Microbial communication –

an interplay of producer, transporter and receptor

Dissertation

zur Erlangung des Grades

„Doktorin der Naturwissenschaften“

im Promotionsfach Chemie

am Fachbereich Chemie, Pharmazie,

Geographie und Geowissenschaften

der Johannes Gutenberg-Universität Mainz

Mai Anh Tran

geb. in Hanoi, Vietnam

Mainz, 2022

Dekanin: _____

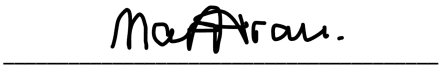
1. Gutachterin: _____

2. Gutachter: _____

Tag der mündlichen Prüfung: 16.03.2023

Urheberschaftserklärung

Ich, Mai Anh Tran, versichere, dass ich die vorliegende Dissertation mit dem Titel „Microbial communication – an interplay of producer, transporter and receptor“ selbst verfasst, nicht andere als die in ihr angegebenen Quellen oder Hilfsmittel benutzt, alle vollständig oder sinngemäß übernommenen Zitate als solche gekennzeichnet sowie die Dissertation in der vorliegenden oder einer ähnlichen Form noch bei keiner anderen in- oder ausländischen Hochschule anlässlich eines Promotionsgesuchs oder zu anderen Prüfungszwecken eingereicht habe.



Danksagung

Gelöscht für die Online-Version

Zusammenfassung

Eine erfolgreiche Kommunikation erfordert eine Nachricht, einen Sender und einen Empfänger. Bei der bakteriellen Kommunikation kann es sich bei der Botschaft um Signalmoleküle handeln, die von multimodularen Enzymen wie nichtribosomalen Peptidsynthetasen (NRPS) und Polyketidsynthasen (PKS) hergestellt werden. Diese synthetisieren Peptidyl- und Polyketidprodukte ohne genetische Transkripte. Membrantransporterproteine wie ATP-Bindekassetten (engl. ATP binding cassette, ABC)-Transporter transportieren solche Signalmoleküle über die Membran, wo sie für Empfängersysteme zugänglich werden. Zwei-Komponenten-Systeme (engl. Two-component systems, TCS) sind typische bakterielle Signalempfänger und Weiterleitungssysteme. Nach der Bindung des Signalmoleküls an eine Histidin-Kinase findet eine intrazelluläre Autophosphorylierungsreaktion statt. Die Phosphatgruppe wird dann auf einen löslichen Reaktionsregulator übertragen, welcher die Genexpression regulieren kann.

Die mikrobielle Kommunikation ist für das bakterielle Überleben von Vorteil. Sie ermöglicht Mikroben auf Umweltveränderungen zu reagieren und Reaktionen, z. B. gegen Konkurrenten, zu koordinieren. Bei infizierten Patienten können Krankheitserreger diese Mechanismen jedoch auch nutzen, um sich an neue Nischen anzupassen, Biofilme zu bilden und Antibiotika zu widerstehen. Da die mikrobielle Kommunikation ein wesentlicher Bestandteil allen (mikrobiellen) Lebens ist, ist es von großer Bedeutung, ihre molekularen Grundlagen im Detail zu verstehen.

Der erste Teil dieser Arbeit befasst sich mit der Struktur und Funktion von multimodularen Enzymen aus verschiedenen Bakterien. Die monomodulare NRPS Pyreudion-Synthetase (Pys) aus *Pseudomonas fluorescens* wurde als Volllängenprotein aufgereinigt und war aktiv *in vitro, d.h.* die Produktion von Pyreudionen konnte nachgewiesen werden. Die Röntgenkristallographie der isolierten Kondensationsdomäne des Pys-Homologs Acylprolinsynthetase (Aps) aus *Pectobacterium betavasculorum* lieferte einen ersten strukturellen Einblick in die Fettsäurekopfgruppenselektivität von Aps und Pys. Die Kernspinresonanzspektroskopie (engl. Nuclear magnetic resonance, NMR) an zwei Acyl-Carrier-Proteinen (ACPs) aus den Polyketidsynthasen von *Burkholderia rhizoxinica* und *Gloeocapsa sp. PCC 7428* ergab für letztere eine unerwartete Struktur. Darüber hinaus wurde mittels NMR die Interaktion der ACPs mit ihren entsprechenden katalytischen Domänen untersucht.

Der zweite Teil dieser Arbeit bestand darin, die Isolierung und Reinigung von Transporterproteinen zu verbessern. Häufig werden Untersuchungen von Membranproteinen durch die schwierige Aufreinigung dieser Proteine behindert. Hier wurden sieben verschiedene Reinigungs-/Rekonstitutionsmethoden in verschiedenen membranmimetischen Systemen am ABC-Transporters BmrA aus *Bacillus subtilis* getestet. Die Ergebnisse zeigen, dass die Wahl der Reinigungsmethode und der Lipidumgebung die ATPase-Aktivität des Transporters stark beeinflusst. Eine Methode zur Isolierung von Membranproteinen mit Erhaltung ihrer nativen Lipidumgebung sind amphiphile Styrol-Maleinsäure (SMA)-Polymere. Aufgrund der hohen Dispersität handelsüblicher Polymere wurden SMA-Polymere mit definierter Molekulargewichtsverteilung durch reversible Additions-Fragmentierungs-Kettentransfer Polymerisation synthetisiert. Mittels dynamischer Lichtstreuung,

Größenausschlusschromatographie und Transmissionselektronenmikroskopie konnte gezeigt werden, dass die Qualität von Lipidpartikeln und Protein-Nanodiscs stark von der strukturellen Homogenität des Polymers beeinflusst wird.

Der dritte Teil der Arbeit konzentrierte sich auf die Strukturuntersuchung eines Rezeptorproteins, der Sensor-Histidin-Kinase (SHK) SpaK aus *Bacillus subtilis*. SpaK ist die SHK eines TCS, dass das Signalmolekül Subtilin in einem Quorum-Sensing-Mechanismus empfängt. Das Membranprotein SpaK wurde durch SMA-Polymere isoliert und mittels Kryo-Elektronenmikroskopie analysiert. Die erhaltenen niedrig aufgelösten Elektronendichte Karten weisen auf Helixbiegung und hohe Domänenflexibilität hin. In einem ersten Versuch, eine isolierte Sensordomäne für Strukturstudien zu reinigen, wurde die Sensordomäne des Homologen SHK NisK aus *Lactococcus lactis* aufgereinigt und erste Zirkulardichroismus-Spektroskopie-Spektren aufgenommen.

Summary

Successful communication requires a message, a transmitter, and a receiver. In bacterial communication, the message can be signal molecules produced by multimodular enzymes such as nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS). These synthesize peptidyl and polyketide products without genetic transcripts. Membrane transporter proteins such as ATP binding cassette (ABC) transporters shuttle such signal molecules across the membrane, where they become accessible for recipient systems. Two component systems (TCS) are typical bacterial signal receivers and relay systems. Upon signal molecule binding to a histidine kinase, an intracellular autophosphorylation reaction takes place. The phosphate group will then be transferred to a soluble response regulator that can regulate gene expression.
Microbial communication is beneficial to bacterial survival because it allows microbes to react to environmental changes and to coordinate responses, e.g., against competitors. However, in infected patients, pathogens can also use these mechanisms to adapt to new niches, form biofilms and evade antibiotic treatment Since microbial communication is an essential part of all (microbial) life, it is of high importance to understand its molecular basis in detail.

The first part of this thesis addresses the structure and function of multimodular producing enzymes from different bacteria. The monomodular NRPS pyreudion synthetase (Pys) from *Pseudomonas fluorescens* was purified as a full-length protein and showed production of pyreudiones *in vitro*. X-ray crystallography of the isolated C-domain from the Pys homolog acylproline synthetase (Aps) from *Pectobacterium betavasculorum* provided a first structural insight into the fatty acid headgroup selectivity of Aps and Pys. Nuclear magnetic resonance (NMR) spectroscopy on two acyl carrier proteins (ACPs) from *Burkholderia rhizoxinica* and *Gloeocapsa sp. PCC 7428* polyketide synthases showed an unexpected structure for the latter. Furthermore, NMR was used to investigate the interaction of ACPs with their corresponding catalytic domains.

The second part of this thesis aimed to improve the isolation and purification of transporter proteins and to develop a general toolbox for their extraction from native environments. Often, membrane protein studies are hampered due to challenging protein purification. Here, seven different purification/reconstitution methods into different membrane mimetic systems were tested using the ABC transporter BmrA from *Bacillus subtilis.* The results show that the choice of purification method and lipid environment highly affects the transporter's ATPase activity. A method to isolate membrane proteins directly in their native lipid environment are amphiphilic styrene maleic acid (SMA) polymers. Due to the high dispersity of commercially available polymers, we synthesized SMA polymers with defined molecular weight distribution by reversible-addition-fragmentation chain-transfer (RAFT) polymerization. By dynamic light scattering, size exclusion chromatography, and transmission electron microscopy it was shown that the quality of lipid particles and protein nanodiscs are highly influenced by the structural homogeneity of the polymer.

The third part of the thesis focused on the structural study of a receptor protein, the sensor histidine kinase (SHK) SpaK from *Bacillus subtilis*. SpaK is the SHK part of a TCS that senses the signaling molecule subtilin in a quorum sensing mechanism. The membrane protein SpaK was isolated by SMA polymers and analyzed by cryo electron microscopy. The obtained low-resolution

electron density maps indicate helix bending and high domain flexibility. In a first attempt to purify an isolated sensor domain for structural studies, the sensor domain of the homolog SHK NisK from *Lactococcus lactis* was purified and first circular spectroscopy spectra were recorded.

Table of contents

List of abbreviations

A domain adenylation domain
AA amino acid
ABC ATP binding cassette
ACP acyl carrier protein
ADP adenosine diphosphate
AF2 AlphaFold2
AHL acylhomoserine lactone
AIP autoinducing peptide
AMP adenosine monophosphate
Aps acylprolin synthetase
AT acyl transferase
ATP adenosine triphosphate

B domain branching domain
B. subtilis Bacillus subtilis
BA benzamidine
Benz benzoate-CoA ligase
BmrA *Bacillus* multidrug-resistance ATP

C domain condensation domain
CA catalytic ATP binding
CD cicular dichroism
CLF chain length factor
CMC critical micelle concentration
CoA acetyl coenzyme A
CryoEM cryo-electron spectroscopy
CSTR continuous stirred tank reaction
CV column volume
CYC cyclase

Đ dispersity
DBD DNA-binding domain
DDM N-dodecyl-β-D-maltoside
DH dehydratase
DHp dimerization and histidine phosphorylation
DIB diisobutylene
DIBMALP diisobutylene maleic acid lipid particle
DMPC 1,2-dimyristoyl-sn-glycerol-3-phosphocholine
DMSO dimethyl sulfoxide

E. coli Escherichia coli
ECF energy-coupling factor
ER enoyl reductase

FAD flavin adenine dinucleotide
FC12 dodecylphosphocholine
FR free radical

GAF cGMP-Specific phosphodiesterase adenylyl cyclases and FhlA
Glo *Gloeocapsa sp. PCC 7428*
GNAT N-acyltransferase

HAMP histidine kinases, adenylyl cyclases, methyl binding proteins, phosphatases
HC heterocyclization
HDX-MS H/D exchange coupled to mass spectrometry

IF inward facing
IO inward occluded
IPTG isopropyl-β-D-thiogalactopyranoside

KR ketoreductase
KS ketosynthase

L. lactis Lactococcus lactis
LB lysogeny broth
LC-MS liquid chromatography mass spectrometry
LDAO lauryl dimethylamine-N-oxide
LDH lactate dehydrogenase
LMNG lauryl maltose neopentyl glycol

MA maleic acid
MAT malonyl transferase
MET methyltransferase

MSP membrane scaffold protein

NADPH nicotinamide adenine dinucleotide phosphate
NaPi sodium phosphate buffer
NBD nucleotide binding domain
Ni-NTA nickel nitrilotriacetic acid
NMR nuclear magnetic resonance
NRP non-ribosomal peptide
NRPS non-ribosomal peptide synthetase

OD_{600} optical density at wavelength 600 nm
OF outward facing
OXY oxygenase

PAS Per-ARNT-SIM
PCP peptidyl carrier protein
PDB protein databank
PDC PhoQ, DcuS and CitA
PEP phosphoenolpyruvate
PK pyruvate kinase
PKS polyketide synthase
Ppant 4'phosphopantethein
PP_i pyrophosphate
PPTase phosphopantetheinyl transferase
Pys pyreudione synthetase

QS quorum sensing

RAFT reversible addition–fragmentation chain transfer
REC receiver domain
RiPP ribosomally synthesized and post-translationally modified peptide
RMSD Root-mean-square deviation of atomic positions
rpm rounds per minute
RR response regulator

S styrene
SapA saposin A
SAXS small angle X-ray scattering
SD sensor domain
SDS Sodium dodecyl sulfate
SDS-PAGE sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SEC size exclusion chromatography
Sfp surfactin-type phosphopantetheinyl transferase
SHK sensor histidine kinase, sensor histidine kinase
SMALP styrene maleic acid lipid particle
SMI styrene maleimide
SNAC N-acetylcysteamine
SRD signal-relay domain

T domain thiolation domain
TB terrific broth
TCS two-component system
TE thioesterase domain
TEV Tobacco Etch Virus
TMD transmembrane domain
tRNA transfer ribonucleic acid

UV-vis ultraviolet-visible

Molecular basis of microbial communication

Communal living provides advantages to bacteria such as resistance to environmental stressors, protection against predation, increased nutrients provision, and efficient usage of available resources (*1, 2*). As a way of communication between microbial neighbors, small molecules are constantly exchanged within microbial populations. This allows immediate adaptation to environmental changes and is a huge evolutionary benefit for microbes.

A special type of communication between microorganisms that allows bacteria to synchronize their actions and act in unison is called quorum sensing (QS) (*3*). Through QS, bacteria are able to adapt their behavior to changes in the cell density and living environment (*4*). The signaling molecules involved in QS, also known as autoinducers, are synthesized intracellularly and cross the membrane by diffusion or active transport. As the bacterial population density increases, the autoinducers accumulate in the environment and can be recognized by specific receptors (*5*). Once a critical autoinducer concentration is reached, a concerted response is triggered throughout the bacterial population that alters gene regulation (*6*).

QS occurs in both Gram-negative and Gram-positive bacteria. Gram-negative bacteria use two mechanisms to communicate: i) cytoplasmic transcription factors to recognize small signaling molecules called acylhomoserine lactones (AHL) and ii) membrane-bound two-component histidine sensor kinases that recognizes signaling molecules. The AHL signaling molecules are produced intracellularly by LuxI-type synthases and diffuse freely across the bacterial membrane (*7*) (Figure 1A). Cytoplasmic LuxR-type transcription factors recognize AHLs produced by LuxI-type synthases, bind the autoinducer, dimerize, and bind to DNA (*7*) (Figure 1A). After LuxR binding to DNA, the target genes are regulated (*7*). The second class of Gram-negative QS receptors are the sensor histidine kinases (SHK) which are part of a two-component system (TCS). The SHK can sense various environmental signals ranging from cations to signaling molecules. The second component of the TCS are cytoplasmic transcription factors that are activated by phosphorylation by the SHK (*4*). A schematic representation of the mechanism can be found in Figure 1B.

Gram-positive bacteria rely on the recognition of peptide-based autoinducing molecules (autoinducing peptides AIPs) by either a TCS mechanism or a membrane bound AIP importer and an intracellular receptor (*6*) (Figure 1C and D). One of the well-described QS mechanisms based on a TCS is found in Gram-positive *Lactococcus lactis* (*L. lactis*) and *Bacillus subtilis (B. subtilis)* that produce the ribosomally synthesized and post-translationally modified peptides (RiPPs) nisin and subtilin, respectively. They are lanthipeptides that, in addition to their roles in intraspecies communication, have antibiotic properties against a broad spectrum of Gram-positive bacteria, including multiresistant clinical isolates (*8, 9*). Once the lanthipeptides nisin and subtilin are produced and modified by different modification enzymes, they are transported to the extracellular region by an ATP biding cassette (ABC) transporter (*10*). Extracellularly, nisin and subtilin bind to the sensor domain of membrane bound SHKs, resulting in intracellular

autophosphorylation of the SHK protein. Intracellularly, phosphorylated SHK transfers its phosphate to the partner response regulator protein and activates it. The activated regulator protein can then bind to DNA to regulate gene expression of the nisin/subtilin gene cluster, which encodes the enzymes for AIP production and modification, the ABC exporters, the two TCS proteins as well as other immunity proteins that prevent lethal action of nisin and subtilin against the producer strains (*11*). The TCS quorum sensing mechanism is schematically shown in Figure 1D.

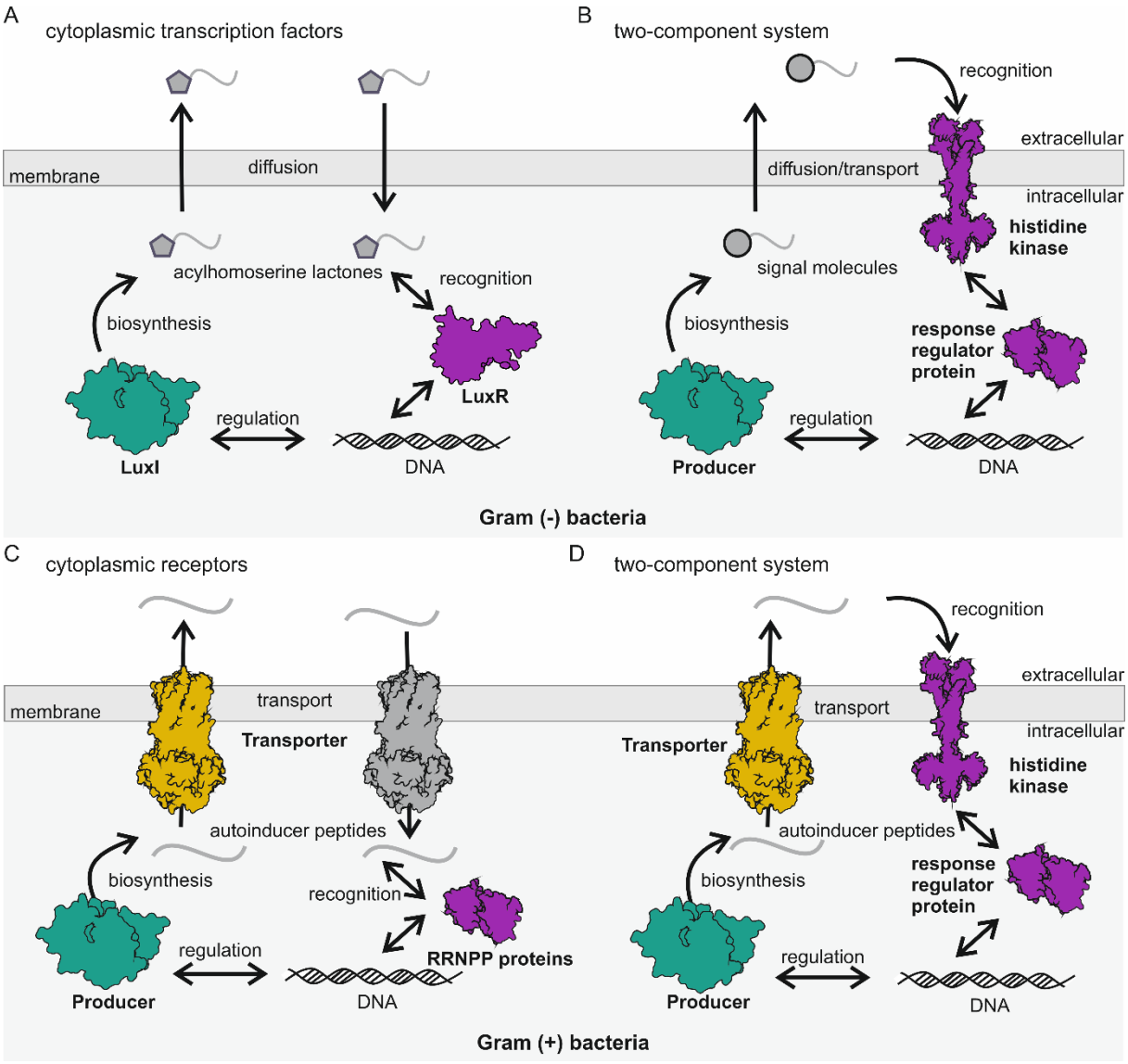


Figure 1: Quorum sensing mechanisms. A: Quorum sensing mechanism in Gram-negative bacteria based on cytoplasmic transcription factors (LuxR-type receptors (purple)) and LuxI-type synthases (cyan). B: Quorum sensing mechanism in Gram-negative bacteria based on the two-component system consists of a histidine kinase and a response regulator protein (purple). C: Quorum sensing mechanism in Gram-positive bacteria based on RRNPP-type receptor proteins (purple) and an importer protein (grey). C: Quorum sensing mechanism in Gram-positive bacteria based on the two-component system consists of a histidine kinase and a response regulator protein (purple). Producer proteins are colored cyan, transporters are yellow and receptor proteins are colored purple.

The second QS system that is found in Gram-positive bacteria is the AIPs recognition by cytoplasmic receptors, the RRNPP proteins. RRNPP proteins are named after the five first experimentally characterized families of such receptors: Rap (*Bacillus* genus), Rgg (*Streptococcus*

genus), NprR (*Bacillus cereus* group), PlcR (*Bacillus cereus* group) and PrgX (pCF10 plasmid of *Enterococcus faecalis*) (*12*). After AIPs have entered the cell via a transport system (Figure 1C), they bind to RRNPP proteins that can act as transcriptional regulator proteins. Similar to the TCS based QS system, the activated regulator proteins subsequently modulate the expression of different target genes (*13*).

As it is shown by the different QS mechanisms, QS thus contains a minimum of three steps: (i) the production of signaling molecules, (ii) the release of these molecules through diffusion or transport, and (iii) the recognition of extracellular signaling molecules that leads to cellular responses (*14, 4*) (Figure 2). Each of these steps will be covered in more detail in one of the result chapters of this thesis.

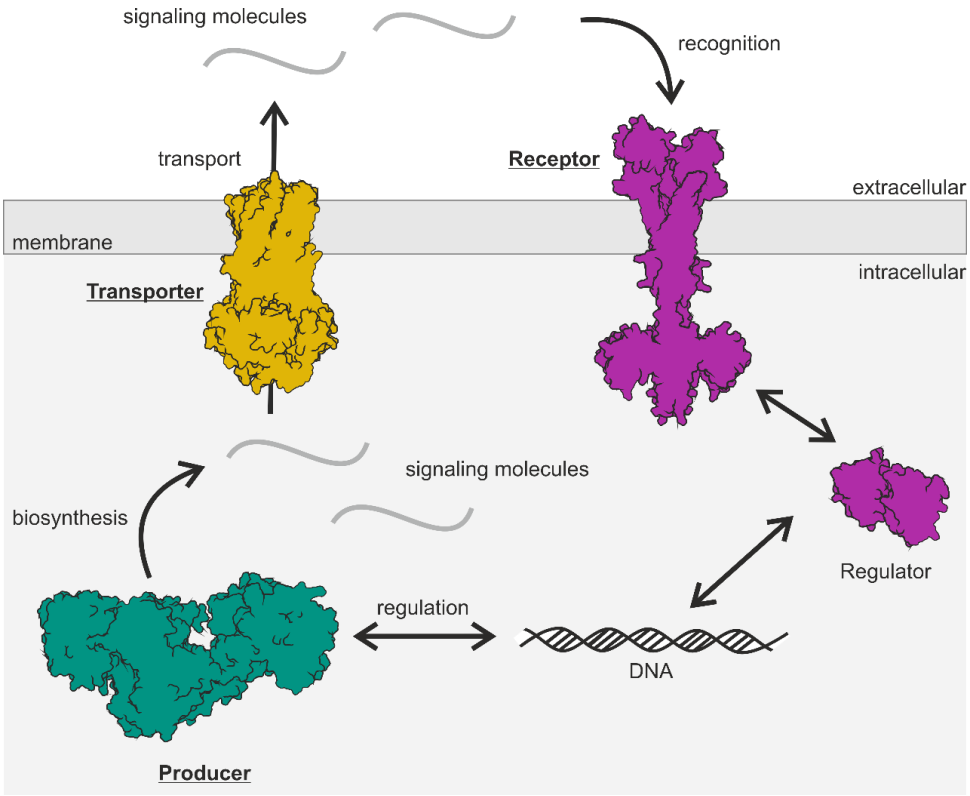


Figure 2: Minimum requirement for a functional microbial communication including production of signaling molecules (grey), transport from intracellular to extracellular, recognition by a receptor protein, e.g., a histidine kinase, signal transduction to a regulator protein, which regulates substrate production through gene regulation (DNA in black). This mechanism is based on the two-component system consists of a membrane-bound receptor and an intracellular regulator protein (purple).

Understanding the mechanism of microbial communication and thus controlling population densities and reducing virulence opens possibilities to combating diseases caused by pathogenic microorganisms. Furthermore, inhibition of microbial communication can be used in environmental protection work, such as sewage, organic pollutant, and heavy metal clearance (*15*). However, the interplay between specific producers, transporters, and receptors has not yet been fully understood, and the complexity of such a communication system makes it difficult to investigate. In addition, it can be technically difficult to access all three systems from the same host for *in vitro* studies. However, a detailed functional and regulation analysis of representative isolated components of such a communication system can shed light on this complex topic.

In the following chapters, different aspects of microbial communication are investigated. In chapter 1, different synthetic mechanisms to produce bioactive compounds are looked at, namely a non-ribosomal peptide synthetase (NRPS) and a polyketide synthase (PKS). After their synthesis, bioactive substances need to be exported from the cell. Finally, they can activate sensors such as histidine kinases to activate cellular responses. Using *B. subtilis* as a model system, these membrane-bound transport and recognition modules were investigated. *B. subtilis* is a nonpathogenic Gram-positive bacterium that has been the go-to model organism for understanding the fundamental principles of communal living (*16*). It produces antimicrobial peptides such as subtilin, also known as lantibiotics, that act as signaling molecules in intraspecies communication, and at the same time show activities against competitive micro-organisms. Here, a *B. subtilis* ABC transporter, BmrA, and its transport activity in different lipid environments are the focus of chapter 2. Finally, the purification, structure, and activity of the subtilin-sensing two-component system, consisting of a histidine kinase and a response regulator protein, are discussed in chapter 3.

Chapter 1

The producer

Chapter 1: The producer

1.1. Theoretical background

Quorum sensing (QS) is a communication mechanism in bacteria that consists of the production and recognition of extracellular signaling molecules. Often, signaling molecules not only play a role in intraspecies communication, but also exhibit antibiotic activities against a wide range of microorganisms to ensure the survival of the producer organism. In Gram-positive bacteria, autoinducing peptides (AIPs) are used as signaling molecules to trigger QS, while Gram-negative bacteria use a variety of signaling molecules ranging from acyl-homoserine lactones to fatty acids, ketone-based, and quinolone derivates (*17*). Structural diversity of signaling molecules is due to the diversity of production mechanisms, and the molecule structures can be divided into i) peptidic and ii) nonpeptidic molecules.

To produce peptidic molecules a peptide bond is formed between a carboxylic acid and an amine. One mechanism to catalyze peptide bond formation is the use of transfer ribonucleic acid (tRNA) and ribosomes. Thereby, the genetic code is used to guide the selection of tRNA and the covalently attached amino acid. Peptide bond formation between to amino acids is then catalyzed by the large subunit of the ribosome (*18*). To increase structural variability beyond the capabilities of the proteinogenic amino acid, post-translational modifications are carried out, resulting in ribosomally synthesized and post-translationally modified peptides (RiPPs). As shown in the previous chapter, AIPs are produced in this manner. Although RiPPs display complex architectures, their diversity is limited by the available RNA-templated building blocks. The incorporation of non-proteinogenic amino acids and their underlying structural flexibility cannot be achieved by the ribosomal synthesis. An enzyme that allows the incorporation of non-proteinogenic amino acids and still forms peptide bonds is the non-ribosomal peptide synthase (NRPS).

1.1.1. Non-ribosomal peptide synthetases (NRPSs) catalyzes peptide formation

NRPSs are mostly found bacteria and fungi, and their products represent a broad spectrum of bioactive compounds (*19*). Modular multienzyme NRPS catalyzes the formation of peptide bonds that lead to non-ribosomal peptides (NRPs). During NRP biosynthesis, each module of an NRPS is responsible for the selection, loading, and condensation of a single amino acid building block. The number of modules corresponds to the number of amino acids incorporated in the final peptide. In addition, the order in which the modules are arranged in the synthase determines the sequence of the peptidic product. Remarkably, NRP production is completely independent of genetic code and RNA-templates. NRPS assembly lines can range from a single module as in the *Pseudomonas fluorescens* pyreudione synthetase up to 18 modules in length as in the peptaibol synthetase from

Trichoderma virens(20, 21) (Figure 3). With a module size between 100 and 200 kDa, depending on domain composition, NRPS assembly lines can reach protein sizes of several MDa (*22, 20*).

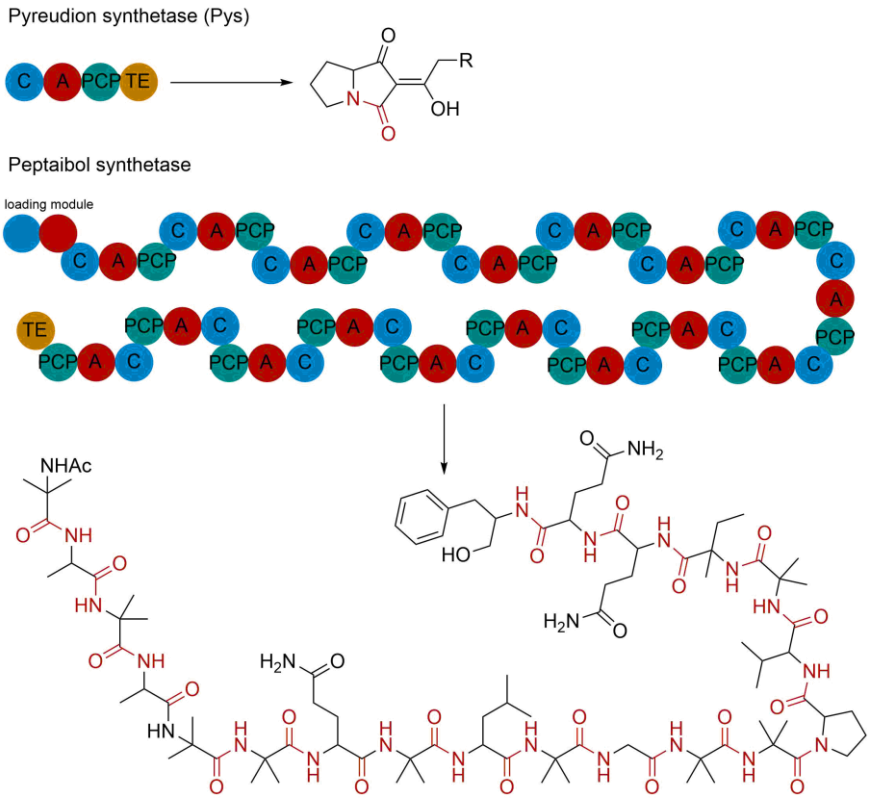


Figure 3: Schematic module composition of the smallest (pyreudion synthetase) and the largest (peptaibol synthetase) NRPS to date and their biosynthetic product. One module consists of three domains: Condensation (C, blue), adenylation (A, red), peptidyl carrier protein (PCP, cyan). The thioesterase domain (TE, yellow) releases the final product from the producing enzymes. Amide bonds are highlighted in red.

As already mentioned, the order and the structure of an NRPS play a vital role in its function. A typical linear synthetase consists of multiple modules, which in turn consist of a set of at least three different domains: the adenylation domain (A domain), the condensation domain (C domain), and peptidyl carrier protein (PCP). The A domain is responsible for selecting and activating specific amino acids. The selection of amino acids is not limited to the 20 proteinogenic L-α-amino acids, but non-proteinogenic amino acids can be selected. The amino acid selection occurs by the specific recognition of the amino acid by the catalytic pocket of the A domain. It has been showed that the specificity of the A domain is defined by approximately ten residues in the active site of the A domain (*23*). These residues act as a code for amino acid substrate recognition by the A domain, and mutation of these residues results in substrate alteration but does not have an impact on the catalytic activity of the A domain (*24*). Based on the analysis of different A domains and their substrate selectivity, web-based programs were created to predict A domain selectivity (NRPSpredictor2 (*25*) or NRPS/PKS substrate predictor (*26*)). After substrate selection, the selected amino acid is activated by adenylation reaction that consumes ATP to form aminoacyl-AMP and pyrophosphate. A schematic adenylation reaction catalyzed by the A domain is shown in Figure 4B. The A domain then loads the activated amino acyl-AMP onto the PCP domain by forming a thioester adduct with the reactive phosphopantetheine arm of the PCP domain (*27*). The thiolation reaction is shown in Figure 4C. Another amino acid is activated in the

same manner in the neighboring module. Once both amino acids are covalently attached to the PCP domain, they are transferred to a C domain located between the two mentioned PCP domains. In the catalytic tunnel of the C domain, a coupling reaction occurs between the two amino acids (Figure 4D) catalyzed by the histidine in the conserved HxxxDG motif (*28*). The catalytic histidine acts as a base, deprotonating and increasing the nucleophilicity of the amino group, and allowing the nucleophilic attack on the carbonyl of the thioester bound amino acid to occur (*29, 30*). The formed amide bond releases a free thiol-terminated phosphopantetheine arm and the extended peptide chain is bound to the PCP of the downstream module (*30*) (Figure 4D).

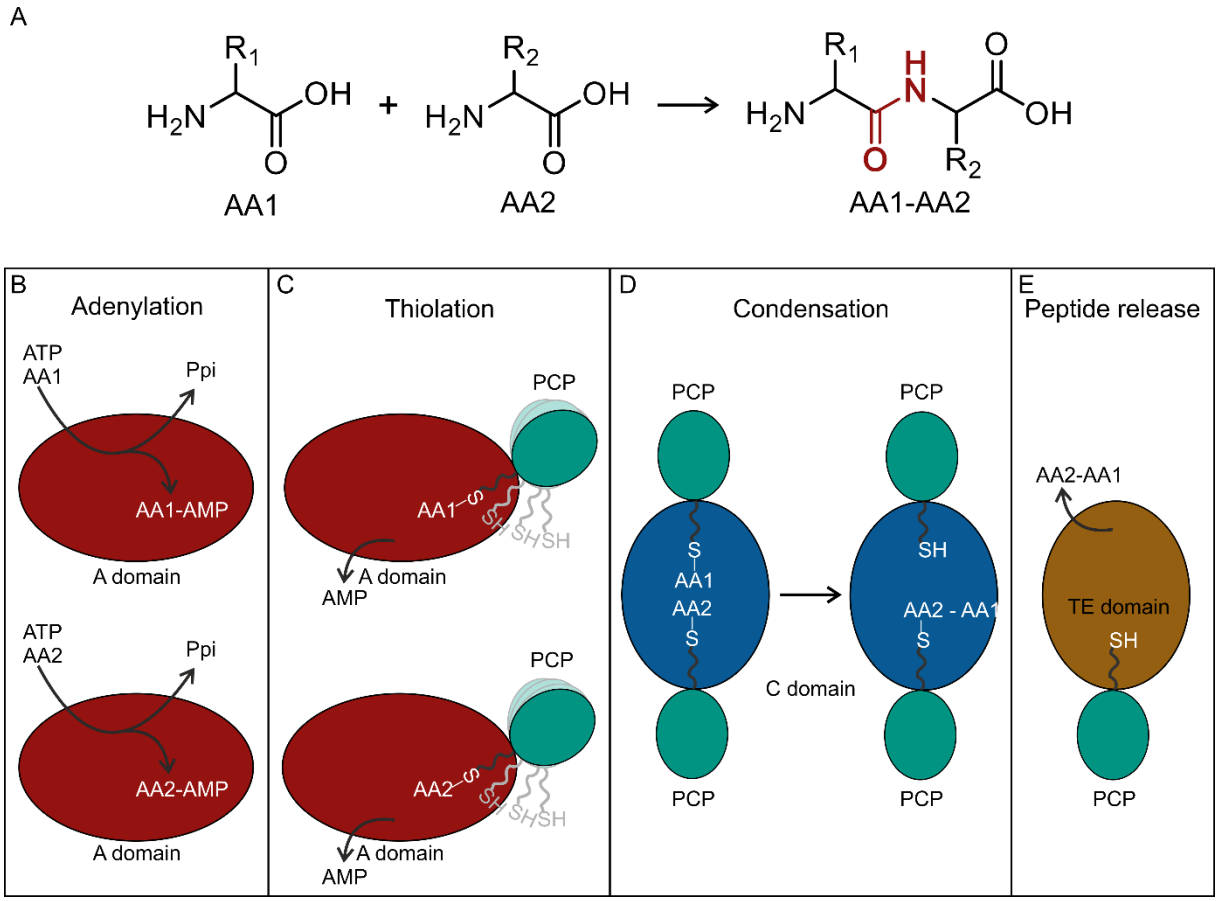


Figure 4: Reaction scheme and mechanism of a NRPS catalyzed peptide bond formation. A: Amino acids 1 and 2 (AA1 and AA2) react to form an amide bond and the dipeptide AA1-AA2. B: Adenylation reaction to activate the amino acids AA1 and AA2 occurs in the adenylation domain (A domain) by consuming ATP and releasing pyrophosphate (Ppi). C: Thiolation reaction to bind the activated amino acids to the phosphopantetheine arm of the PCP domain. AMP is released D: Condensation reaction between the two bound amino acids catalyzed by the condensation domain (C domain). E: product release catalyzed by thioesterase domain (TE domain).

At the end of the peptide extension process, a C-terminal thioesterase (TE) domain catalyzes peptide release (*31*) (Figure 4E). TE domain hydrolyses bound intermediates to carboxylic acids and regenerates the phosphopantetheine arm on the PCP (*30*).

As it has been shown for the 18 module peptaibol synthetase, NRPSs can reach enormous protein size and complexity. Studies of full-length protein structures and dynamics, inter- and intradomain interactions, and understanding the mechanisms of secondary metabolite synthesis are limited. In addition, the development of short constructs that cover only specific modules and domains may encounter problems with protein solubility and/or abrogation of functions because

the full-length enzyme consists of a single amino acid sequence and domain boundaries are not easily predicted. Therefore, monomodular NRPSs are suitable for structural and functional studies, due to their beneficial size. The monomodular NRPS Pys (pyreudione synthetase) from *Pseudomonas fluorescens,* produces amoebicidal pyreudiones, consisting of a proline and 3-oxo fatty acids derivatives (*21*). The minuscule 142 kDa pyreudione synthetase is a combination of a starter and a termination module with a C-A-PCP-Te architecture (*21*).

For the biosynthesis of pyreudione, the amino acid L-proline is selected and activated by the A domain and transported to the C domain by the PCP (*21, 32*). Simultaneously, a 3-oxofatty acid binds to the C domain. Interestingly, the C domain in Pys is not located between two modules but is the first domain in the assembly line. The C domain, which is the first domain in the enzyme and has additional substrate selectivity properties, is a member of the $C_{starter}$ family (*33*). In the active site of the C domain, the histidine in the SHLVVDG motif catalyzes a condensation reaction between 3-oxofatty acid and L-proline. The condensation product is transported to the TE domain, where the Dieckmann cyclization reaction is catalyzed to yield the final product (*32*) (Figure 5).

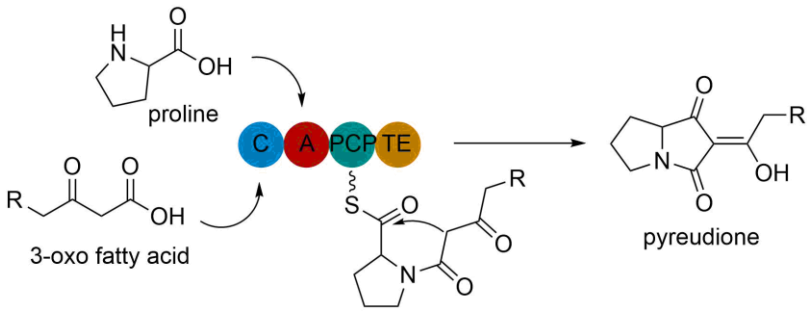


Figure 5: Pyreudione biosynthesis catalyzed by the monomodular NRPS Pys.

Even though Pys has a handily protein size, structural and *in vitro* functional studies have not been performed on this system. In addition, the structure, function and substrate selectivity of $C_{starter}$ domains have not been well studied to date. Pys provides an opportunity to determine the structure on full-length protein and isolated domains, study domain-domain interactions, and investigate the selectivity toward fatty acids in the $C_{starter}$ domain.

Apparently, NRPS can only produce peptides, as the name of this enzyme class implies. For the synthesis of nonpeptidic secondary metabolites, another multimodular enzyme is used: the polyketide synthases (PKS). Similar to NRPS, PKSs use simple building blocks to create complex molecules and the building blocks are transported through the enzyme by carrier proteins.

1.1.2. Polyketide synthases (PKS)

Polyketides is a group of natural products consisting of polyphenols, macrolides, polyenes, enediynes and polyethers (*34*). Beside their roles in microbial communication they also function as pigments, virulence factors or for defense (*35–37*). The biosynthesis of polyketides is catalyzed by the multimodular polyketide synthases (PKS) that consumes precursors, such as acetyl-coenzyme A (CoA) and malonyl-CoA and use them for chain extension (*38*). Malonyl building block selection occurs in an (malonyl)acyl transferase (MAT/AT) and an acyl carrier protein (ACP) with an attached phosphopantetheine arm serves as an anchor for the growing chain. The AT domain in PKS can be compared to the A domain of the NRPS. However, the AT domain receives an activated substrate and only fulfill the role in substrate selection. The decarboxylative Claisen thioester condensation reaction occurs in the ketosynthase (KS) domain that couples an activated acyl starter unit with malonyl-CoA-derived extender units. Substrate translocation and coupling reaction are catalyzed by in total four residues (the active site cysteine, two universally conserved histidines, and a lysin) located in the active site of a KS (*39*). The three named domains (AT, KS and ACP) form the basis and one module in a PKS. This basic module can be repeated resulting in a multimodular, linear PKS. Thereby, the module number corresponds to the number of acetyl-CoA building blocks. To yield a fully saturated acyl backbone, ketoreductase (KR), dehydratase (DH) and an enoyl reductase (ER) can be placed after every chain elongation step (*40*). The reductive steps are optional, and they can be partly or fully omitted before the next round of elongation. At the end of the PKS assembly line, product release is catalyzed by the TE domain. The linearity of the enzyme, the module number, and the order of the modification enzymes (KR, DH and ER) make the product predictable by computational methods (*41*).

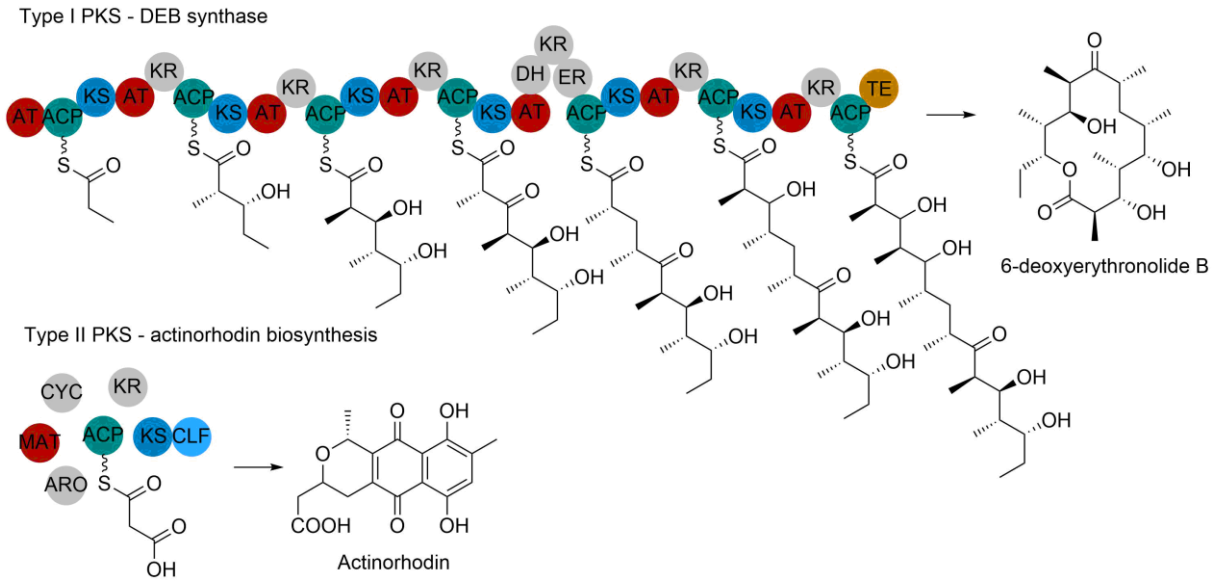


Figure 6: Examples for two PKS types and the produced natural products. 6-deoxyeryhronolide B (DEB) is the product of a linear type I PKS(42). DEB synthase consists of six modules that are built by acyltransferase (ATs), acyl carrier proteins (ACPs), keto synthases (KSs), and modification enzymes ketoreductase (KR), dehydratase (DH), enoyl reductase (ER). The aromatic compound actinorhodin is synthesized by a type II PKS, consisting of the heterodimeric KS and chain length factor (CLF) domains, malonyl acyl transferase (MAT), acyl carrier protein (ACP), and three modification enzymes ketoreductase (KR), aromatase (ARO), cyclase (CYC)(43). Acyltransferases are colored red, ketosynthase blue, acyl carrier proteins are shown in cyan, thioesterase domain is colored yellow. All modifications enzymes are colored grey.

Besides the linear modular architecture, also known as type I PKS, the synthases can also be found in multienzyme assembly architecture, also known as type II PKS. Examples for type I and type II PKS enzymes and their natural product are shown in Figure 6. This PKS type produces mainly aromatic polyketides by catalyzing iterative Claisen condensation reaction using acetate as the starter unit (*44*). The PKS enzyme assemblies are composed of five to eight stand-alone proteins, which form transient complexes with the carrier protein ACP. Thus, ACP is crucial for an efficient biosynthesis (*45*). Characteristic for type II PKS is the heterodimeric complex of the KS with the chain length factor protein (CLF). The role of the CLF protein is to control the size of the polyketide chain length by defining the size of the polyketide channel at the KS/CLF interface (*46*). This interface plays an important role in the protection of the ACP-bound intermediates (*45*).

1.1.3. NRPS-PKS hybrid system

Because of the modular structures of NRPS and PKS systems, a combination of both systems to produce peptide-polyketide hybrid compounds is not unusual. NRPS-PKS hybrid synthases are widely spread among microorganism and can be found in bacteria and fungi (*47*).

One example for a NRPS-PKS hybrid synthase is the biosynthesis of rhizoxin. Rhizoxin, a rice seedling blight causing compound, is produced by *Burkholderia rhizoxinica* bacterial (*48, 49*) endosymbiont of the fungus *Rhizopus microspores (49).* This compound consists of a 16-membered lactone ring connected to an oxazole ring by a long unsaturated chain (*50*) and is produced by a linear modular NRPS-PKS hybrid protein. The organization of the rhizoxin synthase is shown in Figure 7. During rhizoxin biosynthesis, malonyl decarboxylation or acyl channeling is catalyzed by the N-terminal AT-ACP-KS domain, yielding an acetic acid. Due to the N-terminal location of these three domains and the lack of essential conserved motif for elongation reaction in the KS domain, they form the loading module. The first integrated acyl tranferase domain in the loading module is named N-acyltransferase (GNAT). The succeeding NRPS module select the amino acid serin and catalyzed the condensation reaction between the acetic acid and the amino group of the selected serine. The NRPS module also contains a heterocyclization (HC) domain, which catalyzes intramolecular condensation reaction. The following oxidative reaction is catalyzed by the oxygenase (OXY) domain, which yields the methyloxazoline ring. Following the NRPS module, in total eleven PKS modules allow chain elongation via Claisen condensations forming the polyketide backbone. β-keto processing, such as keto reduction to alcohol and dehydration reaction to form unsaturated intermediates are catalyzed by KR and DH domains, respectively. Methylation reaction is catalyzed by MT domains. Downloading and lactonization of the polyketide chain by the TE domain yield the 16-membered macrolide that is further functionalized to form the final rhizoxin product. Interestingly, the rhizoxin NRPS-PKS hybrid synthase lacks integrated AT domains, besides the one in the loading module, which usually can be found in linear modular synthases. Instead, a free-standing AT didomain catalyzes malonyl transfer reaction onto ACP domains in trans (Figure 7). Another unique feature in the rhizoxin biosynthesis is the introduction of a C2 unit to an α,β-unsaturated thioester by an integrated

branching module (consists of a KS domain, a branching domain (B) and an ACP). Usually alkylation at β-position is catalyzed by freestanding enzymes that modify the β–keto group in trans. However, *in vitro* analysis of rhizoxin synthases showed that the vinylogous addition of the malonyl unit is catalyzed by the KS-B didomain in the branching module (*51*). Interestingly, the B module does not have any catalytic function but rather play a structural role in the branching module. Further functional analysis of the branching module revealed possible generation of lactone, lactam and glutarimide heterocycles by this module after cleavage of the KS-bound thioester (*52*). Structural investigation of the branching module has been a major challenge due to the high flexibility of the protein. A crystal structure of the KS-B didomain could be obtained by Bretschneider *et al*. (*51*). However, how the ACP interacts with the KS-B didomain remains unclear. In general, structural insight into KS-ACP interactions in modular PKSs has been limited.

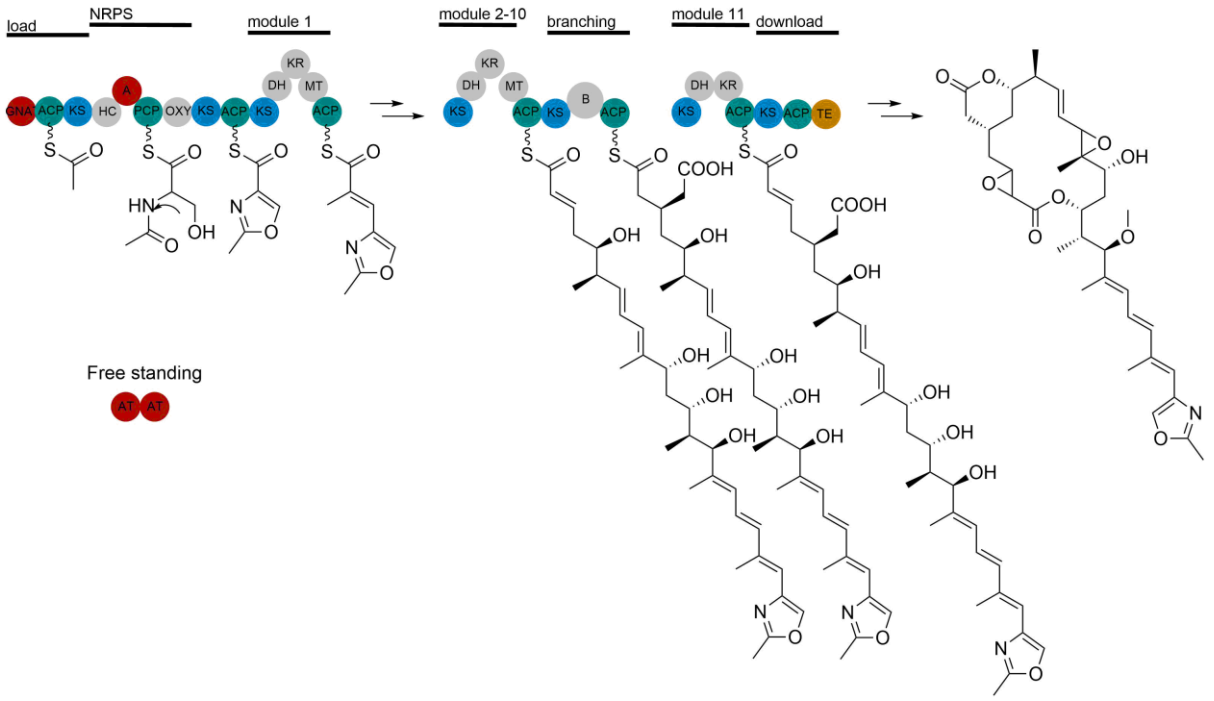


Figure 7: Organization of the rhizoxin synthase. Acyl/acetyl transferase and A domains (AT/GNAT/A) in red, ketosynthase, heterocyclization and condensation domains (KS/HC/C) in blue, carrier proteins (ACP/PCP) in cyan, thioesterase (TE) in yellow and modification enzymes such as ketoreductase (KR), dehydratase (DH), B domain (B), methyltransferase (MT), oxygenase domain (OXY) are in grey.

1.1.4. Carrier proteins – The heart of each synthase

Common to all NRPS and PKS systems is a main component, the carrier protein (peptide carrier protein PCP/ acyl carrier protein ACP). This small protein without catalytic function plays a central role in intra-enzyme communication and shuttling of intermediates through the entire enzyme. According to their function, carrier proteins in NRPS transport peptidyl intermediates have been called peptidyl carrier proteins (PCP), or thiolation (T) domain (*53, 54*), whereas carrier proteins in PKS systems that transport acyl intermediates are named acyl carrier proteins

(ACP). In addition to classification by function, integrated ACPs in linear PKS systems belong to the class of type I ACPs, while free-standing ACPs found in type II PKS belong to type II ACPs (*45*).

In general, carrier proteins form a helical bundle consisting of four with the two N-terminal helices longer than the two C-terminal helices (*55*) (Figure 8A). Helices I and II run (anti-)parallel to one another while the shorter helices III and IV pack against helices I and II and form the back of the domain (Figure 8B). In addition to the four main helices, some PCPs contain a small helix between helices I and II (*19*). The structure is stabilized by inter-helical hydrophobic interactions. Not only the structure remains similar throughout carrier proteins, the electrostatic surfaces show also similar distributions. Only type II ACPs has slightly higher positively charged surface compared to the other two types of carrier proteins (Figure 8C).

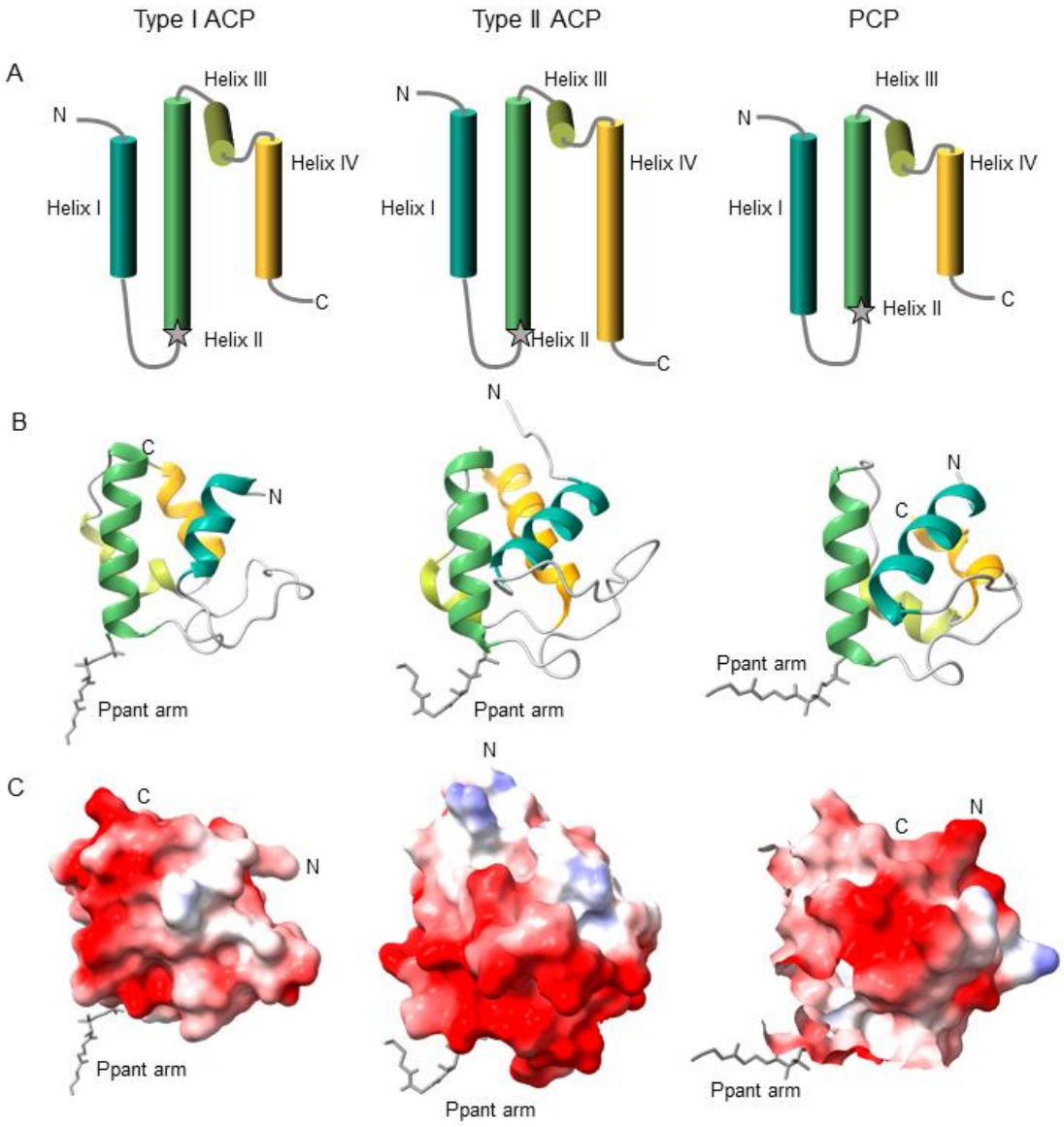


Figure 8: Schematic representation of carrier proteins from NRPS and PKS systems. A: Topology diagram of canonical carrier proteins. The reactive serin and the position of the phosphopantetheinyl modification reaction are marked with grey stars. B: structure of a type I ACP PDB: 5CZD, a type II ACP PDB: 2KOX, a PCP PDB: 5ISX. All structures show the protein in holo state and the 4'phosphopantethein (Ppant) arm is in grey attached to the end of helix II (green). C: electrostatic surface potential of carrier proteins. Positive charges in blue, negative charges in red, neutral charges in white.

To exhibit activity, carrier proteins must be converted from their inactive apo form to an active holo form by attaching the prosthetic 16 – 20 Å long 4'phosphopantethein (Ppant) arm of coenzyme A to a conserved serine residue(*56, 55*). This serine is located at the beginning of helix II within the conserved sequence motif Gx(D/H)S(L/I)(D/K)(*56*). The structures of activated carrier proteins are shown in Figure 8B. The phosphopantetheinyl modification reaction is catalyzed by proteins belonging to the superfamily of phosphopantetheinyl transferase (PPTase) (*57*). A well-known member of this protein family is the surfactin-type phosphopantetheinyl (Sfp) transferase from *B. subtilis* which exhibits broad substrate tolerance and can activate a variety of carrier proteins (*58–60*). It has been shown that the overall structure and conformation of carrier proteins remains unchanged during activation by PPTase, except for type II ACPs. In this specific ACP type, helix III moves closer to helix II upon addition of the PPant arm (*61*). Beside type II ACP, other carrier protein types are rigid (*19*).

During biosynthesis, carrier proteins bind a wide variety of substrates by covalently attaching them to the reactive thiol group of the Ppant arm. The bound compounds can then be transferred to catalytic domains of biosynthetic machineries. Integrated carrier proteins in linear assembly lines must interact with their neighboring domains. In the case of NRPS enzymes, PCPs inevitably interact with the A, C, and Te domains (*19*). Most of the interactions between PCP and neighboring NRPS domains involve PCP helix II and the preceding and subsequent loops (*57*). The complexes between PCP and neighboring domains are stabilized by both hydrophobic and ionic forces. The length and flexibility of Ppant arm perform the function of translocating substrates to the active sites of catalytic A, C and Te domains, where chemical transformation occurs. PCPs in complex with the catalytical domains of NRPS are shown in Figure 9.

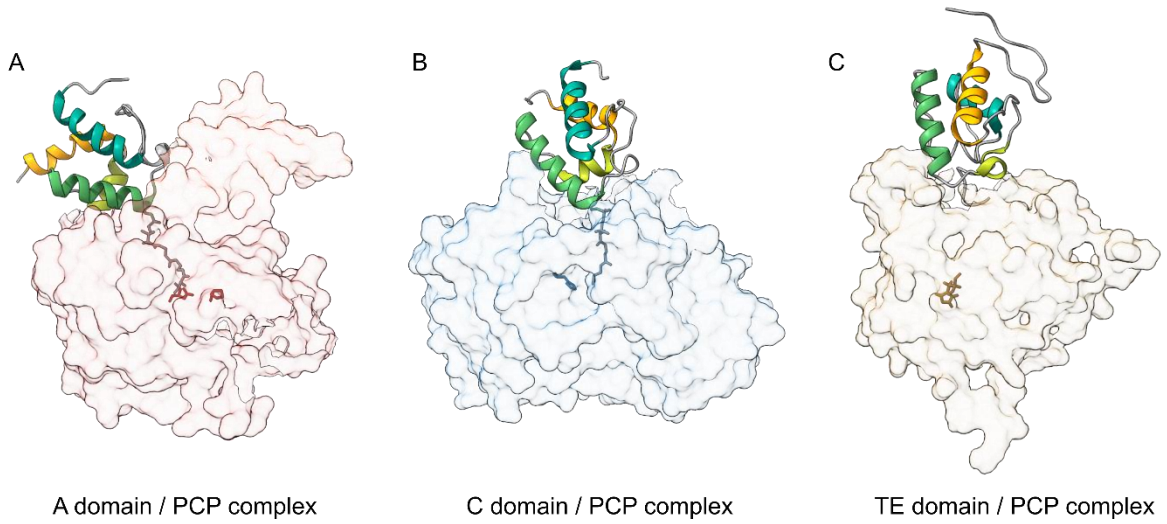


Figure 9: Structures of PCP in complexes with neighboring domains in NRPS. A: PCP-A domain complex (PDB: 5ES8(62)). Catalytic site residues are shown in red sticks. B: Donor PCP-C domain complex (PDB: 4ZXH(63)). Catalytic histidine is shown in blue stick. PCP-TE domain complex (PDB: 2ROQ(64)). Catalytic histidine is shown in yellow stick. Ppant arm, when available, is shown in grey stick. PCP domain is colored as in Figure 8.

In linear PKS systems, the integrated type I ACP interacts with AT, KS and TE domains during the biosynthesis of polyketides. However, not much is known about the complex formation between type I ACP and other catalytic domains. In only one crystal structure, the structural basis of the ACP-AT complex could be investigated, revealing that this complex is formed by the ACP residues

at the beginning of helix II and helix III, and residues in the neighboring loop (*65*). The binding interface is stabilized by salt bridges, hydrogen bonds and hydrophobic interactions.

The most structurally and functionally challenging system to investigate is the type II carrier protein. Since type II PKS consist of free-standing domains, directed and reproducible substrate synthesis is particularly difficult. In this case, the ACP not only enables efficient transfer of the growing product chain from one subunit to the next but is also decisively involved in maintaining the required order of substrate translocation. This is thought to be achieved by conformational changes of the ACP in different states: apo and holo state or acylation state of the Ppant arm. Upon substrate binding to the Ppant arm, the polyketide intermediates and a part of the Ppant arm are displaced into the hydrophobic cavity of the ACP (*66*). Once the ACP binds to an enzymatic partner, the Ppant arm flips over and relocates the bound intermediate into the catalytic center. Sequestration serves to protect the elongating metabolites and deliver them to the appropriate enzymes at the right time (*66*). However, the molecular details of this directing mechanism are still unclear.

In summary, carrier proteins play a dominant role in biosynthetic processes by shuttling substrates and substrate intermediates through the producing enzymes, i.e., NRPS or PKS complexes. Bio engineering strategies and the development of novel bioactive compounds rely on the functionality and directionality of such biosynthetic clusters, which in return rely on their carrier proteins to maintain the proper production pipelines. Therefore, understanding the structure, function, dynamic and regulation of carrier proteins is fundamental for understanding and potentially harnessing the power of nature's molecular producing machineries.

1.2. Aims and objectives

The production of signaling molecules is essential for microbial cell-to-cell communication. Bacteria produce such signaling molecule with the help of complex biosynthetic machineries, nonribosomal peptide synthases (NRPS) and polyketide synthases (PKS). These systems consist of modules, that are arranged like a chemical assembly line, thereby enabling the bacteria to produce a large number of structurally different products. To properly function, the individual components of the NRPS and PKS systems thus need to structurally interact in a predefined manner and the growing substrate chain needs to be trafficked in a pre-ordinated order from module to module. Evolutionary optimized domain-domain interactions are thus necessary for the proper function of multimodular enzymes, while domain disruptions lead to reduced turnover or non-functional enzymes. Therefore, structural, and functional investigations of modular synthases are required to distinguish between obligatory and facultative interactions. Since modular synthases can reach tremendous size and complexity, understanding a small monomodular system is a fundamental starting point. Pyreudione synthetase (Pys) is an optimal candidate for structural and functional analysis of monomodular NRPS.

The first objective of this work was to structurally and functionally characterize the monomodular NRPS Pys from *Pseudomonas fluorescens* using different structural biology techniques and functional assays. Pys synthesizes the lipidylated proline derivative pyreudione. In addition to selectivity toward the amino acid proline, Pys also shows selectivity toward 3-oxo fatty acids. Pys was expressed as a full-length protein, in an inactive apo state and in an activated holo state. In addition, isolated domains to investigate protein and domain structures, intra-domain communication, and protein/domain functionalities were produced. A particular focus was placed on the starter C domain from Pys as well as the functional homologous acylproline synthase (Aps) to understand the selectivity for the fatty acid head group (Figure 10A).

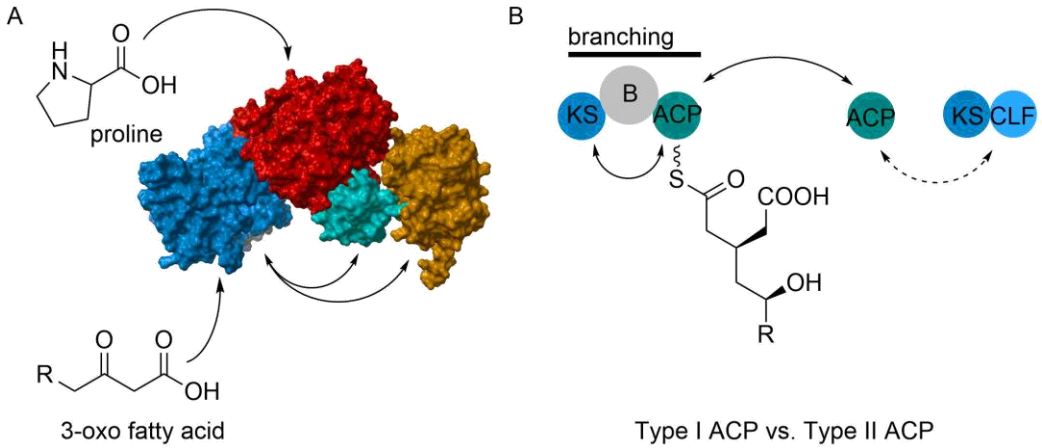


Figure 10: Overview of the aims of this chapter.

