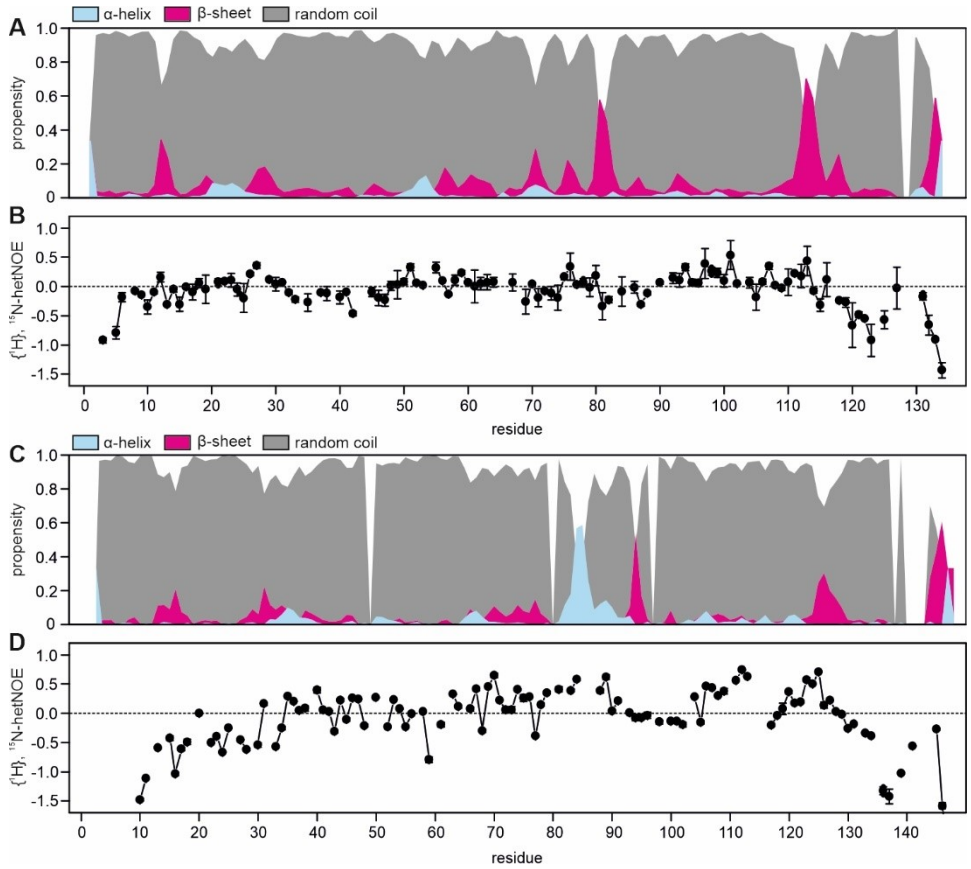


Figure 46: NMR-based structural and dynamic analysis of the TRPV4 IDR. NMR chemical shift-based secondary structure prediction of the (**A**) *G. gallus* TRPV4 IDR using TALOS+ (*380*). Plotted are the propensities for random coil (grey), β-sheet (magenta), and α-helix (light blue) against the residue number. (**B**) Fast backbone dynamics of the *G. gallus* TRPV4 IDR measured via $\{^1H\}$, ^{15}N-hetNOE experiments. Plotted is the hetNOE-value versus the residue number. Low hetNOE values indicate fast dynamics, whereas high values indicate slow dynamics. Similarly, TALOS+ secondary structure prediction (**C**) and $\{^1H\}$, ^{15}N-hetNOE (**D**) were determined for the *H. sapiens* TRPV4 IDR. The error bars in (B) and (D) are the standard deviation from the mean value of two consecutive hetNOE-experiments.

The absence of stable secondary structures paired with the fast backbone dynamics throughout the protein classifies the IDR as a highly flexible intrinsically disordered protein. This observation is entirely consistent with the TRPV4 IDR X-ray scattering profiles, which are commensurate with a random chain-like conformational ensemble (Figure 42). The random chain behaviour of the IDR can be described using conventional polymer theory. Based on Flory's theory, the radius of gyration (R_g) of a polymer follows a power law $R_g = R_0 \cdot N^{\upsilon}$, where N is the number of monomers, R_0 is a constant, and the Flory exponent υ depends on the structural behaviour of the polymer (*441*). The theoretical Flory exponent of a spherical compact globule is $\upsilon = 0.33$, $\upsilon = 0.6$ for a swollen chain, self-avoiding walk, and $\upsilon = 0.5$ for a Gaussian chain (θ-solvent condition) (*441*). Several scaling relationships between R_g and N have been derived empirically from the analysis of published R_g values of unfolded proteins and from the analysis of R_g values of computer-generated IDPs (*407, 442, 443*). Whereas the scaling relationships derived from datasets of denatured proteins yielded υ-values close the theoretical value of swollen chains (Figure 47 A, curve (a) and (b)), the υ-values obtained from computer-generated IDPs are closer to the theoretical value of a Gaussian chain (Figure 47 A, curve (c)). Comparing the SAXS derived R_g values of the 133-residue *G. gallus* and the 147-residue *H. sapiens* IDR with the empirical scaling relationships clearly classifies the TRPV4 IDR as a disordered protein with polymer-like behaviour. The *G. gallus* IDR R_g value is close to what is predicted for a 133-residue chain by scaling law, which has a Flory exponent similar to that of a swollen chain (Figure 47 A, curve (b)) (*443*). The *H. sapiens* IDR, in contrast, has an R_g with almost the expected value of a 147-residue chain described by a scaling law, that has a Flory exponent close to the theoretical value of a Gaussian chain (Figure 47 A, curve (c)) (*407*).

As previously shown, the Flory exponent of a disordered protein can be directly determined from SAXS data (*444*). According to Porod's law, the X-ray scattering intensity I at high scattering wave vectors q shows asymptotic behaviour which follows $I \propto q^{-\zeta}$ (*395*). The Porod exponent ζ can be obtained from the slope in the linear region of a log(I) vs log(q) plot. It is inversely related to the Flory exponent by $\upsilon = 1/\zeta$ *(395)*. For a Gaussian chain, the exponent ζ equals 2, whereas a random chain with solvent-excluded volume exhibits an exponent of $\zeta = 5/3$ (*444*). The log(I) vs log(q) plots of the *G. gallus* and *H. sapiens* TRPV4 IDR were generated from the measured X-ray scattering data (Figure 47 B and C), and the linear mid-q region was identified via visual inspection and fitted via linear regression. The fits yielded $\zeta = 2.18$ and $\upsilon = 0.46$ for *G. gallus* IDR (Figure 47 B) and $\zeta = 2.10$ and $\upsilon = 0.48$ for *H. sapiens* IDR (Figure 47 C). The values place the TRPV4 IDR very close to the θ-solvent condition, thus indicating a highly disordered protein.

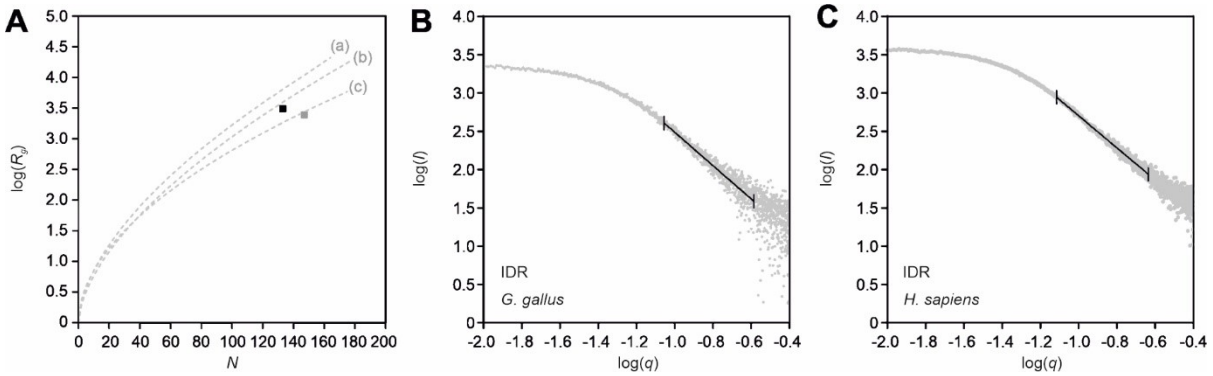


Figure 47: Flory exponent analysis of the *G. gallus* and *H. sapiens* TRPV4 IDR X-ray scattering profile. (**A**) Empirical scaling relationships between the radius of gyration (R_g) and the number of residues in the protein chain (N) found for different datasets of unfolded and disordered proteins (*407, 442, 443*). The scaling law has the form $R_g = R_0 \cdot N^{\upsilon}$ with $R_0 = 0.22$ nm, $\upsilon = 0.60$ for curve (a) (*442*), $R_0 = 0.19$ nm, $\upsilon = 0.58$ for (b) (*443*), and $R_0 = 0.25$ nm, $\upsilon = 0.52$ for (c) (*407*). The *G. gallus* IDR (black box, $R_g = 3.49$ nm, $N = 133$) agrees well with the values for curve (b), which has a Flory exponent close to that of a swollen chain. The *H. sapiens* IDR (grey box, $R_g = 3.39$ nm, $N = 147$), in contrast, agrees best with curve (c), which has a Flory exponent close to that of a Gaussian chain. Log(I) versus log(q) plot of (**B**) *G. gallus* IDR and (**C**) *H. sapiens* IDR. The slope in the linear mid-q region (black line) between log(q) = -1.0 to -0.6 was determined with a linear fit (black line) yielding $\log(I) = (-2.18 \pm 0.03) \cdot \log(q) + (0.27 \pm 0.02)$ for *G. gallus* IDR and $\log(I) = (-2.10 \pm 0.01) \cdot \log(q) + (0.61 \pm 0.01)$ for *H. sapiens*. The Flory exponent was obtained as the negative inverse slope: $\upsilon = 0.46$ for *G. gallus* IDR and $\upsilon = 0.47$ for *H. sapiens* IDR.

The classification as a Gaussian chain implies a random ensemble of IDR conformations with predominantly compact structures. The conformational landscape of the IDR was explored in more detail with a so-called Ensemble Optimization Method (EOM) analysis of the *G. gallus* and *H. sapiens* IDR SAXS data (*394, 419*). As described in section VI2.18 (p. 80), EOM generates 10 000 random structures of disordered chains with N dummy residues and calculates R_g and D_{max} distributions of the resulting random ensemble pool. Using a genetic algorithm, EOM refines the fit of the random pool against the SAXS data, selecting the best sub-set of ensemble-states to represent the experimental profile. For the IDRs, $N = 133$ (*G. gallus*) and $N = 147$ (*H. sapiens*) dummy residues, respectively, were used as input to generate starting structures. The final EOM refined ensembles show good agreement with the experimental data (Figure 48 A, first and third panel). The R_g and D_{max} distributions of the fitted ensembles are approximately as broad as the random pool distributions indicating that the IDR samples a broad range of conformations, i.e. it is extremely flexible. However, the fitted pool distributions indicate that the IDR samples compact conformations more frequently than extended conformations. To validate whether this behaviour is maintained in the context of the NTD, a hybrid-EOM analysis of the *G. gallus* and *H. sapiens* NTDs was carried out. For this, the published X-ray crystal structures of the *G. gallus* and the *H. sapiens* TRPV4 ARD (PDB codes 3w9g and 4dx1, respectively) were used as templates for the ARD-comprising region in the NTD. The rest of the sequence accounting for the IDR was modelled by a chain of dummy residues, in analogy to the EOM analysis of the isolated IDRs. The final EOM refined model of the *G. gallus* TRPV4 NTD showed good agreement with the experimental data (Figure 48 A, second panel). As observed for the isolated IDR, the R_g and D_{max} distribution of the fitted *G. gallus* TRPV4 NTD conformers of are slightly shifted towards smaller R_g and D_{max} values compared to the distributions of the random pool. In contrast to this, the R_g and D_{max} distributions of the fitted *H. sapiens* TRPV4 NTD conformers are shifted to larger values compared to the random pool, indicating a more expanded conformation of the IDR in the context of the *H. sapiens* NTD. The discrepancy between *G. gallus* and *H. sapiens* NTD EOM results might be caused by the presence of 10% glycerol in the human NTD sample, which was required to increase its stability.

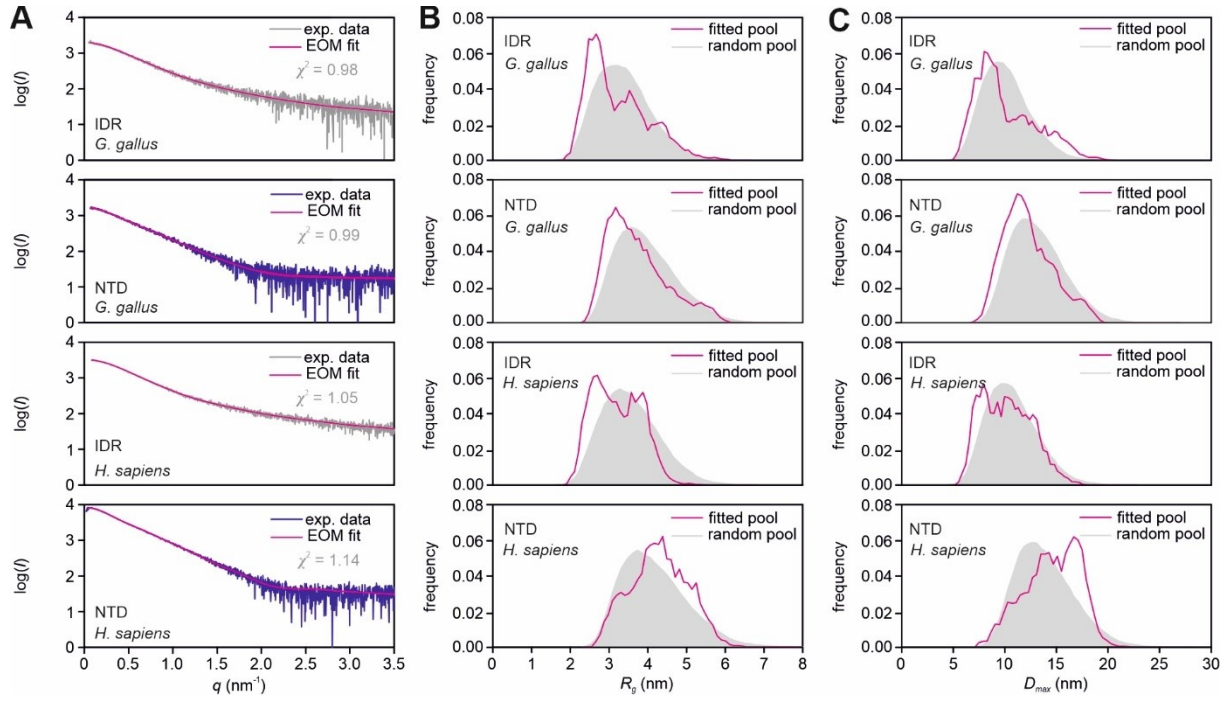


Figure 48: Ensemble optimization method (EOM) analysis of *G. gallus* and *H. sapiens* TRPV4 IDR and NTD. (**A**) Fit of the EOM refined IDR and NTD models (magenta line) to the experimental SAXS data of the *G. gallus* and *H. sapiens* IDR (grey) and NTD (dark blue). Experimental SAXS data are represented as a log(*I*) vs *q* plot. The goodness of the fits is indicated by the χ^2 value. (**B**) R_g and (**C**) D_{max} distributions of the random pool of generated IDR and NTD structures (grey filled curve) compared to the R_g and D_{max} distributions of the EOM refined models (magenta lines). Plotted is the frequency versus the R_g or D_{max} value, respectively.

Based on the EOM refined ensembles of the TRPV4 NTD, a model of a tetrameric TRPV4 assembly with ten IDR conformers in each subunit was generated (Figure 49). Because of the better EOM fit with the experimental data, the *G. gallus* NTD ensembles were used for generating the full-length model. In brief, a structural homology model of a *G. gallus* TRPV4 tetramer was generated by superimposing the *G. gallus* TRPV4 sequence onto the *X. tropicalis* TRPV4 cryo-EM structure (PDB: 6bbj) using the Swiss Model tool (*427*). The EOM refined NTD conformers were aligned along the ARD and afterwards superimposed onto the ARDs of each subunit in the *G. gallus* TRPV4 homology model in PyMOL (*322, 323*). The ARD provides an almost perfect platform for the docking of the NTDs with the TRPV4 subunits, as the ARD in the cryo-EM structure was modelled based on the *H. sapiens* TRPV4 ARD X-ray crystal structure (PDB: 4dx2). The *H. sapiens* and the *G. gallus* ARD X-ray structures (the latter was used for the EOM analysis) are virtually identical (C_α-atom RMSD = 0.349 Å). Conformers of the NTD, which either interfere with other regions of the channel or penetrate the region where the bilayer of the plasma membrane would be expected, were omitted in the final model.

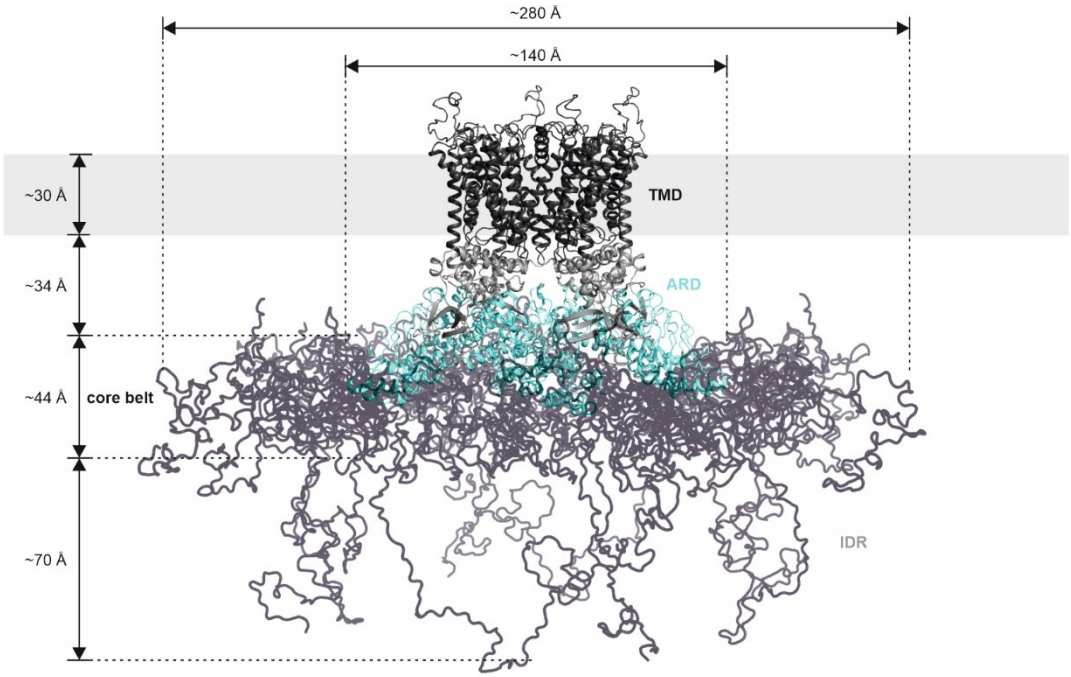


Figure 49: Structural model of a full-length *G. gallus* TRPV4 tetramer. The model was generated by aligning the EOM refined *G. gallus* TRPV4 NTD models with a homology model of the full-length *G. gallus* TRPV4 channel (based on the *X. tropicalis* TRPV4 cryo-EM structure, PDB: 6bbj) along the Ankyrin repeat domain (res. 135-381). Three of the 15 EOM refined conformers in which the IDR clashes with other regions of the channel or with the plasma membrane were omitted. The protein dimensions measured with PyMOL (*322, 323*) are indicated with arrows. The volume filled by most of the IDR conformers is termed the 'core belt'. TMD: transmembrane domain, ARD: ankyrin repeat domain, IDR: intrinsically disordered region.

The generated model of the full-length TRPV4 tetramer impressively demonstrates the structural complexity in the cytosolic domains already imposed by the ensemble of only ten IDR conformers per subunit (Figure 49). The conformational ensemble of the IDRs surrounds the cytosolic ankyrin repeat domains like a belt of protein disorder. The core of this belt is ~44 Å thick and ~34 Å distant from the membrane. It increases the diameter of the cytosolic domain from approximately 140 Å to a maximum of ~280 Å. The volume of the core belt sampled by the IDR increases the overall volume of the cytosolic domain by a factor of ~3. Moreover, the solvent-accessible surface area (SASA) in the cytosolic domain is almost doubled by the disordered nature of the IDR. On average, an IDR conformer has a SASA of ~9500 Å^2, whereas the ARD has a SASA of ~11500 Å^2, even though it contains almost twice as many residues. Notably, conformations of the IDR exist in which the chain extents up to ~150 Å away from the membrane into the cytosol. Such extended conformations, even

though they might be weakly populated in the conformational ensemble of the IDR, further increase the dimension of the TRPV4 cytosolic domain. Notably, the generated model of the full-length TRPV4 tetramer cannot account for potential interactions between IDRs of different TRPV4 subunits or for the interaction of the IDR with the proximal plasma membrane. The initial pool of EOM refined NTD models contained 13 conformers, of which two conformers were dismissed for the alignment with the *G. gallus* TRPV4 homology model due to steric clashes with the region expected to be occupied by the plasma membrane. It can be assumed, that the IDR also samples the space between the core belt and the plasma membrane in a cellular context, particularly when considering that the IDR contains a lipid-binding site, the PBD, close to the ARD. It should be noted that this model assumes a relatively rigid conformation of the TRPV4 regions resolved in cryo-EM studies. However, the cytosolic ARDs may also impose an additional degree of conformational flexibility (see section VI3.2). Moreover, the ~90 putatively disordered residues of the C-terminus are not included in the model and may restrict the space occupied by the IDR in a truly full-length TRPV4 tetramer.

In summary, the divide and conquer-based structural analysis of the TRPV4 N-terminus using SEC, CD, NMR, and SAXS confirm that the region preceding the ARD, i.e. the IDR, forms an intrinsically disordered region. The TRPV4 IDR is highly dynamic and displays random chain-like behaviour. However, EOM analysis of the SAXS data suggests that the conformational landscape of the IDR is not entirely random. In fact, compact conformations are sampled at higher frequencies than expected for a random distribution of conformations. This behaviour is observed both for the isolated IDR as well as for the IDR in tandem with the ARD, i.e. in the context of the NTD. Compact conformations might thus be mediated through long-range interactions within the IDR or between the IDR and the ARD. Alternatively, contraction of the IDR might be achieved through the formation of helical structures. The IDR can adopt extended conformations. Chain expansion in IDPs was previously linked to the formation of polyproline helix (PPII) structures (*395*). The high proline content (~17%) in the IDR might suggest that PPII structures could indeed form and stabilize extended conformers of the TRPV4 IDR. Whether residual or transient secondary structures are present in the IDR cannot be satisfactorily derived from the SAXS, CD and NMR spectroscopic analysis. TFE-induced secondary structure formation demonstrates a low ability of the IDR to form stable secondary structures in hydrophobic environments. Nevertheless, one may hypothesize that transiently formed structures under native conditions are stabilized in response to ligand binding or other changes in the environment of the IDR. Stimulus-induced structural changes in the IDR might be crucial features of channel regulation.

The purified monomeric TRPV4 NTD and IDR constructs provided high-quality SAXS data suitable for the generation of conformational ensembles of the IDR in isolation and in the context with the ARD via EOM analysis. Integrating the EOM refined NTD conformers with available high-resolution data of the full-length TRPV4 channel enabled to build a structural model of a tetrameric TRPV4 assembly with conformational ensembles of the IDR in each subunit. The model exemplifies the architecture of a near full-length TRPV4 tetramer (C-70 terminal residues are still missing) and highlights how the disordered nature of the IDR increases the volume and solvent accessible surface area in the TRPV4 cytosolic domain. The gain of surface area through the presence of the IDR might render the TRPV4 N-terminus more susceptible to the binding of interaction partners. Considering the large volume sampled by the conformational ensemble of the IDR, interaction partners binding to the cytosolic TRPV4 domains are very likely to encounter the IDRs, even when they do not contain the primary binding site. One may even consider that the IDR poses a diffusion barrier that makes it harder for an interaction partner to leave TRPV4 when it reached the periphery of the channel. The presented full-length TRPV4 model can be regarded as a starting point for future refinements and may aid the design of complementary structural and functional studies of TRPV4. An important factor which could not be accounted for in the model is the interaction of the IDR with lipids in the plasma membrane. Such interactions are likely to have huge effects on the conformational landscape of the IDR and the remaining protein. Furthermore, the IDR is not the only region in the cytosolic TRPV4 domain predicted to be disordered. The region comprising the ~70 C-terminal residues of TRPV4 displays similar sequence properties as the IDR and likely exists in a conformational ensemble of disordered states as well (Figure 15). This region will add further complexity to the cytosolic domain and increase the density of disordered protein around the folded core structure of TRPV4.

3.2 The TRPV4 IDR and the ARD are structurally and dynamically coupled

As outlined in section 1.3, the sensitivity of TRPV4 to heat and osmotic stimuli is modulated by the binding of lipids and proteins in the IDR (*76, 77, 128*). It is, however, unclear how ligand-induced structural changes in the IDR can be sensed in the transmembrane domain (TMD) of TRPV4 to modulate the conformational state of the ion channel pore. In view of a potential signal propagation pathway, the ARD is likely to play a key role in linking binding events in the IDR to structural changes in the TMD. The ARD itself acts as a TRPV4 regulatory domain by mediating the binding of ATP (*21, 151*). Due to its position directly underneath the TMD, ligand-induced conformational changes in the ARD can be allosterically converted into conformational changes of the TMD. A structural and dynamic communication between the IDR and the ARD may therefore be efficiently transmitting signals from the IDR via the ARD to the pore region.

To investigate whether the IDR and ARD are structurally and dynamically coupled, a series of deletion constructs of the *G. gallus* TRPV4 N-terminus in which the IDR was successively truncated were generated (Figure 50 A). The constructs are termed according to the number of residues deleted from the N-terminus, i.e. NTD ΔN54, NTD ΔN97, NTD ΔN104, and NTD ΔN120. Similar to wt-NTD, the deletion constructs could be obtained at high purity as soluble proteins (Figure 50 B and C). The SEC elution peaks and the far-UV CD spectra of the deletion mutants indicate monodisperse and properly folded proteins. In analogy to the NTD and ARD (Figure 40), the Stokes radii of the deletion constructs were determined from the SEC elution volumes. According to the scaling laws described in section VI3.1, the deletion constructs NTD ΔN97, NTD ΔN104, and NTD ΔN120 behave similar to native state proteins, whereas the NTD ΔN54 mutant can be classified as a molten-globule like protein. Comparison with the R_s-M_W scaling behaviour of the wt-NTD and the isolated ARD indicates that an increasing length of the IDR shifts the behaviour of the N-terminal constructs from a native to a molten-globule-like state (Figure 50 D). Strikingly, the log(R_s) values of the NTD deletion mutants do not change linearly with the increasing length of the IDR, that is with log(M_W) (Figure 50 D). The SEC elution volume increases from NTD ΔN97 to NTD ΔN104, which translates to a decrease of R_s from 26.0 Å to 25.4 Å. The smaller R_s value of NTD ΔN104 despite the lower M_W protein chain compared to NTD ΔN97 might indicate that the region between residues 97 and 104 enables a more compact protein conformation. Interestingly, this region is very close to the regulatory PIP_2 binding domain (PBD). CD spectroscopic analysis confirms that truncating the IDR does not influence the disordered character of the IDR in the context of the NTD (Figure 50 E). The amplitudes of the α-helix-characteristic minima at 208 and 222 nm decrease almost linearly with the number of IDR residues added to the ARD (Figure 50 F). BeStSel analysis suggests that the spectral changes reflect an increase of disorder at the expense of α-helical content (Figure 50 G). The linear relationship between disorder content and length of the N-terminus is consistent with the observation that the IDR is entirely disordered. The more compact state of NTD ΔN97 compared to NTD ΔN104 is therefore presumably not due to changes in the secondary structure of the IDR when truncating N-terminal residues.

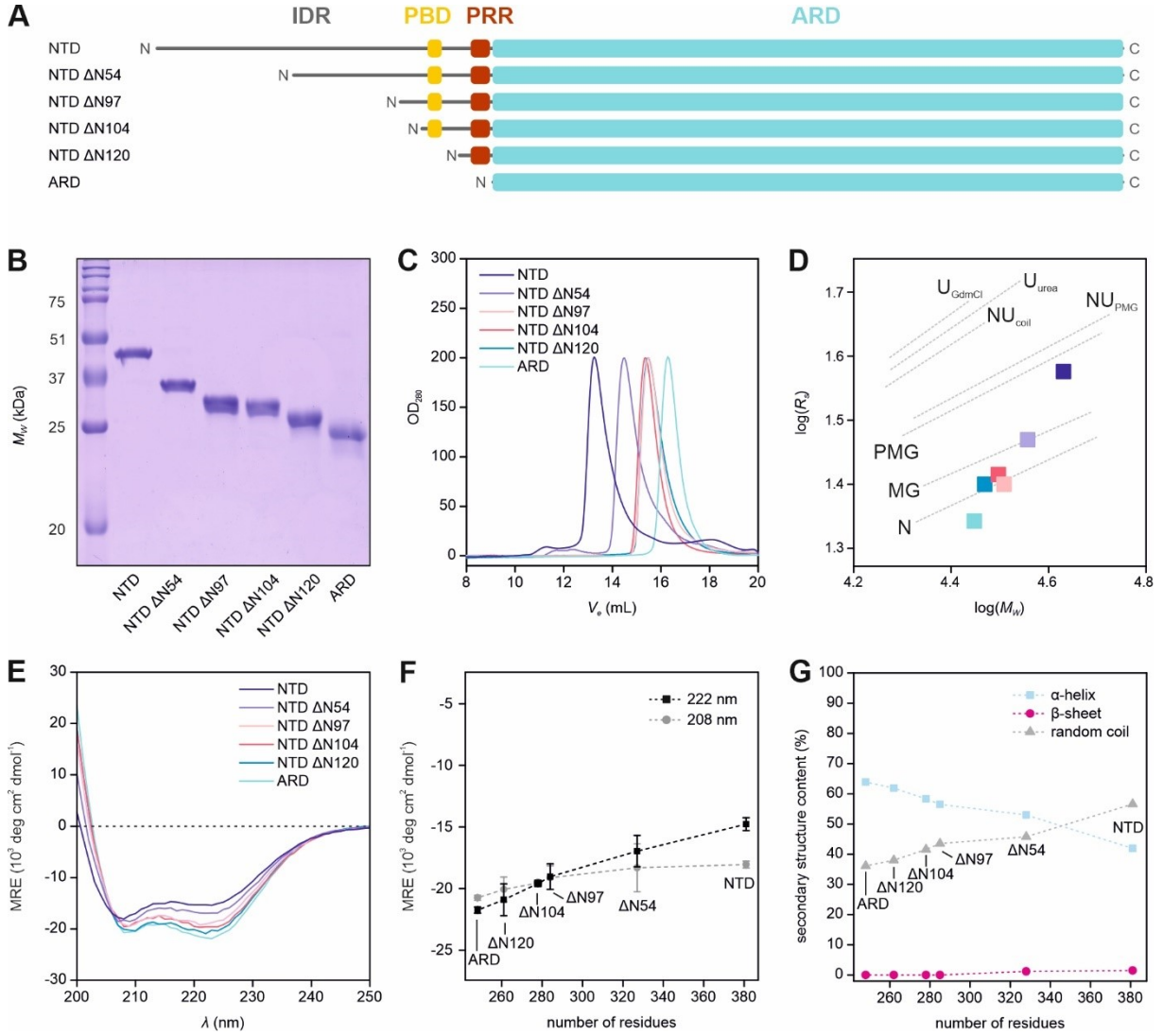


Figure 50: SEC and far-UV CD spectroscopy analysis of *G. gallus* TRPV4 NTD deletion constructs. (**A**) Topology of N-terminally truncated *G. gallus* TRPV4 NTD deletion constructs. (**B**) SDS-PAGE analysis and **C** SEC profiles of NTD deletion constructs. The SEC profiles were normalized to an OD_{280} = 200 of the peak maxima. (**D**) R_s-M_W analysis of NTD deletion constructs. The R_s values were obtained using the calibration curve shown in Figure 40. NTD: V_e = 13.4 mL, R_s = 36.7 Å, M_W = 42.5 kDa; NTD ΔN54: V_e = 14.6 mL, R_s = 29.6 Å, M_W = 37.0 kDa; NTD ΔN97: V_e = 15.4 mL, R_s = 25.4 Å, M_W = 32.3 kDa; NTD ΔN104: V_e = 15.4 mL, R_s = 26.0 Å, M_W = 31.4 kDa; NTD ΔN120: V_e = 15.5 mL, R_s = 25.5 Å, M_W = 29.4 kDa; ARD: V_e = 16.3 mL, R_s = 22.1 Å, M_W = 28.0 kDa. (**E**) Far-UV CD spectra of NTD deletion constructs. (**F**) Mean residue ellipticity (MRE) at 208 nm and 222 nm and (**G**) BeStSel (*434*) based secondary structure analysis plotted versus the number of residues in the NTD constructs. The error bars in (F) are the standard deviation from the mean value of three technical replicates.

The NTD deletion mutants were analyzed via SEC-SAXS to probe the effect of truncating the IDR on the conformation of the NTD in more detail, (Figure 51 A, B, and C) and compared to the wildtype NTD and the isolated ARD (Figure 51 D, E, and F). The deletion mutants NTD ΔN97, NTD ΔN104, and NTD ΔN120 show real-space distance distributions (Figure 51 D) and Kratky plots (Figure 51 E) similar to those of the isolated ARD, thus suggesting rather compact proteins with short and flexible tails. Interestingly, the $p(r)$ curve of the NTD ΔN54 mutant is shifted to a higher R_g value but to a smaller D_{max} than observed for the isolated ARD (Figure 51 D). In line with the Kratky plot, NTD ΔN54 appears to be more compact and more rigid than the isolated ARD (Figure 51 E), despite the ~80 disordered residues contained in the NTD ΔN54 construct. In contrast to the NTD ΔN54 mutant, the wildtype NTD displays a $p(r)$ curve and a Kratky plot typical for a globular domain with a large and flexible tail (Figure 51 E, see also section VI3.1). The NTD contains ~50 additional disordered residues in the IDR than the NTD ΔN54 mutant, which drastically increase the overall size of the protein. Overall, increasing the length of the IDR in the NTD constructs does not manifest as a

linear transition of the $p(r)$ curve or the Kratky plot from the isolated ARD to that of the complete NTD, as one would expect for a globular domain with a disordered chain attached to it (Figure 51 D, E). Instead, the protein dimensions of the NTD constructs, i.e. R_g and D_{max}, change non-linearly with the size of the IDR attached to the ARD (Figure 51 F). The change in R_g and D_{max} observed via SAXS plotted versus the residue number shows a similar trend as the change in R_S observed in SEC experiments (Figure 50, Figure 51 F). The region comprising residues 104-132 seems to increase the protein size, whereas the region comprising residues 54-97 decreases the protein dimensions again.

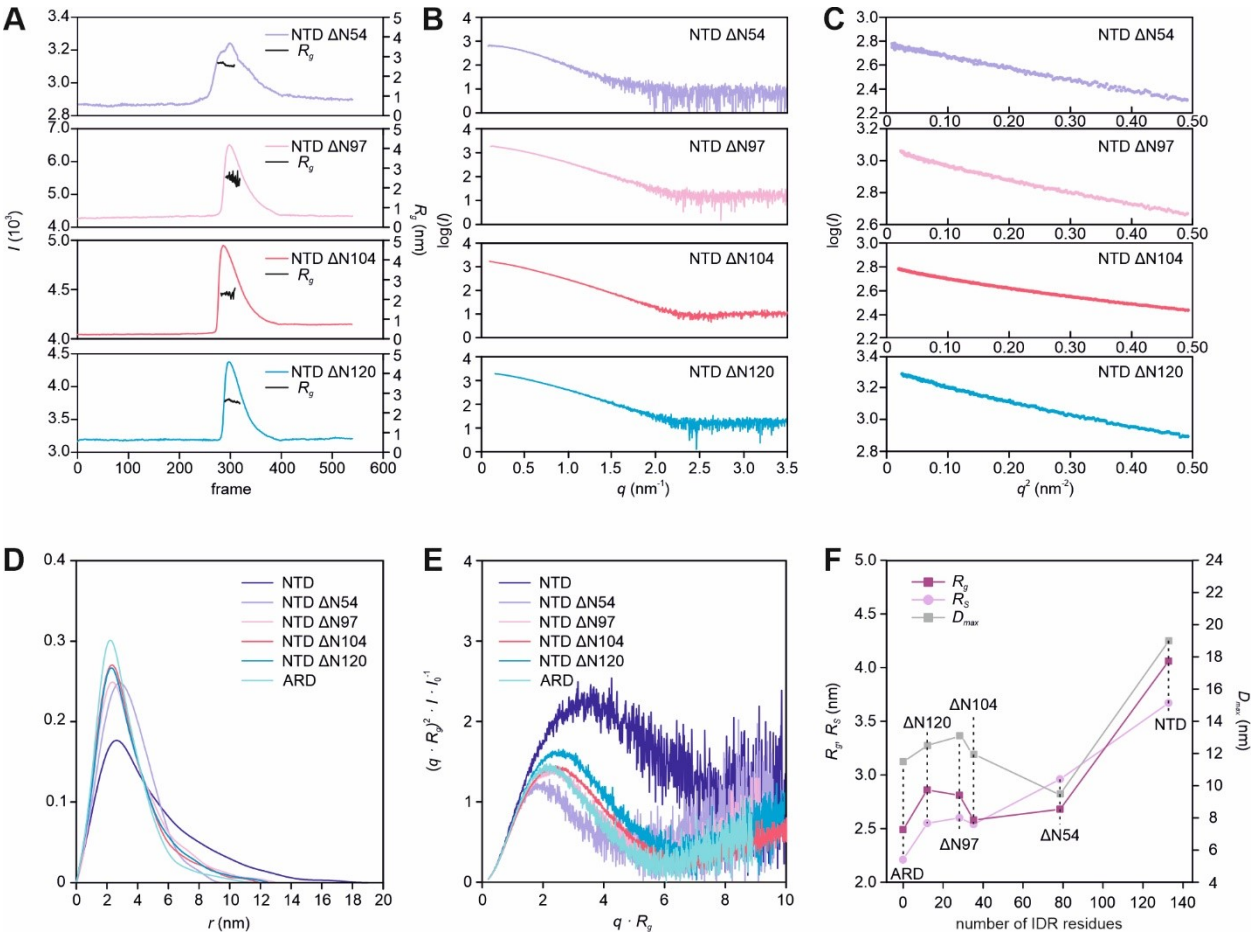


Figure 51: SAXS analysis of *G. gallus* TRPV4 NTD deletion mutants compared to wildtype NTD and the isolated ARD. (**A**) SEC-SAXS profiles of NTD ΔN54 (purple), NTD ΔN97 (pink) NTD ΔN104 (red), and NTD ΔN120 (blue). Plotted is the scattering intensity I at the lowest detected q value (left axis) versus the frame number. The frames included in the final scattering profile and the corresponding radius of gyration (R_g) are indicated in black (right axis). (**B**) Processed X-ray scattering profiles of NTD deletion mutations extracted from the SEC traces shown in (A). Plotted is the logarithmic scattering intensity log(I) versus the scattering vector q. (**C**) Guinier-plot of NTD deletion mutant scattering data shown in (B). Plotted is the logarithmic scattering intensity log(I) versus the squared scattering vector q^2. (**D**) Normalized real-space distance distribution, $p(r)$, of NTD (dark blue), NTD ΔN54 (purple), NTD ΔN97 (pink) NTD ΔN104 (red), NTD ΔN120 (blue), and ARD (cyan). (**E**) Dimensionless Kratky-plot of isolated ARD, NTD, and NTD deletion mutants showing $(q \cdot R_g)^2 \cdot I \cdot I_0^{-1}$ plotted versus $q \cdot R_g$ (see section VI2.18). (**F**) Radius of gyration (R_g) and Stokes radius (R_S) determined from the real-space distance distribution in (D) and the SEC analysis (Figure 50), respectively, plotted versus the number of IDR residues the NTD constructs. The maximum particle dimension (D_{max}) is plotted on a second axis (right axis). The R_g, R_S, and D_{max} values do not increase linearly with IDR residue number as one would expect for a disordered region attached to a globular domain. Instead, adding disordered residues to the ARD leads to compaction when going from ARD to NTD ΔN54. Adding the complete IDR to the ARD leads to an expansion.

Next, the theoretical R_g values of the NTD mutants and the isolated ARD were calculated and compared to the wildtype NTD to evaluate to what extent the protein dimensions of the NTD deletion mutants deviated from what is expected for the ARD with the corresponding amount of IDR residues attached to it, (Figure 52).

The theoretical R_g value of the isolated ARD was calculated via CRYSOL (*411*) from the published X-ray crystal structure (PDB: 3w9g). The R_g values of the NTD deletion mutants were obtained via EOM analysis (*394*). In analogy to the wildtype NTD (see section VI3.1, Figure 48), EOM generated 10 000 conformers of the NTD deletion mutants by modelling the appropriate number of IDR residues as dummy atoms to the N-terminus of the ARD, for which the X-ray crystal structure (PDB: 3w9g) was used as a template. The average values of the random pool R_g distributions were used as a reference for comparison with the experimentally derived values.

Figure 52 A shows the experimental and calculated R_g values of the isolated ARD, the NTD deletion mutants and of the wildtype NTD plotted versus the IDR residue number contained in the NTD constructs. Strikingly, the experimental R_g value of the isolated ARD (Rg^{exp} = 2.49 nm) is increased by ~20% compared to the value calculated with CRYSOL (R_g^{calc} = 2.49). Considering that the ARD is a monomer in solution (see section VI3.1, Figure 40), the increased R_g value might indicate that the ARD samples more extended conformations in solution than the ARD X-ray crystal structure suggests. This is also in line with the assumption that the isolated ARD exhibits a flexible tail (see section VI3.1, Figure 40). Similar to the isolated ARD, the experimental R_g values of the NTD ΔN120 (Rg^{exp} = 2.86 nm) and the NTD ΔN104 mutants (Rg^{exp} = 2.98 nm) are increased by 30% and 12%, respectively, compared to their calculated R_g values (2.19 nm/2.41 nm). Notably, EOM failed to fit the theoretical scattering curves of conformers from the random pool to the experimental SAXS profiles of the NTD ΔN120 and NTD ΔN104 mutants. It is likely that the ARD X-ray structure is not an appropriate template to model the ARD in the context of the NTD ΔN120 and ΔN104 mutants. Considering that the ARD seems to be larger in solution than in the X-ray crystal structure, NTD deletion mutants might systematically display larger experimental R_g values compared to values calculated from models which use the ARD X-ray structure as a template for the ARD. Interestingly, the experimentally obtained R_g value of the NTD ΔN97 mutant is close to the calculated value (2.58 nm versus 2.52 nm, Figure 52 A) even though this mutant differs only by seven additional IDR residues from NTD ΔN104. In contrast to the ΔN120 and ΔN104 mutants, the NTD ΔN97 SAXS profile can be approximated with the scattering curves of EOM generated conformers (Figure 52 B). The R_g and D_{max} distributions of the EOM refined conformers are shifted to overall more expanded conformations compared to the random pool of conformations. This behaviour is reversed in the NTD ΔN54 mutant. The experimental R_g value of NTD ΔN54 (2.68 nm) is reduced by ~20% compared to the calculated R_g value (3.21 nm). The NTD ΔN54 SAXS profile can be approximated with scattering curves of EOM generated conformers which yield R_g and D_{max} distributions that are shifted to very compact conformations compared to the random pool of conformations (Figure 52 B). When the complete IDR is attached to the ARD, i.e. in the NTD, the experimental R_g value (4.06 nm) is close to the calculated R_g value (3.96 nm) (Figure 52 A). According to the EOM analysis, the R_g and D_{max} distributions of the EOM refined wildtype NTD conformers sample a broad range of conformations with a preference for compact states (Figure 52 B).

The comparison of experimental and calculated R_g values suggest that the TRPV4 ARD samples more extended conformations in solution than anticipated from X-ray crystallographic data. Adding IDR residues constituting the PRR and the PBD to the ARD increases the protein size as expected for adding a disordered chain to a globular domain. Remarkably, adding IDR residues beyond the PBD to the ARD N-terminus leads to a compaction of the NTD constructs. Such compaction might indicate an interaction of IDR residues with the ARD. It is worth to consider that an interaction with IDR residues might change the compactness of the ARD itself and thereby further influence the overall protein size. The trend towards more compact NTD structures with the increasing length of the IDR might therefore be due to an interaction of IDR residues with the ARD and consequently the stabilization of a more compact ARD conformation. The fact that NTD mutants with long IDRs yield good EOM fits – these are based on the ARD X-ray structure – might support this hypothesis. Interestingly, the NTD ΔN54 mutant exclusively samples the very compact conformations of the random pool ensemble generated by the EOM algorithm (Figure 52 B). In contrast to this, the wildtype NTD samples almost the entire conformational ensemble of the random pool ensemble, even though compact conformers seem to be preferred (Figure 52 B). This indicates that regions contained in the ~50 N-terminal IDR residues of the NTD decrease the fraction of extremely compact conformations at the cost of more expanded conformations. One may conclude from this that the 50 N-terminal residues influence how the rest of the IDR interacts with the ARD.

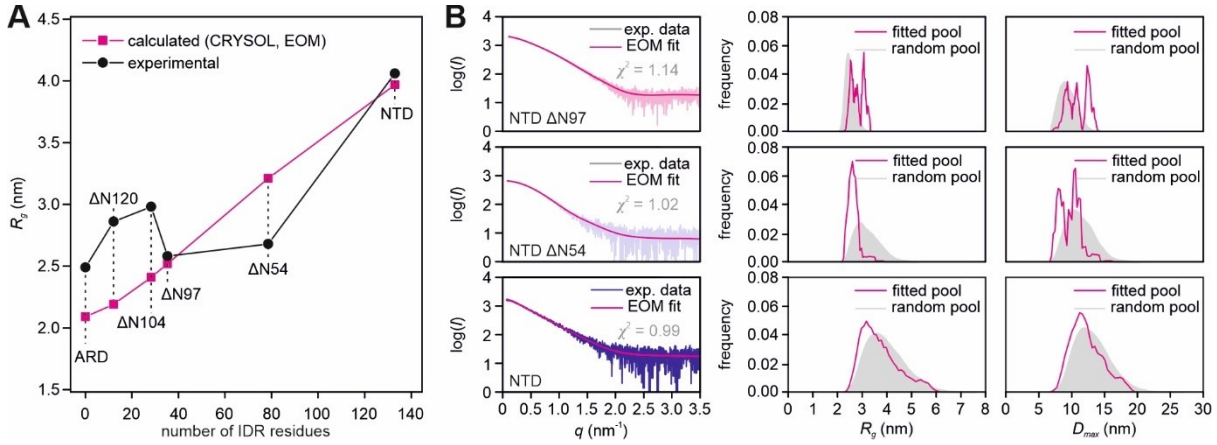


Figure 52: Comparison of experimental and calculated R_g values of NTD deletion mutants obtained from SAXS analysis. (**A**) Experimental (see Figure 51) and calculated radius of gyration (R_g) plotted versus the IDR residue number contained in the NTD deletion mutants. The R_g of the isolated ARD was calculated with CRYSOL (*411*). The R_g of the NTD and the deletion mutants represent the average R_g value of a random pool of conformations calculated via EOM analysis (*394*). It assumes a random chain behaviour of the IDR. (**B**) EOM analysis of NTD ΔN97, NTD ΔN54, and wildtype NTD. Shown is the EOM fit (magenta) to the experimental data in a log(*I*) versus *q* plot (left). The EOM derived R_g and D_{max} distributions of the generated random pool of conformations (grey, filled curved) and the conformations used to fit the experimental data (magenta) are shown on the right. The experimental SAXS profiles of the NTD ΔN104 and NTD ΔN120 mutants could not be fit to a random pool of conformations by the EOM algorithm. Therefore, no EOM plots are shown for both mutants.

The region in which the NTD deletion mutants start to transition from extended to compact conformations in SAXS experiments is close to the PBD. The PBD contains a tryptophan residue (W109) flanked by two basic residues on each side. As the only tryptophan residue in the entire NTD, it is ideally suited to report on the local environment of the PBD using tryptophan fluorescence spectroscopy (see section 2.13). Tryptophan fluorescence is sensitive to the polarity of the environment. Solvent exposed tryptophan residues display red-shifted fluorescence maxima, whereas buried tryptophan residues in hydrophobic environments display blue-shifted fluorescence maxima (*342*). To emulate a fully solvent-exposed tryptophan residue, fluorescence spectra of the isolated amino acid (maximum at 353.7 nm) in buffer were recorded and compared to different TRPV4 N-terminal constructs (Figure 53 A). First, a peptide constituting residues 97-134 of the IDR was measured. The peptide contains the TRPV4 PRR in the C-terminus as well as a Pleckstrin homology (PH) consensus motif in the N-terminus, which is formed by the PBD and a basic residue N-terminal to the PBD. It is accordingly termed PH-PRR. In solution, the PH-PRR peptide is unstructured as described in more detail in section VI3.3 (see Figure 60). The PH-PRR fluorescence wavelength of 353.7 nm is close to that of isolated tryptophan, thus indicating that W109 is in a solvent-exposed position, as expected for a disordered peptide Figure 53 B, C). When the complete IDR was measured, the fluorescence wavelength drops to 351 nm, thus indicating that residues in the IDR preceding the PBD position W109 in a more hydrophobic environment compared to the PH-PRR peptide (Figure 53 B, C). Next, the influence of the ARD on the W109 environment was tested by measuring the NTD deletion mutant NTD ΔN97. This construct corresponds to the PH-PRR peptide in tandem with the ARD. The NTD ΔN97 fluorescence wavelength of 350.1 nm compared to 353.7 nm in the PH-PRR peptide suggests that the ARD increases the hydrophobicity in the environment of W109 in the PBD. Adding 43 residues to the IDR, this yields NTD ΔN54, shifts the fluorescence wavelength to 349.8 nm (Figure 53 B, C). Addition of the remaining amino acids to yield the complete NTD further decreases the fluorescence wavelength to 349.5 nm. The effect of the ARD and the IDR on the fluorescence wavelength of W109 imply that both domains together are responsible to position W109 in a hydrophobic environment.

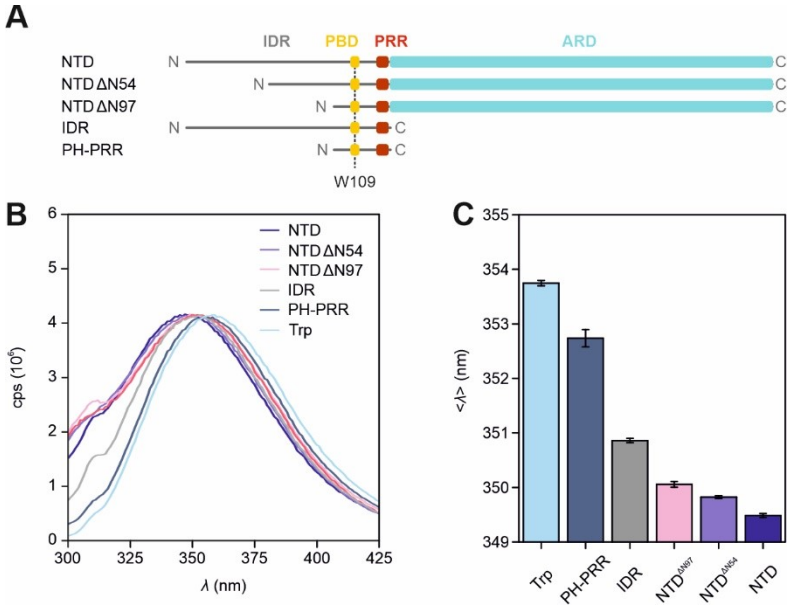


Figure 53: Tryptophan fluorescence spectroscopic analysis of *G. gallus* TRPV4 N-terminal constructs. (**A**) Tryptophan fluorescence spectra of NTD, NTD ΔN54, NTD ΔN97, NTD ΔN104, isolated tryptophan and (**B**) of NTD, IDR, PH-PRR, and isolated tryptophan. Plotted is the fluorescence as counts per second (cps) versus the wavelength λ. The topologies of the investigated proteins are shown in the diagram: cyan: ARD, red: PRR, yellow: PBD, grey: IDR. (**C**) Wavelength of the intensity-weighted tryptophan fluorescence wavelength $<\lambda>$ of investigated TRPV4 N-terminal constructs. The error bars represent the standard deviation from the mean value of three technical replications.

Due to its fortunate location, W109 can be exploited to probe the effects of the IDR on protein unfolding. For this, the tryptophan wavelength shifts of the NTD, the IDR, and the PH-PRR peptide were monitored in a urea titration experiment (Figure 54 A and B). The IDR fluorescence asymptotically shifts to higher wavelengths by ~0.7 nm with increasing urea concentrations, indicating a steady transition of W109 from a slightly buried to a solvent-exposed position. This is in line with previous studies in which urea was described to induce the elongation of IDPs (*445*). In contrast, the tryptophan fluorescence of the PH-PRR peptide upon urea treatment does not change over the entire range of urea concentrations. The NTD fluorescence shifts to higher wavelengths in a sigmoidal manner with a transition point at ~3 M urea. Such behaviour is typically observed for the urea-induced unfolding of globular proteins (*446*). However, W109 resides in the disordered IDR, which in isolation does not show the unfolding behaviour of a globular protein. The sigmoidal character of urea-induced fluorescence shifts of W109 in the NTD might therefore reflect on this residue's ability to "sense" the unfolding of the globular ARD 25-residues away.

The ARD does not contain a native tryptophan residue. To verify that the sigmoidal feature in the tryptophan fluorescence unfolding curve (Figure 54 B, dark blue) nonetheless originates from the unfolding of the ARD, urea titrations of the ARD, as well as the NTD were thus was monitored via CD spectroscopy (Figure 54 C and D). The IDR shows a linear decrease of the CD signal at 222 nm, which agrees well with the almost linear change in tryptophan fluorescence. The ARD and the NTD, in contrast, show the sigmoidal signal change expected for the concerted unfolding of a globular protein. The transition point of the NTD between 2-3 nm is in good agreement with the transition point observed in the tryptophan wavelength shift (~2 nm). This strongly indicates that the sigmoidal shape of the NTD tryptophan fluorescence represents the unfolding of the ARD that is sensed by a residue within the lipid-binding site of the adjacent IDR.

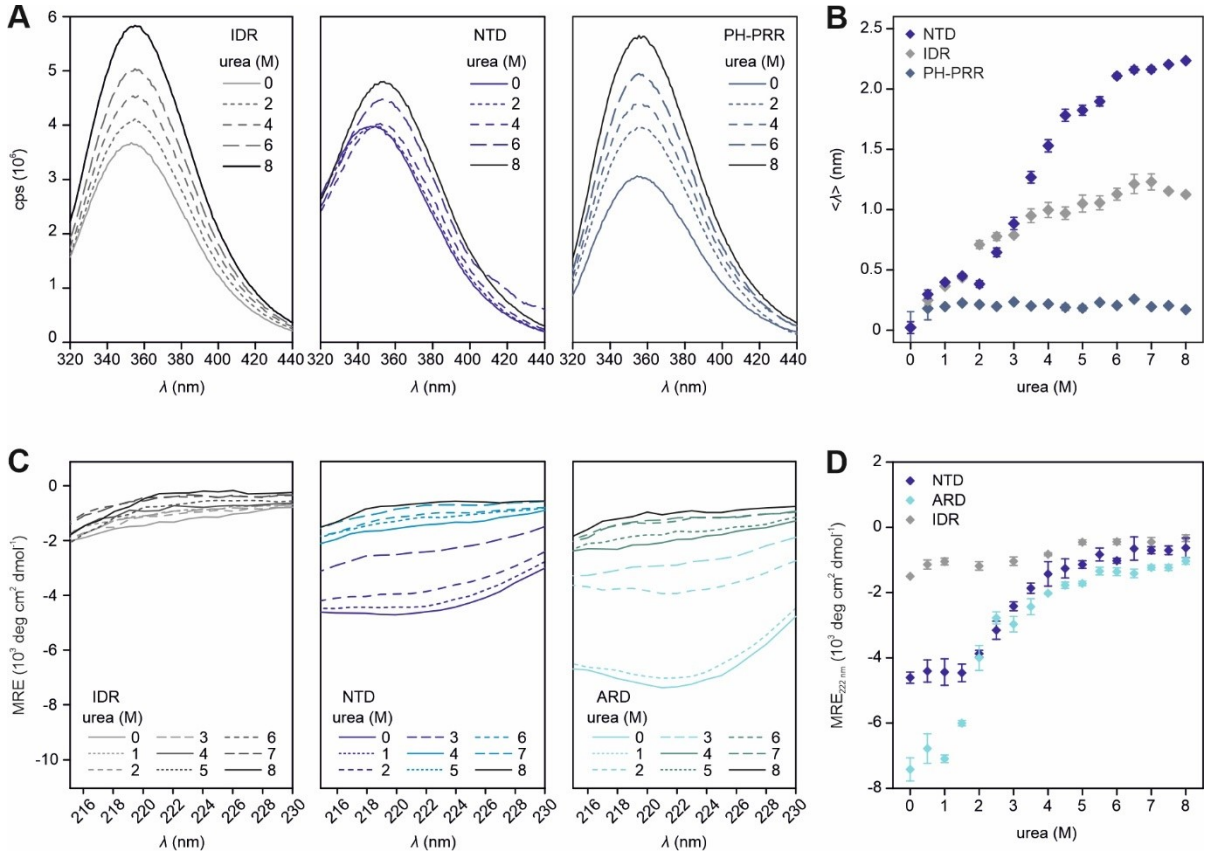


Figure 54: Chemical denaturation of the *G. gallus* TRPV4 N-terminus monitored via tryptophan fluorescence and far-UV CD spectroscopy. (**A**) Tryptophan fluorescence spectra of IDR (left), NTD (centre), and PH-PRR (right) at increasing urea concentrations. Plotted is the fluorescence as counts per second (cps) versus the wavelength λ. **(B)** Change of fluorescence wavelength $\Delta\langle\lambda\rangle$ plotted versus urea concentration. (**C**) Far-UV CD spectra of IDR (left), ARD (centre), and NTD (right) at increasing urea concentrations. Plotted is the mean residue ellipticity (MRE) versus the wavelength λ. (**D**) Change of the mean residue ellipticity at 222 nm ($MRE_{222\ nm}$) of IDR, ARD, and NTD plotted versus the urea concentration. The error bars in (B) and (D) represent the standard deviation from the mean value of three technical replications.

The dependence of the W109 tryptophan fluorescence wavelength on the presence of the ARD or regions preceding the PBD suggests the presence of intra-domain interactions within the IDR as well as interdomain interactions between IDR and ARD. To identify regions involved in such interactions, the spatial proximity between residues in the *G. gallus* TRPV4 NTD was investigated via crosslinking of lysine residues with disuccinimidyl suberate (DSS) (Figure 55). A total of 54 lysine residues are almost evenly distributed throughout the *G. gallus* TRPV4 NTD sequence (Figure 55 A and B). Only the region comprising residues 1-50 in the NTD is devoid of lysine. This makes lysine a well-suited probe to investigate the conformational space available to the IDR in relation to the ARD, i.e. the relative orientation of both domains in the context of the NTD. DSS has a length of 11.4 Å when fully extended (see section VI2.15). Considering the side-chain length of lysine, C_α carbons of DSS cross-linked lysine residues can be ~24 Å apart. Within a commonly accepted tolerance of ~3-6 Å (*351*), identified cross-links can be considered to reflect on lysine C_α-C_α distances up to 30 Å. The NTD was incubated with a mixture of hydrogenated and deuterated DSS (DSS-H12 or DSS-D12, respectively) as a cross-linking reagent. After quenching the cross-linking reaction, the cross-linked protein was proteolytically cleaved into peptide fragments, and DSS cross-linked peptides were subsequently identified via mass spectrometry (see section 2.15). It is important to note the detection of self-links for residues K56 and K116 in the XL-MS data sets. Self-links can either occur through the association of two NTD protomers or represent artefacts due to the non-covalent association of peptides during mass spectrometry measurements (*447*). Other than that, a total of 22 significant cross-links indicating lysine pair distances below 30 Å were identified, ten of which were between the IDR and ARD, three within the ARD, and nine within the IDR (Figure 55 A, B, and C).

The number of possible cross-links in a protein depends on the relative spatial position of lysine residues and the number of pairs with DSS compatible C_α-C_α distances. In the globular ARD, the lysine positions are confined by the defined three-dimensional ARD fold. A certain degree of variability in the lysine positions may be added by intrinsic dynamics of the ARD (also compare SAXS data for the isolated ARD indicative of a flexible protein, Figure 42). Three intra-domain crosslinks, K163-K205, K164-K205, and K205-K246, were identified in the ARD (Figure 55 A and C). They agree well with the X-ray crystal structure (PDB code 3w9g) in which the corresponding lysine C_α-C_α distances are between 10-15 Å (Figure 55 A). This also confirms the structural integrity of the ARD during the cross-linking reaction. The absence of links between other lysine pairs with C_α-C_α distances below 24 to 30 Å in the X-ray crystal structure (21 in total) might be due to low accessibility of lysine side chains or unfavourable steric conditions, such as side chains pointing in opposite directions or steric hindrances.

In contrast to the folded ARD, the IDR adopts a random chain ensemble of conformations in which the positions of lysine residues are not fixed. Even though SAXS-based EOM predicts that the IDR slightly deviates from a purely random chain behaviour (see Figure 48), a broad spectrum of distances between all possible lysine pairs might be sampled in the conformational ensemble. Cross-linking compatible distances are therefore not only expected between sequentially close but might also between sequentially distant lysine pairs. To obtain a rough estimate whether lysine pairs in the IDR conformational ensemble are within reach of the cross-linker, the C_α-C_α distance distributions between lysine residues in the EOM refined models of the *G. gallus* TRPV4 NTD (Figure 48) were calculated. The average distance C_α-C_α distance between all residues in the ensemble are displayed in a heat map (Figure 55 C). All 10 lysine residues in the IDR display cross-linking compatible average C_α-C_α distances (<30 Å) to lysine neighbours within a range of ~20 amino acids throughout the conformational ensemble (Figure 55 C). In agreement with this, cross-links were identified between several lysine pairs with <20 residues separation, including K50-K56, K56-K67, K86-K107, K107-K116, K107-K122, or K116-K133 (Figure 55 A and C). Interestingly, no cross-links were observed between K62 and its neighbours K50, K56, and K67 or between K86 or K87 and their neighbours K62, and K67, despite their proximity in the IDR sequence. Interactions were also identified between lysine pairs separated by more than 50 residues, including cross-links between K50 with K107, K116, and K122, as well as K56 with K107, K116, K122, and K133 (Figure 55 A). In the EOM refined model, all lysine pairs separated by >35 residues exhibit average C_α-C_α distances beyond the reach of the cross-linker (Figure 55 C). Numerous lysine residues within the IDR would not be expected to crosslink based on their pair-wise distances within the primary structure of the protein. However, for every lysine pair deemed too distant in sequence space, the EOM refined NTD conformer yields at least one conformation with a C_α-C_α distance <30 Å (Figure 55 D). The cross-links between sequentially distant lysine residues that are observed in the experiment thus presumably reflect on compact IDR conformations. A more compact IDR conformation was also indicated by the SAXS and tryptophan fluorescence data (Figure 48 and Figure 53).

Besides the intra-domain cross-links within the ARD or IDR, the XL-MS analysis also identified various inter-domain cross-links, that is links between lysine residues in the IDR and lysine residues in the ARD, respectively (Figure 55 B). The two lysine residues K107 and K116 flanking the PBD together show cross-links to lysine residues located in the first three Ankyrin repeats of the ARD, that is $K107^{PBD}$ to$K183^{ARD}$ and $K237^{ARD}$, as well as $K116^{PBD}$ to$K163^{ARD}$, $K205^{ARD}$ and $K237^{ARD}$. The corresponding average C_α-C_α distances between these lysine residues are below 30 Å in the EOM refined NTD ensembles (Figure 55 C). The detected cross-links $K50^{IDR}$-$K237^{ARD}$ as well as $K56^{IDR}$ to ARD residues K178, K183, K237, K338, K341, and K368 reflect on long-range interactions bridging >80 residues (Figure 55 B). These cross-links can only be explained by very compact conformations in which the centre of the IDR comes close to the C-terminal Ankyrin repeats. The corresponding average C_α-C_α distances in the EOM refined NTD ensemble surpass the cross-linkable distance (Figure 55 C). Whereas a few conformations with cross-linking compatible C_α-C_α distances between $K50^{IDR}$-$K237^{ARD}$ and $K56^{IDR}$ to $K178^{ARD}$, $K183^{ARD}$ and$K237^{ARD}$ exist in the EOM refined NTD ensemble, the distances between $K56^{IDR}$ to-K338, K341, and K368 in the ADR are beyond 40 Å in all conformers (Figure 55 D). The identified long-range contacts might therefore indicate that the NTD may at least sample more compact conformations than predicted by the EOM analysis.

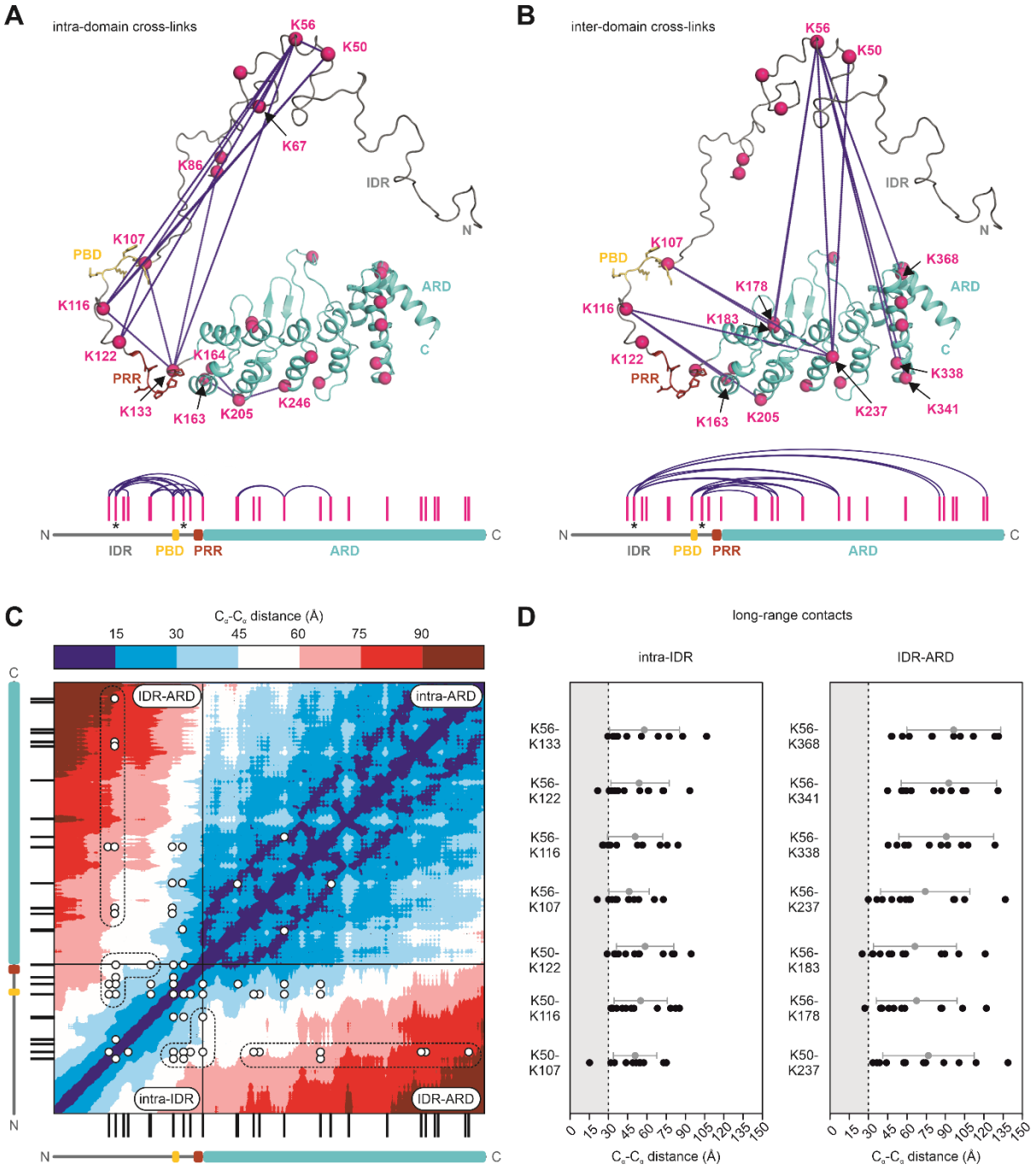


Figure 55: Intra- and interdomain contacts within the *G. gallus* TRPV4 N-terminus identified via cross-linking mass spectrometry. (**A**) Intra-domain contacts and (**B**) inter-domain contacts in *G. gallus* TRPV4 NTD between native lysine residues (highlighted in magenta) crosslinked by disuccinimidyl suberate (DSS). Crosslinks are visualized by blue lines. Putative self-crosslinks, i.e. between the same lysine residue in two different protomers, are indicated with an asterisk in the NTD topology representation. The structural model of the TRPV4 NTD was generated by combining a modified EOM refined conformer for the intrinsically disordered region (IDR) obtained from SAXS measurements (see section 2.18) with the crystal structure of the isolated *G. gallus* ARD (PDB: 3w9g). (**C**) Distance (C_α-C_α) heatmap for all NTD residues generated based on an EOM refined conformational ensemble of 15 NTD conformers determined by SAXS. Cross-links between lysine pairs identified via XL-MS analysis are indicated by white dots. Intra-IDR and IDR-ARD C_α-C_α distances surpassing the cross-linker compatible range are boxed (dashed lines). (**D**) C_α-C_α distances for select lysine pairs in a conformational ensemble determined by SAXS based on 15 EOM refined NTD conformers. Individual (black dots) and average (grey dots incl. standard deviation error bars) distances are shown. Lysine pairs cross-linked in XL-MS experiments but whose average distance within the ensemble exceeds the expected crosslinking range of 30 Å for DSS (the grey area in the plot) are displayed. This indicates large conformational flexibility together with preferential intra- and interdomain interactions within the NTD, since not all lysine residues can be cross-linked (see A, B).

The experimentally determined contacts between the ARD and the IDR are clear indicators of direct ARD-IDR interactions. The distribution of cross-links paired with the fact that not all lysine pairs within the 30 Å range of the DSS-linker are cross-linked suggests that the IDR-ARD contact sites are not solely a product of random conformational sampling. They might rather represent preferential interaction sites, thus implying a hierarchical conformational ensemble of the IDR.

To further explore the details of structural and dynamic crosstalk between the IDR and the ARD, the *G. gallus* TRPV4 NTD, IDR, and ARDs were investigated via hydrogen-deuterium exchange mass spectrometry (HDX-MS) (Figure 56). HDX-MS provides sequence-specific resolution and does not rely on the presence of specific reporters in the proteins. The deuteration of protein backbone amide groups reflects on collective protein motions on the ms-µs timescale (*354*). In brief, the proteins were incubated in a deuterated buffer, and the deuteration of the proteins at different time points was quenched in a low pH buffer. Subsequently, the proteins were proteolytically cleaved, and the deuteration of the obtained peptides was determined via mass spectrometry (see section 2.16). The proteolytic digestion yielded a complete peptide coverage of the ARD sequence. Unfortunately, the PBD and the PRR containing sequence in the IDR was not covered by the proteolytically derived peptides in either the isolated IDR or the NTD, presumably due to insufficient proteolysis (Figure 56).

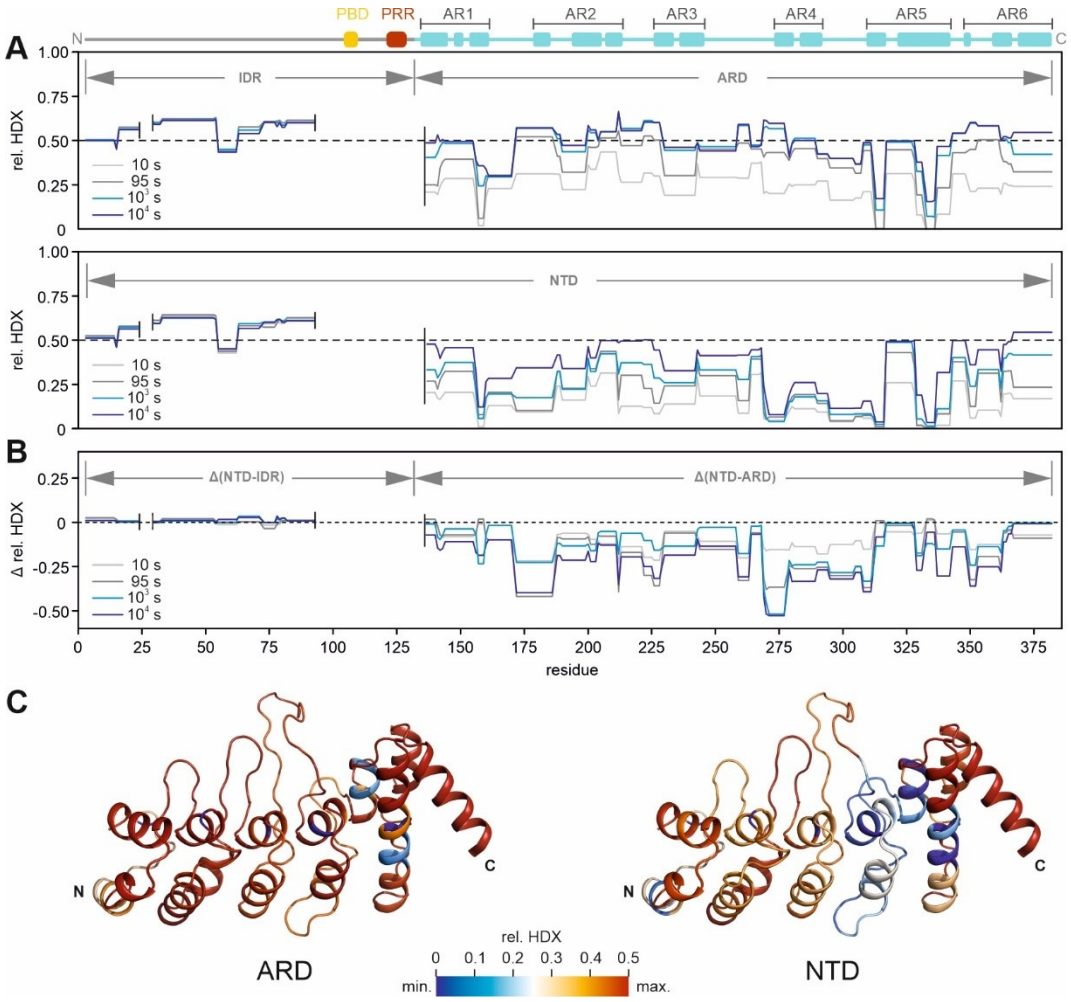


Figure 56: Hydrogen-deuterium exchange analysis of the TRPV4 N-terminus. A topology model of the NTD showing the position of the PBD, the PRR, and the Ankyrin repeat helices in the ARD is shown on top of the diagrams. (**A**) Relative hydrogen-deuterium exchange (rel. HDX) of the isolated IDR and the isolated ARD (upper diagram) as well as the complete NTD (lower diagram) after 10/95/1000/10 000 seconds plotted versus the residue number. (**B**) Difference in relative hydrogen-deuterium exchange (Δ rel. HDX) between the NTD and the isolated IDR, termed Δ(NTD-IDR), and between the NTD and the isolated ARD, termed Δ(NTD-IDR), plotted versus the residue number. Residues 25-28 and 94-135 in the TRPV4 sequence were not covered in the mass spectrometric analysis. A rel. HDX value of 0.5 (indicated by a dashed line) corresponds to complete deuteration. (**C**) Relative hydrogen-deuterium exchange of the isolated ARD (left) and the NTD (right) mapped onto the X-ray crystal structure of the ARD (PDB code 3w9g).

The mass spectrometric analysis revealed the complete deuteration of the IDR within 10 seconds, both in isolation and in the context of the NTD (Figure 56 A). This agrees well with the disordered character and the absence of secondary structures in the IDR, which exposes backbone amide hydrogen atoms to the solvent and makes them highly susceptible to deuterium exchange (*354*). As expected, in the ARD deuteration is slowed substantially in the ARD compared to the IDR (Figure 56 A). Backbone amide hydrogen atoms involved in secondary structure stabilizing hydrogen bonds are protected from exchange against deuterium (*354*). However, deuteration of the isolated ARD is still comparably fast for an α-helical protein and indicates extensive collective motions of the Ankyrin repeats (AR). Complete deuteration is observed throughout almost the entire sequence after 100 seconds. This is somehow surprising, considering the seemingly tight packing of α-helices in the ARD X-ray crystal structure (PDB code 3w9g). Regions of reduced deuteration rates localize in AR 5 and between AR1 and AR2. Strikingly, in the context of the NTD, the deuteration rate in the ARD is slowed by approximately two orders of magnitude (Figure 56 A and C). Complete deuteration within 10^4 seconds is observed for AR1, in the region between AR2 and AR3, the loop of AR5, and AR6 in the C-terminus. AR 4 represents a particularly deuteration-resistant region. Thus, it is not only the deuteration rate *per se* but also the time-dependent isotopic fingerprint of the deuteration exchange that seems to be strongly influenced by the presence of the IDR.

In general, backbone deuteration in proteins can follow two distinct kinetics which depends on the rate at which backbone segments transition between protected (exchange-incompetent) and exposed (exchange-competent) states compared to the HDX rate (*354*). For stably folded proteins, local protein motions are usually faster than HDX. Rapid interconversion between protected and exposed states allows only a fraction of exposed hydrogens to exchange for deuterium. Several opening and closing cycles are therefore required to fully deuterate a segment in a single protein copy. These so-called EX2 deuteration kinetics manifest as unimodal isotopic patterns in the mass spectra that gradually shift to higher masses over time (see VI2.16, p. 57, Figure 26). The rate and extent of deuteration under EX2 kinetics reflects on local structural fluctuations. EX2 kinetics are observed in the ARD in the context of the NTD (Figure 57). In isolation, the ARD displays faster EX2 kinetics throughout the protein sequence. Additionally and quite strikingly, several segments in the isolated ARD, including AR2, AR4, and AR6, are deuterated under so-called EX1 kinetics, or under EX1-EX2 intermediate kinetics (Figure 57, see also Figure 26). This behaviour indicates slow transitions between protected and exposed states compared to the HDX rate or long-lived and highly exchange-competent exposed states. EX1 manifests as bimodal isotopic patterns in the mass spectra. Slow transitions or long-lived exposed states enable the complete deuteration of a segment in a single protein copy within a single opening and closing cycle. Partly deuterated intermediates do not occur under pure EX1 kinetics. Instead of shifting, the isotopic envelope in the mass spectrum of the non-deuterated species diminishes gradually in favour of the isotopic envelope corresponding to the fully deuterated species. Importantly, EX1 kinetics are a strong indicator of local folding and unfolding equilibria in proteins (*354*). The unfolding of a protein region exposes all contained amide hydrogens to the solvent simultaneously, thereby enabling a concerted exchange against deuterium before the region refolds into a protected state. This mechanism does not require a very long-lived unfolded state, as HDX proceeds extremely fast in unfolded protein regions.

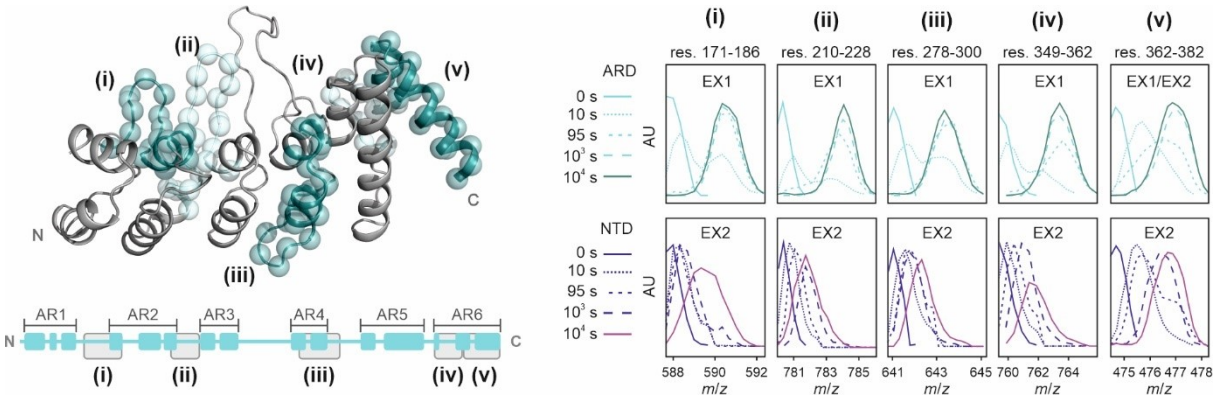


Figure 57: Local hydrogen-deuterium exchange kinetics observed in the ARD in isolation and in the context of the NTD. Local differences in the deuteration kinetics between the isolated ARD and the ARD in the context of the NTD mapped onto the ARD X-ray crystal structure (PDB code 3w9g). Regions displaying EX1 kinetics in the isolated ARD but EX2 kinetics in the context of the NTD comprise residues 171-186, 210-228, 278-300, 349-362, 362-382 and are indicated by spheres. The mass spectra of the corresponding peptides after 0/10/95/1000/10 000 seconds of deuteration illustrate the EX1 and EX2 or intermediate EX1/EX2 characteristic isotopic envelopes. The mass spectra show the intensity (arbitrary units, AU) plotted versus the mass to charge ratio (m/z). The ARD topology showing the position of the six Ankyrin repeats (AR1-6) and position of the peptides with different deuteration kinetics is shown on the bottom.

The increased flexibility of the isolated ARD also becomes evident from the SAXS data (Figure 58). The ARD real-space distance distribution and dimensionless Kratky plot reflect on a globular protein with a flexible extension (Figure 42 D and E). This, however, contrasts with the cryo-EM and X-ray crystallographic structures of the TRPV4 ARD in which the protein was always modelled as a-helical and compact protein. As described above (Figure 52), the theoretical SAXS profile of the ARD crystal structure obtained with the CRYSOL program (*411*) deviates substantially from the experimental data (χ^2 = 3.22) and suggests larger ARD dimensions in solution (R_g = 24.9 Å) than in the crystal structure (R_g = 20.9 Å) (Figure 58 A). This deviation suggests that the solution structure of the isolated TRPV4 ARD differs substantially from the structures reported in X-ray (PDB code 3w9g) or cryo-EM studies (PDB code 6bbj). In contrast to cryo-EM and X-ray crystallography, SAXS captures all coexisting conformations of a protein simultaneously and therefor reports on the population-averaged structure of conformational ensembles. Deviations of SAXS profiles from calculated scattering profiles of crystal structures are often an indicator of extensive flexibility or conformational changes which drastically enlarge a protein or change its shape. The ARD structure observed in the crystal might therefore only represent a substate within a larger ensemble of conformations. The extensive collective motions and potential local unfolding events in the isolated ARD observed via HDX-MS strongly suggest that the Ankyrin repeats are very flexible. The SAXS profile of the isolated ARD was analyzed with SREFLEX using the *G. gallus* TRPv4 ARD X-ray crystal structure (PDB code 3w9g) as a template structure (*418*) (Figure 58 B and C). The SRFEFLEX algorithm partitions a protein structure into pseudo-domains which are then iteratively reassembled to improve the agreement with the experimental SAXS data (see section 2.18 for a detailed description). Applying the SREFLEX algorithm to the isolated ARD improved the agreement with the experimental SAXS data to χ^2 = 1.21 (Figure 58 B). The overall shape of the ARD is preserved in the final SREFLEX-refined model (Figure 58 C). However, all Ankyrin repeats were tilted with respect to the starting structure. This suggests that the deviations between experimental and theoretical SAXS curves are due to substantial flexibility in the TRPV4 ARD.

Next, a low-resolution bead model of the ARD was generated based on the experimental data using the DAMMIN algorithm (*413*)(see section 2.18). The bead model provides an estimate of the overall shape of the ARD conformational ensemble in solution. As expected from the real-space distribution and Kratky-plot (Figure 42 D and E), a short extension protrudes from a globular domain in the final DAMMIN model (Figure 58 D and E). The alignment of the ARD X-ray structure (PDB code 3w9g) with the bead models suggests, that the extension corresponds to the N-terminal region of the ARD, potentially indicating an unwinding of the N-

terminal Ankyrin repeat AR1. To further analyze the structural origin of the extended tail, the ARD was computationally fragmented into four subunits which were then assembled *de novo* to fit the experimental SAXS data using the program SASREF (*413*) (see section 2.18). Based on the superposition of the ARD N-terminus with the extension of the DAMMIN model, the fragmentation operation mainly targeted the first Ankyrin repeat, which was split in two fragments within its loop. An additional fragmentation was carried out between AR3 and AR4 (Figure 58 G). The subunits (sub) for the SASREF-computed *de novo* assembly were as follows: res. 135-147 (sub1), res. 148-163 (sub1), res. 164-250 (sub3), res. 251-382 (sub4) (Figure 58 G). These subunits were reassembled by SASREF to generate six subunit assemblies, which fit the experimental SAXS data with an χ^2 of 1.15-1.25 (Figure 58 F and G). In all assemblies, the first Ankyrin repeat is detached and protrudes away in various directions from the rest of the ARD. This is in good agreement with the DAMMIN model, which can be regarded as an average envelope of the different conformations.

The SAXS-based models (Figure 58) in combination with the HDX-MS analysis (Figure 56) classify the ARD as an extremely flexible protein domain in isolation, something that was unexpected based on available X-ray and cryo-EM models (*21, 42, 150*). The detachment and elongation of AR1 from the rest of the protein in the DAMMIN and SASREF models might explain the EX1 deuteration kinetics observed in the region between AR1 and AR2. The detached AR1 might be a sufficiently long-lived state to allow for complete deuteration before reattaching to the rest of the protein. Importantly, it cannot be ruled out that the flexible extension observed in the DAMMIN model partly or even solely corresponds to the ARD C-terminus, where EX1 deuteration kinetics were also detected in AR6. Unfortunately, the SAXS measurements of the NTD cannot be used to investigate ARD structure and dynamics in the contexts of the NTD. The conformational ensemble of the IDR substantially complicates the data analysis. However, the HDX-MS data strongly suggest that the presence of the IDR reduces the dynamics in the ARD. It is therefore likely, that the ARD in the context of the NTD samples conformations similar as the one observed in the X-ray crystal structure. This is supported by the good agreement between the NTD SAXS profile and the EOM refined NTD models for which the ARD X-ray structure was used as a template (Figure 48).

In summary, the HDX-MS and SAXS analyses highlight the ARD as a highly dynamic protein domain, in which the extent of collective motions depends on the presence (or absence) of the IDR. The absence of the IDR makes the ARD more accessible to deuteration, therefore indicating enhanced collective motions throughout the ARD. Furthermore, the isolated ARD displays EX1 deuteration kinetics at Ankyrin repeats 2 and 4, which might indicate local unfolding and refolding events. To what extent AR2 and AR4 unfold cannot be concluded from the HDX-MS data. CD spectroscopic analysis suggests that the secondary structure content in the ARD in solution is almost identical to the secondary structure content in the ARD X-ray crystal structure (PDB: 3w9g). It is therefore unlikely that the EX1 kinetics observed in HDX-MS originate from the unwinding of α-helices. Instead, the EX1 deuteration kinetics might reflect on the movement of intact AR helices with respect to each other which renders segments of the ARD more solvent accessible. This is partly supported by SAXS models of the isolated ARD, in which a flexible extension is modelled by an unwound but still α-helical AR. Such an unwinding seems to be suppressed in the presence of the IDR, i.e. in the context of the NTD. SAXS, XL-MS, and tryptophan fluorescence suggest that there are interactions between IDR and ARD as well as between residues within the IDR. Such interactions may enable compact NTD conformations and might be required for stabilizing the ARD structurally and dynamically. The IDR emerges as a potential regulator of ARD dynamics. Conversely, the ARD might be a dynamic sensor of conformational changes in the IDR. Interestingly, the PBD, a regulatory site for lipid binding, was identified as a hot spot for intra-IDR and, importantly, also for interactions with the ARD. This fuels speculations on whether channel regulation upon ligand binding in the IDR is based on interference with specific or non-specific IDR-ARD interactions. Accordingly, an IDR ligand might dynamically destabilize and accelerate domain motions in the ARD, which could be passed on to the TMD.

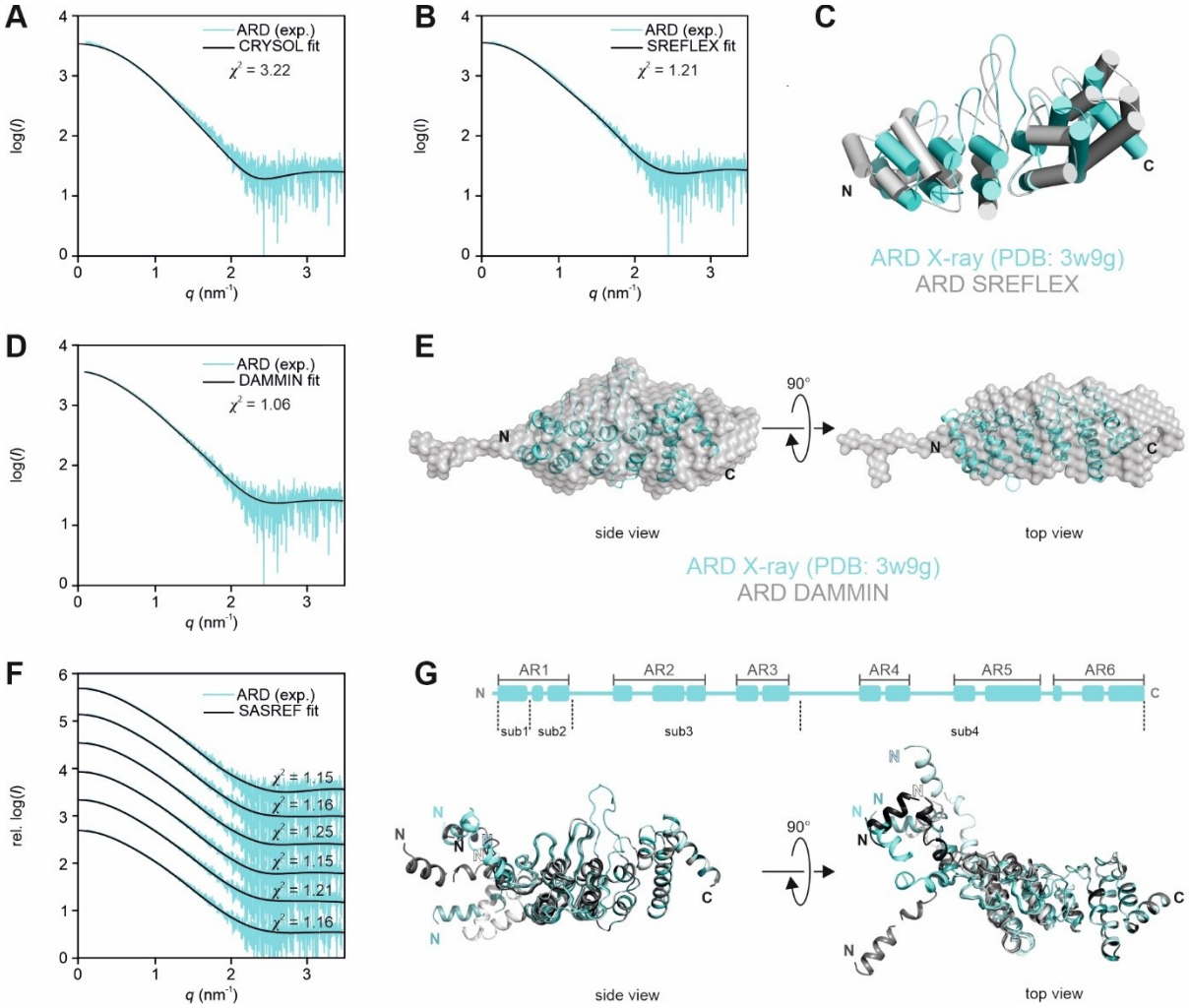


Figure 58: Structural analysis of the conformational flexibility in the ARD based on SAXS experimental data. (**A**) Comparison of the experimental SAXS profile with the theoretical SAXS profile of the ARD X-ray crystal structure (PDB code 3w9g) calculated with CRYSOL (*411*). The theoretical protein dimensions (R_g = 24.9 Å) are substantially larger than in the crystal (R_g = 20.9 Å). Plotted is the logarithmic intensity, log(*I*), versus the scattering vector, *q*. (**B**) Flexibility analysis of the isolated TRPV4 ARD using SREFLEX (*418*). Calculated SAXS profile of the SREFLEX model compared to the experimental data in a plot of the logarithmic intensity, log(*I*), versus the scattering vector, *q*. (**C**) Superposition of ARD X-ray crystal structure (PDB code 3w9g) with the SREFLEX calculated ARD conformer indicates extreme flexibility in the ARD, which reflects in twisted and turned Ankyrin repeats. (D, E) DAMMIN (*413*) bead model analysis of the overall structure of the ARD conformational ensemble. (**D**) Calculated SAXS profile of the DAMMIN model compared to the experimental data in a plot of the logarithmic intensity, log(*I*), versus the scattering vector, *q*. (**E**) Superposition of the ARD X-ray crystal structure (PDB code 3w9g) with the ARD DAMMIN model using the DAMSUP algorithm. The alignment positions the ARD N-terminus towards the extension in the DAMMIN model. (**F**, **G**) SASREF-based assembly of ARD subunits which were generated from fragmenting the ARD X-ray crystal structure (PDB code 3w9g). (F) Calculated SAXS profiles of six SASREF (*413*) models compared to the experimental data in a plot of the logarithmic intensity, log(*I*), versus the scattering vector, *q*. (G) Top: Six Ankyrin repeat topology (AR1-6) of the ARD highlighting the fragmentation positions yielding the four subunits sub1 to sub4 for the *de novo* assembly of the ARD from. Bottom: Superposition of six ARD assemblies calculated by SASREF. The structures were aligned along sub4.

3.3 The TRPV4 PBD non-specifically binds anionic lipids

Previous mutagenesis and functional studies have identified the phosphoinositide binding domain (PBD) in the IDR as the binding site for the lipid phosphatidylinositol-4,5-bisphosphate (PIP_2) (*128*). The PBD is located ~25 residues N-terminal of the ARD and consists of a conserved ++W++ motif (here, '+' is a basic residue) (Figure 59 A and B). PIP_2 binding to the PBD is essential for thermo- and osmosensitivity of TRPV4. Mutation of the basic residues to alanine (++W++ → AAWAA) renders the channel insensitive to thermal and osmotic stimulation, potentially because of disturbed electrostatic interaction between the PBD and the negatively charged PIP_2 head group (*75*).

Since the PBD resides within a disordered region, this raises the question whether its interaction with PIP_2 is specific, and if so, how specificity is achieved. Specific lipid interactions are expected to require some form of a defined three-dimensional orientation of side chains which create a surface complementary to the structure and charges of the lipid head group. Interestingly, the PBD together with neighbouring residues forms a basic residue-rich sequence and harbours a Pleckstrin homology (PH) domain-like motif, (K/R) x_n (K/R) x $(K/R)_2$ (here, x_n are n arbitrary residues) (*448*). PH domains are folded regions which bind specifically to phospho-inositides with adjacent phosphates (*449*). The presence of globular structures in the PBD can be excluded from the NMR and SAXS analysis of the TRPV4 IDR (see VI3.1). Alternatively, the PBD might achieve PIP_2 specificity via lipid-induced folding into an amphiphilic helix, which orients the charged side chains to shape a PIP_2-complementary binding surface. However, a bioinformatic analysis of the *H. sapiens* TRPV4 sequence using the HeliQuest tool (*450*) does not indicate an amino acid pattern in the PBD which is compatible with an amphiphilic helix (Figure 58 B). Instead of being specific for PIP_2, the PBD might simply prefer PIP_2 over other lipids due to the strong negative charge of the PIP_2 head group. This would imply that negatively charged lipids *per se* should be able to bind to the PBD.

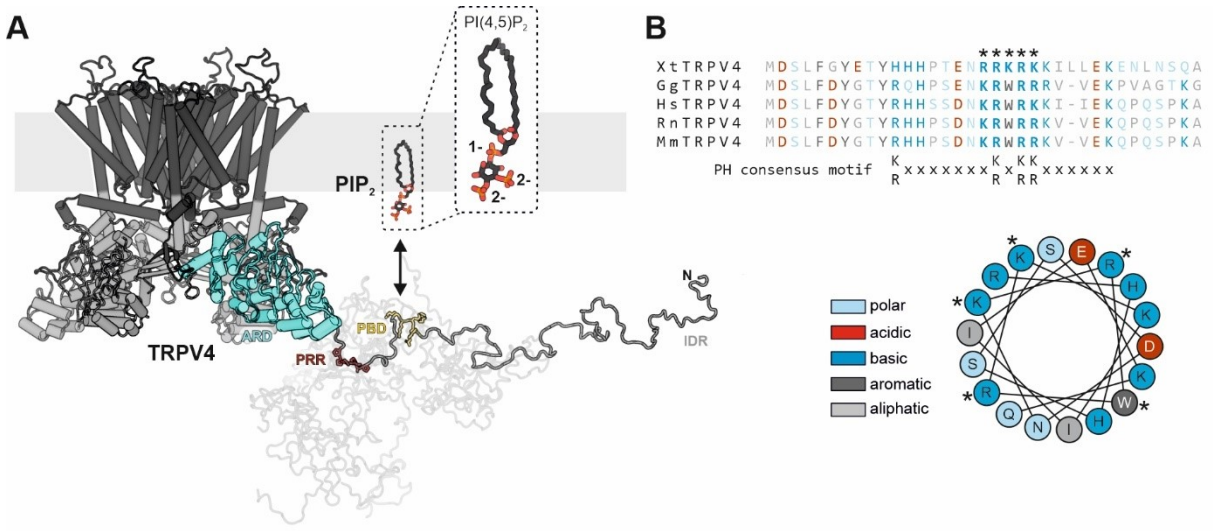


Figure 59: Sequence analysis of the phosphoinositide binding domain and potential lipid-binding sites in the TRPV4 N-terminus. (**A**) Structural model of a TRPV4 monomer containing the full-length IDR. The model was built from a swiss-model (*427*) generated *H. sapiens* TRPV4 homology model based on the *X. tropicalis* TRPV4 cryo-EM structure (PDB: 6bbj) combined with an EOM refined model of the *H. sapiens* TRPV4 NTD. The transmembrane domain (TMD) is indicated in dark grey; the ARD is shown in cyan; the IDR is shown in light grey. The PRR and the PBD are indicated in red and yellow, respectively. An enlargement of the PIP_2 structure is shown in the dashed box. (**B**) Center: Amino acid sequence comparison of the TRPV4 PBD between *X. tropicalis* (Xt), *G. gallus* (Gg), *H. sapiens* (Hs), *R. norvegicus* (Rn), *M. musculus* (Mm) reveals a conserved PH-domain consensus motif, (K/R) x_n (K/R) x $(K/R)_2$ (here, x_n are n arbitrary residues). Bottom: Amphiphilic helix prediction in the *H. sapiens* TRPV4 PBD via HeliQuest (*450*). PBD residues are indicated with an asterisk. The amino acid distribution is <u>not</u> compatible with an amphiphilic helix.

To explore the potential role of the PBD as a general lipid binding domain, the *G. gallus* TRPV4 constructs NTD, IDR, PH-PRR (Figure 60 A, see also section VI3.1 and VI3.2) were analyzed for their ability to bind to anionic lipids. To validate, whether the lipid interaction of the TRPV4 N-terminus is entirely dependent on the positively charged residues within the PBD, mutants of the NTD, IDR, and the PH-PRR peptide were generated. Here, the basic residues in the PBD were replaced by alanine, ^{107}KRWRR^{111} → ^{107}AAWAA^{111} (Figure 60 A). This mutation corresponds to ^{120}KRWRR^{125} → ^{120}AAWAA^{125} in the *H. sapiens* TRPV4 sequence, which was previously described to render TRPV4 insensitive to heat and osmotic stimuli (*128*). The mutants, termed NTD^{AAWAA}, IDR^{AAWAA}, and PH^{AAWAA}-PRR could be expressed and purified to homogeneity in analogy to the wildtype proteins (Figure 60 B). CD spectroscopic analysis confirms that the PBD mutation does not alter the secondary structure in the NTD, IDR, or the PH-PRR peptide (Figure 60 C). Interestingly, SEC analysis reveals that the Stokes radius of the PBD mutants is slightly increased compared to the wt-constructs even though the molecular weight is slightly decreased upon mutation. This effect is most pronounced in the context of the NTD, where the ^{107}AAWAA^{111} mutation increases the R_s by ~3 Å.

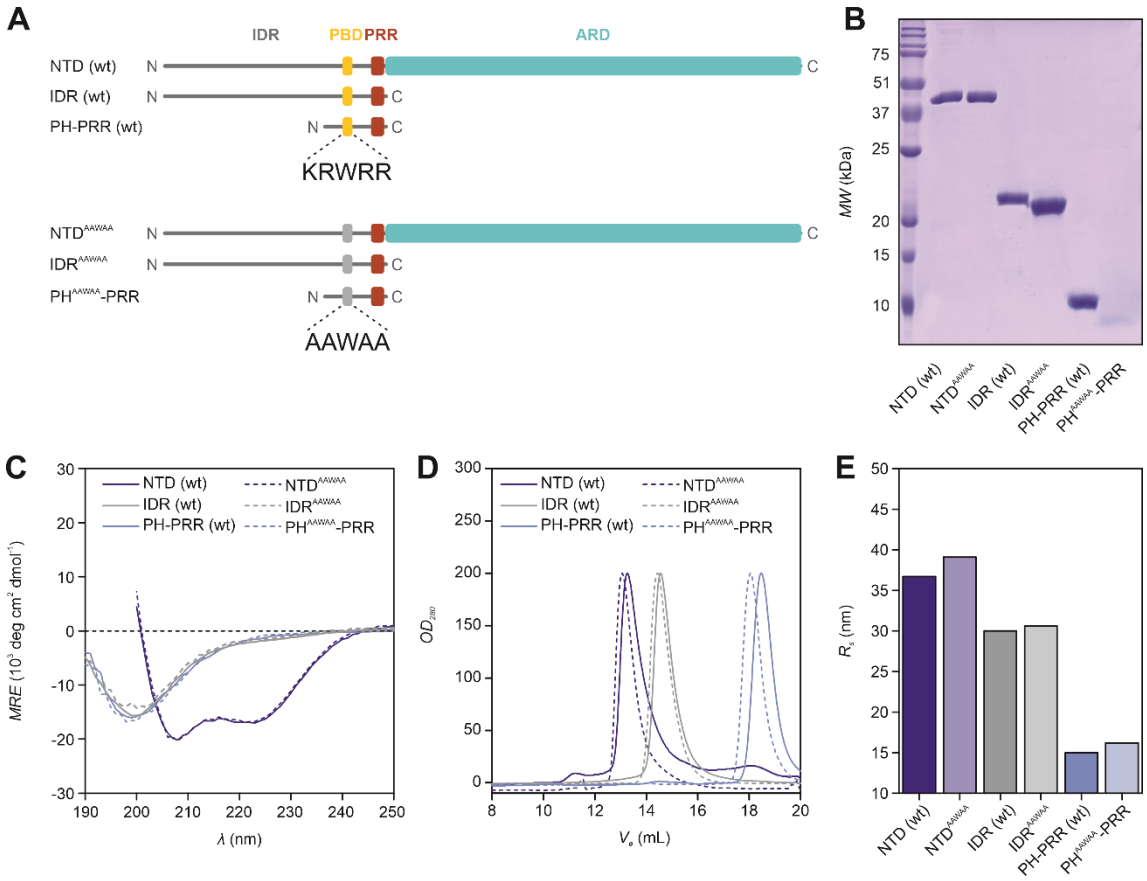


Figure 60: Comparison of wildtype and PBD mutant TRPV4 N-terminal constructs via SEC and far-UV CD spectroscopy. (**A**) Topology of wildtype and PBD-mutants of the *G. gallus* TRPV4 NTD, IDR, and PH-PRR. The wildtype PBD (yellow) comprises residues ^{107}KRWRR^{111}. In the mutant (grey), the basic residues were replaced by alanine, yielding ^{107}AAWAA^{111}. Cyan: ARD, red: PRR, yellow: wt-PBD, grey box: mutant PBD, grey line: IDR. (**B**) SDS-PAGE analysis of constructs shown in (A). (**C**) far-UV CD spectra of purified constructs. Plotted is the mean residue ellipticity *MRE* versus the wavelength λ. (**D**) SEC profiles of purified constructs. Plotted is the OD_{280} versus the elution volume V_e. The SEC profiles were normalized to an OD_{280} = 200 of the peak maxima. (**E**) Stokes radius (R_s) of wt- and ^{107}AAWAA^{111} mutants estimated from the elution volumes observed in (D).

As outlined in section VI3.2, the position of W109 provides a site-specific probe for the environment of the TRPV4 PBD. As such, it is ideally suited for monitoring lipid binding to the PBD via tryptophan fluorescence spectroscopy. Binding of the PBD to lipids transitions W109 from a solvent-exposed position in a polar envi-

ronment to the hydrophobic environment of the membrane. This conformational change manifests as a fluorescence shift to lower wavelengths. To investigate the general ability of the *G. gallus* TRPV4 NTD to interact with lipids, tryptophan fluorescence spectra of the NTD were measured in the presence of increasing liposome concentrations, and the fluorescence wavelength shift was determined as described in section 2.13. Liposomes containing mixtures of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol (POPG) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) were used as mimics of lipid membranes with charged or non-charged surfaces (see section VI2.10, p. 47 for liposome preparation). POPG is an anionic lipid with a net charge of -1, whereas POPC is a zwitterionic lipid with a net charge of 0 (Figure 61 A).

The NTD tryptophan fluorescence spectrum was nearly unaffected upon addition of liposomes containing only POPC (Figure 61 B and C). In contrast, titration of liposomes containing a mixture of 50% POPG and 50% POPC in a 1:1 ratio shifted the tryptophan fluorescence of the NTD to smaller wavelengths in a concentration-dependent manner. The fluorescence wavelength shifts could be fitted with a binding isotherm yielding a dissociation constant of K_d = 211.6 ±7.4 μM (Figure 61 B and C). When liposomes containing only POPG were titrated to the NTD, the dose-response of the fluorescence wavelength was substantially enhanced as evidenced by a ~10-fold increased dissociation constant of K_d = 28.1 ±5.6 μM (Figure 61 B and C). The correlation between liposome surface charge and binding affinity strongly suggests that the TRPV4 PBD binds to anionic lipids via electrostatic interactions. In agreement with this, the interaction of the PBD with lipids is reduced at high ionic strength (Figure 61 D and E). Increasing the salt concentration in the buffer from 100 mM to 500 mM gradually diminishes the fluorescence wavelength shift between free and lipid-bound PBD. This indicates a loss of electrostatic interaction through charge-screening. Similarly, lipid interaction of the PBD is lost when replacing the basic residues by alanine. Tryptophan fluorescence spectra of the NTD^{AAWAA} mutant are almost completely unaffected by the addition of liposomes, regardless of their lipid composition (Figure 61 F and G). These results confirm that the TRPV4 PBD non-specifically interacts with anionic lipids through electrostatic interaction between basic residues and negatively charged lipid headgroups.

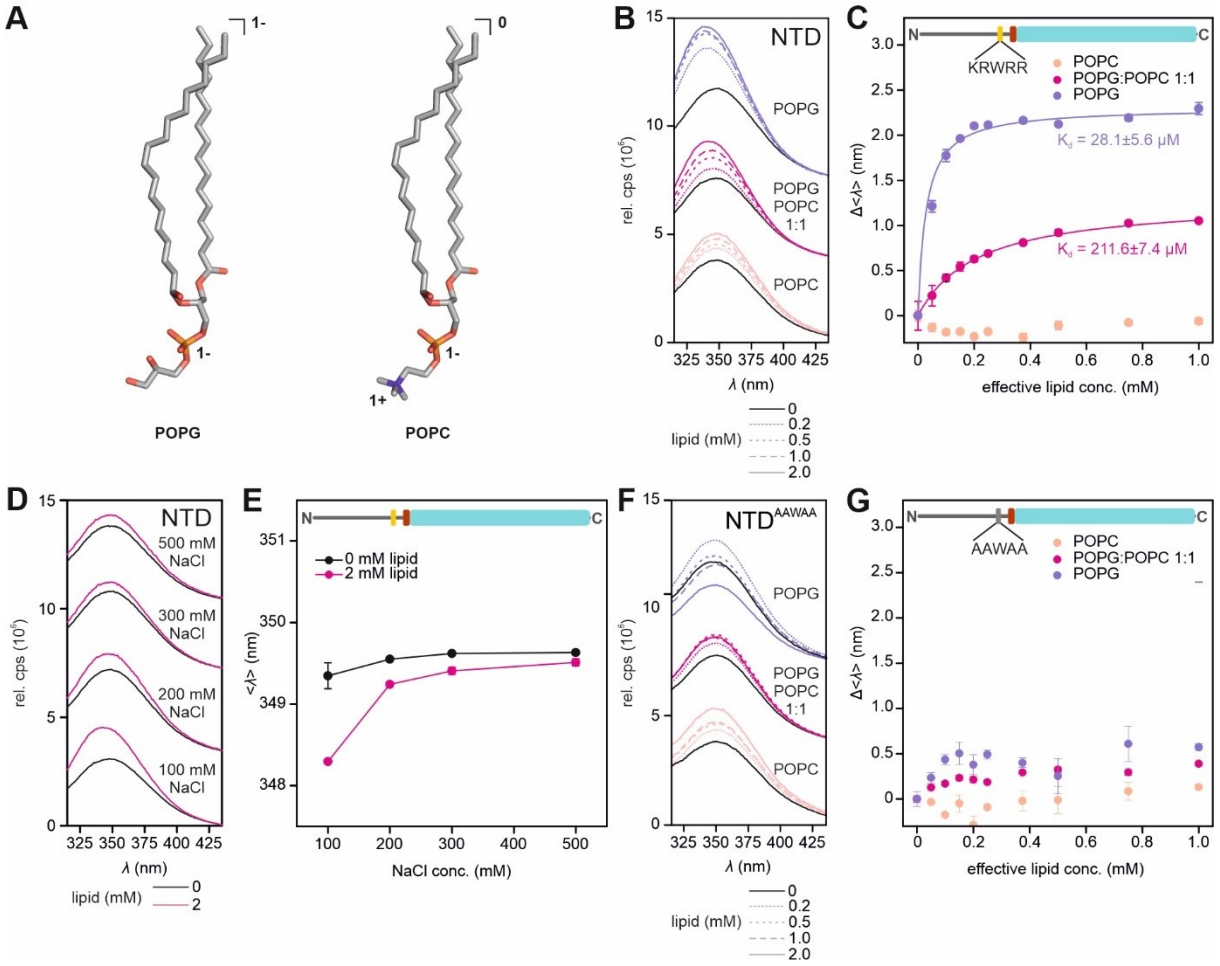


Figure 61: Electrostatic lipid interaction of the *G. gallus* TRPV4 NTD monitored via tryptophan fluorescence spectroscopy. (**A**) Structure of the anionic lipid POPG (net charge 1-) and the zwitterionic lipid POPC (net charge 0). (**B**) Tryptophan fluorescence spectra of the *G. gallus* TRPV4 NTD in the presence of 0/0.2/0.5/1.0/2.0 mM lipid (10 mM Tris pH 7, 100 mM NaCl). Lipids were used in the form of liposomes with 100 nm diameter containing POPC only, POPG:POPC (1:1 mol:mol), or POPG only. (**C**) Change in tryptophan fluorescence wavelength $\Delta\langle\lambda\rangle$ of spectra in (B) plotted against the effective lipid concentration (half of the total lipid concentration, see section VI2.13). The data were fitted with a one-site binding isotherm yielding K_d = 28.1 ± 5.6 µM for POPG-liposomes and K_d = 211.6 ± 7.4 µM for POPG-POPC 1:1 containing liposomes (see section VI2.13, p. 52). (**D**) Tryptophan fluorescence spectra of the *G. gallus* TRPV4 NTD in the absence and presence of 2.0 mM lipid (POPG:POPC 1:1) at 100/200/300/500 mM NaCl. (**E**) Tryptophan fluorescence maximum wavelength $\langle\lambda\rangle$ of spectra in (D) plotted against NaCl concentration. The $\langle\lambda\rangle$ difference between NTD in the presence and NTD in the absence of lipids decreases with increasing NaCl concentration. This strongly suggests electrostatic lipid interactions which are lost at high ionic strength through Debye-Hückel-screening. (**F**) Tryptophan fluorescence spectra of the NTD^{AAWAA} mutant in the presence of 0/0.2/0.5/1.0/2.0 mM lipid (10 mM Tris pH 7, 100 mM NaCl). (**G**) Change in tryptophan fluorescence maximum wavelength $\Delta\langle\lambda\rangle$ of spectra in (F) plotted against the effective lipid concentration. The loss of basic residues in the PBD (^{107}KRWRR^{111} → ^{107}AAWAA^{111}) abolishes lipid interactions in the PBD as evidenced by an almost unchanged $\langle\lambda\rangle$ upon lipid addition. The error bars in(C), (E), and (G) represent the standard deviation from the mean value of three technical replications.

