



Max Planck Graduate Center
mit der Johannes Gutenberg-Universität



Phase Behavior and Condensate Properties of TDP-43

Dissertation for the award of the academic degree of
“Doctor rerum naturalium” (Dr. rer. nat.) of the faculties:

08 - Physics, Mathematics and Computer Science
09 - Chemistry, Pharmacy, Geography and Earth Sciences
10 - Biology
University Medicine
submitted by

Xiaofei PING

Mainz, 2025

D77

Supervisor: Prof. Dr. Lukas Stelzl
Second supervisor: Prof. Dr. Friederike Schmid

Day of the oral examination:

Re-use: Copyright protection (InC-1.0)

Declaration

I hereby declare that I wrote the dissertation submitted without any unauthorized external assistance and used only sources acknowledged in the work. All textual passages which are appropriated verbatim or para-phrased from published and unpublished texts as well as all information obtained from oral sources are duly indicated and listed in accordance with bibliographical rules. In carrying out this research, I complied with the rules of standard scientific practice as formulated in the statutes of Johannes Gutenberg-University Mainz to insure standard scientific practice

Mainz, 2025
Xiaofei PING

Abstract

TAR DNA-binding protein 43 (TDP-43) is a 414-amino acid protein comprising three folded domains—the N-terminal domain (NTD) and two RNA recognition motifs (RRM1 and RRM2)—along with a long intrinsically disordered region (IDR) at the C-terminal. A nuclear localisation signal (NLS) sequence connects the NTD to RRM1. TDP-43 plays crucial roles in RNA metabolism, including transcriptional regulation, splicing, and RNA transport. Misregulation or aggregation of TDP-43 is strongly associated with multiple neurodegenerative diseases, including Alzheimer’s disease (AD), Amyotrophic Lateral Sclerosis (ALS), and frontotemporal dementia (FTD), and it is widely regarded as a hallmark protein in these diseases.

Within cells, native TDP-43 forms liquid-like and dynamic condensates through phase separation, which are thought to contribute to its physiological function. Understanding the molecular determinants of TDP-43 condensate formation is therefore essential for elucidating mechanisms underlying neurodegeneration. A near-atomic resolution coarse-grained model, **Martini 3**, was employed to systematically investigate the roles of both folded and intrinsically disordered regions in TDP-43 condensates. This approach allows exploration of the effects of domain architecture, inter-domain interactions, and conformational flexibility on the dynamic organization of the protein within condensates.

The application of different potentials, including harmonic or Lennard-Jones (LJ) potentials, to restrain the structured domains, combined with tuning of inter-chain interactions within the condensates, can modulate phase behavior, material properties, and dynamic rearrangement of TDP-43. Quantitative analysis of the contributions of different structural elements to condensate stability, internal dynamics, and viscosity-like properties was performed. Comparative analysis across multiple modeling approaches provides insights into how different coarse-grained assumptions influence predicted phase behavior and emergent material characteristics.

Overall, this work provides a detailed molecular-level understanding of the balance between folded and disordered regions in regulating TDP-43 condensation. The study also provides foundational data for applying the Martini 3 model as a transferable framework to study other multi-domain proteins (MDPs) in multi-chain systems.

Zusammenfassung

TAR-DNA-Bindeprotein 43 (TDP-43) ist ein 414 Aminosäuren umfassendes Protein, das aus drei gefalteten Domänen besteht – der N-terminalen Domäne (NTD) und zwei RNA-Erkennungs-Motiven (RRM1 und RRM2) – sowie einer langen intrinsisch ungeordneten Region (IDR) am C-Terminus. Eine nukleäre Lokalisationssequenz (NLS) verbindet die NTD mit RRM1. TDP-43 spielt eine entscheidende Rolle im RNA-Stoffwechsel, einschließlich Transkriptionsregulation, Spleißen und RNA-Transport. Fehlregulation oder Aggregation von TDP-43 ist stark mit mehreren neurodegenerativen Erkrankungen assoziiert, darunter Alzheimer-Krankheit (AD), Amyotrophe Lateralsklerose (ALS) und frontotemporale Demenz (FTD), und gilt als charakteristisches Protein dieser Erkrankungen.

Innerhalb der Zellen bildet natives TDP-43 flüssigartige und dynamische Kondensate durch Phasentrennung, die vermutlich zu seiner physiologischen Funktion beitragen. Das Verständnis der molekularen Determinanten der TDP-43-Kondensatbildung ist daher wesentlich, um die Mechanismen neurodegenerativer Erkrankungen zu erläutern. Ein nah-atomarer Auflösungs-Coarse-Grained-Modell, **Martini 3**, wurde verwendet, um systematisch die Rollen sowohl der gefalteten als auch der intrinsisch ungeordneten Regionen in TDP-43-Kondensaten zu untersuchen. Dieser Ansatz ermöglicht die Analyse der Auswirkungen von Domänenarchitektur, interdomänären Wechselwirkungen und konformationeller Flexibilität auf die dynamische Organisation des Proteins innerhalb der Kondensate.

Die Anwendung verschiedener Potentiale, einschließlich harmonischer oder Lennard-Jones-Potentiale, zur Einschränkung der strukturierten Domänen, kombiniert mit der Anpassung interkettiger Wechselwirkungen innerhalb der Kondensate, kann das Phasenverhalten, die Materialeigenschaften und die dynamische Umorganisation von TDP-43 modulieren. Quantitative Analysen der Beiträge verschiedener Strukturelemente zur Stabilität der Kondensate, deren interner Dynamik und viskositätsähnlichen Eigenschaften wurden durchgeführt. Ein Vergleich mehrerer Modellansätze liefert Einblicke, wie unterschiedliche Coarse-Grained-Annahmen das vorhergesagte Phasenverhalten und die entstehenden Materialeigenschaften beeinflussen.

Insgesamt liefert diese Arbeit ein detailliertes molekularbiologisches Verständnis des Gleichgewichts zwischen gefalteten und ungeordneten Regionen bei der Regulation der TDP-43-Kondensation. Die Studie liefert zudem grundlegende Daten, um das Martini 3-Modell als übertragbares Framework für die Untersuchung anderer Multi-Domänen-Proteine (MDPs) in Mehrkettensystemen anzuwenden.

(Übersetzung erstellt mithilfe von ChatGPT)

Table of Contents

1	Introduction	1
1.1	Liquid-liquid phase separation in cells	1
1.2	Protein TDP-43	2
1.2.1	Endogenous TDP43	2
1.2.2	TDP-43 in diseases	4
1.3	Molecular Dynamics	5
1.3.1	Interactions in molecular dynamics	6
1.3.1.1	Bonded interactions	6
1.3.1.2	Nonbonded interactions	8
1.3.2	Coarse-Grained Simulations	9
1.3.2.1	Chemically-detailed coarse-grained simulations	9
1.3.2.2	Residue-level implicit coarse-grained simulations	11
2	Phase behavior of TDP-43	13
2.1	Dilute-phase behavior of full-length TDP-43	14
2.1.1	Single chain compactness	15
2.1.2	Intra-molecular contacts	17
2.2	Dense-phase behavior of full-length TDP-43	22
2.2.1	Full-length protein phase separation	22
2.2.2	Inter-molecule interactions	26
2.2.2.1	Overall interactions	26
2.2.2.2	LCD in condensates	28
2.3	Phase behavior of TDP-43 Variants	31
2.3.1	Aromatic mutants	33
2.3.2	Serine to Aspartate phosphomimicking mutant	35
2.4	Discussion	38
3	Regulating TDP-43 LLPS via structured domain stabilization and inter-chain bridging	41
3.1	Single chain full-length protein conformation	41
3.1.1	Single chain compactness	41
3.1.2	Residue flexibility	44
3.2	Dilute TDP43-NTD dimerization characterization	47
3.3	Full-length TDP-43 condensation via LJ-enhanced models	51
3.3.1	NTD-dimer within condensates	51
3.3.2	Inter-molecular interactions within condensates	52

Table of Contents

3.3.3	Density distribution within condensates	54
3.3.4	Surface tension and shear viscosity	56
3.4	Discussion	61
4	Helical Structure within the TDP-43 C-Terminal Intrinsically Disordered Region	63
4.1	Residue W334 intra-contacts within the helix and inter-contacts in the con- densate	63
4.2	Residue Ser332 and Ser333 in helix	70
4.2.1	12pSer	72
4.3	Discussion	76
5	Methods	77
5.1	Molecular Dynamics Simulations	77
5.1.1	Atomistic simulations	77
5.1.2	Explicit coarse-grained simulations	77
5.1.2.1	Elastic Network model	78
5.1.2.2	Intra Gō-type potential model	78
5.1.2.3	Intra and inter Gō-type potential model	79
5.1.2.4	Multi-chain simulation setup	79
5.1.3	Implicit coarse-grained simulations	80
5.2	Molecular Dynamics Simulation Analysis	81
5.2.1	Density and phase diagram calculations	81
5.2.2	Interfacial tension calculations	81
5.2.3	Viscosity estimation	82

List of Figures

1.1	Distinct phase behaviors of dipeptides. Representative dipeptides exhibiting no phase separation, liquid–liquid phase separation (LLPS), or liquid–solid phase separation (LSPS). For each case, the left panels show simulation snapshots from molecular dynamics simulations, while the right panels display the corresponding <i>in vitro</i> experimental results. The simulations and experiments consistently distinguish dipeptides that remain dispersed, form liquid-like condensates, or assemble into solid aggregates. (Figure adapted from Tang <i>et al.</i> [1] with modifications.)	1
1.2	Cellular distribution of endogenous TDP-43. Left: Western blot analysis of subcellular fractions showing the distribution of endogenous TDP-43 in the nucleus (N) and cytoplasm (C) of non-transfected HeLa cells. p84 and tubulin were used as nuclear and cytoplasmic markers, respectively, to assess fraction purity. Right: Indirect immunofluorescence of non-transfected U2OS cells showing endogenous TDP-43 localization. TDP-43 was detected using an anti-TDP-43 polyclonal antibody and a FITC-conjugated secondary antibody. The nuclear envelope was visualized using anti-lamin A/C or anti-LAP2 β antibodies with Texas Red- or FITC-conjugated secondary antibodies. (Figure adapted from Ayala <i>et al.</i> [2].)	3
1.3	Disease-associated mutations in TDP-43. Most disease-associated TDP-43 mutations are concentrated in the C-terminal low-complexity domain (LCD) and have been shown to accelerate TDP-43 aggregation. (Figure adapted from Jo <i>et al.</i> [3].)	4
1.4	Molecular bonded interactions. Molecular mechanics interactions of (a) bond stretching, (b) angle bending, (c) dihedral, and (d) improper dihedral torsion (Figure from [4]).	7
1.5	Comparison of structure-bias models used to stabilize coarse-grained protein conformations. Atomistic structure of protein CK1 δ in cartoon. The Elastic Network (EN) model (blue) introduces 1004 harmonic restraints, the GōMartini model (orange) adds 609 native-contact-based potentials, and the OLIVES model (red) includes 374 potentials. While the EN and GōMartini interaction networks connect predominantly backbone beads, the OLIVES model additionally incorporates side-chain bead interactions, providing a more extensive stabilization pattern. (Figure from [5])	10

List of Figures

1.6	Residue-Level bead representation of proteins. All-atom representation is in balls and sticks, and coarse-grained representation in large spheres. (Figure adapted from [6])	11
2.1	Representative structure of full-length TDP-43, with average R_g highlighted in all-atom simulations. Residues colored in green, orange, pink correspond to the N-terminal domain (NTD), RNA-recognition motif 1 (RRM1), and 2 (RRM2), respectively. Cyan areas indicate the protein's disordered regions, and the region colored in purple represents the helical segment within the C-terminal IDR. water and ions are omitted for clarity for all-atom model).	14
2.2	R_g measurement of single chain full-length TDP-43 under all-atom and Martini 3 Elastic Network (EN) model at $\lambda_{1.00}$ and $\lambda_{1.06}$. (A). R_g across three replicates in all-atom simulation, the black dashed line marks the maximum experimental R_g and the orange-like dashed line shows the mean values across replicas. (B). R_g over three replicates in Elastic Network model at $\lambda_{1.00}$, with the same dashed-line conventions. (C). R_g over three replicates in Elastic Network model at $\lambda_{1.06}$, with the same dashed-line conventions.	19
2.3	R_g measurement of single chain full-length TDP-43 6WA under all-atom (AA) simulations. The black dashed line indicates the averaged R_g value in AA simulations.	20
2.4	Representative structure of full-length TDP-43, with average R_g highlighted in Martini 3 Elastic Network (EN) at $\lambda_{1.00}$ and $\lambda_{1.06}$ simulations. (A). Representative structure in EN $\lambda_{1.00}$. (B). Representative structure in EN $\lambda_{1.06}$. Residues colored in green, orange, pink correspond to the N-terminal domain (NTD), RNA-recognition motif 1 (RRM1), and 2 (RRM2), respectively. Cyan areas indicate the protein's disordered regions, and the region colored in purple represents the helical segment within the C-terminal IDR. water and ions are omitted for clarity for EN model.	20
2.5	Averaged intra-chain contact heatmap of single chain full-length TDP-43 under all-atom (AA) and Martini 3 Elastic Network (EN) model at $\lambda_{1.00}$ and $\lambda_{1.06}$. (A). AA simulations. (B). EN model at $\lambda_{1.00}$. (C). EN model at $\lambda_{1.06}$. Upper triangles are the residue-wise contact maps, lower triangles are domain-wise contact maps, which summing up the contact possibility and averaged by the total residue number within the specific domain, colorbar on the right represents the residue contact probabilities.	21

2.6	Full-length TDP-43 multichain condensation in cubic simulation geometry with Martini 3 Elastic Network (EN) at $\lambda_{1.06}$. Initial configuration of the EN model of full-length TDP-43 simulated with protein–water interaction scaling of $\lambda_{1.06}$ in a cubic simulation box. Starting from a dispersed arrangement, the chains assemble into a liquid-like spherical condensate. Na^+ (red) and Cl^- (blue) ions are shown; water beads are omitted for clarity.	22
2.7	Phase separation of full-length TDP-43 in Martini 3 Elastic Network (EN) simulations within slab geometry. (A–C) Time-dependent concentration histograms for simulations with the Martini 3 EN model at protein–water interaction strengths of $\lambda_{1.05}$, $\lambda_{1.06}$, and $\lambda_{1.07}$. Protein concentration is indicated by the color bar. (D–F) Snapshots of the protein condensate along the z-axis for the Martini 3 EN model at $\lambda_{1.05}$, $\lambda_{1.06}$, and $\lambda_{1.07}$, respectively.	24
2.8	Time-dependent concentration histograms for full-length TDP-43 in Martini 3 Elastic Network (EN) simulations within slab geometry, shown for replicates at protein–water interaction strengths of $\lambda_{1.05}$, $\lambda_{1.06}$, and $\lambda_{1.07}$.	25
2.9	Phase concentration quantification in Martini 3 Elastic Network (EN) simulations within slab geometry for replicates. Density along the z-axis fitting results across replicas on $\lambda_{1.05}$, $\lambda_{1.06}$ and $\lambda_{1.07}$	26
2.10	Inter-molecular contacts within condensates. A. Inter-chain contact heatmaps of Elastic Network with $\lambda_{1.06}$ in the slab geometry, upper triangle is the residue-wise contact maps, lower triangle is domain-wise contact maps, residue- and domain-wise contact maps are sharing same colorbar. B. Averaged inter-molecular contact possibility of each residue, tryptophan residues are indicated by red arrows, phenylalanine residues by green arrows, and other aromatic residues by black arrows.	27
2.11	Inter-chain contacts involving the NLS under Martini 3 Elastic Network (EN) simulations at $\lambda_{1.06}$. (A). Snapshot illustrating NLS–NLS inter-chain contacts. (B). Snapshot illustrating NLS–LCD inter-chain contacts. Folded domains (NTD, RRM1, RRM2, and the C-terminal helix) are shown in green, orange, and pink, respectively, while intrinsically disordered regions from different chains are colored in cyan and purple. (C). Residue-type contact heatmap of NLS–LCD inter-chain interactions.	29

2.12	LCD phase separation under Martini 3 Elastic Network (EN) simulations. (A). Protein density profiles along the z-axis. Full-length TDP-43 was simulated with $\lambda_{1.06}$, and LCD chains were simulated with $\lambda_{1.06}$, $\lambda_{1.05}$, and $\lambda_{1.04}$ in the slab geometry. (B). Representative snapshot of LCD condensate at $\lambda_{1.06}$. (C). Representative snapshot of LCD forming a condensate at $\lambda_{1.04}$. (D). Ion and water density profiles along the z-axis for full-length TDP-43 ($\lambda_{1.06}$) and LCD ($\lambda_{1.04}$). (E) <i>In vitro</i> experiments testing phase separation of full-length TDP-43, a TDP-43 variant lacking residues 1–101, and the isolated LCD at different protein concentrations. Experimental data in panel E were kindly provided by Dr. Saskia Hutten (Dormann Group).	30
2.13	Phase separation of aromatic and serine mutants under Martini 3 Elastic Network (EN) simulations at $\lambda_{1.06}$. (A). Transfer free energies of the mutants in slab-geometry simulations at $\lambda_{1.06}$. All mutants display reduced phase separation relative to the wild type (WT). (B). <i>In vitro</i> phase-separation assays for the aromatic mutants under identical experimental conditions, showing diminished condensate formation compared to the WT. Experimental data in panel B were kindly provided by Dr. Saskia Hutten (Dormann Group).	32
2.14	Time-dependent density distributions of aromatic mutants along the z-axis under Martini 3 Elastic Network (EN) simulations at $\lambda_{1.06}$. (A–C) Concentration histograms over time for the wild-type (WT) protein and aromatic mutants W3A and F6W3A. Protein concentration is indicated by the color bar on the right.	33
2.15	Intermolecular contacts within aromatic mutant condensates under Martini 3 Elastic Network (EN) simulations at $\lambda_{1.06}$. (A–C). Inter-chain contact heatmaps for the WT protein and aromatic mutants W3A and F6W3A, respectively. Upper triangles are the residue-wise contact maps, lower triangles are domain-wise contact maps, residue- and domain-wise contact maps are sharing same colorbar. (D). Averaged intermolecular contact probability for each residue. Tryptophan residues are indicated by red arrows, phenylalanine residues by green arrows, and other aromatic residues by black arrows.	34
2.16	Phase separation of the phosphomimetic 12D mutant under Martini 3 Elastic Network (EN) simulations at $\lambda_{1.06}$. (A–B). Time-dependent concentration histograms along the z-axis for the wild-type (WT) protein and the 12D mutant, with protein concentration indicated by the color bar. (C). Ion and water density profiles along the z-axis for WT TDP-43 and the 12D mutant, highlighting differences in ion recruitment and local hydration within the condensate.	36
2.17	Water cavities within the 12D mutant condensate. One of the water-filled cavities is highlighted with a blue circle.	36

2.18	Inter-molecular contacts within the phosphomimetic 12D mutant under Martini 3 Elastic Network (EN) simulations at $\lambda_{1.06}$. (A–B). Inter-chain contact heatmaps of the wild-type (WT) protein and the Ser-to-Asp mutant 12D. (C). Averaged inter-molecular contact probability for each residue, with all serine residues indicated by black arrows.	37
2.19	Time-dependent concentration histograms of mutant replicas under the Martini 3 Elastic Network (EN) model at $\lambda_{1.06}$. Protein concentration is indicated by the color bar.	38
3.1	Radius of gyration (R_g) of single-chain full-length TDP-43 under different models and λ values. (A). R_g across three replicates in the Elastic Network model at $\lambda_{1.06}$; the black dashed line indicates the maximum experimental R_g , and the orange dashed line shows the mean across replicates. (B). R_g across three replicates in the $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.00}$, with the same dashed-line conventions. (C). R_g across three replicates in the $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.06}$, with the same dashed-line conventions.	42
3.2	Averaged intra-chain contact heatmaps of single-chain full-length TDP-43 under different models. (A). Elastic Network model at $\lambda_{1.06}$. (B). $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.00}$. (C). $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.06}$. For each panel, the upper triangle shows the residue-wise contact map, and the lower triangle shows the domain-wise contact map. All contact maps share the same colorbar.	43
3.3	Residue flexibility within the structured domains of TDP-43 under all-atom, Elastic Network and $G\bar{o}_{\text{intra}}$ models at $\lambda_{1.06}$. Root mean square fluctuation (RMSF) of backbone beads in folded regions: NTD, RRM1, RRM2, and the helical segment of the C-terminal disordered region. The left axis measures the RMSF of each residue from its average position. Here both rescaled and unrescaled Martini 3 models are shown for comparison. The right axis in combination with the colorbar track the secondary structure of each residue over the time course of an AA simulation: blue is helix, pink is extended structure (β -sheet), and white is coil.	46
3.4	NTD dimerization. (A). The dimer is stabilized by complementary electrostatic interactions between the negatively charged head of one chain (right; β -sheet in pink and α -helix in green) and the positively charged tail of the partner chain (left; β -sheet in pink). (B). Fraction of native inter-chain contacts observed in all-atom and coarse-grained simulations of the NTD homodimer.	47
3.5	Decomposition of native inter-chain contacts in NTD homodimer. Breakdown of Q_x into individual interface components: β - β , β - α , and other contacts for (A) AA, (B) EN, (C) $G\bar{o}_{\text{intra}}$, and (D) $G\bar{o}_{\text{intra+inter}}$ models.	49
3.6	Inter-chain contact heatmaps of NTD dimerization under different models. (A). EN model at $\lambda_{1.00}$. (B). $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.00}$. (C). EN model at $\lambda_{1.06}$. (D). $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.06}$	50