In a second project, the structural role of carrier proteins in secondary metabolite synthesis has been investigated. Carrier proteins do not possess catalytic properties but play a central role in the biosynthesis of complex bioactive molecules by NRPS and PKS systems. Carrier proteins interact with all catalytically active domains of the biosynthetic machinery and thus ensure proper substrate transport throughout the enzyme. Although carrier proteins are structurally similar in

all known synthases, there are important functional differences between carrier protein classes. Here, a type I and a type II ACP from the mixed PKS/NRPS rhizoxin synthase from *Burkholderia rhizoxinica* and a type II PKS with unknown function from *Gloeocapsa sp. PCC 7428* were structurally investigated by NMR spectroscopy. Interaction studies were performed to gain insight into ACP binding sites and binding affinities for the larger, biosynthetic modules (Figure 10B).

In summary, understanding NRPS synthesis mechanisms as well as the interaction between carrier proteins and neighboring domains, will shed light on the structural basis of the production of chemical mediators and help to distinguish between obligatory and facultative domain-domain interactions. Facultative interactions can be used as a starting point for potential cutting sites within a synthesis enzyme and by that create functional isolated domains. This may potentially allow the functional exchange of modules and thus enable the production of new molecules with yet unexplored biofunctionalities. Indeed, the need for novel bioactive compounds and antibiotics constantly rises due to the increasing resistance of clinical pathogens to existing antibiotics. Therefore, the engineering of NRPS and PKS for the production of novel bioactive compounds provides a promising route to develop new bioactive compounds by exchanging and switching modules in multimodular biosynthetic enzyme complexes.

1.3. Materials and Methods

1.3.1. Text editing, graph plotting, and figure design

This thesis was written with Microsoft Word (Office 365) and all figures were generated with CorelDRAW 2020 Graphics Suite (Corel corporation). All Graphs were plotted in Microsoft Excel (Office 265) and Origin Pro 8 (OriginLab). Structures of biological macromolecules were visualized and rendered with UCSF ChimeraX 1.2.5 (UC San Francisco) using coordinate files obtained from protein data bank (PDB). Molecule structures were drawn using ChemDraw professional 16.0 (PerkinElmer).

1.3.2. Laboratory equipment

All laboratory equipment and analytical measurements are listed in Table 1.

Table 1: Standard laboratory equipment.

Name	Description	Manufacturer
Avanti J-15R	Tabletop centrifuge for 10-50 ml volumes	Beckman Coulter
JS-4.750	Swing out rotor for Avanti J-15R	Beckman Coulter
Avanti JXN-26	Centrifuge	Beckman Coulter
JA 25.50	Fixed-angle rotor for, tube volume 50 ml	Beckman Coulter
JLA 8.100	Fixed-angle rotor for, tube volume 1 l	Beckman Coulter
Allegra X-20R	Centrifuge	Beckman Coulter
Conical C0650	Fixed-angle rotor for Allegra X-20R, tube volume 50 ml	Beckman Coulter
Haraeus Fresco 21	Tabletop centrifuge for 1-2 ml volumes	Fischer Scientific
neoMix 7-0921	Thermoblock for 1-2 ml tubes	neoLabLine
PowerPacBasic	Power supply for gel electrophoresis	BioRad
Mini-PROTEAN Tetra cell	SDS-PAGE chamber and equipment	BioRad
Amersham Imager 680	Blot and gel imager	Cytiva
INCU-Line	Incubation chamber	VWR
Multitron standard	Shaking incubator	Infors HT
Sonoplus MS73	Sonicator	Bandelin
HPL6	High pressure cell disruptor	Maximator
Econo columns	Gravity flow column	BioRad
NGC Quest 10 plus	Chromatography system	BioRad
BioFrac fraction collector	Fraction collector for NGC Quest 10 plus	BioRad
HiLoad 16/60 Superdex 200 pg	Size exclusion chromatography column, column volume 120 ml	Cytiva
HiLoad 16/60 Superdex 75 pg	Size exclusion chromatography column, column volume 120 ml	Cytiva
Superdex 200 10/300 GL	Size exclusion chromatography column, column volume 24 ml	Cytiva
FiveEasy plus	pH-meter	Mettler Toledo
GenPure Pro	Water purification system	Thermo Scientific

NanoDrop One	UV-Vis spectrophotometer	Thermo Scientific
Rotilabo-single-use cells, PS	Spectophotometer cuvettes	Roth
J-1500	Circular dichroism spectrometer	Jasco
110 – Macro cells	Cuvettes for CD spectrometer	Hellma Analytics
FluoroMax	Spectral fluorimeter	Horiba Scientific
119F – Cells for magnetic stirrers	Cuvettes for fluorimeter	Hellma Analytics
AVANCE 600 MHz	Nuclear magnetic resonance spectrometer	Bruker

1.3.3. Chemicals

Chemicals were purchased from Roth (Karlsruhe), Sigma-Aldrich (Munich), New England Biolabs (Frankfurt), Avanti Polar Lipids (Alabama), and VWR (Darmstadt). Chemicals from other manufacturers are listed in Table 2.

Table 2: Chemical list

Name	Manufacturer
Diisobutylen maleic acid copolymer (DIBMA)	Anatrace and as a gift from Prof. Sandro Keller, University Graz
DNase I	AppliChem GmbH
Doxorubicin	Biomol GmbH
Lauryl maltose neopentyl glycol (LMNG)	anatrace
Styrene maleic acid copolymer (SMA) SMALP 300 (SL25010P), SMALP 200 (SL30010P), SMALP 140 (SL40005P)	Orbiscope
Glycerol D8	Cambridge Isotope laboratories
^{15}N ammonium chloride	Eurisotop
^{13}C6 glucose	Cambridge Isotope laboratories

1.3.4. Consumables

Consumables which were used in this thesis are listed in Table 3.

Table 3: List of used consumables.

Name	Description	Manufacturer
Nickel-NTA-Agarose	Affinity column stationary phase	Qiagen
Vivaspin-6 & Vivaspin-20	Centrifugal concentration of proteins with MWCO of 30 and 50, 100 kDa	Satorius
Amicon Ultra-0,5 Amicon Ultra - 4 and Amicon Ultra-15	Centrifugal concentration of proteins with MWCO of 30 and 50 kDa	Merck Milipore
ZelluTrans	Protein dialysis with MWCO 12 kDa	Roth
6x-His Tag Antibody (MA1-21315-HRP)	6x-His Tag Monoclonal Antibody (HIS.H8), HRP for Western Blot	Thermo Fischer Scientific
Roti® Lumin	Solution for chemiluminescence detection of peroxidase coupled molecules	Roth
Roti PVDF	PVDF membrane for western blot	Roth
E.Z.N.A Plasmid Mini Kit	Isolation of plasmid DNA from *E. coli*	Omega Bio-Tek
Precision Plus Protein Dual Color Standard	Molecular marker for SDS-PAGE analysis	BioRad

1.3.5. Expression and purification of Pys proteins and domains from *Pseudomonas fluorescens*

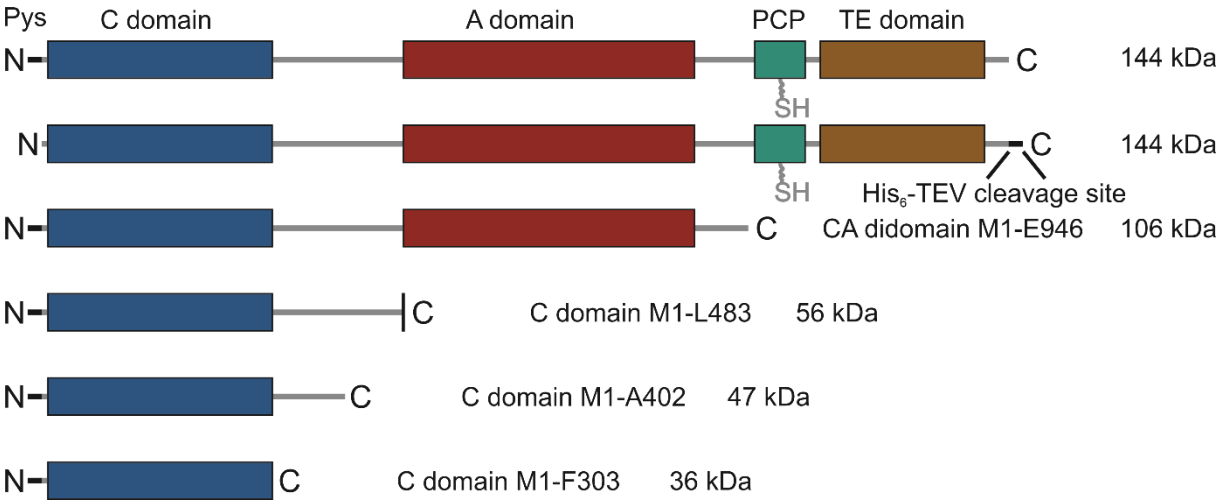


Figure 11: Overview of different Pys constructs from Pseudomonas fluorescens. His_6-tag and TEV cleavage site: black, C domain: blue, A domain: red, PCP: cyan, TE domain: yellow, linker and Ppant arm: grey. Domain boundaries were predicted by AntiSMASH.

All constructs were designed and cloned into pET28a vectors by _____ and _____ (_____).

Constructs in apo state or domains that did not contain the PCP domain were expressed in *E. coli* BL21-Gold (DE3) grown in lysogeny broth (LB) supplemented with 50 ug/ml kanamycin. Cells were grown to optical density at wavelength 600 nm (OD_{600}) = 0.4 – 0.6 at 37 °C, were induced with 500 µM isopropyl-*β*-D-thiogalactopyranoside (IPTG) and were moved to 20 °C for another 16 h incubation. After harvest, cells were stored at -20 °C until further use. Protein purifications were carried out at 4 °C. Harvested cells were resuspended in 50 mM sodium phosphate (NaPi) buffer at pH 8, 300 mM NaCl, 1 mM benzamidine (BA), protease inhibitor mix (1:1000 dilution) a spatula tip of lysozyme, DNase, RNase, and homogenized under high pressure using Maximator HPL6. Cell debris were removed by centrifugation (10.000 x *g*, 30 min, 4 °C) and the supernatant was mixed with nickel nitrilotriacetic acid (Ni-NTA) beads for 1 h on an end-over-end rotator. The mixture was transferred to a gravity column and was washed with 10 column volume (CV) 50 mM NaPi pH 8, 300 mM NaCl, 20 mM imidazole, 5 CV 50 mM NaPi pH 8, 300 mM NaCl, 50 mM imidazole. Proteins were eluted with 50 mM NaPi pH 8, 300 mM NaCl, 500 mM imidazole in 1 CV steps. Since the tag cleavages did not give high protein yields, fractions containing proteins were concentrated to <5 ml and purified via size exclusion chromatography (SEC) using HiLoad prep grade 16 / 60 Superdex 200 column and 50 mM NaPi pH 8, 300 mM NaCl. For long time storage, the proteins were supplemented with 10% glycerol, flash-frozen in liquid nitrogen and stored at -80 °C until further use.

To express holo proteins a two-plasmid transformation including pET28a coding for Pys proteins and pML5-T7 coding for T7 RNA polymerase were carried out into HM0079 *E. coli* strain with genomic integrated 4′-phosphopantetheine transferase gene *sfp (67, 68)*. Due to the lack of T7 RNA polymerase in HM0079 strain the plasmid pML5-T7 is needed to activate the T7 promotor

on pET28a plasmids. Protein expression was activated by the addition of 500 µM IPTG to LB media at a culture OD_{600} = 0.4 – 0.6. Cells were harvested by centrifugation and protein purification was performed with the same procedure as for apo proteins.

1.3.6. Expression and purification of the surfactin-type phosphopantetheinyl transferase Sfp from *Bacillus subtilis*

The plasmid encoding the C-terminal His_6-tagged surfactin-type phosphopantetheinyl transferase R4-4 Sfp from *Bacillus subtilis* was a kind gift from the Gulder group (TU Dresden). After transformation into *E. coli* BL21-Gold (DE3), the cells were grown in terrific broth (TB), supplemented with 1 mg/ml ampicillin at 37 °C until OD_{600} = 0.9 was reached. Cells were moved to 18 °C and induced with 500 µM IPTG at a OD_{600} = 1.2. Protein expression was performed for 16 h and cells were harvested via centrifugation. The yielded cell pellet was resuspended in 50 mM NaPi pH 8, 300 mM NaCl, 10 mM imidazole, 10% glycerol, protease inhibitor mix (1:1000 dilution) a spatula tip of lysozyme, DNase, RNase and cells were lysed using sonification. Cell debris were removed by centrifugation (10.000 x *g*, 30 min, 4 °C) and the supernatant was mixed with Ni-NTA beads for 45 min on an end-over-end rotator. The mixture was transferred to a gravity column and the resins were washed with 10 CV 50 mM NaPi pH 8, 300 mM NaCl, 10 mM imidazole, 10% glycerol and 5 CV 50 mM NaPi pH 8, 300 mM NaCl, 40 mM imidazole, 10% glycerol. Proteins were eluted with 50 mM NaPi pH 8, 300 mM NaCl, 250 mM imidazole, 10% glycerol in 1 CV steps. Fractions containing R4-4 Sfp were concentrated <5 ml and were loaded onto SEC column HiLoad prep grade 16 / 60 Superdex 75 column eluted with 50 mM NaPi pH 8, 300 mM NaCl, 10% glycerol. After SEC, samples were concentrated to required concentrations, flash-frozen in liquid nitrogen and stored at -20 °C until further use.

1.3.7. Expression and purification of the acyl carrier protein from *Gloeocapsa sp. PCC 7428*

The gene fragment encoding the *Gloeocapsa sp. PCC 7428* ACP was designed and cloned into pET28a vector by ____ lab (____). Plasmids containing an N-terminal His_6-tagged ACP with a thrombin cleavage site were transformed into *E. coli* BL21-Gold (DE3) cells and were grow in LB medium with 50 µg/ml kanamycin at 37 °C. At a OD_{600} of 0.4 – 0.6 cells were induced with 312 µM IPTG and transferred to 18 °C for another 16 h of incubation. Cells were harvested via centrifugation and stored at -20 °C until further use. Protein purification was performed at 4 °C. Cell pellets were resuspended in 50 mM NaPi pH 7.6, 300 mM NaCl, 10 mM imidazole, 10% glycerol, protease inhibitor mix (1:1000 dilution) a spatula tip of lysozyme, DNase, RNase and cells were disrupted by sonication. Cell debris were removed by centrifugation (10.000 x *g*, 30 min, 4 °C) and the supernatant was mixed with Ni-NTA beads for 1 h on an end-over-end rotator. The mixture was transferred to a gravity column and was washed with 10 CV 50 mM NaPi pH 7.6,

100 mM NaCl, 10 mM imidazole, 10% glycerol, and 10 CV 50 mM NaPi pH 7.6, 100 mM NaCl, 30 mM imidazole, 10% glycerol. Proteins were eluted stepwise with 50 mM NaPi pH 7.6, 100 mM NaCl, 250 mM imidazole, 10% glycerol. BCA assay was performed to determine fractions that contained proteins. These fractions were pooled together and concentrated to <5 ml. Purification via SEC using HiLoad prep grade 16 / 60 Superdex 75 column and 50 mM NaPi pH 7.6, 100 mM NaCl, 10% glycerol was carried out. After SEC, samples were concentrated to required concentrations, flash-frozen in liquid nitrogen and stored at -20 °C until further use.

Rhizoxin ACP and KSB proteins from *Burkholderia rhizoxinica* were expressed and purified by _____ (_____ Lab, _____).

All plasmids, which have been used in this chapter are listed in Table 4.

Table 4: Plasmids encoding Pys proteins, domains, and Gloeocapsa ACP are listed with their respective plasmid and protein properties. Physicochemical properties of Pys (NCBI Reference Sequence: WP_064118616.1) and Gloeocapsa ACP (from Gloeocapsa sp. PCC7428, complete genome: 537422 – 5374466) were calculated using the ProtParam tool from the ExPasy website (https://web.expasy.org/protparam/).

Plasmid number	**Insert**	**Tag / cleavage site**	**Vector backbone**	**Resistance**	**Origin**	**No. amino acids**	**PI**	**$\varepsilon_{280\ nm}$ [$M^{-1}\ cm^{-1}$]**
P557	Pys CA didomain M1-E946	N-term. His_6 / TEV	pET28a(+)	Kanamycin	_____ lab (_____)	967	5.94	66085
P674	Pys full-length	N-term. His_6 / TEV	pET28a(+)	Kanamycin	_____ lab (_____)	1318	5.65	94755
P675	Pys full-length	C-term. His_6 / TEV	pET28a(+)	Kanamycin	_____ lab (_____)	1318	5.65	94755
P696	Pys C-domain M1-A402	N-term. His_6 / TEV	pET28a(+)	Kanamycin	_____ lab (_____)	423	6.08	29255
P697	Pys C-domain M1-L483	N-term. His_6 / TEV	pET28a(+)	Kanamycin	_____ lab (_____)	504	5.89	36705
P698	Pys C-domain M1-F303	N-term. His_6 / TEV	pET28a(+)	Kanamycin	_____ lab (_____)	324	6.17	26150
P679	Gloeocapsa ACP	N-term. His_6 / Thrombine	pJH1	Kanamycin	_____ lab (_____)	100	6.12	0

1.3.8. Sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE)

Purified proteins were analyzed by sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE). Thereby, proteins were denatured and unraveled according to the protein size. Tris-glycine buffer was used as running buffer (10x buffer contains 0.25 M tris-Cl pH 8.3, 1.92 M glycine, 35 mM SDS). Protein samples were mixed with 4 x sample buffer (200 mM tris pH 6.8, 400 mM DTT, 8% (w/v) SDS, 0.4% (w/v) bromophenol blue, 40% glycerol). Samples containing bacterial cells were mixed with 100 µl 4 x sample buffer. For a complete denaturation, samples were heated for 10 – 15 min at 95 °C. 5 – 15 µl of the samples were loaded onto 10 – 15% gels and run at 100 – 180 V for up to 2 h. The compositions of the different gels are listed in Table 5.

SDS-PAGE gels were stained with Coomassie (1 g/l Coomassie brilliant blue R250, 2% (v/v) phosphoric acid, 40% ethanol) and destained with 30% (v/v) ethanol, 2% phosphoric acid solution.

Table 5: Composition of the stacking and running gel to perform 4 gels for SDS-PAGE

Ingredients	Stacking gel	Running gel		
		10%	12%	15%
ddH_2O	4.6 ml	7.9 ml	6.6 ml	4.6 ml
1.5 mM tris pH 8.8	-	5 ml	5 ml	5 ml
1 mM tris pH 6.8	630 µl	-	-	-
30% acrylamide	830 µl	7.7 ml	8 ml	10 ml
10% (w/v) SDS	50 µl	200 µl	200 µl	200 µl
10% (w/v) APS	50 µl	200 µl	200 µl	200 µl
TEMED	5 µl	20 µl	20 µl	20 µl

1.3.9. *In-vitro* synthesis of pyreudiones

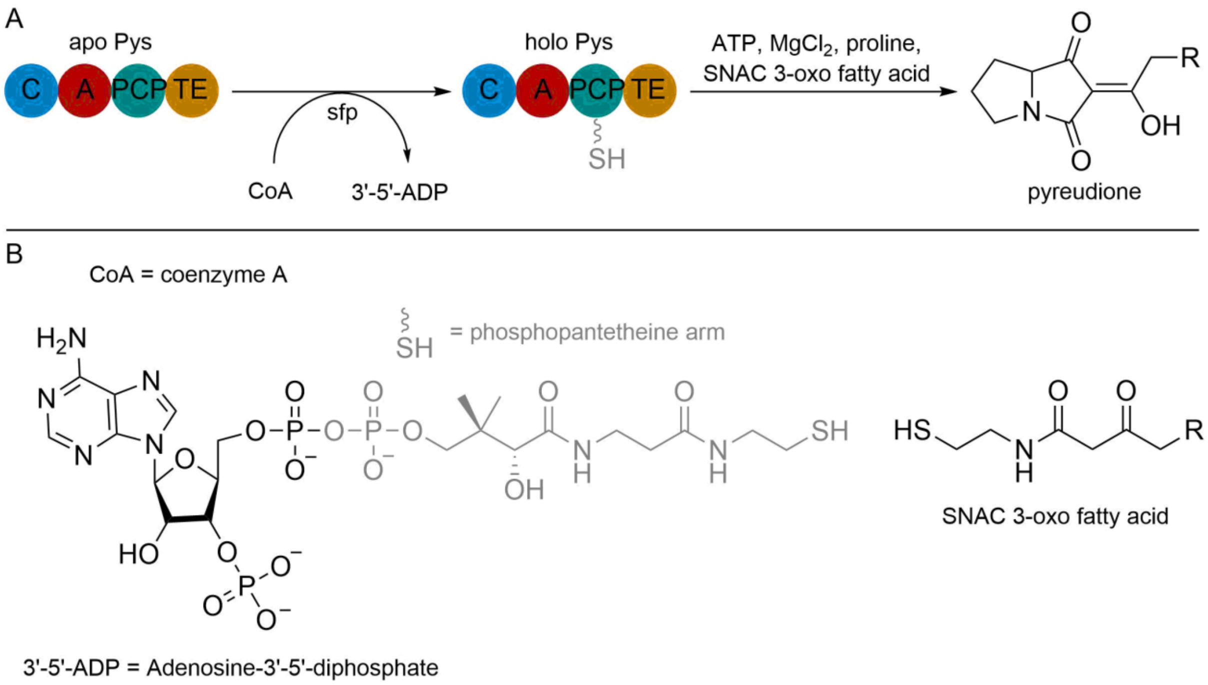


Figure 12: Reaction scheme for an in-vitro synthesis of pyreudiones. A: Pys from Pseudomonas fluorescens in apo state is activated by the enzyme phosphopantetheinyl transferase (Sfp) from Bacillus subtilis consuming CoA and forming holo Pys. Holo Pys catalyzes the pyreudione synthesis with addition of ATP, $MgCl_2$, proline and an activated 3-oxo fatty acid (SNAC), B: Molecule structure of CoA, phosphopantetheine arm, 3'5-ADP and SNAC 3-oxo fatty acid.

For *in-vitro* biosynthesis of nonribosomal peptides, purified apo nonribosomal peptide synthetases were activated with the R4-4 Sfp forming holo proteins. The R4-4 Sfp has three mutations, L28E, T44E, C77Y, that catalyzes carrier protein modification 300-fold higher than wild type Sfp (*59*). All proteins were rebuffered in 100 mM hepes pH 7.8 and 1.5% glycerol before use in this assay. To a final volume of 100 µl 2 mM Coenzyme A, 5 µM R4-4 Sfp, 10 µM apo Pys, 10 mM ATP, 7.5 mM $MgCl_2$, 2 mM proline, 2 mM SNAC 3-oxo fatty acid in DMSO, 10 mM nicotinamide adenine dinucleotide phosphate (NADPH) and 2 mM flavin adenine dinucleotide (FAD) were added. NADPH is important for a hydrolytic off-loading of the substrate in the TE

domain and FAD is relevant for dehydrogenase reactions (*69*). Both components belong to a standard in-vitro NRPS buffer but could be removed in the case of pyreudione synthesis. The mixture was incubated for 16 h at 28 °C, 300 rpm. With the addition of 100 µl MeOH all proteins were precipitated, and the supernatant was filtered through a polyamide membrane. Pyreudiones were detected by LC-MS.

1.3.10. Crystallization of condensation domains

Pyreudion synthetase (Pys) condensation domain from *Pseudomonas fluorescens* and acylprolin synthetase (Aps) from *Pectobacterium betavasculorum* were expressed and purified by _____ (_____ Lab, _____). Pys C domain in 20 mM hepes pH 6.8, 150 mM NaCl, 5% glycerol was rebuffered into 50 mM hepes pH 6.8, 150 mM NaCl and concentrated to 22.5 mg/ml. Aps C domain in 50 mM hepes pH 6.8, 150 mM NaCl was concentrated to 11.4 mg/ml. Both proteins were screened on INDEX HT Screen, testing 10 mg/ml, 5 mg/ml and 3 mg/ml for Pys and 11.4 mg/ml, 5.6 mg/ml and 3.6 mg/ml for Aps, respectively. Crystallization trials were performed in 3 drops 96-well sitting-drop plates. Crystals were formed after 3 days with Aps C domain and 7 days Pys C domain at 20°C.

1.3.11. Nuclear magnetic resonance spectroscopy

In biological macromolecules determination of high-resolution structures is a key to understand molecular details and associated functional mechanisms or dysfunctions. The most extensively applied methods in structural biology include X-ray crystallography, cryo-electron microscopy (CryoEM) and nuclear magnetic resonance (NMR) spectroscopy. X-ray crystallography and CryoEM resolve protein structures with near atomic resolutions, with no limitations on protein sizes. Advantages of these two techniques have been summarized in literature (*70*). However, these two techniques cannot be performed under physiological conditions: for X-ray crystallography, the molecule of interest is packed into highly ordered crystals, whereas for CryoEM the sample is frozen in vitrified ice. Only NMR measurements can be performed in solution, allowing structural and dynamic studies under near-natural conditions. In addition to protein structure determination, solution NMR provides information on protein dynamics, enzyme kinetics, oligomeric states, ligand binding and is a key method in the research on intrinsically disordered proteins (*71–73*). A major disadvantage of NMR spectroscopy is the size limitation of usually less than 100 kDa (*74*).

In general, NMR spectroscopy can be applied to all nuclei with odd proton and neutron numbers. For biomolecular applications, the nuclei commonly used are ^{1}H, ^{13}C, ^{15}N, ^{31}P, and ^{19}F. Although the ^{1}H nucleus is naturally very abundant in proteins, other nuclei must be enriched, can only be found in substrates, or must be selectively labeled (*75*). The combination of ^{1}H, ^{13}C, and ^{15}N nuclei is commonly used in backbone resonance assignment, dynamic studies, and structural calculations.

The assignment of backbone resonance provides insights into structural and dynamic features encoded in the amino acids by their position in the protein sequence. Typical backbone assignments experiments such as HNCA, HNCACB, and HN(CA)CO show correlation between C_α, C_β, and carbonyl C's that allow identification of amide resonance chains.

To enrich ^{13}C and/or ^{15}N nuclei in proteins, bacteria were grown in the minimal medium M9 supplemented with ^{13}C and ^{15}N sources. This medium contains all essential trace elements and salts that is needed for a bacteria growth supplemented with ^{13}C-glucose and/or ^{15}N-ammonium chloride for the uniform labeling (Table 6). Bacteria will synthesize labeled amino acids during growth and incorporate them into protein synthesis.

Table 6: 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 $MgCl_2$ (1 M stock) 10 µM $FeCl_3$ (10 mM stock)
Trace elements dissolved in 10 ml H_2O	2 mg $CaCl_2$-dihydrate 2 mg $ZnSO_4$-heptahydrate 2 mg $MnSO_4$-monohydrate 1 mg biotin 50 mg thiamine and niacin
Magic Mix	1 ml of Centrum® vitamin tablet dissolved in 20 ml H_2O
Isotopes	1,00 g $^{15}NH_4Cl$ 2.00 g $^{13}C_6$-glucose (or 4.00 g $^{12}C_6$-glucose for only ^{15}N-labeling)

^{15}N and $^{13}C/^{15}N$ labeled *Gloeocapsa* ACP were generated by expression in *E. coli* grown in isotope supplemented M9 medium. First, transformed *E. coli* cells were grown in LB medium to an OD_{600} = 0.4 – 0.6 and then centrifuged (room temperature, 5000 x *g*, 10 min). The LB medium was discarded, and the cells were resuspended in M9 medium supplemented with the respective antibiotic and IPTG. All subsequent steps follow the same protocols as for non-labelled protein expressions.

To label specific residues for the backbone assignment, a medium with defined amino acid, nucleobase and salt contents were used (Table 7).

Table 7: Defined medium for selective isotope labeling of amino acids. #Isotope-labeled lysine were added instead of unlabeled lysine.

Component	Ingredients per L medium
Amino acids	0.50 g Ala 0.40 g Arg 0.40 g Asp 0.05 g Cys 0.40 g Gln 0.65 g His 0.23 g Ile 0.23 g Leu 0.42 g #Lys-hydro-Cl 0.25 g Met 0.13 g Phe 0.10 g Pro 2.10 g Ser 0.23 g Thr 0.17 g Tyr 0.23 g Val
Nucleobases	0.50 g adenine 0.65 g guanosine 0.20 g thymine 0.50 g uracil 0.20 g cytosine
Salts	0.50 g NH_4Cl 1.50 g sodium acetate 1.50 g succinic acid 10.5 g K_2HPO4 0.85 g NaOH 4 mM $MgSO_4$ (1 M stock) 10 µM $FeCl_3$ (1 mM stock) 100 ml 20% glucose
Trace elements dissolved in 10 ml H_2O	2 mg $CaCl_2$-dihydrate 2 mg $ZnSO_4$-heptahydrate 2 mg $MnSO_4$-monohydrate 1 mg biotin 50 mg thiamine 50 mg niacin
Magic Mix	1 ml of Centrum® vitamin tablet dissolved in 20 ml H_2O

Similar to $^{13}C/^{15}N$ labeling, transformed cells were grown in LB until OD_{600} reached expression densities. Cells were centrifuged and resuspended in selective medium, induced and grown following the protocols for non-labelled proteins.

The ^{31}P nucleus can be found in lipids, co-factors, and nucleotides and ^{31}P NMR can be used to monitor lipid and nucleotide binding on proteins. Substrate binding can result in changing of the chemical environment of the substrate that influences the peak position in ^{31}P NMR spectra. Further, in the case of an adenosine triphosphate (ATP) consuming proteins, ^{31}P NMR is useful to track the conversion of ATP into adenosine diphosphate (ADP) or adenosine monophosphate (AMP) using the substrate and product as reporters. To study local structural and dynamic properties, site-specific labelling with the ^{19}F nucleus is used. This nucleus does not occur

naturally in proteins, making the displayed signal to be easily assigned to the labelling. The power of ^{19}F nucleus is reflected in its sensitivity to its chemical environment that makes ^{19}F NMR spectroscopy a powerful technique to study structure and dynamic of proteins, even at molecular weight above 30 kDa.

To label the protein with ^{19}F sources two different techniques can be used: labeling of cysteines with ^{19}F molecules or ^{19}F-tryptophan labeling during protein expression. In this work, KSB proteins were labelled with 6-^{19}F-tryptophan by Dr. Maria Dell, using selectively labelled media (Table 7). 45 mg 6-^{19}F-Trp racemate + 5 mg Trp were used for 1 l of media. Spectra were recorded by _____.

Similar to $^{13}C/^{15}N$ labeling, transformed cells were grown in LB until OD_{600} reached expression densities. Cells were centrifuged and resuspended in selective medium, induced and grown following the protocols for non-labelled proteins.

All experiments were carried out on Bruker AVANCE NMR spectrometers. Standard Bruker pulse sequences were used for recording the spectra. Details to the recorded spectra are listed in Table 8 to Table 10.

Table 8: NMR parameters to determine ATP hydrolysis rate of the CA-didomain. Conc. – concentration, NS – number of scans, D1 – relaxation delay, F1 – channel 1, SW – spectral width.

Experiment	Sample	Conc. [µM]	Bruker pulse sequence	Spectrometer frequency [MHz]	NS	D1 [s]	F1		
							Nucleus	SW [ppm]	No. of points
^{31}P-ATP hydrolysis	[^{31}P]-ATP + prolin 10 mM + CA domain 10µM	10 mM	zgdc30	600	1024	1.0	^{31}P	79.79	2048

Table 9: NMR parameters for Rhi ACP backbone assignment and titration experiments with KSB. Conc. – concentration, NS – number of scans, D1 – relaxation delay, F1 – channel 1, SW – spectral width.

Experiments	Sample	Conc. [µM]	Bruker pulse sequence	Spectrometer frequency [MHz]	NS	D1 [s]	F3			F2			F1		
							Nucleus	SW [ppm]	No. of points	Nucleus	SW [ppm]	No. of points	Nucleus	SW [ppm]	No. of points
HSQC	[^{15}N/^{13}C] Rhi ACP	614	fhsqcf3gpph	700	4	1.0				^{1}H	15.9	2048	^{15}N	29	256
HSQC-Trosy	[^{15}N/^{13}C] Rhi ACP	614	b_trosyetf3gpsi.2	700	8	0.3				^{1}H	15.9	2048	^{15}N	29	256
HNCO	[^{15}N/^{13}C] Rhi ACP	614	b_trhncogp3d.2	700	8	0.25	^{1}H	16.0	2048	^{15}N	29	64	^{13}C	10.0	64
HNCA	[^{15}N/^{13}C] Rhi ACP	614	b_trhncagp3d.2	700	8	0.25	^{1}H	16.0	2048	^{15}N	29	64	^{13}C	32.0	160
HNCACB	[^{15}N/^{13}C] Rhi ACP	614	b_trhncacbgp3d.2	700	8	0.3	^{1}H	16.0	2048	^{15}N	29	64	^{13}C	70.0	200
HNCACO	[^{15}N/^{13}C] Rhi ACP	614	b_trhncacogp3d.2	700	8	0.3	^{1}H	16.0	2048	^{15}N	29	64	^{13}C	10.0	64
HNCOCACB	[^{15}N/^{13}C] Rhi ACP	614	b_trhncocacbgp3d.2	700	8	0.3	^{1}H	16.0	2048	^{15}N	29	64	^{13}C	66.0	200
HSQC	[^{15}N] Rhi ACP	84	fhsqcf3gpph	600	8	1.0				^{1}H	18.1	2048	^{15}N	29	200
HSQC	[^{15}N] Rhi ACP + 80 µM KSB	84	fhsqcf3gpph	600	8	1.0				^{1}H	18.1	2048	^{15}N	29	200
HSQC	[^{15}N] Rhi ACP + 125 µM KSB	53	fhsqcf3gpph	600	16	1.0				^{1}H	18.1	2048	^{15}N	29	200
HSQC	[^{15}N] Rhi ACP + 135 µM KSB	45	fhsqcf3gpph	600	32	1.0				^{1}H	18.1	2048	^{15}N	29	200
HSQC	[^{15}N] Rhi ACP + 283 µM KSB	30	fhsqcf3gpph	700	312	1.0				^{1}H	18.1	2048	^{15}N	29	200
^{19}F	[^{19}F]-5Trp Rhi KSB	80	zg	599	10k	1.0							^{19}F	29	1024

^{19}F	[^{19}F]-5Trp Rhi KSB + 536 µM ACP	80	zg	599	10k	1.0								^{19}F	29	1024

Table 10: NMR parameters for Gloeocapsa ACP backbone assignment and structure determination experiments. Conc. – concentration, NS – number of scans, D1 – relaxation delay, F1 – channel 1, SW – spectral width.

							F3			F2			F1		
Experiments	**Sample**	**Conc. [µM]**	**Bruker pulse sequence**	**Spectrometer frequency [MHz]**	**NS**	**D1 [s]**	**Nucleus**	**SW [ppm]**	**No. of points**	**Nucleus**	**SW [ppm]**	**No. of points**	**Nucleus**	**SW [ppm]**	**No. of points**
HSQC	[^{15}N/^{13}C] Gloeo ACP	410	fhsqcf3gpph	600	4	1.0				^{1}H	18	2048	^{15}N	26	256
HSQC-Trosy	[^{15}N/^{13}C] Gloeo ACP	410	b_trosyetf3gpsi.2	600	16	0.3				^{1}H	18	2048	^{15}N	28	256
HSQC-Trosy	[^{15}N-Lys] Gloeo ACP	156	b_trosyetf3gpsi.2	600	8	0.3				^{1}H	18	2048	^{15}N	28	160
HNCO	[^{15}N/^{13}C] Gloeo ACP	410	b_trhncogp3d.2	600	8	0.3	^{1}H	14	1024	^{15}N	28	64	^{13}C	11	64
HNCACO	[^{15}N/^{13}C] Gloeo ACP	410	b_trhncacogp3d.2	600	32	0.3	^{1}H	14	1024	^{15}N	28	64	^{13}C	11	64
HNCA	[^{15}N/^{13}C] Gloeo ACP	410	b_trhncagp3d	600	8	0.3	^{1}H	14	1024	^{15}N	28	64	^{13}C	30	120
HNCACB	[^{15}N/^{13}C] Gloeo ACP	410	b_trhncacbgp3d.2	600	32	0.3	^{1}H	14	1024	^{15}N	28	64	^{13}C	66	180
HBHACONH	[^{15}N/^{13}C] Gloeo ACP	410	hbhaconhgp3d	600	8	1.0	^{1}H	18	2048	^{15}N	24	48	^{1}H	6	100
HNCCCONH	[^{15}N/^{13}C] Gloeo ACP	410	hccconhgp3d2	600	8	1.0	^{1}H	18	2048	^{15}N	23	48	^{1}H	6	128
HNCCCONH	[^{15}N/^{13}C] Gloeo ACP	410	hccconhgp3d3	600	8	1.0	^{1}H	16	2048	^{15}N	23	48	^{13}C	72	160
^{13}C-NOESY	[^{15}N/^{13}C] Gloeo ACP	410	noesyhsqcetgpsi3d	600	8	1.0	^{1}H	14	2048	^{13}C	72	148	^{1}H	11.6	200
NOESY-HSQC	[^{15}N] Gloeo ACP	375	noesyhsqcetf3gp3d	600	8	1.5	^{1}H	18	2048	^{15}N	24	84	^{1}H	12.6	276

1.4. Results & discussion

1.4.1. Expression and purification of Pys full-length proteins and *in-vitro* synthesis of pyreudione

NRPSs are multifunctional enzymes that are involved in microbial communication. They produce a wide range of valuable compounds with applications not only for microbial communication but also for medicine and agriculture. Thus, NRPS systems are important targets for bioengineering. This requires a detailed understanding of NRPS domain and module interactions and regulatory mechanisms. However, cloning and engineering of such biosynthetic assembly lines that can reach enormous sizes is difficult. Thus, analysis of monomodular NRPSs may help to overcome the difficulties of larger systems. The knowledge gained can then be applied to multimodular systems that operate according to the same basic principles.

The smallest NRPS discovered to date is the pyreudione synthetases Pys from *Pseudomonas fluorescens*. Functional studies show that its A domain selects and activates proline and proline derivatives, while the C domain displays selectivity towards 3-oxo fatty acids (*21, 32*). To investigate the *in-vitro* functionality of Pys, full-length proteins were designed by ____.

For the heterologous expression of Pys full-length constructs in the apo state, N-terminal or C-terminal His_6-tags separated from the NRPS by a Tobacco Etch Virus (TEV) protease cleavage were prepared. *E. coli* Bl21 gold (DE3) cells were used for protein expression. During protein purification, only ~10% of the His_6-tagged proteins could be cleaved by TEV protease at a molar protease to protein ratio of 1:15, resulting in a low yield of cleaved proteins. Therefore, the His_6-tag was not cleaved in subsequent analysis. In addition, it was decided to continue with the N-terminally His-tagged Pys for future studies, as this construct yielded higher amounts compared to the C-terminally His-tagged version (Figure 13A).

Nonetheless, using the expression and purification procedure outlined in Materials and Methods, full-length Pys proteins with N-terminal and C-terminal His_6-tag could be obtained at high purity and monodispersity as indicated by SDS-PAGE and SEC analysis (Figure 13A and B). The SEC profiles show the elution volume as expected for globular proteins of this size, but also the aggregation peaks at 45 ml of elution volume, indicating an unstable protein under the respective buffer conditions. To investigate the folding status of the apo Pys constructs, far-UV circular dichroism (CD) spectroscopy was used. The CD spectrum of the N-terminal His_6-tagged Pys full-length protein in apo state shows two minima at ~208 and ~222 nm, which are characteristic for proteins with high α-helical content (Figure 13C).

For expression and purification of *in-vivo* activated holo Pys, an *E. coli* strain with a genomically integrated surfactin-type phosphopantetheinyl transferase Sfp HM0079 was used (*67*). Transformation of either N-terminal or C-terminal His_6-tagged Pys full-length proteins into the *E. coli* HM0079 strain results in the *in vivo* production of the natural product pyreudion, which were detected in ____ Lab using LC-MS (data not shown). This indicated a successful *in vivo*

phosphopantetheinylation of Pys yielding holo proteins. The *E. coli* strain used for expression, HM0079, is optimized for natural product production, since it was designed to produce artificial NRP products in the heterologous host *E. coli*. However, this means that it is not optimal for protein overexpression. Therefore, the yields of holo Pys full-length proteins were 250 µg/l culture.

Comparing the SEC elution profile of apo, expressed in *E. coli* Bl21 gold (DE3) cells, and *in vivo* activated holo Pys, purified from the *E. coli* HM0079 strain, all proteins show similar elution profiles (Figure 13B). In addition, similar secondary structure elements could be observed in CD spectroscopy, although the CD spectrum of holo N-terminal His_6-tagged Pys protein is noisy due to the lower protein concentration (Figure 13C).

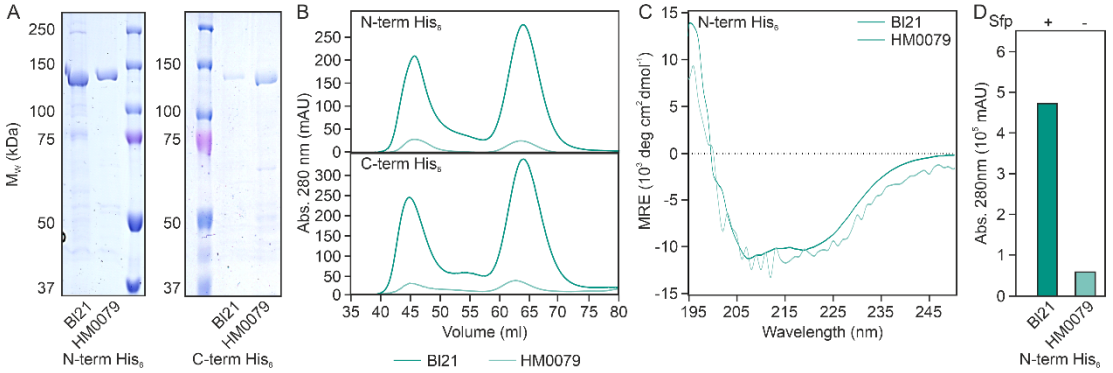


Figure 13: Secondary structural analysis and activity investigation of Pys full-length proteins with N-terminal or C-terminal His_6-tag. A: SDS-PAGE and B: of SEC profiles of heterologously expressed, purified N-terminal and C-terminal His_6-tagged Pseudomonas fluorescens Pys from E. coli Bl21 gold (DE3) and HM0079 cells. C: Far-UV CD spectra of N-terminal His_6-tagged Pys purified from E. coli Bl21 gold (DE3) and HM0079 cells. D: To determine Pys activity, purified N-terminal His_6-tagged Pys protein from E. coli Bl21 gold (DE3) was activated by the phosphopantetheinyl transferase Sfp and the addition of CoA to form holo Pys, while the protein purified from E. coli HM0079 strain was already in-vivo activated. Holo Pys proteins were tested for their enzymatic activity and the absorption of newly formed pyreudione can be detected by the UV/VIS detector of the LC-MS.

To investigate whether the purified Pys full-length proteins show enzymatic activity, an assay to produce pyreudione in vitro was performed. Full-length apo Pys (purified from *E. coli* Bl21 gold (DE3)) was activated by addition of Coenzyme A and the phosphopantetheinyl transferase R4-4 Sfp (Figure 14, first reaction). The transferase Sfp transfers the phosphopantetheinyl group from CoA to the PCP of Pys by catalyzing the covalent bond between the phosphate group of Ppant arm and the reactive serine of the PCP. The activated Pys protein then catalyzes pyreudion synthesis by consuming proline and SNAC-3oxo fatty acids (Figure 14, second part).

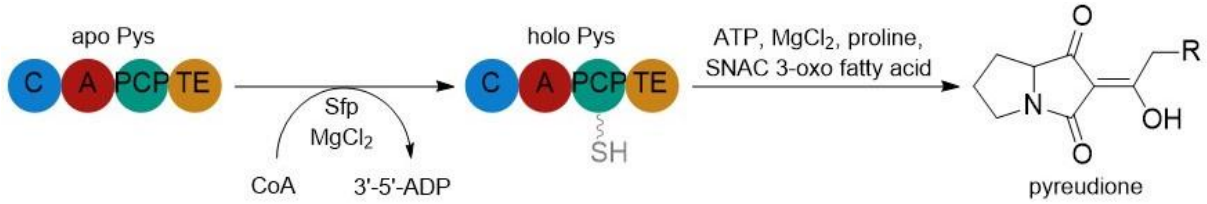


Figure 14: Reaction scheme for the activation of apo Pys catalyzed by Sfp. The formed holo Pys consumes ATP, proline, and SNAC 3-oxo fatty acid to produce pyreudione.

The production of pyreudione can be monitored by LC-MS (Figure 13D). Using *in vivo* activated Pys (purified from *E. coli* strain HM0079), the step of *in vitro* phosphopantetheinylation by Sfp can

be skipped. The holo Pys protein was instead used directly for the synthesis of pyreudione. LC-MS analysis shows a reduced activity of the *in vivo* activated Pys (sample HM0079) compared to the *in vitro* activated protein (sample Bl21) (Figure 13D). Since the Pys protein expressed in *E. coli* strain HM0079 already carried a Ppant arm before performing the assay and no reducing agent was used during purification, it is likely that the reactive thiol group of the Ppant arm was blocked by contaminants during the purification procedure. However, the pyreudion synthesis assay clearly shows the successful *in vitro* activation of apo Pys purified from *E. coli* Bl21 gold (DE3) cells and the production of pyreudion by this protein.

Protein structure determination could be helpful to understand the regulatory mechanisms, specificity and activity of the monomodular NRPS systems. Three methods are employed to gain high resolution structural insights: X-ray crystallography, Cryo-electron microscopy (Cryo-EM) and nuclear magnetic resonance (NMR) spectroscopy. In X-ray crystallography and Cryo-EM the macromolecule must be in a highly ordered crystals or frozen in vitrified ice. In contrast, NMR spectroscopy allows investigating a protein in solution. However, both X-ray crystallography and cryo-EM have no limitations with regard to protein size on the upper boarder. Conventional solution NMR spectroscopy on the other hand reaches its size limit at around ~30-40 kDa, a value exceeded by full-length Pys protein. Therefore, both X-ray crystallography and Cryo-EM were attempted to obtain structural information about the Pys protein. Different crystallization conditions and substrate additions were screened for both full-length Pys proteins with either an N- or a C-terminal His_6-tag. Unfortunately, no diffracting crystals could be obtained, presumably an indication for inherent flexibility in the protein. In both negative stain and single particle Cryo-EM analysis, full-length Pys proteins show large particle heterogeneity and aggregation behavior on the grids. Heterogeneous and aggregated samples complicated the optimization of freezing conditions.

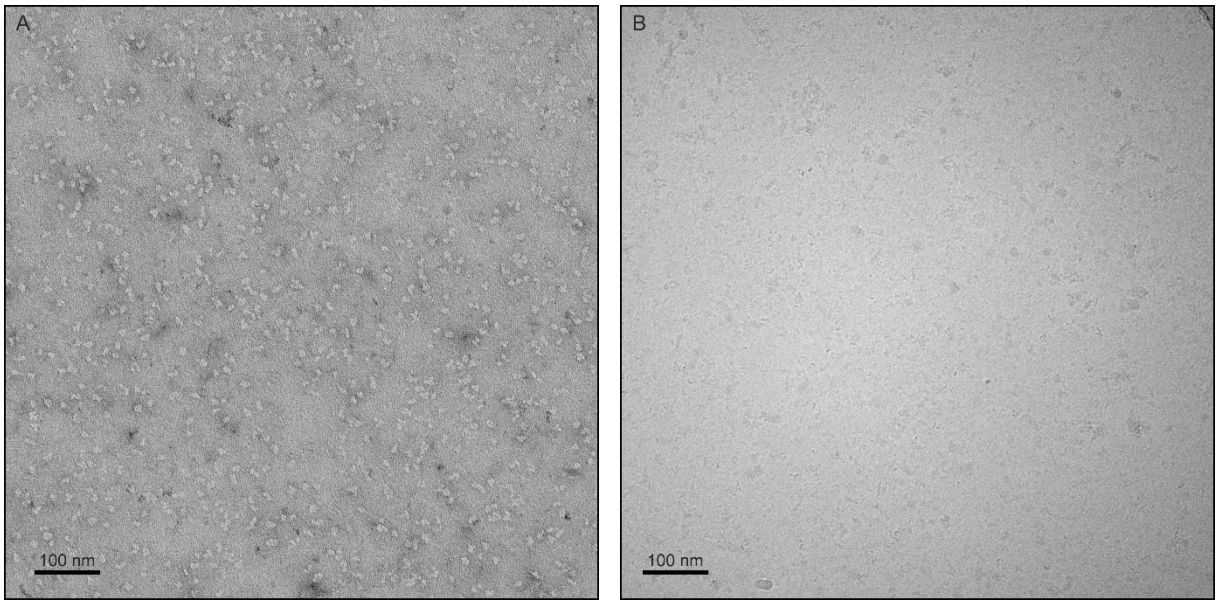


Figure 15: Negative stain and single particle cryo-electron microscopy images of Pys full-length protein. A: Pys full-length with N-terminal His_6-tag (0.125 µM) was negatively stained with uranyl-acetate. B: Representative Cryo-EM micrograph of Pys full-length protein with N-terminal His_6-tag (4 µM). Scale bar shows 100 nm.

In agreement with the lack of success for crystallization, the heterogeneity of the particles implies a wide variation of protein conformations, which cannot be reduced by the addition of substrates alone. This observation also explains the difficulties in structural determination of functional NRPSs and NRPS modules, leading to a lack of structural findings for full-length NRPS proteins in the literature.

1.4.2. With great prediction tools come great responsibility

Since the full-length Pys protein is difficult to handle, a "divide and conquer" approach was followed to obtain structural and functional information about the individual domains of the Pys protein. To define domain boundaries based on amino acid sequence, two prediction tools were used: AntiSMASH 3.0 (*76*) and AlphaFold2 (*77, 78*).

The open-source AntiSMASH 3.0 web server identifies gene clusters by finding open reading frames of genetic information and then perform multiple sequence alignment with already known PKS and NRPS domains. In addition to predicting domains and modules, AntiSMASH can also predict the putative natural product produced PKS and NRPS systems. AntiSMASH prediction has been performed by _____, _____, and the results are shown in Figure 16A.

In the middle of 2021, a novel machine learning technique for structure prediction, as known as AlphaFold2, revolutionized structural biology approaches. The AlphaFold2 network identifies proteins in the protein databank PDB that may have similar structure to the input template and refines the structure using different neural network architectures based on the multiple sequence alignment. The artificial intelligent program is designed as a deep learning system and is open sourced by DeepMind.

For the AlphaFold2 prediction, the Pys sequence with an N-terminal His_6-tag and TEV cleavage site was used. Five structures were predicted, and a global alignment was performed (Sequence alignment score = 6736.5 – 6777.3, 907 – 1051 pruned atom pairs RMSD = 0.484 – 0.837, RMSD = 2.597 – 9.166 all pairs). Superposition of five predicted structures shows high overlap for C and A domains and larger deviations in the N-terminal His_6-tag and in the two C-terminal domains (PCP and TE domain) (Figure 16B). A domain-by-domain alignment result in RMSD = 0.224 – 0.406 for the C domain, RMSD = 0.255 – 0.920 for the A domain, RMSD = 0.270 – 3.027 for PCP domain, and RMSD = 0.337 – 1.203 for TE domain. This suggests that isolated domains and the overall folding can be predicted with high accuracy, but unstructured linkers, the domain-domain distances and orientations are difficult to predict. Less structural overlap of the PCP in global alignment indicates that this domain is subject to large conformational changes and the position of this domain relative to other domains cannot be predicted with high accuracy. The absolute position of the reactive serin on the PCP drifts up to 9.958 Å when comparing all predictions. The TE domain associated with the PCP domain follows the conformational changes of the PCP domain, resulting in repositioning of the TE domain relative to the spatially adjacent A domain. The distance between A and TE domain ranges from 3.648 – 16.921 Å if Q565 (A domain) and T1171

(TE domain) have been taken as anchor points. Considering that the PCP domain and its attached Ppant arm must reach into the active cite of each domain within the Pys protein, large conformational changes of the PCP domain are expected. It remains to be seen whether the orientations of the PCP predicted by AlphaFold2 are physiologically relevant. Likewise, it remains unclear how such structural rearrangements of the PCP domain are influenced by the PPant arm loading status and the other domains within the Pys NRPS.

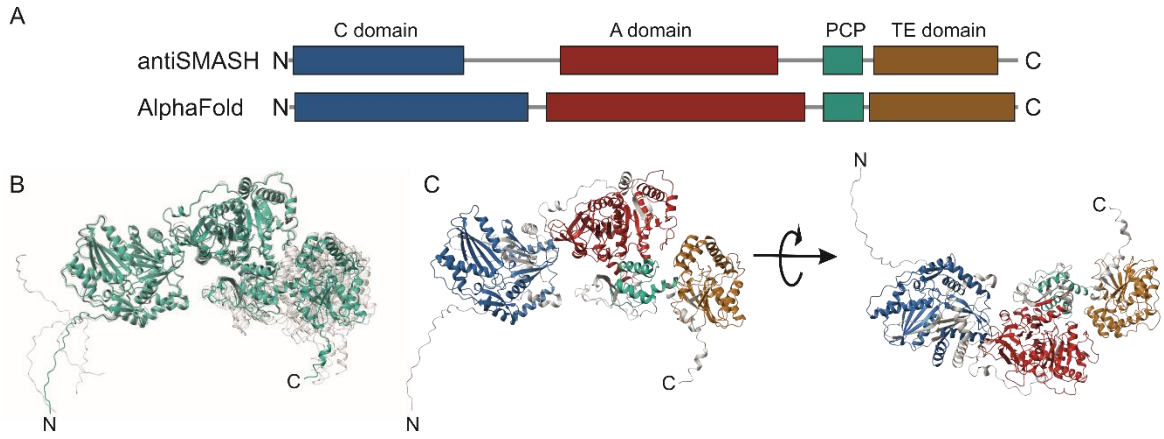


Figure 16: AntiSMASH and AlphaFold2 domain and structure prediction of the Pseudomonas fluorescens NRPS Pys. A: schematic domain boundaries predicted by AntiSMASH 3.0 and AlphaFold2. The C domain region is colored blue, A domain region is colored red, PCP domain is colored cyan, TE domain is colored yellow and linker regions are colored in grey. B: Overlay of five AlphaFold2 structural models aligned to the C domain. C: Domain prediction based on the AntiSMASH 3.0 prediction is colored on one of the AlphaFold2 model. Domains were colored as described in A.

To analyze the accuracy of the previous AntiSMASH domain prediction, it was compared to the AlphaFold2 model (Figure 16A and C). Overall AntiSMASH performed well in predicting the core domains. However, there are notable differences in the domain boundaries of the C domain in the AntiSMASH and AlphaFold predictions. The AntiSMASH model predicts a relatively long linker between C and A domains, while AlphaFold assigns the same region to be a structured part of the C domain. Likewise, the sequence predicted to be a linker between A and PCP domains by AntiSMASH appears to belong to the A domain in the AlphaFold2 model. To verify the results of the two prediction tools and investigating substrate specificity of Pys protein, CA didomain and C domain constructs with different protein sizes were designed.

1.4.3. The condensation and adenylation domain - CA didomain

For pyreudion biosynthesis, the substrates proline and 3-oxo fatty acids need to bind to Pys. Selectivity toward these substrates is mediated by the A domain and the C domain, respectively. Since it has not been shown how the consensus motif in the A domain (D652-M653-L656-V695-M716-G718-V738-F746-A747-K930) coordinate proline binding (*21*) and not much is known about the C domain selectivity, the question about substrate selectivity is raised. To investigate selectivity of C and A domains, a construct consisting of only the CA didomain with an N-terminal His_6-tag was designed, expressed, and purified. SDS-PAGE and SEC profile indicate a globular folded protein with the expected molecular mass of 106 kDa (Figure 17).

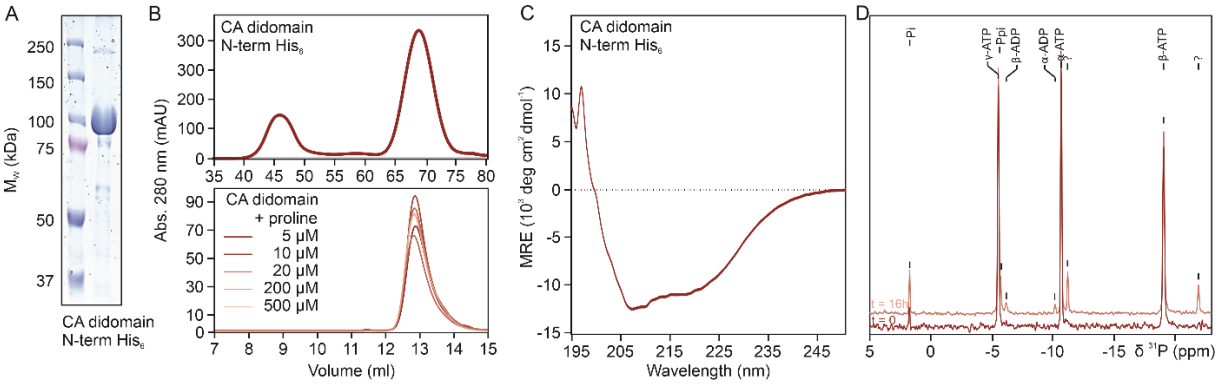


Figure 17: Secondary structural analysis and activity investigation of Pys CA didomain with N-terminal His$_6$-tag. A: SDS-PAGE of N-terminal His$_6$-tagged CA didomain. B: SEC profiles of purified CA didomain with N-terminal His$_6$-tags and the influence of proline to the elution profile. C: Far-UV CD spectra of N-terminal His$_6$-tag Pys CA didomain. D: ^{31}P-NMR spectroscopy monitoring ATP consumption of the CA didomain after 16 hours at 25°C.

The CA didomain shows a CD spectrogram that has characteristic minima at ~208 and ~222 nm suggesting a mainly α-helical fold (Figure 17C). To study whether conformational changes of the CA didomain are induced by the presence of the substrate, the protein was supplemented with different concentrations of proline and the mixtures were analyzed by analytical SEC (Figure 17B). Analytical SEC profiles show elution volumes at ~13 ml independent of the proline concentration. Either proline binding does not result in conformational changes, or the formed complex is not stable in SEC. To analyze whether the CA didomain senses proline binding, tryptophane fluorescence was performed. The CA didomain consists of four tryptophane residues, two in the C domain and two in the A domain. In the A domain one tryptophane is located in the linker region between C and A domain, based on the antiSMASH prediction, and the other one sits close to the consensus signature for proline. Unfortunately, neither the tryptophane fluorescence intensities nor the fluorescence wavelength changed upon addition of proline or ATP to the CA didomain (data not shown). With the method used, it was thus not possible to determine whether and where proline binds to the CA didomain and whether this protein is functional. To address the question of protein functionality, ^{31}P-NMR spectroscopy was used. For proline activation in the A domain, ATP is used to adenylate proline and form prolyl-AMP with elimination of pyrophosphate (PP_i). Taking advantages of the different chemical shifts of the substrate and produced phosphorus species, ^{31}P-NMR spectroscopy can be used to monitor ATP hydrolysis and PP_i elimination over time. Here, 10 µM of purified Pys CA-didomain was incubated with 10 mM ATP and 10 mM proline at 25°C for 16h. ^{31}P-NMR spectra were recorded continuously, and the first and last time points are shown in Figure 17D. The conversion of ATP to PP_i in this system is strikingly low, as shown by the amount of remaining ATP after 16 h. The low activity of the isolated CA didomain is in agreement with what is described in the literature (*27*). In the absence of the PCP domain, the activated prolyl-AMP cannot bind to the Ppant arm of PCP but dissociates from the binding pocket of the A domain. Prolyl-AMP in aqueous solution further reacts with water, which hydrolyzes the prolyl-AMP bond to form proline and AMP. The described chemical shift for AMP in literature is between 0 and 1.3 ppm (*79, 80*). However, no clear signal can be found in this chemical shift range. Another possible reaction is the hydrolysis of another ATP molecule by prolyl-AMP to yield proline and two ADP molecules. The formation of ADP was monitored by ^{31}P-NMR. Two additional signals at ~-10 ppm and ~-23 ppm could not be assigned clearly to the known phosphate species.

This experiment gives first indication to the activity of the designed and purified CA didomain. However, the activity is very slow due to the missing PCP domain.

1.4.4. The condensation domain and its selectivity

To address the open question regarding substrate selectivity, the _____ lab, _____, designed chimeric proteins by exchanging the Pys C domain with the homologous domain from the acylproline synthetase (Aps) from *Pectobacterium betavasculorum*. Aps is also a monomodular NRPS protein that consists of single C, A, PCP and TE domains. Its sequence homology to Pys is 65.2%. Instead of consuming 3-oxo fatty acids, Aps uses unmodified fatty acids to form acylprolines. The _____ lab carried out *in vivo* synthesis assays of pyreudiones and acylprolines using Pys and Aps as well as point mutants and chimeras of the two proteins and could narrow down fatty acid selectivity to two protein regions in the N-terminal part of the C domain and in a region presumed to be the linker between the C and A domains regarding to the AntiSMASH prediction. Referring to the AlphaFold prediction, the presumable linker region forms the C-terminal folded part of the C domain. Based on these observations, _____ (_____ lab) designed four C domain constructs with different lengths and an N-terminal His_6-tag with TEV cleavage site. Mapping the four constructs on the AlphaFold prediction, it can be seen that the C domain M1-L483 has an additional linker C-terminal to the C domain (Figure 18A). C domain M1-A402 lacks the last C terminal helix, which could be crucial for protein solubility (Figure 18B) and the shortest construct, the C domain M1-F303, was designed to be too short and lacked a large C terminal part of the protein (Figure 18C).

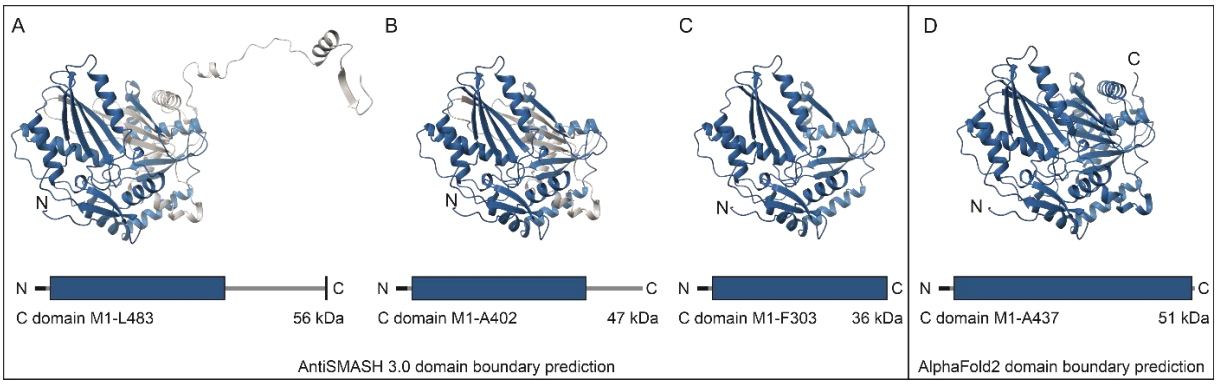


Figure 18: Comparison of AlphaFold2 and AntiSMASH 3.0 structure/domain predictions of the C domain of the Pseudomonas fluorescens Pys NRPS. The AntiSMASH C domain prediction is marked in blue and its linker region prediction nis colored grey on the AlphaFold2 structure prediction. A: C domain M1-L483 construct, B: C domain M1-A402 construct, C: C domain M1-F303 construct, D: C domain M1-A437 predicted by AlphaFold2. N-terminal His_6-TEV tag is shown in black. Molecular mass for all constructs is given including N-terminal His_6-TEV tag.

To investigate the heterologous expression and solubility of the four designed C domain constructs, they were expressed in *E. coli* using the previously established protocols. The harvested cells were lysed to analyze protein content in the supernatant and pellet by SDS-PAGE. During expression and purification of different Pys C domain constructs suggested by the AntiSMASH prediction (Figure 19A), difficulties and solubility issues were faced. While the longest C domain constructs (M1-L504) show good solubility and could be purified (Figure 19C and D),

the two shorter constructs (M1-F303 and M1-A402) were expressed but not soluble (see lane marked “Pe” for pellet in Figure 19B). The region between C and A domains seems not only play a role in substrate selection but seems also to be important for the stability and solubility of the C domain. Changing expression conditions such as expression temperature, time and IPTG concentration did not result in soluble C domain M1-A402 and C domain M1-F303 proteins. Based on these observations, the AlphaFold prediction seems to be more accurate regarding domain boundaries.

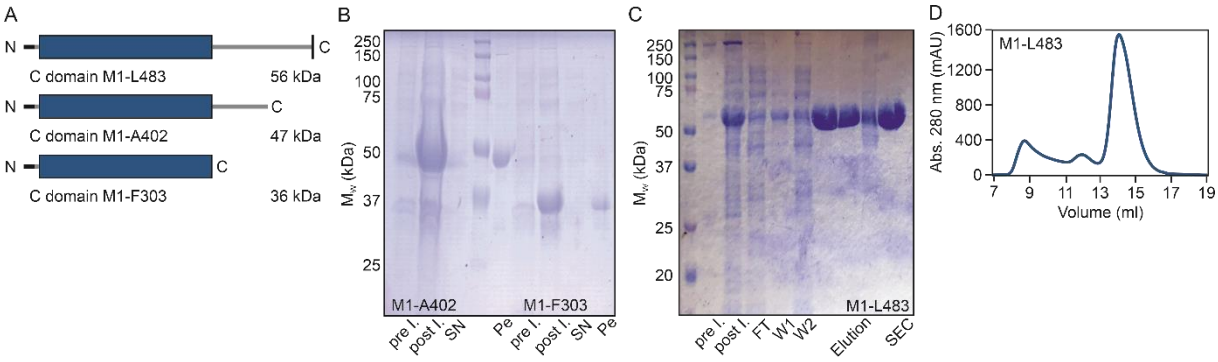


Figure 19: Expression test and purification of different C domain constructs. A: Overview scheme of the different C domain constructs. B: SDS-PAGE of expression test of C domain M1-A402 and C domain M1-F303. Pre and post induction samples are marked with pre I. and post I., supernatant after cell lysis is abbreviated with SN and the pellet sample with Pe. C: SDS-PAGE of N-terminal His_6-tagged C domain M1-L483. Pre and post induction samples are labeled with pre I. and post I. followed by the flow through after NiNTA binding (FT) and two washing steps (W1, W2). C domain M1-L483 is eluted from the column which is shown by elution lanes and the main fraction after SEC (elution volume ~14 ml) is marked with SEC. D: SEC profiles of purified C domain M1-L483 with N-terminal His_6-tags.

Guided by the AlphaFold prediction, Pys C domain M1-A437 and Aps C domain M1-A430 were successfully expressed and purified by ____ (____ Lab). X-ray crystallization screens were prepared, which resulted in single crystals for Aps C domains and microcrystals for Pys C domain. Images of the formed crystals are shown in Figure 20. Aps C domain M1-A430 crystals diffracted to a resolution of 2.1 Å (Figure 20D). An overlay of the crystal structure and the AlphaFold2 model shows a high accuracy of the prediction with RMSD = 0.648 Å between pruned atom pairs and 0.905 Å across all pairs (Figure 20E).