To assess whether the regions outside of the PBD are required for lipid binding, the NTD, the IDR, and the PH-PRR peptide, as well as the corresponding PBD mutants (NTD^{AAWAA}, IDR^{AAWAA}, PH^{AAWAA}-PRR), were titrated with liposomes containing a 1:1 molar mix of POPG and POPC, while monitoring the resulting tryptophan fluorescence wavelength shift (Figure 62). The PBD mutants NTD^{AAWAA}, IDR^{AAWAA}, and PH^{AAWAA}-PRR were used as controls. The wt-TRPV4 constructs displayed clear wavelength shifts when titrated with liposomes, whereas the fluorescence wavelength of the mutants remained almost unchanged (Figure 62 A, B, and C). The wavelength shifts of wt-NTD, IDR, and PH-PRR plotted versus the effective lipid concentration can be fit with a binding isotherm yielding similar association constants in the micromolar range: K_d = 211.6 ± 7.4 µM for wt-NTD, K_d = 95.1 ± 11.6 µM for wt-IDR, and K_d = 79.1 ± 6.9 µM for wt-PH-PRR (Figure 62 D and E). Interestingly, the NTD displays a slightly reduced lipid-affinity compared to IDR and PH-

PRR. This might indicate the existence of NTD conformations in which the ARD reduces lipid binding to the PBD. Because the fluorescence wavelength is a measure of the polarity/hydrophobicity in the environment of tryptophan residues, it can also reflect on how well a tryptophan side chain inserts into the bilayer of lipids. With around 349.7 nm, the fluorescence wavelength of the liposome bound IDR is almost 1.5 nm larger than for the NTD and the PH-PRR (both at around 348.2 nm). The higher fluorescence wavelength of the IDR might therefore point to a weaker insertion of W109 into the lipid bilayer in the context of the IDR than in the NTD and in the PH-PRR peptide.

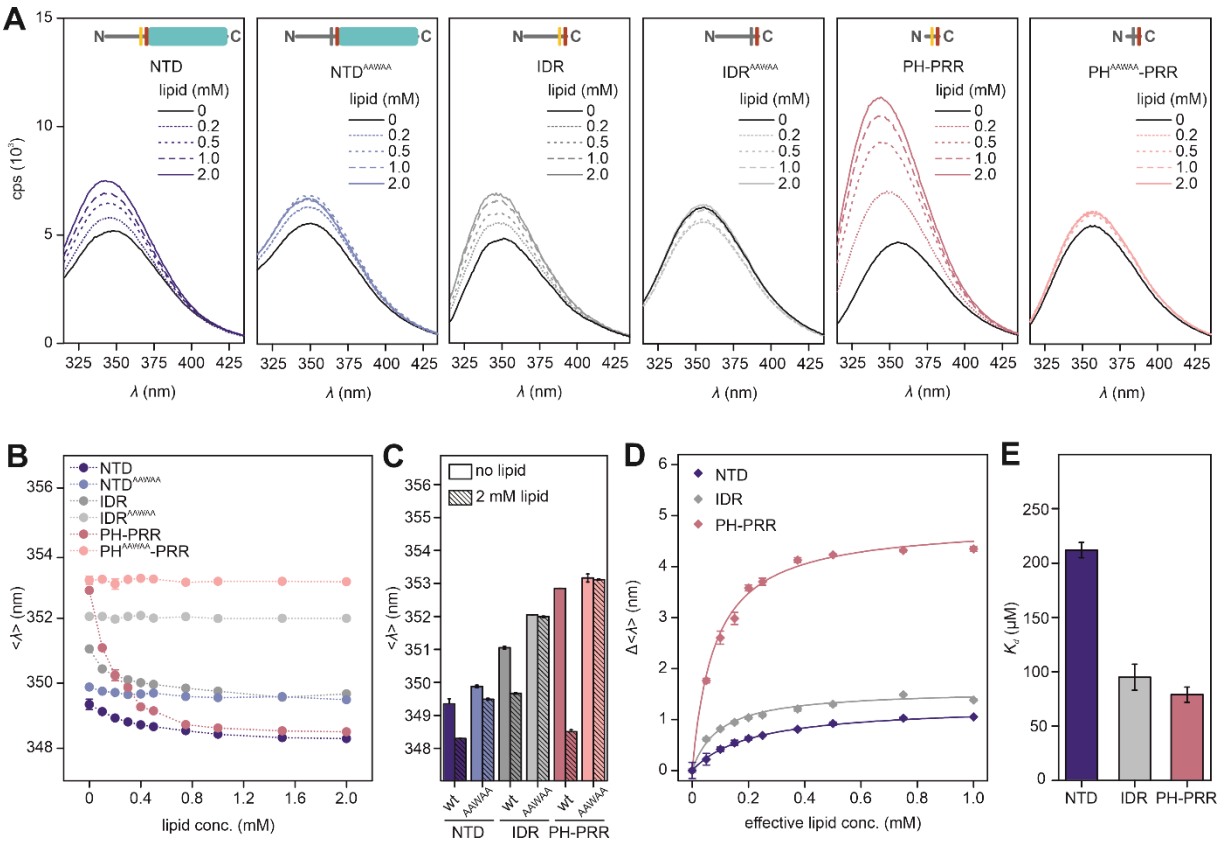


Figure 62: Contribution of ARD and IDR to the lipid interaction of the *G. gallus* TRPV4 NTD analyzed via tryptophan fluorescence spectroscopy. (**A**) Tryptophan fluorescence spectra of *G. gallus* TRPV4 wt- and ^{107}AAWAA^{111} NTD, IDR, and PH-PRR in the presence of 0/0.2/0.5/1.0/2.0 mM lipid. Lipids were used as liposomes with 100 nm diameter containing POPG:POPC 1:1. (**B**) Tryptophan fluorescence maximum wavelength <λ> of spectra in (A) and plotted against the lipid concentration. (**C**) Tryptophan fluorescence wavelength <λ> of TRPV4 NTD constructs in apo state and in the presence of lipids (2 mM lipid). (**D**) Tryptophan fluorescence wavelength change Δ<λ> observed in (B) plotted against the effective lipid concentration. The data were fitted with a one-site binding isotherm yielding K_d = 211.6 ± 7.4 µM for wt-NTD, K_d = 95.1 ± 11.6 µM for wt-IDR, and K_d = 79.1 ± 6.9 µM for wt-PH-PRR. Because the ^{107}AAWAA^{111} mutants displayed no wavelength shifts upon lipid titration (see subfigure B), no Δ<λ> data were plotted for the mutants. (**E**) Comparison of K_d values of wt-constructs obtained in (D). The error bars in (B) and (D) represent the standard deviation from the mean value of three technical replications.

Even though lipid-binding to the PBD seems to be influenced by the ARD and regions in the IDR, the PH-PRR peptide comprises the minimal sequence required for interaction with liposomes. To test whether lipid binding induces secondary structure formation in the PBD, the PH-PRR peptide was investigated via far-UV CD spectroscopy in the presence of TFE, SDS, and liposomes (Figure 63). Even at high TFE concentrations (<60%), the PH-PRR CD spectrum indicates a very disordered peptide (Figure 63 A). A subtle negative increase in the MRE value at 222 nm at TFE concentrations >60% suggests a very low propensity to form α-helical structures in hydrophobic environments. Accordingly, titration with sodium dodecyl sulfate (SDS), a lipid-like detergent with a critical micellar concentration (CMC) of ~8 mM, or with liposomes did not affect the CD spectrum of PH-PRR (Figure 63 B and C). In agreement with the prediction obtained with the HeliQuest tool

(Figure 59 B), the CD spectroscopic analysis confirms that the PBD remains unfolded upon interaction with lipids.

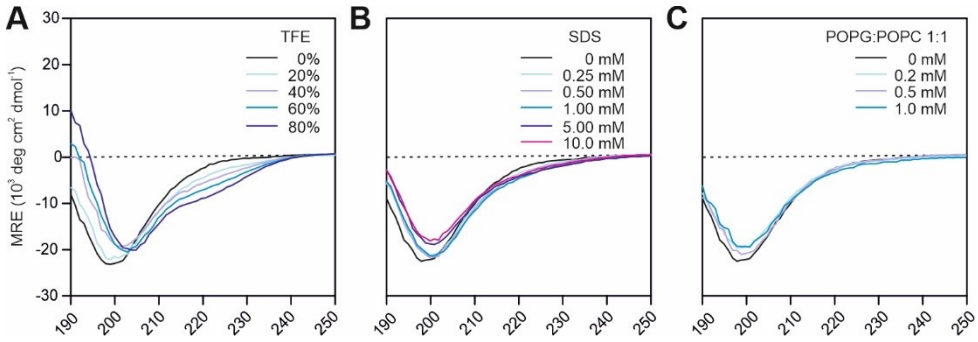


Figure 63: CD spectroscopic analysis of potential lipid-induced folding of the TRPV4 phosphoinositide binding domain. Far-UV CD spectra of the *G. gallus* TRPV4 PH-PRR peptide in the presence of increasing (**A**) TFE, (**B**) SDS, and (**C**) liposome concentrations. The 50 nm diameter liposomes were prepared with a POPG:POPC 1:1 (molar:molar) lipid content. All spectra were recorded in 10 mM Tris pH 7, at 298 K.

As the minimal sequence required for lipid binding, the PH-PRR peptide is well-suited to investigate how lipids interact with the PBD on the per-residue level using NMR spectroscopy. For this, ^{13}C, ^{15}N-labeled PH-PRR and PH^{AAWAA}-PRR peptides were purified to obtain the backbone NMR resonance assignments (Figure 64). As observed for the IDR (Figure 45), the disordered character of the peptides facilitates solvent exchange at pH 7, resulting in severe line broadening. This effect could be substantially suppressed at low pH without changing the secondary structure of the peptides, as confirmed via CD spectroscopy (Figure 65 A). Near-complete NMR backbone assignments of the PH-PRR and PH^{AAWAA}-PRR peptides could be obtained at pH 4.5 using a set of triple resonance NMR experiments (see section 2.17). Backbone resonances (H^N, N, C, C_α, H_α) of ^{13}C, ^{15}N-PH-PRR were assigned to 90.6%, whereas the backbone resonances of ^{13}C, ^{15}N- PH^{AAWAA}-PRR were assigned to 91.1% (see section 2.17). Importantly, the lipid-binding behaviour of the PH-PRR and PH^{AAWAA}-PRR peptides was nearly unaffected by the low pH conditions necessary for NMR spectroscopy (Figure 65 B and C). Tryptophan fluorescence measurements clearly show that the PH^{AAWAA}-PRR peptide does not interact with POPC or POPG:POPC 1:1 liposomes at pH 4.5. The wildtype PH-PRR peptide also does not interact with POPC liposomes but binds to POPG POPC 1:1 liposomes with a similar affinity at pH 4.5 as at pH 7.

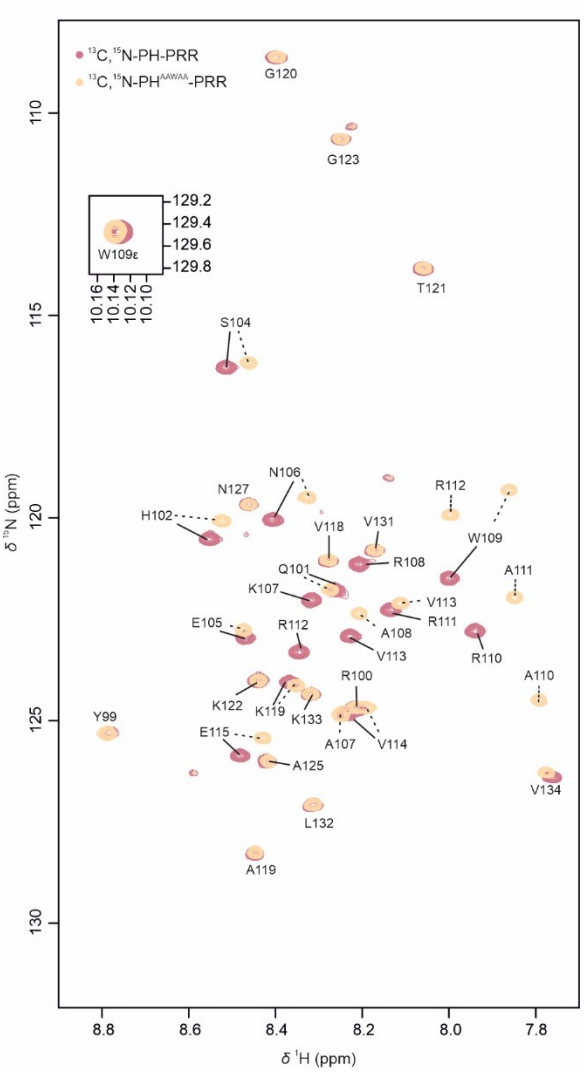


Figure 64: NMR backbone assignment of the *G. gallus* TRPV4 wt-PH-PRR and the PBD mutant PH^{AAWAA}-PRR peptide. (**A**) ^{1}H, ^{15}N-HSQC spectra and NMR backbone assignments of 310 µM ^{13}C, ^{15}N-PH-PRR (dark red) and 240 µM ^{13}C, ^{15}N-PH^{AAWAA}-PRR (light orange). Spectra were recorded at 298 K in 20 mM NaP_i pH 4.5, 150 mM NaCl, 1 mM DTT, 10% D_2O. The narrow peak distribution in the ^{1}H, ^{15}N-HSQC spectrum of the wt-PH-PRR and PH^{AAWAA}-PRR mutant indicates that both peptides are disordered in solution at pH 4.5. Mutating the PBD thus does not introduce structural changes.

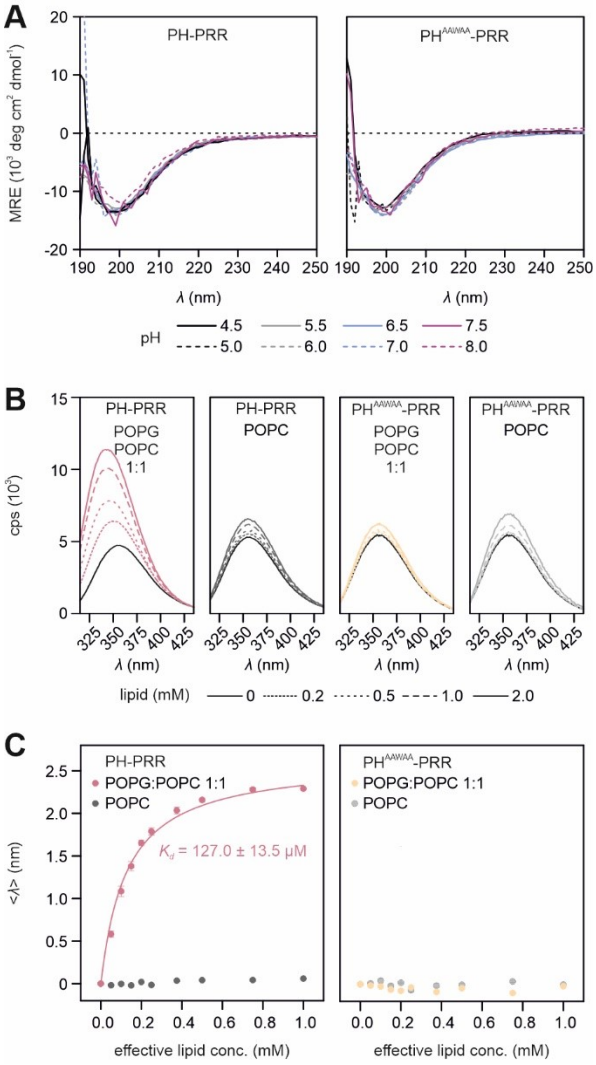


Figure 65: Secondary structure and lipid interaction analysis of the *G. gallus* TRPV4 wt-PH-PRR and the PBD mutant PH^{AAWAA}-PRR peptide at pH 4.5. (**A**) Far-UV CD spectra of PH-PRR and PH^{AAWAA}-PRR at pH 4.5/5.0/5.5/6.0/6.5/7.0/8.0. (**B**) Tryptophan fluorescence spectra of PH-PRR and PH^{AAWAA}-PRR at pH 4.5 in the presence of 0/0.2/0.5/1.0/2.0 mM lipid (POPG POPC 1:1, 100 nm liposomes). (**C**) Plot of fluorescence wavelength shifts $\Delta\langle\lambda\rangle$ in spectra in (C) against the effective lipid concentration. The data for PH-PRR with POPG:POPC 1:1 liposomes were fitted with a one-site binding isotherm yielding K_d = 127.0 ± 13.5 µM. This is in good agreement with the affinity of PH-PRR with POPG:POPC 1:1 liposomes at pH 7 (K_d = 79.1 ± 6.9 µM, see Figure 62). No binding behaviour was observed for PH-PRR with POPC liposomes and for the PH^{AAWAA}-PRR mutant with POPG POPC 1:1 and POPC only liposomes. (A), (B), and (C) confirm that the structure and lipid binding behaviour is preserved between pH 4.5 and pH 7. The error bars in (C) represent the standard deviation from the mean value of three technical replications.

The NMR backbone assignments were taken exploited to elucidate the exact lipid-binding site in the PH-PRR peptide. For this, ^{1}H, ^{15}N-HSQC spectra of ^{15}N-labeled PH-PRR and PH^{AAWAA}-PRR were recorded in the presence of increasing lipid concentrations using liposomes containing POPG POPC 1:1 as membrane mimics. Titration of liposomes severely decreased the intensities of the ^{15}N-PH-PRR resonances, whereas the ^{15}N-PH^{AAWAA}-PRR resonances were almost unaffected (Figure 66 A and B). The intensity loss of the ^{15}N-PH-PRR resonances can be explained by the formation of a large peptide-liposome complex with correlation times beyond the detectability of solution NMR spectroscopy. The relative intensity decrease of the W109 Hε resonance (determined as relative decrease of peak integrals $\Delta V/V_0$) plotted versus the effective lipid concentration could be fitted with a binding isotherm yielding a dissociation constant $K_d = 160.0 \pm 45.1$ μM similar to the one determined from tryptophan fluorescence measurements ($K_d = 127.0 \pm 13.5$ μM) (Figure 66 C and D). Importantly, the intensity losses observed for the ^{15}N-PH-PRR resonances are not equally pronounced throughout the peptide sequence. Plotting the relative intensity decrease between the free and lipid-bound state of ^{15}N-PH-PRR against the amino acid sequence shows a complete signal intensity loss in the region comprising residue 99-118 also encompassing the PBD (Figure 66 E). The region comprising the PRR (residues 121-134) shows only very subtle intensity decreases. This suggests that the residues in and around the PBD tightly stick to the surfaces of the large liposomes and become undetectable. The PRR, in contrast, does not bind to the liposomes and remains sufficiently flexible in the lipid-bound state of the PH-PRR peptide to yield sharp peaks in the NMR spectrum (Figure 66 F). The ^{15}N-PH^{AAWAA}-PRR resonances display almost no intensity loss throughout the amino acid sequence, again reflecting the necessity of an intact PBD for lipid binding.

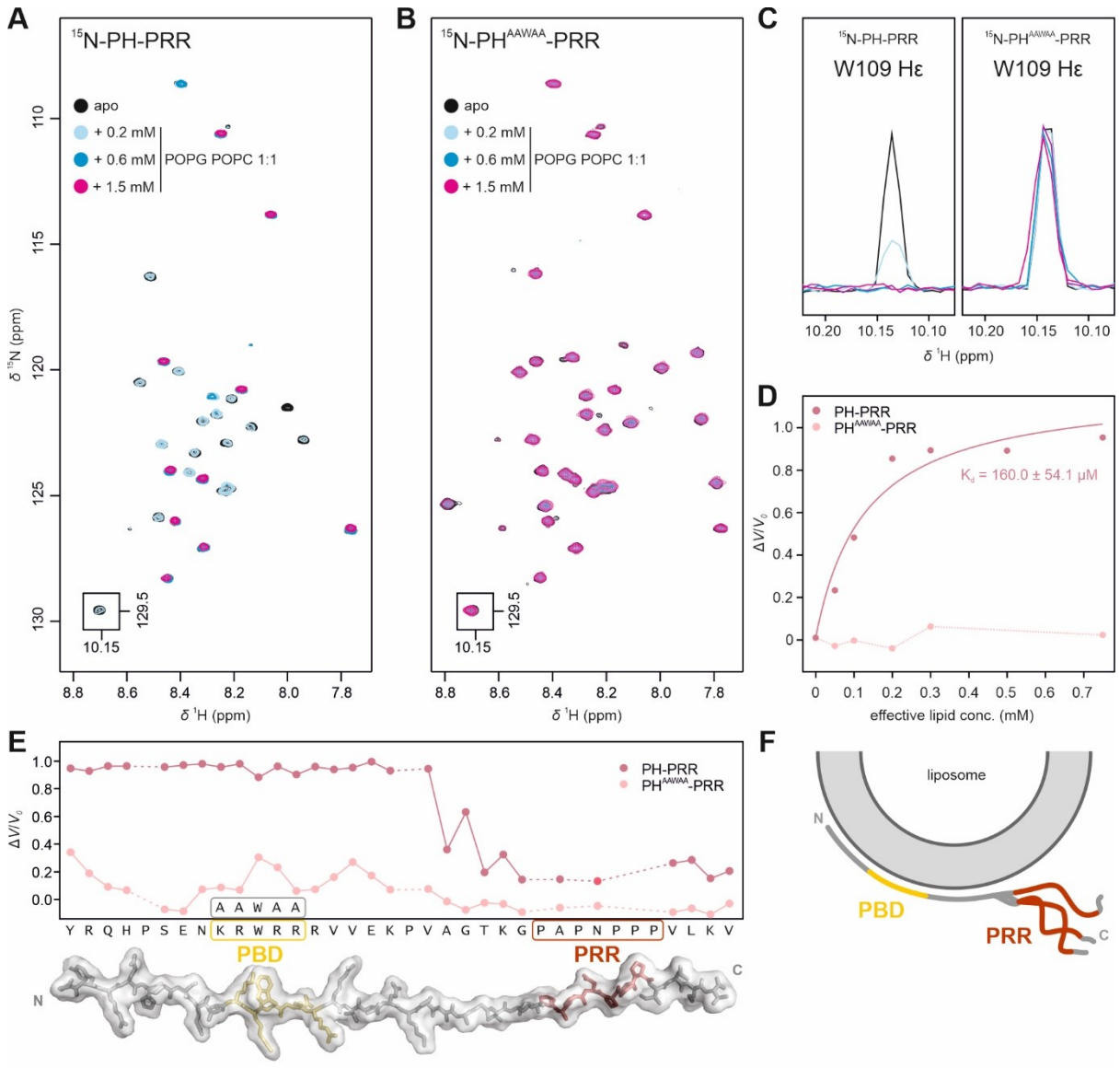


Figure 66: Liposome binding site in the TRPV4 PH-PRR peptide identified via NMR spectroscopy. (**A**, **B**) ^{1}H, ^{15}N-HSQC spectra of (A) 100 µM ^{15}N-PH-PRR and (B) 100 µM ^{15}N-PH^{AAWAA}-PRR in the presence of 0/0.2/0.6/1.5 mM lipid (POPG POPG 1:1 liposomes, 100 nm diameter). Spectra were recorded at 298 K in 20 mM NaP_i pH 4.5, 150 mM NaCl, 1 mM DTT, 10% D_2O. (**C**) 1D ^{1}H projection of W109 Hε side chain resonance of ^{15}N-PH-PRR and ^{15}N-PH^{AAWAA}-PRR in the presence of 0/0.2/0.6/1.5 mM lipid (concentrations and colour code as in (A) and (B)). (**D**) Relative decrease of the W109ε peak integral $\Delta V/V_0$ of ^{15}N-PH-PRR and ^{15}N-PH^{AAWAA}-PRR plotted versus the effective lipid concentration. $\Delta V/V_0$ was calculated as described in section 2.17. The wt-PH-PRR data were fitted with a one-site binding isotherm yielding a K_d = 160.0 ± 45.1 µM, which is in a similar range as the K_d = 79.1 ± 6.9 µM determined under similar conditions via tryptophan fluorescence, see (F) and Figure 65. (**E**) Relative decrease $\Delta V/V_0$ between the apo (0 mM lipid) and lipid-bound state (1.5 mM lipid). (**F**) Schematic model of the PH-PRR peptide bound to liposomes. The N-terminal region comprising residues 99-118 binds tightly to the liposome surface. This results in signal loss of the corresponding resonances in the NMR spectrum. The PRR, in contrast, does not bind to the liposome and remains flexible enough to produce sharp peaks in the NMR spectrum.

To understand whether PIP_2 and anionic lipids share the same binding site in the PH-PRR peptide, chemical shift perturbation (CSP) experiments were carried out using different micelle-forming di-octanoyl (diC_8) lipids as membrane mimics (Figure 67 A). Even though micelles are less representative of native lipid membranes than liposomes, they come with a technical advantage for NMR studies. Peptide-micelle complexes are smaller compared to peptide-liposome complexes and therefore have shorter correlation times which can be within the detectable range of solution NMR spectroscopy (*451*). This facilitates the detection of peptide resonances in the lipid-bound state and opens the path to discern structural details in the lipid-binding sites in the form of

resonance shifts. The extent and direction of resonance shifts between free and lipid-bound states provide parameters to describe lipid-binding quantitatively and qualitatively. Accordingly, ^{1}H, ^{15}N-HSQC-based CSP experiments of ^{15}N-PH-PRR were carried out using diC8-PIP_2, diC_8-PG, and diC_8-PC (Figure 67 A, B, and C). Strong resonance shifts upon titration of diC_8-PIP_2 and diC_8-PG indicate the formation of micelle-peptide complexes. Titration of diC_8-PC, in contrast, had no effect on the PH-PRR resonances suggesting no interaction. This is in line with liposome-based NMR experiments (Figure 65 C and D) as well as the data from tryptophan fluorescence titrations with POPC (Figure 65 B and C). DiC_8-PIP_2 and diC_8-PG-micelle binding to the PH-PRR peptide was accompanied by severe line broadening of resonances corresponding to residues in or close to the PBD (Figure 67 A and B). This effect might be a combination of intermediate exchange rates between free and bound state and increasing correlation times upon formation of micelle-peptide complexes. Even though not visible in the presented NMR spectra (Figure 67 A and B), the chemical shifts of all resonances in the lipid-bound state of the PH-PRR peptide could be determined, and the chemical shift differences (ΔCS) between free and lipid-bound state were plotted versus the residue number of the PH-PRR peptide (Figure 67 D). Chemical shift differences upon titration with diC_8-PIP_2 and diC_8-PG are observed in the same region as the intensity losses observed upon titration with liposomes (Figure 66 E). This suggests that the PH-PRR peptide interacts similarly with lipid micelles bind and liposomes. Surprisingly, the ΔCS pattern is very similar between diC_8-PIP_2 and diC_8-PG with the strongest ΔCS values spanning the sequence ^{105}ENKRWRRRVVE^{115}, which is the region comprising the PBD. The major differences between diC_8-PIP_2 and diC_8-PG binding are the residue-specific ΔCS values, which are higher for PIP_2. The strong ΔCS values for residue V113 and V114 C-terminal to the PBD suggest a substantial contribution of hydrophobic interactions to lipid binding. High ΔCS values implying lipid-binding are also observed in the N-terminal region comprising residues ^{99}YRQH^{102}, which together with the PBD form a PH consensus motif (see Figure 59). PH consensus motifs are commonly found in folded PH domains, which are specificity for phosphoinositide binding. In the TRPV4 PH-PRR peptide, the PH motif, however, cannot provide specificity for PIP_2 over other lipids like PG as the NMR experiment show. This might be attributed to the disordered nature of the region comprising the PH consensus motif in TRPV4. Alternatively, TRPV4 regions which are not contained in the PH-PRR peptide might be required to form a PIP_2 specific binding site together with the PH consensus motif.

To estimate the affinities of diC_8-PIP_2 and diC_8-PG for PH-PRR binding, the ΔCS of the W109 Hε resonance as a representative residue of the PBD was plotted versus the lipid concentration (Figure 67 E). However, the plotted data display a 'delayed' binding behaviour, which was not compatible with a conventional Langmuir-binding isotherm. Below 0.2 mM lipid, the ΔCS value remains nearly unaffected followed by an asymptotic increase towards a ΔCS value >0.15 ppm at high lipid concentrations. Sigmoidal response curves, as observed upon titrating the PH-PRR peptide with diC_8-lipids, are often linked to cooperativity underlying an interaction. Such behaviour is accounted for in the Hill-Langmuir model (see section 2.17). Indeed, the ΔCS of the W109 Hε resonance plotted versus the lipid concentration could be fitted with a Hill-Langmuir isotherm yielding the dissociation constants K_d = 427.4 ± 17.8 μM with n = 2.9 ± 0.3 for diC_8-PIP_2 and K_d = 542.4 ± 30.9 μM with n = 3.0 ± 0.4 for diC_8-PG (Figure 67 E). Hill-coefficients of $n > 1$ are generally indicative of positive cooperativity, which means that binding of one ligand facilitates the binding of additional ligands. The coefficients, however, do not tell about the number of bound ligands. In the context of the PH-PRR peptide, $n > 1$ might indicate that the binding of one lipid molecule promotes the binding of additional lipid molecules.

It is important to note, that the interpretation of the determined K_d values is substantially complicated by the fact, that it is unclear whether diC_8-PIP_2 and diC_8-PG bind as monomers or micelles to the PH-PRR peptide. The CMC published for diC_8-PG is ~1.2 mM§ (*451*), whereas the CMC of diC_8-PIP_2 is estimated to be larger than 4 mM** (*452*). Therefore, both lipids should exist as monomers in the concentration range displaying a sigmoidal behaviour in the ΔCS versus lipid concentration plot. It is, however, possible that an increased local lipid concentration due to the binding of multiple lipids to the PH-PRR peptide triggers micelle formation at concentrations below the actual CMC. Micelle formation might be further aided by the fact that the interaction

§ Value was determined at pH 10 (*451*).

** This estimate is based on the CMC of ~3-4 mM determined for dIC_8-PI(4)P at pH 7.4. (*452*) The additional phosphate group in the lipid head group of diC_8-PI(4,5)P_2 can be expected to increase the CMC.

with basic side chains of PH-PRR can partially screen lipid headgroup charges. The effects lipid binding on the monomer-micelle equilibrium of diC_8-PIP_2 and diC_8-PG might be quite different. The similar K_d values determined for the binding of diC_8-PIP_2 and diC_8-PG should therefore not be interpreted as similar affinities of the PBD for PG or PIP_2 binding.

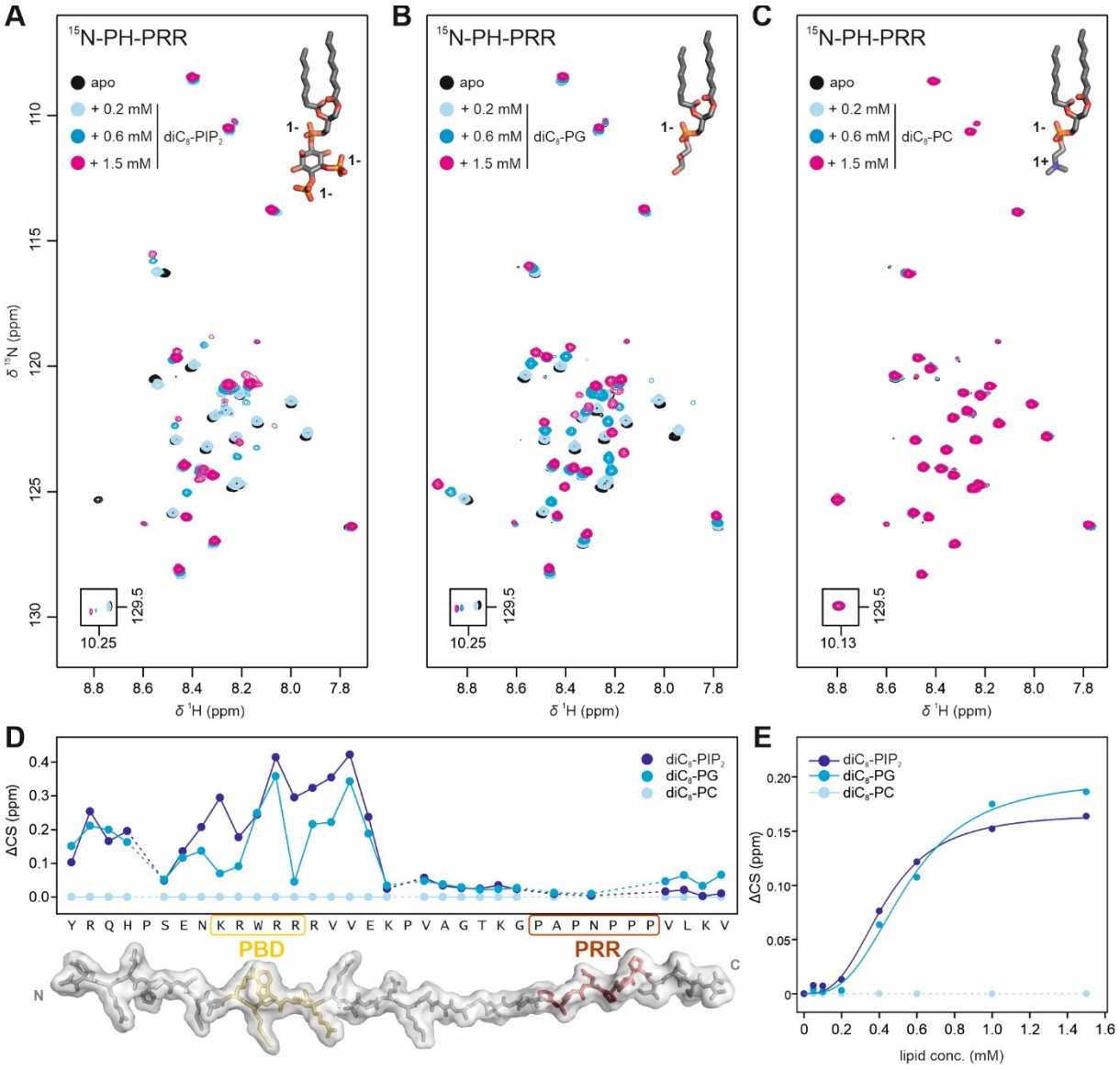


Figure 67: Lipid-specificity of the TRPV4 PBD probed via NMR spectroscopy. ^1H, ^{15}N-HSQC spectra of 100 µM ^{15}N-PH-PRR in the presence of 0/0.2/0.6/1.5 mM diC_8-PIP_2 (**A**), diC_8-PG (**B**), or diC_8-PC (**C**). Spectra were recorded at 298 K in 20 mM NaP_i pH 4.5, 150 mM NaCl, 1 mM DTT, 10% D_2O. (**D**) Chemical shift change (ΔCS) of PH-PRR resonances upon addition of 1.5 mM lipid plotted versus the residue number. (**F**) ΔCS of the W109 Hε resonance plotted versus the lipid concentration. The data were fitted with a Hill-Langmuir isotherm yielding K_d = 427.4 ± 17.8 µM with n = 2.9 ± 0.3 for diC_8-PIP_2 and K_d = 542.4 ± 30.9 µM with n = 3.0 ± 0.4 for diC_8-PG. diC_8-PC did not induce chemical shift changes, thus indicating no binding to the TRPV4 PBD.

The ΔCS versus residue plot of the CSP experiment with diC_8-PIP_2 demonstrates that the PIP_2 binding site exceeds the ++W++ motif of the PBD. The hydrophobic residues V113 and V114 and the sequence ^{99}YRQH^{102}, which contains two basic and one aromatic residue, seem to be strongly involved in lipid binding. Both regions are preserved in the ^{107}AAWAA^{111} mutant. This raises the question, whether the PH^{AAWAA}-PRR peptide exhibits a residual PIP_2 binding capacity. The ^1H, ^{15}N-HSQC-based CSP experiment of [^{5}N-PH^{AAWAA}-PRR with diC_8-PIP_2 reveals that the ^{107}AAWAA^{111} mutation almost completely abolishes the interaction with

PIP_2 (Figure 68 A, B, and C). Residual peak shifts can be observed for S104 and H102 (Figure 68 D). This observation again shows that electrostatic interactions between the basic residues in the PBD and the PIP_2 head group are the driving force for the TRPV4-PIP_2 interaction. The remaining basic and hydrophobic residues outside the PBD cannot compensate for the loss of the ++W++ motif (in the context of the PH-PRR peptide).

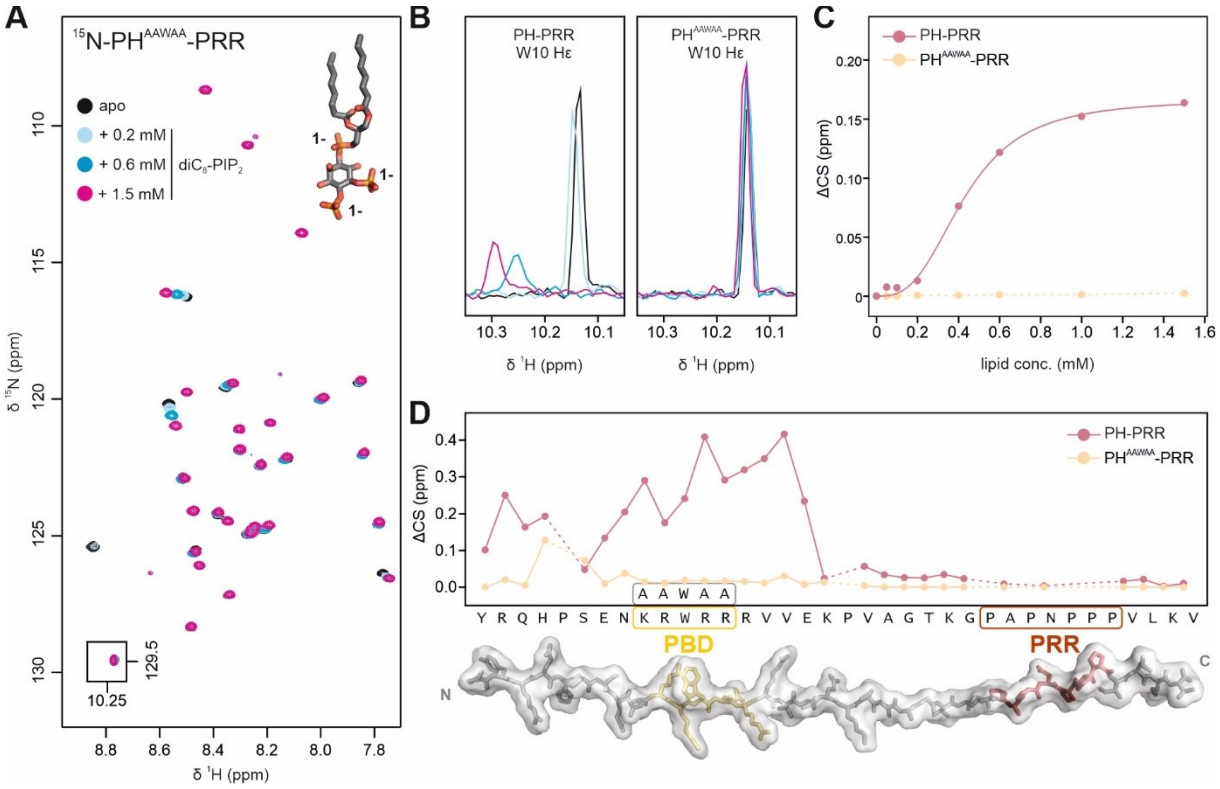


Figure 68: Effect of the ^{107}AAWAA^{111} mutation on PIP_2 binding to the TRPV4 PBD monitored via NMR spectroscopy. (**A**) ^{1}H, ^{15}N-HSQC spectra of 100 µM ^{15}N-PH^{AAWAA}-PRR in the presence of 0/0.2/0.6/1.5 mM dIC_8-PIP_2. Spectra were recorded at 298 K in 20 mM NaP_i pH 4.5, 150 mM NaCl, 1 mM DTT, 10% D_2O. (**B**) 1D-^{1}H projection of W109 Hε resonance of ^{15}N-PH-PRR and ^{15}N-PH^{AAWAA}-PRR in the presence of 0/0.2/0.6/1.5 mM dIC_8-PIP_2 (concentrations and colour code as in (A) and (B)). (**C**) Chemical shift change (ΔCS) of PH-PRR and PH^{AAWAA}-PRR W109ε side chain resonance plotted versus the lipid concentration. (**D**) ΔCS of PH-PRR and PH^{AAWAA}-PRR resonances upon addition of 1.5 mM diC_8-PIP_2 plotted versus the residue number.

Considering the conserved TRPV4 PBD amino acid sequence (Figure 59 B), one may assume that the non-specific binding of anionic lipids to the PBD is conserved in TRPV4 ion channels across species. To confirm this, lipid binding was assessed for both the *G. gallus* and *H. sapiens* TRPV4 IDR via tryptophan fluorescence measurements (Figure 69). Clear fluorescence shifts to lower wavelengths were observed in the IDR spectra upon addition of liposomes with POPG:POPC 1:1 (Figure 69 A and B). No changes are detected when liposomes containing only POPC were titrated to the IDRs (Figure 69 A and B). In addition, fluorescence wavelength changes upon addition of excess concentrations of liposomes containing POPG:POPC 1:1 are substantially suppressed at increasing ionic strength in the buffer (Figure 69 C, D, E, and F). This confirms that the electrostatic interaction of the TRPV4 IDR with anionic lipids is conserved across species. The wavelength shifts upon titration with liposomes containing 50% POPG and 50% POPC were fitted with a binding isotherm yielding a dissociation constant of K_d = 250.0 ± 33.9 µM for the *H. sapiens* TRPV4 IDR (Figure 69 B). For the *G. gallus* TRPV4 IDR, a dissociation constant of K_d = 95.1 ± 11.6 µM was determined (see Figure 62).

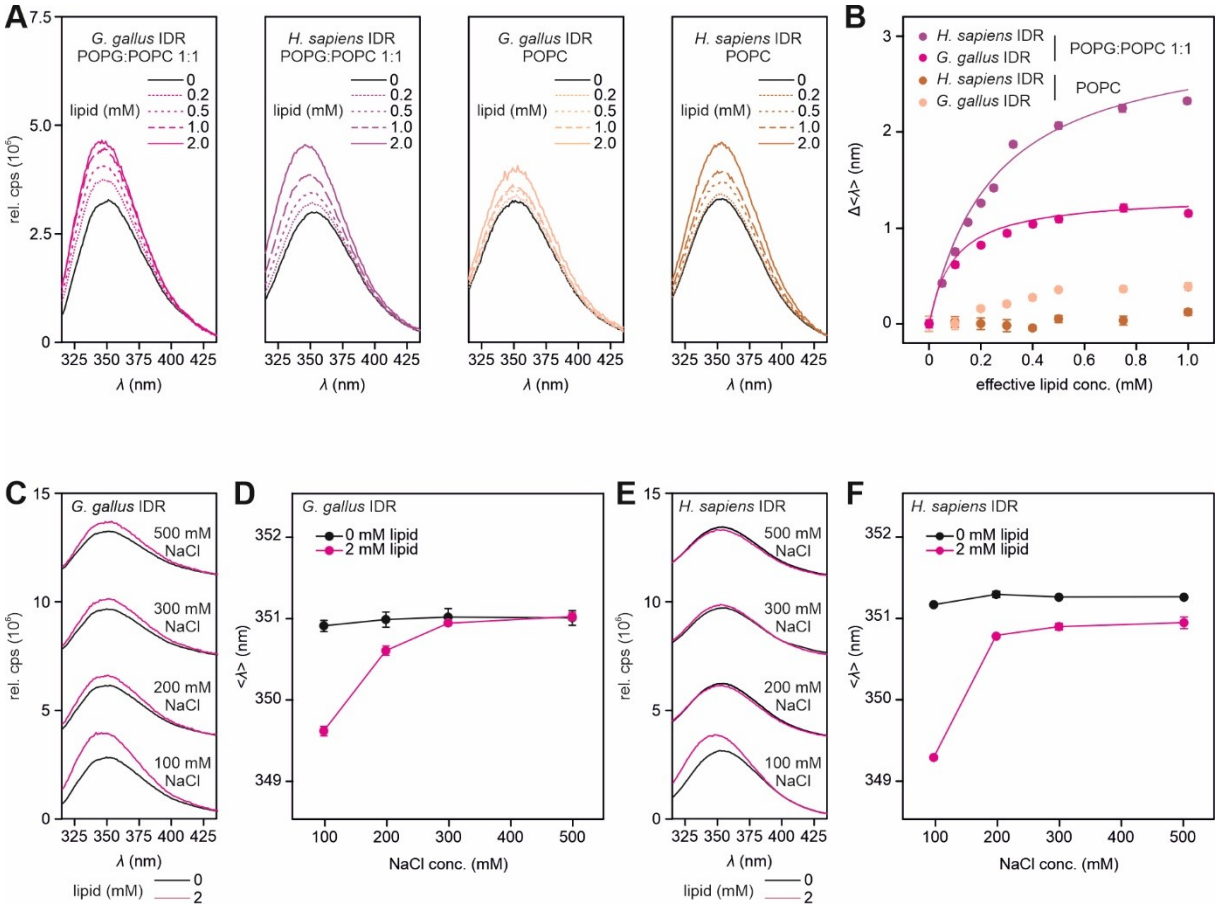


Figure 69: Conservation of lipid binding in the *G. gallus* and *H. sapiens* TRPV4 IDR monitored via tryptophan fluorescence spectroscopy. (**A**) Tryptophan fluorescence spectra of the *G. gallus* and *H. sapiens* TRPV4 IDR in the presence of 0/0.2/0.5/1.0/2.0 mM lipid. Lipids were used as liposomes with 100 nm diameter containing POPG:POPC 1:1 or POPC only. (**B**) Plot of fluorescence wavelength shift $\Delta\langle\lambda\rangle$ obtained from the spectra in (A) against the effective lipid concentration. A dissociation constant of K_d = 250.0 ±33.8 µM could be determined for *H. sapiens* IDR with POPG:POPC 1:1. In comparison to this, the dissociation constant of the *G. gallus* IDR is K_d = 95.1 ± 11.6 µM. No lipid-binding behaviour was observed when POPC-only liposomes were used. (**C**, **E**) Tryptophan fluorescence spectra of the (C) *G. gallus* and the (E) *H. sapiens* TRPV4 IDR in the absence and presence of 2.0 mM lipid (POPG:POPC 1:1) at 100/200/300/500 mM NaCl. (**D**, **F**) Tryptophan fluorescence maximum wavelength $\langle\lambda\rangle$ of spectra in (C) and (E) plotted against the NaCl concentration. The $\langle\lambda\rangle$ difference between IDR in the presence and IDR in the absence of lipids decreases with increasing NaCl concentration. In agreement with the *G. gallus* TRPV4 NTD (Figure 61), this suggests that the IDR lipid interaction is driven by electrostatic forces, a mechanism conserved across species. The error bars in (B), (D) and (F) represent the standard deviation from the mean value of three technical replications.

To complement the tryptophan fluorescence experiments, the *H. sapiens* IDR was fluorescently labelled by attaching an Atto488-maleimide dye to a cysteine side chain for fluorescence correlation spectroscopy (FCS) experiments (Figure 70 A). The *H. sapiens* TRPV4 IDR does not contain a native cysteine residue. Therefore, an artificial cysteine was introduced in the IDR C-terminus by generating an IDR_V148C mutant (see section VI2.14). The IDR_V148C mutant was labelled with Atto488 with an efficiency of 12% as determined via SEC (Figure 70 A). Figure 70 B shows the autocorrelation curves (ACC) obtained from the diffusion of IDR_V148C_Atto488 through the FCS detection volume in the absence and in the presence of 100 nm liposomes containing only POPC or POPG:POPC 1:1. In the absence of liposomes, the IDR ACC can be fit with an autocorrelation function accounting for a single species, i.e. the free IDR, which diffuses through the detection volume in 105 µs ($\tau_{D,1}$). According to equation 15 and 16 (see section VI2.14), this translates to a Stokes radius of R_S = 2.4 nm, which is smaller than R_S of the IDR determined via SEC (R_S^{SEC} = 3.0 nm, see Figure 40). The smaller R_S determined for the IDR_V148C mutant with FCS might be caused by the Atto488-label. It could indicate that the label induces a more compact conformation of the IDR compared to the label-free protein. Upon addition of increasing concentrations of liposomes containing POPG:POPC 1:1, the ACC displays a gradually increasing fraction of a second species which diffuses drastically slower through the detection

volume (Figure 70 B, upper panel). Fitting an autocorrelation function to the ACCs which accounts for two diffusion times – the first diffusion time was set to $\tau_{D,1}$ = 105 µs, the freely diffusing IDR – yields a diffusion time of $\tau_{D,2}$ = 2600 µs for the second species. According to equation 15 and 16 (see section VI2.14), this translates to a particle with a Stokes radius of 50 nm. This particle size fits well to an IDR_V148C_Atto488 species bound to the 100 nm diameter liposomes titrated to the sample. The increase of the fraction of liposome bound IDR in the ACC can be followed in a concentration-dependent manner (Figure 70 B, upper panel). The fraction of lipid-bound IDR plotted versus the lipid concentration could be fitted with a binding isotherm yielding a dissociation constant of K_d = 70.2 ± 23.1 µM (Figure 70 C). This value is in good agreement with the dissociation constant of the *G. gallus* IDR (K_d = 95.1 ± 11.6 µM) determined via tryptophan fluorescence. In the presence of liposomes containing only POPC, the ACC of the IDR_V148C_Atto488 construct is almost identical to the AC curve of the protein alone (Figure 70 B, lower panel). In agreement with tryptophan fluorescence data, this demonstrates that the IDR does not interact with liposomes containing only POPC (Figure 70 B, lower panel).

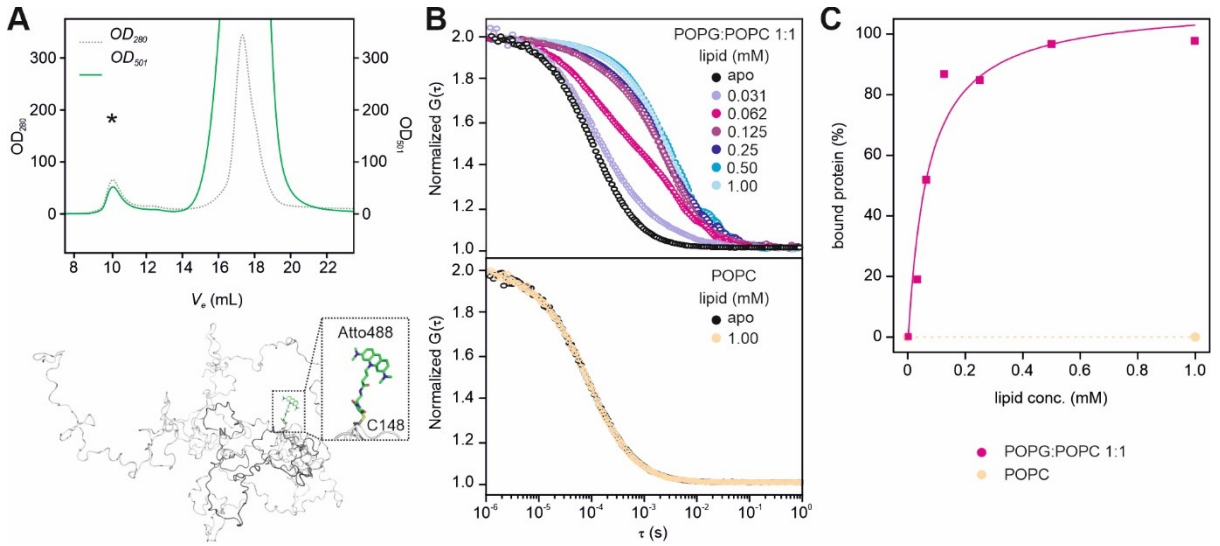


Figure 70: Liposome interaction of *H. sapiens* TRPV4 IDR monitored via fluorescence correlation spectroscopy. (**A**) Top: Size-exclusion chromatogram (using a GE Healthcare Superdex 75 16/60 column) of *H. sapiens* TRPV4 IDR_V148C mutant labelled with Atto488 maleimide. The peak at ~10 mL corresponds to the Atto488-labeled IDR, whereas the peak at ~17.5 mL corresponds to free dye. A labelling efficiency of 12% was determined from the ratio of OD_{280}/OD_{501} (see section VI2.14). Bottom: The modelled structural ensemble of Atto488-maleimide labelled *H. sapiens* TRPV4 IDR demonstrates the size of the Atto488 dye compared to the protein. The IDR ensemble was generated from the EOM refined bead models (see section VI2.20). The Atto488 maleimide label (structure shown in the dashed box) was attached covalently to C148 in PyMOL (*322, 323, 453*). The IDR conformers were aligned along the four C-terminal residues. (**B**) Fluorescence ACCs of IDR_V148_Atto488 in the apo state (0 mM lipid), in the presence of POPG:POPC 1:1 liposomes (0.031/0.063/0.125/0.25/0.50/1.00 mM lipid), and in the presence of POPC only liposomes (1 mM lipid). Plotted is the normalized autocorrelation G(τ) versus the lag time τ. (**C**) Plot of lipid-bound protein versus lipid concentration. The liposome bound protein fraction was obtained by fitting the ACC in (B) with autocorrelation function accounting for two diffusing species. The slower diffusing species corresponds to the IDR bound to liposomes. Fitting the increase of the liposome-bound protein fraction with a binding isotherm yielded a dissociation constant of K_d = 70.2 ± 23.1 µM.

In conclusion, tryptophan fluorescence, CD and NMR spectroscopy demonstrate that the TRPV4 PBD non-specifically interacts with anionic lipids through electrostatic attraction. This behaviour is conserved across species. The PBD does not discriminate between PIP_2 and other anionic lipids such as PG. Instead, PG binds to the exact same residues as PIP_2. This is rather unexpected given the very different lipid head groups of PG and PIP_2. Not only is the PIP_2 head group very bulky, but it also contains two terminal phosphate groups, each with a charge of -1††. That makes PIP_2 more negatively charged than other lipids like PG. The similar lipid-

†† The pKa value of the first protonation step is ~6.7 at phosphate-4 and ~7.7 at phosphate-5 for PIP_2 (*454*). At pH 4.5, the PIP_2 phosphate groups are therefore expected to have a charge of -1, resulting in a total charge of -3.

binding modes of PIP_2 and PG in the PBD are probably a result of the disordered character of the entire region encompassing the PBD. Interaction of the PBD with lipids, therefore, seems to be predominantly dependent on the charge of the lipids but not so much on the structure of their headgroup. A specific lipid interaction presumably requires a domain which is either constitutively folded or folds upon binding to lipids. However, in agreement with sequence-based α-helix predictions, the PBD in the context of the PH-PRR peptide has a low tendency to fold secondary structures and remains entirely disordered upon interaction with lipids. Alternatively, the PBD might act together with other domains in TRPV4, such as the ARD, to form a PIP_2-specific binding site. The ARD or the rest of the IDR, however, did not enhance binding of the PBD to the anionic lipid PG. Whether this is the case for PIP_2 remains to be investigated, e.g. through tryptophan fluorescence spectroscopy of TRPV4 N-terminal constructs with liposomes containing PIP_2. Importantly, the NMR experiments show, that the binding of anionic lipids to the PH-PRR peptide not only involves the conserved ++W++ motif. Hydrophobic and basic residues in the vicinity of the PBD seem to contribute to lipid binding, including the ^{99}YRQH^{102} sequence which forms a PH consensus motif with the PBD. Nevertheless, mutating the basic residues in the PBD to alanine suffices to abolish lipid binding, including PIP_2, to the PBD almost completely. This supports previously published limited proteolysis experiments in which liposomes containing PIP_2 protected the TRPV4 NTD from proteolysis but not the NTD^{AAWAA} mutant (*128*).

In the context of a full-length TRPV4 ion channel in a cell, the PBD is constantly positioned in the vicinity of the plasma membrane. Considering that the PBD in the context of isolated TRPV4 N-terminal constructs has a K_d of ~100 μM for anionic phospholipids, the PBD should be constitutively bound to the plasma membrane since it contains approximately 1-2% phosphoinositides and ~10% other negatively charged phospholipids (*455*). This is somehow expected, as the binding of the PBD to PIP_2 in the plasma membrane is a requirement for TRPV4 to sense heat and osmotic stimuli (*128*). Binding to other anionic lipids might also have a functional role by increasing the residency time of the PBD at the plasma membrane. This would enhance the probability of the PBD to encounter a PIP_2 molecule and thereby indirectly contribute to the PIP_2-dependent TRPV4 sensitization. Mutating the PBD to AAWAA might therefore abolish not only the interaction with PIP_2 but also the interaction with other anionic phospholipids in the plasma membrane. Consequently, the PBD loses contact to the plasma membrane, and TRPV4 remains in a constitutively desensitized state.

3.4 Lipid binding extends throughout the TRPV4 IDR

The fact that the isolated PBD in the context of the PH-PRR peptide does not distinguish between different anionic lipids raises the question whether additional residues in the IDR or in the ARD are needed to enable a specific interaction of TRPV4 with PIP_2. Without such a PIP_2-specific election mechanism, it is difficult to imagine why temperature and osmosensation should be a PIP_2-specific mechanism as had been described in the literature (*128*). Indeed, the ability of other lipids interacting with the same lipid-binding site (and potentially in the same or highly similar binding orientation), should enable them to also act as channel sensitizers.

Inspecting the TRPV4 IDR sequence does reveal several regions other than the PBD with properties suited for lipid binding, such as the stretch of basic and hydrophobic residues in the centre of the IDR (res. 61-75 in *H. sapiens* TRPV4, res. 50-65 in *G. gallus* TRPV4) (Figure 71 A, Figure 15 D). Additionally, patches of charged residues that may interact with lipids can also be identified on the surface of the ARD (Figure 71 B). In fact, a PIP_2 binding site in the ARD was previously proposed in a functional and structural study (*150*). However, as outlined in section VI1.3, the identified PIP_2 binding site is likely an artefact due to the experimental setup of that study where the PIP_2 head group was found to interact with the ARD surface that faces the cytosol and not towards the membrane in the context of the full-length channel. Nevertheless, basic patches also exist in the ARD surface facing the membrane and might thus contribute to PIP_2 binding, either alone or together with IDR residues. Importantly, the basic and hydrophobic patches in the IDR and on the ARD surfaces might also act as interacting sites for the cell membrane in general and thus ultimately influence channel structure and stability.

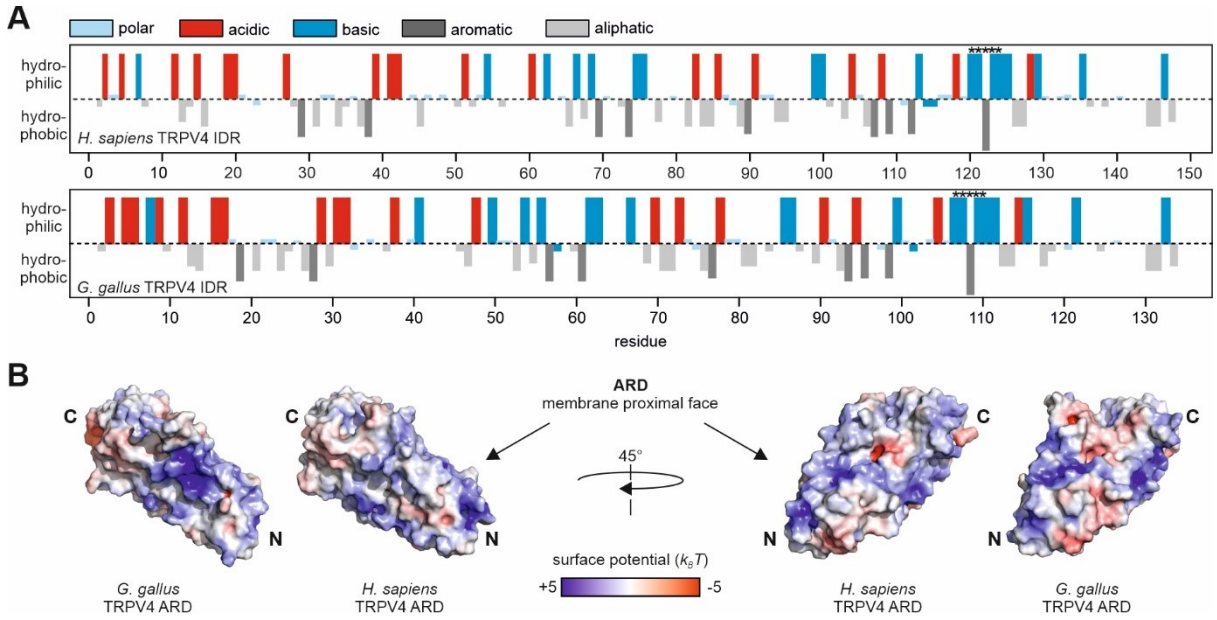


Figure 71: Amino acid property distribution in the TRPV4 NTD. (**A**) The *H. sapiens* and *G. gallus* TRPV4 IDR contains several regions outside of the PBD (indicated with asterisks) with hydrophobic and charged patches, which are potentially suited for lipid interactions (e.g. res. 61-75). (**B**) Surface potential representation from +5 k_BT (blue) to -5 k_BT (red) of the *H. sapiens* TRPV4 ARD (PDB code 4dx2) and the *G. gallus* TRPV4 ARD (PDB: 3w9g) generated with the APBS-tool (*456*) in PyMOL (*322, 323, 453*). Several basic patches facing towards the membrane might indicate potential lipid-binding sites in the ARD.

To assess whether regions in the TRPV4 NTD beyond the PBD can contribute to lipid binding in general, the *G. gallus* TRPV4 constructs NTD, NTD ΔN54, NTD ΔN97, NTD ΔN104, NTD ΔN120, ARD, and NTD^{AAWAA} were subjected to a liposome co-sedimentation assay (Figure 72 A). For this, liposomes containing a mixture of POPG:POPC 1:1 were incubated with TRPV4 N-terminal constructs and subsequently sedimented via ultra-centrifugation. Non-interacting proteins remained in the centrifugation supernatant. Co-localization of the proteins with liposomes was verified via SDS-PAGE analysis (Figure 72 B) and quantified densitometrically via imageJ (Figure 72 C, and section VI2.11). Liposome free samples (protein only) were used as controls to ensure that protein-lipid co-sedimentation is not an artefact caused by protein aggregation. The control experiments demonstrate that all constructs are stable under the experimental conditions (Figure 72 B and C). Therefore, protein co-sedimentation with lipids reflects on protein-lipid interactions.

For the wt-NTD, ~75% of the protein was found to co-sediment with liposomes. Truncating the IDR from the N-terminus increased the fraction of co-sedimented protein to 90-100% in NTD ΔN54, NTD ΔN97, and NTD ΔN104. Further deletion of N-terminal residues including the PBD – as is the case for NTD ΔN120 and the isolated ARD – completely abolished co-sedimentation with lipids. In contrast, co-sedimentation was not lost completely but reduced to ~20% for the NTD^{AAWAA} mutant. These results strongly suggest that the PBD is the major binding site for lipids. The ARD itself does not seem to interact with negatively charged liposomes. The fact that the NTD^{AAWAA} mutants retained a certain ability to sediment with lipids argues for the presence of secondary lipid-binding sites with lower affinity in IDR regions preceding the PBD. Interestingly, deletion of the first 54 residues, a region enriched in negatively charged residues (Figure 73 F), seems to strengthen the liposome interaction of the NTD. It is possible that negatively charged residues compete with the negatively charged liposome surface for binding to the positively charged PBD or other lipid-binding sites in the IDR.

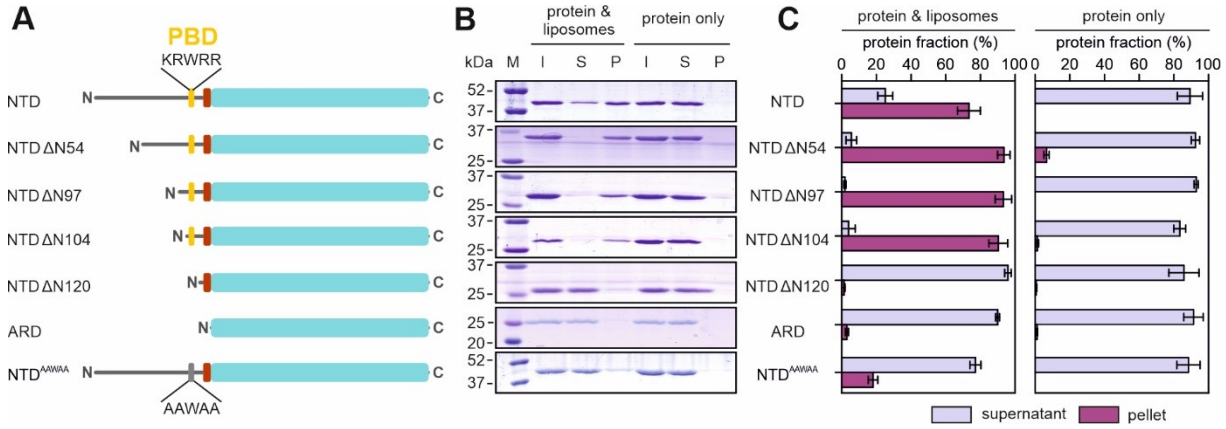


Figure 72: Identification of lipid-binding sites in the *G. gallus* TRPV4 N-terminus via liposome co-sedimentation. (**A**) Topology representation of TRPV4 N-terminal constructs. Cyan: ARD, grey: IDR, red: PRR, yellow: PBD. The mutated PBD in NTD^{AAWAA} is indicated by a grey box. (**B**) SDS-PAGE analysis of liposome co-sedimentation of constructs shown in (A). I: input; S: supernatant; P: Pellet. A lipid mixture of 50% (w/w) POPG and 50% (w/w) POPC was used to prepare liposomes with 100 nm diameter (seeVI2.10). Control experiments were carried out in the absence of liposomes (protein only). (**C**) Protein distribution between pellet or supernatant after centrifugation. The protein fraction in the pellet and supernatant was quantified via densitometry of the SDS-PAGE protein bands in (B) using imageJ (*335*) (see section VI2.11). The error bars in (C) represent the standard deviation from the mean value of three technical replications.

A competition between acidic patches in the IDR and negatively charged lipids for binding to the PBD could reduce the PBD's lipid binding affinity. To test this, tryptophan fluorescence measurements were carried out with the NTD ΔN54, NTD ΔN97 and NTD ΔN104 mutants in comparison to the full-length NTD (Figure 73). In the presence of liposomes, all deletion mutants show similar wavelength shifts and fluorescence wavelengths as the wt-NTD, indicating similar insertion depths of W109 into the lipid bilayer (Figure 73 A, B, and C). The dissociation constants of the mutants determined from the wavelength shifts plotted versus the effective lipid concentration are slightly decreased compared to the wt-NTD (Figure 73 D and E). The lipid-affinity seems to increase in the order NTD < NTD ΔN54 < NTD ΔN97 and slightly decreases again for NTD ΔN104. The increased affinity upon deleting the first 54 amino acids agrees well with the liposome co-sedimentation assay (Figure 72) and supports the assumption that acidic residues in the deleted region might electrostatically interact with the positively charged PBD in the context of the NTD and thus counteract lipid binding. Accordingly, loss of this region through mutation, therefore, renders the PBD more accessible for lipid binding and increases the binding affinity. Deletion of the sequence between residues 54-97 similarly increases the lipid-binding affinity in the PBD. This region contains a mix of basic and acidic patches with the ability for electrostatic interactions. Further truncation of the N-terminus to NTD ΔN104 increases the dissociation constant again. The loss of affinity might be explained by the removal of basic and hydrophobic residues close to the PBD. These residues have been shown via NMR spectroscopy to sense lipid binding in the PH-PRR peptide (Figure 66, Figure 67). Notably, an increase in binding affinity seems to correlate with the fluorescence wavelength in the lipid-bound state. For instance, the NTD ΔN97 mutant not only has a higher lipid affinity than the wt-NTD, but also displays a more blue-shifted fluorescence wavelength which indicates a more hydrophobic environment. This might be explained by a deeper insertion of W109 into the lipid bilayer in the context of the NTD ΔN97 compared to the wt-NTD. Interestingly, deletion of the 97 N-terminal residues seems to reduce the lipid-affinity in the presence of the ARD (NTD ΔN97 mutant), whereas the same mutation had no effect on the binding lipid affinity in the context of the isolated IDR (PH-PRR peptide). This suggests that although it does not bind lipids itself, the ARD influences lipid binding to the PBD (compare Figure 72).

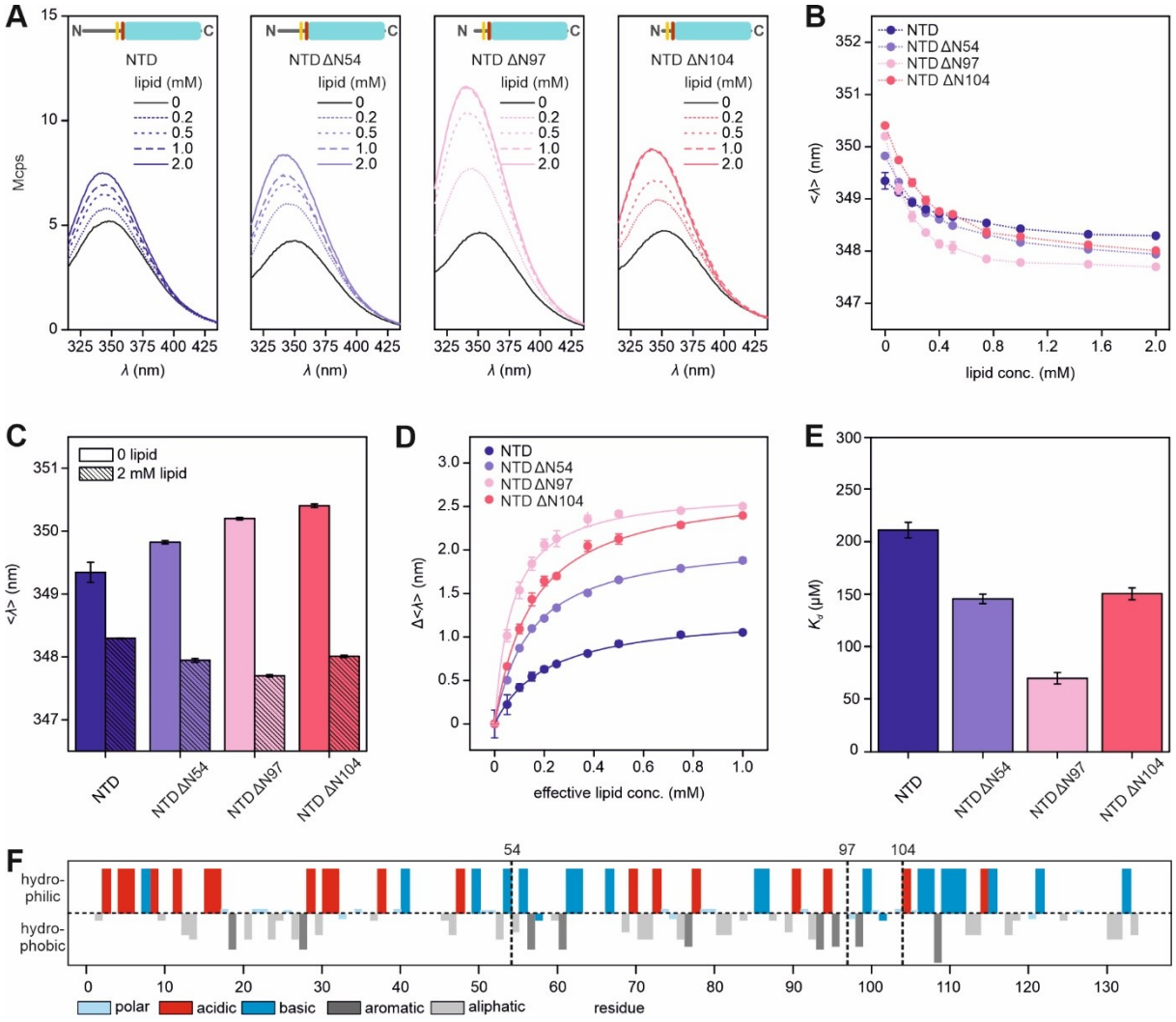


Figure 73: Effect of N-terminal truncations on the lipid interaction of the TRPV4 NTD monitored via tryptophan fluorescence. Tryptophan fluorescence spectra of the *G. gallus* TRPV4 wt-NTD and mutants NTD ΔN54, NTD ΔN97, NTD ΔN104, in the presence of 0/0.2/0.5/1.0/2.0 mM lipid (100 nm liposomes containing POPG-POPC 1:1). (**B**) Tryptophan fluorescence maximum wavelength <λ> of spectra in (A) plotted against the effective lipid concentration. (**C**) Tryptophan fluorescence maximum wavelength <λ> of TRPV4 NTD constructs in apo state (0 mM lipid) and in the presence of liposomes (2 mM lipid). (**D**) Fluorescence wavelength shift Δ<λ> plotted against the effective lipid concentration. The data were fitted with a one-site binding isotherm yielding K_d = 146.3 ± 4.9 μM for NTD ΔN54, K_d = 69.7 ± 5.9 μM for NTD ΔN97, K_d = 151.2 ± 4.6 μM for NTD ΔN104. (**E**) Comparison of determined K_d values in (D) with K_d of wt-NTD determined in Figure 62. (**F**) Amino acid properties plotted versus the residue number. Dashed lines indicate the positions where the IDR was truncated in the NTD deletion mutants. The error bars in (B), (C), and (D) represent the standard deviation from the mean value of three technical replications. The error bars in (E) are the fit errors from (D).

To identify and characterize secondary lipid-binding sites in the IDR in more detail, ^{1}H, ^{15}N-HSQC-based CSP experiments of ^{15}N-IDR and ^{15}N-IDR^{AAWAA} with diC$_8$-PIP$_2$ were carried out. NMR backbone resonance assignment of the IDR^{AAWAA} mutant could be largely obtained from transferring assignments of ^{15}N, ^{13}C-IDR (Figure 45). Peaks corresponding to residues in and around the mutated PBD were strongly shifted between IDR and IDR^{AAWAA} and could not be transferred. However, the peak displacement was similar as observed between the PH-PRR and the PH^{AAWAA}-PRR peptide. The missing assignments in the mutated PBD of IDR^{AAWAA} could therefore be completed via comparison with PH^{AAWAA}-PRR assignments and through procedure of exclusion (Figure 74 A). Prior to the NMR based CSP experiments, tryptophan fluorescence spectroscopy was used to verify that the lipid-binding behaviour of the IDR and the IDR^{AAWAA} mutant is not compromised by the low pH required for NMR experiments (Figure 75). In agreement with the tryptophan fluorescence data at pH 7, the wt-IDR showed strong fluorescence changes when titrated with POPG:POPC 1:1 liposomes at pH 4.5. The determined dissociation constant of K_d = 222.2 ± 20.2 μM is close to the value at pH 7 (see Figure

62). No fluorescence changes were observed upon titration of POPC only liposomes. Also, the fluorescence of the IDR^{AAWAA} mutant is nearly unaffected upon addition of POPG POPC 1:1 or POPC only liposomes.

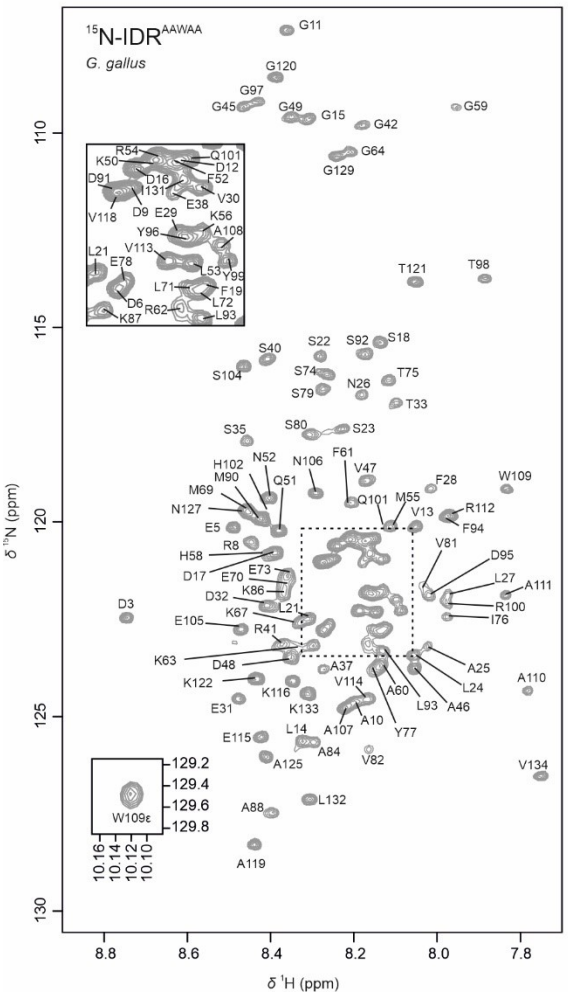


Figure 74: NMR backbone assignment of the *G. gallus* TRPV4 IDR^{AAWAA} mutant. (**A**) ^{1}H, ^{15}N-TROSY NMR spectrum of 60 µM ^{15}N-IDR^{AAWAA}. Assignments could largely be transferred from the wt-IDR assignment (Figure 45 A). Remaining assignments of peaks corresponding to mutated residues in the PBD could be transferred from the PH^{AAWAA}-PRR assignments (Figure 64). Spectra were recorded at 298 K in 20 mM NaP_i pH 4.5, 150 mM NaCl, 1 mM DTT, 10% D_2O. An expanded view of the crowded central region (dashed box) is shown in the bottom right corner.

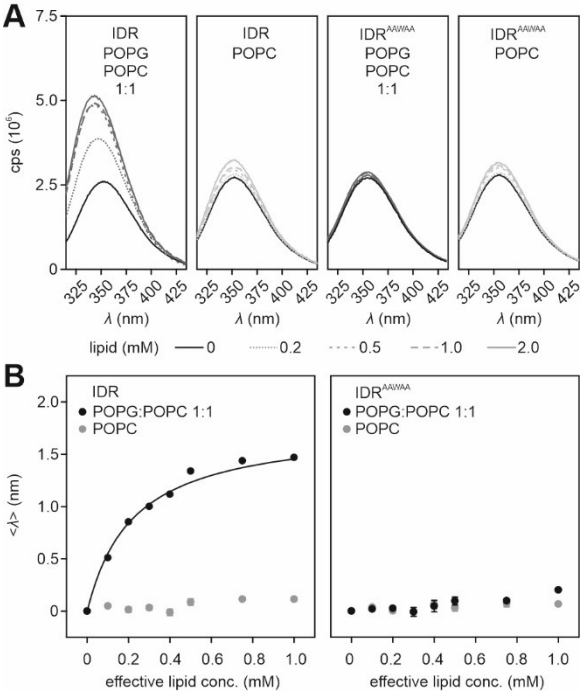


Figure 75: Influence of low pH on the lipid interaction of the *G. gallus* TRPV4 IDR and IDR^{AAWAA} mutant. (**A**) Tryptophan fluorescence spectra of IDR and IDR^{AAWAA} at pH 4.5 in the presence of 0/0.2/0.5/1.0/2.0 mM lipid. Lipids were used as liposomes with 100 nm diameter containing either POPG:POPC 1:1 or POPC only. Plotted is the fluorescence intensity as counts per second (cps) versus the wavelength. (**B**) Fluorescence wavelength shifts $<\lambda>$ obtained from spectra in (A) plotted versus the effective lipid concentration. The wavelength shifts of the IDR upon titration of POPG:POPC 1:1 liposomes could be fitted with a binding isotherm yielding a dissociation constant of K_d = 222.2 ± 20.2 µM. IDR in the presence of POPC liposomes and the IDR^{AAWAA} in the presence of POPG:POPC 1:1 and POPC only liposomes show no binding behaviour. This is in good agreement with the binding behaviour at pH 7 (see Figure 45). The error bars in (B) represent the standard deviation from the mean value of three technical replications.

Titration of diC_8-PIP_2 caused peak shifts throughout the ^{1}H, ^{15}N-TROSY NMR spectrum of ^{15}N-IDR (Figure 76 A). Unfortunately, severe line broadening paired with strong signal overlap prevented the chemical shift determination of various IDR and IDR^{AAWAA} resonances at high lipid concentrations. The identification of lipid-binding sites solely based on the chemical shift changes upon lipid titration was therefore impossible. Instead, a qualitative analysis was carried out on the assumption that IDR resonances, which experience strong line broadening, are direct lipid-binding sites. In contrast, shifting resonances not accompanied by line broadening were assumed to reflect structural changes in the corresponding region but not direct lipid interactions. This assumption is based on observations from CSP experiments with diC_8-PIP_2 using ^{15}N-PH-PRR peptide as the reporter(Figure 67). In this experiment, all resonances displaying strong line broadening upon diC_8-PIP_2

titration also showed the most significant chemical shift changes, which can be attributed to the interaction with PIP_2 molecules. To visualize where lipid-binding occurs in the IDR sequence, the chemical shift difference (ΔCS) between the apo and lipid-bound state of each IDR resonance was plotted versus the residue sequence (Figure 76 D). Regions for which no ΔCS could be determined due to line broadening are indicated in the plot with a maximum ΔCS value (0.45 ppm) and a grey box. The maximum ΔCS value was set close to the highest PIP_2-induced ΔCS in the PH-PRR peptide (Figure 76 A). The ΔCS pattern reveals that diC_8-PIP_2 binding is sensed by ~2/3 of the residues within the IDR and implies the presence of secondary lipid-binding sites outside of the PBD. Strong ΔCS and line broadening are observed throughout a 70-residue region comprising residues K50-V118. This region contains patches enriched in positively charged and hydrophobic residues, i.e. K50-K67 and Y99-K116, which are ideally suited for lipid interaction (Figure 76 E). Interestingly, resonances of the sequence ^{25}ANLFEV^{30,} which also show severe line broadening upon diC_8-PIP_2 titration, even this region is completely devoid of positively charged residues (Figure 76 D and E). It might be possible that line broadening in this region is not due to direct interaction with lipids, but because of association with other lipid-binding regions in the IDR. Such an interaction might be mediated by electrostatic and hydrophobic interactions between the acidic/hydrophobic ^{25}ANLFEV^{30} sequence with a basic/hydrophobic lipid-binding site. Two candidates for interaction partners with ^{25}ANLFEV^{30} might be residues K50-K67 or residues Y99-K116, which encompass the PBD. Overall, the ΔCS pattern suggests that diC_8-PIP_2 binding is sensed almost by the entire IDR, either through direct interaction or in the form of structural rearrangements.

Strikingly, the ΔCS pattern of the wt-IDR is preserved in the IDR^{AAWAA} mutant, including the region comprising the PBD (Figure 76 B). Upon diC_8-PIP_2 titration to the ^{15}N-IDR^{AAWAA}, the resonances corresponding to PBD residues are similarly broadened as for the wt-IDR, thus suggesting that diC_8-PIP_2 lipids can still interact with the mutated PBD in the context of the IDR. This is somewhat surprising, considering that tryptophan fluorescence demonstrated that the ^{107}AAWAA^{111} mutation abolishes binding of anionic lipids to the PBD in the of the context of the NTD, IDR, and PH-PRR constructs (Figure 62). Also, diC_8-PIP_2 binding to the PBD was almost completely suppressed in the PH^{AAWAA}-PRR mutant in CSP experiments (Figure 68). The shifts and line broadening of PBD resonances in the IDR^{AAWAA} mutant are delayed to higher lipid concentrations compared to the wt-IDR, which suggests a reduced lipid binding affinity of the mutated PBD. To compare the lipid-binding affinity of the PBD between the wt-IDR and the IDR^{AAWAA} mutant, the W109 Hε chemical shift changes of the IDR and the IDR^{AAWAA} were plotted versus the diC_8-PIP_2 concentrations and fitted with a Hill-Langmuir isotherm (Figure 76 C). The determined dissociation constant of K_d = 416.9 ± 22.7 μM with n = 2.4 ± 0.3 for the wt-IDR is similar to the values determined for the PH-PRR (K_d = 427.4 ± 17.8 μM with n = 2.9 ± 0.3 for PH-PRR) (Table 8). In contrast, the dissociation constant determined for the IDR^{AAWAA} mutant is reduced to K_d = 980.9 ± 150.8 μM with n = 2.3 ± 0.3 (Table 8). This confirms that lipid binding to the PBD is reduced in the IDR^{AAWAA} mutant, but unlike in the PH^{AAWAA}-PRR mutant, it is not lost completely. A similar reduction of the K_d value can be determined from the resonances of residues close to the PBD such as K122 (Table 8). Notably, the resonances of residues distant from the PBD, such as S35 and G41, respond almost equally to the presence of diC_8-PIP_2 in the wt-IDR and the IDR^{AAWAA} mutant (Table 8, Figure 76 C). This suggests that the ^{107}AAWAA^{111} mutation disturbs lipid binding to the PBD but has no effect on the affinity of the secondary lipid-binding sites in the IDR. The reduced but not abolished ability of the IDR^{AAWAA} mutant to bind lipids in the PBD might be explained by proximal secondary binding sites (see Figure 76 D). In the context of the IDR, the proximal secondary lipid-binding sites increase the lipid concentration in the environment of the mutated PBD and increase its effective affinity. In contrast, the secondary lipid-binding sites in regions preceding the PBD are missing in the PH-PRR peptide, and the affinity for lipids is consequently lost when introducing the ^{107}AAWAA^{111} mutation. Unfortunately, the severe line broadening of most resonances prevented a more detailed analysis of how lipid-binding affinities are distributed throughout the IDR sequence. Therefore, a hierarchy for lipid binding along the IDR sequence cannot be concluded from CSP experiments. Liposome co-sedimentation, however, suggests that the PBD is the primary binding site for the binding of anionic lipids (see Figure 72).

Table 8: Dissociation constants (K_d) and Hill-coefficients (n) determined from resonance shifts of the PH-PRR peptide as well as the wildtype IDR and IDR^{AAWAA} mutant upon diC$_8$-PIP$_2$ titration.

Construct	Resonance	K_d (µM)	n
PH-PRR	W109 Hε	427.4 ± 17.8	2.9 ± 0.3
IDR	W109 Hε	416.9 ± 22.7	2.4 ± 0.3
IDR^{AAWAA}	W109 Hε	980.9 ± 150.8	2.3 ± 0.3
IDR	K122 NH	359.2 ± 58.9	1.8 ± 0.1
IDR^{AAWAA}	K122 NH	621.4 ± 44.7	2.6 ± 0.3
IDR	G41 NH	824.0 ± 85.9	2.0 ± 0.1
IDR^{AAWAA}	G41 NH	767.6 ± 101.8	2.0 ± 0.3
IDR	S35 NH	615.0 ± 42.9	2.0 ± 0.1
IDR^{AAWAA}	S35 NH	606.2 ± 32.3	3.0 ± 0.3

As shown in Figure 15, the amino acid properties in the TRPV4 IDR are highly conserved across species. To test whether this reflects in a conserved lipid-binding behavior in regions beyond the PBD, CSP experiments with diC$_8$-PIP$_2$ were carried out using ^{15}N-IDR from *H. sapiens* TRPV4 as the reporter (Figure 77). Again, peak shifts were accompanied by severe line broadening, which prevented a consistent chemical shift analysis (Figure 77 A). In analogy to Figure 76, regions for which no ΔCS could be determined due to line broadening were assigned with a maximum ΔCS value (0.45 ppm) and are indicated with a grey box in the ΔCS versus residue plot. The ΔCS plot displays a similar pattern as observed for the *G. gallus* TRPV4 IDR (Figure 77 D). Line broadening is observed throughout residues R64-Q131 and in the short sequence ^{37}ANLFE^{41}. This agrees well with the regions K50-V118 and ^{25}ANLFEV^{30} in the *G. gallus* TRPV4 IDR. The amino acid properties of the lipid-sensing regions are highly conserved between *G. gallus* and *H. sapiens* TRPV4 (Figure 77 F). To test whether the identified regions are specific for the binding of diC$_8$-PIP$_2$, the CSP experiments were repeated with diC$_8$-PG and liposomes containing POPG:POPC 1:1(Figure 77 B and C). Titration of diC$_8$-PG caused almost identical peak shifts and line broadening as observed upon diC$_8$-PIP$_2$ titration (Figure 77 D). Similar as observed for the PH-PRR peptides (Figure 66), liposome binding manifested in the loss of resonance intensities due to increased correlation times of liposome bound residues. Plotting the peak intensity decrease versus the residue sequence yielded a pattern similar to the ΔCS distribution upon diC$_8$-PIP$_2$ and diC$_8$-PG titration (Figure 77 E). Almost complete signal loss upon liposome binding is observed in the same regions, which show line broadening upon binding of diC$_8$-lipids. This confirms that the binding sites and structural changes upon lipid binding do not depend on specific lipid structures other than negatively charged head groups or the choice of membrane mimic.

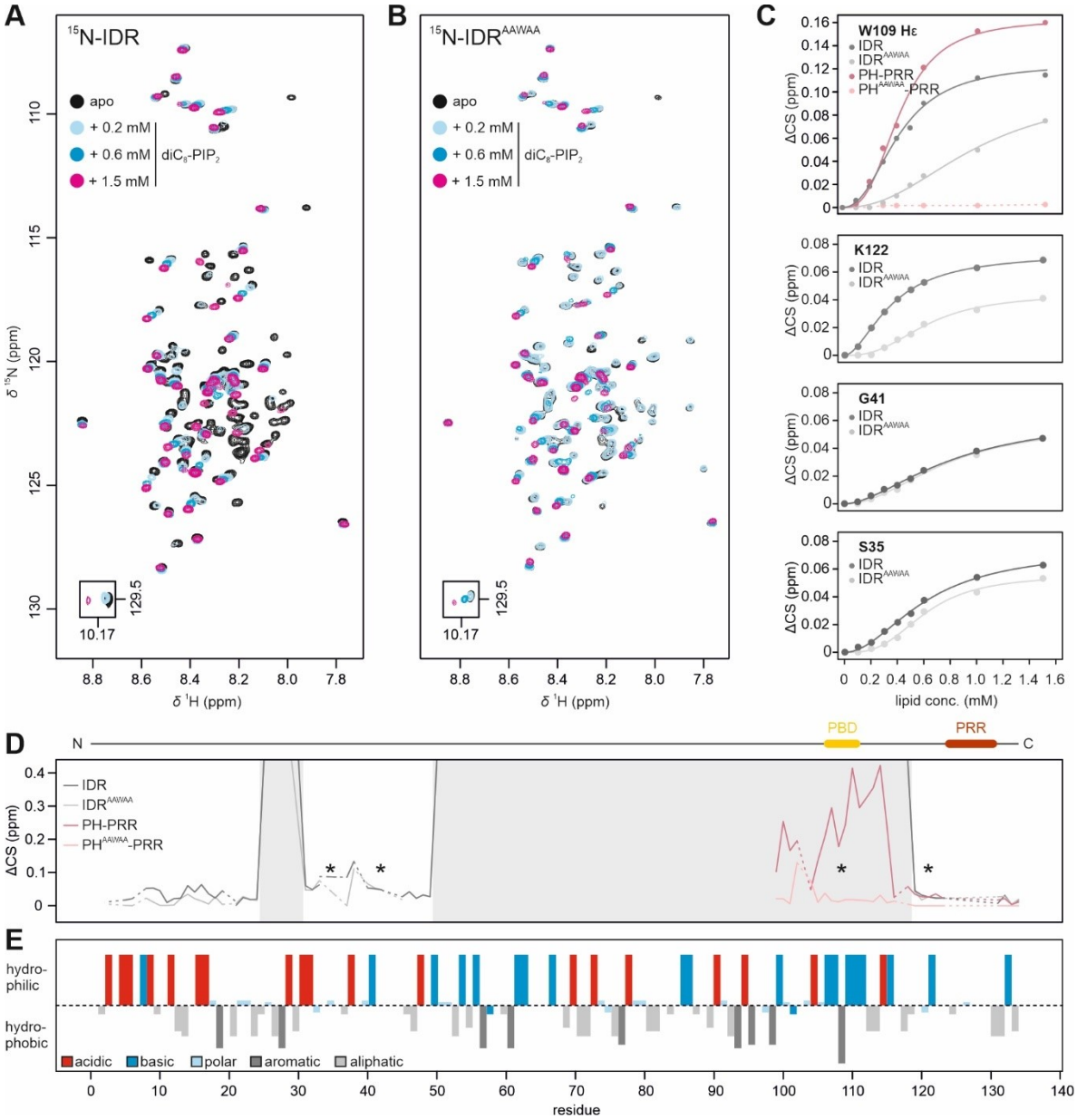


Figure 76: Identification of lipid-binding sites in the *G. gallus* TRPV4 IDR via NMR spectroscopy. ^{1}H, ^{15}N-TROSY NMR spectra of (**A**) 100 µM ^{15}N-IDR and (**B**) 60 µM ^{15}N-IDR^{AAWAA} in the presence of 0/0.2/0.6/1.5 mM diC_8-PIP_2. Spectra were recorded at 298 K in 20 mM NaP_i pH 4.5, 150 mM NaCl, 1 mM DTT, 10% D_2O. (**C**) Chemical shift change (ΔCS) of the W109 Hε resonance of IDR, IDR^{AAWAA}, PH-PRR and PH^{AAWAA}-PRR and chemical shift changes of S35, G4, and K122 resonances of IDR and IDR^{AAWAA} plotted versus the lipid concentration. The ΔCS of W109 Hε were fitted with a Hill-Langmuir isotherm yielding K_d = 427.4 ± 17.8 µM with n = 2.9 ± 0.3 for PH-PRR, K_d = 416.9 ± 22.7 µM with n = 2.4 ± 0.3 for IDR, and K_d = 980.9 ± 150.8 µM with n = 2.3 ± 0.3 for IDR^{AAWAA}. See text for dissociation constants obtained for S35, G41, and K122. (**D**) Chemical shift difference (ΔCS) between the free and lipid-bound state of IDR, IDR^{AAWAA}, PH-PRR and PH^{AAWAA}-PRR plotted versus the residue number. Due to severe line broadening in the spectra of ^{15}N-IDR and ^{15}N-IDR^{AAWAA} the ΔCS of residues 22-28 and 51-119 could not be determined. The missing regions are indicated with a grey box. Severe line broadening is a strong indicator for lipid binding. The asterisks denote the residues plotted in (C). (**E**) Amino acid property distribution in the TRPV4 IDR obtained with the PepCalc tool (*457*).

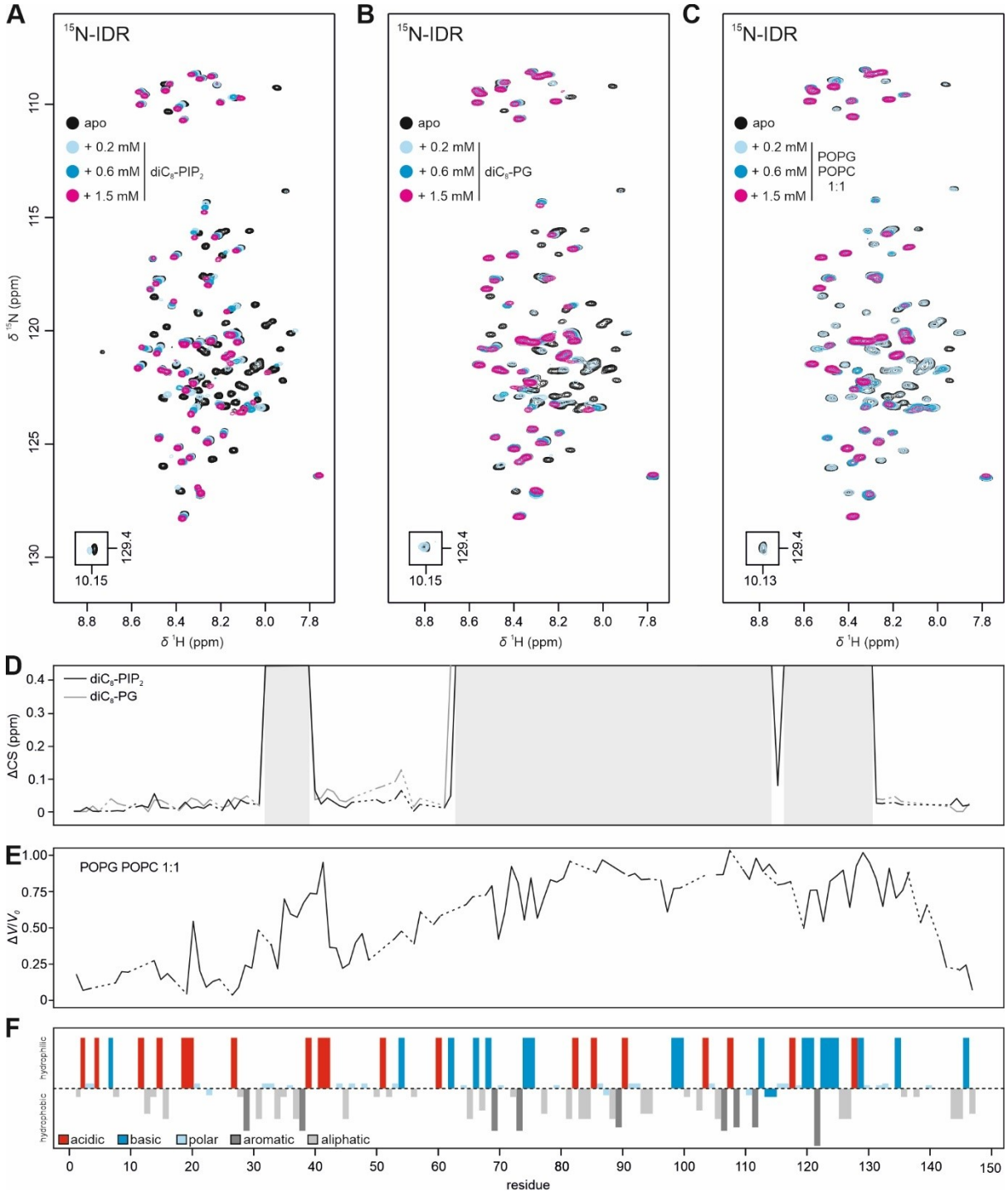


Figure 77: Identification of lipid-binding sites in the *H. sapiens* TRPV4 IDR via NMR spectroscopy. ^{1}H, ^{15}N-TROSY NMR spectra of 100 µM ^{15}N-IDR in the presence of 0/0.2/0.6/1.5 mM (**A**) diC_8-PIP_2, (**B**) diC_8-PG, and (**C**) liposomes containing POPG POPC 1:1. Spectra were recorded at 298 K in 20 mM NaP_i pH 4.5, 150 mM NaCl, 1 mM DTT, 10% D_2O. (**D**) Chemical shift difference (ΔCS) between the free and diC_8-PIP_2 or diC_8-PG bound state of the IDR versus the residue number. Due to severe line broadening in the spectra of ^{15}N-IDR the ΔCS of residues 37-41 and 64-131 could not be determined. The missing regions are indicated with a grey box. Severe line broadening is a strong indicator for lipid binding. (**E**) Resonance intensity decreases between free and liposome bound state of the IDR. The intensity decrease was determined from the peak integral difference ΔV normalized to the peak integral of the free state V_0. (**F**) Amino acid property distribution of in the TRPV4 IDR obtained with the PepCalc tool (*457*).

As a complement to the HSQC-based experiments using the ^{15}N-labeled protein as the reporter, the interaction of PIP_2 with the TRPV4 IDR was monitored from the lipid's perspective via ^{31}P NMR spectroscopy (Figure

78). The lipid diC$_8$-PIP$_2$ gives rise to three peaks in the ^{31}P NMR spectrum, each corresponding to one phosphate moiety in the lipid head group. The assignment of the peaks to the phosphate groups at positions 1, 4, or 5 (termed P1, P4, and P5) was achieved via comparison with the ^{31}P NMR spectra of diC$_8$-PI(3,5)P$_2$ and D-*myo*-I(1,5)P$_2$ (Figure 78 A and B). Using diC$_8$-PIP$_2$ as a reporter, CSP experiments were carried out with PH-PRR, PH^{AAWAA}-PRR, IDR, and IDR^{AAWAA} as ligands (Figure 78 C). An advantage of this approach is the constant lipid concentration. This should maintain a constant monomer-micelle equilibrium of the lipids throughout the titration, thus making the determination of dissociation constants from the quantitative analysis of resonance shifts more feasible. As expected, shifts and line broadening of the PIP$_2$ resonances were observed upon PH-PRR titration, whereas the PIP$_2$ resonances were nearly unaffected upon titrating the PH^{AAWAA}-PRR mutant. Particularly strong shifts were observed for the P4 and P5 resonances. The P1 resonance, in contrast, was less shifted but severely broadened instead. The ΔCS of P4 and P5 plotted against the peptide concentrations could be fitted with binding isotherms yielding K_d = 71.1 ± 38.9 µM for PH-PRR and K_d = 808.1 ± 320.0 µM for PH^{AAWAA}-PRR (Figure 78 D and E). The dissociation constants confirm that the affinity for PIP$_2$ is reduced by a factor of ~10 upon mutating the PBD in the PH-PRR peptide. Titration of the IDR caused similar peak shifts in the same direction as observed during the PH-PRR titration indicating a similar binding mode. The fitted ΔCS data yielded a dissociation constant of K_d = 180.5 ± 72.0 µM (Figure 78 D and E). Within the error of the fit, the dissociation constant is in a similar range as determined for the PH-PRR peptide. Surprisingly, the ^{107}AAWAA^{111} mutation did not reduce PIP$_2$ binding in the context of the IDR. Within the error of the fit, the determined dissociation constant of K_d = 33.6 ± 10.2 µM suggests a similar affinity for PIP$_2$ compared to the wt-IDR or the PH-PRR peptide. However, the P4 and P5 resonances shift in the opposite direction upon titration of IDR^{AAWAA} compared to when the wt-IDR or the PH-PRR peptide were titrated. The inverted direction of resonance shifts implies a different binding mode of the PIP$_2$ head group for the IDR^{AAWAA} mutant versus the wildtype IDR.

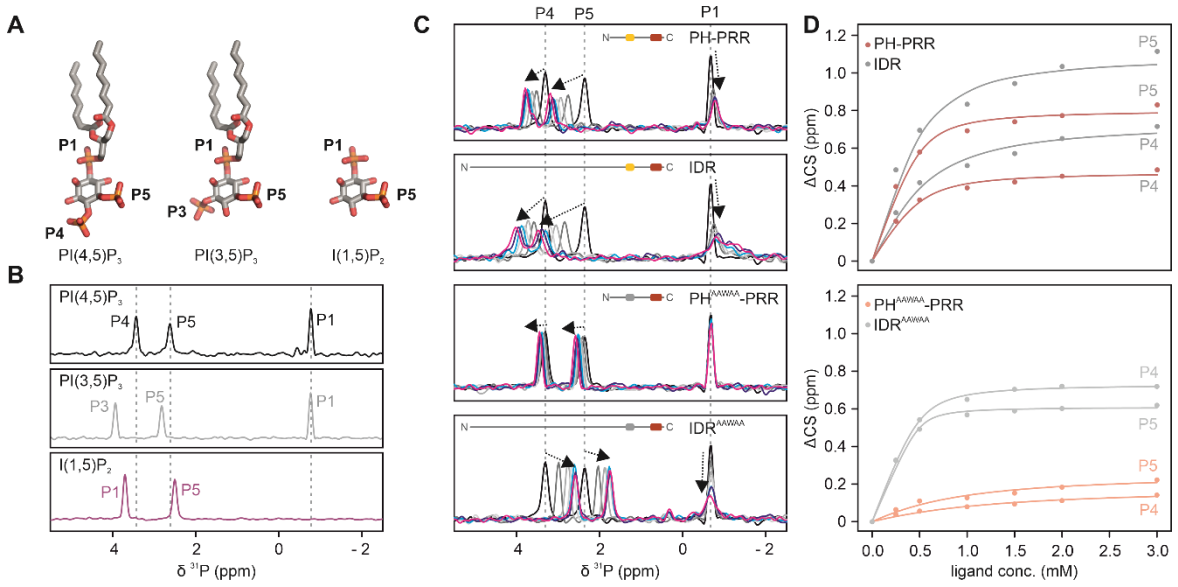


Figure 78: ^{31}P NMR studies of PIP$_2$ interaction with the *G. gallus* TRPV4 IDR. (**A**) Structures and (**B**) ^{31}P NMR spectra of diC$_8$-PI(4,5)P$_2$, diC$_8$-PI(3,5)P$_2$, and D-*myo*-I(1,5)P$_2$ with resonance assignments of phosphate groups (orange). Spectra were recorded at 298 K in 10 mM Tris pH 7, 100 mM NaCl, 10% D_2O at a lipid concentration of 500 µM lipid. (**C**) ^{31}P NMR spectra of diC$_8$-PI(4,5)P2 in the presence of increasing concentrations of PH-PRR, IDR, PH^{AAWAA}-PRR, and IDR^{AAWAA} (0.25/0.50/1.00/1.50/2.00/3.00 mM). (**D**) Chemical shift changes (ΔCS) of P4 and P5 resonances plotted versus the ligand concentration. The ΔCS data were fitted with a binding isotherm yielding K_d = 71.1 ± 38.9 µM for PH-PRR, K_d = 808.1 ± 320.0 µM for PH^{AAWAA}-PRR, K_d = 180.5 ± 72.0 µM for IDR, and K_d = 33.6 ± 10.2 µM for IDR^{AAWAA} (averages from the fits of P4 and P5 resonances).

Together with the HSQC-based CSP experiments (Figure 76), the ^{31}P NMR CSP experiments support the presence of a secondary binding site for diC$_8$-PIP$_2$ in the IDR outside of the PBD, with a different lipid binding

mode but with a similar affinity as the PBD. In contrast to these findings, the NTD^{AAWAA} mutant showed reduced liposome co-sedimentation compared to the wt-NTD, which implies a secondary lipid-binding site with a lower affinity than the PBD (Figure 72). These discrepancies might depend on the different types of lipids used in the experiments. They might, however, also indicate that the ARD influences how lipids bind to the IDR.

The observation that almost the entire IDR binds or senses lipids raises the question of whether lipid-binding outside of the PBD is accompanied by secondary structure formation. This is not possible to determine via NMR spectroscopy as resonances of residues in the lipid-binding sites are severely broadened and thus cannot be detected in the lipid-bound state of the IDR (Figure 76). In the context of the PH-PRR peptide, lipid binding to the PBD did not induce secondary structure formation (Figure 63). To test whether lipid-binding induces secondary structures outside of the PBD, CD spectra of the *G. gallus* and *H. sapiens* TRPV4 NTD and IDR were measured in the absence and presence of POPG or POPC liposomes (Figure 63 A, B, C, and D). No spectral changes are observed upon addition of POPC liposomes. Also, the addition of POPG liposomes does not cause any significant changes in the CD spectra to indicate changes in the secondary structure. Considering the K_d values determined for the binding of POPG in the low μM range (Figure 61), the NTD and the IDR are both expected to be entirely bound to POPG liposomes at 2 mM lipid. The absence of changes in the CD spectra, therefore, suggests that the IDR remains entirely disordered upon binding to lipids.

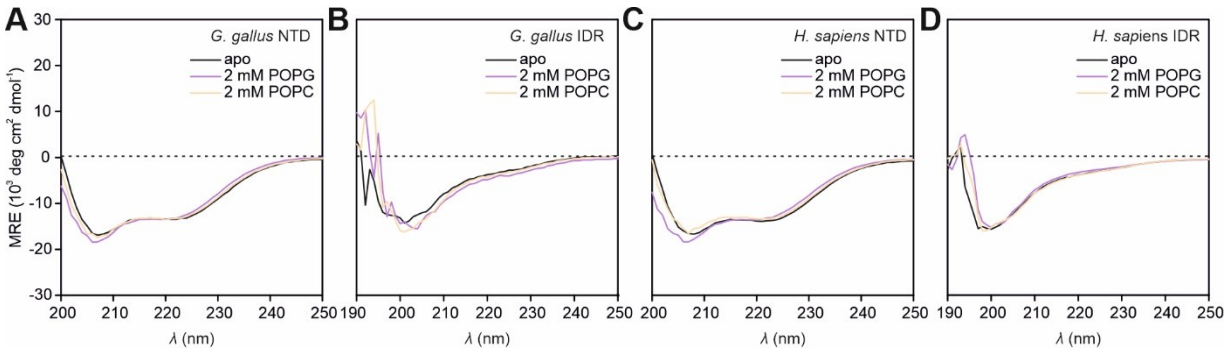


Figure 79: CD spectroscopic secondary structure analysis of the TRPV4 N-terminus upon lipid binding. Far-UV CD spectra of (**A**) *G. gallus* TRPV4 NTD, (**B**) *G. gallus* TRPV4 IDR, (**C**) *H. sapiens* TRPV4 NTD, and (**D**) *H. sapiens* TRPV4 IDR in the apo state (black lines), and in the presence of 2 mM POPG (magenta lines) or 2 mM POPC (orange lines). Plotted is the mean residue ellipticity MRE versus the wavelength λ. Lipids were prepared as liposomes with diameters of 50 nm (see section VI2.10).