3.7	Inter-chain and NTD-dimer-mediated contact pair counts within condensates formed under the Elastic Network (EN), $G\bar{o}_{\text{intra}}$, and $G\bar{o}_{\text{intra+inter}}$ models at $\lambda = 1.06$. (A). Total contact pair counts are shown using blue gradients for the EN model, orange gradients for the $G\bar{o}_{\text{intra}}$ model, and green gradients for the $G\bar{o}_{\text{intra+inter}}$ model. (B). Inter-molecular $Q_{(x)}$ values of full-length chain pairs within the condensate that form NTD dimers dynamically.	51
3.8	Inter-chain contacts within TDP-43 condensates under the EN, $G\bar{o}_{\text{intra}}$, and $G\bar{o}_{\text{intra+inter}}$ models at $\lambda_{1.06}$. (A–C). Inter-chain contact heatmaps for the EN, $G\bar{o}_{\text{intra}}$, and $G\bar{o}_{\text{intra+inter}}$ models, respectively. Upper triangles show residue-wise contact maps, while lower triangles display domain-wise contact maps, calculated by summing the contact probabilities and averaging over the total number of residues within each domain. Residue- and domain-wise contact maps share the same colorbar. (D). Averaged inter-molecular contact probability of each residue across the three models: EN in green, $G\bar{o}_{\text{intra}}$ in orange, and $G\bar{o}_{\text{intra+inter}}$ in dark green. Aromatic residues are indicated with arrows.	53
3.9	Zoomed-in view of residue-level averaged intermolecular contact probabilities within the NLS region. Charged residues are indicated by black dashed lines.	53
3.10	Condensate density analysis under $G\bar{o}_{\text{intra}}$ and $G\bar{o}_{\text{intra+inter}}$ models at $\lambda_{1.06}$. (A, C). Time-dependent density histograms of the condensates for the $G\bar{o}_{\text{intra}}$ and $G\bar{o}_{\text{intra+inter}}$ models, respectively. (B, D). Corresponding averaged density histograms, fitted to a hyperbolic tangent (tanh) function (Eq. 5.2) to extract dense- and dilute-phase concentrations.	55
3.11	Time-dependent density histograms along the Z-axis across replicates of condensates for the $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.05}$, $\lambda_{1.06}$, and $\lambda_{1.07}$ in $G\bar{o}_{\text{intra}}$ model.	56
3.12	Averaged density histograms fitted to a hyperbolic tangent (tanh) function 5.2 to extract dense and dilute phase concentrations under the $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.05}$, $\lambda_{1.06}$ and $\lambda_{1.07}$	57
3.13	A chain trapped at the condensate interface in the $G\bar{o}_{\text{intra+inter}}$ model at $\lambda_{1.06}$	57
3.14	(A). Phase diagram of the Elastic Network (EN) and $G\bar{o}_{\text{intra}}$ models across varied rescaling factors λ , with the dense- and dilute-phase concentrations of $G\bar{o}_{\text{intra+inter}}$ shown at $\lambda_{1.06}$. (B). Surface tension measurements for each model. (C). Interfacial width of condensates across models at $\lambda_{1.06}$. (D). Viscosity estimated from end-to-end distance autocorrelations for each model at $\lambda_{1.06}$	58
3.15	Interfacial width W^2 as a function of block size B. The interface region is divided into rectangular segments, and the interfacial tension γ is determined using $W^2(B) = W_0^2 + \frac{k_B T}{2\pi\gamma} \ln \frac{L}{B}$, where W_0 is the intrinsic interfacial width, k_B is the Boltzmann constant, T is the temperature, and L is the simulation box size.	59

4.1	All-atom simulations of helix starting configurations. Initial structures derived from PDB entries 2N2C (A) and 2N2X (B), with residues Leu330, Trp334, and Met337 highlighted.	64
4.2	Intra-chain interactions within the helix in all-atom simulations. (A–B). Intra-chain contact probability heatmaps for 2N2C (A) and 2N2X (B), with residue Trp334 indicated by dashed lines. (C–D). Zoomed-in views illustrating the local packing and contacts among residues Leu330, Trp334, and Met337 within the helical structures of 2N2C and 2N2X, respectively. Dashed black lines indicate hydrogen bonds between residues Leu330–Trp334 and Trp334–Met337.	65
4.3	Representative structures illustrating side-chain contacts mediated by Trp334 within the helix. A. Representative snapshot of the Phe313–Trp334 side-chain contact within the helix, with the atoms of Phe313 and Trp334 shown. B. Side-chain interaction between residues Gln327 and Trp334. C. Side-chain interaction between residues Gln331 and Trp334. D. Representative configuration showing the combined side-chain contacts involving residues Gln327, Trp334, and Met337.	66
4.4	Intra-chain interactions and helix propensities of the W334A mutant in all-atom simulations. (A). Intra-chain contact probability heat map for the W334A mutant using PDB:2n2c as the starting structure. (B). Intra-chain contact probability heat map for the W334A mutant using PDB:2n2x as the starting structure. (C). Residue-resolved helix propensities for the W334A mutant starting from PDB:2n2c. (D). Residue-resolved helix propensities for the W334A mutant starting from PDB:2n2x.	67
4.5	Representative snapshots illustrating the Leu330–Trp334–Met337 interaction motif and the Trp334–Met337 contact within the helix of the W334A mutant. (A). Leu330–Trp334–Met337 interaction motif; black dashed lines indicate inter-residue distances. (B). Trp334–Met337 contact within the helix; black dashed lines indicate inter-residue distances.	68
4.6	Contact heatmaps of WT and W3A condensates showing inter- and intra-chain interactions under the EN model at $\lambda_{1.06}$. (A–B). Inter-chain contact heatmaps for WT and W3A condensates, respectively. (C–D). Intra-chain contact heatmaps for WT and W3A condensates, respectively.	69
4.7	Phospho-mimicking or phosphorylated Ser residues at positions 332 and 333 and their effect on helix stability in all-atom simulations. (A–B). Intra-chain contact heatmaps for the phospho-mimicking and phosphorylated Ser residues, respectively. (C–D). Residue-resolved helix propensities for the phospho-mimicking and phosphorylated Ser residues, respectively.	70
4.8	Representative snapshots illustrating two phosphorylated Ser residues in all-atom simulations. All atoms of the phosphorylated Ser residues are shown.	71

List of Figures

4.9	Helix stability of the 12 phosphorylated Ser residues (12pSer) in full-length TDP-43 AA simulations. (A). Representative snapshots illustrating 12pSer in all-atom simulations. All atoms of the phosphorylated Ser residues are shown, and the helical region is colored in red. (B). Residue-resolved helix propensities for the 12pSer mutant.	73
4.10	Intra-chain contact heatmaps of WT and 12pSer mutant in all-atom simulations. (A–B). Intra-chain contact heatmaps for WT and 12pSer mutant, respectively. (C–D). Zoomed-in intra-chain contact heatmaps for the LCD region of WT and 12pSer mutant, respectively. . . .	74
4.11	12pSer condensates in implicit coarse-grained simulations. (A). Representative snapshots of the condensates: WT (upper) and 12pSer mutant (lower). (B). Averaged protein density along the z-axis in the slab geometry for WT and 12pSer condensates.	75

List of Acronyms

LLPS Liquid-liquid phase separation

LSPS Liquid-solid phase separation

ALS Amyotrophic Lateral Sclerosis

AD Alzheimer Disease

MDP Multi-domain protein

LCD Low-complexity domain

MD Molecular Dynamics

AA All-atom

CG Coarse-grained

LJ Lennard-Jones

1 Introduction

1.1 Liquid-liquid phase separation in cells

Phase separation describes the process where a homogeneous system spontaneously creates two distinct phases to minimize system free energy. Liquid-liquid phase separation (LLPS) gives rise to dynamic, liquid-like biomolecular condensates that play critical roles in a wide range of physiological processes. Over time, some liquid-like condensates can undergo maturation or aging, transitioning through liquid-solid phase separation (LSPS) into more gel-like or solid-like states (Figure 1.1), this solidification is often associated with pathology, as the irreversible aggregation of proteins into solid fibrils is a hallmark of neurodegenerative diseases such as Amyotrophic Lateral Sclerosis (ALS) and Alzheimer Disease (AD).

Figure 1.1 illustrates the phase behavior of dipeptides (e.g., Ile-Val, Gly-Ala, Phe-Trp, and Phe-Phe) as determined by molecular dynamics simulations and corresponding *in vitro* experiments. The figure represents dipeptides that do not phase separate, those that undergo liquid-liquid phase separation to form liquid condensates, and those that assemble into solid aggregates [1].

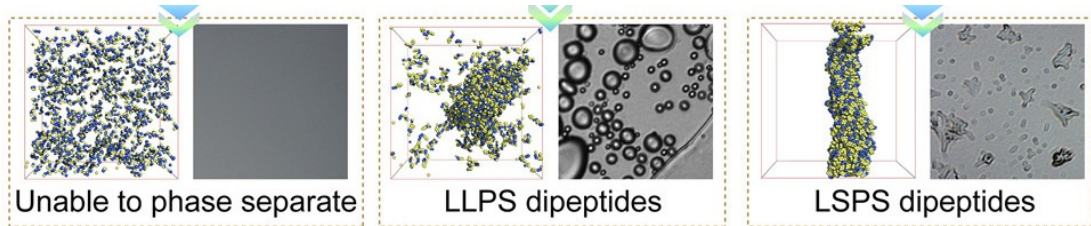


Figure 1.1: Distinct phase behaviors of dipeptides. Representative dipeptides exhibiting no phase separation, liquid-liquid phase separation (LLPS), or liquid-solid phase separation (LSPS). For each case, the left panels show simulation snapshots from molecular dynamics simulations, while the right panels display the corresponding *in vitro* experimental results. The simulations and experiments consistently distinguish dipeptides that remain dispersed, form liquid-like condensates, or assemble into solid aggregates. (Figure adapted from Tang *et al.* [1] with modifications.)

Within cellular environments, homogeneous solution of biomolecules like RNA, DNA, protein and lipids can phase separates into two immiscible liquid phases, a dense phase that is

1 Introduction

enriched for these molecules and a phase that is depleted. Hyphosis "cytoplasm is a liquid emulsion" was proposed by Edmund Beecher Wilson in 1899[7], and in 2009 this hypothesis gained experimental support when Tony Hyman and Cliff Brangwynne demonstrated P granules in *Caenorhabditis elegans* germ cells which localize to the posterior of the one-cell embryo stage exhibit liquid-like properties with measurable viscosity and surface tension [8]. LLPS is a dynamic process that can be triggered by cellular stimulus, like electrochemical gradients (pH, ion concentrations, etc.) which changes the biomolecule local environment. Some liquid-state condensates are transiently formed and have defined cellular functions which named as membrane-less organelles (MLO) which don't have bi-layer phospholipids, like stress granules (SG) in the cytoplasm can store mRNA temporarily to prevent translation and resume translation when stress releases [9].

Liquid state condensates can transform into gel-like or solid-like condensate which refers to the process LSPS. Protein like FUS, hnRNPA1 and Whi3 can undergo LLPS to form liquid-like droplet, whereas over time like 8 hours later for protein FUS, fibers merging from the droplet can be observed by *in vitro* with arrested dynamics [10]. Functions of solid-like or gel-like structures in the living healthy cell are less clear, but their formation is often associated with cellular stress responses and, when dysregulated, can contribute to pathological protein aggregation linked to neurodegenerative diseases.

1.2 Protein TDP-43

1.2.1 Endogenous TDP43

TAR DNA-binding protein 43 (TDP-43) was first identified as a 43-kDa protein that binds to double-stranded DNA within a specific region of the Human Immunodeficiency Virus type 1 (HIV-1) genome [11]. Protein TDP-43 plays a critical role in RNA processing and metabolism. It is expressed in both neuronal and nonneuronal cells, shuttles between the nucleus and the cytoplasm [2, 12]. As shown in Figure 1.2, TDP-43 is predominantly in nucleus, from the immunofluorescence image (Right), the nuclear envelope is highlighted in red to delineate the nucleus, and the Western blot analysis (Left) confirms that only a small fraction of TDP-43 resides in the cytoplasm.

In healthy cells, endogenous TDP-43 is predominantly expressed in the nucleus, with a small portion localized in the cytoplasm [2, 12, 13]. The protein is actively involved in RNA metabolism, including roles in pre-mRNA splicing [14] and RNA stability [13]. Within the nucleus, a small fraction of endogenous TDP-43 forms liquid-like condensates, although the majority is diffusely localized. This diffuse distribution is maintained by the native high nuclear RNA concentration, which prevents extensive condensate formation[15].

In response to cellular stress, TDP-43 can undergo reversible phase separation and form dynamic, liquid-like condensates in both the nucleus and the cytoplasm. Under arsenic-induced stress, TDP-43 assembles into nuclear bodies that are thought to mitigate cytotox-

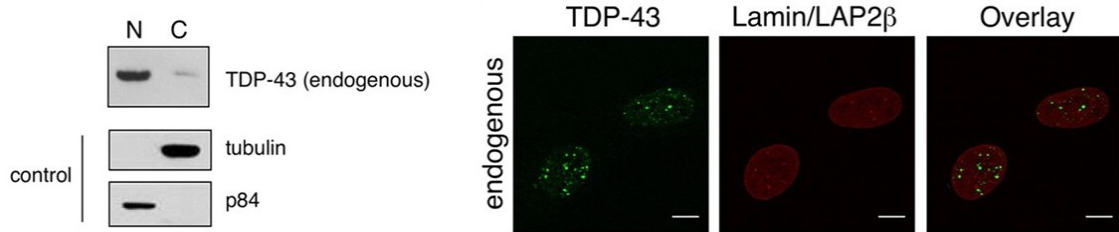


Figure 1.2: Cellular distribution of endogenous TDP-43. Left: Western blot analysis of subcellular fractions showing the distribution of endogenous TDP-43 in the nucleus (N) and cytoplasm (C) of non-transfected HeLa cells. p84 and tubulin were used as nuclear and cytoplasmic markers, respectively, to assess fraction purity. Right: Indirect immunofluorescence of non-transfected U2OS cells showing endogenous TDP-43 localization. TDP-43 was detected using an anti-TDP-43 polyclonal antibody and a FITC-conjugated secondary antibody. The nuclear envelope was visualized using anti-lamin A/C or anti-LAP2 β antibodies with Texas Red- or FITC-conjugated secondary antibodies. (Figure adapted from Ayala *et al.* [2].)

icity, and upon stress removal, these condensates gradually dissolve, restoring the diffuse nuclear distribution characteristic of healthy cells [9]. This reversible behavior highlights the dynamic nature of TDP-43 assemblies and suggests that its phase separation is tightly regulated to maintain nuclear homeostasis.

In the cytoplasm, TDP-43 is similarly recruited to stress granules (SGs) under conditions of oxidative stress. The treatment of sodium arsenite (SA) induces TDP-43 accumulation in the cytoplasm, where it co-localizes with protein HuR, which plays as a SG marker, forming discrete membrane-less organelles. SGs serve as protective compartments, temporarily sequestering non-essential mRNAs, translation factors, and RNA-binding proteins to reduce translational burden and preserve cellular resources during stress. Once the stress is alleviated, these SGs disassemble, releasing their contents back into the cytoplasm and allowing normal cellular function to resume [16–21]. TDP-43 plays a critical role in regulating both the formation and disassembly of SGs. Alterations in TDP-43 protein level or mutations can disrupt SG dynamics, either impairing assembly or slowing disassembly [21].

TDP-43 accumulates in the nucleus and/or cytoplasm under stress conditions, where it forms reversible, liquid-like condensates. These dynamic assemblies are not passive deposits; rather, they serve as adaptive compartments that help the cell cope with stress. To maintain viability under unfavourable conditions, cells rely on mechanisms that temporarily reorganize RNA and protein components without disrupting essential functions. The liquid nature of these assemblies ensures that they can remain dynamic and reversible, once the stress subsides, the sequestered components can be released, and normal cellular functions can resume.

1 Introduction

1.2.2 TDP-43 in diseases

The accumulation of protein aggregates in the central nervous system is a hallmark of many neurodegenerative diseases, including ALS and AD [3, 22]. In affected neurons, these aggregates are predominantly localized in the cytoplasm, while the nucleus is typically depleted of the aggregating proteins [23, 24]. Among the proteins implicated in these pathologies, aggregates of TDP-43 are frequently observed across a broad spectrum of neurodegenerative disorders and represent a defining feature of TDP-43 proteinopathies [25].

In the case of TDP-43, disease-associated mutations or aberrant post-translational modifications (PTMs) can promote irreversible aggregation through a LSPS [15]. Numerous mutations in the TDP-43 gene have been linked to neurodegenerative disorders, with the majority of mutations clustering within the C-terminal low-complexity domain (LCD) [26, 27].

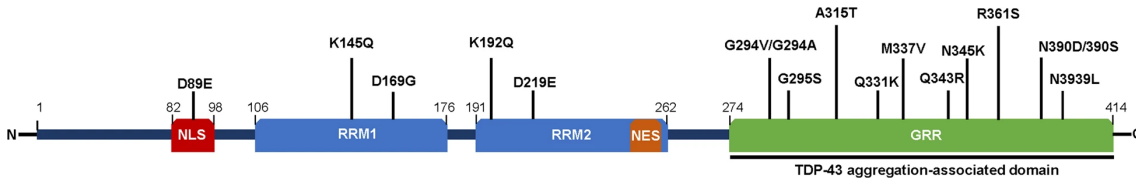


Figure 1.3: Disease-associated mutations in TDP-43. Most disease-associated TDP-43 mutations are concentrated in the C-terminal low-complexity domain (LCD) and have been shown to accelerate TDP-43 aggregation. (Figure adapted from Jo *et al.* [3].)

The expression of disease-associated TDP-43 mutants, including G294A, G294V, G295S, A315T, Q331K, M337V, Q343R, N345K, R361S, N390D, and N390S, some of them alter residue charge or hydrophobicity, has been shown to promote increased TDP-43 aggregation [27–30].

The LCD of TDP-43 mediates numerous protein–protein interactions, and aberrant interactions among these low-complexity regions are hypothesized to promote the maturation and fibrillization of phase-separated droplets [31, 32]. In addition to full-length TDP-43 aggregation, pathological cleavage and cytoplasmic accumulation of toxic TDP-43 C-terminal fragments (CTFs) represent another prominent hallmark of disease [33, 34]. These CTFs vary in size, reflecting proteolytic cleavage at multiple sites within the TDP-43 protein [35]. The C-terminal fragments lack the N-terminal nuclear localization signal (NLS) and RNA recognition motifs, which promotes their cytoplasmic mislocalization and aggregation propensity.

Across the spectrum of neurodegenerative disorders, aggregated TDP-43 is frequently modified by PTMs, such as ubiquitination, SUMOylation, and phosphorylation. Phosphorylated TDP-43 is a highly consistent marker of disease. TDP-43 contains 64 potential phosphorylation sites, comprising 41 serine (Ser), 15 threonine (Thr), and 8 tyrosine (Tyr) residues. Phosphorylation at Ser-379, Ser-403/404, and Ser-409/410 has been closely as-

sociated with TDP-43 proteinopathies [36]. Consistently, a mass spectrometry analysis of ALS spinal cord tissue identified phosphorylation at 14 serine residues, 12 of which reside within the C-terminal LCD of TDP-43 [37].

Various alterations of TDP-43—including cytoplasmic mislocalization, proteolytic cleavage generating C-terminal fragments, aggregation, and phosphorylation at serine residues within the C-terminal region—are hallmarks of neurodegenerative disorders. However, while these pathological features are consistently observed in patient tissues, their presence does not necessarily indicate that they are the primary drivers of disease. Instead, they may represent downstream consequences of upstream cellular perturbations, stress responses, or other molecular events that ultimately contribute to neuronal dysfunction and degeneration.

1.3 Molecular Dynamics

Molecular dynamics (MD) is a computational method used to map the motion of physical particles over time [38]. The motion of particle i is governed by Newton’s second law:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i \quad (1.1)$$

In this equation, m_i denotes the mass of particle i , and $\mathbf{r}_i(t)$ is its position vector as a function of time. The second derivative of the position with respect to time, $d^2 \mathbf{r}_i / dt^2$, represents the acceleration $\mathbf{a}_i$ of the particle. The velocity $\mathbf{v}_i$ of the particle is related to time by

$$\mathbf{v}_i = \mathbf{a}_i t \quad (1.2)$$

In MD simulations, the acceleration is generally not constant because $\mathbf{F}_i$ depends on the instantaneous positions of all particles. The term $\mathbf{F}_i$ is the total force acting on particle i , arising from interactions with other particles in the system. By computing $\mathbf{F}_i$ at each time step and integrating Newton’s equations of motion, the positions and velocities of all particles are updated, generating the trajectory of the system.

For particle i , the force $\mathbf{F}_i$ is computed as the negative gradient of the potential energy U with respect to its position:

$$\mathbf{F}_i = -\nabla_{\mathbf{r}_i} U \quad (1.3)$$

Here, $\nabla_{\mathbf{r}_i}$ is the gradient operator with respect to the spatial coordinates of particle i , defined as

1 Introduction

$$\nabla_{\mathbf{r}_i} = \left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial y_i}, \frac{\partial}{\partial z_i} \right) \quad (1.4)$$

The total potential energy U is a scalar quantity that represents the energy of the entire system, arising from the positions of all particles. The gradient operator converts this scalar energy landscape into vector forces acting on individual particles, and the negative sign ensures that each particle accelerates in the direction that decreases the system’s potential energy.

The total potential energy is typically decomposed into contributions from bonded interactions, which maintain the internal molecular geometry (bonds, angles, dihedrals, and improper torsions), and nonbonded interactions, which govern interactions between particles that are not directly bonded, such as van der Waals and electrostatic forces:

$$U = U_{\text{bonded}} + U_{\text{nonbonded}} \quad (1.5)$$

By integrating these equations of motion over small time steps, MD simulations generate a trajectory of particle positions and velocities, which can be applied to analyze systems as small as atoms undergoing a chemical reaction, and as large as a galaxy [39].

MD simulations are widely used to study proteins, nucleic acids, polymers, membranes, and other biological molecular systems, providing insights into their conformational dynamics, folding, interactions, and self-assembly processes that are often difficult to access experimentally.

1.3.1 Interactions in molecular dynamics

1.3.1.1 Bonded interactions

In MD, bonded interactions describe the forces within a molecule, capturing covalent bonds and the geometry of molecular structures through energy potentials. These interactions are essential for maintaining the integrity and conformational preferences of molecules. Bonded interactions typically include bond stretching, angle bending, dihedral (torsional) rotations, and improper torsions (Figure 1.4).