Analyzing the Aps C domain crystal structure shows differential electron densities in close proximity to the catalytic H141. This observation indicates that a substrate is bound to the protein, which could not be clearly identified due to the low resolution of this electron density (data not shown). It is likely that during protein expression and purification the Aps C domain is partly loaded with substrates. The electron density near H141 could belong to an activated fatty acid but this hypothesis needs to be confirmed by further biochemical assays. Another interesting observation is the modification of S326, which roughly fits to a phosphorylation, but could not identify further due to weak resolution at this position.

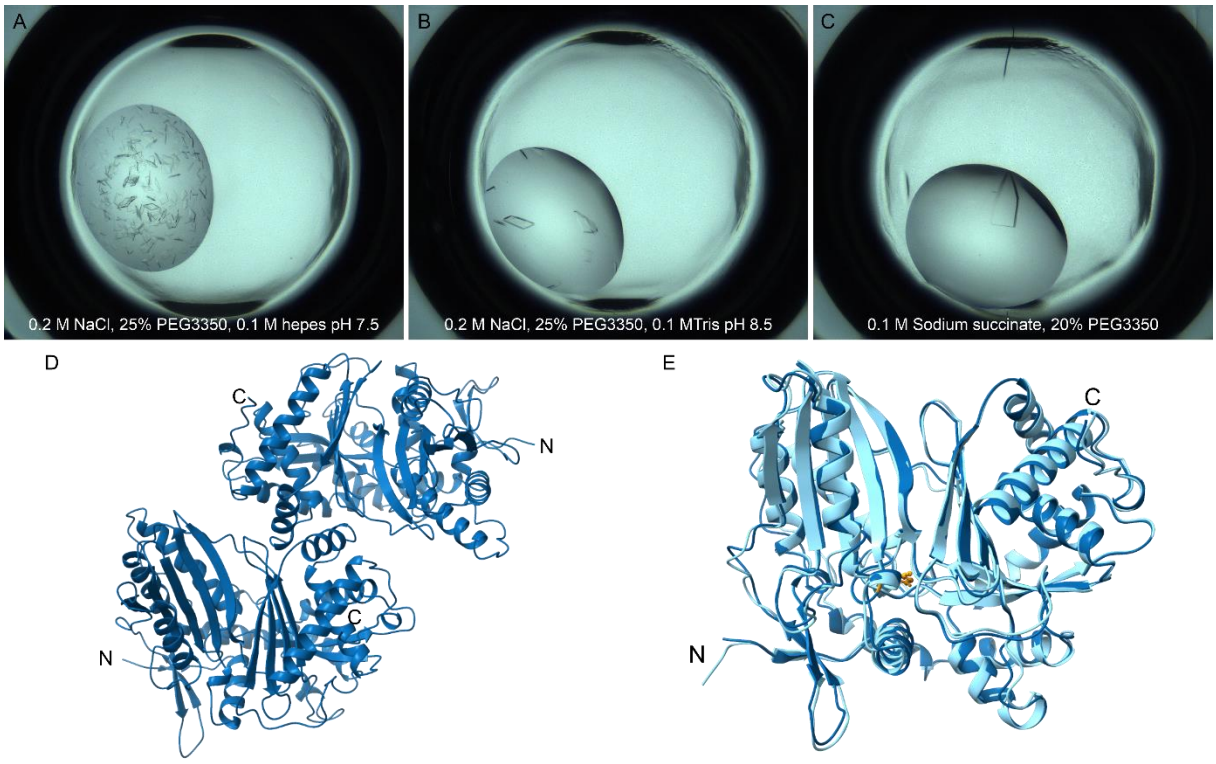


Figure 20: Crystal structure of Aps C domain from Pectobacterium betavasculorum. A-C: Crystals formed by Aps C domain using sitting drop crystallization method. 10 mg/ml Aps C domain protein concentration in 50 mM hepes pH 6.8, 150 mM NaCl and the respective precipitant solutions were used. All crystals were formed within 3 days at 20°C. D: Crystal structure of Aps C domain with 2.1 Å resolution. E: Overlay of the crystal structure and AlphaFold2 model of Aps C domain. Crystal structure is shown in dark blue, AlphaFold2 model in light blue. Active site H141 is shown in yellow balls and sticks. RMSD = 0.648 Å.

To get an idea of where the reaction tunnel could be located, Caver Web 1.0 (*81, 82*) was run on the Aps C domain crystal structure. The web server analyzes protein tunnels and channels based on the protein structure, which can be uploaded from the user or by adding the PDB of interest. The analysis reveals in total four possible tunnels, which are shown in Figure 21. The properties of the calculated tunnels are listed in Figure 21A. For a better orientation the catalytic H141 and the modified S326 are shown in sticks. All tunnels meet at the catalytic H141, which gives evidence to the plausibility of the predicted tunnels. The tunnels #1 and #2 (Figure 21B) fit to the described reactive tunnel for condensation domains (*28, 83*). In the calculation, tunnel #2 has an average length of 27.199 Å, which is longer than the other tunnels, but can be explained be the overlap between tunnel #1 and #2. The cross section close to H141 is also added to the length of tunnel #2. Only based on this model, it is difficult to say from which site the proline loaded holo PCP could enter the condensation domain.

Surprisingly, two other tunnels were predicted (Figure 21C), which have not been described before in literature for condensation domains. These tunnels also come together in close proximity to H141 and allows the hypothesis of a not yet discovered binding site for carrier proteins, or a possible entering site for the (activated) fatty acid. S326 is 9.3 – 13.4 Å apart from the tunnel entrances of the tunnels #3 and #4. It is likely that the S326 plays a role in loading the fatty acid to the condensation domain using either tunnel #3 or #4 to enter the reactive tunnel. However, further experimental evidence is needed to confirm this hypothesis.

A

Tunnel	Average bottleneck radius [Å]	Average tunnel length[Å]
#1	1.94	13.316
#2	1.60	27.199
#3	1.92	12.226
#4	2.55	3.966

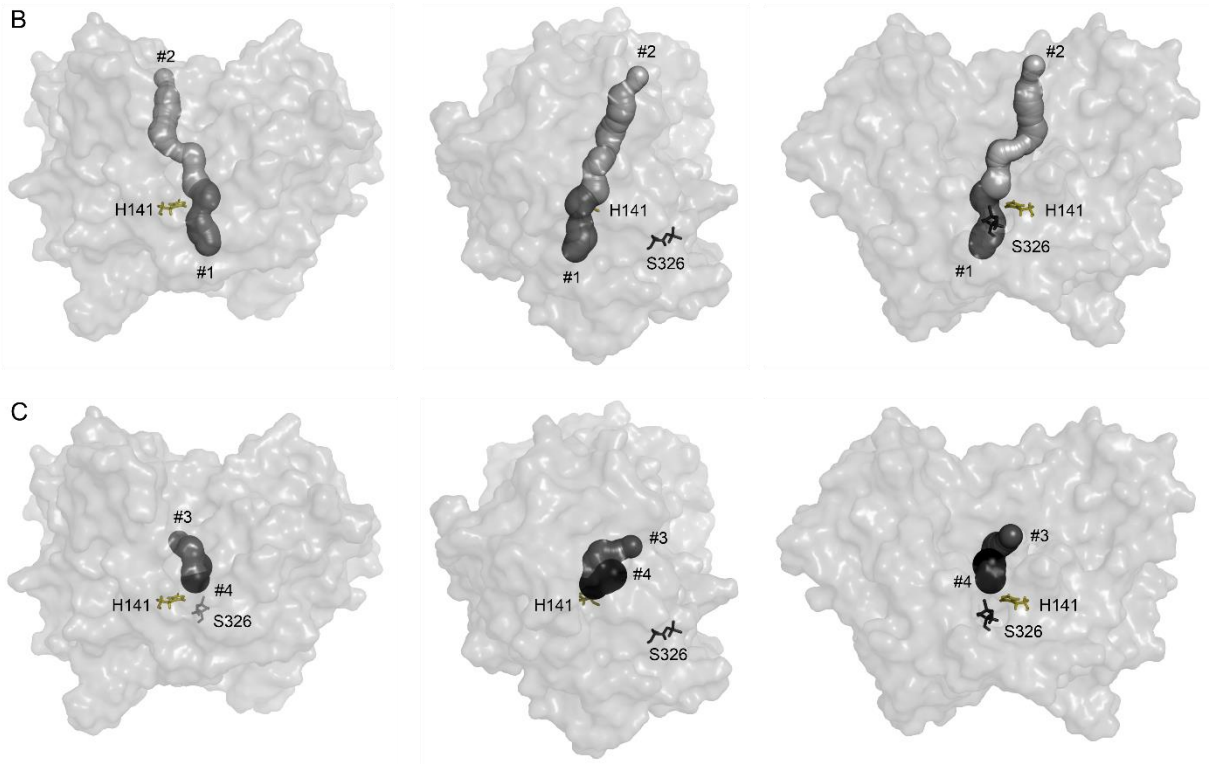


Figure 21: Possible reactive tunnels in the Aps C domain from Pectobacterium betavasculorum were calculated by Caver Web 1.0 (82, 81). A: Properties of the calculated tunnels. B: Tunnels #1 and #2 were mapped on the Aps C domain crystal structure and shown from three different perspectives. C: Tunnels #3 and #4 were mapped on the Aps C domain crystal structure and shown from three different perspectives. H141 marks the active site and is shown in yellow sticks. The modified S326 is shown in black sticks.

Mutation studies performed by _____ (_____ lab) indicate fatty acid headgroup selectivity controlled by a hand full residue in close proximity to the C domain core motif ^{142}SHLVVDGF^{149} in the Pys protein and ^{140}THLIIDGY^{147} in the Aps protein with the catalytic histidine at position 143 and 141, respectively (Figure 22). Based on this observation, it is possible that the fatty acid is located in the tunnel #1.

In a close up of the structural models of the Pys C domain and the structure of the Aps C domain, differences in the catalytic centers of Pys and Aps are noticeable (Figure 22). In the Aps catalytic site, H141 is surrounded by four hydrophilic/charged residues that might stabilize the fatty acid head group of the substrate, and/or a reaction intermediate by the formation of hydrogen bonds. From the perspective shown in Figure 22B part 2, C and E, it appears that the distance between the reactive histidine and the opposing residues is bigger in Pys than compared to Aps. In Pys, rather small residues like D379 and G382 could be found instead of bulkier residues like R377 and N380 in Aps, respectively. It is possible that because of this larger binding site, a more steric head group, such as a 3-oxo fatty acid, could bind to Pys but not to Aps. The fatty acid could be

stabilized by hydrogen bonds formed by hydrophilic residues in the binding pockets (Figure 22C and D). However, this model needs to be verified by experimental or *in silico* data.

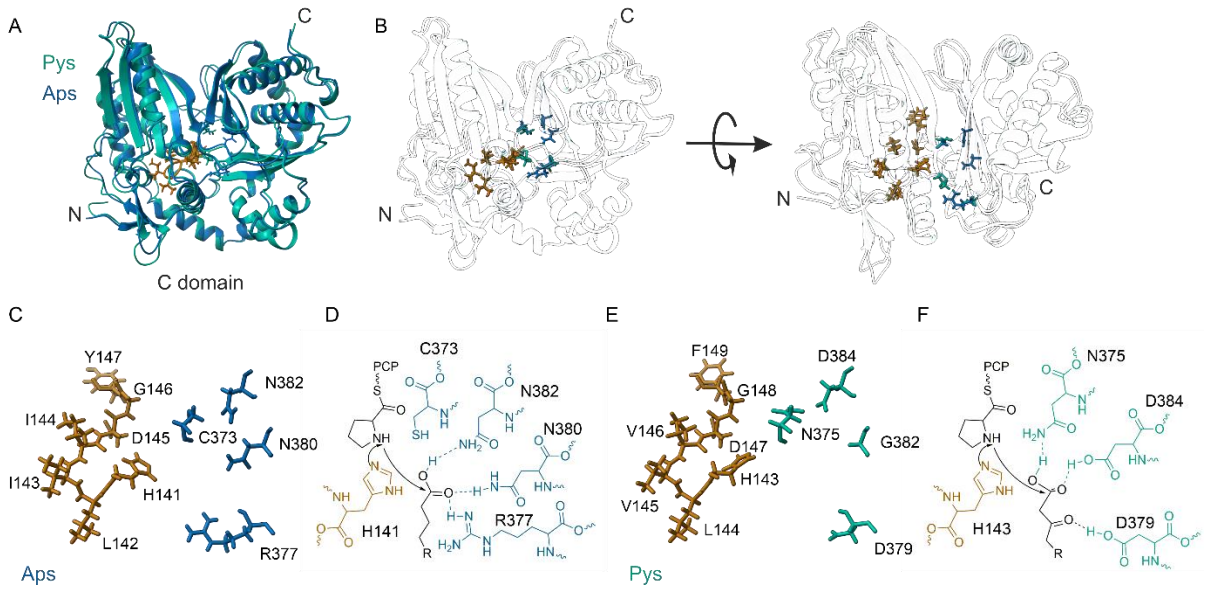


Figure 22: Model of the C domain catalytic site within the Pys protein from Pseudomonas fluorescens and the Aps protein from Pectobacterium betavasculorum. A: Overlay of the AlphaFold models of the Pys and the crystal structure of Aps C domains. Active site is marked in yellow sticks. B: C domain core motif is highlighted yellow and residues that are different in the two proteins but next to the core motif are shown in stick design. Pys residues are colored cyan, Aps residues are colored blue. C and E show only the catalytic center of the C domain. D and F: Model how residue differences might explain fatty acid headgroup selectivity.

Nonetheless, in both Pys and Aps, the binding pocket is very hydrophilic. Proper positioning of the fatty acid is imperative for the condensation reaction with a proline. This condensation reaction could be catalyzed by the deprotonation of the Ppant-bound proline amide by the catalytic His residue, followed by a nucleophilic attack by the carboxyl group of the fatty acid (Figure 22D and F). To confirm or refute this model, biochemical analyses using mutations of the residues R377 (in Aps) and D379 (in Pys) should be performed. Specifically, the effect of such mutations on the ability to interact with fatty acids and product formation should be explored. Investigation of the interaction site between a proline-loaded PCP and the condensation domain would help to determine the position of proline and fatty acid in the reactive tunnel of the condensation domain.

1.4.5. Interaction between its ACP and KSB domains in the rhizoxin branching module

A fundamental understanding of the structure of NRPS and PKS systems should hopefully enable researchers to disassemble and reassemble mega enzymes for the modification and production of novel bioactive compounds. However, due to their size and inherent flexibility, multimodular synthases are difficult systems for structural approaches. Even the monomodular NRPS Pys from *Pseudomonas fluorescens*, as discussed above, is challenging when it comes to the application of classical structural biology approaches. The divide and conquer approach, in which the entire enzyme is cleaved into its separate modules and domains, reduces complexity, and can enable structural and functional studies. Since carrier proteins are key components of synthases, their structural and functional have a direct impact on enzyme activity. Comparing NRPS and PKS enzymes, the carrier proteins show similar overall folds, so that the findings of one carrier protein should in principle be transferable to others.

Two easily accessible but not yet fully characterized acyl carrier proteins were analyzed in this thesis. The first protein is the acyl carrier protein (ACP) in the unique branching module of the *Burkholderia rhizoxinica* rhizoxin synthase. Encoded by the *rhiE* gene, the branching module consists of a ketosynthase (KS) domain, a branching (B) domain without catalytical function, and the beforementioned ACP. The branching module catalyzes vinylogous addition of acetate to an α,β-unsaturated thioester of the growing chain in the rhizoxin synthase. This reaction is unique, since acetate is usually added to the terminal activated thioester, but not to a vinyl group next to the ester. The reaction schemes of the addition and decarboxylation of malonyl-ACP catalyzed by a regular PKS module, and the branching module are shown in Figure 23.

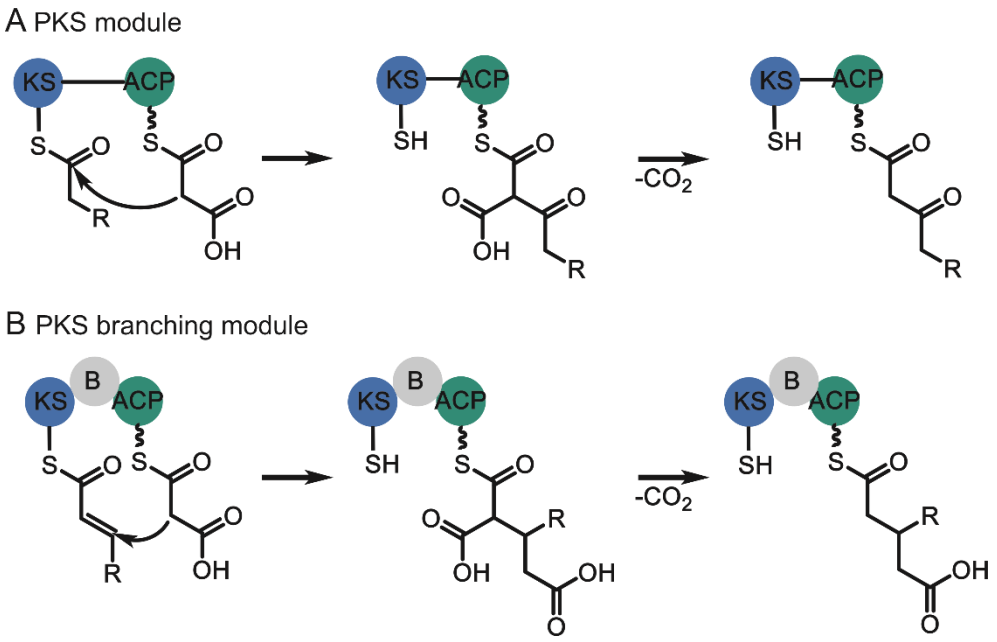


Figure 23: Reaction scheme of the addition and decarboxylation of malonyl-ACP catalyzed by PKS modules. A: Reaction catalyzed by a regular PKS module. B: Reaction catalyzed by the branching module of the rhizoxin synthetases from Burkholderia rhizoxinica. KS: ketosynthase, ACP: acyl carrier protein, B: branching domain. Pantetheine arm is shown in black wavy line.

A crystal structure of the *rhiE*-KSB di-domain was obtained by crystallography by Bretschneider *et al*. that enables structural insights into the KSB didomain (*51*) of the branching module. However, this structure cannot answer questions about the interaction between ACP and KSB domains. In general, structural insight into KS-ACP interaction in modular PKSs has been limited and was addressed in this thesis. Cross-linking mass spectrometric analyses performed by the _____ Lab, _____ and their cooperation partners show interaction between the *rhiE*-KS-B didomain with the ACP from the same module. Unfortunately, structural studies of the entire branching module proved difficult due to the size and flexibility of the module. To nonetheless obtain an idea about where the ACP domain may bind to the KSB didomain, an AlphaFold2 prediction of the *rhiE*-KSB-ACP complex was performed. Overlay of five predicted structures shows a good general overlap of the KSB didomain and little movement in the B domain relative to the KS domain (Figure 24). Surprisingly, two distinct ACP binding sites were predicted. In both cases, the reactive serine at the beginning of helix II was oriented towards the remainder of the KSB domain.

One of the predicted ACP binding sites places the reactive serine on the ACP at a distance of 15.3 Å from the catalytic center of the KS domain, the other binding site increases the distance to 40.4 Å (Figure 24). The binding site shown in Figure 24B with a distance of 15.3 Å between ACP serine and the histidine of the KS catalytic site would allow the growing polyketide chain to bound to the Ppant arm to enter the KS active site. This binding site was also reported in type II PKS systems, such as anthraquinone synthetase from *Photorhabdus luminescens (84)*, and fatty acid synthase from *E. coli (85, 86).*

The second binding pose (shown in Figure 24D) could act as a "parking position" and be populated when malonyl building block that is not transported by the ACP is loaded into the KS domain, as is the case with substrate activation by acyltransferase (AT) domains. It could also be a binding site for a second ACP, since both the upstream and downstream ACP must interact with the KS during the transport cycle. The role of the second ACP binding site is under debate, and it can be found in different KS-ACP structures, such as fatty acid synthase KS (FabB) and the corresponding ACP (AcpP) (*85*), and pikromycin synthetase from *Streptomyces venezuelae (87)*.

When looking at the predicted binding interfaces of ACP and KSB, not many residues are involved in this interaction, besides the reactive serine at the beginning of helix II in the ACP. The "active" binding site, with a short distance between S1061 in the ACP and H419 in the KSB as shown in Figure 24B, is stabilized by possible hydrogen bonds formed by D1069 (ACP) and K121 (KS), and Y1085 (ACP) and N345 (KS) (Figure 24A). I1062 on the ACP in close distance to N286 on the ACP (Figure 24A). This interaction covers 496 $Å^2$ (KS) and 487 $Å^2$ (ACP). The "resting" state binding site, with a large distance between ACP S1061 and KSB H419 as shown in Figure 24C, could be stabilized by electrostatic interaction between KS K146 and ACP D1056, and a van der Waals interaction between KS V150 and ACP L1041. The interaction areas are 320 $Å^2$ (KS) and 324 $Å^2$ (ACP). Since the superposition of the five predicted AlphaFold2 structures shows differences in absolute ACP position, this may be indicative of a general low affinity binding between ACP and KSB. Low affinity binding between carrier proteins and catalytic domains is presumably advantageous in the context of a multimodular enzyme because it allows rapid transport of intermediates and activated substrates. High affinity binding between carrier proteins and

catalytic domains could unnecessarily block the active site, thereby slowing enzymatic turnover. For active multimodular enzymes, selective binding between carrier proteins and catalytic domains is mandatory to bring the activated substrates to the correct active site, but this selective binding must be able to be removed quickly to make room for coupling of the next and subsequent substrate. It remains unclear how the prediction would be changed by using an activated holo ACP. Unfortunately, at the current state, it is not possible to run the prediction with a Ppant arm attached on the ACP.

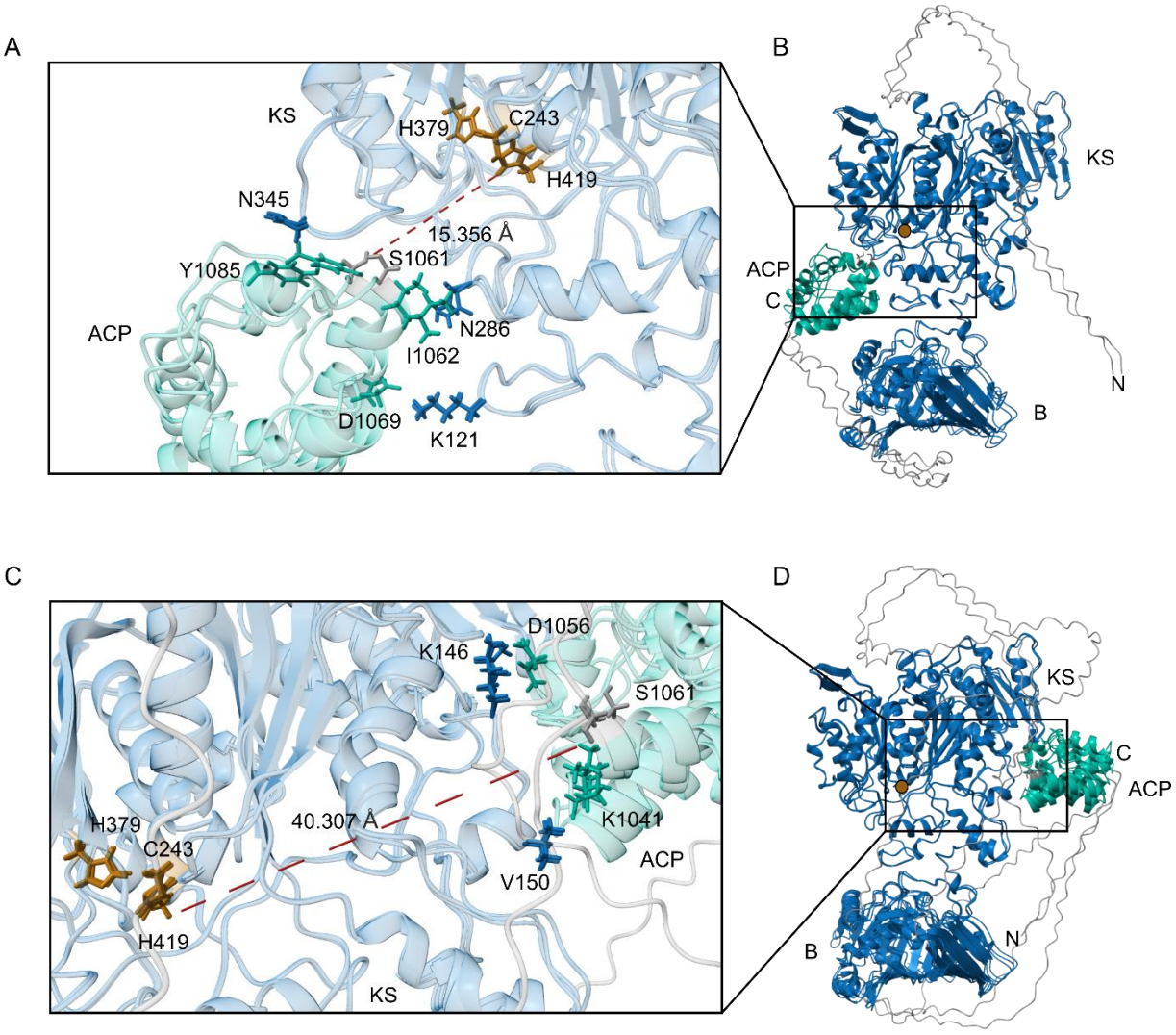


Figure 24: KSB-ACP model from the Burkholderia rhizoxinica rhizoxin synthase predicted by AlphaFold2 reveals two possible binding sites for the ACP domain. B and C: superposition of the predicted five Alphafold2 KSB-ACP models results in two possible binding sites for the ACP domain (cyan) on the KSB didomain (blue). Linker regions are colored grey. A and D: A zoom into the ACP-KS interface with contact residues shown as sticks. The active center of KS is colored yellow and the distance between the ACP reactive serine S1061 residue to the active center of the KSB is marked in red.

To investigate the *rhiE*-ACP-KSB interaction and to verify the AlphaFold2 prediction, nuclear magnetic resonance (NMR) assignments of the holo *rhiE*-ACP domain were performed. The branching module ACP with an N-terminal His_6-tag followed by a thrombin cleavage site was cloned, expressed, and purified by ____ (____ group). During expression, *E. coli* cells were fed with $^{13}C_6$-glucose and $^{15}NH_4Cl$ as the sole carbon and nitrogen sources, resulting in uniformly labeled ^{13}C, ^{15}N-ACP domain. Following the 20 amino acids for purification and cleavage tag, the *rhiE*-ACP

consists of a short stretch of 22 residues that is part of the linker between the B domain and the ACP. The ACP itself extends from residue I1033 to A1100 as indicated by AntiSMASH 3.0 predictions (*76*). The construct contains two proline residues: P1016 in the N-terminal linker region of the ACP and P1081 in the C-terminal part of the ACP. With the exception of the residues belonging to the N-terminal His_6-tag and the thrombin cleavage site, 94.3% of all non-proline residues' backbone peaks could be assigned in the ^{1}H-^{15}N correlation spectra (Figure 25A and E, chapter 1.7., Figure 31). No assignments are available for five residues, the first two residues after the thrombin cleavage site, Q1011 and E1012 as well as residues D1082, Q1088, and V1094 in the ACP domain. Nevertheless, C', Cα and Cβ chemical shifts could be determined for all of the remaining amino acids including the two proline residues except for Q1011.

The backbone chemical shift assignments were then used to determine the secondary structure of the ACP using TALOS-N prediction tool (*88*). Overall, the resulting secondary structure prediction is consistent with the expected overall structure for carrier proteins (Figure 25D and E). The AlphaFold2 structure prediction agrees with the results of the chemical shift-based secondary structure prediction. However, AlphaFold2 tends to slightly overestimate the helix lengths. (Figure 25D and F). *RhiE*-ACP consists of four helices, which are shown in Figure 25D. Residue S1061, at the start of helix II, marks the position for the attachment of the Ppant arm. Interestingly, two helices were observed in the loop region between the helices I and II, which are not typical secondary structure elements of ACP domains but has been observed in PCPs (*19*). In addition, the length of the *rhiE*-ACP helices and their distribution are more consistent with known PCP structures than those of ACPs. It is interesting that a PCP-like carrier protein is found in the PKS part of rhizoxin synthase. This supports the idea that carrier proteins are alike, and the knowledge gained from one carrier protein could be transferred to other carrier proteins.

For efficient substrate transport, the ACP needs to interact with the neighboring KSB domain. NMR spectrometry was used to investigate the ACP-KSB interaction. First, ^{15}N-labeled ACP was titrated with unlabeled KSB. The backbone NMR assignment of *rhiE*-ACP is used to determine the ACP residues involved in KSB binding. Some ACP peaks disappeared upon addition of KSB, but even at almost 10-fold excess of KSB, some residues remain unaffected by KSB interaction (Figure 25C).

When the residues whose chemical shifts disappear upon KSB addition were mapped onto the ACP AlphaFold2 model, a clear binding site on the ACP is difficult to determine. At 2.4-fold molar excess of KSB, the affected residues didn't form a patch but were spread over the ACP (Figure 25G). At 10-fold excess of KSB almost all residues in close proximity to the reactive S1061 were affected by the KSB addition. That these residues are affected the most during the interaction with neighboring PKS domains is in agreement with the observations for different ACP systems in literature (*89*–*91*). It was reported that the major interaction surface of carrier proteins is formed by the serine at the start of the helix II and spatially adjacent residues (*91*). This interaction surface is important to bring the attached Ppant arm into a catalytic cavity of catalytic domains.

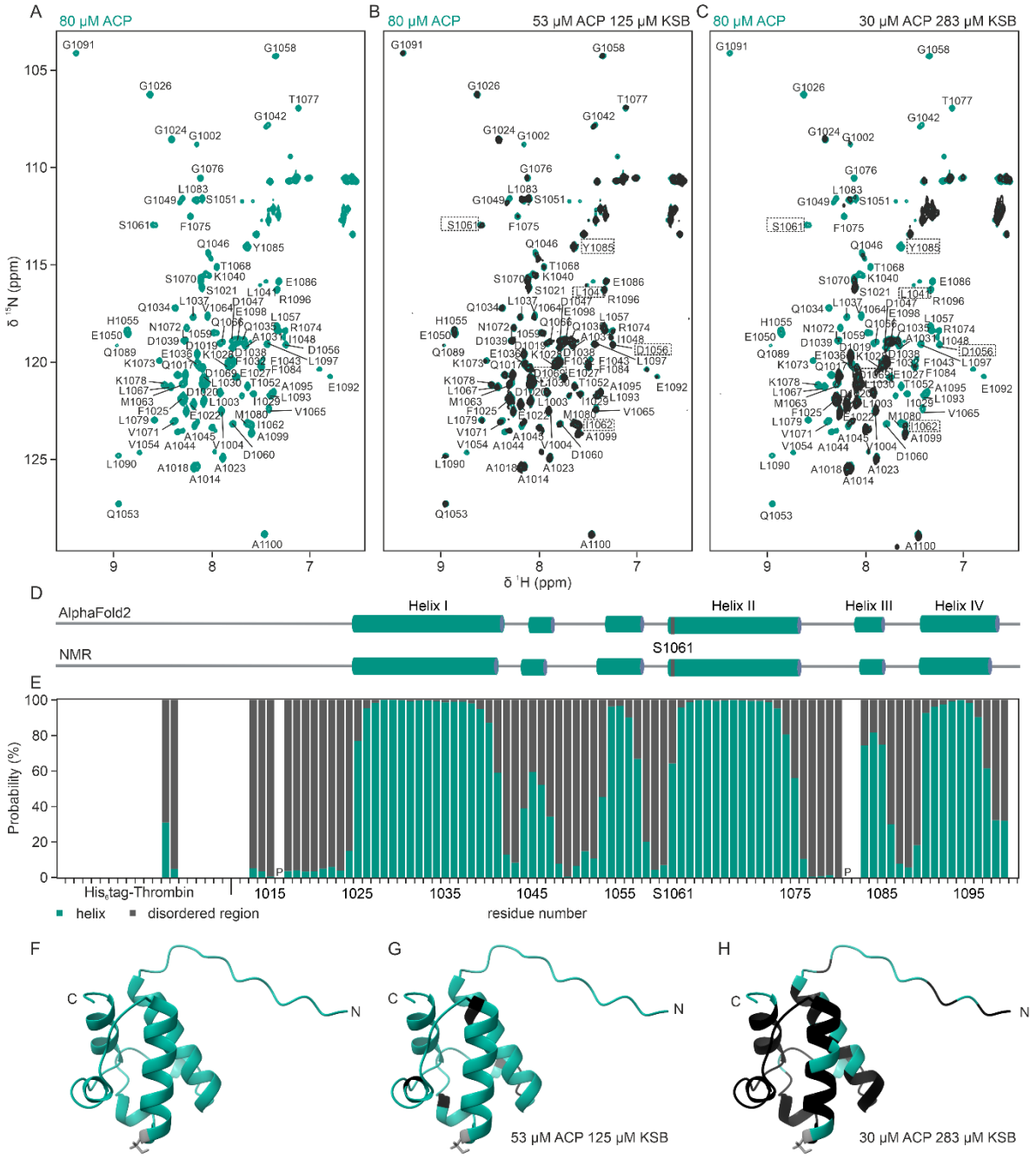


Figure 25: Structural investigation of holo RhiE-ACP from the Burkholderia rhizoxinica rhizoxin synthase. A-C: $^{1}H,^{15}N$-HSQC spectra and backbone assignment of $^{13}C,^{15}N$-labeled RhiE-ACP (cyan) and (B, C) and with addition of 2.4-fold molar excess (B) or 9.4-fold mole excess (C) of KS. Residues, which are predicted by AlphaFold2 to be involved in the KS binding are marked with a dashed box. D: Topology models of RhiE-ACP based on the AlphaFold2 (77, 78) and the NMR chemical shift based secondary structure prediction using TALOS-N (88). Residue S1061 and the position of the Ppant arm is marked with a grey line. E: Chemical shift based secondary structure prediction of the RhiE-ACP using the program TALOS-N. Unassigned peaks are left blank, prolines are marked with P. F: Structure prediction using AlphaFold2. S1061 and the position of the Ppant arm is marked with grey sticks. G-H: NMR chemical shift perturbation induced by KSB (see B, C) mapped on AlphaFold2 RhiE-ACP model. Residues showing line broadening upon KSB addition are marked in black.

After determining the interaction site on the ACP, it was aimed to determine the interaction site on the KSB as well. Unfortunately, with 107.8 kDa, the KSB domain reaches the upper limit of solution NMR spectroscopy. However, reducing the total number of reporters by labeling only specific residues with a nucleus with a high gyromagnetic ratio might help to monitor the interaction area. Hence, we used ^{19}F-tryptophan to label the KSB didomain, which contains 10

tryptophan residues well distributed throughout the protein (Figure 26A). In the best-case scenario, the fluorinated residues can then be used as reporters to monitor the interaction with the ACP domain and potentially even to identify the ACP binding site. The recorded ^{19}F NMR spectra show multiple peak areas with different widths (Figure 26B), indicating that the 10 labeled tryptophans in the KSB didomain perceive slightly different chemical environment and show different mobility. Addition of 6.7-fold molar excess of ACP to the labeled KSB did not change the peak positions or intensity significantly (Figure 26B). Either the tryptophan residues were not suitable reporters because of their location on the protein, or the ACP binding has a weak affinity and does not lead to significant conformational changes in the KSB protein.

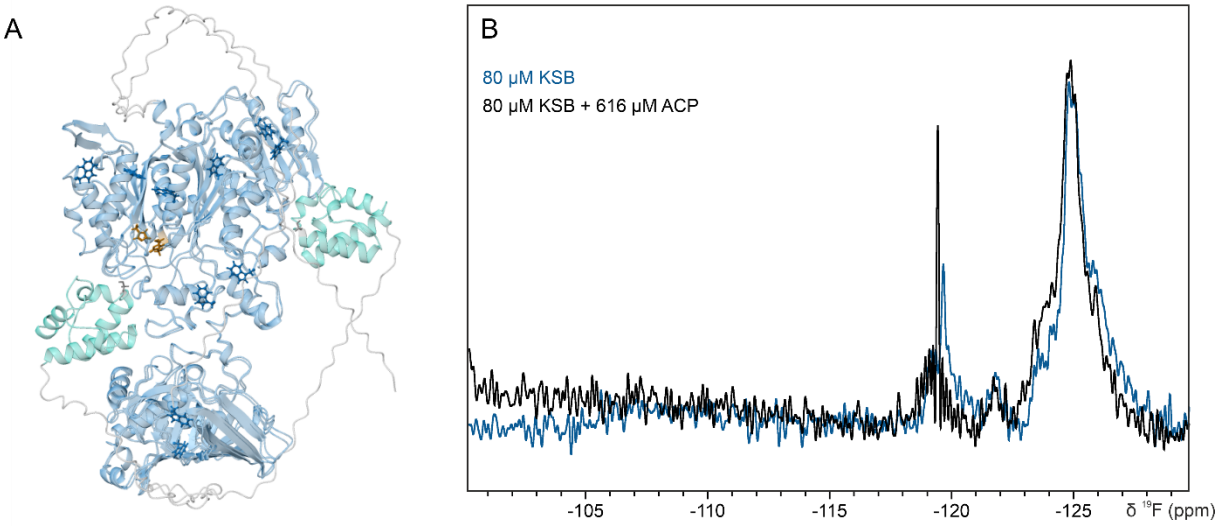


Figure 26: Interaction studies of the rhiE-KSB didomain from the Burkholderia rhizoxinica rhizoxin synthase with the rhiE-ACP from the same module. A: Overlayed AlphaFold2 models showing two possible binding sites of the ACP (cyan) on the KSB didomain (blue). All tryptophans are shown in blue sticks, the KS active site is shown in yellow sticks and the reactive serine on the ACP is shown in gray sticks. The linker region between the KSB didomain and ACP is shown in grey. B: ^{19}F-NMR spectra of the ^{19}F-Trp labelled KSB didomain (blue) and with addition of 6.7-fold molar excess of unlabeled holo ACP (black).

Both NMR experiments indicate no or low affinity binding between the *rhiE*-KSB didomain and *rhiE*-ACP. The low affinity binding between ACPs and catalytic domains has been mentioned in literature (*92*) and agrees with the observed experimental data. This leads to an unexplored intra molecular interaction network between ACPs and catalytic domains and underlines the complexity of these systems.

1.4.6. *Gloeocapsa sp. PCC 7428* ACP – *a* structural investigation of a type II acyl carrier protein

The type II carrier protein is a key player in type II polyketide synthase (PKS) multienzyme assemblies. PKS type II systems form aromatic polyketides by the conversion of malonyl-based building blocks (*93*), a biosynthesis that involves five to eight stand-alone protein components. Carrier proteins guide and control the biosynthetic process by forming transient complexes with

the other catalytic proteins (*45*). Thus, understanding the regulatory mechanism, structure and conformational dynamics of the acyl carrier protein (ACP), as well how it interacts with its partners, is a requirement for the strategic redesigning of type II PKS synthetic machineries. In the long run, this will help to create opportunities to design bioactive compounds.

To date, structural insights into ACP interaction with the catalytic domain ketosynthase (KS) is limited (*87, 94*). The transient interaction between ACP and KS in type I PKS assembly lines makes such investigations particularly challenging (*92, 55*). For type II PKS systems, affinity between ACP and KS domains in low μM range was reported (*95*), which may simplify structural studies of this complex. Furthermore, based on the sequence and functional homology between the PKS systems, the knowledge gained from type II PKSs should be transferable to type I PKSs.

Here, a recently discovered type II ACP from *Gloeocapsa sp. PCC 7428* (GloACP) through bioinformatical predictions in the laboratory of ____ (_____) using AntiSMASH 6.1. (*96*) was chosen as a model system for structural and functional studies of type II ACPs. GloACP is part of a gene cluster encoding a phosphopantetheinyl transferase (PPTase), a ketosynthase (KS), a chain length factor protein (CLF), a cyclase protein (CYC), a methyltransferase (MET), and a benzoate-CoA ligase (Benz). Due to the free-standing character of the type II PKS protein components, it was not possible to predict the natural product produced by the gene cluster with the available computational tools. Thus, structural and functional investigations into the Glo-PKS may not only shed light on the function and regulation of type II PKS in general, but also help to understand the role of individual proteins in this gene cluster and which product they form.

As the heart of every type II PKS system, ACPs direct biosynthesis by moving the growing substrate precursors between the catalytic proteins. For a structural view into the GloACP, AlphaFold2 (AF2) was used (*77, 78*). Interestingly, and in contrast to the typical four-helical ACP fold, the AF2 structure prediction shows only three helices for GloACP. Helix III (according to the nomenclature of archetypical ACPs) is absent from the AF2 model of GloACP (Figure 27A). The interaction of the GloACP and the catalytically active PKS proteins components may be mediated by surface electrostatics. Thus, the surface electrostatic potential was mapped on the AlphaFold2 model. Again, in striking contrast to the negative patch described for archetypical ACP around the reactive serine residue, which is needed for the KS-ACP interaction (97), GloACP shows a positive surface potential near the reactive residue S34 at the beginning of helix II (Figure 27B). To verify the AF2 model, isotope labeled GloACP was purified by IMAC and SEC (Fig. 1C, d) and a 1H, 15N-HSQC spectrum recorded (Figure 27E, chapter 1.7,Figure 32). Both the size exclusion and the HSQC show that purified GloACP is a monomeric, folded protein (Figure 27D and E). The 80 amino-acid construct contains two proline residues: P-5 in the thrombin cleavage site and P16. In total 91.25% (excluding the His_6-tag and thrombin cleavage site) of all non-proline residues backbone resonances could be assigned in the ^{1}H-^{15}N correlation spectrum (Figure 27E). The spectrum shows in total 77 peaks, while 100 were expected, and only three peaks are not assigned. No assignments are available for the backbone amides of the N-terminal His_6-tag, the thrombin cleavage site, and residues M1, S34, T35, N51, S56, R57, as well as residue K58. Nevertheless, Cα and Cβ chemical shifts could be determined for all of the remaining amino acids including the two proline residues except for the active site residue S34, as well as for S56, and R57. M1, I15, K28,

D33, S34, S56, R57 and K58 miss CO chemical shift information. The missing peak for the active site residue S34 could be explained by intermediate exchange between conformations and flexibility at this position (*98*). In future steps, the addition of divalent cations, such as Ca^{2+} or Mg^{2+} may prove useful, as they were described to increase the thermal stability of type II ACPs (*99*). Furthermore, the activation of GloACP with a PPTase may possibly aid the assignment of the active site residue.

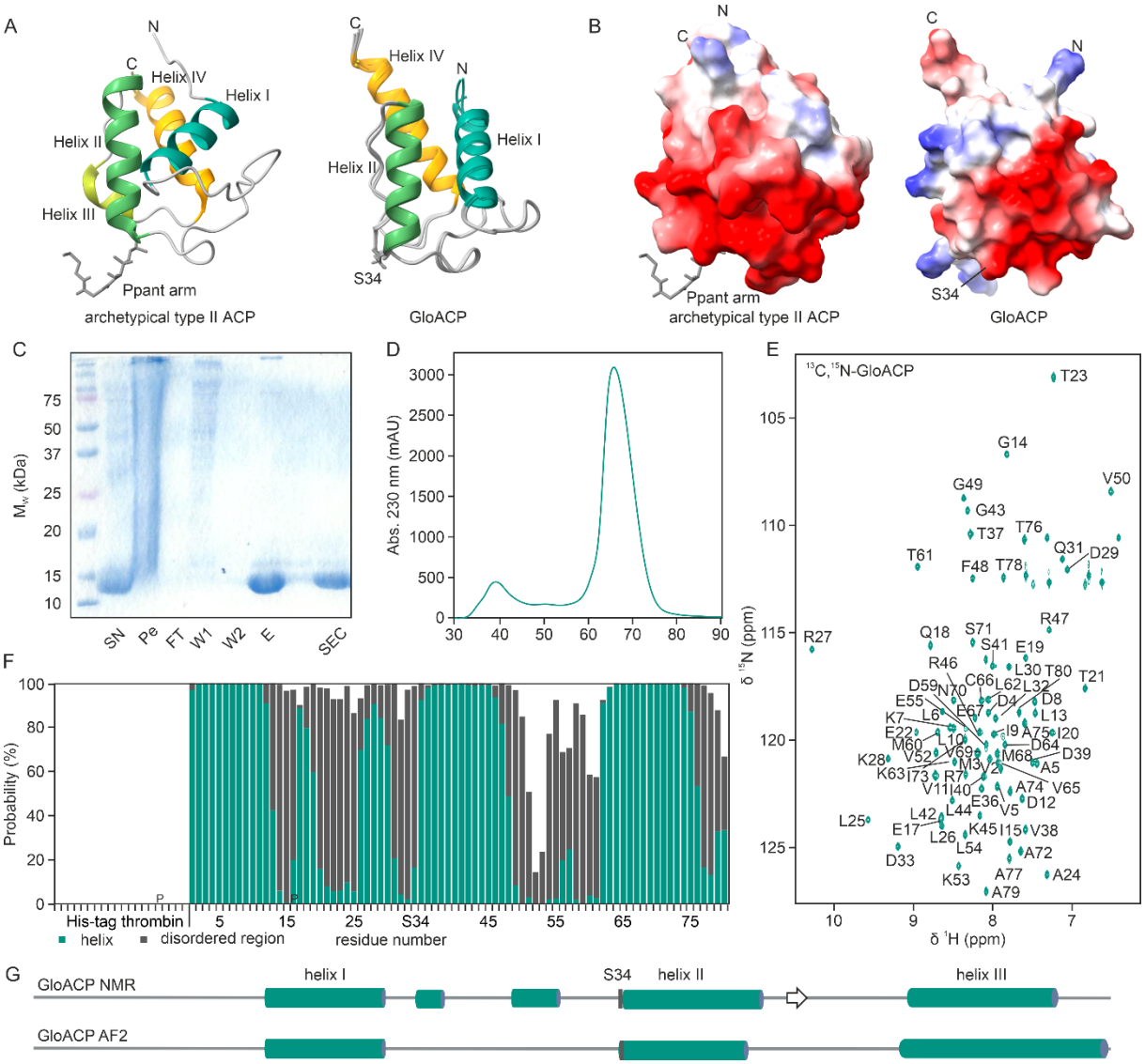


Figure 27: Structural investigation of GloACP from Gloeocapsa sp. PCC 7428. A: An archetypical type II ACP (PDB: 2KOX) compared with superposition of five AlphaFold2 structure predictions of the GloACP protein. The reactive serine residue (S34) and the phosphopantetheine arm (Ppant arm) are marked as grey sticks. B: Electrostatic surface potential of an archetypical type II ACP (PDB: 2KOX) and GloACP based on the AlphaFold2 model. Positive charges in red, negative charges in blue, neutral residues in white. C: SDS-PAGE showing purification of GloACP. SN: supernatant, Pe: pellet, FT: flow through of the affinity column, W1/W2: washing steps, E: elution, SEC: sample after size exclusion chromatography at ~70 mL elution volume. D: Size exclusion chromatogram of the purified GloACP using a Superdex 75 16/60 column. E: $^{1}H,^{15}N$-HSQC spectra and backbone assignment of $^{13}C,^{15}N$ GloACP. F: Chemical shift based secondary structure prediction of the GloACP using TALOS-N (88). Unassigned peaks are left blank, prolines are marked with P. G: Comparison of the GloACP topology models by AlphaFold2 and TALOS-N.

To determine the secondary structure of the GloACP based on the backbone NMR chemical shifts, the TALOS-N prediction tool was used (*88*). GloACP consists of three helices characteristic of carrier proteins, a short beta strand, and two short helices between helices I and II (Figure 27F

and G), which are found in PCPs of NRPS multienzyme complexes but are unusual for ACP-type carrier proteins (*28, 19*). Furthermore, the TALOS prediction based on measured chemical shifts of GloACP in solution does not fully agree with the AF2 model. While AF2 also predicted three core helices, their length and the loop architecture between helix I and II differs.

To activate apo GloACP for functional and structural investigation, PPTase has been used. This enzyme catalyzes the phosphopantetheinylation of ACPs by transferring a 4′-phosphopantetheine moiety from coenzyme A (CoA) to a conserved reactive serine on the carrier proteins (*57*). Accordingly, in the type II PKS gene cluster from *Gloeocapsa sp. PCC 7428*, a PPTase can be found. However, until now, *in vitro* phosphopantetheinylation of GloACP with GloPPTase was not successful, possibly due to the loss of PPTase activity during heterologous expression or purification or because the right conditions for the *in vitro* assay have not yet been established. However, it should be mentioned that the PPTase from *B. subtilis,* Sfp, which is widely used for ACP pantetheinylation due to its promiscuity with regard to both the CoA and receptor protein substrates (*100*), also could not activate GloACP. In the literature, it is described that the PPTase Sfp activates carrier proteins that contain of a GxD/HShxxh motif (x any residues, h hydrophobic residues) (*101*). Yin *et al*. showed that the shortest peptide that can be phosphopantetheinylated by Sfp consists of eleven residues beginning with an N-terminal DSL tripeptide motif (*102*). Accordingly, the residues N-terminal to the D/HSh motif in ACP do not play a crucial role to the Sfp recognition. In addition to the DSL motif, Dötsch and co-workers show the importance of the glycine residue in the GxD/HShxxh motif to be crucial for the correct positioning of the PCP helix II for the Sfp catalyzed pantetheinylation reaction (*101*). In GloACP, residues ^{31}QLDST^{35} form the active site, thus not fully matching the expected Sfp recognition. To gain insights into the interaction between GloACP and Sfp, an AlphaFold2 prediction of the complex was carried out (Figure 28A-C, GloACP cyan, Sfp blue). The interface areas are 791 Å^2 on GloACP and 803 Å^2 on Sfp. Additionally, the double mutant GloACP Q31G, T35L was predicted in complex with Sfp (Figure 28A-C, GloACP Q31G, T35L grey, Sfp black) to investigate the role of these two mutations.

A comparison of the two ACP-Sfp models show that the mutations Q31G and T35L affect the GloACP Sfp interaction. The sidechain of Sfp Y36 faces the GloACP Q31 sidechain when interacting with the wildtype ACP protein but flips by almost 180° when interacting with the GloACP mutant (Figure 28B). It is possible that the orientation of the Sfp Y36 sidechain might influence the orientation of Sfp R34 and H37 sidechains. Thus, the charge of GloACP Q31 at this binding interface might influence the activity of PPTase Sfp.

On the other mutation site, T35L, Sfp K112 and I114 are directly affected when comparing the wildtype GloACP with the GloACP Q31G, T35L (Figure 28C). The orientation of the sidechain I114 might influence the orientation of the residues located in the helix next to it (L116, E117 and R121). The ACP T35L mutation changes the hydrophobicity of the residue at this position. The amino acid property at this position might be important for the stabilization of the Sfp ACP interface. Hydrophobic interactions can be found between Sfp I114 and ACP L35. In contrast, ACP T35 would probably form a hydrogen bridge to Sfp K112. A hydrogen bridge between GloACP and Sfp might result in a stable complex and could possibly negatively influence the Sfp phosphopantheinylation activity and turnover.

In the model with the native GloPPTase the interaction between PPTase and ACP is formed by the ACP helix II with one GloPPTase helix (Figure 28D). This binding interface matches the ACP-AcpS-type PPTase interaction (*103, 104*) and covers the interaction area of 772 Å^2 on the GloACP and 714 Å^2 on the PPTase. Since AcpS-type PPTase forms higher oligomers (*105*), it is possible that the predicted AlphaFold2 model is not complete. Comparing the models formed by GloPPTase and GloACP wt or GloACP Q31G, T35L, only small differences are observed. ACP L35 could probably stabilizes the adherent leucine residues (GloPPTase L41 and L53) by hydrophobic interaction (Figure 28E). Unfortunately, this model does not explain the role of residues Q31 and T35 for the phosphopantheinylation reaction.

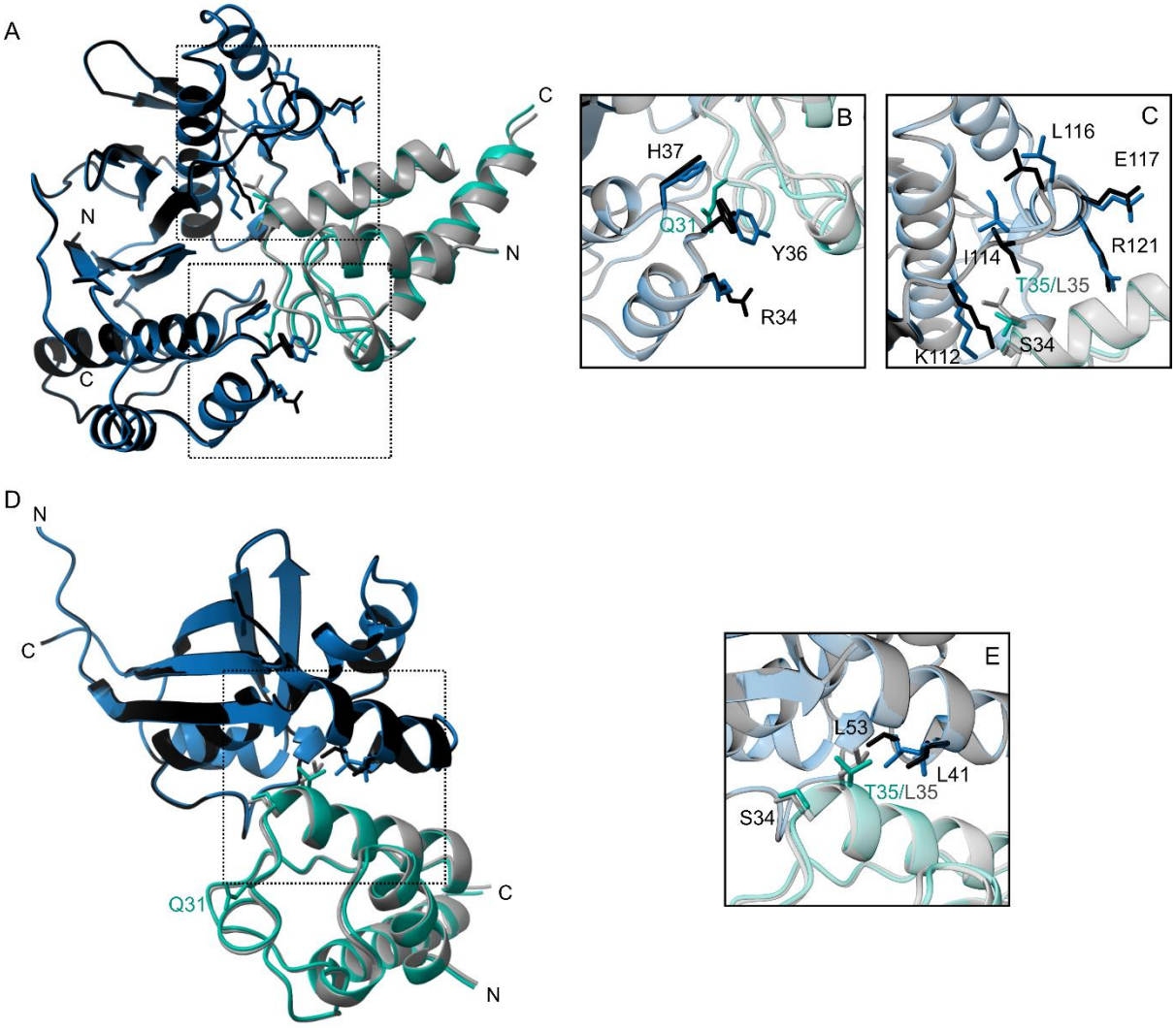


Figure 28: AlphaFold2 models for the interaction between variants of GloACP from Gloeocapsa sp. PCC 7428 with two different PPTases. A: AlphaFold2 model of the B. subtilis PPTase Sfp (blue and black) and GloACP wt (cyan) or GloACP Q31G, T35L (grey). B and C: Close up view of the GloACP wt (cyan) or GloACP Q31G, T35L (grey) binding interface with Sfp (blue and black). D: AlphaFold2 models of the complex between wt GloPPTase with the wt GloACP (blue and cyan, respectively) and the wt GloPPTase and GloACP Q31G, T35L (black, grey). E: Close up view of the GloACP wt (cyan) or GloACP Q31G, T35L (grey) binding interface with GloPPTase (blue and black).

To test whether GloACP may be activated by Sfp when the recognition site adheres to the consensus sequence, the double mutant GloACP Q31G, T35L was created by our collaboration partners (____, ____ Lab, ____). They showed that GloACP Q31G, T35L can be successfully pantetheinylated by the PPTase Sfp (data not shown). Furthermore, phosphopantetheinylated

GloACP Q31G, T35L interacts with other catalytical proteins from the PKS type II gene cluster but also non-native proteins, such as malonyltransferase FabD from *E. coli*. The attachment of the substrates, malonic acid and benzoic acid, to the holo GloACP by transferases and ligases, respectively, could be monitored by LC-MS. The activated substrate is transferred to the ketosynthase/chain length factor protein (KS-CLF) complex by the ACP, resulting in an unloaded holo ACP with a free thiol group at the Ppant arm. These experiments were done by _____and _____ (_____ Lab, _____).

Even though, the _____ lab could show that holo GloACP Q31G, T35L interacts with other catalytical proteins of the *Gloeocapsa sp. PCC 7428* PKS Type II gene cluster, the functional *in vitro* reconstitution of all proteins from the gene cluster could not be recreated until now. Upon mixing of all purified components of the *Gloeocapsa* PKS, the natural product, which is expected to be an aromatic polyketide, could not be detected. Also, non-aromatic polyketides, which indicates a successful repetition of the catalytic PKS cycle, could not be found. It is therefore possible that additional proteins or co-factors are missing. Since all experiments have been performed with GloACP G31G, T35L, it is possible that this carrier protein is not fully suitable for the PKS system. Evolutionary, there must be an explanation for this PKS gene cluster to choose a carrier protein with the unique ^{31}QLDST^{35} motif instead of the highly conserved GxD/HSh motif (x any residues, h hydrophobic residues) (*45, 19*). The hypothesis is that interaction partners of GloACP are selected by the ^{31}QLDST^{35} motif and thus regulates the production of the polyketide produced by this gene cluster.

1.5. Conclusion

Multimodular megaenzymes, such as non-ribosomal peptide synthetases (NRPSs) (*106*) and polyketide synthases (PKSs) (*93*), form production machines for a vast selection of bioactive compounds. These molecules play many important roles in chemical signaling and microbial communication, including e.g., the prevention the growth of competing bacteria strains and predators. Understanding biosynthetic cycle of these megaenzymes, their architecture, organization and the molecular interaction of individual domains and functional modules will unlock the secret language of bacterial communication while opening up opportunities to mimic and optimize synthetic mechanisms for bioengineering approaches.

Multimodular enzymes can reach tremendous protein sizes and are functionally dynamic, thus rendering them challenging for structural biology approaches (*107*). Small monomodular synthases consisting only of a minimal set of functional domains, could make structural studies on a full-length system more feasible, thereby enabling insights into regulatory mechanisms and substrate selectivity.

Nonetheless, even monomodular enzymes can still reach protein sizes of at least 150 kDa, as is the case with the monomodular NRPS Pys from *Pseudomonas fluorescens (21, 32)*. In case the protein additionally undergoes major conformational changes during its synthesis cycle (*108*), even a monomodular NRPS will remain a challenging system for biochemical studies.

One domain that has been shown to move dramatically during biosynthesis is the carrier protein (*109*), among the highest conserved and most important domain in biosynthetic megaenzymes. The carrier protein is responsible to transport activated substrates and intermediates between the catalytic domains, thus the functionality of the entire megaenzyme depends on the carrier protein's ability to correctly interact with partner domains and molecular substrates (*57*). In this thesis, the full-length monomodular NRPS Pys and two different carrier proteins from two PKS were analyzed and compared to find similarities and differences in the structure, binding partners, and function.

1.5.1. Small but mighty – the monomodular NRPS Pys from *Pseudomonas fluorescens*

The amoebicidal compounds pyreudiones A-D are produced by the smallest NRPS known to date, Pys from *Pseudomonas fluorescens (21, 32),* by fusing proline with a short-chain fatty acid. When expressed in a heterologous system, i.e., *E. coli* HM0079 with an integrated surfactin-type phosphopantetheinyl transferase (Sfp-PPTase) gene from *Bacillus subtilis*, the Pys protein shows high efficiency and *in vivo* turnover rates (*21*). In contrast, the purified Pys shows negligible pyreudione production *in vitro*. However, when heterologously expressed and purified apo Pys was activated by the *B. subtilis* Sfp, this also resulted in the successful *in vitro* synthesis of pyreudiones.

Structural studies of full-length Pys with X-ray crystallography and cryo-electron microscopy (CryoEM) were challenging due to the inherent flexibility of the protein. A model to demonstrate the domain movements during pyreudione synthesis is shown in Figure 29. The high flexibility observed is consistent with the dynamics described in the literature for this protein family (*110, 111*) and could not be reduced by adding substrates in the form of proline, ATP, or the combination of both.

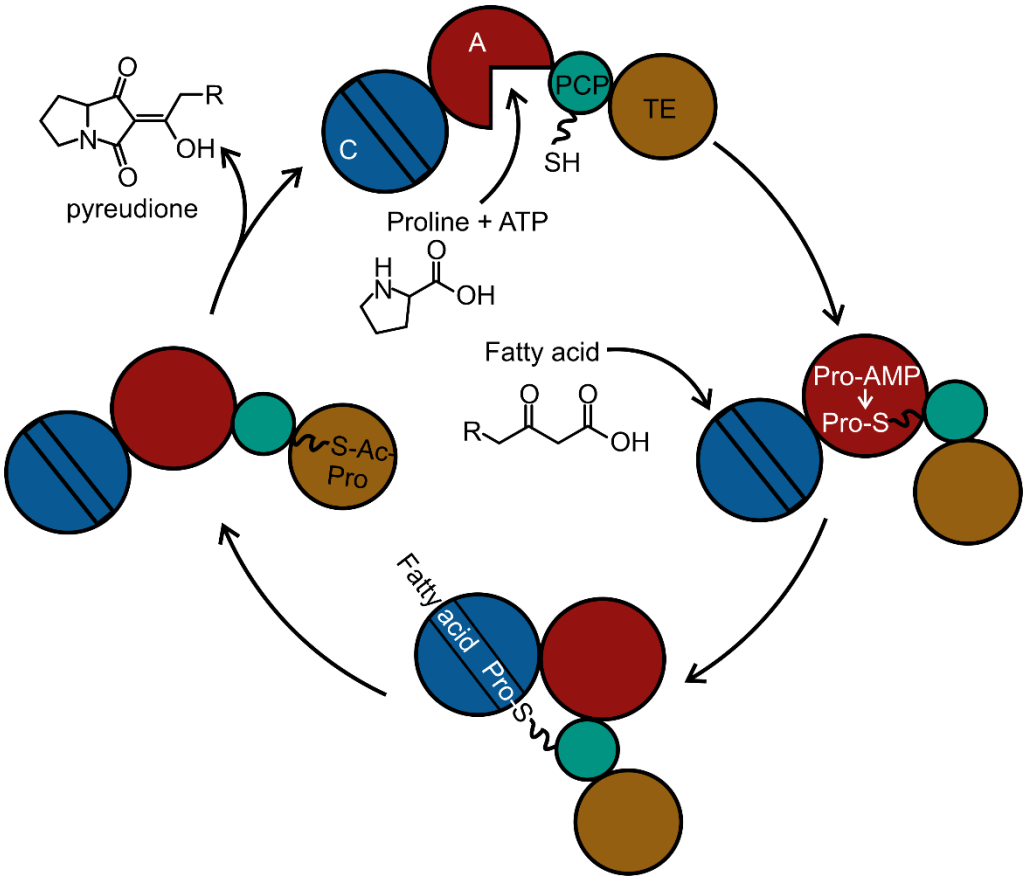


Figure 29: Mechanistic model demonstrating the domain movements of the pyreudione synthetase (Pys) from Pseudomonas fluorescens during pyreudione biosynthesis. The monomodular NRPS Pys consists of four domains: a condensation domain (C, blue), an adenylation domain (A, red), a peptidyl carrier protein (PCP, cyan), and a thioesterase domain (TE, ocher). Proline and ATP bind to the adenylation domain, thereby causing conformational changes and subsequently activating proline via an adenylation reaction forming prolyl-AMP (Pro-AMP). The activated proline is then coupled to the phosphopantetheine arm of the PCP domain. A (modified) fatty acid binds to the condensation domain, where it is transferred onto the activated proline via a condensation reaction forming acetyl proline (Ac-Pro). This intermediate is transferred to the TE domain, where the cyclisation reaction occurs to yield the final product, pyreudione that is then released from the enzyme.

The "divide and conquer" approach has been used successfully in structural biology to overcome hurdles imposed by protein flexibility and complexity (*112–114*). Therefore, in this study, di- and mono-domain constructs of Pys were generated. Guided by the biocomputational prediction tools AntiSMASH 3.0 (*76*) and AlphaFold2 (*77, 78*), constructs consisting of the condensation (C) and adenylation domain (A) di-domain as well as mono-domain constructs of the C domain in varying lengths were designed.

The CA didomain could be expressed and purified and showed the expected adenylation activity, albeit with very low conversion rates. This may presumably be explained by the lack of the peptidyl carrier protein (PCP) in the construct and the premature release and subsequent

hydrolysis of the newly formed prolyl-AMP intermediate by water (*27*). Unfortunately, no structural changes of the CA didomain upon substrate binding and adenylation were discernible by analytical SEC and tryptophan fluorescence measurements. Likewise, neither X-ray crystallography nor cryo-EM led to a successful structure determination of the construct indicating that the di-domain, like the full-length NRPS, is quite flexible.

In vivo experiments performed by _____ (_____ lab, _____) focused on the functional properties of the Pys C domain. Experiments with NRPS chimeras, in which the C domain from Pys was exchanged with the C domain from the homologous protein acylproline synthetase (Aps) from *Pectobacterium betavasculorum,* showed that the C-domain is responsible for the fatty acid head group selectivity. A sequential and structural analysis of the Pys and Aps AF2 models C domain core motives and neighboring residues allowed to propose a possible selectivity mechanism (Figure 22). The unique combination of hydrophilic residues forming a pore for either fatty acids or functionalized 3-oxo fatty acids in either enzyme's C-domain may explain their fatty acid substrate selectivity (Figure 22).

Guided by the structural prediction using AlphaFold2, a Pys C domain construct was cloned and could be successfully purified. Encouragingly, this construct formed microcrystals in a crystallization screen. Furthermore, a 2.1 Å crystal structure of the homologous C-domain from the Aps NRPS was obtained in its apo state.

Based on this structure, the Caver Web 1.0. (*82, 81*) program was used to calculate possible tunnels in the protein that allow substrate entry and exit. Of the four calculated tunnels, two agree with those previously described for NRPS condensation domains (*110, 30*). The other two tunnels have not been described before and seem to be a unique feature for condensation domains with substrate selectivity properties. With the distance 9.3 – 13.4 Å to the newly discovered tunnel entries, a serine residue (S326) is located. This residue has been found to be phosphorylated. It is possible that residue S326 plays a role in loading the fatty acid to the condensation domain. However, this, together with the relevance of the two newly discovered tunnel entries needs to be confirmed by biochemical experiments.