The liposome co-sedimentation and tryptophan fluorescence measurements (Figure 72, Figure 73) indicate that intramolecular interactions of residues within the IDR may compete with lipid binding. Plausible candidates for competitive interactions are regions enriched in acidic residues, e.g. the N-terminal residues 1-49, which might interact with positively charged lipid-binding sites such as the PBD via electrostatic interaction (Figure 76 F). Accordingly, mutations in the PBD would be expected to influence not only lipid interactions but also such long-range intramolecular interaction partners within the IDR. Figure 80 A shows the comparison of the chemical shifts (ΔCS) between the IDR and IDR^{AAWAA} mutant. The ^{107}AAWAA^{111} mutation primarily affects resonances for residues close to the site of the mutation, i.e. within the PBD (Figure 80 C). The ΔCS pattern in the region comprising residues 97-134 is almost identical as the ΔCS between wt-PH-PRR wildtype and the PH^{AAWAA}-PRR mutant. However, several residues distant from the PBD show subtle peak shifts between IDR and IDR^{AAWAA}. The most significant ΔCS values cluster in a hydrophobic stretch comprising residues S22-F28 (Figure 80 C and E). Interestingly, this region overlaps with the sequence ^{25}ANLFEV^{30}, which showed severe line broadening in the CSP experiments with diC$_8$-PIP$_2$ even though it does not contain amino acids favouring lipid binding (Figure 76). The ΔCS of S22-F28 might originate from increased interactions with residues in the mutated PBD or from a loss of interaction with the wt-PBD, respectively. The rather subtle ΔCS values might reflect that only subsets of the conformational IDR ensemble are affected by the PBD mutation. In agreement

with this, the fast backbone dynamics are almost unaffected by the ^{107}AAWAA^{111} mutation as evidenced by {^{1}H}, ^{15}N-hetNOE-measurements (Figure 80 D).

To validate whether the observed ΔCS values for S22-F28 between IDR and IDR^{AAWAA} are due to reduced or strengthened intra-chain interactions, the ΔCS between the IDR and the PH-PRR peptide was compared to the ΔCS between the IDR^{AAWAA} and the PH^{AAWAA}-PRR mutant (Figure 80 B). In the PH-PRR peptide, Y99 is the N-terminal amino acid, whereas a 98-residue peptide chain extends upstream of Y99 in the IDR. Therefore, as expected, large ΔCS values are observed for residues Y99-S104 and can be attributed to the fact that these residues have a very different environment in the PH-PRR peptide compared to the IDR. This effect should be similar for wt-constructs and ^{107}AAWAA^{111} mutants. However, the ΔCS around the PBD is significantly larger between the wildtype IDR and the wt-PH-PRR peptide than between the IDR^{AAWAA} and the PH^{AAWAA}-PRR mutant. This strongly suggests that the PBD interacts with upstream regions of the IDR and these interactions are lost upon removing the first 98 residues, an effect that is subdued by the introduction of the ^{107}AAWAA^{111} mutation.

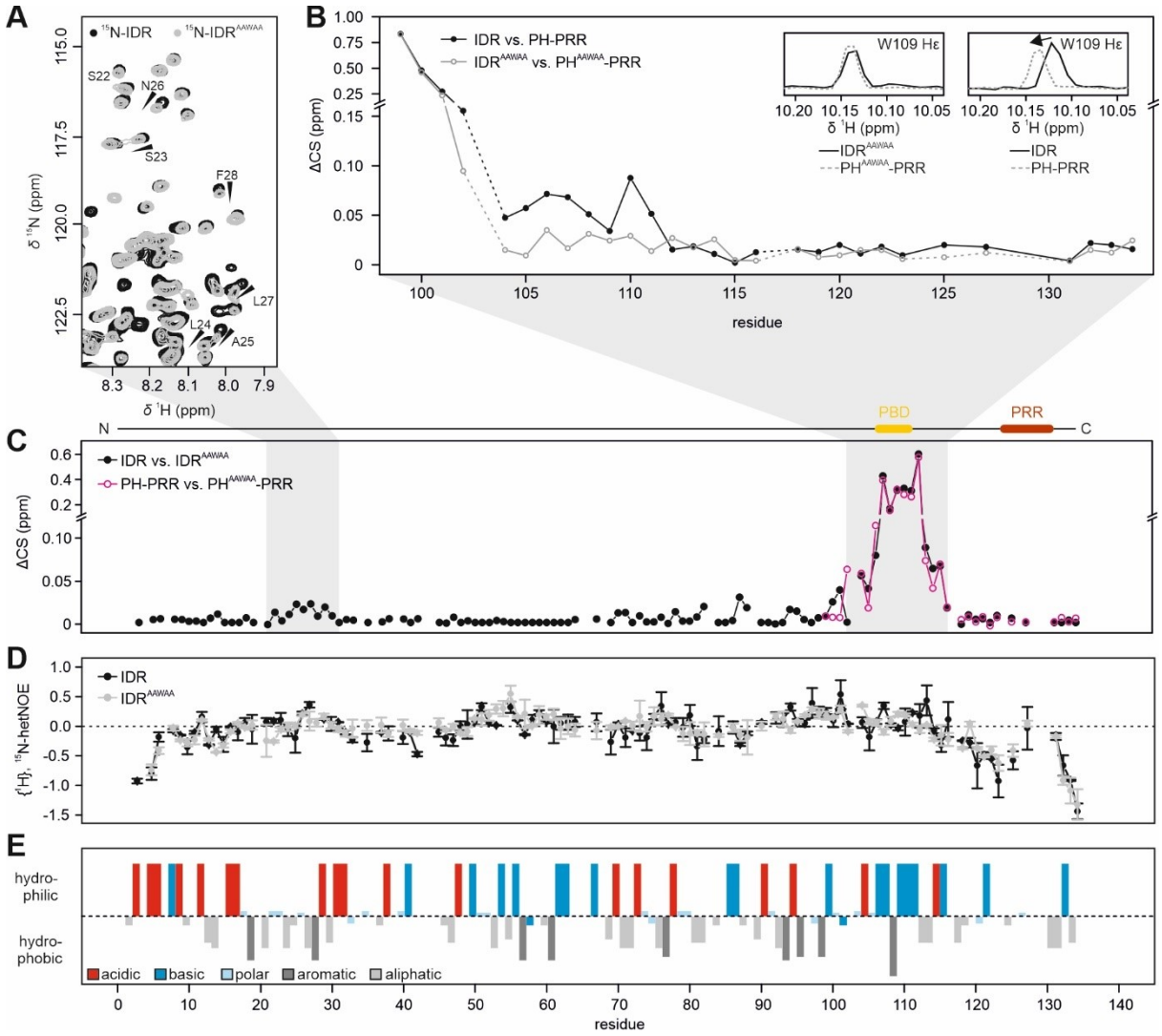


Figure 80: Influence of AAWAA mutation on the structure and dynamics of the *G. gallus* TRPV4 IDR. (**A**) Overlay of ^{1}H, ^{15}N-TOCSY NMR spectra of 100 µM ^{15}N-IDR and 50 µM ^{15}N-IDR^{AAWAA}. (**B**) Chemical shift difference (ΔCS) between IDR and PH-PRR (black) and between IDR^{AAWAA} and PH^{AAWAA}-PRR (grey) plotted versus the residue number. Inset: Overlays of W109 Hε resonances of IDR and PH-PRR (right panel) and IDR^{AAWAA} and PH^{AAWAA}-PRR (left panel). (**C**) Chemical shift difference (ΔCS) between IDR and IDR^{AAWAA} (black) and between PH-PRR and PH^{AAWAA}-PRR (magenta) plotted versus the residue number. (**D**) {^{1}H}, ^{15}N-hetNOE of IDR (lack) and IDR^{AAWAA} (grey) plotted versus the residue number. (**E**) Amino acid properties plotted against the IDR sequence, calculated with the PepCalc tool (*457*).

Loss of intra-chain interactions between the PBD and other IDR regions should also be reflected in wavelength shifts of the W109 fluorescence upon PBD mutation. To verify this, tryptophan fluorescence spectra of the NTD, IDR, PH-PRR wildtype and ^{107}AAWAA^{111} mutants were recorded, and their fluorescence wavelength was determined (Figure 81). In the context of the NTD and the PH-PRR peptide, the shift between the wildtype and the ^{107}AAWAA^{111} mutant is in a similar range of ~0.2-0.4 nm (Figure 81 A, B). In the context of the IDR, however, the fluorescence maximum of W109 is shifted by ~1.3 nm (Figure 81 A, B). Similarly, the wavelength shift between IDR and PH-PRR peptide is increased in the context of the wildtype PBD compared to the AAWAA mutant (Figure 81 B). This agrees well with the ΔCS of the W109 Hε resonance between IDR and PH-PRR in the wt- and mutant proteins (Figure 80 B).

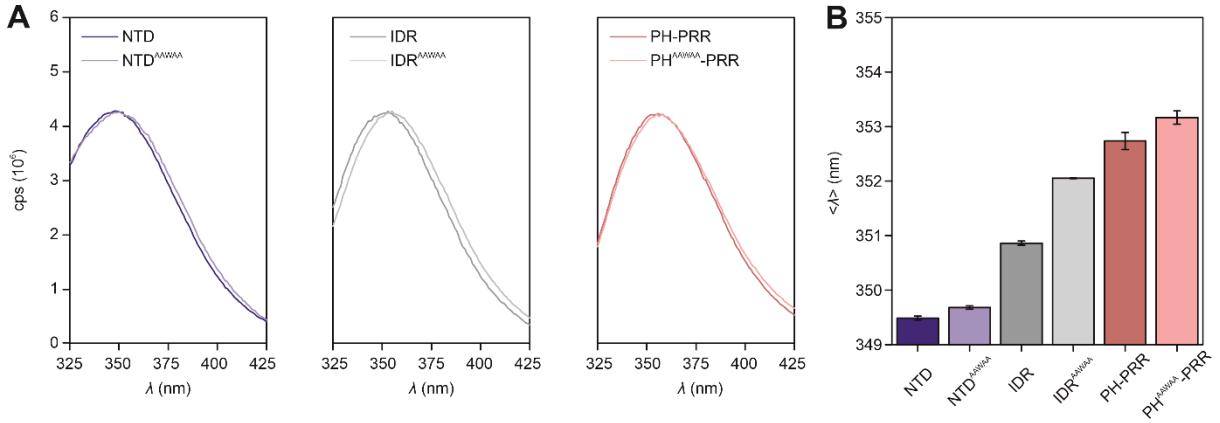


Figure 81: Tryptophan fluorescence spectroscopic analysis of PBD mutation in the *G. gallus* TRPV4 N-terminal constructs. (**A**) Tryptophan fluorescence spectra of NTD and NTD^{AAWAA} (left panel), IDR and IDR^{AAWAA} (middle panel), and PH-PRR and PH^{AAWAA}-PRR (right panel). Plotted is the fluorescence as counts per second (cps) versus the wavelength λ. (**B**) Wavelength of the intensity-weighted tryptophan fluorescence wavelength $<\lambda>$ of wildtype and PBD mutant TRPV4 N-terminal constructs. The error bars in (B) represent the standard deviation from the mean value of three technical replications.

The small effect of the ^{107}AAWAA^{111} mutation on the fluorescence wavelength of the NTD compared to the IDR is somehow surprising. Whereas mutation of the PBD renders W109 more solvent-exposed in the IDR, it remains similarly buried in the context of the NTD. In contrast, the PBD mutation increased the Stokes radius of the NTD by ~3 nm, but the Stokes radius of the IDR was nearly unaffected (Figure 60). To investigate the effect of the ^{107}AAWAA^{111} mutation on the conformation of the NTD and the IDR in more detail, the NTD^{AAWAA} and IDR^{AAWAA} were investigated via SEC-SAXS (Figure 82 A, B, and C). The NTD^{AAWAA} and IDR^{AAWAA} scattering patterns are very similar to those observed for the wildtype proteins. Differences, however, become evident in the real-space distance distribution (Figure 82 D). For both the NTD and the IDR, the ^{107}AAWAA^{111} mutation slightly flattens the distribution and increases the frequencies of larger distances. The R_g and D_{max} values between mutant (R_g = 3.44 nm, D_{max} = 14.5 nm) and wildtype (R_g = 3.50 nm, D_{max} = 14.5 nm) are almost identical in the context of the IDR. Conversely, the mutation in the context of the NTD increased R_g from 3.68 nm to 4.23 nm and D_{max} from 19.0 nm to 19.5 nm. This is in good agreement with the effect of the ^{107}AAWAA^{111} mutation on the Stokes radius (Figure 60 D and E). Notably, the dimensionless Kratky-plots show slightly different effects of the ^{107}AAWAA^{111} mutant on the flexibility of the NTD compared to the IDR. The IDR^{AAWAA} mutant appears to be more flexible than the wildtype IDR, as indicated by higher $(q \cdot R_g)^2 \cdot I \cdot I_0^{-1}$ values at high q-values. This effect is inversed in the NTD. The NTD^{AAWAA} protein shows lower $(q \cdot R_g)^2 \cdot I \cdot I_0^{-1}$ values at high q-values compared to the wildtype NTD.

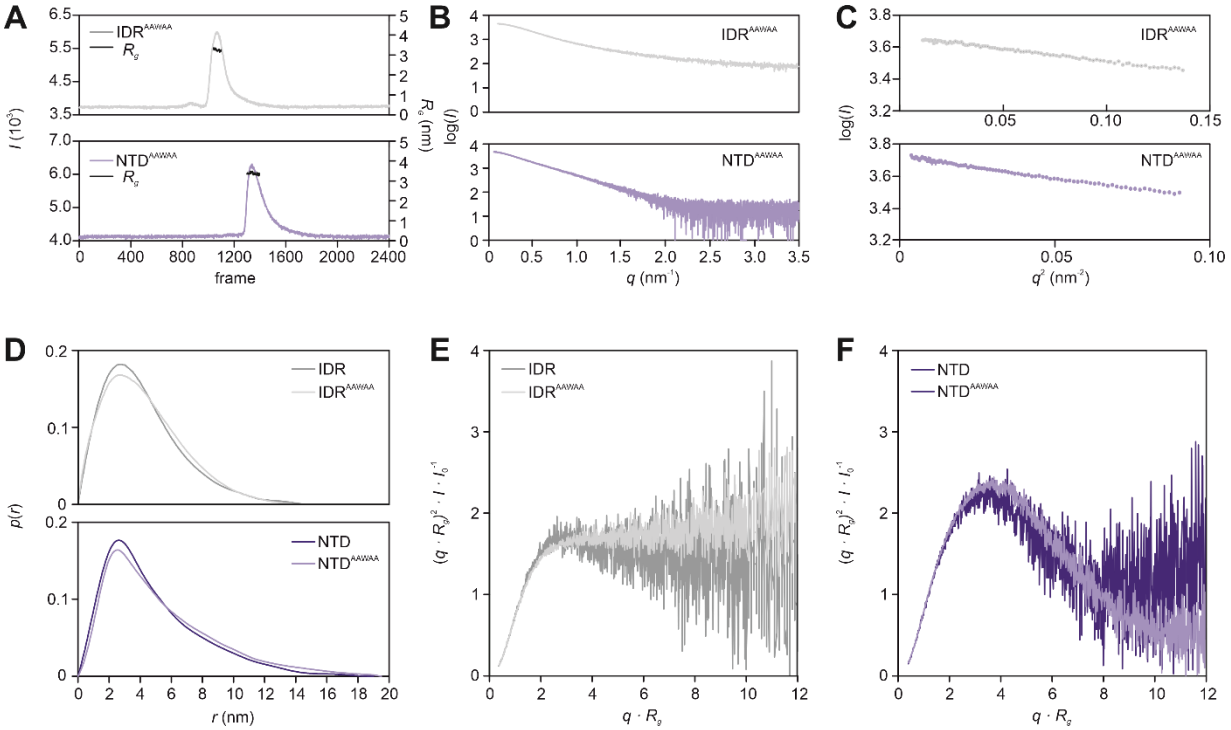


Figure 82: Small-angle X-ray scattering analysis of the PBD mutation in the *G. gallus* TRPV4 N-terminus. (**A**) SEC profiles of IDR^{AAWAA} (light grey) and NTD^{AAWAA} (light blue). Plotted is the scattering intensity I at the lowest detected q value (left axis) versus the frame number. The frames included in the final scattering profile and the corresponding radius of gyration (R_g) are indicated in black (right axis). (**B**) X-ray scattering profiles of IDR^{AAWAA} (light grey) and NTD^{AAWAA} (light blue). Plotted is the logarithmic scattering intensity log(I) versus the scattering vector q. (**C**) Guinier-plot of IDR^{AAWAA} (light grey) and NTD^{AAWAA} (light blue). Plotted is the logarithmic scattering intensity log(I) versus the squared scattering vector q^2. (**D**) Normalized real-space distance distribution of IDR (grey) compared to IDR^{AAWAA} (light grey), upper panel, and NTD (blue) compared to NTD^{AAWAA} (light blue), lower panel. Shown is the distance frequency $p(r)$ plotted versus the length of the distance r. Dimensionless Kratky-plot of (**E**) IDR (grey) and IDR^{AAWAA} (light grey), and (**F**) NTD (blue) and NTD^{AAWAA} (light blue), showing $(q{\cdot}R_g)^2{\cdot}I{\cdot}I_0^{-1}$ plotted versus $q{\cdot}R_g$ (see section VI2.18).

An Ensemble Optimization Method (EOM) analysis of IDR^{AAWAA} and the NTD^{AAWAA} was carried out and compared to the results of the wildtype EOM analyses (Figure 83) (*394, 419*). A random coil chain consisting of $N = 133$ residues was used as input to generate 10 000 starting structures of the IDR^{AAWAA} mutant. NTD^{AAWAA} starting structures were generated in analogy to the wildtype NTD using the *G. gallus* TRPV4 ARD X-ray crystal structure (PDB: 3w9g) as a template for the ARD. The remaining residues accounting for the 133-residue IDR^{AAWAA} mutant were generated as random coil chains. The final EOM refined models of the NTD^{AAWAA} and the IDR^{AAWAA} mutants showed good agreement with the experimental data (Figure 83 A, second and fourth panel). The R_g and D_{max} distribution of the IDR^{AAWAA} and NTD^{AAWAA} models agree very well with those of the random pool conformations. The NTD^{AAWAA} displays slightly elevated frequencies of expanded conformations compared to the random pool. This contrasts with the wildtype constructs for which distributions with a more compact conformation are suggested by EOM. Even though the average R_g and D_{max} values are almost equal between IDR and IDR^{AAWAA}, the EOM refined R_g and D_{max} distributions suggest that the IDR populates compact conformations more frequently than the IDR^{AAWAA} mutant. In the context of the NTD, the ^{107}AAWAA^{111} mutation increases the average R_g and D_{max} values and shifts the EOM refined R_g and D_{max} distributions to more extended conformations compared to the wildtype protein.

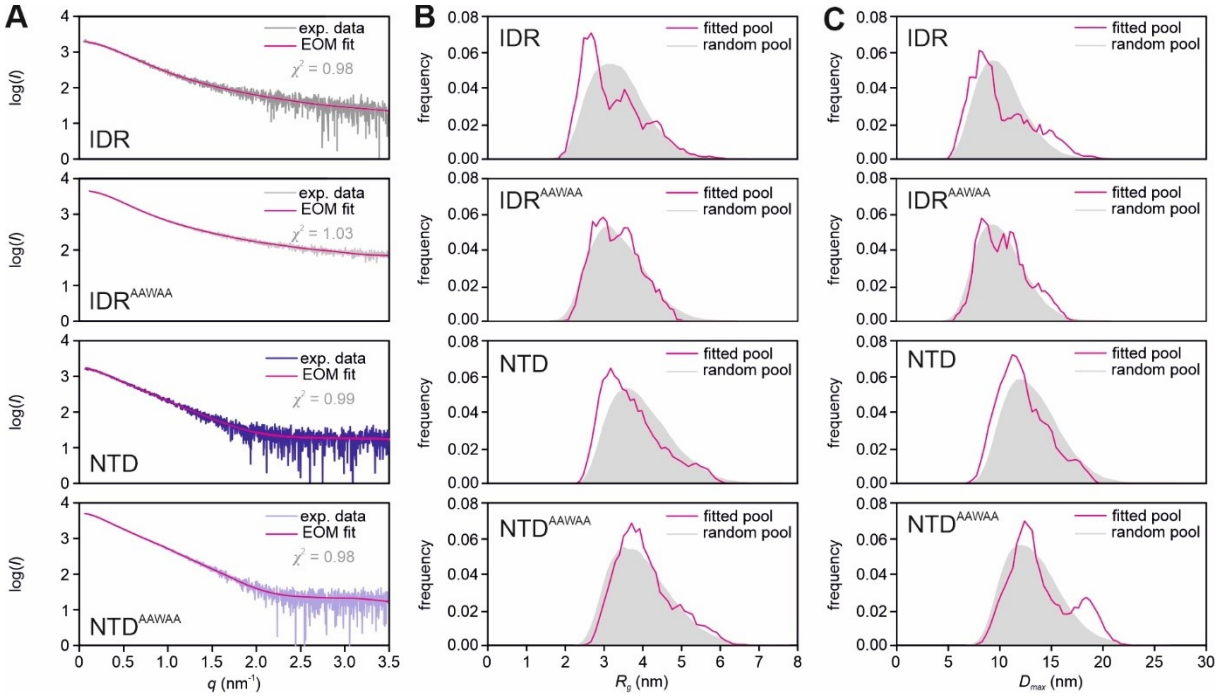


Figure 83: Ensemble optimization method (EOM) analysis of the PBD mutation in the *G. gallus* TRPV4 IDR and NTD. (**A**) Fit of the EOM refined IDR^{AAWAA} and NTD^{AAWAA} models (magenta line) to the experimental SAXS data. Experimental SAXS data are represented as a log(*I*) vs *q* plot. The goodness of the fits is indicated by the χ^2 value. (**B**) R_g and (**C**) D_{max} distributions of the random pool of generated IDR and NTD structures (grey filled curve) compared to the R_g and D_{max} distributions of the EOM refined models (magenta lines). Plotted is the frequency versus the R_g or D_{max} value, respectively. The EOM analysis results of wildtype IDR (first panel) and NTD (third panel) are included for comparison with the PBD mutants.

The effect of the ^{107}AAWAA^{111} mutation on the conformation of the TRPV4 N-terminus further substantiates a complex crosstalk between the IDR and the ARD, which at least partly seems to be mediated by the PBD. Mutating the PBD increases the dimension of the N-terminus and expands the IDR chain (Figure 60, Figure 82, Figure 83). The basic residues in the PBD therefore not only seem to mediate lipid-binding but also promote compact conformations of the TRPV4 N-terminus. In principle, two mechanisms can contribute to compact conformations: association of the IDR with the ARD, or intra-chain interactions within the IDR. NMR, tryptophan fluorescence and chemical cross-linking provide experimental evidence for both mechanisms (Figure 53, Figure 55, Figure 80, Figure 81). Interestingly, the effect of the ^{107}AAWAA^{111} mutation on the compactness of the TRPV4 N-terminus is more pronounced in the context of the NTD than in the context of the isolated IDR (Figure 60, Figure 82). Unexpectedly, the mutation transitions W109 in the PBD to a more solvent-exposed conformation in the isolated IDR, whereas the buried position of W109 in the context of the NTD remains almost unaffected (Figure 81). It seems as if the intra-chain interactions in the IDR are modulated by the presence of the ARD. Importantly, not only does the ARD influence the conformation of the IDR. Conversely, the IDR itself was observed to influence the structural dynamics in the ARD in HDX-MS experiments (Figure 56). This strongly supports the hypothesis that the IDR and ARD communicate to fine-tune how ligands bind to both domains.

In agreement with this hypothesis, the crosstalk between the IDR and the ARD appears to modulate how the TRPV4 N-terminus interacts with lipids. In the context of the NTD, truncation of N-terminal IDR residues increases the binding affinity for lipids (Figure 73). This effect is not observed in the context of the isolated IDR, i.e. in the absence of the ARD (Figure 62). Analysis of the charge distribution in the NTD indicates that the TRPV4 N-terminus is enriched in negatively charged residues. In contrast, the C-terminus of the IDR and the ARD surface are enriched in positively charged residues (Figure 71 and Figure 84). This charge distribution might mediate long-range interactions between distant regions in the N- and the C-terminus of the NTD through electrostatic interactions. The negatively charged N-terminus might compete with negatively charged lipids for the binding to the positively charged PBD. An accumulation of positive charges might be enhanced

through spatial proximity of the PBD and basic patches on the ARD surface. These can attract the negatively charged N-terminus (Figure 84). As a result, the PBD is less accessible for lipid binding, and the lipid-binding affinity is decreased. An accumulation of positive charges in the isolated IDR might be weaker due to the absence of the ARD. Therefore, the competition with lipids through electrostatic attraction between N- and C-terminus is suppressed.

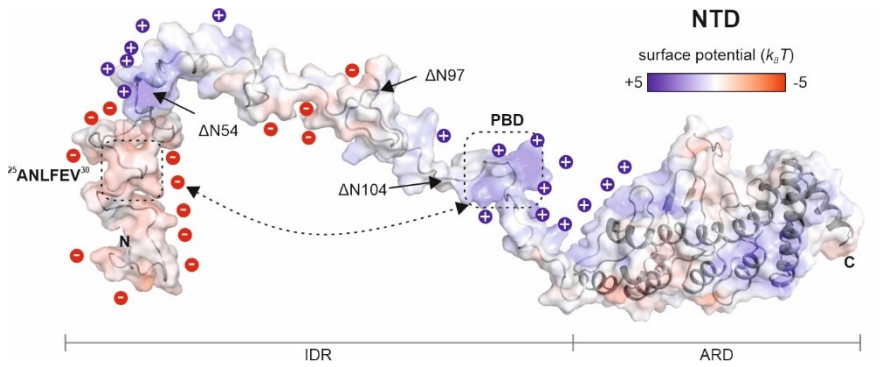


Figure 84: Potential long-range interactions between oppositely charged regions in the TRPV4 N-terminal domain. Surface potential representation in the *G. gallus* TRPV4 NTD generated with APBS (*456*) in PyMOL (*322, 323, 453*). The model of the NTD was obtained as described in section VI2.18. The N-terminal region of the IDR (residues 1-50, including the ^{25}ANLFEV^{30} patch, dashed box) is enriched in acidic residues which results in a high local negative charge. In contrast, the C-terminus of the IDR (residues 97-120) is enriched in basic residues. Together with basic patches on the surface of the ARD, the high local positive charge centred around the PBD (dashed box) might attract the negatively charged N-terminal region in the IDR. This electrostatic interaction might compete with the binding of lipids. Deletion of the negatively charged N-terminus (ΔN54) increases the lipid affinity, potentially because of lost competitive interactions within the NTD. Further deletion of the N-terminus (ΔN97) removes additional acidic patches and by this potentially further suppresses competitive intra-NTD interactions.

Strikingly, the liposome co-sedimentation (Figure 72) and NMR experiments (Figure 76, Figure 77, Figure 78) demonstrate the presence of secondary lipid-binding sites in the TRPV4 IDR. In fact, almost the entire IDR senses the binding of PIP_2 and other anionic lipids. The amino acid properties in the regions involved in lipid binding are conserved between *G. gallus* and *H. sapiens* TRPV4. A potential secondary binding site might be found within the sequence comprising ^{50}KQNLRMKFHGAFRKGPPK^{67} (*G. gallus* TRPV4 sequence, Figure 76). Enriched in basic, hydrophobic, and aromatic residues, this region strongly resembles the characteristics of the PBD and fulfils the criteria for electrostatic interactions with anionic lipids. The corresponding peaks were severely affected in the HSQC-based CSP experiments. An exciting region which showed a strong response to lipid binding is the sequence ^{25}ANLFEV^{30} in the N-terminus of the IDR (*G. gallus* TRPV4 sequence, Figure 76). The region is depleted in basic residues and flanked by very acidic patches. Besides its unexpected response to lipid binding, this region also senses the PBD as evidenced by the NMR analysis of wt-IDR and the IDR^{AAWAA} mutant. This could indicate a weak interaction between the ^{25}ANLFEV^{30} patch and the PBD. Lipid binding to the PBD might attract the N-terminal IDR more tightly resulting in line broadening for residues ^{25}ANLFEV^{30} and substantial chemical shift changes for surrounding residues. This mechanism would agree with the proposed mechanism above (Figure 84). Lipid binding appears to have dramatic consequences for the conformational state of the IDR. This holds a huge potential to modulate ARD conformational dynamics.

3.5 The TRPV4 PRR is the primary binding site for Pacsin SH3 domains

As outlined in section VI1.3, the PIP_2-dependent sensitization of TRPV4 for heat and osmotic stimuli is antagonized by the binding of an isoform of the protein Pacsin (protein kinase C and casein kinase substrate in neurons, also known as Syndapin3) to a highly conserved PAP(Q/P)PPP motif, termed the proline-rich region (PRR), in the TRPV4 IDR (Figure 85) (*76, 77*). Pacsin, also known as Syndapin, is a dimeric F-BAR (Fer-CIP4 homology Bin/Amphiphysin/Rvs) domain protein which occurs in three isoforms, Pacsin1-3, with highly conserved structures (*458*). The N-terminal Pacsin F-BAR domain interacts with membranes to sense or induce curvature (*303, 307*). This function has been implicated in endocytic and vesicle trafficking, regulation of the cytoskeleton during neuronal morphogenesis and cell migration, and receptor signalling (*458*). An SH3 (Src homology 3) domain in the C-terminus of Pacsin serves as the binding module to proline-rich substrates with PxxP motifs, such as dynamin, synaptojanin, synapsin or Wiskott-Aldrich syndrome protein (N-WASP) (*305, 306, 459*). The F-BAR and the SH3 domain within a Pacsin protomer are connected by a long linker that is presumed to be unstructured (Figure 85). The three Pacsin isoforms mainly differ in their tissue-specific expression patterns. Pacsin1 is a neuron-specific protein, Pacsin2 is ubiquitously expressed, and Pacsin3 is mainly present in skeletal muscle and heart. Pull-down and mutagenesis studies revealed that all Pacsin isoforms bind with their SH3 domains to the TRPV4 PRR in the context of the full-length channel (*76*). Expression of Pacsin3 was observed to desensitize TRPV4 to heat and osmotic stimuli and reduce the number of TRPV4 channels presented to the plasma membrane (*77*). Altered subcellular localization of TRPV4 was not observed upon expression of the other Pacsin isoforms 1 and 2, and direct modulation of TRPV4 by Pacsin1 and 2 remains to be explored. The mechanism by which Pacsin3 desensitizes TRPV4 may be encoded in the exact binding mode of the Pacsin SH3 domain to the TRPV4 PRR or in how Pacsin3 competes with PIP_2 for binding to the IDR. Structural details of Pacsin/TRPV4 interactions are, however, currently missing.

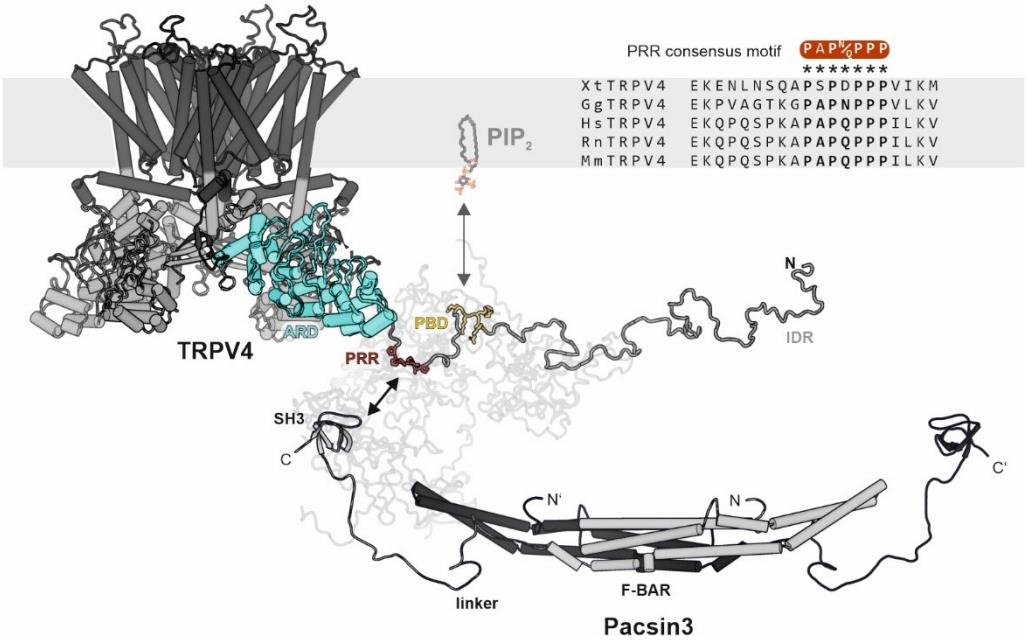


Figure 85: Overview of the interaction of TRPV4 with Pacsin. Structural model of a TRPV4 monomer with a complete IDR (generated as described in section VI2.20). The proline-rich region (PRR, red) resides between the phosphoinositide binding domain (PBD, yellow) and the Ankyrin repeat domain (ARD, cyan). Sequence analysis of the TRPV4 proline-rich region from *X. tropicalis* (Xt), *G. gallus* (Gg), *H. sapiens* (Hs), *R. norvegicus* (Rn), and *M. musculus* (Mm) reveals a PAP(N/Q)PPP consensus motif. The Pacsin3 model was generated from X-ray crystallographic and NMR structural data as described in section VI2.20. Pacsins bind with their SH3 domain to the TRPV4 PRR. The dimeric F-BAR domain can bind to membranes to sense or induce curvature.

The disordered nature of the TRPV4 IDR makes structural studies of the Pacsin-TRPV4 interaction via X-ray crystallography or via cryo-EM difficult. NMR spectroscopy, in contrast, is capable of monitoring transient interactions in unstructured regions. It was recently used to elucidate molecular details of how the Pacsin1 SH3 domain interacts with a proline-rich region of dynamin (*316*). Notably, the Pacsin1 SH3 domain binding affinity for the dynamin PRR was found to be enhanced by short-distance and long-distance elements in the dynamin C-terminus (*316*). This demonstrates that regions outside of proline-rich motifs can principally contribute to Pacsin SH3 domain binding. Such short- and long-range elements might also exist in the TRPV4 N-terminus and influence how the Pacsin SH3 domains bind to the PRR. Here, a similar approach was used to investigate how the Pacsin 1, 2, and 3 SH3 domains, termed PacX SH3 (X is the number of the Pacsin isoform), interact with the TRPV4 PRR. As described in section VI2.6, the *G. gallus* Pacsin1-3 SH3 domains could be expressed in *E. coli* and purified to homogeneity as demonstrated by SDS-PAGE and SEC analysis (Figure 86 A and B). The Stokes radii of the SH3 domains determined from SEC (Pac1 SH3: R_S = 1.7 nm; Pac2 SH3; R_S = 1.8 nm; Pac3 SH3: R_S = 1.7 nm) indicate that they are globular proteins of ~10 kDa in solution and confirm the monomeric states of the SH3 domains. The CD spectroscopic analysis indicates well-folded SH3 domains with very similar secondary structure compositions (Figure 86 C). BeStSel analysis could not find a reasonable fit to deduce the exact secondary structure content potentially owing to the high content of loop, turn and β-barrel structures for which predictions are prone to error (*339*). Nevertheless, the Pac1-3 SH3 domain CD spectra look similar to published CD spectra of SH3 domains from other proteins (*460–462*).

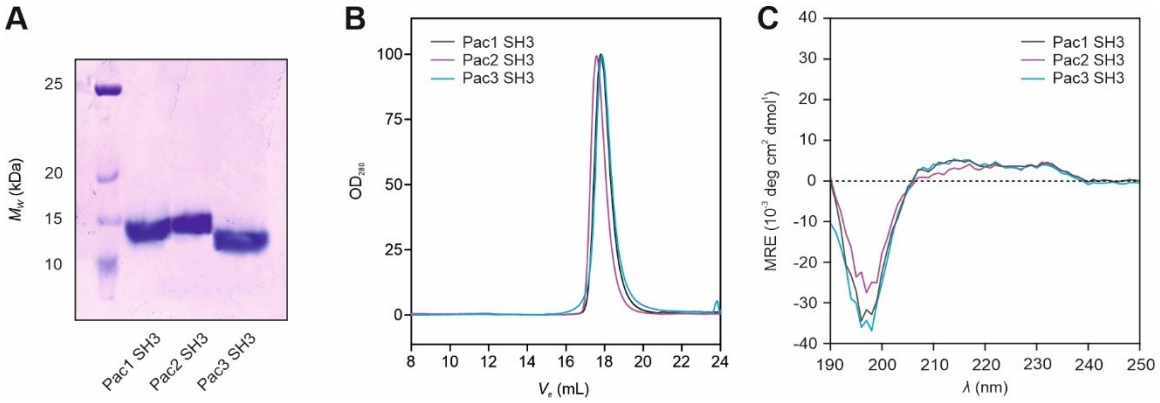


Figure 86: Purification of *G. gallus* Pacsin1-3 SH3 domains. (**A**) SDS-PAGE, (**B**) SEC, and (**C**) CD spectroscopic analysis of the purified Pac1-3 SH3 domains. The SEC analysis yielded Stokes radii of R_S = 1.6-1.7 nm. These values agree well with globular proteins of ~10 kDa. The CD spectra could not be fitted with the BeStSel tool (*339*) but comparison with published CD spectra of SH3 domains indicates proper folding of the Pac1-3 SH3 domains (*460–462*).

For further NMR spectroscopic investigations, isotope-labelled Pac1-3 SH3 constructs were generated as described in section VI2.17. The proteins produced well-dispersed ^{1}H, ^{15}N-HSQC spectra with the typical peak patterns of β-sheet proteins (Figure 87 A, B, and C). Backbone NMR resonance assignments of Pac1-3 SH3 were obtained using a set of triple resonance NMR experiments: HNCA, HNCACB, HNCO, and HNCO (see section VI2.17). Triple resonance assignments (H_N, H, C_α, C_β, C') were completed to 84.7% for Pac1 SH3, 86.6% for Pac2 SH3, and 90.8% for Pac3 SH3. No NMR experiments for the assignment of H_α resonances were carried out. The completeness of the backbone resonance assignments (H_N, H, C_α, C', H_α) was 66.2% for Pac1 SH3, 68.3% for Pac2 SH3, and 71.2% for Pac3 SH3 (see section VI2.17, p. 70). A TALOS+ secondary structure prediction based on the assigned H_N, H, C_α, C' resonances, confirms the high structural conservation between the Pacsin1-3 SH3 domains (Figure 87 D). The general SH3 domain fold is characterized by a five-stranded β-barrel (β1-5) with a so-called Src loop between β1 and β2 and a so-called RT loop between β2 and β3. The Src loop and the RT loop together with β4 and β5 form the binding site for proline-rich substrates (*463*). This topology is predicted with similar propensities for each SH3 domain (Figure 86 D) and also agrees well with the published X-ray crystal structure of the *M. musculus* Pacsin1 SH3 domain (PDB: 2x3w, Figure 86 E) (*307*). Notably, the C-terminus in the Pacsin1 SH3 domain structure does not form a β-strand as would be

expected for an SH3 domain. This might be a crystallization artefact. The alignment of the Pac1-3 SH3 sequences with the TALOS+ secondary structure prediction suggests that the corresponding region forms a β-strand in solution.

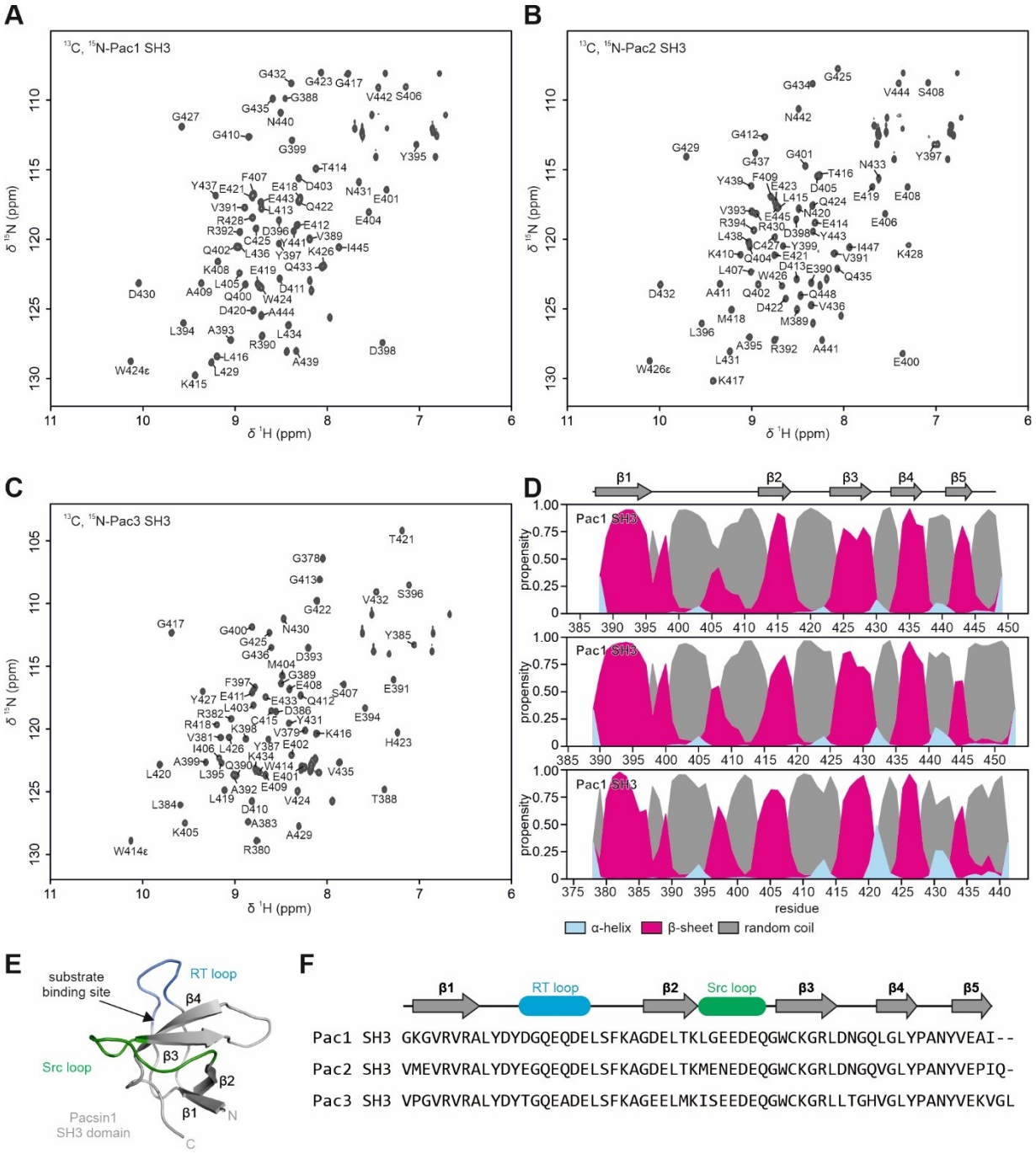


Figure 87: Backbone NMR resonance assignments and chemical shift-based secondary structure prediction of the G. gallus Pacsin1-3 SH3 domains. ^{1}H, ^{15}N-HSQC spectra of (**A**) ^{13}C, ^{15}N-Pac1 SH3, (**B**) ^{13}C, ^{15}N-Pac2 SH3, and (**C**) ^{13}C, ^{15}N-Pac3 SH3 with assigned resonances. All spectra were recorded at 298 K in 10 mM Tris pH 7, 100 mM NaCl, 1 mM DTT, 10% D_2O. (D) TALOS+ secondary structure prediction based on the H_N, H, C_α, and C' chemical shifts (*380*). H_α resonances were not assigned and therefore not included in the prediction. (**E**) X-ray crystal structure of Pacsin1 SH3 domain (PDB: 2x3w) with highlighted β-strands and Src (green) and RT loop (blue). (**F**) Sequences of *G. gallus* Pacsin1, 2, and 3 SH3 domains aligned with the topology derived from the TALOS+ prediction in (D).

To narrow down the minimal region in the TRPV4 N-terminus required for Pacsin binding without loss of affinity, ^{15}N-Pac3 SH3 domain was titrated with different TRPV4 N-terminal constructs (Figure 88). When the complete TRPV4 IDR was added, strong resonance shifts paired with severe line broadening at high IDR concentrations was observed and indicated the formation of a large and slowly tumbling complex (SH3/IDR ~25 kDa) (Figure 88 A). When the PH-PRR peptide (residues 99-134) was added, SH3 domain resonances shifted in the same direction as upon IDR titration (Figure 88 B). However, no line broadening was observed, which can be attributed to the significantly smaller complex size (SH3/PH-PRR~12 kDa). Almost identical spectral changes as upon PH-PRR titration were observed when a shorter peptide, called PBD-PRR (residues 105-134), and the isolated PRR (residues 121-134) were used as ligands (Figure 88 C and D). Titration of the NTD ΔN120 mutant – this construct represents the ARD plus the PRR – caused severe line broadening for various resonances already at low concentrations (Figure 88 E). The line broadening is consistent with the large size of the formed complex (~37 kDa). In contrast, the Ankyrin repeat domain alone does not bind the Pacsin3 SH3 domain as since the ARD titration had no visible effect on the peak positions or line widths in the ^{15}N-Pac3 SH3 domain HSQC spectrum (Figure 88 F).

Based on the backbone NMR resonance assignment, the chemical shift changes (ΔCS) between the apo state and the ligand-bound state of the SH3 domain were analyzed for each residue (Figure 89). The ΔCS versus residue plot reveals that the IDR, PH-PRR, PBD-PRR, and PRR constructs share the same binding site in the Pacsin3 SH3 domain (Figure 89 A). The strongest shifts are observed for residues in the RT loop and residues along β4 and β5. Subtle changes are also observed for backbone amide resonances of residues in the Src loop. A particularly strong shift is observed for the side chain amide resonance of W414 in the Src loop. The resonance shifts in the ^{15}N-Pac3 SH3 domain HSQC spectrum upon titration of NTD ΔN120 could not be monitored due to severe line broadening of shifting residues. Instead, the binding site was evaluated by determining the signal loss between the apo state and the SH3 domain bound to NTD ΔN120 (Figure 89 B). The signal loss distribution is mostly consistent with the chemical shift changes upon IDR, PH-PRR, PBD-PRR, and PRR titration. Strong signal loss can be observed for residues in the RT loop and along β4 and β5. Residues in the Src loop are not affected besides the side chain amide resonance of W414, which shows a complete signal loss (Figure 88 E). Titration of the ARD had almost no effect on the signal intensity or the chemical shift of the ^{15}N-Pac3 SH3 domain resonances.

To determine the dissociation constants of the SH3/ligand complexes, the chemical shift change for each peak with a shift larger than twice the standard deviation over all shifts (2σ) was plotted versus the ligand concentration (Figure 89 C). The data were fitted with a binding isotherm yielding similar dissociation constants in the mid-μM range for PH-PRR, PBD-PRR, and PRR (K_d = 68.6 ± 5.7, 74.2 ± 11.8, and 66.9 ±5 .9 μM, respectively, see section VI2.17, p. 72 for procedure). Surprisingly, the dissociation constant for the SH3-IDR complex was increased to K_d = 775.5 ± 151.2 μM, indicating a loss of affinity by one order of magnitude. No dissociation constant could be determined for the complex formation between the SH3 domain and the NTD ΔN120 mutant. The dissociation constants suggest that the PRR is the minimal region required for Pacsin3 binding. Regions preceding the PRR do not enhance SH3 domain binding. Instead, it seems that TRPV4 regions between residues 2-98 weaken the affinity of the PRR for Pacsin3 binding.

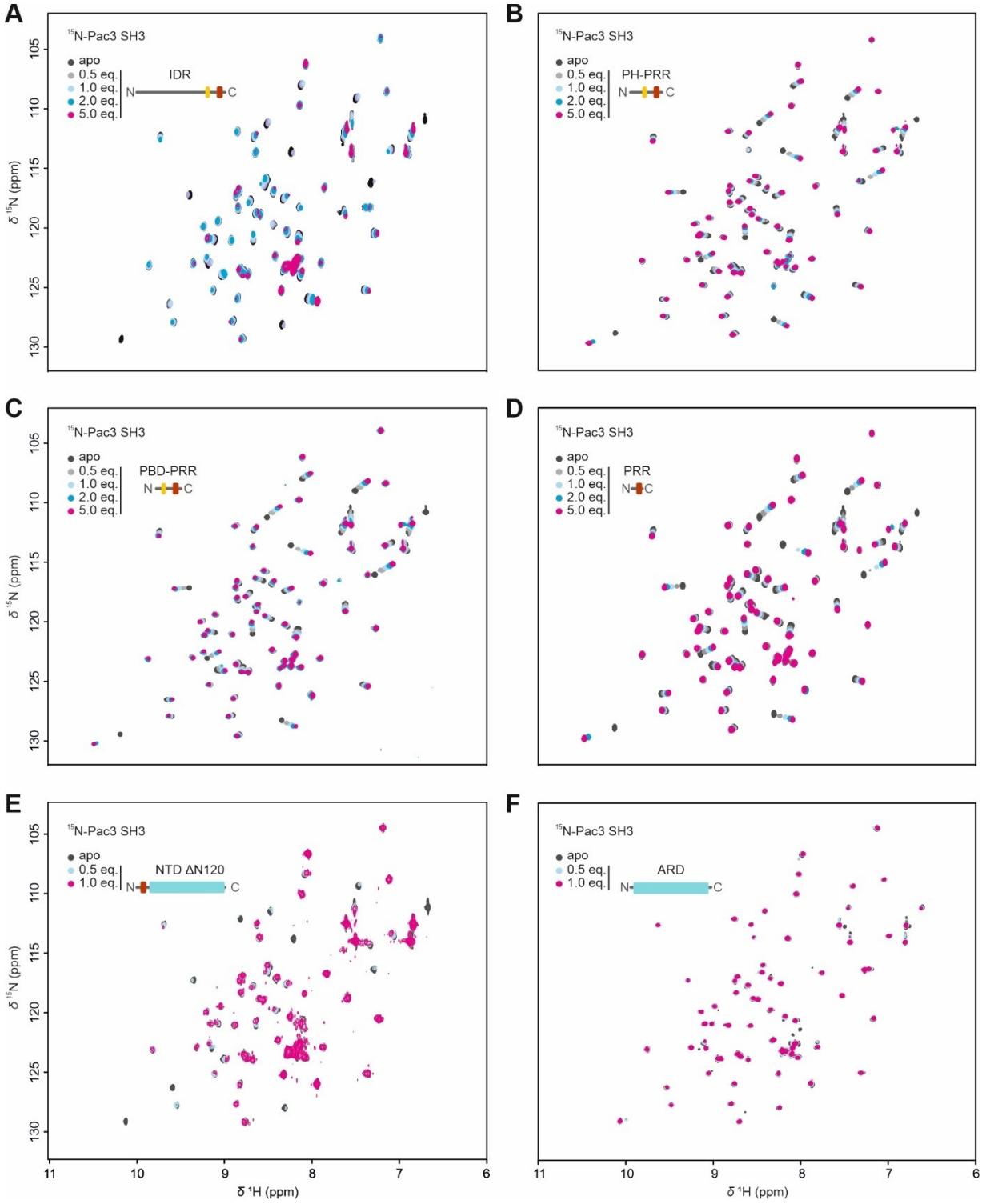


Figure 88: Chemical shift perturbation experiments of Pacsin3 SH3 domain with TRPV4 N-terminal constructs. ^{1}H, ^{15}N-HSQC spectra of 250 µM ^{15}N-Pac3 SH3 domain in the presence of different of TRPV4 N-terminal constructs: (**A**) IDR, (**B**) PH-PRR, (**C**) PBD-PRR, (**D**) PRR, (**E**) NTD ΔN120, (**F**) ARD. All spectra were measured in 10 mM Tris pH 7, 100 mM NaCl, 1 mM DTT, 10% D_2O. The colour code in the topology representation of the titrated constructs is as follows: IDR – grey; PBD – yellow; PRR – red; ARD – cyan.

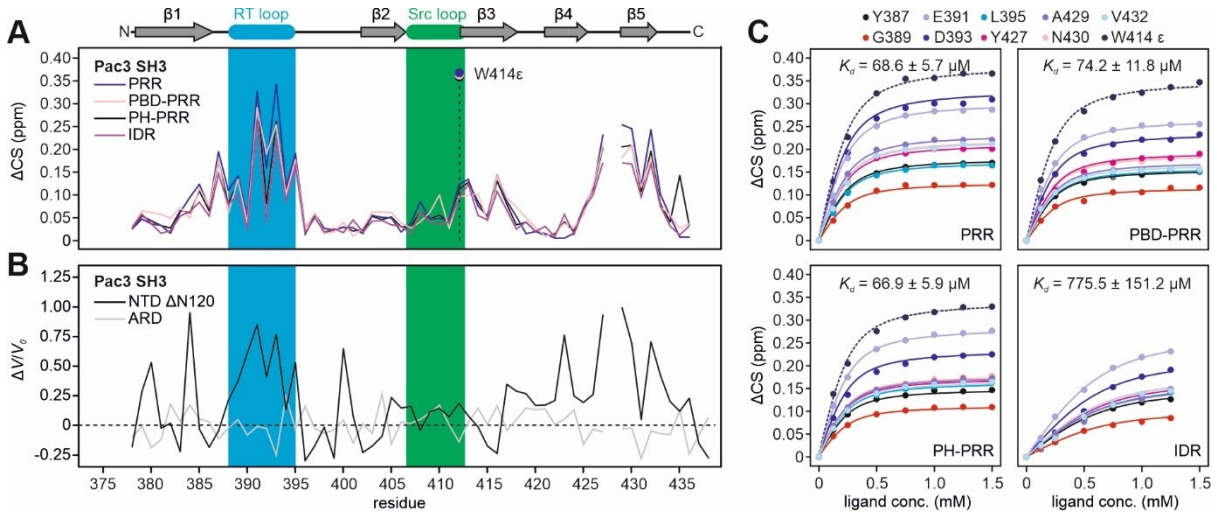


Figure 89: Chemical shift analysis of the *G. gallus* Pacsin3 SH3 domain interaction with TRPV4 ligands. (**A**) Chemical shift change, ΔCS, between the free and the PRR, PBD-PRR, PH-PRR, and IDR bound states of the ^{15}N-Pac3 SH3 domain plotted versus the residue sequence (derived from Figure 89). (**B**) Signal loss of ^{15}N-Pac3 SH3 domain resonances upon addition of the TRPV4 NTD ΔN120 mutant and the isolated ARD plotted versus the residue sequence. (**C**) Chemical shift changes of residues with ΔCS > 2σ plotted versus the ligand concentration. The data were plotted with a binding isotherm for each residue. The final dissociation constant represents the average over the dissociation constants of all residues (see section VI2.17).

To test whether the reduced SH3 domain affinity of the IDR compared to the PRR is conserved across all Pacsin isoforms, chemical shift perturbation experiments were repeated with ^{15}N-Pac1 SH3 and ^{15}N-Pac2 SH3 (Figure 90). Titration of the IDR resulted in peak shifts of Pac1 and Pac2 SH3 domain resonances accompanied by line broadening, presumably owing to the formation of a large complex of ~25 kDa (Figure 90 A and C). Titration of the PRR results in the formation of a smaller complex of ~10 kDa which manifests as peak shifts but no line broadening (Figure 90 B and D). The ΔCS versus residue plot confirms that all Pacsin SH3 domains share the same binding sites for the IDR and the PRR, that is the RT loop and β4-5 (Figure 91 A and B). The dissociation constants for the Pac1 and Pac2 SH3-PRR complexes determined from the ΔCS versus residue plots are in a similar low to mid-μM range as observed for the Pacsin3 SH3 domain (K_d = 51.6 ± 7.6 μM for Pac1 SH3, K_d = 12.7 ± 151.2 μM for Pac2 SH3) (Figure 91 C). Again, the dissociation constant for the Pac1 and Pac2 SH3-IDR complexes are reduced by one order of magnitude (K_d = 876.7 ± 204.0 μM for Pac1 SH3, K_d = 640.0 ± 122.0 μM for Pac2 SH3) (Figure 91 C).

The reduced affinity of all Pacsin SH3 domains for TRPV4 IDR binding despite displaying an identical binding site as for the isolated TRPV4 PRR might suggest that the PRR is less accessible for binding partners in the context of the IDR than in isolation. The long peptide chain preceding the PRR in the context of the IDR might shield the PRR from binding to the SH3 domain when the IDR adopts compact conformations. Moreover, a five-fold molar excess of IDR over the SH3 domains – this ratio is required to form a complex – results in an extremely crowded and viscous solution in the NMR experiment (1.25 mM IDR ≈ 20 mg mL^{-1}). This might hamper the SH3 domain from finding and binding the PRR in the TRPV4 IDR. In comparison, a five-fold PRR excess of PRR over SH3 domain results in a 10-fold decreased overall protein concentration (1.25 mM PRR ≈ 2 mg mL^{-1}). The high concentrations of interaction partners present in the NMR experiments are unlikely to occur in a cellular environment. This raises the question whether the reduced affinity of the IDR for the Pacsin SH3 domains is a physiologically relevant phenomenon or an artefact due to the experimental conditions in the NMR experiment (high concentrations of isolated protein fragments). Nevertheless, the chemical shift perturbation patterns clearly show that all constructs containing the PRR bind to the same binding sites in the Pacsin SH3 domains and that the isolated PRR is sufficient for complex formation.

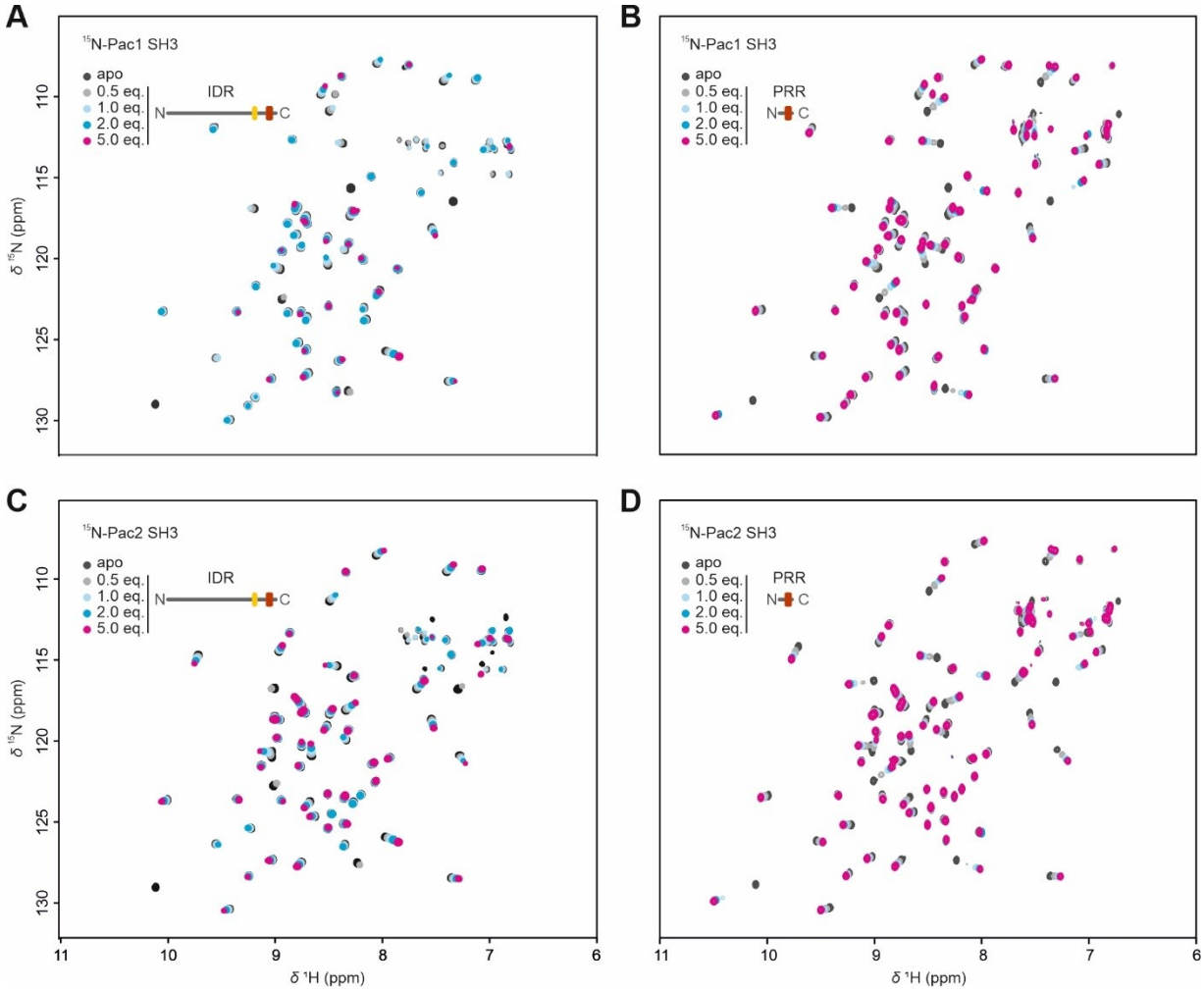


Figure 90: Chemical shift perturbation experiments of the *G. gallus* Pacsin1 and 2 SH3 domains with the TRPV4 IDR and PRR. ^{1}H, ^{15}N-HSQC spectra of 250 µM ^{15}N-Pac1 SH3 domain and of 250 µM ^{15}N-Pac2 SH3 domain in the presence of increasing (**A**, **C**) IDR and (**B**, **D**) PRR All spectra were measured in 10 mM Tris pH 7, 100 mM NaCl, 1 mM DTT, 10% D_2O. The colour code in the topology representation of the titrated constructs is as follows: IDR – grey; PBD – yellow; PRR – red.

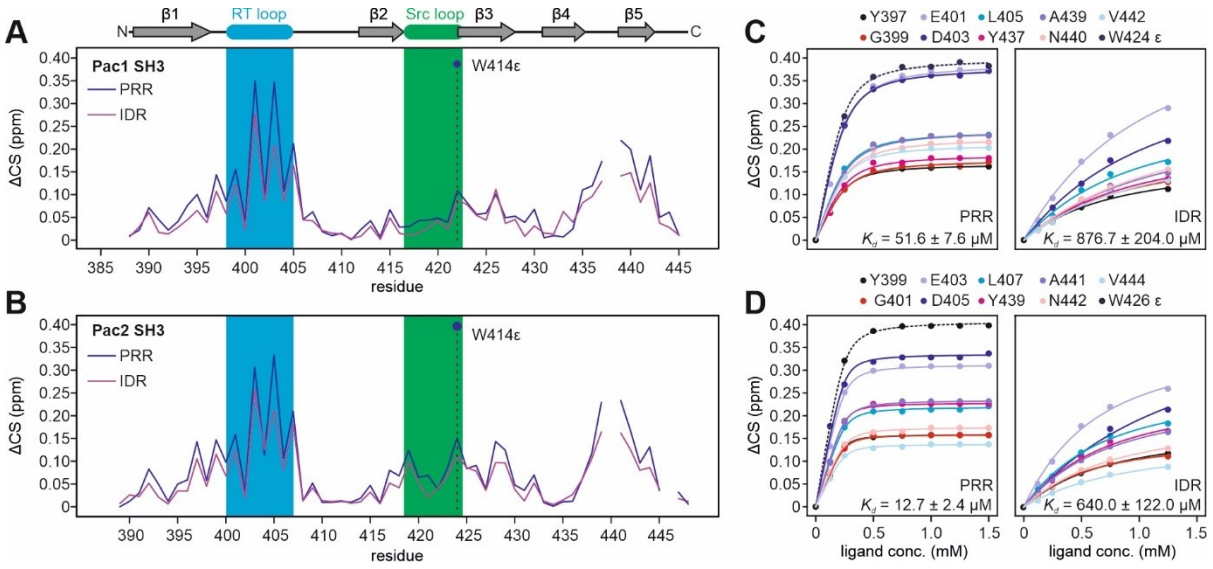


Figure 91: Chemical shift analysis of the *G. gallus* Pacsin1 and 2 SH3 domain interaction with the TRPV4 PRR and IDR. Chemical shift change, ΔCS, between the free and the PRR, or IDR bound states of the (**A**) ^{15}N-Pac1 SH3 and (**B**) ^{15}N-Pac2 SH3 domain plotted versus the residue sequence (derived from Figure 90). Chemical shift changes of (**C**) Pac1 SH3 and (**D**) Pac2 SH3 domain residues with ΔCS > 2σ plotted versus the ligand concentration. The data were plotted with a binding isotherm for each residue. The final dissociation constant represents the average over the dissociation constants of all residues (see section VI2.17).

As a control, chemical shift perturbation experiments were carried out with a ^{15}N-labeled PH-PRR peptide as the reporter and unlabeled Pac1, 2, and 3 SH3 domains as the ligands (Figure 92). The CSP experiments were carried out at pH 7 due to the instability of the SH3 domain at lower pH values. Backbone NMR resonance assignments at pH 4.5 could be largely transferred to the HSQC spectrum at pH 7. Unfortunately, the N-terminal residues Y99-E105 could not be detected at pH 7 due to severe line broadening from solvent exchange. Addition of the SH3 domains to the ^{15}N-PH-PRR sample led to chemical shift changes of nearly all resonances in the HSQC spectrum (Figure 92 A, B, and C). The ΔCS versus residue plot reveals that the strongest chemical shift changes cluster in the PRR from residues K122-V131 (Figure 92 D). The chemical shifts of the G123, A125, N127, and V131 amide resonances shift linearly with increasing SH3 domain concentrations. Plotting the G123, A125, N127, and V131 ΔCS versus the SH3 domain concentration yielded dissociation constants in the low to mid-μM range (K_d = 26.0 ± 3.9 μM for Pac1 SH3, K_d = 17.9 ± 6.2 μM for Pac2 SH3, and K_d = 38.6 ± 4.6 μM for Pac3 SH3) (Figure 92 E). These values are in good agreement with the dissociation constants determined from CSP experiments with ^{15}N-Pac1-3 SH3 domains as a reporter. This further suggests that all Pacsin SH3 domains bind equally to the TRPV4 PRR.

Interestingly, chemical shifts changes were also observed in the PBD (e.g. R108, W109, R110, R111) and in the linker between PRR and PBD (e.g. E115, K116, V118) which suggest that regions outside the PRR sense Pacsin SH3 domain binding (Figure 92 D). However, the resonance shifts do not fit to a simple binding mode with 1:1 stoichiometry between the PH-PRR peptide and the Pacsin SH3 domains. In a simple two-step binding mechanism, a protein transitions from a free to a ligand-bound state and resonances shift linearly with increasing ligand concentrations. All changes in the HSQC spectrum are expected to reach a saturation regime at similar ligand concentrations, regardless of whether the peaks correspond to residues in the ligand-binding site or to residue in regions which undergo ligand-induced conformational changes. In a scenario where the PH-PRR peptide binds to the SH3 domain with a 1:1 stoichiometry, all shifting PH-PRR resonances are expected to follow a linear trajectory and reach a saturation regime at equal SH3 concentrations.

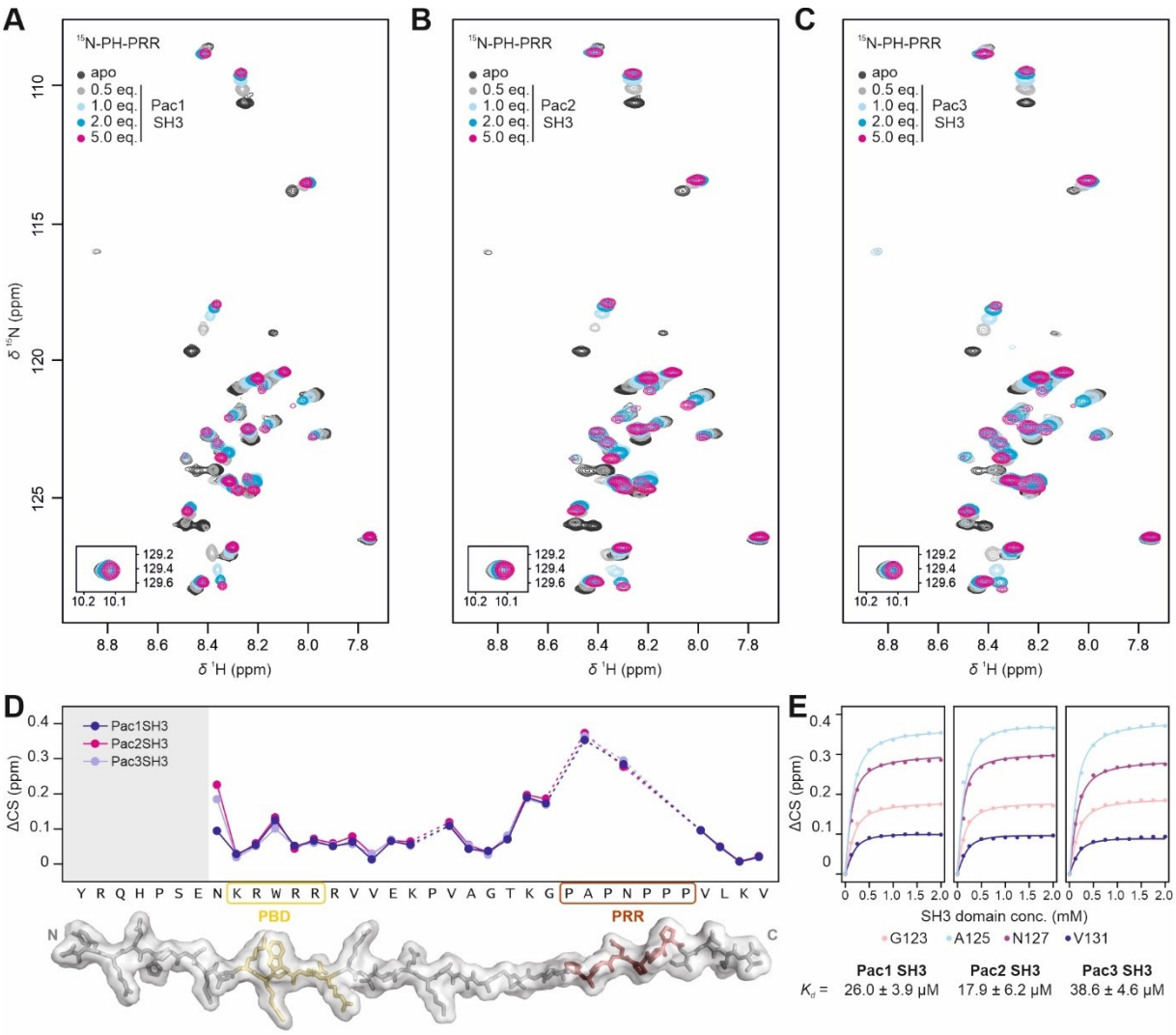


Figure 92: Chemical shift perturbation of the TRPV4 PH-PRR peptide with the Pacsin1, 2, and 3 SH3 domains. ^{1}H, ^{15}N-HSQC spectra of 250 μM ^{15}N-PH-PRR in the presence of increasing concentrations of (**A**) Pac1 SH3, (**B**) Pac2 SH3, and (**C**) Pac3 SH3 domain. All spectra were measured at 298 K in 10 mM Tris pH 7, 100 mM NaCl, 1 mM DTT, 10% D_2O. (**D**) Chemical shift difference, ΔCS, between the resonances of free and SH3 domain bound PH-PRR peptide plotted versus the TRPV4 residues. Resonances corresponding to the N-terminal residues Y99-E105 (grey box) were substantially broadened at pH 7 and therefore not visible in the HSQC spectra. (**E**) Chemical shift changes, ΔCS, of the PH-PRR residues G123, A125, N127, and V131 plotted versus the SH3 domain concentration. The data were fitted with binding isotherms yielding dissociation constants of K_d = 26.0 ± 3.9 μM for Pac1 SH3, K_d = 17.9 ± 6.2 μM for Pac2 SH3, and K_d = 38.6 ± 4.6 μM for Pac3 SH3 (average over all residues, see section VI2.17).

A detailed peak shift analysis reveals that the peak shift trajectories upon SH3 domain titration vary significantly throughout the ^{15}N-PH-PRR peptide HSQC spectrum. Three different peak shift modes can be identified and assigned to various sites in the PH-PRR peptide: The PRR (residues K122-V134), the linker between PBD and PRR (residues V114-T121), and the PBD (residues E105-V113). PRR resonances such as N127 (Figure 93 A) show linear shifts upon SH3 domains addition from which dissociation constants in a range of ~25 μM are determined. In contrast, resonances corresponding to the PBD, e.g. W109 (Figure 93 A), shift more slowly to a saturation regime which reflects in dissociation constants in a range of ~500 μM. Remarkably, amide resonances corresponding to residues in the linker region between PBD and PRR, including E115 (Figure 93 A), shift in a nonlinear manner at PH-PRR to SH3 domain ratios above 1:1. The site-specific peak trajectories in the PH-PRR peptide cannot be explained by a simple 1:1 SH3 domain/PH-PRR peptide stoichiometry upon complex formation. Instead, it indicates that the interaction of the Pacsin SH3 domains with the PH-PRR peptide involves multiple binding events, of which each has a different effect on the chemical shift of the PH-PRR resonances.

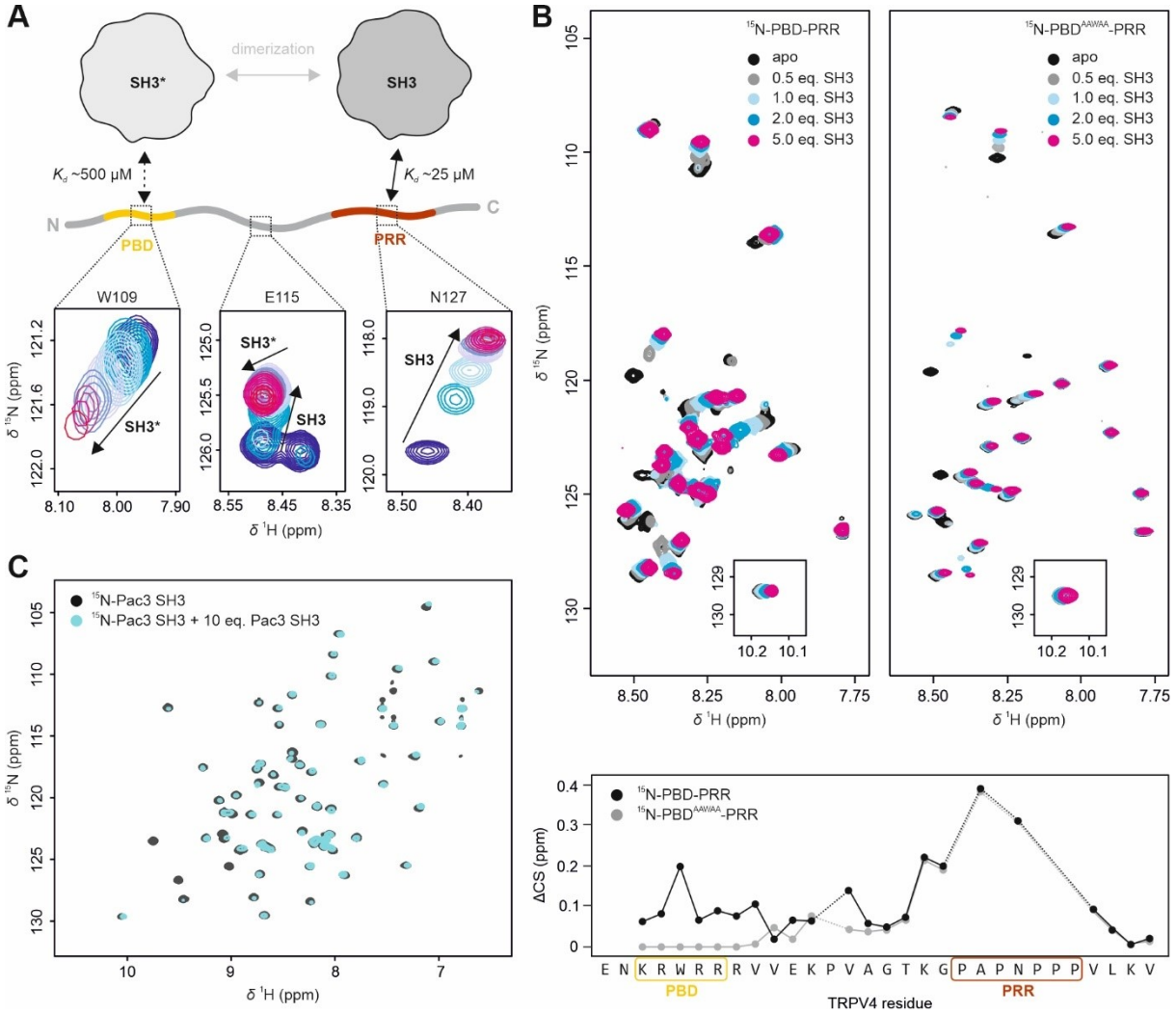


Figure 93: Indicators for a low-affinity Pacsin SH3 domain binding site in the TRPV4 PBD. (**A**) Resonances in the ^{1}H, ^{15}N-HSQC spectrum of ^{15}N-PH-PRR peptide (also see Figure 92) display different peak shift trajectories depending on their location in the peptide. The amide resonances of residues in the PRR (N127) and the PBD (W414) shift linearly whereas resonances corresponding to the linker region (E115) shift in a curved manner. Moreover, the dissociation constants determined from the PRR are in a range of ~25 µM, whereas the PBD yields constant at ~500 µM. This agrees with a model in which two SH3 domain copies (SH3 and SH3*) bind to one PH-PRR peptide simultaneously. SH3 binds to the PRR with a high affinity followed by the binding of SH3* to the PBD with low affinity. An alternative or additional contributor to the diverse peak shift modes might be SH3 domain dimerization at high concentrations (see subfigure C). (**B**) The ^{107}KRWRR^{111} to ^{107}AAWAA^{111} mutation abolishes SH3 domain binding to the PBD. Top: Overlay of ^{1}H, ^{15}N-HSQC spectra of 300 µM ^{15}N-PBD-PRR (left) and ^{15}N-PBD^{AAWAA}-PRR (right) at increasing Pac3 SH3 domain concentrations. The spectra were recorded at 293 K in 10 mM Tris pH 7, 100 mM NaCl, 10% D_2O. Bottom: Chemical shift change, ΔCS, of ^{15}N-PBD-PRR and ^{15}N-PBD^{AAWAA}-PRR resonances between the apo and SH3 domain bound state. Mutation of the PBD abolishes all SH3 domain-induced peak shifts in the C-terminus of the PBD-PRR peptide. The N-terminus, constituting the PRR, shows the same peak shifts in the wt-PBD-PRR peptide as in the PBD mutant. (**C**) Overlay of ^{1}H, ^{15}N-HSQC spectra of 250 µM ^{15}N-Pac3 SH3 domain (black) and 250 µM ^{15}N-Pac3 SH3 domain in the presence of 2.5 mM unlabeled Pac3 SH3 domain. The spectra were recorded at 298 K in 10 mM Tris pH 7, 100 mM NaCl, 10% D_2O. Line broadening of resonances at elevated SH3 domain concentrations indicates an increase in molecular weight of the studied protein due to the formation of dimers or higher oligomers. SH3 domain dimerization at elevated concentrations might be a contributor or even solely be responsible for the curved peak shift trajectories shown in (A).

A plausible explanation for the diverse peak shift trajectories throughout the ^{15}N-PH-PRR HSQC spectrum is a model in which two SH3 copies bind to the PH-PRR peptide simultaneously (Figure 93 A). In this model, the PRR acts as a high-affinity binding site with a K_d~25 µM whereas the PBD acts as a low-affinity binding site with K_d~500 µM. At early steps in the titration of ^{15}N-PH-PRR with the SH3 domains, the peak shifts are dominated by the binding of the first SH3 copy to the PRR. As the PRR becomes saturated, a second SH3

copy binds to the PBD which results in shifts of the corresponding resonances. Because the linker between the PRR and the PBD senses binding of both the first and the second SH3 protein, its resonances shift in a curved manner. It is, of course, questionable whether SH3 domain binding at a secondary binding site in the PBD is physiologically relevant considering the very low affinity determined from the chemical shift changes in this region. The PBD and the linker to the PRR do not contain the classical proline-rich sequences commonly required for SH3 domain binding. However, a recent study revealed that SH3 binding is not strictly dependent on proline residues (*464*). Various non-canonical SH3 binding motifs with low proline content exist of which several are compatible with the PBD and the linker region. For instance, the non-canonical SH3 consensus binding motif RxxxxxP (x is any amino acid) aligns well with ^{111}RxxxxxP^{117} motif in the PBD and the linker. Strikingly, when the PBD is mutated (^{107}KRWRR^{111} to ^{107}AAWAA^{111}), the HSQC spectrum displays uniform peak shifts in the linker between PBD and PRR upon SH3 domain titration, that is linear shifts from which very similar dissociation constants can be determined (Figure 93 B). Moreover, resonances corresponding to the mutated PBD show no chemical shift changes at all. The ^{107}KRWRR^{111} to ^{107}AAWAA^{111} mutation might therefore remove the low-affinity binding site of the SH3 domain and restrict SH3 binding entirely to the PRR.

The peak trajectories in the HSQC spectrum of ^{15}N-PH-PRR upon titration with SH3 domains might also be a result of SH3 domain dimerization at high concentrations, which is sensed in the PBD and the linker between the PBD and the PRR (Figure 93 A). The ability of the SH3 domain to form higher oligomers is supported by the observation that titration of unlabeled Pac3 SH3 domain to ^{15}N-Pac3 SH3 domain results in severe line broadening of resonances in the HSQC spectrum (Figure 93 C). Dimerization of the Pacsin SH3 domains at elevated concentrations might similarly affect the chemical shifts of residues in the PBD and produce curved peak shift trajectories in the linker region as a direct binding of SH3 domains. The observation that the ^{107}KRWRR^{111} to ^{107}AAWAA^{111} mutation abolishes the SH3 domain-induced resonance shifts in the PBD makes the dimerization mechanism less likely because mutations in the PBD should not influence the dimerization of the SH3 domains.

Previous studies on *H. sapiens* TRPV4 have demonstrated that depletion of PIP_2-levels in HEK293 cells abolishes channel response to heat and osmotic stimuli similarly as mutating the ++W++ motif in the PBD to AAWAA (*128*). The loss of channel response was accompanied by a structural change of the cytosolic domains to a more compact conformation observed by FRET measurements (*128*). Therefore, the interaction of PIP_2 seems essential to keep TRPV4 in a sensitized state with a more expanded cytosolic domain. Remarkably, co-expression of TRPV4 with Pacsin3 not only desensitized TRPV4 to heat and osmotic stimuli, but it also led to a compaction of the cytosolic domains, similar as PIP_2-depletion or PBD mutation (*77, 128*). It is tempting to conclude from these observations that Pacsin3 desensitizes TRPV4 by directly competing with PIP_2 for the binding sites in the TRPV4 N-terminus. The NMR chemical shift perturbation studies of ^{15}N-labelled TRPV4 PH-PRR peptide with the Pacsin SH3 domain show that SH3 domain binding is largely restricted to the PRR (Figure 92). The PBD, however, seems to act as a secondary low-affinity binding site for the Pacsin SH3 domains. Considering the partial overlap between the PBD and Pacsin SH3 domain binding in the TRPV4 N-terminus, binding of the Pacsin3 SH3 domain to the PRR might directly influence how PIP_2 binds to the PBD.

Chemical shift perturbation experiments were carried out using the ^{15}N-labeled TRPV4 PBD-PRR peptide as a reporter to test whether the Pacsin SH3 domain and PIP_2 compete for their binding sites or bind to the TRPV4 N-terminus simultaneously (Figure 94 A and B). First, unlabeled Pac3 SH3 domain was titrated to the ^{15}N-PBD-PRR, which caused chemical perturbations indicating complex formation between both interaction partners (Figure 94 A). Resonances of residues in the PRR such as N127 are strongly shifted whereas resonances of residues in the PBD, e.g. W109, are not or only slightly shifted (Figure 94 C). Similar as was seen when the ^{15}N-PH-PRR peptide was titrated with Pacsin3 SH3 domain (Figure 92, section VI3.5), the peak shift pattern agrees with SH3 domain binding to the PRR. When diC_8-PIP_2 was titrated to the PBD-PRR/SH3 complex, almost all ^{15}N-PBD-PRR resonances showed severe line broadening beyond the detection limit (Figure 94 A and C). The detachment of the SH3 domain from the PRR upon PIP_2 binding would yield the spectrum of the lipid-bound PBD-PRR peptide, in which resonances of the PRR display sharp peaks (compare Figure 94 B). Instead, the severe line broadening of PRR and PBD resonances suggests that diC_8-PIP_2 binds to the PBD

while the SH3 domain remains bound to the PRR. This results in a large and slowly tumbling PBD-PRR/SH3/PIP_2 tripartite complex (Figure 94 B).

Similar observations were made when the titration order of SH3 domain and diC_8-PIP_2 was reversed (Figure 94 B and D). Titration of diC_8-PIP_2 caused resonance shifts and line broadening of residues in the PBD such as W109 whereas resonances of PRR residues like N127 were almost unaffected (Figure 94 A and B). Subsequent addition of SH3 domain to the PBD-PRR/diC_8-PIP_2 complex resulted in a signal loss of almost all ^{15}N-PBD-PRR resonances (Figure 94 B and D). Displacement of PIP_2 from the PBD through SH3 domain titration would be expected to reproduce the HSQC spectrum of the PBD-PRR/SH3 domain complex which shows a different chemical shift pattern and line width of the ^{15}N-PBD-PRR resonances (Figure 94 B). The severe line broadening throughout the HSQC spectrum, therefore, again indicates the formation of a large and slowly tumbling PBD-PRR/SH3/PIP_2 tripartite complex.

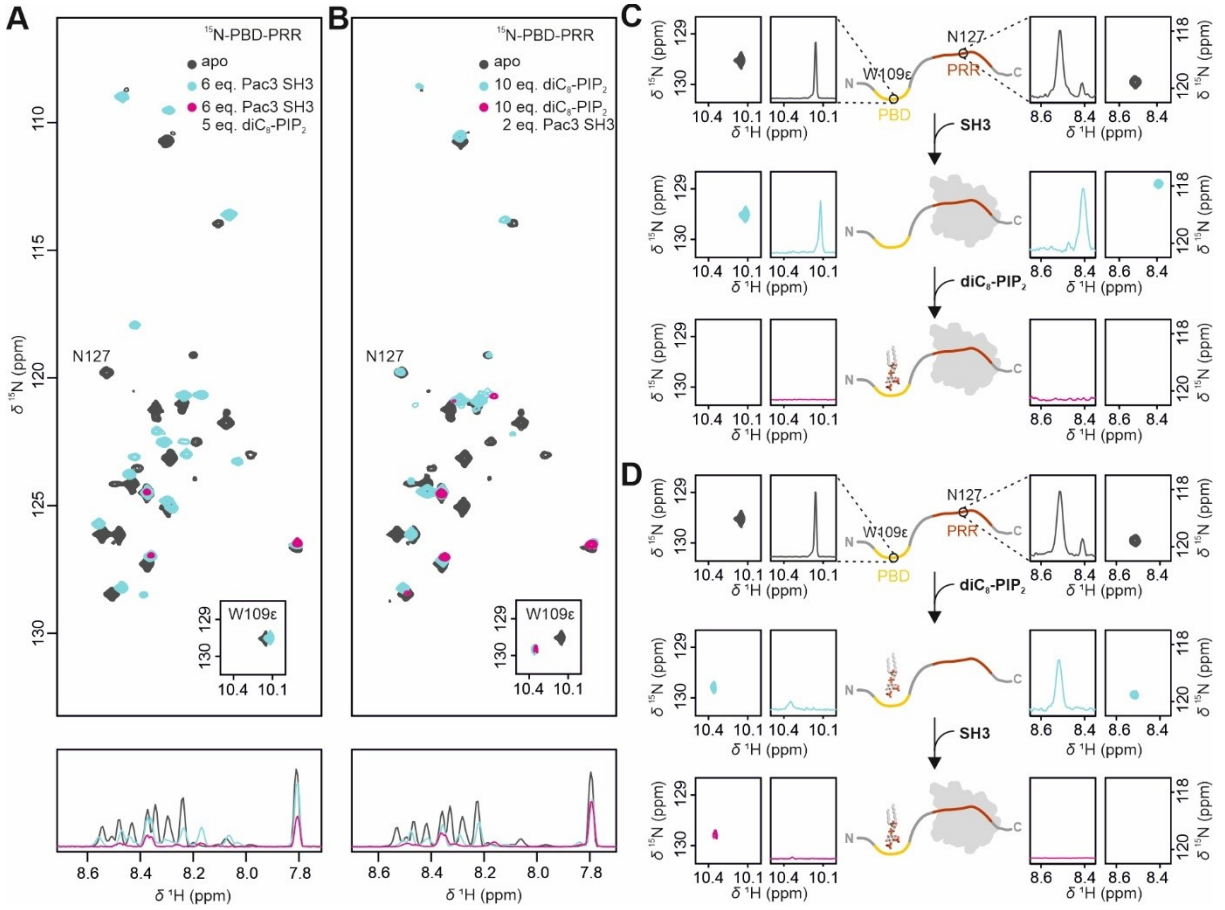


Figure 94: Chemical shift perturbation experiment probing the simultaneous binding of the Pacsin3 SH3 domain and PIP_2 to the TRPV4 PBD-PRR peptide. ^{1}H, ^{15}N-HSQC spectra of 300 µM ^{15}N-PBD-PRR titrated with diC_8-PIP_2 followed by Pac3 SH3 domain (**A**), and in the reverse titration order (**B**). All spectra were recorded at 293 K in 10 mM Tris pH 7, 100 mM NaCl, 10% D_2O. 1D-projections of the ^{1}H resonances indicate the line broadening during lipid and SH3 domain titration. (**C**) and (**D**) Effect of titrating diC_8-PIP_2 and the Pac3 SH3 domain to ^{15}N-PBD-PRR in different orders on the chemical shift and line width of the W109ε resonance in the PBD and the N127 resonance in the PRR. A section from the 2D ^{1}H, ^{15}N-HSQC spectrum shows the W109ε and N127 chemical shift, the 1D projection of the ^{1}H resonance shows the peak line width. (**C**) Titration of the SH3 domain causes a peak shift of the N127 resonance, whereas the W109ε resonance remains mostly unaffected. This indicates SH3 binding predominantly to the PRR. Subsequent addition of diC_8-PIP_2 leads to line broadening of both the W109ε and the N127 resonances. This is consistent with the formation of a large PBD-PRR/SH3/PIP_2 tripartite complex. (**D**) When diC_8-PIP_2 is first titrated to the PBD-PRR, the W109ε resonance is severely broadened and shifted. The N127 resonances, in contrast, is almost unaffected. This is indicative of PIP_2 binding to the PBD. Subsequent addition of the SH3 domain causes additional line broadening of the N127 resonances, thus indicating binding of the SH3 domain to the PRR which results in a large PBD-PRR/SH3/PIP_2 tripartite complex.

In an alternative chemical shift perturbation experiment, ^{15}N-Pac3 SH3 was used as the reporter, and diC_8-PIP_2 and unlabeled PBD-PRR were titrated in different orders (Figure 95 A and B). The effects of PIP_2 and PBD-PRR titration on the ^{15}N-Pac3 SH3 domain resonances in the HSQC spectrum is well represented by the Y387 resonance (Figure 95 C). When diC_8-PIP_2 was first titrated to the SH3 domain, no changes of peak positions or line width of ^{15}N-Pac3 SH3 resonances were observed in the HSQC spectrum which indicates that the SH3 domain itself does not interact with diC_8-PIP_2 (Figure 95 C, lower panel). Subsequent addition of the PBD-PRR to the ^{15}N-Pac3 SH3/diC_8-PIP_2 mixture caused the same peak shifts which previously observed in the absence of lipid (see section VI3.5, Figure 88 C). This demonstrates that the presence of diC8-PIP_2 does not prevent the SH3 domain from binding to the PBD-PRR peptide. Strikingly, the peak shifts were accompanied by severe line broadening, an effect which was not observed in the lipid-free titration experiment (see section VI3.5, Figure 88 C). This strongly suggests that the Pac3 SH3 domain becomes part of a slowly tumbling Pac3 SH3/PBD-PRR/diC_8-PIP_2 tripartite complex. When the order of titration was reversed, the addition of unlabeled PBD-PRR peptide to ^{15}N-Pac3 SH3 caused resonance shifts and subtle line broadening (Figure 95 C, upper panel), thus indicating the formation of a Pac3 SH3/PBD-PRR complex as described in section VI3.5. Subsequent addition of diC_8-PIP_2 caused severe line broadening throughout the HSQC spectrum. This again suggests that the ^{15}N-Pac3 SH3 domain becomes part of a larger complex with the PBD-PRR peptide and PIP_2. Notably, both titration orders yield almost identical peak positions and line widths of the ^{15}N-Pac3 SH3 domain resonances in the HSQC spectrum (Figure 95 C, right panel). Moreover, the PBD-PRR peptide binds to the Pac3 SH3 domain with almost the same affinity regardless of the presence of diC_8-PIP_2. Together, this strongly indicates that binding of the Pac3 SH3 domain to the PRR is not influenced by PIP_2 binding to the PBD and *vice versa*.

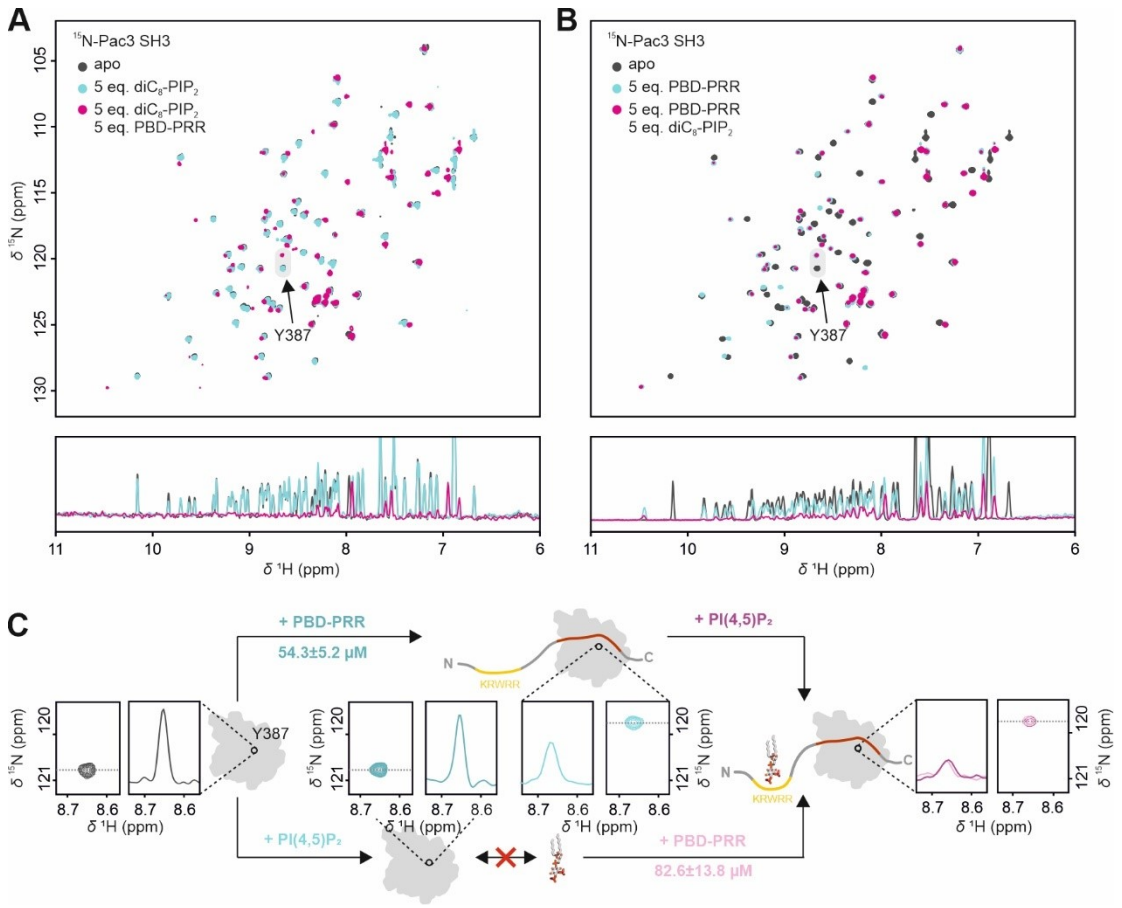


Figure 95: Chemical shift perturbation experiment probing the interaction of the Pacsin3 SH3 domain with the TRPV4 PBD-PRR peptide in the presence of lipids. ^{1}H, ^{15}N-HSQC spectra of 250 µM ^{15}N-Pac3 SH3 titrated with diC_8-PIP_2 followed by the PBD-PRR peptide (**A**), and in the reverse titration order (**B**). All spectra were recorded at 298 K in 10 mM Tris pH 7, 100 mM NaCl, 10% D_2O. 1D-projections of the ^{1}H resonances indicate the line broadening during lipid and peptide titration. (**C**) Effect of titrating diC_8-PIP_2 and the PBD-PRR peptide to ^{15}N-Pac3 SH3 domain in different orders on the chemical shift and line width of the Y387 resonance. A section from the 2D ^{1}H, ^{15}N-HSQC spectrum shows the Y387 chemical shift, the 1D projection of the ^{1}H resonance shows the peak line width. Addition of diC_8-PIP_2 has no effect on the line width or chemical shift of the Y387 resonance (lower panel), thus demonstrating that the SH3 domain PIP_2 do not interact. Conversely, the addition of the PBD-PRR peptide causes peak shift and concomitant line broadening, indicating PBD-PRR/SH3 domain complex formation (upper panel). In a subsequent step, the respective second binding partner (PIP_2 in the upper panel, PBD-PRR in the lower panel) is added. Both routes lead to the same Y387 chemical shift and line width, showing that the PBD-PRR forms a tripartite complex with the SH3 domain and PIP_2. Notably, the dissociation constant for PBD-PRR/SH3 domain complex formation determined from the Y387 peak shifts is almost unaffected by the presence of PIP_2: K_d = 54.3 ± 5.2 µM no PIP_2 vs K_d = 82.6 ± 13.8 µM with PIP_2.

As a control, the chemical shift perturbation experiment was repeated with the PBD^{AAWAA}-PRR mutant (^{107}KRWRR^{111} to ^{107}AAWAA^{111}) instead of the PBD-PRR peptide (Figure 96 A and B). The PBD^{AAWAA}-PRR peptide can still bind to the Pac3 SH3 domains with its intact PRR (Figure 93, section VI3.5), but the interaction with PIP_2 is abolished (compare Figure 68, section VI3.3). Consistent with this, the titration of PBD^{AAWAA}-PRR mutant to ^{15}N-Pac3 SH3 produced the same peak shifts and line broadening as observed for titration of the wt-PBD-PRR peptide (Figure 95). However, titration of diC_8-PIP_2 either prior or subsequently to PBD^{AAWAA}-PRR addition did not affect the line width of the ^{15}N-Pac3 SH3, as observed in the titrations with wt-PBD-PRR peptide (Figure 96 A, B, and C). Consistently, the affinity of the PBD^{AAWAA}-PRR mutant for SH3 domain binding is not affected by the presence of PIP_2 (Figure 96 C). This suggests that the PBD^{AAWAA}-PRR peptide is unable to form a tripartite complex with PIP_2 and the SH3 domain.

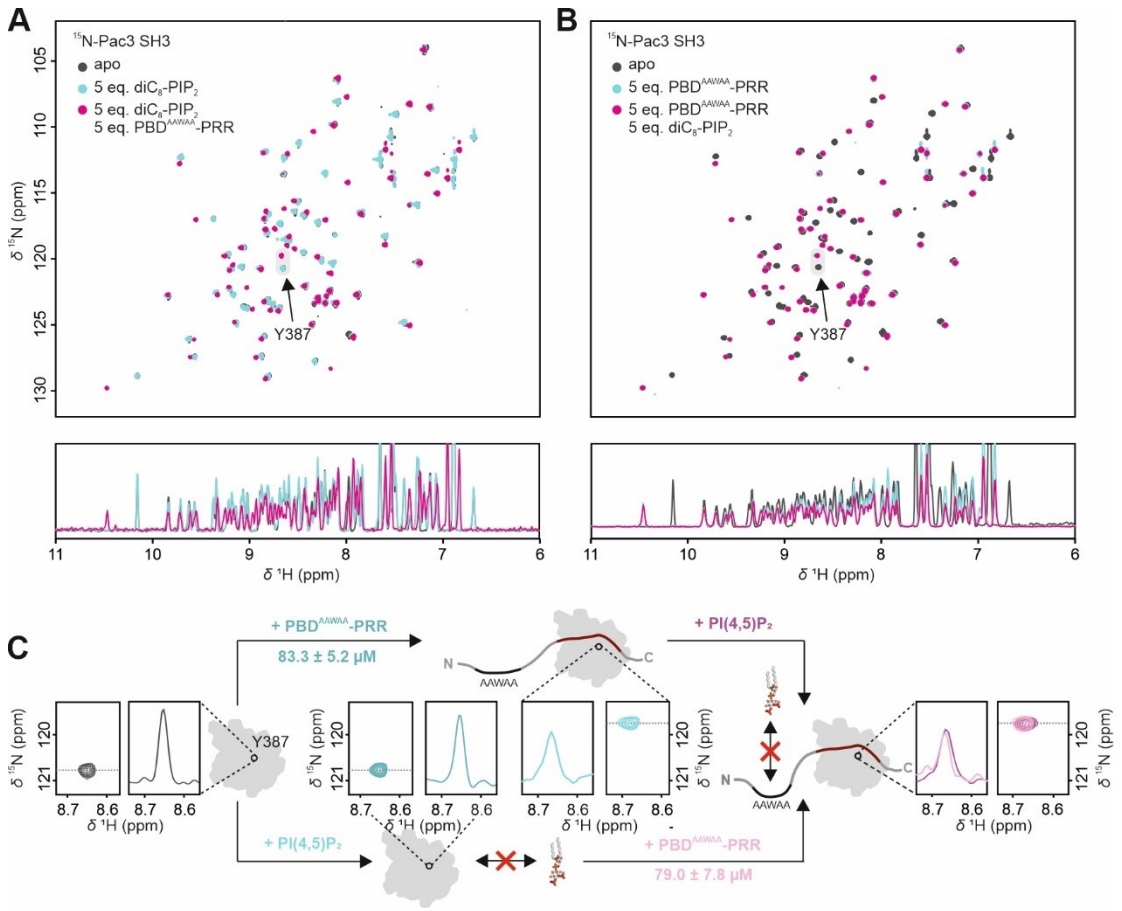


Figure 96: Chemical shift perturbation (control) experiment probing the interaction of the Pacsin3 SH3 domain with the TRPV4 PBD^{AAWAA}-PRR mutant in the presence of lipids. ^{1}H, ^{15}N-HSQC spectra of 250 µM ^{15}N-Pac3 SH3 titrated with diC$_8$-PIP$_2$ followed by the PBD^{AAWAA}-PRR mutant (**A**), and in the reverse titration order (**B**). All spectra were recorded at 298 K in 10 mM Tris pH 7, 100 mM NaCl, 10% D_2O. 1D-projections of the ^{1}H resonances indicate the line broadening during lipid and peptide titration. (**C**) Effect of titrating diC$_8$-PIP$_2$ and the PBD^{AAWAA}-PRR mutant to ^{15}N-Pac3 SH3 domain in different orders on the chemical shift and line width of the Y387 resonance. A section from the 2D ^{1}H, ^{15}N-HSQC spectrum shows the Y387 chemical shift, the 1D projection of the ^{1}H resonance shows the peak line width. As shown in Figure 95, the addition of diC$_8$-PIP$_2$ has no effect on the line width or chemical shift of the Y387 resonance (lower panel). Addition of the PBD^{AAWAA}-PRR peptide causes almost identical peak shifts and concomitant line broadening as the wt-PBD-PRR peptide (Figure 95), indicating PBD^{AAWAA}-PRR/SH3 domain complex formation (upper panel). In a subsequent step, the respective second binding partner (PIP$_2$ in the upper panel, PBD^{AAWAA}-PRR in the lower panel) is added. Both routes lead to the same Y387 chemical shift and line width. Importantly, the addition of PIP$_2$ to the PBD^{AAWAA}-PRR/SH3 domain complex has no effect on the peak position or line width. As expected for the AAWAA mutant, this indicates that PIP$_2$ does not bind to the PBD^{AAWAA}-PRR/SH3 complex. Accordingly, the dissociation constant for PBD^{AAWAA}-PRR/SH3 domain complex formation determined from the Y387 peak shifts is almost unaffected by the presence of PIP$_2$: K_d = 83.3 ± 5.2 µM no PIP$_2$ vs K_d = 79.0 ± 7.8 µM with PIP$_2$ (the error is the fit standard error).

In summary, the PAP(Q/N)PPP consensus motif of the proline-rich region represents the minimal TRPV4 sequence required for the binding of all Pacsin SH3 domains. Unlike what was described previously for the Pacsin1-dynamin interaction (*316*), the TRPV4 IDR does not seem to contain short- or long-distance elements from the PRR, enhancing SH3 domain binding specificity. In contrast, the complete IDR showed a lower affinity for SH3 domain binding compared to the isolated PRR or the PBD-PRR and PH-PRR peptides which additionally contain the PBD. However, as pointed out above, the reduced affinity of the IDR might be an artefact due to the experimental conditions (high concentrations and isolated protein fragments) required for NMR. Similarly, the high protein concentrations are likely responsible for the binding of the SH3 domain to the TRPV4 PBD. In a cellular context, it is more than unlikely that such a high SH3 domain concentration occurs in the vicinity of the IDR to allow for simultaneous binding of two SH3 domains to the PBD and the PRR. However, a low-affinity binding site next to a high-affinity binding site still might frequently bind a Pacsin SH3 domain. Whether this has any functional relevance can only be speculated at this moment. It is possible

that the Pacsin SH3 domains can switch between the low and high-affinity binding site and thereby increase the residency time of the SH3 domain close to the PRR. The ability of the Pacsin SH3 domain to bind to the PBD, however, does not interfere with the binding of lipids. Considering the determined lipid-affinities in the mid-μM range (see section VI3.3), it is likely that the PBD is mostly bound to the membrane in a cellular context. Accordingly, SH3 domain binding to the low-affinity site in the PBD will be outcompeted by the binding of lipids in most cases. The observation that the Pacsin SH3 domain and PIP_2 can bind to the TRPV4 PBD-PRR peptide suggests that the SH3 domain alone does not compete with PIP_2 binding to the TRPV4 N-terminus. This is consistent with published experiments in which co-expression of the isolated Pacsin3 SH3 domain in HEK293 cells is not sufficient to desensitize TRPV4 (*128*). Potential competition between PIP_2 and Pacsin3 binding to TRPV4, therefore, seems to require additional factors that are not present in the system used in this study. Finally, the observation that all Pacsin SH3 domains bind equally to the TRPV4 PRR may indicate that, even though Pacsin1 and Pacsin2 do not regulate the subcellular localization of TRPV4, they desensitize TRPV4 similar to Pacsin3.

3.6 Pacsin SH3 domain binding requires a *cis* proline conformation in the TRPV4 PRR

Structural details determine both the specificity and the functional consequences of SH3 binding to a target (*463*). Most SH3 domains bind to proline-rich substrates with left-handed polyproline helix (PPII) structures. SH3 domains exhibit two proline-pockets, termed PP1 and PP2, between the Src and the RT loop to recognize PxxP motifs (x can be any amino acid). Because of the PPII helix pseudo-symmetry, SH3 domain substrate binding sites can recognize PxxP motifs in two different orientations, termed class I and class II (Figure 97 A). The orientation of the substrate relative to the SH3 domain is determined by a specificity pocket (SP) formed by the RT and Src loop adjacent to the PxxP-binding site. In most canonical SH3 domains, the specificity pocket is negatively charged and recognizes positively charged residues in the substrate next to the PxxP motif (*464*). In a class I substrate, the positively charged residue is N-terminal to the PxxP motif, in a class II substrate it is C-terminal. The canonical class I and II consensus motifs can be written as +xϕPxxP and xPxϕPx+, respectively. Here, '+' is arginine or lysine, 'ϕ' is a hydrophobic residue. The proline residues in the PxxP motif (underlined) bind to the SH3 domain proline pockets PP1 and PP2. The basic residue (+) binds to the specificity pocket between the Src and RT loops (*463*).

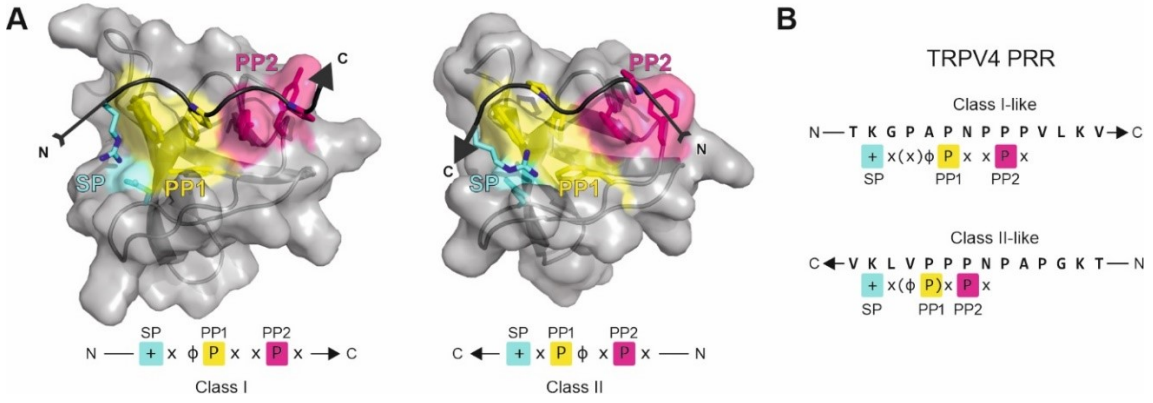


Figure 97: Canonical SH3 domain substrate interactions. (**A**) Representative structures of SH3 domains in complex with a class I (PDB: 1rlq) and class II (PDB: 1sem) substrate. The peptide backbone is shown in black with the direction from N- to C-terminus indicated by an arrow. Side chains binding to the selectivity pocket (SP, cyan), the first proline pocket (PP1, yellow), and the second proline pocket (PP2, magenta) are shown as s coloured sticks. The SH3 domains are shown as a transparent surface with the SP, PP1, and PP2, with colours matching the bound residues in the substrate. The class I and class II consensus binding motifs are shown below the structures. Here, 'x' represents any amino acid, P is proline, '+' is lysine or arginine, and 'ϕ' is a hydrophobic residue. (**B**) The TRPV4 PRR is compatible with a class I-like and a class II-like SH3 domain interaction. Brackets indicate where the PRR sequence deviates from the canonical class I or class II motifs.

The orientation of the TRPV4 PRR in complex with the Pacsin SH3 domain determines the overall orientation of the Pacsin proteins with respect to the TRPV4 channel in the context of the full-length proteins. It is, therefore, a crucial detail for understanding the molecular mechanism of Pacsin binding to TRPV4. For most SH3 domain substrates, the binding mode can be predicted from the amino acid sequence. The TRPV4 PRR sequence, however, shows compatibility with both a class I and class II motif. Nonetheless, in either case, a single amino acid mismatch prevents a perfect agreement between TRPV4 and the consensus sequences (Figure 97 C). The exact agreement with a class I motif is perturbed by two (G123 and P124) instead of one residue between the 'ϕ'-residue and the SP binder. For an exact match with a class II motif, P130 directed at PP1 needs to be switched with its 'ϕ'-neighbour V131. Various non-canonical motifs have been reported in which the number of 'x'-residues between the proline and the '+'-residue or the position of the hydrophobic 'ϕ'-residue deviate from the canonical class I and II motifs (*464*). The TRPV4 PRR might therefore bind the Pacsin SH3 domain in a class I-like or a class II-like mode.