They are commonly modeled using harmonic potentials (1.6), which approximate the energy associated with small deviations from equilibrium values.

$$U(x) = \frac{1}{2}k(x - x_0)^2 \quad (1.6)$$

where x represents the instantaneous value of a bond length, bond angle, or dihedral

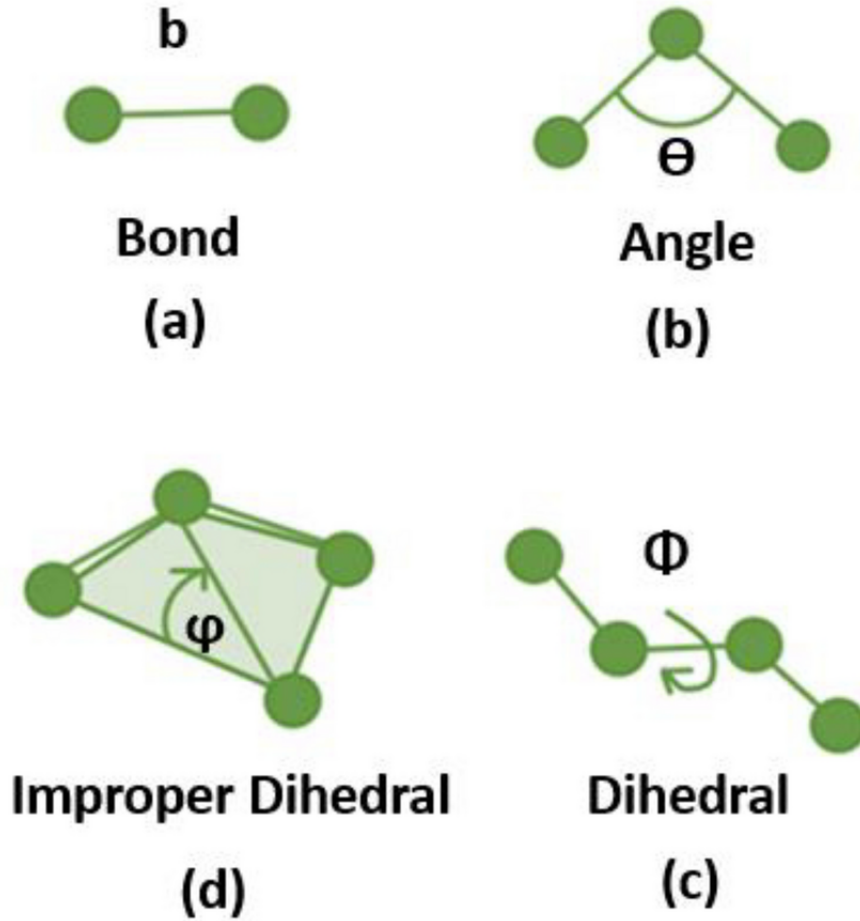


Figure 1.4: Molecular bonded interactions. Molecular mechanics interactions of (a) bond stretching, (b) angle bending, (c) dihedral, and (d) improper dihedral torsion (Figure from [4]).

angle, x_0 is the corresponding equilibrium value, and k is the force constant determining the stiffness of the interaction.

The total bonded potential energy of a molecule is given by:

$$U_{\text{bonded}} = \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \frac{1}{2} k_\varphi (\varphi - \varphi_0)^2 + \sum_{\text{impropers}} \frac{1}{2} k_\omega (\omega - \omega_0)^2 \quad (1.7)$$

where:

1 Introduction

- $r, \theta, \varphi, \omega$ are the instantaneous bond lengths, bond angles, dihedral angles, and improper angles, respectively,
- $r_0, \theta_0, \varphi_0, \omega_0$ are their equilibrium values,
- $k_b, k_\theta, k_\varphi, k_\omega$ are the corresponding force constants defined in the force field.

Bonded interactions maintain molecular geometry and flexibility, providing a physically realistic description of molecular structure while allowing the system to explore its conformational space.

1.3.1.2 Nonbonded interactions

Nonbonded interactions describe the forces between particles that are not connected by covalent bonds, primarily comprising van der Waals (VDW) interactions and electrostatic interactions. In the context of proteins, nonbonded interactions are critical for maintaining protein structure, modulating stability, and governing interactions with the surrounding solvent or other biomolecules.

VDW interactions arise from temporary or induced dipoles between atoms or molecules, which produce weak attractive forces at intermediate distances and strong repulsion when particles approach too closely. In MD simulations, VDW interactions are commonly modeled using the Lennard-Jones (LJ) potential:

$$U_{\text{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (1.8)$$

where r is the distance between two particles, ϵ is the depth of the potential well, representing the interaction strength (taken from the chosen force field), and σ is the distance at which the potential is zero, also provided by the force field. The first term (r^{-12}) accounts for short-range repulsion, while the second term (r^{-6}) accounts for long-range attraction.

Electrostatic interactions occur between charged or partially charged particles, including ion-ion, ion-dipole, and dipole-dipole interactions. These forces are typically described by Coulomb's law:

$$U_{\text{Coulomb}}(r) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r} \quad (1.9)$$

where q_i and q_j are the charges of the interacting particles, r is the distance between them, and ϵ_0 is the permittivity of free space, a physical constant. In biological systems, electrostatic interactions are often long-range and can be modulated by the surrounding

solvent and ionic strength, referring to protein folding, ligand binding, and membrane formation.

The total nonbonded potential energy in a system is the sum of all pairwise van der Waals and electrostatic interactions:

$$U_{\text{nonbonded}} = \sum_{i < j} [U_{\text{LJ}}(r_{ij}) + U_{\text{Coulomb}}(r_{ij})] \quad (1.10)$$

where the sum runs over all particle pairs i, j .

1.3.2 Coarse-Grained Simulations

In all-atom (AA) simulations, every atom in the system is explicitly represented, providing a detailed and accurate description of molecular interactions. Coarse-grained (CG) simulations reduce the total number of particles by grouping multiple atoms into single interaction sites or “beads.” This simplification makes the simulation of larger systems and longer timescales computationally feasible.

In CG models, the potential energy uses similar functional forms as in all-atom simulations:

$$U_{\text{CG}} = U_{\text{bonded}}^{\text{CG}} + U_{\text{nonbonded}}^{\text{CG}} \quad (1.11)$$

where $U_{\text{bonded}}^{\text{CG}}$ includes interactions such as bonds, angles, and dihedrals between beads, and $U_{\text{nonbonded}}^{\text{CG}}$ includes van der Waals and electrostatic interactions between beads that are not directly connected.

The reduction in degrees of freedom enables the study of mesoscale phenomena such as phase separation, protein aggregation, membrane remodeling, and self-assembly over microseconds to milliseconds, which are often inaccessible to all-atom simulations.

Force fields provide the bead types, bonded parameters (bonds, angles, dihedrals), and nonbonded interaction parameters required for coarse-grained simulations. Popular coarse-grained force fields include Martini [40–42], UNRES [43] and HPS [6], which are widely used in biomolecular and polymer simulations.

1.3.2.1 Chemically-detailed coarse-grained simulations

Chemically detailed coarse-grained simulation models reduce atomistic detail by grouping atoms into coarse-grained sites, while still preserving chemically distinct backbone and side-chain representations.

1 Introduction

The Martini force field, with explicitly represented water molecules, was initially developed for simulations of lipid and surfactant systems [44]. Subsequently, it was extended and refined to model proteins amino acids, enabling coarse-grained simulations of biomolecular systems [40, 41].

Martini 3 represents a further refinement of the force field, featuring improved interaction balance, the introduction of new bead types, and an updated water model, resulting in enhanced accuracy and transferability across diverse molecular systems [45].

In Martini 3, the secondary, tertiary, and quaternary structures of proteins often require stabilization through the application of additional interaction potentials to preserve native conformations during simulations. Harmonic potentials are commonly implemented in the form of an Elastic Network (EN) model, Lennard-Jone (LJ) potentials are employed in structure-based approaches such as GōMartini and OLIVES models.

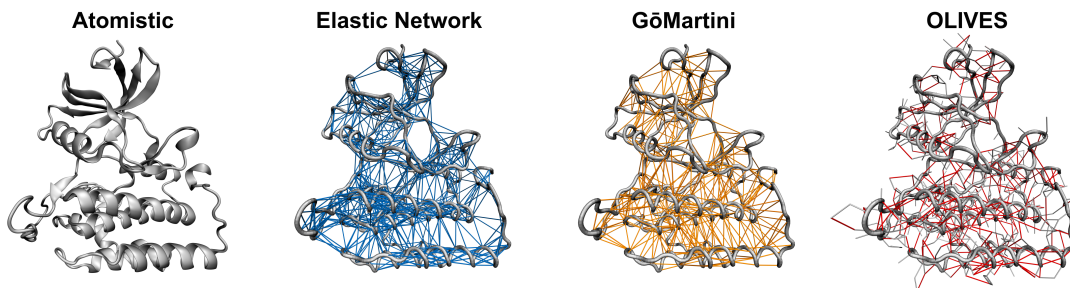


Figure 1.5: Comparison of structure-bias models used to stabilize coarse-grained protein conformations. Atomistic structure of protein CK1 δ in cartoon. The Elastic Network (EN) model (blue) introduces 1004 harmonic restraints, the GōMartini model (orange) adds 609 native-contact-based potentials, and the OLIVES model (red) includes 374 potentials. While the EN and GōMartini interaction networks connect predominantly backbone beads, the OLIVES model additionally incorporates side-chain bead interactions, providing a more extensive stabilization pattern. (Figure from [5])

Recently, to improve simulations of proteins containing both IDRs and structured domains (multi-domain protein, MDP), it has become common to release structural restraints on the IDRs, as described by Wang et al. (2025) [46]. The following command illustrates a typical setup, where an Elastic Network (EN) is applied to stabilize the structured domain (residues 1–292), while the IDR (residues 293–415) remains fully flexible using the `-idr-regions` and `-idr-tune` options:

```
1 martinize2 -f input_all_atom.pdb -x input_cg.pdb
2   -o protein_only.top
3   -ff martini3IDP -p backbone -dssp -elastic
4   -el 0 -eu 0.85 -eunit 1:292
5   -idr-regions 293:415 -idr-tune
```

The Hybrid Resolution (HyRes) protein model and its updated versions are chemically detailed coarse-grained approaches that preserve an atomistic representation of the protein backbone, while side chains are coarse-grained into up to five beads and solvent effects are treated implicitly. This level of resolution enables the HyRes model to capture secondary-structure elements such as α -helices and β -sheets within intrinsically disordered regions during liquid–liquid phase separation, as well as the internal structural organization of folded protein domains [47, 48].

1.3.2.2 Residue-level implicit coarse-grained simulations

Hydropathy Scale (HPS) and CALVADOS are residue-level coarse-grained (CG) simulation frameworks that treat solvent effects implicitly, with each amino acid represented by a single bead [49] (Figure 1.6). The updated versions, CALVADOS 2 and CALVADOS 3, incorporate improved residue parameters [49, 50].

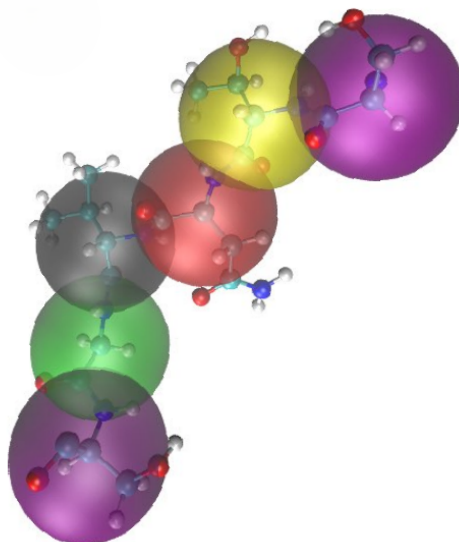


Figure 1.6: Residue-Level bead representation of proteins. All-atom representation is in balls and sticks, and coarse-grained representation in large spheres. (Figure adapted from [6])

These parameter sets were initially developed to model liquid–liquid phase separation (LLPS) of disordered proteins, capturing essential physicochemical properties such as residue hydropathy and charge distributions. With further development, the frameworks have been MDPs, where the structured domains can be stabilized either by treating them as rigid bodies or by applying an Elastic Network (EN) to maintain their native fold while allowing the disordered regions to remain flexible.

2 Phase behavior of TDP-43

Structurally, TDP-43 is a 414-amino acid protein, comprising three folded domains the N-terminal domain (NTD) and two RNA recognition motifs domains (RRM1 and RRM2) along with a long intrinsically disordered region (IDR) at the C-terminal and a nuclear localisation signal (NLS) sequence links the NTD to RRM1 (Figure 2.1).

Within the C-terminal IDR, also known as the low-complexity domain (LCD), a short helical structure has been identified as a critical structural element modulating TDP-43 liquid–liquid phase separation [51]. This helix forms transiently, with residues sampling both helical and coil conformations as evidenced by nuclear magnetic resonance (NMR) and molecular dynamics (MD) studies. The structured region encompasses residues 320 to 343, with residues 320–332 exhibiting higher helical propensity.

The near atomic coarse-grained Martini 3 with explicit solvent simulation model requires additional potentials to model the protein secondary, tertiary and quaternary structures during simulations [5]. In this chapter, to preserve the native three-dimensional folds, harmonic potentials between backbone beads within certain distance are applied, this approach is known as the Elastic Network (EN) model. However, in its baseline form, the force field underestimates global dimensions of multi-domain proteins, as shown by discrepancies with small-angle X-ray scattering (SAXS) measurements. Rescaling all protein–water interaction strength by a factor λ , including side chain beads (typically $\lambda_{1.06}$), is an approach to improve the representation of multi-domain protein conformations in accordance with experimental benchmarks [52].

This chapter focuses on the application of an adapted Martini 3 coarse-grained force field incorporating an EN model to elucidate the phase behavior underlying LLPS in the full-length protein and key variants. The multi-domain architecture of TDP-43, comprising both structured domains and extensive IDRs, provides an ideal framework for probing the complex interplay of interactions that govern its phase behavior.

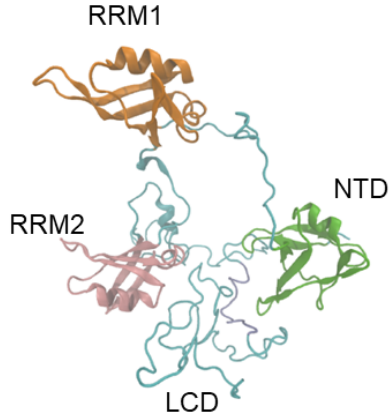


Figure 2.1: Representative structure of full-length TDP-43, with average R_g highlighted in all-atom simulations. Residues colored in green, orange, pink correspond to the N-terminal domain (NTD), RNA-recognition motif 1 (RRM1), and 2 (RRM2), respectively. Cyan areas indicate the protein’s disordered regions, and the region colored in purple represents the helical segment within the C-terminal IDR. water and ions are omitted for clarity for all-atom model).

2.1 Dilute-phase behavior of full-length TDP-43

A single TDP-43 chain was simulated in either a cubic or dodecahedral simulation box to characterize chain behavior in the dilute phase, using both all-atom (AA) and coarse-grained (CG) models. For AA simulations, dodecahedral boxes were employed to reduce the number of solvent molecules, thereby limiting computational cost while maintaining adequate solvation of the protein. The use of both AA and CG simulations enabled the characterization of protein conformational properties at different levels of resolution, providing complementary insights into the structural dynamics and flexibility of the chain.

The definitions of the NTD, RRM1, RRM2, and C-terminal helix used in this work may differ from those reported in published studies [53–57]. Rather than strictly adhering to literature-defined domain boundaries, this study adopts a structure-based definition. Across all simulations presented in this thesis, folded regions with confidently identified structures, specifically residues 4–76, 105–176, 192–259, and 320–332 as predicted by AlphaFold2 [58], were selected and treated as structured domains. These regions were stabilized using additional restraints to preserve their folded conformations during CG simulations.

For the purpose of simulation analysis, the following analysis-specific domain definitions were applied: the NTD (residues 1–80), RRM1 (residues 105–176), and RRM2 (residues 192–260). These definitions facilitate consistent quantitative analysis across different systems and do not necessarily coincide exactly with previously reported domain boundaries.

2.1.1 Single chain compactness

Protein chain conformational compactness was quantified using the radius of gyration (R_g), which measures the root-mean-square distance of all protein atoms from the center of mass (COM) [59]. This metric provides a global descriptor of the spatial extent of the protein chain and is widely used to assess changes in chain compaction or expansion.

R_g at time t is defined as:

$$R_g(t) = \sqrt{\frac{\sum_{i=1}^N m_i |\mathbf{r}_i(t) - \mathbf{r}_{cm}(t)|^2}{\sum_{i=1}^N m_i}}, \quad (2.1)$$

where N is the total number of atoms, m_i is atom weight mass, $\mathbf{r}_i(t)$ is the position of atom i at time t , and $\mathbf{r}_{cm}(t)$ is the center of mass of the protein at time t .

$$\mathbf{r}_{cm} = \frac{\sum_{i=1}^N m_i \mathbf{r}_i}{\sum_{i=1}^N m_i} \quad (2.2)$$

The R_g is mathematically equivalent to the mass-weighted standard deviation of the atomic positions around the COM, value provides a quantitative measure of the protein's overall spatial extent, a lower R_g indicates that atoms lie closer to the COM, corresponding to a more compact conformation, while a higher R_g signifies that the atoms are more widely dispersed, characteristic of an extended or partially unfolded state.

In AA molecular dynamics simulations, a single TDP-43 chain was simulated for 1 μ s in each of three replicas using the Amber99SB*-ILDN-q force field together with the TIP4P-D water model. The resulting conformational ensemble displayed substantial variability across replicas (Figure 2.2A), reflecting the inherent flexibility of TDP-43 as a partially disordered, multi-domain protein. This variability highlights the broad range of conformations that the protein can adopt even over microsecond timescales, encompassing both compact and more extended states. While these simulations provide valuable insights into the conformational preferences and structural dynamics of the protein, achieving full convergence of the ensemble would likely require considerably longer simulation times. Such extended sampling would be necessary to capture rare but potentially important conformational transitions and to obtain a more comprehensive representation of the protein's equilibrium behavior in solution.

Two of the three trajectories sampled predominantly extended conformations, with R_g values exceeding 4 nm, reflecting the protein's capacity to explore a broad conformational space. In contrast, the first trajectory remained largely restricted to more compact states, with $R_g \approx 2.5\text{--}3.0$ nm. Across all replicas, the mean R_g was 2.95 nm, indicating that, on average, the atomistic simulations may be somewhat biased toward overly compact

2 Phase behavior of TDP-43

conformations. This variability among replicas highlights the slow convergence inherent to simulations of partially disordered, multi-domain proteins on microsecond timescales. A representative structure with $R_g = 2.95$ nm is shown in Figure 2.1, illustrating a relatively compact arrangement of the IDR and providing a visual reference for the average conformational state sampled in these simulations.

Experimental SAXS measurements (black dashed line in Figure 2.2) provide additional context. One study reported the R_g of wild-type TDP-43 at 40°C, where the protein is expected to adopt more expanded conformations than at room temperature [60]. A separate SAXS study measured a TDP-43 mutant in which all tryptophan residues were substituted with alanine (6WA) [61]. Several Trp residues reside within the folded domains, these substitutions may destabilize domain packing and artificially increase the apparent R_g , which might partly explain the higher experimental value.

The 6WA mutant has also been simulated in AA simulations, which showed more expanded conformations with an average R_g of 3.07 ± 0.017 nm (Figure 2.3). However, whether the folded domains in this mutant are fully unpacked or partially disrupted has not yet been systematically studied. Further investigation is needed to determine how the Trp-to-Ala substitutions specifically affect the stability of folded domains and the overall conformational ensemble in dilute solution condition.

In the semi-coarse-grained simulations using the unrescaled Martini 3 force field, full-length TDP-43 predominantly remained in collapsed conformations throughout the trajectories, a trend that was consistently observed across independent replica simulations for the EN model at $\lambda_{1.00}$ (Figure 2.2B). As a representative structure from Figure 2.4A. These collapsed states indicate that, in the absence of modification, the original Martini 3 protein–water interactions favor overly compact conformations, which may not fully reflect the conformational flexibility expected for a multi-domain protein with extensive disordered regions. When protein–water interactions were rescaled, with strengthened protein–water interactions ($\lambda_{1.06}$), TDP-43 sampled more expanded conformations relative to the compact states observed in the unscaled ($\lambda_{1.00}$) or atomistic simulations (Figure 2.2B). The mean R_g value for the unscaled coarse-grained model ($\lambda_{1.00}$, 2.41 ± 0.021 nm) clearly showed the overly compact structures were favoured in absence of the rescaling of protein–water interactions, as illustrated in Figure 2.4B. By contrast, the rescaled model with $\lambda_{1.06}$ captures more extended conformations and exhibit greater variability, as higher mean R_g values of 3.55 ± 0.13 nm.

Rescaling protein–water interactions ($\lambda_{1.06}$) in the Martini 3 force field shifted TDP-43 toward more extended conformations in the dilute phase, effectively alleviating the excessive compactness observed in both the unscaled coarse-grained model and atomistic simulations. By strengthening protein–solvent interactions relative to protein–protein interactions, the rescaled model allows the intrinsically disordered regions to sample a broader range of conformations, better capturing the balance between folded domains and flexible regions.

2.1.2 Intra-molecular contacts

Variations in R_g reflect differences in intramolecular interactions and conformational flexibility, enabling quantitative comparison of chain behavior across different simulation models.

The three simulation models represented distinct levels of chain compactness and, therefore, produced clearly different intramolecular contact landscapes within full-length TDP-43 (Figure 2.5). Intrachain contacts were defined using a physical distance cutoff, enabling direct comparison across models with different resolutions and conformational ensembles.

In the AA simulations, TDP-43 adopted a relatively compact yet heterogeneous conformational ensemble, with the LCD playing a dominant role in mediating long-range intramolecular interactions (Figure 2.5A). Within the LCD, the strongest intra-domain contacts were observed, encompassing both short- and long-distance interactions, highlighting its intrinsic propensity to form transiently stabilized structures. Beyond domain self-contacts, the LCD frequently interacted with RRM2 as well as the second linker connecting RRM1 and RRM2, indicating that it repeatedly folded back toward the structured domains. Additional long-range interactions between the LCD and the NTD further reinforced a compact global architecture. The NLS displayed a characteristic pattern of strong self-contacts, consistent with a locally collapsed state, along with recurrent interactions with RRM1 and occasional contacts with the second linker. Together, these features indicated that the all-atom simulations produced a contact network dominated by extensive LCD-mediated interactions and stabilizing cross-domain contacts that compacted the chain.