In summary, this study provides the first structural insights into the fatty acid headgroup selective NRPS systems Pys and Aps and provides testable hypotheses for the molecular basis of fatty acid selectivity. The flexibility of the Pys full-length protein makes it difficult to study with classical structural biology approaches. As a complementary tool, the Alphafold2 protein structure prediction algorithm can be used to anticipate the fold and domain boundaries of proteins (*77, 78*). To reduce the complexity of the Pys NRPS for structural studies isolated domains and didomains were cloned and successfully purified. Both condensation domains from Pys and Aps crystallized. The Aps domain yielded satisfactory diffraction data and its crystal structure of Aps was solved at 2.1 Å resolution. Based on the structure and the structure predictions of Aps and Pys, it could be speculated that the substrate selectivity in the condensation domain may be regulated by the core motif and neighboring residues that form the reactive tunnel. Thus, the investigation of isolated domains can contribute to the understanding of enzyme structure and

function, even when the full-length protein may be inaccessible or too flexible for structural studies.

1.5.2. Interaction between carrier protein and ketosynthase domain might be transient in polyketide synthases

Rhizoxin produced by *Burkholderia rhizoxinica* is a potent phytotoxin that has been considered as an antitumoral therapeutic agent (*115*). Recently, it has been shown to help to protect the bacterium against protozoan and metazoan predators (*116*). Rhizoxin is synthetized by a NRPS-PKS type I fusion protein consisting of 11 modules (*49*). A unique branching domain catalyzes the vinylogous addition of acetate to a α,β-unsaturated thioester (*117*). This vinylogous addition step is critical for the subsequent formation of the δ-lactone ring, which is essential for the antimitotic activity of the product (*118*).

The branching module is encoded by the *rhiE* gene. It consists of a ketosynthase (KS), a branching domain (B), and an acyl carrier protein (ACP) (*49*). A KSB didomain crystal structure is available (*51*), but the entire branching domain was not accessible to both X-ray crystallography and CryoEM. AlphaFold2 predictions of the branching module show that there are two possible binding sites for the ACP domain on the KSB didomain (Figure 24). In one binding stance, the distance between ACP reactive serine and the KS catalytic site are 15.356 Å, while in the other orientation, these two sites are 40.307 Å apart. The existence of both binding sites needs to be verified with experimental data. NMR spectroscopy was used to shed light onto the ACP structure of the branching domain and to investigate the interaction between ACP and KSB didomain. To investigate which ACP residues are involved in the KSB binding, a backbone NMR assignment of $^{13}C,^{15}N$ labeled *rhiE*-ACP was carried out and the effect KSB binding was monitored by chemical perturbations upon KSB didomain titration. The interaction hot spot clearly centers around the reactive serine residue (S1061) of the ACP domain, while other residues showed no or only negligible chemical shift perturbations upon KSB addition. The ACP reactive serine is important to deliver substrates bound to the attached pantetheine arm to the active site of the KS domain (*86*).

To investigate the interaction of the KSB didomain with ACP, a different approach was chosen due to the size of the KSB domain (107.8 kDa). Since this is too large for conventional ^{1}H, ^{15}N NMR approaches, the protein was instead labeled with ^{19}F-tryptophan. KSB contains ten native tryptophan residues in its sequence, and the 19F NMR spectrum of the protein shows that the label was successfully incorporated. The KSB di-domain was titrated with 6.7-fold molar excess of ACP. However, no chemical shift perturbations were observed upon ACP addition. Either the tryptophan residues are not suitable reporters because of their location, as their chemical environment may not be perturbed by ACP binding, or the affinity between isolated KSB and ACP domains is exceptionally low. The first argument matches the AlphaFold2 model. Low affinity

binding between KS and ACP in type I PKS system has been observed by others (*92*), which underlines the second argument.

In the context of a full-length PKS type I system, low affinity binding between carrier proteins and catalytical domains might allow rapid exchange of interaction partners without blocking the catalytic sites. Since the interaction partners are covalently linked, the binding between domains might not be precise and need only serve the function of bringing reactive substrates to the active sites.

1.5.3. Not all ACPs are the same

Despite lacking a catalytic function, NRPS/PKS carrier proteins fulfil a critical functional role in intra-enzyme communication by transporting substrates and intermediates throughout the entire enzyme (*55*). All currently described carrier proteins share a common architecture of a four α-helix bundle (*45*). At the N-terminus of the second helix, a conserved reactive serine is modified by a phosphopantetheine arm (Ppant arm) in a reaction catalyzed by a phosphopantetheinyl transferase (PPTase) (*55*). Equipped with the Ppant arm, active carrier proteins covalently attach substrates to the arm's terminal thiol group

Intriguingly, GloACP, a carrier protein from a *Gloeocapsa sp. PCC 7428* type II PKS system from studied in this thesis shows interesting structural features that distinguish it from canonical carrier domains. GloACP only has three helices, as predicted by AlphaFold2 and confirmed by an NMR chemical shift based secondary structure prediction.

Furthermore, and in strong contrast to the other ACP studied here, establishing *in vitro* phosphopantetheinylation was a big challenge with GloACP. The wildtype protein could not be activated with the universal PPTase Sfp from *B. subtilis* or its native PPTase from *Gloeocapsa sp. PCC 7428*. It was hypothesized that this was due to missing co-factors or because *Gloeocapsa sp. PCC 7428* PPTAse was not functionally expressed or purified. However, looking at the sequence of GloACP, it became apparent that this domain lacks the Sfp recognition motif GxD/HShxxh (x any residues, h hydrophobic residues) (*101, 102*). Rather, GloACP features an ^{31}QLDST^{35} motif at its phosphopantetheinylation site. Encouragingly, the introduction of two point mutations (Q31G, T35L) led to the formation of holo GloACP by Sfp activation thus showing that GloACP is structurally intact when obtained from heterologous expression sources. *In vitro* protein-protein interaction experiments performed by the _____ (_____) indicated that activated GloACP can interact with catalytic enzymes from the same gene cluster as well as malonyltransferase (MAT) from *E. coli*. However, until now, the *Gloeocapsa sp. PCC 7428* minimal PKS system could not be reconstituted *in vitro,* and polyketide products could not be detected to date. A schematic overview of the phosphopantetheinylation and substrate loading experiments is shown in Figure 30.

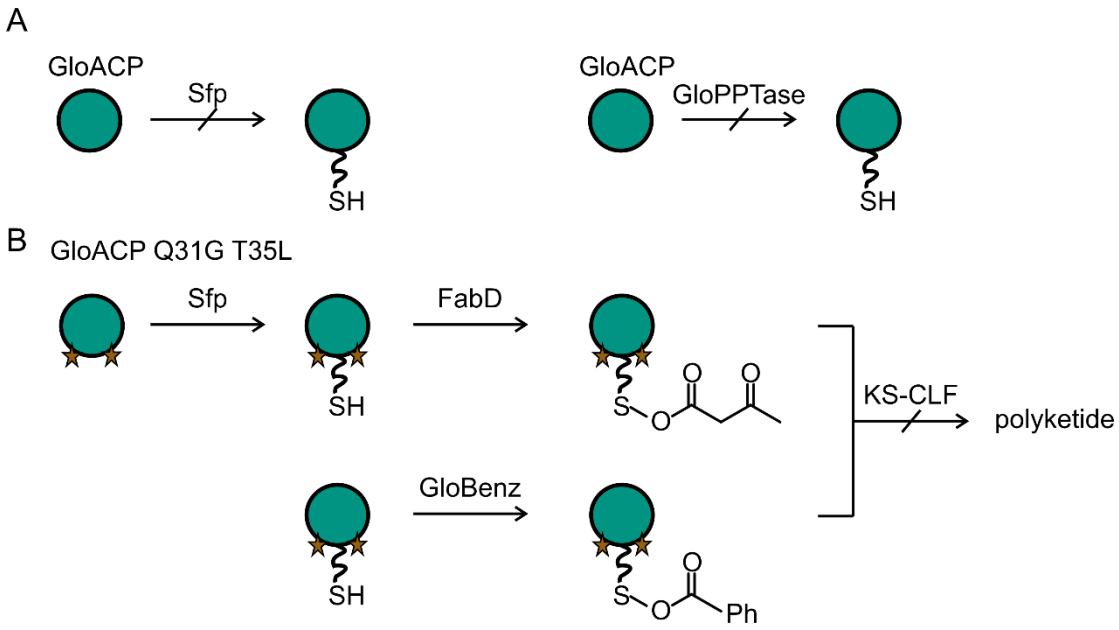


Figure 30: Schematic overview of the phosphopantetheinylation experiments carried out to obtain holo GloACP, an acyl carrier protein type II from Gloeocapsa sp. PCC 7428. A: Wildtype GloACP could not be phosphopantetheinylated by the phosphopantetheine transferase (PPTase) Sfp from Bacillus subtilis or the native PPTase (GloPPTase) in vitro. B: the double mutant GloACP Q31G T35L could be activated by the PPTase Sfp thus yielding holo GloACP. The location of the mutated residues near the phosphopantetheine arm (~SH) are indicated with stars. Substrates, such as malonic acid and benzoic acid could possibly be loaded onto the GloACP by FabD (a malonyl transferase from Escherichia coli) or by Gloeocapsa sp. PCC 7428 benzoic CoA ligase. To date, a polyketide product could not yet be detected after adding the ketosynthase with chain-length-factor domain (KS-CLF) to a substrate loaded GloACP indicating that other co-factors may be missing or that the double mutation has far reaching consequences.

It may be possible that the point mutations currently needed to activate GloACP *in vitro* also disrupt the protein's interaction with downstream partners. This demonstrates that the activation of the carrier protein is essential for polyketide biosynthesis, but not sufficient to keep the system running. In particular in a "free floating" PKS type II system, where the protein components are not covalently attached, the carrier protein is a major player. It directs the catalytic proteins, selectively binds to them in a specific order and thus controls product biosynthesis.

The three examples studied here demonstrate the complexity and uniqueness of carrier proteins in their respective biosynthetic system. To understand the function of an entire NRPS/PKS enzyme complex, the interaction between carrier proteins and catalytic domains must be explored in great detail.

1.6. Outlook

1.6.1. Structural and functional investigation of pyreudione and acyl proline synthetases

Structural and functional investigations on full-length non ribosomal peptide synthetases (NRPS) are challenging due to their size and flexibility. The handily size of monomodular NRPSs should thus confer a significant advantage when studying producing enzymes. However, here I found that even monomodular NRPSs such as the pyreudione synthetase (Pys) from *Pseudomonas fluorescent* and the acyl proline synthetase (Aps) from *Pectobacterium betavasculorum* are still very challenging systems. Both proteins catalyze the coupling between proline and fatty acids, resulting in pyreudione or acyl proline products, respectively. However, their only differences rely in the selection of fatty acid headgroups, which is used to synthetizing the fatty acyl proline products. Their inherent flexibility precluded structure determination of the full-length proteins by X-ray crystallography and CryoEM in the course of this thesis. To reduce the complexity of the system, a divide and conquer strategy was used. This approach enabled me to obtain the crystal structure of the apo state Aps condensation (C) domain and first micro crystals for the homologous Pys C domain, also in the apo state. Building up on these crystallization screens, the next step will be to attempt to crystallize the C domains of both proteins in their substrate bound states. The substrate proline is loaded on the C domain by the peptidyl carrier protein (PCP). A structure of the C domain in complex with the PCP domain loaded with the proline cargo attached to the phosphopantetheine (Ppant) arm would explain how proline is stabilized in the reactive tunnel of the C domain. A known problem when investigating PCP-C domain complexes is the transient PCP interactions with partner proteins (*57*). To overcome this issue, the substrates could be cross-linked within the C domain. This has been successfully demonstrated using the siderophore enterobactin synthetase from *Escherichia coli (119)*.

The co-substrate to which activated proline is fused is fatty acid. To identify the fatty acid headgroup orientation in the C domain active site, C domain loaded with a fatty acid should be structurally investigated. However, in which form the (modified) fatty acid is loaded onto the C domain remains unclear. Therefore, CoA activated fatty acid and ACP bound fatty acid should be tested. Knowledges of how the substrates are bound to the C domain might explain the role of the additional tunnels found in the Aps C domain crystal structure.

The way the substrates interact with the individual NRPSdomains, as well as interdomain contacts implied by structures should also be verified by mutation experiments and biophysical analyses. The role of specific residues should be tested *in vivo* and *in* vitro through the use of site-specific mutations. Unfortunately, a condensation assay has not been established for Pys and Aps condensation domains to date. Two approaches to investigate the condensation activity have been described in literature (*120*). The differences between these approaches rely in the form in which the substrates are used and whether the formed product is detected in solution or attached to the carrier protein. In the assay described by Bian and coworkers, aminoacyl-SNAC and acyl-CoA were mixed with the C domain and the formed product was analyzed by mass spectrometry (*120, 28*).

In the assay described by Cryle and coworkers C domain is covalently linked to an activated PCP loaded with the amino acid cargo. By adding the second substrate, the condensation product is still attached to the Ppant arm of the PCP, thus can be analyzed by mass spectrometry (*28*).

Substrate binding affinity, the protein's thermal stability and its dynamics should be investigated to understand the role of the C domain for the NRPS. A recent NMR study performed by Früh and coworkers revealed widespread structural fluctuation of the cyclization domain, a protein of the C domain family (*110*). The authors hypothesized that structural dynamics of C domains might be important for substrate loading and protein functionality (*110*). However, little is known about C domain dynamics and should be addressed.

The other domains of the NRPS, adenylation (A) domain, PCP and thioesterase (TE) domain, should also be studied with regard to their structure. The domain-domain communication should be investigated by determining the binding interfaces and changes therein between adjacent domains in a substrate-dependent manner. With the Pys and Aps proteins as two related NRPS with distinct substrate specificities in hand, structural and regulatory (dis)similarities should be detectable that will not only shed light on the mechanism of monomodular NRPS systems and details of substrate specificity but also provide insights into multimodular NRPSs. Ultimately, understanding the mechanism of NRPS will enable the rational design of therapeutic peptides, which are used in the treatment of intestinal and gastric diseases, and in peptide vaccines against viral infections (*121*).

1.6.2. Exploring the carrier protein interaction with catalytic partner proteins

In linear producing enzymes, the order of catalytic events is regulated by the protein domain arrangement in a specific structural sequence (*122*). Shuttled by carrier proteins, substrates and intermediates are transported through the entire enzyme (*123*), thus making carrier proteins key players in producing enzymes. However, the spatiotemporal regulation of domain crosstalk in complex biosynthetic machineries remains unclear. Here, I have seen that carrier proteins bind enzymatic domains with low affinity as seen for the branching module in the rhizoxin synthase from *Burkholderia rhizoxinica* (chapter 1.4.5). Similar observations have been described in literature for linear producing enzymes (*55*). To investigate these transient interactions, chemical probes to crosslink carrier proteins with enzymatic partners could be used. This approach could be helpful to stabilize transient complexes for X-ray crystallography and CryoEM. Examples of successful linkage between catalytic domains and carrier proteins using chemical probes are summarized in a review by Gulick and Aldrich (*55*). In the case of the *rhiE*-acyl carrier protein studied here, which interacts with a ketosynthase domain, crosslinking with phosphopantetheine or thiocyanate probes could confirm or refute the existence of the proposed binding sites as predicted by AlphaFold2 (Figure 24).

Conformational changes in producing enzymes could be monitored by Hydrogen deuterium exchange mass spectrometry (HDX-MS) and Förster resonance energy transfer (FRET)

spectroscopy to understand how enzymatic turnover and conformational dynamics are coupled (*123*). SAXS has been shown to be a powerful method to investigate structure and conformational changes of multidomain producing enzymes in solution (*124, 125, 108*). These techniques should be performed with Pys, Aps and the branching module of the rhizoxin synthase to obtain a better understanding of domain motions and the mechanism during biosynthesis of natural products.

Comparing the interaction site(s) and dynamics of linear producing enzymes with free standing multienzyme assemblies, the underlying mechanisms of protein-protein interactions during the catalytic cycle of NRPS and PPKs could potentially be proposed. This knowledge might allow the mixing of producing systems to engineer enzymes with controlled function.

1.6.3. Exploring the regulatory mechanism of free-standing type II carrier proteins and their role in natural product synthesis

In contrast to linear producing enzymes, where the order of catalytical events is regulated by the arrangement of the protein domains in a structural sequence, the synthesis process of free standing multienzyme assemblies is controlled by its carrier protein. How the carrier proteins conduct the natural product syntheses is still unclear. Here, a unique carrier protein has been discovered in PKS type II system from *Gloeocapsa sp. PCC 7428* that does not follow the canonical 4-helix bundle fold typically seen for these domains. The carrier protein in this system seems to be highly selective with respect to its interaction partners. However, although it has been shown to interact with different enzymes, a natural product or polyketide-like fragment could not be detected to date. Before polyketide synthesis, the inactive apo ACP is activated by phosphopantetheine transferase yielding holo ACP. Holo ACP then takes part to the polyketide synthesis cycle and interacts with malonyl acyl transferase or benzoic acid CoA ligase to yield substrate-bound ACP. Then the cargo-loaded ACP binds to the ketosynthase (KS) in complex with chain-length factor domain. The substrate is transferred from the ACP to the KS, which allows the ACP to be recycled for another substrate loading step. In this context, many open questions remain about the specific binding sites between the carrier protein and its different binding partners, the respective binding affinities, and the sequence of interaction events. Docking experiments could be performed to get an idea of the interaction interfaces. However, *in silico* experiments should always be paired with wet lab experiments. NMR titration experiments could be performed to investigate the interaction interface between the ACP and different catalytic proteins. Protein dynamics, binding affinities and thermal stability should be studied to understand the ACP – enzyme complexes better.

Using free standing multienzyme assemblies, it may be possible to mix proteins from different systems to create a toolbox of different enzymes with known function. Since all proteins involved in the multienzyme assembly are already free-standing, product synthesis can be carried out entirely *in vitro* without the need to genetically modify the enzymatic proteins, as is the approach for linear systems (*126, 127*). The enzyme toolbox will facilitate sustainable synthesis that is

capable to produce aromatic polyketides while being competitive with chemical synthesis methods.

1.7. Appendix I

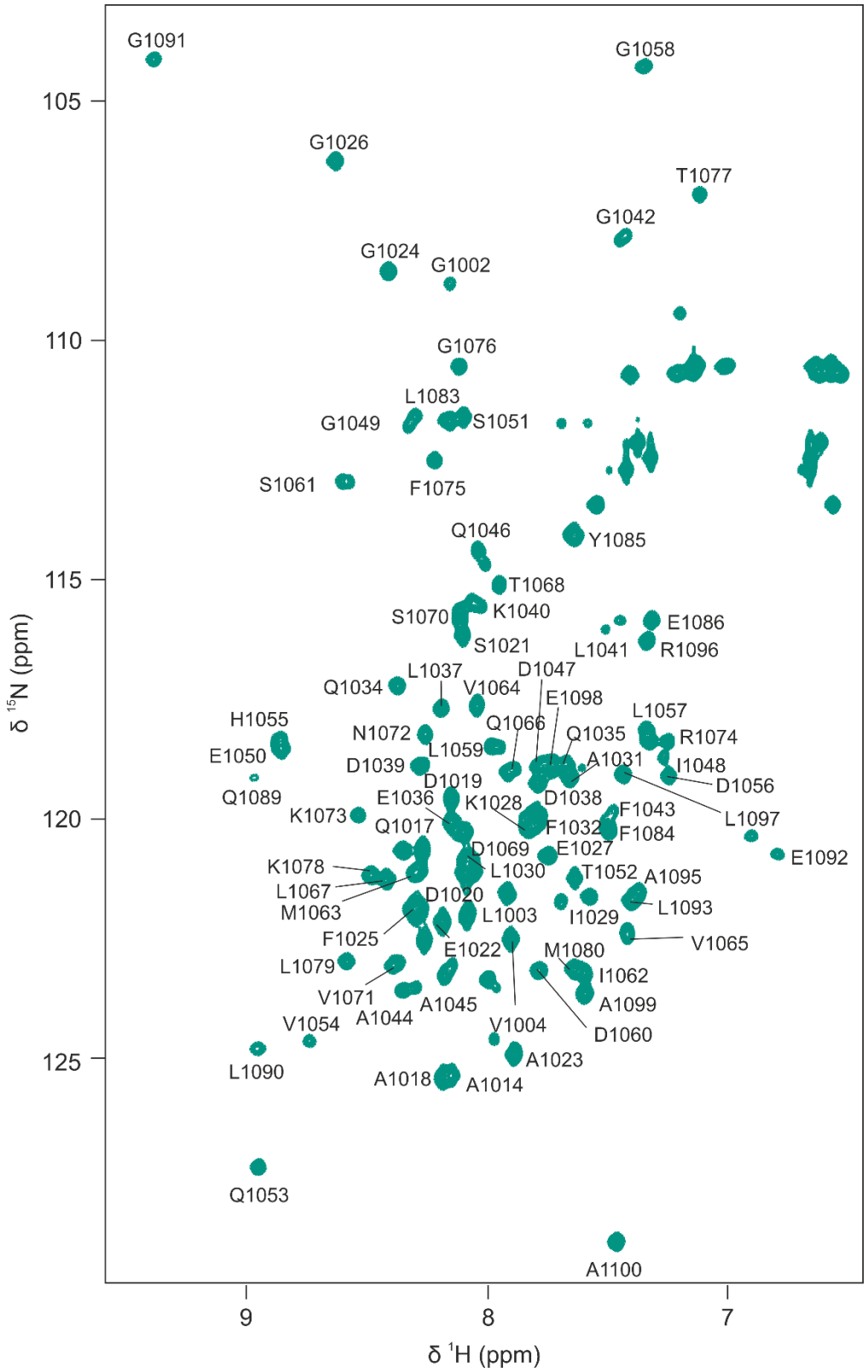


Figure 31: ^{1}H-^{15}N HSQC and assignment of the rhiE-ACP from Burkholderia rhizoxinica.

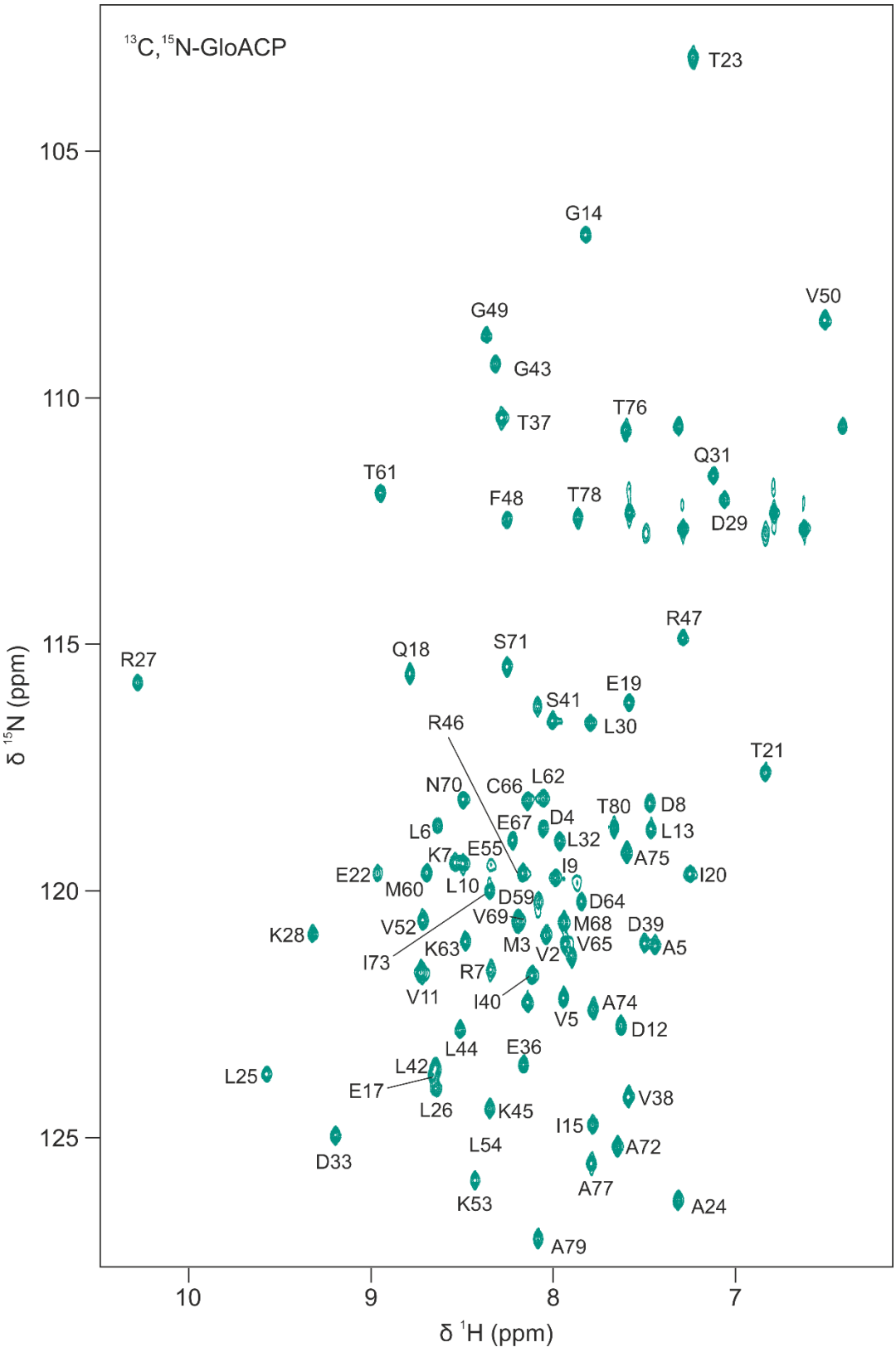


Figure 32: ^{1}H-^{15}N HSQC and assignment of the Glo-ACP from Gloeocapsa sp. PCC 7428.

Pys N-terminal His_6-tag TEV from *Pseudomonas fluorescens*

atg ggc agc agc cat cat cat cat cat cac agc agc ggc gaa aat cta tat ttt caa ggc

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

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

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

Pys C-terminal His_6-tag TEV from *Pseudomonas fluorescens*

```
atg cgt gac gta cac ccg tgc ccg ctt tcc acg gcg caa cag gaa gtt tgc ctg gcg atc
 M   R   D   V   H   P   C   P   L   S   T   A   Q   Q   E   V   C   L   A   I
agc cag agc cac agc aat ctc aac tac cac ttg tgc gac gtc atc gaa ctg cgc ggc gag
 S   Q   S   H   S   N   L   N   Y   H   L   C   D   V   I   E   L   R   G   E
```

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

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

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

Gloeocapsa sp. PCC 7428 ACP

```
atg gtc atg gat gcg cta aaa gac atc tta gtt gat tta gga att ccc gaa caa gag att
 M   V   M   D   A   L   K   D   I   L   V   D   L   G   I   P   E   Q   E   I
aca gaa aca gca ctg ctg cgc aaa gac ttg caa ctc gac tca aca gaa acc gtc gat att
 T   E   T   A   L   L   R   K   D   L   Q   L   D   S   T   E   T   V   D   I
tcc ctt gga ctc aaa cgc cgc ttt ggt gtc aat gtc aaa cta gaa tcc cgc aag gac atg
 S   L   G   L   K   R   R   F   G   V   N   V   K   L   E   S   R   K   D   M
aca tta aaa gat gtt tgc gag atg gtt aac agc gcg atc gcg gct acg gct acg gcg aca
 T   L   K   D   V   C   E   M   V   N   S   A   I   A   A   T   A   T   A   T
```

tga
 -

Bacillus subtilis R4-4 sfp

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

Chapter 2

The transporter

Chapter 2: The transporter

2.1. Theoretical background

2.1.1. ABC transporter and its role in substrate transport

As seen in chapter I, bacteria use elaborate cytosolic molecular machines to synthesize numerous bioactive compounds. These molecules then need to be exported from the cell in order to play a role in e.g., cell-cell communication. In addition, many natural products are toxins or antibiotics, and thus the producing cell needs to protect itself against these compounds. ATP-binding cassette (ABC) transporters are an important protein family that play a role in e.g., bacterial antibiotic resistance due to the enhanced extrusion of such molecules from the cell (see below). Members of the superfamily of ABC transporters are found in all kingdoms of life where they can act as both exporters and importers for a wide variety of molecules against a concentration gradient using ATP hydrolysis as the energy source.

As shown for the Quorum-Sensing (QS) mechanism in Gram-positive *B. subtilis*, the homodimeric ABC exporter SpaT is responsible for the export of the intracellularly synthesized and modified precursor lanthipeptide of subtilin (*128*) (Figure 33, left). Extracellularly, the precursor peptide is cleaved by proteases to give rise to a fully active lanthipeptide (*129*). To ensure self-immunity, *B. subtilis* expresses the heterodimeric ABC exporter SpaFEG, which exports the active lantibiotic when it is found intracellularly (*130*) (Figure 33, middle). Although bacteria can use various secretion systems to export toxic molecules from cells, antimicrobial peptides, especially microcins and bacteriocins, are exported only by ABC exporters (*10*). Interestingly, substrate immunity and resistance can also be provided by multidrug transporters that transport a wide range of substrates through the membrane. In *B. subtilis*, cervimycin C antibiotic resistance is caused by the expression of the homodimeric multidrug transporter *B*acillus multidrug-resistance ATP protein (BmrA) (*131*). In addition to cervimycin C transport, BmrA is able to transport a variety of drugs and hydrophobic molecules, e.g., Hoechst 33342, doxorubicin, and 7-amino-actinomycin-D (*131*) (Figure 33, right). In this thesis, *B. subtilis* BmrA was used as a model ABC transporter to investigate its function in different membrane mimicking environment.

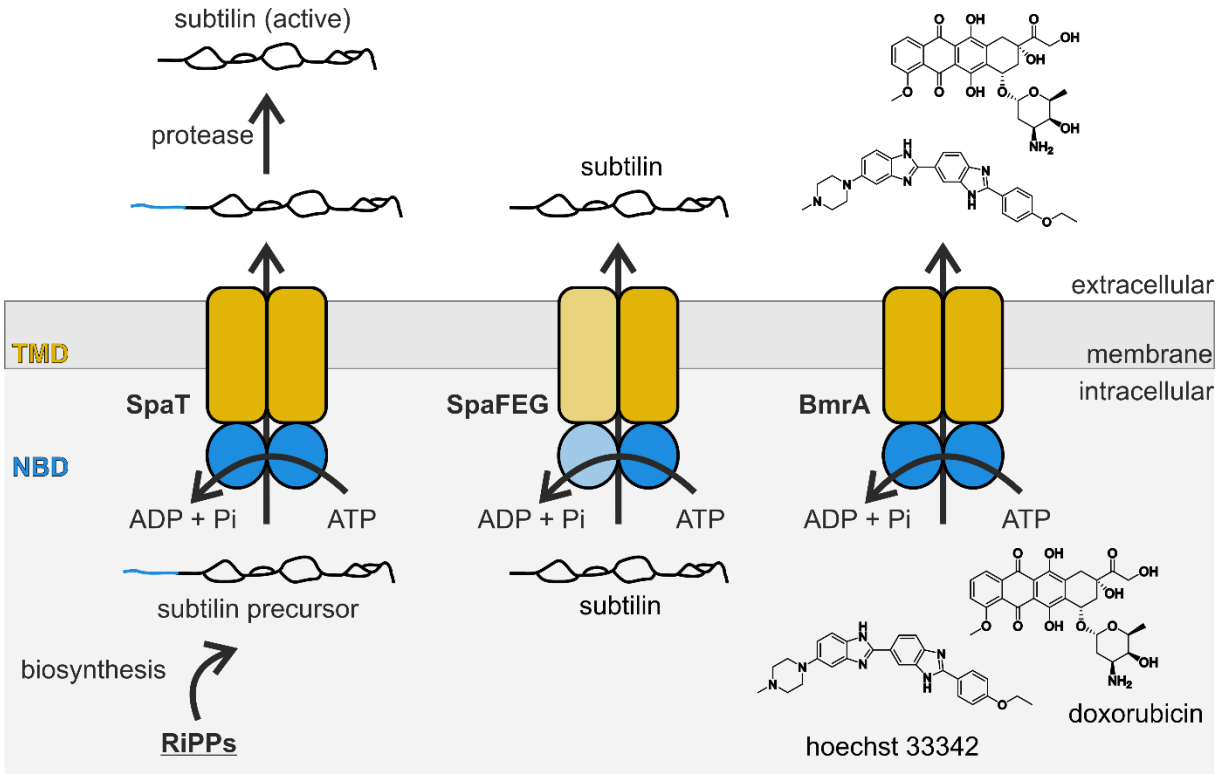


Figure 33: Overview over three ABC exporters in B. subtilis involved in lanthipeptide extrusion and multidrug resistance. SpaT transports an intracellularly synthesized subtilin precursor peptide to the extracellular region, where it can be cleaved by a protease to form active subtilin. SpaFEG protects the cell from subtilin by exporting it when it is found intracellularly. BmrA is a multidrug transporter, that can recognize and extrude more than one substrate. Transmembrane domains are colored yellow, nucleotide binding domains are colored blue. Homo and heterodimers are colored either in the same or in different color shades, respectively.

2.1.2. ABC transporter structure and function

ABC transporters share a common architecture containing two variable transmembrane domains (TMD) (Figure 33) with a substrate binding site, and two highly conserved cytoplasmic nucleotide-binding domains (NBD) that bind and hydrolyze ATP (*132*) (Figure 33). Bacterial ABC transporters are often "half transporters", i.e., one TMD is fused to one NBD. These half-transporters then form functional homodimers or heterodimers (Figure 33). The NBDs consist of a RecA-type ATP-binding core with a six-stranded β-sheet surrounded by four α-helices (*132*)(Figure 34A). This ATP-binding core is supplemented with an ABC-specific-three-stranded antiparallel β-sheet and an α-helical subdomain (*132*) (Figure 34A). In the NBD the following sequence motifs and structural elements can be found (*131*): the Walker A (GxxGxGK(X/T – motif, where x denotes any residue) (*133*), Walker B (hhhhDE, where h denotes a hydrophobic residue) (*133, 134*), signature motifs, also known as C-loop (LSGGQ(K/R)Q) (*135*), A- (F/KxY), D- (SALD), Q- (hVS/PQ) (*136*), X- (TRVGDKGTQ) (*137*) and the H-loop (hAHRL). All conserved motifs and their location in the NBD are shown in Figure 34B.

The nucleotide's phosphate groups are coordinated by the Walker A motif (*138, 139*), and the A-loop interacts with the nucleotide base through π-π-stacking (*140*). The Walker B motif coordinates the crucial Mg^{2+} ion through an aspartate side chain and catalyzes ATP hydrolysis by polarizing the attacking water molecule with its catalytic glutamate (*138, 134*). The histidine of the H-loop stabilizes the catalytic transition state of the hydrolysis (*140–142*), while the glutamine of the Q-loop coordinates the Mg^{2+} ion, senses the ATP's γ-phosphate, and plays an important role in NBD-TMD communication (*140, 136*). The D-loop plays a role in NBD-NBD communication (*140, 143*) and the C-loop contributes to ATP-binding-induced NBD dimerization (*135*). The X-loop is only present in exporters and contributes to NBD-TMD communication (*140, 137, 144*).

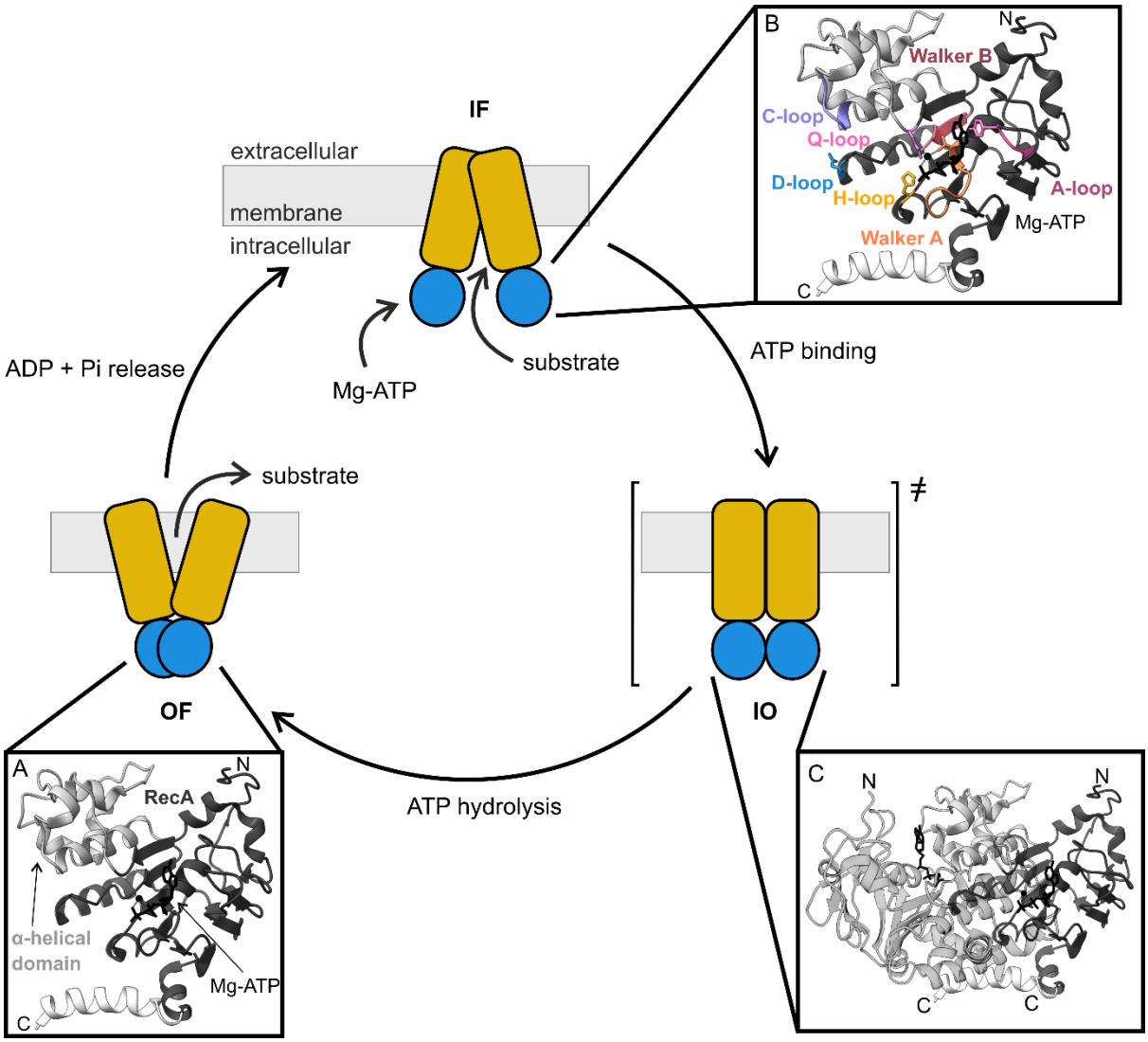


Figure 34: Transport cycle of an ABC exporter and conserved structural elements of the NBD. The model of the transport cycle of ABC exporters in the center shows that substrate and Mg-ATP binding occur in the inward facing conformation (IF). ATP binding causes NBD dimerization, leading to an intermediate inward occluded state (IO). ATP hydrolysis causes conformational changes yielding an outward facing (OF) state, where substrate release occurs. ADP and phosphate (Pi) release resets the transporter to the IF state. A: BmrA NBD (PDB: 7ow8) with the RecA-type domain marked in dark grey, α-helical domain in light grey, Mg-ATP colored black and C-terminal helix in white. B: BmrA NBD with the conserved motifs in color. Mg-ATP is colored black. C: Side view of the dimerized NBD. Mg-ATP colored black, sub domains of one NBD are colored as in panel A. The left NBD is transparent for clarity.

At the beginning of each transport cycle, the ABC exporter is in an inward-facing (IF) conformation, allowing the protein to bind substrates (Figure 34). After stochiometric ATP binding to the NBDs, the NBDs dimerize (Figure 34C). The dimerization results in structural rearrangements in the TMDs to reach an inward occluded (IO) state followed by the outward-facing (OF) conformation (Figure 34). The coupling interface between NBDs and TMDs are short TMD helices, also known as coupling helices, which interact with a groove on the NBD surface (*145*). The transmission from IF to OF allows substrate transport and ATP hydrolysis to occur. The release of Pi and/or ADP destabilizes the NBD dimer and restarts the transport cycle (Figure 34).

The alternation between an inward and an outward facing state with different substrate affinities (high affinity in the IF, and low affinity in the OF state for exporters) is named 'alternating access mechanism'. Although the transport cycle includes substrate recognition and translocation, ATP binding and hydrolysis, and the global structural changes have been elucidated, the mechanism remains unclear at the molecular level.

However, before open questions on mechanistic details and protein dynamics can be investigated, protein purification is the bottleneck in membrane protein research, thus also in ABC transporter research.

2.1.3. How to isolate membrane proteins?

A successful protein purification requires that protein stability and functionality are maintained throughout the purification process. Membrane proteins are particularly challenging to purify as their hydrophobic region needs to be maintained in a hydrophobic environment at all times to prevent aggregation. Historically, detergents have been the most used tool for membrane protein purification.

2.1.3.1. With detergents

Detergents are amphiphilic molecules composed of a hydrophobic hydrocarbon tail and a hydrophilic head group. In solution detergents form discoidal structures, also known as micelles. The minimum concentration needed to form micelles is called "critical micelle concentration" (CMC). The CMC depends on the molecular structure of the surfactant, the solvents, and additives, such as salt (*146*). During membrane protein purification, detergent micelles stabilize the hydrophobic transmembrane region of the protein with their hydrophobic tails, leaving the hydrophilic regions exposed to solvent. Detergents are classified into three classes regarding their head groups: ionic, nonionic and zwitterionic detergents (*147*). Ionic detergents, that include negative and positive charged head groups, have been the most efficient group in extracting membrane proteins from lipid bilayers. However, due to their adverse effects on protein-protein

interaction, they often lead to denaturation of membrane proteins. Sodium dodecyl sulfate (SDS) (Figure 35B) and sodium cholate are two common examples of ionic detergents. The most popular detergents for both functional and structure determination purposes are nonionic detergents. With their nondisruptive nature, these detergents only interfere protein-lipid interaction, while protein-protein interactions remain intact. *N*-dodecyl-β-D-maltoside (DDM) (Figure 35B) and lauryl maltose neopentyl glycol (LMNG) are prominent examples of nonionic detergents. Zwitterionic detergents with an intermediate level of harshness between ionic and non-ionic detergents, are used as alternative detergents to nonionic detergents. One of the most successful molecules from this class is lauryl dimethylamine-N-oxide (LDAO) (Figure 35B).

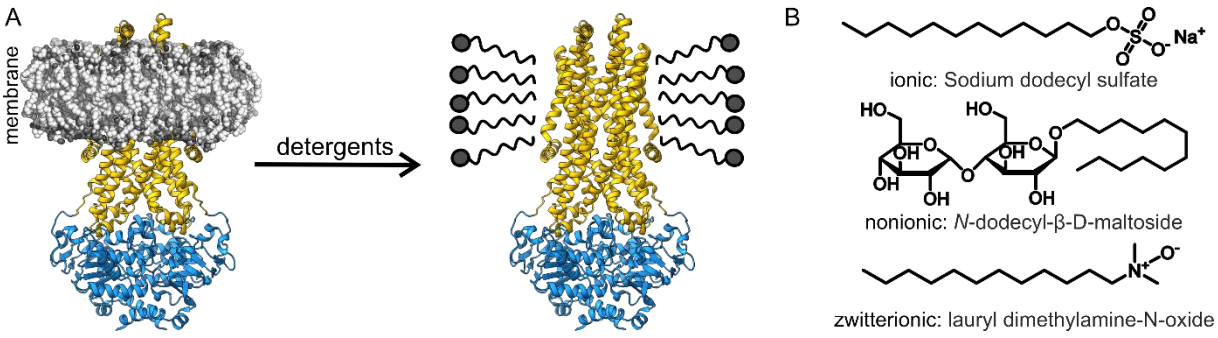


Figure 35: Scheme for membrane protein purification with detergents. A: A membrane protein is solubilized from its native environment into detergent micelles. B: Examples for the three detergent types: ionic, nonionic and zwitterionic. PDB: 7ow8.

Although detergents have been shown to be good candidates for membrane protein purification, the native membrane is disrupted and replaced by a detergent micelle (Figure 35A). The membrane protein activity is highly dependent on the choice of the detergent. A comprehensive detergent screening study reveals the preference of the ABC transporter LmrA for zwitterionic detergents but not for with the ionic detergent DDM (*148*). LmrA purified with the zwitterionic detergent FC-16 shows high basal ATPase activity, which could be stimulated by a substrate. The opposite behavior is observed in DDM micelles (*148*). Artificial structures are observed with the energy-coupling factor (ECF) type ABC transporter when using detergent purification (*149*) and the dynamic of the maltose transporter differs in detergent or lipid environment (*150*). It is known that the activities of membrane proteins are also affected by the lipid environment (*151*). A sufficient substitute for the complex environment of the cell membrane cannot be provided by detergent micelles. Membrane mimetic systems, such as proteoliposomes and lipid nanodiscs, stabilized by either protein or polymers, provide an alternative for membrane protein purification and stabilization.

2.1.3.2. By reconstitution into liposomes

After purification in detergent, membrane proteins can be returned into a lipid bilayer by reconstituting them into liposomes (Figure 36). For the first step, lipids, either synthetic lipids

with a precise composition or extracts, such as *E. coli* lipid extract without a clearly defined composition, are used. The choice of the lipids can have an impact on the protein activity. The effect of lipids composition on transporter activity is summarized in the literature (*152, 153*). Thus, lipid screening is a necessary and critical step. For the second step, detergents can be removed by using nonpolar, microporous, styrene beads, also known as Bio-Beads, by dilution, gel filtration, or by dialysis against detergent-free buffer (*154*). Detergent adsorption with Bio-Beads is effective for detergents with a low CMC such as DDM, while the other named techniques work best for detergents with a high CMC (*154*).

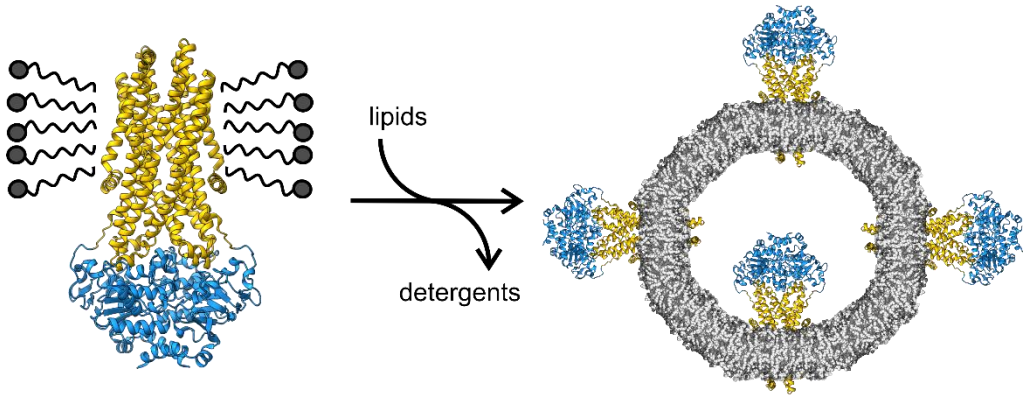


Figure 36: Scheme for membrane protein reconstitution into liposomes.

Reconstitution into liposomes offers a particular advantage due to their membrane curvature, which allows curvature-dependent effects to be studied (*154*). In addition, both structure and function of membrane proteins in a lipid environment can be investigated in proteoliposomes. However, major drawbacks of proteoliposomes are the lack of control over liposome size, protein amount, and protein orientation during reconstitution. To circumvent these difficulties but to still investigate the protein in a lipid environment, membrane proteins can be reconstituted into lipid nanodiscs stabilized by proteins or polymers.

2.1.3.3. By reconstitution into protein stabilized lipid nanodiscs

A nanodisc is built by a discoidal phospholipid bilayer stabilized by a membrane scaffold protein (MSP) or a saposin protein. The currently used MSPs are genetically engineered derivates of the human apolipoprotein-A1 (Apo-A1) (*155*). This human plasma lipoprotein forms complexes with lipids, proteins, and cholesterol derivatives and is involved in lipid metabolism (*156*). The saposin used as an alternative to form lipid nanodiscs is the human saposin A (SapA) protein. This protein is found in lysosomes, where it plays a role in sphingolipid catabolism, transport, and forms discoidal lipoprotein particle upon lipid binding (*157*).

To form lipid nanodics, detergent-solubilized membrane proteins, lipids and MSPs or SapA protein are mixed, followed by removal of detergent molecules by dialysis or Bio-Beads (*158*). By self-assembly, protein-stabilized nanodiscs are formed, resulting in MSP nanodiscs or saposin

lipid particles (named salipro), respectively (Figure 37). Following the reconstitution step with affinity and size-exclusion chromatography allows the separation of membrane proteins encapsuled in nanodiscs, empty nanodiscs without membrane protein, and free-floating scaffold proteins.

The size of the lipid particles formed is determined by the length of the selected engineered MSP variant. Short MSPs stabilize small lipid particles and long MSPs stabilize larger lipid particles. The size of the lipid nanodiscs can vary between 7 and 90 nm (*155*). In addition to screening the MSP variant, the lipid composition must also to be optimized, since this has an influence on the activity of the membrane protein, as already mentioned for the reconstitution into liposomes. Optimizing reconstitution conditions by screening the MSP size and lipids for the protein of interest can be a time-consuming factor that needs to be considered when working with this system.

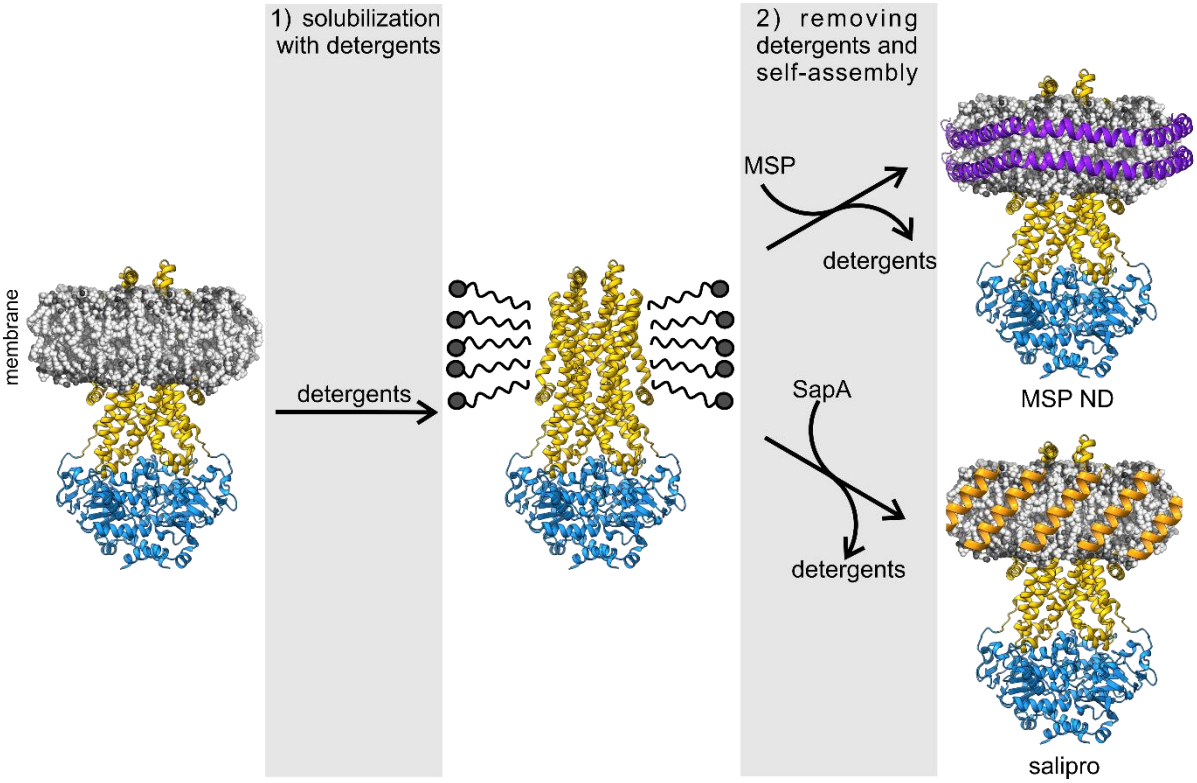


Figure 37: Scheme for membrane protein reconstitution into nanodiscs stabilized by A: membrane scaffold protein (MSP, PDB: 1av1), or B: Saposin A protein (SapA, PDB: 4ddj).

In contrast to MSP stabilized nanodiscs, the salipro particles adapt to the size of the incorporated membrane proteins (*159*). Remarkably, the size variation of the salipro nanoparticles does not negatively affect functional and structural investigations; rather, the salipro nanodiscs are homogeneous and can be studied by a wide range of biochemical and biophysical methods (*159–164*). In addition to the reconstitution protocol, which requires membrane protein to be purified by detergent and then reconstitute into nanodiscs, salipro particles can also be formed from a crude cell membrane in a single reconstitution step (*165*). Here, crude membranes are mixed with SapA and 1% digitonin (a non-ionic detergent). Insoluble material is removed by centrifugation,

and the protein of interest is further purified by affinity column and SEC. In the one-step reconstitution technique, the endogenous lipids present in the native membrane are incorporated into the lipid nanodiscs (*165*). The main advantages of this one-step purification method are the reduction in time to obtain lipid nanodiscs and the use of native lipids for nanodisc formations. Important protein-lipid interactions, which are crucial for structural and functional integrity, are preserved during purification so that the influence of native lipids can be studied. Although the one-step preparation of salipro nanodiscs has advantages, the reconstitution method is not free of detergent. It is likely that even a small amount of detergent could interfere with the important protein-lipid interaction, which in turn affects protein activity. However, this effect has not been well investigated because the one-step purification of membrane protein in salipro particles has not been widely described in literature.

To obtain isolated membrane proteins in the most native environment possible and still be able to perform biochemical and biophysical analyses, a detergent-free one-step nanodisc formation is needed.

2.1.3.4. By forming polymer-stabilized nanodiscs without the need of detergents

To obtain lipid nanoparticles with encapsulated membrane protein in the native lipid bilayer in a single step, amphiphilic copolymers are used. These polymers successfully combine the strengths of detergents and nanodisc stabilization of proteins. Direct solubilization from native membrane sources is very convenient for structural studies, as reflected by the increasing number of cryo-EM structures using this method (*166–168*). Copolymers composed of a hydrophobic styrene (S) or diisobutylene (DIB) and a hydrophilic maleic acid (MA) (Figure 38B) are used to form styrene maleic acid lipid particles (SMALP) or diisobutylene maleic acid lipid particles (DIBMALP), respectively. The self-assembled nanodisc structures (Figure 38) form particle diameters between 10 and 60 nm (*169*). The particle size of SMALPs and DIBMALPs can be tuned varying the polymer/lipid stoichiometric ratios. A high polymer/lipid ratio results in small nanodiscs, while a low polymer/lipid ratio yields large nanodiscs (*170, 171*). Remarkably, membrane protein extraction with polymers is not limited to any type of lipids, making this purification method widely applicable to many different membrane types and integral membrane proteins (*172*). In addition, lipidomic investigation on polymer stabilized nanodiscs to understand the native lipid environment of a membrane protein is possible (*173*). The formation of polymer-stabilized nanodiscs works best at physiological pH of 7-8. Below pH 6 the maleic acid groups of the polymers are protonated resulting in a decrease in polymer solubility in aqueous solution (*174, 175*). A decrease in polymer solubility can also be observed in the presence of divalent cations, which form complexes with the maleic acid groups. SMA polymers tolerate a divalent cation concentration up to 5 mM (*176, 177*), while DIBMA polymers precipitates at a concentration of 30 mM (*178*). Another advantage of using DIBMA over SMA is the absence of absorption of the

polymer at 280 nm. This property allows for the determination of protein concentration in DIBMA by ultraviolet-visible (UV-vis) spectroscopy.

Commercially available SMA polymers have a ratio of S to MA of 1.2-3:1, an average chain length of 2000 – 4000 g/mol, and a chain length dispersity Đ = 2.5 – 2.6 (*179*). DIBMA polymers are commercially available with a 1:1 DIB to MA ratio, an average chain length of 8500 g/mol and a narrower distribution than SMA with Đ = 1.4 (*179*). Polymer dispersity affects the size distribution of formed lipid particles and solubilization efficacy (*179*). The size inhomogeneity seems to be a limitation factor for the successful application in e.g., solution-state NMR studies (*180*). Strategies to overcome the drawbacks of solubility and dispersity are summed up in the following section.

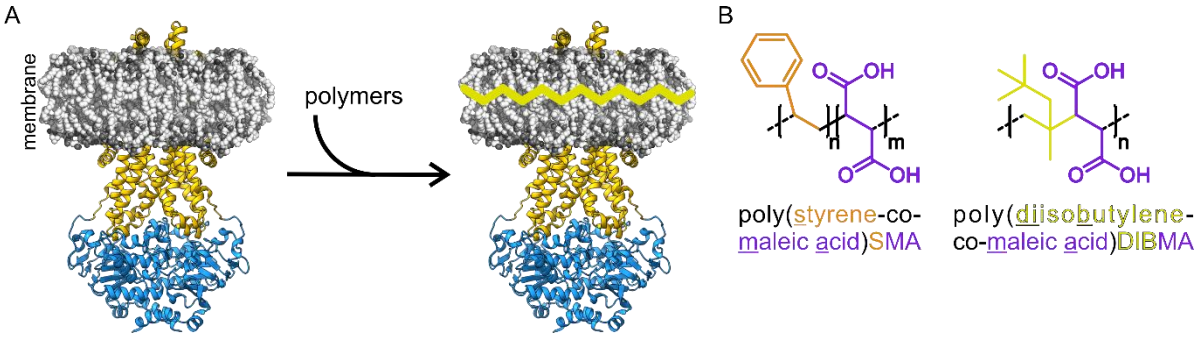


Figure 38: Scheme for protein purification into polymer stabilized nanodiscs (A). B: Molecule structure of the polymers for membrane protein extraction.

2.1.4. Control the polymer to control the nanodisc – strategies to improve polymer composition and quality

The goal in polymer design for biological applications is to create polymers that remain soluble at all physiologically relevant pH values, tolerate a high concentration of divalent cations, have homogeneous molecular weight, and form homogeneous lipid particles, without compromising membrane protein purification efficiency. New SMA and DIBMA derivatives need to be developed to improve polymer functionality and performance spectrum. The simplest way to prepare SMA and DIBMA derivates is to modify the maleic acid groups of these polymers by adding functional groups. This modification reaction can be performed on the anhydride version of the polymer, the precursor polymer of SMA and DIBMA. The anhydride group of the copolymers can be functionalized by (substituted) amines via a nucleophilic ring opening reaction resulting in a half-amide derivative, also called maleamic acid (*181*). On a SMA backbone, amide functionalization was performed by different working groups using different substituted amines (*174, 182*) (Figure 39A). Styrene maleamic acid polymers enhance viability under lower pH values and show more tolerance to increased divalent metal concentration compared to SMA (*155*). A ring-closing procedure upon maleamic acid functionalization yields styrene maleimide (SMI) polymers (Figure 39A). SMI lipid particles (SMILP) are smaller in diameter than SMALPs, thus imposing limits on the size of imbedded membrane proteins. Depends on the substituent of the maleimide,

SMIs show a broad tolerant towards pH and buffer conditions but are less efficient at solubilization from native membrane than SMA polymers (*183, 184*).

A glycosyl amide derivative of DIBMA (Glyco-DIBMA, Figure 39B), which is more hydrophobic and less charged compared to DIBMA, was developed on a DIBMA backbone and exhibits high solubilization efficiency while being stable in a wide range of buffer compositions (*185*). By functionalizing the maleic acid groups of the SMA polymers, a new generation of polymers has been created that successfully overcomes the problems with the pH range and concentration of divalent cations (*186, 187*). However, the dispersity of the polymer cannot be improved by functionalizing already heterogeneous polymers. High polymer dispersity is a general challenge in polymer synthesis, and further workup and polymer purification can only minimize the dispersity, but do not lead to well-defined molecular weights.

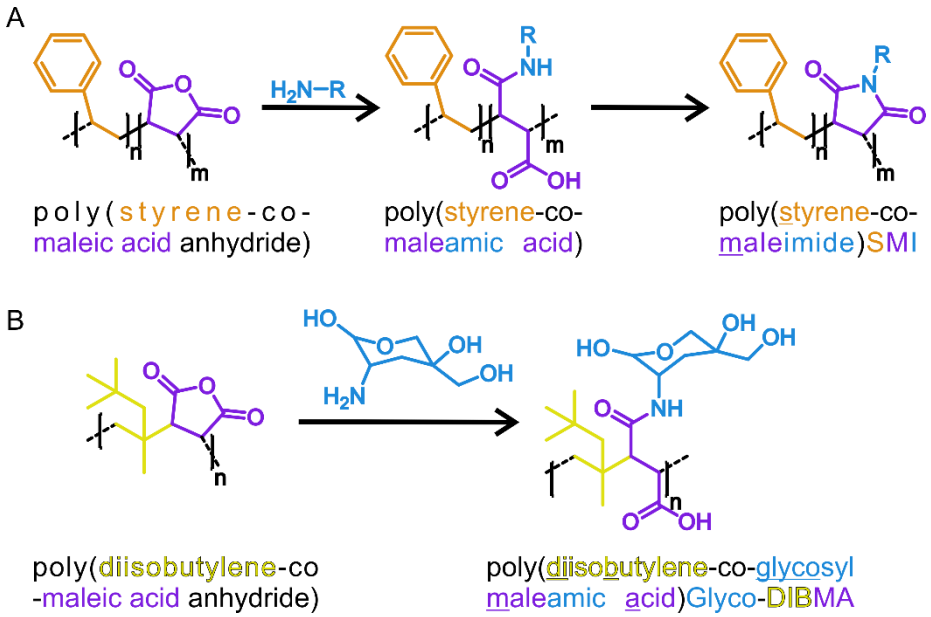


Figure 39: Scheme for the modification of SMA and DIBMA polymers. A: SMA polymer precursor is functionalized with amine, forming maleamic acid. Ring-closing reaction occurs on maleamic acid resulting in styrene maleimide polymers. B: DIBMA precursor reacts with a glycosyl-substituted amine to Glyco-DIBMA.

Commercially available polymers are synthesized by a continuous stirred tank reaction (CSTR) based on conventional free-radical polymerization (*188*)(Figure 40A). Since it is difficult to control the degree of polymerization and the polymer composition during free-radical polymerization, the resulting polymers vary from batch to batch, with variations in exact distribution of the monomers in addition to the high dispersity (Figure 40A and B). To gain control over the length and composition of the polymer, the increasingly popular reversible addition–fragmentation chain transfer (RAFT) polymerization method is used. This polymerization technique can be used to produce tailored polymers with high precision (Figure 40C and D).

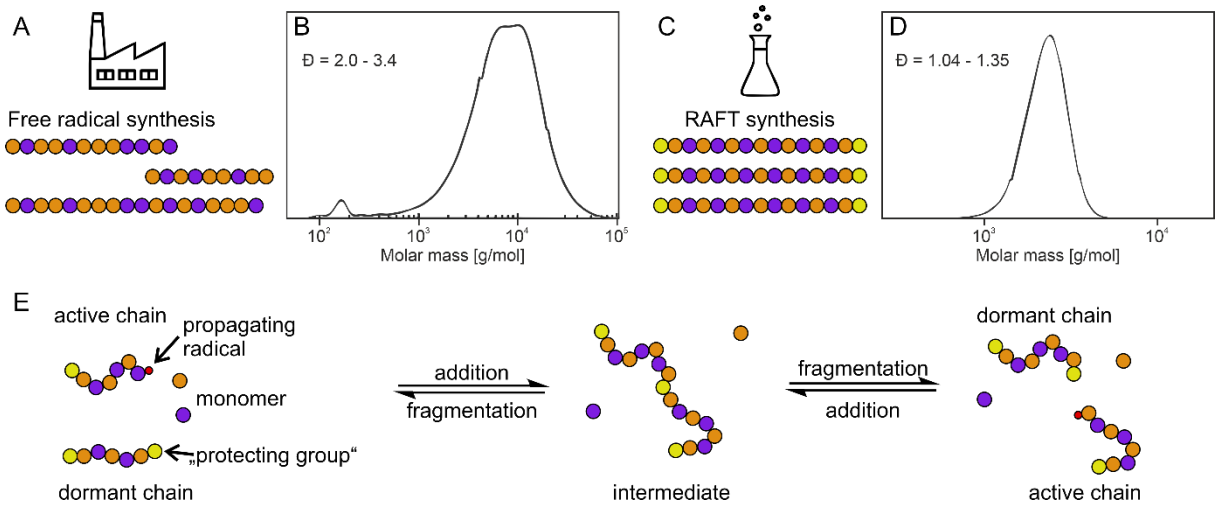


Figure 40: Reversible addition-fragmentation chain transfer polymerization and the dispersity advantage over free radical synthesis. A: Scheme of the product mixture in the free radical synthesis. B: Molar mass distribution of a free radical synthesized polymer and the dispersity range of this polymerization technique. C: Scheme of the product mixture for the RAFT synthesis. D: molar mass distribution of a RAFT synthesized polymer and the dispersity range of this technique. E: Reaction mechanism during RAFT synthesis. Different monomers are colored orange and purple. Protecting group / RAFT-agent is colored yellow. The radical is shown in a red dot.

RAFT is a controlled radical polymerization that relies on the reversible transfer of a "radical protecting" group between a dormant species and a propagating radical, forming a new dormant species and a new propagating radical (*189, 190*)(Figure 40E). Thereby, the dormant chain and the active propagating chain are in equilibrium during the reaction (Figure 40E). With this polymerization technique, SMA polymers with distribution Đ = 1.04-1.35 are reached (*191, 177, 192, 188*). SMALPs prepared from polymers with low dispersity also have more homogeneous particles than SMALPs prepared from commercially available polymers with broad distribution (*188*). Various studies have examined polymer sizes ranging from 0.8 – 28 kDa and styrene to maleic acid ratio ranging from 1:1 to 4:1 (*193, 191*). However, each study covers a specific range of the weight-average molecular mass and/or composition, and a systematic investigation of the influence of polymer chain length and composition influences on particle formation is still missing. A first indication that the molecular weight of the polymer influences SMALP particle size was provided by Smith *et al*. (*188*) with RAFT SMA polymers and Fiori *et al*. (*177*) with functionalized RAFT SMA polymers.

2.1.5. BmrA in membrane mimetic systems

The aforementioned multidrug ABC transporter BmrA from *B. subtilis* is a well-characterized model to study the structural dynamics and molecular mechanism of multidrug ABC exporters. To determine the functional integrity of BmrA, two types of activities can be monitored: substrate transport and ATP hydrolysis. In transport activity, fluorescent substrates are used as reporters. This assay can only be performed in proteoliposomes or in inverted membrane vesicles, as these are the only two system in which the substrate is not accessible on both sides of the protein.