Prior to investigating the exact mode of TRPV4 PRR binding to the Pacsin SH3 domains, the PRR was probed for its ability to form a PPII helix in solution. Because proline does not have a backbone amide proton, PPII helices are not held together by hydrogen bonding. As a result of the absence of hydrogen bonds, PPII helices produce similar CD spectra as disordered proteins (*465*). In contrast to disordered proteins, however, stable PPII helices refrain from transitioning to α-helical structures upon addition of secondary structure inducing agents such as TFE (*466*). This behaviour is observed for the TRPV4 PRR (Figure 98 A). When titrated with TFE, the secondary structure of the TRPV4 PRR remains unaffected. This is a strong indicator that the PRR forms a PPII helix in solution. As a control, a TFE titration was carried out with a peptide which corresponds to the N-terminal region preceding the TRPV6 ARD, termed V6H. In the TRPV6 cryo-EM structure, this region forms an α-helix (see Figure 4). In agreement with this, the V6H peptide displays the spectrum of a disordered peptide in solution but transitions to an α-helix upon TFE titration (Figure 98 B).

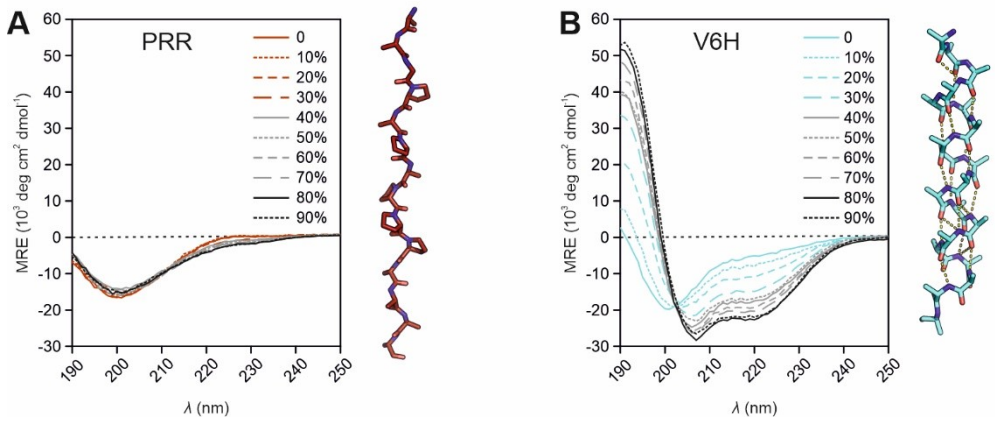


Figure 98: CD spectroscopic secondary structure analysis of the TRPV4 PRR. Far-UV CD spectra of the (**A**) TRPV4 PRR and the (**B**) TRPV6 N-terminal helix (V6H) at increasing TFE concentrations. Plotted is the mean residue ellipticity, MRE, versus the wavelength, λ. The absence of TFE-induced secondary-structure formation in the PRR strongly indicates a PPII conformation. PPII conformations are characterized by left-handed helices which are not held together by hydrogen bonds. The V6H peptide transitions from a disordered state to an α-helix. PyMOL generated models of the PRR in a PPII conformation and the V6H helix from the TRPV6 cryo-EM model are shown next to the spectra.

The observation that the PRR refrains from folding into α-helical structures in the presence of TFE is a good indicator for a PPII conformation (*466*). However, the PRR might also be merely disordered, and folding is inhibited by the high proline content and the lack of amide hydrogen atoms required for secondary structure formation. A PPII helix forms when sequential residues adopt *trans* peptide bonds with dihedral angles of $\phi = -75°$ and $\psi = 145°$ (*467*). Dihedral angles of the TRPV4 PRR have been partially derived from the H_N, H_α, C_α and C' chemical shifts of the TRPV4 IDR via TALOS+ in section VI3.1 (see Figure 46). However, the

backbone resonances of several proline residues in the PRR could not be entirely assigned in the assignment process of the IDR, which results in incomplete and potentially inaccurate dihedral angle predictions. Proline residues are not directly detected in triple resonance experiments due to the lack of amide protons. The assignments of proline H_α, C_α, and C' shifts are usually achieved indirectly via the assignment of $H_{\alpha-1}$, $C_{\alpha-1}$, and C'_{-1} resonances of the proline successor. However, this is not possible for P128 and P129 in the ^{128}PPP^{130} triple proline motif of the PRR because both have a proline successor. Only the P130 resonances could be obtained from $H_{\alpha,-1}$, $C_{\alpha,-1}$, $C_{\beta,-1}$ and C'_{-1} of V131.

To obtain the H_α and C_α resonances of P128 and P129 in the triple proline motif, an NMR backbone assignment of the ^{13}C,^{15}N-labeled PBD-PRR was carried out using HNCACB, HNCACO, HAHBCONH triple resonance experiments (Figure 99 A). Importantly, high-quality spectra could be obtained at pH 6 compared to pH 4.5 used for the complete IDR. The P117, P124, P126 and P130 H_α, C_α, and C' resonances could be assigned indirectly via neighbouring residues. The P128 H_α and C_α chemical shifts were obtained from the ^{1}H, ^{13}C-HSQC spectrum of a selectively ^{13}C-P128-labelled PRR peptide. The P129 H_α and C_α resonances could be identified in the ^{1}H, ^{13}C-HSQC spectrum of ^{13}C,^{15}N-PBD-PRR by procedure of exclusion (Figure 99 B). The assigned H_N, N, H_α, C_α, and C' resonances were used for a TALOS+ backbone torsion angle prediction of the PXPXPPP motif of the PRR (residues P124-P130 in the *G. gallus* TRPV4 sequence). All proline residues are predicted to adopt backbone dihedral angles compatible with PPII structures, i.e. close to $\phi = -75°$ and $\psi = 145°$ (Figure 99 C). Also, the torsion angles of non-proline residues in the PRR, such as A125 and N127, are predicted to be close to the dihedral angles required for PPII structures. This strongly suggests that the PAP(N/Q)PPP consensus motif of the PRR adopts PPII or PPII-like conformations in solution.

A remarkable feature of proline compared to other amino acids is its high energy barrier (E_a ~ 10-20 kcal/mol) and therefore slow interconversion rate between the *trans* and the *cis* peptide bond conformation (*468*). Under physiological conditions, the *trans* conformer ($\omega = 0°$) is energetically favoured over the *cis* conformer ($\omega = 180°$) (Figure 99 D). In the PRR, the interconversion between *cis* and *trans* peptide bonds is sufficiently slow such that residues with proline neighbours like G123, A125, or A127 display two peaks in the PBD-PRR ^{1}H, ^{15}N HSQC spectrum (Figure 99 E). The more intense peak corresponds to the residue with a *trans* proline neighbour, whereas the less intense peak corresponds to the same residue with a *cis* proline neighbour. Integration of the Gly123, Ala125, or Asn127 peaks reveals an equilibrium between approximately 20% *cis* conformer and 80% *trans*. This is in good agreement with published data of *cis-trans* equilibria in proline-containing peptides (*469, 470*). Notably, the PRR was observed to undergo fast backbone fluctuations on the ns-ps timescale in hetNOE measurements (see Figure 46, section VI3.1). This is a bit surprising considering the rigidity of the proline peptide bond and slow interconversion rates between *trans* and *cis*. The measured fast dynamics, however, agree well with previous single molecule FRET studies on polyproline peptides (*471*). The fast backbone fluctuations in the PRR and the presence of *cis-trans* peptide bond equilibria indicate that the PRR adopts a continuum of interconverting conformations which differ by their *cis* and *trans* peptide bond content. A sequence of *cis* peptide bonds results in righthanded polyproline I helices (*467*). But because the *cis* conformation is energetically unfavored, a PPI helix conformation likely represents only a very small subpopulation in the conformational ensemble of proline-rich region (*468, 471*). The lefthanded PPII helix is much more likely to occur as it is formed when all bonds are in the energetically more favourable *trans* conformation (*472*). Therefore, it might represent the preferred, but importantly, not the only structure of the PRR in solution. Assuming that the peptide bonds of all proline residues in the PRR are 80% in *trans* and 20% in *cis* conformers, only ~30% of all PRR conformers are a continuous PPII helix.

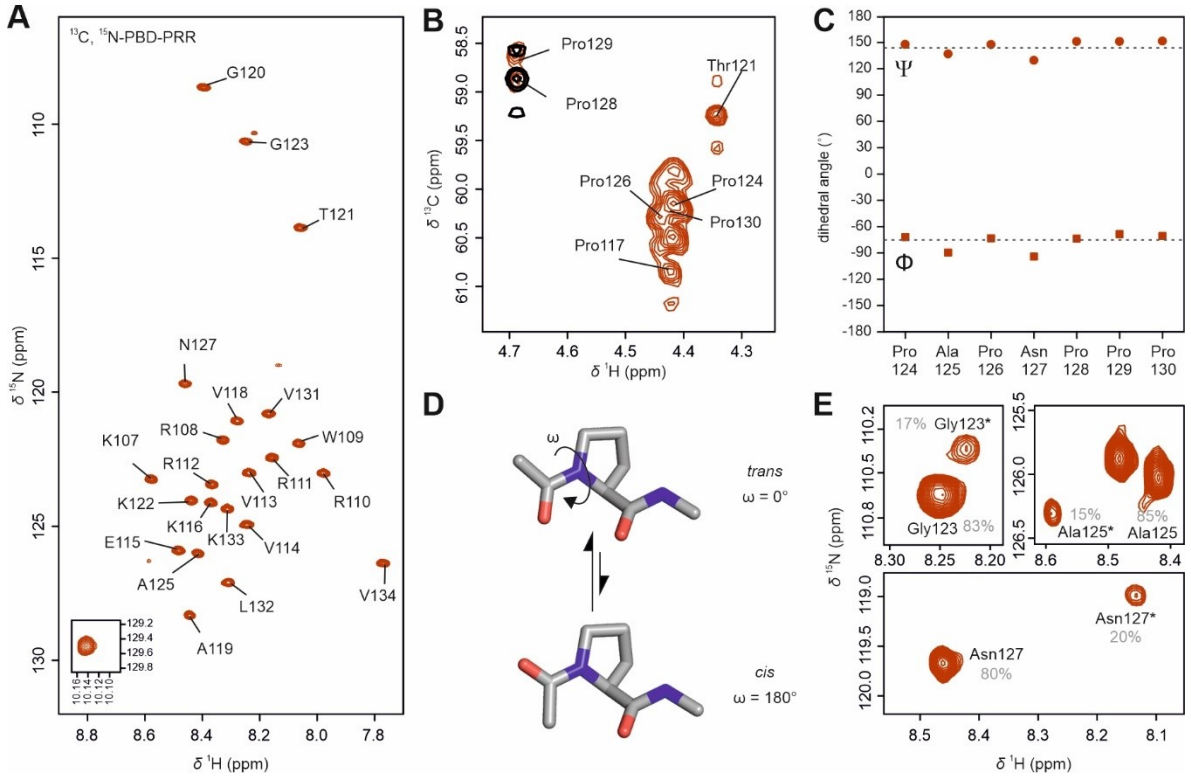


Figure 99: NMR spectroscopic analysis of proline helix conformations in the TRPV4 PRR. (**A**) ^{1}H, ^{15}N-HSQC spectrum of 250 µM ^{13}C, ^{15}N-PBD-PRR with backbone NMR resonance assignments recorded at 293 K in 50 mM KP_i pH 6, 500 mM NaCl, 10% D_2O. (**B**) Section of the ^{1}H, ^{13}C-HSQC spectrum of ^{13}C, ^{15}N-PBD-PRR (red) and the selectively ^{13}C-P128-labeled peptide (black) showing the proline H_α/C_α cross-peaks in the PRR. (**C**) TALOS+ predicted backbone torsion angles ϕ and ψ in the PAPNPPP consensus motif of the PRR. The ϕ and ψ angles in a PPII helix are indicated with dashed lines. (**D**) *Cis-trans* isomerization of a peptidyl-prolyl bond in acetyl-prolyl-N-methyl amide. (**E**) Section of the ^{13}C, ^{15}N-PBD-PRR ^{1}H, ^{15}N-HSQC spectrum showing the G123, A125, and N127 resonances. The residues display two peaks in the HSQC spectrum, one corresponding to a residue with a proline neighbour in a *trans* configuration and one residue with a proline neighbour in a *cis* conformation (indicated by an asterisk). The percentage of *trans* and *cis* populations are indicated in light grey.

To resolve the details of how the TRPV4 PRR binds to the Pacsin SH3 domains, a solution NMR structure of the *G. gallus* Pacsin3 SH3 domain in complex with the isolated TRPV4 PRR was determined (Table 9, Figure 100, see section VI2.17 for technical details). Of note, a commercially obtained *G. gallus* TRPV4 PRR peptide with a V131I mutation was used for structure determination. The *H. sapiens* TRPV4 carries isoleucine in the V131 position, and the valine to isoleucine substitution substantially facilitated side chain assignments for the PRR prior to structure determination. Importantly, titration of the PRR V131I peptide to the ^{15}N-Pac3 SH3 produced almost identical chemical shift perturbations (CSPs) in the ^{1}H, ^{15}N-HSQC spectrum as the wt-PRR peptide (Figure 100 A). This verifies that the V131I mutation does not influence the complex formation and is suitable for structure determination. Deviations in the CSPs occur mainly at pH-sensitive residues and can be attributed to residual trifluoroacetate (TFA) in the synthetic PRR V131I peptide sample. The structure of the Pacsin3 SH3 domain shows the typical five-stranded β-barrel SH3 domain fold (Figure 100 B and C). The PRR substrate is wedged between the Src and RT loops and stretches along β4 and β5. The long RT loop acts as the 'lid' for recognition and binding of the TRPV4 proline-rich region.

Table 9: Structural statistics for the NMR solution structure of the *G. gallus* Pacsin3/TRPV4 PRR complex.

Conformational restricting restraints	
Total NOE distance restraints	948
intra-residue $\|i = j\|$	98
sequential $\|i - j\| = 1$	230
medium-range $1 < \|i - j\| < 5$	143
long-range $\|i - j\| \geq 5$	477
Dihedral angle restraints (Talos-N)	108
No. of restraint per residue	13.5
No. of long-range restraints per residue	6.1
Residual restraint violations[a]	
Average no. of distance violations per structure	
0.1-0.2 Å	3.8
0.2-0.5 Å	0.3
>0.5 Å	1.5
Average no. of dihedral angle violations per structure	
1-10°	7.8
>10°	0
Model quality (ordered residues)[a]	
RMSD backbone atoms (Å)	0.4
RMSD heavy atoms (Å)	0.9
RMSD bond lengths (Å)	0.01
RMSD bond angles (°)	2.3
MolProbity Ramachandran statistics[a]	
Most favored regions	93.5
Allowed regions	5.5
Disallowed regions	1.1
Global quality scores (raw/Z score)[a]	
Verify3D	0.23/-3.69
Prosall	0.17/-1.99
PROCHECK (φ-ψ)	-0.76/-2.68
PROCHECK (all)	-0.89/-5.26
MolProbity clash score	2.80/0.68
Model contents	
Ordered residue ranges (HetNOE > 0.6)	6-62, 103-111
Total no. of residues	82
BMRB accession number	34211
PDB ID code	6f55

[a] calculated using PSVS 1.5 for using ordered residues (HetNOE > 0.6) (Bhattacharya et al., 2007). Average distance violations were calculated using the sum over r^{-6}

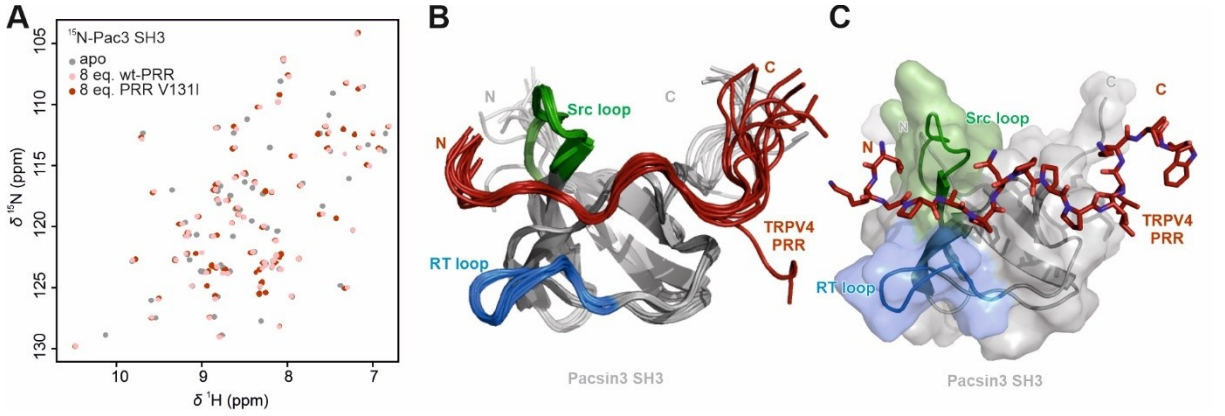


Figure 100: Solution NMR structure of the *G. gallus* Pacsin3 SH3 domain in complex with the *G. gallus* TRPV4 PRR (PDB: 6F55). (**A**) Overlay of ^{1}H, ^{15}N-HSQC spectra of 250 µM ^{15}N-Pac3 SH3 domain in the apo state (grey), bound to the wt-PRR (light pink), and bound to the PRR V131I peptide (red). (**B**) Overlay of the ten lowest energy SH3-PRR conformers calculated from NOE-based distance restraints and (**C**) average structure of the SH3-PRR complex. The Pacsin3 SH3 domain is shown in grey with the RT loop highlighted in blue and the Src loop in green. The TRPV4 PRR is shown in red.

The structure of the complex reveals that the TRPV4-PRR binds between the PACSIN3 SH3 Src and RT-loops in a class I-like binding mode (Figure 101 A). K122 in the PRR interacts with the conserved E394 residue forming the specificity pocket in the RT loop of the SH3 domain. P126 in the PRR occupies the first proline pocket of the SH3 domain formed by W414 and P428. Strikingly, the second SH3 domain proline pocket formed by Y385, Y387, and Y431 does not bind PRR residue P129 as would be expected for a canonical class I binding mode (compare Figure 97 C). Instead, P130 occupies PP2. Remarkably, the PRR is bound to the SH3 domain with the peptide bond between N127 and P128 in an energetically unfavourable *cis* conformation. The *cis* peptide bond configuration, in fact, seems to be a requirement for PRR binding in a class I-like orientation to the SP, PP1, and PP2 simultaneously. Modelling the N127-P128 bond in the SH3-PRR complex in a *trans* configuration either bends the PRR chain away from PP2 (Figure 101 B) or results in a steric clash in the SP and PP1 (Figure 101 C). In complex with the Pacsin3 SH3 domain, residues T121 to N127 in the TRPV4 PRR adopt a left-handed PPII helix conformation with *trans* configured peptide bonds. At P128, the N127-P128 *cis* peptide bond introduces a kink in the PRR backbone and breaks the PPII helix conformation. P128 is followed by P129, P130, both having *trans* peptide bonds configurations. This triple-proline motif is conserved across TRPV4 from various species (Figure 85). The skipped residue and *cis* proline peptide bond leading to a perturbed PPII helix conformation in the Pacsin3 SH3/TRPV4 PRR complex is a so far undescribed binding mode of an SH3 domain to a proline-rich substrate. Considering the almost identical chemical shift perturbation patterns of the Pacsin1-3 SH3 domains (see section VI3.5), the Pacsin3 SH3/TRPV4 PRR complex structure can be regarded as representative for the TRPV4 PRR interactions with all Pacsin SH3 domains.

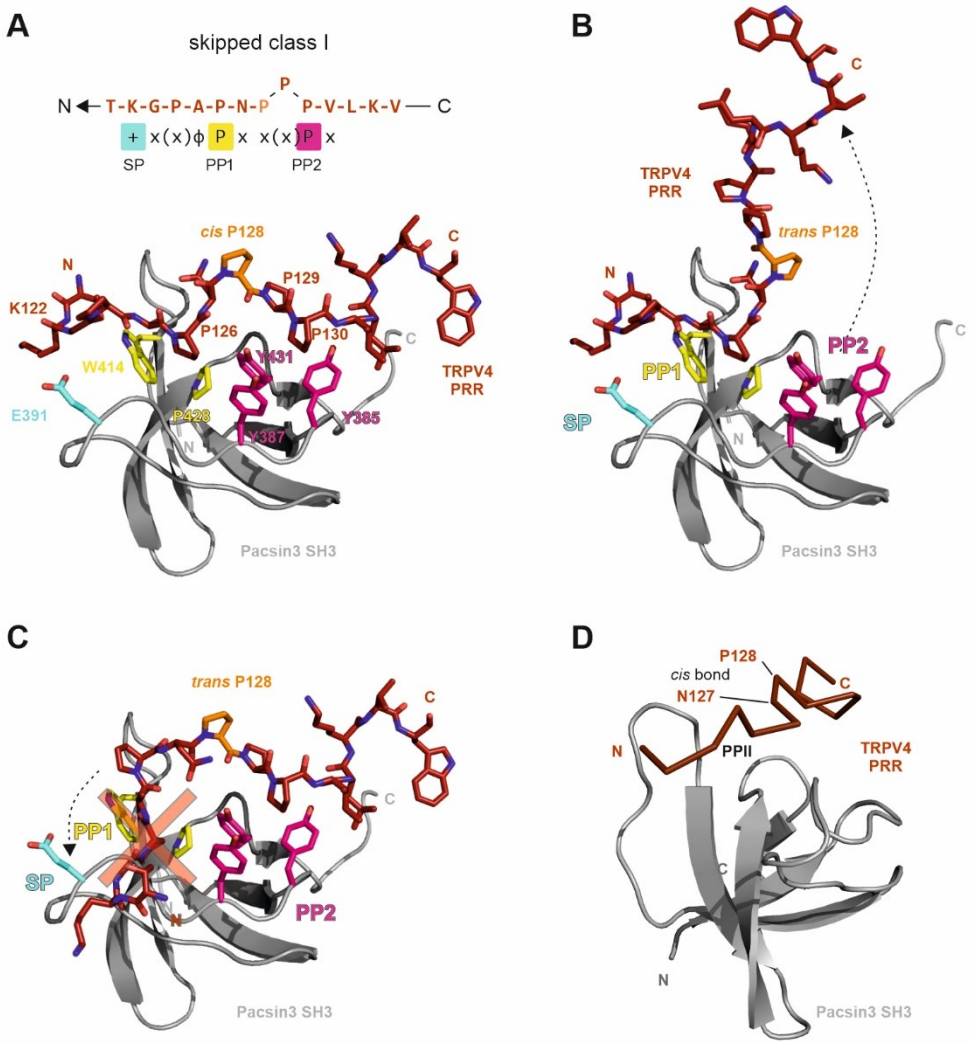


Figure 101: Implications of the P128 *cis* peptide bond conformation in the TRPV4 PRR on the binding to the Pacsin3 SH3 domain. (**A**) Skipped class I binding mode in the complex of the Pacsin3 SH3 domain with the TRPV4 PRR. Binding of PRR residue K122 to the SH3 domain specificity pocket (SP) defines a class I orientation. The PRR residue P126 occupies the first proline pocket (PP1) in the SH3 domain. According to a canonical class I interaction, P129 would occupy the second proline pocket (PP2). Instead, P130 moves into PP2. Importantly, the N127-P128 peptide bond of the PRR is required in a *cis* conformation to maintain contact of the PRR with the SH3 domain substrate binding site. (**B** and **C**) PyMOL generated models show that the N127-P128 peptide bond in a *trans* conformation is incompatible with PRR binding to the SP, PP1, and PP2 in the Pacsin3 SH3 domain simultaneously. With the PRR bound to the SP and PP1, a *trans*-N127-P128 conformation would tilt residues P129-V134 away from the SH3 domain, thus leaving PP2 empty. Conversely, binding of the PRR to PP2 with a *trans* N127-P128 peptide bond conformation results in a steric clash at residues preceding P127. (**D**) The *cis* N127-P128 peptide bond introduces a kink in the PRR backbone and breaks the PPII helix conformation in the SH3-PRR complex.

In isolation, the TRPV4 PRR predominantly adopts a PPII conformation in which all peptide bonds are in a *trans* conformation. The *trans* conformation in P128 is, however, incompatible with PRR binding to the Pacsin SH3 domain. To occupy the SP, PP1, and PP2 simultaneously, P128 needs to be in a *cis* conformation (Figure 102 A, B and C). The presence of stable PRR subpopulations with proline residues in the thermodynamically unfavourable *cis* peptide bond conformation is evidenced by the PBD-PRR ^{1}H, ^{15}N-HSQC spectrum. Because the ^{1}H, ^{15}N-HSQC analysis of the PBD-PRR only indirectly reports on *trans-cis* equilibria of peptide bonds, a site-specifically labelled ^{13}C-P128 PRR was used to test whether P128 *cis* conformations occur in solution. In the ^{1}H, ^{13}C-HSQC spectrum of ^{13}C-P128 PRR, the chemical shifts for the H_β/C_β and the H_γ/C_γ pairs for the P128 peptide bond in either the *trans* or the *cis* conformation can be distinguished clearly (Figure 102 D) (*473,*

474). In the 1D projections of the *cis* and *trans* H_γ/C_γ peaks, the population differences between both conformers show a clear preference for the *trans* configuration in the free PR (Figure 102 E). Nevertheless, subpopulations with *cis* P128 bonds can be detected. Upon addition of the Pacsin1-3 SH3 domains, the *cis* population becomes dominant (Figure 102 E). Thus, in agreement with the NMR structure of the Pacsin3 SH3/TRPV4 PRR complex, the Pacsin SH3 domains bind the PRR with the P128 peptide bond in a *cis* conformation.

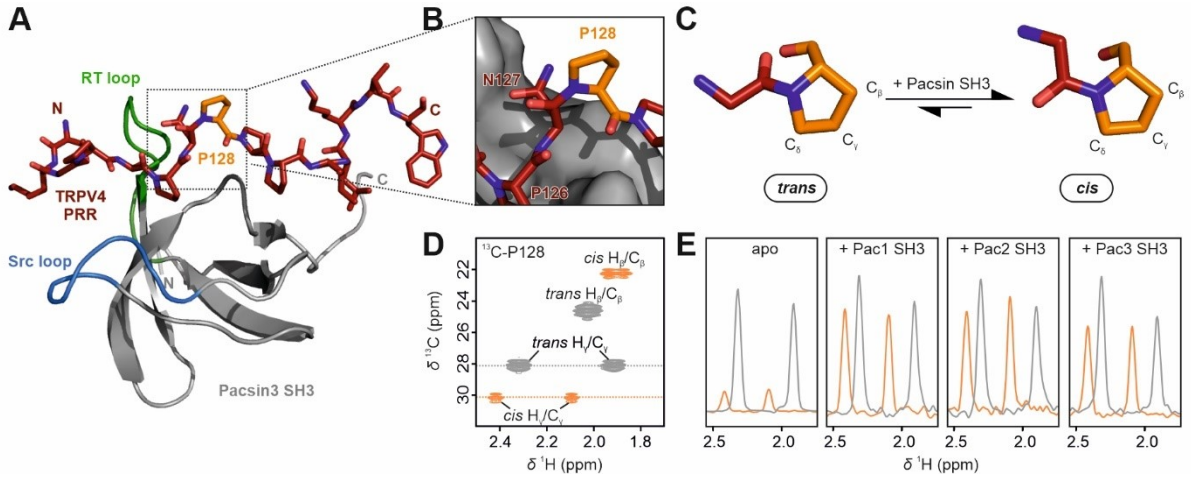


Figure 102: P128 peptide bond conformational changes between the free and the SH3 domain bound TRPV4 PRR. (**A**) Solution NMR structure of the Pacsin3 SH3 domain in complex with the TRPV4 PRR. Proline P128 in the cis conformation is shown in orange. (**B**) Expanded view of the *cis* peptide bond between N127 and P128 in the TRPV4 PRR/Pacsin3 SH3 complex structure. (**C**) Equilibrium between P128 in the *trans* and the *cis* conformation. Pacsin SH3 domain binding stabilizes the *cis* conformation. (**D**) Section of the ^{1}H, ^{13}C-HSQC spectrum of site-specifically labelled ^{13}C-P128 PRR in isolation (unbound) showing the H_β/C_β and the H_γ/C_γ pairs of the P128 *trans* (grey) or the *cis* (orange) conformation, respectively. (**E**) 1D projections of the *trans* (grey) and the *cis* (orange) H_γ resonances.

Interestingly, P128 was previously shown to be essential for the interaction of TRPV4 with Pacsin in co-immunoprecipitation assays (*76*). Mutation of P128 to alanine in combination with a P129L in *H. sapiens* TRPV4 abrogated co-immunoprecipitation with Pacsin1-3 and the stimulus-specific desensitization of TRPV4 by Pacsin3 (*76, 77*). To probe whether residue P128 is essential for the interaction between the TRPV4 PRR and the Pacsin SH3 domains, CSP experiments of the ^{15}N-Pac3 SH3 domain were carried out with wt-PBD-PRR, PBD-PRR P128A, and PBD-PRR P128A/P129L peptides as ligands (Figure 103 A and B). Surprisingly, neither the P128A single mutant nor the P128A/P129L double mutant fully abolished the interaction with the Pacsin3 SH3 domain, but in both cases, reduced the binding affinity by a factor of ~2 compared to the wt-PBD-PRR peptide (Figure 103 C). The P128A mutant produced a very similar chemical shift perturbation pattern in the ^{15}N-Pac3 SH3 domain HSQC spectrum but with overall lower ΔCS values of RT loop and β4 and β5 resonances compared to the wt-PBD-PRR peptide (Figure 103 B). The P128A/P128L double mutant, in contrast, severely reduced the resonance shifts of residues in the RT loop and in the β4 and β5 region.

In the Pac3 SH3 domain/TRPV4 PRR complex structure (Figure 101 A), P128 and P129 of the PRR do not occupy the binding pockets of the SH3 domain. Instead, their side chains are oriented away from the SH3 domain. Replacing P128 and P129 by alanine and leucine, respectively, might be tolerated in the SH3/PRR complex because both alanine and leucine can in principle adopt the same backbone angles as P128 and P129, even though *cis* peptide bond conformations are less frequently observed in non-proline residues (*475*). Interestingly, the shift direction of various ^{15}N-Pac3 SH3 resonances in the PRR binding site changes upon titrating the P128A/P129L double mutant compared to the wt-PBD-PRR peptide. This is not only observed for residues directly contacting the mutated positions of the PRR but also more distant, e.g. in the specificity pocket or proline pocket 1. This might suggest, that even though the P128A and P129L mutations do not abolish the

interaction with the SH3 domain, they change the overall SH3-PRR binding mode. Potentially, several binding modes between the Pacsin3 SH3 domain and the PBD-PRR P128A/P129L peptide with similar affinities in the ~150 µM range exist. The HSQC spectrum might therefore represent the ^{15}N-Pac3 SH3 domain chemical shifts averaged over all binding modes which result in overall low ΔCS values between free and PRR bound state.

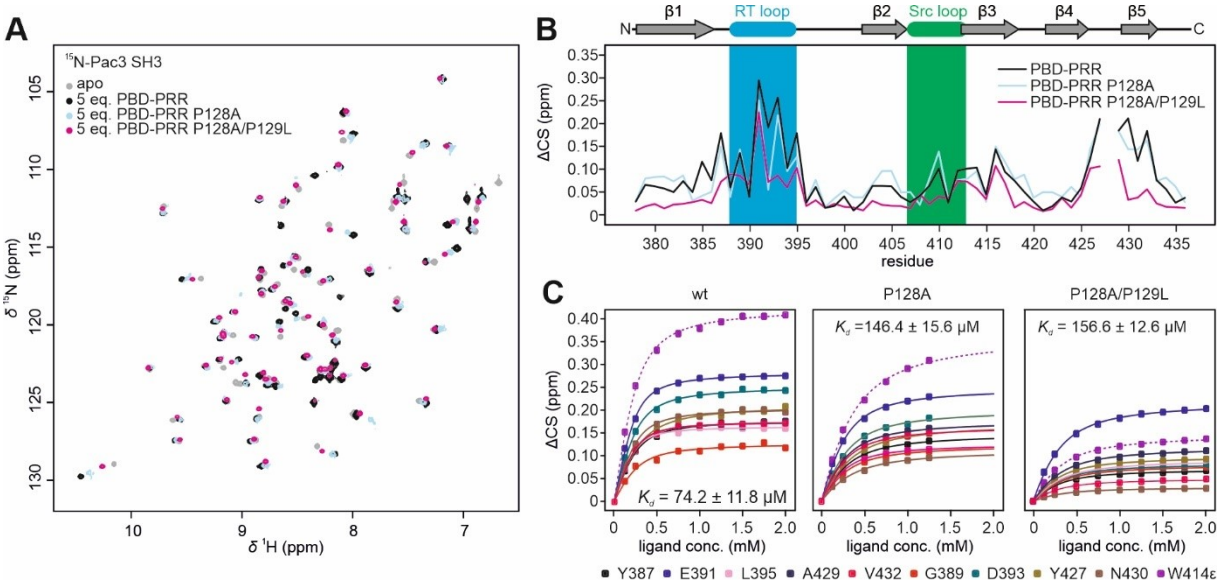


Figure 103: Effect of mutations in the TRPV4 PRR triple-proline motif on complex formation with the Pacsin3 SH3 domain. (**A**) Overlay of ^{1}H, ^{15}N-HSQC spectra of 250 µM ^{15}N-Pac3 SH3 in the apo state (grey) and bound to wt-PBD-PRR (black), PBD-PRR P128A (light blue), and PBD-PRR P128A P129L (magenta). Spectra were recorded at 298 K in 10 mM Tris pH 7, 100 mM NaCl, 10% D_2O. (**B**) Chemical shift changes, ΔCS, of ^{15}N-Pac3 SH3 resonances between apo state and bound to wt-PBD-PRR (black), PBD-PRR P128A (light blue), and PBD-PRR P128A P129L (magenta) plotted versus the residue number. (**C**) Chemical shift changes of ^{15}N-Pac3 SH3 residues plotted versus the peptide concentration. The data were fitted with binding isotherms yielding dissociation constants of K_d = 74.2 ± 11.8 µM for wt-PBD-PRR (left), K_d = 146.4 ± 15.6 µM for PBD-PRR P128A (centre), and K_d = 156.6 ± 12.6 µM for PBD-PRR P128A/P129L. The dissociation constants are averaged over all residues (see section VI2.17).

The structural analysis of the TRPV4 PRR in isolation and in complex with the Pacsin SH3 domain reveals crucial mechanistic details underlying the Pacsin-TRPV4 interaction. In solution, all proline peptide bonds of the PRR exist in a thermodynamic equilibrium between *trans* and *cis*. Because the *trans* conformer is energetically favourable, the TRPV4 PRR preferentially adopts a lefthanded PPII helix conformation, a structure which is commonly recognized as a binding partner by SH3 domains. The PPII helix in the PRR, however, is not compatible with the Pacsin SH3 domain binding site between the RT and the Src loop. In fact, PRR binding to the Pacsin SH3 specificity pocket and proline simultaneously requires P128 to be in a *cis* conformation. The Pacsin SH3 domains likely select PRRs with *cis*-P128 bonds primed for binding from a pool of conformations characterized by different *cis-trans* bond content. In such a scenario, which is commonly referred to as conformational selection (*476*), only a subfraction of the PRR conformational ensemble is available for SH3 domain binding and thereby reduces the affinity for complex formation.

Importantly, P128 isomerization between the *trans* and the *cis* configuration introduces a kink in the IDR backbone and might have large implications for the conformation of the TRPV4 N-terminus. Within the conformational ensemble of the TRPV4 IDR, a *cis*-P128 might define a small subset of conformations in which the channel is desensitized. Accordingly, Pacsin might select and stabilize TRPV4 N-terminus conformations with a *cis*-P128 bond and thereby extend the lifetime of a desensitized state. It is unlikely that the binding of the SH3 domain facilitates the isomerization of P128 from *trans* to *cis* via induced fit. According to the current understanding, protein-mediated *cis-trans* prolyl isomerization requires an own class of enzymes called peptidylprolyl-

cis/trans-isomerases (PPIases) (*477, 478*). The *cis*-P128 conformation in the Pacsin SH3/TRPV4 PRR complex, however, might fuel speculations whether PPIases are involved in the regulation of TRPV4 by Pacsin.

3.7 TRPV4 releases Pacsin3 from its 'clamped' conformation

The molecular mechanism by which an interaction partner regulates the function of a target protein is generally determined by structural details. Therefore, elucidating the structural details of the TRPV4/Pacsin interaction should explain why Pacsin3 acts as a TRPV4 desensitizer (*76, 77*). As outlined in section VI1.3, an autoinhibition mechanism was proposed to regulate Pacsin1 (*303, 307*). According to this mechanism, an association of the Pacsin SH3 domain with the F-BAR domain impedes an F-BAR/membrane interaction and thereby inhibits Pacsin's tubulation activity. In an X-ray crystal structure of Pacsin1 (PDB: 2x3x), the Pacsin1 SH3 domain associates via residues Q396, E397, and E400 in the RT loop with a patch of basic residues (K141, K145, K148) in the F-BAR domain hinge region (Figure *104* A, B, and C). The binding site was confirmed via pull-down experiments in combination with mutagenesis (*307*). The autoinhibitory function of the SH3 domain was concluded from the observation that membrane binding and the lipid tubulation activity of Pacsin1 were enhanced upon deleting the SH3 domain (*303*). Importantly, the SH3 domain substrate-binding site between the RT and Src loop overlaps with the SH3/F-BAR binding interface. Interaction of the Pacsin SH3 domain with a proline-rich substrate thus requires dissociation of the SH3 domain from the F-BAR domain. Once bound to a substrate via the SH3 domain, Pacsin remains in an activated state and can bind with its F-BAR domain to curved membranes or induce membrane curvature. In agreement with this assumption, a peptide constituting the dynamin-PRR was observed to release Pacsin1 from its autoinhibited state and trigger Pacsin1-mediated lipid tubulation (*307, 479*). The structure of the Pacsin SH3 and F-BAR domain is highly conserved within the Pacsin family, including the position of the acidic QEE motif in the SH3 domain and the KKK motif in the F-BAR domain. It is therefore likely that the autoinhibition mechanism elucidated for Pacsin1 is a feature conserved among Pacsin isoforms. Aligning the Pacsin3 SH3/TRPV4 PRR complex structure with the SH3 domain in the Pacsin1 X-ray crystal structure shows that the SH3/PRR interaction interferes with an SH3/F-BAR association (Figure *104* D). The TRPV4 PRR is therefore a potential ligand to release Pacsin proteins from their autoinhibited conformation. Downstream effects of Pacsin activation such as membrane remodeling in the direct vicinity of the TRPV4 TMD might be crucial for the regulation of TRPV4 activity. The downstream effects of Pacsin activation might determine why some Pacsin isoforms regulate TRPV4 for certain stimuli whereas others do not.

For understanding the Pacsin-mediated regulation of TRPV4, it is crucial to identify which factors besides the Pacsin SH3/TRPV4 PRR interaction contribute to the overall complex formation of Pacsins with TRPV4. This not only requires a detailed understanding of the structural dynamics of the Pacsin binding site, the PRR, in the disordered TRPV4 N-terminus and how it is connected to the rest of the TRPV4 channel (see section VI3.1, VI3.2). Moreover, it relies on an in-depth structural characterization of the Pacsin proteins. Whereas this has been done extensively for full-length Pacsin1 and Pacsin2 and the isolated Pacsin1-3 F-BAR domains (*303, 307, 311, 317, 479*), structural investigations on full-length Pacsin3, the only Pacsin isoform currently known to affect TRPV4 function, are missing. Here, an integrated structural biology approach was applied to investigate the structure of the full-length Pacsin3 protein and how it interacts with the TRPV4 N-terminal domain.

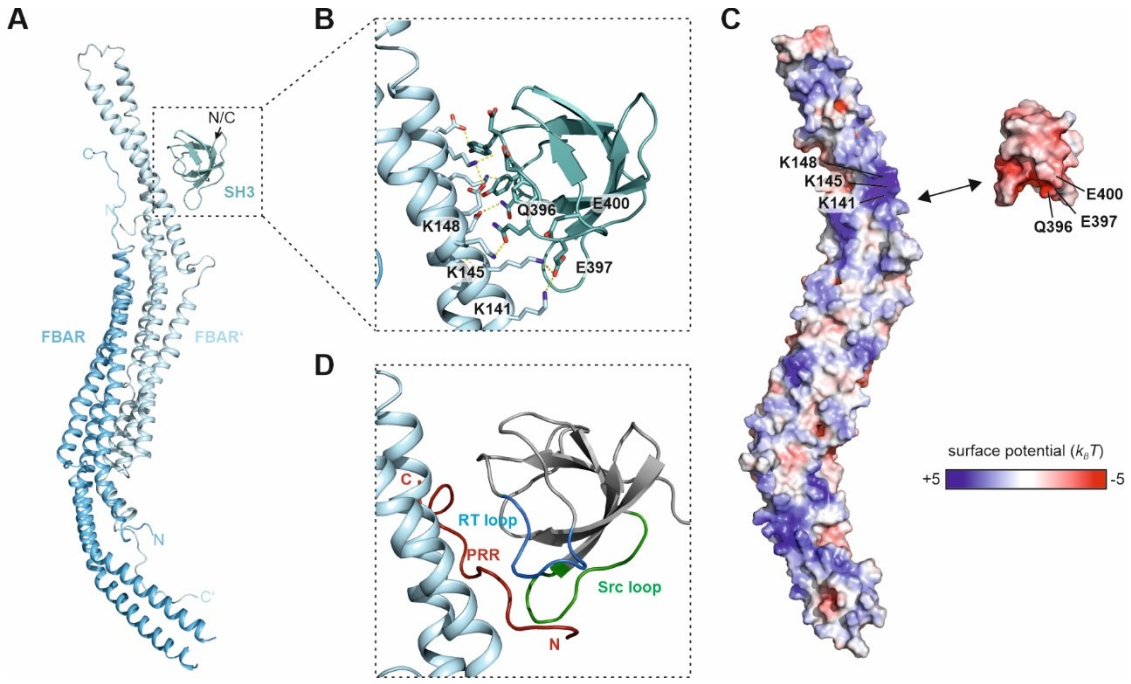


Figure 104: Structural characteristics of the proposed Pacsin SH3/F-BAR domain association interface. (**A**) X-ray crystal structure of *M. musculus* Pacsin1 (PDB: 2x3x). The structure shows the dimeric F-BAR domain and one SH3 domain bound to the hinge region of one F-BAR subunit. The linker region connecting the SH3 domain to the F-BAR domain was not resolved in the structure. (**B**) Close up view of the SH3/F-BAR binding interface observed in the Pacsin1 X-ray crystal structure. Residues K141, K145, K148 in the F-BAR domain form salt bridges to residues Q396, E397, and E400 in the RT loop of the SH3 domain. Mutating these residues was shown to abrogate the autoinhibition of Pacsin1. (**C**) Surface potential representation of the Pacsin1 structure showing the positively charged patch formed by K141, K145, and K148 in the F-BAR domain which mediates an electrostatic interaction with the negatively charged patch consisting of Q396, E397, and E400 in the SH3 domain. (**D**) NMR structure of Pacsin3 SH3/TRPV4 PRR complex (PDB: 6f55) aligned with the SH3 domain in the Pacsin1 X-ray crystal structure shown in (B). The Pacsin1 SH3 domain was removed for better visualization. The alignment demonstrates that the Pacsin SH3/F-BAR binding interface coincides with the SH3 domain substrate binding site. Binding of the TRPV4 PRR to the SH3 domain interferes with the SH3/F-BAR interaction and thus might release Pacsin from a clamped and autoinhibited conformation.

For a structural investigation of Pacsin3, *G. gallus* Pacsin3 constructs (Figure 105 A) comprising the full-length protein sequence (termed fl-Pac3, residues 2-437), the F-BAR domain (F-BAR, residues 2-320), or the linker region with the SH3 domain (termed l-SH3, residues 321-437) were generated via expression in *E. coli* (see section VI2.6). All proteins were expressed as N-terminally tagged His_6SUMO-fusion proteins and purified via Ni-NTA IMAC. The His_6SUMO-tag was cleaved during the purification procedure with the protease Ulp-1, which yielded tag-less and monodisperse Pacsin3 constructs at high purity as demonstrated by SDS-PAGE and SEC (Figure 105 B and C). For estimating the oligomeric state of the constructs, the Stokes radii of the fl-Pac3, F-BAR, and l-SH3 domain were determined from their SEC elution volumes (Figure 105 C) using the calibration curve shown in Figure 40 (section VI3.1). The experimental Stokes radius (R_S^{exp}) of 4.57 nm of the isolated F-BAR domain is close to the theoretical value (R_S^{calc}) of 4.35 nm calculated from the *M. musculus* Pacsin3 F-BAR dimer X-ray crystal structure (PDB: 3qe6) using the HYDROPRO tool (*480*). The slightly increased R_S^{exp} compared to R_S^{calc} might be due to residues contained in the purified F-BAR construct which are missing in the X-ray structure (residues 2-11 and 302-320). The F-BAR domain, therefore, can be assumed to be a dimer, as one would expect from previous SAXS studies on the Pacsin1 F-BAR domain (*303, 479*). The R_S^{exp} of 5.70 nm determined for fl-Pac3 is increased by ~30% compared to the value determined for the F-BAR domain and can be attributed to the increased size of the protein by the additional linker and SH3 domain in each Pacsin3 subunit. It is nearly impossible to correlate the R_S^{exp} value to the oligomeric state of the fl-Pac3 due to the elongated shape of the protein and the lack of structural information of the full-length protein (regarding an R_S calculation with HYDROPRO). Considering that the isolated F-BAR domain is a dimer in solution, the fl-Pac3 is also likely a dimer. The isolated SH3 domain with an R_S^{exp} = 1.6 nm can be unambiguously classified as a monomer (see also section VI3.5). Attaching the linker to the SH3 domain (l-SH3) increases R_S^{exp} to

2.63 nm. This value is close to the R_S^{PMG} of 2.53 nm (premolten-globule state) calculated for l-SH3 according to the scaling laws described in section VI2.8 (*328*), and thus reflects the disordered state of the linker. The isolated SH3 domain and the l-SH3 construct, therefore, are monomers in solution.

To verify that the purified Pacsin3 constructs are properly folded, the proteins' secondary structure was analyzed via far-UV circular dichroism spectroscopy followed by BeStSel analysis (Figure 105 D and E) (*339*). The fl-Pac3 and F-BAR CD spectra (Figure 105 D) display the typical features of α-helical proteins, i.e. two minima at 208 and 222 nm. The BeStSel analysis of fl-Pac3 reveals a secondary structure content of approximately 50% α-helix, 7% β-sheet, and 43% random coil (Figure 105 E). This agrees well with the secondary structure found in high-resolution structures of Pacsin3 domains deposited in the PDB (PDB: 3qe6 for F-BAR, PDB: 6F55 for SH3, the linker is assumed to be entirely disordered). The BeStSel analysis of the isolated F-BAR domain yields an increased α-helix content (~65%) slightly decreased content of random coil (33%) compared to fl-Pac3. Almost no β-sheet structures are predicted (<2%). This is consistent with the secondary structure content contained in the *M. musculus* Pacsin3 F-BAR domain X-ray crystal structure (PDB: 3qe6) (*317*). The CD spectrum of l-SH3 resembles that of a disordered protein, with a minimum at ~197 nm (Figure 105 D). This is likely a result of the disordered linker, which makes up for approximately 50% of the l-SH3 sequence, and because the isolated SH3 domain alone shows similar features as a disordered protein. BeStSel analysis of the l-SH3 CD spectrum did not yield a reasonable fit, similar to the isolated SH3 domain (see section VI3.5, Figure 86). However, in combination with the SEC analysis (Figure 105 C), it can be concluded that the SH3 domain in the l-SH3 construct is folded and that the linker region is disordered.

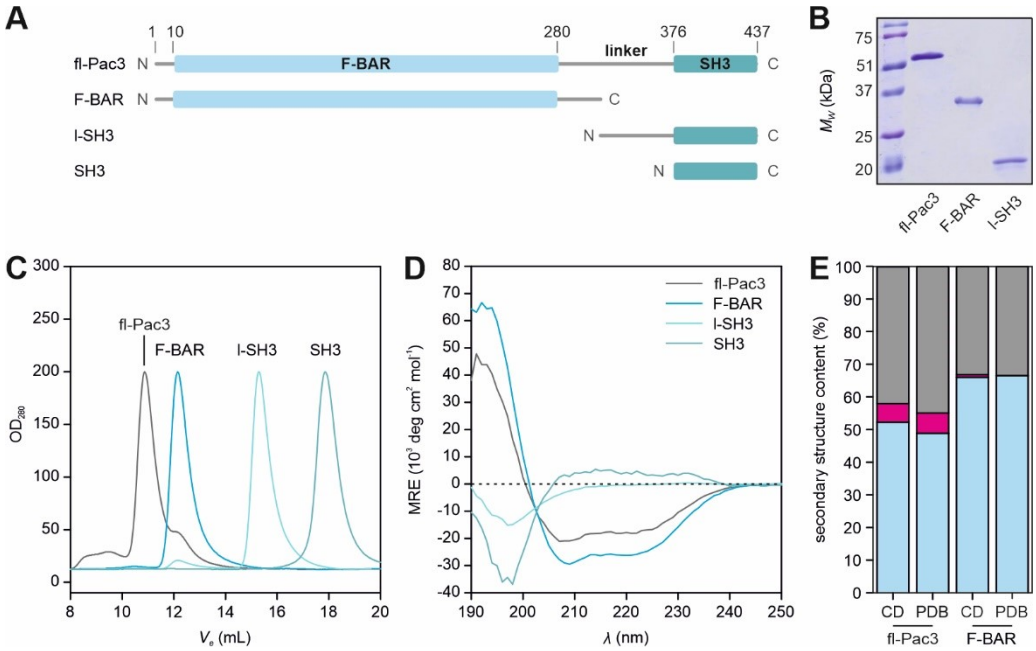


Figure 105: Purification and secondary structure analysis of *G. gallus* Pacsin3 constructs. (**A**) Topology and (**B**) SDS-PAGE analysis of full-length Pacsin3 (fl-Pac3), Pacsin3 F-BAR domain (F-BAR), Pacsin3 flexible linker with SH3 domain (l-SH3), and Pacsin3 SH3 domain (SH3). See Figure 86 for SDS-PAGE of SH3. (**C**) SEC traces and (**D**) far-UV CD spectra of Pacsin3 constructs. SEC experiments were carried out in 20 mM Tris pH 7, 100 mM NaCl, 1 mM DTT. CD spectra were measured in 2 mM Tris pH 7, 5 mM NaCl. (**D**) BeStSel (*339*) analysis of fl-Pac3 and F-BAR ('CD') and comparison with the secondary structure content of available Pacsin3 data deposited in the PDB ('PDB'). Blue: α-helix, magenta: β-sheet, grey: random coil. The 'PBD' data of F-BAR were calculated from the published *M. musculus* Pacsin3 F-BAR domain X-ray crystal structure (PDB: 3qe6). For the 'PDB' data of fl-Pac3, the secondary structure content accounting for the SH3 domain was derived from the NMR structure of the Pacsin3 SH3/TRPV4 PRR complex (PDB:6F55). Residues accounting for the linker region connecting the SH3 and F-BAR domain were assumed to be entirely disordered.

In an effort to obtain a high-resolution structure of the full-length *G. gallus* Pacsin3 protein and of the isolated Pacsin3 F-BAR domain, the constructs fl-Pac3 and F-BAR were subjected to crystallization using the sitting

drop technique (see VI2.19). Crystals from screenings with both fl-Pac3 and F-BAR were obtained after two weeks and were used for X-ray diffraction data collection yielding several data sets with resolutions between 2.1-3.0 Å. Electron density maps were obtained via molecular replacement using the *M. musculus* Pacsin3 X-ray crystal structure (PDB: 3qe6) as a search model. Unfortunately, in the electron density maps obtained from crystals grown with either fl-Pac3 or F-BAR, only the F-BAR domain was resolved (Figure 106 A). No electron density accounting for the linker region or the SH3 domain could be identified. The best data set collected from crystals obtained with the fl-Pac3 construct diffracted to a resolution of 2.1 Å and was used for model building and refinement (Figure 106, Table 10, see section VI2.19 for experimental details).

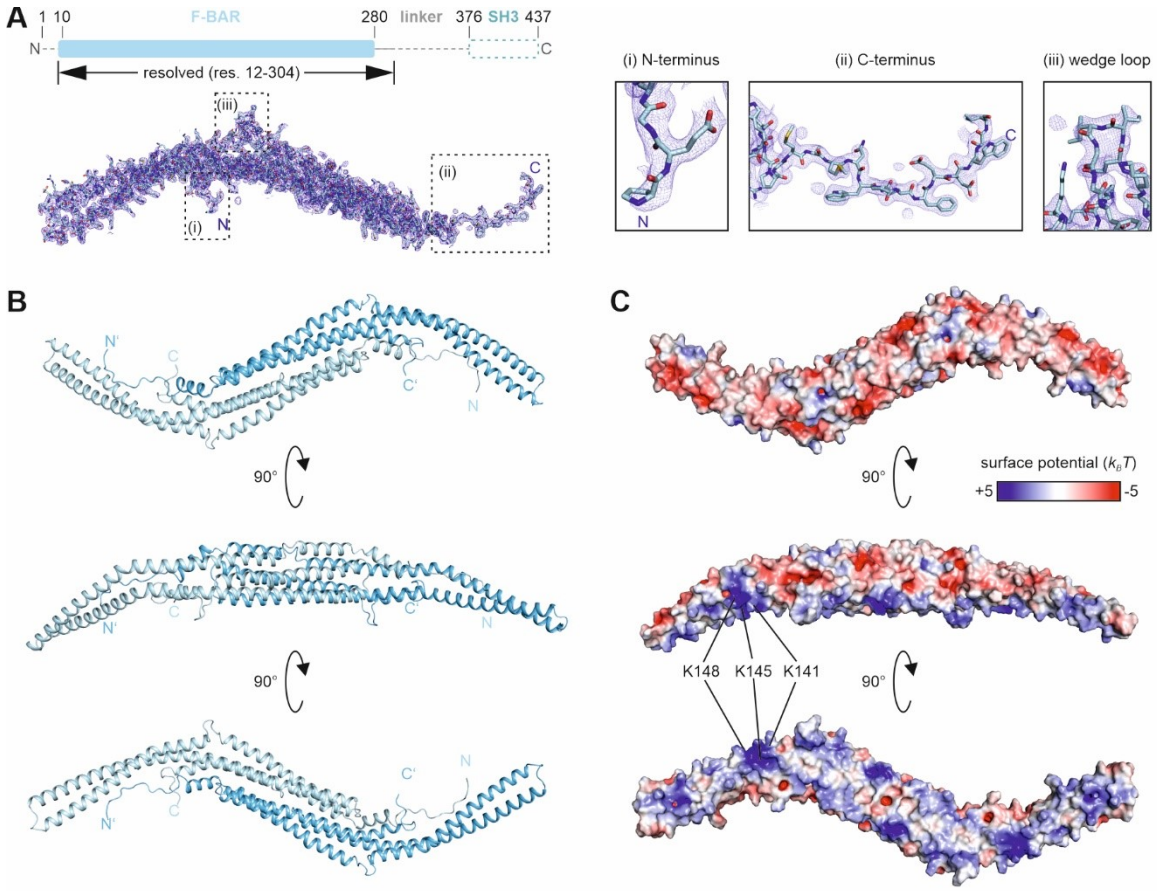


Figure 106: X-ray crystal structure of the *G. gallus* Pacsin3 F-BAR domain. (**A**) $2F_o$-F_c electron density map (at 3σ, shown as a blue mesh) with the modelled structure of a Pacsin3 F-BAR monomer. X-ray diffraction data obtained from crystals grown with the fl-Pac3 protein yielded a high resolution (2.1 Å) electron density map in which only the Pacsin3 F-BAR domain was resolved (residues 12-304). Magnified views of the N-terminal and the C-terminal residues modelled into the $2F_o$-F_c electron density map are shown in (i) and (ii), respectively. (iii) shows a magnified view of the modelled wedge loop. (**B**) Cartoon representation of the dimeric *G. gallus* Pacsin3 F-BAR domain. (**C**) The surface potential representation highlights the accumulation of positively charged amino acid side chains on the concave surface of a *G. gallus* Pacsin3 F-BAR dimer. The residues K141, K145, and K148 which were identified as the SH3 domain binding site in the full-length *M musculus* Pacsin1 X-ray crystal structure (PBD: 2x2w) are conserved in the *G. gallus* Pacsin3 F-BAR domain and form a cluster of positive charges in the F-BAR domain hinge region. The convex surface of the F-BAR dimer is enriched in negatively charged amino acid side chains. Surface potentials were calculated using the APBS tool (*456*).

Based on the electron density map obtained from the diffraction data of fl-Pac3, a model of the F-BAR domain encompassing residues 12-304 could be generated (Figure 106 A, B, C). No electron density was observed for the 10 N-terminal residues (res. 2-11) or the linker region with the SH3 domain (res. 305-437), even though they were contained in the fl-Pac3 construct used for crystallization screenings (Figure 106 A). Notably, the flexible wedge loop was well resolved (Figure 106 A). The wedge loop was shown to be the main structural

feature which enables F-BAR domains to bend membranes by inserting into the membrane bilayer upon F-BAR/membrane interaction (*303, 317*). The *G. gallus* Pacsin3 F-BAR domain displays the typical crescent (side view) and sigmoidal (top view) shape observed for other Pacsin F-BAR domains (Figure 106 B). The concave surface of the F-BAR domain is enriched in basic residues which mediate electrostatic interactions with anionic phospholipids in the plasma membrane(Figure 106 C). The convex surface is enriched in negatively charged side chains.

Table 10: Structural statistics for *G. gallus* Pacsin3 F-BAR domain X-ray crystal structure.

Data collection	
Wavelength (Å)	0.97626 Å
Space group	P32 2 1
Unit cell parameters	
a, *b*, *c* (Å)	73.085, 73.085, 232.629
α, β, γ (°)	90, 90, 120
Resolution range (overall/inner shell/outer shell)	63.6-2.3 Å/63.6-8.0 Å/2.3-2.1 Å
R_{meas} (overall/inner shell/outer shell)	0.143/0.060/1.266
Total number of reflections	134064/8277/5196
Number of unique reflections	18173/908/910
Completeness (overall/inner shell/outer shell)	73.8/99.9/53.3
Multiplicity (overall/inner shell/outer shell)	7.4/9.1/5.7
$CC_{1/2}$ (overall/inner shell/outer shell)	0.998
$[I/\sigma]$	1.20
Refinement	
Resolution range	2.1 Å
R_{work}/R_{free}	0.212/0.254
Number of atoms	
Total	4860
Ligand	10
Waters	94
B-factor (Å^2)	
Protein	47.92
Ligand	74.00
Waters	44.51
Model quality (ordered residues)[a]	
RMSD bond lengths (Å)	0.014
RMSD bond angles (°)	2.0
MolProbity Ramachandran statistics[a]	
Most favoured regions	98.6%
Allowed regions	1.4%
Disallowed regions	0%
Global quality scores (raw/Z score)[a]	
Verify3D	0.39/2.01
Prosall	0.80/0.62
PROCHECK (φ-ψ)	0.43/-3.85
PROCHECK (all)	0.16/0.95
MolProbity clash score	0.69/1.41
Model contents	
α-helices	23A-70A, 75A-103A, 105A-117A, 127A-174A, 182A-253A, 259A-273A, 277A-288A
β-strands	none
Total no. of residues	293
PDB ID code	To be submitted

[a] calculated using PSVS 1.5 (*389*)

The *G. gallus* Pacsin3 F-BAR domain aligns well (heavy atom RMSD = 0.51 Å) with the *M. musculus* Pacsin3 F-BAR domain structure (PDB: 3qe6) (*317*) (Figure 107 A). The *M. musculus* and the *G. gallus* F-BAR domains display deviations between the helices α4 and α5, which might reflect a certain degree of flexibility in this region. This is consistent with recent computational studies which demonstrated substantial structural flexibility of the Pacsin1 F-BAR domain in the hinge regions and the tips (*481*). Intriguingly, the chicken PACSIN3 F-BAR domain wedge loop does not contain the proline residue that was associated with enhanced wedge loop rigidity and subsequently with differences in lipid tubulation between PACSIN1/2 and PACSIN3 (Figure 107 B). Instead, it contains the glutamine residue associated with higher wedge loop flexibility in mouse Pacsin1 and Pacsin2 (*317*). The second stabilizing factor of the wedge loop, a salt bridge between the loop's histidine and glutamate residue (H121 and E130 in the *G. gallus* Pacsin3 sequence) are conserved. Alignment of the wedge loops (backbone RMSD = 0.623 Å) shows that despite the absence of the proline residue in the chicken protein, the two loops are highly similar.

Inspection of the crystal packing shows that contacts between F-BAR dimers are mediated by interactions of the wedge loop of one dimer with the α4 and α5 helices of another dimer (Figure 107 C). This results in a C_2 symmetric contact site with the hinge region in the centre. The hinge region with lysine residues K141, K145, and K148 constitutes the site expected to interact with the SH3 domain in full-length Pacsin proteins based on what was observed in the F-BAR/SH3 X-ray crystal structure of *M. musculus* Pacsin1 (PDB: 2x3w) (*307*). The crystal contacts seem to compete with the SH3/F-BAR interaction and potentially leave the SH3 domain attached to the linker tumbling in the water-filled space of the F-BAR domain crystal lattice. The packing of F-BAR domains has previously been linked to their membrane bending modes and the type of membrane deformation they produce (*303, 311, 317, 482*). In this context, tip-to-tip interactions between F-BAR dimers and parallel alignments of F-BAR dimers mediated by the wedge loops were proposed to mediate the assembly of a tight F-BAR domain coat around lipid tubules (*482*). No such F-BAR dimer contacts are observed in the crystal of the *G. gallus* Pacsin3 F-BAR domain. Instead, the crystal contacts mediated by the wedge loops result in a right-handed helix composed of three F-BAR dimers per turn (~230 Å) with an inner diameter of ~40 Å (Figure 107 C). The inner part of the F-BAR helix is constituted by the concave surface of the F-BAR dimer, i.e. the lipid-binding surface. It is tempting to speculate that Pacsin3-mediated lipid tubulation is facilitated by the Pacsin3 F-BAR domain wrapping around lipid tubules in a helical fashion, thus stabilizing the tubes (Figure 107 C). However, converting a lipid bilayer with an average thickness of ~30 Å to ~40 Å thin tubular shape creates a highly bent membrane surface for which high energies are required. According to the currently accepted mechanism by which F-BAR domains bend membranes, the bending energy is provided by the insertion of the F-BAR wedge loops into the lipid leaflets (*303, 311, 317*). In the helical *G. gallus* Pacsin3 F-BAR assembly of the crystal, the wedge loops mediate inter F-BAR contacts and thus would not be available for membrane insertion. It is unclear whether such an F-BAR assembly can provide the energy required to form a 40 Å lipid tubule. So far, lipid tubules with diameters below 10 Å formed by any of the Pacsin F-BAR domains have not been observed (*303, 311, 317*). Thus, it is possible that the observed crystal contacts do not represent physiologically relevant F-BAR domain interactions.

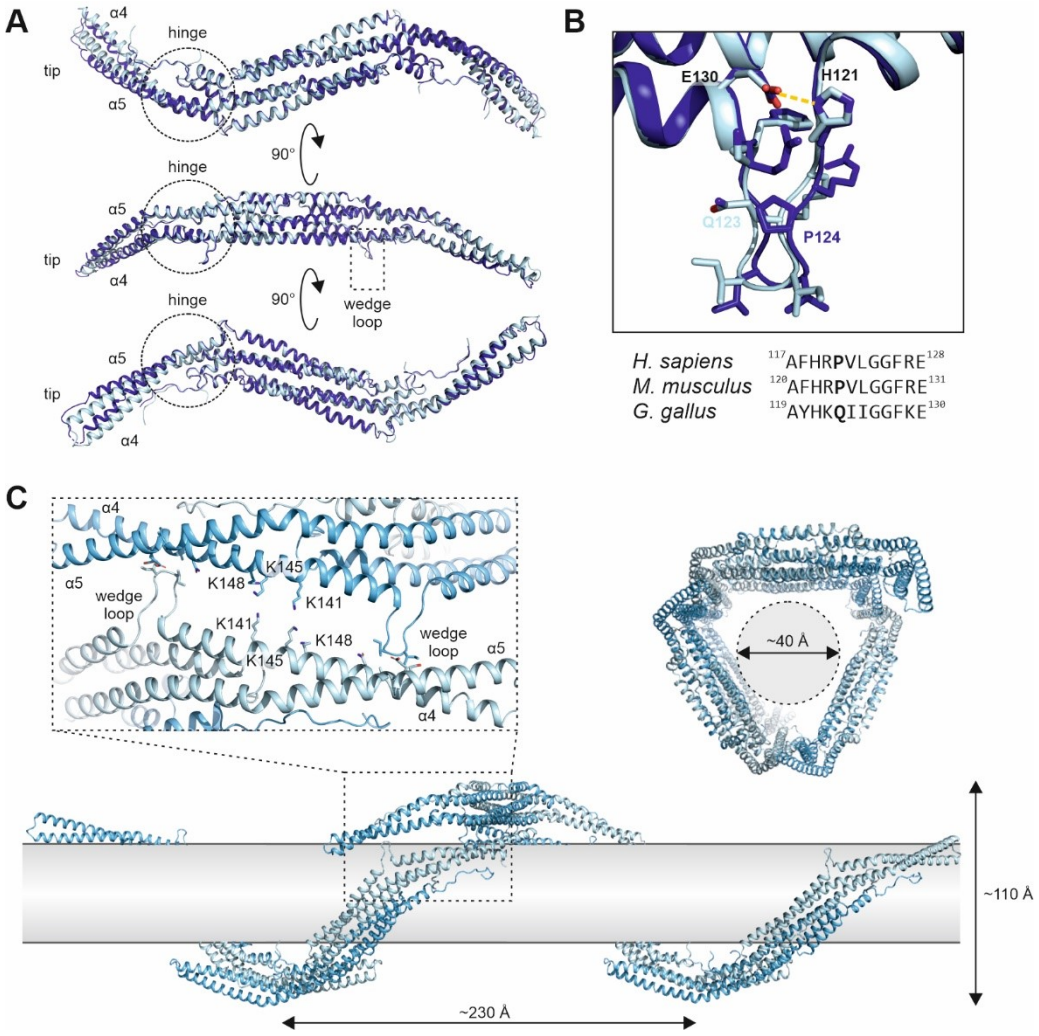


Figure 107: Structural features observed in the X-ray crystal structure of the *G. gallus* Pacsin3 F-BAR domain. (**A**) All-atom alignment of *M. musculus* and *G. gallus* Pacsin3 F-BAR domain X-ray crystal structures (backbone RMSD = 0.623 Å). (**B**) Structural alignment of the wedge loop from the *M. musculus* and *G. gallus* Pacsin3 F-BAR domain X-ray crystal structures and sequence alignment of the wedge loop from *H. sapiens*, *M. musculus*, and *G. gallus* Pacsin3 F-BAR domains. The *G. gallus* Pacsin3 wedge loop misses the proline residue which was previously linked to wedge loop flexibility in *M. musculus* and *H. sapiens* Pacsin3. Instead, it contains a glutamine residue. Nevertheless, the structure of the wedge loop is conserved between *M. musculus* and *G. gallus* Pacsin3. (**C**) Analysis of the crystal contacts between F-BAR dimers in the *G. gallus* Pacsin3 F-BAR domain X-ray crystal structure. Interactions between the wedge loop of one F-BAR dimer with the hinge region of another F-BAR dimer create a C_2 symmetric crystal contact site. The putative SH3 domain binding site (K141, K145, K148) is in the centre of the crystal contact site and thus not available for SH3 domain binding. This may explain why the SH3 domain is not resolved in the X-ray crystal structure. The wedge-loop/hinge region contacts between different F-BAR dimers result in a helical assembly with an inner diameter of ~40 Å. In theory, the F-BAR dimer helix can wrap around a lipid tubule with ~40 Å diameter. The formation of such highly bent lipid tubules would require extreme energies. Whether the helical F-BAR assembly is a physiologically relevant state remains unclear.

The determination of a high-resolution structure of the full-length *G. gallus* Pacsin3 protein via X-ray crystallography does not seem feasible. All crystallization trials using the fl-Pac3 constructs to date only yielded structures of the isolated F-BAR domain. SAXS, however, can be a powerful tool to determine the structure of full-length proteins in solution when only high-resolution structures of isolated fragments are available (*318*). In this sense, the determined F-BAR domain X-ray crystal structure and the SH3 domain NMR structure are versatile templates for SAXS modelling of the full-length Pacsin3 protein. The constructs fl-Pac3, F-BAR, l-SH3, and SH3 domain were subjected to SEC-SAXS and yielded high-quality SAXS data (Figure 108 A, B, C). Notably, the SAXS measurements confirmed that SH3 are l-SH3 are monomers in solution, whereas F-BAR and fl-Pac3 form dimers. In agreement with the NMR structure of the Pac3 SH3/TRPV4 PRR complex, the real-space distance distributions and Kratky-plots of the isolated SH3 domain (R_g = 1.57 nm, D_{max} = 6.2 nm)

indicate a globular protein (Figure 108 D and F). Attaching the linker region to SH3 (l-SH3) generates a globular protein with a disordered and flexible chain (R_g = 2.53 nm, D_{max} = 10.1 nm) (Figure 108 D). The F-BAR (R_g = 5.42 nm, D_{max} = 21.1 nm) and fl-Pac3 (R_g = 6.33 nm, D_{max} = 23.4 nm) real-space distance distributions and Kratky-plots are commensurate with elongated protein domains (Figure 108 E and F). The fl-Pac3 SAXS data moreover may suggest the presence of flexible regions which can be assigned to the flexible linker connecting the F-BAR and SH3 domains.

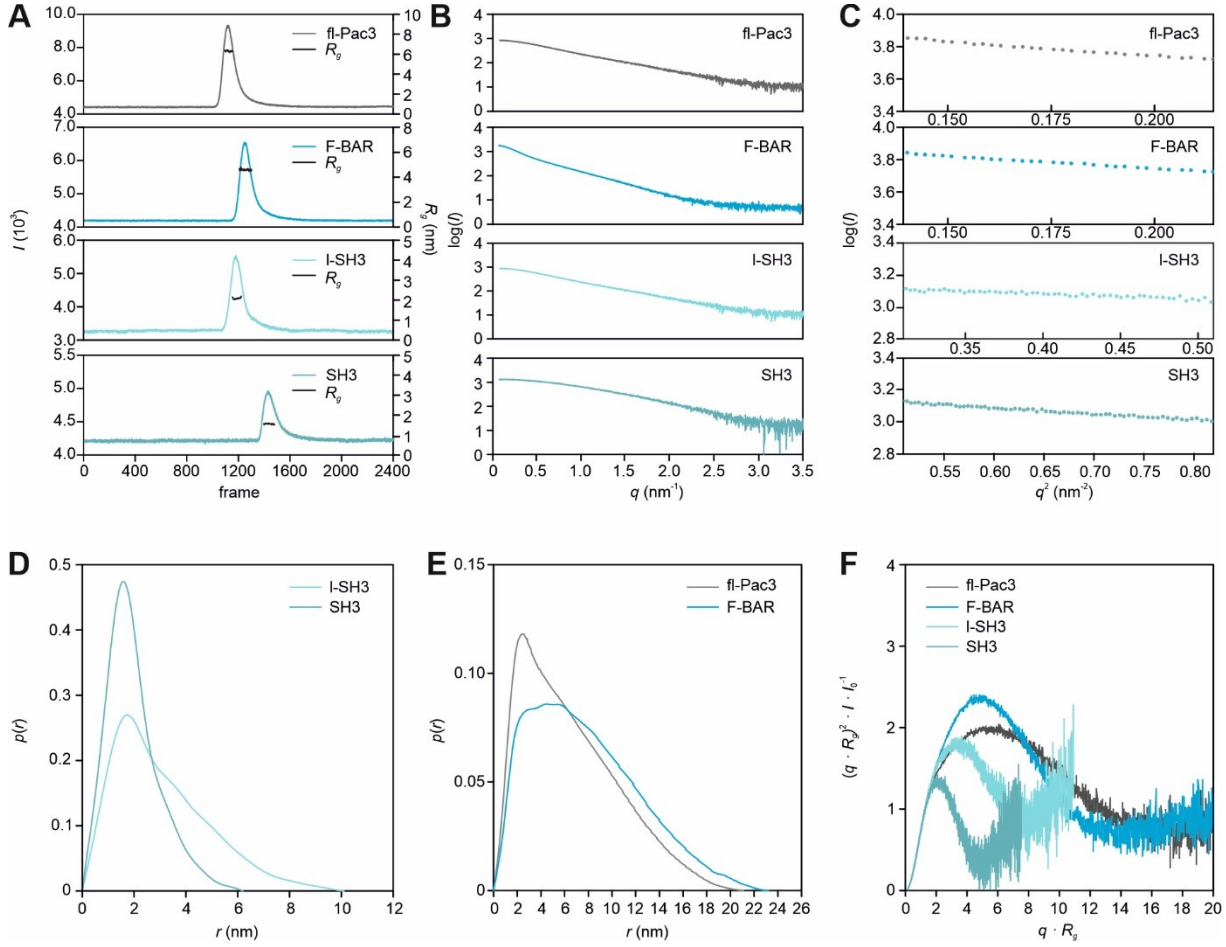


Figure 108: Small-angle X-ray scattering analysis of *G. gallus* Pacsin3. (**A**) SEC profiles of fl-Pac3 (grey), F-BAR (blue), l-SH3 (light cyan), and SH3 (dark cyan). Plotted is the scattering intensity I at the lowest detected q value (left axis) versus the frame number. The frames included in the final scattering profile and the corresponding radius of gyration (R_g) are indicated in black (right axis). SEC runs of fl-Pac3 and F-BAR were carried out with an S200 column, SEC runs of l-SH3 and SH3 with an S75 column (see VI2.18, Table 6). (**B**) X-ray scattering profiles of fl-Pac3 (grey), F-BAR (blue), l-SH3 (light cyan), and SH3 (dark cyan). Plotted is the logarithmic scattering intensity log(I) versus the scattering vector q. (**C**) Guinier-plot of fl-Pac3 (grey), F-BAR (blue), l-SH3 (light cyan), and SH3 (dark cyan). Plotted is the logarithmic scattering intensity log(I) versus the squared scattering vector q^2. (**D**) Normalized real-space distance distribution, $p(r)$, of fl-Pac3 (grey), F-BAR (blue), and (**E**) of l-SH3 (light cyan), and SH3 (dark cyan). (**F**) Dimensionless Kratky-plots of fl-Pac3 (grey), F-BAR (blue), and (**E**) of l-SH3 (light cyan), and SH3 (dark cyan) showing $(q \cdot R_g)^2 \cdot I \cdot I_0^{-1}$ plotted versus $q \cdot R_g$ (see section VI2.18).

Based on the experimental SAXS data, *ab initio* bead models of the Pacsin3 F-BAR domain and fl-Pac3 were calculated with P2 symmetry imposed as the only structural restraint (Figure 109 A and B). F-BAR domain models were generated with the GASBOR tool (χ^2 = 1.07) (*405*)and yielded an elongated envelope which aligns well with the determined X-ray crystal structure of the *G. gallus* Pacsin3 F-BAR domain dimer (Figure 109 A). Bead model generation of fl-Pac3 was performed with an extended version of the GASBOR tool, called MONSA (*413*). MONSA can read multiple SAXS data sets and subsequently generates multiphase bead models with the appropriate volume fractions of the corresponding phases. For generating a full-length Pacsin3 *ab initio*

bead model, the fl-Pac3 SAXS data set was used as the first phase and the F-BAR SAXS data set as the second phase. This yielded a MONSA model with two phases, one accounting for the volume occupied by the dimeric F-BAR domain (χ^2 = 1.26) and the other accounting for the volume occupied by two SH3 domains with the flexible linkers (χ^2 = 1.23) (Figure 109 B). As expected, the envelope corresponding to the F-BAR domain phase is nearly identical to that in the GASBOR model calculated from the SAXS data of the isolated F-BAR domain. It aligns well with the determined X-ray crystal structure of the *G. gallus* Pacsin3 F-BAR domain dimer. MONSA generated two envelopes accounting for the linker-SH3 phases on opposite ends of the F-BAR domain envelope. This suggests that the SH3 domains in the full-length Pacsin3 protein are attached to the F-BAR domain. Notably, MONSA places the linker-SH3 domain envelope on the opposite side of the F-BAR domain compared to where one would expect it from the SH3/F-BAR domain binding interface in Pacsin1 (grey surface in Figure 109 B). This might be explained by the physical limitations of SAXS to discriminate between enantiomers (*414*). The sigmoidal shape of the Pacsin3 F-BAR domain envelope in the MONSA model is not as pronounced as in the F-BAR domain crystal structures. Thus, positioning the linker-SH3 envelopes on either side of the F-BAR domain envelope yields very similar shapes with an enantiomer-like relationship, and may produce indistinguishable SAXS patterns.

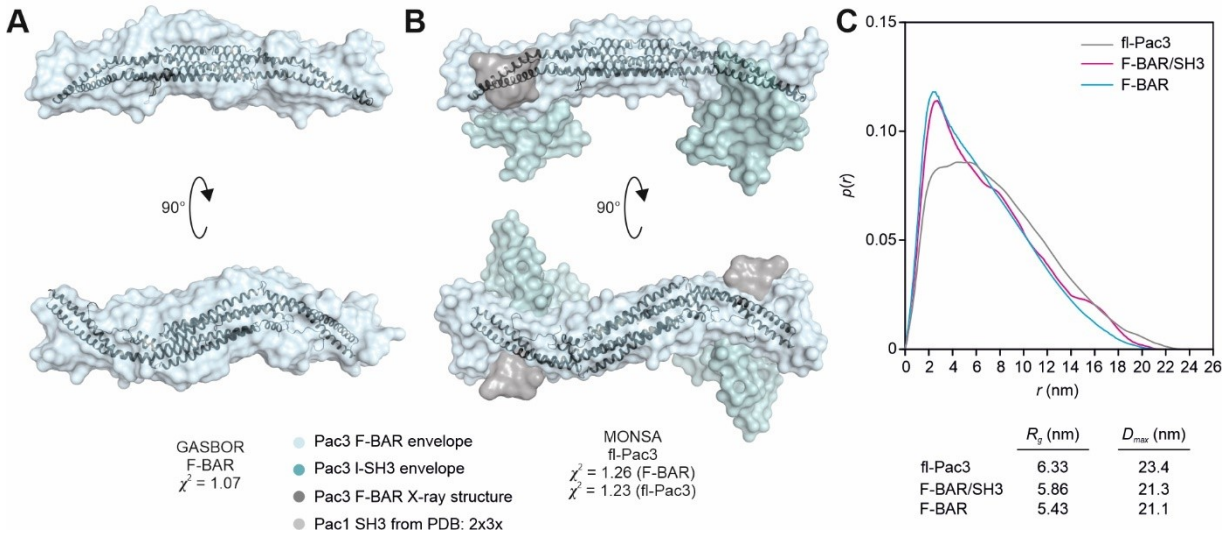


Figure 109: SAXS based *ab initio* structural models of the *G. gallus* Pacsin3 F-BAR domain and full-length protein. (**A**) GASBOR modeled envelope of the Pacsin3 F-BAR domain (light blue) aligned with the X-ray crystal structure. The envelope represents the sum of four GASBOR models with an average fit of χ^2 = 1.07 to the experimental data. (**B**) MONSA modelled envelope of the full-length Pacsin3 protein by integrating the scattering data of fl-Pac3 and F-BAR. The light blue envelope indicates the volume occupied by the F-BAR domain (res. 2-320). The cyan volume indicates the volume occupied by the remaining linker and SH3 domain (res. 321-437). The envelope represents the sum of nine GASBOR models with an average fit of χ^2 = 1.26 for the F-BAR phase and χ^2 = 1.23 for the linker-SH3 domain phase. The X-ray crystal structure of the Pacsin3 F-BAR domain was aligned with the envelope of the F-BAR domain. The grey surface shows the putative position of the SH3 domain bound to the F-BAR domain, as is observed in the full-length Pacsin1 X-ray crystal structure (PDB: 2x3x) (*307*). Alignments of X-ray structures of with SAXS envelope were performed using the SUBCOMP tool. (**C**) Real-space-distance distribution of full-length Pacsin3 (fl-Pac3, res. 2-437) and the isolated F-BAR domain (F-BAR, res. 2-320) compared to the CRYSOL-derived pair-distance distribution of the Pacsin3 F-BAR/SH3 domain complex (F-BAR/SH3) as it is expected based on the full-length Pacsin1 X-ray crystal structure (PDB: 2x3x). The table below the diagram shows the R_g and D_{max} values determined from the real-space-distance distributions.

Despite the discrepancies between the MONSA model and the putative F-BAR/SH3 binding mode in full-length Pacsin3 (Figure 109 B), the SAXS-based *ab initio* model of fl-Pac3 corroborates that the SH3 domain is proximal to the F-BAR domain instead of tumbling around randomly while being attached to the F-BAR domain via the linker. This is also supported by the differences in R_g and D_{max} between the isolated F-BAR domain and fl-Pac3 (Figure 109 C). The experimental R_g value increases by 17% from 5.43 nm in F-BAR to 6.33 nm in fl-Pac3. This is somewhat expected for attaching linker-SH3 domains to the ends of the elongated F-BAR

domain. By contrast, the D_{max} value increases only by 9% going from F-BAR (21.1 nm) to fl-Pac3 (23.4 nm) (Figure 109 C). This suggests that the added linker-SH3 domains stay close to the ends of the F-BAR domain instead of protruding away from it. Moreover, the SH3 domains likely attach to the sides of the F-BAR domain. Positioning the SH3 domains to the F-BAR tips would otherwise increase D_{max} significantly. SAXS measurements of the l-SH3 construct revealed that the SH3 domain with linker extends to a D_{max} of 10.1 nm (Figure 108 D). Thus, if the linker-SH3 was tumbling around randomly while being attached to the F-BAR domain in the full-length Pacsin3 protein, a much larger D_{max} of around ~35-40 nm# would be expected. Modelling a Pacsin3 SH3/F-BAR complex based on the X-ray crystal structure of the *M. musculus* Pacsin1 F-BAR/SH3 complex (PDB: 2x3x) and calculating the R_g and D_{max} values with CRYSOL reveals that SH3 domain binding to its putative binding site in the F-BAR domain would only increase R_g (5.86 nm) but not D_{max} (21.3 nm) (Figure 109 C). The increase in R_g while D_{max} is barely affected when going from F-BAR to fl-Pac3 may indicate that the SH3 domain attaches to the F-BAR domain. At the same time, the linker region can sample several conformations of which some lead to a slight increase of the particle dimension. It is, however, impossible to exclude that the increased D_{max} between F-BAR and fl-Pac3 originates from subpopulations of Pacsin3 molecules in open conformations. In these conformations, the linker may remain compact such that the SH3 domains stay in proximity of the F-BAR domain.

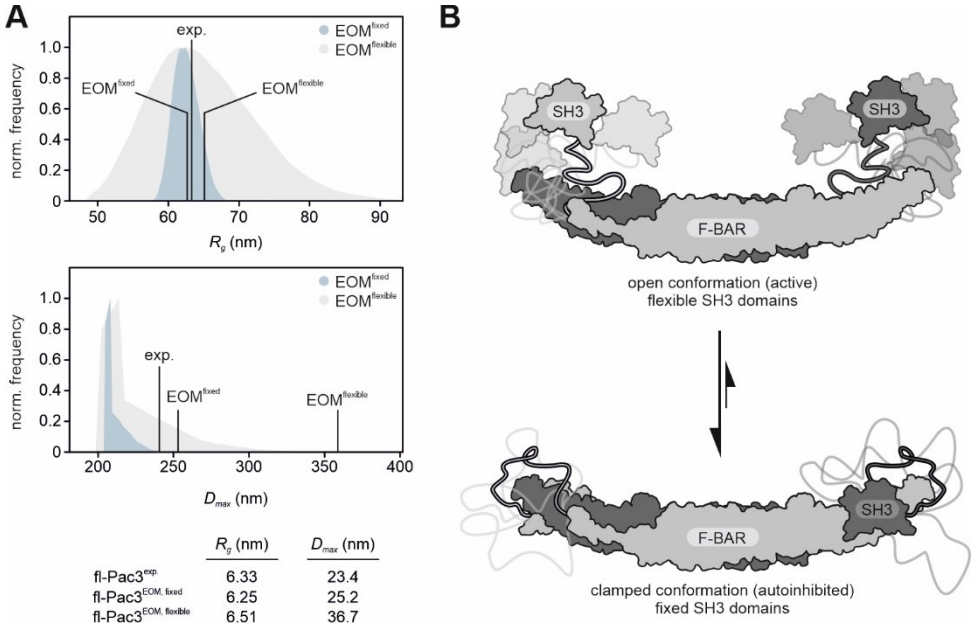
To further estimate whether Pacsin3 adopts an open or clamped conformation in solution, the fl-Pac3 SAXS data were subjected to an EOM analysis. The determined *G. gallus* Pacsin3 F-BAR domain X-ray structure and the NMR structure of the *G. gallus* Pacsin3 SH3 domain (PDB: 6f55) were used as templates to calculate pools of 10 000 structures in which the missing residues of the linker region were modelled as random chains of dummy atoms (see section VI2.18). For the generation of a random pool of Pacsin3 conformers in an open conformation, the F-BAR domain was fixed in position, and the SH3 domains were allowed to move freely. For generating a random pool of Pacsin3 conformers in clamped conformations, the Pacsin3 SH3/F-BAR complex was model based on the X-ray crystal structure of the *M. musculus* Pacsin1 F-BAR/SH3 complex (PDB: 2x3x). The SH3 domains and the F-BAR dimer were kept in fixed positions during structure calculation. Importantly, for structure generation, it was assumed that the SH3/F-BAR interaction occurs in *trans*, i.e. the SH3 domain of one subunit attaches to the F-BAR domain of the other F-BAR domain. A *cis* interaction (i.e. within the same subunit) is physically not possible as the C-terminus of an F-BAR subunit is ~14 nm distant from the putative SH3 domain binding in the same subunit. The D_{max} value of l-SH3 is only ~10 nm. Thus, even in a fully extended linker conformation, the SH3 domain should not be able to bind the F-BAR domain in *trans*. For the *ab initio* models, Pacsin3 was assumed to have P2 symmetry. However, the linker regions between the SH3 domains and the F-BAR dimer can alternate between different conformers independently. In an open conformation, the SH3 domains additionally move independently from each other. Pacsin3 is therefore pseudo-P2 symmetric. To account for the asymmetry imposed by the linker, the Pacsin3 dimer was modelled as a single chain, and the EOM calculations were performed with P1 symmetry (see section VI2.18).

The EOM analysis could not approximate the experimental data with the generated conformers beyond $\chi^2 < 2$, neither with the random pool of clamped conformers nor with that of open conformers. Nevertheless, the R_g and D_{max} distributions of the generated random pools of clamped and open Pacsin3 conformers are versatile references for a comparison with the experimentally determined values (Figure 110 A). The R_g value of a protein in solution determined via SAXS represent the volume fraction weighted R_g value of all coexisting conformers. Thus, the experimentally determined R_g value of a protein should match the random pool average R_g value if the random pool and the structural restraints used for its calculation are a valid representation of the protein's actual behaviour. As expected, the random pool R_g distribution of the calculated open Pacsin3 conformers is relatively broad, with an average R_g of 6.51 nm (Figure 110 A, top diagram, grey curve). By contrast, the random pool R_g distribution of the clamped conformers is quite narrow and has an average R_g value of 6.25 nm (Figure

This estimate is based on attaching the N-terminus of the ~10 nm l-SH3 construct to the C-terminus to the F-BAR domain. The C-termini in the F-BAR dimer are ~15 nm apart in the Pacsin3 F-BAR domain X-ray crystal structure. The 16 C-terminal residues of the F-BAR construct which are not resolved in the X-ray crystal structure can increase the distance between the C-termini by ~5 nm assuming a random chain behavior. Thus, D_{max} of a full-length Pacin3 dimer with detached SH3 domains and strechted linker regions is approx. 35-40 nm. This estimate is supported by EOM calculations (see Figure 110).

110 A, top diagram, grey-blue curve). The experimental R_g value of 6.33 nm is closer to the average R_g value of the random pool of clamped conformers than to that of open conformers (Figure 110 A, bottom). Unlike R_g, the experimentally determined D_{max} value does not refer to the volume fraction weighted D_{max} value of all coexisting conformers. Instead, the experimental D_{max} corresponds to the D_{max} of the largest conformer of the measured particle, even when this conformer makes up only for a subfraction of all coexisting conformers. The experimentally determined D_{max} value of 23.4 nm for fl-Pac3 is closer to D_{max} of the random pool of clamped conformers (25.2 nm, see Figure 110 A, middle diagram, blue-grey curve) than to the that of the random pool of open conformers (36.7 nm, see Figure 110 A, middle diagram, grey curve).

In terms of both R_g and D_{max}, the experimental SAXS results of fl-Pac3 suggest that full-length Pacsin3 predominantly adopts a clamped conformation in solution (Figure 110 A). However, it is not unlikely that Pacsin3 occasionally switches from a clamped to an open conformation to enable substrate binding to the SH3 domain (Figure 110 B). Because this switch needs to be reversible, the SH3 domain may not diffuse too far away from the F-BAR domain to be able to reattach when no substrate is present. Thus, even in the absence of a substrate, Pacsin3 can be expected to exist in an equilibrium between clamped and open conformations, but with a clear preference for the clamped over the open conformations. The sampling of extended open conformations might be prevented by the SH3 domain reattaching to the F-BAR domain right after it has detached. This might be aided by interactions between the linker and the F-BAR domain. Therefore, the experimental R_g and D_{max} values of fl-Pac3 may agree with clamped conformations and compact open conformations. Previous SAXS studies on Pacsin1 have shown that deletion of the SH3 domain in Pacsin1 almost retains R_g and D_{max} compared to full-length Pacsin1 (*303, 479*). This suggests that the linker does not adopt completely random conformations in Pacsin1 and strongly argues for interactions between the F-BAR domain and the linker. An F-BAR/linker interaction might also exist in Pacsin3 and help to stabilize the clamped conformation. Such an interaction, besides the fact that the Pacsin3 subunits probably switch between clamped and open conformations independently, substantially complicate a description of the conformational ensemble via EOM. Accounting for multiple coexisting conformational ensembles (i.e. coexistence of fully clamped, fully open, one subunit clamped/another subunit open) is currently beyond the scope of EOM.



	R_g (nm)	D_{max} (nm)
fl-Pac3$^{exp.}$	6.33	23.4
fl-Pac3$^{EOM,\ fixed}$	6.25	25.2
fl-Pac3$^{EOM,\ flexible}$	6.51	36.7

Figure 110: EOM analysis of full-length *G. gallus* Pacsin3 SAXS data. (**A**) EOM calculated R_g and D_{max} distributions of Pacsin3 in open (EOM^{flexible}) and clamped conformations (EOM^{fixed}). EOM could not fit the generated random pool of conformations to the experimental data. This might be due to limitations of EOM to account for the coexistence of clamped and open Pacsin3 conformers. The average R_g values and the largest D_{max} values of the random pool ensemble are indicated in the plots and compared to the experimentally determined R_g and D_{max} values of fl-Pac3. The EOM calculated R_g and D_{max} values of the random pool of clamped Pacsin3 conformers are closer to the experimental values than those derived from the random pool of open Pacsin3 conformers. This suggests that Pacsin3 predominantly adopts a clamped conformation in solution, similar to what was proposed for Pacsin1 (*303, 307*). (**B**) Cartoon representation of a potential equilibrium between open and clamped conformations of Pacsin3 in solution. SAXS analysis suggests that Pacsin3 predominantly adopts the clamped conformation in the absence of a substrate. The SH3 might occasionally detach from the F-BAR domain and thus increase the overall particle dimension in solution. However, before protruding too far away from the F-BAR domain in the open conformation, the SH3 domain might reattach to the F-BAR domain. Therefore, the overall particle dimension of Pacsin3 in the absence of a substrate stays close to that of the clamped conformation.

The SAXS analysis of the Pacsin3 constructs cannot unambiguously confirm that the SH3 domain binds to the F-BAR domain at the same site in the hinge region (K141, K145, K148) which were observed in the *M. musculus* Pacsin1 F-BAR/SH3 domain X-ray crystal structure. Nevertheless, the low D_{max} differences between fl-Pac3 and F-BAR and the comparison of experimental R_g and D_{max} values with those predicted by EOM suggest that the SH3 domain does not tumble around randomly in the full-length Pacsin3 protein (Figure 110). The SH3 domain appears to be rather close to the F-BAR domain, thus supporting an interaction between F-BAR and SH3 domain. To further validate a putative Pacsin3 SH3 and F-BAR domain interaction, [^{1}H, ^{15}N]s-HSQC NMR spectra of ^{15}N-labeled *G. gallus* Pacsin3 SH3 domain were recorded in the presence of increasing concentrations of the F-BAR domain (Figure 111 A). Binding of the F-BAR domain to ^{15}N-Pac3 SH3 resulted in substantial peak broadening, thus indicating the formation of a large complex with unfavourable tumbling behaviour (SH3 ~8 kDa, F-BAR dimer~75 kDa). Strikingly, the signal decrease of the ^{15}N-SH3 domain resonances in response to the presence of the F-BAR domain is relatively weak (Figure 111 A). At a fourfold molar excess of F-BAR domain over the SH3 domain, the ^{15}N-SH3 domain peaks still have ~50% of the initial signal intensity. Furthermore, the signal intensity decreases almost linearly with the F-BAR domain concentration and does not reach a plateau within the used concentration range. This suggests a relatively low affinity between the isolated SH3 domain and the isolated F-BAR domain.

As a control, a mutant of the Pacsin3 SH3 domain was generated in which residues Q398, E391, and E392 in the RT loop were replaced by three arginine residues (termed SH3 QEE/RRR). The ^{1}H, ^{15}N-HSQC spectrum of the ^{15}N-SH3 QEE/RRR mutant confirms that the QEE/RRR mutation does not perturb the structure of the SH3 domain (Figure 111 B). In Pacsin1, the corresponding QEE motif in the SH3 domain RT loop was identified to interact with lysine side chains in the F-BAR domain (K141, K145, and K148) (*307*). In pull-down

experiments, the QEE → RRR Pacsin1 SH3 domain mutant no longer interacted with the F-BAR domain (*307*). In agreement with this observation, no line broadening was observed when the F-BAR domain was titrated to the ^{15}N-Pac3 SH3 QEE/RRR mutant (Figure 111 B). This suggests that the QEE/RRR mutation in the Pacsin3 SH3 domain almost completely abolishes the interaction with the Pacsin3 F-BAR domain. Thus, the Pacsin3 SH3 domain interacts with the F-BAR domain via its QEE motif in the RT loop.

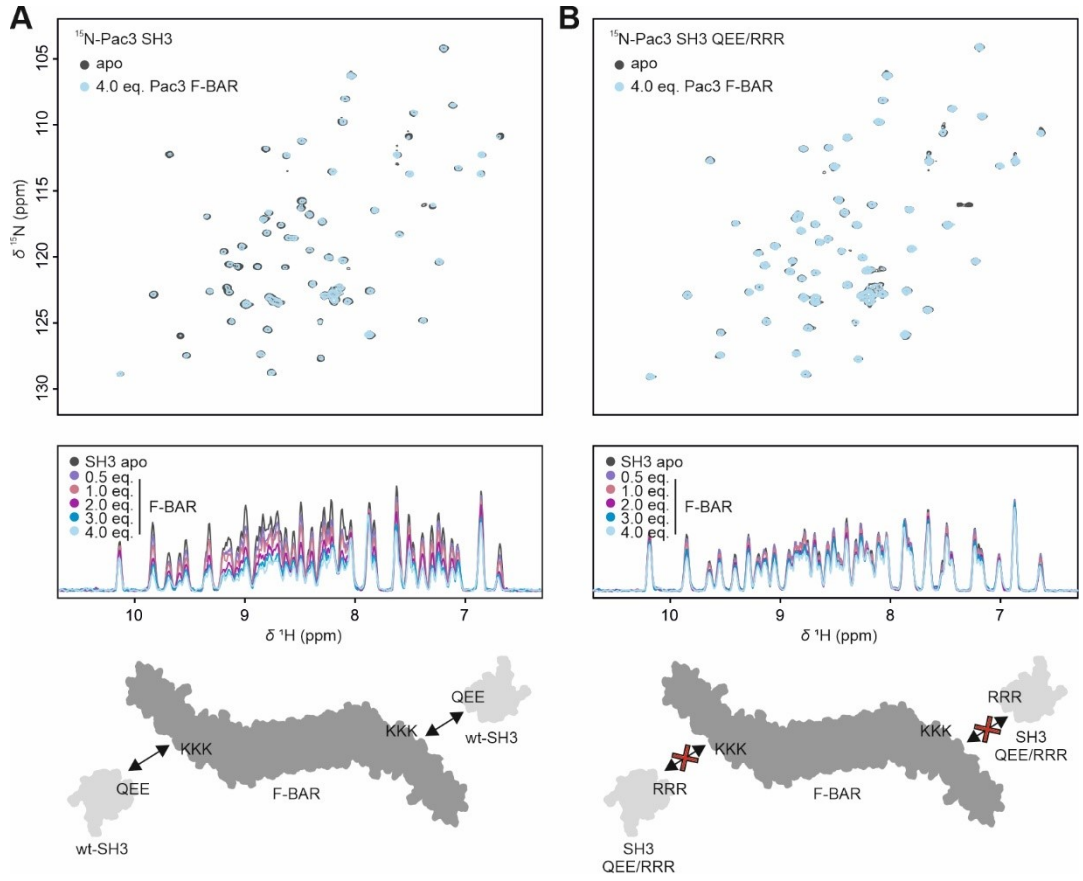


Figure 111: Interaction of Pacsin3 SH3 domain with F-BAR domain probed via NMR spectroscopy. ^{1}H, ^{15}N-HSQC spectra of (**A)** 75 µM ^{15}N-Pac3 SH3 and (**B**) 75 µM ^{15}N-Pac3 SH3 QEE/RRR mutant titrated with Pacsin3 F-BAR domain. All spectra were recorded at 298 K in 10 mM Tris pH 7, 100 mM NaCl, 10% D_2O. 1D-projections of the ^{1}H resonances indicate the line broadening/signal decrease during F-BAR titration. Titration of the F-BAR domain leads to line broadening/signal loss of ^{15}N-Pac3 wt-SH3 domain resonances whereas the ^{15}N-Pac3 SH3 QEE/RRR resonances are almost unaffected. This suggests that the SH3 domain binds via the QEE motif in the RT loop to the F-BAR domain.

The association of the Pacsin3 SH3 domain with the F-BAR domain was further investigated via HDX-MS (Figure 112 A and B). In the fl-Pac3 construct, low HDX rates are detected in the F-BAR domain helices, whereas loops, such as the wedge loop, show comparably fast HDX rates (Figure 112 A and B). Fast deuteration is also detected in the flexible linker between F-BAR and SH3 domain (<10 s), thus reflecting its disordered character (Figure 112 A and B). HDX in the SH3 domain is substantially increased compared to the F-BAR domain (Figure 112 A and B). This might be explained by the high content of disordered loop segments in the SH3 domain. A potential SH3 domain binding site in the F-BAR domain should become visible from local differences in the relative HDX between the isolated F-BAR domain or fl-Pac3. However, the HDX profile of the F-BAR domain in fl-Pac3 is almost entirely preserved in the isolated F-BAR domain (Figure 112 A). Importantly, the F-BAR region corresponding to the SH3 domain binding site in Pacsin1 (i.e. K141, K145, and K145) was not covered by proteolytic peptides and is thus not accessible to HDX-MS analysis (Figure 112 A). Nevertheless, also regions in close proximity to the putative SH3 domain binding site do not show any relative HDX differences between the isolated F-BAR domain or fl-Pac3. This might suggest that the SH3 domain only

weakly associates with the F-BAR domain and thus does not alter the local dynamics in the sensitivity range detectable via HDX-MS. A weak SH3/F-BAR association is supported by the EOM-based SAXS analysis (Figure 110) and by the NMR titration experiments (Figure 111).

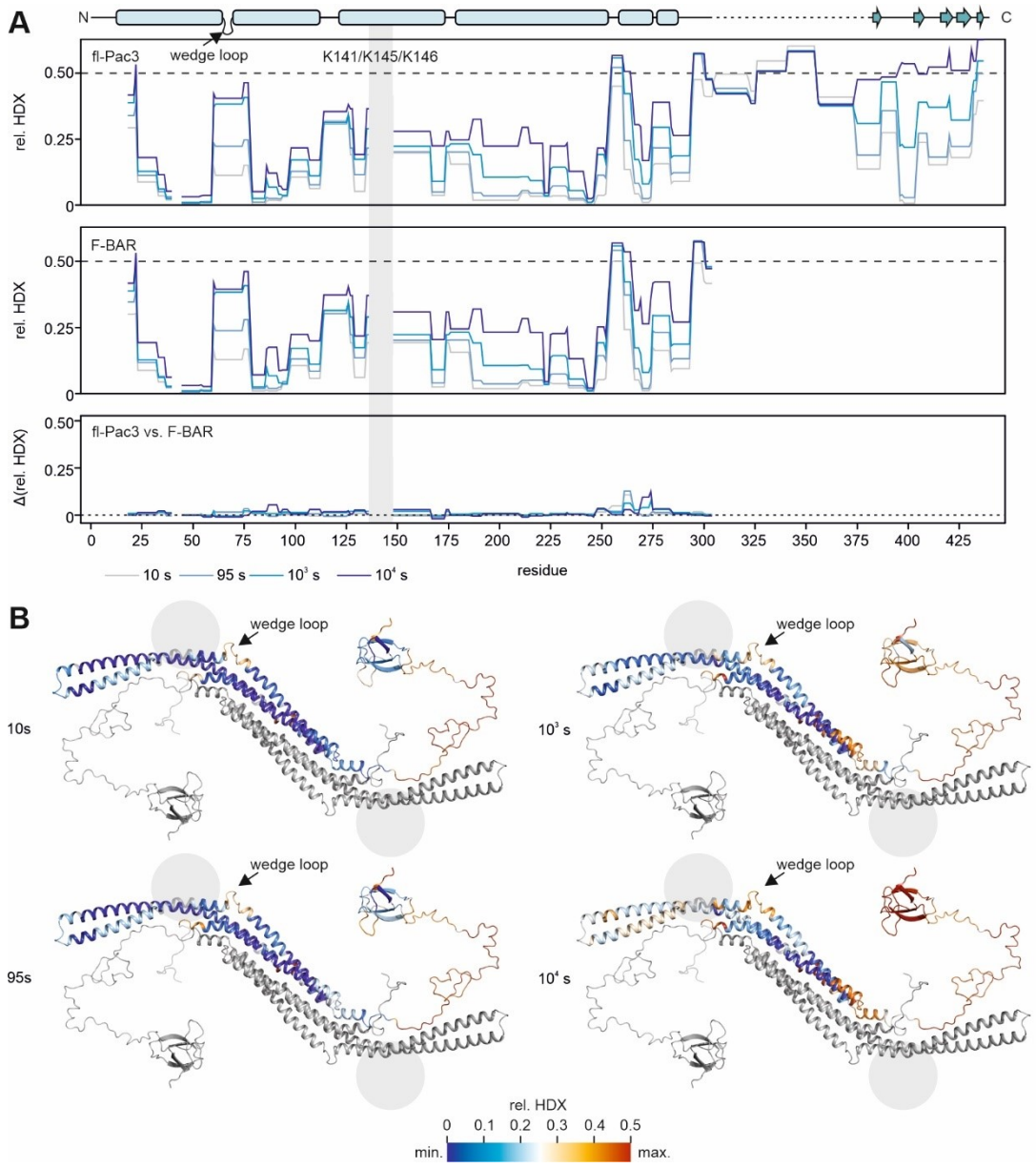


Figure 112: Hydrogen-deuterium exchange analysis of the full-length *G. gallus* Pacsin3 compared to the isolated F-BAR domain. A topology model of full-length Pacsin3 showing the position of the F-BAR domain, the wedge loop, the putative SH3 domain binding site (K141, K145, and K148), the linker, and the SH3 domain is shown on top of the diagrams. (**A**) Relative hydrogen-deuterium exchange (rel. HDX) of fl-Pac3 (upper diagram) and F-BAR (middle diagram), as well as the difference in relative hydrogen-deuterium exchange (Δ rel. HDX) between fl-Pac3 and F-BAR (lower diagram) after 10/95/1000/10 000 seconds, plotted versus the residue number. The amino acid sequence containing the putative SH3 domain binding site (K141, K145, and K148) was not covered in the HDX-MS analysis. (**B**) Relative hydrogen-deuterium exchange of fl-Pac3 after 10/95/1000/10 000 seconds mapped onto a model of the full-length Pacsin3 protein. The putative SH3 domain binding site in the F-BAR domain is indicated by a grey sphere. The full-length Pacsin3 model was generated, as described in section VI2.20.

As a complement to the NMR spectroscopic and HDX-MS analysis, the association of the Pacsin3 SH3 domain with the F-BAR domain was investigated via XL-MS (Figure 113). For this, fl-Pac3 was incubated with the lysine cross-linker DSS and cross-links were subsequently identified via mass spectrometric analysis (see section

VI2.15). A total of 30 cross-links between lysine residues within the F-BAR domain and six cross-links between lysine residues within the linker-SH3 domain region (intra-domain cross-links) could be identified (Figure 113 A). In addition to intra-domain cross-links, ten cross-links between lysine residues in the F-BAR domain and lysine residues in the SH3 domain could be identified (Figure 113 A).

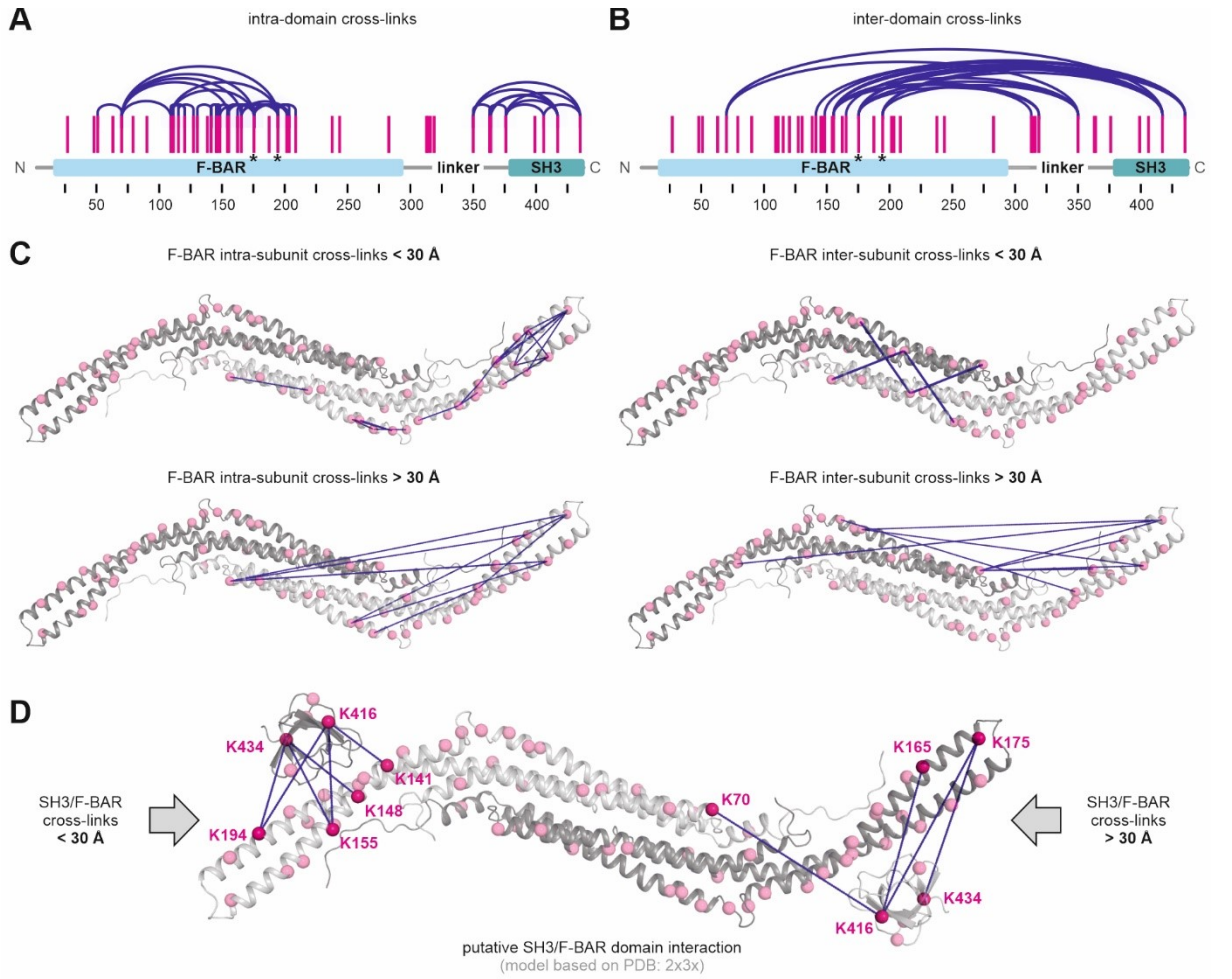


Figure 113: Intra- and interdomain contacts within full-length *G. gallus* Pacsin3 identified via cross-linking mass spectrometry. Intra-domain contacts and (**B**) inter-domain contacts in *G. gallus* Pacsin3 between native lysine residues (highlighted in magenta) crosslinked by disuccinimidyl suberate (DSS). Crosslinks are visualized by blue lines. Putative self-crosslinks, i.e. between the same lysine residue in two different protomers, are indicated with an asterisk in the NTD topology representation. The numbers represent the residue numbers. (**C**) Identified F-BAR domain intra- and inter-subunit cross-links between lysine residues with crystallographic C_α-C_α distances below 30 Å (tolerated by DSS-linker) or beyond 30 Å (beyond DSS-linker range). (**D**) Identified interdomain cross-links, i.e. between F-BAR domain and SH3 domain, mapped onto a structural model of the putative Pacsin3 F-BAR/SH3 domain complex. The model is based on the X-ray crystal structure of the *M. musculus* Pacsin1 F-BAR/SH3 complex (PDB: 2x3x). Cross-links between lysine residues with C_α-C_α distances below 30 Å in the structural model are shown on the left. They indicate that the Pacsin3 SH3 domain binds to the F-BAR domain at the same site, which was identified as the SH3/F-BAR domain-binding interface in Pacsin1 (*307*). Cross-links between F-BAR and SH3 domain lysine residues with C_α-C_α distances beyond 30 Å in the structural model are shown on the right. They may suggest that the SH3 domains can detach from the F-BAR domain and sample the space around the binding site, including the F-BAR tips.