The Martini 3 EN model at $\lambda_{1.00}$ preserved several qualitative features of the all-atom contact map but also introduced systematic differences (Figure 2.5B). LCD–LCD interactions remained highly enriched and formed the core of the intramolecular contact landscape. The LCD interacted with RRM2 and RRM1 more frequently than in the all-atom simulations, suggesting an even stronger tendency for the LCD to associate with the structured domains under the unscaled coarse-grained force field. This enhanced cross-domain engagement likely contributed to the exaggerated compactness observed in the $\lambda_{1.00}$ ensemble. Consistent with the AA simulations, the second linker exhibited recurring contacts with both the LCD and RRM2, indicating that this region functioned as a flexible connector across models. The NLS also maintained pronounced self-contacts and persistent interactions with RRM1 and the second linker. Overall, the $\lambda_{1.00}$ EN model amplified several compacting interactions present in the all-atom system, resulting in an intramolecular contact pattern biased toward excessive chain collapse.

By contrast, the $\lambda_{1.06}$ rescaled model produced a markedly different intramolecular contact network consistent with its more extended conformations (Figure 2.5C). Long-range cross-domain interactions were substantially reduced, and the contact map became dominated by local or domain-proximal interactions. LCD–LCD contacts were weakened but remained detectable, indicating that the LCD sampled flexible but less collapsed states.

2 Phase behavior of TDP-43

Strong NLS self-contacts persisted across all models, and NLS–RRM1 interactions continued to occur. Contacts involving the second linker remained frequent. Overall, the $\lambda_{1.06}$ model exhibited a more diffuse and sparse intramolecular contact pattern, reflecting an ensemble in which stabilizing long-range contacts were diminished. This reduction in cross-domain interactions was fully consistent with the expanded conformational states indicated by the R_g measurements.

All-atom and $\lambda_{1.00}$ Martini 3 models both favored compact conformations driven by extensive LCD-mediated and cross-domain contacts, whereas the $\lambda_{1.06}$ rescaled model disrupted many of these long-range interactions, producing a more extended, less interconnected contact network. The intramolecular contact patterns reflected the degree of chain compactness imposed by each simulation model.

2.1 Dilute-phase behavior of full-length TDP-43

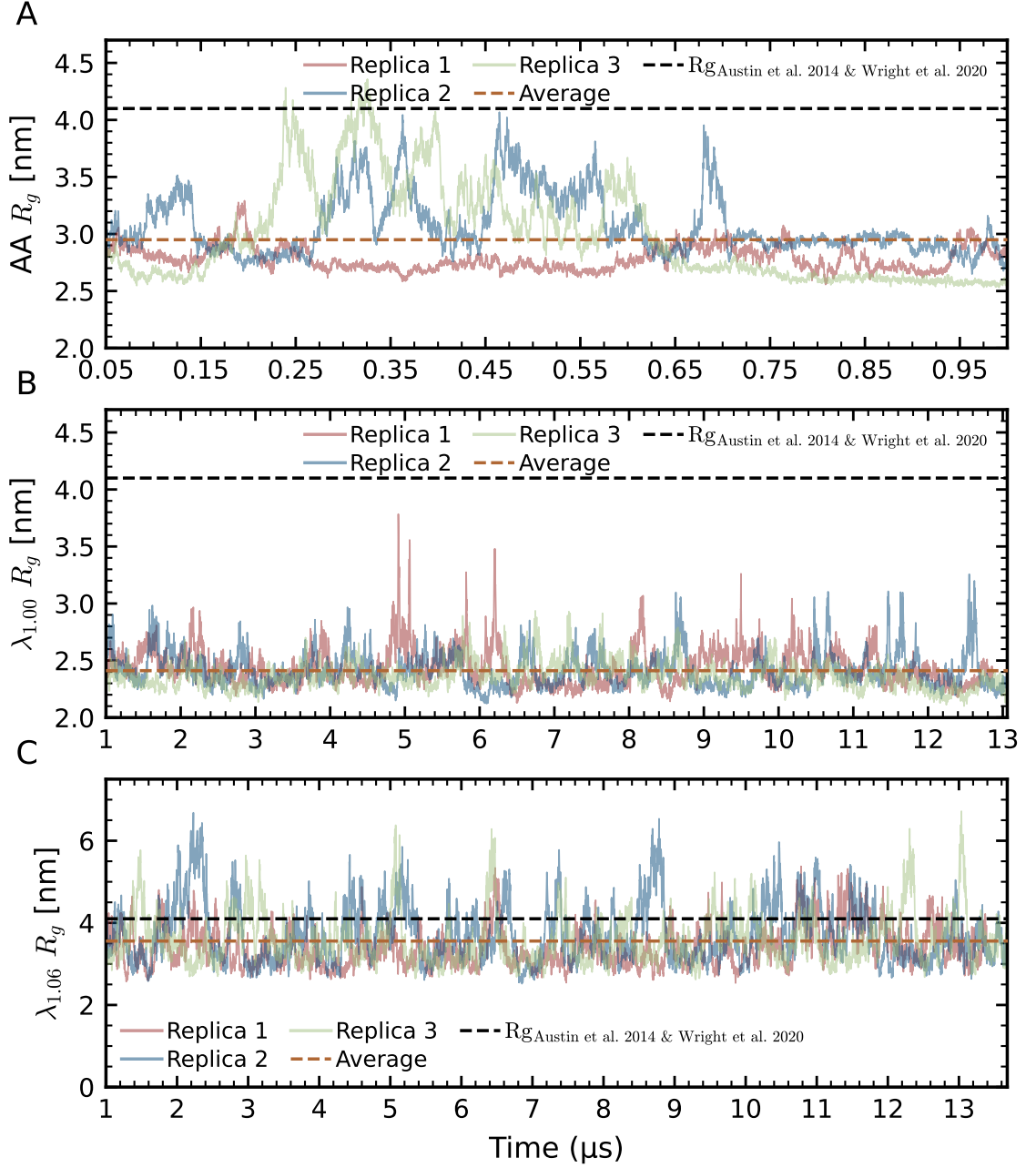


Figure 2.2: R_g measurement of single chain full-length TDP-43 under all-atom and Martini 3 Elastic Network (EN) model at $\lambda_{1.00}$ and $\lambda_{1.06}$. (A). R_g across three replicates in all-atom simulation, the black dashed line marks the maximum experimental R_g and the orange-like dashed line shows the mean values across replicas. (B). R_g over three replicates in Elastic Network model at $\lambda_{1.00}$, with the same dashed-line conventions. (C). R_g over three replicates in Elastic Network model at $\lambda_{1.06}$, with the same dashed-line conventions.

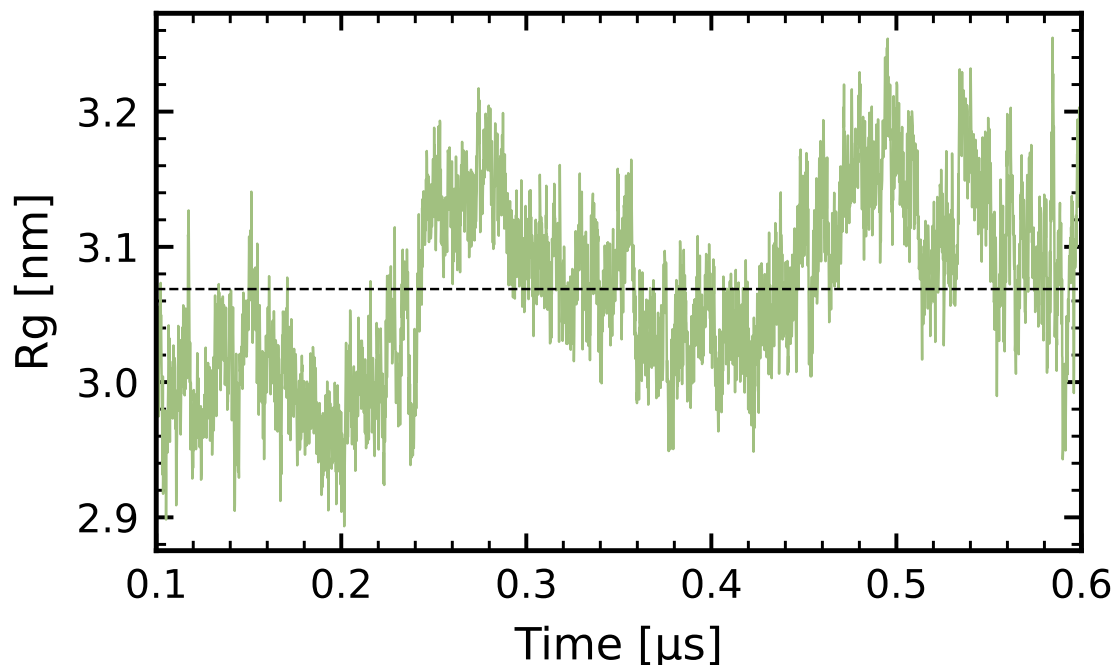


Figure 2.3: R_g measurement of single chain full-length TDP-43 6WA under all-atom (AA) simulations. The black dashed line indicates the averaged R_g value in AA simulations.

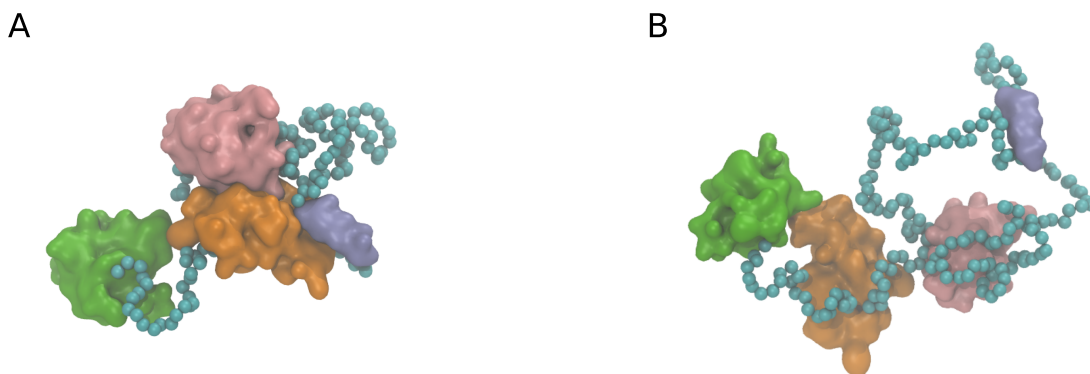


Figure 2.4: Representative structure of full-length TDP-43, with average R_g highlighted in Martini 3 Elastic Network (EN) at $\lambda_{1.00}$ and $\lambda_{1.06}$ simulations. (A). Representative structure in EN $\lambda_{1.00}$. (B). Representative structure in EN $\lambda_{1.06}$. Residues colored in green, orange, pink correspond to the N-terminal domain (NTD), RNA-recognition motif 1 (RRM1), and 2 (RRM2), respectively. Cyan areas indicate the protein's disordered regions, and the region colored in purple represents the helical segment within the C-terminal IDR. water and ions are omitted for clarity for EN model.

2.1 Dilute-phase behavior of full-length TDP-43

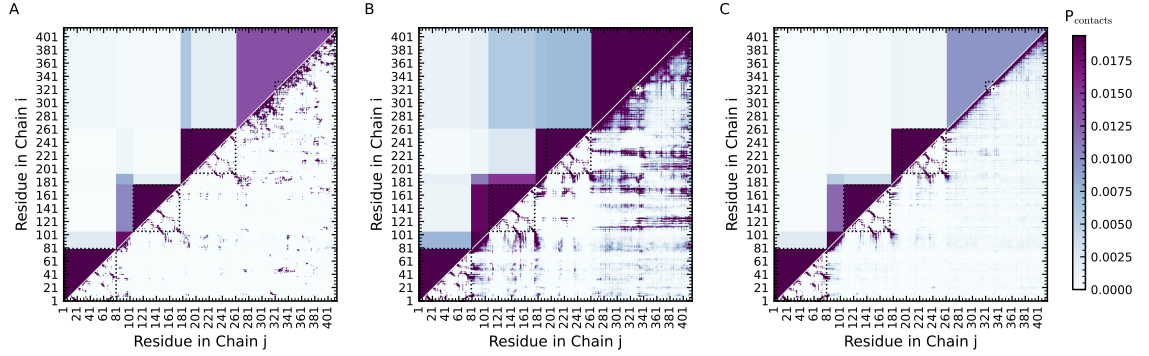


Figure 2.5: Averaged intra-chain contact heatmap of single chain full-length TDP-43 under under all-atom (AA) and Martini 3 Elastic Network (EN) model at $\lambda_{1.00}$ and $\lambda_{1.06}$. (A). AA simulations. (B). EN model at $\lambda_{1.00}$. (C). EN model at $\lambda_{1.06}$. Upper triangles are the residue-wise contact maps, lower triangles are domain-wise contact maps, which summing up the contact possibility and averaged by the total residue number within the specific domain, colorbar on the right represents the residue contact probabilities.

2.2 Dense-phase behavior of full-length TDP-43

2.2.1 Full-length protein phase separation

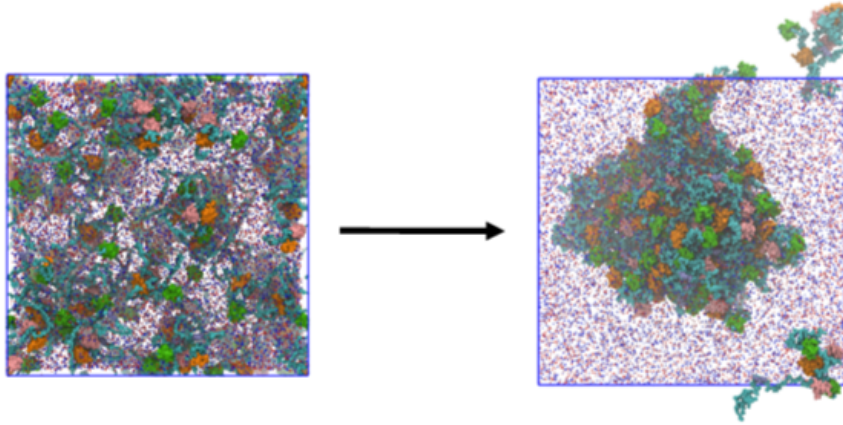


Figure 2.6: Full-length TDP-43 multichain condensation in cubic simulation geometry with Martini 3 Elastic Network (EN) at $\lambda_{1.06}$. Initial configuration of the EN model of full-length TDP-43 simulated with protein–water interaction scaling of $\lambda_{1.06}$ in a cubic simulation box. Starting from a dispersed arrangement, the chains assemble into a liquid-like spherical condensate. Na^+ (red) and Cl^- (blue) ions are shown; water beads are omitted for clarity.

To investigate multi-chain phase separation, Martini 3 simulations were performed with $\lambda_{1.06}$ rescaling of protein–water interactions to model 70 full-length TDP-43 chains in a 50 nm cubic box. The simulation was initialized with chains dispersed randomly throughout the simulation volume, ensuring no preformed aggregates. Over the course of the simulation, the system consistently underwent phase separation, ultimately producing a stable, spherical condensate (Figure 2.6). Initial clustering of chains became apparent after approximately 5 μs , indicating the onset of coalescence and nucleation of the condensed phase. By around 8 μs , the condensate had adopted a well-defined, liquid-like spherical morphology, as shown in Figure 2.6 (right). The condensate remained stable throughout the remainder of the simulation.

To quantify phase behavior, simulations were performed using a slab geometry [62], which facilitates faster equilibration between the dense and dilute phases by providing a large interface for phase coexistence. The Martini 3 force field was employed with varied protein–water interaction strengths (λ) to mimic changes in solvent conditions, such as effective decreases or increases in temperature. By systematically adjusting λ , the simulations allowed exploration of how changes in solvent quality influence the formation, stability, and composition of the protein condensates.

Simulations were performed with 60 chains in a slab geometry of dimensions $20 \times 20 \times 90$ nm.

2.2 Dense-phase behavior of full-length TDP-43

To expedite chain association, systems were initially condensed in a smaller simulation box before being expanded to the target slab dimensions. All systems were neutralized with 150 mM NaCl to approximate physiological ionic conditions. The protein–water interaction parameter (λ) was set to 1.05, 1.06, and 1.07, with three independent replicates conducted for each condition to ensure statistical robustness. Production runs were carried out for about 20 μ s, with some simulations extending beyond 30 μ s. Analyses were performed on the final 15 μ s of each trajectory, focusing on equilibrated behavior and stable properties of the phase-separated systems.

Tracking the concentration histograms over time provides a dynamic perspective on condensate stability and molecular exchange between the dense and dilute phases (Figure 2.7). In condensates simulated with $\lambda_{1.06}$, a fraction of molecules exchanged into the dilute phase over the simulation timescale of tens of μ s, indicating that the dense phase maintained liquid-like dynamics while allowing ongoing molecular turnover (Figure 2.7B). Figure 2.7E illustrates the corresponding condensate morphology. The exchange dynamics can also be observed in other replicates (Figure 2.8).

When protein–water interactions were reduced to $\lambda_{1.05}$, the number of proteins in the dilute phase decreased, reflecting stronger intermolecular attractions, more stable phase separation, and reduced chain exchange between phases (Figure 2.7A and D). Conversely, increasing the protein–water interaction strength to $\lambda_{1.07}$ weakened protein–protein interactions, producing a dense-phase region that appeared more diffuse and heterogeneous and compromising overall condensate stability (Figure 2.7C and F).

These observations highlight that subtle changes in protein–water interactions can modulate both the extent of phase separation and the dynamics of molecular exchange within and across phases. Such tunability underscores the critical role of solvent-mediated interactions in controlling the structural and dynamical properties of biomolecular condensates, with important implications for understanding how environmental influences phase behavior.

The equilibrium concentration of TDP-43 in the dilute phase, corresponding to the saturation concentration (C_{sat}) measured in *in vitro* experiments, was quantified by analyzing the simulation density histograms. Specifically, the time-averaged protein density profile along the z -axis was fitted to a hyperbolic tangent (tanh) function to parameterize the phase boundaries (Method 5.2.1). This fitting procedure yielded quantitative estimates of both the dilute- and dense-phase protein concentrations (Figure 2.9).

At $\lambda_{1.05}$, phase separation was highly stable and only limited molecular exchange was observed. Among the three independent replicates, only replicate 3 exhibited a single dissociation event, in which one TDP-43 chain transiently escaped from the condensate into the dilute phase before eventually being reabsorbed. This rare event contributed to the estimated dilute-phase concentration. The experimentally reported saturation concentration of TDP-43 at 150 mM NaCl, 0.0041 mM [63], is in reasonable agreement with

2 Phase behavior of TDP-43

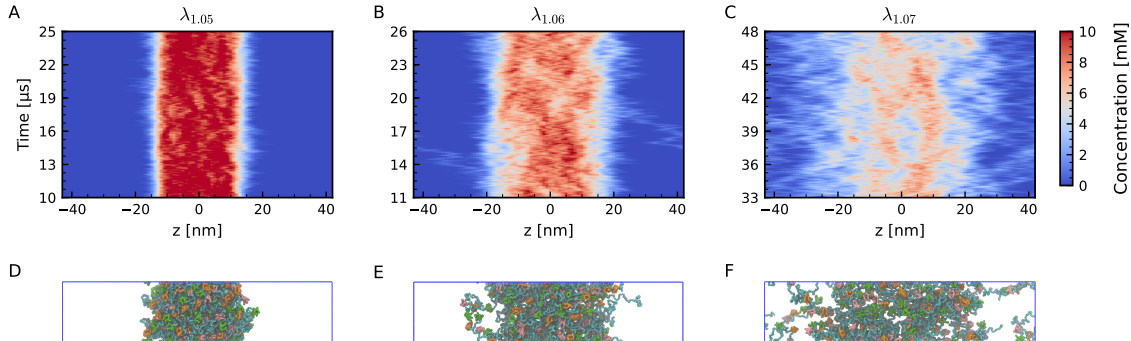


Figure 2.7: Phase separation of full-length TDP-43 in Martini 3 Elastic Network (EN) simulations within slab geometry. (A–C) Time-dependent concentration histograms for simulations with the Martini 3 EN model at protein–water interaction strengths of $\lambda_{1.05}$, $\lambda_{1.06}$, and $\lambda_{1.07}$. Protein concentration is indicated by the color bar. (D–F) Snapshots of the protein condensate along the z-axis for the Martini 3 EN model at $\lambda_{1.05}$, $\lambda_{1.06}$, and $\lambda_{1.07}$, respectively.

the dilute-phase concentration obtained from simulations at $\lambda_{1.05}$ (0.0020 ± 0.0018 mM). However, because this estimate is dominated by a single dissociation event occurring in only one replicate, the resulting C_{sat} value suffers from limited statistical robustness and should be interpreted with caution.

In contrast, simulations performed at $\lambda_{1.06}$ displayed consistent and reproducible chain exchange dynamics across all replicates. TDP-43 molecules repeatedly exited and re-entered the condensate on microsecond timescales, indicative of a liquid-like dense phase with ongoing molecular turnover. This dynamic equilibrium enabled a well-defined separation between dense and dilute phases and yielded a reproducible estimate of the dilute-phase concentration, making $\lambda_{1.06}$ the most reliable condition for quantitative analysis.

At $\lambda_{1.07}$, the condensate exhibited clear signs of destabilization. The dense phase became more diffuse and heterogeneous, and the distinction between dense and dilute regions was progressively blurred. As a result, identifying phase boundaries and extracting reliable concentration profiles became increasingly challenging. This behavior is consistent with weakened effective intermolecular attractions and reflects a shift toward partial dissolution of the condensate.

Although the simulated C_{sat} at $\lambda_{1.05}$ shows encouraging agreement with experimental measurements, the lack of sustained molecular exchange limits confidence in this apparent correspondence. By contrast, $\lambda_{1.06}$ produced the most robust and reproducible exchange dynamics across all replicates, and therefore all subsequent analyses in this study were conducted under this condition.