ATPase activity can be determined by the conversion of ATP to ADP and the release of Pi during the transport cycle and can therefore be measured by transporters in all existing purification and reconstitution methods. However, finding the right purification method is the bottleneck in membrane protein research. Each technique has its own advantages and disadvantages that affect the solubilization behavior and/or activity of the proteins. Using the BmrA protein as an example, it has been shown that the choice of detergent leads to differences in activity. As described by Mathieu *et al*., BmrA expressed in *E. coli* cells shows higher ATPase activity when purified with LMNG detergent than in dodecylphosphocholine (FC12) micelles (*194*). In a follow up study, the authors described an increased in ATPase activity by adding cholate detergent to the DDM purification protocol (*195*). After reconstitution in proteoliposomes, the ATPase activity of BmrA increased, with no differences whether the protein had been previously purified with LMNG or FC12 detergents (*194*). These observations clearly highlight the importance of membrane protein environment and, in particular, the lipid environment for the membrane protein activity. BmrA structures of the protein reconstituted into MSP stabilized nanodiscs and purified with SMA and DIBMA polymers have also been reported in literature (*195, 196*). However, ATPase activities have not been determined in these systems, perhaps due to the drawbacks of these systems with regard to their structural sensitivity towards divalent cations. BmrA has not been successfully reconstituted or purified into salipro nanodiscs to date.

To investigate the influence of the different environments on BmrA activity, BmrA is purified with detergents and polymers and reconstituted in different membrane mimicking systems.

2.2. Aims and objectives

Substrate transport from a cell is a fundamental step for microbial cell-to-cell communication. At the same time, substrate export can also provide protection against toxic substances that have unintentionally entered the cell. ATP binding cassette (ABC) transporters, which are found in all kingdoms of life, are an important class of bacterial export pumps. These transporters are integral membrane proteins and consist of two highly conserved cytoplasmic nucleotide-binding domains (NBD) that bind and hydrolyze ATP, and two transmembrane domains (TMD) with substrate binding sites. The energy source for substrate transport is provided by ATP hydrolysis. During substrate transport, ABC transporters undergo dramatic conformational changes. However, it is difficult to study transporters *in vitro* to investigate the transport mechanism in molecular detail because studies of membrane proteins often fail due to unsuccessful protein purification. In addition, the performance of membrane proteins, especially ABC transporters, is strongly influenced by their environment. The choice of purification and/or reconstitution method can drastically affect the activity of the protein. To date, there is no purification or reconstitution method that is equally effective for all membrane proteins, but it may be speculated that all membrane proteins should perform best in their native environment, the lipid bilayer of their host organism. Therefore, a systematic investigation of different purification and membrane mimicking methods on a transporter protein is required to find the best optimized system for this protein family. Since protein purification is the bottleneck of membrane protein research, the generally well-characterized ABC exporter BmrA from *B. subtilis* was chosen as a model protein. The multidrug transporter BmrA can be heterologously expressed in *E. coli* and detergent micelle and liposome reconstitution protocols are available. Transport and ATPase activity can be monitored, making this protein an optimal candidate for studying the influence of the (lipid) environment on protein activity.

The objective of this work was to study the functional and structural implications of lipid environment on the ABC exporter BmrA. Therefore, BmrA is purified in detergent micelles and reconstituted into proteoliposomes, nanodiscs stabilized by different membrane scaffold proteins (MSPs), saposin lipid particles (Salipro), or is purified with amphiphilic poly(styrene-co-maleic acid) (SMA) and poly(diisobutylene-co-maleic acid) (DIBMA) polymers directly from the native membrane. Ideally, the protein remains in its natural environment, hopefully achieved by the formation of styrene maleic acid lipid particles (SMALPs). The ATPase activity of BmrA was determined of all purified and reconstituted samples to investigate the influence of different membrane mimicking system on this activity (Figure 41).

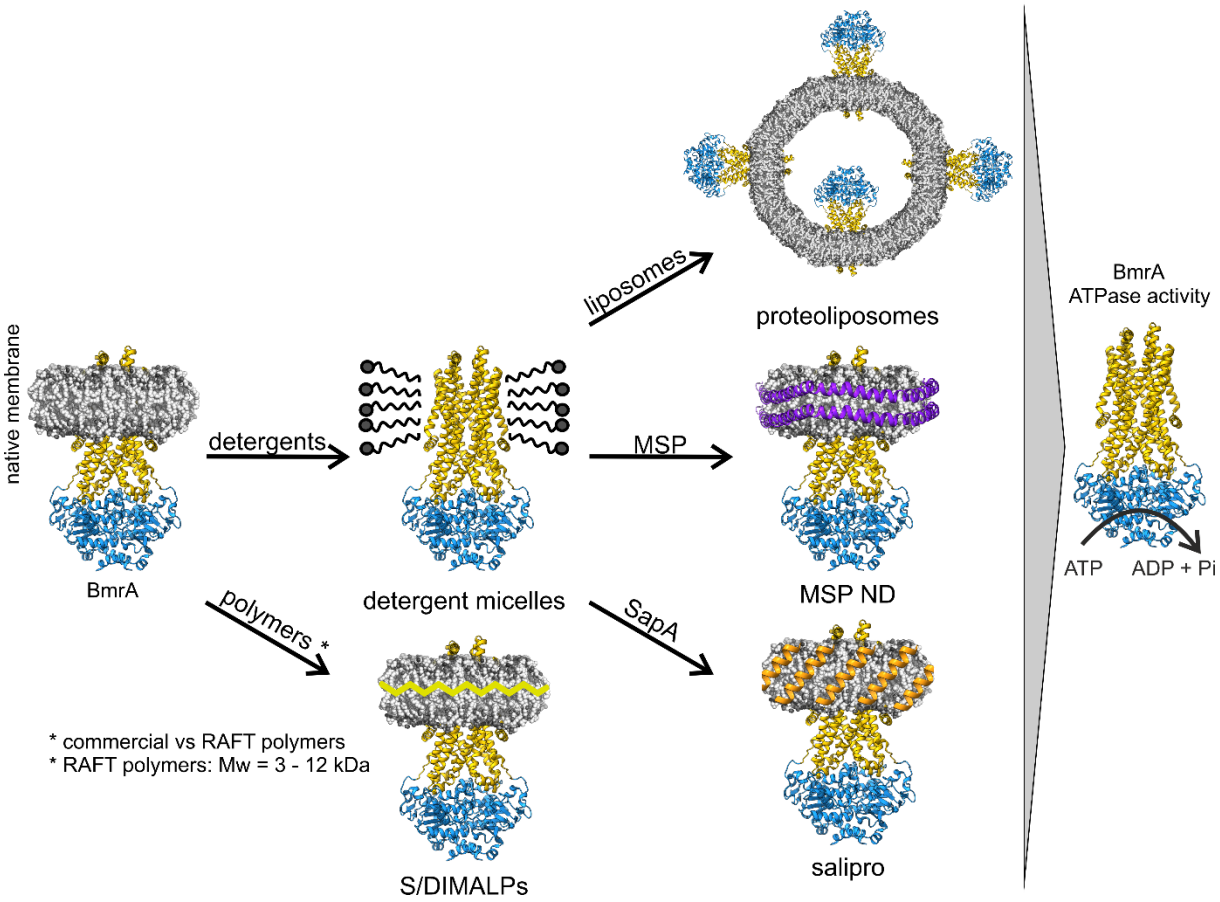


Figure 41: Overview of the aims of this chapter.

Since commercially available SMA polymers exhibit a broad distribution of weight-average molecular mass, resulting in heterogeneous lipid particle formation, reversible addition–fragmentation chain transfer (RAFT) polymerization was used to obtain homogenous polymers. SMA polymers with chain lengths ranging from 3 – 12 kDa were synthesized by the ____ group, _____, and applied to form lipid particles with and without embedded membrane protein. The effect of polymer molecular weight on solubilization efficiency and particle size was investigated.

In summary, finding the best optimized purification and reconstitution system for BmrA provides insights into the influence of the lipid environment on protein activity. The most native system to date is given by the membrane protein purification with SMA and DIBMA polymers. However, these polymers face challenges with regard to homogeneity and divalent cation tolerance. Optimizing polymer synthesis allow this purification method to become a universal tool in the membrane protein purification toolbox. The expansion of purification tools is in great demand in membrane protein research.

2.3. Materials and Methods

Xiran 25010, 30010 and 40005 were a kind gift from Orbiscope and was supplied as a ready to use solution (20% w/v).
DIBMA polymer was a kind gift from Prof. Sandro Keller and was supplied as hydrolyzed polymer in solid material.

2.3.1. Styrene maleic acid polymer synthesis

Styrene maleic acid copolymers with variation in polymer chain-length and styrene to maleic acid ratio were synthesized by _____ and _____, _____ Lab, _____. Free radical and reversible-addition-fragmentation chain-transfer polymerization have been used to couple styrene monomers to maleic acid anhydride. The formed polymer intermediate was hydrolyzed with KOH to yield the final potassium salt of the styrene malic acid polymer (SMA).

2.3.2. Expression of BmrA full-length constructs

All constructs were cloned by Dr. Dania Rose-Sperling into pET-11a vector with an N-terminal His_6-tag and a cleavage site for tobacco etch virus (TEV) protease. As shown by Khadja Mathieu et.al., BmrA full-length protein expressions and their transport activity were strongly influenced by the expression systems (*194*). Referring to the described observation, N-terminal His_6-tagged BmrA full-length protein was expressed in *E. coli* C41 cells. The transformed cells were inoculated in 25 ml 2x YT-medium (2,5% of the main culture) with ampicillin (0,1 mg/ml) and were grown at 37 °C, 200 rpm for 16 h. Overnight cultures were used to inoculate 1 l of 2x YT media supplemented with ampicillin. This culture was incubated at 37 °C and 180 rpm until OD_{600} reached 0.6-0.8. After addition of 700 µM IPTG, cells were incubated for additional 4 hours at 25 °C, 180 rpm. Cells were harvested via centrifugation at 4 °C, 9000 x *g* for 10 minutes. The resulting pellet was frozen in liquid nitrogen and stored at 20 °C until further use.

2.3.3. Membrane preparation and inside-out vesicle formation

To obtain inside-out vesicles (IOVs) containing the BmrA full-length construct, a membrane preparation was performed. First, harvested *E. coli* cells from 1 l main culture were resuspended in 40 ml lysis buffer (50 mM tris, pH 8, 5 mM $MgCl_2$), supplemented with 1 mM DNase I and 1 mM DTT. After 20 minutes of incubation with overhead rotation, the cells were lysed via high pressure homogenization (Maximator HPL6, 2 bar). The Lysis was followed by a low-speed centrifugation

at 15.000 x *g* for 30 minutes, 4 °C, to remove cell debris. The first high-speed centrifugation at 150.000 x *g* for 90 minutes, 4 °C, was performed with the received supernatant in the low-speed centrifugation step. At this step the membrane fraction was harvested. Afterward, the pellet was washed with 50 mM tris, pH 8, 1.5 mM EDTA, followed by a second high-speed centrifugation at 150.000 x *g* for 60 minutes, 4 °C. The pellet was homogenized in 20 mM tris, pH 8, 1 mM EDTA, 300 mM sucrose and frozen in liquid nitrogen. The inside-out membrane vesicles were stored at -20 °C until further use, and the total protein concentration in the IOVs was analyzed by bicinchoninic acid (BCA)-assay.

2.3.4. Transport assay

The transport activity of the BmrA full-length constructs was measured with a spectral fluorimeter. Thereby, the fluorescence decrease of doxorubicin (590 nm emission and 480 nm excitation) after its transport into the inside-out vesicles was detected. At the beginning of the measurement, 4 mM PEP, 60 µg PK, 2 mM $MgCl_2$ and 100 µg/ml membrane sample were filled up to 1 ml final volume with 50 mM hepes-KOH, pH 8, 8.5 mM NaCl. The baseline was measured for 125 seconds, then 10 µM doxorubicin was added. After another 125 seconds, 2 mM ATP were added to the mixture, and the spectra were recorded for another 600 seconds. The sample was stirred during the entire measurement at 25 °C.

2.3.5. Expression and purification of saposin A (SapA) protein

pNIC-Bsa4 plasmid coding for Saposin A with a N-terminal His_6-tag (MHHHHHHSSGVDLGTENLYFQSMGSLPCDICKDVVTAAGDMLKDNATEEEILVYLEKTCDWLPKPN MSASCKEIDSYLPVILDIIKGEMSRPGEVCSALNLCES) was transformed to *E. coli* Rosetta gami-2 cells. 50 ml of TB medium containing 20 µg/ml of chloramphenicol, 15 µg/ml of tetracycline-hydrochloride and 50 µg/ml of kanamycin was inoculated with the transformed cells and was incubated at 37°C, 180 rpm for 16 h. 2 l TB medium with the same antibiotics composition as described before was inoculated and induced with 0.7 mM IPTG once the OD_{600} = 1. After three hours of additional incubation at 37°C the OD600 was measured to be 1,77 and the cells were harvested by centrifugation for 10 min at 10000 x *g*. The obtained pellet was frozen in liquid nitrogen and stored at -80°C until further use.

For the purification of the protein, the frozen cell pellet from 1 l culture was suspended in 80 ml buffer containing 20 mM hepes, pH 7,5, 150 mM NaCl, 20 mM imidazole, supplemented with RNase and lysozyme and then subjected to 10 sonification cycles of 60% intensity at the Sonotrode VS 70T, each cycle lasting 45 s consisting of 1 s pulse followed by 1 s pause. Cell debris were removed via centrifugation at 9400 x *g* for 30 min. The supernatant was heated for 10 min

in a water bath at ~80°C until a white precipitate was formed, which was then removed by another centrifugation step at 3000 x *g* for 25 min. The supernatant was mixed with equilibrated NiNTA resin (10 ml for 1 l culture) and incubated with overhead-rotation for 1 h at 4°C. The mixture was transferred to a gravity flow column. The NiNTA beads were washed firstly with 15 CV 20 mM hepes, pH 7,5, 150 mM NaCl, 20 mM imidazole and secondly with 15 CV 20 mM hepes pH 7.5, 150 mM NaCl, 40 mM imidazole. The protein was eluted with 50 ml 20 mM hepes pH 7.5, 150 mM NaCl, 400 mM imidazole. Fractions containing SapA protein were mixed with TEV (1:40 protease:protein) and dialyzed against 20 mM hepes pH 7.5, 150 mM NaCl for 16 h. TEV protease was removed via centrifugation at 3000 rpm for 10 min and the supernatant was run twice through a second Ni-IMAC column with a bed volume of 5 ml. After that the column was then washed 3 times with 10 ml 20 mM hepes pH 7.5, 150 mM NaCl. The flow-through and fractions containing SapA were concentrated, and a size exclusion chromatography was performed at 4°C using HiLoad Superdex 200 16/60 (20 mM hepes, pH 7,5, 150 mM NaCl).

2.3.6. Expression and purification of membrane scaffold proteins (MSP)

Plasmids containing N-terminal His_6-tagged MSP variants were transformed in *E. coli* Bl21 (DE3) gold. 50 ml LB_{Kan} was inoculated with the transformed cells and incubated for 16 h, 130 rpm at 37°C. 500 ml TB_{Kan} was inoculated with the preculture and induced with 0.7 mM IPTG once OD_{600} = 0.6. After 4 h of incubation at 37°C, 130 rpm, cells were harvested via centrifugation (15 min, 5000 x *g*, 4°C), frozen in liquid nitrogen, and stored at -20°C until use.

Harvested cells were resuspended in 35 ml 40 mM tris pH 7.4, 100 mM NaCl, 1% Triton X-100, 0.5 mM EDTA, supplemented with DNase I, protease inhibitor mix, and lysed via sonification. To remove cell debris the mixture was centrifugated for 30 min, 30000 x *g*, 4°C, and the supernatant was incubated with 5 ml NiNTA resin for 45 min at 4°C. After transferring to a gravity flow column, the resin was washed with 10 CV 40 mM tris pH 8, 300 mM NaCl, 1% Triton X-100, 5 CV with 40 mM tris pH 8, 300 mM NaCl, 50 mM Na-cholate, 20mM imidazole, 5 CV with 40 mM tris pH 8, 300 mM NaCl, 50 mM imidazole, and 2 CV 40 mM tris pH 8, 300 mM NaCl. MSP proteins were eluted with 5 CV (divided in 2 x 2 CV, 1 CV) 40 mM tris pH 8, 300 mM NaCl, 500 mM imidazole. Fractions containing MSP proteins were dialyzed against 20 mM tris pH 8, 100 mM NaCl, 0.5 mM EDTA, 0.5 mM DTT, supplemented with TEV protease (1:40 protease:protein) for 16 h. After dialysis and a centrifugation step 10 min, 3000 rpm, 4°C, the supernatant was supplemented with 25 mM imidazole and transferred to a NiNTA affinity column. The resin was washed with 2 x 2 CV 40 mM tris pH 8, 300 mM NaCl and non-cleaved MSP proteins were eluted with 5 CV 40 mM tris pH 8, 300 mM NaCl, 500 mM imidazole. Fractions containing MSP protein were dialyzed against buffer with 20 mM tris pH8, 100 mM NaCl, 0.5 mM EDTA for 16 h, at 4°C. After dialysis, the sample was concentrated, aliquoted and frozen in liquid nitrogen. Purified MSP proteins were stored at -20°C until further use.

2.3.7. Purification of BmrA full-length protein in detergent micelles

For the detergent purification of BmrA full-length protein, the membrane suspension was diluted to 2 mg/ml total protein concentration using 50 mM tris pH 8, 10% glycerol, 100 mM NaCl, 10 mM imidazole, 1% DDM, 1 mM DTT and incubated for 1 h at 4°C. The insoluble part was removed by ultracentrifugation at 150000 x *g* for 1 h at 4°C. The supernatant was incubated with pre-equilibrated NiNTA resins (2ml for 1 l culture) for 1 h at 4°C under rotation. The mixture was transferred into a gravity column and was subsequently washed with 5 CV 50 mM tris pH 8, 10% glycerol, 500 mM NaCl and 5 times with 5 CV 50 mM tris pH 8, 10% glycerol, 100 mM NaCl, 25 mM imidazole, 0,0675% DDM, 0,05% Na-cholate. The protein was eluted with six times 1CV 50 mM tris pH 8, 10% glycerol, 100 mM NaCl, 250 mM imidazole, 0,0675% DDM, 0,05% Na-cholate. Fraction containing BmrA full-length protein were dialyzed against 50 mM tris pH 8, 10% glycerol, 100 mM NaCl, 0,0675% DDM, 0,05% Na-cholate for 16 h at 4°C and stored at 4°C until further use.

2.3.8. Reconstitution of BmrA in Liposomes

For the reconstitution of BmrA into liposomes and by this forming proteoliposomes 1 mg of *E. coli* lipid extract was suspended in 40 µl 40 mM hepes pH 8 and mixed with DDM (0,4% w/v final concentration) for 1 h at room temperature. To the clear lipid solution purified BmrA full-length protein in detergent (200 µg/ml) was then added and filled up to 250 µl with 40 mM hepes pH 8. The protein lipid mix was gently stirred for 1 h at room temperature. Afterwards 20 mg of dried Bio-Beads were added and mixed for 1 hour at room temperature. The procedure of adding Bio-Beads was repeated 2 more times before they were separated from the solution by centrifugation at 6000 x *g* for 1 min.

2.3.9. Reconstitution of BmrA in membrane scaffold protein stabilized nanodiscs

18.6 µl of the prepared *E. coli* lipid extract solution (25 mg/ml in 50 mM hepes pH 8, 50 mM NaCl) is mixed with 30 µl of a 10% DDM stock solution and stirred at room temperature for 1 h before adding purified BmrA full-length protein in detergent (400 µg/ml) and purified MSP (384 µg/ml). The mixture is then filled up to 500 µl with 50 mM hepes pH 8, 50 mM NaCl and mixed for 1 h at room temperature. 340 mg of the Bio-Beads were added during 3 h of stirring at room temperature before getting separated from the solution by centrifugation at 6000 x *g* for 1 min.

2.3.10. BmrA full-length protein reconstituted in Salipro nanodiscs

40 µl of *E. coli* extract in 50 mM hepes pH 7.5, 10% glycerol, 50mM NaCl, 0.28% DDM with a lipid concentration of 5 mg/ml was incubated for 5 min at 37°C at 450 rpm. To this solution 240 µl of detergent purified BmrA full-length protein (0,765 mg/ml protein concentration) is added and incubated for another 5 min at 37°C and 450 rpm. Then 114 µl of purified SapA protein (0,969 mg/ml) was added for a last incubation step for 5 min at room temperature, 450 rpm before the mixture was subjected to the size exclusion chromatography using a Superdex200 Increase 10/300 GL column. Protein concentration was increased by centrifugation to 0.336 mg/ml. BmrA Salipro particles were stored at 4°C for further analysis.

2.3.11. BmrA full-length protein solubilization with SMA/DIBMA polymers

Inverted *E. coli* membrane vesicles were mixed with SMA or DIBMA polymers and 20 mM tris pH 8, 150 mM NaCl buffer to yield final membrane concentration of 30 mg/ml and polymer concentration of 2.5% (w/v). The mixture was incubated for 1 h, 850 rpm at 25 °C. Insoluble membranes were removed by ultracentrifugation (125000 x *g*, 45 min, 4°C). The supernatant containing solubilized protein was analyzed by silver staining and western blotting, using an anti-his_6 antibody.

2.3.12. ATPase-assay

To determine the ATPase activity of the purified and reconstituted BmrA full-length protein in different lipid mimicking environments, an enzymatic coupled ATPase assay was performed. The detection of activity is based on an enzymatic chain reaction initiated by the hydrolysis of added ATP by the protein of interest. BmrA catalyzes the reaction from ATP to ADP, which leads to the reaction from phosphoenolpyruvate (PEP) to pyruvate catalyzed by pyruvate kinase (PK). In this step ADP is formed to ATP. Pyruvate, in turn, reacts in a lactate dehydrogenase (LDH) catalyzed reaction to lactate. During this reaction NADH + H^+ oxidates to NAD+, which leads to the decrease of the absorption at 340 nm wavelength. A reaction scheme is shown in Figure 42.

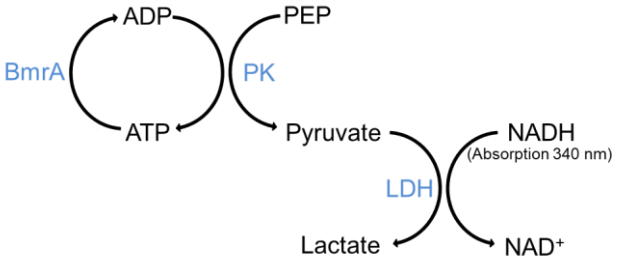


Figure 42: Reaction scheme showing the enzymatic coupled reaction for the ATPase-assay. BmrA full-length protein catalyzes the ATP hydrolyzation to ADP. PEP is converted to pyruvate by PK, while ADP reacts back to ATP. Pyruvate reacts to lactate catalyzed by LDH. Thereby, NADH+H⁺ oxidases to NAD+, which leads to a decrease of the absorption at 340 nm.

In a quartz cuvette, 1 mM $MgCl_2$, 1 mM ascorbic acid, 3 mM PEP, 3.5 µl PK/LDH-mix, 0.21 mM NADH were mixed with protein sample with a final BmrA full-length concentration of 5 µg/ml (15 ug/ml for detergent solubilized BmrA full-length protein) and filled up to 630 µl final volume with 50 mM hepes-KOH pH 8. 70 µl (10 mM) ATP was added to start the reaction and the measurement was performed at 37°C under stirring.

2.3.13. Western Blot

After SDS-PAGE (see chapter 1.3.8) proteins were transferred onto a PVDF membrane with another electrophoresis step to detect His-tagged proteins with an antibody mediated chemiluminescence reaction. The blotting step was performed with the tank blot method, using tank blot buffer (3 g/l tris base, 14.4 g glycine, 20% (v/v) ethanol), 110 V for 1 h. After the blotting step, the membrane was incubated with a 5% milk-block solution in PBS/T followed by the incubation with the α-His antibody solution (1:2000 dilution in 1% milk powder in PBS/T) for 16 h. After three wash steps, each for 5 min at room temperature with PBS/T, chemiluminescence was induced with a 1:1-mixture of Roti®-Lumin reagents 1 and 2 and detected with Cytiva imager AI680.

1 l 10 x PBS buffer contains 80 g NaCl (137 mM), 2 g KCl (2.7 mM), 14.4 g $Na_2HPO_4 * 2\ H_2O$ (10 mM), and 2.4 g $KH_2PO_4 * H_2O$ (1.8 mM). The buffer is adjusted to pH 7.4. To prepare the PBS/T buffer, 50 ml of the 10 x PBS buffer, supplemented with 0.5 g Tween20, was filled up to 450 ml.

2.3.14. Silver staining

An alternative protocol to Coomassie staining after SDS-PAGE is the silver staining method. Solution A was prepared by mixing 1 ml 3.6% NaOH with 700 µl 29.7% NH_3 and filled up to 48 ml. 400 mg $AgNO_3$ were dissolved in 2 ml H_2O and added dropwise to the prepared $NaOH/NH_3$ mixture. The SDS-PAGE gel after electrophoresis, was incubated with solution A for 15 min. In this incubation time the gel was kept in the dark. The silver solution was discarded, and the gel was

washed three times with H_2O. Solution B, containing 500 µl formaldehyde, 50 mg citric acid in 1 l H_2O, was added and the staining reaction was stop by adding 1% acetic acid solution.

2.3.15. Lipid only SMALPs formation

DMPC (1,2-dimyristoyl-sn-glycerol-3-phosphocholine (di-14:0 PC) lipid was purchased from Avanti Polar Lipids. Lipids in chloroform were dried under nitrogen flow and under vacuum. The dried lipid film was resuspended in 20 mM tris pH 8, 150 mM NaCl buffer with a lipid concentration of 2 % (w/v). A polymer solution in 20 mM tris pH 8, 150 mM NaCl was added to the lipid suspension to yield a final polymer concentration of 2.5% (w/v) and a final lipid concentration of 1% (w/v). After vigorous mixing with the pipette, the mixture was incubated at room temperature for another 30 min. 100 µl of the clear solution was analyzed by SEC, using Superdex 200 10/300 column.

2.3.16. Dynamic light scattering (DLS)

DLS data were recorded using Malvern Zetasizer Nano ZS instrument with a scattered angle of 173°. Measurements were taken with 30s equilibration time and 11 accumulations in a 100 µl quartz cuvette at 25°C. Each measurement was repeated three times and the average is analyzed by number to estimate the average hydrodynamic diameter.

2.3.17. Transmission electron microscopy

Electron microscope images were recorded by _____ and _____ (_____ Lab, _____ and _____).

2.4. Results and Discussion

2.4.1. The effect of (lipid) environment on the activity of the ABC transporter BmrA

ATP binding cassette (ABC) exporters play a crucial role in the transport of across cellular lipid bilayers thereby enabling e.g., the secretion of biosynthetic natural products or the extrusion of toxic compounds. So-called multidrug transporters can transport an exceptionally diverse selection of substrates (*131*). Such relatively non-specific substrate transport may confer an advantage for cellular detoxification, however, at the same time, this mechanism can lead to resistance to antibiotics or cancer treatment in bacteria or human cells, respectively (*197*). ABC transporters are the only primary active transporters involved in MDR. Despite their importance, their investigation is not always straight forward since they may be difficult to purify. Additionally, it has been shown that the membrane environment is crucial for the function of these proteins (*150, 151*), and conventional purification approaches may strip the protein of important lipid-protein interactions. Thus, for each transporter, a systematic screen of different purification methods and membrane mimicking agents is required to find the optimal working conditions. Of particular interest would be the ability to compare the same protein in diverse membrane-mimetic environments, including the native membrane.

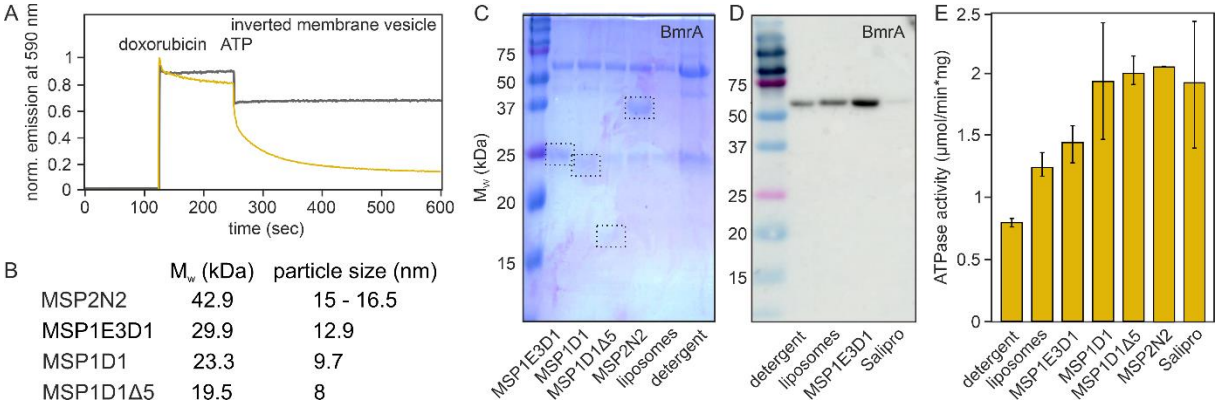
To investigate the effect of the lipid environment on the function of on ABC exporter, a well-studied model multidrug transporter, BmrA from *B. subtilis*, was chosen as a model system. N-terminally His_6-tagged BmrA was heterologously expressed in *E. coli* C41 (DE3) cells. To investigate the transport function of BmrA protein mimicking a cellular environment, membrane vesicles were prepared for transport assays with the fluorescent substrate doxorubicin (Figure 43A). BmrA was then purified from these vesicles using a detergent mixture of dodecyl-β-D-maltosid (DDM) and Na-cholate. Furthermore, the protein was reconstituted in lipid environments, i.e., in liposomes and lipid nanodiscs stabilized by the membrane scaffold protein (MSP) or saposin A (SapA). Different MSP sizes ranging from 19.5 – 42.9 kDa were used that formed particle sizes between 8 – 16.5 nm (Figure 43B).

The successful reconstitution of BmrA protein into membrane mimicking systems is indicated by a clear solution after detergent removing (data not shown) and the presence of both BmrA and MSP protein in SDS-PAGE (Figure 43C). All MSP variants used run slightly lower in the SDS-PAGE (Figure 43C) than expected for their protein sizes, which may be due to the fact that they are not fully denatured by SDS and therefore run faster than the reference protein with the same size. This observation has not been described in literature and might be caused by the way the SDS-PAGE samples were handled for this experiment. BmrA reconstituted in Salipro particles could not be stained with coomassie, due to the low yield and concentration (336 µg/ml) of the sample. Therefore, the presence of BmrA in Salipro particles was detected by Western Blot (Figure 43D).

The basal ATPase activity of all purified and reconstituted BmrA samples was measured. Detergent solubilized BmrA showed a lower ATPase activity compared to BmrA in a lipid environment (Figure 43E). This observation is consistent with what was described by Mathieu *et al*. (*194*). Once BmrA was reconstituted in a *E. coli* lipid environment, the basal ATPase activity increased by a factor of around two to four depended on the lipid environment (Figure 43E). This observation is consistent what has been described in the literature and suggests that the protein's ATPase activity is strongly dependent on its interaction with lipids (*194*).

When comparing the basal ATPase activity of all membrane mimicking systems (proteoliposomes and MSP or Saposin-based nanodiscs), liposome-reconstituted BmrA displayed the lowest basal ATPase activity (Figure 43E). This can be explained by the random orientation of the protein in proteoliposomes, thereby only taking those molecules into account, whose NBDs point outward and are accessible to ATP. BmrA proteins with NBD facing the liposome core do not respond to the addition of ATP outside of the liposomes but are included in the protein concentration. For a precise conclusion, the protein orientation should be determined, which has not been done in this work. At this stage, a clear comparison between BmrA in liposomes and nanodiscs is not possible due to the lack of information on protein orientation in liposomes.

In MSP stabilized nanodiscs with the BmrA : MSP : *E. coli* lipid mass ratio 1:0.96:2.3 , the lowest basal BmrA ATPase activity was observed with MSP1E3D1. Surprisingly, BmrA reconstituted in larger (MSP1D1, MSP1D1D5) as well as smaller MSP nanodiscs (MSP2N2) showed a higher basal ATPase activity (Figure 43E). BmrA reconstituted in Salipro-stabilized lipid particles showed similar basal ATPase activity compared to that seen for MSP nanodiscs, with the exception of MSP1E3D1 (Figure 43E).



	M_w (kDa)	particle size (nm)
MSP2N2	42.9	15 - 16.5
MSP1E3D1	29.9	12.9
MSP1D1	23.3	9.7
MSP1D1Δ5	19.5	8

Figure 43: Activity of BmrA in dependence of its environment. A: Normalized transport activity of BmrA protein in inverted E. coli C41 (DE3) membrane vesicles. Transport activity is monitored by the ATP-dependent extrusion of the fluorescent substrate doxorubicin (590 nm emission and 480 nm excitation). A negative control is given by the BmrA catalytic inactive Walker A mutant K380A (133) (grey). B: Molecular weight and particle size of the MSP variant used here. C: 12% SDS-PAGE of BmrA (66.4 kDa, 10 µg per lane) solubilized by detergent (DDM Na-Cholate mixture) and reconstituted in liposomes (BmrA:E. coli lipid mass ratio 1:20) and nanodiscs (BmrA:MSP:E. coli lipid mass ratio 1:0.96:2.3). MSP bands are highlighted in dashed boxes. D: Anti His_6-tag Western Blot of BmrA proteins (10 µg per lane) solubilized in detergent (DDM Na-Cholate mixture) and reconstituted in liposomes and nanodiscs. E: Basal ATPase activity of full-length BmrA proteins solubilized in detergent (DDM Na-Cholate mixture) or reconstituted in liposomes using or nanodiscs, both using E. coli lipid extracts (n=3, technical replicates, errors represent standard deviation of the measurements).

Comparing the results obtained from BmrA with homologous ABC transporters, different levels of ATPase activity in detergent micelles were reported from different authors for the homologous MsbA from *E. coli (198, 199)* and murine, rat and human P-glycoprotein (Pgp) (*200, 201*). This indicates that not all ABC transporters are equally sensitive to the effect of detergent micelles and a comparison with absolute numbers is difficult. Once the transporters were reconstituted into a membrane mimicking system, an increase in basal ATPase activity in dependence of the lipid environment was observed (*198, 199*). Binding of lipids to the membrane protein might hold and stabilize the TMDs of these ABC transporters, creating a more native environment than detergent micelles. Another possible explanation for the increased ATPase activity in lipid environment could be that lipids might bind to specific residues in the TMDs and thereby regulate protein activity of the transporter.

When being reconstituted into nanodiscs, the homologous MsbA protein seems to prefer MSP1E3D1 over MSP1D1 (*198*) but no significant differences in the choice between MSP or saposin stabilized nanodiscs were noticed (*164*). Since there are no clear preferences which MSP protein to choose when ABC exporters are reconstituted into nanodiscs, this highlights the need to screen MPS sizes for the protein of interest.

For all membrane reconstitution approaches discussed so far, the ABC transporter nonetheless needed to be purified in detergent first. However, complete removal of detergent is not guaranteed during proteoliposome or nanodisc formation. Moreover, synthetic lipids or lipid extracts are used, rather than the native membrane lipid composition which provides a significantly more complex environment. It has been shown that the protein environment has a huge impact on protein dynamics and conformation (*202*), e.g., in the case of MsbA (*203*). Ultimately, the goal of functional and structural studies is to investigate membrane proteins in an environment that resembles the native conditions as closely as possible, including the native lipid-protein interactions. To date, the only method to purify membrane proteins directly from cellular membranes are amphiphilic polymers (styrene maleic acid, SMA and diisobutanyl maleic acid, DIBMA polymers). In contrast to a purification in detergent and subsequent membrane reconstitution, these polymers allow a one-step purification from native membranes. However, a major throwback of this purification method is the instability of the formed nanodiscs in the presence of divalent cations (*176, 177*) and the structural inhomogeneity of commercially available polymers (*179*) due to the commercially used continually stirring tank reactor (CSTR) synthesis method (*204*). The dispersity of commercially available polymers affected the homogeneity of the formed lipid particles, as shown by the size exclusion chromatograms of DMPC lipid particles formed by these polymers (Figure 44). Nevertheless, these polymers have been described to successfully isolate different types of membrane proteins (*205, 206*).

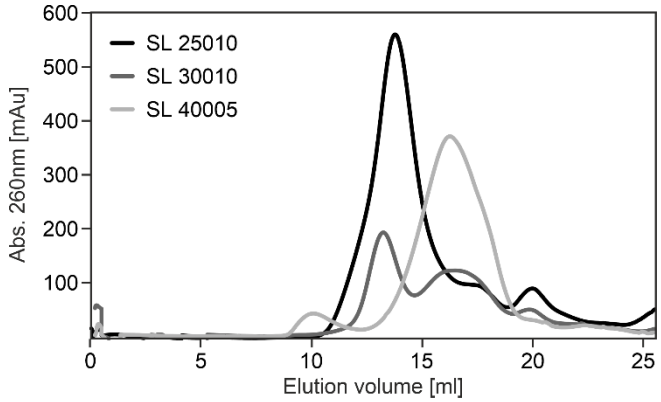


Figure 44: Size exclusion chromatography traces of DMPC lipid particles formed by commercially available polymers (Xiran SL25010, SL30010, and SL40005). Samples were run on the BioRad ENrich650 column.

Here, BmrA was successfully purified using either SMA (Xiran SL25010) or DIBMA polymers (Figure 45A). However, due to the instability of the isolated lipid particle in the presence of divalent cations, the basal ATPase activity of BmrA could not be measured. Therefore, the effect of a native environment to BmrA ATPase activity has not been studied and could not be compared with other mentioned membrane mimicking systems.

Before the effect of the native lipid environment can be analyzed, the polymer dispersity should be improved to obtain homogeneous particles during purification. Particle size homogeneity is a requirement for biophysical techniques such as solution-state nuclear magnetic resonance spectroscopy and small angle x-ray scattering (*180, 207*). To address the problem of particle size heterogeneity, the polymer synthesis of SMA polymers was optimized.

2.4.2. Production of defined SMA polymers by RAFT synthesis

Commercially available SMA and DIBMA polymers are synthesized by CSTR reaction using the process of radical polymerization (*188*). These polymers can be purchased with styrene:maleic acid ratio ranging from 1:1 to 4:1, and weight-average molecular mass of 5 – 10 kDa. Because of the CSTR polymerization method, it is difficult to precisely control the polymer distribution, and as a result, a range of weight-average molecular mass and high dispersity are produced (*179*). Defined polymers were produced as an alternative synthesis approach, and to enable a systematic study of polymer composition affecting lipid particle formation. To investigate the effect of polymer dispersity on the homogeneity of the formation of lipoparticles, SMA polymers were synthesized by reversible addition–fragmentation chain transfer (RAFT) polymerization. As a control, a free radical polymerization as is used in the production of commercially available SMA polymers was also carried out. The synthesis was carried out in collaboration with the ____ laboratory (____).

Three SMA polymers with an S:Ma ratio of 2:1 and weight-average molecular mass between 6 and 15 kDa were synthesized by radical polymerization. This specific S:Ma ratio was chosen due to the

described thermal stability of particles formed by this polymer and higher membrane protein solubilization efficiency compared to other S:Ma ratios in the literature (*208, 209*). The polymers produced by free radical synthesis showed a dispersity between Đ = 2.4 – 2.8 and were designated FR S_2Ma (free radically synthesized styrene maleic acid polymer with S:Ma ratio of 2:1). In addition to the radical synthesis, an SMA polymer with an S:Ma ratio of 1:1, a molecular mass of 6 kDa, and a dispersity Đ = 1.07 was prepared by RAFT synthesis. A higher S:Ma ratio could not be achieved, probably due to the incomplete reaction with the RAFT agent 2-cyanopropan-2yl benzodithioate. The reaction scheme for radical and RAFT polymerization is shown in Figure 45B.

To investigate the membrane protein solubilization efficiency of commercial SMA (SL25010), DIBMA and the newly synthesized polymers, BmrA was overexpressed in *E. coli* and inverted membrane vesicles were generated. Upon treatment with the different polymers, the respective solubilization efficiency was analyzed by SDS-PAGE. For comparison, the detergent DDM was used (Figure 45A). Commercially available SL25010 showed good solubilization behavior for various membrane protein sizes. However, the yield of solubilized BmrA was significantly lower compared to the sample solubilized with DDM. The commercial DIBMA polymer is forming a smear across all protein sizes which precludes a more detailed analysis (*210*). To avoid smearing, the SDS-PAGE sample should be boiled or treated with divalent cation to precipitate the polymer. This has not been done to have a better comparison between all used polymers that are treated in the same way. The SDS-PAGE lane where the sample solubilized by the radically synthesized polymers was loaded shows a similar smearing. This may indicate that these polymers affect the charge of the SDS – protein micelles, resulting in poor separation of proteins by size. It is hypothesized that this effect might be caused by heterogeneity of the polymer sample, as clear bands were observed when the RAFT synthesized polymer was used. The yield of solubilized BmrA was slightly higher compared to a sample solubilized with SL25010. In addition, low molecular weight proteins (15 – 20 kDa) were efficiently selected by the RAFT polymer, whereas these proteins were less enriched when membranes are lysed with SL25010.

To reduce the complexity of the SMALP system and investigating the polymer - lipid interaction, nanodiscs without proteins were formed. Using only DMPC lipids, the polymer nanodiscs were analyzed by DLS to determine the respective particle sizes. A correlation was found between the length of the polymer chain and the size of the formed particle, with longer polymer forming larger and shorter polymers forming smaller particles (Figure 45C). Comparing polymers produced by the two different synthesis approaches (FR and RAFT) with the same chain length (6 kDa), the DMPC particles formed with both polymers had a size of ~15-20 nm regardless of the synthesis method. However, the SEC chromatograms of the DMPC lipid particles showed a narrower particle size distribution with the RAFT synthesized SMa-6 polymers, and a significantly broader particle size distribution for the free radical synthesized polymers (Figure 45D). The heterogeneity of the FR S_2Ma DMPC particles was also visible in the corresponding negative stained electron microscope image (Figure 45F). When RAFT SMa polymers were used to form lipid particles, more homogeneous particles were formed compared to the counterpart synthesized by free radicals. However, it was also noted that the RAFT SMa DMPC lipid particles tended to aggregate with each

other, resulting in stacked particles (Figure 45G). Although it was found that the polymer synthesis approach and the associated dispersity of the polymer did not have a major impact on the particle size of DMPC lipid particles, the overall homogeneity of the sample can be controlled by using defined polymers. For some experiments, such as solution NMR and SAXS, a homogeneous sample is highly desirable RAFT synthesized polymers may thus constitute a way to modulate the overall sample homogeneity.

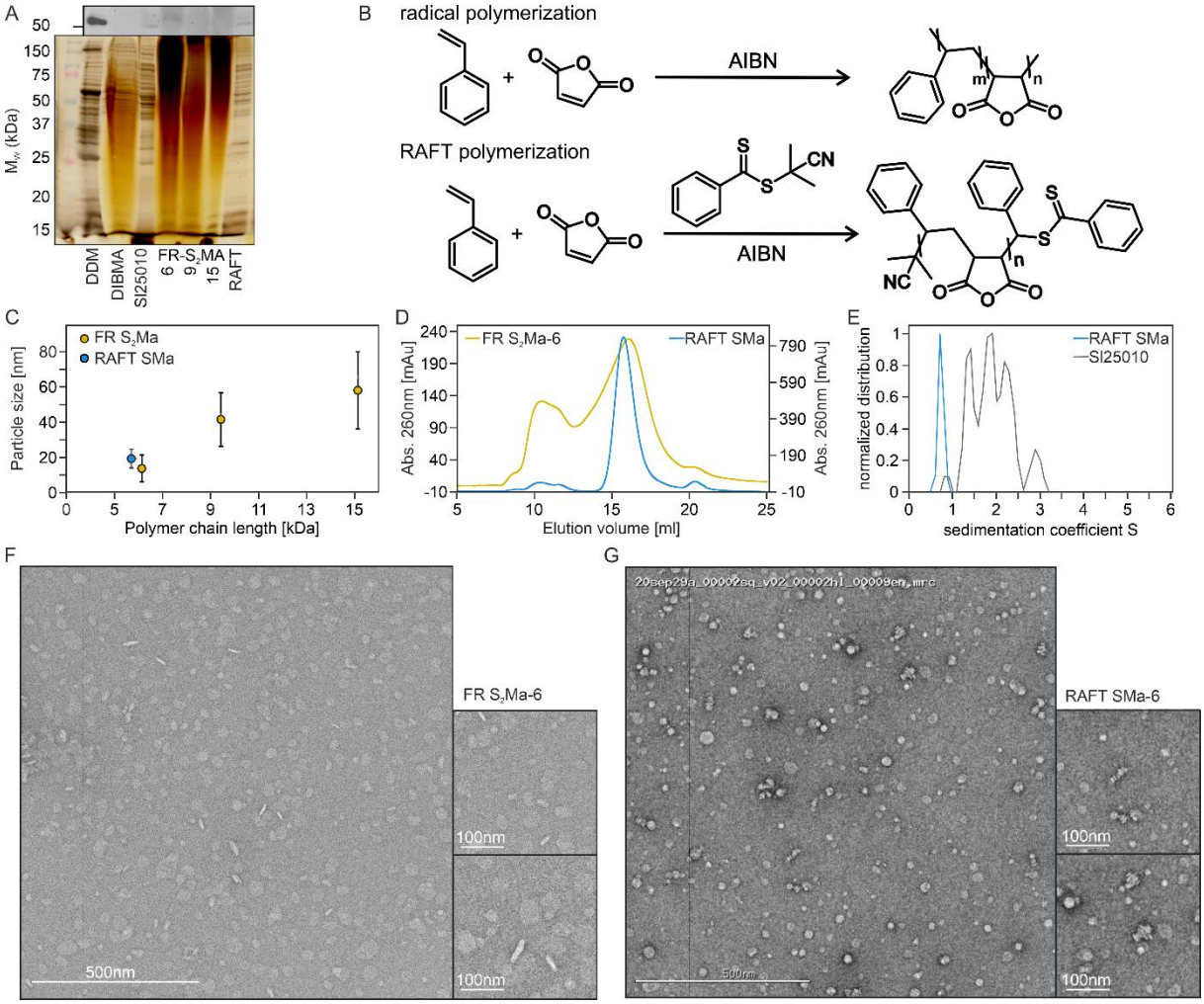


Figure 45: Comparison of the lipid and membrane protein solubilization behavior of free radical/RAFT synthesized SMA polymers with commercially available polymers. A: Western Blot (top) – anti-His$_6$ antibody showing BmrA and silver stained SDS-PAGE of E. coli membranes treated with the detergent DDM, commercially available polymers (DIBMA and SL25010), free radically (FR) synthesized SMA polymers with molecular weights of 6, 9, and 15 kDa, or RAFT synthesized SMA polymer with a molecular weight of 6 kDa (labeled with RAFT).B: Reaction scheme for the SMA polymer with free radical polymerization and RAFT synthesis C: DLS analysis of DMPC lipid nanodiscs formed by free radical (yellow) and RAFT (blue) SMA polymers. D: SEC analysis of DMPC lipid particles formed in presence of free radical (yellow) and RAFT (blue) SMA polymers with molecular weight of 6 kDa. E: Sedimentation coefficient obtained by analytical ultra-centrifugation of the RAFT SMA polymer (6 kDa) and the commercially available SMA polymer SL25010. Negative stained electron microscopy images from DMPC lipid particles formed by the free radical (FR) polymer (average-weight molecular mass = 6 kDa) (F) and particles formed by the RAFT polymer with the same chain length (G).

To compare the homemade RAFT polymer with commercially available Xiran SL25010, analytical ultra-centrifugation has been carried out. Figure 45E shows the distribution of the sedimentation coefficient of the SL25010 polymer compared to the defined RAFT SMa-6 polymer. The SL25010 polymer is clearly heterogeneous in terms of molecular mass and exhibits three main

sedimentation coefficients. Analysis of the sedimentation coefficients revealed weight-average molecular mass ranging from 15 to 31 kDa.

With DMPC, SL25010 resulted in the formation of very small particles with <10 nm diameter. The electron microscopy images show 2-3 nm dots rather than the expected 10 nm discs shaped particles (chapter 2.7., Figure 47). These tiny particles have not been described in literature and did not fit the SEC profile shown in Figure 44. It is possible that the already formed DMPC lipid particles might disassemble on the support grid used for electron microscope and what is seen on the electron micrographs might be tangled polymers. However, this hypothesis needs to further investigation.

This study shows that the chosen polymer synthesis approach can be used to influence the molecular weight and distribution, which results in the possibility to fine-tune the homogeneity of lipid particles formed using these polymers. When used to solubilize *E. coli* membranes, the RAFT synthesized polymers showed improved solubilization efficiency compared to commercially available polymers. Overall, the creation of a defined set of polymers with controlled polymer composition and molecular weight enables the study of the influence of polymer composition on nanodiscs formation and membrane protein purification. Only after successful membrane protein isolation, further application and analysis can be performed.

2.5. Conclusion

ATP binding cassette (ABC) transporters play important roles in microbial communication by transporting signaling molecules produced within a cell into the medium. In addition, multidrug ABC exporters protect the cell from toxic compounds that enter the organism unintentionally. Because ABC transporters are membrane proteins and their purification is challenging, they are difficult to study *in vitro.* Importantly, the choice of purification method can drastically affect the activity of the protein. In this part of the thesis, the multidrug ABC exporter BmrA from *B. subtilis* was isolated by detergents and reconstituted in different membrane mimetic systems. The activity of BmrA in these systems was compared with regard to the influence of the environment on protein function.

2.5.1. A lipid environment increases the BmrA ATPase activity

The *Bacillus* multidrug resistance ATP (BmrA) protein is a *B. subtilis* multidrug membrane transporter belonging to the ATP binding cassette (ABC) superfamily (*211*). Similar to other membrane proteins, the initial purification of BmrA was challenging and involved intensive detergent screening (*211*). In this thesis, it was found that the purification of the transporter in a detergent mix consisting of dodecyl-β-D-maltoside DDM with the sterol-derivative Na-cholate yielded pure protein with the lowest basal ATPase activity compared to all other membrane mimetics tested (Figure 11E). This observation matches the results described in literature for BmrA (*194*) and homologous ABC exporters (*198–200*). Once reconstituted into a lipid environment, proteoliposomes, an increased basal ATPase activity of BmrA compared to detergent purification was observed (Figure 11E). This is in agreement with the observation described in literature and suggests that lipids support the function of BmrA or contribute to protein stability (*212, 213*).

In addition to the proteoliposome reconstitution, I showed that BmrA full-length protein can be reconstituted into different types of lipid nanodiscs. These nanodiscs are stabilized by different types of membrane scaffold proteins (MSPs) and SapA protein. In our experiments, the BmrA nanodiscs showed slightly higher basal ATPase activity, compared to the liposome reconstituted protein. However, in literature similar levels of BmrA ATPase activity have been observed in liposomes and nanodiscs (*214*). The lower ATPase activity in proteoliposomes can be explained by the random orientation of the protein in liposomes, thereby only taking those molecules into account, whose NBDs point outward and are accessible to ATP. Since the orientation of the membrane protein reconstituted into proteoliposomes cannot be controlled and was not determined in the experiments, the readout of the ATPase activity assay might be misleading at this point. To fully understand the effect of proteoliposome vs nanodiscs to BmrA function, further investigations are needed.

The comparison between reconstituted BmrA in nanodiscs and detergent solubilization shows that all reconstituted BmrA proteins had significantly higher basal ATPase activity than the detergent solubilized protein. This underlines the effect of lipids to the basal ATPase activity. Comparing all reconstituted nanodiscs samples with each other shows that the choice of the scaffold protein does not affect the BmrA ATPase activity. This indicates that the size of the obtained nanodiscs might not affect BmrA function significantly. The results of this part of the thesis are summarized in Figure 46.

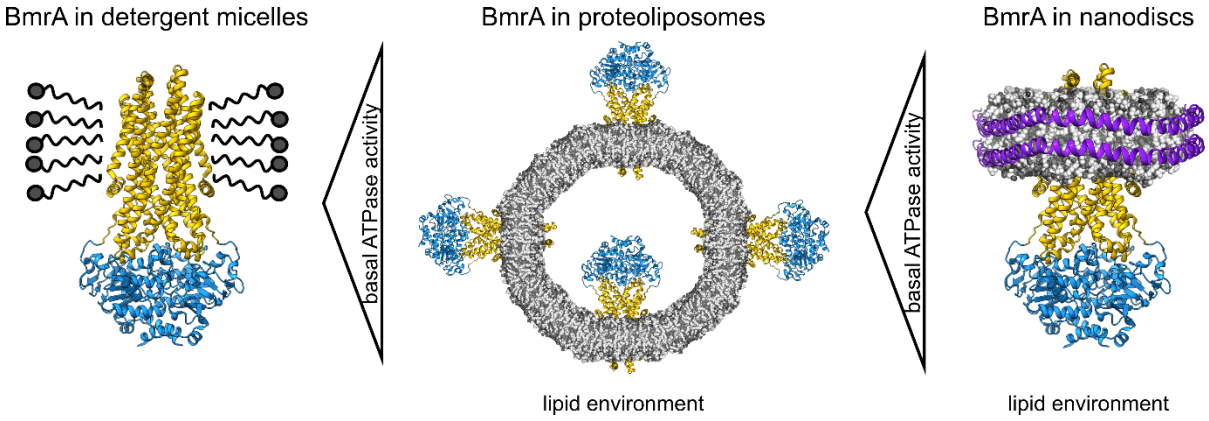


Figure 46: Relative basal ATPase activity of the ABC transporter BmrA from B. subtilis in different membrane mimetic systems. In detergent micelles (here DDM Na-cholate), the lowest basal ATPase activity was observed. While BmrA reconstituted in nanodiscs (membrane scaffold protein- or saposin-stabilized) showed the highest relative activity.

In conclusion, all purification methods used are helpful to obtain pure and active protein. This study shows that the presence of a lipid environment affects the activity of the membrane protein of interest but small lipid environmental changes, such as nanodiscs sizes do not have a significant impact on protein activity.

The aim of membrane protein research is to study the membrane protein of interest in the most native environment to avoid effects that are caused by artificial membrane systems. Although nanodiscs reconstitution is a great tool to mimic native membranes, membrane proteins must be isolated from their native lipid environment with detergents. It is possible that during protein reconstitution into nanodiscs the removing of detergents is not completed, as described in literature (*214*). As shown in this work, the lipid environment is crucial for membrane protein activity. Therefore, the study of membrane proteins in the native cell membrane is of great importance.

2.5.2. Polymer distribution affects the size distribution of the formed lipid particles

To date, the only method to purify membrane proteins in their native lipid environment is to use amphiphilic polymers (*185, 215*). Styrene maleic acid (SMA) and diisobutylene maleic acid (DIBMA) copolymers have been used extensively for the formation of nanodiscs using a one-step protocol (*204, 205, 196*). However, the nanodisc-forming polymers also have some disadvantages

such as instability in the presence of divalent cations and pH limitations (*176, 177*). Furthermore, commercially available polymers are inhomogeneous with regard to their polymer chain-length (*179*). In collaboration with the _____ group (_____), a controlled polymer synthesis using the RAFT polymerization has been established. For the best comparison with the best described RAFT SMA polymer in literature (*192*), the SMA polymer with a S:Ma ratio of 1:1, weight-average molecular mass of 6 kDa, and dispersity Đ = 1.07 (termed RAFT SMa-6) was synthesized. To compare its dispersity with the commercially available polymer, analytical ultra-centrifugation was used (Figure 12E). RAFT SMa-6 formed homogeneous lipid particles with DMPC as gauged by dynamic light scattering, size exclusion chromatography and negative stain electron microscope (Figure 12). Using RAFT SMa-6, membrane proteins from *E. coli* membranes could be successfully solubilized. The solubilization behavior of the synthesized polymer was found to be comparable to the commercially available SZ25010 (now distributed as SMALP 300) from Orbiscope.

For some biophysical techniques, such as nuclear magnetic resonance spectroscopy and small angle X-ray spectroscopy, homogeneous samples are mandatory. Here, it was shown that the quality of lipid particles and protein nanodiscs are highly influenced by the quality, i.e., the structural homogeneity of the used polymer. With the successful synthesis of RAFT SMa-6, a series of polymers can be envisioned that may be useful to overcome the challenges in membrane protein purification and could become a universal tool in the membrane protein field.

2.6. Outlook

2.6.1. Towards an understanding of the structural dynamics of BmrA in different membrane mimetic systems

To date, the structural investigation of the *B. subtilis* BmrA ABC transporter is limited to a crystal structure of the ATP inactive mutant E504A (PDB 6R72), and a Cryo electron microscope (CryoEM) structure of the same mutant in an outward-facing conformation (PDB 7OW8, 7BG4, and 6R81) (*195*). This conformation was trapped by using nucleotide and substrates (*195*). All structures have been obtained with BmrA in detergent micelles. However, for a description of the conformational space of the protein in a (near) native environment, BmrA should be investigated in other membrane mimetic systems.

The effect of membrane mimetic systems on BmrA function and protein stabilization have been studied by different working groups (*216, 195, 214*). Reconstitution of BmrA in proteoliposomes resulted in stable protein for couple of weeks (*216*) and BmrA in nanodiscs can be trapped in different conformation using nucleotides, nucleotide analoga, and substrates (*195, 214*). The protein stability of BmrA was improved in lipid nanodisc environment, which have been shown by higher melting temperatures in MSP stabilized nanodisc than in detergent micelles. To investigate BmrA in the most native environment and investigate the effect of native lipids, the lipid environment should be improved for protein reconstitution. One way to reconstitute a membrane protein such as BmrA to a native lipid like environment is to use a lipid extract from native host (here *B. subtilis* lipid extract) for protein reconstitution in proteoliposomes and nanodiscs. However, in this case the reconstitution protocol of membrane proteins into proteoliposomes or nanodiscs still starts with a detergent-based solubilization and purification. During the detergent purification step, crucial lipid-protein interactions might be disrupted, and the recovery of these interactions might be difficult, even when lipids are supplemented later. A possible approach to overcome this problem could be the isolation of membrane proteins with amphiphilic polymers directly from their native membrane and the subsequent reconstitution in liposomes and protein stabilized nanodiscs. Since the amphiphilic polymers (styrene maleic acid (SMA) and diisobutylene maleic acid (DIBMA) copolymers) are not stable in the presence of divalent cations, it may be possible to precipitate them by increasing the cation concentration. If membrane scaffold proteins are added to the system at the same time, the polymers might be replaced by the membrane scaffold proteins and the nanodiscs might remain intact.

The obtained BmrA in native lipid nanodiscs should treated with with nucleotides, nucleotide analogues and/or substrates. This would allow the capture of specific conformations in the transport cycle. Ultimately, the structures of BmrA may be obtained under turnover conditions, as demonstrated for the heterodimeric ABC exporter TmrAB in MSP nanodiscs (*217*). For the TmrAB protein, eight high-resolution structures were obtained and revealed the translocation

and ATP-hydrolysis cycle of this ABC exporter in detail (*217*). Comparing the possibly obtained structures of BmrA proteins under turnover conditions with the results obtained for TmrAB, a general concept for substrate transport can be proposed.

However, understanding the mechanism for substrate transport should not be only limited by structural findings. H/D exchange coupled to mass spectrometry (HDX-MS) (*195*) and nuclear magnetic resonance (NMR) spectroscopy (*218, 219*) are powerful techniques for studying protein dynamics. In order to understand the transport cycle of an ABC transporter such as BmrA, the influence of the lipid environment its dynamics should be addressed, and the role of the native lipids for protein structure and function should be investigated. As mentioned earlier, native lipid nanodiscs formed by amphiphilic polymers have not been fully optimized. However, styrene maleic acid lipid particles (SMALPs) have been shown to be suitable for HDX-MS (*215*). Dynamic analysis of membrane proteins in SMALPs has not been yet performed, likely due to the heterogeneity of commercially available polymers (*220*).

2.6.2. Creating a polymer library for membrane protein research

As I have shown, controlling the dispersity of the polymers used to prepare native lipid particles is essential for the resulting homogeneity of the membrane protein sample. Here, a synthesis route to produce styrene maleic acid copolymers with controlled length, composition and narrow chain-length distribution was introduced. Moving forward, polymers with variation in styrene and maleic acid composition, chain length and functionalization can be produced for membrane protein research. A polymer library consists of SMA polymers with S:Ma ration from 1:1 to 3:1, and polymer molecular weight ranging from 3000 – 12000 g/mol should be created to systematically study the effect of polymer size and composition on membrane protein solubilization behavior.

To date, it is still under debate whether membrane protein size correlates with polymer length (*188, 177*) and thus influence the yield of the membrane protein purification, since the published studies only covered a small range of polymer molecular weights (*191, 193*) or used polymers with high dispersity (*209*). With a polymer library in hand, protein size preferences, lipid composition and the quality of the formed nanodiscs can be addressed systematically. Thereby, structural biology approaches should be coupled with particle analysis using dynamic light scattering, SAXS, mass spectrometry, as well as functional assays to monitor membrane protein integrity.

Furthermore, the synthesis of polymers that can be easily modified by simple click chemistry can be envisioned. This would allow biochemists to modify the polymers with dyes or labels tailored to their use without needing to address the polymer synthesis itself. A widely used click chemistry reaction for modifying proteinogenic cysteines with labels is the thiol-ene reaction (*221*). This reaction has been shown to link fluorophores, affinity tags, spin labels, or inhibitors to the protein

of interest (*222*). To label a polymer with maleimide-functionalized molecules, the polymer must have a thiol-group. This can be easily achieved by using RAFT polymers and forming thiols from RAFT end-groups. The thiol group can be generated by nucleophilic cleavage of the thiocarbonyl group with nucleophiles, such as amines (*223*). In this approach, each polymer chain would have a thiol group that can be further functionalized. To increase the number of functionalization sites, SMA polymers could be modified with thiol amines to form thiol-functionalized SMA (SMA-SH) (*224*).

However, to create a powerful and versatile tool for membrane protein purification and structural biology the current issues of the polymers with pH tolerance (*175*) and cation concentration (*176, 177*) need to be addressed.

2.7. Appendix II

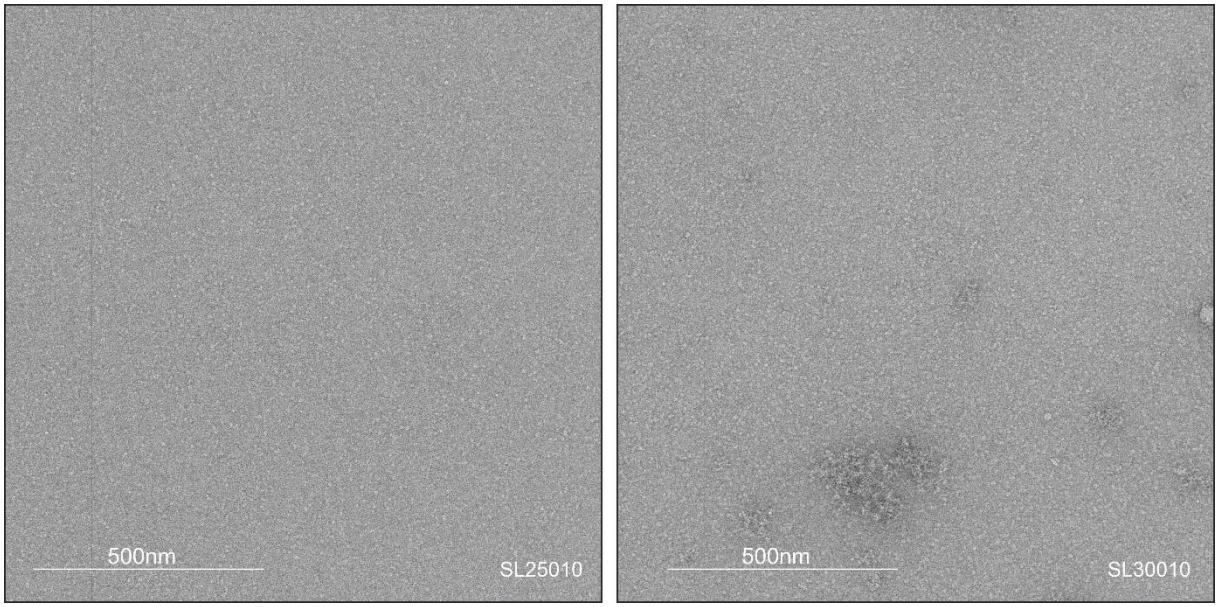


Figure 47: Negative staining micrographs of styrene maleic acid lipid particles (SMALPs) with DMPC lipids and commercially available Xiran SL25010 and SL30010 polymers.