The determined X-ray crystal structure of the Pacsin3 F-BAR domain was used to verify whether the lysine pairs of the detected F-BAR domain inter-domain cross-links have crystallographic C_α-C_α distances below 30 Å (DSS cross-linkable range) (Figure 113 C). The dimeric state of Pacsin3, however, makes it impossible to assign some of the identified cross-links to pairs of lysine residues from the same or from different subunits. 22 of the 30 intra-F-BAR cross-links may be assigned to lysine pairs within the same subunit (Figure 113 C, top/left

panel), whereas four of the 22 intra-subunit F-BAR cross-links may also account for inter-subunit cross-links (Figure 113 C, top/right panel). Notably, eight of the identified cross-links are between lysine pairs with crystallographic distances beyond 30 Å, both when intra- or inter-subunit lysine pairs are considered (Figure 113 C, lower panels). It is difficult to explain the origin of these cross-links, but they might point to transient interactions between different Pacsin3 molecules, resulting in intermolecular cross-links. Intradomain cross-links were also identified for lysine pairs in the SH3 domain (K405-K434). The corresponding C_α-C_α distances are below 30 Å in the Pacsin3 SH3 domain NMR structure (PDB: 6F55). Structural information to validate the crosslinks between lysine residues within the linker or between the SH3 domain and linker lysine residues are currently not available. However, the disordered linker can be expected to adopt multiple conformations and thus sample various cross-linking compatible lysine distances of intra-linker or linker-SH3 lysine pairs.

The identified interdomain cross-links partially support the SH3/F-BAR domain-binding interface as previously observed in the *M. musculus* Pacsin1 F-BAR/SH3 domain X-ray crystal structure (Figure 113 D). In the modelled Pacsin3 F-BAR/SH3 complex, six of the crosslinked F-BAR/SH3 lysine pairs have C_α-C_α distances below 30 Å, among them, $K416^{SH3}$ to $K141^{F\text{-}BAR}$, $K155^{F\text{-}BAR}$, $K194^{F\text{-}BAR,}$ or $K434^{SH3}$ with $K148^{F\text{-}BAR}$, $K155^{F\text{-}BAR}$, $K194^{F\text{-}BAR}$ (Figure 113 D, left side). However, four of the cross-linked F-BAR and SH3 domain lysine residues have C_α-C_α distances larger than 30 Å in the FBAR/SH3 complex model (Figure 113 D, right side). The corresponding lysine pairs are $K416^{SH3}$-$K70^{F\text{-}BAR}$, $K416^{SH3}$-$K165^{F\text{-}BAR}$, $K416^{SH3}$-$K175^{F\text{-}BAR}$, and $K434^{SH3}$-$K175^{F\text{-}BAR}$. They may indicate that the SH3 domain is not constitutively bound to the F-BAR domain but exists in an equilibrium between F-BAR-associated and dissociated states. Upon dissociation, the SH3 domain seems to sample a volume in the proximity of the tip regions. The cross-links of SH3 lysine residues to $K70^{F\text{-}BAR}$, $K165^{F\text{-}BAR}$, and $K175^{F\text{-}BAR}$ suggest that the SH3 domains can move along the tips of the F-BAR domains.

A comparison of the X-ray crystal structure of the *M. musculus* Pacsin1 F-BAR/SH3 complex with the NMR structure of the Pacsin3 SH3/TRPV4 PRR complex reveals that the putative Pacsin3 F-BAR and the TRPV4 PRR binding sites in the Pacsin3 SH3 domain overlap (Figure 104). Before binding the TRPV4 PRR, the Pacsin SH3 domain needs to dissociate from the F-BAR domain to expose the substrate-binding site. NMR, SAXS, HDX-MS, and XL-MS suggest that the SH3 domain is in an equilibrium between an F-BAR associated and dissociated state. Upon dissociation, the TRPV4 PRR should compete with the reassociation of the Pacsin3 SH3 domain to the F-BAR domain. To test this, ^{15}N-labeled Pacsin3 SH3 domain was titrated with the Pacsin3 F-BAR domain and the TRPV4 PRR peptide in varying order. The TRPV4 PRR peptide had been used previously for NMR structure determination (see section VI3.6) (Figure 114 A and B). In the first titration experiment, ^{15}N-labeled Pacsin3 SH3 domain was titrated with the Pacsin3 F-BAR domain, followed by titration of the TRPV4 PRR peptide (Figure 114 A). In the reversed experiment, ^{15}N-labeled Pacsin3 SH3 was first titrated with TRPV4 PRR, followed by the addition of the Pacsin3 F-BAR domain (Figure 114 B). The Y387 resonance adequately represents the chemical shift changes and line broadening/signal loss of the SH3 domain resonances in response to the TRPV4 PRR peptide or the Pacsin3 F-BAR domain (Figure 114 C). Titration of the F-BAR domain to ^{15}N-SH3 results in signal loss of the Y387 resonance due to the formation of a large SH3/F-BAR complex (Figure 114 C, lower panel). Subsequent titration with TRPV4 PRR shifts the resonance frequency, thus indicating the SH3/PRR complex formation. Moreover, titration of the TRPV4 PRR to the SH3/F-BAR complex leads to a signal intensity increase of the Y387 resonance. The gain in signal intensity indicates that the smaller Pacsin3 SH3/TRPV4 PRR complex (~10 kDa) forms while the large Pacsin3 SH3/F-BAR complex disassembles (~75 kDa). The Y387 resonance shift upon titrating TRPV4 PRR to SH3/F-BAR is identical to the Y387 resonance shift induced by the TRPV4 PRR in the absence of the F-BAR domain (Figure 114 C, upper panel). This suggests that TRPV4 PRR binding to the SH3 domain is independent of the F-BAR domain. This reflects in very similar dissociation constants of the SH3/PRR complex in the absence or presence of F-BAR domain (see Figure 114 C). Titration of the F-BAR domain to the SH3/PRR complex did not cause line broadening/signal loss, as observed for the SH3 domain alone (Figure 114 C, lower panel, and D). Thus, the TRPV4 PRR prevents SH3/F-BAR complex formation similar as the QEE/RRR mutation (see Figure 114 D). Notably, the final chemical shift and line width/signal intensity of the Y378 resonance (and of all other SH3 domain resonances) is independent of the order of the TRPV4 PRR and Pacsin3 F-BAR domain titration. This suggests a higher affinity of the Pacsin3 SH3 domain for the TRPV4 PRR than for the Pacsin3 F-BAR domain. The TRPV4 PRR thus releases the SH3 domain from the F-BAR domain by preventing their reassociation.

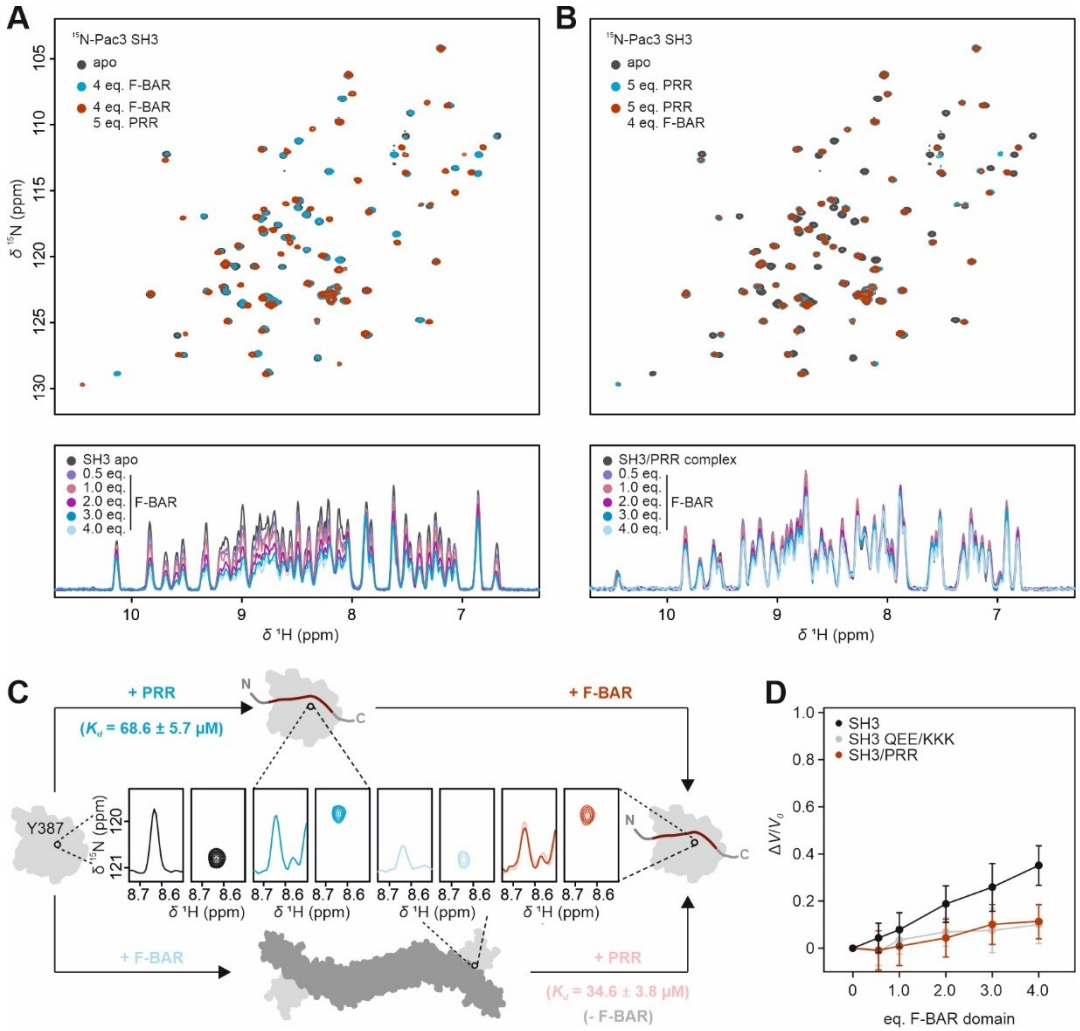


Figure 114: Effect of the TRPV4 PRR on the interaction of the Pacsin3 SH3 domain with the Pacsin3 F-BAR domain probed via NMR spectroscopy. (**A**) ^{1}H, ^{15}N-HSQC spectra of 75 µM ^{15}N-Pac3 SH3 titrated with Pacsin3 F-BAR domain followed by titration of the TRPV4 PRR peptide. (**B**) Reversed titration order as in (A). All spectra were recorded at 298 K in 10 mM Tris pH 7, 100 mM NaCl, 10% D_2O. 1D-projections of the ^{1}H resonances indicate the line broadening/signal decrease during F-BAR titration. Titration of the F-BAR domain leads to line broadening/signal loss of ^{15}N-Pac3 SH3 domain resonances. This effect is suppressed when titrating the F-BAR domain to the ^{15}N-Pac3 SH3 domain in the presence of an excess of TRPV4 PRR peptide. (**C**) Effect of titrating Pacsin3 F-BAR domain and the TRPV4 PRR peptide to ^{15}N-Pac3 SH3 domain in different orders on the chemical shift and line width of the Y387 resonance. A section from the 2D ^{1}H, ^{15}N-HSQC spectrum shows the Y387 chemical shift, the 1D projection of the ^{1}H resonance shows the peak line width. Addition of the TRPV4 PRR causes a shift of the Y387 resonance (lower panel), indicating the SH3 domain/PRR complex formation. Subsequent addition of the F-BAR domain neither affects the chemical shift nor the line with of the Y387 resonance (upper panel). This suggests that the F-BAR domain cannot bind to the SH3 domain when in complex with the TRPV4 PRR peptide. In the reversed titration order (lower panel), titration of the F-BAR domain causes line broadening but no chemical shift change of the Y387 resonance. This is commensurate with the formation of a large SH3/F-BAR complex. The subsequent addition of the TRPV4 PRR leads to chemical shift changes and increases the signal intensity. The increased signal intensity strongly suggests that the SH3 domain is no longer associated with the F-BAR domain and instead forms a complex with the TRPV4 PRR. Importantly, the final chemical shift and signal intensity/line width of the Y387 resonance is independent of the order of the titration. This suggests that the TRPV4 PRR has a substantially higher affinity for SH3 domain binding than the F-BAR domain. Accordingly, the dissociation constant of the Pacsin3 SH3/TRPV4 PRR complex is not influenced by the presence of the Pacsin3 F-BAR domain (K_d = 68.3 ± 5.7 µM no F-BAR domain vs K_d = 34.6 ± 3.8 µM with F-BAR domain). Of note, the effects of ligand titration on the Y387 resonance chemical shift and line width are consistent for all resonances throughout the ^{1}H, ^{15}N-HSQC spectrum. (**D**) Peak integral decrease $\Delta V/V_0$ of the ^{15}N-Pac3 SH3 domain resonances plotted versus the number of F-BAR domain equivalents. The plotted values represent the average peak integral decrease of all ^{15}N-Pac3 SH3 domain resonances in the ^{1}H, ^{15}N-HSQC spectrum. The error bars are the standard deviations from the average values. Strikingly, the F-BAR domain induced signal decrease observed for the wt-SH3 domain (SH3, black) is equally suppressed for the SH3 QEE/RRR mutant and the SH3/PRR complex. This strongly suggests that the TRPV4 PRR prevents the Pacsin3 SH3 domain from binding to the F-BAR domain.

To test whether the TRPV4 PRR also binds to the Pacsin3 SH3 domain in the full-length protein, HDX-MS experiments of Pac3-fl in the absence presence of the TRPV4 PRR peptide were performed (Figure 115). Differences in the rel. HDX profile of Pac3-fl caused by the presence of the TRPV4 PRR peptide exclusively cluster in the SH3 domain, which is the direct PRR binding domain in Pacsin3. HDX changes in the Pacsin3 F-BAR domain or in the linker region, indicating the site from which the SH3 domain detaches are not detected. This is in agreement with the observation that the isolated Pacsin3 F-BAR domain has the same relative HDX pattern as the full-length Pacsin3 protein. The SH3/F-BAR domain interaction seems to be too weak to influence the HDX pattern in the F-BAR domain. Thus, no changes in the relative HDX of the F-BAR domain are observed when the SH3 domain binds the TRPV4 PRR peptide and thus remains detached from the F-BAR domain. The observations from HDX-MS are in support of the NMR experiments which suggest that the Pacsin3 SH3/TRPV4 PRR interaction is stronger than the Pacsin3 SH3/F-BAR domain association.

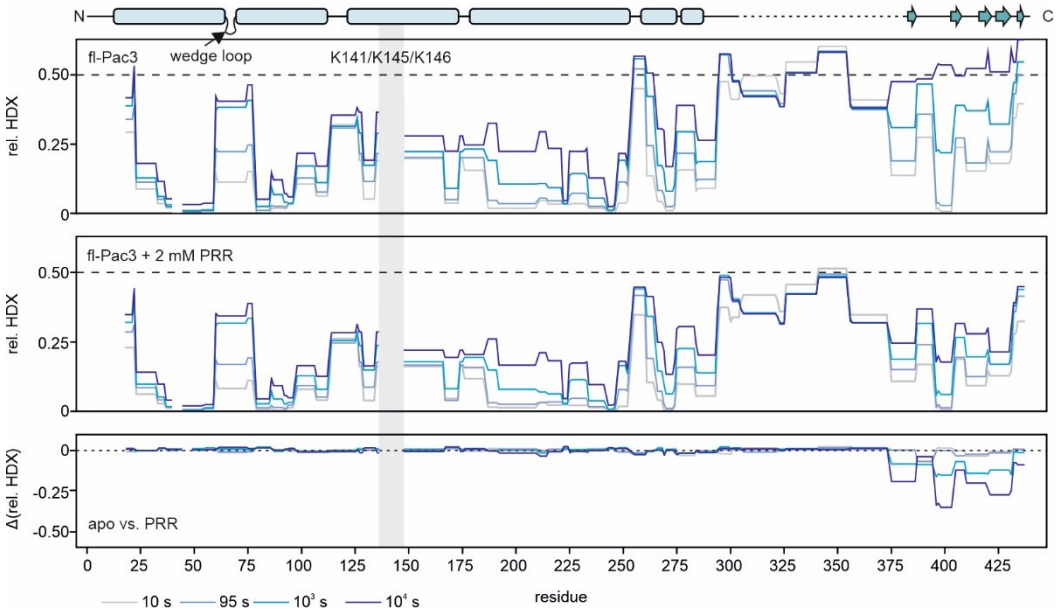


Figure 115: Effect of PRR binding on the hydrogen/deuterium-exchange in full-length Pacsin3. A topology model of full-length Pacsin3 showing the position of the F-BAR domain, the wedge loop, the putative SH3 domain binding site (K141, K145, and K148), the linker, and the SH3 domain is shown on top of the diagrams. (**A**) Relative hydrogen-deuterium exchange (rel. HDX) of fl-Pac3 in the absence (upper diagram) and in the presence of 2 mM TRPV4 PRR peptide (middle diagram) after 10/95/1000/10 000 seconds plotted versus the residue number. The difference in relative hydrogen-deuterium exchange (Δ rel. HDX) between fl-Pac3 with and without TRPV4 PRR is shown in the bottom diagram. The amino acid sequence containing the putative SH3 domain binding site (K141, K145, and K148) was not covered in the HDX-MS analysis.

NMR and HDX-MS demonstrate that the TRPV4 PRR alone can bind to the Pacsin3 SH3 domain both in isolation and in the context of full-length Pacsin3. The PRR is not only the minimal region in TRPV4 required for Pacsin3 binding (see section VI3.5) but also to compete with the Pacsin3 F-BAR domain for SH3 domain binding (Figure 114). Thus, the TRPV4 PRR seems to be sufficient to release Pacsin3 from a clamped conformation. To understand whether TRPV4 N-terminal regions beyond the PRR contribute to the interaction with Pacsin3, full-length Pacsin3 was cross-linked with the TRPV4 NTD via DSS and subjected to XL-MS analysis (Figure 116). A total of 18 intermolecular cross-links between Pacsin3 and TRPV4 NTD lysine residues could be identified.

Several factors complicate the interpretation of the XL-MS data. Both Pacsin3 and the TRPV4 NTD contain large disordered regions for which only limited structural information is available. They presumably exist in many interconverting conformations in which the number of cross-linkable lysine pairs may differ. Moreover,

a Pacsin3 dimer has two SH3 domains and thus can bind two TRPV4 NTDs simultaneously. The XL-MS data do not tell at which complex stoichiometries the cross-linking reactions occurred. The dimeric structure of Pacsin3 also makes it impossible to say whether cross-links from F-BAR and SH3 domain lysine residues to the lysine residues in the TRPV4 NTD are from the same Pacsin3 subunit or different subunits. It is, therefore, almost impossible to model the organization of the NTD relative to a Pacsin dimer based on XL-MS. Nevertheless, the identified cross-links provide valuable insights into the contact sites between TRPV4 and Pacsin3.

The two most extensively cross-linked lysines are $K107^{TRPV4}$ and $K116^{TRPV4}$ in the TRPV4 NTD (Figure 116 A). Both residues are close to the PRR, i.e. the Pacsin3 SH3 domain binding site in the TRPV4 NTD. TRPV4 residue $K116^{TRPV4}$ shows cross-links to the Pacsin3 SH3 lysine residues $K375^{Pac3}$/$K416^{Pac3}$/$K434^{Pac3}$, and $K107^{TRPV4}$ shows one cross-link to $K416^{Pac3}$. Strikingly, $K107^{TRPV4}$ and $K116^{TRPV4}$ also show cross-links to lysine residues in the Pacsin3 F-BAR domain ($K70^{Pac3}$, $K148^{Pac3}$, $K155^{Pac3}$, $K162^{Pac3}$, $K165^{Pac3}$, $K175^{Pac3}$, and $K194^{Pac3}$). These Pacsin3 F-BAR lysine residues show cross-links to SH3 domain lysine residues $K375^{Pac3}$/$K416^{Pac3}$/$K434^{Pac3}$ which themselves are cross-linked to lysine $K107^{TRPV4}$ and $K116^{TRPV4}$, thus creating cross-linking triangles between TRPV4 NTD, Pacsin3 SH3 domain, and Pacsin3 F-BAR domain (Figure 116 B). In other words, the TRPV4 NTD not only makes contacts with the Pacsin3 SH3 domain but with the Pacsin3 SH3/F-BAR binding interface. Cross-links from the F-BAR domain to the TRPV4 NTD suggest the SH3 domain in complex with the TRPV4 NTD remains close to the F-BAR domain. Interestingly, two lysine residues in the TRPV4 NTD distant from the PRR, $K50^{TRPV4}$ and $K56^{TRPV4}$, were cross-linked with lysine residues in the Pacsin3 F-BAR and SH3 domain (Figure 116 A and B). The same Pacsin3 lysine residues were also identified in cross-links between F-BAR and SH3 domain and in cross-links with $K107^{TRPV4}$ and $K116^{TRPV4}$. Further, $K50^{TRPV4}$ and $K56^{TRPV4}$ show intramolecular cross-links with $K107^{TRPV4}$ and $K116^{TRPV4}$. This suggests an interaction network with contacts between four components: The Pacsin3 SH3 domain, the Pacsin3 F-BAR domain, the TRPV4 PBD-PRR, and TRPV4 K50/K56. The cross-links from $K50^{TRPV4}$ and $K56^{TRPV4}$ to the Pacsin3 SH3 and F-BAR domains may be mediated by intramolecular contacts with the TRPV4 PBD-PRR region which itself was shown in NMR to bind the Pacsin3 SH3 domain.

Interestingly, no cross-links were identified between Pacsin3 and the TRPV4 ARD, even though the ARD has multiple lysine side chains on its surface available for cross-linking with another lysine. Either the distance between the ARD and Pacsin3 in the Pacsin3/TRPV4 NTD complex is beyond the range of the DSS linker, or unfavourable steric conditions prevent the formation of ARD-Pacsin3 cross-links. Interestingly, no cross-links were identified between the lysine residues K123 and K133 in the TRPV4 PRR and Pacsin3, even though this is the major Pacsin3 binding site. These residues might not be available for cross-linking when the PRR binds to the Pacsin3 SH3 domain.

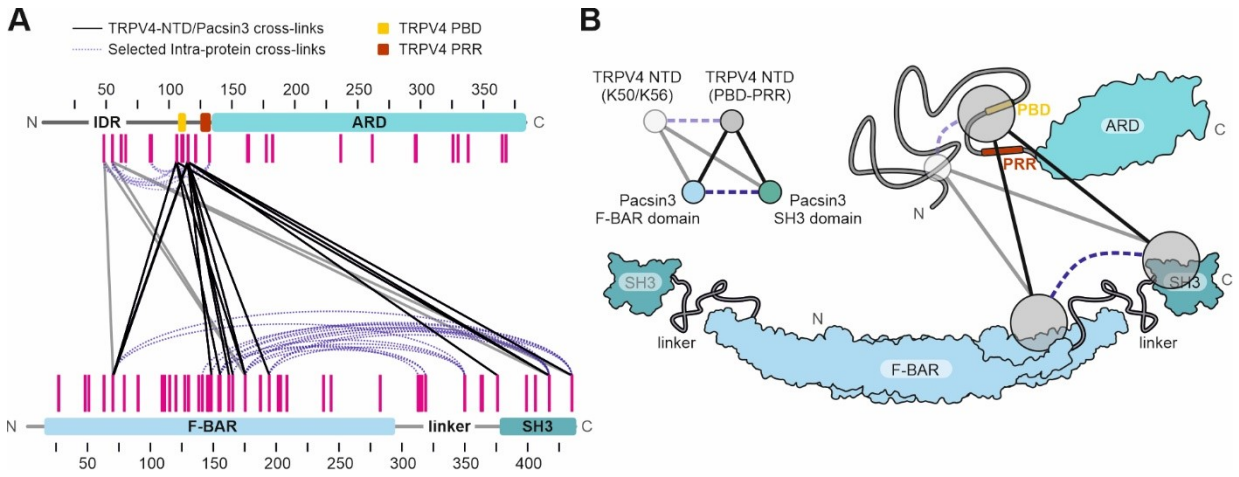


Figure 116: Contacts between full-length *G. gallus* Pacsin3 the G. gallus TRPV4 N-terminal domain identified via cross-linking mass spectrometry. (**A**) Identified cross-links (dark lines) between native lysine residues (magenta bars) of *G. gallus* Pacsin3 and the *G. gallus* TRPV4 NTD. DSS was used as the cross-linking agent. Selected intra-molecular cross-links in Pacsin3 and the TRPV4 NTD are indicated with blue dotted lines. The numbers represent the residue numbers. (**B**) Schematic representation of the interaction network between Pacsin3 and the TRPV4 NTD derived from the XL-MS analysis in (A). Intermolecular cross-links between the Pacsin3 F-BAR domain and the SH3 domain illustrate the SH3/F-BAR association required for the autoinhibition of Pacsin3. Cross-links between the Pacsin3 SH3 domain and the TRPV4 NTD are a consequence of the Pacsin3 SH3/TRPV4 PRR interaction (see section VI3.5). The SH3 domain might remain close to its binding site in the F-BAR domain when in complex with the TRPV4 NTD. This enables cross-links between the SH3 binding site in the F-BAR domain with residues in the vicinity of the PRR in the TRPV4 NTD. The TRPV4 PBD shows cross-links to lysine residues K50 and K56, which might be explained by intramolecular interactions within the TRPV4 IDR. These interactions bring K50/K56 close to the PRR and close to the Pacsin3 SH3 and F-BAR domain.

In summary, an integrated structural biology approach was used to investigate the structure of full-length *G. gallus* Pacsin3 and how it interacts with the *G. gallus* TRPV4 NTD. Using X-ray crystallography, a structure of the Pacsin3 F-BAR domain could be obtained which shows a novel crystal packing mode in which the F-BAR dimers assemble in a helical orientation (Figure 107). It is, however, unclear whether this crystal packing mode has physiological relevance. A structure of the full-length Pacsin3 protein could not be obtained via X-ray crystallography. In the final electron density map of crystals obtained from the full-length Pacsin3 construct, only the F-BAR domain was resolved. Previous crystallization attempts with full-length Pacsin have similarly failed to yield a full-length structure, with the exception of a study on full-length *M. musculus* Pacsin1 which revealed an F-BAR dimer with one SH3 domain bound (*307*). In all other cases, only the F-BAR domain could be resolved, even when full-length Pacsin constructs were used (*303, 307, 311, 317, 479*). Either the SH3 domain is cleaved from the F-BAR domain via limited proteolysis in the linker during crystal growth, or crystal contacts interfere with an SH3/F-BAR association. In the *G. gallus* Pacsin3 F-BAR domain X-ray structure determined in this study, crystal contacts between F-BAR dimers overlap substantially with the putative SH3/F-BAR binding interface (Figure 107). Also, in F-BAR domain X-ray crystal structures of Pacsin1 and Pacsin2, the putative SH3/F-BAR binding site is either directly involved in crystal contacts or sterically blocked by the crystal packing (*303, 307, 311, 317, 479*). Assuming that the F-BAR domain crystal contacts reflect physiological oligomerization interfaces, one may speculate that the release of Pacsin3 from the clamped conformation also exposes binding sites that enable F-BAR oligomerization at the plasma membrane.

The determined *G. gallus* Pacsin3 F-BAR domain X-ray crystal structure in combination with the NMR structure of the Pacsin3 SH3/TRPV4 PRR complex could be used to generate models of Pacsin3 in the putative clamped and open conformation for subsequent SAXS analyses (Figure 109 and Figure 110). A comparison of the theoretical R_g and D_{max} values of the open and clamped conformation with the SAXS-derived values strongly suggest that the SH3 domain is close to the F-BAR domain in full-length Pacsin3 in solution. This is consistent with previous SAXS studies on full-length Pacsin1 which showed that the SH3 domain must be close to the F-BAR domain in solution (*303, 479*). NMR spectroscopic experiments demonstrate that the isolated SH3 domain can interact with the F-BAR domain (Figure 111). However, the interaction appears to be comparably weak as it requires high F-BAR concentrations. The assumption that the SH3/F-BAR interaction is relatively

weak is also supported by the observation that the dynamics in the F-BAR domain are not affected by the association of the SH3 domain (Figure 112). As a consequence of a weak SH3/F-BAR assembly, Pacsin3 can be expected to exist in an equilibrium between a clamped and open conformations rather than being constitutively trapped in a clamped conformation. A weak interaction of the Pacsin3 SH3 domain with the F-BAR domain makes sense in terms of the proposed Pacsin autoinhibition mechanism. Considering the moderate affinity (K_d ~100 μM) of the SH3 domain for the TRPV4 PRR (see section VI3.5), the PRR would not compete with a robust SH3/F-BAR domain interaction and release the autoinhibition.

The XL-MS analysis indicates that the SH3 domain binds to the F-BAR domain in the hinge region, in analogy to how it was observed for Pacsin1 (Figure 116) (*307*). Further, the observation that the QEE/RRR mutation in the SH3 domain abolishes interaction with the Pacsin3 F-BAR domain in NMR experiments is a strong indicator that the SH3/F-BAR binding mode in Pacsin3 is the same as in Pacsin1 (Figure 114). In this binding mode, the substrate binding site in the SH3 domain overlaps with the binding site of the F-BAR domain. Accordingly, the SH3 domain cannot bind to the F-BAR domain when in complex with the TRPV4 PRR as evidenced in NMR experiments (Figure 111). Similarly, the TRPV4 PRR peptide can bind and release the Pacsin3 SH3 domain from an F-BAR/SH3 domain complex (Figure 111). HDX-MS experiments demonstrate that the TRPV4 PRR peptide is also sufficient to bind the Pacsin3 SH3 domain in the context of the full-length protein (Figure 112). Cross-links identified in the XL-MS analysis of the TRPV4 NTD/Pacsin3 complex further demonstrated that the TRPV4 PRR contacts the SH3/F-BAR domain association interface and thus presumably releases Pacsin3 from a clamped conformation (Figure 116). Interestingly, also contacts between Pacsin3 and regions outside of the PRR in the TRPV4 NTD were identified in the XL-MS. Further structural and functional studies are, however, required to verify whether and how these contacts contribute to the interaction between Pacsin3 and TRPV4.

The findings of this study strongly suggest that the Pacsin3 SH3 domain interacts with the hinge region of the F-BAR domain. This interaction seems to be weak enough such that TRPV4 can release Pacsin3 from the clamped conformation by binding with its PRR to the Pacsin3 SH3 domain. The Pacsin3/TRPV4 interaction in a cellular context is likely to become extremely complex upon releasing Pacsin3 from the clamped/autoinhibited state. Even though lipid tubulation studies were not included in this study, activated Pascin3 is expected to bind and remodel the plasma membrane via its F-BAR domain. Thus, it may ultimately modulate the membrane environment in the direct vicinity of the TRPV4 transmembrane domain. Furthermore, a Pacsin3 dimer has two SH3 domains, and a TRPV4 tetramer has four PRRs. Consequently, several stoichiometries of the Pacsin3/TRPV4 complex are possible, and each stoichiometry might affect TRPV4 function differently. In principle, Pacsin3 may bind with one SH3 domain to TRPV4 and with the second SH3 domain to an additional binding partner. An additional binding partner may also be a second TRPV4 channel. Pacsin3 may therefore serve as a module to assemble multiple components in a large signalling complex. Thus, it remains to be seen whether the binding of Pacsin3 itself regulates TRPV4 activity or whether it serves as the connective link to yet unknown TRPV4 regulators.

4 Conclusion

Many TRP channels are polymodally regulated, most notably by proteins and lipids (*170, 241*). TRPV4, as a transducer of diverse stimuli in various physiological processes, is a prime example of this. TRPV4-mediated Ca^{2+}-influx needs to be tightly regulated to maintain cellular homeostasis and to prevent the increase of intracellular Ca^{2+} concentrations to cell-toxic levels. Deregulation of TRPV4, e.g. through mutations in the *trpv4* gene, affects multiple cellular pathways and leads to a plethora of pathological conditions (*204*). Several pathways for TRPV4 regulation have been identified, but their molecular details have remained mostly unclear, partly due to a lack of structural information on interactions of TRPV4 with regulatory binders (*76, 77, 128, 151, 227, 251*). The recently determined cryo-EM structures of near full-length TRPV4 have made a massive leap towards a better mechanistic understanding of TRPV4 (*42*). However, the TRPV4 N- and C-termini have refrained from all structure determination attempts, presumably because they are disordered. These regions contain sites for regulatory interaction partners (*75–77, 128, 242, 250, 262*) and PTMs (*233, 483*) or cause channelopathies when specific residues are mutated (*204*). Thus, they seem to be crucial for TRPV4 regulation. The disordered region of the TRPV4 N-terminus contains binding sites for PIP_2, a lipid sensitizer, and Pacsin3, a protein desensitizer (*75–77, 128*). It has remained unclear how PIP_2 and Pacsin3 binding in the TRPV4 N-terminus can be sensed in the distant pore region of the channel. The structural characterization of the disordered TRPV4 N-terminus reveals novel insights into how it is structurally and dynamically coupled to the rest of the channel. Together with detailed analyses of how PIP_2 and Pacsin3 interact with their binding sites in the TRPV4 N-terminus, this thesis provides a foundation to understand for how ligand binding in a disordered region of a TRPV channel may regulate the conformation of the distant pore region.

4.1 The ARD-IDR interaction is a potential regulatory switch for TRPV4 sensitivity

Cryo-EM has been the workhorse of TRP channel structural biology, and high-resolution cryo-EM structures have been determined for all members of the TRPV family to date (*13, 43*). However, the distal N-terminal regions preceding the ARD are consistently missing in all TRPV1-4 structures. Only in a single TRPV3 cryo-EM structure, a short fragment of the distal N-terminus folded onto the ARD was resolved (*134*). Bioinformatic analysis of the TRPV sequences predicts high disorder propensities for the distal TRPV1-4 N-termini. Thus, these regions have mostly been deleted in the constructs used for structure determination. In cases where full-length proteins were used, the distal N-terminal regions were not resolved in the final structures (*40, 42, 49, 50, 134–138, 140–142, 144, 147*). The conventional high-resolution structure determination methods used on membrane proteins, i.e. X-ray crystallography and cryo-EM, thus, seem inappropriate to investigate the structure of the N-terminal regions in full-length TRPV1-4.

The divide and conquer strategy used in this theses proved extremely valuable to investigate the structural disorder in the TRPV4 N-terminus (section VI3.1). Constructs of the TRPV4 N-terminus could be expressed and purified from *E. coli* in high amounts. This made the entire TRPV4 N-terminus, including the region preceding the ARD, accessible to several biophysical techniques, which are expected to become impossible or extremely complicated when using full-length TRPV4 constructs. The structural analyses of recombinantly expressed TRPV4 N-terminal constructs via SEC, CD spectroscopy, NMR spectroscopy, and SAXS confirm that the distal TRPV4 N-terminus is indeed disordered (section VI3.1). The structural disorder of the TRPV4 intrinsically disordered region (IDR) manifests in high flexibility and a broad ensemble of conformations as confirmed via NMR spectroscopy, HDX-MS, and EOM analysis. In isolation and in tandem with the ARD, the IDR behaves almost like a random chain, even though compact conformations of the IDR are slightly preferred to extended conformations (Figure 48). The consistent experimental observations between TRPV4 constructs from *G. gallus* and *H. sapiens* paired with the consistently high disorder propensities predicted for the distal N-terminus of *G. gallus*, *H. sapiens*, and *X. tropicalis* TRPV4 are a strong indicator that the disorder in the TRPV4 IDR is conserved across species.

Combining the EOM refined TRPV4 NTD conformers derived from SAXS studies with the available near-full length TRPV4 cryo-EM structure enabled to model a full-length TRPV4 tetramer (Figure 49, p. 102). It is referred to as full-length TRPV4 hereafter even though it does not contain the disordered C-terminus. In the full-length TRPV4 model, the IDR ensemble occupies almost the entire space in the periphery of the cytosolic domain. This steric demand drastically increases the reach of binding sites in the IDR for its ligands. It enables the binding of interaction partners in the IDR in different orientations and positions with respect to the rest of the channel. Another consequence of the large space occupied by the IDR conformational ensemble might be that cytosolic interaction partners of TRPV4 are likely to encounter the IDR even when their primary binding site is in the ARD. The IDR might act like a scavenger that holds interaction partners in the periphery of the cytosolic regions, e.g. through non-specific electrostatic interactions between the charged IDR and charged surface patches of interaction partners. The IDR provides a substantial degree of plasticity to the N-terminus of TRPV4. Such plasticity is ideal for accommodating multiple overlapping binding sites which can adjust their conformation to the environment to enable the binding of specific ligands under defined conditions (*83, 484*). The IDR is therefore ideal for enlarging and diversifying the interactome of TRPV4. As such, it might substantially contribute to the pronounced polymodality of TRPV4. Notably, the full-length TRPV4 model neglects potential intersubunit interactions and expected interactions between the IDRs and the proximal plasma membrane in a cellular context (see sections VI3.3 and VI3.4). Nevertheless, the TRPV4 model may represent a conformational ensemble in which the IDR detaches from the plasma membrane. It is worth to consider that the 90 C-terminal residues in the TRPV4 sequence are also predicted to be disordered (Figure 15, p. 30) and are missing in the TRPV4 cryo-EM structure (*42*). The C-terminus may substantially increase the density of disordered protein regions in the cytosolic domain. As described in section VI1.3, the C-terminus also contains several sites for regulatory ligand binding or PTMs. It may, thus, serve a similar function as the N-terminal IDR in terms of enlarging and diversifying the TRPV4 interactome.

Conformational changes of an ion-conducting pore commonly mediate ion channel activity. Many channel modulators bind to regions distant from the ion-conducting pore. The local conformational changes induced by the ligands need to propagate through allosteric pathways to the pore region (*485*). Accordingly, the binding of regulatory ligands in the TRPV4 IDR should elicit local conformational changes which propagate to the TRPV4 pore region. In a full-length TRPV4 subunit, the IDR directly connects to the ARD which itself is underneath the TMD. The ARD connects to the pore region via an interaction network passing through the linker region, the pre-S1 helix, the TRP box and the S4-S5 linker. Therefore, it is possible to assume, that ligand binding in the IDR is coupled to conformational changes in the ARD which propagate to the pore region (*42, 61, 156, 265*). Thus, the ARD can be expected to play a crucial role in regulating channel activity upon ligand binding in the IDR. In all X-ray crystal structures of isolated TRPV4 ARDs in apo states or in complexes with ligands as well as in the cryo-EM structure of full-length TRPV4, the ARD is a compact α-helical protein (Figure 117 A) (*21, 42, 150*). Strikingly, the SAXS analysis reveals that the ARD in solution is more expanded than the ARD high-resolution structures from X-ray crystallography suggest (Figure 58, p. 117). In fact, the SAXS data suggest that a flexible tail protrudes from the ARD and the dimensions of the ARD in solution is larger than in high-resolution structures. HDX-MS analysis revealed EX1 deuteration kinetics in AR1, AR2, AR4, and AR6 of the isolated ARD, indicating a very labile assembly (Figure 57, p. 115). EX1 deuteration kinetics are an indicator of transiently unfolding globular structures (see section VI2.16). Thus, the isolated TRPV4 ARD appears to populate transiently unfolded structures in solution. Similar behaviour was previously observed for the ARD of IκBα (*486*). As in TRPV4, the IκBα ARD encompasses six repeats (Figure 117 A). Whereas AR1-4 form a stable globular structure in solution, AR5 and AR6 populate transiently disordered states. Previous studies on the stability of ARDs identified an AR-consensus motif of residues which stabilize ARDs by strengthening inter-repeat interactions (*487, 488*). A sequence alignment shows that the TRPV4 ARD deviates strongly from the ARD-stabilizing consensus motif in AR1 and AR2, whereas the IκBα ARD deviates from the consensus motif in AR5 and 6 (Figure 117 B). The apparent correlation between AR-consensus motif and AR-dynamics gives reason to assume that the deviations from the AR-consensus motif alter ARD dynamics up to a point where ARs populate unfolded states. Notably, the ARD amino acid sequences are highly conserved among thermoTRPV channels and, thus, all thermoTRPV ARDs show similar deviations from the AR-consensus motif (Appendix, Figure 126). The high flexibility and the potential transient unfolding of the N-terminal

ankyrin repeats observed in TRPV4 might be a general feature of thermoTRPV channel ARDs. In the context of full-length channels, AR3-6 in the ARDs of TRPV channels make intersubunit contacts and directly connect to the linker region. Thus, they are probably more stable in the context of a full-length channel than in isolated ARDs. AR1 and AR2, which also have the least sequence similarity with the AR-consensus motif, may also undergo unfolding or at least be highly dynamic in full-length channels. High conformational flexibility was, in fact, recently described in a computational study for AR1 and AR2 in the context of full-length TRPV1 (*489*). The extreme flexibility of TRPV ARDs, particularly in AR1 and AR2, could explain why the N-terminal ARs in cryo-EM studies on full-length TRPV channels are usually less well-resolved than AR3-6 (*40, 42, 49, 50, 134–138, 140–142, 144, 147*) .

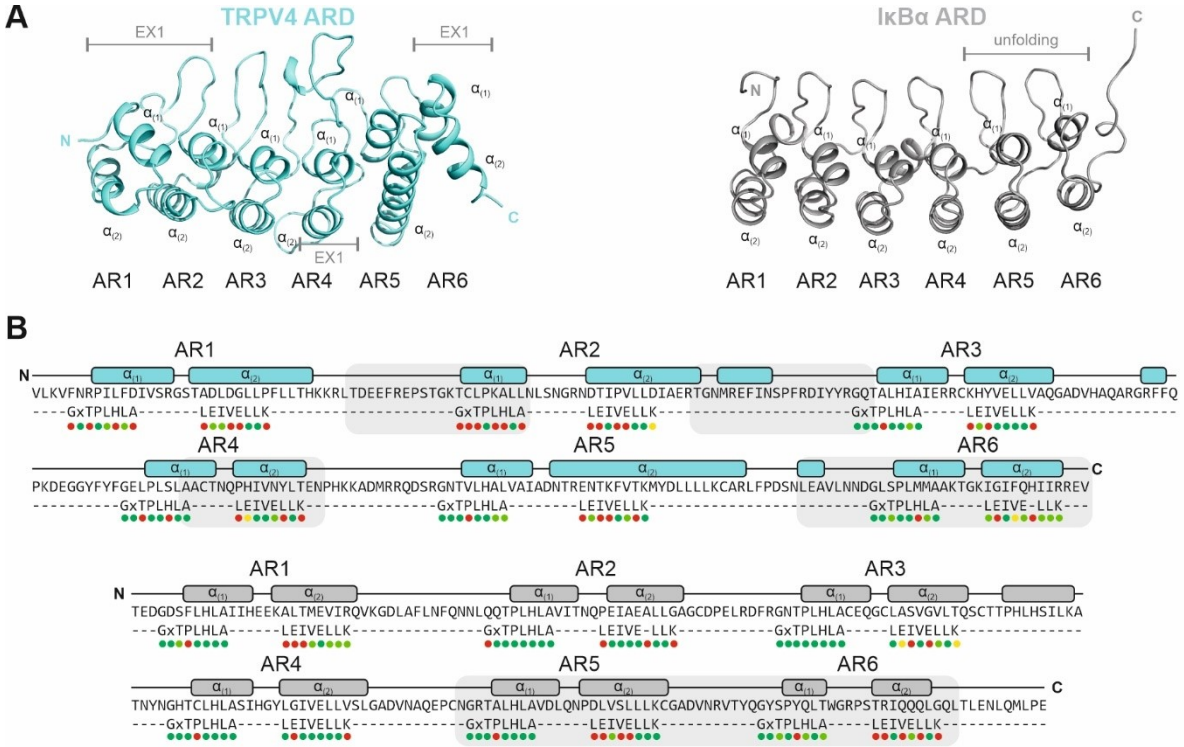


Figure 117: Consensus motif analysis of the TRPV4 ARD and the IκBα ARD. (**A**) X-ray crystal structures of the *H. sapiens* TRPV4 ARD (PDB: 4dx2) and the *H. sapiens* IκBα ARD (PDB: 1nfi). (**B**) Sequence alignment of the *H. sapiens* TRPV4 and IκBα ARDs with AR-consensus motifs previously described to confer stability to ARDs. The consensus motif in the $\alpha_{(1)}$-helix is GxTPLHLA, where x is any amino acid besides C, G, and P. The consensus motif in the $\alpha_{(2)}$-helix is LEIVELLK (*487*). Substantial deviations from the consensus motif are indicated in red, medium deviations in yellow, subtle deviations in light green, and identical residues in dark green. The amino acid sequence in AR1 and AR2 of the TRPV4 ARD deviates substantially from the ARD consensus motif. Together with SAXS data, the HDX-MS suggest that the N-terminal repeats in the TRPV4 ARD detach from the rest of the ARD (see section VI3.2). AR4 and AR6 also seem to undergo unfolding in the isolated TRPV4 ARD as evidenced by HDX-MS (indicated with light grey boxes, see Figure 57). Similar to AR1 and AR2 in the TRPV4 ARD, AR5 and AR6 of the IκBα ARD populate transiently unfolded states (*486, 488*). These repeats also show the largest deviation between the IκBα ARD amino acid sequence and the AR-consensus motifs.

The structural analysis of the TRPV4 NTD deletion constructs with gradually truncated IDRs revealed an unexpected structural and dynamic coupling between the ARD and the IDR (see section VI3.2). As expected for an entirely disordered IDR, adding IDR residues to the ARD increases the disorder in the protein almost linearly (Figure 50, p. 105). One would expect that this comes with a linear increase of the protein size. However, SEC and SAXS analysis showed that that the addition of IDR residues beyond the PBD induces the sudden compaction of the protein (Figure 50 and Figure 51, p. 105-106). The compaction may originate from the stabilization of the ARD to a more compact state through an association between IDR and ARD. Crucially, the HDX kinetics in the ARD change dramatically when the complete IDR is part of the construct (Figure 57, p. 115). Deuteration was substantially reduced throughout the ARD in the context of the NTD compared to

the ARD in isolation. No EX1 deuteration kinetics occurred. The IDR appears to structurally and dynamically stabilize the ARD, even though HDX shows that the disorder and flexibility in the IDR are retained (Figure 57, p. 115). A ligand-dependent stabilization of ankyrin repeat structure and dynamics was also observed for the IκBα ARD upon binding of its ligand NFκB (*486, 488*). While AR5 and AR6 transiently unfold in the isolated IκBα ARD, the repeats form the typical ankyrin repeat fold in complex with NFκB (*490*). In TRPV4, the IDR seems to have a similar effect on the ARD as NFκB on the IκBα ARD. However, whereas the TRPV4 ARD and IDR are part of the same protein, NFκB and IκBα are individual proteins. The most striking difference is that NFkB is folded whereas the TRPV4 IDR is entirely disordered. The stabilization of the ARD through the IDR implies some form of physical interaction between both domains. Tryptophan fluorescence demonstrates that the PBD tryptophan residue W109 is in a much more hydrophobic environment in the presence of the ARD compared to isolated tryptophan (Figure 53, p. 109). Similarly, the IDR itself buries W109 in a hydrophobic environment. Therefore, residues in the IDR and between ARD and IDR seem to interact and bury W109 in a hydrophobic environment. This observation is in good agreement with crosslinks identified via XL-MS (Figure 55, p. 112). Intra-IDR and ARD-IDR interactions, involving sequentially distant pairs, suggest that the NTD must sample compact conformations (Figure 55, p. 112). The observation that not all residues expected to cross-link showed up in the XL-MS analysis suggests a certain degree of hierarchy and order underlying the IDR-ARD interaction. This order and hierarchy might point to specific ARD-IDR interactions.

The finding that a disordered region stabilizes a globular domain structurally and dynamically as observed for the TRPV4 IDR and ARD is a hitherto undescribed phenomenon. In TRPV4, this structural and dynamic coupling between IDR and ARD might form the mechanistic basis for channel regulation through ligand binding in the IDR. IDR-ARD association and dissociation may directly influence the structural and dynamic state of the ARD. In the associated state, the ARD is relatively rigid, whereas in the dissociated state, the ARD is very flexible and transiently unfolds in the N-terminus. Being structurally coupled to the pore via the linker/pre-S1/TRP box/S4-S5 linker, the ARD is also ideal for linking its dynamics to the pore region. Ligand binding or changes in the environment such as ionic strength or temperature may influence how the IDR interacts with the ARD. The modulation of the ARD/IDR communication can directly impact the ARD conformational and dynamic state, which in turn controls the pore conformation and dynamics. It is difficult to predict the effect of a ligand on the ARD/IDR interaction. A ligand may act similar to a glue that enhances an ARD/IDR interaction. It may also prevent the IDR from associating with the ARD. Both scenarios can occur through ligands with a primary binding site in the IDR or the ARD. It remains to be seen whether a rigid or a dynamic ARD favours an open/sensitized or a closed/desensitized pore conformation.

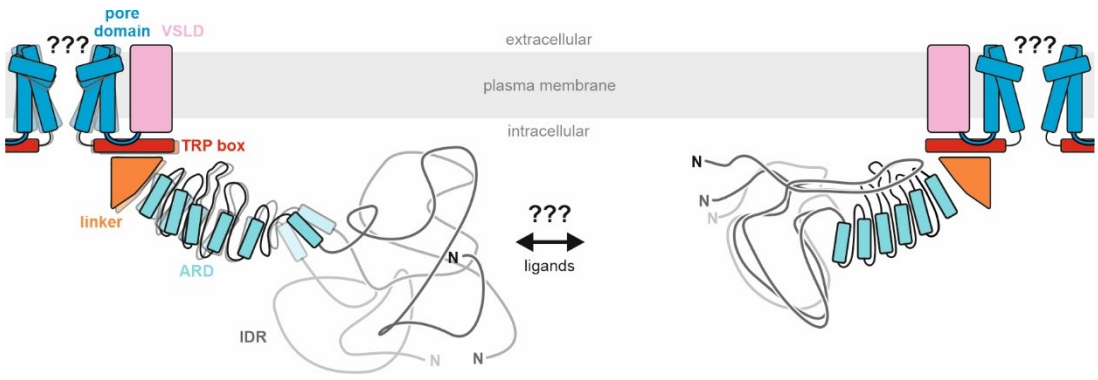


Figure 118: Possible implications of the conformational and dynamic coupling between the IDR and the ARD in the TRPV4 N-terminus on the TRPV4 pore domain. In a full-length TRPV4 subunit, the ARD connects to the TMD via a linker/pre-S1/TRP box interaction network (the pre-S1 helix is not indicated in the cartoon). The TRP box itself controls the conformation of the pore domain via an interaction with the S4-S5 linker. Conformational changes and the structural dynamics in the ARD may propagate along the linker/pre-S1/TRP box/S4-S5 linker network to the pore domain. In a state where the ARD dissociates from the IDR, the ARD fluctuates strongly and transiently unfolds in AR1 and AR2. This flexibility may propagate to the pore domain. Association of the IDR with the ARD stabilizes the ARD structurally and dynamically and might also reduce the flexibility in the pore domain. Ligand binding in the ARD and the IDR might influence the IDR/ARD interaction and thus regulate the ARD conformational dynamics and ultimately the pore domain. Whether a flexible or a rigid ARD favours an open/sensitized or closed/desensitized TRPV4 state is unclear.

In a recent cryo-EM study, an ARD-IDR interaction is a crucial feature in the sensitization mechanism of TRPV3 (see section VI1.2, Figure 8, p. 19). According to this mechanism, the residues succeeding the C-terminal β-strand bind in a coil conformation to a cavity formed by the ARD fingers of one subunit and the β-sheet of an adjacent subunit (termed ARD/β-sheet cavity) in the desensitized state of TRPV3. Upon transitioning to the sensitized state, the C-terminus folds into a helix and moves out of the ARD/β-sheet cavity. A fragment of the TRPV3 IDR (called NTD in the corresponding TRPV3 study) subsequently occupies the vacant position in the ARD/β-sheet cavity (*134*). The ARD/IDR association in the sensitized TRPV3 state is reminiscent of the ARD-IDR interaction observed in the TRPV4 NTD in this study. However, which residues in the TRPV3 IDR bind to the ARD, are as unclear as for TRPV4. Nevertheless, observing an ARD-IDR interaction in two thermoTRPV channels (TRPV3 and 4) could point to a general role of IDRs for sensitizing thermoTRPVs (Figure 119). An associated IDR in the ARD/β-sheet cavity might stabilize a sensitized state. In contrast, the dissociation of the IDR might favour a desensitized state in which the C-terminus binds to the ARD/β-sheet cavity (Figure 119).

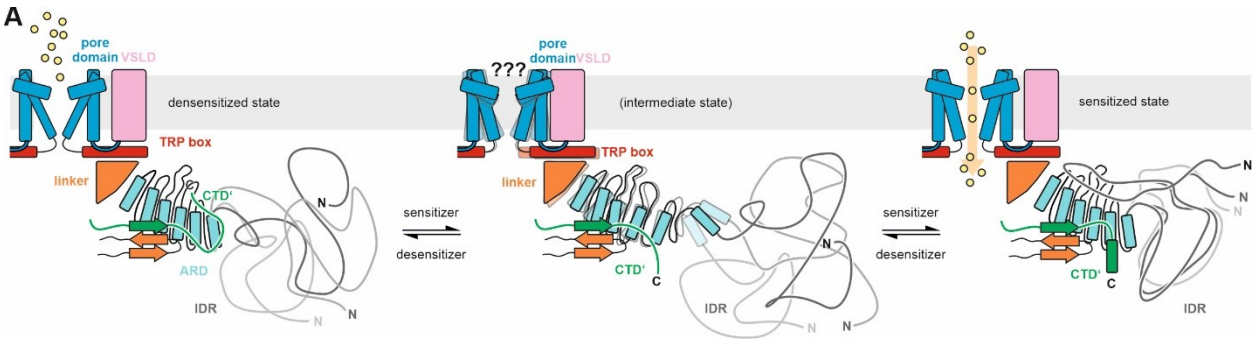


Figure 119: Potential regulatory mechanism of TRPV4 and other thermoTRPV channels involving an ARD-IDR interaction. In a desensitized state of a thermoTRPV channel, the C-terminal domain of an adjacent subunit (CTD') binds to the ARD/β-sheet cavity in the interface between two neighbouring subunits. Upon sensitization, the CTD detaches from the ARD/β-sheet cavity and forms an α-helix. The vacant position in the ARD/β-sheet cavity is subsequently occupied by the IDR. In both the desensitized state and the sensitized state, the ARD might be dynamically stabilized through binding of the CTD or the IDR, respectively. In an intermediate state between the desensitized and the sensitized state, the ARD becomes extremely dynamic due to the vacant ARD/β-sheet cavity. The desensitized, intermediate, and the sensitized states exist in a dynamic equilibrium which is shifted by the binding of sensitizing or desensitizing ligands. The details of the CTD-IDR switch might differ between thermoTRPV channels and may be rooted in the TRPV subtype-specific IDR properties.

The architecture of the cytosolic intersubunit interface which forms the ARD/β-sheet cavity and the residues keeping the C-terminus bound in the ARD/β-sheet cavity are conserved in all thermoTRPV channels (V1-4). Even though their IDR amino acid sequences are weakly conserved (Figure 7, p. 18; and Appendix, Figure 123), the disordered character of the region preceding the ARD is a thermoTRPV specific feature and might be crucial for a reversible association with the ARD as observed for TRPV3 (*134*). It is thus possible that the C-terminus/IDR switch proposed for TRPV3 (termed CTD-NTD switch in the corresponding study) exists in all thermoTRPVs. The structural and dynamic coupling between the IDR and the ARD as observed in TRPV4 via SAXS and HDX-MS might be a mechanistic feature of the C-terminus/IDR switch, which is not observable via cryo-EM. Strikingly, the EX1 deuteration kinetics indicating the unfolding of AR1 and AR2 in the TRPV4 ARD cluster in the ARD fingers which also bind the C-terminus and the IDR fragment in the TRPV3 cryo-EM structures (Figure 56, p. 113) (*134*). Thus, both the C-terminus and the IDR might structurally and dynamically stabilize the ARD but potentially in different ways. In an intermediate state between a sensitized and desensitized state, in which the C-terminus and the IDR are both detached from the ARD/β-sheet cavity, the ARD might become extremely flexible to the point where AR1 and AR2 transiently unfold (Figure 119). The flexibility of the ARD in the intermediate state might differ between the thermoTRPVs, either due to different intrinsic ARD dynamics or due to different affinities between the ARD and the IDR or the C-terminus. The intrinsic ARD dynamics might be similar, considering the conserved ARD amino acid sequences among thermoTRPVs (Appendix, Figure 126). By contrast, the IDRs are the least conserved regions in the thermoTRPVs

(Appendix, Figure 123). They could determine the details of the transitions between the different states, e.g. through the binding of specific ligands or PTMs (Figure 119). It is tempting to hypothesize, that thermoTRPVs have evolved different IDRs to fine-tune their sensitization mechanisms and thus to achieve different gating properties. Due to their disordered characters, IDRs are expected to be more susceptible to mutagenesis than the folded regions of thermoTRPVs, i.e. the TMD and the ARD. Multiple iterations of mutations in the IDR sequence are thus possible to meet the specific needs of a thermoTRPV channel, e.g. ligand binding sites or PTMs, without perturbing the remaining channel regions. In other words, IDRs undergo faster evolutionary changes due to the lack of structural constraints (*491, 492*).

Notably, the residues which mediate the binding of the C-terminus in the ARD/β-sheet cavity in the desensitized state of TRPV3 and presumably also in other thermoTRPVs are not conserved in TRPV5/6. Moreover, the TRPV5/6 ARD is not preceded by a disordered region as in TRPV1-4 (Figure 7, p. 18). Instead, an α-helix directly precedes the ARD and folds on top of the ARD where it occupies the ARD/β-sheet cavity similar as the IDR in the sensitized TRPV3 state (Figure 4 and Figure 8) (*134*). A C-terminus/IDR switch as proposed for thermoTRPV channels thus seems to be impossible in TRPV5/6. This is consistent with the observation that TRPV5/6 channels are constitutively open. One might speculate that the N-terminal helix substantially increases the ARD stability in TRPV5/6 compared to the stability of the TRPV1-4 ARDs. The increased ARD stability might be a crucial determinant of the different gating behaviours between TRPV5/6 and TRPV1-4.

In summary, this study provides structural and dynamic insights into a disordered TRPV channel region which is mostly inaccessible to structure determination via cryo-EM and X-ray crystallography. Seeing that the IDR modulates the conformational dynamics in the TRPV4 ARD points to a regulatory mechanism which may explain how ligand binding, PTMs, or disease mutations in the IDR can alter channel activity. The structural and dynamic stability in the ARD induced by its preceding unstructured protein region is an intriguing yet undescribed phenomenon: disorder promotes order! This finding may be a motivation to rethink the construct design of TRPV channels or other proteins with disordered regions for future structure determination attempts. Commonly, protein regions which are predicted to be disordered based on amino acid analysis are deleted in constructs used for structure determination via cryo-EM and X-ray crystallography. Even though this seems counterintuitive, it might be beneficial to leave IDRs in protein constructs intact, thus potentially stabilizing ordered regions. In this sense, future cryo-EM studies of TRPV4 and other TRPV channels might yield better resolutions in the cytosolic domains when disordered residues of the IDR are part of the protein constructs.

4.2 An IDR plasma membrane association fine-tune TRPV4 sensitivity

An essential regulatory binding partner of TRPV channels is the lipid second messenger PIP_2 (*170, 173*). With around 1-2% of the total lipid content, PIP_2 is a ubiquitous component of the plasma membrane (*455*). It is tightly regulated by receptor signalling and thus links TRPV channel gating to several signalling pathways (*171*). The exact mechanism by which PIP_2 modulates TRPV channel activity is unclear and seems to diverge within the TRPV family (see section VI1.2) (*112, 181–183, 188, 189*). Recent studies propose a PIP_2 binding site in the TMD of TRPV channels in a pocket formed by the TRP box and the S4-S5 linker (*188, 189*). The TRP box and the S4-S5 linker have been assigned crucial roles in controlling the conformation of the pore domain (*60, 61, 265*). Lipid binding in the TRP box/S4-S5 linker region thus may directly exert a force on the pore region and induce conformational changes. The presumed residues of the PIP_2 binding pocket are relatively conserved in TRPV channels with minor deviations between the thermoTRPVs and TRPV5/6 (*189*). The high-resolution structures of TRPV channels reveal substantial differences in the TMD architecture of TRPV4 compared to the other TRPV channels (see section VI1.3, Figure 12). Unlike in all other TRPV4 channels, the TRPV4 S4-S5 linker is not an α-helix and the VSLD is tilted by ~90° (*42*). Consequently, the TRP box and the S4-S5 linker residues presumed to shape the PIP_2 binding pocket are not in a suitable position to bind the PIP_2 head group. It is therefore questionable whether the TRP box and S4-S5 linker in TRPV4 form a PIP_2 binding site as it is proposed for other TRPV channels.

Instead, a phosphoinositide binding domain (PBD) constituted by a ++W++ motif (+ is a basic residue) in the IDR was identified as the PIP_2 binding site (*128*). PIP_2 binding to the PBD is required for TRPV4 to sense temperature and osmotic stimuli. The interaction is supposed to be mediated by the electrostatic interaction between positively charged side chains of the PBD and the negative head group of PIP_2 (*75*). Mutating the ++W++ motif in TRPV4 to AAWAA rendered TRPV4 insensitive to stimuli (*128*). The observation that a specific lipid is required to achieve a regulatory effect might suggest that PIP_2 binding to the PBD is specific. Phosphoinositide-specific lipid-binding domains are commonly known as PH domains, folded proteins with a consensus motif (K/R) x_n (K/R) x $(K/R)_2$ (here, x_n are n arbitrary residues) (*449*). The ++W++ motif of the PBD forms such a consensus motif with an upstream arginine side chain (see section VI3.3, Figure 59). However, the structural analysis of the TRPV4 N-terminus demonstrated that the IDR is entirely disordered, including the region encompassing the PBD (see sections VI3.1 and VI3.3). It is therefore not surprising that the PBD does not show specificity for PIP_2 in the *in vitro* system used in this study. Instead, it binds to anionic lipids irrespective of their exact structure (see section VI3.3) and remains fully disordered in the lipid-bound state (Figure 63 and Figure 79).

The measured affinities of the TRPV4 PBD for liposomes containing phosphatidylglycerol (PG), as a representative for membranes containing anionic phospholipids, are in the low micromolar range (Figure 62). The affinity of the PBD measured for short-chain PIP_2 micelles (Figure 78) is in a similar range as that for liposomes containing PG. At physiological pH, PIP_2 has a charge between -3 to -4 *(454)*. It is supposed to have a significantly increased affinity to polybasic protein stretches such as the PBD compared to a monovalent phospholipid competitor (*171*). The similar affinities of the PBD for PG and PIP_2 are therefore a bit surprising and could be explained by the different lipid types used in the experiments. It is likely that the affinity of the PBD for PIP_2 is higher than for anionic lipids such as PG. However, an exact quantification of lipid binding affinities in the PBD is currently missing. Considering that the PM contains about 1-2% PIP_2, it is reasonable to assume that the PBD will almost always bind to PIP_2 in the two-dimensional, diffusion-limited context of the plasma membrane. This might be expected since TRPV4 cannot be efficiently activated by heat or osmotic shock in the absence of PIP_2 (*128*). Notably, the plasma membrane contains around 10% anionic phospholipids (*455*). Thus, even in the absence of PIP_2, e.g. upon PLC activation, the PBD can be expected to bind to the plasma membrane.

When mutating the ++W++ motif in the PBD to AAWAA, PIP_2 and PG binding to the PBD is almost entirely abolished (Figure 68). This is consistent with the observation that the AAWAA mutation rendered TRPV4 insensitive to temperature and osmotic stimuli in cell-based functional assays (*128*). However, the NTD^{AAWAA} mutant retained a particular ability to co-sediment with negatively charged liposomes (Figure 72), thus suggesting that regions in the IDR other than the PBD may function as lipid-binding sites as well. Indeed, NMR experiments demonstrate that in both the wild-type IDR and the IDR^{AAWAA} mutant multiple patches of basic and hydrophobic residues in the region preceding the PBD can bind PIP_2 and PG, even though with a lower affinity (Figure 76, Figure 77, Figure 78). Approximately two-thirds of the IDR residues are involved in membrane binding. Unlike what is observed for many IDPs (*436*), no secondary structure forms in the TRPV4 IDR upon binding to membranes. Instead, the IDR remains entirely disordered in a lipid-bound state (Figure 79). These observations indicate that even though the PBD is the primary lipid-binding site in the IDR, regions outside of the PBD may weakly and non-specifically associate with anionic lipids. In a physiological context of the cell, the IDR might continuously sample the inner leaflet of the plasma membrane for negatively charged patches.

Interestingly, the PBD and other lipid-binding regions in the IDR seem to be involved in intra-chain interactions (Figure 80, Figure 82, Figure 83, Figure 84). Some of these intra-chain interactions may explain why the TRPV4 N-terminal deletion constructs show subtle differences in their lipid-binding behaviours (Figure 62, Figure 72, and Figure 73). Membrane interaction of the IDR presumably interferes with these intra-chain interactions and may lead to significant conformational rearrangements throughout the IDR. The conformational ensemble of the IDR in solution presumably differs significantly from those of membrane-bound IDRs. The structural differences between both the free and the lipid-bound state of the IDR could have functional implications for TRPV4 activity. The details of lipid-mediated structural rearrangements, however, remain to be

elucidated. A recent study showed that TRPV4 activity depends on membrane bulk properties, such as membrane fluidity (*129*). One may speculate that an association of the IDR with the PM not only depends on the local surface charge but also membrane fluidity. Besides, the association of the IDR with the PM may alter the bulk membrane properties. In this sense, the IDR could be a modulator and sensor of membrane dynamics and fluidity in the direct vicinity of the TRPV4 TMD and thus regulate TRPV4 function. Notably, IDPs have recently emerged as sensors and inducers of membrane curvature (*493*). One may hypothesize that the membrane curvature around TRPV4 modulates how the IDR interacts with the PM. Induction of membrane curvature by the TRPV4 IDR might also aid to control the functional state of TRPV4 or to attract other curvature sensing TRPV4 interactors such as Pacsin (see next section, VI4.3). Several studies localized TRPV4 in small plasma membrane invaginations, so-called caveolae (*494, 495*). A potential ability of the TRPV4 IDR to sense or induce membrane curvature could thus not only directly regulate TRPV4 activity but help to anchor TRPV4 in specific PM micro-domains.

The observation that the PBD non-specifically binds to anionic lipids and, as the rest of the IDR, remains disordered raises the question how PIP_2 produces a stimulus-specific sensitization of TRPV4 whereas other anionic lipids which are abundant in the PM do not (*128*). It is possible that folded regions in the TRPV4 cytosolic domains such as the ARD or regions not included in the constructs used in this study contribute to a TRPV4-PIP_2 interaction and are needed to mediate PIP_2 specific regulatory effects. In a previous X-ray crystallographic study, the isolated ARD directly interacted with the PIP_2 headgroup (IP_3) in AR4 (*150*). However, in the context of a full-length TRPV4 channel, the identified binding site resides in the inner vestibule of the cytosolic domain and does not seem to make physiological sense. It may instead be a crystallization artefact. Molecular dynamics simulations identified another potential PIP_2 binding site in the ARD on the side of the finger loops between AR3 and AR4 (*150*). This binding site is also unlikely to contact the headgroup of a PIP_2 molecule in the PM. In all available full-length TRPV structures, the ARD is separated by a gap of ~15-20 Å from the region where the PM would be expected (*40, 42, 49, 50, 134–138, 140–142, 144, 147*). The ARD would have to undergo a substantial conformational change towards the membrane directly contact the headgroup of a PIP_2 molecule in the PM. In all available TRPV channel structures, no ARD conformation is compatible with a membrane interaction. One would expect that the presence of a PIP_2 binding pocket formed by basic side chains at least enables a weak interaction with anionic lipids in general. In this study, however, liposomes containing high amounts of the anionic lipid PG did not interact with the ARD (Figure 72, p. 134). It is therefore questionable whether the ARD is a direct lipid-binding site in TRPV4. Nevertheless, the ARD may indirectly contribute to lipid binding. The structural analysis of the TRPV4 N-terminal deletion constructs revealed that residues in and close to the PBD not only mediate interactions to residues in the IDR but also with the ARD (see section VI3.3). Because the PBD is the primary binding site for lipids in the IDR, it is possible that lipid binding to the PBD affects the ARD/IDR interaction which in turn regulates the ARD conformational dynamics (Figure 118). This mechanism might also link a PIP_2/PBD interaction to the proposed mechanism of thermoTRPV sensitization (Figure 119). According to this mechanism, PIP_2 binding to the PBD would facilitate or mediate binding of the IDR in the ARD/β-sheet cavity.

A recent study identified a hydrogen bond between L596 (S4-S5 linker) and W733 (TRP box) which holds TRPV4 in a closed state (*60, 61, 265*). Lipids close to L596 can act as a hydrophobic tug on the L596-W733 hydrogen bond by interacting with the L596 side chain. Hydrophobic tug leads to a movement of the TRP box and the subsequent opening of the pore (Figure 14). The hydrophobic tug mechanism rests on from MD simulations on a TRPV4 tetramer embedded in a PC bilayer. Instead of PC, a PIP_2 molecule may at least equally well-suited for destabilizing the L596-W733 hydrogen bond and prime the TRPV4 pore for opening in response to a stimulus. The region where a lipid needs to bind in order to interact with the L596 side chain (termed TMD lipid pocket, TMD^{LP}) differs from the PIP_2 binding site at the TRP box/S4-S5 linker proposed for TRPV channels (Figure 5) (*188, 189*). However, the TMD^{LP} also exhibits basic side chains in the Trp box and pre-S1 helix, which may be suitable to coordinate the charged PIP_2 headgroup (Figure 120 A). Considering the position of the PBD in the IDR, it is theoretically possible that the PBD binds to the head group of a PIP_2 molecule in the TMD^{LP}. In such a scenario, the trivalency of PIP_2 would be an important requirement for concerted electrostatic interactions of the lipid headgroup with basic side chains in the TMD^{LP} and the PBD. In this sense, the PBD and the TMD^{LP} might together form a lipid-binding pocket with high specificity for PIP_2. Strikingly,

the TMD^{LP} is close to the binding site for the osmotransducer 5', 6'-EET in the VSLD (*264*). Thus, PIP_2-mediated destabilization of the L596-W733 hydrogen bond and the subsequent TRP box movement may rearrange the VSLD to facilitate the binding of a 5', 6'-EET molecule. This structural rearrangement may also facilitate TRPV4 activation by warm temperatures (Figure 120 B). Notably, the hydrophobic tug and binding of the IDR to the ARD/β-sheet cavity may not be mutually exclusive mechanisms to sensitize TRPV4. An ARD/IDR interaction may guide the PBD to the TMD^{LP} where it binds to the headgroup of a resident PIP_2 molecule. Conversely, keeping the PBD bound close to the TMD through a PIP_2/PBD/TMD^{LP} complex could facilitate the interaction of IDR residues with the ARD/β-sheet cavity.

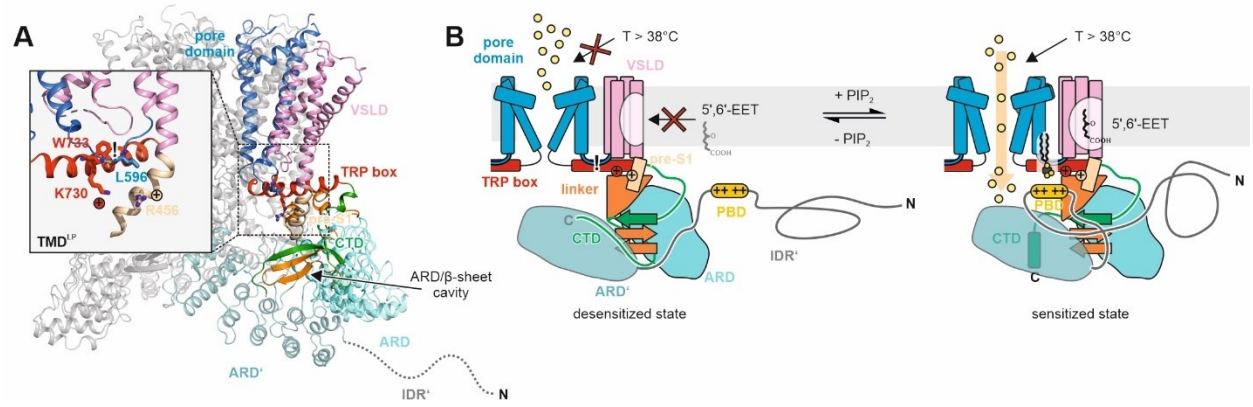


Figure 120: Mechanistic model for PIP_2-dependent sensitization of TRPV4 to temperature and osmotic stimuli. (**A**) A hydrogen bond between L596 in the S4-S5 linker and W733 in the TRP box stabilizes a closed pore conformation in TRPV4. This hydrogen bond can be destabilized by a competing interaction of the L596 side chain with a lipid. The corresponding lipid-binding site, termed the TMD^{LP} might be formed by the pre-S1 helix, the TRP box, the S4-S5 linker, and the VSLD). The positively charged side chains of K730 (TRP box) and R456 (pre-S1 helix) may interact with a negatively charged headgroup of an anionic lipid such as PIP_2. (**B**) In a desensitized state, TRPV4 cannot respond to temperature or the binding of the osmotransducer 5', 6'-EET. Based on the mechanism proposed in Figure 119, the desensitized state is characterized by the CTD binding to the ARD/β-sheet cavity. The IDR is dissociated from the ARD, and the hydrogen bond between L596 in the S4-S5 linker and W733 in the TRP box (indicated by an "**!**") holds the pore in a closed conformation. In the sensitized state, a PIP_2 molecule binds to the TMD^{LP} and the PBD. This creates a high-affinity PIP_2/TRPV4 complex held together by electrostatic interactions between the lipid headgroup and the basic side chains of the PBD and the residues in the TRP box (K730) and pre-S1 (R456). The competitive interaction between PIP_2 and the L596 side chain releases or destabilizes the L596-W733 hydrogen bond. The subsequent dislocation of the TRP box may lead to a structural change in the TMD to enable 5', 6'-EET-binding in the VSLD. 5', 6'-EET is a direct TRPV4 activator. The structural change in the TMD may also enable activation of TRPV4 by elevated temperatures. Notably, binding of the PBD to a PIP_2 molecule in the TMD^{LP} requires the IDR to fold onto the ARD. This might facilitate the insertion of an IDR region into the ARD/β-sheet cavity, as proposed in Figure 119. Both, the release of the L596-W733 hydrogen bond and the IDR binding in the ARD/β-sheet cavity may elicit structural and dynamic changes in the pore region which enable activation by temperature and osmotic stimuli (i.e. 5', 6'-EET).

In summary, this study establishes the structural basis of lipid binding in the disordered TRPV4 N-terminus. The PBD acts as a potent lipid-binding site for anionic lipids. Unlike what one might expect for the stimulus-specific sensitization of TRPV4 elicited by PIP_2, the PBD does not exhibit binding specificity for PIP_2. Instead, the PBD can bind monovalent anionic lipids such as PG. The observation that the minimalistic system used in this study does not achieve PIP_2 specificity fuels speculations that a regulatory interaction of TRPV4 with PIP_2 requires the synergism of multiple regions, e.g. the PBD together with the TMD. Importantly, this study finds that the PBD is not the only region in the IDR capable of interacting with anionic lipids. Multiple IDR regions outside of the PBD can interact with lipids as well, though with lower affinity. Thus, the IDR may not only associate with the PM via a PIP_2/PBD interaction. Moreover, the entire IDR seems to be able to associate with the PM to create a high density of membrane-bound disordered protein around the channel. The functional implications of the potent membrane binding behaviour of the IDR for TRPV4, however, remain to be established.

4.3 Pacsin3 might desensitize TRPV4 by reorienting the N-terminus and reshaping the plasma membrane

PIP_2 rarely acts alone on proteins; it almost always co-plays with another regulator (*171*). In the case of TRPV4, Pacsin3 antagonizes PIP_2 function by reducing channel sensitivity to temperature and hypotonicity (*77*). The Pacsin binding site in TRPV4 was previously revealed through mutagenesis and pull-downs as a proline-rich region (PRR) between the PBD and ARD (*76*). It is intuitive to hypothesize that the interaction of the TRPV4 PRR with Pacsin3 removes PIP_2 from the PBD and desensitizes TRPV4 due to the absence of a channel sensitizer. The NMR experiments carried out in this study confirm that the PRR is the primary binding site of the Pacsin1-3 SH3 domains with a similar low micromolar affinity as PIP_2 binding to the PBD (see sections VI3.5 and VI3.3). Interestingly, a secondary Pacsin SH3 domain binding site with a ~10-fold lower affinity than the PRR colocalizes with the PBD (Figure 93). Interaction of the SH3 domain, however, does not prevent the PBD from binding to PIP_2 (Figure 94). In the isolated TRPV4 peptide constituting the PBD and the PRR, the Pacsin3 SH3 domain and PIP_2 can bind simultaneously to their binding sites to form a tripartite complex with the peptide (Figure 94 and Figure 95). This suggests that the functional regulation of TRPV4 by PIP_2 and Pacsin3 does not depend on a straight-forward competition mechanism.

SH3 domains commonly bind to proline-rich sequences which form polyproline type II (PPII) helix structures (*463*). Unlike α-helices, PPII helices are not held together by hydrogen bonds and are similarly flexible as random coil structures (*468, 471*). CD and NMR spectroscopic analysis suggested that also the isolated TRPV4 PRR predominantly adopts a PPII helix structure (Figure 98, Figure 99). The proline peptide bonds in the PRR are in equilibrium between a *trans* (~80%) and a *cis* conformation (~20%). However, only the *trans* conformation is compatible with a PPII helix conformation. Thus, the PRR continuously switches between multiple conformations of which only a fraction is a continuous PPII helix. HetNOE measurements showed that the PRR exhibits rapid backbone fluctuations (Figure 46). This flexibility may explain why the PRR was not resolved in the TRPV4 cryo-EM structure even though it was part of the *X. tropicalis* TRPV4 construct used for structure determination (*42*).

Two SH3 consensus binding modes, class I and class II, with opposite orientations of the substrate to the SH3 domain, are known (Figure 97) (*463*). The determined solution NMR structure of the Pacsin3 SH3/TRPV4 PRR complex revealed that TRPV4 interacts with Pacsin3 in a class I SH3 domain binding mode (Figure 101). This orientation could not have been predicted solely from the PRR amino acid sequence as it is commensurate with a class I and a class II consensus binding mode. In complex with the Pacsin3 SH3 domain, the TRPV4 PRR adopts a canonical PPII fold between residues T121 to N127. Strikingly, the N127-P128 peptide bond is in a *cis* conformation that enables P130 to occupy the second proline pocket, thus skipping one residue compared to the typical class I binding mode (Figure 101). This binding mode of the PRR to the Pacsin SH3 domain might explain why the single P128A and the double P128A/P129L mutants equally reduce the Pacsin3 SH3 binding affinity by a factor of three (Figure 103). Reduction of interaction strength may well explain why TRPV4 channels carrying mutations in P128A and P128A/P129L did not co-immunoprecipitate with Pacsin3 and were still activated by heat when co-expressed with Pacsin3 in cell-based experiments (*77*).

NMR chemical shift perturbation experiments reveal that unlike almost the entire IDR, the PRR is entirely unaffected by titration of PIP_2 (Figure 66, Figure 67). This makes sense from a structural point of view since the PRR sequence is not particularly well-positioned to encounter the PM. The PRR predominantly has prolylpeptide bonds in *trans* when no ligand is bound. Considering that the PRR is unaffected by lipids (section Figure 76, Figure 77), one may assume that in the PIP_2 bound and sensitized state of TRPV4, the P128 peptide bond is in a *trans* conformation (Figure 121 A). Binding of the Pacsin3 SH3 domain stabilizes the N-terminal part of the PRR (residues 121-126) in a PPII conformation, and the *cis* conformation at Pro128 will introduce a kink in the PRR amino acid chain (Figure 121 B). Pro128 is in a triple proline stretch that is conserved in TRPV4 channels and directly precedes the ARD. The presence of the SH3 domain and the forced *cis* conformation at Pro128 changes the relative orientation of the PRR and thus also the preceding PBD to the ARD. The titration of the Pacsin3 SH3 domain with the TRPV4 NTD ΔN120 construct, i.e. the ARD and PRR (30.5 kDa in size), suggests that the SH3 domain movement bound to the PRR is severely restricted. Here,

peaks showed strong line broadening, a behaviour that agrees with the SH3 domain now constituting part of a bigger complex rather than moving like a bead on a loose string (Figure 88). A reorientation of the PRR may ultimately pull the PBD away from the membrane. Thus, instead of a direct binding competition between Pacsin3 and PIP_2, Pacsin3 might desensitize TRPV4 by stabilizing a conformation of the IDR in which the PBD loses contact to the PM and thus to the sensitizer PIP_2 (Figure 121). Given the mechanism proposed in Figure 120, loss of the PBD-PIP_2 interaction may weaken PIP_2 binding to the TMD^{LP} and thereby stabilize a closed pore conformation. Reorienting the PRR may also desensitize TRPV4 by displacing the IDR from the ARD/β-sheet cavity (compare with Figure 119 and Figure 120)

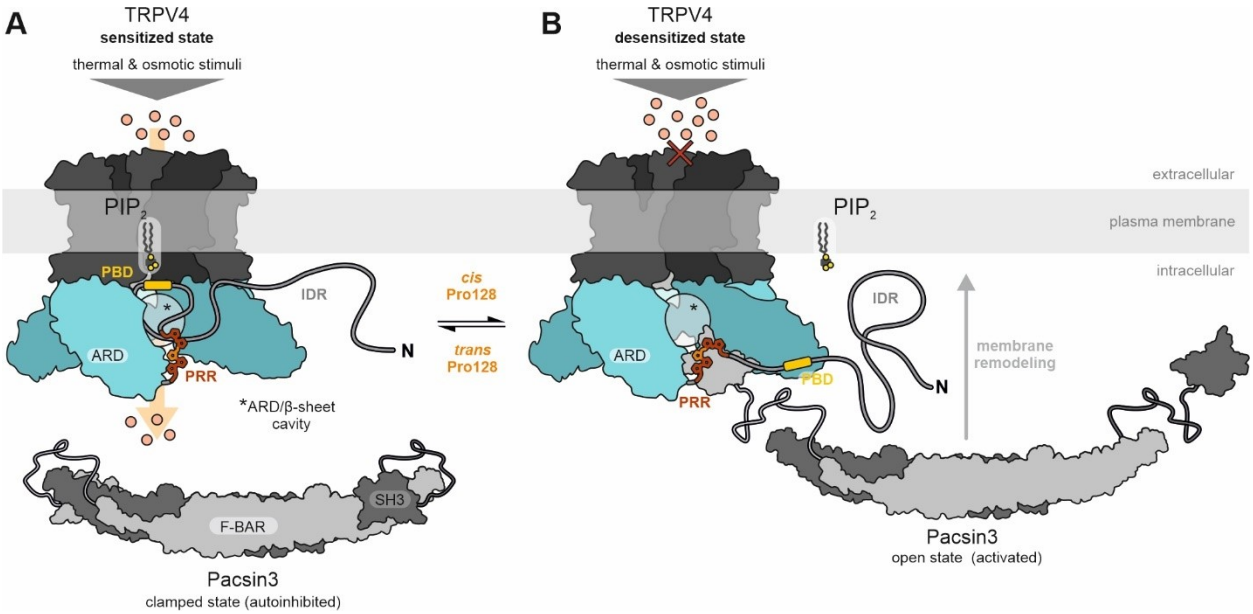


Figure 121: Mechanistic model for the Pacsin3-dependent desensitization of TRPV4 to thermal and osmotic stimuli. (**A**) In the sensitized state, TRPV4 can be activated by thermal and osmotic stimuli. Residue P128 in the PRR presumably adopts a *trans* peptide bond conformation which orients the PBD towards the plasma membrane where it can bind to the head group of a PIP_2 molecule. The PIP_2 molecule may also bind to the TRPV4 TMD to destabilize the TRP box/S4-S5 linker interaction and favour an open pore conformation or guide the IDR to the ARD/β-sheet cavity (indicated by an asterisk) as described in Figure 120. Pacsin3 is autoinhibited by an association of the SH3 domain with the F-BAR domain and does not bind to the TRPV4 PRR. (**B**) In the desensitized state, Pacsin3 binds to the TRPV4 PRR with P128 in a *cis* peptide conformation. The *trans*/*cis* switch of P128 reorients the region preceding the PRR such that the PBD loses contact to the PIP_2 head group, thus desensitizing TRPV4. This may also lead to the diffusion of the PIP_2 molecule away from the TMD such that the TRP box/S4-S5 linker interaction is stabilized and closes the pore region. A reorientation of the region preceding the PRR may also displace the IDR from the ARD/β-sheet cavity. Binding of the Pacsin3 SH3 domain to the TRPV4 PRR releases Pacsin3 from an autoinhibited state, thus leading to F-BAR domain binding to the plasma membrane followed by membrane remodelling.

The structural analysis of full-length Pacsin3 suggests that a Pacsin3-TRPV4 interaction may not only cause a reorientation of the TRPV4 N-terminus but also influence the conformational and functional state of Pacsin3 (see sectionVI3.7). A recent study on Pacsin1 proposed that Pacsin proteins exist in an autoinhibited state in which the SH3 domain interacts with the hinge region of the F-BAR domain. Binding of the SH3 domain to a substrate releases the F-BAR domain activates its membrane binding and remodelling function (*307*). The residues which mediate the SH3/F-BAR association in Pacsin1 are conserved among the Pacsin family. The observations of this study suggest that an autoinhibition mechanism, as observed for Pacsin1, also exists in Pacsin3. The integrative structural biology approach based on NMR, X-ray crystallography, SAXS, and XL-MS confirms that Pacsin3 in the apo state is in a clamped conformation in which the SH3 domain attaches to the hinge region of the F-BAR domain (see sectionVI3.7). The structural models of the Pacsin3 SH3/F-BAR and the Pacsin3 SH3/TRPV4-PRR complex offer an elegant structural explanation for the structural consequences of TRPV4 PRR binding to Pacsin3 (Figure 104). The PRR binding site in the SH3 domain overlaps with the binding site of the F-BAR domain. Thus, the Pacsin3 SH3 domain should either bind to the PRR or to the F-

BAR domain. This is indicated by NMR titration experiments in which the PRR and the isolated F-BAR domain compete for SH3 domain binding (Figure 114). Importantly, the PRR affinity for the SH3 domain appears to be higher than for the F-BAR domain. This enables the PRR to efficiently bind to the Pacsin SH3 domain. Based on the low affinity of the SH3 domain to F-BAR domain indicated in NMR titration experiments and in agreement with SAXS measurements, Pacsin3 appears to exist an equilibrium between clamped and open (SH3 detached from F-BAR) state even in the absence of a substrate (Figure 110). The TRPV4-PRR presumably stabilizes the open Pacsin3 state by introducing a steric clash between SH3 and F-BAR domain. This is experimentally supported by XL-MS experiments (Figure 116). A logical consequence of stabilizing an open Pacsin3 state is F-BAR domain binding to the plasma membrane and subsequent membrane remodelling (Figure 121 B). The afore-mentioned reorientation of the TRPV4 IDR upon Pacsin3 SH3 domain binding would potentially make room for the Pacsin3 F-BAR domain to attach to the plasma membrane.

According to the proposed Pacsin3-dependent TRPV4 desensitization mechanism, the SH3/PRR binding mode is primarily responsible for antagonizing the sensitization of TRPV4 by PIP_2 (Figure 121). However, co-expression of the isolated Pacsin3 SH3 domain with TRPV4 in HEK293 cells did not desensitize TRPV4 to temperature and hypotonicity as observed when full-length Pacsin3 was expressed (*77, 128*). This may suggest that the Pacsin3 linker and the F-BAR domain are required for TRPV4 desensitization. Alternatively, the effective Pacsin SH3 domain concentrations may be too low to achieve sufficient SH3/PRR complex formation (K_d in a low micromolar range) and consequently to desensitize TRPV4. It is important to consider that the full-length TRPV4 and Pacsin3 proteins are both membrane-associated. The increased local concentrations due to the diffusion-limited context of the PM is a feature which would have been lost when using an isolated SH3 domain. In addition, TRPV4 is a homotetramer (*42*), and Pacsin proteins are homodimers (*307, 317*). Thus, the avidity of the TRPV4/Pacsin interaction may be quite high in a physiological system. On the other hand, the functional effect of Pacsin3 on TRPV4 is most likely of a transient nature, and thus it is presumably not desirable to have a too strong interaction between these proteins if frequent switching between sensitized and desensitized ion channel states should be achieved as a response to changing environments. A *cis/trans* isomerization may be an elegant solution for a short-lived regulatory response. Proline *cis/trans* isomerization reactions have been shown to act as biological regulatory switches and even as timers in several proteins including TRP channels (*496–498*). P128 in the TRPV4 N-terminus resides in a critical region where the interactions of multiple functional modulators converge, and this proline residue is highly conserved across TRPV4 sequences (Appendix, Figure 124). Notably, the Pacsin3 SH3 domain is unlikely to catalyze a *trans/cis* isomerization of P128 itself to form a complex with the PRR. *Cis/trans* isomerization is reserved for peptidylprolyl-*cis/trans*-isomerases (PPIases), which have an entirely different fold than SH3 domains (*499*). In the afore proposed desensitization mechanism of TRPV4 (Figure 121), P128 in the PRR adopts a *trans* conformation in the sensitized state. With P128 in *trans*, the Pacsin SH3 domain cannot bind to the TRPV4 PRR. Thus, Pacsin3 binding must be preceded by isomerization of P128 from *trans* to *cis* either spontaneously (slow) or through a yet unknown PPIase-mediated enzymatic reaction (fast). A PPIase offers the possibility to fine-tune under which conditions and how fast P128 *trans/cis* isomerization and thus desensitization of TRPV4 occurs.

The structure and domain organization between Pacsin1-3 is highly conserved (*458*) and all Pacsin isoforms co-immunoprecipitate with TRPV4 (*76*) and bind equally with their SH3 domains to the TRPV4 PRR with P128 in *cis* (Figure 90 and Figure 91). One may thus expect that all Pacsin isoforms equally regulate TRPV4 function. However, there seem to exist some variations within the Pacsin family regarding their functional roles for TRPV4. Pacsin3 increased the amount of TRPV4 channels presented to the plasma membrane in HEK293 cells, presumably by inhibiting TRPV4 endocytosis. Additionally, Pacsin3 desensitizes TRPV4 to heat and osmotic stimuli. By contrast, Pacsin1 and Pacsin2 did not influence the subcellular localization of TRPV4 in HEK293 cells. The effect of Pacsin1 and Pacsin2 on TRPV4 activity was thus not investigated (*76, 77*). Pacsin1-3 and TRPV4 have overlapping tissue expression patterns (*206, 459*) and seem to be involved in related cellular processes, e.g. in the genesis and function of cilia (*98, 211, 500*) and caveolae (*494, 495, 501*) or viral infectivity (*227, 502*). Therefore, it seems reasonable to assume that all Pacsins interact with TRPV4 in specific physiological contexts to regulate TRPV4 function in yet unexplored ways. Functional variations of the three Pacsins may be rooted in different tissue distributions and isoform- or tissue-specific Pacsin interactomes. Moreover, the observation that the isolated Pacsin3 SH3 domain is not sufficient for TRPV4 desensitization supports the

notion that the linker and the F-BAR domains of Pacsin proteins may be crucial to modulate TRPV4 functionally. The F-BAR domains of the Pacsin isoforms induce different degrees of membrane curvature (*303, 307, 317, 479*). Consequently, activation of Pacsin upon binding to TRPV4 may lead to isoform-specific structural and dynamic changes in the local membrane environment of TRPV4. Bulk membrane properties are crucial for the function of membrane proteins and have recently been shown to determine the functional state of TRPV4 (*129*). Therefore, Pacsin-induced membrane curvature may be a regulatory mechanism of TRPV4 activity. Strikingly, BAR domains were recently shown to bind and accumulate phosphatidylinositides in highly curved membrane regions (*503–506*). Thus, as a side-effect of reshaping the PM, the Pacsin F-BAR domains might scavenge PIP_2 from the periphery of the TRPV4 TMD. Structural features likely to contribute to the membrane interaction of Pacsins and therefore to TRPV4 regulation are the flexible linker regions which show the strongest sequence deviations among the Pacsin family (see Appendix, Figure 125) (*458*). Disordered linkers have recently been shown to fine-tune how BAR domains sense and induce membrane curvature (*507, 508*). They may thus also determine Pacsin isoform-specific membrane curvature. Moreover, the linker regions contain binding sites for various interaction partners which may also regulate TPRV4 (*458*). The potential involvement of additional binding partners in the regulation of TRPV4 by Pacsin proteins is a yet unexplored scenario. A Pacsin dimer has two SH3 domains and two linkers which theoretically enable the simultaneous binding of multiple interaction partners to a single dimer. As an example, the Pacsin1 and Pacsin2 SH3 domains have recently been demonstrated to bind a proline-rich sequence in the E3 ubiquitin ligase ITCH (*509*), a modulator of TRPV4 subcellular localization (see section VI1.3) (*251*). This may suggest that Pacsin1 and Pacsin2 dimers can bind to TRPV4 and ITCH simultaneously. Accordingly, Pacsins might act as scaffold proteins which mediate the assembly of multicomponent TRPV4 signalling complexes in curved membrane regions. Due to the different tissue distributions between the Pacsin isoforms and subtle differences in their amino acid sequences, Pacsins also seem to exhibit isoform-specific interactomes (*458*). Tissue- and isoform-specific Pacsin interaction partners may be required to regulate TRPV4. In this sense, the increased TRPV4 presentation in the plasma membrane by Pacsin3 may be rooted in Pacsin3 specific interactions with endocytic proteins. These interactions may not occur for Pacsin1 and Pacsin2.

In summary, the results of this study complement and enhance existing structural and functional data on how Pacsin3 desensitizes TRPV4. The structural characterization of the TRPV4 PRR in isolation and in complex with the Pacsin3 SH3 highlights P128 as a potential regulatory switch for TRPV4 desensitization. Mechanistically, Pascin3 may not sensitize TRPV4 by competing with PIP_2 but by inducing a conformation of the IDR in which the PBD cannot bind to the headgroup of a PIP_2 molecule. The observation that the SH3 domains of Pacsin isoforms equally bind the TRPV4 PRR with P128 in *cis* may suggest that Pacsin1 and Pacsin2 can also desensitize TRPV4. The Pacsin3-specific increase of TRPV4 channels in the PM may depend on additional proteins of the endocytosis machinery that exclusively interact with Pacsin3. Importantly, Pacsin3 binding to TRPV4 may not only induce structural changes in the channel but also release Pacsin3 from its autoinhibited state, thus leading to a remodelling of the plasma membrane. This mechanism is likely to be conserved among all Pacsin isoforms. The Pacsin isoform-specific remodelling of the plasma membrane may directly influence the TRPV4 functional state or may link TRPV4 to specific subcellular micro-environments, e.g. caveolae or cilia, where it assembles into larger signalling complexes with additional proteins. Variations in the structural characteristics and interactomes between the Pacsin isoforms may be crucial for TRPV4 regulation and require further exploration.

5 Outlook

5.1 Decoding the PIP_2 specificity of TRPV4

Cell-based studies pointed out that TRPV4 requires the binding of PIP_2 to the PBD in the disordered TRPV4 N-terminus to sense temperature and osmotic stimuli (*128*). This thesis shows that the PBD is not restricted to PIP_2 as a ligand (sections VI3.3 and VI3.4). It can bind to anionic lipids in general. How much the affinity of PIP_2 for binding to the PBD differs from the affinity of other anionic lipids is currently unclear. In the ^{31}P NMR experiments of this study, the short-chained diC_8-PIP_2 lipid showed an affinity of ~100 μM to the IDR and the PH-PRR peptide. This affinity is comparably low, considering that an almost identical affinity of the PG-containing liposomes was determined for the same constructs. Due to the disordered state of the PBD and the non-specific lipid-binding behaviour, lipid interaction is probably primarily determined by electrostatic attraction between the PBD's basic side chains and the lipids' anionic head groups. PIP_2 is supposed to have a negative charge between -3 and -4 at physiological pH (*454*). PG has a negative charge of -1. PIP_2 should therefore exhibit a higher affinity for the PBD than PG, for instance. The discrepancies of this study may owe to the different lipid types. For a reliable assessment of lipid affinity, binding experiments need to be carried out under comparable conditions. Liposomes are presumably well-suited membrane mimics for a comparative study as they yielded reproducible binding curves in tryptophan fluorescence measurements (sections VI3.3). They are easy to handle, and their lipid compositions are broadly variable. The TRPV4 NTD should be the most appealing construct as it contains the ARD and the IDR. Tryptophan fluorescence might be the method of choice as there is no need for additional protein labelling. For studies on PIP_2-binding, it can be exploited that the TRPV4 NTD does not bind to zwitterionic lipids like PC. Thus, PC-liposomes can serve as a membrane matrix into which specific lipids can be titrated. Accordingly, one may screen conditions with physiological PIP_2 content, e.g. 2% PIP_2/98% PC and as reference 2% PG/98%PC. Titration experiments in analogy to those of this thesis with liposomes containing 100% POPG or 50% POPG with 50% POPC (sections VI3.3) should enable to deduce the affinity of the PBD to PIP2. However, it should be verified that the ARD does not bind to PIP_2-containing liposomes, as it was observed previously (*150*). This will require a complementary liposome co-sedimentation study as the ARD is not amenable to tryptophan fluorescence due to the absence of tryptophan residues in the protein. Importantly, other lipids may be included in the study such as phosphatidylserine, an abundant lipid of the PM, or cholesterol to tune membrane fluidity (*455*).

Suppose no substantial difference between PIP2 and monovalent anionic lipids are observed in liposome-based studies. In that case, it may be likely that PIP_2-specificity requires additional binding sites in TRPV4 which are not contained in the NTD, e.g. the TMD. PIP_2 has an arachidoyl chain which distinguishes it from other lipids and might help to achieve a degree of selectivity in the TMD. A potential binding site could be the proposed TMD^{LP} (section VI4.2). Validating such a hypothesis would require a more extensive effort than liposome-interaction assays. The currently most potent tool to identify lipids bound to transmembrane proteins is mass spectrometry (MS) (*510*). MS experiments would require full-length channels embedded in membranes, which can be obtained from expression in HEK293 cells or from an expression system like yeast. Full-length channels, ideally reconstituted in nanodisc systems, may then be applied to mass spectrometric analyses to identify co-purified lipids. However, instrumentation access is a rate-limiting factor for these kinds of studies. Once lipids are identified in MS, mutagenesis is required to locate the lipids' binding pockets. Here, IDR truncations may be a key to explore potentially shared lipids between the TMD and the IDR.

A relatively simple experiment to explore the regulation of TRPV4 by PIP_2 might be a domain shift of the PBD further away from the ARD. This may rule out that the PBD does not only serve to anchor the IDR to the plasma membrane by interacting with a PIP_2 molecule. If the PBD/PIP_2 interaction stabilizes a specific channel conformation which is responsible for TRPV4 sensitization, then it will presumably depend on the precise position of the PBD relative to other domains. A changed position should therefore render TRPV4 insensitive to temperature and osmotic stimuli.

An intriguing hypothesis discussed in section VI4.2 is a potential function of the IDR as a sensor of membrane structure and dynamics. It may be worth to investigate whether the lipid interaction of the IDR depends on membrane curvature or membrane fluidity. Liposome-based studies may be promising methods for this. The membrane curvature in liposomes can quite reliably be tuned by adjusting the liposome diameter. Fluidity can be adjusted by the right choice of lipid types, e.g. the acyl chain length and saturation or incorporating cholesterol. The affinity to differently curved liposomes of varying lipid fluidity can be assessed by tryptophan fluorescence and the changes in the membrane bilayer upon binding of the IDR can be monitored with a technique called Laurdan fluorescence (*511*).

5.2 Exploring the physiological function of Pacsin-TRPV4 interactions

The NMR spectroscopic investigation of the Pacsin SH3 domain interaction with the TRPV4 PRR demonstrated that all Pacsin isoforms bind equally to the PRR. Most strikingly, all Pacsin isoform SH3 domains bind to the PRR with P128 in *cis* (see section VI3.6). They might therefore stabilize the same IDR conformation. It is obvious to assume that due to their almost identical domain organisations and binding modes with TRPV4, the Pacsin isoforms should regulate TRPV4 similarly. Concerning channel trafficking, only Pacsin3 was observed to increase the number of TRPV4 channels in the PM. Because Pacsin1 and Pacsin2 did not alter TRPV4 activity, they have been disregarded in further studies on Pacsin/TRPV4 interactions (*76, 77*). However, channel trafficking and modulation of the gating behaviour presumably are rooted in different mechanisms. Given the findings of this thesis that the Pacsin SH3 domains equally stabilize cis P128 in the PRR, a proposed PIP_2-antagonizing switch, it is only logical to test a potential functional modulation of TRPV4 by Pacsin1 and Pacsin2 in cell-based systems, e.g. HEK293 cells paired with calcium imaging. A domain swap or shift of the PRR may be a facile complementary experiment to test whether Pacsins are direct modulators of TRPV4 activity or just need an interaction with TRPV4 to be released from an autoinhibited state.

This raises the next question: how is the membrane sculpting activity of Pacsins related to TRPV4? As pointed out in section VI4.3, Pacsin participates in membrane remodelling processes such as the formation of cilia and caveolae (*310, 501*). Interestingly, TRPV4 seems to localize to such environments (*47, 98, 500*). This could indicate that the membrane sensing ability of Pacsins enables the targeting of TRPV4 to specific membrane micro-environments or that curved membrane areas serve as venues for TRPV4/Pacsin interactions in physiological contexts. In such micro-environments, Pacsin proteins may act as adaptor proteins for TRPV4 because they have two SH3 domains per dimer. They could therefore bind with one SH3 domain to TRPV4 and with the other domain to a second binding partner. Mediating the assembly of TRPV4 signalling complexes might thus be another possible role for Pacsins. Identifying such signalling complex, however, is a tremendous challenge, and it will probably rely on sophisticated proteomics approaches (*512*).

The membrane sculpting activity of Pacsins may also directly regulate TRPV4 activity by remodelling the plasma membrane in the periphery of TRPV4. TRPV4 is a mechanosensitive channel, and its function was shown to be modulated by the membrane environment (*129*). It may be interesting to test how the membrane remodelling ability of Pacsin affects TRPV4 activity in HEK293 cells. One could switch the Pacsin F-BAR domain for another type of BAR domain such as N-BAR (N-terminal amphipathic helix) or I-BAR (inverse BAR domain) which are supposed to induce different membrane morphologies (*513*). Instead of BAR domains, epsin or AP180 N-terminal homology domains (ENTH or ANTH) domains might be an alternative to BAR domains (*514*). These domains not dimerize and might enable to probe the effect of a monomeric membrane-bound Pacsin SH3 domain on TRPV4 function. The combination of Pacsin chimaeras might explain whether the morphology of the membrane induced upon Pacsin binding is crucial or whether membrane binding only serves to increase the local concentration of both interaction partners.

5.3 A roadmap for elucidating the molecular mechanisms of ligand-binding, post-translational modifications, and channelopathy mutations in the TRPV4 NTD

This thesis establishes the TRPV4 IDR as a direct modulator of ARD dynamics. The ARD is a crucial regulatory domain, and its dynamics may ultimately propagate to the pore region to control ion influx. Therefore, the ARD acts as the coupling module between the IDR and the rest of the channel (section VI4.1). Modulating the communication between the ARD and the IDR might be a general mechanism in TRPV4 for sensing changes in the cytosolic domains in the transmembrane domain. The ARD/IDR communication provides a particularly appealing explanation for how ligand binding, post-translational modification (PTMs), or channelopathy mutations in the TRPV4 IDR may be coupled to TRPV4 activity. Considering the essential physiological role of TRPV4 (*206*), it is of high interest to understand the signal propagation pathway from the IDR to the TMD via the ARD in detail and how ligands, PTMs, and disease-mutations modulate it. The divide and conquer approach of this thesis provides the methodological basis to explore this. NMR and SAXS are ideal for probing the conformation and the dynamics of the IDR, whereas SAXS, HDX-MS, and XL-MS may be suitable to probe ARD dynamics and ARD-IDR communication.

Several TRPV4 interaction partners like Pacsin1/2/(3), DDX3X, RhoA, ITCH, or lipids such as PIP_2 are known to bind in the TRPV4 cytosolic domain and are attractive candidates for interaction studies with the TRPV4 NTD constructs (*76, 77, 128, 227, 251*). The exact binding modes of most interaction partners are still mostly elusive and require a proper structural characterization. Divide and conquer approaches analogous to that used for the structural and dynamic analysis of the TRPV4 NTD and how it interacts with Pacsin3 may be transferred to other interaction partners. Some of the interaction partner candidates are multidomain proteins, similar to Pacsin1-3, and contain disordered regions. It may therefore be possible to identify the TRPV4 NTD binding domains in these proteins through an additional layer of divide and conquer, i.e. testing deletion mutants for their ability to bind to the available set of TRPV4 deletion constructs. XL-MS analysis may be versatile as support for the divide and conquer strategy. The domains identified as the interacting modules can then be used for NMR spectroscopic studies provided they are within the NMR compatible size range. For more complex systems, integrative structural biology approaches combining SAXS, X-ray crystallography, NMR spectroscopy, XL-MS, HDX-MS, and potentially cryo-EM might be useful to structurally characterize interaction partners in isolation and in complex with TRPV4 fragments.

The effect of phosphorylations on the IDR-ARD communication and on how binding partners influence it can be assessed by pseudo-phosphorylation mimics, most commonly aspartate/glutamate mutants. The TRPV4 IDR contains three potential sites for phosphorylation: S88, Y110, and S134 (see section VI1.3, p. 29). They are close to binding sites of regulatory lipids and proteins, and phosphorylation may have a direct impact on how these interactors bind to the TRPV4 NTD. They may also directly affect how the IDR communicates with the ARD. The analysis of phosphorylation mimics in the NTD deletion constructs may further deepen the understanding of how PTMs contribute to the regulation of TRPV4. In analogy to PTMs, channelopathy mutations have the potential to alter the structural dynamics in the NTD. To date, approximately 20 ARD residues and six IDR residues are known to be affected by channelopathy mutations (see section VI1.3, p. 31) (*204*). The mechanisms by which these mutations deregulate TRPV4 mostly remain unclear. Besides disturbing protein folding, trafficking, and ligand binding, channelopathy mutations may also target the ARD/IDR communication as a potential key feature of channel regulation. The TRPV4 N-terminal constructs used in this thesis are ideal for testing the effect of channelopathy mutations on the structure and dynamics of the ARD, of the IDR, and how both domains communicate.

To understand whether the structural and dynamic coupling between the ARD and IDR observed *in vitro* is physiologically relevant, the constructs of the divide and conquer approach need to be transferred to cell-based investigations. TRPV4 N-terminal deletion mutants have been expressed in HEK293 cells in previous studies, including a ΔN30 and a ΔN130 mutant (*128*). These channels appeared to be correctly targeted to the plasma membrane and could be investigated for their ability to gate calcium in response to different stimuli. The TRPV4 ΔN30 mutant was fully functional whereas the ΔN130 mutant lacked sensitivity to temperature and hypotonicity due to the deleted PBD (res. 12-125 in *H. sapiens* TRPV4). This example demonstrates that it

should be possible to implement deletion constructs used in this thesis into full-length TRPV4 constructs suitable for expression and detection in HEK293 cells and subsequent functional studies, e.g. through calcium imaging or patch-clamp. An important question will be at which IDR truncation the TRPV4 gating behaviour begins to change. This may depend on the experimental setup, and a given TRPV4 IDR deletion may only alter TRPV4 activity in response to a certain stimulus or in combination with a particular interaction partner. The combination of TRPV4 deletion mutants with phosphorylation mimics and channelopathy mutations as well as with different interaction partners in cell-based setups will be a key to understand TRPV4 regulation.

The structural and functional observations from *in vitro* and cell-based experiments can be combined with *in silico* methods like structural modelling of TRPV4/ligand complexes or MD simulations of the TRPV4 NTD. The biophysical analysis of TRPV4 NTD constructs and their interaction partners through complementary and integrative methods may yield information which cannot straightforwardly be processed and translated into structural models. In such cases, computational methods can aid by combining structural restraints from XL-MS, NMR, X-ray crystallography, SAXS etc. with dynamic information from NMR, HDX-MS, or other methods which were not yet used, e.g. single molecule FRET. Furthermore, *in silico* approaches may not only help to integrate structural and functional information into models. They may also be used to explore the structural and dynamic features of ligand binding, PTMs, or channelopathy mutations in MD simulations. Structural models derived from *in vitro* studies and *in silico* modelling could provide suitable inputs for MD simulations.

The interplay between *in vitro*, *in cellulo*, and *in silico* methods as an iterative scheme holds an immense potential to understand the regulation of TRPV4 in the cytosolic domains from a structural and physiological perspective. Observations from *in vitro* studies can be directly implemented into the cell-based system and in parallel approached by *in silico* methods. An appealing advantage of the scheme is its unidirectionality, i.e. *in vitro* data can feed *in silico* and *in cellulo* studies but also *vice versa* (Figure 122).

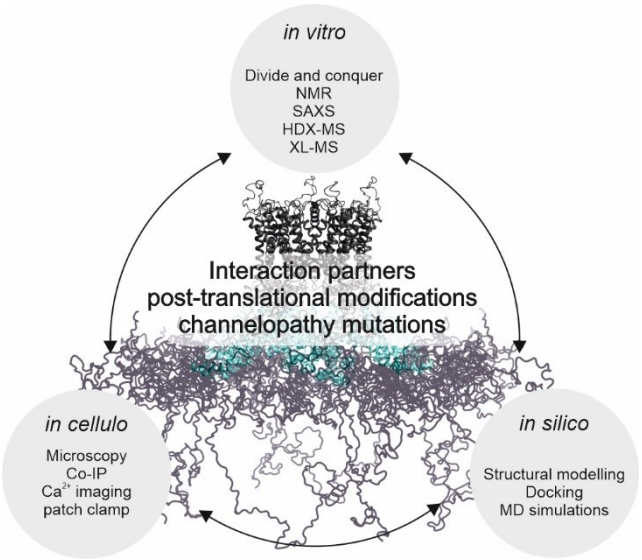


Figure 122: Iterative *in vitro*, *in cellulo*, and *in silico* approach to explore the effect of ligand-binding and modifications in the NTD on its structural dynamics and TRPV4 function. *In vitro* studies of TRPV4 NTD constructs can be based on the divide and conquer approach presented in this thesis. It exploits the soluble and monomeric state of the isolated TRPV4 NTD for structural and dynamics studies, e.g. via NMR, SAXS, HDX-MS, and XL-MS. This approach can be combined with studies on interaction partners, post-translational modifications, or channelopathy-causing mutations. The complementation of *in vitro* by cell-based studies will be critical to link structural and dynamic effects elicited by ligands, PTMs, or mutations to functional roles in the physiological context. For this, TRPV4 channels need to be expressed as wildtype proteins, deletion mutants, or phosphorylation mimics and channelopathy mutations either alone or co-expressed with an interaction partner. Channel trafficking can be assessed via microscopy, interactions with co-immunoprecipitation (Co-IP), and channel activity via calcium imaging or patch-clamp recordings. *In vitro* and *in cellulo* experiments can be complemented by *in silico* methods, e.g. the structural modelling and docking of protein interactions or molecular dynamics (MD) simulations.