As protein–water interactions were progressively strengthened, the condensate became

2.2 Dense-phase behavior of full-length TDP-43

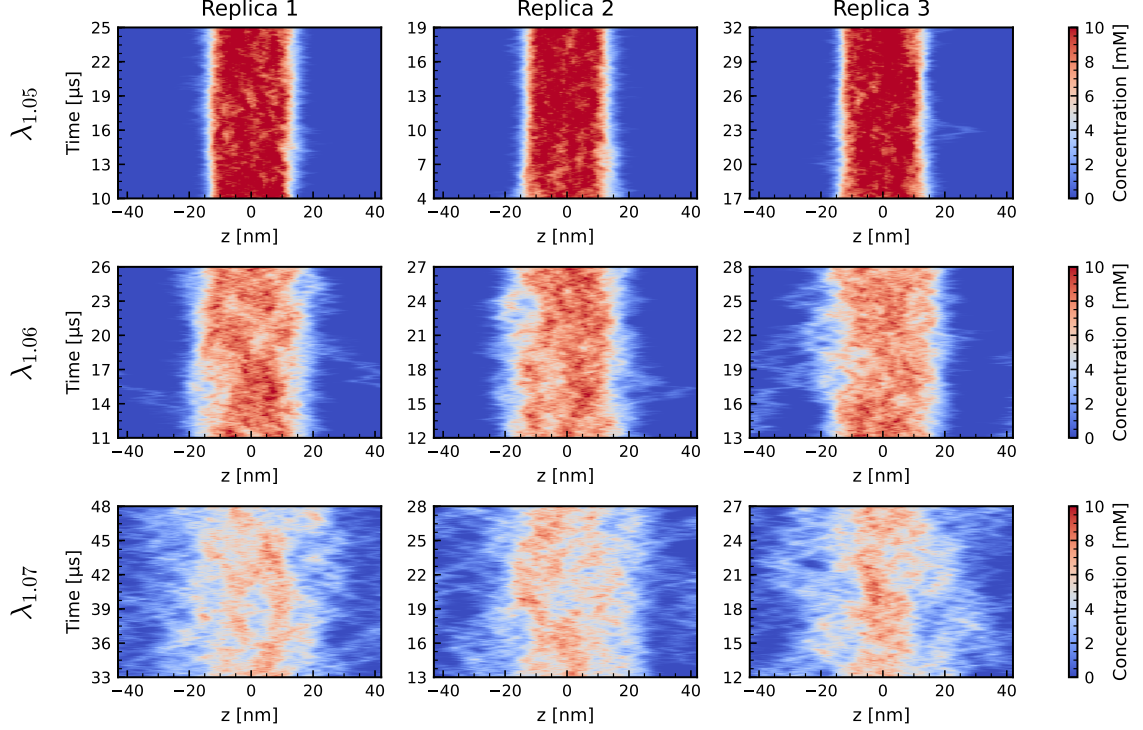


Figure 2.8: Time-dependent concentration histograms for full-length TDP-43 in Martini 3 Elastic Network (EN) simulations within slab geometry, shown for replicates at protein–water interaction strengths of $\lambda_{1.05}$, $\lambda_{1.06}$, and $\lambda_{1.07}$.

increasingly unstable and ultimately showed a tendency toward dissolution. Among the conditions tested, $\lambda_{1.06}$ yielded the most consistent, stable, and reproducible phase behavior, characterized by well-defined dense and dilute phases and clear evidence of ongoing molecular exchange between them. On this basis, $\lambda_{1.06}$ was selected as the reference condition for all subsequent analyses of TDP-43 variants.

Regardless of whether cubic or slab simulation geometry was employed, the Martini 3 Elastic Network model with $\lambda_{1.06}$ consistently produced liquid-like condensates. In both geometries, the condensates exhibited characteristic features of a liquid state, including a well-defined dense phase, dynamic molecular rearrangements, and continuous exchange of protein chains between the dense and dilute phases. The reproducibility of liquid-like behavior across distinct simulation setups demonstrates that the observed phase separation is not an artifact of the simulation geometry but instead reflects an intrinsic property of the model under this interaction scaling. These results further support $\lambda_{1.06}$ as a robust parameter choice for capturing physiologically relevant condensate behavior of full-length TDP-43.

2 Phase behavior of TDP-43

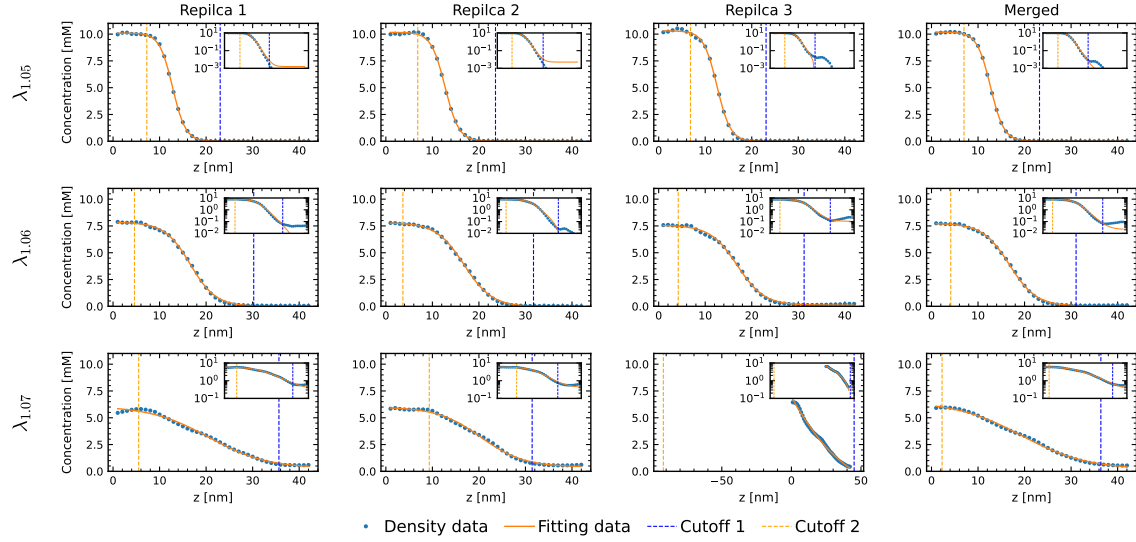


Figure 2.9: Phase concentration quantification in Martini 3 Elastic Network (EN) simulations within slab geometry for replicates. Density along the z-axis fitting results across replicates on $\lambda_{1.05}$, $\lambda_{1.06}$ and $\lambda_{1.07}$.

2.2.2 Inter-molecule interactions

2.2.2.1 Overall interactions

The intermolecular interaction network across full-length TDP-43 plays a central role in driving condensate formation and maintaining its stability. To systematically characterize these interactions, inter-chain contacts were quantified using a defined physical distance cutoff and analyzed across equilibrated portions of the trajectories. The resulting contact probabilities were visualized as heatmaps, enabling the identification of specific residues and protein regions that contribute most strongly to intermolecular association within the condensate (Figure 2.10A and B).

Residue-wise contact heatmaps were generated by computing the probability that each residue forms contacts with residues on neighboring chains, providing high-resolution insight into interaction hotspots along the protein sequence. In parallel, domain-wise contact heatmaps were constructed by summing all intermolecular contacts within each predefined domain and normalizing by the number of residues in that domain. This domain-wise representation facilitates comparison of interaction strengths between structured domains and disordered regions, allowing assessment of their relative contributions to condensate cohesion. Together, these complementary analyses provide a comprehensive view of the intermolecular interaction architecture that stabilizes TDP-43 condensates.

The majority of intermolecular contacts were localized within intrinsically disordered regions (IDRs), most prominently the low-complexity domain (LCD) and the nuclear local-

2.2 Dense-phase behavior of full-length TDP-43

ization sequence (NLS; residues 80–104) (Figure 2.10A). In contrast, residues within the folded domains exhibited relatively weak intermolecular interactions and rarely engaged in direct contacts with other structured regions, indicating that structured–structured domain interactions do not substantially contribute to condensate stabilization.

The LCD played a dominant role in mediating intermolecular interactions within the condensate. It engaged extensively in homotypic LCD–LCD contacts between neighboring chains and also formed heterotypic interactions, albeit to a lesser extent, with the folded domains. Within the LCD, a helical core region spanning residues 320–332 emerged as a prominent interaction hotspot, forming frequent contacts with equivalent helical regions on adjacent chains (Figure 2.10A). In addition to these homotypic interactions, this helical segment also contacted residues within the RRM2 domain, suggesting a potential role in bridging disordered and structured regions across chains.

Aromatic residues within the LCD were particularly enriched at intermolecular interfaces (Figure 2.10B). Specifically, eight phenylalanine residues (Phe276, Phe283, Phe289, Phe313, Phe316, Phe367, Phe397, and Phe401) and three tryptophan residues (Trp334, Trp401, and Trp412) frequently participated in inter-chain contacts. The prevalence of these aromatic interactions is consistent with their known roles in driving phase separation through π – π and hydrophobic interactions [64, 65], underscoring the importance of sequence-encoded interaction motifs within the LCD for condensate formation and stability.

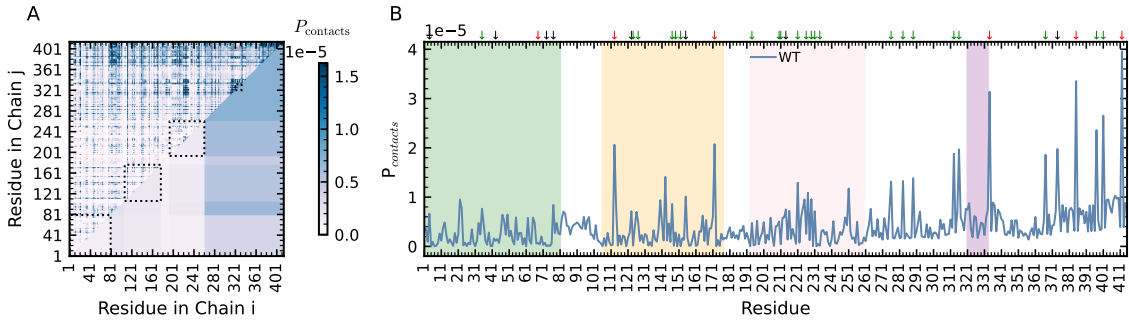


Figure 2.10: Inter-molecular contacts within condensates. A. Inter-chain contact heatmaps of Elastic Network with $\lambda_{1.06}$ in the slab geometry, upper triangle is the residue-wise contact maps, lower triangle is domain-wise contact maps, residue- and domain-wise contact maps are sharing same colorbar. B. Averaged inter-molecular contact possibility of each residue, tryptophan residues are indicated by red arrows, phenylalanine residues by green arrows, and other aromatic residues by black arrows.

Similarly, the NLS formed effective intermolecular contacts not only with NLS regions on neighboring chains but also with the extended LCDs, highlighting its role as a mediator of multivalent interactions within the condensate. Although NLS–NLS interactions were not strongly enriched in the domain-wise contact map, residues within the NLS exhibited

2 Phase behavior of TDP-43

distinct peaks in the residue-wise contact profile (Figure 2.10B), indicating recurrent interactions at the residue level. A representative structure of NLS–NLS contacts is shown in Figure 2.11A, illustrating how these regions can transiently associate across chains. While these spikes within NLS region were less pronounced than those of the aromatic residues, they were consistently higher than those of other non-aromatic residues within the LCD, indicating a measurable contribution of the NLS to intermolecular interactions.

In addition to NLS–NLS interactions, the NLS frequently engaged with the LCD of neighboring chains, forming a network of heterotypic interactions that likely stabilizes the condensate. Representative NLS–LCD contacts are depicted in Figure 2.11B. Within these NLS–LCD interactions, Residue Met in the NLS contributed noticeably, while Trp residues within LCD formed frequent contacts with positively charged residues (Lys and Arg) in the NLS (Figure 2.11C). These observations indicate that, although the NLS is shorter and less aromatic than the LCD, it plays a complementary role in mediating inter-chain associations, contributing to both the structural integrity and dynamic behavior of TDP-43 condensates. This highlights the importance of specific residue types within IDRs, aromatic and charged residues, in establishing multivalent contacts that drive liquid-like condensate formation.

Overall, the LCD and NLS emerged as the primary drivers of inter-chain interactions in TDP-43 condensates. Within these intrinsically disordered segments, aromatic residues (Trp and Phe) and positively charged residues (Arg and Lys) exhibited the highest contact frequencies, forming recurrent interaction hotspots that facilitate multivalent associations across chains. The aromatic residues likely contribute through a combination of π – π stacking and hydrophobic interactions, while the positively charged residues promote electrostatic and cation– π contacts. Together, these residues establish a dense and interconnected network of inter-chain contacts that stabilizes the condensed phase, providing both structural cohesion within the condensate.

2.2.2.2 LCD in condensates

Although residues within the LCD contributed extensively to inter-chain interactions in the context of full-length TDP-43, the isolated LCD fragment alone was insufficient to drive phase separation under the same solvent conditions ($\lambda_{1.06}$) (Figure 2.12A). In contrast, full-length TDP-43 readily formed condensates under these conditions (Figure 2.12B), indicating that the cooperative action of additional domains, including the NTD, RRM1, RRM2, and NLS, is required to achieve robust phase separation. These observations highlight that, while the LCD provides key multivalent interaction motifs, the presence of other structured and disordered regions within the full-length protein is critical for establishing a network of interactions capable of driving and stabilizing liquid-like condensates.

Within the simulation system, a single full-length TDP-43 protein contains more beads than an isolated LCD fragment. In all simulations, the total number of protein beads and the overall system size were kept constant to enable direct comparisons. The helical

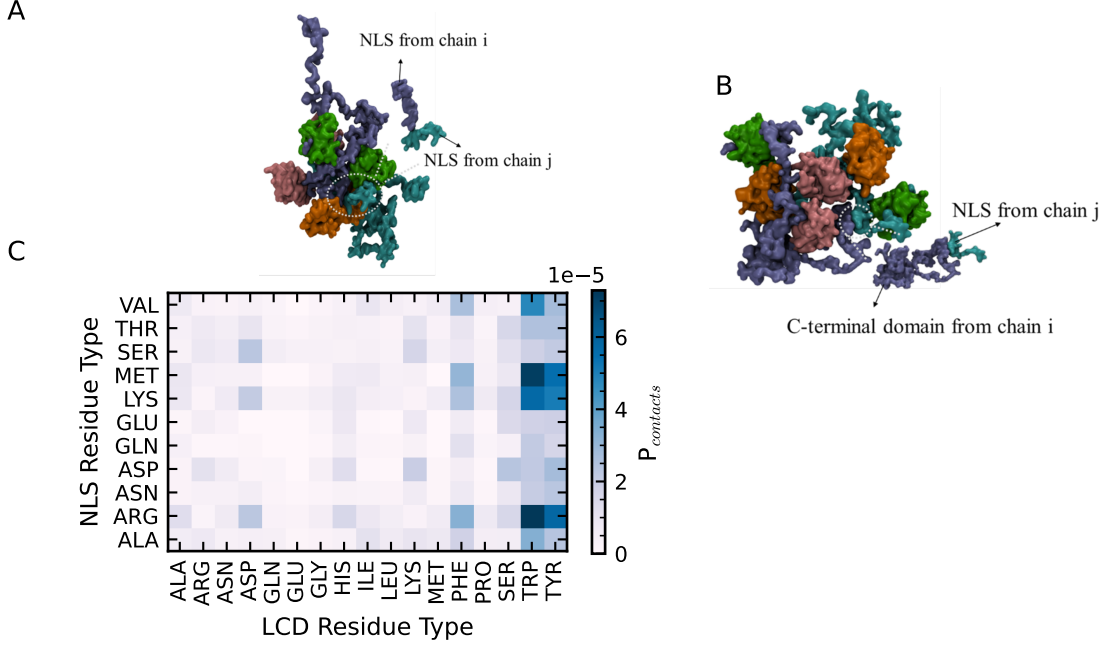


Figure 2.11: Inter-chain contacts involving the NLS under Martini 3 Elastic Network (EN) simulations at $\lambda_{1.06}$. (A). Snapshot illustrating NLS–NLS inter-chain contacts. (B). Snapshot illustrating NLS–LCD inter-chain contacts. Folded domains (NTD, RRM1, RRM2, and the C-terminal helix) are shown in green, orange, and pink, respectively, while intrinsically disordered regions from different chains are colored in cyan and purple. (C). Residue-type contact heatmap of NLS–LCD inter-chain interactions.

structure within the LCD was stabilized using the EN model. Under $\lambda_{1.06}$ conditions, the LCD fragments did not accumulate to form a clearly detectable condensate. In contrast, at $\lambda_{1.04}$, LCD fragments readily aggregated, forming a well-defined dense phase (Figure 2.12A and C). At these lower values of λ , stable LCD condensates were consistently observed. However, upon increasing the interaction parameter to $\lambda_{1.06}$, LCD condensates that had formed at $\lambda_{1.04}$ dissolved (Figure 2.12B), indicating that the LCD alone is insufficient to sustain condensates under conditions that favor full-length TDP-43 phase separation.

Despite having an identical number of beads in the simulation box, full-length TDP-43 readily formed a condensate under $\lambda_{1.06}$, whereas the isolated LCD failed to do so. This observation was mirrored experimentally, under identical buffer conditions and protein concentration, the isolated LCD fragment failed to form condensates, whereas the full-length protein (WT) phase separated (Figure 2.12E). This direct comparison indicates that the phase behavior of the full-length protein cannot be recapitulated by its LCD alone and suggests that additional regions of the protein enhance condensate stability and resilience. Consistently, experimental deletion of residues 1–101 abolished condensate formation, and partial removal of other regions also prevented condensate formation, highlighting the

2 Phase behavior of TDP-43

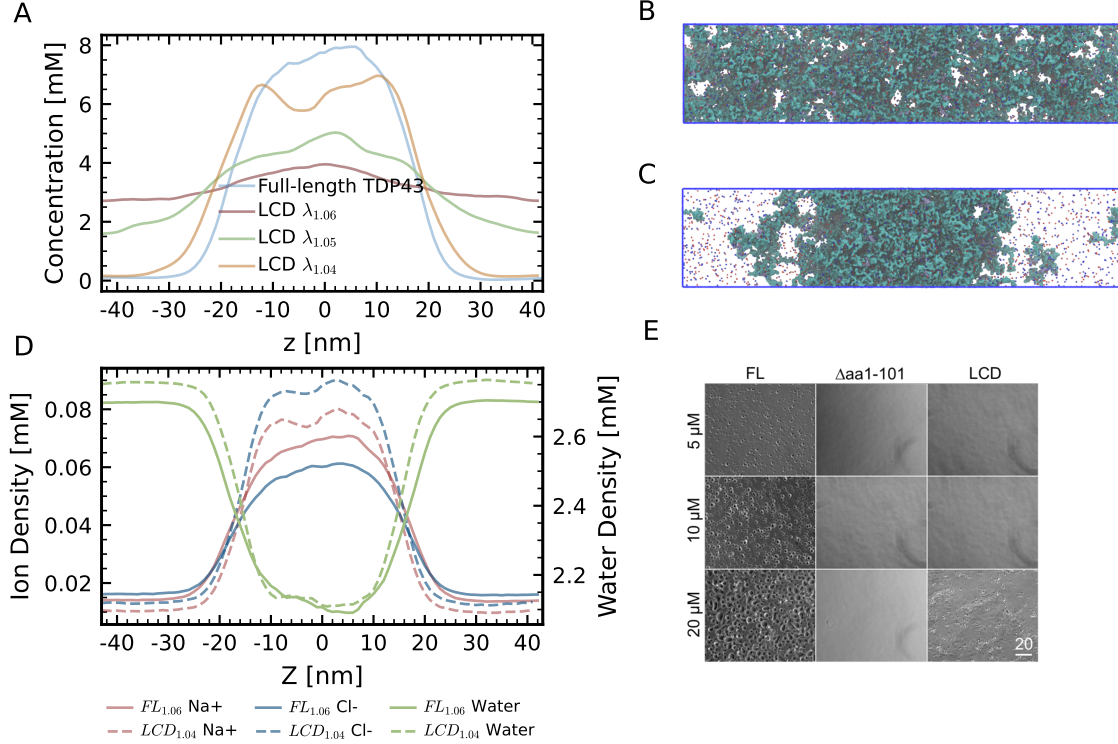


Figure 2.12: LCD phase separation under Martini 3 Elastic Network (EN) simulations. (A). Protein density profiles along the z-axis. Full-length TDP-43 was simulated with $\lambda_{1.06}$, and LCD chains were simulated with $\lambda_{1.06}$, $\lambda_{1.05}$, and $\lambda_{1.04}$ in the slab geometry. (B). Representative snapshot of LCD condensate at $\lambda_{1.06}$. (C). Representative snapshot of LCD forming a condensate at $\lambda_{1.04}$. (D). Ion and water density profiles along the z-axis for full-length TDP-43 ($\lambda_{1.06}$) and LCD ($\lambda_{1.04}$). (E) *In vitro* experiments testing phase separation of full-length TDP-43, a TDP-43 variant lacking residues 1–101, and the isolated LCD at different protein concentrations. Experimental data in panel E were kindly provided by Dr. Saskia Hutten (Dormann Group).

cooperative contribution of multiple protein segments to robust phase separation.

The LCD is enriched in aromatic residues, which are key drivers of phase separation [66], and participates in numerous inter-chain interactions, including LCD–LCD contacts as well as interactions with the RRM2, RRM1, NTD, and NLS regions of neighboring chains. This indicates that the stabilizing contributions from the folded regions cannot be overlooked. Although the isolated LCD possesses intrinsic multivalency, its interactions may be too weak or transient to maintain a stable condensate under the more stringent conditions simulated at $\lambda_{1.06}$. In contrast, the full-length protein presents a composite interaction network in which structured domains provide crucial, complementary adhesive motifs. These folded regions might likely act not merely as passive spacers but as active partici-

pants, organizing and reinforcing the LCD-mediated network.

A key difference between the LCD and full-length TDP-43 lies in their net charge per residue (NCPR). The LCD, spanning residues 261–414, contains 3 Glu, 1 Asp, 5 Arg, and 2 Lys residues, resulting in a net charge of +3. It also contains 1 His, which is treated as neutral in the Martini 3 force field, giving the LCD an NCPR of approximately +0.019. In contrast, the full-length protein (414 residues) contains 5 His, 20 Arg, 20 Lys, 22 Glu, and 22 Asp residues, resulting in a net charge of −4 and an NCPR of roughly −0.001. Consequently, LCD condensates bind more ions than condensates of the full-length protein (Figure 2.12D), likely weakening effective intermolecular protein–protein interactions and disfavoring phase separation.

This observation can be understood within a broader biophysical framework. While it is generally true that, under identical conditions, increasing chain length generally favors phase separation [67]. For protein TDP-43, folded regions not only increase effective chain length but also act as essential organizers of the condensation network. Their presence is required to generate specific interactions that nucleate and stabilize the condensate, going beyond the simple collapse of the isolated LCD.

Full-length TDP-43 undergoes phase separation under $\lambda_{1.06}$ conditions, whereas its isolated LCD fragment fails to form stable condensates despite possessing intrinsic multivalency and aromatic residues that mediate inter-chain interactions. Simulations and experiments consistently show that additional regions of the protein, including folded domains, are required to stabilize condensates. Differences in net charge per residue between the LCD and full-length protein also modulate ion binding and effective intermolecular interactions, further influencing phase behavior. These results highlight that both disordered and structured regions cooperatively drive TDP-43 condensation, demonstrating that phase separation is governed not only by intrinsic LCD properties but also by contributions from the full-length protein architecture.