Bacillus subtilis BmrA full-length

```
atg ggg cat cat cac cac cac cac gga tca gaa aat cta tat ttt caa ggg ccc acc aag
 M   G   H   H   H   H   H   H   G   S   E   N   L   Y   F   Q   G   P   T   K
aag cag aaa agc aag agc aag ctg aaa ccg ttc ttt gcg ctg gtt cgt cgt acc aac ccg
 K   Q   K   S   K   S   K   L   K   P   F   F   A   L   V   R   R   T   N   P
agc tac ggt aaa ctg gcg ttt gcg ctg gcg ctg agc gtg gtt acc acc ctg gtt agc ctg
 S   Y   G   K   L   A   F   A   L   A   L   S   V   V   T   T   L   V   S   L
ctg atc ccg ctg ctg acc aaa cag ctg gtt gac ggc ttt agc atg agc aac ctg agc ggc
 L   I   P   L   L   T   K   Q   L   V   D   G   F   S   M   S   N   L   S   G
acc cag atc ggt ctg att gcg ctg gtt ttc ttt gtg caa gcg ggt ctg agc gcg tac gcg
 T   Q   I   G   L   I   A   L   V   F   F   V   Q   A   G   L   S   A   Y   A
acc tat gcg ctg aac tac aac ggc caa aag atc att agc ggt ctg cgt gag ctg ctg tgg
 T   Y   A   L   N   Y   N   G   Q   K   I   I   S   G   L   R   E   L   L   W
aag aaa ctg atc aaa ctg ccg gtt agc tat ttc gac acc aac gcg agc ggc gaa acc gtg
 K   K   L   I   K   L   P   V   S   Y   F   D   T   N   A   S   G   E   T   V
agc cgt gtg acc aac gat acc atg gtg gtt aag gaa ctg atc acc acc cac att agc ggc
 S   R   V   T   N   D   T   M   V   V   K   E   L   I   T   T   H   I   S   G
ttc atc acc ggt atc att agc gtg atc ggt agc ctg acc atc ctg ttt att atg aac tgg
 F   I   T   G   I   I   S   V   I   G   S   L   T   I   L   F   I   M   N   W
aag ctg acc ctg ctg gtt ctg gtg gtt gtt ccg ctg gcg gcg ctg att ctg gtg ccg att
 K   L   T   L   L   V   L   V   V   V   P   L   A   A   L   I   L   V   P   I
ggt cgt aaa atg ttc agc att agc cgt gaa acc cag gac gaa acc gcg cgt ttt acc ggt
 G   R   K   M   F   S   I   S   R   E   T   Q   D   E   T   A   R   F   T   G
```

```
ctg ctg aac caa att ctg ccg gag atc cgt ctg gtt aag gcg agc aac gcg gag gat gtg
 L   L   N   Q   I   L   P   E   I   R   L   V   K   A   S   N   A   E   D   V
gaa tat ggc cgt ggt aaa atg ggc atc agc agc ctg ttc aag ctg ggt gtt cgt gaa gcg
 E   Y   G   R   G   K   M   G   I   S   S   L   F   K   L   G   V   R   E   A
aaa gtt cag agc ctg gtg ggc ccg ctg atc agc ctg gtg ctg atg gcg gcg ctg gtt gcg
 K   V   Q   S   L   V   G   P   L   I   S   L   V   L   M   A   A   L   V   A
gtg att ggt tac ggt ggc atg caa gtt agc agc ggt gaa ctg acc gcg ggt gcg ctg gtt
 V   I   G   Y   G   G   M   Q   V   S   S   G   E   L   T   A   G   A   L   V
gcg ttc atc ctg tac ctg ttc cag atc att atg ccg atg ggc caa atc acc acc ttc ttt
 A   F   I   L   Y   L   F   Q   I   I   M   P   M   G   Q   I   T   T   F   F
acc cag ctg caa aag agc att ggt gcg acc gag cgt atg atc gaa att ctg gcg gag gaa
 T   Q   L   Q   K   S   I   G   A   T   E   R   M   I   E   I   L   A   E   E
gag gaa gac acc gtt acc ggc aaa cag att gaa aac gcg cac ctg ccg atc caa ctg gac
 E   E   D   T   V   T   G   K   Q   I   E   N   A   H   L   P   I   Q   L   D
cgt gtg agc ttt ggt tac aag ccg gat cag ctg att ctg aaa gag gtt agc gcg gtg atc
 R   V   S   F   G   Y   K   P   D   Q   L   I   L   K   E   V   S   A   V   I
gaa gcg ggc aaa gtt acc gcg att gtg ggt ccg agc ggt ggc ggt aaa acc acc ctg ttc
 E   A   G   K   V   T   A   I   V   G   P   S   G   G   G   K   T   T   L   F
aaa ctg ctg gag cgt ttt tat agc ccg acc gcg ggt acc atc cgt ctg ggt gac gaa ccg
 K   L   L   E   R   F   Y   S   P   T   A   G   T   I   R   L   G   D   E   P
gtt gat acc tac agc ctg gag agc tgg cgt gaa cac att ggt tat gtg agc cag gag agc
 V   D   T   Y   S   L   E   S   W   R   E   H   I   G   Y   V   S   Q   E   S
ccg ctg atg agc ggc acc att cgt gag aac atc tgc tac ggt ctg gaa cgt gac gtt acc
 P   L   M   S   G   T   I   R   E   N   I   C   Y   G   L   E   R   D   V   T
gat gcg gag atc gaa aag gcg gcg gag atg gcg tat gcg ctg aac ttc att aaa gaa ctg
 D   A   E   I   E   K   A   A   E   M   A   Y   A   L   N   F   I   K   E   L
ccg aac cag ttt gat acc gag gtg ggt gag cgt ggt atc atg ctg agc ggc ggt cag cgt
 P   N   Q   F   D   T   E   V   G   E   R   G   I   M   L   S   G   G   Q   R
caa cgt atc gcg att gcg cgt gcg ctg ctg cgt aac ccg agc att ctg atg ctg gac gag
 Q   R   I   A   I   A   R   A   L   L   R   N   P   S   I   L   M   L   D   E
gcg acc agc agc ctg gat agc cag agc gaa aag agc gtt cag caa gcg ctg gaa gtg ctg
 A   T   S   S   L   D   S   Q   S   E   K   S   V   Q   Q   A   L   E   V   L
atg gaa ggt cgt acc acc atc gtt att gcg cac cgt ctg agc acc gtt gtg gac gcg gat
 M   E   G   R   T   T   I   V   I   A   H   R   L   S   T   V   V   D   A   D
caa ctg ctg ttc gtg gag aaa ggc gaa atc acc ggc cgt ggt acc cac cac gaa ctg atg
 Q   L   L   F   V   E   K   G   E   I   T   G   R   G   T   H   H   E   L   M
gcg agc cac ggt ctg tac cgt gat ttt gcg gaa caa caa ctg aag atg aat gcg gac ctg
 A   S   H   G   L   Y   R   D   F   A   E   Q   Q   L   K   M   N   A   D   L
gag aat aag gcg ggc taa
 E   N   K   A   G   -
```

MSP1

```
atg ggc agc agc cat cat cat cat cat cat gaa aac ctg tat ttt cag ggc ctg aaa ctg
```

```
 M   G   S   S   H   H   H   H   H   H   E   N   L   Y   F   Q   G   L   K   L
ctg gat aac tgg gat agc gtg acc agc acc ttt agc aaa ctg cgt gaa cag ctg ggc ccg
 L   D   N   W   D   S   V   T   S   T   F   S   K   L   R   E   Q   L   G   P
gtg acc cag gaa ttt tgg gat aac ctg gaa aaa gaa acc gaa ggc ctg cgt cag gaa atg
 V   T   Q   E   F   W   D   N   L   E   K   E   T   E   G   L   R   Q   E   M
agc aaa gat ctg gaa gag gtg aaa gcg aaa gtg cag ccg tat ctg gat gac ttt cag aaa
 S   K   D   L   E   E   V   K   A   K   V   Q   P   Y   L   D   D   F   Q   K
aaa tgg cag gaa gag atg gaa ctg tat cgt cag aaa gtg gaa ccg ctg cgt gcg gaa ctg
 K   W   Q   E   E   M   E   L   Y   R   Q   K   V   E   P   L   R   A   E   L
cag gaa ggc gcg cgt cag aaa ctg cat gaa ctg cag gaa aaa ctg agc ccg ctg ggc gaa
 Q   E   G   A   R   Q   K   L   H   E   L   Q   E   K   L   S   P   L   G   E
gag atg cgt gat cgt gcg cgt gcg cat gtg gat gcg ctg cgt acc cat ctg gcg ccg tat
 E   M   R   D   R   A   R   A   H   V   D   A   L   R   T   H   L   A   P   Y
agc gat gaa ctg cgt cag cgt ctg gcg gcc cgt ctg gaa gcg ctg aaa gaa aac ggc ggt
 S   D   E   L   R   Q   R   L   A   A   R   L   E   A   L   K   E   N   G   G
gcg cgt ctg gcg gaa tat cat gcg aaa gcg acc gaa cat ctg agc acc ctg agc gaa aaa
 A   R   L   A   E   Y   H   A   K   A   T   E   H   L   S   T   L   S   E   K
gcg aaa ccg gcg ctg gaa gat ctg cgt cag ggc ctg ctg ccg gtg ctg gaa agc ttt aaa
 A   K   P   A   L   E   D   L   R   Q   G   L   L   P   V   L   E   S   F   K
gtg agc ttt ctg agc gcg ctg gaa gag tat acc aaa aaa ctg aac acc cag taa
 V   S   F   L   S   A   L   E   E   Y   T   K   K   L   N   T   Q   -
```

MSP1D1Δ5

```
atg ggc agc agc cat cat cat cat cat cat gaa aac ctg tat ttt cag ggc agc acc ttt
 M   G   S   S   H   H   H   H   H   H   E   N   L   Y   F   Q   G   S   T   F
agc aaa ctg cgt gaa cag ctg ggc ccg gtg acc cag gaa ttt tgg gat aac ctg gaa aaa
 S   K   L   R   E   Q   L   G   P   V   T   Q   E   F   W   D   N   L   E   K
gaa acc gaa ggc ctg cgt cag gaa atg agc aaa gat ctg gaa gag gtg aaa gcg aaa gtg
 E   T   E   G   L   R   Q   E   M   S   K   D   L   E   E   V   K   A   K   V
cag ccg tat ctg gat gac ttt cag aaa aaa tgg cag gaa gag atg gaa ctg tat cgt cag
 Q   P   Y   L   D   D   F   Q   K   K   W   Q   E   E   M   E   L   Y   R   Q
aaa gtg gaa ccg ctg ggc gaa gag atg cgt gat cgt gcg cgt gcg cat gtg gat gcg ctg
 K   V   E   P   L   G   E   E   M   R   D   R   A   R   A   H   V   D   A   L
cgt acc cat ctg gcg ccg tat agc gat gaa ctg cgt cag cgt ctg gcg gcc cgt ctg gaa
 R   T   H   L   A   P   Y   S   D   E   L   R   Q   R   L   A   A   R   L   E
gcg ctg aaa gaa aac ggc ggt gcg cgt ctg gcg gaa tat cat gcg aaa gcg acc gaa cat
 A   L   K   E   N   G   G   A   R   L   A   E   Y   H   A   K   A   T   E   H
ctg agc acc ctg agc gaa aaa gcg aaa ccg gcg ctg gaa gat ctg cgt cag ggc ctg ctg
 L   S   T   L   S   E   K   A   K   P   A   L   E   D   L   R   Q   G   L   L
ccg gtg ctg gaa agc ttt aaa gtg agc ttt ctg agc gcg ctg gaa gag tat acc aaa aaa
 P   V   L   E   S   F   K   V   S   F   L   S   A   L   E   E   Y   T   K   K
ctg aac acc cag taa
 L   N   T   Q   -
```

MSP2N2

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

MSP1E3D1

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

Chapter 3

The receptor

Chapter 3: The receptor

3.1. Theoretical background

One of the best studied signaling pathways between bacteria are two-component systems (TCSs) (*225, 226*). TCSs comprise a sensor histidine kinase (SHK) and a response regulator (RR). The membrane-bound histidine kinase senses external stimuli, such as nutrient concentration, osmotic pressure, pH, redox states, signaling molecules or membrane stress (*225*). Sensing of external stimuli causes autophosphorylation of the SHK in the cytoplasm. With this mechanism the signal crosses the membrane. Intracellularly, signal transduction is achieved by phosphoryl-transfer reaction between the SHK and the RR proteins (*225*).

In the Gram-positive soil-bacterium *B. subtilis*, the histidine kinase SpaK senses the lanthipeptide subtilin. Subtilin is a ribosomally synthesized, and post translationally modified peptide (RiPP) (*227, 129*) produced by *B. subtilis* to act as antibiotics against a wide range of gram-positive bacteria (*228, 229*). Subtilin is transported out of the cell by an ATP binding cassette (ABC) transporter (*230*)(chapter 2.1).

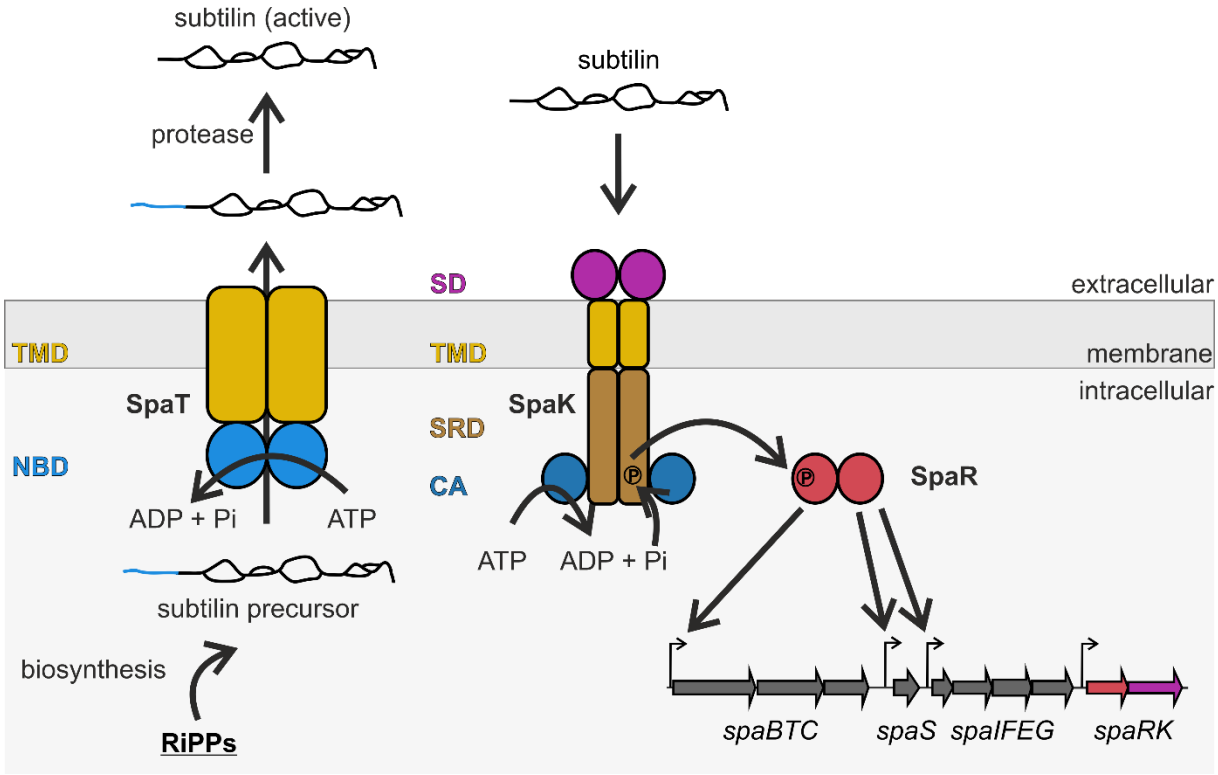


Figure 48: Schematic overview of subtilin production and sensing in B. subtilis. The subtilin precursor is a ribosomally synthesized and post-translationally modified peptide (RiPP) that is exported by the ATP binding cassette (ABC) transporter SpaT. SpaT consumes adenosine triphosphate (ATP) via its nucleotide binding domains (NBDs). The subtilin precursor peptide is cleaved by B. subtilis proteases to form active subtilin that can bind to the sensor domain (SD) of the histidine kinase SpaK. The signal is transduced from the SD via the histidine kinase's transmembrane domain (TMD) and signal-relay domain (SRD) to the catalytic ATP binding (CA) domain. The CA domain uses ATP to catalyze the autophosphorylation at a conserved histidine residue (circled P). Next, the phosphate group is transferred to the response regulator protein SpaR. Phosphorylated SpaR then binds to the spa gene cluster and regulates the expression of genes mediating and regulating subtilin production, transport, and recognition.

The binding of subtilin to SpaK results in the autophosphorylation of SpaK at the conserved histidine residue H247 (*231*). This phosphate is subsequently transferred to the conserved aspartate residue D51 of the RR protein SpaR (*231*). Phosphorylated SpaR then binds to the promoter of the Spa gene cluster, thus activating the expression of biosynthesis genes for the signaling molecule subtilin, all transporter proteins (SpaT, SpaFEG), as well as immunity proteins, and also the receptor system SpaK and SpaR (*232*). A schematic overview is in Figure 48.

Although the basic principles of TCSs are well studied (*233–236*), gaps remain in our understanding of the molecular details of the signaling pathways and the connection between sensory perception and signal output (*237*). In this thesis, structural studies on SpaK were carried out look into the molecular mechanism of TCS signaling.

3.1.1. Sensor histidine kinase structure and function

A canonical sensor histidine kinase (SHK) composes an extracytoplasmic or periplasmic sensor (SD) domain, a transmembrane domain (TMD), a signal-relay module (or transduction module TRA) (*237*), and a kinase domain composed of a dimerization and histidine phosphorylation (DHp) domain, and a catalytic ATP binding (CA, or AB) domain (*237, 238*). Intracellular signal-relay modules are HAMP (named after histidine kinases, adenylyl cyclases, methyl binding proteins, phosphatases), PAS (Per-ARNT-SIM), and GAF domains, found in cGMP-Specific phosphodiesterase adenylyl cyclases and FhlA (*238*). The majority of SHKs are homodimeric (*226*). SHKs are bifunctional enzymes that can perform kinase and phosphatase activities in a signal-dependent manner (*226*). In the kinase mode, the protein performs the autophosphorylation of a conserved histidine residue and transfers the phosphoryl group to the response regulator protein (*226*). The phosphatase reaction catalyzes the dephosphorylation of the phosphorylated response regulator protein (*226*).

To date, structural data on full-length membrane SHKs are missing, due to many technical challenges during protein purification (*239*)(chapter 2.1). However, the "divide and conquer" approach allowed the structural and functional investigation of individual SHK domains, multi-domains and domain hybrids (*239*). The role of each domain is summarized in the following sections.

3.1.1.1. The sensor domain – tailored to recognize environmental signals

The N-terminal periplasmic sensor domain of many SHKs is responsible for sensing environmental signals (*238*). This information is then transmitted to the cytoplasmic kinase domain via the transmembrane domain and intracellular signal domain(s) (*238, 239*). The structural features of sensor domains are as diverse as the ligands that can be sensed (*239*). However, common folds have been identified amongst different sensor domains (*239*): The mixed

α-helix and β-sheet (in the following α,β-fold), and the α-helical fold. All α,β-folded domain consist of an antiparallel β-sheet and several flanking α-helices (Figure 49A and B) and are described under various names in the literature: PDC (after the three founding members, PhoQ, DcuS and CitA, Figure 49A), Per/Arnt/Sim (PAS) fold or Cache fold (*240*). Although the secondary structure of α,β-folded domains is highly conserved, the protein sequence is diverse. Although all α,β-folded SDs share the same secondary structure ligand specificity is present (*238*). Ligands bind to α,β-folded sensor domains within a conserved cleft and in some proteins that sense metal ligands, binding occurs at the surface of the domain (*241*). The dimerization behavior of α,β-folded proteins is essential for protein function and modulates protein-protein interactions (*241*).

The α-helical folded sensor domain is found in different types of SHKs, such as nitrate- and nitrite SHKs (*242*), e.g., NarQ (Figure 49C) and NarX, trimethylamine-N-oxide SHKs (*243, 244*), and chemoreceptors, e.g., McpS (*245*). This sensor domain type consists of either a four-helical bundle (Figure 49C and D) or a duplicated four-helical bundle that allows the binding of multiple ligands at the same time (*236, 245*). In α-helical folded sensor domain a conserved periplasmic box (P-box, Figure 49C, GxxR motif, x any residues) was discovered that binds ligands, resulting in the dimerization of the sensor domains (*242*). A piston-like motion of the four-helical bundle accompanied with helical rotations have been proposed to transfer the signal sensed by the sensor domain to the transmembrane domain (TMD) (*236*).

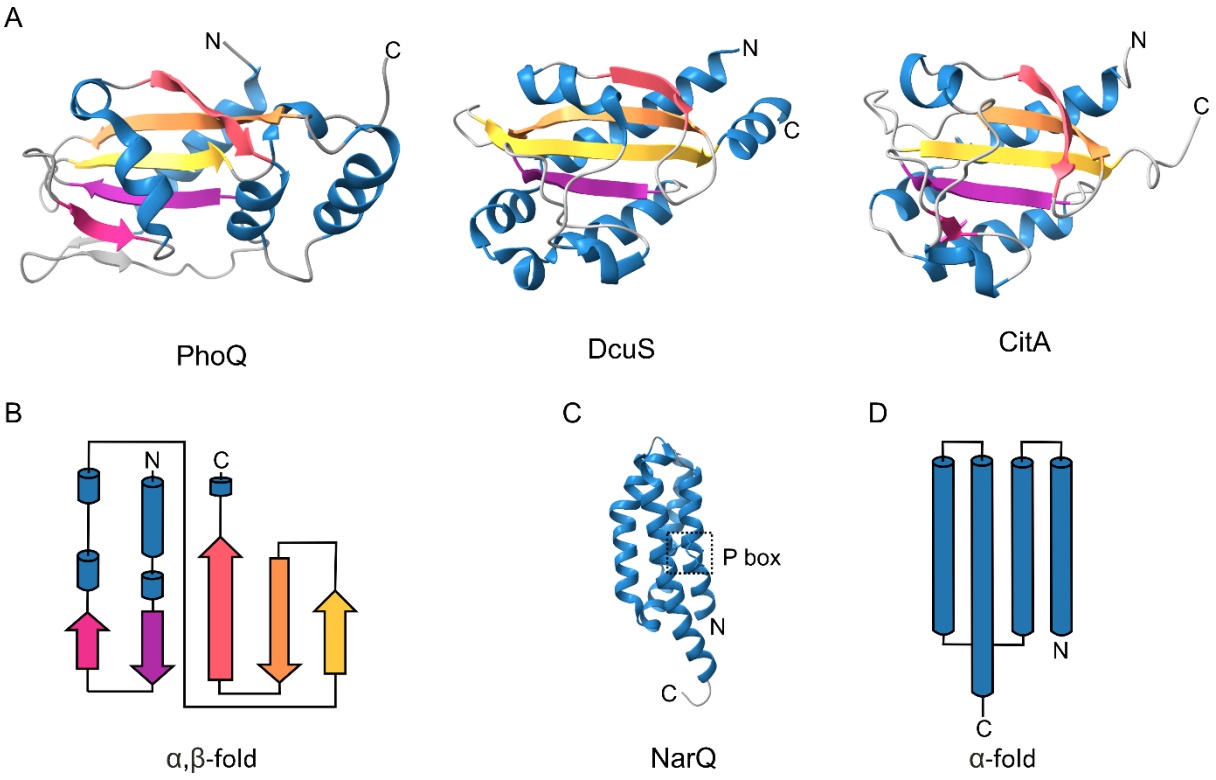


Figure 49: Representative structures and topology diagrams for α,β-folded and a-helical folded sensor domains. A: α,β–folded domains from PhoQ (PDB: 1YAX) and DcuS (PDB: 3BY8) from Escherichia coli, and CitA (PDB: 1P0Z) from Klebsiella pneumoniae. The antiparallel β-sheets are colored from purple to yellow from N- to C-terminus. α-helices are colored blue, and loop and variable secondary elements in grey. B: Topology diagram for the α,β-fold. C: NarQ from Escherichia coli represents the α-helical folded sensor domains. The conserved P box (GxxR-motif) and the substrate binding site is marked with a box. D: Topology diagram for the α-helical fold found in bacterial sensor domains.

Each monomer of a homodimeric SHK forms a two-helix coiled coil that is flanked by the helices from the opposing monomer thereby forming a four-transmembrane helix bundle (*226*). Only few atomic structures of SHK transmembrane domains are available and how a signal crosses the membrane barrier has not been fully understood. However, several mechanisms have been proposed involving piston shift motions, diagonal scissoring, and helix twisting of the TMD upon ligand binding in the sensor domain (*226, 236*) (Figure 50). The piston model has been suggested for the fumarate SHK DcuS from *E. coli (246, 247)*, citrate sensor kinase CitA from *Klebsiella pneumoniae (248)*, and BvgS from *Bordetella pertussis (249)*. It was shown by cross-linking (*246*), cysteine accessibility (*249*), and solid state nuclear magnetic resonance (NMR) spectroscopy experiments (*248*) that substrate binding to the periplasmic sensor domain forces a piston-like upward motion of two of the four transmembrane helices (*248, 246*).

The scissoring model was proposed for the TMD movements for the *E. coli* SHK PhoQ and VxRA from *Vibrio cholerae (250)*. TMD motions of these systems were monitored by cross-linking experiments, statistical modeling (*251*), and crystallization of the isolated sensor domains (*250*). Upon substrate binding to the SD, scissor-like movement in the TMD were proposed to influence the reordering of the cytosolic signal-relay domain and to induce kinase activity.

The last model, helical rotation of the TMD, was proposed for LuxQ from *Vibro harveyi (252),* DesK from *B. subtilis (253)*, and AgrC from *Staphylococcus aureus (254)*. The rotational motion of the TMD is then transmitted to the catalytical domain (*255*).

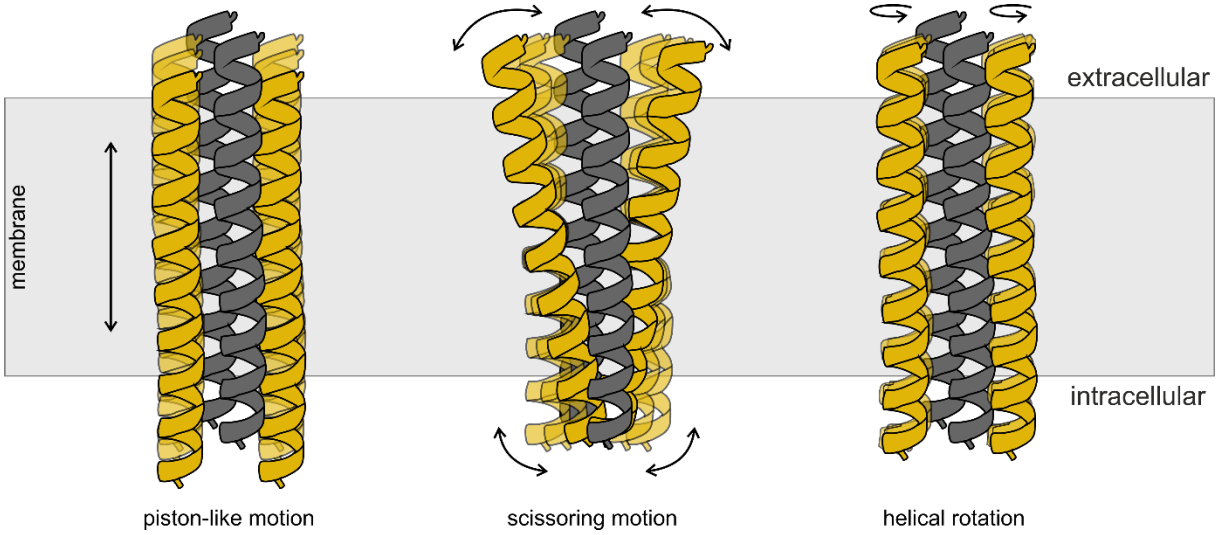


Figure 50: Schematic models to demonstrate movements of the transmembrane domain. The transmembrane domain is shown here as a four-helical bundle with grey and yellow helices. Motions are only shown for yellow helices and show relative movements to the grey helices.

3.1.1.2. Signal transduction in the cytoplasm and the autophosphorylation reaction

In SHKs, HAMP, PAS or GAF domains are found between the transmembrane domain and catalytical domain. HAMP domains are part of exceptionally many signal transduction systems, and couple sensory and output domains with each other (*256*). Cytoplasmic PAS and GAF domain

are found in almost half of all SHKs (*226*). Similar to extracellular PAS domains, cytoplasmic PAS domains are formed by a central β-sheet flanked by a-helices. PAS and GAF domain act as additional cytoplasmic sensors and are often found C-terminally to HAMP domain (*257*). However, signal transduction through PAS and GAF domains is not well studied (*236, 226*) although the input signal might cause structural reorganization of the domain and thus affect the SHK dimer interface or the intra-protomer crosstalk (*258–260*).

The HAMP domain is formed by a four-helical bundle with two helices per monomer (*261*). The two helices carry a characteristic coiled coil structure and are joined by a flexible linker (*261*). Similar to the mechanisms proposed for the TMD, scissoring, pistoning and helical rotation models describe the conformational changes of the HAMP domain (*236*). Regardless of the type of conformational motion involved, HAMP domains show symmetric conformations and are responsible for maintaining a kinase-off state (*262*). Mutations in the HAMP domain can result in a decoupling of the kinase activity from substrate binding at the sensor domain. Mutations in the HAMP domain can influence the kinase-off state, resulting in a constitutive active SHK (*263*). In the SHK PhoQ from *S. Typhimurium*, the constitutive active phenotype resulted in increased phosphorylation of the response regulator protein PhoP and a decreased in phosphatase activity of the histidine kinase PhoQ. Constitutive active PhoQ mutants are avirulent, while wildtype proteins are essential for full virulence survival of the host within macrophages (*264*). This underlining the importance of the induced activation of the protein (*263*).

Once an extracellular signal has migrated via the TMD and HAMP domain, it reaches the highly conserved C-terminal kinase domain with its functional histidine residue, that gives the protein family its name (*255*). The structure of the CA domain is highly conserved. The minimal CA core structure, also known as the Bergerat fold is formed by five stranded β-sheet opposed to a layer of three α-helices (*237, 265*) (Figure 51A) and contain various conserved motifs shown in Figure 50B. ATP binding to the CA domain is stabilized by a conserved aspartate that forms a hydrogen bond with the adenine base. This aspartate in a glycine rich region is part of a conserved motif (DxGxGh motif, with x any residues, h hydrophobic residues) named box G1 (*266*). The asparagine and a glutamine in the signature motif N box (NhhxNAh motif) form hydrogen bond interactions with the β-phosphate of ATP and coordinate a Mg^{2+} ion (*267, 265, 268*). The F box (hFxxF-motif) and a second glycine rich region (GGxLGLx-motif) named G2 box, form a so-called ATP lid that interacts with the nucleotide (*265*). The ATP lid varies in length and conformation depending on the histidine kinase (*268*). Mediating autophosphorylation, the CA domain transfers the γ-phosphate of ATP to the conserved histidine residue in the DHp domain.

Each monomer of the dimeric dimerization and histidine phosphorylation DHp domain consists of two α-helices that together form an antiparallel coiled-coil structure (Figure 51C). This four-helix bundle is flanked by the two CA domains thus forming the entire kinase domain (Figure 51C). The conserved histidine residue is found in the middle of the first DHp domain helix, in the conserved H-box (*267*)(Figure 51B) with the signature motif h(S/T/A)H(D/E)h(R/K)TPL, h hydrophobic residues. The autophosphorylation reaction follows the rules of a nucleophilic substitution mechanism in which the histidine acts as a phosphor-acceptor and nucleophilcally

attacks the γ-phosphate of ATP. The nucleophilicity of the histidine is enhanced by adjacent acidic residues (*269*).

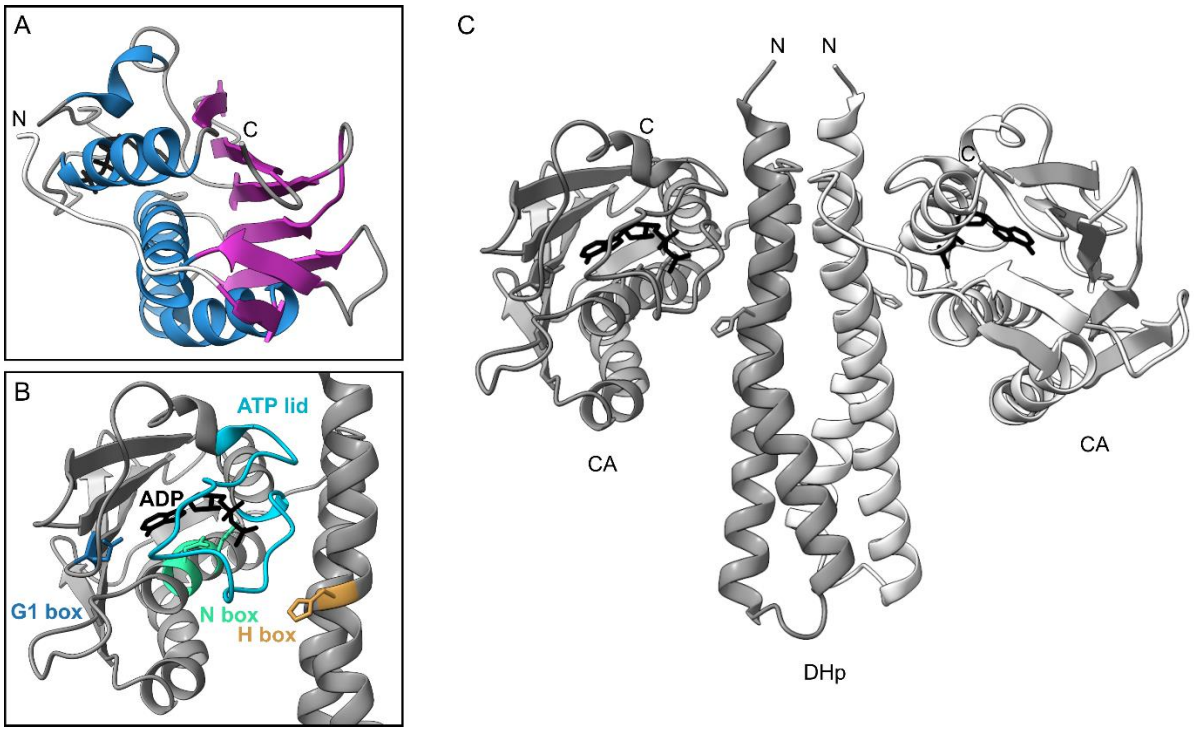


Figure 51: Structure and conserved motifs of the kinase domain (PDB: 3DGE) from Thermotoga maritima. A: The Bergerat fold of the catalytic ATP binding (CA) domain. β-sheets are colored purple, α-helices are colored blue. B: The conserved motifs in the CA and DHp domains are shown in colors. The CA domain interacts with nucleotides, here ADP (black). C: The homodimeric kinase domain (PDB: 3DGE) consists of DHp and CA domain. The monomers are shown in grey tone for clarity. ADP is shown in black sticks.

Although signal-relay domains and kinase domains were shown in a symmetrical conformation, an asymmetric conformation has been observed when the two domains are connected (*270, 271*). Kink positions were found along the HAMP and DHp helices resulting in DHp helix bending (*272, 273*). As a result, only one CA-domain-bound ATP molecule reaches phosphorylatable histidine residue of the DHp domain at any given time (*226, 274, 260*). This CA domain is in a so-called active state, while the CA domain from the other protomer is in an inactivated state (*260*). ATP binds to both active and inactive CA domains (*273, 275*). Autophosphorylation can occur intramolecularly, i.e., in *cis* within one monomer or in *trans (257, 276)*, i.e., intermolecularly across the dimer. It has been hypothesized that the length and structure of the loop between the two helices of each DHp monomer define the mechanism for the autophosphorylation reaction (*276*). A short loop between the two DHp helices of one monomer might create a large angle between these helices and the CA domain that is bound N-terminally to the second DHp helix might interact with the histidine in first helix of the other protomer. This results in a *trans* phosphorylation reaction (*276*). With a long loop connecting the two helices of one monomer, the CA domain from the same monomer interacts within the monomer due to close distance between the two DHp helices (*276*).

Once the conserved histidine is phosphorylated, the interaction with the RR protein enables a phosphotransfer reaction and subsequent activation of the RR protein.

3.1.1.3. Interaction with the response regulator proteins and the phosphotransfer reaction

After autophosphorylation of the conserved histidine in the SHK's Dhp domain, a complex between DHp domain and RR protein needs to be formed for phosphate transfer. During autophosphorylation reaction the active CA domain binds the conserved histidine at the DHp domain. Since the phosphate group at this histidine must interact with the response regulator protein, the active CA domain must make space to allow the RR protein to reach the DHp surface. It is proposed that CA domains alternate between active and inactive conformations in a cyclic manner (*277*). A back and forth bending motion of the DHp domain is required to switch the CA domain between an active and an inactive state (*277*). Inactive CA domain allows the receiver domain (REC) of the RR protein to bind at the DHp domain and phosphotransfer reaction occurs (*237*). At the same time a CA domain binds to the second protomer and autophosphorylation takes place (*277*).

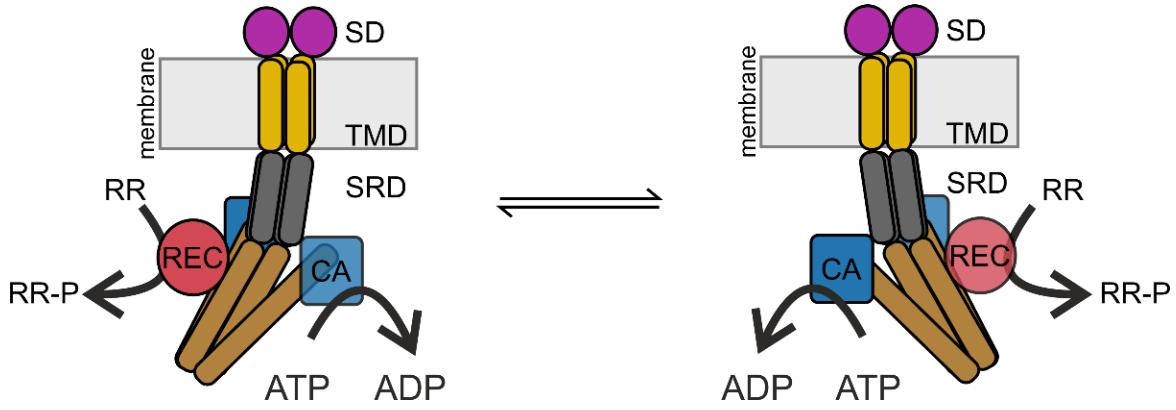


Figure 52: Proposed allosteric switching mechanism for the autophosphorylation and phosphotransfer activity in two-component system-mediated signal transduction. The catalytic ATP (CA) binding domain of the sensor histidine kinase (SHK) alternates between inactive and active states cyclically and in synchrony with the binding and release of the receiver (REC) domain of the response regulator (RR) protein. SD: sensor domain, TMD: transmembrane domain, SRD: signal-relay domain. Figure modified after Mechaly et al. (277).

RR proteins always consist of a REC domain with an α/β Rossmann-like fold composed of five-stranded parallel β–sheet surrounded by five α–helices (*278*). The REC domain carries two conserved aspartates in the active site required for coordinating Mg^{2+} ions and nucleophilic attack of the phosphor-histidine (*226, 275*). Phosphotransfer from the DHp histidine to the REC aspartate takes place. After phosphate transfer, the phosphate oxygens are coordinated by a threonine or serine and a lysine residue. However, these findings rely on beryllofluoride (BeF_3-) experiments due to the instability of the acyl phosphate group at the phosphorylated aspartate (*279*). With the exception of the phosphorylation site, REC domains share only up to 30% sequence identity (*278*). This allows RRs to couple phosphorylation to diverse effector functions. Furthermore, the REC domain may be coupled to additional effector domains, that are responsible for e.g., DNA, ligand, and protein binding (*278*).

Two thirds of RRs contain a DNA-binding domain (DBD) as an effector domain. This reflects the importance of transcriptional regulation as a response to environmental changes (*278*). Due to

the lack of full-length RR structures in complex with SHKs, it is unclear how effector domains might be positioned relative to their own REC as well as the corresponding SHK domains. A model of the interaction between the *Bradyrhizobium japonicum* SHK FixL and its RR protein suggest free movements of RR effector domains, which does not participate in binding to the SHK (*201*). Upon phosphorylation, the RR protein undergoes dramatic conformational changes (*278*). As shown for the DNA binding RR VraR from *Staphylococcus aureus*, the monomeric unphosphorylated VraR exist in a closed conformation that inhibits DNA binding (*280*). A similar closed conformation has been observed with the RR protein NarL from *E. coli (281).* Once activated by phosphotransfer or the phosphoryl analog beryllofluoride, the protein dimerized that promotes DNA binding at target promoters (*280, 282*).

While the general concepts of the operation of two-component system is understood, our knowledge of the molecular details for signal sensing, transduction, and transfer is still limited. The major challenge remains the purification of the full-length proteins of both components, the membrane bound histidine kinase and the soluble response regulator in their respective active forms. This part of the thesis deals with the structural investigation of a membrane bound histidine kinase, SpaK from *B. subtilis*. Structural findings might contribute to understand the function of this protein class that is highly important for microbial communication.

3.2. Aims and objectives

Two-component systems (TCSs) play a crucial role in microbial communication. TCSs consist of a membrane bound sensor histidine kinase (SHK) and a cytoplasmic response regulator (RR) protein. The system senses external stimuli, relays the signal into the cytoplasm, and regulates gene expression in response to environmental changes. Extracellularly, the sensor domain (SD) of the SHK binds substrates, such as nutrients, signaling molecules, and bioactive compounds (*225*). This interaction is converted into a signal that is transduced through the transmembrane domains (TMD), the signal-relay domain(s) (SRD), and into the kinase domains leading to the autophosphorylation at a conserved histidine residue. This phosphate group is then transferred to a conserved aspartate in the RR protein, which enables DNA binding of the cytoplasmic protein. To date, structural and functional investigations of TCSs are limited to isolated domains, as studies on full-length proteins often fail due to unsuccessful protein purification. Accordingly, not a single full-length membrane bound SHK structure has been reported.

The aim of this work was the structural characterization of the SHK SpaK from *Bacillus subtilis.* SpaK and its RR protein SpaR form the TCS that responds to the lanthipeptide subtilin. SpaK was expressed as a full-length protein and purified directly from the native membrane using amphiphilic styrene maleic acid (SMA) polymers. This technique allows to maintain the native protein-lipid interaction. SpaK in polymer-based lipid particles was structurally characterized by cryo-electron microscope (Cryo-EM). Special attention was paid to the SD from SpaK to understand substrate selectivity. Interestingly, the lanthipeptide sensor is highly specific and molecules similar to subtilin, such as nisin produced by *Lactococcus lactis* (*L. lactis*), cannot activate SpaK (*283, 284*). Due to challenges with protein expression and purification of the isolated SpaK SD, the homologous NisK SD from *L. lactis* was purified with a refolding procedure and analyzed structurally.

In summary, structural findings for SpaK provides insight into signal transduction in TCSs. This might help in the understanding of how bacteria communicate by sensing the presence of lanthipeptides.

3.3. Materials and methods

Xiran 30010, with the new name SMALP 200, was a kind gift from Orbiscope and was also supplied as a ready to use solution (20% w/v).
Subtilin was purified by _____ (_____lab, _____) and dialyzed against 20 mM tris, pH 8.5, 150 mM NaCl buffer.
All plasmids used in this chapter were cloned by _____ (_____lab, _____) with a C-terminal His_{10}-Tag and a cleavage site for the human rhinovirus (HRV) 3C protease.

All plasmids, which have been used in this chapter are listed in Table 11.

Table 11: Plasmids coding for SpaK full-length proteins and NisK sensor domain are listed with their respective plasmid and protein properties. Physicochemical properties of SpaK and NisK proteins were calculated using the ProtParam tool from the ExPasy website (https://web.expasy.org/protparam/).

Plasmid number	**Insert**	**Tag / cleavage site**	**Vector backbone**	**Resistance**	**Origin**	**No. amino acids**	**PI**	**$\varepsilon_{280\ nm}$ [$M^{-1}\ cm^{-1}$]**
P941	SpaK full-length	C-term. His_{10} / 3C	pBXC3H	ampicillin	_____ lab (_____)	459	7.73	43780
P1024	NisK sensor domain	N-term. His_{10} / 3C	pBXNH3	ampicillin	_____ lab (_____)	91	6.12	8940
P943	SpaK full-length L179P	C-term. His_{10} / 3C	pBXC3H	ampicillin	_____ lab (_____)	459	7.73	43780
P944	SpaK full-length L198P	C-term. His_{10} / 3C	pBXC3H	ampicillin	_____ lab (_____)	459	7.73	43780

3.3.1. Expression of the histidine kinase SpaK from *Bacillus subtilis*

Plasmid coding for the SpaK full-length protein was transformed in *E. coli* MC1061 strains. A colony was used to inoculated terrific broth (TB) medium supplemented with 2 mM $MgSO_4$ and 100 µg/ml ampicillin. Thereby, a 5 l flask should only filled with 1 l medium. The culture was incubated at 37 °C until an OD_{600} = 0.2 was reached. Then, the culture was transferred to 25 °C and induced with 0.1% w/v once the OD_{600} = 0.4 – 0.6. After 14 – 16 h at room temperature, the cells were harvested by centrifugation (10 min, 4 °C, 10.000 x *g*). Cell pellets were stored at -20 °C until further use.

3.3.2. Membrane preparation

Harvested cells from 1 l culture volume were resuspended in 100 ml lysis buffer (50 mM tris pH 8.8 (adjusted at 4 °C), 300 mM NaCl, 2 mM DTT), supplemented with 1 mM DNase I, 1:1000 dilution of a protease inhibitor mix, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM phenylmethylsulfonyl fluoride

(PMSF) and a spatula tip RNase. Cells were lysed by the high-pressure homogenizer HPL6 (Maximator, 2 bar). The lysis was followed by a low-speed centrifugation at 15.000 x *g* for 30 min, 4 °C to remove cell debris. A high-speed centrifugation step (100.000 x *g*, 1 h, 4 °C) with the supernatant of the low-speed centrifugation step was performed to harvest the membrane. The obtained membrane was resuspended in 50 mM tris pH 8.8 (adjusted at 4 °C), 300 mM NaCl, 2 mM DTT buffer with a wet pellet concentration of 90 mg/ml. *E. coli* membrane can be used immediately or frozen with 10% glycerol and stored at -20 °C.

3.3.3. SpaK full-length protein solubilization with styrene maleic acid (SMA) polymers

E. coli membranes (90 mg/ml) were mixed with styrene maleic acid (SMA) polymers (here Xiran 30010/SMALP 200) to obtain a final polymer concentration of 2.5% w/v. The mixture was incubated for 1 h at 4 °C with overhead rotation. Insoluble membranes were removed by ultracentrifugation (125000 x *g*, 45 min, 4°C). The supernatant was then mixed with pre-equilibrated nickel nitrilotriacetic acid (NiNTA) agarose beads (1 ml NiNTA/5 ml membrane polymer mixture) and incubated for 14 – 16 h with overhead rotation at 4 °C. To avoid unselective binding to the NiNTA beads, 10 mM imidazole were added.

3.3.4. Purification of SpaK full-length protein in styrene maleic acid lipid particles (SMALPs)

The mixture from the previous step with SpaK proteins in styrene maleic acid lipid particles (SMALPs) were transferred to a gravity flow column. The column was washed with 10 column volume (CV) 50 mM tris pH 8.8 (adjusted at 4 °C), 300 mM NaCl, 15 mM imidazole, 10 CV with 50 mM tris pH 8.8 (adjusted at 4 °C), 300 mM NaCl, 30 mM imidazole, 10 CV with 50 mM tris pH 8.8 (adjusted at 4 °C), 300 mM NaCl, 45 mM imidazole and finally elute with 50 mM tris pH 8.8 (adjusted at 4 °C), 300 mM NaCl, 250 mM imidazole. Fractions containing SpaK SMALPs were concentrated with a centrifugation concentrator and loaded to a size exclusion column (Superdex 200 Highload 16/60). Protein concentration was determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using a 12% gel and a bovine serum albumin standard.

3.3.5. Transmission electron microscopic analyses of SpaK in SMALP nanodiscs

Transmission electron microscopic analyses were performed in collaboration with the _____ lab (_____). 3 µl of Spak SMALPs (0.2 µM or 29,5 ug/ml) were loaded on to a pre-glow discharged quantifoil™ R 1.2/1.3 on 400 copper mesh grid (Jena Bioscience) and incubate for 30 second before blotted with a whatman filter paper. The grids were washed with 5 µl H_2O and stained with

3 µl 2% uranyl acetate. After removing the excess of uranyl acetate, the grids were dried at room temperature for 2 – 5 min. Images were recorded at the FEI Tecnai T12 (120 kV) electron microscope using the software SerialEM.

For the subtilin sample, the SpaK SMALP protein sample was supplemented with 20 µM subtilin and then loaded to the copper support grid.

Following settings have been use for negative stain electron microscope micrographs:

TiltAngle = -0.000488	Magnification = 52000
Intensity = 0.460162	ExposureDose = 0
PixelSpacing = 4.402	SpotSize = 4 - 5
ExposureTime = 2	Binning = 2
CameraIndex = 0	DividedBy2 = 1
MagIndex = 30	TargetDefocus = -1

3.3.6. Cryo electron microscopic analyses of SpaK in SMALP nanodiscs

For cryo electron microscopic (CryoEM) analyses of SpaK, 3.5 µl SpaK in SMALPs (250 µg/ml) were loaded to a pre-glow discharged Cu 300 copper grid with carbon support films (Jena Bioscience) and plunge freeze with liquid ethane (~ -163 °C). The Plunge freezer FEI Vitrobot Mark IV was used with following settings:

Humidity: 100%	Temperature: 4 °C
Blot time: 7 s	Blot force: 25
Wait time: 0 s	

Frozen grids were stored in a grid box in liquid nitrogen until further use.

Low resolution micrographs were recorded at the FEI Tecnai T12 (120 kV) electron microscope using the software SerialEM and following settings:

TiltAngle = -0.000488	Magnification = 52000
Intensity = 0.451822	ExposureDose = 0
PixelSpacing = 4.402	SpotSize = 4
ExposureTime = 2	Binning = 2
CameraIndex = 0	DividedBy2 = 1
MagIndex = 30	TargetDefocus = -2 - -3

For the subtilin sample, the SpaK SMALP protein was supplemented with 20 µM subtilin and then loaded to the copper support grid.

Data collections were performed at the CryoEM facility of the Rudolf-Virchow institute, Würzburg. FEI Titan Krios electron microscope (300 kV) was used, and the data collections were run in counting-mode for 2 days.

3.3.7. Expression of NisK sensor domain from *Lactococcus lactis*

Similar to the SpaK transformation, plasmid coding for the NisK sensor domain was transformed in *E. coli* MC1061 strains. A colony was picked to inoculate a preculture and that was used to inoculated terrific broth (TB) medium supplemented with 2 mM $MgSO_4$ and 100 µg/ml ampicillin. The culture was incubated at 37 °C until an OD_{600} = 0.8 – 1.0 was reached. Another 5 h at 37 °C was followed. After that, the cells were harvested by centrifugation (10 min, 4 °C, 10.000 x *g*). Cell pellets were stored at -20 °C until further use.

3.3.8. Purification of NisK sensor domain

Harvested cells from 1 l culture volume were resuspended in 50 ml lysis buffer (50 mM tris pH 7.5 (adjusted at 4 °C), 300 mM NaCl), supplemented with 1 mM DNase I, 1:1000 dilution of a protease inhibitor mix, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM phenylmethylsulfonyl fluoride (PMSF) and a spatula tip RNase. Cells were lysed by sonication. A low-speed centrifugation step (10.000 x *g*, 4 °C, 30 min) was performed to remove cell debris. The pellet was resuspended in 50 mM tris pH 7.5 (adjusted at 4 °C), 300 mM NaCl supplemented with 8 M urea. The mixture was incubated for 1 h at 4 °C with continuous shaking. A centrifugation step at 10.000 x *g*, 4°C for 30 min removed insoluble parts and the supernatant was mixed with equilibrated NiNTA resin for at least 1 h. The protein resin mixture was transferred to a gravity column and washed with 5 CV 50 mM tris pH 7.5 (adjusted at 4 °C), 300 mM NaCl, 10 mM imidazole, 4 M urea. The washing step was followed by 5 CV 50 mM tris pH 7.5 (adjusted at 4 °C), 300 mM NaCl, 20 mM imidazole, 2 M urea, 5 CV 50 mM tris pH 7.5 (adjusted at 4 °C), 300 mM NaCl 30 mM imidazole, 1 M urea, 5 CV 50 mM tris pH 7.5 (adjusted at 4 °C), 300 mM NaCl, 40 mM imidazole, 0.5 M urea and finally 5 CV 50 mM tris pH 7.5 (adjusted at 4 °C), 300 mM NaCl, 50 mM imidazole. Protein was eluted with 5x1 CV 50 mM tris pH 7.5 (adjusted at 4 °C), 300 mM NaCl, 250 mM imidazole. Fraction containing NisK SD were collected and loaded to the size exclusion column Superdex 75 Highload 16/60.

3.3.9. Circular dichroism spectroscopy

To investigate the secondary structure of the NisK SD, circular dichroism spectroscopy was performed. NisK SD was dialyzed against 50 mM tris pH 7.5 or pH 8.0 (adjusted at 4 °C) and 300 mM NaCl. The sample was diluted with H_2O to yield a protein concentration ≥10 µM in 5 mM tris and 30 mM NaCl. The CD spectrum was measured from 190 – 260 nm, with 1.0 nm data pitch, a

scanning speed of 50 nm/min, bandwidth = 5.0 nm, three accumulation and in continuous scanning mode. The Beta Structure Selection (BeStSel) analysis was used to determine the secondary structure propensities (*285*).

3.3.10. NisK sensor domain stability test

The protein stability of NisK SD was tested at two temperatures (room temperature and 4 °C), in two buffers (pH 7.5 and pH 8.0) over a time span of 8 days. Therefore 3 samples of each 1 ml have been prepared for each temperature and buffer. In addition, one sample for each buffer condition were frozen and thawed. After 1, 2 and 8 days, the samples at room temperature and 4 °C were centrifugated at 20.000 x *g*, 4 °C for 15 minutes. SDS-PAGE samples were prepared from the supernatant and pellet fractions after centrifugation. All SDS-PAGE samples were loaded to 15% SDS-PAGE gels with the protein marker Precision Plus Protein Dual Color Standards (BioRad). The gels were stained with Coomassie, and band intensities were analyzed by the software image j.

3.4. Results and discussion

3.4.1. Structural studies on the *Bacillus subtilis* sensor histidine kinase SpaK

Bacteria can respond to external stimuli via a sophisticated signal transduction mechanism that depends on two-component systems (TCSs). The sensor histidine kinase (SHK) SpaK from *Bacillus subtilis* (*B. subtilis*) is part of the TCS that senses the lantibiotic subtilin in a quorum-sensing mechanism (*286*). When subtilin binds to SpaK, the histidine kinase autophosphorylates. Upon transfer of the phosphate group to the respective response regulator (RR) SpaR allows this protein to bind to the promoter and induce the transcription of the subtilin gene cluster (*286*). Of note, lantibiotic sensing seems to be highly specific, i.e., lantibiotics with high structural similarity to subtilin such as nisin cannot activate the SHK SpaK (*284, 283*). Elucidating the molecular details of signal transduction between the sensor and output domains is incredibly difficult given the size, flexibility, and dynamics of histidine kinases.

To date, structural insights into full-length SHK is limited to soluble SHKs (blue light photoreceptor LOV histidine kinase from *Brucella abortus (150)* and phytochrome from *Deinoccocus radiodurans (287)*). Full-length membrane protein purification in general is challenging and requires extensive screening of different purification methods. To obtain functional full-length SpaK protein, multiple membrane mimicking systems were tested by _____ (_____Labs, _____). Only the native nanodiscs purification method using styrene maleic acid (SMA) polymers resulted in functional SpaK proteins. Using laser induced liquid bead ion desorption (LILBID) mass spectrometry, it was shown that SpaK in styrene maleic acid lipid particles (SMALPs) is homodimeric and can bind to the lantibiotic subtilin. Further, autophosphorylation of SpaK in SMALPs and phosphotransfer to SpaR have been shown by _____ using autoradiography and [γ^{32}]-ATP.

Here, an AlphaFold2 (AF2) model of homodimeric SpaK was generated to provide a first structural view of the protein (*77, 78*) (Figure 53A). The global alignment of five predicted models yields RMSD values between 0.522 – 0.823 Å (RMSD of pruned atom pairs, without loops, RMSD of all pairs = 4.099 – 6,770 Å, when comparing two models). Superposition of the predicted models shows that the largest structural deviations are observed in the C-terminal catalytic ATP binding (CA) domain and the relative position of the sensor domain (SD) (Figure 53A). The AF2 models show that the SpaK (SD) domain has an α,β mixed fold consisting of five β-sheets and four α-helices (Figure 53B-D). This fold has been reported to be the dominant sensing module among SHK from TCS in *B. subtilis (288).* Typically, it contains four to five antiparallel β-sheets flanked by an undefined number of α-helices (*240*).

In addition to the structural model of SpaK by AF2, the domain boundaries of the protein's transmembrane helices were predicted by the online tool DeepTMHMM (*289*). This suggested that two helices span residues M19 – I36, and S156 – A176 (Figure 53B, highlighted in yellow). Between the TMD and the SD, SpaK has a short helix formed by residues ^{143}PLLRKY^{148} (Figure 53B, colored white) that belongs neither to the predicted TMD nor the α,β-core fold of the SD. A similar

structural feature was found in the SHK PhoQ from *E. coli* (*290*). PhoQ senses polycationic molecules such as antimicrobial peptides and here, this short α-helix consists of several acidic residues which were proposed to sense divalent cations and adenosine monophosphate (AMP) at the membrane surface (*290, 262*). In SpaK, the role of this additional α-helical elbow is still unclear, but subtilin is a hydrophobic molecule suggesting a possible involvement of this helix in SpaK for ligand binding. However, with its two positively charges residues, R147 and K147 and its proximity to the membrane surface, this helix could also bind negatively charged lipids that may regulate kinase activity.These hypotheses need to be tested in future experiments. Interestingly, other than for PhoQ, this helix has not been described for SHKs. In the homologous SHK NisK from *L. lactis* this α-helical elbow is missing.

SpaK's signal relay domain between the TMD and cytoplasmic kinase domain consists of a four-helix bundle (Figure 53B, colored grey). This fold fits to the described secondary structure found in HAMP domains (a domain found in histidine kinases, adenylyl cyclases, methyl binding proteins, phosphatases) (*261*). The helices of the HAMP domain merge into the four-helix bundle of the dimerization and histidine phosphorylation (DHp) domain. The boundary between HAMP and DHp domains is not clearly defined. The histidine residue important for autophosphorylation, H247, sits in the DHp domain (Figure 53B and E). The DHp domain is flanked by two catalytical ATP (CA) binding domains (Figure 53B and E) that show the typical Bergerat fold formed by a five stranded β-sheet opposing a layer of three α-helices (*237, 265*).

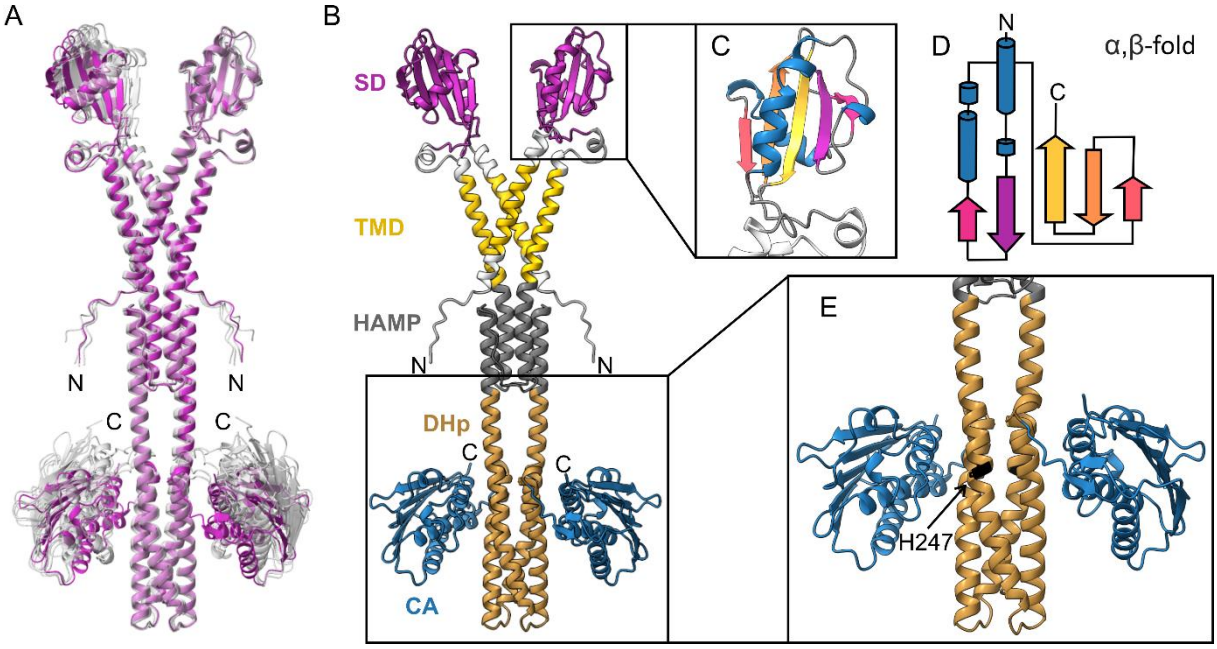


Figure 53: AlphaFold2 (AF2) models of the sensor histidine kinase SpaK from Bacillus subtilis*. A: Superposition of five AF2 models. The whole protein was aligned. B: SpaK functional domains are highlighted on representative AF2 model. SD: sensor domain, puple, TMD: transmembrane domain, yellow, HAMP (histidine kinases, adenylyl cyclases, methyl binding proteins, phosphatases), dark grey, DHp (dimerization and histidine phosphorylation) domain, beige, CA (catalytic ATP binding) domain, blue. The boundaries of the TMD regions were predicted by the online tool DeepTMHMM (289) . α-helices that were not predicted to be membrane-spanning by DeepTMHMM are colored white. C: Close-up view of the SD. α-helices, β-sheets are colored from purple to yellow from N- to C-terminus, loops are colored grey. D: Topology model of the SD shows a canonical α,β-fold, also known as PDC (acronym of three founding members, PhoQ, DcuS and CitA), Per/Arnt/Sim (PAS) fold or Cache fold (240). E: close-up view of the DHp and CA domains. H247 marks the position for autophosphorylation.*

To obtain a detailed structural view of the SHK in the absence and presence of subtilin, full-length SpaK was heterologously expressed in *E. coli,* purified into SMALPs and analyzed by electron microscope (EM). Based on a purification protocol originally developed by ____ (____ labs, ____), *E. coli* membranes with C-terminally His_{10}-tagged SpaK were solubilized by the commercially available Xiran SL 30010 polymer. The solubilized protein was then purified by nickel affinity column and size exclusion chromatography (SEC) (Figure 54A and B. The elution profile of the SEC indicates some protein aggregation (~45 ml) as well as dimeric protein (~58 ml) regarding to the column manual.

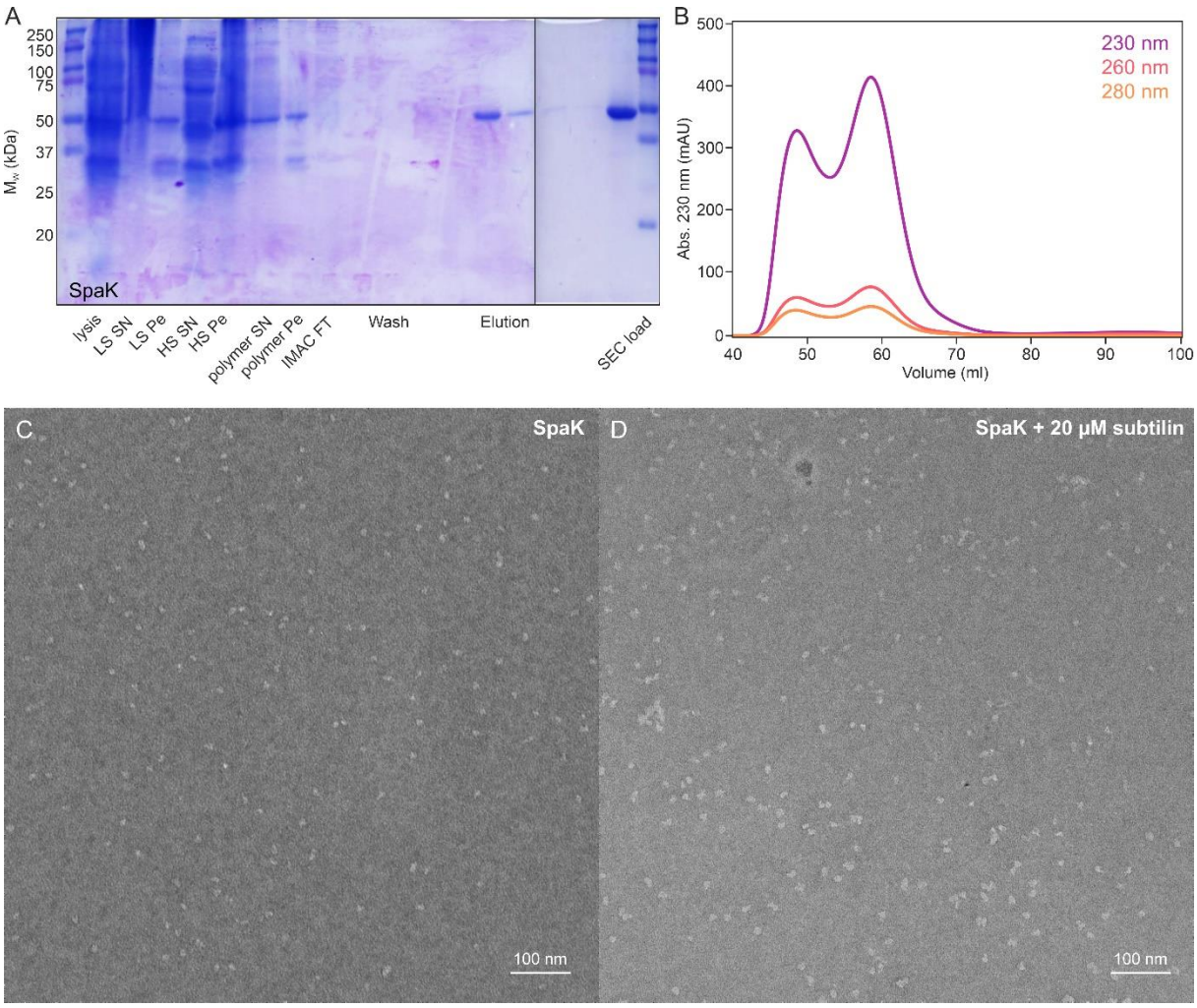


Figure 54: Purification and electron microscopic analysis of SpaK in styrene maleic acid lipid particles (SMALPs). A: SDS-PAGE of the SpaK purification into SMALPs nanodiscs. N-terminal His_{10}-tagged SpaK overexpressing E. coli were lysed with high pressure (lysis). Cell debris was removed by a low speed (LS) centrifugation at 20.000 x g, 4 °C for 30 min. The membrane was harvested via a high speed (HS) centrifugation step using the supernatant from the LS centrifugation. The pellet fraction that contained E. coli membranes was solubilized with Xiran SL 30010 polymer and centrifuged again. The supernatant contains the successfully solubilized membrane proteins while the insoluble polymers and protein aggregates remained in the pellet fraction. Solubilized protein was load onto an affinity column, washed with increasing imidazole concentrations, and eluted with a buffer containing 300 mM imidazole. The fraction containing SpaK SMALPs were concentrated (sample SEC load) and load onto a size exclusion chromatography (SEC) column (Superdex 200 16/600). B: size exclusion chromatogram of SpaK in SMALPs. C: Electron microscope images of negatively stained (uranyl acetate) SpaK SMALPs at a concentration of 29.5 µg/ml on quantifoil™ R 1.2/1.3 400 copper mesh grid. D: negative stain EM of SpaK (29.5 µg/ml, 0.163 µM)) with 20 µM subtilin in 20 mM Tris-HCl, pH 8.5, 150 mM NaCl buffer.

For electron microscopic investigations, the purified, SMALPs containing SpaK were then loaded onto copper grids and stained with uranyl acetate. Negative stained SpaK nanodiscs show good

particle distribution on the grid with particle diameters of ~10 nm (Figure 54C). Particle size and distribution remain unchanged upon addition of 20 µM subtilin (Figure 54D).

After having established a satisfactory sample quality by negative stain EM, both samples were loaded on copper grids and frozen in liquid ethane for analysis by single-particle cryo-EM. Due to difficulties with automatic particle picking in both samples, the total number of particles was not yet sufficient to obtain a high-resolution structure of SpaK. The resolution of the current CryoEM structure stays above 11 Å. The 2D classification of the picked particles showed a preferential orientation (top and bottom view) of the protein on the grid. A side view of the protein was not detected. The missing side views were clearly represented in the 3D classification of the CryoEM micrographs (Figure 55). For a comparison with the AF2 model, it was mapped into the electron density volume obtained by 3D classes (Figure 55). When comparing the CryoEM electron density maps with the AF2 model there are striking differences in the distance between sensor domain and the tip of the DHp domain.In the CryoEM electron density maps, the distance from SD to DHp domain is slightly shorter (151.3 Å) than in the AF2 model (161.4 Å) (Figure 55A). In contrast, additional electron densities were observed in the cytoplasmatic regions of SpaK, in the HAMP domain and around the DHp domain (Figure 55A).