5.4 Towards understanding the structure and dynamics of the IDR in full-length TRPV4

The investigation of the structural disorder in the TRPV4 N-terminus in this thesis was based on a divide and conquer approach. It exploited the ability to purify the TRPV4 NTD as a soluble and monomeric protein from *E. coli*, thus substantially facilitating the use of biophysical methods such as SAXS and NMR, key tools for the structural analysis of IDPs. However, full-length TRPV4 channels are homotetrameric transmembrane proteins (*42*). Therefore, the NMR- and SAXS-based description of the structural and dynamic ensemble of the monomeric TRPV4 NTD and IDR may not reflect the behaviour of the corresponding domains in a TRPV4 tetramer embedded in a lipid bilayer. Not only does the TRPV4 IDR interact with negatively charged membranes, it may also interact with other domains not present in the NTD, e.g. the linker region of the disordered C-terminal domain succeeding the TRP box. Moreover, the extreme flexibility observed in the isolated ARD in isolation may be suppressed in the context of a full-length TRPV channel owing to interactions with the linker region and with adjacent subunits. Nevertheless, approaching the structural disorder of the TRPV4 N-terminus in a divide and conquer approach has its eligibility as analogous investigations using full-length TRPV4 constructs face several technical and theoretical challenges. Not only is the expression of eukaryotic membrane proteins notoriously difficult and financially demanding, particularly for structural studies that require large sample amounts. Moreover, the methodologies and the experiment design has to be adapted when switching from monomeric and soluble proteins with sizes below 30 kDa to large, oligomeric membrane proteins. For instance, solution NMR spectroscopy as a standard tool for studying the structure and dynamics of IDPs with per residue resolution becomes unfeasible for a large TRPV4 tetramer.

Despite the challenges of membrane protein and IDP structural studies, recent developments in SAXS technology may be suited to approach the structural disorder in complex targets such as full-length TRPV4 channels. Unlike NMR spectroscopy, SAXS is not restricted by the size of the studied particle and applicable to membrane proteins and lipid nanodisc systems (*515–517*). SAXS may therefore be a potential method to investigate the conformational ensemble of the TRPV4 IDR in the context of the full-length protein reconstituted in lipid nanodiscs with varying lipid compositions. Because SAXS is extremely sensitive to the size of the studied particle, broad size distributions of the membrane mimic in which TRPV4 is reconstituted can make data analysis impossible. Thus, the bottleneck for SAXS studies on full-length TRPV4 will likely be the reliable and reproducible reconstitution of the protein in lipid nanodiscs, yielding narrow size distributions. N- and C-terminal deletion mutants may enable to validate the contribution of the individual domains such as the TMD, the ARD, the IDR, to the overall conformational flexibility of TRPV4. Changes in the SAXS patterns when using different lipid compositions for nanodisc reconstitutions of TRPV4 constructs may reveal how different TRPV4 domains respond to the local lipid environment. This might enable to investigate whether and under which conditions, e.g. absence and presence of PIP_2, the IDR attaches to the membrane.

The structural investigation of the IDR in full-length TRPV4 via SAXS can be complemented by probing the dynamics of full-length TRPV4 reconstituted in nanodiscs with differing lipid compositions via HDX-MS (*518*). This could be used to validate ARD and IDR dynamics concerning a potential ARD-IDR interaction in the context of full-length TRPV4 and how it depends on the lipid composition of the nanodiscs or potential protein interaction partners. HDX-MS offers a unique way to understand how conformational and dynamic changes propagate from the IDR to the pore. This method, thus, will probably be essential to elucidate the functional roles of the disordered N-terminus.

The TRPV IDRs are consistently predicted to be disordered based on their amino acid sequences (see section VI1.2). In classical structural biology approaches, disordered regions are deleted in the constructs used for structure determination. Accordingly, the IDRs were almost always deleted in cryo-EM studies of TRPV channels. A potentially groundbreaking observation of this thesis is that the disordered TRPV4 IDR stabilizes the folded ARD. Therefore, it may be beneficial to preserve the IDRs in constructs intended for cryo-EM studies. The IDR might stabilize the ARD, thus yielding a higher-resolution in the cytosolic domains. Moreover, it might be possible to trap the IDR in a state, where it associates with other regions of the channel, as observed for TRPV3 (*134*). A way for trapping the IDR might be a TRPV4 nanodisc reconstitution such that the IDR associates with the lipids in the periphery of the TRPV4 TMD. A trapped IDR would enable to investigate

conformational substates of the IDR in a full-length TRPV4 tetramer at atomic resolution. A high-resolution TRPV4 structure with at least a partly resolved IDR could serve as a template for SAXS-based modelling approaches.

5.5 Exploring the structural disorder in the TRPV subfamily

Besides TRPV4, all other thermoTRPV channels (TRPV1-4) contain large intrinsically disordered regions in their distal N-termini (see section VI1.2). It is possible that the structural and dynamic coupling between the IDR and the ARD observed in TRPV4 is a conserved feature in thermoTRPV channels, even though their IDR sequences are diverse. The divide and conquer approach used to investigate the structural disorder in the TRPV4 N-terminus may serve as a guideline for future investigations on the TRPV1-3 N-termini. Previous structural studies have demonstrated that the feasibility to express the ARD as a soluble and monomeric protein in *E. coli* is not exclusive to TRPV4. The isolated ARDs of all thermoTRPV channels can be expressed and purified in high amounts for structural studies (*18, 20, 23, 24, 161*). It is very likely that the TRPV1-3 ARDs can also be expressed in tandem with their IDRs, i.e. as the TRPV1-3 NTDs. In addition, the TRPV1-3 IDRs may be obtained analogously to the TRPV4 IDR. With these constructs in hand, the methodology including CD, NMR, SAXS, HDX-MS, XL-MS applied to the TRPV4 NTD holds excellent promises to answer questions concerning the structure and dynamics of the disordered N-termini and how they interact with the ARDs in TRPV1-3. A comparative structural and dynamic analysis of the TRPV N-terminal domains might reveal general concepts of TRPV regulation by their IDRs, or rather TRPV-specific structural and dynamic features which may explain their different polymodalities. In this regard, tt may be tempting to investigate the behaviour of thermoTRPV NTD chimaeras, i.e. exchanging the IDR in the NTD of one thermoTRPV channel by the IDRs of the other thermoTRPVs. These chimaeras may then also be tested in cell-based functional assays in the context of full-length thermoTRPV channels. IDR-chimaera TRPV channels could give exciting insights to what extent the IDRs determine the function of a thermoTRPV channel. One may also study the thermoTRPV IDRs with scrambled amino acid compositions both *in vitro* from a structural and dynamic perspective, and *in cellulo* from a functional perspective. This may reveal whether the distribution of amino acid properties or rather the overall amino acid composition in the IDRs is relevant to the TRPV4 gating behaviour.

An intriguing question ideal to address with TRPV NTD constructs is how the ARD dynamics differ between thermoTRPVs and TRPV5/6. In TRPV5/6, the N-terminal helix preceding the ARD might have a strong stabilizing effect on the ARD dynamics. The TRPV6 sARD has been used in isolation for structural studies via X-ray crystallography (*24*). However, the N-terminal helix was deleted in these studies. In the TRPV6 cryo-EM structure, the N-terminal helix not only folds on top of the ARD, but it binds the ARD/β-sheet cavity in the intersubunit regions (*41, 139, 143*). Therefore, it cannot be precluded that the TRPV5/6 ARD in tandem with the N-terminal helix is not stable in isolation. It is unclear whether the N-terminal helix is constitutively bound to the ARD, or whether it exists in various conformational states, e.g. folded onto the ARD or protruding away from it. Intriguingly, a peptide encompassing the sequence of the N-terminal TRPV6 helix was investigated via CD spectroscopy in this thesis. Unlike in the TRPV6 cryo-EM structure, the peptide does not form an α-helix. However, it could be stabilized as an α-helix by the addition of trifluoroethanol. The N-terminal helix is also an attractive target for generating chimaera constructs. In this sense, one might test whether it is possible to stabilize a thermoTRPV ARD by replacing its preceding IDR with the N-terminal TRPV5/6 helix. A comparative investigation of the structure and dynamics of the N-terminal domains between TRPV5/6 and thermoTRPVs might further explain the different gating behaviours between both subfamilies of TRPV channels.

6 References

1. D. J. Cosens, A. Manning, Abnormal electroretinogram from a Drosophila mutant. *Nature*. **224**, 285–287 (1969), doi:10.1038/224285a0.
2. C. Montell, G. M. Rubin, Molecular characterization of the drosophila trp locus: A putative integral membrane protein required for phototransduction (1989), doi:10.1016/0896-6273(89)90069-X.
3. D. E. Clapham, TRP channels as cellular sensors. *Nature*. **426**, 517–524 (2003), doi:10.1038/nature02196.
4. A. Huber *et al.,* The transient receptor potential protein (Trp), a putative store-operated Ca2+ channel essential for phosphoinositide-mediated photoreception, forms a signaling complex with NorpA, InaC and InaD. *The EMBO journal*. **15**, 7036–7045 (1996), doi:10.1002/j.1460-2075.1996.tb01095.x.
5. J. B. Startek, B. Boonen, K. Talavera, V. Meseguer, TRP Channels as Sensors of Chemically-Induced Changes in Cell Membrane Mechanical Properties. *International journal of molecular sciences*. **20** (2019), doi:10.3390/ijms20020371.
6. I. Vandewauw *et al.,* A TRP channel trio mediates acute noxious heat sensing. *Nature*. **555** (2018), doi:10.1038/nature26137.
7. R. Paricio-Montesinos *et al.,* The Sensory Coding of Warm Perception. *Neuron*. **106**, 830-841.e3 (2020), doi:10.1016/j.neuron.2020.02.035.
8. L. Moparthi, P. M. Zygmunt, Human TRPA1 is an inherently mechanosensitive bilayer-gated ion channel. *Cell calcium*. **91**, 102255 (2020), doi:10.1016/j.ceca.2020.102255.
9. R. Giniatullin, Ion Channels of Nociception. *International journal of molecular sciences*. **21** (2020), doi:10.3390/ijms21103553.
10. G. Owsianik, K. Talavera, T. Voets, B. Nilius, Permeation and selectivity of TRP channels. *Annual review of physiology*. **68** (2006), doi:10.1146/annurev.physiol.68.040204.101406.
11. I. S. Ramsey, M. Delling, D. E. Clapham, An introduction to TRP channels. *Annual review of physiology*. **68** (2006), doi:10.1146/annurev.physiol.68.040204.100431.
12. B. Nilius, TRP channels in disease. *Biochimica et biophysica acta*. **1772**, 805–812 (2007), doi:10.1016/j.bbadis.2007.02.002.
13. M. G. Madej, C. M. Ziegler, Dawning of a new era in TRP channel structural biology by cryo-electron microscopy. *Pflugers Archiv : European journal of physiology*. **470**, 213–225 (2018), doi:10.1007/s00424-018-2107-2.
14. N. J. Himmel, D. N. Cox, Transient receptor potential channels: current perspectives on evolution, structure, function and nomenclature. *Proceedings. Biological sciences*. **287**, 20201309 (2020), doi:10.1098/rspb.2020.1309.
15. H. Li, in *Transient Receptor Potential Canonical Channels and Brain Diseases*, Y. Wang, Ed. (Springer Netherlands; Imprint; Springer, Dordrecht, 2017), pp. 1–8.
16. Y. Yu *et al.,* Structural and molecular basis of the assembly of the TRPP2/PKD1 complex. *Proceedings of the National Academy of Sciences of the United States of America*. **106**, 11558–11563 (2009), doi:10.1073/pnas.0903684106.
17. Y. Yu *et al.,* Molecular mechanism of the assembly of an acid-sensing receptor ion channel complex. *Nature communications*. **3**, 1252 (2012), doi:10.1038/ncomms2257.
18. P. V. Lishko, E. Procko, X. Jin, C. B. Phelps, R. Gaudet, The ankyrin repeats of TRPV1 bind multiple ligands and modulate channel sensitivity. *Neuron*. **54**, 905–918 (2007), doi:10.1016/j.neuron.2007.05.027.
19. G. Landouré *et al.,* Mutations in TRPV4 cause Charcot-Marie-Tooth disease type 2C. *Nature genetics*. **42**, 170–174 (2010), doi:10.1038/ng.512.
20. X. Jin, J. Touhey, R. Gaudet, Structure of the N-terminal ankyrin repeat domain of the TRPV2 ion channel. *The Journal of biological chemistry*. **281**, 25006–25010 (2006), doi:10.1074/jbc.C600153200.
21. H. Inada, E. Procko, M. Sotomayor, R. Gaudet, Structural and biochemical consequences of disease-causing mutations in the ankyrin repeat domain of the human TRPV4 channel. *Biochemistry*. **51**, 6195–6206 (2012), doi:10.1021/bi300279b.
22. M. Ihara *et al.,* Molecular bases of multimodal regulation of a fungal transient receptor potential (TRP) channel. *The Journal of biological chemistry*. **288**, 15303–15317 (2013), doi:10.1074/jbc.M112.434795.
23. C. J. McCleverty, E. Koesema, A. Patapoutian, S. A. Lesley, A. Kreusch, Crystal structure of the human TRPV2 channel ankyrin repeat domain. *Protein science : a publication of the Protein Society*. **15**, 2201–2206 (2006), doi:10.1110/ps.062357206.
24. C. B. Phelps, R. J. Huang, P. V. Lishko, R. R. Wang, R. Gaudet, Structural analyses of the ankyrin repeat domain of TRPV6 and related TRPV ion channels. *Biochemistry*. **47**, 2476–2484 (2008), doi:10.1021/bi702109w.
25. Y. Fujiwara, D. L. Minor, X-ray crystal structure of a TRPM assembly domain reveals an antiparallel four-stranded coiled-coil. *Journal of Molecular Biology*. **383**, 854–870 (2008), doi:10.1016/j.jmb.2008.08.059.
26. P. R. Tsuruda, D. Julius, D. L. Minor, Coiled coils direct assembly of a cold-activated TRP channel. *Neuron*. **51**, 201–212 (2006), doi:10.1016/j.neuron.2006.06.023.
27. I. Díaz-Franulic, N. Raddatz, K. Castillo, F. D. González-Nilo, R. Latorre, A folding reaction at the C-terminal domain drives temperature sensing in TRPM8 channels. *Proceedings of the National Academy of Sciences of the United States of America*. **117**, 20298–20304 (2020), doi:10.1073/pnas.2004303117.
28. G. Q. Martinez, S. E. Gordon, Multimerization of Homo sapiens TRPA1 ion channel cytoplasmic domains. *PloS one*. **14** (2019), doi:10.1371/journal.pone.0207835.
29. M. Bycroft *et al.,* The structure of a PKD domain from polycystin-1: implications for polycystic kidney disease. *The EMBO journal*. **18**, 297–305 (1999), doi:10.1093/emboj/18.2.297.
30. E. T. Petri *et al.,* Structure of the EF-hand domain of polycystin-2 suggests a mechanism for Ca2+-dependent regulation of polycystin-2 channel activity. *Proceedings of the National Academy of Sciences of the United States of America*. **107**, 9176–9181 (2010), doi:10.1073/pnas.0912295107.
31. F. H. Schumann *et al.,* NMR-assignments of a cytosolic domain of the C-terminus of polycystin-2. *Biomolecular NMR assignments*. **3**, 141–144 (2009), doi:10.1007/s12104-009-9160-x.

32. T. L. Cvetkov, K. W. Huynh, Cohen, V. Y. Moiseenkova-Bell, Molecular architecture and subunit organization of TRPA1 ion channel revealed by electron microscopy. *The Journal of biological chemistry*. **286** (2011), doi:10.1074/jbc.M111.288993.
33. K. W. Huynh *et al.,* Structural insight into the assembly of TRPV channels. *Structure (London, England : 1993)*. **22** (2014), doi:10.1016/j.str.2013.11.008.
34. K. Mio *et al.,* The TRPC3 channel has a large internal chamber surrounded by signal sensing antennas. *Journal of Molecular Biology*. **367** (2007), doi:10.1016/j.jmb.2006.12.043.
35. Y. Maruyama *et al.,* Three-dimensional reconstruction using transmission electron microscopy reveals a swollen, bell-shaped structure of transient receptor potential melastatin type 2 cation channel. *The Journal of biological chemistry*. **282** (2007), doi:10.1074/jbc.M705694200.
36. V. Y. Moiseenkova-Bell, La Stanciu, Serysheva, II, B. J. Tobe, T. G. Wensel, Structure of TRPV1 channel revealed by electron cryomicroscopy. *Proceedings of the National Academy of Sciences of the United States of America*. **105** (2008), doi:10.1073/pnas.0711835105.
37. H. Shigematsu, T. Sokabe, R. Danev, M. Tominaga, K. Nagayama, A 3.5-nm structure of rat TRPV4 cation channel revealed by Zernike phase-contrast cryoelectron microscopy. *The Journal of biological chemistry*. **285**, 11210–11218 (2010), doi:10.1074/jbc.M109.090712.
38. W. Kühlbrandt, The resolution revolution. *Science (New York, N.Y.)*. **343**, 1443–1444 (2014), doi:10.1126/science.1251652.
39. M. Liao, E. Cao, D. Julius, Y. Cheng, Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature*. **504**, 107–112 (2013), doi:10.1038/nature12822.
40. E. Cao, M. Liao, Y. Cheng, D. Julius, TRPV1 structures in distinct conformations reveal activation mechanisms. *Nature*. **504**, 113–118 (2013), doi:10.1038/nature12823.
41. K. Saotome, A. K. Singh, M. V. Yelshanskaya, A. I. Sobolevsky, Crystal structure of the epithelial calcium channel TRPV6. *Nature*. **534**, 506–511 (2016), doi:10.1038/nature17975.
42. Z. Deng *et al.,* Cryo-EM and X-ray structures of TRPV4 reveal insight into ion permeation and gating mechanisms. *Nature structural & molecular biology*. **25**, 252–260 (2018), doi:10.1038/s41594-018-0037-5.
43. K. E. Huffer, A. A. Aleksandrova, A. Jara-Oseguera, L. R. Forrest, K. J. Swartz, Global alignment and assessment of TRP channel transmembrane domain structures to explore functional mechanisms. *eLife*. **9** (2020), doi:10.7554/eLife.58660.
44. J. Kalia, K. J. Swartz, Exploring structure-function relationships between TRP and Kv channels. *Scientific reports*. **3**, 1523 (2013), doi:10.1038/srep01523.
45. L. Hofmann *et al.,* The S4-S5 linker - gearbox of TRP channel gating. *Cell calcium*. **67**, 156–165 (2017), doi:10.1016/j.ceca.2017.04.002.
46. T. Kalstrup, R. Blunck, S4-S5 linker movement during activation and inactivation in voltage-gated K+ channels. *Proceedings of the National Academy of Sciences of the United States of America*. **115**, E6751-E6759 (2018), doi:10.1073/pnas.1719105115.
47. Q. Wang *et al.,* Lipid Interactions of a Ciliary Membrane TRP Channel: Simulation and Structural Studies of Polycystin-2. *Structure (London, England : 1993)*. **28**, 169-184.e5 (2020), doi:10.1016/j.str.2019.11.005.
48. C. Fan, W. Choi, W. Sun, J. Du, W. Lü, Structure of the human lipid-gated cation channel TRPC3. *eLife*. **7** (2018), doi:10.7554/eLife.36852.
49. Y. Gao, E. Cao, D. Julius, Y. Cheng, TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature*. **534**, 347–351 (2016), doi:10.1038/nature17964.
50. Z. Deng *et al.,* Gating of human TRPV3 in a lipid bilayer. *Nature structural & molecular biology*. **27**, 635–644 (2020), doi:10.1038/s41594-020-0428-2.
51. J. Duan *et al.,* Cryo-EM structure of TRPC5 at 2.8-Å resolution reveals unique and conserved structural elements essential for channel function. *Science advances*. **5**, eaaw7935 (2019), doi:10.1126/sciadv.aaw7935.
52. K. Venkatachalam, C. Montell, TRP channels. *Annual review of biochemistry*. **76**, 387–417 (2007), doi:10.1146/annurev.biochem.75.103004.142819.
53. C. E. Paulsen, J.-P. Armache, Y. Gao, Y. Cheng, D. Julius, Structure of the TRPA1 ion channel suggests regulatory mechanisms. *Nature*. **520**, 511–517 (2015), doi:10.1038/nature14367.
54. W. C. Tseng *et al.,* TRPA1 ankyrin repeat six interacts with a small molecule inhibitor chemotype. *Proceedings of the National Academy of Sciences of the United States of America*. **115** (2018), doi:10.1073/pnas.1808142115.
55. J. Guo *et al.,* Structures of the calcium-activated, non-selective cation channel TRPM4. *Nature*. **552**, 205–209 (2017), doi:10.1038/nature24997.
56. C. B. Phelps, R. Gaudet, The role of the N terminus and transmembrane domain of TRPM8 in channel localization and tetramerization. *The Journal of biological chemistry*. **282**, 36474–36480 (2007), doi:10.1074/jbc.M707205200.
57. L. Wang *et al.,* Structures and gating mechanism of human TRPM2. *Science (New York, N.Y.)*. **362** (2018), doi:10.1126/science.aav4809.
58. Y. Yin *et al.,* Structure of the cold- and menthol-sensing ion channel TRPM8. *Science (New York, N.Y.)*. **359**, 237–241 (2018), doi:10.1126/science.aan4325.
59. J. Duan *et al.,* Structure of the mammalian TRPM7, a magnesium channel required during embryonic development. *Proceedings of the National Academy of Sciences of the United States of America*. **115**, E8201-E8210 (2018), doi:10.1073/pnas.1810719115.
60. J. Teng, A. Anishkin, C. Kung, P. Blount, Human mutations highlight an intersubunit cation-π bond that stabilizes the closed but not open or inactivated states of TRPV channels. *Proceedings of the National Academy of Sciences of the United States of America*. **116**, 9410–9416 (2019), doi:10.1073/pnas.1820673116.
61. J. Teng, S. H. Loukin, A. Anishkin, C. Kung, L596-W733 bond between the start of the S4-S5 linker and the TRP box stabilizes the closed state of TRPV4 channel. *Proceedings of the National Academy of Sciences of the United States of America*. **112**, 3386–3391 (2015), doi:10.1073/pnas.1502366112.

62. M. Hirschi *et al.,* Cryo-electron microscopy structure of the lysosomal calcium-permeable channel TRPML3. *Nature.* **550**, 411–414 (2017), doi:10.1038/nature24055.
63. P. Schmiege, M. Fine, G. Blobel, X. Li, Human TRPML1 channel structures in open and closed conformations. *Nature.* **550**, 366–370 (2017), doi:10.1038/nature24036.
64. K. K. Viet *et al.,* Structure of the Human TRPML2 Ion Channel Extracytosolic/Lumenal Domain. *Structure (London, England : 1993).* **27**, 1246-1257.e5 (2019), doi:10.1016/j.str.2019.04.016.
65. M. Grieben *et al.,* Structure of the polycystic kidney disease TRP channel Polycystin-2 (PC2). *Nature structural & molecular biology*. **24**, 114–122 (2017), doi:10.1038/nsmb.3343.
66. A. L. Brill, B. E. Ehrlich, Polycystin 2: A calcium channel, channel partner, and regulator of calcium homeostasis in ADPKD. *Cellular signalling.* **66** (2020), doi:10.1016/j.cellsig.2019.109490.
67. K. Venkatachalam, C. O. Wong, M. X. Zhu, The role of TRPMLs in endolysosomal trafficking and function. *Cell calcium.* **58** (2015), doi:10.1016/j.ceca.2014.10.008.
68. M. Fine, P. Schmiege, X. Li, Structural basis for PtdInsP2-mediated human TRPML1 regulation. *Nature communications.* **9**, 4192 (2018), doi:10.1038/s41467-018-06493-7.
69. X. Zhang, Xinran Li, H. Xu, Phosphoinositide isoforms determine compartment-specific ion channel activity. *PNAS*. **109**, 11384–11389 (2012), doi:10.1073/pnas.1202194109.
70. M. D. Allen, S. Qamar, M. K. Vadivelu, R. N. Sandford, M. Bycroft, A high-resolution structure of the EF-hand domain of human polycystin-2. *Protein science : a publication of the Protein Society*. **23** (2014), doi:10.1002/pro.2513.
71. P. G. DeCaen, X. Liu, S. Abiria, D. E. Clapham, Atypical calcium regulation of the PKD2-L1 polycystin ion channel. *eLife*. **5** (2016), doi:10.7554/eLife.13413.
72. I. Y. Kuo *et al.,* The number and location of EF hand motifs dictates the calcium dependence of polycystin-2 function. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. **28** (2014), doi:10.1096/fj.13-247106.
73. A. Giamarchi *et al.,* A polycystin-2 (TRPP2) dimerization domain essential for the function of heteromeric polycystin complexes. *The EMBO journal.* **29**, 1176–1191 (2010), doi:10.1038/emboj.2010.18.
74. E. Nwanochie, V. N. Uversky, Structure Determination by Single-Particle Cryo-Electron Microscopy: Only the Sky (and Intrinsic Disorder) is the Limit. *International journal of molecular sciences*. **20** (2019), doi:10.3390/ijms20174186.
75. B. Goretzki *et al.,* Structural Basis of TRPV4 N Terminus Interaction with Syndapin/PACSIN1-3 and PIP2. *Structure (London, England : 1993).* **26**, 1583-1593.e5 (2018), doi:10.1016/j.str.2018.08.002.
76. M. P. Cuajungco *et al.,* PACSINs bind to the TRPV4 cation channel. PACSIN 3 modulates the subcellular localization of TRPV4. *The Journal of biological chemistry*. **281**, 18753–18762 (2006), doi:10.1074/jbc.M602452200.
77. D. D'hoedt *et al.,* Stimulus-specific modulation of the cation channel TRPV4 by PACSIN 3. *The Journal of biological chemistry*. **283**, 6272–6280 (2008), doi:10.1074/jbc.M706386200.
78. S.-Y. Lau, E. Procko, R. Gaudet, Distinct properties of Ca2+-calmodulin binding to N- and C-terminal regulatory regions of the TRPV1 channel. *The Journal of general physiology*. **140**, 541–555 (2012), doi:10.1085/jgp.201210810.
79. K. T. O'Neil, W. F. DeGrado, How calmodulin binds its targets: sequence independent recognition of amphiphilic α-helices. *Trends in Biochemical Sciences*. **15**, 59–64 (1990), doi:10.1016/0968-0004(90)90177-D.
80. B. B. Singh, X. Liu, J. Tang, M. X. Zhu, I. S. Ambudkar, Calmodulin Regulates Ca2+-Dependent Feedback Inhibition of Store-Operated Ca2+ Influx by Interaction with a Site in the C Terminus of TrpC1. *Molecular Cell*. **9**, 739–750 (2002), doi:10.1016/s1097-2765(02)00506-3.
81. Q. Tong *et al.,* Regulation of the transient receptor potential channel TRPM2 by the Ca2+ sensor calmodulin. *The Journal of biological chemistry*. **281**, 9076–9085 (2006), doi:10.1074/jbc.M510422200.
82. W. Zheng *et al.,* Direct Binding between Pre-S1 and TRP-like Domains in TRPP Channels Mediates Gating and Functional Regulation by PIP2. *Cell reports*. **22**, 1560–1573 (2018), doi:10.1016/j.celrep.2018.01.042.
83. M. Kjaergaard, B. B. Kragelund, Functions of intrinsic disorder in transmembrane proteins. *undefined* (2017).
84. B. Mészáros, G. Erdos, Z. Dosztányi, IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic acids research*. **46**, W329-W337 (2018), doi:10.1093/nar/gky384.
85. S. Vucetic, C. J. Brown, A. K. Dunker, Z. Obradovic, Flavors of protein disorder. *Proteins*. **52**, 573–584 (2003), doi:10.1002/prot.10437.
86. V. N. Uversky, J. R. Gillespie, A. L. Fink, Why are ?natively unfolded? proteins unstructured under physiologic conditions? *Proteins*. **41**, 415–427 (2000), doi:10.1002/1097-0134(20001115)41:3<415::AID-PROT130>3.0.CO;2-7.
87. P. Romero *et al.,* Sequence complexity of disordered protein. *Proteins*. **42**, 38–48 (2001), doi:10.1002/1097-0134(20010101)42:1<38::AID-PROT50>3.0.CO;2-3.
88. Romero, Obradovic, Dunker, Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. *Genome informatics. Workshop on Genome Informatics*. **8**, 110–124 (1997).
89. Garner, Romero, Dunker, Brown, Obradovic, Predicting Binding Regions within Disordered Proteins. *Genome informatics. Workshop on Genome Informatics*. **10**, 41–50 (1999).
90. Li, Romero, Rani, Dunker, Obradovic, Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome informatics. Workshop on Genome Informatics*. **10**, 30–40 (1999).
91. B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, V. N. Uversky, PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochimica et biophysica acta*. **1804**, 996–1010 (2010), doi:10.1016/j.bbapap.2010.01.011.
92. D. T. Jones, D. Cozzetto, DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics (Oxford, England)*. **31**, 857–863 (2015), doi:10.1093/bioinformatics/btu744.
93. A. Patapoutian, A. M. Peier, G. M. Story, V. Viswanath, ThermoTRP channels and beyond: mechanisms of temperature sensation. *Nature reviews. Neuroscience*. **4**, 529–539 (2003), doi:10.1038/nrn1141.

94. D. N. Cortright *et al.,* The tissue distribution and functional characterization of human VR1. *Biochemical and biophysical research communications*. **281**, 1183–1189 (2001), doi:10.1006/bbrc.2001.4482.
95. S. Ständer *et al.,* Expression of vanilloid receptor subtype 1 in cutaneous sensory nerve fibers, mast cells, and epithelial cells of appendage structures. *Experimental dermatology*. **13**, 129–139 (2004), doi:10.1111/j.0906-6705.2004.0178.x.
96. A. J. Stokes, L. M. N. Shimoda, M. Koblan-Huberson, C. N. Adra, H. Turner, A TRPV2-PKA signaling module for transduction of physical stimuli in mast cells. *The Journal of experimental medicine*. **200**, 137–147 (2004), doi:10.1084/jem.20032082.
97. J. J. Toledo Mauriño *et al.,* TRPV Subfamily (TRPV2, TRPV3, TRPV4, TRPV5, and TRPV6) Gene and Protein Expression in Patients with Ulcerative Colitis. *Journal of immunology research*. **2020**, 2906845 (2020), doi:10.1155/2020/2906845.
98. I. M. Lorenzo, W. Liedtke, M. J. Sanderson, M. A. Valverde, TRPV4 channel participates in receptor-operated calcium entry and ciliary beat frequency regulation in mouse airway epithelial cells. *PNAS*. **105**, 12611–12616 (2008), doi:10.1073/pnas.0803970105.
99. L. Birder *et al.,* Activation of urothelial transient receptor potential vanilloid 4 by 4alpha-phorbol 12,13-didecanoate contributes to altered bladder reflexes in the rat. *The Journal of pharmacology and experimental therapeutics*. **323**, 227–235 (2007), doi:10.1124/jpet.107.125435.
100. M. J. Caterina *et al.,* The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature*. **389**, 816–824 (1997), doi:10.1038/39807.
101. M. Tominaga *et al.,* The Cloned Capsaicin Receptor Integrates Multiple Pain-Producing Stimuli. *Neuron*. **21**, 531–543 (1998), doi:10.1016/s0896-6273(00)80564-4.
102. M. Raisinghani, R. M. Pabbidi, L. S. Premkumar, Activation of transient receptor potential vanilloid 1 (TRPV1) by resiniferatoxin. *The Journal of physiology*. **567**, 771–786 (2005), doi:10.1113/jphysiol.2005.087874.
103. Y. Dong *et al.,* A distinct structural mechanism underlies TRPV1 activation by piperine. *Biochemical and biophysical research communications*. **516**, 365–372 (2019), doi:10.1016/j.bbrc.2019.06.039.
104. L. J. Macpherson *et al.,* The pungency of garlic: activation of TRPA1 and TRPV1 in response to allicin. *Current biology : CB*. **15**, 929–934 (2005), doi:10.1016/j.cub.2005.04.018.
105. C. J. Bohlen *et al.,* A bivalent tarantula toxin activates the capsaicin receptor, TRPV1, by targeting the outer pore domain. *Cell*. **141**, 834–845 (2010), doi:10.1016/j.cell.2010.03.052.
106. H. Xu, N. T. Blair, D. E. Clapham, Camphor activates and strongly desensitizes the transient receptor potential vanilloid subtype 1 channel in a vanilloid-independent mechanism. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. **25**, 8924–8937 (2005), doi:10.1523/JNEUROSCI.2574-05.2005.
107. R. A. Ross, Anandamide and vanilloid TRPV1 receptors. *British journal of pharmacology*. **140**, 790–801 (2003), doi:10.1038/sj.bjp.0705467.
108. A. Dhaka *et al.,* TRPV1 is activated by both acidic and basic pH. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. **29**, 153–158 (2009), doi:10.1523/JNEUROSCI.4901-08.2009.
109. S. Mustafa, H. N. Ismael, Ethanol potentiates heat response in the carotid artery via TRPV1. *Life sciences*. **188**, 83–86 (2017), doi:10.1016/j.lfs.2017.08.037.
110. L. Liu *et al.,* Nicotine inhibits voltage-dependent sodium channels and sensitizes vanilloid receptors. *Journal of neurophysiology*. **91**, 1482–1491 (2004), doi:10.1152/jn.00922.2003.
111. O. Gouin *et al.,* TRPV1 and TRPA1 in cutaneous neurogenic and chronic inflammation: pro-inflammatory response induced by their activation and their sensitization. *Protein & cell*. **8**, 644–661 (2017), doi:10.1007/s13238-017-0395-5.
112. E. D. Prescott, D. Julius, A modular PIP2 binding site as a determinant of capsaicin receptor sensitivity. *Science (New York, N.Y.)*. **300**, 1284–1288 (2003), doi:10.1126/science.1083646.
113. G. Bhave *et al.,* Protein kinase C phosphorylation sensitizes but does not activate the capsaicin receptor transient receptor potential vanilloid 1 (TRPV1). *Proceedings of the National Academy of Sciences of the United States of America*. **100**, 12480–12485 (2003), doi:10.1073/pnas.2032100100.
114. H. H. Chuang *et al.,* Bradykinin and nerve growth factor release the capsaicin receptor from PtdIns(4,5)P2-mediated inhibition. *Nature*. **411**, 957–962 (2001), doi:10.1038/35082088.
115. M. J. Caterina, T. A. Rosen, M. Tominaga, A. J. Brake, D. Julius, A capsaicin-receptor homologue with a high threshold for noxious heat. *Nature*. **398**, 436–441 (1999), doi:10.1038/18906.
116. A. Penna *et al.,* PI3-kinase promotes TRPV2 activity independently of channel translocation to the plasma membrane. *Cell calcium*. **39**, 495–507 (2006), doi:10.1016/j.ceca.2006.01.009.
117. K. Muraki *et al.,* TRPV2 is a component of osmotically sensitive cation channels in murine aortic myocytes. *Circulation research*. **93**, 829–838 (2003), doi:10.1161/01.RES.0000097263.10220.0C.
118. G. D. Smith *et al.,* TRPV3 is a temperature-sensitive vanilloid receptor-like protein. *Nature*. **418**, 186–190 (2002), doi:10.1038/nature00894.
119. S.-E. Jordt, D. Julius, Molecular Basis for Species-Specific Sensitivity to "Hot" Chili Peppers. *Cell*. **108**, 421–430 (2002), doi:10.1016/s0092-8674(02)00637-2.
120. J. F. Doerner, H. Hatt, I. S. Ramsey, Voltage- and temperature-dependent activation of TRPV3 channels is potentiated by receptor-mediated PI(4,5)P2 hydrolysis. *The Journal of general physiology*. **137**, 271–288 (2011), doi:10.1085/jgp.200910388.
121. H.-Z. Hu *et al.,* Potentiation of TRPV3 channel function by unsaturated fatty acids. *Journal of cellular physiology*. **208**, 201–212 (2006), doi:10.1002/jcp.20648.
122. M. A. Sherkheli *et al.,* Monoterpenoids induce agonist-specific desensitization of transient receptor potential vanilloid-3 (TRPV3) ion channels. *Journal of pharmacy & pharmaceutical sciences : a publication of the Canadian Society for Pharmaceutical Sciences, Societe canadienne des sciences pharmaceutiques*. **12**, 116–128 (2009), doi:10.18433/j37c7k.
123. H. Xu, M. Delling, J. C. Jun, D. E. Clapham, Oregano, thyme and clove-derived flavors and skin sensitizers activate specific TRP channels. *Nature neuroscience*. **9**, 628–635 (2006), doi:10.1038/nn1692.

124. W. Liedtke *et al.,* Vanilloid Receptor–Related Osmotically Activated Channel (VR-OAC), a Candidate Vertebrate Osmoreceptor. *Cell.* **103**, 525–535 (2000), doi:10.1016/S0092-8674(00)00143-4.
125. R. Strotmann, C. Harteneck, K. Nunnenmacher, G. Schultz, T. D. Plant, OTRPC4, a nonselective cation channel that confers sensitivity to extracellular osmolarity. *Nature cell biology*. **2**, 695–702 (2000), doi:10.1038/35036318.
126. H. Watanabe *et al.,* Anandamide and arachidonic acid use epoxyeicosatrienoic acids to activate TRPV4 channels. *Nature*. **424**, 434–438 (2003), doi:10.1038/nature01807.
127. H. Watanabe *et al.,* Activation of TRPV4 channels (hVRL-2/mTRP12) by phorbol derivatives. *The Journal of biological chemistry*. **277**, 13569–13577 (2002), doi:10.1074/jbc.M200062200.
128. A. Garcia-Elias *et al.,* Phosphatidylinositol-4,5-biphosphate-dependent rearrangement of TRPV4 cytosolic tails enables channel activation by physiological stimuli. *Proceedings of the National Academy of Sciences of the United States of America*. **110**, 9553–9558 (2013), doi:10.1073/pnas.1220231110.
129. R. Caires *et al.,* Omega-3 Fatty Acids Modulate TRPV4 Function through Plasma Membrane Remodeling. *Cell reports*. **21**, 246–258 (2017), doi:10.1016/j.celrep.2017.09.029.
130. J. G. J. Hoenderop, B. Nilius, R. J. M. Bindels, Molecular mechanism of active Ca2+ reabsorption in the distal nephron. *Annual review of physiology*. **64**, 529–549 (2002), doi:10.1146/annurev.physiol.64.081501.155921.
131. A. K. Singh, L. L. McGoldrick, E. C. Twomey, A. I. Sobolevsky, Mechanism of calmodulin inactivation of the calcium-selective TRP channel TRPV6. *Science advances*. **4**, eaau6088 (2018), doi:10.1126/sciadv.aau6088.
132. J. van der Wijst *et al.,* A Gate Hinge Controls the Epithelial Calcium Channel TRPV5. *Scientific reports*. **7**, 45489 (2017), doi:10.1038/srep45489.
133. T. T. Lambers, A. F. Weidema, B. Nilius, J. G. J. Hoenderop, R. J. M. Bindels, Regulation of the mouse epithelial Ca2(+) channel TRPV6 by the Ca(2+)-sensor calmodulin. *The Journal of biological chemistry*. **279**, 28855–28861 (2004), doi:10.1074/jbc.M313637200.
134. L. Zubcevic, W. F. Borschel, A. L. Hsu, M. J. Borgnia, S.-Y. Lee, Regulatory switch at the cytoplasmic interface controls TRPV channel gating. *eLife*. **8** (2019), doi:10.7554/eLife.47746.
135. L. Zubcevic *et al.,* Cryo-electron microscopy structure of the TRPV2 ion channel. *Nature structural & molecular biology*. **23**, 180–186 (2016), doi:10.1038/nsmb.3159.
136. L. Zubcevic *et al.,* Conformational ensemble of the human TRPV3 ion channel. *Nature communications*. **9**, 4773 (2018), doi:10.1038/s41467-018-07117-w.
137. L. Zubcevic, A. L. Hsu, M. J. Borgnia, S.-Y. Lee, Symmetry transitions during gating of the TRPV2 ion channel in lipid membranes. *eLife*. **8** (2019), doi:10.7554/eLife.45779.
138. L. Zubcevic, S. Le, H. Yang, S.-Y. Lee, Conformational plasticity in the selectivity filter of the TRPV2 ion channel. *Nature structural & molecular biology*. **25**, 405–415 (2018), doi:10.1038/s41594-018-0059-z.
139. A. K. Singh, K. Saotome, L. L. McGoldrick, A. I. Sobolevsky, Structural bases of TRP channel TRPV6 allosteric modulation by 2-APB. *Nature communications*. **9**, 2465 (2018), doi:10.1038/s41467-018-04828-y.
140. A. K. Singh, L. L. McGoldrick, A. I. Sobolevsky, Structure and gating mechanism of the transient receptor potential channel TRPV3. *Nature structural & molecular biology*. **25**, 805–813 (2018), doi:10.1038/s41594-018-0108-7.
141. A. K. Singh *et al.,* Structural basis of temperature sensation by the TRP channel TRPV3. *Nature structural & molecular biology*. **26**, 994–998 (2019), doi:10.1038/s41594-019-0318-7.
142. R. A. Pumroy *et al.,* Molecular mechanism of TRPV2 channel modulation by cannabidiol. *eLife*. **8** (2019), doi:10.7554/eLife.48792.
143. L. L. McGoldrick *et al.,* Opening of the human epithelial calcium channel TRPV6. *Nature*. **553**, 233–237 (2018), doi:10.1038/nature25182.
144. K. W. Huynh *et al.,* Structure of the full-length TRPV2 channel by cryo-EM. *Nature communications*. **7**, 11130 (2016), doi:10.1038/ncomms11130.
145. T. E. Hughes *et al.,* Structure-based characterization of novel TRPV5 inhibitors. *eLife*. **8** (2019), doi:10.7554/eLife.49572.
146. T. E. T. Hughes *et al.,* Structural insights on TRPV5 gating by endogenous modulators. *Nature communications*. **9**, 4198 (2018), doi:10.1038/s41467-018-06753-6.
147. T. L. Dosey *et al.,* Structures of TRPV2 in distinct conformations provide insight into role of the pore turret. *Nature structural & molecular biology*. **26**, 40–49 (2019), doi:10.1038/s41594-018-0168-8.
148. T. E. T. Hughes *et al.,* Structural basis of TRPV5 channel inhibition by econazole revealed by cryo-EM. *Nature structural & molecular biology*. **25**, 53–60 (2018), doi:10.1038/s41594-017-0009-1.
149. S. Dang *et al.,* Structural insight into TRPV5 channel function and modulation. *Proceedings of the National Academy of Sciences of the United States of America*. **116**, 8869–8878 (2019), doi:10.1073/pnas.1820323116.
150. N. Takahashi *et al.,* TRPV4 channel activity is modulated by direct interaction of the ankyrin domain to PI(4,5)P2. *Nature communications*. **5**, 4994 (2014), doi:10.1038/ncomms5994.
151. C. B. Phelps, R. R. Wang, S. S. Choo, R. Gaudet, Differential regulation of TRPV1, TRPV3, and TRPV4 sensitivity through a conserved binding site on the ankyrin repeat domain. *The Journal of biological chemistry*. **285**, 731–740 (2010), doi:10.1074/jbc.M109.052548.
152. M. Arniges, J. M. Fernández-Fernández, N. Albrecht, M. Schaefer, M. A. Valverde, Human TRPV4 channel splice variants revealed a key role of ankyrin domains in multimerization and trafficking. *The Journal of biological chemistry*. **281**, 1580–1586 (2006), doi:10.1074/jbc.M511456200.
153. P. Doñate-Macián, E. Álvarez-Marimon, F. Sepulcre, J. L. Vázquez-Ibar, A. Perálvarez-Marín, The Membrane Proximal Domain of TRPV1 and TRPV2 Channels Mediates Protein¯Protein Interactions and Lipid Binding In Vitro. *International journal of molecular sciences*. **20** (2019), doi:10.3390/ijms20030682.

154. M. Kim *et al.,* Evidence that the TRPV1 S1-S4 membrane domain contributes to thermosensing. *Nature communications*. **11**, 4169 (2020), doi:10.1038/s41467-020-18026-2.
155. M. K. van Goor, L. de Jager, Y. Cheng, J. van der Wijst, High-resolution structures of transient receptor potential vanilloid channels: Unveiling a functionally diverse group of ion channels. *Protein Science*. **29**, 1569–1580 (2020), doi:10.1002/pro.3861.
156. A. Garcia-Elias *et al.,* Interaction between the Linker, Pre-S1, and TRP Domains Determines Folding, Assembly, and Trafficking of TRPV Channels. *Structure (London, England : 1993)*. **23**, 1404–1413 (2015), doi:10.1016/j.str.2015.05.018.
157. F. Yang, Y. Cui, K. Wang, J. Zheng, Thermosensitive TRP channel pore turret is part of the temperature activation pathway. *PNAS*. **107**, 7083–7088 (2010), doi:10.1073/pnas.1000357107.
158. J. Yao, B. Liu, F. Qin, Pore turret of thermal TRP channels is not essential for temperature sensing. *PNAS*. **107**, E125; author reply E126-7 (2010), doi:10.1073/pnas.1008272107.
159. G. Du *et al.,* A specialized pore turret in the mammalian cation channel TRPV1 is responsible for distinct and species-specific heat activation thresholds. *The Journal of biological chemistry*. **295**, 9641–9649 (2020), doi:10.1074/jbc.RA120.013037.
160. M. Geron *et al.,* TRPV1 pore turret dictates distinct DkTx and capsaicin gating. *PNAS*. **115**, E11837-E11846 (2018), doi:10.1073/pnas.1809662115.
161. D.-J. Shi, S. Ye, X. Cao, R. Zhang, K. Wang, Crystal structure of the N-terminal ankyrin repeat domain of TRPV3 reveals unique conformation of finger 3 loop critical for channel function. *Protein & cell*. **4**, 942–950 (2013), doi:10.1007/s13238-013-3091-0.
162. H. Shimada *et al.,* The structure of lipid nanodisc-reconstituted TRPV3 reveals the gating mechanism. *Nature structural & molecular biology*. **27**, 645–652 (2020), doi:10.1038/s41594-020-0439-z.
163. M. Tanaka *et al.,* Structure determination of the human TRPV1 ankyrin-repeat domain under nonreducing conditions. *Acta crystallographica. Section F, Structural biology communications*. **76**, 130–137 (2020), doi:10.1107/S2053230X20001533.
164. K. Elokely *et al.,* Understanding TRPV1 activation by ligands: Insights from the binding modes of capsaicin and resiniferatoxin. *PNAS*. **113**, E137-45 (2016), doi:10.1073/pnas.1517288113.
165. F. Yang, S. Vu, V. Yarov-Yarovoy, J. Zheng, Rational design and validation of a vanilloid-sensitive TRPV2 ion channel. *PNAS*. **113**, E3657-66 (2016), doi:10.1073/pnas.1604180113.
166. F. Zhang *et al.,* Engineering vanilloid-sensitivity into the rat TRPV2 channel. *eLife*. **5** (2016), doi:10.7554/eLife.16409.
167. F. Zhang, K. J. Swartz, A. Jara-Oseguera, Conserved allosteric pathways for activation of TRPV3 revealed through engineering vanilloid-sensitivity. *eLife*. **8** (2019), doi:10.7554/eLife.42756.
168. M. Numazaki *et al.,* Structural determinant of TRPV1 desensitization interacts with calmodulin. *Proceedings of the National Academy of Sciences of the United States of America*. **100**, 8002–8006 (2003), doi:10.1073/pnas.1337252100.
169. H.-Z. Hu *et al.,* 2-aminoethoxydiphenyl borate is a common activator of TRPV1, TRPV2, and TRPV3. *The Journal of biological chemistry*. **279**, 35741–35748 (2004), doi:10.1074/jbc.M404164200.
170. F. J. Taberner, G. Fernández-Ballester, A. Fernández-Carvajal, A. Ferrer-Montiel, TRP channels interaction with lipids and its implications in disease. *Biochimica et biophysica acta*. **1848**, 1818–1827 (2015), doi:10.1016/j.bbamem.2015.03.022.
171. D. W. Hilgemann, S. Feng, C. Nasuhoglu, The complex and intriguing lives of PIP2 with ion channels and transporters. *Science's STKE : signal transduction knowledge environment*. **2001**, re19 (2001), doi:10.1126/stke.2001.111.re19.
172. B. Hille, E. J. Dickson, M. Kruse, O. Vivas, B.-C. Suh, Phosphoinositides regulate ion channels. *Biochimica et biophysica acta*. **1851**, 844–856 (2015), doi:10.1016/j.bbalip.2014.09.010.
173. T. Rohacs, Phosphoinositide regulation of TRP channels. *Handbook of experimental pharmacology*. **223**, 1143–1176 (2014), doi:10.1007/978-3-319-05161-1_18.
174. V. Lukacs, J.-M. Rives, X. Sun, E. Zakharian, T. Rohacs, Promiscuous activation of transient receptor potential vanilloid 1 (TRPV1) channels by negatively charged intracellular lipids: the key role of endogenous phosphoinositides in maintaining channel activity. *The Journal of biological chemistry*. **288**, 35003–35013 (2013), doi:10.1074/jbc.M113.520288.
175. V. Lukacs *et al.,* Dual regulation of TRPV1 by phosphoinositides. *J. Neurosci.* **27**, 7070–7080 (2007), doi:10.1523/JNEUROSCI.1866-07.2007.
176. A. Y. Kim *et al.,* Pirt, a phosphoinositide-binding protein, functions as a regulatory subunit of TRPV1. *Cell*. **133**, 475–485 (2008), doi:10.1016/J.CELL.2008.02.053.
177. R. M. Klein, C. A. Ufret-Vincenty, Li Hua, S. E. Gordon, Determinants of molecular specificity in phosphoinositide regulation. Phosphatidylinositol (4,5)-bisphosphate (PI(4,5)P2) is the endogenous lipid regulating TRPV1. *The Journal of biological chemistry*. **283**, 26208–26216 (2008), doi:10.1074/jbc.M801912200.
178. B. Liu, C. Zhang, F. Qin, Functional recovery from desensitization of vanilloid receptor TRPV1 requires resynthesis of phosphatidylinositol 4,5-bisphosphate. *J. Neurosci.* **25**, 4835–4843 (2005), doi:10.1523/JNEUROSCI.1296-05.2005.
179. H. Poblete *et al.,* Molecular determinants of phosphatidylinositol 4,5-bisphosphate (PI(4,5)P2) binding to transient receptor potential V1 (TRPV1) channels. *The Journal of biological chemistry*. **290**, 2086–2098 (2015), doi:10.1074/jbc.M114.613620.
180. C. A. Ufret-Vincenty *et al.,* Mechanism for phosphoinositide selectivity and activation of TRPV1 ion channels. *The Journal of general physiology*. **145**, 431–442 (2015), doi:10.1085/jgp.201511354.
181. C. A. Ufret-Vincenty, R. M. Klein, Li Hua, J. Angueyra, S. E. Gordon, Localization of the PIP2 sensor of TRPV1 ion channels. *The Journal of biological chemistry*. **286**, 9688–9698 (2011), doi:10.1074/jbc.M110.192526.
182. E. Cao, J. F. Cordero-Morales, B. Liu, F. Qin, D. Julius, TRPV1 channels are intrinsically heat sensitive and negatively regulated by phosphoinositide lipids. *Neuron*. **77**, 667–679 (2013), doi:10.1016/j.neuron.2012.12.016.
183. O. F. Harraz, T. A. Longden, D. Hill-Eubanks, M. T. Nelson, PIP2 depletion promotes TRPV4 channel activity in mouse brain capillary endothelial cells. *eLife*. **7** (2018), doi:10.7554/eLife.38689.
184. J. Mercado, A. Gordon-Shaag, W. N. Zagotta, S. E. Gordon, Ca2+-dependent desensitization of TRPV2 channels is mediated by hydrolysis of phosphatidylinositol 4,5-bisphosphate. *J. Neurosci.* **30**, 13338–13347 (2010), doi:10.1523/JNEUROSCI.2108-10.2010.

185. C. Cao, E. Zakharian, I. Borbiro, T. Rohacs, Interplay between calmodulin and phosphatidylinositol 4,5-bisphosphate in Ca2+-induced inactivation of transient receptor potential vanilloid 6 channels. *The Journal of biological chemistry*. **288**, 5278–5290 (2013), doi:10.1074/jbc.M112.409482.
186. E. Zakharian, C. Cao, T. Rohacs, Intracellular ATP supports TRPV6 activity via lipid kinases and the generation of PtdIns(4,5) P_2. *FASEB j.* **25**, 3915–3928 (2011), doi:10.1096/fj.11-184630.
187. B. Thyagarajan, V. Lukacs, T. Rohacs, Hydrolysis of phosphatidylinositol 4,5-bisphosphate mediates calcium-induced inactivation of TRPV6 channels. *The Journal of biological chemistry*. **283**, 14980–14987 (2008), doi:10.1074/jbc.M704224200.
188. S. Brauchi *et al.,* Dissection of the components for PIP2 activation and thermosensation in TRP channels. *Proceedings of the National Academy of Sciences of the United States of America*. **104**, 10246–10251 (2007), doi:10.1073/pnas.0703420104.
189. P. Velisetty *et al.,* A molecular determinant of phosphoinositide affinity in mammalian TRPV channels. *Scientific reports*. **6**, 27652 (2016), doi:10.1038/srep27652.
190. F. Yang *et al.,* Structural mechanism underlying capsaicin binding and activation of the TRPV1 ion channel. *Nature chemical biology*. **11**, 518–524 (2015), doi:10.1038/nchembio.1835.
191. B. Holakovska, L. Grycova, J. Bily, J. Teisinger, Characterization of calmodulin binding domains in TRPV2 and TRPV5 C-tails. *Amino acids*. **40**, 741–748 (2011), doi:10.1007/s00726-010-0712-2.
192. I. Derler *et al.,* Dynamic but not constitutive association of calmodulin with rat TRPV6 channels enables fine tuning of Ca2+-dependent inactivation. *The Journal of physiology*. **577**, 31–44 (2006), doi:10.1113/jphysiol.2006.118661.
193. M. Bödding, V. Flockerzi, Ca2+ dependence of the Ca2+-selective TRPV6 channel. *The Journal of biological chemistry*. **279**, 36546–36552 (2004), doi:10.1074/jbc.M404679200.
194. T. de Groot *et al.,* Molecular mechanisms of calmodulin action on TRPV5 and modulation by parathyroid hormone. *Molecular and cellular biology*. **31**, 2845–2853 (2011), doi:10.1128/MCB.01319-10.
195. N. V. Kovalevskaya, F. M. Bokhovchuk, G. W. Vuister, The TRPV5/6 calcium channels contain multiple calmodulin binding sites with differential binding properties. *Journal of structural and functional genomics*. **13**, 91–100 (2012), doi:10.1007/s10969-012-9128-4.
196. F. M. Bokhovchuk *et al.,* The Structural Basis of Calcium-Dependent Inactivation of the Transient Receptor Potential Vanilloid 5 Channel. *Biochemistry*. **57**, 2623–2635 (2018), doi:10.1021/acs.biochem.7b01287.
197. J. Btesh, M. J. M. Fischer, K. Stott, P. A. McNaughton, Mapping the binding site of TRPV1 on AKAP79: implications for inflammatory hyperalgesia. *J. Neurosci.* **33**, 9184–9193 (2013), doi:10.1523/JNEUROSCI.4991-12.2013.
198. M. J. M. Fischer, J. Btesh, P. A. McNaughton, Disrupting sensitization of transient receptor potential vanilloid subtype 1 inhibits inflammatory hyperalgesia. *J. Neurosci.* **33**, 7407–7414 (2013), doi:10.1523/JNEUROSCI.3721-12.2013.
199. K. Mack, M. J. M. Fischer, Disrupting sensitization of TRPV4. *Neuroscience*. **352**, 1–8 (2017), doi:10.1016/j.neuroscience.2017.03.037.
200. S. Tajada *et al.,* Distance constraints on activation of TRPV4 channels by AKAP150-bound PKCα in arterial myocytes. *The Journal of general physiology*. **149**, 639–659 (2017), doi:10.1085/jgp.201611709.
201. H.-C. Fan, X. Zhang, P. A. McNaughton, Activation of the TRPV4 ion channel is enhanced by phosphorylation. *The Journal of biological chemistry*. **284**, 27884–27891 (2009), doi:10.1074/jbc.M109.028803.
202. A. Lundby *et al.,* Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. *Nature communications*. **3**, 876 (2012), doi:10.1038/ncomms1871.
203. E. Vázquez, M. A. Valverde, A review of TRP channels splicing. *Seminars in cell & developmental biology*. **17**, 607–617 (2006), doi:10.1016/j.semcdb.2006.11.004.
204. B. Nilius, T. Voets, The puzzle of TRPV4 channelopathies. *EMBO reports*. **14**, 152–163 (2013), doi:10.1038/embor.2012.219.
205. K. Bugge *et al.,* A combined computational and structural model of the full-length human prolactin receptor. *Nature communications*. **7**, 11578 (2016), doi:10.1038/ncomms11578.
206. J. P. M. White *et al.,* TRPV4: Molecular Conductor of a Diverse Orchestra. *Physiological reviews*. **96**, 911–973 (2016), doi:10.1152/physrev.00016.2015.
207. E. Guatteo *et al.,* Temperature sensitivity of dopaminergic neurons of the substantia nigra pars compacta: involvement of transient receptor potential channels. *Journal of neurophysiology*. **94**, 3069–3080 (2005), doi:10.1152/jn.00066.2005.
208. A. D. Güler *et al.,* Heat-Evoked Activation of the Ion Channel, TRPV4. *J. Neurosci.* **22**, 6408–6414 (2002), doi:10.1523/JNEUROSCI.22-15-06408.2002.
209. C. A. Lowry, S. L. Lightman, D. J. Nutt, That warm fuzzy feeling: brain serotonergic neurons and the regulation of emotion. *Journal of psychopharmacology (Oxford, England)*. **23**, 392–400 (2009), doi:10.1177/0269881108099956.
210. K. Shibasaki, M. Suzuki, A. Mizuno, M. Tominaga, Effects of body temperature on neural activity in the hippocampus: regulation of resting membrane potentials by transient receptor potential vanilloid 4. *J. Neurosci.* **27**, 1566–1575 (2007), doi:10.1523/JNEUROSCI.4284-06.2007.
211. J. Fernandes *et al.,* IP3 sensitizes TRPV4 channel to the mechano- and osmotransducing messenger 5'-6'-epoxyeicosatrienoic acid. *The Journal of cell biology*. **181**, 143–155 (2008), doi:10.1083/jcb.200712058.
212. N. S. Delany *et al.,* Identification and characterization of a novel human vanilloid receptor-like protein, VRL-2. *Physiological genomics*. **4**, 165–174 (2001), doi:10.1152/physiolgenomics.2001.4.3.165.
213. A. D. Grant *et al.,* Protease-activated receptor 2 sensitizes the transient receptor potential vanilloid 4 ion channel to cause mechanical hyperalgesia in mice. *The Journal of physiology*. **578**, 715–733 (2007), doi:10.1113/jphysiol.2006.121111.
214. J. Vriens *et al.,* Modulation of the Ca2 permeable cation channel TRPV4 by cytochrome P450 epoxygenases in vascular endothelium. *Circulation research*. **97**, 908–915 (2005), doi:10.1161/01.RES.0000187474.47805.30.
215. J. Vriens, G. Owsianik, A. Janssens, T. Voets, B. Nilius, Determinants of 4 alpha-phorbol sensitivity in transmembrane domains 3 and 4 of the cation channel TRPV4. *The Journal of biological chemistry*. **282**, 12796–12803 (2007), doi:10.1074/jbc.M610485200.

216. P. Zhang *et al.,* Nitric oxide and protein kinase G act on TRPC1 to inhibit 11,12-EET-induced vascular relaxation. *Cardiovascular research*. **104**, 138–146 (2014), doi:10.1093/cvr/cvu190.
217. K. S. Thorneloe *et al.,* N-((1S)-1-{4-((2S)-2-{(2,4-dichlorophenyl)sulfonylamino}-3-hydroxypropanoyl)-1-piperazinylcarbonyl}-3-methylbutyl)-1-benzothiophene-2-carboxamide (GSK1016790A), a novel and potent transient receptor potential vanilloid 4 channel agonist induces urinary bladder contraction and hyperactivity: Part I. *The Journal of pharmacology and experimental therapeutics*. **326**, 432–442 (2008), doi:10.1124/jpet.108.139295.
218. R. N. Willette *et al.,* Systemic activation of the transient receptor potential vanilloid subtype 4 channel causes endothelial failure and circulatory collapse: Part 2. *The Journal of pharmacology and experimental therapeutics*. **326**, 443–452 (2008), doi:10.1124/jpet.107.134551.
219. W. Tian *et al.,* Renal expression of osmotically responsive cation channel TRPV4 is restricted to water-impermeant nephron segments. *American journal of physiology. Renal physiology*. **287**, F17-24 (2004), doi:10.1152/ajprenal.00397.2003.
220. R. Fian *et al.,* The contribution of TRPV4-mediated calcium signaling to calcium homeostasis in endothelial cells. *Journal of Receptors and Signal Transduction*. **27**, 113–124 (2007), doi:10.1080/10799890701402446.
221. R. Masuyama *et al.,* Calcium/calmodulin-signaling supports TRPV4 activation in osteoclasts and regulates bone mass. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*. **27**, 1708–1721 (2012), doi:10.1002/jbmr.1629.
222. C. J. O'Conor, H. A. Leddy, H. C. Benefield, W. B. Liedtke, F. Guilak, TRPV4-mediated mechanotransduction regulates the metabolic response of chondrocytes to dynamic loading. *PNAS*. **111**, 1316–1321 (2014), doi:10.1073/pnas.1319569111.
223. S. Muramatsu *et al.,* Functional gene screening system identified TRPV4 as a regulator of chondrogenic differentiation. *The Journal of biological chemistry*. **282**, 32158–32167 (2007), doi:10.1074/jbc.M706158200.
224. Y. Jang *et al.,* Axonal neuropathy-associated TRPV4 regulates neurotrophic factor-derived axonal growth. *The Journal of biological chemistry*. **287**, 6014–6024 (2012), doi:10.1074/jbc.M111.316315.
225. D. Becker, C. Blase, J. Bereiter-Hahn, M. Jendrach, TRPV4 exhibits a functional role in cell-volume regulation. *Journal of cell science*. **118**, 2435–2440 (2005), doi:10.1242/jcs.02372.
226. H. Pivonkova *et al.,* The Contribution of TRPV4 Channels to Astrocyte Volume Regulation and Brain Edema Formation. *Neuroscience*. **394**, 127–143 (2018), doi:10.1016/j.neuroscience.2018.10.028.
227. P. Doñate-Macián *et al.,* The TRPV4 channel links calcium influx to DDX3X activity and viral infectivity. *Nature communications*. **9**, 2307 (2018), doi:10.1038/s41467-018-04776-7.
228. A. Mizuno, N. Matsumoto, M. Imai, M. Suzuki, Impaired osmotic sensation in mice lacking TRPV4. *American journal of physiology. Cell physiology*. **285**, C96-101 (2003), doi:10.1152/ajpcell.00559.2002.
229. W. Liedtke, TRPV4 plays an evolutionary conserved role in the transduction of osmotic and mechanical stimuli in live animals. *The Journal of physiology*. **567**, 53–58 (2005), doi:10.1113/jphysiol.2005.088963.
230. V. Benfenati *et al.,* An aquaporin-4/transient receptor potential vanilloid 4 (AQP4/TRPV4) complex is essential for cell-volume control in astrocytes. *PNAS*. **108**, 2563–2568 (2011), doi:10.1073/pnas.1012867108.
231. X. Liu *et al.,* A role for AQP5 in activation of TRPV4 by hypotonicity: concerted involvement of AQP5 and TRPV4 in regulation of cell volume recovery. *The Journal of biological chemistry*. **281**, 15485–15495 (2006), doi:10.1074/jbc.M600549200.
232. V. Hartmannsgruber *et al.,* Arterial response to shear stress critically depends on endothelial TRPV4 expression. *PLOS ONE*. **2**, e827 (2007), doi:10.1371/journal.pone.0000827.
233. S. Baratchi, M. Knoerzer, K. Khoshmanesh, A. Mitchell, P. McIntyre, Shear Stress Regulates TRPV4 Channel Clustering and Translocation from Adherens Junctions to the Basal Membrane. *Scientific reports*. **7**, 15942 (2017), doi:10.1038/s41598-017-16276-7.
234. C. Goswami, J. Kuhn, P. A. Heppenstall, T. Hucho, Importance of non-selective cation channel TRPV4 interaction with cytoskeleton and their reciprocal regulations in cultured cells. *PLOS ONE*. **5**, e11654 (2010), doi:10.1371/journal.pone.0011654.
235. J. D. Levine, N. Alessandri-Haber, TRP channels: targets for the relief of pain. *Biochimica et biophysica acta*. **1772**, 989–1003 (2007), doi:10.1016/j.bbadis.2007.01.008.
236. H. Todaka, J. Taniguchi, J. Satoh, A. Mizuno, M. Suzuki, Warm temperature-sensitive transient receptor potential vanilloid 4 (TRPV4) plays an essential role in thermal hyperalgesia. *The Journal of biological chemistry*. **279**, 35133–35138 (2004), doi:10.1074/jbc.M406260200.
237. N. Alessandri-Haber *et al.,* Transient receptor potential vanilloid 4 is essential in chemotherapy-induced neuropathic pain in the rat. *J. Neurosci.* **24**, 4444–4452 (2004), doi:10.1523/JNEUROSCI.0242-04.2004.
238. S. Yu *et al.,* Transient receptor potential ion-channel subfamily V member 4: a potential target for cancer treatment. *Cell death & disease*. **10**, 497 (2019), doi:10.1038/s41419-019-1708-9.
239. K.-T. Lu, T.-C. Huang, Y.-H. Tsai, Y.-L. Yang, Transient receptor potential vanilloid type 4 channels mediate Na-K-Cl-co-transporter-induced brain edema after traumatic brain injury. *Journal of neurochemistry*. **140**, 718–727 (2017), doi:10.1111/jnc.13920.
240. M. Suzuki, A. Mizuno, K. Kodaira, M. Imai, Impaired pressure sensation in mice lacking TRPV4. *The Journal of biological chemistry*. **278**, 22664–22668 (2003), doi:10.1074/jbc.M302561200.
241. Y.-C. Shin, S.-Y. Shin, I. So, D. Kwon, J.-H. Jeon, TRIP Database: a manually curated database of protein-protein interactions for mammalian TRP channels. *Nucleic acids research*. **39**, D356-61 (2011), doi:10.1093/nar/gkq814.
242. R. Strotmann, G. Schultz, T. D. Plant, Ca2+-dependent potentiation of the nonselective cation channel TRPV4 is mediated by a C-terminal calmodulin binding site. *The Journal of biological chemistry*. **278**, 26541–26549 (2003), doi:10.1074/jbc.M302590200.
243. J. Chun, S. H. Shin, S. S. Kang, The negative feedback regulation of TRPV4 Ca2+ ion channel function by its C-terminal cytoplasmic domain. *Cellular signalling*. **24**, 1918–1922 (2012), doi:10.1016/j.cellsig.2012.06.008.
244. X. Ma *et al.,* Apigenin, a plant-derived flavone, activates transient receptor potential vanilloid 4 cation channel. *British journal of pharmacology*. **166**, 349–358 (2012), doi:10.1111/j.1476-5381.2011.01767.x.

245. D. Peixoto-Neves, Q. Wang, J. H. Leal-Cardoso, L. V. Rossoni, J. H. Jaggar, Eugenol dilates mesenteric arteries and reduces systemic BP by activating endothelial cell TRPV4 channels. *British journal of pharmacology*. **172**, 3484–3494 (2015), doi:10.1111/bph.13156.
246. L. de Petrocellis *et al.,* Cannabinoid actions at TRPV channels: effects on TRPV3 and TRPV4 and their potential relevance to gastrointestinal inflammation. *Acta physiologica (Oxford, England)*. **204**, 255–266 (2012), doi:10.1111/j.1748-1716.2011.02338.x.
247. P. L. Smith, K. N. Maloney, R. G. Pothen, J. Clardy, D. E. Clapham, Bisandrographolide from Andrographis paniculata activates TRPV4 channels. *The Journal of biological chemistry*. **281**, 29897–29904 (2006), doi:10.1074/jbc.M605394200.
248. F. Vincent, M. A. J. Duncton, TRPV4 agonists and antagonists. *Current topics in medicinal chemistry*. **11**, 2216–2226 (2011), doi:10.2174/156802611796904861.
249. R. Strotmann, M. Semtner, F. Kepura, T. D. Plant, T. Schöneberg, Interdomain interactions control Ca2+-dependent potentiation in the cation channel TRPV4. *PLOS ONE*. **5**, e10580 (2010), doi:10.1371/journal.pone.0010580.
250. S. H. Loukin, J. Teng, C. Kung, A channelopathy mechanism revealed by direct calmodulin activation of TrpV4. *PNAS*. **112**, 9400–9405 (2015), doi:10.1073/pnas.1510602112.
251. T. Wegierski, K. Hill, M. Schaefer, G. Walz, The HECT ubiquitin ligase AIP4 regulates the cell surface expression of select TRP channels. *The EMBO journal*. **25**, 5659–5669 (2006), doi:10.1038/sj.emboj.7601429.
252. Y. Wang *et al.,* OS-9 regulates the transit and polyubiquitination of TRPV4 in the endoplasmic reticulum. *The Journal of biological chemistry*. **282**, 36561–36570 (2007), doi:10.1074/jbc.M703903200.
253. S. H. Shin, E. J. Lee, J. Chun, S. Hyun, S. S. Kang, Phosphorylation on TRPV4 Serine 824 Regulates Interaction with STIM1. *The open biochemistry journal*. **9**, 24–33 (2015), doi:10.2174/1874091X01509010024.
254. S. Cao *et al.,* Transient receptor potential vanilloid 4 (TRPV4) activation by arachidonic acid requires protein kinase A-mediated phosphorylation. *The Journal of biological chemistry*. **293**, 5307–5322 (2018), doi:10.1074/jbc.M117.811075.
255. H. Peng *et al.,* Identification of a Protein Kinase C-dependent phosphorylation site involved in sensitization of TRPV4 channel. *Biochemical and biophysical research communications*. **391**, 1721–1725 (2010), doi:10.1016/j.bbrc.2009.12.140.
256. S. H. Shin *et al.,* Phosphorylation on the Ser 824 residue of TRPV4 prefers to bind with F-actin than with microtubules to expand the cell surface area. *Cellular signalling*. **24**, 641–651 (2012), doi:10.1016/j.cellsig.2011.11.002.
257. T. Wegierski, U. Lewandrowski, B. Müller, A. Sickmann, G. Walz, Tyrosine phosphorylation modulates the activity of TRPV4 in response to defined stimuli. *The Journal of biological chemistry*. **284**, 2923–2933 (2009), doi:10.1074/jbc.M805357200.
258. H. Xu *et al.,* Regulation of a transient receptor potential (TRP) channel by tyrosine phosphorylation. SRC family kinase-dependent tyrosine phosphorylation of TRPV4 on TYR-253 mediates its response to hypotonic stress. *The Journal of biological chemistry*. **278**, 11520–11527 (2003), doi:10.1074/jbc.M211061200.
259. E. J. Lee *et al.,* The modulation of TRPV4 channel activity through its Ser 824 residue phosphorylation by SGK1. *Animal Cells and Systems*. **14**, 99–114 (2010), doi:10.1080/19768354.2010.486939.
260. N. Alessandri-Haber, O. A. Dina, E. K. Joseph, D. Reichling, J. D. Levine, A transient receptor potential vanilloid 4-dependent mechanism of hyperalgesia is engaged by concerted action of inflammatory mediators. *J. Neurosci.* **26**, 3864–3874 (2006), doi:10.1523/JNEUROSCI.5385-05.2006.
261. D. M. Kurrasch-Orbaugh, J. C. Parrish, V. J. Watts, D. E. Nichols, A complex signaling cascade links the serotonin2A receptor to phospholipase A2 activation: the involvement of MAP kinases. *Journal of neurochemistry*. **86**, 980–991 (2003), doi:10.1046/j.1471-4159.2003.01921.x.
262. A. Garcia-Elias, I. M. Lorenzo, R. Vicente, M. A. Valverde, IP3 receptor binds to and sensitizes TRPV4 channel to osmotic stimuli via a calmodulin-binding site. *The Journal of biological chemistry*. **283**, 31284–31288 (2008), doi:10.1074/jbc.C800184200.
263. J. Vriens *et al.,* Cell swelling, heat, and chemical agonists use distinct pathways for the activation of the cation channel TRPV4. *Proceedings of the National Academy of Sciences of the United States of America*. **101**, 396–401 (2004), doi:10.1073/pnas.0303329101.
264. A. Berna-Erro *et al.,* Structural determinants of 5',6'-epoxyeicosatrienoic acid binding to and activation of TRPV4 channel. *Scientific reports*. **7**, 10522 (2017), doi:10.1038/s41598-017-11274-1.
265. J. Teng, S. H. Loukin, A. Anishkin, C. Kung, A competing hydrophobic tug on L596 to the membrane core unlatches S4-S5 linker elbow from TRP helix and allows TRPV4 channel to open. *PNAS*. **113**, 11847–11852 (2016), doi:10.1073/pnas.1613523113.
266. G. Askarieh *et al.,* Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. *Nature*. **465**, 236–238 (2010), doi:10.1038/nature08962.
267. N. Kronqvist *et al.,* Sequential pH-driven dimerization and stabilization of the N-terminal domain enables rapid spider silk formation. *Nature communications*. **5**, 3254 (2014), doi:10.1038/ncomms4254.
268. B. Nilius, J. Vriens, J. Prenen, G. Droogmans, T. Voets, TRPV4 calcium entry channel: a paradigm for gating diversity. *American journal of physiology. Cell physiology*. **286**, C195-205 (2004), doi:10.1152/ajpcell.00365.2003.
269. S. L. Zhang *et al.,* STIM1 is a Ca2+ sensor that activates CRAC channels and migrates from the Ca2+ store to the plasma membrane. *Nature*. **437**, 902–905 (2005), doi:10.1038/nature04147.
270. E. J. Lee, S. H. Shin, S. Hyun, J. Chun, S. S. Kang, Mutation of a putative S-nitrosylation site of TRPV4 protein facilitates the channel activates. *Animal Cells and Systems*. **15**, 95–106 (2011), doi:10.1080/19768354.2011.555183.
271. A. Patapoutian, S. Tate, C. J. Woolf, Transient receptor potential channels: targeting pain at the source. *Nature reviews. Drug discovery*. **8**, 55–68 (2009), doi:10.1038/nrd2757.
272. N. Alessandri-Haber, O. A. Dina, E. K. Joseph, D. B. Reichling, J. D. Levine, Interaction of transient receptor potential vanilloid 4, integrin, and SRC tyrosine kinase in mechanical hyperalgesia. *J. Neurosci.* **28**, 1046–1057 (2008), doi:10.1523/JNEUROSCI.4497-07.2008.
273. X. Liu *et al.,* Activation of PTEN by inhibition of TRPV4 suppresses colon cancer development. *Cell death & disease*. **10**, 460 (2019), doi:10.1038/s41419-019-1700-4.

274. W. H. Lee *et al.,* TRPV4 plays a role in breast cancer cell migration via Ca2+-dependent activation of AKT and downregulation of E-cadherin cell cortex protein. *Oncogenesis*. **6**, e338 (2017), doi:10.1038/oncsis.2017.39.
275. W. H. Lee *et al.,* TRPV4 Regulates Breast Cancer Cell Extravasation, Stiffness and Actin Cortex. *Scientific reports*. **6**, 27903 (2016), doi:10.1038/srep27903.
276. K. Suresh, mtROS-Induced TRPV4 Activation in Traumatic Brain Injury. *Journal of neurotrauma*. **36**, 639 (2019), doi:10.1089/neu.2018.5847.
277. G. Nishimura *et al.,* TRPV4-associated skeletal dysplasias. *American journal of medical genetics. Part C, Seminars in medical genetics*. **160C**, 190–204 (2012), doi:10.1002/ajmg.c.31335.
278. S. S. Kang, S. H. Shin, C.-K. Auh, J. Chun, Human skeletal dysplasia caused by a constitutive activated transient receptor potential vanilloid 4 (TRPV4) cation channel mutation. *Experimental & molecular medicine*. **44**, 707–722 (2012), doi:10.3858/emm.2012.44.12.080.
279. T.-J. Cho *et al.,* TRPV4-pathy manifesting both skeletal dysplasia and peripheral neuropathy: a report of three patients. *American journal of medical genetics. Part A*. **158A**, 795–802 (2012), doi:10.1002/ajmg.a.35268.
280. E. Andreucci *et al.,* TRPV4 related skeletal dysplasias: a phenotypic spectrum highlighted byclinical, radiographic, and molecular studies in 21 new families. *Orphanet journal of rare diseases*. **6**, 37 (2011), doi:10.1186/1750-1172-6-37.
281. H. A. Leddy *et al.,* Follistatin in chondrocytes: the link between TRPV4 channelopathies and skeletal malformations. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. **28**, 2525–2537 (2014), doi:10.1096/fj.13-245936.
282. S. Aharoni *et al.,* Striking phenotypic variability in familial TRPV4-axonal neuropathy spectrum disorder. *American journal of medical genetics. Part A*. **155A**, 3153–3156 (2011), doi:10.1002/ajmg.a.34327.
283. M. McEntagart, TRPV4 axonal neuropathy spectrum disorder. *Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia*. **19**, 927–933 (2012), doi:10.1016/j.jocn.2011.12.003.
284. S. R. Lamandé *et al.,* Mutations in TRPV4 cause an inherited arthropathy of hands and feet. *Nature genetics*. **43**, 1142–1146 (2011), doi:10.1038/ng.945.
285. W. Tian *et al.,* A loss-of-function nonsynonymous polymorphism in the osmoregulatory TRPV4 gene is associated with human hyponatremia. *Proceedings of the National Academy of Sciences of the United States of America*. **106**, 14034–14039 (2009), doi:10.1073/pnas.0904084106.
286. H.-X. Deng *et al.,* Scapuloperoneal spinal muscular atrophy and CMT2C are allelic disorders caused by alterations in TRPV4. *Nature genetics*. **42**, 165–169 (2010), doi:10.1038/ng.509.
287. C. J. Klein *et al.,* TRPV4 mutations and cytotoxic hypercalcemia in axonal Charcot-Marie-Tooth neuropathies. *Neurology*. **76**, 887–894 (2011), doi:10.1212/WNL.0b013e31820f2de3.
288. F. Fecto *et al.,* Mutant TRPV4-mediated toxicity is linked to increased constitutive function in axonal neuropathies. *The Journal of biological chemistry*. **286**, 17281–17291 (2011), doi:10.1074/jbc.M111.237685.
289. J. Dai *et al.,* Novel and recurrent TRPV4 mutations and their association with distinct phenotypes within the TRPV4 dysplasia family. *Journal of medical genetics*. **47**, 704–709 (2010), doi:10.1136/jmg.2009.075358.
290. D.-H. Chen *et al.,* CMT2C with vocal cord paresis associated with short stature and mutations in the TRPV4 gene. *Neurology*. **75**, 1968–1975 (2010), doi:10.1212/WNL.0b013e3181ffe4bb.
291. G. Nishimura *et al.,* Spondylo-epiphyseal dysplasia, Maroteaux type (pseudo-Morquio syndrome type 2), and parastremmatic dysplasia are caused by TRPV4 mutations. *American journal of medical genetics. Part A*. **152A**, 1443–1449 (2010), doi:10.1002/ajmg.a.33414.
292. N. Camacho *et al.,* Dominant TRPV4 mutations in nonlethal and lethal metatropic dysplasia. *American journal of medical genetics. Part A*. **152A**, 1169–1177 (2010), doi:10.1002/ajmg.a.33392.
293. S. Unger *et al.,* Fetal akinesia in metatropic dysplasia: The combined phenotype of chondrodysplasia and neuropathy? *American journal of medical genetics. Part A*. **155A**, 2860–2864 (2011), doi:10.1002/ajmg.a.34268.
294. M. J. Rock *et al.,* Gain-of-function mutations in TRPV4 cause autosomal dominant brachyolmia. *Nature genetics*. **40**, 999–1003 (2008), doi:10.1038/ng.166.
295. M. Zimoń *et al.,* Dominant mutations in the cation channel gene transient receptor potential vanilloid 4 cause an unusual spectrum of neuropathies. *Brain : a journal of neurology*. **133**, 1798–1809 (2010), doi:10.1093/brain/awq109.
296. K. A. Fawcett *et al.,* Comprehensive analysis of the TRPV4 gene in a large series of inherited neuropathies and controls. *Journal of neurology, neurosurgery, and psychiatry*. **83** (2012), doi:10.1136/jnnp-2012-303055.
297. C. Fiorillo *et al.,* TRPV4 mutations in children with congenital distal spinal muscular atrophy. *Neurogenetics*. **13** (2012), doi:10.1007/s10048-012-0328-7.
298. J. Velilla *et al.,* Homozygous TRPV4 mutation causes congenital distal spinal muscular atrophy and arthrogryposis. *Neurology. Genetics*. **5** (2019), doi:10.1212/NXG.0000000000000312.
299. M. Suzuki, A. Hirao, A. Mizuno, Microtubule-associated corrected protein 7 increases the membrane expression of transient receptor potential vanilloid 4 (TRPV4). *The Journal of biological chemistry*. **278**, 51448–51453 (2003), doi:10.1074/jbc.M308212200.
300. J. W. Putney, T. Tomita, Phospholipase C signaling and calcium influx. *Advances in biological regulation*. **52**, 152–164 (2012), doi:10.1016/j.advenzreg.2011.09.005.
301. M. M. Kessels, B. Qualmann, The syndapin protein family: linking membrane trafficking with the cytoskeleton. *Journal of cell science*. **117**, 3077–3086 (2004), doi:10.1242/jcs.01290.
302. B. M. Woolums *et al.,* TRPV4 disrupts mitochondrial transport and causes axonal degeneration via a CaMKII-dependent elevation of intracellular Ca2. *Nature communications*. **11**, 2679 (2020), doi:10.1038/s41467-020-16411-5.

303. Q. Wang *et al.,* Molecular mechanism of membrane constriction and tubulation mediated by the F-BAR protein Pacsin/Syndapin. *Proceedings of the National Academy of Sciences of the United States of America*. **106**, 12700–12705 (2009), doi:10.1073/pnas.0902974106.
304. N. Naslavsky, S. Caplan, C-terminal EH-domain-containing proteins: consensus for a role in endocytic trafficking, EH? *Journal of cell science*. **118**, 4093–4101 (2005), doi:10.1242/jcs.02595.
305. B. Qualmann, R. B. Kelly, Syndapin isoforms participate in receptor-mediated endocytosis and actin organization. *The Journal of cell biology*. **148**, 1047–1062 (2000), doi:10.1083/jcb.148.5.1047.
306. B. Qualmann, J. Roos, P. J. DiGregorio, R. B. Kelly, Syndapin I, a synaptic dynamin-binding protein that associates with the neural Wiskott-Aldrich syndrome protein. *Molecular biology of the cell*. **10**, 501–513 (1999), doi:10.1091/mbc.10.2.501.
307. Y. Rao *et al.,* Molecular basis for SH3 domain regulation of F-BAR-mediated membrane deformation. *Proceedings of the National Academy of Sciences of the United States of America*. **107**, 8213–8218 (2010), doi:10.1073/pnas.1003478107.
308. V. Anggono *et al.,* Syndapin I is the phosphorylation-regulated dynamin I partner in synaptic vesicle endocytosis. *Nature neuroscience*. **9**, 752–760 (2006), doi:10.1038/nn1695.
309. A. Quan *et al.,* Phosphorylation of syndapin I F-BAR domain at two helix-capping motifs regulates membrane tubulation. *PNAS*. **109**, 3760–3765 (2012), doi:10.1073/pnas.1108294109.
310. Y. Senju, S. Suetsugu, Possible regulation of caveolar endocytosis and flattening by phosphorylation of F-BAR domain protein PACSIN2/Syndapin II. *Bioarchitecture*. **5**, 70–77 (2015), doi:10.1080/19490992.2015.1128604.
311. X. Bai, X. Zheng, Tip-to-tip interaction in the crystal packing of PACSIN 2 is important in regulating tubulation activity. *Protein & cell*. **4**, 695–701 (2013), doi:10.1007/s13238-013-3041-x.
312. N. Kassem *et al.*, *Order and disorder – an integrative structure of the full-length human growth hormone receptor* (2020).
313. A. Prestel, K. Bugge, L. Staby, R. Hendus-Altenburger, B. B. Kragelund, Characterization of Dynamic IDP Complexes by NMR Spectroscopy. *Methods in enzymology*. **611**, 193–226 (2018), doi:10.1016/bs.mie.2018.08.026.
314. R. Schneider, M. Blackledge, M. R. Jensen, Elucidating binding mechanisms and dynamics of intrinsically disordered protein complexes using NMR spectroscopy. *Current opinion in structural biology*. **54**, 10–18 (2019), doi:10.1016/j.sbi.2018.09.007.
315. A. Schramm *et al.,* An arsenal of methods for the experimental characterization of intrinsically disordered proteins - How to choose and combine them? *Archives of biochemistry and biophysics*. **676**, 108055 (2019), doi:10.1016/j.abb.2019.07.020.
316. L. Luo *et al.,* The Binding of Syndapin SH3 Domain to Dynamin Proline-rich Domain Involves Short and Long Distance Elements. *The Journal of biological chemistry*. **291**, 9411–9424 (2016), doi:10.1074/jbc.M115.703108.
317. X. Bai, G. Meng, M. Luo, X. Zheng, Rigidity of wedge loop in PACSIN 3 protein is a key factor in dictating diameters of tubules. *The Journal of biological chemistry*. **287**, 22387–22396 (2012), doi:10.1074/jbc.M112.358960.
318. M. L. Karlsen *et al.,* Structure of Dimeric and Tetrameric Complexes of the BAR Domain Protein PICK1 Determined by Small-Angle X-Ray Scattering. *Structure (London, England : 1993)*. **23**, 1258–1270 (2015), doi:10.1016/j.str.2015.04.020.
319. S. Bhattacharya, C. B. Stanley, W. T. Heller, P. A. Friedman, Z. Bu, Dynamic structure of the full-length scaffolding protein NHERF1 influences signaling complex assembly. *The Journal of biological chemistry*. **294**, 11297–11310 (2019), doi:10.1074/jbc.RA119.008218.
320. A. Grishaev, J. Wu, J. Trewhella, A. Bax, Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *Journal of the American Chemical Society*. **127**, 16621–16628 (2005), doi:10.1021/ja054342m.
321. K. T. Debiec, M. J. Whitley, L. M. I. Koharudin, L. T. Chong, A. M. Gronenborn, Integrating NMR, SAXS, and Atomistic Simulations: Structure and Dynamics of a Two-Domain Protein. *Biophysical Journal*. **114**, 839–855 (2018), doi:10.1016/j.bpj.2018.01.001.
322. L. L. Schrödinger, "The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint, Version 1.8". (2015).
323. L. L. Schrödinger, "The PyMOL Molecular Graphics System, Version 1.8". (2015).
324. D. G. Gibson, Enzymatic assembly of overlapping DNA fragments. *Methods in enzymology*. **498**, 349–361 (2011), doi:10.1016/B978-0-12-385120-8.00015-2.
325. M. P. Malakhov *et al.,* SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *Journal of structural and functional genomics*. **5**, 75–86 (2004), doi:10.1023/B:JSFG.0000029237.70316.52.
326. S. Kalva, J. D. Boeke, P. Mita, Gibson Deletion: a novel application of isothermal in vitro recombination. *Biological procedures online*. **20**, 2 (2018), doi:10.1186/s12575-018-0068-7.
327. H. Liu, J. H. Naismith, An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC biotechnology*. **8**, 91 (2008), doi:10.1186/1472-6750-8-91.
328. V. N. Uversky, Size-exclusion chromatography in structural analysis of intrinsically disordered proteins. *Methods in molecular biology (Clifton, N.J.)*. **896**, 179–194 (2012), doi:10.1007/978-1-4614-3704-8_11.
329. D.-M. Smilgies, E. Folta-Stogniew, Molecular weight-gyration radius relation of globular proteins: a comparison of light scattering, small-angle X-ray scattering and structure-based data. *Journal of applied crystallography*. **48**, 1604–1606 (2015), doi:10.1107/S1600576715015551.
330. D. Some, H. Amartely, A. Tsadok, M. Lebendiker, Characterization of Proteins by Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS). *Journal of visualized experiments : JoVE* (2019), doi:10.3791/59615.
331. H. Zhao, P. H. Brown, P. Schuck, On the distribution of protein refractive index increments. *Biophysical Journal*. **100**, 2309–2317 (2011), doi:10.1016/j.bpj.2011.03.004.
332. B. H. Zimm, The Scattering of Light and the Radial Distribution Function of High Polymer Solutions. *The Journal of Chemical Physics*. **16**, 1093–1099 (1948), doi:10.1063/1.1746738.
333. J. M. Walker, *The Proteomics Protocols Handbook* (Humana Press, Totowa, NJ, 2005).
334. M. M. Julkowska, J. M. Rankenberg, C. Testerink, Liposome-binding assays to assess specificity and affinity of phospholipid-protein interactions. *Methods in molecular biology (Clifton, N.J.)*. **1009**, 261–271 (2013), doi:10.1007/978-1-62703-401-2_24.

335. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nature methods*. **9**, 671–675 (2012), doi:10.1038/nmeth.2089.
336. N. J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols*. **1**, 2876–2890 (2006), doi:10.1038/nprot.2006.202.
337. S. M. Kelly, T. J. Jess, N. C. Price, How to study proteins by circular dichroism. *Biochimica et biophysica acta*. **1751**, 119–139 (2005), doi:10.1016/j.bbapap.2005.06.005.
338. L. A. Abriata, A Simple Spreadsheet Program To Simulate and Analyze the Far-UV Circular Dichroism Spectra of Proteins. *J. Chem. Educ.* **88**, 1268–1273 (2011), doi:10.1021/ed200060t.
339. A. Micsonai *et al.,* Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*. **112**, E3095-103 (2015), doi:10.1073/pnas.1500851112.
340. A. B. T. Ghisaidoobe, S. J. Chung, Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Förster resonance energy transfer techniques. *International journal of molecular sciences*. **15**, 22518–22538 (2014), doi:10.3390/ijms151222518.
341. C. A. Kraft, J. L. Garrido, L. Leiva-Vega, G. Romero, Quantitative analysis of protein-lipid interactions using tryptophan fluorescence. *Science signaling*. **2**, pl4 (2009), doi:10.1126/scisignal.299pl4.
342. N. Hellmann, D. Schneider, Hands On: Using Tryptophan Fluorescence Spectroscopy to Study Protein Structure. *Methods in molecular biology (Clifton, N.J.)*. **1958**, 379–401 (2019), doi:10.1007/978-1-4939-9161-7_20.
343. R. Kazlauskas, Engineering more stable proteins. *Chemical Society reviews*. **47**, 9026–9045 (2018), doi:10.1039/c8cs00014j.
344. B. J. Bennion, V. Daggett, The molecular basis for the chemical denaturation of proteins by urea. *Proceedings of the National Academy of Sciences of the United States of America*. **100**, 5142–5147 (2003), doi:10.1073/pnas.0930122100.
345. T. Weidemann, Application of fluorescence correlation spectroscopy (FCS) to measure the dynamics of fluorescent proteins in living cells. *Methods in molecular biology (Clifton, N.J.)*. **1076**, 539–555 (2014), doi:10.1007/978-1-62703-649-8_24.
346. A. M. Melo, M. Prieto, A. Coutinho, Quantifying lipid-protein interaction by fluorescence correlation spectroscopy (FCS). *Methods in molecular biology (Clifton, N.J.)*. **1076**, 575–595 (2014), doi:10.1007/978-1-62703-649-8_26.
347. F. P. Schäfer, J. P. Toennies, W. Zinth, R. Rigler, E. S. Elson, *Fluorescence Correlation Spectroscopy* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2001).
348. Y. Kim *et al.,* Efficient site-specific labeling of proteins via cysteines. *Bioconjugate chemistry*. **19**, 786–791 (2008), doi:10.1021/bc7002499.
349. A. N. Holding, XL-MS: Protein cross-linking coupled with mass spectrometry. *Methods (San Diego, Calif.)*. **89**, 54–63 (2015), doi:10.1016/j.ymeth.2015.06.010.
350. A. Sinz, Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein-Protein Interactions: Where Are We Now and Where Should We Go from Here? *Angewandte Chemie (International ed. in English)*. **57**, 6390–6396 (2018), doi:10.1002/anie.201709559.
351. E. D. Merkley *et al.,* Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein science : a publication of the Protein Society*. **23**, 747–759 (2014), doi:10.1002/pro.2458.
352. A. Leitner, T. Walzthoeni, R. Aebersold, Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nature protocols*. **9**, 120–137 (2014), doi:10.1038/nprot.2013.168.
353. C. W. Combe, L. Fischer, J. Rappsilber, xiNET: cross-link network maps with residue resolution. *Molecular & cellular proteomics : MCP*. **14**, 1137–1147 (2015), doi:10.1074/mcp.O114.042259.
354. E. A. Hodge, M. A. Benhaim, K. K. Lee, Bridging protein structure, dynamics, and function using hydrogen/deuterium-exchange mass spectrometry. *Protein science : a publication of the Protein Society*. **29**, 843–855 (2020), doi:10.1002/pro.3790.
355. O. Vadas, M. L. Jenkins, G. L. Dornan, J. E. Burke, Using Hydrogen-Deuterium Exchange Mass Spectrometry to Examine Protein-Membrane Interactions. *Methods in enzymology*. **583**, 143–172 (2017), doi:10.1016/bs.mie.2016.09.008.
356. D. D. Weis, T. E. Wales, J. R. Engen, M. Hotchko, L. F. ten Eyck, Identification and characterization of EX1 kinetics in H/D exchange mass spectrometry by peak width analysis. *Journal of the American Society for Mass Spectrometry*. **17**, 1498–1509 (2006), doi:10.1016/j.jasms.2006.05.014.
357. R. K. Chitta, D. L. Rempel, M. L. Gross, The gramicidin dimer shows both EX1 and EX2 mechanisms of H/D exchange. *Journal of the American Society for Mass Spectrometry*. **20**, 1813–1820 (2009), doi:10.1016/j.jasms.2009.05.017.
358. P. Pausch *et al.,* Structural basis for (p)ppGpp-mediated inhibition of the GTPase RbgA. *The Journal of biological chemistry*. **293**, 19699–19709 (2018), doi:10.1074/jbc.RA118.003070.
359. X. Han *et al.,* A kiwellin disarms the metabolic activity of a secreted fungal virulence factor. *Nature*. **565**, 650–653 (2019), doi:10.1038/s41586-018-0857-9.
360. T. E. Wales, K. E. Fadgen, G. C. Gerhardt, J. R. Engen, High-speed and high-resolution UPLC separation at zero degrees Celsius. *Analytical chemistry*. **80**, 6815–6820 (2008), doi:10.1021/ac8008862.
361. S. J. Geromanos *et al.,* The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics*. **9**, 1683–1695 (2009), doi:10.1002/pmic.200800562.
362. G.-Z. Li *et al.,* Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics*. **9**, 1696–1719 (2009), doi:10.1002/pmic.200800564.
363. G. S. Rule, T. K. Hitchens, *Fundamentals of protein NMR spectroscopy* (Springer, Dordrecht, 2006).
364. A. E. Derome, *Modern NMR techniques for chemistry research* (Pergamon Press, Oxford, England, 1987).
365. A. G. Palmer, NMR characterization of the dynamics of biomacromolecules. *Chemical reviews*. **104**, 3623–3640 (2004), doi:10.1021/cr030413t.
366. M. P. Williamson, Using chemical shift perturbation to characterise ligand binding. *Progress in nuclear magnetic resonance spectroscopy*. **73**, 1–16 (2013), doi:10.1016/j.pnmrs.2013.02.001.

367. M. Sattler, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in nuclear magnetic resonance spectroscopy*. **34**, 93–158 (1999), doi:10.1016/S0079-6565(98)00025-9.
368. V. Kharchenko, M. Nowakowski, M. Jaremko, A. Ejchart, Ł. Jaremko, Dynamic 15N{1H} NOE measurements: a tool for studying protein dynamics. *Journal of biomolecular NMR* (2020), doi:10.1007/s10858-020-00346-6.
369. R. Verardi, N. J. Traaseth, L. R. Masterson, V. V. Vostrikov, G. Veglia, Isotope labeling for solution and solid-state NMR spectroscopy of membrane proteins. *Advances in experimental medicine and biology*. **992**, 35–62 (2012), doi:10.1007/978-94-007-4954-2_3.
370. R. Keller, *Computer aided resonance assignment tutorial* (Cantina, 2004).
371. W. F. Vranken *et al.,* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*. **59**, 687–696 (2005), doi:10.1002/prot.20449.
372. A. L. Breeze, Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Progress in nuclear magnetic resonance spectroscopy*. **36**, 323–372 (2000), doi:10.1016/S0079-6565(00)00020-0.
373. K. Ogura, H. Terasawa, F. Inagaki, An improved double-tuned and isotope-filtered pulse scheme based on a pulsed field gradient and a wide-band inversion shaped pulse. *Journal of biomolecular NMR*. **8**, 492–498 (1996), doi:10.1007/BF00228150.
374. M. Piotto, V. Saudek, V. Sklenár, Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of biomolecular NMR*. **2**, 661–665 (1992), doi:10.1007/BF02192855.
375. V. Sklenar, M. Piotto, R. Leppik, V. Saudek, Gradient-Tailored Water Suppression for 1H-15N HSQC Experiments Optimized to Retain Full Sensitivity. *Journal of Magnetic Resonance, Series A*. **102**, 241–245 (1993), doi:10.1006/jmra.1993.1098.
376. M. Ikura, A. Bax, Isotope-filtered 2D NMR of a protein-peptide complex: study of a skeletal muscle myosin light chain kinase fragment bound to calmodulin. *Journal of the American Chemical Society*. **114**, 2433–2440 (1992), doi:10.1021/ja00033a019.
377. M. Kjaergaard, S. Brander, F. M. Poulsen, Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *Journal of biomolecular NMR*. **49**, 139–149 (2011), doi:10.1007/s10858-011-9472-x.
378. S. Schwarzinger *et al.,* Sequence-dependent correction of random coil NMR chemical shifts. *Journal of the American Chemical Society*. **123**, 2970–2978 (2001), doi:10.1021/ja003760i.
379. S. P. Mielke, V. V. Krishnan, Characterization of protein secondary structure from NMR chemical shifts. *Progress in nuclear magnetic resonance spectroscopy*. **54**, 141–165 (2009), doi:10.1016/j.pnmrs.2008.06.002.
380. Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of biomolecular NMR*. **44**, 213–223 (2009), doi:10.1007/s10858-009-9333-z.
381. T. Herrmann, P. Güntert, K. Wüthrich, Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA. *Journal of Molecular Biology*. **319**, 209–227 (2002), doi:10.1016/S0022-2836(02)00241-3.
382. J. M. Würz, S. Kazemi, E. Schmidt, A. Bagaria, P. Güntert, NMR-based automated protein structure determination. *Archives of biochemistry and biophysics*. **628**, 24–32 (2017), doi:10.1016/j.abb.2017.02.011.
383. C. Zwahlen *et al.,* Methods for Measurement of Intermolecular NOEs by Multinuclear NMR Spectroscopy: Application to a Bacteriophage λ N-Peptide/boxB RNA Complex. *J. Am. Chem. Soc.* (1997).
384. J. Iwahara, J. M. Wojciak, R. T. Clubb, Improved NMR spectra of a protein-DNA complex through rational mutagenesis and the application of a sensitivity optimized isotope-filtered NOESY experiment. *Journal of biomolecular NMR*, 231–241 (2001).
385. K. Ogura, H. Terasawa, F. Inagaki, An improved double-tuned and isotope-filtered pulse scheme based on a pulsed field gradient and a wide-band inversion shaped pulse. *Journal of biomolecular NMR*, 492–498 (1996).
386. Y. Shen, A. Bax, Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of biomolecular NMR*. **56**, 227–241 (2013), doi:10.1007/s10858-013-9741-y.
387. R. Konradi, M. Billeter, P. Güntert, Point-centered domain for decomposition for parallel moleculara dynamics simulation. *Computer Physics Communications*, 139–147 (2000).
388. J. W. Ponder, D. A. Case, "Force fields for protein simulations". *Adv Protein Chem*, 2003, 27–85 (2003).
389. A. Bhattacharya, R. Tejero, G. T. Montelione, Evaluating protein structures determined by structural genomics consortia. *Proteins*. **66** (2007), doi:10.1002/prot.21165.
390. F. A. Mulder, D. Schipper, R. Bott, R. Boelens, Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins. *Journal of Molecular Biology*. **292**, 111–123 (1999), doi:10.1006/jmbi.1999.3034.
391. M. P. Williamson, Using chemical shift perturbation to characterise ligand binding. *Progress in nuclear magnetic resonance spectroscopy*. **73**, 1–16 (2013), doi:10.1016/j.pnmrs.2013.02.001.
392. J. N. Weiss, The Hill equation revisited: uses and misuses. *FASEB j.* **11**, 835–841 (1997), doi:10.1096/FASEBJ.11.11.9285481.
393. B. N. Chaudhuri, Emerging applications of small angle solution scattering in structural biology. *Protein Science*. **24**, 267–276 (2015), doi:10.1002/pro.2624.
394. P. Bernadó, E. Mylonas, M. V. Petoukhov, M. Blackledge, D. I. Svergun, Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society*. **129**, 5656–5664 (2007), doi:10.1021/ja069124n.
395. V. Receveur-Brechot, D. Durand, How random are intrinsically disordered proteins? A small angle scattering perspective. *Current protein & peptide science*. **13**, 55–75 (2012), doi:10.2174/138920312799277901.
396. D. A. Jacques, J. Trewhella, Small-angle scattering for structural biology—expanding the frontier while avoiding the pitfalls. *Protein Science*. **19**, 642–657 (2010), doi:10.1002/pro.351.
397. C. D. Putnam, M. Hammel, G. L. Hura, J. A. Tainer, X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics*. **40**, 191–285 (2007), doi:10.1017/S0033583507004635.
398. L. A. Feigin, D. I. Svergun, *Structure Analysis by Small-Angle X-Ray and Neutron Scattering* (Springer US, Boston, MA, 1987).
399. J. Drenth, *Principles of Protein X-Ray Crystallography* (Springer New York, New York, NY, 2007).

400. A. Panjkovich, D. I. Svergun, CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics (Oxford, England)*. **34**, 1944–1946 (2018), doi:10.1093/bioinformatics/btx846.
401. A. Guinier, La diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques. *Ann. Phys.* **11**, 161–237 (1939), doi:10.1051/anphys/193911120161.
402. D. I. Svergun, Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr.* **25**, 495–503 (1992), doi:10.1107/S0021889892001663.
403. A. G. Kikhney, D. I. Svergun, A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS letters*. **589**, 2570–2577 (2015), doi:10.1016/j.febslet.2015.08.027.
404. C. E. Blanchet, D. I. Svergun, Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. *Annual review of physical chemistry*. **64**, 37–54 (2013), doi:10.1146/annurev-physchem-040412-110132.
405. D. I. Svergun, M. V. Petoukhov, M. H. Koch, Determination of Domain Structure of Proteins from X-Ray Solution Scattering. *Biophysical Journal*. **80**, 2946–2953 (2001), doi:10.1016/S0006-3495(01)76260-1.
406. T. W. Gräwert, D. I. Svergun, Structural Modeling Using Solution Small-Angle X-ray Scattering (SAXS). *Journal of Molecular Biology*. **432**, 3078–3092 (2020), doi:10.1016/j.jmb.2020.01.030.
407. P. Bernadó, M. Blackledge, A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophysical Journal*. **97**, 2839–2845 (2009), doi:10.1016/j.bpj.2009.08.044.
408. C. E. Blanchet *et al.,* Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J Appl Crystallogr.* **48**, 431–443 (2015), doi:10.1107/S160057671500254X.
409. N. R. Hajizadeh, D. Franke, D. I. Svergun, Integrated beamline control and data acquisition for small-angle X-ray scattering at the P12 BioSAXS beamline at PETRAIII storage ring DESY. *Journal of synchrotron radiation*. **25**, 906–914 (2018), doi:10.1107/S1600577518005398.
410. D. Franke, A. G. Kikhney, D. I. Svergun, Automated acquisition and analysis of small angle X-ray scattering data. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. **689**, 52–59 (2012), doi:10.1016/j.nima.2012.06.008.
411. D. Franke *et al.,* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of applied crystallography*. **50**, 1212–1225 (2017), doi:10.1107/S1600576717007786.
412. D. Franke, C. M. Jeffries, D. I. Svergun, Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nature methods*. **12**, 419–422 (2015), doi:10.1038/nmeth.3358.
413. D. I. Svergun, Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophysical Journal*. **76**, 2879–2886 (1999), doi:10.1016/S0006-3495(99)77443-6.
414. A. T. Tuukkanen, G. J. Kleywegt, D. I. Svergun, Resolution of ab initio shapes determined from small-angle scattering. *IUCrJ*. **3**, 440–447 (2016), doi:10.1107/S2052252516016018.
415. V. V. Volkov, D. I. Svergun, Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr.* **36**, 860–864 (2003), doi:10.1107/S0021889803000268.
416. M. B. Kozin, D. I. Svergun, Automated matching of high- and low-resolution structural models. *J Appl Crystallogr.* **34**, 33–41 (2001), doi:10.1107/S0021889800014126.
417. A. Panjkovich, D. I. Svergun, Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis. *Physical chemistry chemical physics : PCCP*. **18**, 5707–5719 (2016), doi:10.1039/C5CP04540A.
418. M. V. Petoukhov, D. I. Svergun, Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophysical Journal*. **89**, 1237–1250 (2005), doi:10.1529/biophysj.105.064154.
419. G. Tria, H. D. T. Mertens, M. Kachala, D. I. Svergun, Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ*. **2**, 207–217 (2015), doi:10.1107/S205225251500202X.
420. G. J. Kleywegt, Validation of protein models from Calpha coordinates alone. *Journal of Molecular Biology*. **273**, 371–376 (1997), doi:10.1006/jmbi.1997.1309.
421. W. Kabsch, XDS. *Acta crystallographica. Section D, Biological crystallography*. **66**, 125–132 (2010), doi:10.1107/S0907444909047337.
422. W. Kabsch, Integration, scaling, space-group assignment and post-refinement. *Acta crystallographica. Section D, Biological crystallography*. **66**, 133–144 (2010), doi:10.1107/S0907444909047374.
423. A. J. McCoy *et al.,* Phaser crystallographic software. *J Appl Crystallogr.* **40**, 658–674 (2007), doi:10.1107/S0021889807021206.
424. M. D. Winn *et al.,* Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography*. **67**, 235–242 (2011), doi:10.1107/S0907444910045749.
425. P. D. Adams *et al.,* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography*. **66**, 213–221 (2010), doi:10.1107/S0907444909052925.
426. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography*. **66**, 486–501 (2010), doi:10.1107/S0907444910007493.
427. A. Waterhouse *et al.,* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*. **46**, W296-W303 (2018), doi:10.1093/nar/gky427.
428. B. L. Moore, L. A. Kelley, J. Barber, J. W. Murray, J. T. MacDonald, High-quality protein backbone reconstruction from alpha carbons using Gaussian mixture models. *Journal of computational chemistry*. **34**, 1881–1889 (2013), doi:10.1002/jcc.23330.
429. G. W. Haxholm *et al.,* Intrinsically disordered cytoplasmic domains of two cytokine receptors mediate conserved interactions with membranes. *The Biochemical journal*. **468**, 495–506 (2015), doi:10.1042/BJ20141243.
430. R. Gaudet, Divide and conquer: high resolution structural information on TRP channel fragments. *The Journal of general physiology*. **133**, 231–237 (2009), doi:10.1085/jgp.200810137.
431. S. P. Graether, Troubleshooting Guide to Expressing Intrinsically Disordered Proteins for Use in NMR Experiments. *Frontiers in Molecular Biosciences*. **5**, 118 (2018), doi:10.3389/fmolb.2018.00118.

432. N. Goda *et al.,* An optimized Npro-based method for the expression and purification of intrinsically disordered proteins for an NMR study. *Intrinsically disordered proteins*. **3**, e1011004 (2015), doi:10.1080/21690707.2015.1011004.
433. T. Panavas, C. Sanders, T. R. Butt, SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems. *Methods in molecular biology (Clifton, N.J.)*. **497** (2009), doi:10.1007/978-1-59745-566-4_20.
434. A. Micsonai *et al.,* BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic acids research*. **46**, W315-W322 (2018), doi:10.1093/nar/gky497.
435. Y. Peng *et al.,* A Metastable Contact and Structural Disorder in the Estrogen Receptor Transactivation Domain. *Structure (London, England : 1993)*. **27**, 229-240.e4 (2019), doi:10.1016/j.str.2018.10.026.
436. V. N. Uversky, Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *The protein journal*. **28**, 305–325 (2009), doi:10.1007/s10930-009-9201-4.
437. M. Arbesú, G. Iruela, H. Fuentes, J. M. C. Teixeira, M. Pons, Intramolecular Fuzzy Interactions Involving Intrinsically Disordered Domains. *Frontiers in Molecular Biosciences*. **5** (2018), doi:10.3389/fmolb.2018.00039.
438. S. P. Graether, Troubleshooting Guide to Expressing Intrinsically Disordered Proteins for Use in NMR Experiments. *Frontiers in Molecular Biosciences*. **5**, 118 (2018), doi:10.3389/fmolb.2018.00118.
439. R. Konrat, NMR contributions to structural dynamics studies of intrinsically disordered proteins. *Journal of magnetic resonance (San Diego, Calif. : 1997)*. **241**, 74–85 (2014), doi:10.1016/j.jmr.2013.11.011.
440. N. Y. Haba *et al.,* NMR Determines Transient Structure and Dynamics in the Disordered C-Terminal Domain of WASp Interacting Protein. *Biophysical Journal*. **105**, 481–493 (2013), doi:10.1016/j.bpj.2013.05.046.
441. S. M. Bhattacharjee, A. Giacometti, A. Maritan, Flory theory for polymers. *Journal of physics. Condensed matter : an Institute of Physics journal*. **25**, 503101 (2013), doi:10.1088/0953-8984/25/50/503101.
442. I. S. Millett, S. Doniach, K. W. Plaxco, in *Advances in protein chemistry, Vol.62*, F. M. Richards, D. S. Eisenberg, J. Kuriyan, G. D. Rose, Eds. (Academic Press, Amsterdam, London, 2002), vol. **62,** pp. 241–262.
443. D. K. Wilkins *et al.,* Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry*. **38**, 16424–16431 (1999), doi:10.1021/bi991765q.
444. Y. Wang, J. Trewhella, D. P. Goldenberg, Small-angle X-ray scattering of reduced ribonuclease A: effects of solution conditions and comparisons with a computational model of unfolded proteins. *Journal of Molecular Biology*. **377**, 1576–1592 (2008), doi:10.1016/j.jmb.2008.02.009.
445. W. Zheng *et al.,* Probing the Action of Chemical Denaturant on an Intrinsically Disordered Protein by Simulation and Experiment. *Journal of the American Chemical Society*. **138**, 11702–11713 (2016), doi:10.1021/jacs.6b05443.
446. E. Monsellier, H. Bedouelle, Quantitative measurement of protein stability from unfolding equilibria monitored with the fluorescence maximum wavelength. *Protein engineering, design & selection : PEDS*. **18**, 445–456 (2005), doi:10.1093/protein/gzi046.
447. S. H. Giese, A. Belsom, L. Sinn, L. Fischer, J. Rappsilber, Noncovalently Associated Peptides Observed during Liquid Chromatography-Mass Spectrometry and Their Effect on Cross-Link Analyses. *Analytical chemistry*. **91**, 2678–2685 (2019), doi:10.1021/acs.analchem.8b04037.
448. B. Nilius, G. Owsianik, T. Voets, Transient receptor potential channels meet phosphoinositides. *The EMBO journal*. **27**, 2809–2816 (2008), doi:10.1038/emboj.2008.217.
449. M. A. Lemmon, Membrane recognition by phospholipid-binding domains. *Nature reviews. Molecular cell biology*. **9**, 99–111 (2008), doi:10.1038/nrm2328.
450. R. Gautier, D. Douguet, B. Antonny, G. Drin, HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics (Oxford, England)*. **24**, 2101–2102 (2008), doi:10.1093/bioinformatics/btn392.
451. J. H. Kleinschmidt, L. K. Tamm, Structural Transitions in Short-Chain Lipid Assemblies Studied by 31P-NMR Spectroscopy. *Biophysical Journal*. **83**, 994–1003 (2002), doi:10.1016/S0006-3495(02)75225-9.
452. M. D. Collins, S. E. Gordon, Short-chain phosphoinositide partitioning into plasma membrane models. *Biophysical Journal*. **105**, 2485–2494 (2013), doi:10.1016/j.bpj.2013.09.035.
453. L. L. Schrödinger, "The JyMOL Molecular Graphics Development Component, Version 1.8". (2015).
454. P. A. van Paridon, B. de Kruijff, R. Ouwerkerk, K. W. Wirtz, Polyphosphoinositides undergo charge neutralization in the physiological pH range: a 31P-NMR study. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*. **877**, 216–219 (1986), doi:10.1016/0005-2760(86)90137-2.
455. G. van Meer, D. R. Voelker, G. W. Feigenson, Membrane lipids: where they are and how they behave. *Nature reviews. Molecular cell biology*. **9**, 112–124 (2008), doi:10.1038/nrm2330.
456. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*. **98**, 10037–10041 (2001), doi:10.1073/pnas.181342398.
457. S. Lear, S. L. Cobb, Pep-Calc.com: a set of web utilities for the calculation of peptide and peptoid properties and automatic mass spectral peak assignment. *Journal of computer-aided molecular design*. **30**, 271–277 (2016), doi:10.1007/s10822-016-9902-7.
458. A. Quan, P. J. Robinson, Syndapin—a membrane remodelling and endocytic F-BAR protein. *The FEBS journal*. **280**, 5198–5212 (2013), doi:10.1111/febs.12343.
459. J. Modregger, B. Ritter, B. Witter, M. Paulsson, M. Plomann, All three PACSIN isoforms bind to endocytic proteins and inhibit endocytosis. *Journal of cell science*. **113 Pt 24**, 4511–4521 (2000).
460. B.-K. Koo, M.-H. Kim, S.-T. Lee, W. Lee, Purification and spectroscopic characterization of the human protein tyrosine kinase-6 SH3 domain. *Journal of biochemistry and molecular biology*. **35**, 343–347 (2002), doi:10.5483/bmbrep.2002.35.3.343.
461. Y. Matsumura *et al.,* Structural study of hNck2 SH3 domain protein in solution by circular dichroism and X-ray solution scattering. *Biophysical chemistry*. **0**, 39–46 (2013), doi:10.1016/j.bpc.2013.02.005.