2.3 Phase behavior of TDP-43 Variants

Aromatic residues have been shown to exhibit high contact probabilities in intermolecular interactions. In addition, twelve serine residues within the LCD (Ser373, Ser375, Ser379, Ser387, Ser389, Ser393, Ser395, Ser403, Ser404, Ser407, Ser409, and Ser410) have been identified as critical drivers of phase separation [68]. To assess how mutations at these residues affect phase behavior, the transfer free energy, $\Delta G_{\text{transfer}}$, was introduced to quantify the thermodynamic preference of a protein to reside in the dense (condensed) phase relative to the dilute phase. This quantity is defined from the ratio of protein concentrations in the two coexisting phases at equilibrium,

$$\Delta G_{\text{transfer}} = -RT \ln \left(\frac{c_{\text{dense}}}{c_{\text{dilute}}} \right), \quad (2.3)$$

2 Phase behavior of TDP-43

where c_{dilute} and c_{dense} are the protein concentrations in the dense and dilute phases, respectively. A more negative $\Delta G_{\text{transfer}}$ indicates a stronger thermodynamic driving force for phase separation, reflecting favorable intermolecular interactions that stabilize the condensed phase. Conversely, a less negative (or positive) $\Delta G_{\text{transfer}}$ corresponds to weakened condensation and a reduced propensity for phase separation.

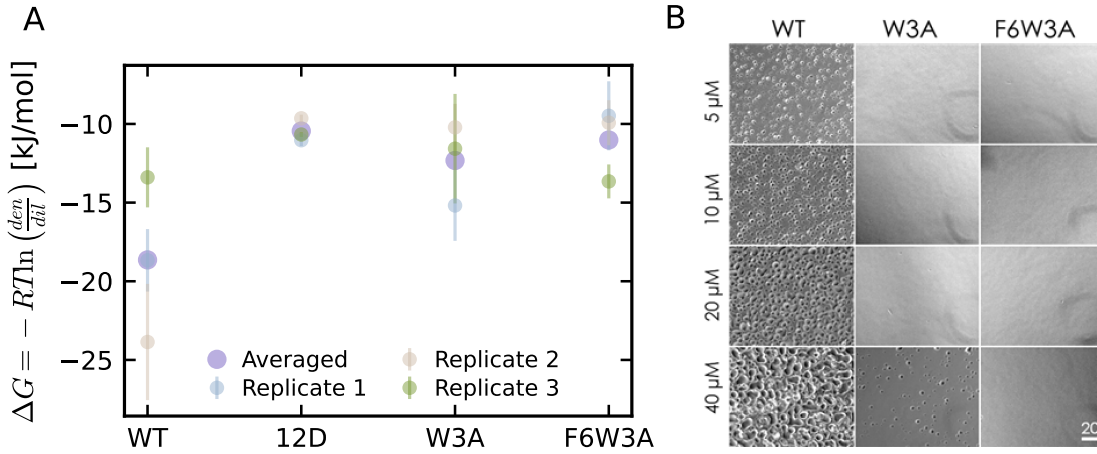


Figure 2.13: Phase separation of aromatic and serine mutants under Martini 3 Elastic Network (EN) simulations at $\lambda_{1.06}$. (A). Transfer free energies of the mutants in slab-geometry simulations at $\lambda_{1.06}$. All mutants display reduced phase separation relative to the wild type (WT). (B). *In vitro* phase-separation assays for the aromatic mutants under identical experimental conditions, showing diminished condensate formation compared to the WT. Experimental data in panel B were kindly provided by Dr. Saskia Hutten (Dormann Group).

Substituting these twelve serines with charged aspartates (12D) disrupted condensate formation both experimentally and in simulations, reducing the transfer free energy relative to the wild type (WT) (Figure 2.13A) and producing less favorable values than WT across all three replica simulations of the 12D variant (Figure 2.19). Similarly, replacement of three tryptophans with alanine (W3A), as well as a combined substitution of six phenylalanines (Phe276, Phe283, Phe289, Phe367, Phe397, and Phe401) together with the three tryptophans (F6W3A), also reduced condensation, as indicated by the transfer free energy. These effects were further confirmed in *in vitro* experiments (Figure 2.13B).

In this framework, the mutations decreased the ratio $\left(\frac{c_{\text{dense}}}{c_{\text{dilute}}}\right)$, thereby increasing $\Delta G_{\text{transfer}}$ and diminishing the stability of the condensed phase. Because $\Delta G_{\text{transfer}}$ depends only on the equilibrium concentrations of the two phases, it provides a direct and quantitative measure for comparing phase behavior across different protein variants.

2.3.1 Aromatic mutants

In the study of LCD aromatic mutants, six of the eight phenylalanine residues were substituted with alanine, while Phe313 and Phe316 were retained. Previous work by Rizuan et al. [69] demonstrated that single-point mutations F313A and F316A can each significantly reduce phase separation. To avoid confounding effects arising from these critical residues, Phe313 and Phe316 were retained. Consequently, the remaining six phenylalanine residues, together with three tryptophan residues, were substituted with alanine to selectively perturb aromatic-mediated interactions while minimizing dominant effects from Phe313 and Phe316.

From the time-dependent density histogram plots, W3A exhibited a lower dense-phase concentration and more frequent exchanges between the dense and dilute phases compared to the WT across all replicas (Figure 2.19A–C). F6W3A displayed an even stronger tendency toward dissolution than W3A, consistent with experimental observations (Figure 2.13B). Experimentally, the WT formed liquid droplets at all tested concentrations (5, 10, 20, and 40 μM), with droplet size increasing with concentration. The W3A mutant produced only a few small droplets at the highest concentration (40 μM), whereas F6W3A failed to form droplets at any concentration. These experimental trends were mirrored in simulations, which showed a clear reduction in the transfer free energy ($\Delta G_{\text{transfer}}$) for the mutants (Figure 2.13A).

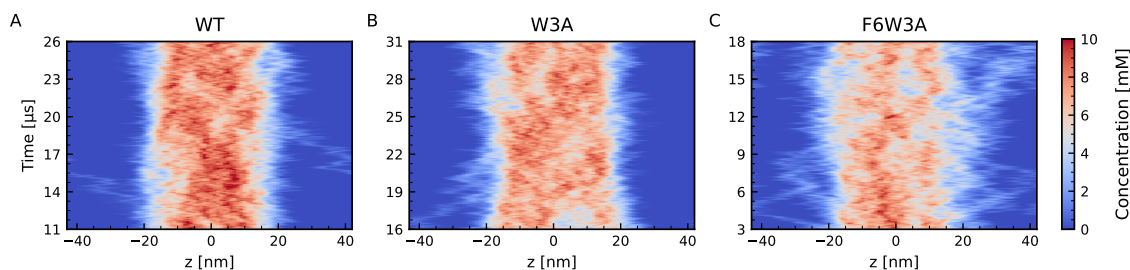


Figure 2.14: Time-dependent density distributions of aromatic mutants along the z-axis under Martini 3 Elastic Network (EN) simulations at $\lambda_{1.06}$. (A–C) Concentration histograms over time for the wild-type (WT) protein and aromatic mutants W3A and F6W3A. Protein concentration is indicated by the color bar on the right.

Aromatic residues with high inter-chain contact probabilities were replaced with alanine, resulting in marked changes in inter-chain interactions (Figure 2.15). Overall, contacts between the LCD and neighboring LCDs, as well as interactions with the RRM2 and NLS regions, were substantially reduced in the mutants compared to the wild-type protein. At the specific sites of mutation, the corresponding residues became largely silent in the contact maps, indicating a loss of local intermolecular interactions. In the W3A variant, residues 334, 401, and 412 exhibited markedly diminished contact probabilities (Figure 2.15D), while in the F6W3A mutant, all substituted aromatic positions showed similarly reduced interaction signals. Inter-chain contacts near the C-terminal region were

2 Phase behavior of TDP-43

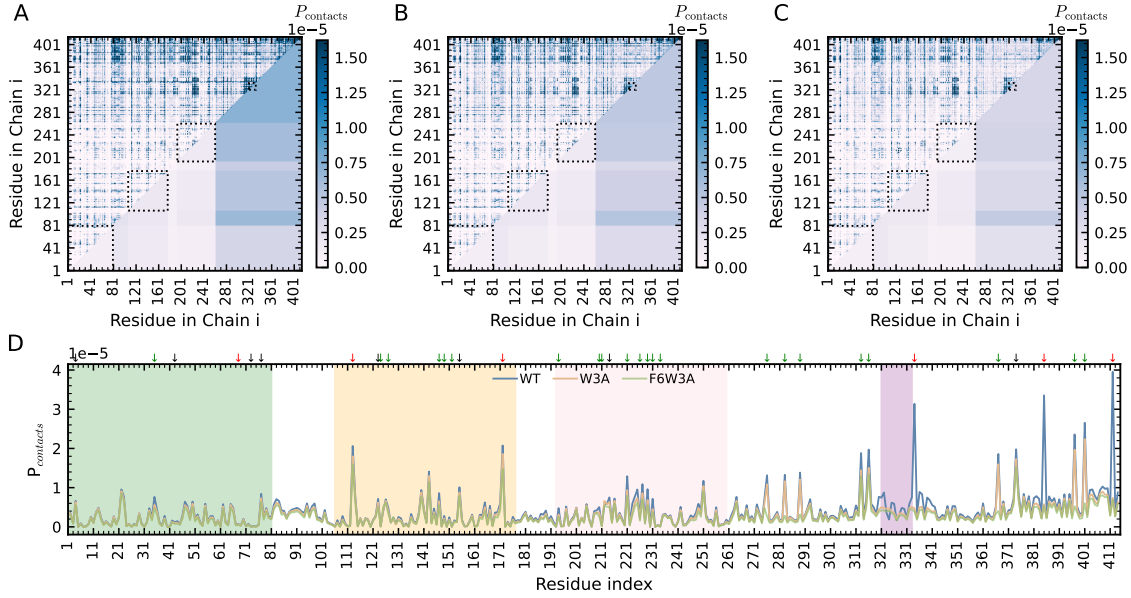


Figure 2.15: Intermolecular contacts within aromatic mutant condensates under Martini 3 Elastic Network (EN) simulations at $\lambda_{1.06}$. (A–C). Inter-chain contact heatmaps for the WT protein and aromatic mutants W3A and F6W3A, respectively. Upper triangles are the residue-wise contact maps, lower triangles are domain-wise contact maps, residue- and domain-wise contact maps are sharing same colorbar. (D). Averaged intermolecular contact probability for each residue. Tryptophan residues are indicated by red arrows, phenylalanine residues by green arrows, and other aromatic residues by black arrows.

particularly affected, showing a pronounced overall reduction.

In these simulations, the core helical structure within the LCD was not explicitly stabilized by harmonic restraints. As a result, the contribution of helix stability to condensate formation and maintenance was not directly assessed. To disentangle whether the observed loss of inter-chain contacts arises primarily from reduced aromatic-mediated interactions or from destabilization of the helical structure itself, future simulations incorporating explicit helical stabilization would be informative. Such simulations could clarify whether maintaining helix integrity rescues intermolecular contacts and condensate stability, or whether aromatic residue identity and side-chain interactions are the dominant determinants.

The loss of inter-chain contacts destabilized the condensed phase, increasing $\Delta G_{\text{transfer}}$ and weakening the intermolecular interactions that support phase separation. Time-dependent density histograms reflected this weakening, aromatic mutants maintained a lower dense-phase concentration and exhibited more frequent exchanges between dense and dilute phases. While W3A displayed instability, F6W3A showed more severe dissolution, directly

linking disrupted contact networks to unfavorable transfer thermodynamics and a loss of condensate stability.

2.3.2 Serine to Aspartate phosphomimicking mutant

Residues Ser within the LCD are frequently phosphorylated, a hallmark of pathological TDP-43 inclusions observed in neurodegenerative diseases such as ALS and FTLD [36, 70, 71]. Phosphorylation introduces additional negative charges (-2) into the LCD, altering its electrostatic landscape and weakening intermolecular interactions that support condensate formation. In this context, the 12D mutant, where twelve serine residues were substituted with Asp amino acids to mimic phosphorylation, provides a mechanistic model for assessing the impact of serine phosphorylation on phase behavior [68].

The 12D mutant exhibited a markedly reduced propensity for phase separation (Fig. 2.13A), in line with experimental evidence that serine-mimicking diminishes TDP-43 condensation [68]. Long-timescale simulations exceeding $30\ \mu\text{s}$ enabled quantitative characterization of how these mutations reshaped protein partitioning between the dense and dilute phases as equilibrium was established, revealing continuous molecular exchange between the two phases (Figure 2.16B and Figure 2.19G–I).

The 12D mutant exhibited a markedly reduced propensity for phase separation, as reflected by lower dense-phase concentrations and more frequent exchanges between the dense and dilute phases (Figure 2.16B and Figure 2.19G–I). Correspondingly, the transfer free energy, $\Delta G_{\text{transfer}}$, was less favorable than that of the wild-type protein, quantitatively capturing the destabilization of the condensed phase (Fig. 2.13A). The 12D variant carried a net charge of -16 and a net charge per residue (NCPR) of -0.039 , resulting in enhanced recruitment of Na^+ ions within the condensate relative to the wild-type protein (Figure 2.16C). In addition, regions of “water holes” were observed within the condensate (Figure 2.17); one such water hole is highlighted by a blue outline. These observations indicate that the serine-to-aspartate phosphomimetic mutations not only weaken inter-chain interactions and reduce phase separation propensity but also perturb the internal organization of the condensate, including ion distribution and local solvent structure, which may further contribute to its instability.

In terms of inter-chain interactions, substitution of the twelve serines with aspartates (12D) led to a pronounced reduction in contacts at the mutated positions, reflecting a local loss of intermolecular interactions. A few aromatic residues (Phe313 and Phe316) and the helical region within the LCD displayed slightly enhanced inter-chain contacts (Figure 2.18B–C), partially compensating for the loss of serine-mediated interactions. Overall, LCD–LCD contacts were modestly increased, whereas interactions between the LCD and other domains, such as the RRM1, RRM2, or NTD regions, were weakened. This combination of increased ionic content, resulting from the negatively charged aspartates, and the local loss of inter-chain interactions contributed to reduced condensate stability, reflected in a

2 Phase behavior of TDP-43

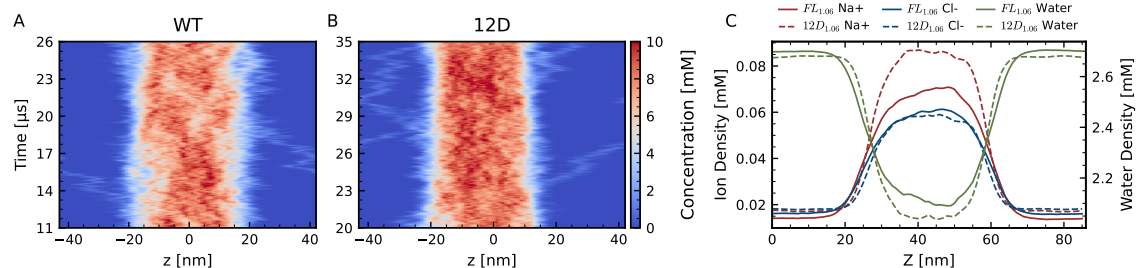


Figure 2.16: Phase separation of the phosphomimetic 12D mutant under Martini 3 Elastic Network (EN) simulations at $\lambda_{1.06}$. (A-B). Time-dependent concentration histograms along the z-axis for the wild-type (WT) protein and the 12D mutant, with protein concentration indicated by the color bar. (C). Ion and water density profiles along the z-axis for WT TDP-43 and the 12D mutant, highlighting differences in ion recruitment and local hydration within the condensate.

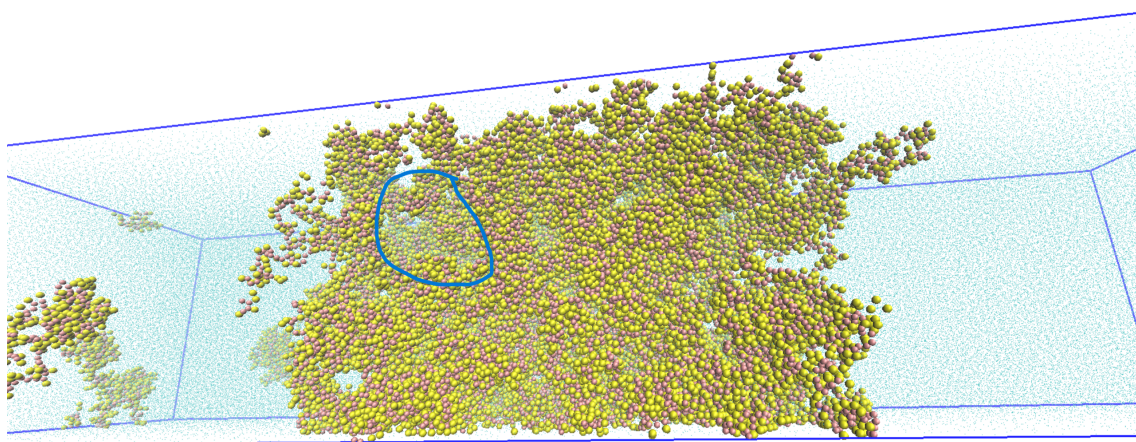


Figure 2.17: Water cavities within the 12D mutant condensate. One of the water-filled cavities is highlighted with a blue circle.

lower dense-phase concentration and weakened protein–protein interactions driving phase separation.

This case illustrates how charge can modulate condensate stability. The introduction of negatively charged residues Asp in place of residues Ser increased the net charge of the protein, promoting greater recruitment of counterions into the condensate and altering the electrostatic environment. These charge-mediated effects weakened inter-chain interactions with other domains, diminishing the overall network of contacts that supported

2.3 Phase behavior of TDP-43 Variants

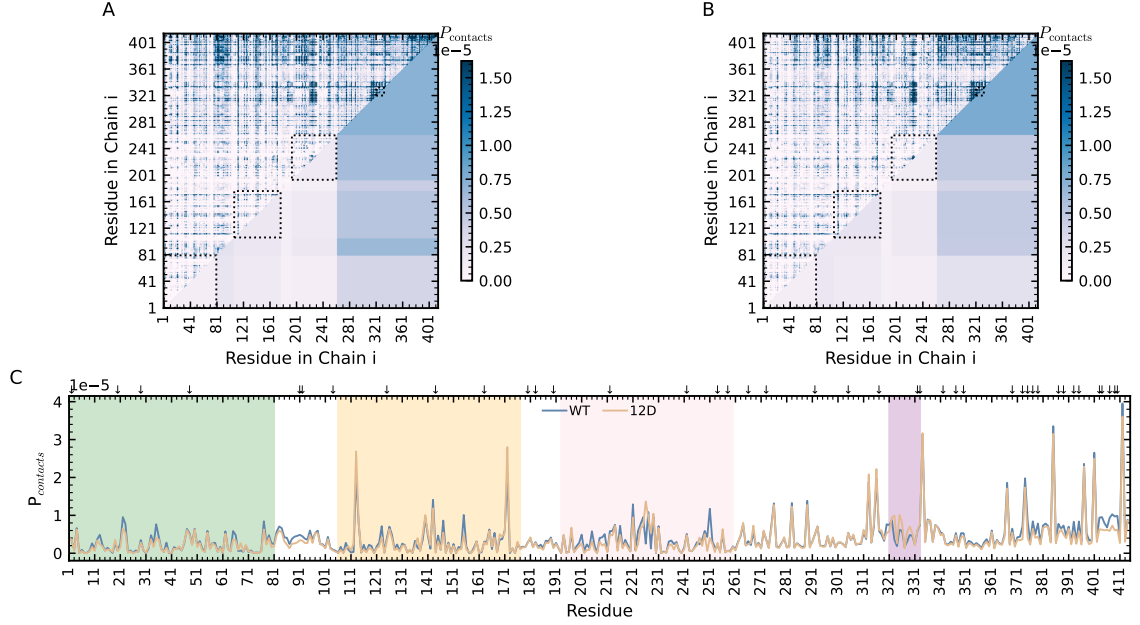


Figure 2.18: Inter-molecular contacts within the phosphomimetic 12D mutant under Martini 3 Elastic Network (EN) simulations at $\lambda_{1.06}$. (A–B). Inter-chain contact heatmaps of the wild-type (WT) protein and the Ser-to-Asp mutant 12D. (C). Averaged inter-molecular contact probability for each residue, with all serine residues indicated by black arrows.

condensate stability. Some local contacts were enhanced due to the altered electrostatics, the net effect favored ion accumulation and condensate dissolution, highlighting the delicate balance between specific interactions and electrostatic contributions in determining condensate stability.

2 Phase behavior of TDP-43

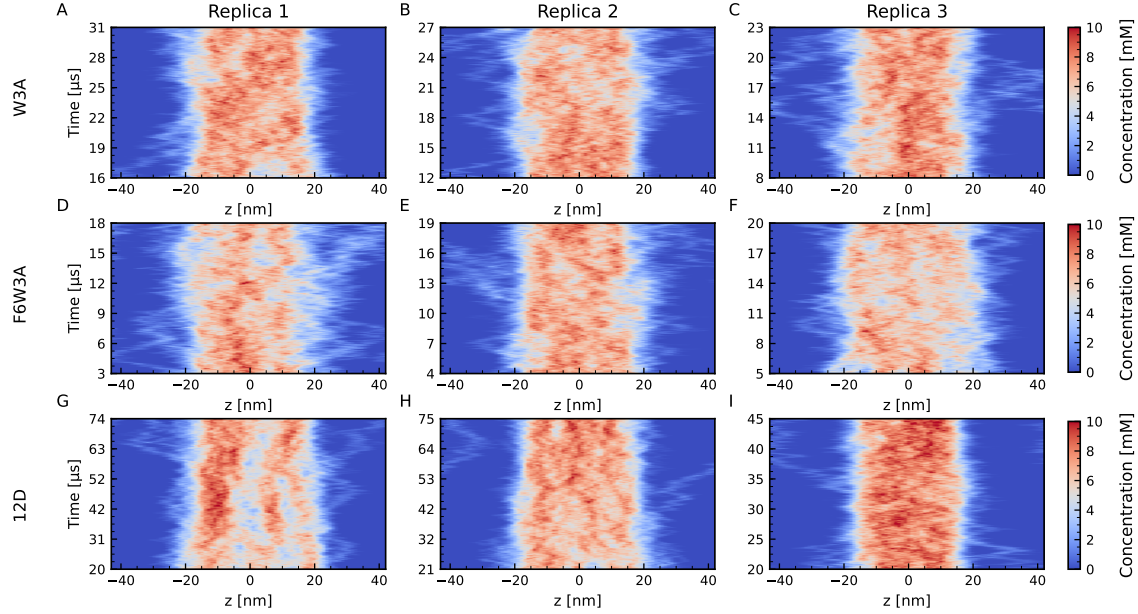


Figure 2.19: Time-dependent concentration histograms of mutant replicas under the Martini 3 Elastic Network (EN) model at $\lambda_{1.06}$. Protein concentration is indicated by the color bar.