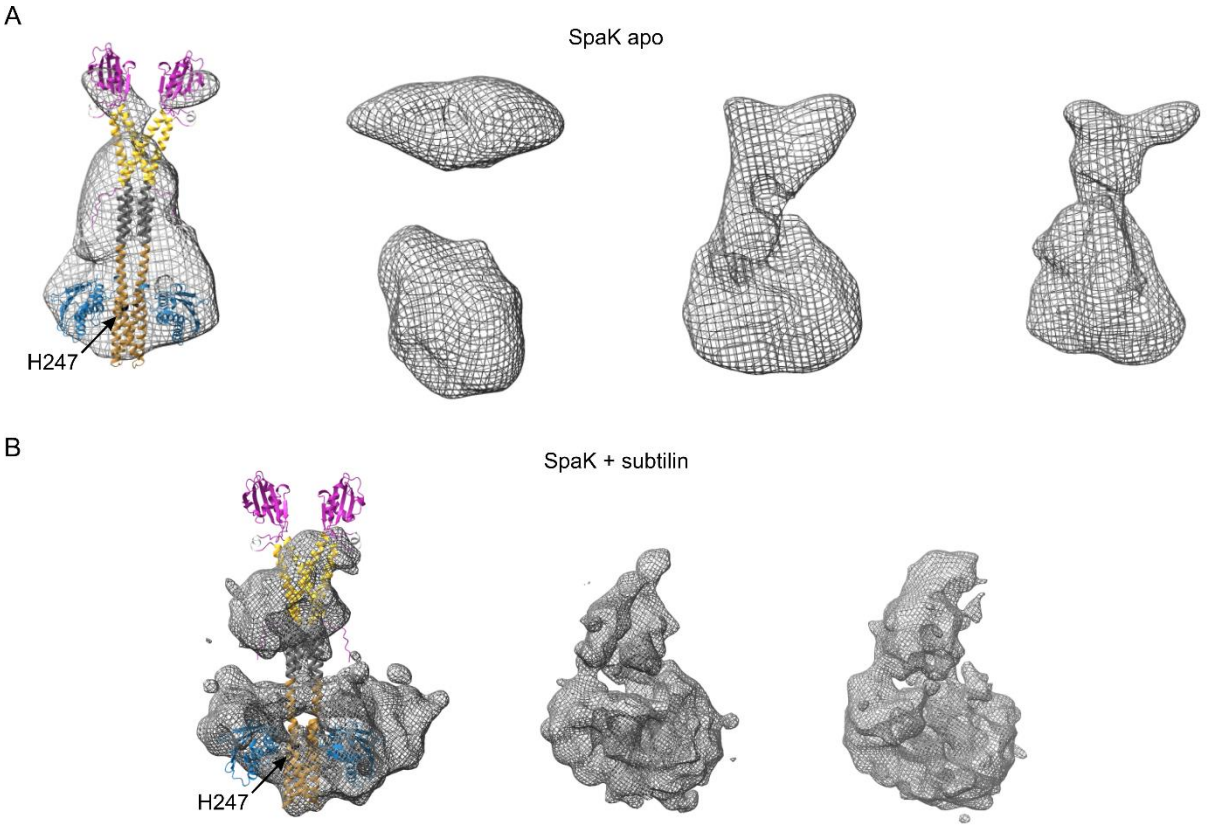


Figure 55: Electron density maps obtained by 3D classification of the CryoEM micrographs from SpaK in SMALP nanodiscs. A: 3D classes were extracted from the micrographs recorded from SpaK in SMALPs without addition of substrate and with the addition of subtilin (B). The AF2 model of SpaK was mapped into the density map for orientation. SpaK domains shown in colors (color code as in Fig. 7A), H247 is marked black.

In both SpaK samples (apo state and with subtilin), the transmembrane regions with the surrounding nanodiscs are missing (Figure 55). Rather than stabilizing the SHK, the addition of the substrate subtilin resulted in the "disappearance" of the sensor domain in addition to the

already poorly resolved TMD (Figure 55B). The resolution in the subtilin cryoEM electron density maps are worse (18 Å) compared to the apo state. Additional electron densities around the DHp domain, especially around the region of the phosphorylable histidine were observed (Figure 55B). It is possible that a conformation is shown by the Figure 55B, in which the CA domain binds to the DHp domain at the position of the conserved histidine. This allows the hypothesis that CA domain undergoes conformational changes upon subtilin binding.

In summary, based on the available cryo-Em density maps, the elongated SpaK model predicted by AF2 might not capture the confirmation of the protein in solution. It seems likely that the helical regions between TMD and DHp domain are slightly bent in the apo SpaK cryo-EM density maps, which is not represented by the AF2 model. This is supported by molecular dynamics simulations (____, ____ lab, ____, data not shown). Interestingly, helix bending has also been observed in literature for the SHKs CpxA and CusS from *E. coli* (*273, 291*). In the crystal structure of the isolated cytoplasmic regions (including the HAMP domain) of these SHKs, the authors saw two kink positions along the helix formed by the HAMP and DHp domains (Figure 56). The first kink is in a polar residue patch and has been proposed to avoid steric clash (Figure 56B). The second kink is caused by a conserved proline residue in the proximity of the phosphorylatable histidine (*273, 291*). Both kink sites are possibly also present in SpaK. The polar residue patch could be found in the ^{235}EKNKKEQ^{241} in SpaK and the proline marking the second kink could be induced by P253 (Figure 56B).

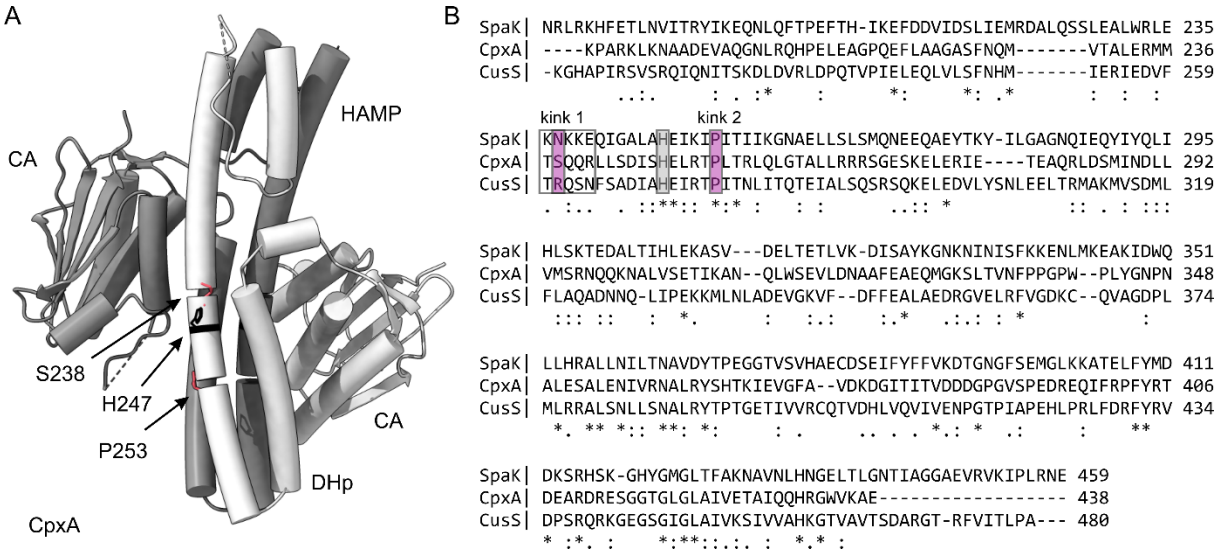


Figure 56: Structural model and sequence alignment illustrating possible kink position in the cytoplasmic domain. A: Structure of E. coli *CpxA cytoplasmic domain (PDB: 4BIU). Kink positions (S238, kink 1, and P253, kink 2) are shown in sticks. H247 is colored black. HAMP (histidine kinases, adenylyl cyclases, methyl binding proteins, phosphatases), DHp (dimerization and histidine phosphorylation) domain, CA (catalytic ATP binding) domain. B: Sequence alignment of the entire cytoplasmic domain of* B. subtilis *SpaK (residues 177 – 459) with* E. coli *CusS (residues 208 - 480) and* E. coli *CpxA (residues 188 - 455). Kink positions are marked with purple boxes. The proposed polar residue patch is mark with a grey frame. The conserved histidine is mark with a grey box.*

Another important difference between the cryo-EM density maps and the AF2 prediction pertains the relative position of the two sensor domains. They seem to be connected to the TMD by a flexible linker indicating a structural flexibility that is seen in the density maps, but not the static AF2 model. The two SDs might move closer to each other or closer to the membrane depending on protein conformation or whether substrate is present. A similar conformational flexibility

might be observed for the CA domains. In the literature, an allosteric switching mechanism is proposed, where the DHp domain moves back and forth in bending motions (*277*) (Figure 52). This mechanism allows the two CA domains switch between an active and inactive state where they bind and release H247 in the Dhp domain asymmetrically. In the unbound/inactive state of the CA domain, the response regulator protein can then bind to the phosphorylated histidine at the DHp domain. The switching mechanism might explain the additional electron densities around the DHp domain obtained by CryoEM analysis, which would then represent CA domains in various conformations. However, these hypotheses need to be verified by improving the resolution of the cryoEM structures and with orthogonal biophysical experiments that look into SpaK structural dynamics.

3.4.2. Structural investigation of the sensor domain of the *Lactococcus lactis* NisK histidine kinase

To understand substrate binding and lanthipeptide selectivity of SpaK, a structure of the SpaK SD is needed that could show molecular details of its interaction with subtilin. However, as seen above, addition of subtilin did not improve the resolution of the SpaK CryoEM electron density map (Figure 55B). Additionally, expression and purification trials of the isolated SpaK SD were challenging due to problems with protein solubility. Our collaboration partners from _____ labs (_____) hypothesized that the homologous SD from the *Lactococcus lactis* SHK NisK shared similar secondary structure arrangements with SpaK (*283*). Similar to SpaK SD, NisK SD has been proposed to have an α,β-fold. NisK and SpaK sensor domains share ~50% sequence homology and ~20% sequence identity. The multiple sequence alignment is shown in Figure 57A. The AF2 prediction of NisK SD shows an α,β-fold with five antiparallel β-sheets flanked by two α-helices (Figure 57B and C). A comparison of NisK and SpaK SDs shows similar secondary structure elements arrangement in both proteins (Figure 57C-E). Significant differences were observed at the N-terminus of the proteins and the total number of a-helices (Figure 57C and D). In NisK, the N-terminus starts with an unstructured loop followed by an α-helix. In contrast, SpaK SD contains two N-terminal helices. In total SpaK has four α-helices. Structural overlay of NisK and SpaK SD shows high overlaps (Figure 57E) with a sequence alignment score = 100.6 and RMSD = 1.075 Å (of pruned pairs without loops, RMSD of all pairs = 10.781 Å). The comparison of the AF2 models of NisK and SpaK SDs indicates structural homology.

Guided by the AF2 model, a C-terminal His_{10}-tagged NisK SD construct containing the residues S43-T134 was cloned, heterologously expressed and purified in small amounts by _____ (_____ groups, _____). However, a major part of the expressed protein remained in the insoluble pellet fraction/inclusion bodies. To increase the protein yield, a refolding protocol was established, to purify the protein from inclusion bodies.

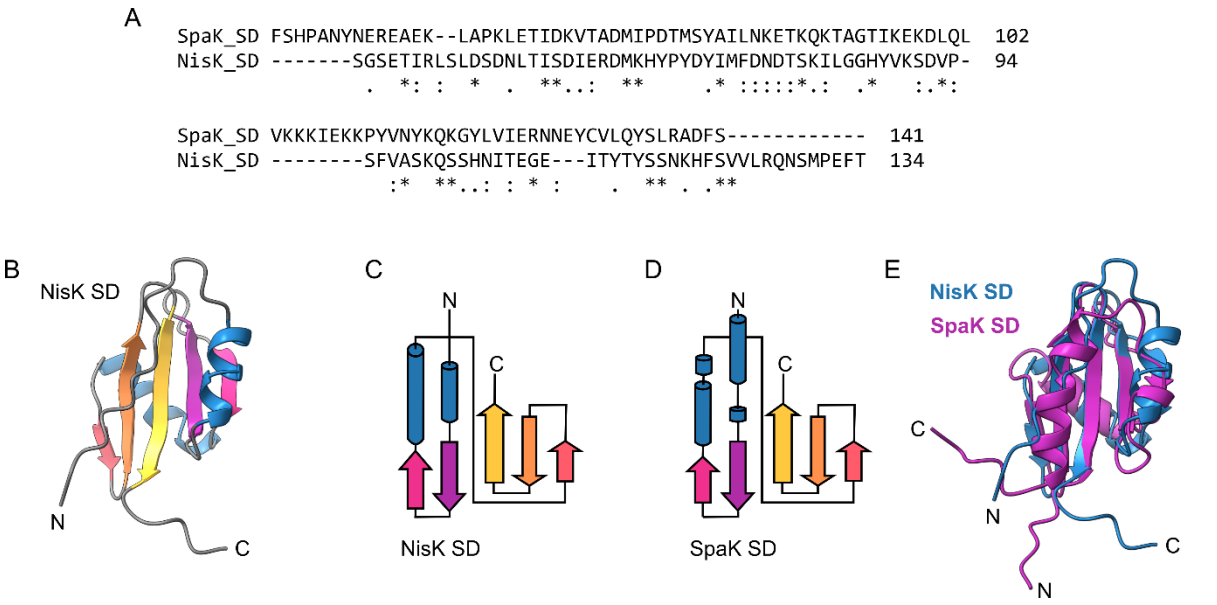


Figure 57: Structural comparison of the B. subtilis *SpaK sensor domain with the* L. lactis *NisK sensor domain (SD). A: Sequence alignment of SpaK (F45 – S141) and NisK (S43 – T134) SDs. B: AlphaFold2 model of NisK SD. A-helices are colored blue, linker/loops are colored grey, b-sheets are colored from purple to yellow from N- to C-terminus. C: Topology model of the NisK SD shows a canonical α,β-fold, also known as PDC (acronym of three founding members, PhoQ, DcuS and CitA), Per/Arnt/Sim (PAS) fold or Cache fold (240). D: Topology model of the SpaK SD. E: Superposition of NisK and SpaK SDs AF2 models.*

First, proteins in the pellet fraction after cell lysis were fully denatured with 8 M urea and then loaded onto a NiNTA column. The protein was then refolded by stepwise decreasing the urea concentration during the washing steps. The refolded protein was eluted from the column with an imidazole containing buffer and analyzed by SEC and circular dichroism (CD) spectroscopy (Fig. 9A-C). The SDS-PAGE of the NisK SD shows a relatively pure, albeit diffuse protein band running at roughly the expected molecular weight. The SEC elution profile shows protein aggregation at the elution volume ~40 ml as well as monomeric globular protein regarding to the column manual. The CD spectrum indicates a folded protein consisting of both α-helices and β-sheets (Figure 58C). The Beta Structure Selection (BeStSel (*285*)) analysis of the data shows 7% α-helix, 34% antiparallel and 11% parallel β-sheets, and 48% loops and turns (Figure 58D). These secondary structure propensities were compared with the AF2 model of NisK SD. The superposition of five AF2 models is shown in Figure 58D. The AF2 model contains 25% α-helix, 32% β-sheets, and 44% loops and turns (Figure 58D and E).

NisK SD could be expressed and purified in high amounts (~25 mg from 1 l culture). Protein stability tests were performed to monitor protein recipitation at room temperature and 4 °C, over a time span of 8 days. In addition, two different pH values (pH 7.5 and pH 8.0) were tested. At room temperature and pH = 7.5, the protein precipitated within a couple of minutes (data not shown) but showed improved stability at 4 °C, where 50% of NisK SD was found to still be soluble after 8 days (Figure 58F and G). Freezing the protein in liquid nitrogen seems to be the best option for long time storage. Here, ~60% of the sample was still soluble after thawing. Protein stability in the first two days could be slightly improved by using a buffer with pH = 8.0 and a 4 °C storage temperature (Figure 58F and G). However, a slightly basic pH did not affect long time stability, when comparing the samples after 8 days at either pH 7.5 or pH 8.0 (Figure 58F and G). Unfortunately, protein stability could not be improved further by the addition of glycerol or salt

concentration (data not shown) and protein precipitation was enhanced by increasing the protein concentration using centrifugation concentrators. Further biophysical and structural approaches should thus always be performed with freshly purified NisK SD proteins stored at 4 °C, unless more favorable storage conditions can be identified.

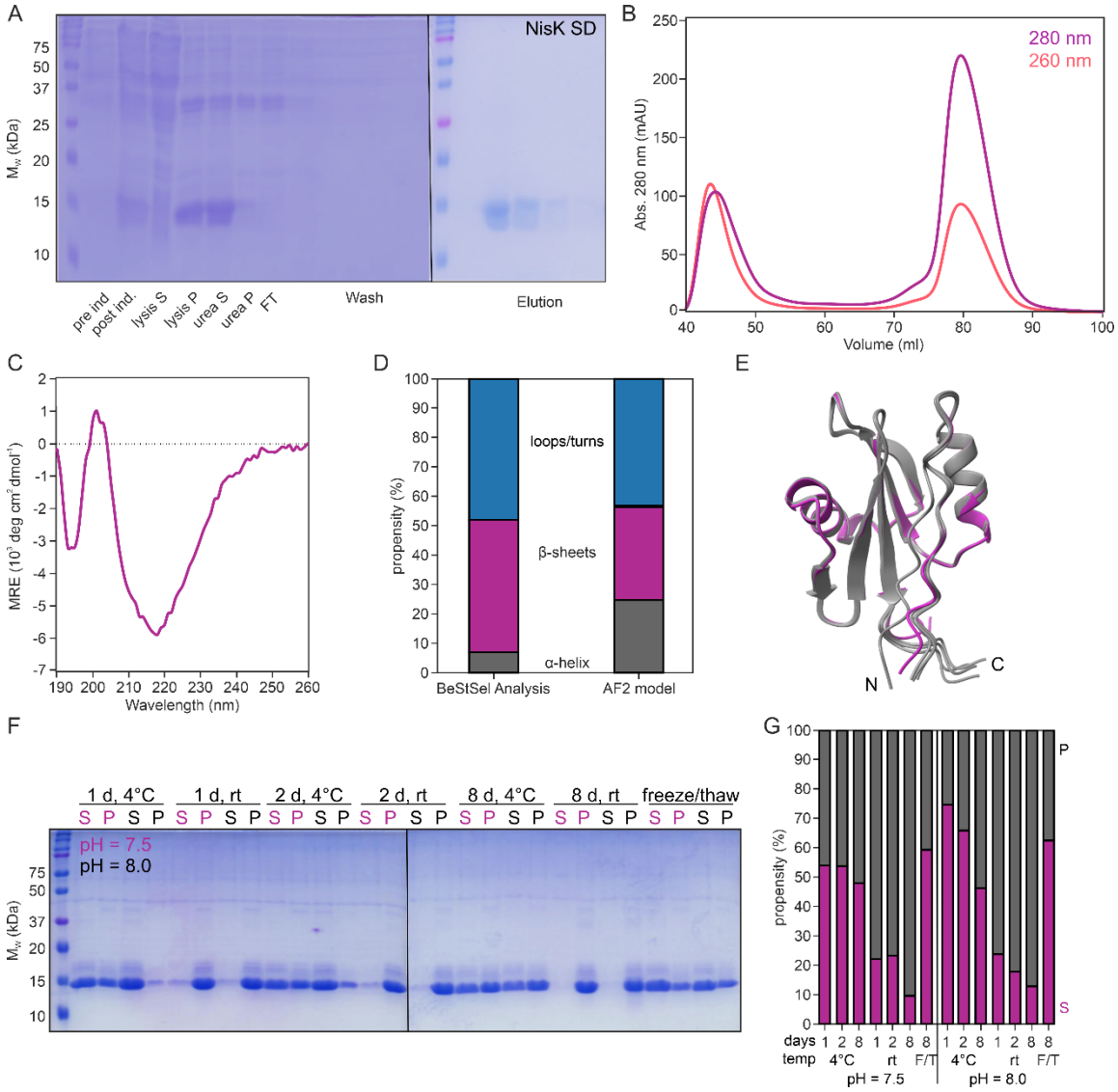


Figure 58: Purification, secondary structure determination and protein stability test of the NisK sensor domain (SD) from Lactococcus lactis*. A: SDS-PAGE of the NisK SD purification via refolding. Samples were taken from the pre induced (pre ind.) liquid culture, before harvesting of the culture (post induction, post ind.), after the centrifugation step after cell lysis (lysis supernatant (lysis S) and pellet (lysis P)), after the solubilization of the lysis pellet with 8 M urea (supernatant (urea S), and pellet (urea P)), of the frow-through of the nickel affinity column (FT), from the washing steps (Wash) with decreased urea concentration and increased imidazole concentration, and from the elution fractions eluted with imidazole containing buffers (Elution). B: Size exclusion chromatogram of the purified NisK SD run on a Superdex 75 high load 16/60 column. C: Circular dichroism spectrum of the refolded NisK SD. D: secondary structure content according to the CD spectrum (analyzed with the BeStSel online tool) and calculated secondary structure content in the AF2 models. E: Superposition of five AlphaFold2 models of the NisK SD. F: SDS-PAGE of the stability test of NisK SD in two different buffers (pH = 7.5 and pH = 8.0) under different storage conditions (RT or 4°C for 1-8 days). Precipitated proteins were removed by centrifugation prior to gel loading. A reduction in band intensity of the soluble fraction (S) and an increase in the pellet (P) should thus reflect a decrease in protein stability under the tested condition. Protein samples after freeze-thaw were collected after shock freezing with liquid nitrogen, stored for 8 days at -80 °C and thawed on ice. G: image j band intensity analysis of the SDS-PAGE bands at 12.8 kDa that correspond to NisK SD. Band intensity of each lane in the SDS-PAGE was normalized to the total band intensities in the samples S + P for the same condition (time, temperature, and pH).*

3.5. Conclusion

Two component systems (TCSs) sense external stimuli and transmit the signal into the cell, thereby playing the role of receiver in microbial communication. In response to the received information, the microbial gene expression can be regulated. TCSs consist of a membrane bound sensor histidine kinase (SHK) and a response regulator (RR) protein. Substrate binding in the exterior to the SHK triggers autophosphorylation in the cytosolic kinase domain. The phosphotransfer from the SHK kinase domain to the soluble response regulator (RR) results in conformational changes of the RR and allows its interaction with DNA. Molecular details of substrate recognition, signal transmission through the membrane and signal transfer in the cytoplasm are poorly understood. One of the major challenges for biophysical studies on TCS is the purification of the membrane protein SHK.

In this part of the thesis, the full-length SHK SpaK from *B. subtilis* was purified and used for structural studies. SpaK binds the lantibiotic subtilin, which is a ribosomally synthesized and post translationally modified peptide (RiPP). Substrate selectivity is encoded by the sensor domain (SD). Unfortunately, the isolated SpaK SD faces solubility problems and could not be purified. Instead, the homologous SD from the nisin SHK NisK from *L. lactis* was purified and structurally investigated.

3.5.1. SpaK in styrene maleic acid lipid particles undergoes structural changes with subtilin

The homodimeric histidine kinase SpaK is involved in intraspecies communication of *B. subtilis (284, 286)*. and could could be expressed both homologously and heterologously. It can be purified in a membrane-bound state using styrene maleic acid polymers. Functional analyses performed by _____ (_____, _____) first confirmed the interaction of the substrate subtilin with SMA-purified SpaK and its autophosphorylation activity. The structural examination of SpaK SMALP using cryo-electron microscopy (CryoEM) was difficult due to the preferential orientation of the particles on the support grid and the protein's apparent structural flexibility. This resulted in a low resolved electron density map (11 Å) of SpaK in SMALPs. Overlaying the AlphaFold2 (AF2) model of SpaK with the electron density map obtained from CryoEM, showed that electron densities in the transmembrane domain (TMD) were partly missed (Figure 55) and additional electron densities were found between the dimerization and histidine phosphorylation (DHp) domain and the four-helix bundle of the HAMP (histidine kinases, adenylyl cyclases, methyl binding proteins, phosphatases) domain (Figure 55). The distance between SD and the kinase domain was slightly shorter in the electron density maps compared to the AF2 prediction. Based on these observations the question arises whether the elongated helix core shown by the AF2 model is a good representation of SpaK. It is likely that there might be kinks along the helix from the TMD to the DHp, that allow helix bending. In the literature, it has been proposed that segmental helical

motions can promote histidine kinase autophosphorylation reaction (*273, 291*). The described kink position between the HAMP and kinase domains (*273, 291*) was also identified in the SpaK sequence. Thus, I propose a model for spaK including segmental helical motions and domain movement, that probably reflects its structure better than the AF2 prediction (Figure 59).

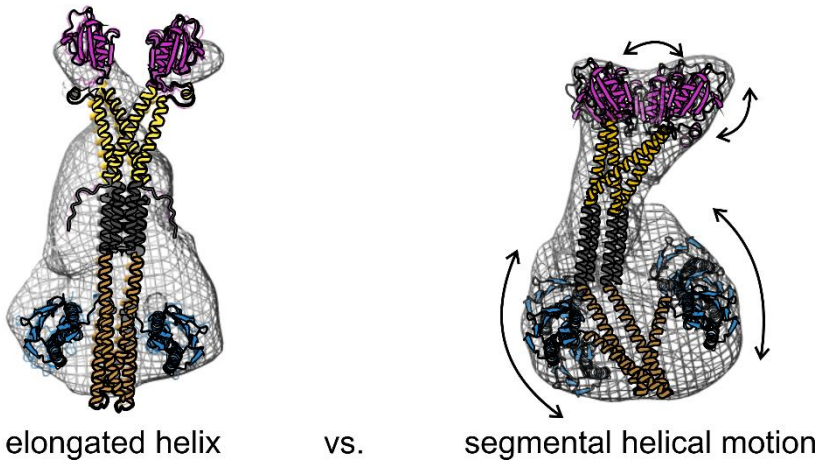


Figure 59: Structural models of apo and subtilin-bound SpaK based on the electron density maps obtained by cryo electron microscope (cryoEM). AlphaFold2 model (left) vs a schematic model demonstrating segmental helical motions (right) fits to the cryoEM electron density maps.

The addition of the lanthipeptide subtilin resulted in a decreased resolution of the cryoEM electron density maps. Subtilin binding resulted in the "disappearance" of the sensor domain in addition to the already poorly resolved TMD (Figure 55B). Additional electron densities were found around the DHp domain, which indicates conformational changes of the CA domain induced by substrate binding.

In summary, this study provides the first structural insights into the full-length histidine kinase SpaK. Although the resolution of the CryoEM density maps were poor, the electron density maps still gave indication for possible segmental helical motions and structural flexibility of SD and catalytical domains. This in total is a good starting point for further structural investigation of SpaK full-length protein.

3.5.2. SpaK and NisK sensor domains share an α,β-fold

When difficulties arise with determining the full-length protein structure, the "divide and conquer" approach can be helpful in obtaining structural information of isolated domains (*112–114*). Accordingly, the structure of the SpaK SD and the homologous NisK SD were addressed. The AF2 models indicated a mixed α,β-fold for SpaK SD and NisK SD. Since NisK SD showed improved solubility compared to SpaK SD, the purification of an isolated SD focused on this protein. However, even for NisK SD, the majority of the heterologously expressed protein was found in inclusion bodies. A refolding protocol was optimized for NisK SD that allowed NisK SD purification with high yields (25 mg from 1 l liquid culture). Analyzing the secondary structure of NisK SD showed slightly differences in the secondary structural element distribution compared to the AF2

model (Figure 58E). Nonetheless, the α,β-fold was confirmed by circular dichroism spectroscopy. Unfortunately, the protein was instable and thus could not be used for further structural investigations. At room temperature, NisK SD precipitated within minutes and 20 – 40% precipitated within 24 h at 4 °C. This makes structure determination and functional investigation with this protein extremely difficult.

Altogether, the structural investigation of membrane bound sensor histidine kinases is challenging. The current data confirm the high flexibility of these proteins, e.g. through segmental helical motions. In additions, the globular domains may also adopt different conformations. Future studies will hopefully elucidate the structural basis of lanthipeptide binding, SD selectivity and signal transduction through the membrane. This work lays to foundation for such studies.

3.6. Outlook

3.6.1. Structural and functional investigation of the sensor histidine kinase SpaK from *Bacillus subtilis*

To understand signal transduction through the membrane, multiple protein structures and conformations need to be elucidated. The conformational space of SpaK could be investigated by electron paramagnetic resonance (EPR) spectroscopy to analyze domain distances and distance changes. Cross-linking experiments and substrate binding might help to enhance a specific protein conformation for subsequential structural biology analyses with small angle X-ray scattering (SAXS) or CryoEM. Molecular dynamic simulations might help to understand protein motions and trajectories in great detail that could not be visualized by protein snapshots obtained by structural biology approaches. Knowledge gained from structural analyses and *in silico* simulation should always be paired with functional data. Therefore, site-specific mutations should be used to investigate the role of specific residues involved in signal transduction and kinase activity.

Preliminary *in vivo* mutation experiments performed by _____ (____labs, ____) showed that mutations in the four-helix bundle of the HAMP domain (L179P and L198P) can result in an increased autoinduction of SpaK or a constitutively active protein. Initial purification experiments were performed and SpaK L198P could be purified into SMALPs nanodiscs (chapter 3.7, Figure 60). In contrast, L179P tends to form aggregates (chapter 3.7, Figure 60). The effect of these mutations on the kinase activity should be investigated by autoradiography of SpaK using [γ-^{32}P]-ATP. It was hypothesized that the wildtype HAMP domain can give rise to kinase-off phenotypes, while the kinase domain maintains a kinase-on state (*263*). Mutations in the HAMP domain were proposed to disconnect the HAMP domain from the kinase domain, which results in constitutive active protein locked in a kinase-on state (*263, 262*). This hypothesis should be tested in context of the full-length SpaK protein, with the two already obtained HAMP mutants (L179P and L198P) but also other not yet discovered mutants along the HAMP domain.

3.6.2. Exploring the substrate selectivity of the SpaK and NisK sensor domain

The SHK SpaK can only be stimulated by the lanthipeptide subtilin, while the homologous lanthipeptide nisin only activates the SHK NisK from *Lactococcus lactis.* Although the two lanthipeptides are structurally and functionally related, a cross activation is not possible (*283*). To date, the mechanism by which SHKs specifically sense and bind lanthipeptides is still poorly understood. Understanding substrate specificity might be useful in the design of new lanthipeptides that can selectively be used against antibiotic resistant pathogens.

To understand substrate selectivity, the sensor domain (SD) where lanthipeptide binds should be addressed structurally. This chapter showed that SpaK and NisK SDs were difficult to purify and

were expressed in inclusion bodies. With a refolding purification protocol, NisK SD could be purified in big amount but showed only short-term stability when being stored at 4 °C. Therefore, buffer conditions should be screened to find the suitable environment for NisK SD. It is possible that the SD domain could be stabilized in presence of the respective lanthipeptide and is needed during protein purification. Once having stable NisK SD in hand, structural biology approaches should be performed to determine the NisK SD structure with and without substrate. Biophysical analyses to determine binding affinity and thermal stability of the formed substrate - SD complex should be performed. Probably, the gained knowledge of NisK SD could be transferred to the SpaK SD.

Since the lanthipeptide cannot cross-activate the histidine kinases, the question of selectivity exchange by designing chimeric construct arises. Protein chimeras with SpaK SD on NisK backbone and vice versa should be designed and tested *in vivo* and *in vitro.* Experiments with chimeric proteins would identify domains and protein regions responsible for substrate selectivity.

Ultimately, we would have the protein set containing SpaK and NisK in full-length and isolated domains to investigate substrate binding, signal transduction through the histidine kinase, and autophosphorylation mechanism. To this protein set, the two response regulator proteins SpaR and NisR with the respective DNA fragments should be added to fully understand the two-component system from substrate binding to gene regulation.

3.7. Appendix III

SpaK full-length wildtype

```
atg agt gga ata ggt ttt aag ggc cgg aaa acg ctt ctt aga gaa ctg gta aag tat atg
 M   S   G   I   G   F   K   G   R   K   T   L   L   R   E   L   V   K   Y   M
gtg acc cta tgt att agc ttg gtc gta ctt gct ctg cta tat ata ttt ata aat acg att
 V   T   L   C   I   S   L   V   V   L   A   L   L   Y   I   F   I   N   T   I
gca atg aat aca ggt ttc tca cac cca gcg aat tat aat gaa cga gaa gcg gaa aaa ctt
 A   M   N   T   G   F   S   H   P   A   N   Y   N   E   R   E   A   E   K   L
gcg ccg aag ctg gaa acc atc gac aaa gta act gcc gat atg att cct gat aca atg tca
 A   P   K   L   E   T   I   D   K   V   T   A   D   M   I   P   D   T   M   S
tat gcc att ctt aac aaa gaa aca aaa caa aaa aca gct gga act att aaa gaa aaa gat
 Y   A   I   L   N   K   E   T   K   Q   K   T   A   G   T   I   K   E   K   D
ctg cag ctg gta aaa aag aag att gaa aaa aaa cca tat gtc aat tat aag caa aaa ggc
 L   Q   L   V   K   K   K   I   E   K   K   P   Y   V   N   Y   K   Q   K   G
tat ttg gtt ata gaa aga aac aat gaa tat tgt gtt ttg caa tat tca ttg cgt gcc gat
 Y   L   V   I   E   R   N   N   E   Y   C   V   L   Q   Y   S   L   R   A   D
ttc agt tcg cct tta cta aga aaa tac ctt cct aat tat gaa tta acg agt att tgt att
 F   S   S   P   L   L   R   K   Y   L   P   N   Y   E   L   T   S   I   C   I
                                                                       ccg

                                                                     L179P

ttg att atc ttg ctc att ata gtt atc tct att ata aca act tac ttt gcg aat aga tta
 L   I   I   L   L   I   I   V   I   S   I   I   T   T   Y   F   A   N   R   L
                                                                         ccg

                                                                     L198P

aga aaa cat ttt gag aca ttg aac gta atc act cga tat ata aaa gag caa aat ttg caa
 R   K   H   F   E   T   L   N   V   I   T   R   Y   I   K   E   Q   N   L   Q
ttc act ccc gaa ttt aca cat att aag gaa ttc gat gat gtg att gac tcg ttg att gaa
 F   T   P   E   F   T   H   I   K   E   F   D   D   V   I   D   S   L   I   E
atg aga gat gct ctc cag tcc tca ctt gag gca cag tgg cgt tta gaa aaa aac aag aaa
 M   R   D   A   L   Q   S   S   L   E   A   Q   W   R   L   E   K   N   K   K
gaa caa ata ggt gct ttg gca cat gat atc aag att ccg atc acc att att aag ggg aat
 E   Q   I   G   A   L   A   H   D   I   K   I   P   I   T   I   I   K   G   N
gct gaa tta ctc agc ttg tca atg caa aac gaa gaa caa gct gag tat aca aag tat att
 A   E   L   L   S   L   S   M   Q   N   E   E   Q   A   E   Y   T   K   Y   I
ctt gga gca gga aat caa att gaa cag tat att tat cag ctg att cat ctt tcg aaa acg
 L   G   A   G   N   Q   I   E   Q   Y   I   Y   Q   L   I   H   L   S   K   T
gag gat gcc tta acg att cat ttg gag aaa gca tct gta gat gag ttg aca gaa act cta
 E   D   A   L   T   I   H   L   E   K   A   S   V   D   E   L   T   E   T   L
gtg aaa gat ata tcc gcc tat aaa gga aat aaa aat ata aac att tcg ttc aaa aaa gaa
 V   K   D   I   S   A   Y   K   G   N   K   N   I   N   I   S   F   K   K   E
aat cta atg aaa gaa gcg aaa ata gac tgg caa ttg tta cat agg gca tta ttg aat att
 N   L   M   K   E   A   K   I   D   W   Q   L   L   H   R   A   L   L   N   I
```

```
tta aca aat gca gtt gat tat aca ccg gaa ggc ggc acc gtc tcg gta cat gca gaa tgc
 L   T   N   A   V   D   Y   T   P   E   G   G   T   V   S   V   H   A   E   C
gat tct gag ata ttt tat ttt ttt gta aaa gac acg gga aat gga ttt tcg gaa atg ggg
 D   S   E   I   F   Y   F   F   V   K   D   T   G   N   G   F   S   E   M   G
tta aaa aaa gcc aca gaa ctt ttt tat atg gat gac aag agc cgt cat tct aaa ggt cat
 L   K   K   A   T   E   L   F   Y   M   D   D   K   S   R   H   S   K   G   H
tat gga atg gga ttg act ttt gca aaa aat gca gtg aat ctt cat aat gga gaa ctt act
 Y   G   M   G   L   T   F   A   K   N   A   V   N   L   H   N   G   E   L   T
ctt ggg aac act ata gcc ggt gga gca gag gtt cga gta aaa ata cct tta aga aac gaa
 L   G   N   T   I   A   G   G   A   E   V   R   V   K   I   P   L   R   N   E
gca tta gaa gtt ttg ttt caa ggt cca cac cat cat cat cac cat cat cat cat cat cat
 A   L   E   V   L   F   Q   G   P   H   H   H   H   H   H   H   H   H   H
```

NisK SD

```
cac cat cat cat cac cat cat cat cat cat tta gaa gtt ttg ttt caa ggt cca agt ggt
 H   H   H   H   H   H   H   H   H   H   L   E   V   L   F   Q   G   P   S   G
tca gaa act att cgt tta tct tta gat tca gat aat tta act att tct gat atc gaa cgt
 S   E   T   I   R   L   S   L   D   S   D   N   L   T   I   S   D   I   E   R
gat atg aaa cac tac cca tat gat tat att att ttt gac aat gat aca agt aaa att ttg
 D   M   K   H   Y   P   Y   D   Y   I   I   F   D   N   D   T   S   K   I   L
gga gga cat tat gtc aag tcg gat gta cct agt ttt gta gct tca aaa cag tct tca cat
 G   G   H   Y   V   K   S   D   V   P   S   F   V   A   S   K   Q   S   S   H
aat att aca gaa gga gaa att act tat act tat tca agc aat aag cat ttt tca gtt gtt
 N   I   T   E   G   E   I   T   Y   T   Y   S   S   N   K   H   F   S   V   V
tta aga caa aac agt atg cct gaa ttt aca
 L   R   Q   N   S   M   P   E   F   T
```

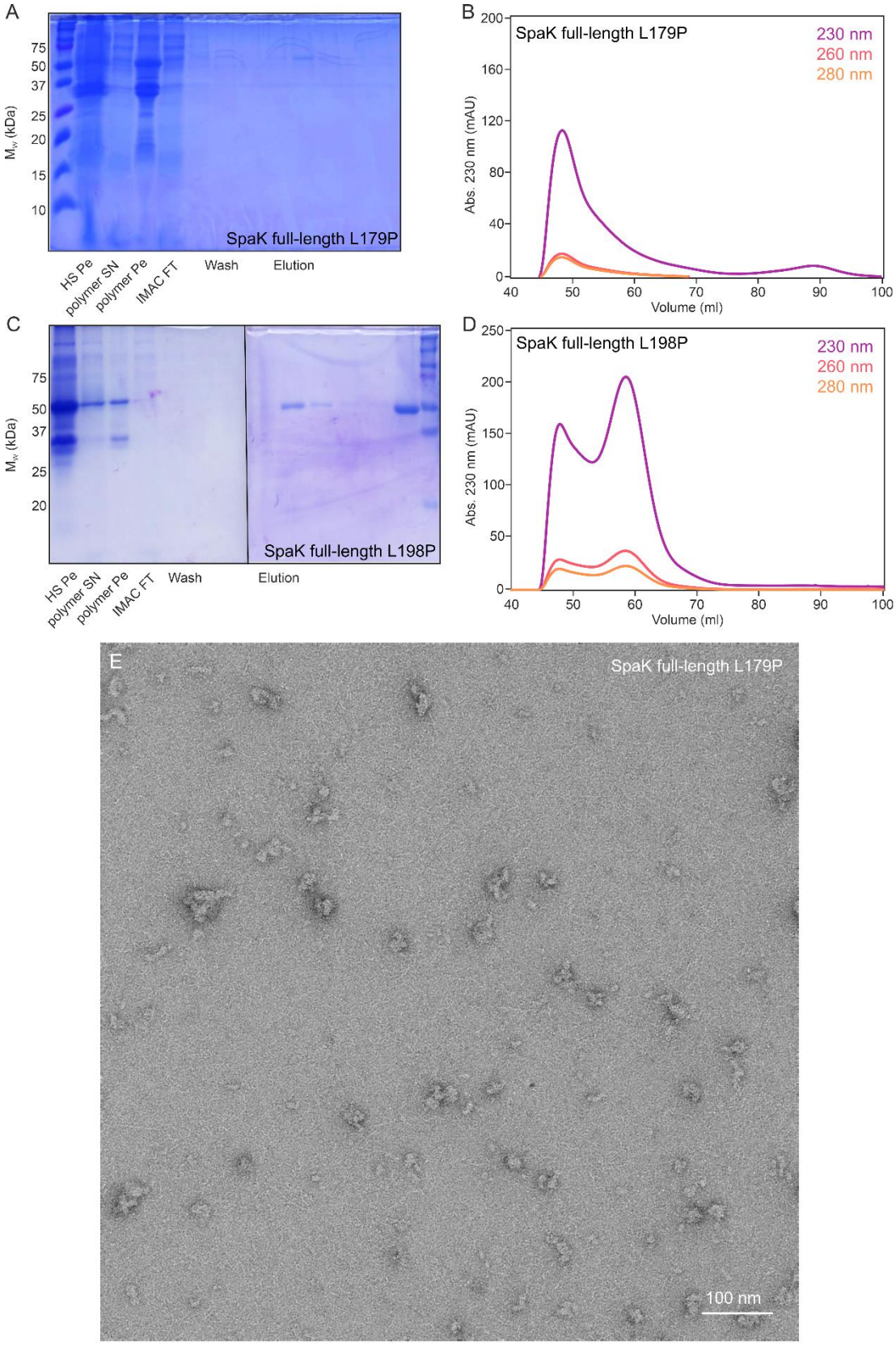


Figure 60: Purification of SpaK full-length L179P and L198P constructs. A and C: SDS-PAGE gel for the purification of SpaK L179P and L198P full-length protein, respectively. B and D: Size exclusion chromatograph of the samples containing SpaK L179P and L198P protein, respectively, run on Superdex 75 Highload 16/60. E: Electron micrograph of 0.03 µM SpaK L179P protein in smalps stained with uranyl acetate on a copper support grid.

References

1. Lyons, N. A.; Kolter, R. On the evolution of bacterial multicellularity. *Current opinion in microbiology* **2015,** *24*, 21–28. DOI: 10.1016/j.mib.2014.12.007.
2. Hengge, R. Linking bacterial growth, survival, and multicellularity - small signaling molecules as triggers and drivers. *Current opinion in microbiology* **2020,** *55*, 57–66. DOI: 10.1016/j.mib.2020.02.007.
3. Urvoy, M.; Labry, C.; L'Helguen, S.; Lami, R. Quorum Sensing Regulates Bacterial Processes That Play a Major Role in Marine Biogeochemical Cycles. *Front. Mar. Sci.* **2022,** *9*, 47. DOI: 10.3389/fmars.2022.834337.
4. Papenfort, K.; Bassler, B. L. Quorum sensing signal-response systems in Gram-negative bacteria. *Nature Reviews Microbiology* **2016,** *14* (9), 576–588. DOI: 10.1038/nrmicro.2016.89.
5. Winzer, K.; Hardie, K. R.; Williams, P. Bacterial cell-to-cell communication: sorry, can't talk now - gone to lunch! *Current opinion in microbiology* **2002,** *5* (2), 216–222. DOI: 10.1016/s1369-5274(02)00304-1.
6. McBrayer, D. N.; Cameron, C. D.; Tal-Gan, Y. Development and utilization of peptide-based quorum sensing modulators in Gram-positive bacteria. *Org. Biomol. Chem.* **2020,** *18* (37), 7273–7290. DOI: 10.1039/D0OB01421D.
7. Liu, L.; Zeng, X.; Zheng, J.; Zou, Y.; Qiu, S.; Dai, Y. AHL-mediated quorum sensing to regulate bacterial substance and energy metabolism: A review. *Microbiological research* **2022,** *262*, 127102. DOI: 10.1016/j.micres.2022.127102.
8. Reiners, J.; Lagedroste, M.; Gottstein, J.; Adeniyi, E. T.; Kalscheuer, R.; Poschmann, G.; Stühler, K.; Smits, S. H. J.; Schmitt, L. Insights in the Antimicrobial Potential of the Natural Nisin Variant Nisin H. *Frontiers in microbiology* **2020,** *11*, 573614. DOI: 10.3389/fmicb.2020.573614.
9. Fuchs, S. W.; Jaskolla, T. W.; Bochmann, S.; Kötter, P.; Wichelhaus, T.; Karas, M.; Stein, T.; Entian, K.-D. Entianin, a novel subtilin-like lantibiotic from Bacillus subtilis subsp. spizizenii DSM 15029T with high antimicrobial activity. *Applied and environmental microbiology* **2011,** *77* (5), 1698–1707. DOI: 10.1128/AEM.01962-10.
10. Beis, K.; Rebuffat, S. Multifaceted ABC transporters associated to microcin and bacteriocin export. *Research in Microbiology* **2019,** *170* (8), 399–406. DOI: 10.1016/j.resmic.2019.07.002.
11. Simons, A.; Alhanout, K.; Duval, R. E. Bacteriocins, Antimicrobial Peptides from Bacterial Origin: Overview of Their Biology and Their Impact against Multidrug-Resistant Bacteria. *Microorganisms* **2020,** *8* (5). DOI: 10.3390/microorganisms8050639.
12. Bernard, C.; Li, Y.; Bapteste, E.; Lopez, P. RRNPP_detector: a tool to detect RRNPP quorum sensing systems in chromosomes, plasmids and phages of gram-positive bacteria. *bioRxiv* **2021**, 2021.08.18.456871. DOI: 10.1101/2021.08.18.456871.

13. Wynendaele, E.; Debunne, N.; Verbeke, F.; Janssens, Y.; Spiegeleer, A. de; Spiegeleer, B. de. Quorum Sensing Peptides and Their Interactions with the Host. *Quorum sensing : microbial rules of life* **2020,** *1374*, 157–176. DOI: 10.1021/bk-2020-1374.ch009.

14. Boedicker, J. and Nealson, K. Microbial communication via quorum sensing. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications* **2015** (4), 310–320. DOI: 10.1109/TMBMC.2016.2587629.

15. Yi, L.; Dong, X.; Grenier, D.; Wang, K.; Wang, Y. Research progress of bacterial quorum sensing receptors: Classification, structure, function and characteristics. *Science of The Total Environment* **2021,** *763*, 143031. DOI: 10.1016/j.scitotenv.2020.143031.

16. Kalamara, M.; Spacapan, M.; Mandic-Mulec, I.; Stanley-Wall, N. R. Social behaviours by Bacillus subtilis: quorum sensing, kin discrimination and beyond. *Molecular microbiology* **2018,** *110* (6), 863–878. DOI: 10.1111/mmi.14127.

17. Rémy, B.; Mion, S.; Plener, L.; Elias, M.; Chabrière, E.; Daudé, D. Interference in Bacterial Quorum Sensing: A Biopharmaceutical Perspective. *Front. Pharmacol.* **2018,** *9*, 203. DOI: 10.3389/fphar.2018.00203.

18. Hiller, D. A.; Singh, V.; Zhong, M.; Strobel, S. A. A two-step chemical mechanism for ribosome-catalysed peptide bond formation. *Nature* **2011,** *476* (7359), 236–239. DOI: 10.1038/nature10248.

19. Izoré, T.; Cryle, M. J. The many faces and important roles of protein-protein interactions during non-ribosomal peptide synthesis. *Natural Product Reports* **2018,** *35* (11), 1120–1139. DOI: 10.1039/C8NP00038G.

20. Wiest, A.; Grzegorski, D.; Xu, B.-W.; Goulard, C.; Rebuffat, S.; Ebbole, D. J.; Bodo, B.; Kenerley, C. Identification of peptaibols from Trichoderma virens and cloning of a peptaibol synthetase. *Journal of Biological Chemistry* **2002,** *277* (23), 20862–20868. DOI: 10.1074/jbc.M201654200.

21. Klapper, M.; Braga, D.; Lackner, G.; Herbst, R.; Stallforth, P. Bacterial Alkaloid Biosynthesis: Structural Diversity via a Minimalistic Nonribosomal Peptide Synthetase. *Cell Chemical Biology* **2018,** *25* (6), 659-665.e9. DOI: 10.1016/j.chembiol.2018.02.013.

22. Bode, H. B.; Brachmann, A. O.; Jadhav, K. B.; Seyfarth, L.; Dauth, C.; Fuchs, S. W.; Kaiser, M.; Waterfield, N. R.; Sack, H.; Heinemann, S. H.; Arndt, H.-D. Structure Elucidation and Activity of Kolossin A, the D-/L-Pentadecapeptide Product of a Giant Nonribosomal Peptide Synthetase. *Angewandte Chemie International Edition* **2015,** *54* (35), 10352–10355. DOI: 10.1002/anie.201502835.

23. Eppelmann, K.; Stachelhaus, T.; Marahiel, M. A. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **2002,** *41* (30), 9718–9726. DOI: 10.1021/bi0259406.

24. D'Ambrosio, H. K.; Derbyshire, E. R. Investigating the Role of Class I Adenylate-Forming Enzymes in Natural Product Biosynthesis. *ACS Chemical Biology* **2020,** *15* (1), 17–27. DOI: 10.1021/acschembio.9b00865.

25. Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. *Nucleic acids research* **2011,** *39* (Web Server issue), W362-7. DOI: 10.1093/nar/gkr323.

26. Khayatt, B. I.; Overmars, L.; Siezen, R. J.; Francke, C. Classification of the adenylation and acyl-transferase activity of NRPS and PKS systems using ensembles of substrate specific hidden Markov models. *PLOS ONE* **2013,** *8* (4), e62136. DOI: 10.1371/journal.pone.0062136.

27. Stanišić, A.; Kries, H. Adenylation Domains in Nonribosomal Peptide Engineering. *ChemBioChem* **2019,** *20* (11), 1347–1356. DOI: 10.1002/cbic.201800750.

28. Izoré, T.; Candace Ho, Y. T.; Kaczmarski, J. A.; Gavriilidou, A.; Chow, K. H.; Steer, D. L.; Goode, R. J. A.; Schittenhelm, R. B.; Tailhades, J.; Tosin, M.; Challis, G. L.; Krenske, E. H.; Ziemert, N.; Jackson, C. J.; Cryle, M. J. Structures of a non-ribosomal peptide synthetase condensation domain suggest the basis of substrate selectivity. *Nat Commun* **2021,** *12* (1), 2511. DOI: 10.1038/s41467-021-22623-0.

29. Chu Yuan Kee, M.-J.; Bharath, S. R.; Wee, S.; Bowler, M. W.; Gunaratne, J.; Pan, S.; Zhang, L.; Song, H. Structural insights into the substrate-bound condensation domains of non-ribosomal peptide synthetase AmbB. *Scientific reports* **2022,** *12* (1), 5353. DOI: 10.1038/s41598-022-09188-8.

30. Bloudoff, K.; Schmeing, T. M. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochimica et biophysica acta. Proteins and proteomics* **2017,** *1865* (11 Pt B), 1587–1604. DOI: 10.1016/j.bbapap.2017.05.010.

31. Little, R. F.; Hertweck, C. Chain release mechanisms in polyketide and non-ribosomal peptide biosynthesis. *Natural Product Reports* **2022,** *39* (1), 163–205. DOI: 10.1039/d1np00035g.

32. Klapper, M.; Götze, S.; Barnett, R.; Willing, K.; Stallforth, P. Bacterial Alkaloids Prevent Amoebal Predation. *Angewandte Chemie International Edition* **2016,** *55* (31), 8944–8947. DOI: 10.1002/anie.201603312.

33. Dekimpe, S.; Masschelein, J. Beyond peptide bond formation: the versatile role of condensation domains in natural product biosynthesis. *Natural Product Reports* **2021,** *38* (10), 1910–1937. DOI: 10.1039/d0np00098a.

34. Xu, X.; Qu, R.; Wu, W.; Jiang, C.; Shao, D.; Shi, J. Applications of microbial co-cultures in polyketides production. *Journal of applied microbiology* **2021,** *130* (4), 1023–1034. DOI: 10.1111/jam.14845.

35. Pavesi, C.; Flon, V.; Mann, S.; Leleu, S.; Prado, S.; Franck, X. Biosynthesis of azaphilones: a review. *Natural Product Reports* **2021,** *38* (6), 1058–1071. DOI: 10.1039/d0np00080a.

36. Ruocco, M.; Baroncelli, R.; Cacciola, S. O.; Pane, C.; Monti, M. M.; Firrao, G.; Vergara, M.; Di Magnano San Lio, G.; Vannacci, G.; Scala, F. Polyketide synthases of Diaporthe helianthi and involvement of DhPKS1 in virulence on sunflower. *BMC genomics* **2018,** *19* (1), 27. DOI: 10.1186/s12864-017-4405-z.

37. Flórez, L. V.; Scherlach, K.; Miller, I. J.; Rodrigues, A.; Kwan, J. C.; Hertweck, C.; Kaltenpoth, M. An antifungal polyketide associated with horizontally acquired genes supports symbiont-mediated defense in Lagria villosa beetles. *Nat Commun* **2018,** *9* (1), 2478. DOI: 10.1038/s41467-018-04955-6.

38. Hwang, S.; Lee, N.; Cho, S.; Palsson, B.; Cho, B.-K. Repurposing Modular Polyketide Synthases and Non-ribosomal Peptide Synthetases for Novel Chemical Biosynthesis. *Frontiers in molecular biosciences* **2020,** *7*, 87. DOI: 10.3389/fmolb.2020.00087.

39. Robbins, T.; Kapilivsky, J.; Cane, D. E.; Khosla, C. Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. *Biochemistry* **2016,** *55* (32), 4476–4484. DOI: 10.1021/acs.biochem.6b00639.

40. Musiol-Kroll, E. M.; Wohlleben, W. Acyltransferases as Tools for Polyketide Synthase Engineering. *Antibiotics* **2018,** *7* (3). DOI: 10.3390/antibiotics7030062.

41. Yadav, G.; Gokhale, R. S.; Mohanty, D. Towards prediction of metabolic products of polyketide synthases: an in silico analysis. *PLoS computational biology* **2009,** *5* (4), e1000351. DOI: 10.1371/journal.pcbi.1000351.

42. Kornfuehrer, T.; Eustáquio, A. S. Diversification of polyketide structures via synthase engineering. *MedChemComm* **2019,** *10* (8), 1256–1272. DOI: 10.1039/c9md00141g.

43. Kim, J.; Yi, G.-S. PKMiner: a database for exploring type II polyketide synthases. *BMC microbiology* **2012,** *12*, 169. DOI: 10.1186/1471-2180-12-169.

44. Wang, J.; Zhang, R.; Chen, X.; Sun, X.; Yan, Y.; Shen, X.; Yuan, Q. Biosynthesis of aromatic polyketides in microorganisms using type II polyketide synthases. *Microbial cell factories* **2020,** *19* (1), 110. DOI: 10.1186/s12934-020-01367-4.

45. Sulpizio, A.; Crawford, C. E. W.; Koweek, R. S.; Charkoudian, L. K. Probing the structure and function of acyl carrier proteins to unlock the strategic redesign of type II polyketide biosynthetic pathways. *The Journal of biological chemistry* **2021,** *296*, 100328. DOI: 10.1016/j.jbc.2021.100328.

46. Tang, Y.; Tsai, S.-C.; Khosla, C. Polyketide chain length control by chain length factor. *Journal of the American Chemical Society* **2003,** *125* (42), 12708–12709. DOI: 10.1021/ja0378759.

47. Miyanaga, A.; Kudo, F.; Eguchi, T. Protein-protein interactions in polyketide synthase-nonribosomal peptide synthetase hybrid assembly lines. *Natural Product Reports* **2018,** *35* (11), 1185–1209. DOI: 10.1039/c8np00022k.

48. Partida-Martinez, L. P.; Hertweck, C. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* **2005,** *437* (7060), 884–888. DOI: 10.1038/nature03997.

49. Partida-Martinez, L. P.; Hertweck, C. A gene cluster encoding rhizoxin biosynthesis in "Burkholderia rhizoxina", the bacterial endosymbiont of the fungus Rhizopus microsporus. *ChemBioChem* **2007,** *8* (1), 41–45. DOI: 10.1002/cbic.200600393.

50. Iwasaki, S.; Kobayashi, H.; Furukawa, J.; Namikoshi, M.; Okuda, S.; Sato, Z.; Matsuda, I.; Noda, T. Studies on macrocyclic lactone antibiotics. VII. Structure of a phytotoxin "rhizoxin" produced by Rhizopus chinensis. *The Journal of antibiotics* **1984,** *37* (4), 354–362. DOI: 10.7164/antibiotics.37.354.

51. Bretschneider, T.; Heim, J. B.; Heine, D.; Winkler, R.; Busch, B.; Kusebauch, B.; Stehle, T.; Zocher, G.; Hertweck, C. Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature* **2013,** *502* (7469), 124–128. DOI: 10.1038/nature12588.

52. Heine, D.; Bretschneider, T.; Sundaram, S.; Hertweck, C. Enzymatic polyketide chain branching to give substituted lactone, lactam, and glutarimide heterocycles. *Angewandte Chemie International Edition* **2014,** *53* (43), 11645–11649. DOI: 10.1002/anie.201407282.

53. Beck, C.; Garzón, J. F. G.; Weber, T. Recent Advances in Re-engineering Modular PKS and NRPS Assembly Lines. *Biotechnology and Bioprocess Engineering* **2020,** *25* (6), 886–894. DOI: 10.1007/s12257-020-0265-5.

54. Farmer, R.; Thomas, C. M.; Winn, P. J. Structure, function and dynamics in acyl carrier proteins. *PLOS ONE* **2019,** *14* (7), e0219435. DOI: 10.1371/journal.pone.0219435.

55. Gulick, A. M.; Aldrich, C. C. Trapping interactions between catalytic domains and carrier proteins of modular biosynthetic enzymes with chemical probes. *Natural Product Reports* **2018,** *35* (11), 1156–1184. DOI: 10.1039/c8np00044a.

56. Copp, J. N.; Neilan, B. A. The phosphopantetheinyl transferase superfamily: phylogenetic analysis and functional implications in cyanobacteria. *Applied and environmental microbiology* **2006,** *72* (4), 2298–2305. DOI: 10.1128/AEM.72.4.2298-2305.2006.

57. Corpuz, J. C.; Sanlley, J. O.; Burkart, M. D. Protein-protein interface analysis of the non-ribosomal peptide synthetase peptidyl carrier protein and enzymatic domains. *Synthetic and systems biotechnology* **2022,** *7* (2), 677–688. DOI: 10.1016/j.synbio.2022.02.006.

58. Quadri, L. E.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T. Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **1998,** *37* (6), 1585–1595. DOI: 10.1021/bi9719861.

59. Sunbul, M.; Marshall, N. J.; Zou, Y.; Zhang, K.; Yin, J. Catalytic turnover-based phage selection for engineering the substrate specificity of Sfp phosphopantetheinyl transferase. *Journal of molecular biology* **2009,** *387* (4), 883–898. DOI: 10.1016/j.jmb.2009.02.010.

60. Yasgar, A.; Foley, T. L.; Jadhav, A.; Inglese, J.; Burkart, M. D.; Simeonov, A. A strategy to discover inhibitors of Bacillus subtilis surfactin-type phosphopantetheinyl transferase. *Molecular bioSystems* **2010,** *6* (2), 365–375. DOI: 10.1039/b913291k.

61. Evans, S. E.; Williams, C.; Arthur, C. J.; Burston, S. G.; Simpson, T. J.; Crosby, J.; Crump, M. P. An ACP structural switch: conformational differences between the apo and holo forms of the

actinorhodin polyketide synthase acyl carrier protein. *ChemBioChem* **2008,** *9* (15), 2424–2432. DOI: 10.1002/cbic.200800180.

62. Alonzo, D. A.; Chiche-Lapierre, C.; Tarry, M. J.; Wang, J.; Schmeing, T. M. Structural basis of keto acid utilization in nonribosomal depsipeptide synthesis. *Nature chemical biology* **2020,** *16* (5), 493–496. DOI: 10.1038/s41589-020-0481-5.

63. Drake, E. J.; Miller, B. R.; Shi, C.; Tarrasch, J. T.; Sundlov, J. A.; Allen, C. L.; Skiniotis, G.; Aldrich, C. C.; Gulick, A. M. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **2016,** *529* (7585), 235–238. DOI: 10.1038/nature16163.

64. Frueh, D. P.; Arthanari, H.; Koglin, A.; Vosburg, D. A.; Bennett, A. E.; Walsh, C. T.; Wagner, G. Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature* **2008,** *454* (7206), 903–906. DOI: 10.1038/nature07162.

65. Miyanaga, A.; Ouchi, R.; Ishikawa, F.; Goto, E.; Tanabe, G.; Kudo, F.; Eguchi, T. Structural Basis of Protein-Protein Interactions between a trans-Acting Acyltransferase and Acyl Carrier Protein in Polyketide Disorazole Biosynthesis. *Journal of the American Chemical Society* **2018,** *140* (25), 7970–7978. DOI: 10.1021/jacs.8b04162.

66. Chen, A.; Re, R. N.; Burkart, M. D. Type II fatty acid and polyketide synthases: deciphering protein-protein and protein-substrate interactions. *Natural Product Reports* **2018,** *35* (10), 1029–1045. DOI: 10.1039/c8np00040a.

67. Gruenewald, S.; Mootz, H. D.; Stehmeier, P.; Stachelhaus, T. In vivo production of artificial nonribosomal peptide products in the heterologous host Escherichia coli. *Applied and environmental microbiology* **2004,** *70* (6), 3282–3291. DOI: 10.1128/AEM.70.6.3282-3291.2004.

68. Drepper, T.; Arvani, S.; Rosenau, F.; Wilhelm, S.; Jaeger, K.-E. High-level transcription of large gene regions: a novel T(7) RNA-polymerase-based system for expression of functional hydrogenases in the phototrophic bacterium Rhodobacter capsulatus. *Biochemical Society transactions* **2005,** *33* (Pt 1), 56–58. DOI: 10.1042/BST0330056.

69. Süssmuth, R. D.; Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angewandte Chemie International Edition* **2017,** *56* (14), 3770–3821. DOI: 10.1002/anie.201609079.

70. Brito, J. A.; Archer, M. Structural biology techniques: X-ray crystallography, cryo-electron microscopy, and small-angle X-ray scattering. In: *Practical Approaches to Biological Inorganic Chemistry*; Elsevier, 2020, pp 375–416. DOI: 10.1016/B978-0-444-64225-7.00010-9.

71. Puthenveetil, R.; Vinogradova, O. Solution NMR: A powerful tool for structural and functional studies of membrane proteins in reconstituted environments. *The Journal of biological chemistry* **2019,** *294* (44), 15914–15931. DOI: 10.1074/jbc.REV119.009178.

72. Wiedemann, C.; Goretzki, B.; Merz, Z. N.; Tebbe, F.; Schmitt, P.; Hellmich, U. A. Extent of intrinsic disorder and NMR chemical shift assignments of the distal N-termini from human

TRPV1, TRPV2 and TRPV3 ion channels. *Biomolecular NMR assignments* **2022,** *16* (2), 289–296. DOI: 10.1007/s12104-022-10093-4.

73. Goretzki, B.; Tebbe, F.; Mitrovic, S.-A.; Hellmich, U. A. Backbone NMR assignments of the extensive human and chicken TRPV4 N-terminal intrinsically disordered regions as important players in ion channel regulation. *Biomolecular NMR assignments* **2022,** *16* (2), 205–212. DOI: 10.1007/s12104-022-10080-9.

74. Tugarinov, V.; Muhandiram, R.; Ayed, A.; Kay, L. E. Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase G. *Journal of the American Chemical Society* **2002,** *124* (34), 10025–10035. DOI: 10.1021/ja0205636.

75. Verardi, R.; Traaseth, N. J.; Masterson, L. R.; Vostrikov, V. V.; Veglia, G. Isotope labeling for solution and solid-state NMR spectroscopy of membrane proteins. *Advances in experimental medicine and biology* **2012,** *992*, 35–62. DOI: 10.1007/978-94-007-4954-2_3.

76. Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic acids research* **2015,** *43* (W1), W237-43. DOI: 10.1093/nar/gkv437.

77. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021,** *596* (7873), 583–589. DOI: 10.1038/s41586-021-03819-2.

78. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; Žídek, A.; Green, T.; Tunyasuvunakool, K.; Petersen, S.; Jumper, J.; Clancy, E.; Green, R.; Vora, A.; Lutfi, M.; Figurnov, M.; Cowie, A.; Hobbs, N.; Kohli, P.; Kleywegt, G.; Birney, E.; Hassabis, D.; Velankar, S. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research* **2022,** *50* (D1), D439-D444. DOI: 10.1093/nar/gkab1061.

79. Bhinderwala, F.; Evans, P.; Jones, K.; Laws, B. R.; Smith, T. G.; Morton, M.; Powers, R. Phosphorus NMR and Its Application to Metabolomics. *Analytical chemistry* **2020,** *92* (14), 9536–9545. DOI: 10.1021/acs.analchem.0c00591.

80. Fu, Y.; Kang, Z.; Cao, W.; Yin, J.; Tu, Y.; Li, J.; Guan, H.; Wang, Y.; Wang, Q.; Kong, X. Defect-Assisted Loading and Docking Conformations of Pharmaceuticals in Metal-Organic Frameworks. *Angewandte Chemie International Edition* **2021,** *60* (14), 7719–7727. DOI: 10.1002/anie.202010231.

81. Stourac, J.; Vavra, O.; Kokkonen, P.; Filipovic, J.; Pinto, G.; Brezovsky, J.; Damborsky, J.; Bednar, D. Caver Web 1.0: identification of tunnels and channels in proteins and analysis of

ligand transport. *Nucleic acids research* **2019,** *47* (W1), W414-W422. DOI: 10.1093/nar/gkz378.

82. Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS computational biology* **2012,** *8* (10), e1002708. DOI: 10.1371/journal.pcbi.1002708.

83. Patteson, J. B.; Fortinez, C. M.; Putz, A. T.; Rodriguez-Rivas, J.; Bryant, L. H.; Adhikari, K.; Weigt, M.; Schmeing, T. M.; Li, B. Structure and Function of a Dehydrating Condensation Domain in Nonribosomal Peptide Biosynthesis. *Journal of the American Chemical Society* **2022,** *144* (31), 14057–14070. DOI: 10.1021/jacs.1c13404.

84. Bräuer, A.; Zhou, Q.; Grammbitter, G. L. C.; Schmalhofer, M.; Rühl, M.; Kaila, V. R. I.; Bode, H. B.; Groll, M. Structural snapshots of the minimal PKS system responsible for octaketide biosynthesis. *Nature chemistry* **2020,** *12* (8), 755–763. DOI: 10.1038/s41557-020-0491-7.