462. J. Tanaka, H. Yanagawa, N. Doi, Comparison of the frequency of functional SH3 domains with different limited sets of amino acids using mRNA display. *PloS one*. **6**, e18034 (2011), doi:10.1371/journal.pone.0018034.
463. K. Saksela, P. Permi, SH3 domain ligand binding: What's the consensus and where's the specificity? *FEBS letters*. **586** (2012), doi:10.1016/j.febslet.2012.04.042.
464. J. Teyra *et al.,* Comprehensive Analysis of the Human SH3 Domain Family Reveals a Wide Variety of Non-canonical Specificities. *Structure*. **25**, 1598-1610.e3 (2017), doi:10.1016/j.str.2017.07.017.
465. J. L. S. Lopes, A. J. Miles, L. Whitmore, B. A. Wallace, Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: applications in secondary structure analyses. *Protein science : a publication of the Protein Society*. **23**, 1765–1772 (2014), doi:10.1002/pro.2558.
466. S. Kakinoki, M. Kitamura, Y. Noguchi, Y. Arichi, Effect of residue insertion on the stability of polyproline-I and II structures: circular dichroism spectroscopic analyses of block-type oligo-prolines (Pro) m -Gly/Ala-(Pro) n. *Pept Sci* (2020), doi:10.1002/pep2.24170.
467. A. A. Adzhubei, M. J. E. Sternberg, A. A. Makarov, Polyproline-II helix in proteins: structure and function. *Journal of Molecular Biology*. **425**, 2100–2132 (2013), doi:10.1016/j.jmb.2013.03.018.
468. M. Moradi, V. Babin, C. Roland, T. A. Darden, C. Sagui, Conformations and free energy landscapes of polyproline peptides. *Proceedings of the National Academy of Sciences of the United States of America*. **106**, 20746–20751 (2009), doi:10.1073/pnas.0906500106.
469. Y. K. Kang, Cis−Trans Isomerization and Puckering of Pseudoproline Dipeptides. *J. Phys. Chem. B*. **106**, 2074–2082 (2002), doi:10.1021/jp013608i.
470. I. Z. Steinberg, W. F. Harrington, A. Berger, M. Sela, E. Katchalski, The Configurational Changes of Poly-L-proline in Solution. *Journal of the American Chemical Society*. **82**, 5263–5279 (1960), doi:10.1021/ja01505a001.
471. S. Doose, H. Neuweiler, H. Barsch, M. Sauer, Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. *Proceedings of the National Academy of Sciences of the United States of America*. **104**, 17400–17405 (2007), doi:10.1073/pnas.0705605104.
472. Mahmoud Moradi, Volodymyr Babin, Christopher Roland, Celeste Sagui, A classical molecular dynamics investigation of the free energy and structure of short polyproline conformers. *The Journal of Chemical Physics*. **133**, 125104 (2010), doi:10.1063/1.3481087.
473. M. Schubert, D. Labudde, H. Oschkinat, P. Schmieder, A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics. *Journal of biomolecular NMR*. **24**, 149–154 (2002), doi:10.1023/a:1020997118364.
474. Y. Shen, A. Bax, Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. *Journal of biomolecular NMR*. **46**, 199–204 (2010), doi:10.1007/s10858-009-9395-y.
475. P. Craveur, A. P. Joseph, P. Poulain, A. G. de Brevern, J. Rebehmed, Cis-trans isomerization of omega dihedrals in proteins. *Amino acids*. **45**, 279–289 (2013), doi:10.1007/s00726-013-1511-3.
476. C.-S. Goh, D. Milburn, M. Gerstein, Conformational changes associated with protein-protein interactions. *Current opinion in structural biology*. **14**, 104–109 (2004), doi:10.1016/j.sbi.2004.01.005.
477. K. P. Lu, G. Finn, T. H. Lee, L. K. Nicholson, Prolyl cis-trans isomerization as a molecular timer. *Nature chemical biology*. **3**, 619–629 (2007), doi:10.1038/nchembio.2007.35.
478. P. A. Schmidpeter, F. X. Schmid, Prolyl isomerization and its catalysis in protein folding and protein function. *Journal of Molecular Biology*. **427** (2015), doi:10.1016/j.jmb.2015.01.023.
479. Shih Lin Goh, Qi Wang, Laura J. Byrnes, Holger Sondermann, Versatile Membrane Deformation Potential of Activated Pacsin. *PLOS ONE*. **7**, e51628 (2012), doi:10.1371/journal.pone.0051628.
480. A. Ortega, D. Amorós, García de la Torre, J., Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophysical Journal*. **101**, 892–898 (2011), doi:10.1016/j.bpj.2011.06.046.
481. M. I. Mahmood, H. Noguchi, K.-I. Okazaki, Curvature induction and sensing of the F-BAR protein Pacsin1 on lipid membranes via molecular dynamics simulations. *Scientific reports*. **9**, 14557 (2019), doi:10.1038/s41598-019-51202-z.
482. A. Frost *et al.,* Structural basis of membrane invagination by F-BAR domains. *Cell*. **132**, 807–817 (2008), doi:10.1016/j.cell.2007.12.041.
483. D. P. Poole *et al.,* Protease-activated receptor 2 (PAR2) protein and transient receptor potential vanilloid 4 (TRPV4) protein coupling is required for sustained inflammatory signaling. *The Journal of biological chemistry*. **288**, 5790–5802 (2013), doi:10.1074/jbc.M112.438184.
484. R. Appadurai, V. N. Uversky, A. Srivastava, The Structural and Functional Diversity of Intrinsically Disordered Regions in Transmembrane Proteins. *The Journal of membrane biology*. **252**, 273–292 (2019), doi:10.1007/s00232-019-00069-2.
485. J.-P. Changeux, A. Christopoulos, Allosteric Modulation as a Unifying Mechanism for Receptor Function and Regulation. *Cell*. **166**, 1084–1102 (2016), doi:10.1016/j.cell.2016.08.015.
486. J. A. Lamboy, H. Kim, K. S. Lee, T. Ha, E. A. Komives, Visualization of the nanospring dynamics of the IkappaBalpha ankyrin repeat domain in real time. *PNAS*. **108**, 10178–10183 (2011), doi:10.1073/pnas.1102226108.
487. A. Kohl *et al.,* Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proceedings of the National Academy of Sciences of the United States of America*. **100**, 1700–1705 (2003), doi:10.1073/pnas.0337680100.
488. C. H. Croy, S. Bergqvist, T. Huxford, G. Ghosh, E. A. Komives, Biophysical characterization of the free IkappaBalpha ankyrin repeat domain in solution. *Protein science : a publication of the Protein Society*. **13**, 1767–1777 (2004), doi:10.1110/ps.04731004.
489. E. Ladrón-de-Guevara *et al.,* The Contribution of the Ankyrin Repeat Domain of TRPV1 as a Thermal Module. *Biophysical Journal*. **118**, 836–845 (2020), doi:10.1016/j.bpj.2019.10.041.
490. M. D. Jacobs, S. C. Harrison, Structure of an IκBα/NF-κB Complex. *Cell*. **95**, 749–758 (1998), doi:10.1016/S0092-8674(00)81698-0.

491. C. J. Brown, A. K. Johnson, G. W. Daughdrill, Comparing models of evolution for ordered and disordered proteins. *Molecular biology and evolution*. **27**, 609–621 (2010), doi:10.1093/molbev/msp277.
492. C. J. Brown *et al.,* Evolutionary rate heterogeneity in proteins with long disordered regions. *Journal of molecular evolution*. **55**, 104–110 (2002), doi:10.1007/s00239-001-2309-6.
493. M. A. A. Fakhree, C. Blum, M. M. A. E. Claessens, Shaping membranes with disordered proteins. *Archives of biochemistry and biophysics*. **677**, 108163 (2019), doi:10.1016/j.abb.2019.108163.
494. T. Lu *et al.,* Role of the endothelial caveolae microdomain in shear stress-mediated coronary vasorelaxation. *The Journal of biological chemistry*. **292**, 19013–19023 (2017), doi:10.1074/jbc.M117.786152.
495. S. Goedicke-Fritz *et al.,* Evidence for functional and dynamic microcompartmentation of Cav-1/TRPV4/K(Ca) in caveolae of endothelial cells. *European journal of cell biology*. **94**, 391–400 (2015), doi:10.1016/j.ejcb.2015.06.002.
496. S. C. R. Lummis *et al.,* Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature*. **438**, 248–252 (2005), doi:10.1038/nature04130.
497. F. Zosel, D. Mercadante, D. Nettels, B. Schuler, A proline switch explains kinetic heterogeneity in a coupled folding and binding reaction. *Nature communications*. **9**, 3332 (2018), doi:10.1038/s41467-018-05725-0.
498. P. Sarkar, T. Saleh, S.-R. Tzeng, R. B. Birge, C. G. Kalodimos, Structural basis for regulation of the Crk signaling protein by a proline switch. *Nature chemical biology*. **7**, 51–57 (2011), doi:10.1038/nchembio.494.
499. J. Fanghänel, G. Fischer, Insights into the catalytic mechanism of peptidyl prolyl cis/trans isomerases. *Frontiers in bioscience : a journal and virtual library*. **9**, 3453–3478 (2004), doi:10.2741/1494.
500. Y. N. Andrade *et al.,* TRPV4 channel is involved in the coupling of fluid viscosity changes to epithelial ciliary activity. *The Journal of cell biology*. **168**, 869–874 (2005), doi:10.1083/jcb.200409070.
501. E. Seemann *et al.,* Deciphering caveolar functions by syndapin III KO-mediated impairment of caveolar invagination. *eLife*. **6** (2017), doi:10.7554/eLife.29854.
502. L. P. Nguyen, S. C. Tran, S. Suetsugu, Y.-S. Lim, S. B. Hwang, PACSIN2 Interacts with Nonstructural Protein 5A and Regulates Hepatitis C Virus Assembly. *Journal of virology*. **94** (2020), doi:10.1128/JVI.01531-19.
503. L. Picas, F. Gaits-Iacovoni, B. Goud, The emerging role of phosphoinositide clustering in intracellular trafficking and signal transduction. *F1000Research*. **5** (2016), doi:10.12688/f1000research.7537.1.
504. L. Picas *et al.,* BIN1/M-Amphiphysin2 induces clustering of phosphoinositides to recruit its downstream partner dynamin. *Nature communications*. **5**, 5647 (2014), doi:10.1038/ncomms6647.
505. T. B. Stanishneva-Konovalova, O. S. Sokolova, Effects of PI(4,5)P2 concentration on the F-BAR domain membrane binding as revealed by coarse-grained simulations. *Proteins*. **87**, 561–568 (2019), doi:10.1002/prot.25678.
506. H. Zhao *et al.,* Membrane-sculpting BAR domains generate stable lipid microdomains. *Cell Reports*. **4**, 1213–1223 (2013), doi:10.1016/j.celrep.2013.08.024.
507. W. F. Zeno, W. T. Snead, A. S. Thatte, J. C. Stachowiak, Structured and intrinsically disordered domains within Amphiphysin1 work together to sense and drive membrane curvature. *Soft matter*. **15**, 8706–8717 (2019), doi:10.1039/c9sm01495k.
508. D. J. Busch *et al.,* Intrinsically disordered proteins drive membrane curvature. *Nature communications*. **6**, 7875 (2015), doi:10.1038/ncomms8875.
509. G. Desrochers *et al.,* Molecular basis of interactions between SH3 domain-containing proteins and the proline-rich region of the ubiquitin ligase Itch. *The Journal of biological chemistry*. **292**, 6325–6338 (2017), doi:10.1074/jbc.M116.754440.
510. K. Gupta *et al.,* Identifying key membrane protein lipid interactions using mass spectrometry. *Nature protocols*. **13**, 1106–1120 (2018), doi:10.1038/nprot.2018.014.
511. T. Parasassi, E. Gratton, Membrane lipid domains and dynamics as detected by Laurdan fluorescence. *Journal of fluorescence*. **5**, 59–69 (1995), doi:10.1007/BF00718783.
512. K. Yugandhar, S. Gupta, H. Yu, Inferring Protein-Protein Interaction Networks From Mass Spectrometry-Based Proteomic Approaches: A Mini-Review. *Computational and structural biotechnology journal*. **17**, 805–811 (2019), doi:10.1016/j.csbj.2019.05.007.
513. P. J. Carman, R. Dominguez, BAR domain proteins-a linkage between cellular membranes, signaling pathways, and the actin cytoskeleton. *Biophysical reviews*. **10**, 1587–1604 (2018), doi:10.1007/s12551-018-0467-7.
514. V. Legendre-Guillemin, S. Wasiak, N. K. Hussain, A. Angers, P. S. McPherson, ENTH/ANTH proteins and clathrin-mediated membrane budding. *Journal of cell science*. **117**, 9–18 (2004), doi:10.1242/jcs.00928.
515. D. S. Molodenskiy, H. D. T. Mertens, D. I. Svergun, An automated data processing and analysis pipeline for transmembrane proteins in detergent solutions. *Scientific reports*. **10**, 8081 (2020), doi:10.1038/s41598-020-64933-1.
516. T. Bengtsen *et al.,* Structure and dynamics of a nanodisc by integrating NMR, SAXS and SANS experiments with molecular dynamics simulations. *eLife*. **9** (2020), doi:10.7554/eLife.56518.
517. I. Josts *et al.,* Structural Kinetics of MsbA Investigated by Stopped-Flow Time-Resolved Small-Angle X-Ray Scattering. *Structure (London, England : 1993)*. **28**, 348-354.e3 (2020), doi:10.1016/j.str.2019.12.001.
518. M. Redhair, A. F. Clouser, W. M. Atkins, Hydrogen-deuterium exchange mass spectrometry of membrane proteins in lipid nanodiscs. *Chemistry and physics of lipids*. **220**, 14–22 (2019), doi:10.1016/j.chemphyslip.2019.02.007.

VII Appendix I:

1.1 DNA and amino acid sequences

The genes encoding for the TRPV4 and Pacsin3 constructs were sequenced at *GENter-prise* GENOMICS (StarSEQ Gmbh, Mainz). In the following, only the DNA sequences of the full length-constructs and point mutants are shown. Mutated sites are highlighted with bold letters. The DNA sequences of deletion constructs can be obtained by removing the deleted regions. Artificial His_6- and SUMO-tags which are cleaved off during purification are shown in italic.

TRPV4 constructs

G. gallus TRPV4 NTD (this construct was used to derive IDR, ARD, and NTD ΔN120/104/97/54):

atg gct agc ggt cat cac cat cat cat cac ggg tcc gag gtc aag cca gaa gtc aag cct
M A S G H H H H H H G S E V K P E V K P
gag act cac atc aat tta aag gtg tcc gat gga tct tca gag atc ttc ttc aag atc aag
E T H I N L K V S D G S S E I F F K I K
aaa acc acc cca ctg cgt cgg ctg atg gaa gcg ttc gct aaa cgt cag ggt aaa gag atg
K T T P L R R L M E A F A K R Q G K E M
gac agc ctg cgc ttt ctg tat gat ggc att cgc att cag gcg gat cag acc ccg gaa gac
D S L R F L Y D G I R I Q A D Q T P E D
ctg gac atg gaa gat aac gat atc att gag gcc cat cgt gaa cag att ggt ggt gca gac
L D M E D N D I I E A H R E Q I G G A D
ccc gaa gac ccc cgt gat gct ggg gac gtg ctg ggg gat gac tcc ttc ccg ctc tcc tcg
P E D P R D A G D V L G D D S F P L S S
ctg gcc aac ctg ttt gag gtg gag gac acc ccg tct cct gct gag cca tcc cgg ggt ccc
L A N L F E V E D T P S P A E P S R G P
ccc ggc gct gtg gat gga aag caa aac ctc cgg atg aaa ttc cac ggg gcg ttc cgg aaa
P G A V D G K Q N L R M K F H G A F R K
ggc ccc ccg aaa ccc atg gag ctg ctg gag tcc acc atc tac gag tca tcg gtg gtc cct
G P P K P M E L L E S T I Y E S S V V P
gca ccc aaa aag gcc ccc atg gat tcc ctc ttc gac tac ggc acc tac agg cag cac ccc
A P K K A P M D S L F D Y G T Y R Q H P
agc gag aac aaa cgc tgg cgc agg agg gtc gtg gag aag ccg gtg gct ggc acc aag ggg
S E N K R W R R R V V E K P V A G T K G
ccg gca ccc aac ccg ccg ccc gtc ctc aag gtg ttc aac aga ccc atc ctc ttc gac atc
P A P N P P P V L K V F N R P I L F D I
gtc tcc cgg ggg tcc cca gat ggc ctg gag ggt ctc ctc tcc ttc ctg ctc acc cac aag
V S R G S P D G L E G L L S F L L T H K
aag cgt ctg acg gat gag gag ttc cga gag ccc tca aca ggg aag acg tgc ctg ccc aaa
K R L T D E E F R E P S T G K T C L P K
gcg ctg ctc aac ctg agt gct ggc cgg aat gac acc atc ccc atc ctc ctc gac atc gcc
A L L N L S A G R N D T I P I L L D I A
gag aag acg gga aac atg cgg gag ttc atc aac tca ccc ttc cgt gat gtc tac tac aga
E K T G N M R E F I N S P F R D V Y Y R
ggt cag aca gcg ctg cac atc gcc att gag cgc cgc tgc aag cac tac gtg gag ctg ctg
G Q T A L H I A I E R R C K H Y V E L L
gtg gag aag ggt gcc gac gtg cac gcc cag gcc cgt ggt cgc ttc ttc cag ccc aag gat
V E K G A D V H A Q A R G R F F Q P K D
gag ggc ggc tac ttc tat ttc gga gag ctg ccc ctc tca ctg gcc gcc tgc acc aac cag
E G G Y F Y F G E L P L S L A A C T N Q
ccc cac atc gtg cac tat ctg acg gag aat ggg cac aag cag gca gac ctg cgg cgc cag
P H I V H Y L T E N G H K Q A D L R R Q
gac tcg cgc ggc aac acc gtg ctg cat gct ctg gtt gcc atc gct gac aac acc cgt gag
D S R G N T V L H A L V A I A D N T R E
aac acc aag ttt gtc acc aag atg tac gac ctg ctc ctc atc aag tgc gcc aag ctc ttc
N T K F V T K M Y D L L L I K C A K L F
ccc gac acc aac ctc gag gcc ctg ctc aac aac gat ggc ctc tct cca ctc atg atg gct
P D T N L E A L L N N D G L S P L M M A
gcc aag act ggc aag atc ggg atc ttc cag cac atc atc cgc cgg gag atc gca gat taa
A K T G K I G I F Q H I I R R E I A D -

G. gallus TRPV4 NTD^{AAWAA} (this construct was used to derive ID^{RAAWAA}):

```
atg gct agc ggt cat cac cat cat cat cac ggg tcc gag gtc aag cca gaa gtc aag cct
 M   A   S   G   H   H   H   H   H   H   G   S   E   V   K   P   E   V   K   P
gag act cac atc aat tta aag gtg tcc gat gga tct tca gag atc ttc ttc aag atc aag
 E   T   H   I   N   L   K   V   S   D   G   S   S   E   I   F   F   K   I   K
aaa acc acc cca ctg cgt cgg ctg atg gaa gcg ttc gct aaa cgt cag ggt aaa gag atg
 K   T   T   P   L   R   R   L   M   E   A   F   A   K   R   Q   G   K   E   M
gac agc ctg cgc ttt ctg tat gat ggc att cgc att cag gcg gat cag acc ccg gaa gac
 D   S   L   R   F   L   Y   D   G   I   R   I   Q   A   D   Q   T   P   E   D
ctg gac atg gaa gat aac gat atc att gag gcc cat cgt gaa cag att ggt ggt gca gac
 L   D   M   E   D   N   D   I   I   E   A   H   R   E   Q   I   G   G   A   D
ccc gaa gac ccc cgt gat gct ggg gac gtg ctg ggg gat gac tcc ttc ccg ctc tcc tcg
 P   E   D   P   R   D   A   G   D   V   L   G   D   D   S   F   P   L   S   S
ctg gcc aac ctg ttt gag gtg gag gac acc ccg tct cct gct gag cca tcc cgg ggt ccc
 L   A   N   L   F   E   V   E   D   T   P   S   P   A   E   P   S   R   G   P
ccc ggc gct gtg gat gga aag caa aac ctc cgg atg aaa ttc cac ggg gcg ttc cgg aaa
 P   G   A   V   D   G   K   Q   N   L   R   M   K   F   H   G   A   F   R   K
ggc ccc ccg aaa ccc atg gag ctg ctg gag tcc acc atc tac gag tca tcg gtg gtc cct
 G   P   P   K   P   M   E   L   L   E   S   T   I   Y   E   S   S   V   V   P
gca ccc aaa aag gcc ccc atg gat tcc ctc ttc gac tac ggc acc tac agg cag cac ccc
 A   P   K   K   A   P   M   D   S   L   F   D   Y   G   T   Y   R   Q   H   P
agc gag aac gca gcc tgg gcc gcg agg gtc gtg gag aag ccg gtg gct ggc acc aag ggg
 S   E   N   A   A   W   A   A   R   V   V   E   K   P   V   A   G   T   K   G
ccg gca ccc aac ccg ccg ccc gtc ctc aag gtg ttc aac aga ccc atc ctc ttc gac atc
 P   A   P   N   P   P   P   V   L   K   V   F   N   R   P   I   L   F   D   I
gtc tcc cgg ggg tcc cca gat ggc ctg gag ggt ctc ctc tcc ttc ctg ctc acc cac aag
 V   S   R   G   S   P   D   G   L   E   G   L   L   S   F   L   L   T   H   K
aag cgt ctg acg gat gag gag ttc cga gag ccc tca aca ggg aag acg tgc ctg ccc aaa
 K   R   L   T   D   E   E   F   R   E   P   S   T   G   K   T   C   L   P   K
gcg ctg ctc aac ctg agt gct ggc cgg aat gac acc atc ccc atc ctc ctc gac atc gcc
 A   L   L   N   L   S   A   G   R   N   D   T   I   P   I   L   L   D   I   A
gag aag acg gga aac atg cgg gag ttc atc aac tca ccc ttc cgt gat gtc tac tac aga
 E   K   T   G   N   M   R   E   F   I   N   S   P   F   R   D   V   Y   Y   R
ggt cag aca gcg ctg cac atc gcc att gag cgc cgc tgc aag cac tac gtg gag ctg ctg
 G   Q   T   A   L   H   I   A   I   E   R   R   C   K   H   Y   V   E   L   L
gtg gag aag ggt gcc gac gtg cac gcc cag gcc cgt ggt cgc ttc ttc cag ccc aag gat
 V   E   K   G   A   D   V   H   A   Q   A   R   G   R   F   F   Q   P   K   D
gag ggc ggc tac ttc tat ttc gga gag ctg ccc ctc tca ctg gcc gcc tgc acc aac cag
 E   G   G   Y   F   Y   F   G   E   L   P   L   S   L   A   A   C   T   N   Q
ccc cac atc gtg cac tat ctg acg gag aat ggg cac aag cag gca gac ctg cgg cgc cag
 P   H   I   V   H   Y   L   T   E   N   G   H   K   Q   A   D   L   R   R   Q
gac tcg cgc ggc aac acc gtg ctg cat gct ctg gtt gcc atc gct gac aac acc cgt gag
 D   S   R   G   N   T   V   L   H   A   L   V   A   I   A   D   N   T   R   E
aac acc aag ttt gtc acc aag atg tac gac ctg ctc ctc atc aag tgc gcc aag ctc ttc
 N   T   K   F   V   T   K   M   Y   D   L   L   L   I   K   C   A   K   L   F
ccc gac acc aac ctc gag gcc ctg ctc aac aac gat ggc ctc tct cca ctc atg atg gct
 P   D   T   N   L   E   A   L   L   N   N   D   G   L   S   P   L   M   M   A
gcc aag act ggc aag atc ggg atc ttc cag cac atc atc cgc cgg gag atc gca gat taa
 A   K   T   G   K   I   G   I   F   Q   H   I   I   R   R   E   I   A   D   -
```

G. gallus TRPV4 PRR (V131I):

```
atg gct agc ggt cat cac cat cat cat cac ggg tcc gag gtc aag cca gaa gtc aag cct
 M   A   S   G   H   H   H   H   H   H   G   S   E   V   K   P   E   V   K   P
gag act cac atc aat tta aag gtg tcc gat gga tct tca gag atc ttc ttc aag atc aag
 E   T   H   I   N   L   K   V   S   D   G   S   S   E   I   F   F   K   I   K
aaa acc acc cca ctg cgt cgg ctg atg gaa gcg ttc gct aaa cgt cag ggt aaa gag atg
 K   T   T   P   L   R   R   L   M   E   A   F   A   K   R   Q   G   K   E   M
gac agc ctg cgc ttt ctg tat gat ggc att cgc att cag gcg gat cag acc ccg gaa gac
 D   S   L   R   F   L   Y   D   G   I   R   I   Q   A   D   Q   T   P   E   D
ctg gac atg gaa gat aac gat atc att gag gcc cat cgt gaa cag att ggt ggt act aaa
 L   D   M   E   D   N   D   I   I   E   A   H   R   E   Q   I   G   G   T   K
ggt ccg gcg ccg aac ccg ccg cca atc ctg aaa gtg tgg taa
 G   P   A   P   N   P   P   P   I   L   K   V   W   -
```

G. gallus TRPV4 PBD-PRR:

```
atg gct agc ggt cat cac cat cat cat cac ggg tcc gag gtc aag cca gaa gtc aag cct
 M   A   S   G   H   H   H   H   H   H   G   S   E   V   K   P   E   V   K   P
gag act cac atc aat tta aag gtg tcc gat gga tct tca gag atc ttc ttc aag atc aag
 E   T   H   I   N   L   K   V   S   D   G   S   S   E   I   F   F   K   I   K
aaa acc acc cca ctg cgt cgg ctg atg gaa gcg ttc gct aaa cgt cag ggt aaa gag atg
 K   T   T   P   L   R   R   L   M   E   A   F   A   K   R   Q   G   K   E   M
gac agc ctg cgc ttt ctg tat gat ggc att cgc att cag gcg gat cag acc ccg gaa gac
 D   S   L   R   F   L   Y   D   G   I   R   I   Q   A   D   Q   T   P   E   D
ctg gac atg gaa gat aac gat atc att gag gcc cat cgt gaa cag att ggt ggt gaa aac
 L   D   M   E   D   N   D   I   I   E   A   H   R   E   Q   I   G   G   E   N
aaa cgc tgg cgt cgc cgt gtg gtt gaa aaa cca gtg gcg ggc acc aaa ggt ccg gct cct
 K   R   W   R   R   R   V   V   E   K   P   V   A   G   T   K   G   P   A   P
aac ccg cca ccg gta ctg aaa gtt taa
 N   P   P   P   V   L   K   V   -
```

G. gallus TRPV4 PBD-PRR_P128A:

```
atg gct agc ggt cat cac cat cat cat cac ggg tcc gag gtc aag cca gaa gtc aag cct
 M   A   S   G   H   H   H   H   H   H   G   S   E   V   K   P   E   V   K   P
gag act cac atc aat tta aag gtg tcc gat gga tct tca gag atc ttc ttc aag atc aag
 E   T   H   I   N   L   K   V   S   D   G   S   S   E   I   F   F   K   I   K
aaa acc acc cca ctg cgt cgg ctg atg gaa gcg ttc gct aaa cgt cag ggt aaa gag atg
 K   T   T   P   L   R   R   L   M   E   A   F   A   K   R   Q   G   K   E   M
gac agc ctg cgc ttt ctg tat gat ggc att cgc att cag gcg gat cag acc ccg gaa gac
 D   S   L   R   F   L   Y   D   G   I   R   I   Q   A   D   Q   T   P   E   D
ctg gac atg gaa gat aac gat atc att gag gcc cat cgt gaa cag att ggt ggt gaa aac
 L   D   M   E   D   N   D   I   I   E   A   H   R   E   Q   I   G   G   E   N
aaa cgc tgg cgt cgc cgt gtg gtt gaa aaa cca gtg gcg ggc acc aaa ggt ccg gct cct
 K   R   W   R   R   R   V   V   E   K   P   V   A   G   T   K   G   P   A   P
aac gcg cca ccg gta ctg aaa gtt taa
 N   A   P   P   V   L   K   V   -
```

G. gallus TRPV4 PBD-PRR_P128A/P129L:

```
atg gct agc ggt cat cac cat cat cat cac ggg tcc gag gtc aag cca gaa gtc aag cct
 M   A   S   G   H   H   H   H   H   H   G   S   E   V   K   P   E   V   K   P
gag act cac atc aat tta aag gtg tcc gat gga tct tca gag atc ttc ttc aag atc aag
 E   T   H   I   N   L   K   V   S   D   G   S   S   E   I   F   F   K   I   K
aaa acc acc cca ctg cgt cgg ctg atg gaa gcg ttc gct aaa cgt cag ggt aaa gag atg
 K   T   T   P   L   R   R   L   M   E   A   F   A   K   R   Q   G   K   E   M
gac agc ctg cgc ttt ctg tat gat ggc att cgc att cag gcg gat cag acc ccg gaa gac
 D   S   L   R   F   L   Y   D   G   I   R   I   Q   A   D   Q   T   P   E   D
ctg gac atg gaa gat aac gat atc att gag gcc cat cgt gaa cag att ggt ggt gaa aac
 L   D   M   E   D   N   D   I   I   E   A   H   R   E   Q   I   G   G   E   N
aaa cgc tgg cgt cgc cgt gtg gtt gaa aaa cca gtg gcg ggc acc aaa ggt ccg gct cct
 K   R   W   R   R   R   V   V   E   K   P   V   A   G   T   K   G   P   A   P
aac gcg ctg ccg gta ctg aaa gtt taa
 N   A   P   P   V   L   K   V   -
```

G. gallus TRPV4 PBD^{AAWAA}-PRR:

```
atg gct agc ggt cat cac cat cat cat cac ggg tcc gag gtc aag cca gaa gtc aag cct
 M   A   S   G   H   H   H   H   H   H   G   S   E   V   K   P   E   V   K   P
gag act cac atc aat tta aag gtg tcc gat gga tct tca gag atc ttc ttc aag atc aag
 E   T   H   I   N   L   K   V   S   D   G   S   S   E   I   F   F   K   I   K
aaa acc acc cca ctg cgt cgg ctg atg gaa gcg ttc gct aaa cgt cag ggt aaa gag atg
 K   T   T   P   L   R   R   L   M   E   A   F   A   K   R   Q   G   K   E   M
gac agc ctg cgc ttt ctg tat gat ggc att cgc att cag gcg gat cag acc ccg gaa gac
 D   S   L   R   F   L   Y   D   G   I   R   I   Q   A   D   Q   T   P   E   D
ctg gac atg gaa gat aac gat atc att gag gcc cat cgt gaa cag att ggt ggt gaa aac
 L   D   M   E   D   N   D   I   I   E   A   H   R   E   Q   I   G   G   E   N
```

gca gcc tgg gcc gcg cgt gtg gtt gaa aaa cca gtg gcg ggc acc aaa ggt ccg gct cct
A A W A A R V V E K P V A G T K G P A P
aac ccg cca ccg gta ctg aaa gtt taa
N P P P V L K V -

G. gallus TRPV4 PH-PRR:

atg gct agc ggt cat cac cat cat cat cac ggg tcc gag gtc aag cca gaa gtc aag cct
M A S G H H H H H H G S E V K P E V K P
gag act cac atc aat tta aag gtg tcc gat gga tct tca gag atc ttc ttc aag atc aag
E T H I N L K V S D G S S E I F F K I K
aaa acc acc cca ctg cgt cgg ctg atg gaa gcg ttc gct aaa cgt cag ggt aaa gag atg
K T T P L R R L M E A F A K R Q G K E M
gac agc ctg cgc ttt ctg tat gat ggc att cgc att cag gcg gat cag acc ccg gaa gac
D S L R F L Y D G I R I Q A D Q T P E D
ctg gac atg gaa gat aac gat atc att gag gcc cat cgt gaa cag att ggt ggt acc tat
L D M E D N D I I E A H R E Q I G G T Y
cgt cag cat ccg agc gag aac aaa cgc tgg cgt cgc cgt gtg gtt gaa aaa cca gtg gcg
R Q H P S E N K R W R R R V V E K P V A
ggc act aaa ggt ccg gcg ccg aac ccg ccg cca gtg ctg aaa gtg
G T K G P A P N P P P V L K V

G. gallus TRPV4 PH^{AAWAA}-PRR:

atg gct agc ggt cat cac cat cat cat cac ggg tcc gag gtc aag cca gaa gtc aag cct
M A S G H H H H H H G S E V K P E V K P
gag act cac atc aat tta aag gtg tcc gat gga tct tca gag atc ttc ttc aag atc aag
E T H I N L K V S D G S S E I F F K I K
aaa acc acc cca ctg cgt cgg ctg atg gaa gcg ttc gct aaa cgt cag ggt aaa gag atg
K T T P L R R L M E A F A K R Q G K E M
gac agc ctg cgc ttt ctg tat gat ggc att cgc att cag gcg gat cag acc ccg gaa gac
D S L R F L Y D G I R I Q A D Q T P E D
ctg gac atg gaa gat aac gat atc att gag gcc cat cgt gaa cag att ggt ggt acc tat
L D M E D N D I I E A H R E Q I G G T Y
cgt cag cat ccg agc gag aac **gca gcc tgg gcc gcg** cgt gtg gtt gaa aaa cca gtg gcg
R Q H P S E N **A A W A A** R V V E K P V A
ggc act aaa ggt ccg gcg ccg aac ccg ccg cca gtg ctg aaa gtg
G T K G P A P N P P P V L K V

H. sapiens TRPV4 NTD:

atg gct agc ggt cat cac cat cat cat cac ggg tcc gag gtc aag cca gaa gtc aag cct
M A S G H H H H H H G S E V K P E V K P
gag act cac atc aat tta aag gtg tcc gat gga tct tca gag atc ttc ttc aag atc aag
E T H I N L K V S D G S S E I F F K I K
aaa acc acc cca ctg cgt cgg ctg atg gaa gcg ttc gct aaa cgt cag ggt aaa gag atg
K T T P L R R L M E A F A K R Q G K E M
gac agc ctg cgc ttt ctg tat gat ggc att cgc att cag gcg gat cag acc ccg gaa gac
D S L R F L Y D G I R I Q A D Q T P E D
ctg gac atg gaa gat aac gat atc att gag gcc cat cgt gaa cag att ggt ggt gcg gat
L D M E D N D I I E A H R E Q I G G A D
tcc agc gaa ggc ccc cgc gcg ggg ccc ggg gag gtg gct gag ctc ccc ggg gat gag agt
S S E G P R A G P G E V A E L P G D E S
ggc acc cca ggt ggg gag gct ttt cct ctc tcc tcc ctg gcc aat ctg ttt gag ggg gag
G T P G G E A F P L S S L A N L F E G E
gat ggc tcc ctt tcg ccc tca ccg gct gat gcc agt cgc cct gct ggc cca ggc gat ggg
D G S L S P S P A D A S R P A G P G D G
cga cca aat ctg cgc atg aag ttc cag ggc gcc ttc cgc aag ggg gtg ccc aac ccc atc
R P N L R M K F Q G A F R K G V P N P I
gat ctg ctg gag tcc acc cta tat gag tcc tcg gtg gtg cct ggg ccc aag aaa gca ccc
D L L E S T L Y E S S V V P G P K K A P

```
atg gac tca ctg ttt gac tac ggc acc tat cgt cac cac tcc agt gac aac aag agg tgg
 M   D   S   L   F   D   Y   G   T   Y   R   H   H   S   S   D   N   K   R   W
agg aag aag atc ata gag aag cag ccg cag agc ccc aaa gcc cct gcc cct cag ccg ccc
 R   K   K   I   I   E   K   Q   P   Q   S   P   K   A   P   A   P   Q   P   P
ccc atc ctc aaa gtc ttc aac cgg cct atc ctc ttt gac atc gtg tcc cgg ggc tcc act
 P   I   L   K   V   F   N   R   P   I   L   F   D   I   V   S   R   G   S   T
gct gac ctg gac ggg ctg ctc cca ttc ttg ctg acc cac aag aaa cgc cta act gat gag
 A   D   L   D   G   L   L   P   F   L   L   T   H   K   K   R   L   T   D   E
gag ttt cga gag cca tct acg ggg aag acc tgc ctg ccc aag gcc ttg ctg aac ctg agc
 E   F   R   E   P   S   T   G   K   T   C   L   P   K   A   L   L   N   L   S
aat ggc cgc aac gac acc atc cct gtg ctg ctg gac atc gcg gag cgc acc ggc aac atg
 N   G   R   N   D   T   I   P   V   L   L   D   I   A   E   R   T   G   N   M
agg gag ttc att aac tcg ccc ttc cgt gac atc tac tat cga ggt cag aca gcc ctg cac
 R   E   F   I   N   S   P   F   R   D   I   Y   Y   R   G   Q   T   A   L   H
atc gcc att gag cgt cgc tgc aaa cac tac gtg gaa ctt ctc gtg gcc cag gga gct gat
 I   A   I   E   R   R   C   K   H   Y   V   E   L   L   V   A   Q   G   A   D
gtc cac gcc cag gcc cgt ggg cgc ttc ttc cag ccc aag gat gag ggg ggc tac ttc tac
 V   H   A   Q   A   R   G   R   F   F   Q   P   K   D   E   G   G   Y   F   Y
ttt ggg gag ctg ccc ctg tcg ctg gct gcc tgc acc aac cag ccc cac att gtc aac tac
 F   G   E   L   P   L   S   L   A   A   C   T   N   Q   P   H   I   V   N   Y
ctg acg gag aac ccc cac aag aag gcg gac atg cgg cgc cag gac tcg cga ggc aac aca
 L   T   E   N   P   H   K   K   A   D   M   R   R   Q   D   S   R   G   N   T
gtg ctg cat gcg ctg gtg gcc att gct gac aac acc cgt gag aac acc aag ttt gtt acc
 V   L   H   A   L   V   A   I   A   D   N   T   R   E   N   T   K   F   V   T
aag atg tac gac ctg ctg ctg ctc aag tgt gcc cgc ctc ttc ccc gac agc aac ctg gag
 K   M   Y   D   L   L   L   L   K   C   A   R   L   F   P   D   S   N   L   E
gcc gtg ctc aac aac gac ggc ctc tcg ccc ctc atg atg gct gcc aag acg ggc aag att
 A   V   L   N   N   D   G   L   S   P   L   M   M   A   A   K   T   G   K   I
ggg atc ttt cag cac atc atc cgg cgg gag gtg acg gat gag taa gga tcc ggc tgc taa
 G   I   F   Q   H   I   I   R   R   E   V   T   D   E   -
```

Pacsin3 constructs

G. gallus Pacsin3 (this construct was used to derive F-BAR and l-SH3):

```
atg gct agc ggt cat cac cat cat cat cac ggg tcc gag gtc aag cca gaa gtc aag cct
 M   A   S   G   H   H   H   H   H   H   G   S   E   V   K   P   E   V   K   P
gag act cac atc aat tta aag gtg tcc gat gga tct tca gag atc ttc ttc aag atc aag
 E   T   H   I   N   L   K   V   S   D   G   S   S   E   I   F   F   K   I   K
aaa acc acc cca ctg cgt cgg ctg atg gaa gcg ttc gct aaa cgt cag ggt aaa gag atg
 K   T   T   P   L   R   R   L   M   E   A   F   A   K   R   Q   G   K   E   M
gac agc ctg cgc ttt ctg tat gat ggc att cgc att cag gcg gat cag acc ccg gaa gac
 D   S   L   R   F   L   Y   D   G   I   R   I   Q   A   D   Q   T   P   E   D
ctg gac atg gaa gat aac gat atc att gag gcc cat cgt gaa cag att ggt ggt agc agc
 L   D   M   E   D   N   D   I   I   E   A   H   R   E   Q   I   G   G   S   S
gaa gat gat gcg ggt ggt gaa gcg ccg gaa ggc agc ttt tgg gaa gcg ggc aac tac cgt
 E   D   D   A   G   G   E   A   P   E   G   S   F   W   E   A   G   N   Y   R
cgt acc gtg aag cgt gtg gac gat ggc tac cgt ctg tgc aac gac ctg atc agc tgc ttc
 R   T   V   K   R   V   D   D   G   Y   R   L   C   N   D   L   I   S   C   F
cag gaa cgt gcg aag att gag aaa aac tat gcg cag caa ctg acc gag tgg agc cgt aag
 Q   E   R   A   K   I   E   K   N   Y   A   Q   Q   L   T   E   W   S   R   K
tgg cgt acc aac gtg gaa aaa ggc ccg caa tac ggt acc ctg gag aaa gcg tgg cat gcg
 W   R   T   N   V   E   K   G   P   Q   Y   G   T   L   E   K   A   W   H   A
ttt ctg acc gcg gcg gat aaa ctg agc gaa atc cac ctg gtg gtt cgt aac cac ctg gcg
 F   L   T   A   A   D   K   L   S   E   I   H   L   V   V   R   N   H   L   A
ggc gaa gac agc gat aag att aaa gcg tgg cag aaa gag gcg tac cac aaa caa atc att
 G   E   D   S   D   K   I   K   A   W   Q   K   E   A   Y   H   K   Q   I   I
ggt ggc ttc aag gaa acc aaa gag gcg gaa gac ggt ttt cgt aag gcg cag aaa ccg tgg
 G   G   F   K   E   T   K   E   A   E   D   G   F   R   K   A   Q   K   P   W
gtg aag aaa atg aag gat gtt gaa acc agc aag aaa aac tat cac gcg gcg cgt aaa gag
 V   K   K   M   K   D   V   E   T   S   K   K   N   Y   H   A   A   R   K   E
gag aaa acc gcg cac acc cgt gaa aac cac gcg aaa gcg gac agc agc gtg agc cag gag
 E   K   T   A   H   T   R   E   N   H   A   K   A   D   S   S   V   S   Q   E
caa ctg cgt aag ctg cag gag cgt gtt gaa aaa tgc acc caa gag gcg gaa aag tgc aaa
 Q   L   R   K   L   Q   E   R   V   E   K   C   T   Q   E   A   E   K   C   K
gac cag tat gag aag atg ctg gag gaa ctg aac cgt tac aac ccg cgt tat atg gag gat
 D   Q   Y   E   K   M   L   E   E   L   N   R   Y   N   P   R   Y   M   E   D
```

```
atg gaa cag gtg ttc gaa ggt tgc caa gag gcg gag cgt aag cgt ctg tgc ttc ttt aaa
 M   E   Q   V   F   E   G   C   Q   E   A   E   R   K   R   L   C   F   F   K
gag atg ttc ctg aac ctg cac cag cac ctg aac ctg agc acc agc gaa agc ttt caa gcg
 E   M   F   L   N   L   H   Q   H   L   N   L   S   T   S   E   S   F   Q   A
ctg tac cgt gat ctg tat cag gtt atc atg gcg gcg gac aac caa gaa gat ctg aaa tgg
 L   Y   R   D   L   Y   Q   V   I   M   A   A   D   N   Q   E   D   L   K   W
tgg cgt aac acc cac ggc ccg ggt atg gcg atg aac tgg ccg cag ttt gag gaa tgg agc
 W   R   N   T   H   G   P   G   M   A   M   N   W   P   Q   F   E   E   W   S
ctg gaa acc caa cgt ccg atc acc aag aaa gag aag agc ggc aaa gtg gcg gac gat gtt
 L   E   T   Q   R   P   I   T   K   K   E   K   S   G   K   V   A   D   D   V
acc ctg acc agc att ctg ccg acc cgt gac ggt gtg gtt agc caa acc ccg ctg cag acc
 T   L   T   S   I   L   P   T   R   D   G   V   V   S   Q   T   P   L   Q   T
cgt caa ggt ggc ggt aaa gag gac atc agc gaa tgg agc gac gag gat acc ccg aag aaa
 R   Q   G   G   G   K   E   D   I   S   E   W   S   D   E   D   T   P   K   K
tac ctg gac gcg aac ggc cac gag gaa gat att aaa gtg ccg ggt gtg cgt gtt cgt gcg
 Y   L   D   A   N   G   H   E   E   D   I   K   V   P   G   V   R   V   R   A
ctg tac gac tat acc ggc cag gaa gcg gat gaa ctg agc ttc aag gcg ggc gag gaa ctg
 L   Y   D   Y   T   G   Q   E   A   D   E   L   S   F   K   A   G   E   E   L
atg aaa att agc gag gaa gac gag caa ggc tgg tgc aag ggt cgt ctg ctg acc ggc cac
 M   K   I   S   E   E   D   E   Q   G   W   C   K   G   R   L   L   T   G   H
gtt ggt ctg tat ccg gcg aat tac gtt gag aaa gtt ggc ctg taa
 V   G   L   Y   P   A   N   Y   V   E   K   V   G   L   -
```

G. gallus Pacsin1 SH3:

```
atg ggg aag gga gtc cgt gtc cgt gcc ctt tat gat tac gat ggg cag gaa caa gat gag
 M   G   K   G   V   R   V   R   A   L   Y   D   Y   D   G   Q   E   Q   D   E
ctg tcg ttc aaa gct ggt gac gag ctg aca aaa tta ggc gaa gag gac gag cag ggc tgg
 L   S   F   K   A   G   D   E   L   T   K   L   G   E   E   D   E   Q   G   W
tgt aaa ggg aga ttg gac aat ggt cag ctt ggt tta tat cct gca aat tat gta gag gcg
 C   K   G   R   L   D   N   G   Q   L   G   L   Y   P   A   N   Y   V   E   A
att gct gct gct cac cac cac cat cat cac taa
 I   A   A   A   H   H   H   H   H   H   -
```

G. gallus Pacsin2 SH3:

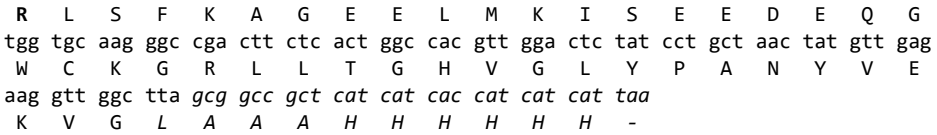
```
atg ccg gtt atg gaa gtg cgt gta cgt gct ctt tat gat tat gag ggxt cag gag cag gat
 M   P   V   M   E   V   R   V   R   A   L   Y   D   Y   E   G   Q   E   Q   D
gaa ttg tcg ttt aag gcg gga gat gaa ctg aca aag atg gaa aat gag gac gag cag ggg
 E   L   S   F   K   A   G   D   E   L   T   K   M   E   N   E   D   E   Q   G
tgg tgt aag gga aga ttg gac aat ggg cag gta ggt ttg tac ccg gct aat tac gtc gaa
 W   C   K   G   R   L   D   N   G   Q   V   G   L   Y   P   A   N   Y   V   E
ccc ata caa gca gcc gca cac cat cac cac cat cat taa
 P   I   Q   A   A   A   H   H   H   H   H   H   -
```

G. gallus Pacsin3 SH3:

```
atg aag gta cca ggt gta cgg gta cga gca ctg tat gat tac aca gga cag gag gca gat
 M   K   V   P   G   V   R   V   R   A   L   Y   D   Y   T   G   Q   E   A   D
gag ctg agc ttt aaa gca ggt gaa gaa cta atg aag att agt gag gaa gat gag cag gga
 E   L   S   F   K   A   G   E   E   L   M   K   I   S   E   E   D   E   Q   G
tgg tgc aag ggc cga ctt ctc act ggc cac gtt gga ctc tat cct gct aac tat gtt gag
 W   C   K   G   R   L   L   T   G   H   V   G   L   Y   P   A   N   Y   V   E
aag gtt ggc tta gcg gcc gct cat cat cac cat cat cat taa
 K   V   G   L   A   A   A   H   H   H   H   H   H   -
```

G. gallus Pacsin3 SH3 QEE/RRR:

```
atg aag gta cca ggt gta cgg gta cga gca ctg tat gat tac aca gga cgt cgt gca gat
 M   K   V   P   G   V   R   V   R   A   L   Y   D   Y   T   G   R   R   A   D
cgt ctg agc ttt aaa gca ggt gaa gaa cta atg aag att agt gag gaa gat gag cag gga
```



1.2 List of figures

1.3 List of tables

1.4 Supplementary figures

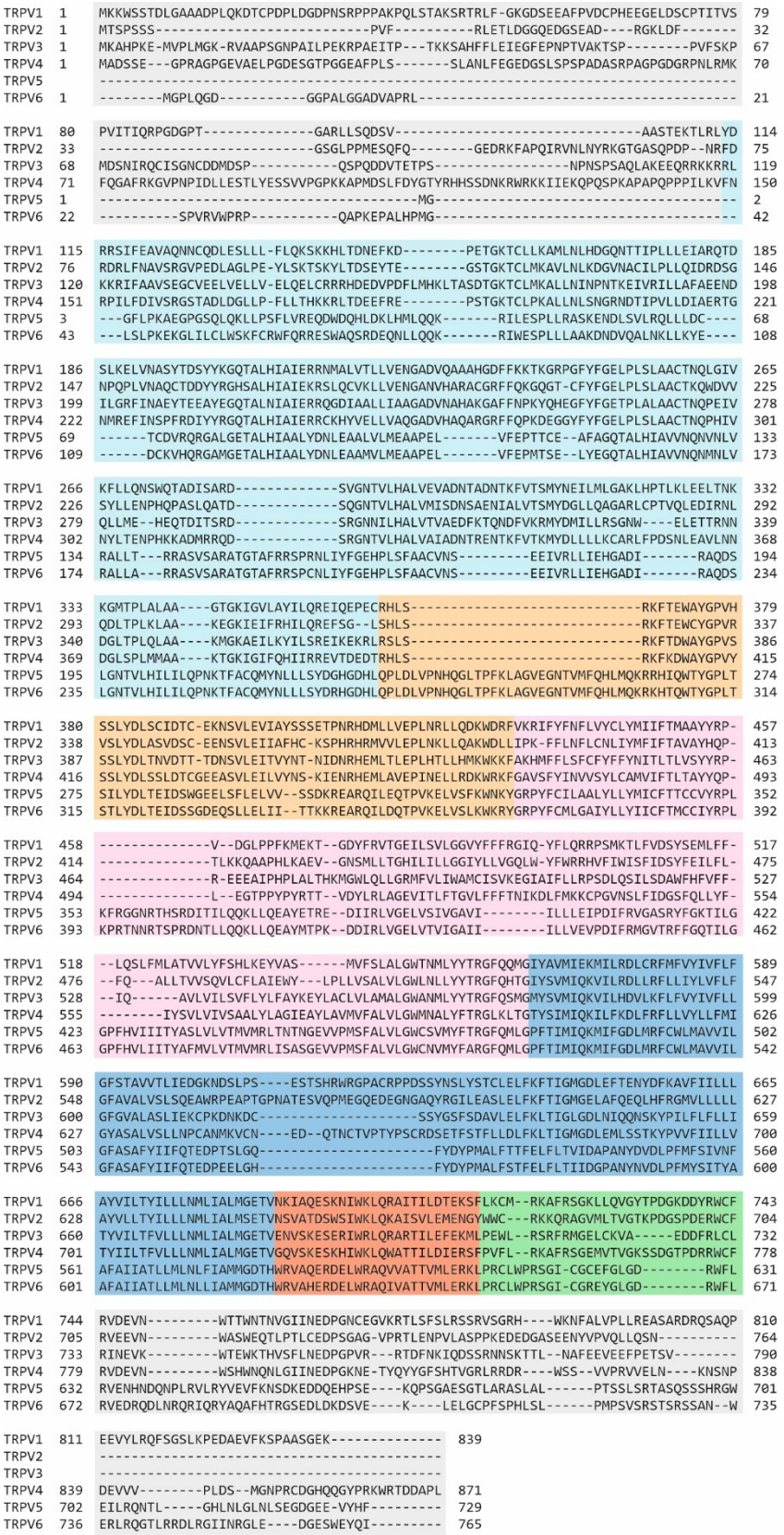


Figure 123: Sequence alignment of human TRPV channels with highlighted domains. ARD: Ankyrin repeat domain (cyan), Pre-S1 linker region (orange), Voltage sensor like domain (S1-S4, pink), pore helices (S5-S6, blue), TRP helix (red), C-terminus (green). The color code is indicated in the figure. Regions which were not resolved in the respective structures are indicated with dashed lines.

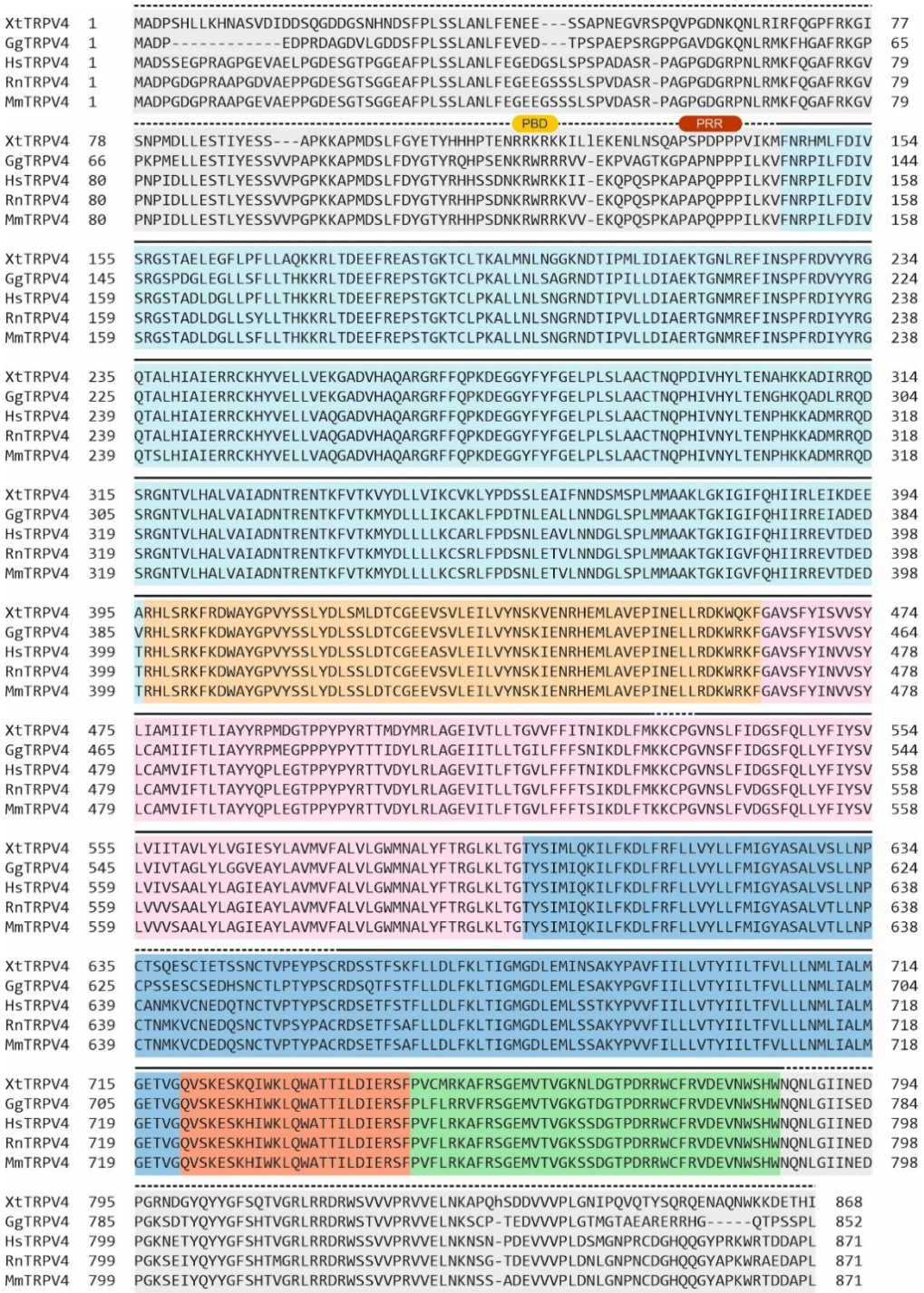


Figure 124: Sequence alignment of TRPV4 from different species with highlighted domains. Sequence alignment of *X. tropicalis* (Xn), *H. sapiens* (Hs), *M. musculus* (Mm), *R. norvegicus* (Rn), and *G. gallus* (Gg) TRPV4. Domains were highlighted based on the XtTRPV4 cryo-EM structure (PDB: 6bbj). ARD: Ankyrin repeat domain (cyan), Pre-S1 linker region (orange), Voltage sensor like domain (S1-S4, pink), pore helices (S5-S6, blue), TRP helix (red), C-terminus (green). The color code is indicated in the figure. Regions which were not resolved in the XtTRPV4 structure are indicated with dashed lines. The proline-rich region (PRR, red) and the phosphoinositide binding domain (PBD, yellow) are shown as boxes.

```
PACSIN1     MSGSYDESAAAAEETTDSFWEVGNYKRTVKRIDDGHRLCNDLMNCVHERAKIEKSYAQQL     60
PACSIN2     MSGSYDDSVG-VEVSSDSFWEVGNYKRTVKRIDDGHRLCNDLMNCIHERARIEKVYAQQL     59
PACSIN3     MSS---EDDAGGEAPEGSFWEAGNYRRTVKRVDDGYRLCNDLISCFQERAKIEKNYAQQL     57

PACSIN1     TDWSKRWRQLIEKGPQYGSLERAWGAIMTEADKVSELHQEVKNSLLNDDFEKVKNWQKDA     120
PACSIN2     TEWAKRWKQLVEKGPQYGTVERAWCAFMSEAEKVSELHLEVKGSLMNEDFEKIKNWQKEA     119
PACSIN3     TEWSRKWRTNVEKGPQYGTLEKAWHAFLTAADKLSEIHLVVRNHLAGEDSDKIKAWQKEA     117

PACSIN1     YHKQIMGGFKEAKEAEDGFRKAQKPWAKKLKELETAKKAYHLACKEEKLAMTREANSKAD     180
PACSIN2     FHKQMMGGFKETKEAEDGFRKAQKPWAKKLKEVEAAKKAYHAACKEEKLAISRETNSKAD     179
PACSIN3     YHKQIIGGFKETKEAEDGFRKAQKPWVKKMKDVETSKKNYHAARKEEKTAHTRENHAKAD     177

PACSIN1     QSNTPEQQKKLQDKVEKCKQDVQKTQEKYEKVLDELNKCTPQYMESMEQVFEQCQQFEEK     240
PACSIN2     PALNPEQLKKLQDKVERSKQDVLKTKAKYEKSLKELDNATPQYMENMEQVFEQCQQFEEK     239
PACSIN3     SSVSQEQLRKLQERVEKCTQEAEKCKDQYEKMLEELNRYNPRYMEDMEQVFEGCQEAERK     237

PACSIN1     RLNFLKEVLLDIKRHLNLAESSSYANVYRELEQTIRLSDAQEDLRWFRSTSGPGMPMNWP     300
PACSIN2     RLRFFREVLLEVQKHLDLSNVASYKNIYRELEQNIKTADAVEDLRWFRANQGPGMSMNWP     299
PACSIN3     RLCFFKEMFLNLHQHLNLSTSESFQALYRDLYQVIMAADNQEDLKWWRNTHGPGMAMNWP     297

PACSIN1     QFE--EWNPDLTHTITRKEKMK-KGEGVTLTNASSTGESGASAGERGSVSSHDRGQTYSA     357
PACSIN2     QFEDDEWSADLNRTLSRREKKK-ASDGVTLTGINQTGDQVSQPNKHSSVSSYEKNQSYPT     358
PACSIN3     QFE--EWSLETQRPITKKEKSGKVADDVTLTSILPTRDGVVSQTPL--QTRQGGGKEDIS     353

PACSIN1     EWSDDEGSNSFNASEANGGTNPFDEDSA-GKGVRVRALYDYDGQEQDELSFKAGDELTKL     416
PACSIN2     DWSDEESNNPFSSTDAKGDTNPFDEDTSPVMEVRVRALYDYEGQEQDELSFKAGDELTKM     418
PACSIN3     EWSDEDTPKKYLD--ANGH-----EEDIKVPGVRVRALYDYTGQEADELSFKAGEELMKI     406

PACSIN1     GEEDEQGWCKGRLDNGQLGLYPANYVEAI--     445
PACSIN2     ENEDEQGWCKGRLDNGQVGLYPANYVEPIQ-     448
PACSIN3     SEEDEQGWCKGRLLTGHVGLYPANYVEKVGL     437
```

Figure 125: Sequence alignment of *G. gallus* Pacsin isoforms 1, 2, and 3. The N-terminal F-BAR domain is colored in light blue, the C-terminal SH3 domain is colored in light purple. The flexible linker between F-BAR and SH3 domain is shown in grey.

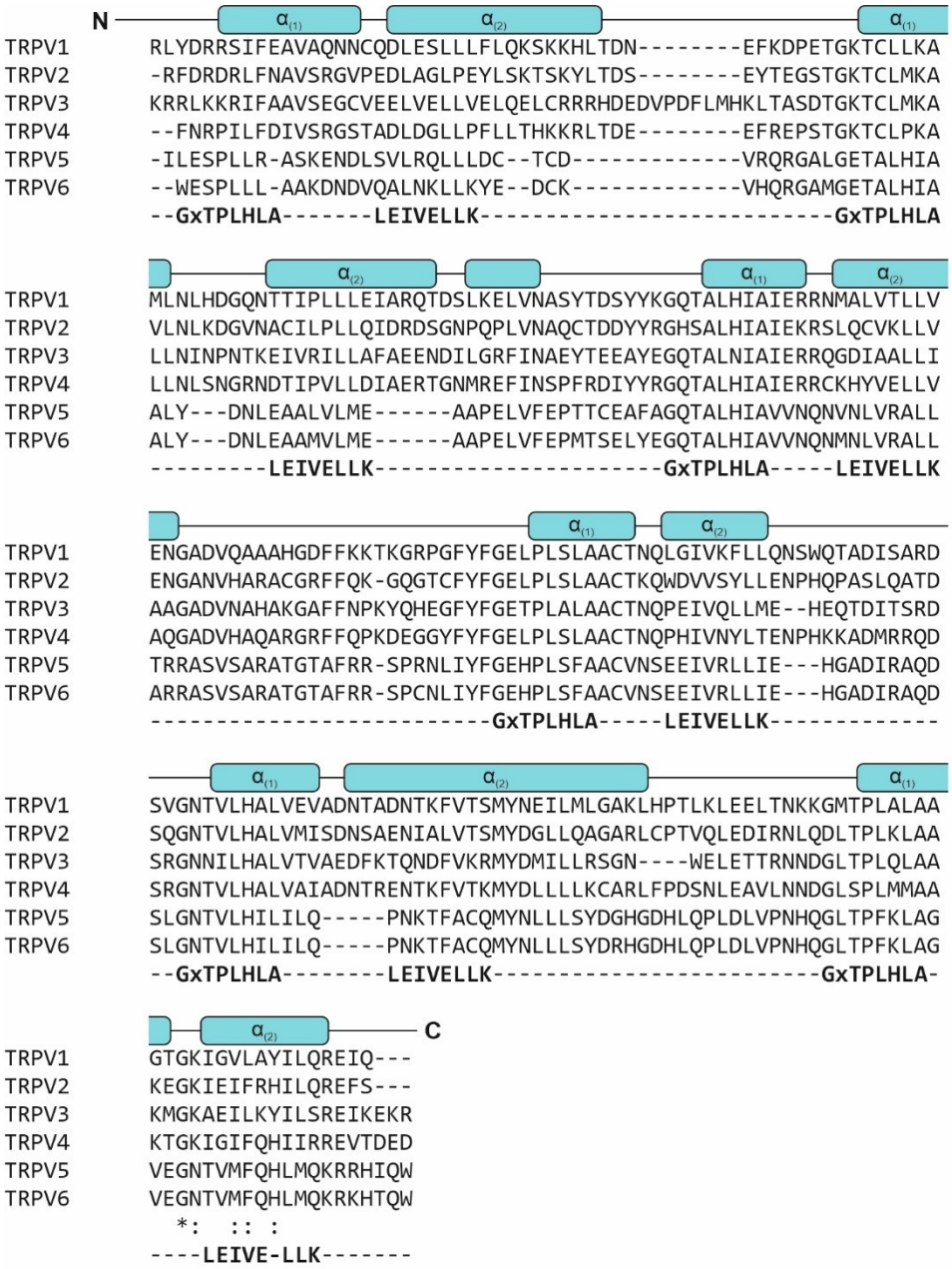


Figure 126: Sequence alignment of TRPV ankyrin repeat domains and comparison with the ARD consensus motif.. The consensus motif (shown in bold letters on the p´bottom of the alignment) in the $\alpha_{(1)}$-helix is GxTPLHLA, where x is any amino acid besides C, G, and P. The consensus motif in the $\alpha_{(2)}$-helix is LEIVELLK. Strong deviations are shown in red, medium deviations in yellow, subtle deviations in light green, and identical residues in dark green. The ARD helices are indicated as cyan boxes on top of the alignment.

1.5 Collaboration partner affiliations

To protect the privacy of the collaboration partners, this section is not available in the online version of the thesis.

VIII Appendix II: other published data from this thesis

1 Hydrophobic core methionine drives tight interactions required for the assembly of spider silk proteins

Parts of this chapter were published in:

J. C. Heiby[#], B. Goretzki[#], C. M. Johnson, U. A. Hellmich, H. Neuweiler, Methionine in a protein hydrophobic core drives tight interactions required for assembly of spider silk. **Nature communications. 10**, 4378 (2019), doi:10.1038/s41467-019-12365-5. ([#] equal contribution)

B. Goretzki, J. C. Heiby, C. Hacker, H. Neuweiler, U. A. Hellmich, NMR assignments of a dynamically perturbed and dimerization inhibited N-terminal domain variant of a spider silk protein from *E. australis*. **Biomolecular NMR assignments. 14**, 67–71 (2020), doi:10.1007/s12104-019-09922-w.

2 Introduction

2.1 Spider silk formation is based on the polymerization of spidroin proteins

Web-weaving spiders use up to seven specialized glands to produce different silk fibers. These fibers have precisely tailored properties suitable for tasks as diverse as prey capture, reproduction, and shelter (*1*). Dragline silk from the major ampullate gland exhibits extraordinary toughness, which makes it perfectly suited as the main structural web fiber and the spider's lifeline (*2*). The remarkable mechanical properties of dragline silk have attracted substantial interest in biotechnology and biomimetic material sciences (*3*). To date, however, it has remained challenging to produce natural spider silk in large quantities or to reproduce the properties of natural spider silk synthetically. A main problem is the limited biochemical understanding of the silk formation process within the spider's spinning duct. It is, therefore, crucial to elucidate the molecular details of the silk constituting proteins, their structural transitions in the spider's spinneret, and their interactions in fibres (*4*).

The major components of spider silk are so-called spidroins, soluble proteins stored as a highly concentrated dope in the spider's ampulla (Figure 127 A). Spidroins have a tripartite composition of a non-repetitive N-terminal domain (NTD, ~130 amino acids), a massive region containing silk-specific repetitive domains (RD, ~2000-4000 amino acids), and a non-repetitive C-terminal domain (CTD, ~110 amino acids) (Figure 127 B). The repetitive domain is composed alanine-rich motifs which can stack as β-sheets to form a crystalline structure, and of glycine-rich regions which seem to form 3_1-helices in an amorphous region. The amount of glycine- and alanine-rich domains is silk-specific and believed to be the primary contributor to the mechanical properties of silk (*5–11*). The NTD and CTD are both five-helix bundles (*12, 13*) and critical modules to polymerize spidroins into silk fibres. During storage in the ampulla of the spinning gland, spidroins exist as dimers via disulfide-linked CTDs (Figure 127 B) (*13, 14*). On demand, the spidroins are passed through a spinning duct in which the changing biochemical environment paired with mechanical forces initiates spidroin polymerization as well as phase and structural transitions. As spidroin dimers pass along the tapering duct, acidification triggers the dimerization of the NTD and thereby induces spidroin assembly into higher oligomers (*12, 15–17*) The process of NTD dimerization is conserved across glands and species and involves site-specific protonation of surface-charges paired with conformational changes of helices in the NTD dimer interface (see section VIII2.2) (*17, 18*). Whereas the role of protonation for NTD dimerization has been studied excessively, the mechanism of conformational changes required to achieve complementarity in the dimer interface is only partly explored. Spidroins finally leave the spinning gland through a spinneret and exit the spider's abdomen as solid fibres (*4*).

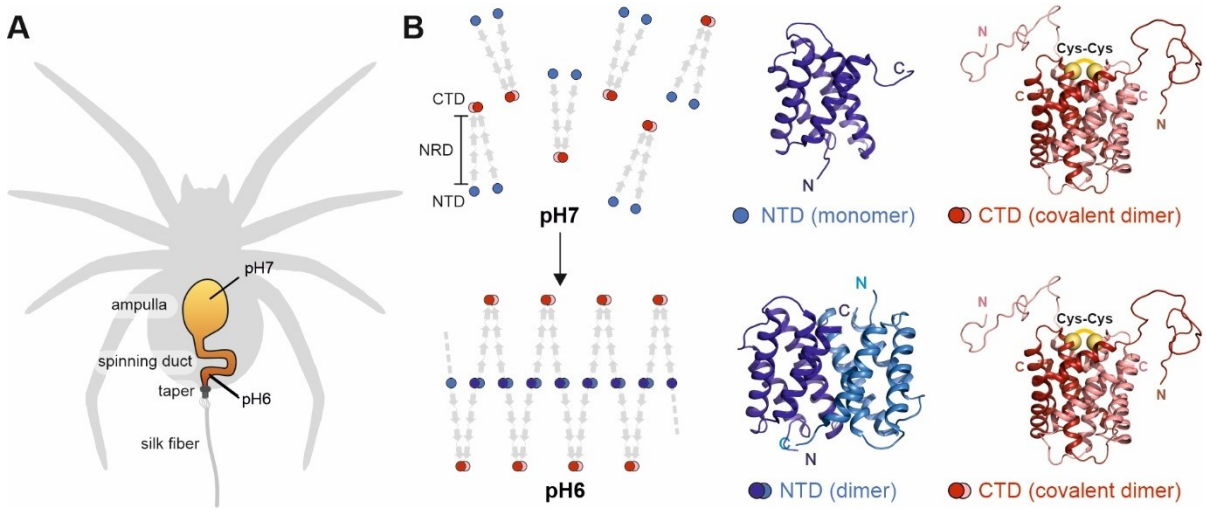


Figure 127: Basic mechanism of the spidroin polymerization process during silk formation in spiders. (**A**) Schematic depiction of the major ampullate gland where spidroins are stored as dimers in a concentrated dope in the ampulla at pH 7. Upon being passed along the spinning duct, changing chemical conditions and shear-forces trigger a yet poorly understood process involving phase-separation and polymerization of spidroins. In the tapering duct, acidification mediates NTD dimerization and enables a tight interaction between spidroins. Additional water removal leads to spidroins being pulled out of the spider's abdomen as sold fibres. (**B**) Spidroin domain architecture comprising the disulfide-linked C-terminal domain (CTD) dimer (red), the repetitive domain (NR) consisting of amorphous and crystalline regions (grey), and the N-terminal domain (NTD, blue). In the ampulla, at pH 7, the NTD is a monomer. Acidification to pH 6 in the taper leads to the formation of anti-parallel NTD dimers, thus causing spidroin to polymerize. NTD monomer: PDB: 3lr2, dimer PDB: 2lpj, CTD dimer PDB: 2khm.

2.2 Spidroin NTD dimerization is a critical step in silk formation

The dimerization of the NTDs is a critical step in spidroin polymerization, and its pH-dependence provides temporal and spatial control over the tight assembly of spidroins in the spider's duct (*12*). According to a recently proposed mechanism (*17*), early dimer formation occurs through long-range electrostatic interactions between acidic and basic residues in the dimer interface (Figure 128, steps **i** and **ii**). Due to a steric mismatch in the dimer interface, the NTD assembly is loosely connected. Subsequently, protonation and conformational changes lead to self-complementary dimer interfaces, thus enabling a more compact NTD dimer (Figure 128, steps **iii**). This pH-dependent dimerization step was proposed to occur via a conformational selection mechanism. According to this mechanism, the NTD is in a dimerization-incompetent conformation at pH 7, but it has a slight tendency to adopt a dimer-competent conformation (Figure 128, steps **i** and **ii**). Acidification to pH 6 shifts the equilibrium from the dimerization-incompetent to the competent conformation and increases the readiness of the NTD to dimerize (*19*). A third step was proposed to include further protonation of a glutamate residue which leads to a highly stable NTD dimer (Figure 128, steps **iv**).

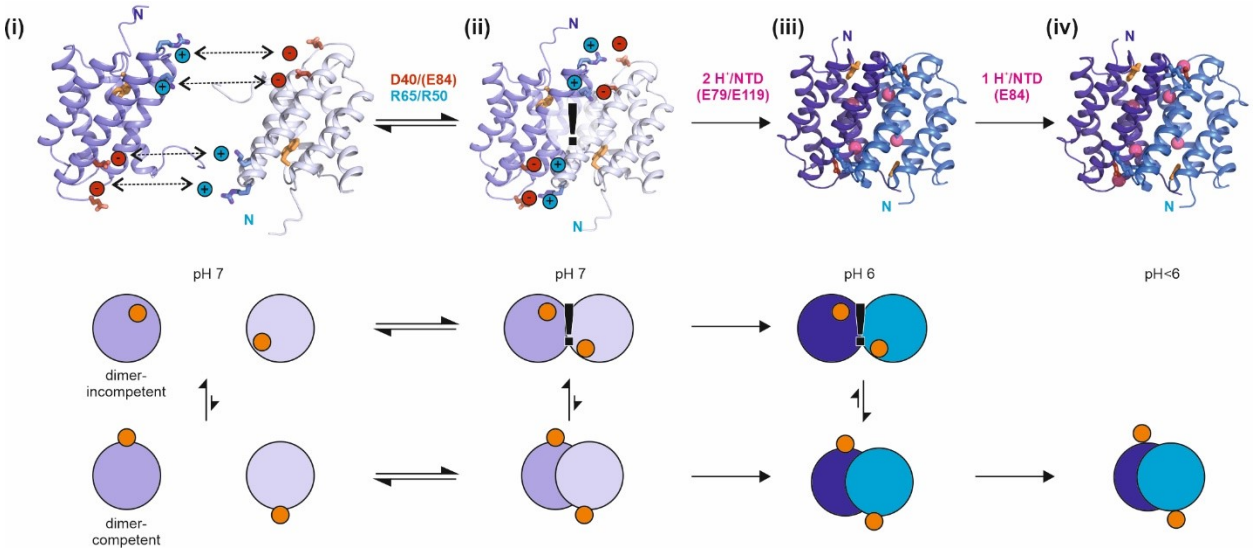


Figure 128: Proposed pH-dependent dimerization mechanism of the *Euprosthenops australis* major ampullate spidroin 1 N-terminal domain (Ea-MaSp1 NTD). (**i**) NTDs at pH 7 in a monomer conformation (dimer-incompetent) with W10 in a buried in the hydrophobic core can transiently populate dimer-competent conformations in which W10 transitions to a solvent-exposed position. (**ii**) Long-range electrostatic attraction mediated by D40, K65, R60 and potentially by E84 enables a weak dimer association. A steric clash in the dimer interface (**!**) prevents a tight interaction. The p*Ka* values of residues E79 and E119 are elevated and enable their protonation at pH 6.5. (**iii**) Protonation of E79 and E119 shifts the equilibrium of the NTD to the dimer-competent conformation, which leads to the formation of a compact and stable dimer. This conformational change requires W10 (orange) to transition from a buried to a solvent-exposed position (**iv**) A drop of pH below 6 causes protonation at E84 and presumably leads to further stabilization of the dimer. The mechanism was adapted from refs (*19*) and (*17*).

A structural alignment of the NTD from the *Euprosthenops australis* (Australian nursery web spider) major ampullate spidroin 1 protein in its monomeric and dimeric state reveals that its interfacial helices α2 and α3 undergo drastic rearrangements during dimerization (Figure 129 A). Tilting of helix α3 towards the centre of the five-helix bundle provides the steric requirements in the dimer interface for tight NTD assembly (Figure 129 B). This movement is accompanied by the reorientation of a conserved tryptophan residue (W10) from a buried to a more solvent-exposed position, which subsequently creates the necessary space for the movement of α3. (*18*). Whereas the driving force for early complex formation can be ascribed to the electrostatic attraction between charged side chains in the dimer interface (*15*), the driving force for the subsequent conformational changes of α2 and α3 and their energetic contributions to the overall dimerization mechanism of the NTD are currently unknown. Without these conformational changes, the NTD dimer cannot form due to major sterical clashes in the interface.

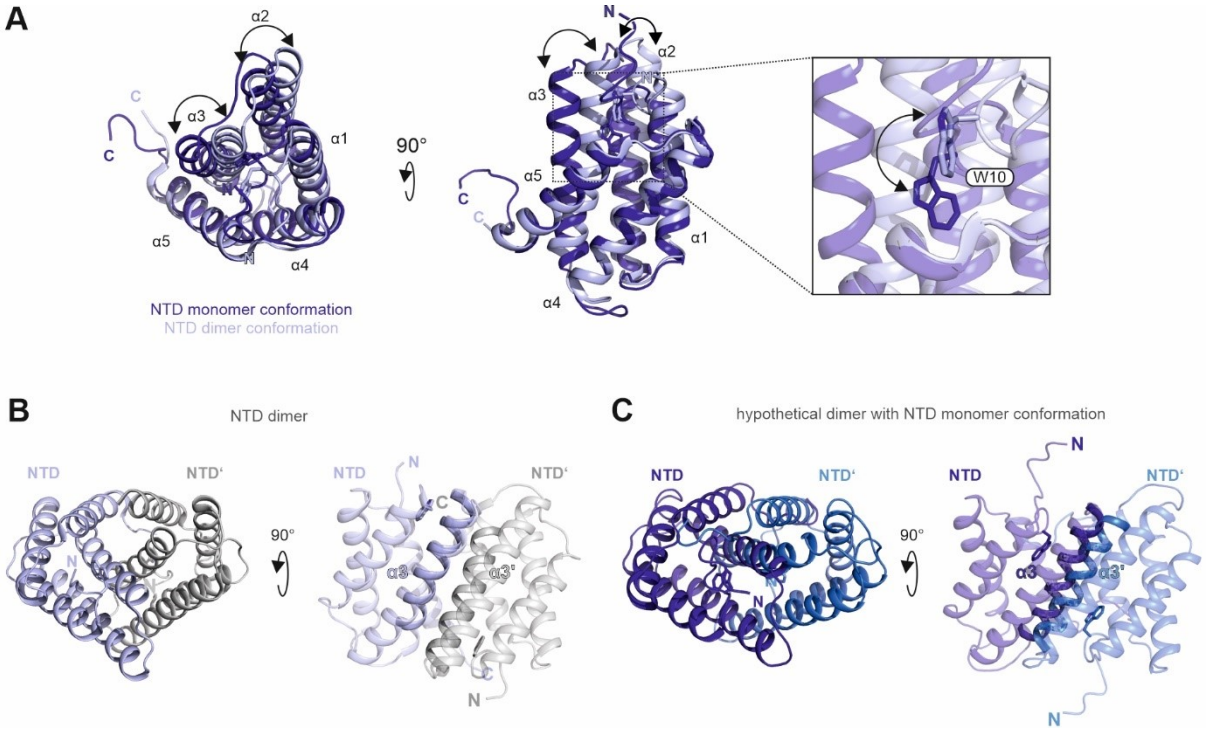


Figure 129: Conformational changes in the *Euprosthenops australis* major ampullate spidroin 1 N-terminal domain (Ea-MaSp1 NTD). (**A**) Structural alignment of Ea-MaSp1 NTD in the monomeric (PDB: 2lpj) and dimeric form (PDB: 3lr2) highlights the tilting of helices α2 und α3 and the movement of W10 from a buried to a solvent-exposed position (inset on the right) upon dimerization. The structures were aligned along α4 und α5. (**B**) Structure of the Ea-MaSp1 NTD dimer (left) and structure of a hypothetical dimer with two NTDs in a monomer conformation (right). (**C**) The hypothetical NTD dimer with monomer conformation was obtained by aligning the NTD monomer structure 2lpj with the two NTDs in the dimer structure 3lr2 along helices α4 und α5. Whereas the dimer conformation exhibits perfectly self-complementary dimerization interfaces required for NTD association, a hypothetical association of NTDs in the monomer conformations results in a steric clash along α3.

2.3 The hydrophobic core of MaSp NTDs is enriched in methionine

Major ampullate spidroin (MaSps) NTDs share highly conserved amino acid sequences which are reflected in their high structural and functional conservation (*12, 13, 16, 18, 20, 21*). An alignment of MaSp sequences from *Euprosthenops australis* (Ea), *Latrodectus geometricus* (Lg), *Latrodectus Hesperus* (Lh), *Nephila clavipes* (Nc), *Argiope trifasciata* (At), and *Nephila inaurata madagascariensis* (Nim) reveals several peculiarities in the NTD's amino acid compositions (Figure 130 A) (*12*). The NTDs contain 13.8% charged residues, which is substantially lower than the 24.9% found on average in proteins from *Bacteria*, *Archaea*, and *Eukaryota (22)*. Charged residues are usually located on protein surfaces to provide water-solubility. The absence of charged residues in MaSp NTDs is compensated by an increase in polar residues (33.3% compared to 19.1% typically found in proteins). The overall content of hydrophobic amino acids in MaSp NTDs (44.8%) is comparable to the content found on average in proteins from all phyla of life (43.1%) (*22*). Hydrophobic amino acids are crucial for the structural integrity of proteins as they form the hydrophobic core. Interestingly, with 15.5% alanine is substantially overrepresented among the hydrophobic amino acids in MaSp NTDs compared to the average 8.1% alanine typically found in proteins (*22*). Alanine is known to stabilize α-helical structures (*23*). As such, the high alanine content in NTD presumably drives the formation of the five-helix bundle. In contrast, leucine, tyrosine, and valine are slightly underrepresented. Strikingly, the methionine frequency (5.7%) is twice as high as in the average protein methionine content, where it is a comparably infrequent amino acid (2.5%) (*22*). In contrast, throughout their overall sequences, dragline silk spidroins contain only 0.2-0.4% methionine (*24*). It is thus surprising that the methionine content in dragline silk mainly clusters in the NTD. An analysis of the methionine positions in the NTDs of various spider species in combination with the three-dimensional methionine

placement within available NTD structures reveals that around two-thirds of the methionine residues are positioned within the protein hydrophobic core (Figure 130 B, C, and D). The high abundance of methionine in MaSp NTD hydrophobic core positions might point to a yet unknown structural and functional role of this residue. (*25*)

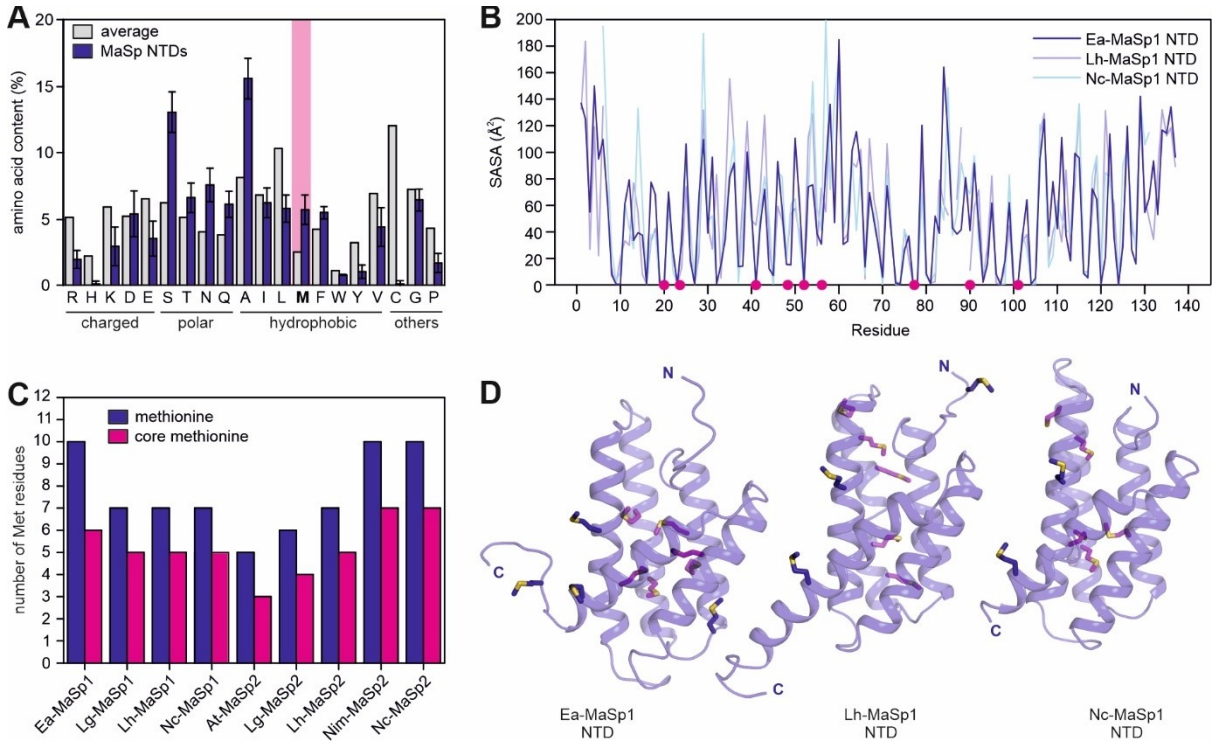


Figure 130: Amino acid sequence analysis of major ampullate spidroin NTDs. (**A**) Average amino acid distribution in MaSp NTDs from *Euprosthenops australis* (Ea), *Latrodectus geometricus* (Lg), *Latrodectus hesperus* (Lh), *Nephila clavipes* (Nc), *Argiope trifasciata* (At), and *Nephila inaurata madagascariensis* (Nim) (dark blue) compared to the amino acid distribution found on average in proteins of all phyla of life (grey) (*22*). (**B**) High-resolution MaSp1 NTD structures were inspected for hydrophobic core methionine residues by computing the solvent-accessible surface area (SASA) with 'SurfRacer5.0' (*26*) from PDB structures 2lpj (Ea-MaSp1 NTD) (*18*), 2n3e (Lh-MaSp1 NTD) (*21*), and 5iz2 (Nc-MaSp1 NTD) (*20*). Positions in which core-methionine residues (low SASA) occur in at least one of the three of the MaSp NTDs are indicated with a magenta sphere. (**C**) Number of overall methionine (dark blue) and core methionine (magenta) residues found in the analyzed MaSp NTD sequences. Core methionine residues in MaSp NTDs for which no structural data are available were identified through sequence alignment with MaSp NTDs from Ea, Lh, and Nc of which structures are available (see sequence alignment Appendix, Figure 145). (**D**) NTD structures from Ea-MaSp1 (*18*), Lh-MaSp1 (*21*), and Nc-MaSp1 NTD (*20*) with surface-exposed methionine (dark blue) and core methionine residues (magenta) highlighted as sticks (Figure 130 C is modified from (*27*)).

3 Material & Methods

Cloning, expression, and purification of the *E. australis* major ampullate spidroin 1 N-terminal domain (Ea-MaSp1 NTD), folding and dimerization experiments, and PET-FCS studies were carried out by Dr. Julia Heiby together with Dr. Hannes Neuweiler (Julius-Maximilians-University Würzbg). SEC-MALS measurements were carried out by Christopher M. Johnson (Medical Research Council Laboratory of Molecular Biology, Cambridge, UK). A detailed protocol of the methods is described in ref. (*25*).

3.1 Nuclear magnetic resonance spectroscopy

Isotope-labeling of proteins

Uniformly ^{15}N- or ^{13}C, ^{15}N-labeled wt-NTD and L6-NTD were generated via expression in *E. coli* using ^{15}N-ammonium chloride and 13C- glucose supplemented M9 medium (see section VI2.17, p. 69, Table 5). 19F-Trp-labelled wt-NTD and L6-NTD were generated via expression in *E. coli* using a defined medium (supplemented 5-fluorotryptophan (Sigma Aldrich), hereafter termed ^{19}F-Trp.

Table 11: Composition of defined medium for ^{19}F-Trp-labeling.

Salts, Additives (per L medium)	**Amino acids (per L medium)**
0.5 g adenine	0.5 g alanine
0.65 g guanosine	0.4 g arginine
0.2 g thymine	0.4 aspartate
0.5 g uracil	0.05 g cysteine
0.2 g cytosine	0.4 g glutamine
1.5 g sodium acetate	0.65 g glutamate
1.5 g succinic acid	0.55 g glycine
0.5 g NH_4Cl	0.1 g histidine
0.85 g NaOH	0.23 g isoleucine
10.5 g K_2HPO_4	0.23 g leucine
20 mL 20% (w/v) glucose	0.42 g lysine-HCl
4 mL 1 M $MgSO_4$	0.25 g methionine
1 mL 10 mM $FeCl_3$	0.13 g phenylalanine
1 mL Magic Mix*	0.1 g proline
10 mL trace elements#	2.1 g serine
	0.23 g threonine
	0.17 g 0.17 tyrosine
	0.23 valine
#Trace elements	***Magic Mix**
2 mg $CaCl_2$-dihydrate	1 mL of Centrum® vitamin tablet
2 mg $ZnSO_4$-heptahydrate	dissolved in 20 mL ddH_2O
2 mg $MnSO_4$-monohydrate	
1 mg biotin	
50 mg thiamine and niacin	

Backbone and sidechain NMR assignments of L6-NTD:

Backbone resonance assignments of ^{13}C, ^{15}N-L6-NTD under monomer conditions (50 mM NaP_i pH 7, 200 mM NaCl, 300 μM protein, 298 K), were carried out using HNCO, HN(CA)CO, HNCA and HNCACB spectra whereas sidechain assignments were carried out using (H)C(CO)NH (100 ms mixing time), H(CCO)NH (100 ms mixing time), HCCH-TOCSY (100 ms mixing time), CBCACONH, HBHACONH experiments and recorded with standard Bruker pulse sequences including water suppression with WATERGATE (*28*)(*29*). Additional 3D experiments to assign the side chains were ^{13}C-NOESY-HSQC (mixing time 150 ms, aliphatic carbons) and ^{15}N-NOESY-HSQC (mixing time 150 ms) experiments (*30*). The assignment of the protein backbone was completed to 96%, aliphatic side chain proton and carbon assignments were completed to 85% and 84%, respectively. The aromatic protons were assigned with 33%.

Assignments of ^{13}C, ^{15}N-L6-NTD under dimer conditions (20 mM MES pH 6, 8 mM NaCl, 298 K) were carried out using solely an HNCA spectrum. The sample concentration was increased to 800 μM to promote dimer formation. Substantial line broadening of amide resonances prevented a more elaborate backbone assignment. The final overall backbone resonances could be assigned to 37% (including 58% of backbone amide resonances).

NMR structure calculation

Structure calculation of the L6-NTD: First, peak picking and NOE assignment of the L6-NTD was performed with the ATNOS/CANDID module in UNIO (*26*) in combination with CYANA (*27*) using the 3D NOESY spectra listed above. Peak lists were reviewed manually and corrected in case of artefacts. Distance restraints within the protein were obtained using the automated NOE assignment and structure calculation protocol available in CYANA (*27*). Backbone H, N, Cα, Cβ chemical shifts were used to calculate torsion angle restraints using TALOS+ (*31*). 100 conformers were calculated with CYANA using the three sets of distance restraints. 20 structures with the lowest target function were submitted to a restrained energy refinement with OPALp (*32*) and the AMBER94 force field (*33*). Structure validation was carried out with the Protein Structure Validation Software suite 1.5 (*34*) restricted to residues with hetNOE values >0.6. (Table 12).

Heteronuclear NOE measurements

{^{1}H},^{15}N-hetNOE experiments of ^{15}N-wild-type Ea-Masp1 NTD and ^{15}N-L6 Ea-MaSp1 NTD: {^{1}H},^{15}N-heteronuclear nuclear Overhauser effect (hetNOE) experiments for ^{15}N-labeled wt- NTD and L6-NTD were recorded using Bruker standard pulse sequences. Experiments were run in an interleaved fashion with and without proton saturation during the recovery delay. Peak integrals were obtained using Bruker TopSpin 3.2. The {^{1}H},^{15}N-hetNOE data were recorded on a Bruker 600 MHz spectrometer in an interleaved manner with a ^1H saturation period of 5 s duration on resonant or 10000 Hz off-resonant for the cross- and reference experiment, respectively. The relaxation delay was set to 3 s. Error bars were obtained from to consecutive {^{1}H},^{15}N-hetNOE experiments of both wt- and L6-NTD.

Hydrogen-deuterium exchange experiments

H/D exchange: ^{15}N-labeled wild-type NTD and L6-NTD mutant were lyophilized in an NMR tube to remove H_2O. The dried protein was dissolved in an equivalent amount of D_2O and immediately placed into the NMR spectrometer to record a series of ^{1}H,^{15}N-HSQC spectra with 12 min intervals. The relative hydrogen-deuterium exchange (rel. HDX) of each backbone amide group at each time point t_i was derived from the signal loss of the corresponding peak between time point t_i and time zero (t_0), i.e. the signal in non-deuterated buffer under equal conditions. The final relative HDX values were calculated as the ratio between the integral V_i, and the integral V_0.

^{19}F-NMR experiments

^{19}F-NMR spectra of ^{19}F-Trp-labeled wt-NTD and L6-NTD were recorded on a Bruker 600 MHz spectrometer at 298 K using a protein concentration of 110 µM. Buffers contained 25 mM BisTris at pH 5.5, 6.0, 6.5, 7.0, 7.5 each at NaCl concentrations of 60, 100, 140, 180, 220, or 260 mM. Each buffer was supplemented with 10% D_2O. Spectra of ^{19}F-Trp-labeled wt-NTD and L6-NTD were recorded in each buffer using standard 1D Bruker pulse sequences.

4 Results & discussion

4.1 The five-helix bundle NTD fold is preserved upon methionine-depletion

The so-called hydrophobic effect is the driving factor for protein folding into three-dimensional structures and leads to the formation of a hydrophobic protein core (*31*). Within this hydrophobic core, the tight packing of hydrophobic amino acid side chains enables extensive van-der-Waals (vdW) contacts, which hold a protein together (*32*). Disturbances in the core interaction network typically destabilize a protein and cause misfolding and aggregation. Modifications of hydrophobic protein cores, e.g. through mutagenesis, are thus frequently not tolerated since they are detrimental to a protein's structural integrity, particularly if the replacement is not isosteric and exhibits a different hydrophobicity (*33*).

To investigate the impact of hydrophobic core methionine residues on the structure, dynamics, and functionality of MaSp NTDs, a mutagenesis study was carried out using the major ampullate spidroin 1 (MaSp1) NTD from *Euprosthenops australis* (Ea) as a representative spider silk NTD. Among the NTDs included in the sequence analysis (Figure 130), this variant is the most intensively studied protein and various structural and dynamic information are available (*12, 15, 17–19, 34*). The Ea-MaSp1 NTD (hereafter referred to as wt-NTD) is a 137-residue protein with six hydrophobic core methionine residues at positions M20, M24, M41, M48, M77 and M101 (Figure 130, Appendix Figure 145). The aliphatic and branched leucine side chain has a similar hydrophobicity and size as methionine and therefore seems to be a suitable replacement for a mutagenesis study on hydrophobic core methionine residues (*35*). Additionally, the sequence alignment of MaSp NTDs of different species indicates that methionine and leucine in hydrophobic core positions are almost interchangeable (see Appendix Figure 145). Consequently, a methionine-depleted NTD protein in which the six methionine residues were successively replaced by leucine was generated via site-directed mutagenesis. The construct is accordingly termed L6-NTD.

Both the wt-NTD and the L6-NTD protein construct could be expressed in *E. coli* in high amounts and purified as soluble and α-helical proteins as demonstrated via circular dichroism (CD) spectroscopy (Figure 131). BeSt-Sel analysis of the CD spectra confirms that the secondary structure content of the wt-NTD is preserved in the L6-NTD. In agreement with the five-helix bundle structure observed in the high-resolution structure (PDB code 2lpj), the wt-NTD and L6-NTD proteins contain almost 60% α-helical secondary structures.

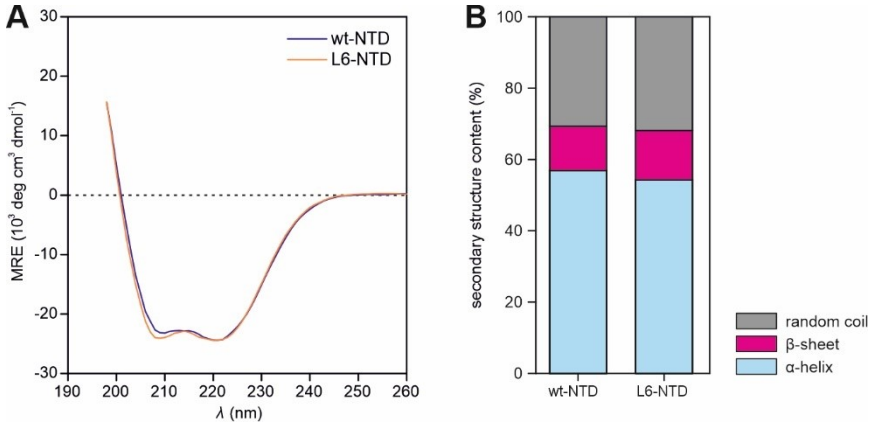


Figure 131: Secondary structure analysis of the wt- and L6-NTD. (**A**) Far-UV CD spectra of wt-NTD (orange) and L6-NTD (blue). The spectra were recorded at 298 K and at a protein concentration of 10 µM in 50 mM phosphate buffer, pH 7.0, with the ionic strength adjusted to 200 mM using potassium chloride. (B) BeStSel secondary structure analysis of the spectra in (A). The secondary structure content of the wt-NTD is preserved in the L6-NTD.

To verify that the NTD five-helix bundle fold is preserved in the methionine-depleted construct, L6-NTD was additionally investigated via NMR spectroscopy. For this, ^{15}N-labeled wt- and L6-NTD were generated in *E. coli*, purified (see ref. (*25*) for details) and ^{1}H, ^{15}N-HSQC NMR spectra were recorded (Figure 133 A). The spectra show a peak dispersion typical for α-helical proteins. NMR backbone assignments of wt-NTD could be transferred from previously published assignments (*18*), which are deposited in the Biomagnetic Resonance Databank (BMRB, entry 18262). However, the peak positions in the L6-NTD spectrum are drastically shifted compared to the peaks in the wt-NTD spectrum. It was, therefore, necessary to obtain a *de novo* NMR backbone assignment of L6-NTD under monomer conditions (pH 7, 60 mM ionic strength). Using ^{13}C, ^{15}N-L6-NTD and a set of triple-resonance NMR experiments, i.e. HNCO, HN(CA)CO, HNCA and HNCACB (described in section VIII3.1, see also section VI2.17) (*27*), the L6-NTD backbone amide resonances could be assigned to 93%. The N-terminal residues G1 and S2 of L6-NTD are artificial remnants from a thrombin cleavage site. The resonances of both residues were not assigned. As previously observed for the wt-NTD, no amide backbone peaks could be identified for G3, H6, N64 and N113. However, C_α/H_α backbone resonances assignments for G3, H6, N64 and N113 could be obtained from HNCACB, CBCACONH and HBHACONH experiments and were confirmed through NOE connectivities with neighbouring residues in ^{13}C- and ^{15}N-NOESY-HSQC experiments. The final assignments of H, H_N, H_α, C_α, C' backbone resonances were completed to 96%, including proline and cleavage tag residues.

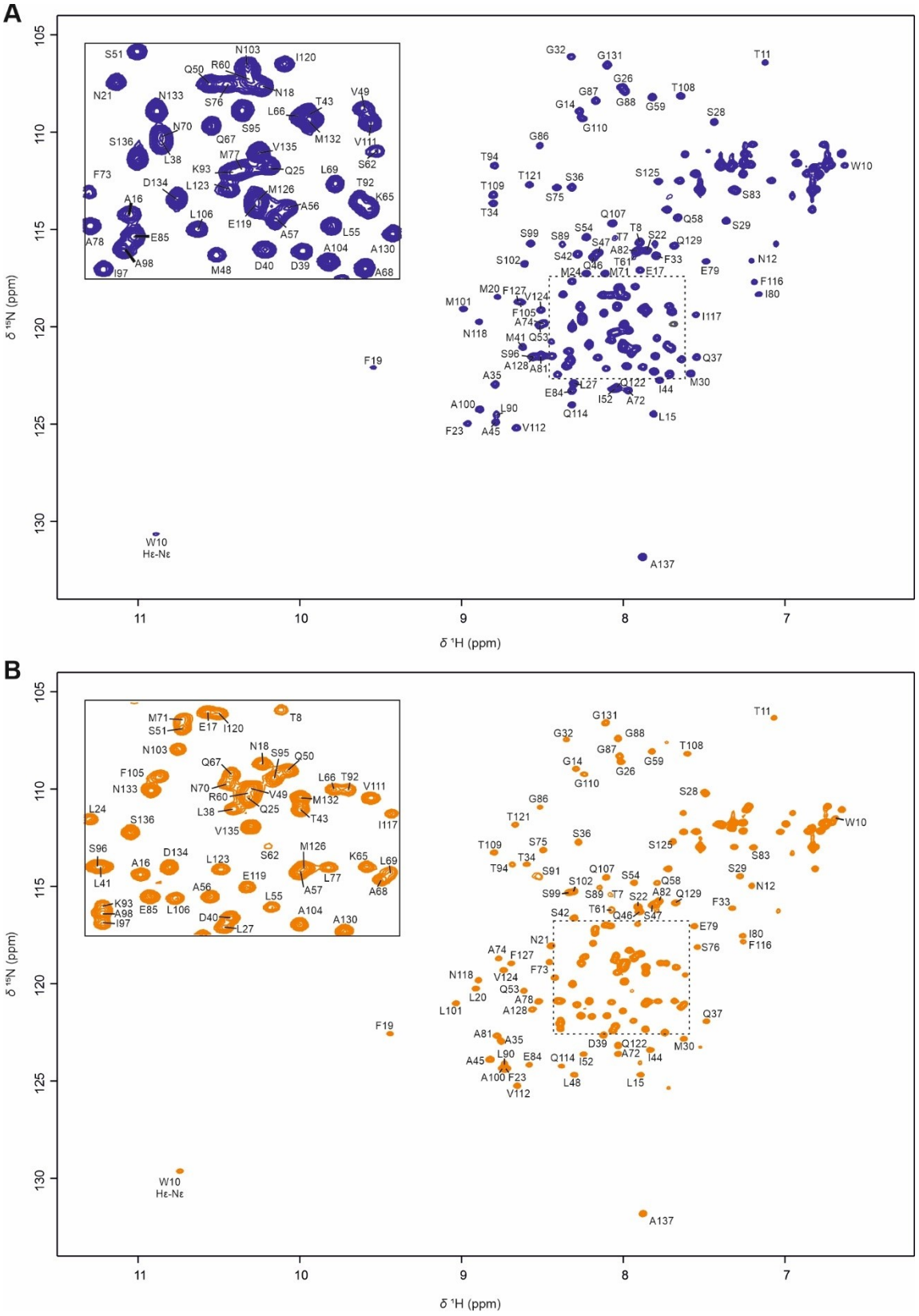


Figure 132: NMR backbone assignments of wt- and L6-NTD at pH 7. $^{1}H,^{15}N$-HSQC NMR spectrum of (**A**) ^{15}N-wt-NTD and (**B**) ^{15}N-L6-NTD under monomer conditions (300 µM protein concentration, pH 7, 200 mM ionic strength). L6-NTD assignments were obtained *de novo* (see section VI2.17), whereas wt-NTD assignments were transferred from BMRB entry 18262.

A comparison of the chemical shifts obtained for ^{15}N-wt-NTD and ^{15}N-L6-NTD reveals moderate as well as very strong differences in the chemical shifts of the protein backbone resonances throughout the protein sequence (Figure 133 B). As is expected for mutations in a hydrophobic core, the strongest chemical shift changes can be assigned to residues located in the centre of the five-helix bundle. To confirm that the five-helix bundle topology is preserved when exchanging core methionine residues by leucine, TALOS+ (*36*) was used for predicting the secondary structure of L6-NTD (Figure 133 C). This was then compared to the results obtained with the chemical shifts of wt-NTD (BMRB, entry 18262) (*18*). Even though the NTD backbone resonances are substantially perturbed through methionine to leucine exchange, TALOS+ predicts an identical topology for both L6-NTD and wt-NTD. This indicating that the exchange of methionine by leucine in the hydrophobic core of NTD does not alter the five-helix bundle topology.

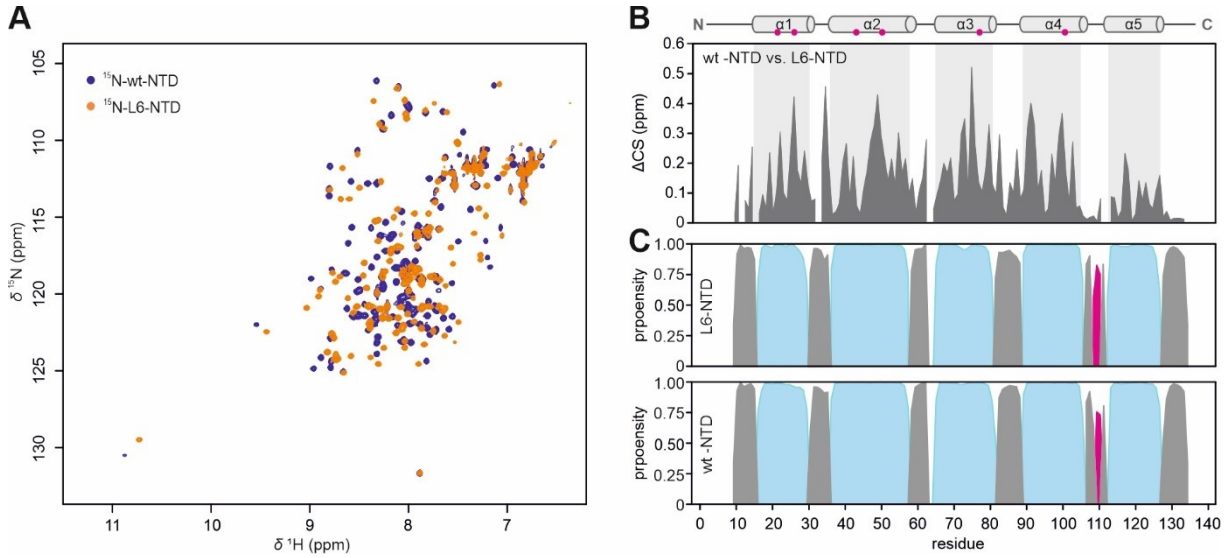


Figure 133: NMR chemical shift analysis of wt- and L6-NTD. (**A**) Overlay of ^{15}N-wt-NTD and ^{15}N-L6-NTD ^{1}H, ^{15}N-HSQC NMR spectra under monomer conditions (300 µM protein, pH 7, 200 mM NaCl, 10% D_2O). (**B**) Chemical shift difference (ΔCS) between backbone amide resonances of wt- and L6-NTD plotted against the residue number. On top of the diagram is a topology representation of the Ea-MaSp1 NTD showing the position of the helices α1-5 (derived from the available solution NMR structure, PDB: 2lpj). The positions of the methionine-leucine mutations are indicated with magenta spheres. (**C**) TALOS+ (*36*) secondary structure prediction of L6-NTD and wt-NTD based on H_N, N, H_α, C_α, and C' chemical shifts. Grey: random coil, light blue: α-helix, magenta: β-sheet.

To further verify that the mutations in L6-NTD do not cause local structural changes such as tilting of helices or reorientation of sides which alters core vdW-contacts, the solution NMR structure of L6-NTD was determined (Figure 134, Table 12, see section VI2.17). The structure was deposited in the PDB with the accession number 6qjy. The overlay of the ten lowest energy structural models calculated for L6-NTD shows high convergence, particularly along the five helices (Figure 134 A). The low convergence of N- and C-terminus is a common observation in solution NMR structures and reflects their flexibility and dynamic properties. The L6-NTD structure bundle confirms that the three-dimensional fold of the five-helix bundle is conserved upon replacing all core methionine with leucine residues. Importantly, the L6-NTD structure aligns well with the published NMR structure of wt-NTD (PDB: 2lpj) (*18*). An all-atom-alignment of residues 9-129 (omitting the flexible N- and C-termini) yields a heavy-atom RMSD of 1.2 Å and a backbone heavy-atom RMSD of 1.2 Å (Figure 134 C). Crucially, the positions and orientations of the leucine side chains in L6-NTD superimpose extremely well with the methionine side chains in wt-NTD (Figure 134 B). Additionally, the buried position of W10 in the protein core is preserved in L6-NTD (Figure 134 B). In conclusion, the solution NMR structure of L6-NTD confirms the structural integrity of the Ea-MaSp1 NTD upon the replacement of methionine residues in the hydrophobic core.