2.4 Discussion

This chapter primarily examines the full-length TDP-43 protein, which comprises both structured and intrinsically disordered regions, with a focus on its conformational compactness in the dilute phase and its phase-separation behavior. By employing the Martini 3 Elastic Network (EN) model with a protein–water interaction rescaling factor of $\lambda_{1.06}$, these properties were characterized through time-resolved protein density profiles and inter-chain contact analyses, while time-dependent density histogram plots were used to visualize molecular exchange between coexisting dense and dilute phases.

A central finding of this work is that the Martini 3 EN model at $\lambda_{1.06}$ provides a physically consistent description of TDP-43 condensation. Under these conditions, individual protein chains remain sufficiently expanded in the dilute phase while still forming a stable, liquid-like condensate characterized by dynamic chain exchange. This balance is particularly important for multi-domain proteins, where overly compact chains can artificially suppress intermolecular contacts, while excessive expansion can destabilize condensates altogether. The ability of the EN-stabilized Martini 3 model to reproduce both phase coexistence and exchange dynamics highlights its suitability for studying biomolecular condensates formed by proteins with mixed ordered–disordered architectures.

Beyond reproducing wild-type behavior, this simulation framework was successfully ex-

tended to examine a range of TDP-43 mutants, including aromatic substitutions and phosphomimetic variants. These mutant simulations revealed how subtle sequence-level perturbations reshape intermolecular interaction networks, alter ion partitioning, and ultimately modulate condensate stability. Importantly, the simulation results showed strong qualitative agreement with available *in vitro* experimental data, reinforcing the relevance of the chosen parameterization and analysis strategy. This demonstrates that the model is not merely descriptive but can serve as a predictive tool for probing how sequence features control phase separation in multi-domain proteins.

Recent methodological advances, such as the approach introduced by Wang *et al.* (2025) [46], provide an alternative route for simulating phase separation by releasing angular and dihedral restraints within intrinsically disordered regions (IDRs) while maintaining Elastic Network stabilization for folded domains. Notably, in both the method of Wang *et al.* and the approach used in this chapter, angular and dihedral restraints within IDRs are released, allowing for chain flexibility.

However, a key distinction lies in how condensate dynamics are tuned. In the present work, rescaling protein–water interactions provides an additional control parameter for modulating intermolecular attraction strength and condensate fluidity. For example, at lower rescaling values (e.g., $\lambda_{1.05}$), condensates become noticeably less dynamic, exhibiting reduced molecular exchange despite the absence of explicit IDR restraints. This highlights that solvent-mediated interactions can strongly influence condensate dynamics even when chain flexibility is preserved.

In contrast, the method of Wang *et al.* relies primarily on force-field reparameterization without an explicit solvent-interaction tuning knob analogous to λ . This raises one open question, whether such an approach, without further adaptation, can robustly capture the full range of condensate dynamics observed, particularly for systems in which subtle changes in interaction strength strongly regulate molecular exchange.

A direct and systematic comparison between these approaches, benchmarked against experimental observables such as dense phase concentration, condensate viscosity, and molecular exchange rates, will be essential to determine how best to balance physical realism, transferability, and computational efficiency when modeling phase separation of complex, multi-domain proteins.

Another consideration is whether applying protein–water interaction rescaling selectively to intrinsically disordered regions (IDRs), rather than uniformly across the entire protein, could provide a more balanced description of phase separation. In principle, such a targeted rescaling strategy might preserve the native stability and specific contacts of folded domains while still enhancing the driving forces for condensation mediated by disordered regions.

This question is particularly relevant because global rescaling inevitably weakens interactions involving structured domains as well. Indeed, in next chapter 3, protein–water

2 Phase behavior of TDP-43

rescaling was found to reduce NTD dimerization propensity, indicating that folded–folded interactions are sensitive to the same tuning parameter used to promote condensate dynamics. Selective rescaling of IDRs might therefore mitigate this unintended effect while maintaining sufficient folded-domain intermolecular attraction to support liquid–liquid phase separation.

At present, this idea remains speculative, as such a region-specific rescaling scheme has not yet been systematically tested. Nevertheless, it represents a promising direction for future methodological development, especially for multi-domain proteins in which both structured domains and IDRs contribute distinct and competing roles to phase behavior.

From a methodological perspective, simulations in slab geometry required careful preparation to ensure efficient equilibration. Pre-compaction of protein chains into a smaller simulation box was necessary to accelerate nucleation and promote condensate formation within accessible simulation timescales. To avoid artifacts associated with early-stage nucleation, all quantitative analyses were restricted to the final 15 μ s of each trajectory, during which the systems exhibited stable phase coexistence and steady exchange dynamics. While this strategy proved effective, it also highlights an ongoing challenge in condensate simulations, achieving sufficient statistical sampling at equilibrium remains computationally demanding, particularly for large protein systems.

Nevertheless, the results presented in this chapter demonstrate that the Martini 3 EN model with appropriate interaction rescaling offers a good framework for dissecting the molecular determinants of phase separation in TDP-43. More broadly, this approach provides a foundation for studying liquid-like condensate formation in other multi-domain, thereby contributing to a deeper mechanistic understanding of biomolecular condensation in both physiological and pathological contexts.

3 Regulating TDP-43 LLPS via structured domain stabilization and inter-chain bridging

In addition to employing an Elastic Network (EN) model to stabilize the three-dimensional structures of proteins, Lennard–Jones (LJ) potentials have also been incorporated to capture conformational flexibility within the folded domains, as implemented in the $G\bar{o}_{\text{intra}}$ model. The native NTD dimerization interaction was underestimated in the baseline models, including both the Elastic Network and $G\bar{o}_{\text{intra}}$ models. To address this, an additional LJ potential describing inter-chain contacts between folded domains was introduced in the $G\bar{o}_{\text{intra}+\text{inter}}$ model. This chapter will focus on how condensate formation and material properties are shaped by the interplay between specific interactions among TDP-43 folded domains and their intrinsic conformational flexibility.

3.1 Single chain full-length protein conformation

With less restrictive restraints provided by the LJ potentials, individual chains adopted more extended and flexible conformations under the $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.06}$. The definition of the structured regions, i.e., the residues forming stable secondary structures, was kept identical to that used in the Elastic Network Model (Section 2.1). However, the number of applied bonds differed from the bond count used in the Elastic Network implementation, reflecting the more reduced rigidity in the $G\bar{o}_{\text{intra}}$ setup.

3.1.1 Single chain compactness

Using the $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.06}$, single-chain simulations of full-length TDP-43 were performed using the same system setup as in the Elastic Network model. All replicas were simulated for more than 10 μs . In the Elastic Network model, the protein exhibited an average radius of gyration of $R_g = 3.55 \pm 0.13$ nm at $\lambda_{1.06}$, reflecting an expanded conformational ensemble, as discussed in Chapter 2.1.

The $G\bar{o}_{\text{intra}}$ model exhibited a degree of chain compactness comparable to that of the Elastic Network (EN) model when using a stronger protein–water interaction strength (Figure 3.1A and C). At the baseline rescaling factor $\lambda_{1.00}$, the single TDP-43 chain remained

3 Regulating TDP-43 LLPS via structured domain stabilization and inter-chain bridging

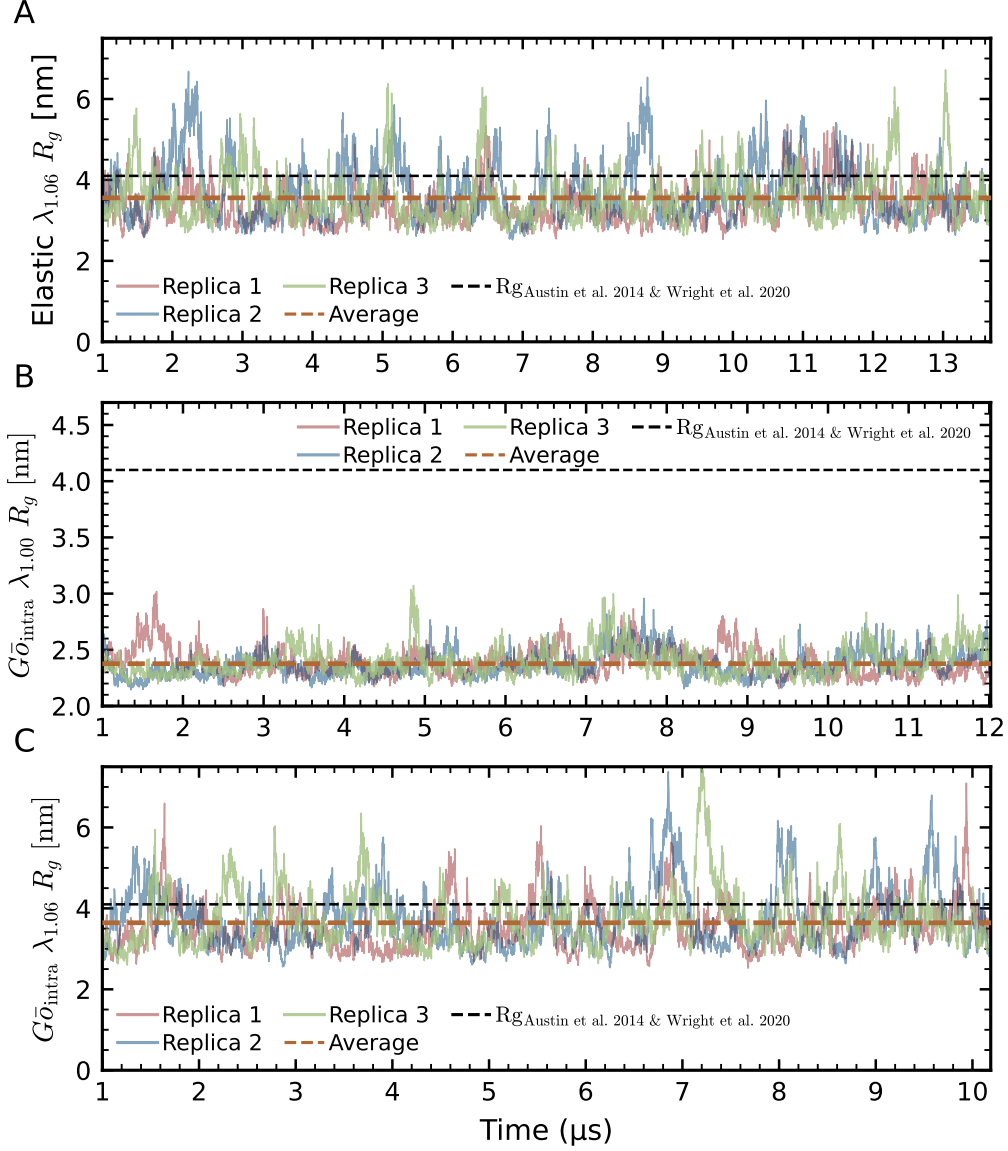


Figure 3.1: Radius of gyration (R_g) of single-chain full-length TDP-43 under different models and λ values. (A). R_g across three replicates in the Elastic Network model at $\lambda_{1.06}$; the black dashed line indicates the maximum experimental R_g , and the orange dashed line shows the mean across replicates. (B). R_g across three replicates in the $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.00}$, with the same dashed-line conventions. (C). R_g across three replicates in the $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.06}$, with the same dashed-line conventions.

relatively compact, with a mean radius of gyration $R_g = 2.38 \pm 0.013$ nm (Figure 3.1B).

3.1 Single chain full-length protein conformation

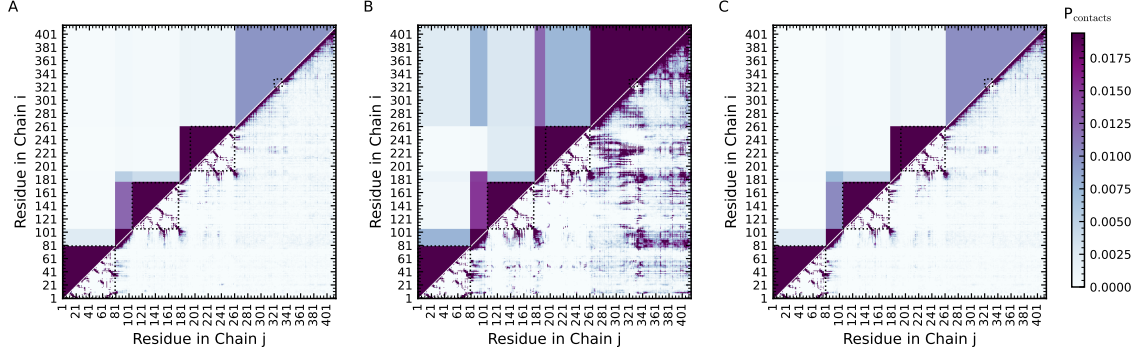


Figure 3.2: Averaged intra-chain contact heatmaps of single-chain full-length TDP-43 under different models. (A). Elastic Network model at $\lambda_{1.06}$. (B). $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.00}$. (C). $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.06}$. For each panel, the upper triangle shows the residue-wise contact map, and the lower triangle shows the domain-wise contact map. All contact maps share the same colorbar.

Increasing the interaction strength to $\lambda_{1.06}$ led to a substantial expansion of the protein, with the average radius of gyration increasing to $R_g = 3.64 \pm 0.073$ nm. This value was slightly smaller than the experimental upper bound (black dashed line in Figure 3.1) and was markedly larger than that observed at $\lambda_{1.00}$.

The distribution of R_g values at $\lambda_{1.06}$ closely resembled that obtained from the EN model at the same interaction strength, suggesting that both modeling strategies sample similarly extended single-chain conformations under equivalent solvent conditions. These results demonstrate the $G\bar{o}_{\text{intra}}$ model is capable of capturing the essential balance between chain flexibility and residual intramolecular interactions that govern TDP-43’s conformational ensemble in the dilute phase.

Compared to the Elastic Network model at $\lambda_{1.06}$, the $G\bar{o}_{\text{intra}}$ model also exhibited a reduction in long-range intramolecular contacts. The $G\bar{o}_{\text{intra}}$ model was dominated by short-range, domain-local contacts, with long-distance interactions largely absent (Figure 3.2A and C). NLS–RRM1 contacts were present in both models but occurred less frequently in the $G\bar{o}_{\text{intra}}$ model. Interactions between the LCD and other domains were strongly suppressed in the $G\bar{o}_{\text{intra}}$ model, particularly at $\lambda_{1.00}$ compared to $\lambda_{1.06}$ (Figure 3.2B and C).

Overall, under the $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.06}$, single-chain expansion was captured, showing slightly greater extension compared to the Elastic Network model under the same solvent conditions.

3.1.2 Residue flexibility

Residue flexibility of the folded domains of TDP-43, as observed in atomistic molecular dynamics, was captured by both the rescaled Martini 3 Elastic Network and $G\bar{o}_{\text{intra}}$ models. This flexibility was quantified using the root-mean-square fluctuation (RMSF), which measures how much each particle fluctuates around its ensemble-averaged position over the course of the simulation. For particle i , the ensemble-averaged position is defined as

$$\langle \mathbf{x}_i \rangle = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_i(t), \quad (3.1)$$

where $\mathbf{x}_i(t)$ is the position of particle i at time t and T is the total number of frames. The RMSF is then calculated as

$$\rho_i = \sqrt{\langle (\mathbf{x}_i(t) - \langle \mathbf{x}_i \rangle)^2 \rangle}. \quad (3.2)$$

Trajectory frames were first aligned to the average structure to remove overall translation and rotation. Backbone beads were used in coarse-grained simulations, whereas C α atoms were used in all-atom simulations. Higher RMSF values indicate residues with greater flexibility, whereas lower values correspond to more rigid residues.

Atomistic molecular dynamics simulations in general provided a very accurate description of the conformational ensembles of folded protein domains[72–74] and thus can serve as a benchmark for Martini3 models[72, 74]. Figure 3.3 applied dual-axes to depict the residue fluctuations from different models within sub-regions of the over-all folded domain. The primary (left) y-axis represented the RMSF value of each residue. Simultaneously, the secondary (right) y-axis represented simulation time, and in combination with the color bar illustrated changes in secondary structure over time for individual residues in all-atom simulations. The color at different time points represented whether the particular residue was part of a helical, an extended (β -sheet), or an unstructured sub-region (coil). Residues maintaining stable secondary structures within the sub-folded region showed low RMSF, while the loops maintain flexibility across various coarse-grained models, resembling the behavior observed in all-atom model. For instance, residues 216-221 in RRM2 stably formed an extended β -strand like conformation in the atomistic molecular dynamics simulations and RMSF values were below 2 Å for these residues. By contrast, focusing on the α -helix formed by residues 236-242, some residues towards its N- and C-terminal ended transiently lose their helical conformations; nevertheless, the RMSF values remained low as the segment kept its structure. The N-terminal residues of the adjacent β -strand residues 246-250 sometimes lost their extended conformation and the RMSF values for these residues rose to values close to 2 Å.

Collectively, the rescaled Martini 3 force field ($\lambda_{1.06}$), implemented with either the Elas-

3.1 Single chain full-length protein conformation

tic Network or $G\bar{o}_{\text{intra}}$ model, accurately captured single-chain conformations as well as residue-level flexibility.

3 Regulating TDP-43 LLPS via structured domain stabilization and inter-chain bridging

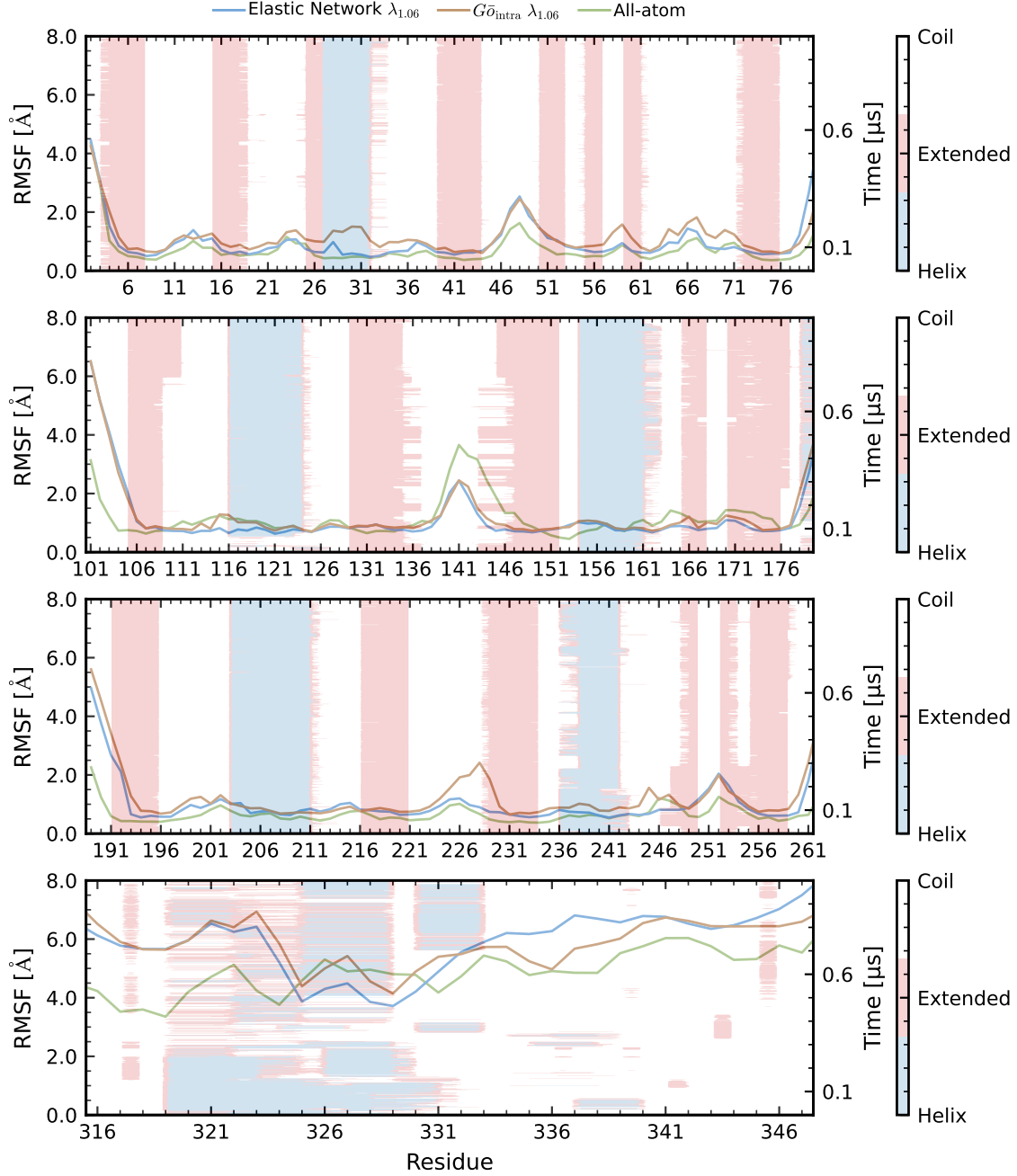


Figure 3.3: Residue flexibility within the structured domains of TDP-43 under all-atom, Elastic Network and $G\bar{o}_{\text{intra}}$ models at $\lambda_{1.06}$. Root mean square fluctuation (RMSF) of backbone beads in folded regions: NTD, RRM1, RRM2, and the helical segment of the C-terminal disordered region. The left axis measures the RMSF of each residue from its average position. Here both rescaled and unrescaled Martini 3 models are shown for comparison. The right axis in combination with the colorbar track the secondary structure of each residue over the time course of an AA simulation: blue is helix, pink is extended structure (β -sheet), and white is coil.