85. Mindrebo, J. T.; Patel, A.; Kim, W. E.; Davis, T. D.; Chen, A.; Bartholow, T. G.; La Clair, J. J.; McCammon, J. A.; Noel, J. P.; Burkart, M. D. Gating mechanism of elongating β-ketoacyl-ACP synthases. *Nature Communications* **2020,** *11* (1), 1727. DOI: 10.1038/s41467-020-15455-x.

86. Milligan, J. C.; Lee, D. J.; Jackson, D. R.; Schaub, A. J.; Beld, J.; Barajas, J. F.; Hale, J. J.; Luo, R.; Burkart, M. D.; Tsai, S.-C. Molecular basis for interactions between an acyl carrier protein and a ketosynthase. *Nature chemical biology* **2019,** *15* (7), 669–671. DOI: 10.1038/s41589-019-0301-y.

87. Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R. H.; Håkansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. Structure of a modular polyketide synthase. *Nature* **2014,** *510* (7506), 512–517. DOI: 10.1038/nature13423.

88. Shen, Y.; Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of biomolecular NMR* **2013,** *56* (3), 227–241. DOI: 10.1007/s10858-013-9741-y.

89. Du, D.; Katsuyama, Y.; Horiuchi, M.; Fushinobu, S.; Chen, A.; Davis, T. D.; Burkart, M. D.; Ohnishi, Y. Structural basis for selectivity in a highly reducing type II polyketide synthase. *Nature chemical biology* **2020,** *16* (7), 776–782. DOI: 10.1038/s41589-020-0530-0.

90. Viegas, M. F.; Neves, R. P. P.; Ramos, M. J.; Fernandes, P. A. Modeling of Human Fatty Acid Synthase and in Silico Docking of Acyl Carrier Protein Domain and Its Partner Catalytic Domains. *The journal of physical chemistry. B* **2018,** *122* (1), 77–85. DOI: 10.1021/acs.jpcb.7b09645.

91. Maloney, F. P.; Gerwick, L.; Gerwick, W. H.; Sherman, D. H.; Smith, J. L. Anatomy of the β-branching enzyme of polyketide biosynthesis and its interaction with an acyl-ACP substrate. *Proceedings of the National Academy of Sciences of the United States of America* **2016,** *113* (37), 10316–10321. DOI: 10.1073/pnas.1607210113.

92. Passmore, M.; Gallo, A.; Lewandowski, J. R.; Jenner, M. Molecular basis for acyl carrier protein-ketoreductase interaction in trans-acyltransferase polyketide synthases. *Chemical science* **2021,** *12* (41), 13676–13685. DOI: 10.1039/d1sc03478b.

93. Lin, Z.; Qu, X. Emerging diversity in polyketide synthase. *Tetrahedron Letters* **2022,** *110*, 154183. DOI: 10.1016/j.tetlet.2022.154183.

94. Herbst, D. A.; Huitt-Roehl, C. R.; Jakob, R. P.; Kravetz, J. M.; Storm, P. A.; Alley, J. R.; Townsend, C. A.; Maier, T. The structural organization of substrate loading in iterative polyketide synthases. *Nature chemical biology* **2018,** *14* (5), 474–479. DOI: 10.1038/s41589-018-0026-3.

95. Acheampong, K. K.; Kokona, B.; Braun, G. A.; Jacobsen, D. R.; Johnson, K. A.; Charkoudian, L. K. Colorimetric Assay Reports on Acyl Carrier Protein Interactions. *Scientific reports* **2019,** *9* (1), 15589. DOI: 10.1038/s41598-019-51554-6.

96. Blin, K.; Shaw, S.; Kloosterman, A. M.; Charlop-Powers, Z.; van Wezel, G. P.; Medema, M. H.; Weber, T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic acids research* **2021,** *49* (W1), W29-W35. DOI: 10.1093/nar/gkab335.

97. Chen, A.; Jiang, Z.; Burkart, M. D. Enzymology of standalone elongating ketosynthases. *Chemical science* **2022,** *13* (15), 4225–4238. DOI: 10.1039/d1sc07256k.

98. East, K. W.; Skeens, E.; Cui, J. Y.; Belato, H. B.; Mitchell, B.; Hsu, R.; Batista, V. S.; Palermo, G.; Lisi, G. P. NMR and computational methods for molecular resolution of allosteric pathways in enzyme complexes. *Biophysical reviews* **2020,** *12* (1), 155–174. DOI: 10.1007/s12551-019-00609-z.

99. Rivas, M.; Courouble, V. C.; Baker, M. C.; Cookmeyer, D. L.; Fiore, K. E.; Frost, A. J.; Godbe, K. N.; Jordan, M. R.; Krasnow, E. N.; Mollo, A.; Ridings, S. T.; Sawada, K.; Shroff, K. D.; Studnitzer, B.; Thiele, G. A. R.; Sisto, A. C.; Nawaal, S.; Huff, A. R.; Fairman, R.; Johnson, K. A.; Beld, J.; Kokona, B.; Charkoudian, L. K. The Effect of Divalent Cations on the Thermostability of Type II Polyketide Synthase Acyl Carrier Proteins. *American Institute of Chemical Engineers* **2018,** *64* (12), 4308–4318. DOI: 10.1002/aic.16402.

100. Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Natural Product Reports* **2014,** *31* (1), 61–108. DOI: 10.1039/c3np70054b.

101. Tufar, P.; Rahighi, S.; Kraas, F. I.; Kirchner, D. K.; Löhr, F.; Henrich, E.; Köpke, J.; Dikic, I.; Güntert, P.; Marahiel, M. A.; Dötsch, V. Crystal structure of a PCP/Sfp complex reveals the structural basis for carrier protein posttranslational modification. *Chemistry & biology* **2014,** *21* (4), 552–562. DOI: 10.1016/j.chembiol.2014.02.014.

102. Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (44), 15815–15820. DOI: 10.1073/pnas.0507705102.

103. Parris, K. D.; Lin, L.; Tam, A.; Mathew, R.; Hixon, J.; Stahl, M.; Fritz, C. C.; Seehra, J.; Somers, W. S. Crystal structures of substrate binding to Bacillus subtilis holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure* **2000,** *8* (8), 883–895. DOI: 10.1016/s0969-2126(00)00178-7.

104. Nikolopoulos, N.; Matos, R. C.; Courtin, P.; Ayala, I.; Akherraz, H.; Simorre, J.-P.; Chapot-Chartier, M.-P.; Leulier, F.; Ravaud, S.; Grangeasse, C. DltC acts as an interaction hub for AcpS, DltA and DltB in the teichoic acid D-alanylation pathway of Lactiplantibacillus plantarum. *Scientific reports* **2022,** *12* (1), 13133. DOI: 10.1038/s41598-022-17434-2.

105. Marcella, A. M.; Culbertson, S. J.; Shogren-Knaak, M. A.; Barb, A. W. Structure, High Affinity, and Negative Cooperativity of the Escherichia coli Holo-(Acyl Carrier Protein):Holo-(Acyl Carrier Protein) Synthase Complex. *Journal of molecular biology* **2017,** *429* (23), 3763–3775. DOI: 10.1016/j.jmb.2017.10.015.

106. Mitra, S.; Dhar, R.; Sen, R. Designer bacterial cell factories for improved production of commercially valuable non-ribosomal peptides. *Biotechnology advances* **2022,** *60*, 108023. DOI: 10.1016/j.biotechadv.2022.108023.

107. Katsuyama, Y.; Miyanaga, A. Recent advances in the structural biology of modular polyketide synthases and nonribosomal peptide synthetases. *Current opinion in chemical biology* **2022,** *71*, 102223. DOI: 10.1016/j.cbpa.2022.102223.

108. Reimer, J. M.; Eivaskhani, M.; Harb, I.; Guarné, A.; Weigt, M.; Schmeing, T. M. Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science (New York, N.Y.)* **2019,** *366* (6466). DOI: 10.1126/science.aaw4388.

109. Reimer, J. M.; Aloise, M. N.; Harrison, P. M.; Schmeing, T. M. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature* **2016,** *529* (7585), 239–242. DOI: 10.1038/nature16503.

110. Mishra, S. H.; Kancherla, A. K.; Marincin, K. A.; Bouvignies, G.; Nerli, S.; Sgourakis, N.; Dowling, D. P.; Frueh, D. P. Global protein dynamics as communication sensors in peptide synthetase domains. *Science advances* **2022,** *8* (28), eabn6549. DOI: 10.1126/sciadv.abn6549.

111. Wang, J.; Li, D.; Chen, L.; Cao, W.; Kong, L.; Zhang, W.; Croll, T.; Deng, Z.; Liang, J.; Wang, Z. Catalytic trajectory of a dimeric nonribosomal peptide synthetase subunit with an inserted epimerase domain. *Nature Communications* **2022,** *13* (1), 592. DOI: 10.1038/s41467-022-28284-x.

112. Gaudet, R. Divide and conquer: high resolution structural information on TRP channel fragments. *The Journal of general physiology* **2009,** *133* (3), 231–237. DOI: 10.1085/jgp.200810137.

113. Strelkov, S. V.; Herrmann, H.; Geisler, N.; Lustig, A.; Ivaninskii, S.; Zimbelmann, R.; Burkhard, P.; Aebi, U. Divide-and-conquer crystallographic approach towards an atomic structure of intermediate filaments. *Journal of molecular biology* **2001,** *306* (4), 773–781. DOI: 10.1006/jmbi.2001.4442.

114. Wiegand, T.; Gardiennet, C.; Cadalbert, R.; Lacabanne, D.; Kunert, B.; Terradot, L.; Böckmann, A.; Meier, B. H. Variability and conservation of structural domains in divide-and-conquer approaches. *Journal of biomolecular NMR* **2016,** *65* (2), 79–86. DOI: 10.1007/s10858-016-0039-8.

115. Sundaram, S.; Kim, H. J.; Bauer, R.; Thongkongkaew, T.; Heine, D.; Hertweck, C. On-Line Polyketide Cyclization into Diverse Medium-Sized Lactones by a Specialized Ketosynthase Domain. *Angewandte Chemie International Edition* **2018,** *57* (35), 11223–11227. DOI: 10.1002/anie.201804991.

116. Richter, I.; Radosa, S.; Cseresnyés, Z.; Ferling, I.; Büttner, H.; Niehs, S. P.; Gerst, R.; Scherlach, K.; Figge, M. T.; Hillmann, F.; Hertweck, C. Toxin-Producing Endosymbionts Shield Pathogenic Fungus against Micropredators. *mBio* **2022,** *13* (5), e0144022. DOI: 10.1128/mbio.01440-22.

117. Kusebauch, B.; Busch, B.; Scherlach, K.; Roth, M.; Hertweck, C. Polyketide-chain branching by an enzymatic Michael addition. *Angewandte Chemie International Edition* **2009,** *48* (27), 5001–5004. DOI: 10.1002/anie.200900277.

118. Prota, A. E.; Bargsten, K.; Diaz, J. F.; Marsh, M.; Cuevas, C.; Liniger, M.; Neuhaus, C.; Andreu, J. M.; Altmann, K.-H.; Steinmetz, M. O. A new tubulin-binding site and pharmacophore for microtubule-destabilizing anticancer drugs. *Proceedings of the National Academy of Sciences of the United States of America* **2014,** *111* (38), 13817–13821. DOI: 10.1073/pnas.1408124111.

119. Shi, C.; Miller, B. R.; Alexander, E. M.; Gulick, A. M.; Aldrich, C. C. Design, Synthesis, and Biophysical Evaluation of Mechanism-Based Probes for Condensation Domains of Nonribosomal Peptide Synthetases. *ACS Chemical Biology* **2020,** *15* (7), 1813–1819. DOI: 10.1021/acschembio.0c00411.

120. Zhong, L.; Diao, X.; Zhang, N.; Li, F.; Zhou, H.; Chen, H.; Bai, X.; Ren, X.; Zhang, Y.; Wu, D.; Bian, X. Engineering and elucidation of the lipoinitiation process in nonribosomal peptide biosynthesis. *Nature Communications* **2021,** *12* (1), 296. DOI: 10.1038/s41467-020-20548-8.

121. Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. Therapeutic peptides: current applications and future directions. *Signal transduction and targeted therapy* **2022,** *7* (1), 48. DOI: 10.1038/s41392-022-00904-4.

122. Duban, M.; Cociancich, S.; Leclère, V. Nonribosomal Peptide Synthesis Definitely Working Out of the Rules. *Microorganisms* **2022,** *10* (3). DOI: 10.3390/microorganisms10030577.

123. Rüschenbaum, J.; Steinchen, W.; Mayerthaler, F.; Feldberg, A.-L.; Mootz, H. D. FRET Monitoring of a Nonribosomal Peptide Synthetase Elongation Module Reveals Carrier Protein Shuttling between Catalytic Domains. *Angewandte Chemie International Edition* **2022,** *61* (48), e202212994. DOI: 10.1002/anie.202212994.

124. Davison, J.; Dorival, J.; Rabeharindranto, H.; Mazon, H.; Chagot, B.; Gruez, A.; Weissman, K. J. Insights into the function of trans-acyl transferase polyketide synthases from the SAXS

structure of a complete module. *Chemical science* **2014,** *5* (8), 3081–3095. DOI: 10.1039/C3SC53511H.

125. Bunnak, W.; Winter, A. J.; Lazarus, C. M.; Crump, M. P.; Race, P. R.; Wattana-Amorn, P. SAXS reveals highly flexible interdomain linkers of tandem acyl carrier protein-thioesterase domains from a fungal nonreducing polyketide synthase. *FEBS letters* **2021,** *595* (1), 133–144. DOI: 10.1002/1873-3468.13954.

126. Wenski, S. L.; Thiengmag, S.; Helfrich, E. J. N. Complex peptide natural products: Biosynthetic principles, challenges and opportunities for pathway engineering. *Synthetic and systems biotechnology* **2022,** *7* (1), 631–647. DOI: 10.1016/j.synbio.2022.01.007.

127. Neves, R. P. P.; Ferreira, P.; Medina, F. E.; Paiva, P.; Sousa, J. P. M.; Viegas, M. F.; Fernandes, P. A.; Ramos, M. J. Engineering of PKS Megaenzymes—A Promising Way to Biosynthesize High-Value Active Molecules. *Topics in Catalysis* **2022,** *65* (1-4), 544–562. DOI: 10.1007/s11244-021-01490-5.

128. Izaguirre, G.; Hansen, J. N. Use of alkaline phosphatase as a reporter polypeptide to study the role of the subtilin leader segment and the SpaT transporter in the posttranslational modifications and secretion of subtilin in Bacillus subtilis 168. *Applied and environmental microbiology* **1997,** *63* (10), 3965–3971. DOI: 10.1128/aem.63.10.3965-3971.1997.

129. Zhang, Q.; Kobras, C. M.; Gebhard, S.; Mascher, T.; Wolf, D. Regulation of heterologous subtilin production in Bacillus subtilis W168. *Microbial cell factories* **2022,** *21* (1), 57. DOI: 10.1186/s12934-022-01782-9.

130. Geiger, C.; Korn, S. M.; Häsler, M.; Peetz, O.; Martin, J.; Kötter, P.; Morgner, N.; Entian, K.-D. LanI-Mediated Lantibiotic Immunity in Bacillus subtilis: Functional Analysis. *Applied and environmental microbiology* **2019,** *85* (11). DOI: 10.1128/AEM.00534-19.

131. Orelle, C.; Mathieu, K.; Jault, J.-M. Multidrug ABC transporters in bacteria. *Research in Microbiology* **2019,** *170* (8), 381–391. DOI: 10.1016/j.resmic.2019.06.001.

132. Thomas, C.; Tampé, R. Structural and Mechanistic Principles of ABC Transporters. *Annual review of biochemistry* **2020,** *89*, 605–636. DOI: 10.1146/annurev-biochem-011520-105201.

133. Orelle, C.; Gubellini, F.; Durand, A.; Marco, S.; Lévy, D.; Gros, P.; Di Pietro, A.; Jault, J.-M. Conformational change induced by ATP binding in the multidrug ATP-binding cassette transporter BmrA. *Biochemistry* **2008,** *47* (8), 2404–2412. DOI: 10.1021/bi702303s.

134. Orelle, C.; Dalmas, O.; Gros, P.; Di Pietro, A.; Jault, J.-M. The conserved glutamate residue adjacent to the Walker-B motif is the catalytic base for ATP hydrolysis in the ATP-binding cassette transporter BmrA. *Journal of Biological Chemistry* **2003,** *278* (47), 47002–47008. DOI: 10.1074/jbc.M308268200.

135. Fetsch, E. E.; Davidson, A. L. Vanadate-catalyzed photocleavage of the signature motif of an ATP-binding cassette (ABC) transporter. *Proceedings of the National Academy of Sciences of the United States of America* **2002,** *99* (15), 9685–9690. DOI: 10.1073/pnas.152204499.

136. Jones, P. M.; George, A. M. Mechanism of ABC transporters: a molecular dynamics simulation of a well characterized nucleotide-binding subunit. *Proceedings of the National Academy of Sciences of the United States of America* **2002,** *99* (20), 12639–12644. DOI: 10.1073/pnas.152439599.

137. Kluth, M.; Stindt, J.; Dröge, C.; Linnemann, D.; Kubitz, R.; Schmitt, L. A mutation within the extended X loop abolished substrate-induced ATPase activity of the human liver ATP-binding cassette (ABC) transporter MDR3. *The Journal of biological chemistry* **2015,** *290* (8), 4896–4907. DOI: 10.1074/jbc.M114.588566.

138. Ford, R. C.; Hellmich, U. A. What monomeric nucleotide binding domains can teach us about dimeric ABC proteins. *FEBS letters* **2020,** *594* (23), 3857–3875. DOI: 10.1002/1873-3468.13921.

139. Lapinski, P. E.; Neubig, R. R.; Raghavan, M. Walker A lysine mutations of TAP1 and TAP2 interfere with peptide translocation but not peptide binding. *Journal of Biological Chemistry* **2001,** *276* (10), 7526–7533. DOI: 10.1074/jbc.M009448200.

140. Stockner, T.; Gradisch, R.; Schmitt, L. The role of the degenerate nucleotide binding site in type I ABC exporters. *FEBS letters* **2020,** *594* (23), 3815–3838. DOI: 10.1002/1873-3468.13997.

141. Buchaklian, A. H.; Klug, C. S. Characterization of the LSGGQ and H motifs from the Escherichia coli lipid A transporter MsbA. *Biochemistry* **2006,** *45* (41), 12539–12546. DOI: 10.1021/bi060830a.

142. Zaitseva, J.; Jenewein, S.; Jumpertz, T.; Holland, I. B.; Schmitt, L. H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. *The EMBO journal* **2005,** *24* (11), 1901–1910. DOI: 10.1038/sj.emboj.7600657.

143. Grossmann, N.; Vakkasoglu, A. S.; Hulpke, S.; Abele, R.; Gaudet, R.; Tampé, R. Mechanistic determinants of the directionality and energetics of active export by a heterodimeric ABC transporter. *Nature Communications* **2014,** *5*, 5419. DOI: 10.1038/ncomms6419.

144. Dawson, R. J. P.; Locher, K. P. Structure of a bacterial multidrug ABC transporter. *Nature* **2006,** *443* (7108), 180–185. DOI: 10.1038/nature05155.

145. Furuta, T. Structural dynamics of ABC transporters: molecular simulation studies. *Biochemical Society transactions* **2021,** *49* (1), 405–414. DOI: 10.1042/BST20200710.

146. Li, H.; Hu, D.; Liang, F.; Huang, X.; Zhu, Q. Influence factors on the critical micelle concentration determination using pyrene as a probe and a simple method of preparing samples. *Royal Society open science* **2020,** *7* (3), 192092. DOI: 10.1098/rsos.192092.

147. Kermani, A. A. A guide to membrane protein X-ray crystallography. *The FEBS journal* **2021,** *288* (20), 5788–5804. DOI: 10.1111/febs.15676.

148. Infed, N.; Hanekop, N.; Driessen, A. J. M.; Smits, S. H. J.; Schmitt, L. Influence of detergents on the activity of the ABC transporter LmrA. *Biochimica et biophysica acta* **2011,** *1808* (9), 2313–2321. DOI: 10.1016/j.bbamem.2011.05.016.

149. Guo, Y. Be Cautious with Crystal Structures of Membrane Proteins or Complexes Prepared in Detergents. *Crystals* **2020,** *10* (2). DOI: 10.3390/cryst10020086.

150. Alvarez, F. J. D.; Orelle, C.; Huang, Y.; Bajaj, R.; Everly, R. M.; Klug, C. S.; Davidson, A. L. Full engagement of liganded maltose-binding protein stabilizes a semi-open ATP-binding cassette dimer in the maltose transporter. *Molecular microbiology* **2015,** *98* (5), 878–894. DOI: 10.1111/mmi.13165.

151. Bolla, J. R.; Agasid, M. T.; Mehmood, S.; Robinson, C. V. Membrane Protein-Lipid Interactions Probed Using Mass Spectrometry. *Annual review of biochemistry* **2019,** *88*, 85–111. DOI: 10.1146/annurev-biochem-013118-111508.

152. Stieger, B.; Steiger, J.; Locher, K. P. Membrane lipids and transporter function. *Biochimica et biophysica acta. Molecular basis of disease* **2021,** *1867* (5), 166079. DOI: 10.1016/j.bbadis.2021.166079.

153. Neumann, J.; Rose-Sperling, D.; Hellmich, U. A. Diverse relations between ABC transporters and lipids: An overview. *Biochimica et biophysica acta. Biomembranes* **2017,** *1859* (4), 605–618. DOI: 10.1016/j.bbamem.2016.09.023.

154. Skrzypek, R.; Iqbal, S.; Callaghan, R. Methods of reconstitution to investigate membrane protein function. *Methods* **2018,** *147*, 126–141. DOI: 10.1016/j.ymeth.2018.02.012.

155. Farrelly, M. D.; Martin, L. L.; Thang, S. H. Polymer Nanodiscs and Their Bioanalytical Potential. *Chemistry* **2021,** *27* (51), 12922–12939. DOI: 10.1002/chem.202101572.

156. Sligar, S. G.; Denisov, I. G. Nanodiscs: A toolkit for membrane protein science. *Protein science : a publication of the Protein Society* **2021,** *30* (2), 297–315. DOI: 10.1002/pro.3994.

157. Popovic, K.; Holyoake, J.; Pomès, R.; Privé, G. G. Structure of saposin A lipoprotein discs. *Proceedings of the National Academy of Sciences of the United States of America* **2012,** *109* (8), 2908–2912. DOI: 10.1073/pnas.1115743109.

158. Yokogawa, M.; Fukuda, M.; Osawa, M. Nanodiscs for Structural Biology in a Membranous Environment. *Chemical & pharmaceutical bulletin* **2019,** *67* (4), 321–326. DOI: 10.1248/cpb.c18-00941.

159. Frauenfeld, J.; Löving, R.; Armache, J.-P.; Sonnen, A. F.-P.; Guettou, F.; Moberg, P.; Zhu, L.; Jegerschöld, C.; Flayhan, A.; Briggs, J. A. G.; Garoff, H.; Löw, C.; Cheng, Y.; Nordlund, P. A saposin-lipoprotein nanoparticle system for membrane proteins. *Nature methods* **2016,** *13* (4), 345–351. DOI: 10.1038/nmeth.3801.

160. Chien, C.-T. H.; Helfinger, L. R.; Bostock, M. J.; Solt, A.; Tan, Y. L.; Nietlispach, D. An Adaptable Phospholipid Membrane Mimetic System for Solution NMR Studies of Membrane Proteins. *Journal of the American Chemical Society* **2017,** *139* (42), 14829–14832. DOI: 10.1021/jacs.7b06730.

161. Flayhan, A.; Mertens, H. D. T.; Ural-Blimke, Y.; Martinez Molledo, M.; Svergun, D. I.; Löw, C. Saposin Lipid Nanoparticles: A Highly Versatile and Modular Tool for Membrane Protein Research. *Structure* **2018,** *26* (2), 345-355.e5. DOI: 10.1016/j.str.2018.01.007.

162. Nguyen, N. X.; Armache, J.-P.; Lee, C.; Yang, Y.; Zeng, W.; Mootha, V. K.; Cheng, Y.; Bai, X.-C.; Jiang, Y. Cryo-EM structure of a fungal mitochondrial calcium uniporter. *Nature* **2018,** *559* (7715), 570–574. DOI: 10.1038/s41586-018-0333-6.

163. Kanonenberg, K.; Smits, S. H. J.; Schmitt, L. Functional Reconstitution of HlyB, a Type I Secretion ABC Transporter, in Saposin-A Nanoparticles. *Scientific reports* **2019,** *9* (1), 8436. DOI: 10.1038/s41598-019-44812-0.

164. Kehlenbeck, D.-M.; Traore, D. A. K.; Josts, I.; Sander, S.; Moulin, M.; Haertlein, M.; Prevost, S.; Forsyth, V. T.; Tidow, H. Cryo-EM structure of MsbA in saposin-lipid nanoparticles (Salipro) provides insights into nucleotide coordination. *The FEBS journal* **2022,** *289* (10), 2959–2970. DOI: 10.1111/febs.16327.

165. Lloris-Garcerá, P.; Klinter, S.; Chen, L.; Skynner, M. J.; Löving, R.; Frauenfeld, J. DirectMX - One-Step Reconstitution of Membrane Proteins From Crude Cell Membranes Into Salipro Nanoparticles. *Frontiers in bioengineering and biotechnology* **2020,** *8*, 215. DOI: 10.3389/fbioe.2020.00215.

166. Sun, C.; Gennis, R. B. Single-particle cryo-EM studies of transmembrane proteins in SMA copolymer nanodiscs. *Chemistry and physics of lipids* **2019,** *221*, 114–119. DOI: 10.1016/j.chemphyslip.2019.03.007.

167. Yoder, N.; Gouaux, E. The His-Gly motif of acid-sensing ion channels resides in a reentrant 'loop' implicated in gating and ion selectivity. *eLife* **2020,** *9. DOI:* 10.7554/eLife.56527.

168. Yu, J.; Zhu, H.; Lape, R.; Greiner, T.; Du, J.; Lü, W.; Sivilotti, L.; Gouaux, E. Mechanism of gating and partial agonist action in the glycine receptor. *Cell* **2021,** *184* (4), 957-968.e21. DOI: 10.1016/j.cell.2021.01.026.

169. Ravula, T.; Hardin, N. Z.; Ramamoorthy, A. Polymer nanodiscs: Advantages and limitations. *Chemistry and physics of lipids* **2019,** *219*, 45–49. DOI: 10.1016/j.chemphyslip.2019.01.010.

170. Harding, B. D.; Dixit, G.; Burridge, K. M.; Sahu, I. D.; Dabney-Smith, C.; Edelmann, R. E.; Konkolewicz, D.; Lorigan, G. A. Characterizing the structure of styrene-maleic acid copolymer-lipid nanoparticles (SMALPs) using RAFT polymerization for membrane protein spectroscopic studies. *Chemistry and physics of lipids* **2019,** *218*, 65–72. DOI: 10.1016/j.chemphyslip.2018.12.002.

171. Oluwole, A. O.; Danielczak, B.; Meister, A.; Babalola, J. O.; Vargas, C.; Keller, S. Solubilization of Membrane Proteins into Functional Lipid-Bilayer Nanodiscs Using a Diisobutylene/Maleic Acid Copolymer. *Angewandte Chemie International Edition* **2017,** *56* (7), 1919–1924. DOI: 10.1002/anie.201610778.

172. Dominguez Pardo, J. J.; Dörr, J. M.; Iyer, A.; Cox, R. C.; Scheidelaar, S.; Koorengevel, M. C.; Subramaniam, V.; Killian, J. A. Solubilization of lipids and lipid phases by the styrene-maleic acid copolymer. *European biophysics journal : EBJ* **2017,** *46* (1), 91–101. DOI: 10.1007/s00249-016-1181-7.

173. Teo, A. C. K.; Lee, S. C.; Pollock, N. L.; Stroud, Z.; Hall, S.; Thakker, A.; Pitt, A. R.; Dafforn, T. R.; Spickett, C. M.; Roper, D. I. Analysis of SMALP co-extracted phospholipids shows distinct membrane environments for three classes of bacterial membrane protein. *Scientific reports* **2019,** *9* (1), 1813. DOI: 10.1038/s41598-018-37962-0.

174. Ravula, T.; Hardin, N. Z.; Di Mauro, G. M.; Ramamoorthy, A. Styrene maleic acid derivates to enhance the applications of bio-inspired polymer based lipid-nanodiscs. *European polymer journal* **2018,** *108*, 597–602. DOI: 10.1016/j.eurpolymj.2018.09.048.

175. Scheidelaar, S.; Koorengevel, M. C.; van Walree, C. A.; Dominguez, J. J.; Dörr, J. M.; Killian, J. A. Effect of Polymer Composition and pH on Membrane Solubilization by Styrene-Maleic Acid Copolymers. *Biophysical journal* **2016,** *111* (9), 1974–1986. DOI: 10.1016/j.bpj.2016.09.025.

176. Kopf, A. H.; Dörr, J. M.; Koorengevel, M. C.; Antoniciello, F.; Jahn, H.; Killian, J. A. Factors influencing the solubilization of membrane proteins from Escherichia coli membranes by styrene-maleic acid copolymers. *Biochimica et biophysica acta. Biomembranes* **2020,** *1862* (2), 183125. DOI: 10.1016/j.bbamem.2019.183125.

177. Fiori, M. C.; Jiang, Y.; Altenberg, G. A.; Liang, H. Polymer-encased nanodiscs with improved buffer compatibility. *Scientific reports* **2017,** *7* (1), 7432. DOI: 10.1038/s41598-017-07110-1.

178. Danielczak, B.; Meister, A.; Keller, S. Influence of Mg2+ and Ca2+ on nanodisc formation by diisobutylene/maleic acid (DIBMA) copolymer. *Chemistry and physics of lipids* **2019,** *221*, 30–38. DOI: 10.1016/j.chemphyslip.2019.03.004.

179. Brown, C. J.; Trieber, C.; Overduin, M. Structural biology of endogenous membrane protein assemblies in native nanodiscs. *Current opinion in structural biology* **2021,** *69*, 70–77. DOI: 10.1016/j.sbi.2021.03.008.

180. Günsel, U.; Hagn, F. Lipid Nanodiscs for High-Resolution NMR Studies of Membrane Proteins. *Chemical reviews* **2022,** *122* (10), 9395–9421. DOI: 10.1021/acs.chemrev.1c00702.

181. Trivedi, B. C.; Culbertson, B. M. *Maleic Anhydride;* Springer US, 1982. DOI: 10.1007/978-1-4757-0940-7.

182. Burridge, K. M.; Harding, B. D.; Sahu, I. D.; Kearns, M. M.; Stowe, R. B.; Dolan, M. T.; Edelmann, R. E.; Dabney-Smith, C.; Page, R. C.; Konkolewicz, D.; Lorigan, G. A. Simple Derivatization of RAFT-Synthesized Styrene-Maleic Anhydride Copolymers for Lipid Disk Formulations. *Biomacromolecules* **2020,** *21* (3), 1274–1284. DOI: 10.1021/acs.biomac.0c00041.

183. Ravula, T.; Hardin, N. Z.; Ramadugu, S. K.; Ramamoorthy, A. pH Tunable and Divalent Metal Ion Tolerant Polymer Lipid Nanodiscs. *Langmuir : the ACS journal of surfaces and colloids* **2017,** *33* (40), 10655–10662. DOI: 10.1021/acs.langmuir.7b02887.

184. Hall, S. C. L.; Tognoloni, C.; Charlton, J.; Bragginton, É. C.; Rothnie, A. J.; Sridhar, P.; Wheatley, M.; Knowles, T. J.; Arnold, T.; Edler, K. J.; Dafforn, T. R. An acid-compatible co-

polymer for the solubilization of membranes and proteins into lipid bilayer-containing nanoparticles. *Nanoscale* **2018,** *10* (22), 10609–10619. DOI: 10.1039/c8nr01322e.

185. Danielczak, B.; Rasche, M.; Lenz, J.; Pérez Patallo, E.; Weyrauch, S.; Mahler, F.; Agbadaola, M. T.; Meister, A.; Babalola, J. O.; Vargas, C.; Kolar, C.; Keller, S. A bioinspired glycopolymer for capturing membrane proteins in native-like lipid-bilayer nanodiscs. *Nanoscale* **2022,** *14* (5), 1855–1867. DOI: 10.1039/d1nr03811g.

186. Esmaili, M.; Acevedo-Morantes, C.; Wille, H.; Overduin, M. The effect of hydrophobic alkyl sidechains on size and solution behaviors of nanodiscs formed by alternating styrene maleamic copolymer. *Biochimica et biophysica acta. Biomembranes* **2020,** *1862* (10), 183360. DOI: 10.1016/j.bbamem.2020.183360.

187. Ravula, T.; Ramadugu, S. K.; Di Mauro, G.; Ramamoorthy, A. Bioinspired, Size-Tunable Self-Assembly of Polymer-Lipid Bilayer Nanodiscs. *Angewandte Chemie International Edition* **2017,** *56* (38), 11466–11470. DOI: 10.1002/anie.201705569.

188. Smith, A. A. A.; Autzen, H. E.; Laursen, T.; Wu, V.; Yen, M.; Hall, A.; Hansen, S. D.; Cheng, Y.; Xu, T. Controlling Styrene Maleic Acid Lipid Particles through RAFT. *Biomacromolecules* **2017,** *18* (11), 3706–3713. DOI: 10.1021/acs.biomac.7b01136.

189. Semsarilar, M.; Abetz, V. Polymerizations by RAFT: Developments of the Technique and Its Application in the Synthesis of Tailored (Co)polymers. *Macromolecular Chemistry and Physics* **2021,** *222* (1), 2000311. DOI: 10.1002/macp.202000311.

190. Moad, C. L.; Moad, G. Fundamentals of reversible addition–fragmentation chain transfer (RAFT). *Chemistry Teacher International* **2021,** *3* (2), 3–17. DOI: 10.1515/cti-2020-0026.

191. Craig, A. F.; Clark, E. E.; Sahu, I. D.; Zhang, R.; Frantz, N. D.; Al-Abdul-Wahid, M. S.; Dabney-Smith, C.; Konkolewicz, D.; Lorigan, G. A. Tuning the size of styrene-maleic acid copolymer-lipid nanoparticles (SMALPs) using RAFT polymerization for biophysical studies. *Biochimica et biophysica acta* **2016,** *1858* (11), 2931–2939. DOI: 10.1016/j.bbamem.2016.08.004.

192. Fiori, M. C.; Zheng, W.; Kamilar, E.; Simiyu, G.; Altenberg, G. A.; Liang, H. Extraction and reconstitution of membrane proteins into lipid nanodiscs encased by zwitterionic styrene-maleic amide copolymers. *Sci Rep* **2020,** *10* (1), 9940. DOI: 10.1038/s41598-020-66852-7.

193. Cunningham, R. D.; Kopf, A. H.; Elenbaas, B. O. W.; Staal, B. B. P.; Pfukwa, R.; Killian, J. A.; Klumperman, B. Iterative RAFT-Mediated Copolymerization of Styrene and Maleic Anhydride toward Sequence- and Length-Controlled Copolymers and Their Applications for Solubilizing Lipid Membranes. *Biomacromolecules* **2020,** *21* (8), 3287–3300. DOI: 10.1021/acs.biomac.0c00736.

194. Mathieu, K.; Javed, W.; Vallet, S.; Lesterlin, C.; Candusso, M.-P.; Ding, F.; Xu, X. N.; Ebel, C.; Jault, J.-M.; Orelle, C. Functionality of membrane proteins overexpressed and purified from E. coli is highly dependent upon the strain. *Scientific reports* **2019,** *9* (1), 2654. DOI: 10.1038/s41598-019-39382-0.

195. Chaptal, V.; Zampieri, V.; Wiseman, B.; Orelle, C.; Martin, J.; Nguyen, K.-A.; Gobet, A.; Di Cesare, M.; Magnard, S.; Javed, W.; Eid, J.; Kilburg, A.; Peuchmaur, M.; Marcoux, J.; Monticelli, L.; Hogbom, M.; Schoehn, G.; Jault, J.-M.; Boumendjel, A.; Falson, P. Substrate-bound and substrate-free outward-facing structures of a multidrug ABC exporter. *Science advances* **2022,** *8* (4), eabg9215. DOI: 10.1126/sciadv.abg9215.

196. Gulamhussein, A. A.; Uddin, R.; Tighe, B. J.; Poyner, D. R.; Rothnie, A. J. A comparison of SMA (styrene maleic acid) and DIBMA (di-isobutylene maleic acid) for membrane protein purification. *Biochimica et biophysica acta. Biomembranes* **2020,** *1862* (7), 183281. DOI: 10.1016/j.bbamem.2020.183281.

197. Juan-Carlos, P.-D. M.; Perla-Lidia, P.-P.; Stephanie-Talia, M.-M.; Mónica-Griselda, A.-M.; Luz-María, T.-E. ABC transporter superfamily. An updated overview, relevance in cancer multidrug resistance and perspectives with personalized medicine. *Molecular biology reports* **2021,** *48* (2), 1883–1901. DOI: 10.1007/s11033-021-06155-w.

198. Mi, W.; Li, Y.; Yoon, S. H.; Ernst, R. K.; Walz, T.; Liao, M. Structural basis of MsbA-mediated lipopolysaccharide transport. *Nature* **2017,** *549* (7671), 233–237. DOI: 10.1038/nature23649.

199. Arana, M. R.; Fiori, M. C.; Altenberg, G. A. Functional and structural comparison of the ABC exporter MsbA studied in detergent and reconstituted in nanodiscs. *Biochemical and biophysical research communications* **2019,** *512* (3), 448–452. DOI: 10.1016/j.bbrc.2019.03.069.

200. Dong, M.; Penin, F.; Baggetto, L. G. Efficient purification and reconstitution of P-glycoprotein for functional and structural studies. *Journal of Biological Chemistry* **1996,** *271* (46), 28875–28883. DOI: 10.1074/jbc.271.46.28875.

201. Wright, G. S. A.; Saeki, A.; Hikima, T.; Nishizono, Y.; Hisano, T.; Kamaya, M.; Nukina, K.; Nishitani, H.; Nakamura, H.; Yamamoto, M.; Antonyuk, S. V.; Hasnain, S. S.; Shiro, Y.; Sawai, H. Architecture of the complete oxygen-sensing FixL-FixJ two-component signal transduction system. *Science signaling* **2018,** *11* (525). DOI: 10.1126/scisignal.aaq0825.

202. Frey, L.; Lakomek, N.-A.; Riek, R.; Bibow, S. Micelles, Bicelles, and Nanodiscs: Comparing the Impact of Membrane Mimetics on Membrane Protein Backbone Dynamics. *Angewandte Chemie International Edition* **2017,** *56* (1), 380–383. DOI: 10.1002/anie.201608246.

203. Bonifer, C.; Glaubitz, C. MsbA: an ABC transporter paradigm. *Biochemical Society transactions* **2021,** *49* (6), 2917–2927. DOI: 10.1042/BST20211030.

204. Stroud, Z.; Hall, S. C. L.; Dafforn, T. R. Purification of membrane proteins free from conventional detergents: SMA, new polymers, new opportunities and new insights. *Methods* **2018,** *147*, 106–117. DOI: 10.1016/j.ymeth.2018.03.011.

205. Unger, L.; Ronco-Campaña, A.; Kitchen, P.; Bill, R. M.; Rothnie, A. J. Biological insights from SMA-extracted proteins. *Biochemical Society transactions* **2021,** *49* (3), 1349–1359. DOI: 10.1042/BST20201067.

206. Krishnarjuna, B.; Ramamoorthy, A. Detergent-Free Isolation of Membrane Proteins and Strategies to Study Them in a Near-Native Membrane Environment. *Biomolecules* **2022,** *12* (8). DOI: 10.3390/biom12081076.

207. Jeffries, C. M.; Graewert, M. A.; Blanchet, C. E.; Langley, D. B.; Whitten, A. E.; Svergun, D. I. Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. *Nature protocols* **2016,** *11* (11), 2122–2153. DOI: 10.1038/nprot.2016.113.

208. Swainsbury, D. J. K.; Scheidelaar, S.; Foster, N.; van Grondelle, R.; Killian, J. A.; Jones, M. R. The effectiveness of styrene-maleic acid (SMA) copolymers for solubilisation of integral membrane proteins from SMA-accessible and SMA-resistant membranes. *Biochimica et biophysica acta. Biomembranes* **2017,** *1859* (10), 2133–2143. DOI: 10.1016/j.bbamem.2017.07.011.

209. Domínguez Pardo, J. J.; Koorengevel, M. C.; Uwugiaren, N.; Weijers, J.; Kopf, A. H.; Jahn, H.; van Walree, C. A.; van Steenbergen, M. J.; Killian, J. A. Membrane Solubilization by Styrene-Maleic Acid Copolymers: Delineating the Role of Polymer Length. *Biophysical journal* **2018,** *115* (1), 129–138. DOI: 10.1016/j.bpj.2018.05.032.

210. Ball, L. E.; Riley, L. J.; Hadasha, W.; Pfukwa, R.; Smith, C. J. I.; Dafforn, T. R.; Klumperman, B. Influence of DIBMA Polymer Length on Lipid Nanodisc Formation and Membrane Protein Extraction. *Biomacromolecules* **2021,** *22* (2), 763–772. DOI: 10.1021/acs.biomac.0c01538.

211. Wiseman, B.; Kilburg, A.; Chaptal, V.; Reyes-Mejia, G. C.; Sarwan, J.; Falson, P.; Jault, J.-M. Stubborn contaminants: influence of detergents on the purity of the multidrug ABC transporter BmrA. *PLOS ONE* **2014,** *9* (12), e114864. DOI: 10.1371/journal.pone.0114864.

212. Steinfels, E.; Orelle, C.; Fantino, J.-R.; Dalmas, O.; Rigaud, J.-L.; Denizot, F.; Di Pietro, A.; Jault, J.-M. Characterization of YvcC (BmrA), a multidrug ABC transporter constitutively expressed in Bacillus subtilis. *Biochemistry* **2004,** *43* (23), 7491–7502. DOI: 10.1021/bi0362018.

213. Fribourg, P. F.; Chami, M.; Sorzano, C. O. S.; Gubellini, F.; Marabini, R.; Marco, S.; Jault, J.-M.; Lévy, D. 3D cryo-electron reconstruction of BmrA, a bacterial multidrug ABC transporter in an inward-facing conformation and in a lipidic environment. *Journal of molecular biology* **2014,** *426* (10), 2059–2069. DOI: 10.1016/j.jmb.2014.03.002.

214. Waqas Javed. *Study of the conformational states of a bacterial multidrug ABC transporter BmrA. Structural Biology*, 2020.

215. Reading, E.; Hall, Z.; Martens, C.; Haghighi, T.; Findlay, H.; Ahdash, Z.; Politis, A.; Booth, P. J. Interrogating Membrane Protein Conformational Dynamics within Native Lipid Compositions. *Angewandte Chemie International Edition* **2017,** *56* (49), 15654–15657. DOI: 10.1002/anie.201709657.

216. Kunert, B.; Gardiennet, C.; Lacabanne, D.; Calles-Garcia, D.; Falson, P.; Jault, J.-M.; Meier, B. H.; Penin, F.; Böckmann, A. Efficient and stable reconstitution of the ABC transporter BmrA

for solid-state NMR studies. *Frontiers in molecular biosciences* **2014,** *1*, 5. DOI: 10.3389/fmolb.2014.00005.

217. Hofmann, S.; Januliene, D.; Mehdipour, A. R.; Thomas, C.; Stefan, E.; Brüchert, S.; Kuhn, B. T.; Geertsma, E. R.; Hummer, G.; Tampé, R.; Moeller, A. Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature* **2019,** *571* (7766), 580–583. DOI: 10.1038/s41586-019-1391-0.

218. Lacabanne, D.; Wiegand, T.; Di Cesare, M.; Orelle, C.; Ernst, M.; Jault, J.-M.; Meier, B. H.; Böckmann, A. Solid-State NMR Reveals Asymmetric ATP Hydrolysis in the Multidrug ABC Transporter BmrA. *Journal of the American Chemical Society* **2022,** *144* (27), 12431–12442. DOI: 10.1021/jacs.2c04287.

219. Lacabanne, D.; Orelle, C.; Lecoq, L.; Kunert, B.; Chuilon, C.; Wiegand, T.; Ravaud, S.; Jault, J.-M.; Meier, B. H.; Böckmann, A. Flexible-to-rigid transition is central for substrate transport in the ABC transporter BmrA from Bacillus subtilis. *Communications biology* **2019,** *2*, 149. DOI: 10.1038/s42003-019-0390-x.

220. Klöpfer, K.; Hagn, F. Beyond detergent micelles: The advantages and applications of non-micellar and lipid-based membrane mimetics for solution-state NMR. *Progress in nuclear magnetic resonance spectroscopy* **2019,** *114-115*, 271–283. DOI: 10.1016/j.pnmrs.2019.08.001.

221. Stump, B. Click Bioconjugation: Modifying Proteins Using Click-Like Chemistry. *ChemBioChem* **2022,** *23* (16), e202200016. DOI: 10.1002/cbic.202200016.

222. Renault, K.; Fredy, J. W.; Renard, P.-Y.; Sabot, C. Covalent Modification of Biomolecules through Maleimide-Based Labeling Strategies. *Bioconjugate chemistry* **2018,** *29* (8), 2497–2513. DOI: 10.1021/acs.bioconjchem.8b00252.

223. Roth, P. J.; Boyer, C.; Lowe, A. B.; Davis, T. P. RAFT polymerization and thiol chemistry: a complementary pairing for implementing modern macromolecular design. *Macromolecular rapid communications* **2011,** *32* (15), 1123–1143. DOI: 10.1002/marc.201100127.

224. Lindhoud, S.; Carvalho, V.; Pronk, J. W.; Aubin-Tam, M.-E. SMA-SH: Modified Styrene-Maleic Acid Copolymer for Functionalization of Lipid Nanodiscs. *Biomacromolecules* **2016,** *17* (4), 1516–1522. DOI: 10.1021/acs.biomac.6b00140.

225. Hirakawa, H.; Kurushima, J.; Hashimoto, Y.; Tomita, H. Progress Overview of Bacterial Two-Component Regulatory Systems as Potential Targets for Antimicrobial Chemotherapy. *Antibiotics* **2020,** *9* (10). DOI: 10.3390/antibiotics9100635.

226. Jacob-Dubuisson, F.; Mechaly, A.; Betton, J.-M.; Antoine, R. Structural insights into the signalling mechanisms of two-component systems. *Nature Reviews Microbiology* **2018,** *16* (10), 585–593. DOI: 10.1038/s41579-018-0055-7.

227. Klein, C.; Kaletta, C.; Entian, K. D. Biosynthesis of the lantibiotic subtilin is regulated by a histidine kinase/response regulator system. *Applied and environmental microbiology* **1993,** *59* (1), 296–303. DOI: 10.1128/aem.59.1.296-303.1993.

228. Stein, T.; Heinzmann, S.; Düsterhus, S.; Borchert, S.; Entian, K.-D. Expression and functional analysis of the subtilin immunity genes spaIFEG in the subtilin-sensitive host Bacillus subtilis MO1099. *Journal of bacteriology* **2005,** *187* (3), 822–828. DOI: 10.1128/JB.187.3.822-828.2005.

229. Stein, T. Bacillus subtilis antibiotics: structures, syntheses and specific functions. *Molecular microbiology* **2005,** *56* (4), 845–857. DOI: 10.1111/j.1365-2958.2005.04587.x.

230. Kiesau, P.; Eikmanns, U.; Gutowski-Eckel, Z.; Weber, S.; Hammelmann, M.; Entian, K. D. Evidence for a multimeric subtilin synthetase complex. *Journal of bacteriology* **1997,** *179* (5), 1475–1481. DOI: 10.1128/jb.179.5.1475-1481.1997.

231. Chakicherla, A.; Ecale Zhou, C. L.; Dang, M. L.; Rodriguez, V.; Hansen, J. N.; Zemla, A. SpaK/SpaR two-component system characterized by a structure-driven domain-fusion method and in vitro phosphorylation studies. *PLoS computational biology* **2009,** *5* (6), e1000401. DOI: 10.1371/journal.pcbi.1000401.

232. Stein, T.; Heinzmann, S.; Kiesau, P.; Himmel, B.; Entian, K.-D. The spa-box for transcriptional activation of subtilin biosynthesis and immunity in Bacillus subtilis. *Molecular microbiology* **2003,** *47* (6), 1627–1636. DOI: 10.1046/j.1365-2958.2003.03374.x.

233. Macek, B.; Forchhammer, K.; Hardouin, J.; Weber-Ban, E.; Grangeasse, C.; Mijakovic, I. Protein post-translational modifications in bacteria. *Nature Reviews Microbiology* **2019,** *17* (11), 651–664. DOI: 10.1038/s41579-019-0243-0.

234. Liu, C.; Di Sun; Zhu, J.; Liu, W. Two-Component Signal Transduction Systems: A Major Strategy for Connecting Input Stimuli to Biofilm Formation. *Frontiers in microbiology* **2018,** *9*, 3279. DOI: 10.3389/fmicb.2018.03279.

235. Shaw, C.; Hess, M.; Weimer, B. C. Two-component systems regulate bacterial virulence in response to the host gastrointestinal environment and metabolic cues. *Virulence* **2022,** *13* (1), 1666–1680. DOI: 10.1080/21505594.2022.2127196.

236. Gushchin, I.; Gordeliy, V. Transmembrane Signal Transduction in Two-Component Systems: Piston, Scissoring, or Helical Rotation? *BioEssays : news and reviews in molecular, cellular and developmental biology* **2018,** *40* (2). DOI: 10.1002/bies.201700197.

237. Buschiazzo, A.; Trajtenberg, F. Two-Component Sensing and Regulation: How Do Histidine Kinases Talk with Response Regulators at the Molecular Level? *Annual review of microbiology* **2019,** *73*, 507–528. DOI: 10.1146/annurev-micro-091018-054627.

238. Ishii, E.; Eguchi, Y. Diversity in Sensing and Signaling of Bacterial Sensor Histidine Kinases. *Biomolecules* **2021,** *11* (10). DOI: 10.3390/biom11101524.

239. Ma, P.; Phillips-Jones, M. K. Membrane Sensor Histidine Kinases: Insights from Structural, Ligand and Inhibitor Studies of Full-Length Proteins and Signalling Domains for Antibiotic Discovery. *Molecules* **2021,** *26* (16). DOI: 10.3390/molecules26165110.

240. Upadhyay, A. A.; Fleetwood, A. D.; Adebali, O.; Finn, R. D.; Zhulin, I. B. Cache Domains That are Homologous to, but Different from PAS Domains Comprise the Largest Superfamily of

Extracellular Sensors in Prokaryotes. *PLoS computational biology* **2016,** *12* (4), e1004862. DOI: 10.1371/journal.pcbi.1004862.

241. Stuffle, E. C.; Johnson, M. S.; Watts, K. J. PAS domains in bacterial signal transduction. *Current opinion in microbiology* **2021,** *61*, 8–15. DOI: 10.1016/j.mib.2021.01.004.

242. Gushchin, I.; Aleksenko, V. A.; Orekhov, P.; Goncharov, I. M.; Nazarenko, V. V.; Semenov, O.; Remeeva, A.; Gordeliy, V. Nitrate- and Nitrite-Sensing Histidine Kinases: Function, Structure, and Natural Diversity. *International journal of molecular sciences* **2021,** *22* (11). DOI: 10.3390/ijms22115933.

243. Moore, J. O.; Hendrickson, W. A. An asymmetry-to-symmetry switch in signal transmission by the histidine kinase receptor for TMAO. *Structure* **2012,** *20* (4), 729–741. DOI: 10.1016/j.str.2012.02.021.

244. Moore, J. O.; Hendrickson, W. A. Structural analysis of sensor domains from the TMAO-responsive histidine kinase receptor TorS. *Structure* **2009,** *17* (9), 1195–1204. DOI: 10.1016/j.str.2009.07.015.

245. Pineda-Molina, E.; Reyes-Darias, J.-A.; Lacal, J.; Ramos, J. L.; García-Ruiz, J. M.; Gavira, J. A.; Krell, T. Evidence for chemoreceptors with bimodular ligand-binding regions harboring two signal-binding sites. *Proceedings of the National Academy of Sciences of the United States of America* **2012,** *109* (46), 18926–18931. DOI: 10.1073/pnas.1201400109.

246. Stopp, M.; Steinmetz, P. A.; Schubert, C.; Griesinger, C.; Schneider, D.; Unden, G. Transmembrane signaling and cytoplasmic signal conversion by dimeric transmembrane helix 2 and a linker domain of the DcuS sensor kinase. *The Journal of biological chemistry* **2021,** *296*, 100148. DOI: 10.1074/jbc.RA120.015999.

247. Monzel, C.; Unden, G. Transmembrane signaling in the sensor kinase DcuS of Escherichia coli: A long-range piston-type displacement of transmembrane helix 2. *Proceedings of the National Academy of Sciences of the United States of America* **2015,** *112* (35), 11042–11047. DOI: 10.1073/pnas.1507217112.

248. Salvi, M.; Schomburg, B.; Giller, K.; Graf, S.; Unden, G.; Becker, S.; Lange, A.; Griesinger, C. Sensory domain contraction in histidine kinase CitA triggers transmembrane signaling in the membrane-bound sensor. *Proceedings of the National Academy of Sciences of the United States of America* **2017,** *114* (12), 3115–3120. DOI: 10.1073/pnas.1620286114.

249. Lesne, E.; Dupré, E.; Locht, C.; Antoine, R.; Jacob-Dubuisson, F. Conformational Changes of an Interdomain Linker Mediate Mechanical Signal Transmission in Sensor Kinase BvgS. *Journal of bacteriology* **2017,** *199* (18). DOI: 10.1128/JB.00114-17.

250. Tan, K.; Teschler, J. K.; Wu, R.; Jedrzejczak, R. P.; Zhou, M.; Shuvalova, L. A.; Endres, M. J.; Welk, L. F.; Kwon, K.; Anderson, W. F.; Satchell, K. J. F.; Yildiz, F. H.; Joachimiak, A. Sensor Domain of Histidine Kinase VxrA of Vibrio cholerae- A Hairpin-swapped Dimer and its Conformational Change. *Journal of bacteriology* **2021,** *203* (11). DOI: 10.1128/JB.00643-20.

251. Molnar, K. S.; Bonomi, M.; Pellarin, R.; Clinthorne, G. D.; Gonzalez, G.; Goldberg, S. D.; Goulian, M.; Sali, A.; DeGrado, W. F. Cys-scanning disulfide crosslinking and bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ. *Structure* **2014,** *22* (9), 1239–1251. DOI: 10.1016/j.str.2014.04.019.

252. Neiditch, M. B.; Federle, M. J.; Pompeani, A. J.; Kelly, R. C.; Swem, D. L.; Jeffrey, P. D.; Bassler, B. L.; Hughson, F. M. Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. *Cell* **2006,** *126* (6), 1095–1108. DOI: 10.1016/j.cell.2006.07.032.

253. Saita, E.; Abriata, L. A.; Tsai, Y. T.; Trajtenberg, F.; Lemmin, T.; Buschiazzo, A.; Dal Peraro, M.; Mendoza, D. de; Albanesi, D. A coiled coil switch mediates cold sensing by the thermosensory protein DesK. *Molecular microbiology* **2015,** *98* (2), 258–271. DOI: 10.1111/mmi.13118.

254. Wang, B.; Zhao, A.; Novick, R. P.; Muir, T. W. Activation and inhibition of the receptor histidine kinase AgrC occurs through opposite helical transduction motions. *Molecular cell* **2014,** *53* (6), 929–940. DOI: 10.1016/j.molcel.2014.02.029.

255. Alvarez, A. F.; Georgellis, D. The role of sensory kinase proteins in two-component signal transduction. *Biochemical Society transactions* **2022. DOI:** 10.1042/BST20220848.

256. Airola, M. V.; Sukomon, N.; Samanta, D.; Borbat, P. P.; Freed, J. H.; Watts, K. J.; Crane, B. R. HAMP domain conformers that propagate opposite signals in bacterial chemoreceptors. *PLoS biology* **2013,** *11* (2), e1001479. DOI: 10.1371/journal.pbio.1001479.

257. Zschiedrich, C. P.; Keidel, V.; Szurmant, H. Molecular Mechanisms of Two-Component Signal Transduction. *Journal of molecular biology* **2016,** *428* (19), 3752–3775. DOI: 10.1016/j.jmb.2016.08.003.

258. Diensthuber, R. P.; Bommer, M.; Gleichmann, T.; Möglich, A. Full-length structure of a sensor histidine kinase pinpoints coaxial coiled coils as signal transducers and modulators. *Structure* **2013,** *21* (7), 1127–1136. DOI: 10.1016/j.str.2013.04.024.

259. Garcia, D.; Watts, K. J.; Johnson, M. S.; Taylor, B. L. Delineating PAS-HAMP interaction surfaces and signalling-associated changes in the aerotaxis receptor Aer. *Molecular microbiology* **2016,** *100* (1), 156–172. DOI: 10.1111/mmi.13308.

260. Wang, C.; Sang, J.; Wang, J.; Su, M.; Downey, J. S.; Wu, Q.; Wang, S.; Cai, Y.; Xu, X.; Wu, J.; Senadheera, D. B.; Cvitkovitch, D. G.; Chen, L.; Goodman, S. D.; Han, A. Mechanistic insights revealed by the crystal structure of a histidine kinase with signal transducer and sensor domains. *PLoS biology* **2013,** *11* (2), e1001493. DOI: 10.1371/journal.pbio.1001493.

261. Aravind, L.; Ponting, C. P. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS microbiology letters* **1999,** *176* (1), 111–116. DOI: 10.1111/j.1574-6968.1999.tb13650.x.

262. Mensa, B.; Polizzi, N. F.; Molnar, K. S.; Natale, A. M.; Lemmin, T.; DeGrado, W. F. Allosteric mechanism of signal transduction in the two-component system histidine kinase PhoQ. *eLife* **2021,** *10. DOI:* 10.7554/eLife.73336.

263. Matamouros, S.; Hager, K. R.; Miller, S. I. HAMP Domain Rotation and Tilting Movements Associated with Signal Transduction in the PhoQ Sensor Kinase. *mBio* **2015,** *6* (3), e00616-15. DOI: 10.1128/mBio.00616-15.

264. Miller, S. I.; Mekalanos, J. J. Constitutive expression of the phoP regulon attenuates Salmonella virulence and survival within macrophages. *Journal of bacteriology* **1990,** *172* (5), 2485–2490. DOI: 10.1128/jb.172.5.2485-2490.1990.

265. Trajtenberg, F.; Graña, M.; Ruétalo, N.; Botti, H.; Buschiazzo, A. Structural and enzymatic insights into the ATP binding and autophosphorylation mechanism of a sensor histidine kinase. *The Journal of biological chemistry* **2010,** *285* (32), 24892–24903. DOI: 10.1074/jbc.M110.147843.

266. Tanaka, T.; Saha, S. K.; Tomomori, C.; Ishima, R.; Liu, D.; Tong, K. I.; Park, H.; Dutta, R.; Qin, L.; Swindells, M. B.; Yamazaki, T.; Ono, A. M.; Kainosho, M.; Inouye, M.; Ikura, M. NMR structure of the histidine kinase domain of the E. coli osmosensor EnvZ. *Nature* **1998,** *396* (6706), 88–92. DOI: 10.1038/23968.

267. Kim, D.-J.; Forst, S. Genomic analysis of the histidine kinase family in bacteria and archaea. *Microbiology* **2001,** *147* (Pt 5), 1197–1212. DOI: 10.1099/00221287-147-5-1197.

268. Celikel, R.; Veldore, V. H.; Mathews, I.; Devine, K. M.; Varughese, K. I. ATP forms a stable complex with the essential histidine kinase WalK (YycG) domain. *Acta crystallographica. Section D, Biological crystallography* **2012,** *68* (Pt 7), 839–845. DOI: 10.1107/S090744491201373X.

269. Willett, J. W.; Kirby, J. R. Genetic and biochemical dissection of a HisKA domain identifies residues required exclusively for kinase and phosphatase activities. *PLoS genetics* **2012,** *8* (11), e1003084. DOI: 10.1371/journal.pgen.1003084.

270. Casino, P.; Miguel-Romero, L.; Marina, A. Visualizing autophosphorylation in histidine kinases. *Nat Commun* **2014,** *5*, 3258. DOI: 10.1038/ncomms4258.

271. Cai, Y.; Su, M.; Ahmad, A.; Hu, X.; Sang, J.; Kong, L.; Chen, X.; Wang, C.; Shuai, J.; Han, A. Conformational dynamics of the essential sensor histidine kinase WalK. *Acta crystallographica. Section D, Structural biology* **2017,** *73* (Pt 10), 793–803. DOI: 10.1107/S2059798317013043.

272. Rinaldi, J.; Fernández, I.; Shin, H.; Sycz, G.; Gunawardana, S.; Kumarapperuma, I.; Paz, J. M.; Otero, L. H.; Cerutti, M. L.; Zorreguieta, Á.; Ren, Z.; Klinke, S.; Yang, X.; Goldbaum, F. A. Dimer Asymmetry and Light Activation Mechanism in Brucella Blue-Light Sensor Histidine Kinase. *mBio* **2021,** *12* (2). DOI: 10.1128/mBio.00264-21.

273. Mechaly, A. E.; Sassoon, N.; Betton, J.-M.; Alzari, P. M. Segmental helical motions and dynamical asymmetry modulate histidine kinase autophosphorylation. *PLoS biology* **2014,** *12* (1), e1001776. DOI: 10.1371/journal.pbio.1001776.

274. Ferris, H. U.; Coles, M.; Lupas, A. N.; Hartmann, M. D. Crystallographic snapshot of the Escherichia coli EnvZ histidine kinase in an active conformation. *Journal of structural biology* **2014,** *186* (3), 376–379. DOI: 10.1016/j.jsb.2014.03.014.

275. Trajtenberg, F.; Imelio, J. A.; Machado, M. R.; Larrieux, N.; Marti, M. A.; Obal, G.; Mechaly, A. E.; Buschiazzo, A. Regulation of signaling directionality revealed by 3D snapshots of a kinase:regulator complex in action. *eLife* **2016,** *5. DOI:* 10.7554/eLife.21422.

276. Ashenberg, O.; Keating, A. E.; Laub, M. T. Helix bundle loops determine whether histidine kinases autophosphorylate in cis or in trans. *Journal of molecular biology* **2013,** *425* (7), 1198–1209. DOI: 10.1016/j.jmb.2013.01.011.

277. Mechaly, A. E.; Soto Diaz, S.; Sassoon, N.; Buschiazzo, A.; Betton, J.-M.; Alzari, P. M. Structural Coupling between Autokinase and Phosphotransferase Reactions in a Bacterial Histidine Kinase. *Structure* **2017,** *25* (6), 939-944.e3. DOI: 10.1016/j.str.2017.04.011.

278. Gao, R.; Bouillet, S.; Stock, A. M. Structural Basis of Response Regulator Function. *Annual review of microbiology* **2019,** *73*, 175–197. DOI: 10.1146/annurev-micro-020518-115931.

279. Yan, D.; Cho, H. S.; Hastings, C. A.; Igo, M. M.; Lee, S. Y.; Pelton, J. G.; Stewart, V.; Wemmer, D. E.; Kustu, S. Beryllofluoride mimics phosphorylation of NtrC and other bacterial response regulators. *Proceedings of the National Academy of Sciences of the United States of America* **1999,** *96* (26), 14789–14794. DOI: 10.1073/pnas.96.26.14789.

280. Leonard, P. G.; Golemi-Kotra, D.; Stock, A. M. Phosphorylation-dependent conformational changes and domain rearrangements in Staphylococcus aureus VraR activation. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110* (21), 8525–8530. DOI: 10.1073/pnas.1302819110.

281. Baikalov, I.; Schröder, I.; Kaczor-Grzeskowiak, M.; Grzeskowiak, K.; Gunsalus, R. P.; Dickerson, R. E. Structure of the Escherichia coli response regulator NarL. *Biochemistry* **1996,** *35* (34), 11053–11061. DOI: 10.1021/bi960919o.

282. He, X.; Wang, L.; Wang, S. Structural basis of DNA sequence recognition by the response regulator PhoP in Mycobacterium tuberculosis. *Scientific reports* **2016,** *6*, 24442. DOI: 10.1038/srep24442.

283. Geiger, C.; Spieß, T.; Korn, S. M.; Kötter, P.; Entian, K.-D. Specificity of Subtilin-Mediated Activation of Histidine Kinase SpaK. *Applied and environmental microbiology* **2017,** *83* (18). DOI: 10.1128/AEM.00781-17.

284. Spieß, T.; Korn, S. M.; Kötter, P.; Entian, K.-D. Activation of Histidine Kinase SpaK Is Mediated by the N-Terminal Portion of Subtilin-Like Lantibiotics and Is Independent of Lipid II. *Applied and environmental microbiology* **2015,** *81* (16), 5335–5343. DOI: 10.1128/AEM.01368-15.

285. Micsonai, A.; Moussong, É.; Wien, F.; Boros, E.; Vadászi, H.; Murvai, N.; Lee, Y.-H.; Molnár, T.; Réfrégiers, M.; Goto, Y.; Tantos, Á.; Kardos, J. BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy. *Nucleic acids research* **2022,** *50* (W1), W90-8. DOI: 10.1093/nar/gkac345.

286. Stein, T.; Borchert, S.; Kiesau, P.; Heinzmann, S.; Klöss, S.; Klein, C.; Helfrich, M.; Entian, K.-D. Dual control of subtilin biosynthesis and immunity in Bacillus subtilis. *Molecular microbiology* **2002,** *44* (2), 403–416. DOI: 10.1046/j.1365-2958.2002.02869.x.

287. Wahlgren, W. Y.; Claesson, E.; Tuure, I.; Trillo-Muyo, S.; Bódizs, S.; Ihalainen, J. A.; Takala, H.; Westenhoff, S. Structural mechanism of signal transduction in a phytochrome histidine kinase. *Nat Commun* **2022,** *13* (1), 7673. DOI: 10.1038/s41467-022-34893-3.

288. Chang, C.; Tesar, C.; Gu, M.; Babnigg, G.; Joachimiak, A.; Pokkuluri, P. R.; Szurmant, H.; Schiffer, M. Extracytoplasmic PAS-like domains are common in signal transduction proteins. *Journal of bacteriology* **2010,** *192* (4), 1156–1159. DOI: 10.1128/JB.01508-09.

289. Hallgren, J.; Tsirigos, K. D.; Pedersen, M. D.; Almagro Armenteros, J. J.; Marcatili, P.; Nielsen, H.; Krogh, A.; Winther, O. *DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks*, 2022. DOI: 10.1101/2022.04.08.487609.

290. Lemmin, T.; Soto, C. S.; Clinthorne, G.; DeGrado, W. F.; Dal Peraro, M. Assembly of the transmembrane domain of E. coli PhoQ histidine kinase: implications for signal transduction from molecular simulations. *PLoS computational biology* **2013,** *9* (1), e1002878. DOI: 10.1371/journal.pcbi.1002878.

291. Affandi, T.; McEvoy, M. M. Mechanism of metal ion-induced activation of a two-component sensor kinase. *The Biochemical journal* **2019,** *476* (1), 115–135. DOI: 10.1042/BCJ20180577.

List of figures

List of tables

Curriculum vitae

Gelöscht für die Online-Version

Gelöscht für die Online-Version

Gelöscht für die Online-Version