Table 12: Structural statistics for the NMR solution structure of MaSp1 L6-NTD.

Conformational restricting restraints	
Total NOE distance restraints	1673
intra-residue $\|i = j\|$	448
sequential $\|i - j\| = 1$	463
medium-range $1 < \|i - j\| < 5$	347
long-range $\|i - j\| \geq 5$	415
Dihedral angle restraints (Talos-N)	216
No. of restraint per residue	14.4
No. of long-range restraints per residue	3.2
Residual restraint violations[a]	
Average no. of distance violations per structure	
0.1-0.2 Å	6.95
0.2-0.5 Å	0
>0.5 Å	1
Average no. of dihedral angle violations per structure	
1-10°	4.4
>10°	0
Model quality (ordered residues)[a]	
RMSD backbone atoms (Å)	2.9
RMSD heavy atoms (Å)	2.9
RMSD bond lengths (Å)	0.02
RMSD bond angles (°)	0.25
MolProbity Ramachandran statistics[a]	
Most favored regions	98.4
Allowed regions	1.6
Disallowed regions	0
Global quality scores (raw/Z score)[a]	
Verify3D	0.39
Prosall	0.84
PROCHECK (φ-ψ)	0.47
PROCHECK (all)	-0.11
MolProbity clash score	2.80
Model contents	
Ordered residue ranges (HetNOE > 0.6)	10-60, 63-82, 90-108, 110-129
Total no. of residues	137
BMRB accession number	27683
PDB ID code	6qjy

[a] calculated using PSVS 1.5 for using ordered residues (HetNOE > 0.6) (Bhattacharya et al., 2007). Average distance violations were calculated using the sum over r^{-6}

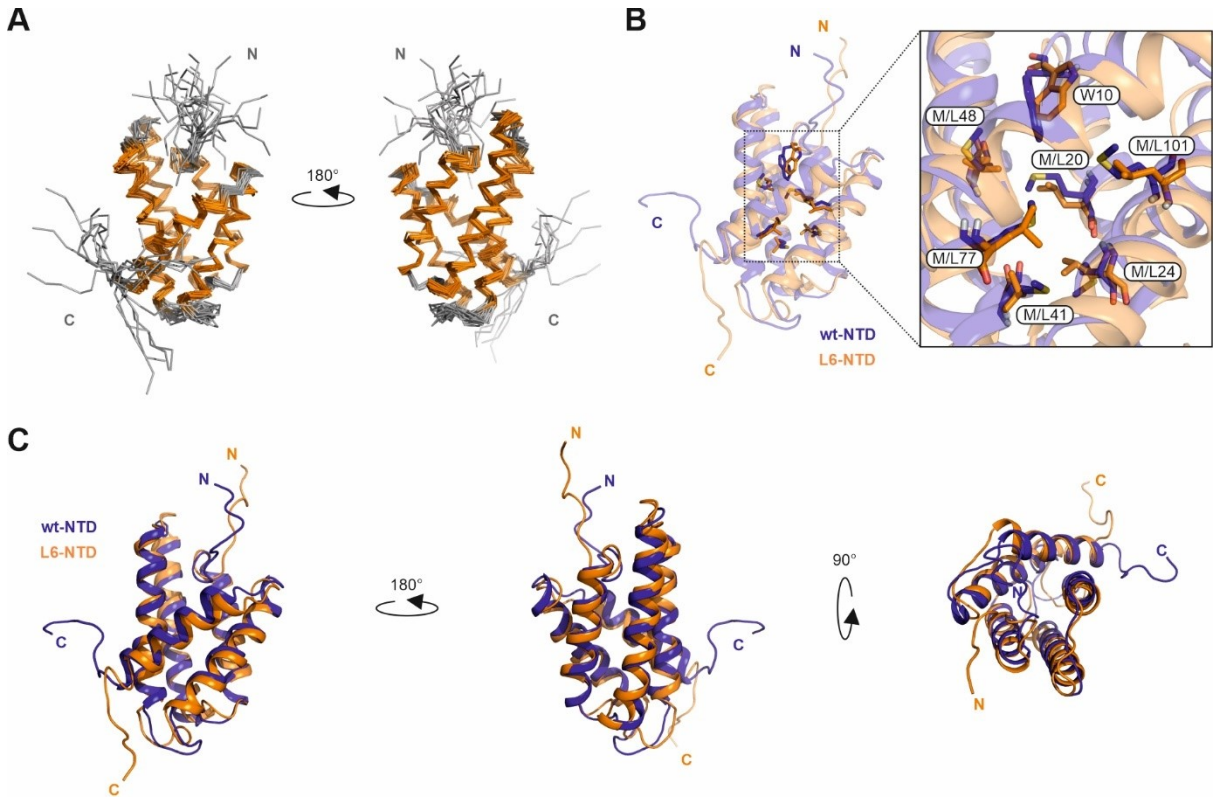


Figure 134: Solution NMR structure of L6-NTD. (**A**) Overlay of the ten lowest energy L6-NTD conformers calculated from NOE-based distance restraints. (**B**) Superposition of wt-NTD (PDB: 2lpj) and L6-NTD (PDB: 6qjy) core methionine/leucine side chains and W10 reveal the structural integrity of the hydrophobic core upon core methionine-depletion. (**C**) Superposition of wt-NTD and L6-NTD aligned along residues 9-129 viewed from different perspectives (heavy-atom RMSD of 1.2 Å).

4.2 Replacement of core methionine by leucine increases NTD stability and impairs dimerization

Having established the structural integrity of the methionine-depleted L6-NTD protein, the effect of the hydrophobic core mutations on the stability of the NTD was probed via chemical and thermal denaturation using far-UV CD and tryptophan fluorescence spectroscopy (Figure 135 A and B, Table 13). Equilibrium denaturation experiments under monomer conditions at pH 7 revealed that the free energy of folding increases from 5.6 ± 0.1 kcal/mol for wt-NTD to 8.0 ± 0.3 kcal/mol for L6-NTD. The equilibrium *m*-value ($m_{D\text{-}N}$) as the first derivative of change in stabilization free energy upon heating is slightly reduced for L6-NTD compared to wt-NTD. The $m_{D\text{-}N}$-value correlates with changes in the solvent-accessible surface area differences between native and folded state (*31*). Considering that the monomer structures of wt- and L6-NTD are virtually identical, the reduction in $m_{D\text{-}N}$ indicates a more compact denatured state in L6-NTD compared to wt-NTD. Consistent with the increase in folding free energy, the thermal stability of NTD increased from T_m = 61.0 ± 0.1°C for wt-NTD to T_m = 81.0 ± 0.2°C for L6-NTD.

To test whether the increase in stability is an additive effect of the methionine-leucine mutations (*37*), five cumulative mutants termed L1-L5[§§], were generated and analyzed with regards to their thermal stability (Figure 135 C). Indeed, the melting point increased almost linearly from L1- to L6-NTD. Only the M48L mutation showed little effect as indicated by the similar melting points of L3- and L4-NTD. This might be explained by residue M48 being less buried in the core than the other methionine residues and thus participating in fewer

[§§] L1: M20L; L2: M20L, M24L; L3: M20L, M24L, M41L; L4: M20L, M24L, M41L, M48L; L5: M20L, M24L, M41L, M48L, M77L; L6: M20L, M24L, M41L, M48L, M77L, M101L

stabilizing vdW-contacts. On the other hand, the M48 side chain is sufficiently close to the W10 indole ring to form a potentially stabilizing interact with the tryptophan indole ring (*38*). The loss of stability by removing the M48-W10 interaction might be compensated by a gain of stability through substitution with leucine.

The enhanced stability in the NTD through methionine-leucine substitution was further investigated by determining folding and unfolding rate constants using stopped-flow tryptophan fluorescence spectroscopy (Figure 135 D and E, Table 13). A recent study on a set of homologous MaSp NTDs revealed rapid folding of wt-NTDs on a time scale of ~100 µs through a barrier limited two-state transition (*16*). Strikingly L6-NTD was found to fold 40 times faster (k_f = 519 000 ± 213 000 s^{-1}) compared to wt-NTD (k_f = 13 000 ± 6 000 s^{-1}). The theoretical folding speed limit for a single domain protein with N-residues is τ_f = N/100 µs (*39*), which translates to ~1.4 µs for the 137-residue NTD. From the determined folding rate of L6-NTD it follows that τ_f = 1.9± 0.8 µs. Thus, L6-NTD folds at the theoretical speed limit. In addition to an increased folding rate, the unfolding rate constant was 15-fold decreased in L6-NTD (k_u = 0.2 ± 0.1 s^{-1} compared to k_u = 3 ± 1 s^{-1} for wt-NTD). From the increase in folding rate and decrease in unfolding rate follows a decrease of the free energy barrier between denatured and transition state by -2.2 kcal/mol, and an increase of the free energy barrier between the native and transition state by ~1.6 kcal/mol (Figure 135 F, Table 13).

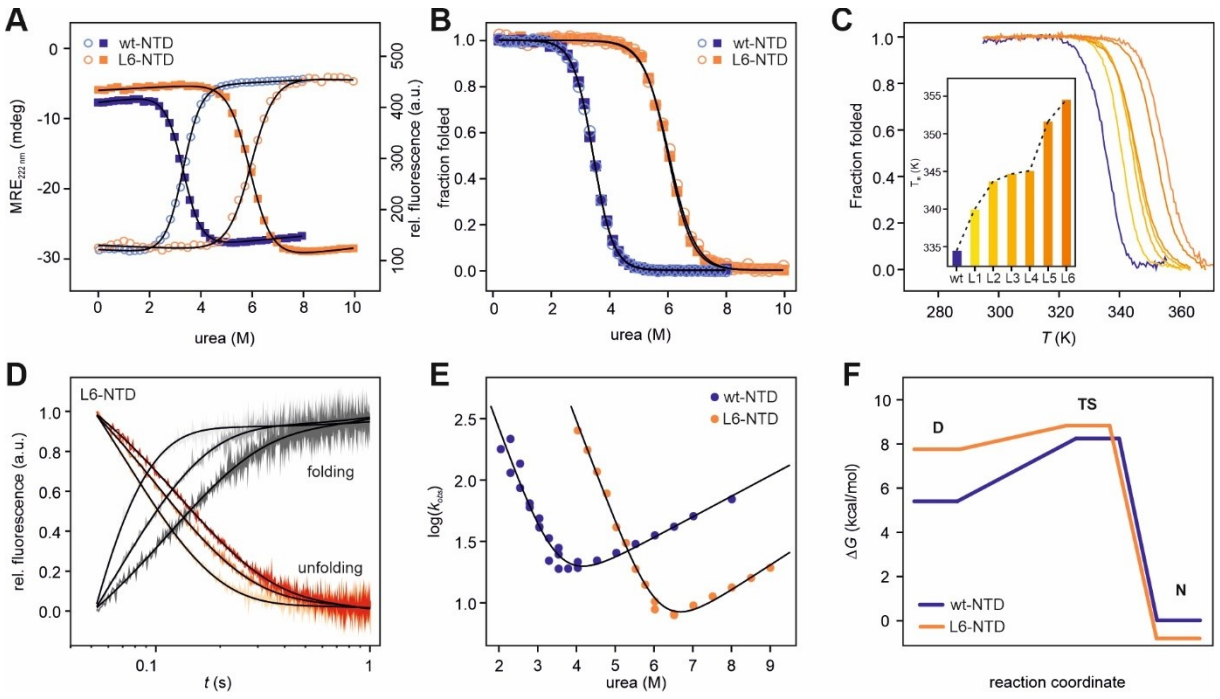


Figure 135: Influence of methionine-to-leucine exchange on NTD stability. (**A**) Chemical denaturation of wt- and L6 NTD monitored via far-UV CD spectroscopy (open circles) and tryptophan fluorescence (squares). (**B**) Data from (A) normalized to the fraction of folded protein. (**C**) Thermal denaturation of wt-NTD and the six cumulative NTD mutants L1-L6§§ monitored via CD spectroscopy (see Figure 131). The inset shows the gradual increase of the melting temperatures T_m as a consequence of methionine-to-leucine exchange in the NTD hydrophobic core. (**D**) Normalized kinetic transients of folding (light grey, grey, dark grey; 5.0, 5.5, 6.0 M urea, respectively) and unfolding (red, orange, light orange; 6.5, 7.5, 8.5 M urea, respectively) of L6-NTD measured via stopped-flow tryptophan fluorescence spectroscopy. (**E**) Observed rate constants of wt- and L6-NTD plotted versus urea concentration. Data were fitted using a model for a barrier-limited two-state transition. See ref. (*27*) for details. (**F**) Free-energy diagram of wt- and L6-NTD folding and unfolding. D: denatured state, TS: transition state, N: native state. The free energy of the wt-NTD native state was set to zero as reference. The figure was modified from ref. (*27*).

The preserved structure with a concomitant increase in stability of L6-NTD raises the question whether the protein functionality, i.e. the ability to dimerize upon pH decrease (*12*), is maintained or altered. At pH 7, NTDs exist as monomers, whereas at pH 6, they are tight dimers with dissociation constants in a low nanomolar range (*16*). The ability of L6-NTD to dimerize was investigated via SEC-MALS (Figure 136 A). At pH 7 and high

ionic strength (200 mM), the protein elutes as a monomer, similar to wt-NTD. At pH 6 and low ionic strength, the protein elutes at a molecular weight (M_W) between that of a dimer and a monomer. This might indicate the existence of a rapid and dynamic monomer-dimer equilibrium for L6-NTD at pH 6. Wt-NTD at pH 6 eluted at an M_W which was only 7% lower than the theoretical dimer-M_W. This is well within the error of the measurement (±3–5%) and may be explained by the small amounts of salt (60 mM ionic strength) in the buffer which are needed to prevent protein interactions with the column material. High concentrations of salt cause Debye-Hückel screening of electrostatic interactions and thus weakening electrostatic attraction in the dimerization interface (*40*). To strengthen dimerization, the ionic strength in the pH 6 buffer was reduced to 8 mM which yielded an M_W of L6-NTD closer to that of the expected dimer in SEC-MALS experiments (Figure 136 B). The gradual decrease of the L6-NTD protein concentration, from 12 μM to 0.4 μM reduced the M_W. This indicates that the dissociation constant at pH 6 is in a μM range and thus substantially reduced compared to Wt-NTD. However, it was not possible to reach a saturation point within the probed concentration range due to to the low sensitivity of SEC-MALS below 0.4 μM L6-NTD.

Table 13: Thermodynamic and kinetic data of folding obtained for wt-NTD and L6-NTD. Values were calculated according to ref. (*27*). Errors are s.e. from regression analysis and propagated s.e.

Parameter	wt-NTD	L6-NTD
$m_{D\text{-}N}^{CD}$ (kcal mol^{-1} M^{-1})	1.69 ± 0.03	1.30 ± 0.05
$m_{D\text{-}N}^{F}$ (kcal mol^{-1} M^{-1})	1.69 ± 0.02	1.37 ± 0.03
$[\text{urea}]_{50\%}^{CD}$ (M)	3.30 ± 0.01	5.99 ± 0.02
$[\text{urea}]_{50\%}^{F}$ (M)	3.31 ± 0.01	5.95 ± 0.01
ΔG_{unfold}^{CD} (kcal mol^{-1})	5.6 ± 0.1	7.8 ± 0.3
ΔG_{unfold}^{F} (kcal mol^{-1})	5.6 ± 0.1	8.2 ± 0.2
k_f (s^{-1})	13 000 ±6 000	519 000 ± 213 000
k_u (s^{-1})	3 ± 1	0.2 ± 0.1
ΔG_{unfold} (kcal mol^{-1}) [a]	4.96 ± 0.08	8.75 ± 0.05
m_f (kcal mol^{-1} M^{-1})	1.2 ± 0.1	1.13 ± 0.05
m_u (kcal mol^{-1} M^{-1})	0.23 ± 0.03	0.30 ± 0.04
β_T	0.84 ± 0.02	0.79 ± 0.09
$\Delta\Delta G_{D\text{-}TS}^{wt\text{-}L6}$ (kcal mol^{-1}) [b]	2.18 ± 0.03	
$\Delta\Delta G_{N\text{-}TS}^{wt\text{-}L6}$ (kcal mol^{-1}) [c]	1.60 ± 0.10	

[a] determined from k_u and k_f via $\Delta G_{unfold} = -RT\ln(k_u/k_f)$; [b] determined from k_u^{wt} and k_u^{L6} via $\Delta\Delta G_{D\text{-}TS}^{wt\text{-}L6} = -RT\ln(k_u^{wt}/k_u^{L6})$; [c] determined from k_u and k_f via $\Delta\Delta G_{N\text{-}TS}^{wt\text{-}L6} = -RT\ln(k_f^{wt}/k_f^{L6})$

To estimate the L6-NTD dimer affinity, a dilution series of L6-NTD at pH 6 was carried out with high-resolution SEC in combination with tryptophan fluorescence, which enables detection in sub-micromolar concentrations (Figure 136 C). The shift in elution volume in dependence of the protein concentration could be fitted with a binding isotherm from which a K_d of 1.1 ± 0.2 μM for L6-NTD dimerization could be determined (Figure 136 D). In contrast, the elution volume of wt-NTD was nearly unaffected by the protein concentration, consistent with its reported dimer affinity in the subnanomolar range (1.1 ± 0.1 nM) (*15*). The L6-NTD dimer affinity is thus three orders of magnitude lower than that reported for wt-NTD. Notably, the elution volume of the L6-NTD dimer was significantly lower than for the wt-NTD dimer indicating larger dimensions of the L6-NTD dimer (see Figure 136 D). SEC analysis of the L1-to L5-NTDs demonstrates that the perturbed ability to dimerize is an additive effect of the methionine-leucine mutations (Figure 136 E). With increasing methionine-depletion, the NTD's dimer affinity decreases (Figure 136 F). Concomitantly, the SEC elution volume of the NTD dimer increased almost linearly to the amount of methionine-leucine mutations, indicating increasing dimer dimensions. The dimensions of the monomeric NTDs were unaffected by the mutations.

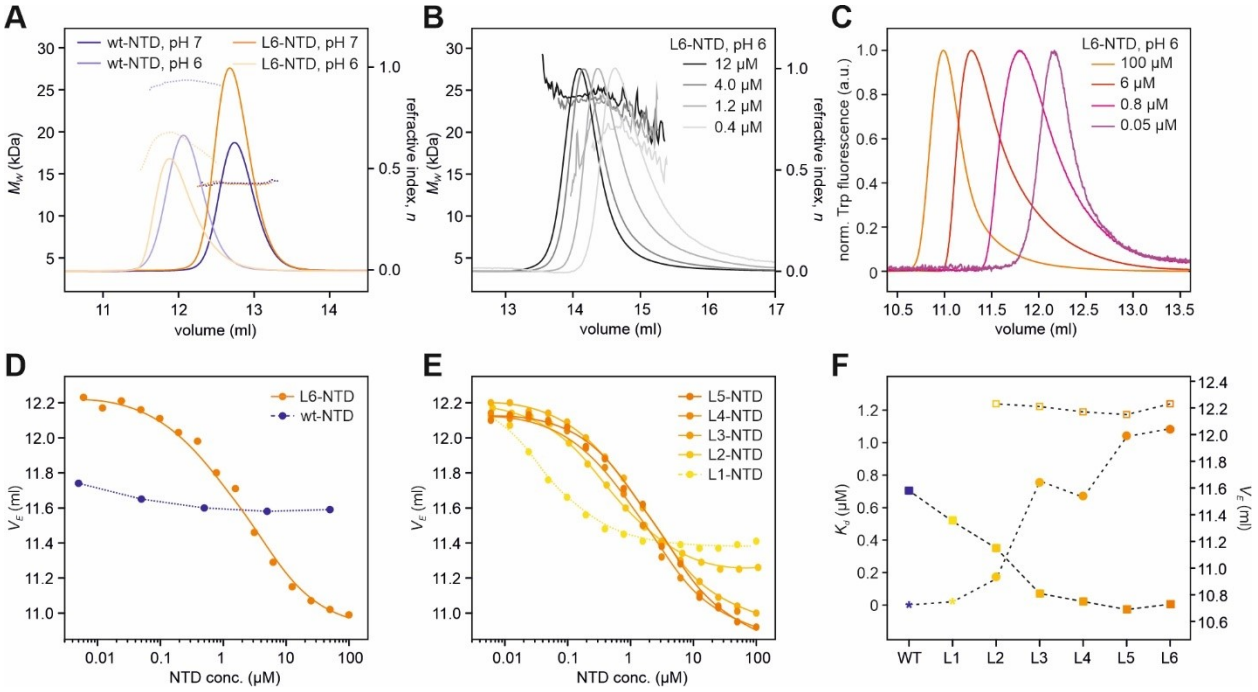


Figure 136: Effect of methionine-depletion on NTD dimerization. (**A**) SEC-MALS analysis of wt- and L6-NTD under monomer (pH 7, 200 mM ionic strength) and dimer conditions (pH 6, 60 mM ionic strength). Solid lines represent the SEC elution bands whereas the broken lines represent the molecular weight M_W of the eluted particles. (**B**) Protein concentration-dependent SEC-MALS measurements of L6-NTD under dimer conditions (pH 6, 60 mM ionic strength). (**C**)Normalized tryptophan fluorescence high-resolution SEC traces of 100, 6, 0.8, and 0.05 µM L6-NTD. (**D**) SEC elution volume V_E plotted versus protein concentration of wt-NTD (blue) and L6-NTD (orange). The solid orange line is a fit of the L6-NTD data using a thermodynamic model for a monomer-dimer equilibrium. (**D**) SEC elution volumes V_E plotted versus protein concentration of L1- to L5-NTD. (**F**) Dimer affinities K_d (spheres) of L1- to L6-NTD mutants determined via high-resolution SEC. The wt-NTD dimer affinity value was adapted from the literature (*15*) (denoted by a blue asterisk). The L1-NTD dimer affinity value is an estimate because of the non-defined monomer baseline (denoted by a yellow asterisk). Closed and open squares represent the end-point elution volume values of dimer and monomer NTDs, respectively. Wt- and L1-NTD monomer end-point values could not be obtained because their K_d was too low to enter the baseline regions. The figure is modified from (*27*).

4.3 Substituting core methionine by leucine conformationally restricts the NTD

The impaired ability of L6-NTD to dimerize compared to wt-NTD is not due to a changed three-dimensional structure (see Figure 134) and might therefore originate from a reduced ability to undergo conformational changes. Dimerization of the NTDs requires the concerted movement of interfacial helices to form a self-complementary dimerization interface. This structural change involves the movement of residue W10 from the protein's hydrophobic core to a more solvent-exposed position, which can be monitored as a red-shift and quenching of tryptophan fluorescence (*12, 15–18*). Interestingly, while this is indeed observed for wt-NTD when changing the buffer from monomer (pH 7, 200 mM NaCl) to dimer conditions (pH 6, 60 mM NaCl), the tryptophan fluorescence red-shift and quenching is almost lost completely in L6-NTD, indicating that W10 remains buried under buffer conditions typically required for dimer formation (Figure 137). Comparison of the tryptophan fluorescence shifts and quenching under monomer and dimer buffer conditions between wt-NTD and the L1- to L6-NTD variants indicates that the W10 conformational change becomes increasingly restricted with increasing methionine-depletion (Figure 137 B).

In addition to Trp fluorescence, the conformational switching of W10 was further investigated via ^{19}F NMR spectroscopy using wt- and L6-NTD labelled with ^{19}F-Trp (Figure 137 C, D and E). Because W10 is the only tryptophan in the NTD (Figure 137 C), the ^{19}F NMR spectra of both wt- and L6-NTD display a single peak. The peak position and line width is indicative of the structure and dynamics of W10. For wt-NTD, monomer buffer conditions (high pH and ionic strength) result in a sharp peak at ~-127.6 ppm corresponding to W10 in

the buried conformation (Figure 137 D). A decrease in pH and ionic strength (approx. pH 6-7, and 100-180 mM ionic strength) results in line broadening, which indicates the onset of weak dimer association. In this dynamic monomer-dimer equilibrium, W10 interconverts between a buried and solvent-exposed position at intermediate rates. At pH < 6 and low ionic strength, the dimer affinity is in the low nanomolar range, and wt-NTD exists exclusively as a dimer with W10 locked in the solvent-exposed conformation. This produces a sharp peak at around -127.1 ppm). Under monomer buffer conditions, L6-NTD also displays a sharp peak in the ^{19}F NMR spectrum indicating W10 is in a buried position (Figure 137 E). In contrast to wt-NTD, a change in buffer conditions to pH 6 and 100 mM salt has only subtle effects on the peak line widths, and the W10 peak only shows a slight shift. The small chemical shift perturbations can be explained with the gradual formation of weakly associated dimers with affinities in the high micromolar range. Further decrease of the buffer pH and salt concentration (pH 5.5 and 60 mM ionic strength) leads to NMR line broadening and indicates an increased dimer affinity in the low micromolar range under these conditions. Considering that the protein was used at high micromolar concentrations for ^{19}F NMR experiments, a significant fraction of L6-NTD is expected to exist as a dimer at low pH and low ionic strength (SEC-MALS suggests an affinity of the L6-NTD dimer in the μM range at 60 mM ionic strength and pH 6, Figure 136 B). The severe line broadening in the weakly associated L6-NTD dimer compared to the sharp dimer peak of the Wt-NTD dimer indicates substantially altered conformational dynamics of W10 in L6-NTD.

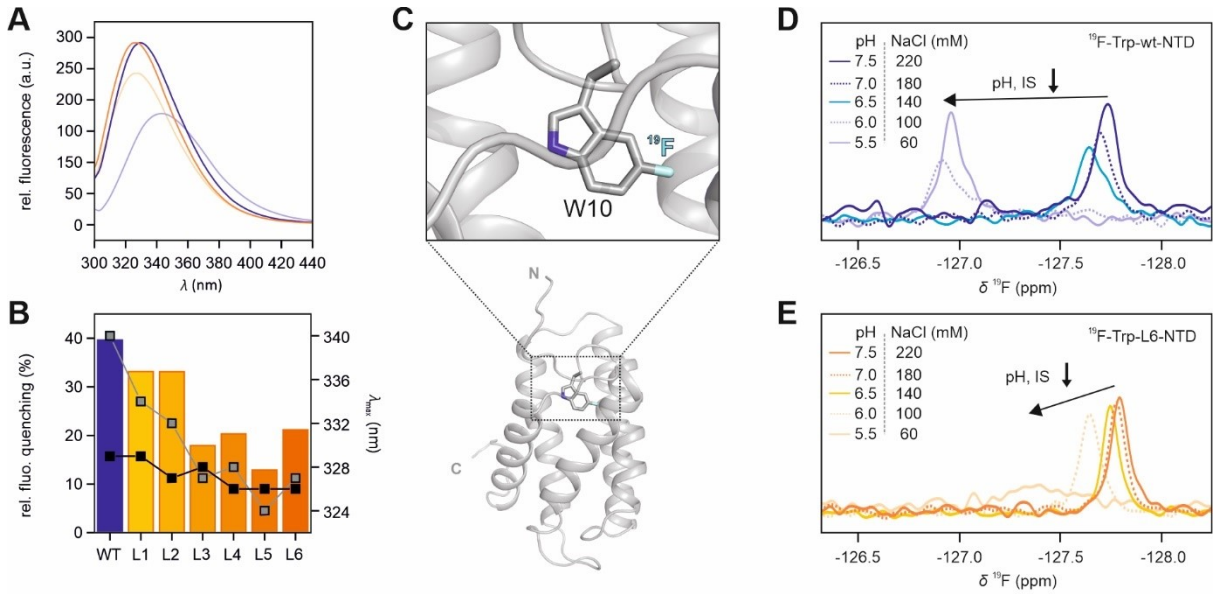


Figure 137: Tryptophan fluorescence and ^{19}F NMR spectroscopic analysis of W10 conformational changes in wt- and L6 NTD. (**A**) Tryptophan fluorescence spectra of wt- and L6-NTD under monomer (pH 7, 200 mM ionic strength) and dimer buffer conditions (pH 6, 8 mM ionic strength). (**B**) Relative tryptophan fluorescence quenching of cumulative methionine-to-leucine NTD mutants between monomer and dimer buffer conditions (coloured bars). Tryptophan fluorescence emission wavelength maximum λ_{max} of wt-NTD and NTD mutants under monomer (black squares) and dimer buffer conditions (grey squares). (**C**) Cartoon representation of NTD (PDB structure 2lpj) with ^{19}F-5-Trp labelled W10 in stick representation. ^{19}F NMR spectra of (**D**) ^{19}F-Trp-wt-NTD (blue, top) and (**E**) ^{19}F-Trp-L6-NTD (orange, bottom) at decreasing pH and ionic strength (IS). Dark blue/orange solid line: pH 7.5, 220 mM NaCl; dark blue/orange dashed line: pH 7.0, 180 mM NaCl, blue/yellow solid line: pH 6.5, 140 mM NaCl, light blue/light orange dashed line: pH 6.0, 100 mM NaCl, light blue solid line: pH 6, 60 mM NaCl. Subfigures (A) and (B) were modified from (*27*).

To investigate the effect of methionine-depletion on NTD conformational changes for each residue, uniformly ^{15}N-isotope-labelled wt- and L6-NTD constructs were generated and ^{1}H, ^{15}N-HSQC NMR spectra were recorded under monomer and dimer buffer conditions (Figure 138 A and B). In agreement with previous NMR studies (*17, 18*), the peaks in the HSQC spectrum of wt-NTD shift severely between monomer and dimer conditions (Figure 139 A). The L6-NTD spectra, in contrast, show comparably weak peak shifts between mon-

omer and dimer conditions (Figure 139 B). Instead, the L6-NTDspectrum under dimer buffer conditions displays substantial peak broadening. Importantly, the number of peaks observed under dimer conditions is comparable to the number of residues in the protein, which indicates that the symmetry observed in the wt-NTD dimer is maintained in L6-NTD. Using previously published NMR backbone assignments (BMRB codes 18262 and 18480, respectively) (*17, 18*), the peaks of wt-NTD could be assigned under monomer and dimer conditions (Figure 138). The NMR backbone assignments of L6-NTD under monomer conditions were obtained *de novo* and subsequently used to obtain the NMR structure of the L6-NTD under monomer conditions (Figure 132 and Figure 134). (*25, 27*) *De novo* NMR backbone assignments could also partially be obtained for L6-NTD under dimer conditions; however, severe peak broadening prevented a complete L6-NTD dimer NMR backbone assignment. Particularly residues in helices α2 and α3, the main constituents of the dimer interface, could not be assigned. The backbone amide resonances in L6-NTD could be assigned to 51%, and the overall H, H_N, H_α, C_α, C' backbone resonance assignments were completed to 37%%, including proline and cleavage tag residues. The different line shapes between wt- and L6-NTD dimer spectra are consistent with observations made in ^{19}F NMR experiments (Figure 137 D and E). The line broadening of dimer L6-NTD peaks indicates a monomer-dimer exchange on the intermediate timescale and reflects the reduced affinity of the mutant dimer compared to wt-NTD. The low nanomolar dimer affinity of wt-NTD results in a slow monomer-dimer exchange rate, which gives rise to sharp peaks.

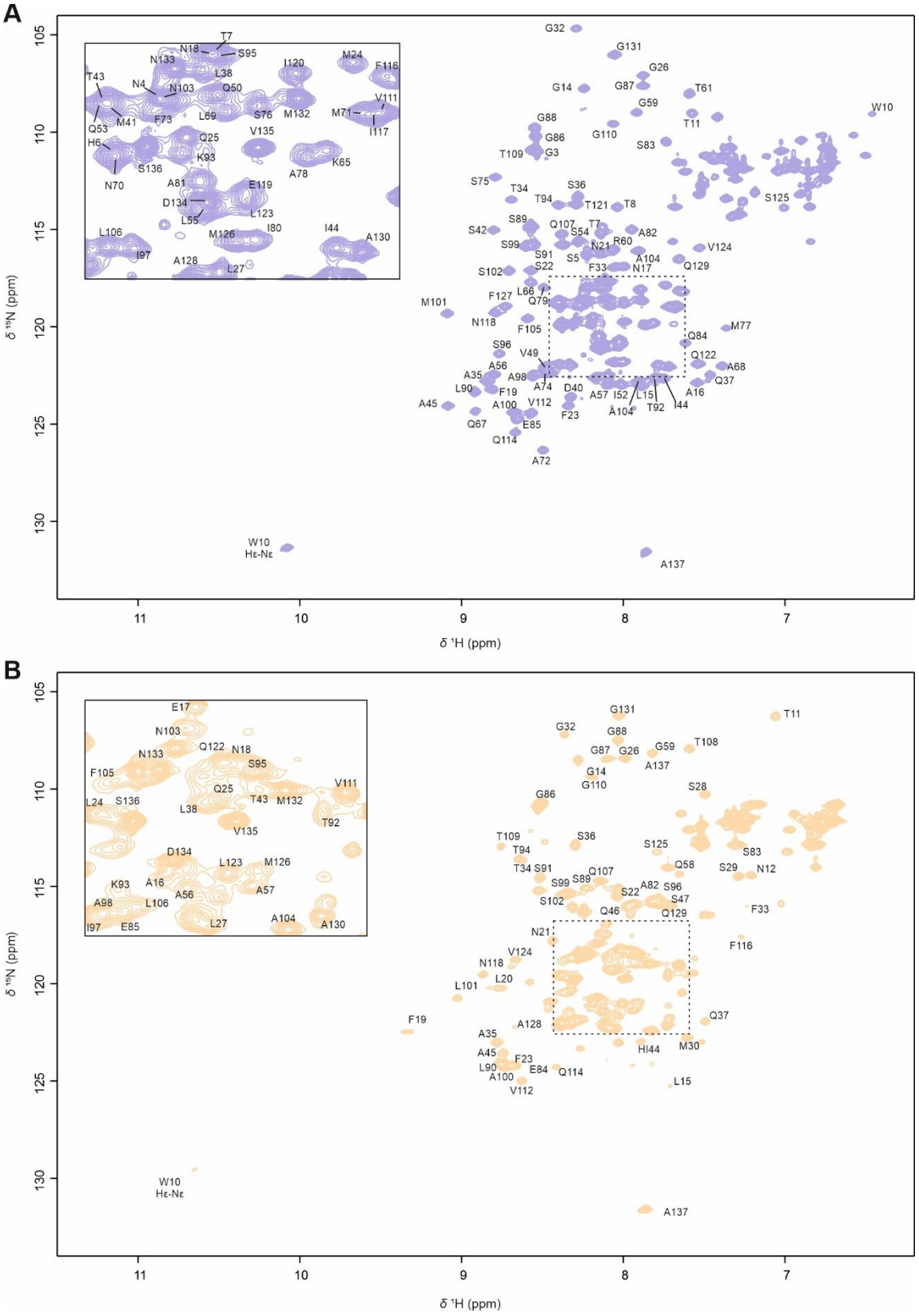


Figure 138: NMR backbone assignments of wt- and L6-NTD at pH 6. $^{1}H,^{15}N$-HSQC NMR spectrum of (**A**) ^{15}N-wt-NTD and (**B**) ^{15}N-L6-NTD under dimer conditions (300 µM protein concentration, pH 6, 8 mM ionic strength). L6-NTD assignments were obtained *de novo* (see section 60), whereas wt-NTD assignments were transferred from BMRB entry 18480.

Based on the NMR backbone assignments, the chemical shift differences of each peak between monomer and dimer conditions were determined for wt- and L6-NTD (Figure 139 C). To link chemical shift changes observed in the wt-NTD spectra to structural changes during NTD dimerization, the dimer contact sites and the conformational changes expressed as changes in the relative solvent accessibility and C_α-C_α distance difference, respectively, were determined from the monomer (2lpj) and dimer (3lr2) structures of wt-NTD (Figure 139 D). The effect from protonation was neglected, as a recent study showed only little effect of pH on the chemical shifts of a dimerization inhibited NTD A72R mutant (*18*). Strong chemical shift changes in wt-NTD predominantly cluster in helices α1-3 and the NTD C-terminus, regions strongly involved in conformational changes or NTD inter-subunit contacts during dimerization as assessed by comparing the wt-NTD monomer and dimer structures (Figure 139 C and D). As evident from the wt-NTD monomer and dimer structures (Figure 139 E), the W10 side chain undergoes a transition from a buried to a solvent-exposed position during dimerization. This is reflected by a strong shift of the W10ε peak in the ^{1}H, ^{15}N-HSQC spectrum (Figure 139 E) and is consistent with the strong ^{19}F-W10 peak shift of wt-NTD in ^{19}F NMR experiments (Figure 137 D). The L6-NTD, in contrast, shows very weak chemical shift changes in the ^{1}H, ^{15}N-HSQC spectrum between monomer and dimer conditions throughout the assigned regions (Figure 139 C), even in regions involved in dimer formation in the wt-NTD structures. Importantly, the W10ε side chain amine peak of L6-NTD is barely shifted but severely broadened between monomer and dimer conditions (Figure 139 A, B, and E). This is consistent with the ^{19}F NMR experiments (Figure 137 E) where the ^{19}F-W10 peak shift of L6-NTD was less shifted than observed in the wt-NTD but severely broadened. The comparison of the chemical shift changes from monomer to dimer conditions between wt-NTD and L6-NTD in the^{1}H, ^{15}N-HSQC and the ^{19}F NMR spectra are commensurate with impaired conformational changes in the L6-NTD, thus explaining its reduced ability to dimerize.

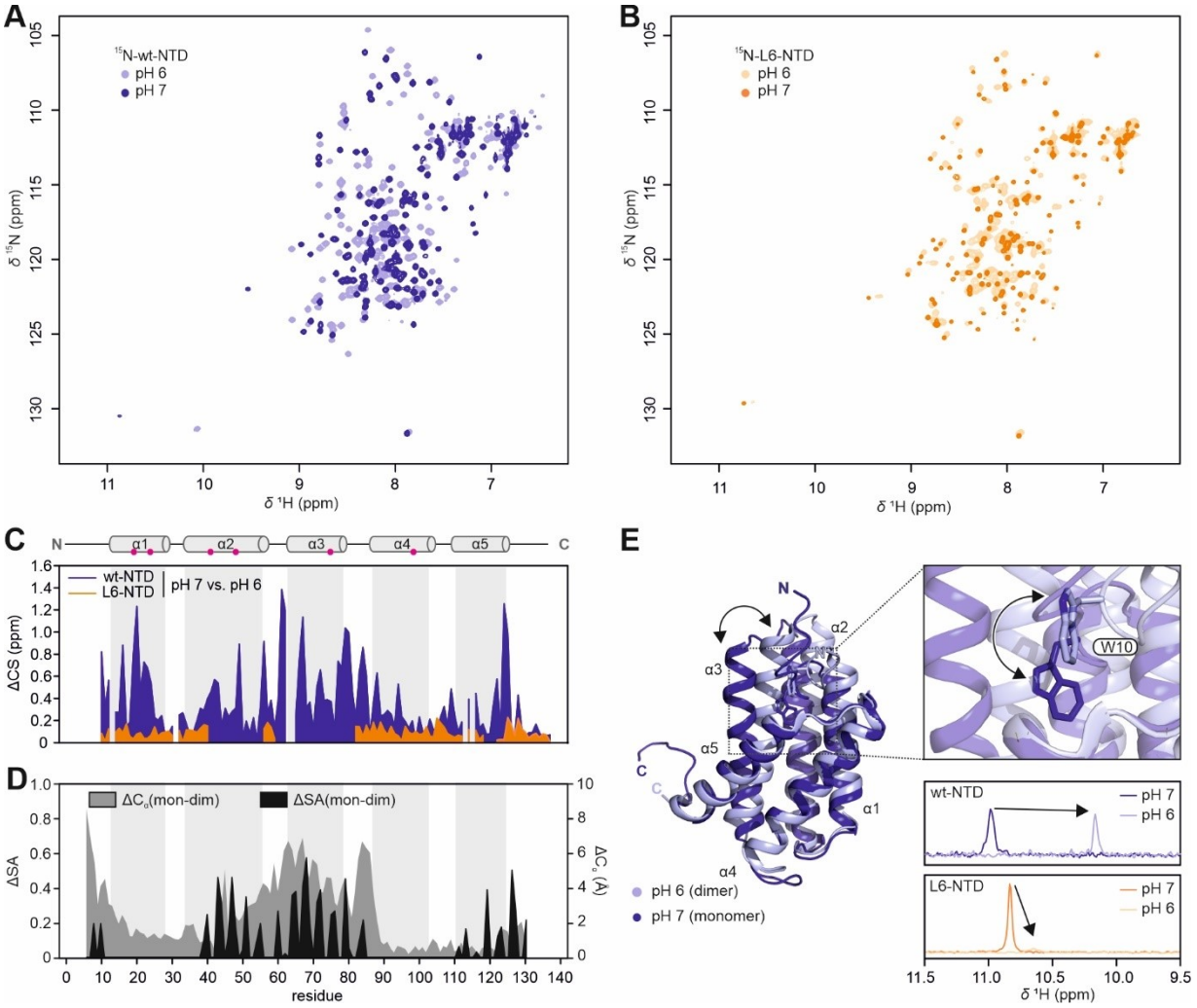


Figure 139: HSQC NMR spectroscopic analysis of wt- and L6-NTD conformational changes. Overlay of ^{1}H, ^{15}N-HSQC spectra of (**A**) ^{15}N-wt-NTD and (**B**) ^{15}N-L6-NTD under monomer (300 µM NTD, pH 7, 200 mM ionic strength) and dimer conditions (300 µM NTD, pH 6, 8 mM ionic strength). (**C**) Chemical shift changes upon dimerization observed for wt-NTD (blue) and L6-NTD (orange). A topology model of the NTD with positions of the methionine-leucine mutations highlighted as magenta dots is shown on top of the diagram. (**D**) The change in relative solvent accessibility ΔSA (black) between the NTD in the monomer and the dimer conformation (calculated via SurfRacer5.0 (*26*) from PDB structures 2lpj and 3lr2) highlights the residues forming the dimerization interface. The displacement of the NTD C_{α} positions between the monomer and the dimer conformation (grey), termed ΔC_{α}(mon-dim), indicates where conformational changes in the protein sequence occur upon dimerization (calculated from PDB structures 2lpj and 3lr2). (**E**) 1D-^{1}H projection of the W10ε amide peak of wt- and L6-NTD under monomer and dimer conditions.

4.4 Methionine to leucine mutations drastically change the NTD native-state dynamics

The structural changes of the NTD between monomer and dimer conformation involve the reorientation of three interfacial helices. Such collective motions typically occur on microsecond to millisecond timescales. However, conformational changes can also depend on local dynamics such as backbone or side chain fluctuations on the picosecond to nanosecond time scale (*41*). Changes in the dynamics of local and collective motions, e.g. due to mutations, can therefore strongly affect protein functionality. The perturbed ability of L6-NTD to undergo conformational changes required for dimerization thus might originate from drastically altered native-state dynamics.

Based on the NMR backbone assignments of wt- and L6-NTD, hydrogen/deuterium-exchange (HDX) and heteronuclear Overhauser effect (hetNOE) NMR experiments were carried out to probe the influence of methionine-leucine mutations in the NTD on the protein's collective and local backbone dynamics under monomer conditions on the per-residue level (Figure 140) (*42, 43*). HDX measurements were limited to a temporal resolution of ~12 min, which corresponds to the acquisition time of an HSQC spectrum. Within the first ~12 min, most amide protons in the wt-NTD were deuterated completely (Figure 140 A, and C). This was particularly pronounced for residues in the solvent-exposed helix-connecting loops or for residues in the flexible N- and C-termini. Deuteration was substantially slower for residues in the centre of the five-helix bundle, as also previously observed via HDX-MS for the dimerization inhibited NTD A72R mutant (*18*). Strikingly, deuteration in the five-helix bundle of the L6-NTD was ~10-fold slower compared to the wt-NTD (Figure 140 A, B, C). This affected not just the positions where the methionine-leucine mutations were introduced, but also residues 18-28, 41-51, 72-81, 98-107, and 108-128 (Figure 140 A, and C). These HDX results indicate drastically altered collective motions in the L6-NTD compared to the wt-protein. In contrast, the local backbone dynamics observed via hetNOE measurements were almost identical for wt- and L6-NTD (Figure 140 D and E). Differences, however, were observed at the N-terminal end of helix α3 (residues 67-70). Interestingly, this region undergoes a drastic conformational change during dimerization, which might be facilitated by fast local backbone dynamics.

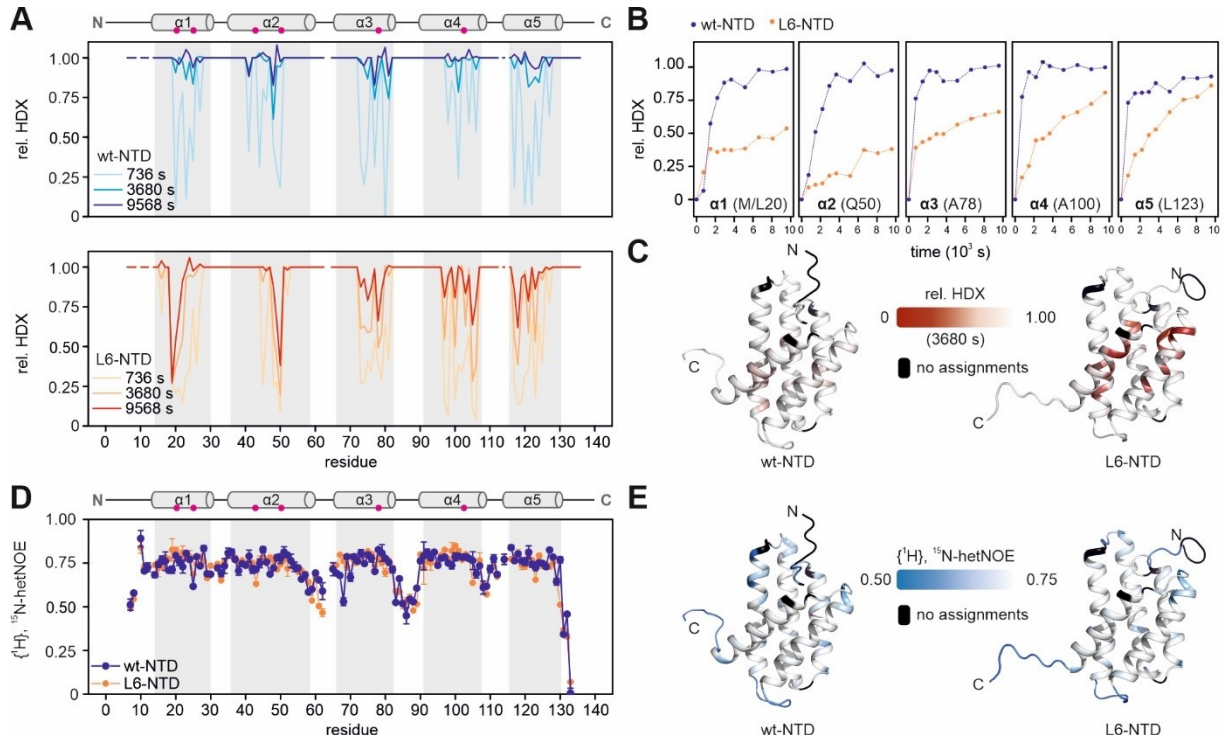


Figure 140: Collective and local monomer NTD dynamics. (**A**) Hydrogen/deuterium-exchange (HDX) NMR analysis of wt-NTD (top) and L6-NTD (bottom). Plotted is the relative Hydrogen/deuterium-exchange (rel. HDX) based on peak integral for each residue in a 1H, 15N-HSQC spectrum after 736, 3680 or 9568 s against the residue number. The NTD topology with the positions of the methionine-leucine mutations indicated as magenta dots is shown on the top of the diagram. (**B**) Comparison of time-dependent rel. HDX of representative residues from helices α1-6 between wt-NTD (blue) and L6-NTD (orange). (**C**) Rel. HDX after 3680 s mapped onto the structures of wt-NTD (PDB: 2lpj) and L6-NTD (PDB: 6qjy). (**D**) {^{1}H}, ^{15}N-hetNOE values of wt-NTD (blue) and L6-NTD (orange) plotted versus the residue number. (**E**) {^{1}H}, ^{15}N-hetNOE plotted onto the structures of wt-NTD (PDB: 2lpj) and L6-NTD (PDB: 6qjy).

In a complementary approach, the dynamics of wt- and L6-NTD were investigated site-specifically using photoinduced electron transfer (PET) fluorescence correlation spectroscopy (FCS) by the Neuweiler group, Uni-

versität Würzburg. Here, the NTD G3C mutant was fluorescently labeled with AttoOxa11 via a cysteine-maleimide coupling (Figure 141 A). The labeling scheme had previously been used to study the influence of charges and protonation on the dynamics of the NTD (*19*). It is ideally suited to detect movements of residue W10 from the buried to the solvent exposed position, as this movement alters the distance between W10 and the Atto-dye and results in fluctuations of PET-induced fluorescence quenching. During the FCS measurement, the signal fluctuations caused by PET give rise to an additional decay in the autocorrelation trace, whose amplitude and time constant informs on both structural and kinetic aspects of the relative motions between Atto-dye and W10.

PET-FCS measurements were carried out under monomer conditions at pH 7 and high ionic strength. The obtained auto-fluorescence functions showed three decays in the ms and sub-ms time scales (Figure 141 B). The slow, millisecond decay accounts for fluorescence fluctuations caused by diffusion of NTDs in and out of the detection volume. The two sub-millisecond decays account for intramolecular PET-induced fluorescence fluctuations caused by W10 conformational switching and are dominated by a single-exponential kinetic phase with a relaxation time constant τ_1 and amplitude a_1 with an underlying second kinetic phase with τ_2 and a_2. The dominating phase displays values of $\tau_1 = 250 \pm 19$ µs and $a_1 = 0.64 \pm 0.03$ for wt-NTD and $\tau_1 = 529 \pm 19$ µs and $a_1 = 0.35 \pm 0.01$ for L6-NTD already indicating a decrease in dynamics and dampening of the amplitude of motions for L6-NTD. As previously described, this phase was assigned to the transitioning of W10 between buried and solvent exposed positions (*19*). The second phase has values of $\tau_2 = 2 \pm 1$ µs and $a_2 = 0.19 \pm 0.04$ for wt-NTD and $\tau_1 = 4 \pm 1$ µs and $a_1 = 0.05 \pm 0.01$ for L6-NTD. It was, however, not possible to assign this phase to a particular motion of W10. The decrease in the main decay time constant and amplitude indicate that the motion of W10 is substantially restricted in L6-NTD compared to wt-NTD. The τ_1 and a_1 values were used to estimate the time constants of the W10 conformational change, which yielded $\tau_{out}^{wt} = 0.61 \pm 0.06$ ms and $\tau_{out}^{L6} = 2.1 \pm 0.1$ ms (Figure 141 C), indicating that the movement of W10 from the buried to the solvent exposed conformation is much slower in L6-NTD compared to wt-NTD. It thus follows, that the energy required for native-state conformational changes ($\Delta G_{conf,nat}$) increases from 0.26 ± 0.09 kcal mol^{-1} for wt-NTD to 0.655 ± 0.16 kcal mol^{-1} for L6-NTD (Table 14).

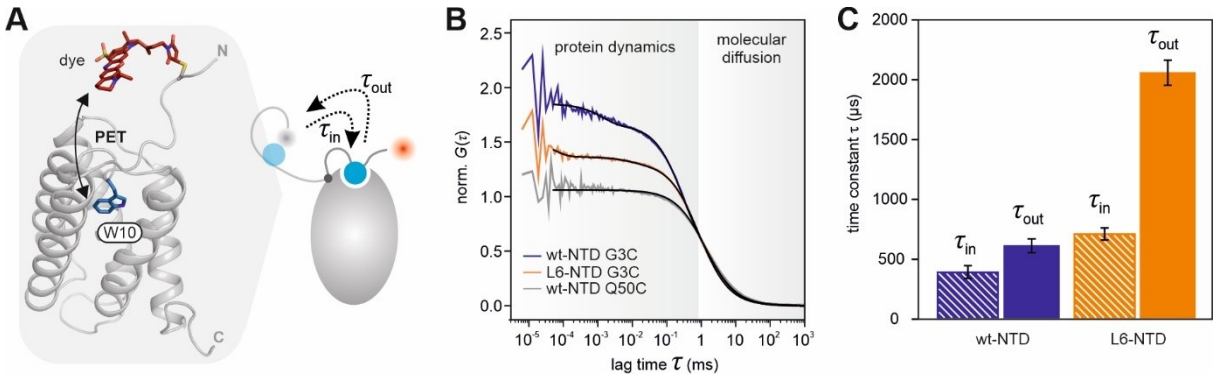


Figure 141: PET-FCS analysis of NTD conformational dynamics under monomer conditions. (**A**) PET-FCS reporter design. Left: Cartoon representation of NTD G3C mutant with attached AttoOxa11 maleimide dye (modeled with PyMOL on PDB structure 2lpj). The wedged W10 residue (blue) and the Atto-dye (red) are shown as stick representations in blue and red, respectively. Right: Conformational changes of W10 detected via PET-FCS. W10 moving out of the buried (τ_{out}) and into the core (τ_{in}) switches PET-induced fluorescence quenching off and on. (**B**) PET-FCS autocorelation functions $G(\tau)$ recorded from wt-NTD (blue) and L6-NTD (orange) G3C mutants labeled with AttoOxa11. An L6-NTD labeled with AttoOxa11 via mutation Q50C was used as a control (grey). The Q50C mutations carries the AttoOxa11 label in a distant position from W10 judging from on the available NMR and X-ray crystal structures. The Q50C mutant autocorelation function should thus not decay due to PET but only due to molecular diffusion. (**C**) Time constants of W10 moving in (τ_{in}) and out (τ_{out}) of the buried position determined from PET-FCS data. Figure is modified from (*27*).

Table 14: Kinetic parameter of W10 conformational switching derived via PET-FCS. Errors are s.e. from regression analysis and propagated s.e.

Kinetic parameter	**wt-NTD**	**L6-NTD**
τ_1 (µs)	240 ± 17	529 ± 19
k_1 (ms^{-1})	4.16 ± 0.30	1.89 ± 0.07
a_1	0.64 ± 0.03	0.35 ± 0.01
τ_2 (µs)	2 ± 1	4 ± 1
a_2	0.19 ± 0.04	0.05 ± 0.01
τ_{in} (ms)	0.393 ± 0.022	0.704 ± 0.001
τ_{out} (ms)	0.61 ± 0.06	2.1 ± 0.07
k_{in} (ms^{-1})	2.53 ±0.14	1.42 ± 0.15
k_{out} (ms^{-1})	1.63 ± 0.16	0.47 ± 0.08
$\Delta G_{conf,nat}$ (kcal mol^{-1}) *	0.26 ± 0.09	0.66 ± 0.16

* determined from k_{in} and k_{out} via $\Delta G_{conf,nat} = -RT\ln(k_{in}/k_{out})$

5 Conclusion

Web weaving spiders have evolved a highly efficient silk formation machinery in which soluble spidroin proteins are assembled into solid fibres at very fast rates. To ensure high silk quality, the spidroins need to orient and align properly before strong inter-molecular interactions tightly assemble them (*4*). The multi-step mechanism underlying the dimerization of spidroin NTDs along a pH and ionic strength gradient in the spinning duct is perfectly designed to control the spatial and temporal tightening of spidroin interactions and thereby prevent premature silk formation (*15, 17, 44*). Sequential protonation in the NTD and a reduction in ionic strength triggers early NTD association via electrostatic attraction. A steric mismatch in the dimer interface leaves the early NTD dimer loosely assembled. Late structural rearrangements in the NTD subunits lead to self-complementary dimer interfaces, which in combination with further protonation enable tight dimer interactions. The exact chronology of weak NTD association, protonation, and conformational changes and whether or how the steps are coupled are still under debate. The unexpected observation that core methionine-depletion leads to impaired NTD dimerization reveals novel insights into the role of conformational changes in the NTD dimer assembly process. Moreover, it highlights a yet unknown function of methionine in a protein hydrophobic core.

5.1 Hydrophobic core methionine residues render the NTD malleable

The defined three-dimensional structures of proteins arise from the tight packing of hydrophobic side chains into a protein core where they adopt low energy conformations and form excessive vdW-contacts (*45*). The shape of the hydrophobic core and the extent of intra-core vdW-contacts has direct implications on the structure, dynamics, and stability of a protein (*31, 32, 41, 46*). Consequently, core mutations which alter the sterics and vdW-interaction network are prone to disturb the structural and functional integrity of a protein. Structural dynamics play a crucial role in protein function. The spider silk NTD is no exception to this rule (*15, 19*). It needs to undergo precise conformational changes to enable tight interaction between two subunits (*18*). The composition of the hydrophobic core in MaSp NTDs needs to be correctly adjusted to enable conformational changes at fast rates (*15, 16*). Sequence analysis reveals that the amino acid methionine is substantially overrepresented in the hydrophobic core of MaSp NTDs. Despite its similar hydrophobicity and size compared to other aliphatic amino acid side chains such as Leu or Ile (*35*), Met is on average present much less frequent in the core of proteins (Figure 130 A) (*22*). The spider silk NTD, therefore, presented a unique opportunity to study the role of Met for the structural, dynamic and functional properties of an important protein. All six core methionine residues in NTD can be replaced by leucine and yield a structurally intact protein with a substantially increased stability. However, the methionine-depletion not only drastically impairs NTD native-state dynamics but also leads to an impairment of its ability to undergo the conformational changes required for dimerization.

This is reflected by three orders of magnitude reduction in dimer affinity. This highlights methionine in the hydrophobic core as a key ingredient to providing high plasticity required for tight dimerization of NTDs.

The effect of hydrophobic core methionine residues on the structural dynamics of the NTD might originate from the unique side chain properties of methionine (*47*). An early study on the conformational states of amino acids in proteins pointed out, that besides being the only non-branched hydrophobic chain, methionine exhibits extreme side chain flexibility (*48*). This flexibility predominantly manifests as almost unhindered rotation around the $C_\beta C_\gamma$-$S_\delta C_\varepsilon$ dihedral angle, termed χ_3 (Figure 142 A). Within two groups of analyzed protein structures in that study, the methionine X_3 angles were almost equally distributed over the entire possible range of conformations (*48*). Only the eclipsed conformation with aligned C_ε and C_β was not observed. This suggests very similar energies of the staggered conformations with low energy barriers between them, which allows the ε-methyl group to rotate almost freely (Figure 142 B). In contrast, the rotational freedom along a C-C bond (as in the all-carbon methionine analogue norleucine) is much more restricted (Figure 142 C, D). In Norleucine, the hydrogen atoms on the C_γ atom lead to higher free energies for the eclipsed conformations compared to methionine, which contains two free electron pairs on the S_γ atom (*47*). Therefore, the higher energy barriers imposed by the staggered conformations of norleucine restrict the X_3 rotation. Side chain rotations are even more restricted in the branched side chains of leucine or isoleucine.

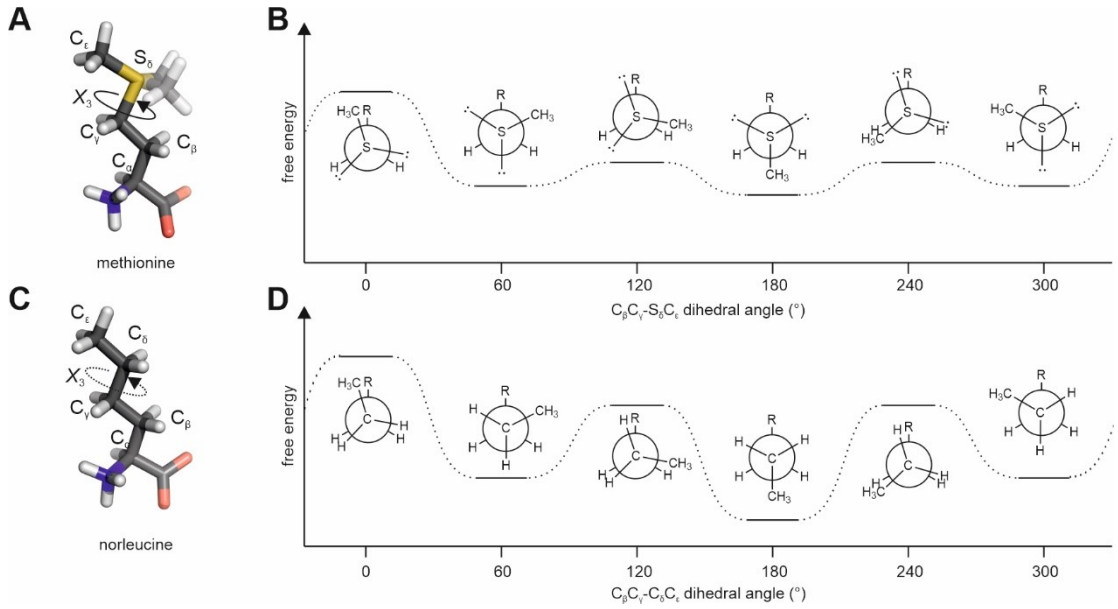


Figure 142: Rotational flexibility of the methionine side chain. (**A**) Stick representation of methionine with three staggered X_3 conformers. (**B**) Newman projection along methionine X_3 angle. Plotted is the free energy versus the $C_\beta C_\gamma$-$S_\delta C_\varepsilon$ dihedral angle. (**C**) Stick representation of norleucine, the carbon-only analogue of methionine. The restricted side chain flexibility around X_3 is indicated with a dotted arrow. (**D**) Newman projection along norleucine X_3 angle. Plotted is the free energy versus the $C_\beta C_\gamma$-$C_\delta C_\varepsilon$ dihedral angle (*47*).

The rotational flexibility of the ε-methyl group enables methionine to adapt its side chain conformation to various structural environments. In this sense, methionine has been implicated in rendering protein surfaces ductile thereby enabling them to rapidly reshape and fit the binding site of an interaction partner (*49–51*). This principle can be transferred to the hydrophobic core of the NTD, where the side chain flexibility of methionine renders the core malleable. The ability for rapid side chain reorientation enables methionine to compensate disturbances in the tight packing of the hydrophobic core upon changes in a protein's shape through the motion of secondary and tertiary structure. This is evident from NMR hydrogen/deuterium-exchange and PET-FCS experiments. Replacement of methionine by leucine in the hydrophobic core substantially reduced the collective motions of the NTD and the conformational dynamics of its conserved tryptophan residue (Figure 140 and Figure 141) which results in an impaired dimerization. Methionine side chains in the NTD hydrophobic core

are therefore key components to provide high overall structural flexibility required for conformational changes in the dimer interface.

5.2 Methionine-depletion in the NTD reveals novel insights into the dimerization mechanism

The currently most prominent mechanism proposed for the dimerization of Ea-MaSp1 NTD is divided into three steps. Early NTD association via electrostatic attraction is mediated by residues D40, K65, R60 and potentially E84 in their charged states. This causes an elevation of the typical pK_a values for acidic side chains of glutamates from ~4 into a range of pH 6.5 for NTD residues E79 and E119 due to local environmental changes between the loosely associated monomer subunits. Thus, a drop in pH from 7 to below 6.5 (e.g. within the spider spinning duct) results in the protonation of the sidechains of E79 and E119 and leads to structural rearrangements of the NTD subunits and subsequent tight dimer association. Further pH decreases to below 5.7 in the distal spinning duct additionally enables the protonation of the sidechain of residue E84, which further stabilizes the complex (*17*). In a classical 'induced fit' model, the dimerization of NTDs triggers the conformational change when going from pH 7 to pH 6.5. However, a recent study suggests that this step rather occurs via a 'conformational selection' mechanism (*52*). According to this mechanism, the NTD exists in a dynamic equilibrium between various conformational states. From this conformational ensemble, the binding partner selects the dimerization-competent state, i.e. the conformation with a self-complementary association interface (*19*). The high abundance of methionine in the NTD hydrophobic core likely provides access to transient dimer-competent states through increased conformational dynamics. The NTD dimerization process is very complex and requires multiple steps. These may be dependent on an early conformational selection step enabled by a dynamic ensemble of the met-containing NTD, while the later tightening of the dimer is obtained via an induced fit mechanism.

Both conformational changes and pH-dependent electrostatic interactions are crucial components of NTD dimerization and seem to be coupled to some extent (*15, 17–19*). With the conformationally restricted L6-NTD, a new tool to study the contribution of both events independently, was provided. Methionine-depletion in the NTD through replacement with leucine blocks conformational changes required for tight dimerization but preserves the electrostatics of the domain. The L6-NTD can therefore be used to abort the multi-step dimerization mechanism before structural rearrangements consolidate the dimer and stabilize the intermediate state in which the NTD dimer is held together predominantly by electrostatic forces. Assuming that the overall dimerization process is only driven by electrostatic attraction and conformational changes, the free energy contribution of both processes can be estimated for the wt-NTD by taking the thermodynamic data obtained for L6-NTD into account (Figure 143) (*31*). The free energy of dimerization (ΔG_{dim}) can be calculated from the equilibrium dimer dissociation constants of the wt-NTD via $\Delta G = -RT\ln(K_d)$. The K_d of 1.1 ± 0.1 nM (*15*) yields a free energy of $\Delta G_{dim} = -12.3 \pm 0.1$ kcal/mol. The energetic contribution from electrostatics (ΔG_{ele}) can be obtained from the equilibrium dissociation constant of the conformationally inhibited L6-NTD. The K_d of 1.1 ± 0.2 μM yields a free energy of $\Delta G_{ele} = -8.2 \pm 0.1$ kcal/mol. The free energy contribution from conformational changes is derived via simple subtraction as $\Delta G_{conf} = \Delta G_{dim} - \Delta G_{ele} = -4.1 \pm 0.1$ kcal/mol. This reveals a substantial contribution of the conformational changes to the free energy of the dimer formation process. Nevertheless, electrostatic attraction is the dominant driving force.

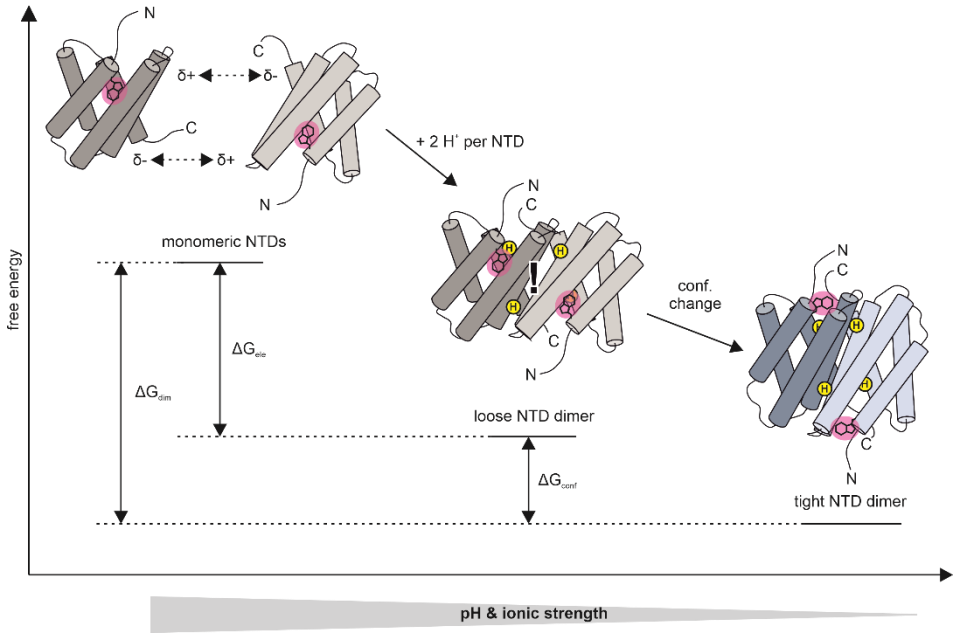


Figure 143: Contribution of electrostatic interactions and conformational changes to tight NTD dimer assembly. Free energy diagram showing the first two steps along the dimerization pathway in a range of pH 7 to pH 6. The first step involves protonation and electrostatic attraction of NTD subunits resulting in an early dimer complex. The early dimer is only weakly associated due to a steric mismatch in the dimer interface (indicated by an exclamation mark). W10 remains buried in the early dimer. Conformational changes in the second step reshape the dimer interface and enable tight subunit interactions. For this, W10 needs to transitions to a more solvent-exposed position. ΔG_{dim}: free energy of dimerization, ΔG_{elc}: free energy contribution of electrostatic attraction, ΔG_{conf}: free energy contribution of conformational changes.

The energetic contributions calculated above apply to the dimerization events occurring in a pH range of 7 to 6. The proposed three-step mechanism, however, includes a final protonation event at ~pH 5.7, probably affecting the E84 side chain, which might be coupled with additional conformational changes to further tighten the dimer (*17*). In ^{19}F NMR experiments, the decrease from pH 6.0 to pH 5.5 led to a substantially sharper wt-NTD W10 peak, which indicates a further tightening of the dimer coupled to a conformational change of W10 (Figure 137 D). This agrees very well with the proposed three-step mechanism. A decrease from pH 6.0 to pH 5.5 also indicated strengthening of the L6-NTD dimer and altered exchange dynamics of W10 based on ^{19}F NMR measurements (Figure 137 E). Assuming that the L6-NTD is conformationally restricted but that titratable residues remain similarly accessible for protonation as in wt-NTD would suggest that the final tightening of the NTD dimer below pH 5.7 is dominated by changes in electrostatics rather than structural rearrangements. The conformational changes, therefore, seem to play a major role for dimer assembly in the main part of the spinning duct between pH 7 and pH 6. The fact that L6-NTD can be protonated between pH 6 and pH 5.5 further indicates that the structural rearrangements required for dimerization are not a prerequisite for the protonation below pH 5.7.

5.3 Hydrophobic core methionine residues are promising targets for protein engineering

Protein stability usually refers to a protein's resistance against unfolding, meaning the loss of secondary and tertiary structure. In their folded states, proteins form distinct structures through precise interactions between amino acids. Those interactions are lost upon unfolding, which results in ensembles of undefined structures in the denatured state. The free energy between the native state and the ensemble of denatured states is the determinant of protein stability. Protein stability can be increased either through destabilizing the denatured state, by stabilizing the folded state or through a combination of both. Single domain proteins usually have a free energy of unfolding (ΔG_{unfold}) in the range of +10 kcal/mol, which means that unfolding events occur rarely but reversibly (*31*). The rate at which unfolding and folding occur depends on the free energy of the native and denatured states with respect to the transition state. The free energies of the individual states themselves are determined by their structures.

Exchanging hydrophobic core methionine residues against leucine in the MaSp1 NTD lead to a drastic increase in stability. The kinetic measurements reveal that this is a result of a decelerated unfolding rate paired with an accelerated folding rate close to the theoretical speed limit. The changed folding rate must originate from changes in the free energies between the denatured and the transition state, which in turn means that their structures are also changed in L6-NTD compared to wt-NTD. The thermodynamic and kinetic data reveal that the free energy of the denatured L6-NTD state is increased and thus close to the free energy of the transition state (Figure 144). This might be explained by the large polarizability of the sulfur atom which makes the methionine side chain more "sticky" at vdW-contact sites than carbon-only side chains and potentially enables residual stabilizing interactions in the denatured state of wt-NTD (*49*). Mutation of the methionine residues to leucine disturbs the stabilizing interactions in the denatured state and thus increase its free energy. Moreover, the reduced free energy of the wt-NTD denatured state with respect to the L6-NTD denatured state might have an entropic origin. The denatured state of a protein is characterized by an ensemble of flexible micro-states, of which each contributes to the free energy through an entropic term $-T\Delta S$, where ΔS is the entropy difference between native and denatured state. The entropic increase from generating a large number of micro-states is a major driving force for unfolding (*31*). The methionine side chain exhibits a higher side chain flexibility than leucine and therefore generates more micro-states in the denatured state of wt-NTD compared to L6-NTD. Substitution of methionine through leucine reduces this entropic stabilization and consequently increases the denatured state free energy. The destabilization of the NTD denatured state free energy through methionine-leucine exchange might thus result from the combined loss of residual vdW-interactions and a reduction in the number of microstates. This almost diminishes the barrier for folding in L6-NTD and accelerates it significantly (in this case coming close to the speed limit of folding). In contrast, exchanging methionine to leucine stabilizes the native state of the NTD. This might be due to the branched side chain of leucine, which has the potential to form a tighter vdW-network in the hydrophobic core of the NTD, similar to a jigsaw puzzle. The linear methionine side chain might be less suited for intertwined hydrophobic core networks. The reduction of the native state free energy increases the energy barrier for unfolding and consequently decelerates the rate of unfolding. The increased stability of NTD obtained via the introduction of hydrophobic core methionine to leucine mutations is thus due to a stabilized native state and a destabilized denatured state (Figure 144).

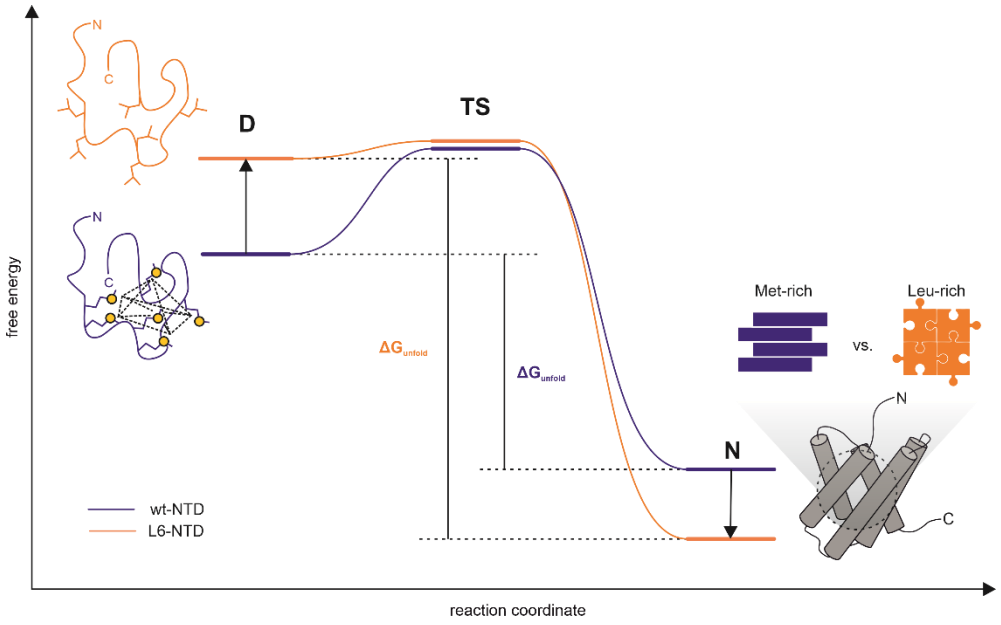


Figure 144: Increased stability of MaSp1 NTD through methionine-to-leucine exchange in the protein hydrophobic core. The increased stability of L6-NTD compared to wt-NTD is a combined effect from an accelerated folding and a decelerated unfolding rate. This, in turn, results in an energetically increased denatured state (D) and an energetically lowered native state (N). The assumption underlying the schematic energy diagram is that the transition state (TS) free energy between wt- and L6-NTD remain unchanged. The comparative stabilization of the denatured state for the wt-NTD compared to the L6-NTD is due to the high abundance of methionine. Core methionine residues might form stronger residual vdW-contacts (dotted lines) due to the large polarizability of the methionine sulfur atom compared to leucine. Residual vdW-contacts are weakened/abolished through replacement with leucine. In the native state, the branched leucine side chain might enable a more extensive vdW-network, as in a jigsaw puzzle. This stabilizes the native state. The vdW-network formed by linear methionine side chains is less intertwined.

Amino acid alignments from major (MaSp) and minor ampullate (MiSp), aciniform (AcSp), tubuliform (TuSp), and cylindrical spidroins (CySp) indicate that methionine and leucine and isoleucine occur almost interchangeably in hydrophobic core positions of NTDs (Figure 145) (*24*). The sum of methionine, leucine, and isoleucine in spidroin NTDs is conserved across gland types and spider species. However, methionine is overrepresented in MaSp, MiSp and AcSp NTDs. All three spidroin forms are used to generate extremely tough silk fibres (*53*). MaSp silk is used as a structural fibre for web building and as the spider's lifeline. MiSp fibres serve as auxiliary spirals in the spider's web. AcSp silks are used to wrap the prey and to insulate the spider egg cases. TuSp/CySp, in contrast, form flocculent silk fibres which are used for breeding (*53*). Their NTDs are completely devoid of methionine. This strongly suggests that the tough mechanical properties of MaSp, MiSp and AcSp silk fibres are achieved through an increased methionine content in the spidroin NTDs, which enable stronger intermolecular spidroin interactions through tight NTD dimerization. Phylogenetic studies show that Met is among the most frequently gained amino acid in protein evolution (*22*). DNA codons of methionine, leucine and isoleucine differ only by a single nucleobase. The probability for evolutionary transformation of isoleucine or leucine into methionine is therefore extremely likely. It is tempting to speculate that methionine-rich NTDs required for strong fibre formation has evolved from leucine-rich ancestors. Protein evolution often runs in the direction of decreased stability in favour of improved function. Ancestors are, therefore, usually more stable than their descendants (*54*). As such, the hydrophobic core of L6-NTD might represent an ancestor-like form of the methionine-rich core.

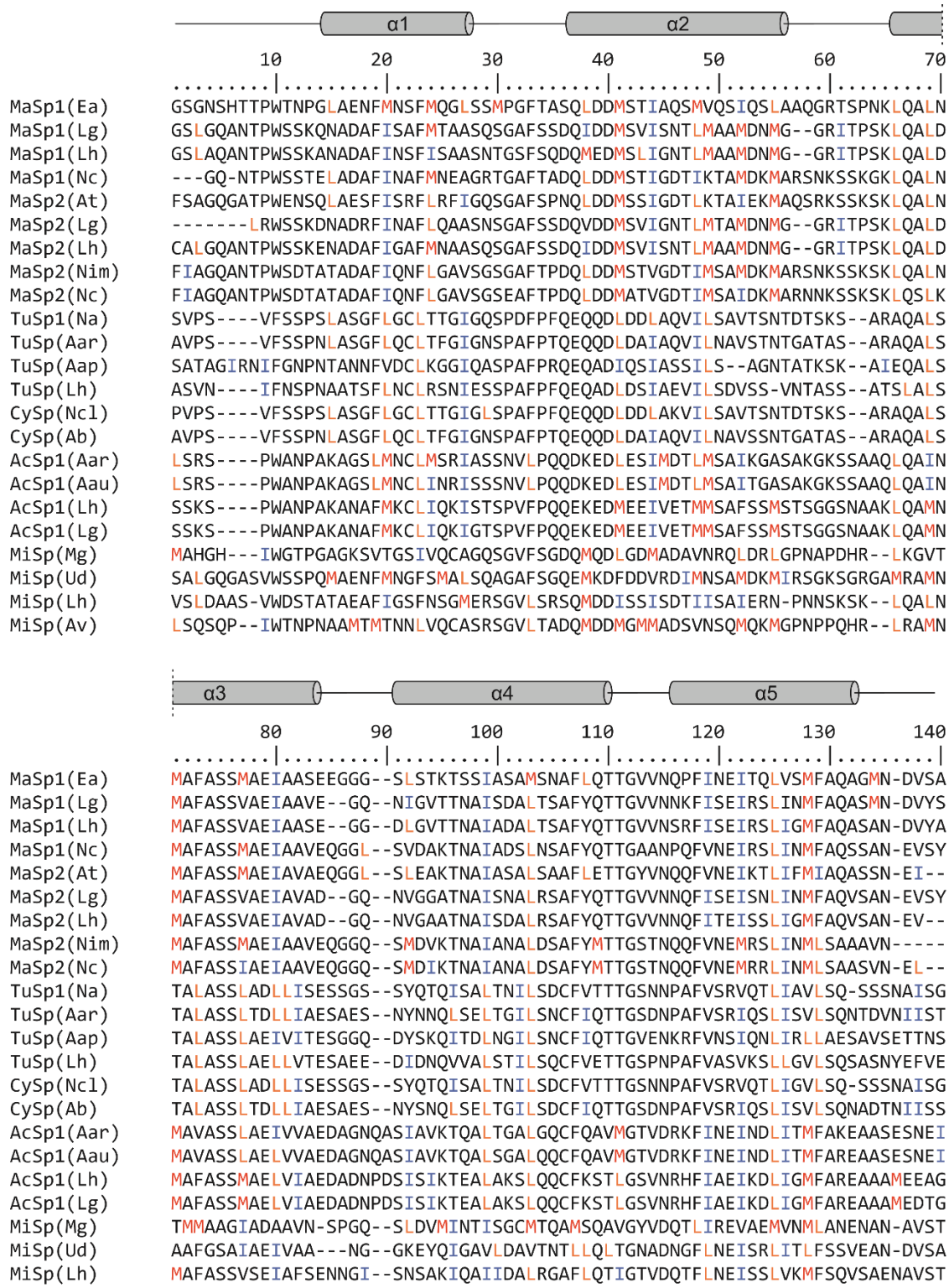


Figure 145*:* Amino acid alignment of spidroin NTDs from diverse species and glands. Ea: *Euprosthenops australis*, Lg: *Latrodectus geometricus*, Lh *Latrodectus hesperus*, Nc: *Nephila clavipes*, At: *Argiope trifasciata*, Nim: *Nephila inaurata madagascariensis*, Na: *Nephila antipodiana*, Aar: *Argiope argentata*, Aap: *Agelenopsis aperta*, Ncl: *Nephila clavata*, Ab: *Argiope bruennichi*, Aau: *Argiope aurantia*, Mg: *Metepeira grandiosa*, Ud: *Uloborus diversus*, Av: *Araneus ventricosus*; CySp/TuSp: cylindrical silk/tubuliform spidroin, MiSp: minor ampullate spidroin, MaSp: major ampullate spidroin, AcSp: aciniform spidroin. The five-helix topology is depicted on top of the sequence. Hydrophobic core methionine residues in the NTD sequence were identified from available NTD structures (see main text) and colored in magenta. Depending on the species and spidroin type, core methionine residues (red) are replaced by the isosteric residues leucine (orange) or isoleucine (blue). The figure is modified from *(27)*

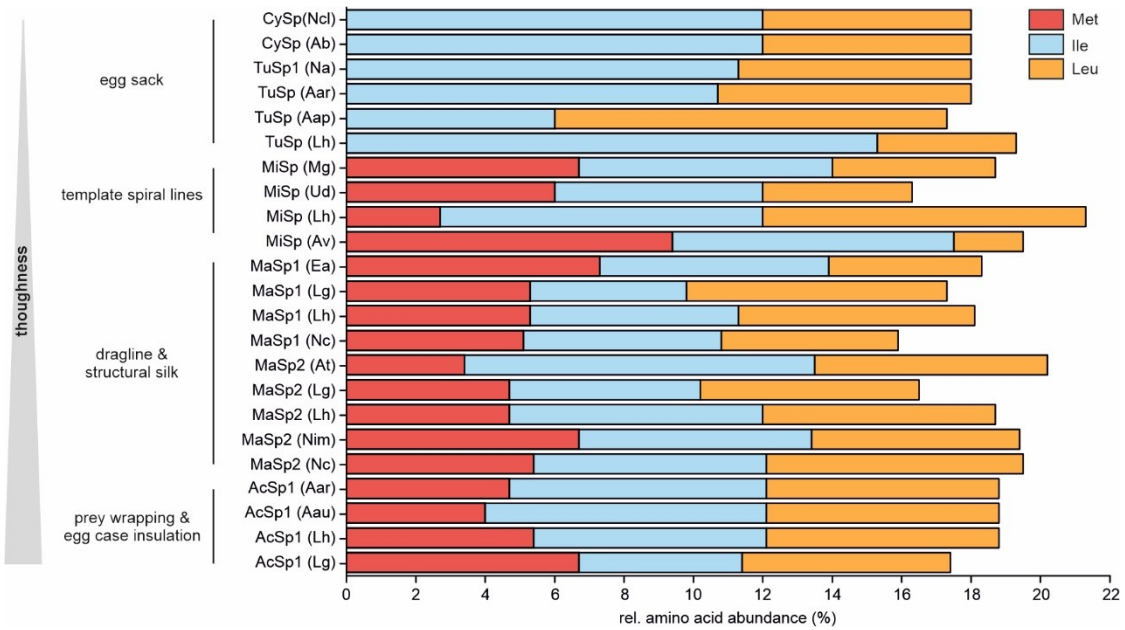


Figure 146: Correlation of hydrophobic core composition in spidroin NTDs with the toughness and function of silks from different glands. Relative abundance of methionine (red), isoleucine (blue) and leucin (orange) in spidroin NTDs from various species and glands. Ea: *Euprosthenops australis,* Lg: *Latrodectus geometricus, Lh Latrodectus hesperus,* Nc: *Nephila clavipes,* At: *Argiope trifasciata,* Nim: *Nephila inaurata madagascariensis,* Na: *Nephila antipodiana,* Aar: *Argiope argentata,* Aap: *Agelenopsis aperta,* Ncl: *Nephila clavata,* Ab: *Argiope bruennichi,* Aau: *Argiope aurantia,* Mg: *Metepeira grandiosa,* Ud: *Uloborus diversus,* Av: *Araneus ventricosus;* CySp/TuSp: cylindrical silk/tubuliform spidroin, MiSp: minor ampullate spidroin, MaSp: major ampullate spidroin, AcSp: aciniform spidroin. The NTD methionine content is correlated to the toughness of the silk fibres (*53*). The figure modified from (*27*).

The gain of functionality at the expense of stability through the methionine-leucine exchange in protein hydrophobic cores might be a general feature in protein evolution. This, in turn, would suggest that protein stability and function can be tuned artificially by hydrophobic core mutations. Exchanging hydrophobic core methionine by leucine or isoleucine might thus present a novel approach to engineer more stable proteins, which can resist harsh conditions. Conversely, the exchange of leucine against methionine might improve a protein's function through facilitated conformational dynamics. The consecutive introduction of methionine-to-leucine mutations in the NTD shows a gradual trade-off between stability and function. Thus, the amount of gained or lost function and stability in a protein might be tunable by the extent to which hydrophobic core residues are modified. These observations are extremely valuable for engineering proteins with improved properties for biotechnological applications or mechanistic studies of proteins such as the MaSp1 NTD.

6 Outlook

6.1 L6-NTD: a future standard tool to investigate the NTD dimerization mechanism

In this study, the conformational dynamics of L6-NTD under monomer conditions at pH 7 and high ionic strength were found to be substantially reduced compared to wt-NTD. This suggests that NTD assembly via conformational selection is impaired in L6-NTD, which reduces dimer affinity. Dimerization of NTDs does not occur at pH 7 but requires protonation of acidic sidechains at pH values below 6.5. Strikingly, protonation of wt-NTD was found to accelerate the interconversion between dimer-incompatible and dimer-compatible conformations, which indicates that protonation and conformational changes are to some extent coupled (*15, 17, 19*). Whether this coupling is preserved or lost in the dynamically perturbed L6-NTD may be addressed in

the future via PET-FCS measurements in combination with the dimerization inhibiting A72R mutation (*19*). This might clarify whether an additional layer of dynamic perturbation exists in L6-NTD, which presents an insensitivity towards the pH-induced acceleration of conformational switching between dimer-incompetent and dimer-competent states.

A central assumption in the presented study is that the L6-NTD structure is virtually the same in the monomer as in the dimer. This is strongly supported by NMR and Trp fluorescence experiments. The L6-NTD dimer can therefore be regarded structurally as identical the early dimer in the wt-NTD dimerization mechanism, in which the subunits are held together only by electrostatics. However, a high-resolution structure of the L6-NTD dimer would be necessary to ascertain that L6-NTD does not undergo larger structural rearrangements and maintains the monomer conformation also in the dimer. Moreover, such a structure would also clarify, which inter-subunit contacts mediate the weak dimer assembly. A high-resolution structure of L6-NTD dimer is not feasible with NMR due to the strong line broadening observed for the dimer interface under dimer conditions. A suitable alternative might be X-ray crystallography. It is, however, difficult to control the sample conditions such as pH and ionic strength in a crystal and to prevent artificial conformations of the NTD due to crystal packing. The high structural similarity of the wt-NTD dimer in the crystal and in solution (determined via NMR) might nevertheless encourage crystallization trials of L6-NTD at pH 6 (*12, 18*).

In a recent study, the replacement of the pH relay glutamate residues E79, E84, and E119 with glutamine gave rise to a constitutive NTD dimer (*17*). The glutamine residues mimic protonated glutamate residues and thereby seem to promote the dimer conformation. The effects of the E/Q mutation in the context of L6-NTD might reveal further insights into the relationship and hierarchy between protonation and conformational changes in the NTD dimerization mechanism. Several scenarios upon E/Q mutation for the L6-NTD are imaginable: The formation of tight constitutive L6-NTD dimers with subunits in dimer conformation; the formation of tight constitutive L6-NTD dimers with subunits in the monomer conformation; the formation of weak L6-NTD dimers with subunits in the monomer conformation; or no effect on the monomer-dimer equilibrium. The effects can be easily observed via NMR spectroscopy and compared to wt-NTD E/Q mutant. The formation of a tighter L6-NTD E/Q dimer with monomer conformations would hold great potential to improve the NMR spectral quality of a L6-NTD dimer and open the way to structure determination via NMR. With the wt-NTD E/Q mutant as a reference, the L6-NTD E/Q mutant would also allow more detailed dynamic analyses of the NTD dimer intermediate and its conformational dynamics via NMR and PET-FCS.

The L6-NTD has demonstrated its unique versatility to discern contributions from conformational changes and electrostatic interaction in the first two steps of the NTD dimerization mechanism. Beyond that, the dynamically perturbed L6-NTD is also a promising tool to study potential conformational changes in the final step of the dimerization mechanism between pH 6.0-5.5 (*17*). This final step is presumed to involve a third protonation event at E84, which consolidates NTD dimers before silk fibres leave the spider's abdomen. Whether or not further conformational changes within the NTD or changes of its conformational dynamics accompany E84 protonation is currently unclear. The peak position and line shape analysis in ^{19}F NMR experiments indicate that the W10 conformational dynamics are indeed affected between pH 6.0-5.5, in both wt- and L6-NTD. However, a more detailed analysis of W10 conformational dynamics and the overall NTD conformational state in dependence of changing pH and ionic strength are required to elucidate the molecular details of the final dimerization step. Preliminary ^{19}F NMR based screenings of wt-NTD and L6-NTD at multiple pH values and ionic strengths demonstrate the potential of ^{19}F NMR for future studies of W10 (Figure 147). Together with temperature-dependent ^{19}F NMR relaxation measurements, it might be a key to address additional dynamic time scales of W10 conformational changes (*55*). Moreover, NMR, PET-FCS, and tryptophan fluorescence spectroscopy experiments in analogy to this or previous studies will be required to deduce the structural and dynamic features as well as the energetic contributions of the final protonation step to the overall NTD dimer assembly (*15, 17–19*). Not only will a more elaborate study in a broader pH and ionic strength range validate the currently proposed three-step mechanism, but it will also provide a picture of the NTD conformational and dynamic landscape during its assembly into tight dimers along the spider's spinning duct.

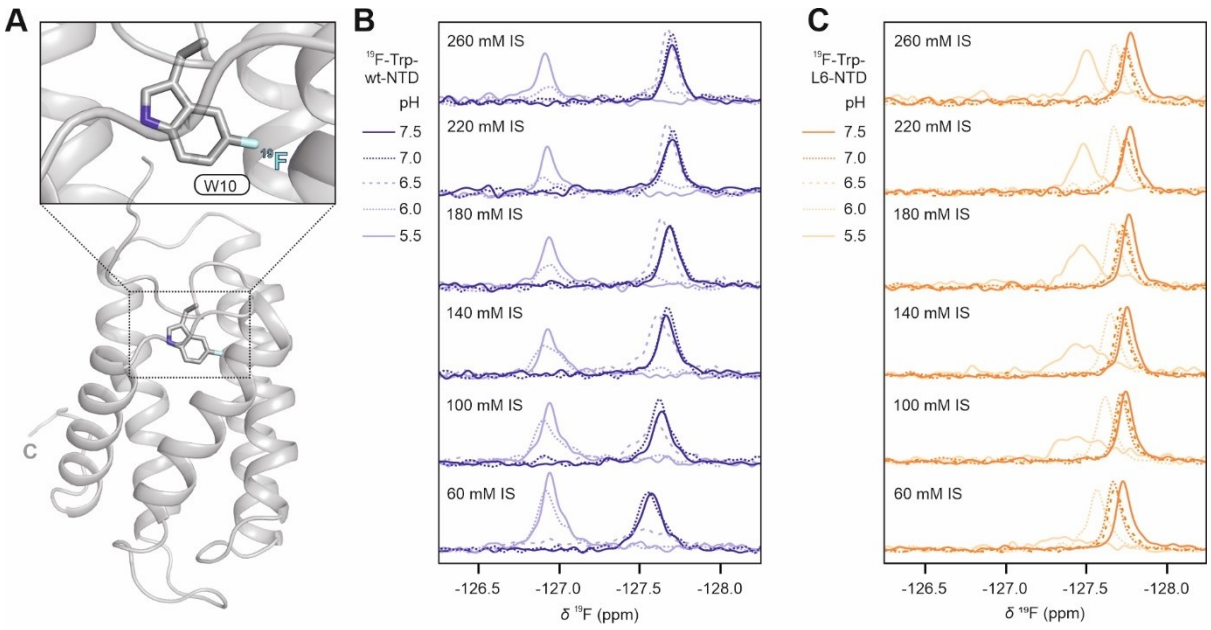


Figure 147: ^{19}F NMR based screen of W10 conformational dynamics in wt- and L6-NTD in a broad pH and ionic strength range. (**A**) Cartoon representation of ^{19}F-Trp-labelled Ea-MaSp1-NTD. ^{19}F NMR spectra of (**B**) ^{19}F-Trp-wt-NTD and (**C**) ^{19}F-Trp-L6-NTD 260, 220, 180, 140, 100, and 60 mM ionic strength (IS) in the range of pH 7.5 to pH 5.5 in steps of 0.5 (for legend see figure).

6.2 A general role of core-methionine residues as determinants of protein stability and function

The observation that methionine-depletion of *Ea*-MaSp1 NTD increases the protein stability but reduces its ability to dimerize points to a role of hydrophobic-core methionine residues in the evolution of silk fibres with improved functions. As outline above, there seems to be a link between the amino acid composition of the NTD hydrophobic core of spidroins and the toughness of silk fibres (*53*). Spidroins with a high content of methionine in the hydrophobic cores of their NTDs form tough silk fibres, whereas spidroins with no methionine in the hydrophobic cores of their NTDs form soft silk fibres. This correlation is presumably encoded in the biophysical properties of methionine, which enables tight NTD dimers and thereby promotes strong inter-spidroin interactions in silk fibres. According to this hypothesis, the properties of silk fibres can be tuned by modifying the composition of spidroin NTD hydrophobic cores, i.e. by replacing methionine with leucine and *vice versa*. In *Ea*-MaSp1, this modification dramatically reduces the NTD dimer affinity but increases NTD stability. In the context of full-length spidroins in an *in vivo* system, an L6-NTD mutant would therefore be expected to produce silk fibres with reduced toughness compared to a wildtype protein. It is tempting to validate this concept by modifying the hydrophobic cores of other methionine-rich NTDs from other spidroins such as MaSp, MiSp, or AcSp. Replacement of methionine by leucine should yield more stable MaSp, MiSp, and AcSp NTDs with concomitantly reduced abilities to dimerize. Consequently, silk fibres formed by methionine-depleted MaSp, MiSp, and AcSp should show a reduced toughness. Conversely, incorporation of methionine into the hydrophobic cores of CySp and TuSp NTDs should reduce protein stability but enable the formation of tight dimers, which should translate into tougher silk fibres. Regarding the biotechnological production of spider silk fibres, the hydrophobic cores of spidroin NTDs present an additional and extremely promising target in spidroin proteins to tune the properties of silk fibres.

The concept of increased functionality at the cost of stability through the incorporation of hydrophobic-core methionine residues might not be limited to spidroin NTDs. Instead, it might be a general concept in the evolution of protein function. In future studies, this should be validated by modifying the hydrophobic cores of other proteins, in which functionality is coupled to large conformational changes. The incorporation of

methionine into cores should improve functionality, whereas methionine-depletion should reduce it. Protein stability, in contrast, should behave inversely. This would have far-reaching implications on the general understanding of how stability and functionality can be improved through protein engineering protein hydrophobic cores.

6.3 A potential role of methionine in mediating long-range interactions in denatured proteins

The thermodynamic analysis of the denatured states of the wt- and L6-NTD suggests that residual interactions mediated by the methionine sulfur atoms stabilize the denatured state of wt-NTD compared to L6-NTD. This shifts the denatured state free energy away from the transition state energy and decelerates folding. NMR spectroscopy is ideally suited to capture the conformational ensemble of proteins in their denatured states at a per residue level. It is therefore ideal for studying the unfolded states of wt- and L6-NTD in molecular detail with respect to potential residual long-range interactions. A study on the denatured state of hen egg lysozyme demonstrates how NMR can be used to elucidate residual structure in denatured proteins (*56*). In this study, the authors measured H^N, C_α chemical shifts, R_2 relaxation rates and HDX of denatured and folded lysozyme. Deviations from random coil chemical shifts, the distribution of R_2 rates across the sequence, and hydrogen exchange protection in the denatured state revealed long-range interactions between sequentially distant regions. A tryptophan residue was identified as a key residue for the formation of a hydrophobic cluster in the denatured state. Mutation of this tryptophan residue to glycine completely abolished the hydrophobic cluster formation. Such an approach might be transferable to study potential long-range interactions and hydrophobic clusters in the methionine-rich NTD and validate the hypothesis that methionine residues promote residual interactions, presumably by their sticky sulfur atom. Considering the role of tryptophan for long-range interactions in denatured lysozyme, residue W10 in the NTD might similarly contribute to the formation of hydrophobic clusters in the denatured NTD state. As pointed out above, methionine residues are known to form stabilizing interactions with tryptophan residues in hydrophobic protein cores (*38*). These stabilizing interactions are presumably less pronounced for leucine-tryptophan pairs. Mutating methionine to leucine might therefore abolish residual interactions and destabilize the denatured state of the NTD. Elucidating the molecular basis of methionine residues for stabilizing denatured protein states would deepen our understanding of how hydrophobic core side chains contribute to the protein folding process and protein stability.

7 References

1. F. Vollrath, P. Selden, The Role of Behavior in the Evolution of Spiders, Silks, and Webs. *Annu. Rev. Ecol. Evol. Syst.* **38**, 819–846 (2007), doi:10.1146/annurev.ecolsys.37.091305.110221.
2. F. Vollrath, D. P. Knight, Liquid crystalline spinning of spider silk. *Nature.* **410**, 541–548 (2001), doi:10.1038/35069000.
3. C. Vendrely, T. Scheibel, Biotechnological production of spider-silk proteins enables new applications. *Macromolecular bioscience.* **7**, 401–409 (2007), doi:10.1002/mabi.200600255.
4. M. Heim, D. Keerl, T. Scheibel, Spider silk: from soluble protein to extraordinary fiber. *Angewandte Chemie (International ed. in English).* **48**, 3584–3596 (2009), doi:10.1002/anie.200803341.
5. X. Li, C.-H. Shi, C.-L. Tang, Y.-M. Cai, Q. Meng, The correlation between the length of repetitive domain and mechanical properties of the recombinant flagelliform spidroin. *Biology open.* **6**, 333–339 (2017), doi:10.1242/bio.022665.
6. J. Kümmerlen, J. D. van Beek, F. Vollrath, B. H. Meier, Local Structure in Spider Dragline Silk Investigated by Two-Dimensional Spin-Diffusion Nuclear Magnetic Resonance †. *Macromolecules.* **29**, 2920–2928 (1996), doi:10.1021/ma951098i.
7. A. Simmons, E. Ray, L. W. Jelinski, Solid-State 13C NMR of Nephila clavipes Dragline Silk Establishes Structure and Identity of Crystalline Regions. *Macromolecules.* **27**, 5235–5237 (1994), doi:10.1021/ma00096a060.
8. J. D. van Beek, S. Hess, F. Vollrath, B. H. Meier, The molecular structure of spider dragline silk: folding and orientation of the protein backbone. *Proceedings of the National Academy of Sciences of the United States of America*. **99**, 10266–10271 (2002), doi:10.1073/pnas.152162299.
9. J. E. Jenkins, G. P. Holland, J. L. Yarger, High resolution magic angle spinning NMR investigation of silk protein structure within major ampullate glands of orb weaving spiders. *Soft matter.* **8**, 1947–1954 (2012), doi:10.1039/C2SM06462F.
10. J. E. Jenkins *et al.,* Characterizing the secondary protein structure of black widow dragline silk using solid-state NMR and X-ray diffraction. *Biomacromolecules.* **14**, 3472–3483 (2013), doi:10.1021/bm400791u.
11. J. M. Gosline, M. W. Denny, M. E. DeMont, Spider silk as rubber. *Nature.* **309**, 551–552 (1984), doi:10.1038/309551a0.
12. G. Askarieh *et al.,* Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. *Nature.* **465**, 236–238 (2010), doi:10.1038/nature08962.
13. F. Hagn *et al.,* A conserved spider silk domain acts as a molecular switch that controls fibre assembly. *Nature.* **465**, 239–242 (2010), doi:10.1038/nature08936.
14. C. Rat, J. C. Heiby, J. P. Bunz, H. Neuweiler, Two-step self-assembly of a spider silk molecular clamp. *Nature communications.* **9**, 4779 (2018), doi:10.1038/s41467-018-07227-5.
15. S. Schwarze, F. U. Zwettler, C. M. Johnson, H. Neuweiler, The N-terminal domains of spider silk proteins assemble ultrafast and protected from charge screening. *Nature communications.* **4**, 2815 (2013), doi:10.1038/ncomms3815.
16. J. C. Heiby, S. Rajab, C. Rat, C. M. Johnson, H. Neuweiler, Conservation of folding and association within a family of spidroin N-terminal domains. *Scientific reports.* **7**, 16789 (2017), doi:10.1038/s41598-017-16881-6.
17. N. Kronqvist *et al.,* Sequential pH-driven dimerization and stabilization of the N-terminal domain enables rapid spider silk formation. *Nature communications.* **5**, 3254 (2014), doi:10.1038/ncomms4254.
18. K. Jaudzems *et al.,* pH-dependent dimerization of spider silk N-terminal domain requires relocation of a wedged tryptophan side chain. *Journal of Molecular Biology.* **422**, 477–487 (2012), doi:10.1016/j.jmb.2012.06.004.
19. J. Ries, S. Schwarze, C. M. Johnson, H. Neuweiler, Microsecond folding and domain motions of a spider silk protein structural switch. *Journal of the American Chemical Society.* **136**, 17136–17144 (2014), doi:10.1021/ja508760a.
20. J. H. Atkison, S. Parnham, W. R. Marcotte, S. K. Olsen, Crystal Structure of the Nephila clavipes Major Ampullate Spidroin 1A N-terminal Domain Reveals Plasticity at the Dimer Interface. *The Journal of biological chemistry.* **291**, 19006–19017 (2016), doi:10.1074/jbc.M116.736710.
21. J. Bauer *et al.,* Acidic Residues Control the Dimerization of the N-terminal Domain of Black Widow Spiders' Major Ampullate Spidroin 1. *Scientific reports.* **6**, 34442 (2016), doi:10.1038/srep34442.
22. I. K. Jordan *et al.,* A universal trend of amino acid gain and loss in protein evolution. *Nature.* **433**, 633–638 (2005), doi:10.1038/nature03306.
23. N. C. Pace, M. J. Scholtz, A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophysical Journal.* **75**, 422–427 (1998), doi:10.1016/S0006-3495(98)77529-0.
24. S. J. Lombardi, D. L. Kaplan, The Amino Acid Composition of Major Ampullate Gland Silk (Dragline) of Nephila Clavipes (Araneae, Tetragnathidae). *The Journal of Arachnology*, 297–306 (1990).
25. J. C. Heiby, B. Goretzki, C. M. Johnson, U. A. Hellmich, H. Neuweiler, Methionine in a protein hydrophobic core drives tight interactions required for assembly of spider silk. *Nature communications.* **10**, 4378 (2019), doi:10.1038/s41467-019-12365-5.
26. O. V. Tsodikov, M. T. Record, Y. V. Sergeev, Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *Journal of computational chemistry.* **23**, 600–609 (2002), doi:10.1002/jcc.10061.
27. B. Goretzki, J. C. Heiby, C. Hacker, H. Neuweiler, U. A. Hellmich, NMR assignments of a dynamically perturbed and dimerization inhibited N-terminal domain variant of a spider silk protein from E. australis. *Biomolecular NMR assignments.* **14**, 67–71 (2020), doi:10.1007/s12104-019-09922-w.
28. S. Grzesiek, A. Bax, The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. *Journal of the American Chemical Society.* **115**, 12593–12594 (1993), doi:10.1021/ja00079a052.
29. S. Grzesiek, A. Bax, The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.*, 12539–12594 (1993).
30. M. Sattler, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in nuclear magnetic resonance spectroscopy.* **34**, 93–158 (1999), doi:10.1016/S0079-6565(98)00025-9.

31. R. Kazlauskas, Engineering more stable proteins. *Chemical Society reviews*. **47**, 9026–9045 (2018), doi:10.1039/c8cs00014j.
32. C. Chothia, Principles that determine the structure of proteins. *Annual review of biochemistry*. **53**, 537–572 (1984), doi:10.1146/annurev.bi.53.070184.002541.
33. J. U. Bowie, J. F. Reidhaar-Olson, W. A. Lim, R. T. Sauer, Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science (New York, N.Y.)*. **247**, 1306–1310 (1990), doi:10.1126/science.2315699.
34. W. Jiang, G. Askarieh, A. Shkumatov, M. Hedhammar, S. D. Knight, Structure of the N-terminal domain of Euprosthenops australis dragline silk suggests that conversion of spidroin dope to spider silk involves a conserved asymmetric dimer intermediate. *Acta crystallographica. Section D, Structural biology*. **75**, 618–627 (2019), doi:10.1107/S2059798319007253.
35. G. D. Rose, A. R. Geselowitz, G. J. Lesser, R. H. Lee, M. H. Zehfus, Hydrophobicity of amino acid residues in globular proteins. *Science (New York, N.Y.)*. **229**, 834–838 (1985), doi:10.1126/science.4023714.
36. Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of biomolecular NMR*. **44**, 213–223 (2009), doi:10.1007/s10858-009-9333-z.
37. Q. Wang, A. M. Buckle, N. W. Foster, C. M. Johnson, A. R. Fersht, Design of highly stable functional GroEL minichaperones. *Protein science : a publication of the Protein Society*. **8**, 2186–2193 (1999), doi:10.1110/ps.8.10.2186.
38. D. Pal, P. Chakrabarti, Non-hydrogen bond interactions involving the methionine sulfur atom. *Journal of biomolecular structure & dynamics*. **19**, 115–128 (2001), doi:10.1080/07391102.2001.10506725.
39. J. Kubelka, J. Hofrichter, W. A. Eaton, The protein folding 'speed limit'. *Current opinion in structural biology*. **14**, 76–88 (2004), doi:10.1016/j.sbi.2004.01.013.
40. P. Debye, E. Hückel, The interionic attraction theory and deviations from ideal behavior in solution. *Phys. Z.*, 185–206 (1923).
41. K. Henzler-Wildman, D. Kern, Dynamic personalities of proteins. *Nature*. **450**, 964–972 (2007), doi:10.1038/nature06522.
42. Y. Bai, Protein folding pathways studied by pulsed- and native-state hydrogen exchange. *Chemical reviews*. **106**, 1757–1768 (2006), doi:10.1021/cr040432i.
43. V. A. Jarymowycz, M. J. Stone, Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chemical reviews*. **106**, 1624–1671 (2006), doi:10.1021/cr040421p.
44. M. Andersson *et al.,* Carbonic anhydrase generates CO2 and H+ that drive spider silk formation via opposite effects on the terminal domains. *PLoS biology*. **12**, e1001921 (2014), doi:10.1371/journal.pbio.1001921.
45. E. P. Baldwin, B. W. Matthews, Core-packing constraints, hydrophobicity and protein design. *Current Opinion in Biotechnology*. **5**, 396–402 (1994), doi:10.1016/0958-1669(94)90048-5.
46. W. A. Lim, A. Hodel, R. T. Sauer, F. M. Richards, The crystal structure of a mutant protein with altered but improved hydrophobic core packing. *Proceedings of the National Academy of Sciences of the United States of America*. **91**, 423–427 (1994), doi:10.1073/pnas.91.1.423.
47. J. C. Aledo, Methionine in proteins: The Cinderella of the proteinogenic amino acids. *Protein science : a publication of the Protein Society*. **28**, 1785–1796 (2019), doi:10.1002/pro.3698.
48. J. Janin, S. Wodak, M. Levitt, B. Maigret, Conformation of amino acid side-chains in proteins. *Journal of Molecular Biology*. **125**, 357–386 (1978), doi:10.1016/0022-2836(78)90408-4.
49. S. H. Gellman, On the role of methionine residues in the sequence-independent recognition of nonpolar protein surfaces. *Biochemistry*. **30**, 6633–6636 (1991), doi:10.1021/bi00241a001.
50. H. D. Bernstein *et al.,* Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle. *Nature*. **340**, 482–486 (1989), doi:10.1038/340482a0.
51. K. T. O'Neil, W. F. DeGrado, How calmodulin binds its targets: sequence independent recognition of amphiphilic α-helices. *Trends in Biochemical Sciences*. **15**, 59–64 (1990), doi:10.1016/0968-0004(90)90177-D.
52. C.-S. Goh, D. Milburn, M. Gerstein, Conformational changes associated with protein-protein interactions. *Current opinion in structural biology*. **14**, 104–109 (2004), doi:10.1016/j.sbi.2004.01.005.
53. T. A. Blackledge, C. Y. Hayashi, Silken toolkits: biomechanics of silk fibers spun by the orb web spider Argiope argentata (Fabricius 1775). *The Journal of experimental biology*. **209**, 2452–2461 (2006), doi:10.1242/jeb.02275.
54. D. L. Trudeau, M. Kaltenbach, D. S. Tawfik, On the Potential Origins of the High Stability of Reconstructed Ancestral Proteins. *Molecular biology and evolution*. **33**, 2633–2641 (2016), doi:10.1093/molbev/msw138.
55. M. Lu, R. Ishima, T. Polenova, A. M. Gronenborn, 19F NMR relaxation studies of fluorosubstituted tryptophans. *Journal of biomolecular NMR*. **73**, 401–409 (2019), doi:10.1007/s10858-019-00268-y.
56. J. Klein-Seetharaman *et al.,* Long-range interactions within a nonnative protein. *Science (New York, N.Y.)*. **295**, 1719–1722 (2002), doi:10.1126/science.1067680.

IX Acknowledgements

To protect the privacy of the author, the acknowledgements are not available in the online version of this thesis.

X Declaration

I hereby declare that I wrote the dissertation submitted without any unauthorized external assistance and used only sources acknowledged in the work. All textual passages which are appropriated verbatim or paraphrased from published and unpublished texts as well as all information obtained from oral sources are duly indicated and listed in accordance with bibliographical rules. In carrying out this research, I complied with the rules of standard scientific practice as formulated in the statutes of Johannes Gutenberg University Mainz to insure standard scientific practice.

_____

Benedikt Goretzki

Mainz, den _____

XI Curriculum vitae

To protect the privacy of the author, the CV is not available in the online version of the thesis.