3.2 Dilute TDP43-NTD dimerization characterization

The N-terminal domain (NTD) of TDP-43 is capable of forming homodimers, predominantly mediated through interactions between the head region of one monomer and the tail region of a neighboring monomer [75]. The dimerization interface is characterized by electrostatic complementarity, where a negatively charged head region interacts with a positively charged tail region, facilitating stable association. Structural insights from the crystal structure (Figure 3.4A) show that the tail β -sheet (β_{51-58} , colored pink) establishes contacts not only with the head β -sheet (β_{16-20} , pink) but also with the α -helix (α_{26-35} , green) of the adjacent chain. These interactions collectively stabilize the NTD dimer, creating a well-defined interface that is critical for oligomerization and may contribute to the multivalent interactions that drive higher-order assembly of full-length TDP-43 [76]. The combination of electrostatic complementarity and secondary-structure contacts underscores the importance of both sequence and structural elements in guiding NTD-mediated dimerization.

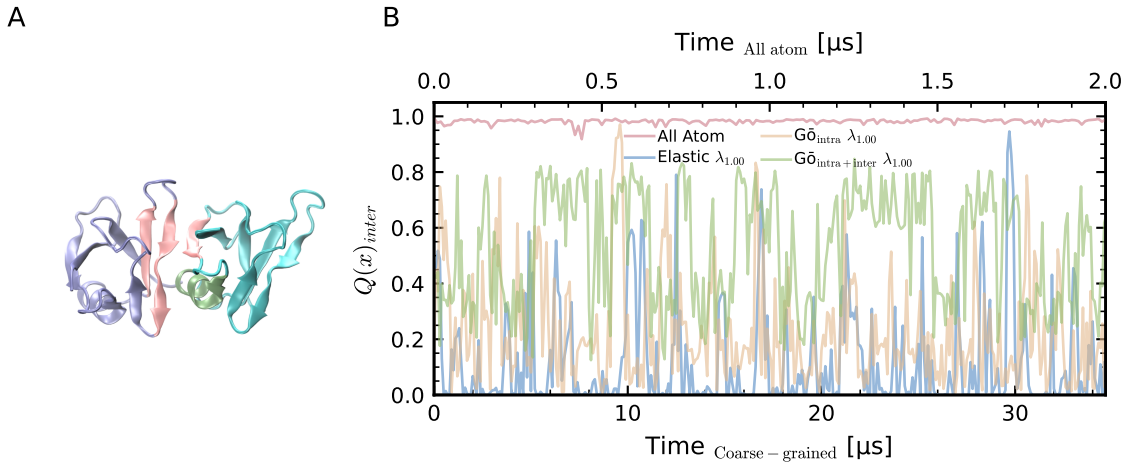


Figure 3.4: NTD dimerization. (A). The dimer is stabilized by complementary electrostatic interactions between the negatively charged head of one chain (right; β -sheet in pink and α -helix in green) and the positively charged tail of the partner chain (left; β -sheet in pink). (B). Fraction of native inter-chain contacts observed in all-atom and coarse-grained simulations of the NTD homodimer.

The NTD homodimer was simulated using all-atom models, the EN model, and both coarse-grained $G\bar{o}_{intra}$ and $G\bar{o}_{intra+inter}$ models. The fraction of native contacts, $Q_{(x)}$ [77], was used to quantify the preservation of native inter-molecular contacts within the NTD dimer. In all-atom simulations, the dimer remained stably associated over the full 2 μ s trajectory, with Q_x values consistently close to 1 (Figure 3.4B), in excellent agreement with the experimental crystal structure (PDB: 5mdi) reported in Ref. 75.

All three coarse-grained models were able to capture NTD dimer formation, though with

3 Regulating TDP-43 LLPS via structured domain stabilization and inter-chain bridging

varying degrees of accuracy in reproducing the native inter-chain contacts (Figure 3.4B and 3.5). The decomposition of native contacts focused on the specific interactions between the $\beta_{51-58}-\beta_{16-20}$ and $\beta_{51-58}-\alpha_{26-35}$ interfaces, as well as other residue contacts across the dimer. The EN model successfully captured the overall formation of the dimer; however, it rarely reproduced both the $\beta_{51-58}-\beta_{16-20}$ and $\beta_{51-58}-\alpha_{26-35}$ interfacial contacts with consistently high $Q_{(x)}$ values. This indicates that, while the EN model is sufficient to maintain global dimer architecture, it may not fully capture the detailed residue-level interactions that stabilize the interface.

The $G\bar{o}_{\text{intra}}$ model also maintained the dimer and sampled interfacial interactions more frequently than the EN model, suggesting improved flexibility in capturing transient contacts. Nevertheless, the two critical contact interfaces remained somewhat less stable than in the all-atom reference, highlighting limitations in describing inter-chain specificity when only intra-chain interactions are included. In contrast, the $G\bar{o}_{\text{intra+inter}}$ model, which incorporates additional inter-chain Lennard–Jones potentials to better describe native contacts, consistently preserved dimer formation and produced relatively high $Q_{(x)}$ values across both interfaces. Despite this improvement, the $\beta_{51-58}-\alpha_{26-35}$ contacts were still somewhat underestimated, suggesting that even the enhanced coarse-grained model cannot fully recapitulate all atomistic details of the NTD dimer interface.

Rescaling the protein–water interaction did not improve the capture of inter-chain contacts or their frequency, as evidenced by the inter-chain contact heatmaps (Figure 3.6). Without rescaling, the $G\bar{o}_{\text{intra}}$ model (Figure 3.6B) captures more inter-chain interactions than the EN model (Figure 3.6A). Upon rescaling ($\lambda_{1.06}$), some inter-chain contacts persisted in the EN model (Figure 3.6C), whereas the corresponding contacts were largely lost in the $G\bar{o}_{\text{intra}}$ model (Figure 3.6D). It should be noted that rescaling effects on NTD monomer conformations have not been quantified using the $Q_{(x)}$ metric.

This evaluation focuses on the integrity of the two main interfacial contacts, rather than on individual residue-pair interactions. Among the coarse-grained models, the $G\bar{o}_{\text{intra+inter}}$ model remains the most reliable in representing NTD dimerization.

Overall, for monomer association, without scaling up the protein–water interaction strength, the Martini 3 $G\bar{o}_{\text{intra}}$ model performed better than the Elastic Network model in capturing the contact geometry observed in the crystal structure, and the additional inter-chain potential in the $G\bar{o}_{\text{intra+inter}}$ model enhanced both the frequency and accuracy of native contacts.

3.2 Dilute TDP43-NTD dimerization characterization

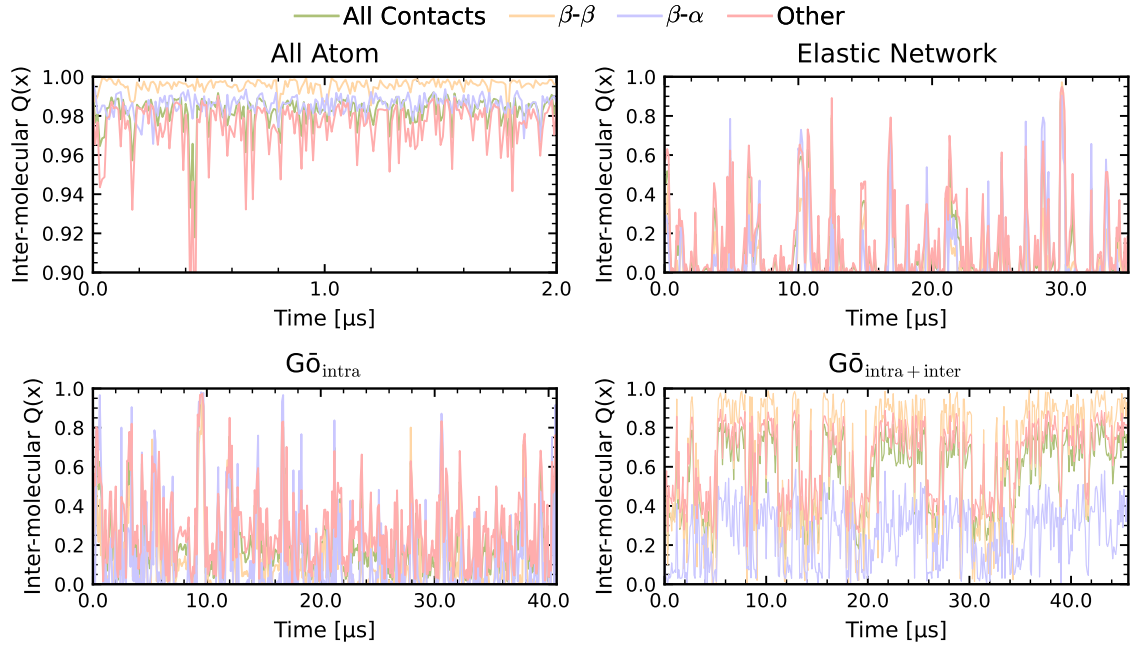


Figure 3.5: Decomposition of native inter-chain contacts in NTD homodimer. Breakdown of Q_x into individual interface components: β - β , β - α , and other contacts for (A) AA, (B) EN, (C) $G\bar{o}_{\text{intra}}$, and (D) $G\bar{o}_{\text{intra}+\text{inter}}$ models.

3 Regulating TDP-43 LLPS via structured domain stabilization and inter-chain bridging

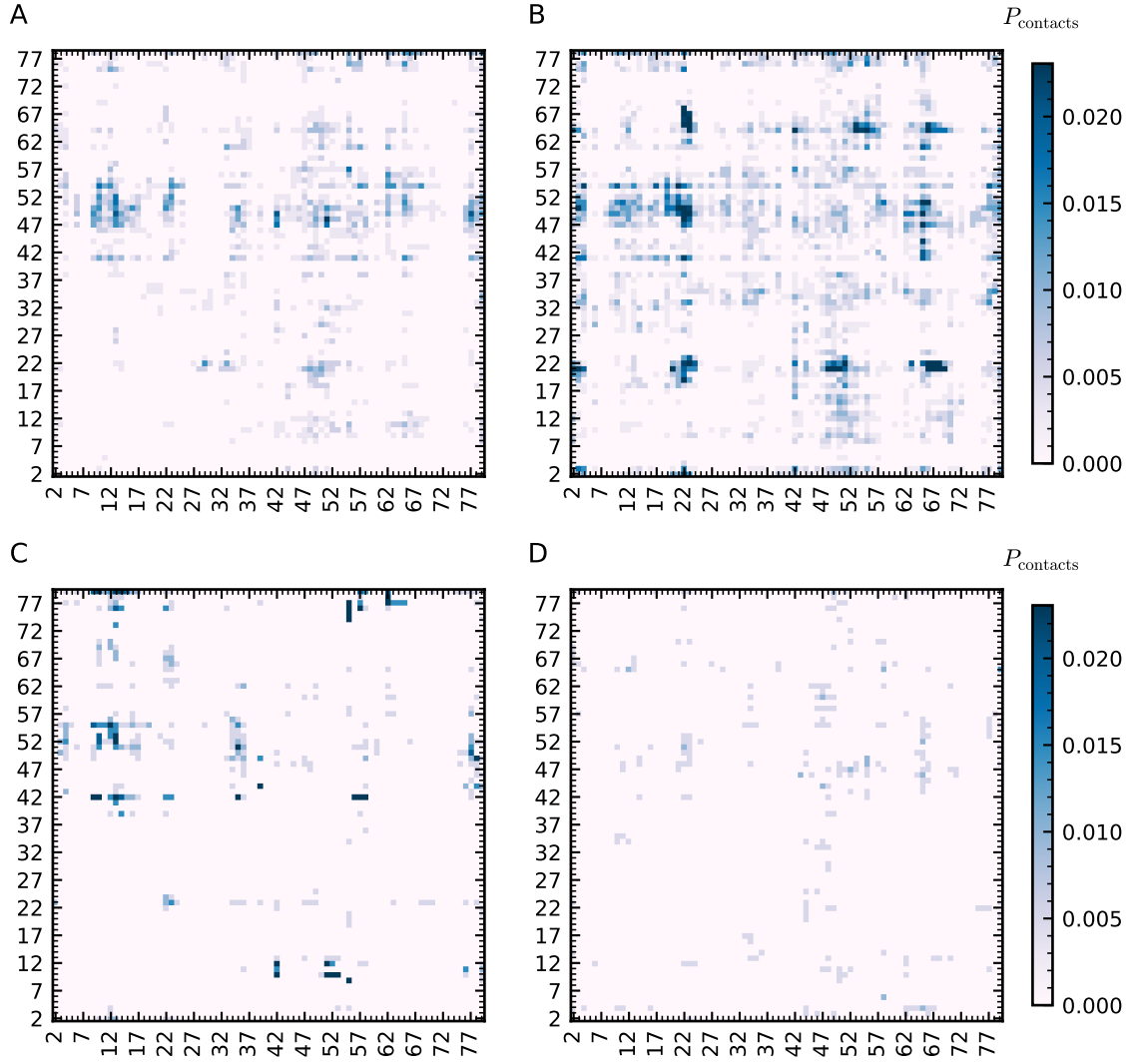


Figure 3.6: Inter-chain contact heatmaps of NTD dimerization under different models. (A). EN model at $\lambda_{1.00}$. (B). $G\bar{\delta}_{\text{intra}}$ model at $\lambda_{1.00}$. (C). EN model at $\lambda_{1.06}$. (D). $G\bar{\delta}_{\text{intra}}$ model at $\lambda_{1.06}$.

3.3 Full-length TDP-43 condensation via LJ-enhanced models

Similar to the condensation simulation methods used in basic Elastic Network models within the slab geometry, full-length TDP-43 simulations were performed with the protein–water interaction parameter (λ) set to 1.05, 1.06, and 1.07 in the $G\bar{o}_{\text{intra}}$ model, and to 1.06 in the $G\bar{o}_{\text{intra}+\text{inter}}$ model.

3.3.1 NTD-dimer within condensates

NTD dimerization and higher-order oligomerization are essential for full-length TDP-43 liquid-liquid phase separation (LLPS) [75]. The Elastic Network model, without enhanced protein–water interaction strength, underestimated the association propensity of the isolated monomeric NTD fragment. In full-length protein condensation simulations, a key question was whether NTD dimerization could be maintained under the adjusted interaction parameters that define the condensate environment.

To quantify inter-chain associations, the total number of contacting chain pairs as well as NTD-dimer-mediated chain pairs were counted in condensates across three different models at $\lambda_{1.06}$ (Figure 3.7A). The $G\bar{o}_{\text{intra}}$ model (orange gradient lines) slightly increased the total number of contacting chain pairs relative to the baseline Elastic Network model (blue gradient lines) within the condensate, while both models maintained a similar number of NTD-mediated contact pairs. The $G\bar{o}_{\text{intra}+\text{inter}}$ model (green line) showed increases in both overall inter-chain associations and NTD-dimer-mediated interactions (Figure 3.7A). NTD dimers dynamically associated and dissociated within the TDP-43 condensate (Figure 3.7B), contributing to internal rearrangements and fluidity of the condensed phase.

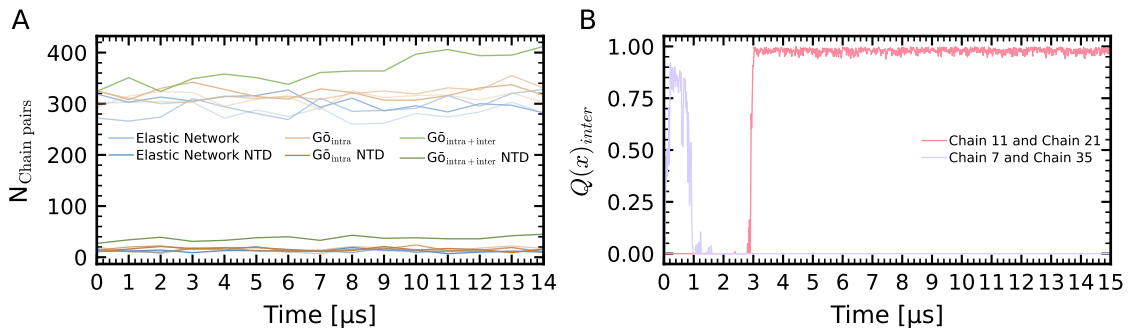


Figure 3.7: Inter-chain and NTD-dimer-mediated contact pair counts within condensates formed under the Elastic Network (EN), $G\bar{o}_{\text{intra}}$, and $G\bar{o}_{\text{intra}+\text{inter}}$ models at $\lambda = 1.06$. (A). Total contact pair counts are shown using blue gradients for the EN model, orange gradients for the $G\bar{o}_{\text{intra}}$ model, and green gradients for the $G\bar{o}_{\text{intra}+\text{inter}}$ model. (B). Inter-molecular $Q(x)$ values of full-length chain pairs within the condensate that form NTD dimers dynamically.

3 Regulating TDP-43 LLPS via structured domain stabilization and inter-chain bridging

Together, condensates formed from rigid and flexible folded domains (EN and $G\bar{o}_{\text{intra}}$ models) maintained comparable levels of NTD-mediated and total inter-chain contacts. The enhanced inter-chain potential in the $G\bar{o}_{\text{intra}+\text{inter}}$ model allowed dynamic exchange of chains while preserving the NTD-mediated interaction network within the condensate.

3.3.2 Inter-molecular interactions within condensates

Similar to the protein inter-chain interactions observed in the condensate under the EN model at $\lambda_{1.06}$, the intrinsically disordered regions (IDRs), including the NLS and LCD, actively mediated inter-chain interactions in both $G\bar{o}$ -based models (Figure 3.8A–C). These flexible regions contributed extensively to the network of multivalent contacts, with LCD–LCD and NLS–LCD interactions being particularly abundant. This observation suggests that, despite the differences in modeling approaches, the disordered regions retained their central role as primary drivers of intermolecular associations and condensate assembly. Notably, within the LCD, aromatic residues displayed the highest inter-chain contact probabilities (Figure 3.8D), highlighting their critical contribution to stabilizing dense-phase interactions. Together, these findings emphasize that both the sequence composition of IDRs and the presence of aromatic residues govern the multivalent interactions underlying phase separation, even when the folded domains are modeled differently.

Furthermore, in the $G\bar{o}_{\text{intra}}$ and $G\bar{o}_{\text{intra}+\text{inter}}$ models, a broadening of the inter-chain interaction network was observed. NLS–NLS interactions became more pronounced in both coarse-grained models compared to the EN model baseline. The NLS region was defined as residues 80–104. In the $G\bar{o}_{\text{intra}}$ model, residues 89–95 displayed higher contact probabilities than in the EN and $G\bar{o}_{\text{intra}+\text{inter}}$ models (Figure 3.9). With the inclusion of enhanced NTD-dimer interactions in the $G\bar{o}_{\text{intra}+\text{inter}}$ model, charged residues within the NLS exhibited contact probabilities that equaled or exceeded those observed in the $G\bar{o}_{\text{intra}}$ model, and clearly surpassed the EN model baseline (Figure 3.9). These results highlight that both the $G\bar{o}$ -based models promote a more extensive and dynamic network of interactions within the NLS, with subtle differences in the specific residues contributing to contacts depending on the presence of additional inter-chain potentials.

Beyond the disordered regions, both $G\bar{o}$ -based models promoted increased contact formation among structured domains. Notably, there was a significant rise in inter-chain interactions involving RRM2 domains, as well as contacts between RRM2 and the C-terminal helix, highlighting a coordinated role for both flexible and folded domains in stabilizing the condensate network. Within the RRM2 domain, residues 321–331 exhibited particularly high contact probabilities (Figure 3.8D).

This expansion of the interaction network was strongly associated with enhanced NTD-mediated contacts. In the $G\bar{o}_{\text{intra}+\text{inter}}$ model, the data demonstrated a clear trend where strengthening NTD-dimerization drove a global increase in inter-chain interactions, indicating that NTD–NTD contacts act as a scaffold that recruits and organizes both flexible and folded domains within the condensate.

3.3 Full-length TDP-43 condensation via LJ-enhanced models

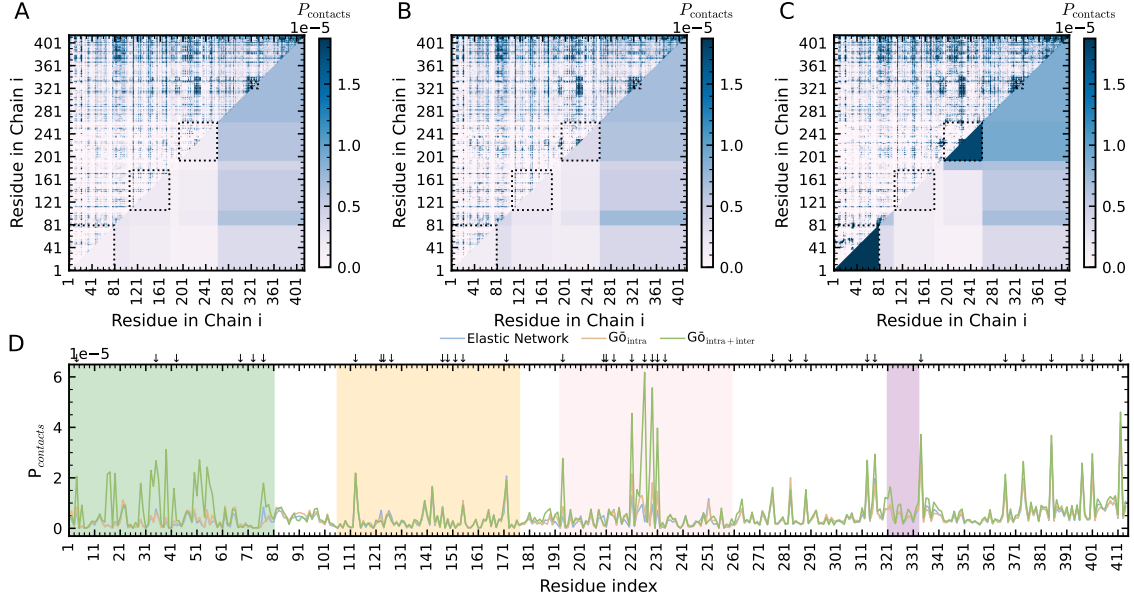


Figure 3.8: Inter-chain contacts within TDP-43 condensates under the EN, $G\bar{o}_{\text{intra}}$, and $G\bar{o}_{\text{intra}+\text{inter}}$ models at $\lambda_{1.06}$. (A–C). Inter-chain contact heatmaps for the EN, $G\bar{o}_{\text{intra}}$, and $G\bar{o}_{\text{intra}+\text{inter}}$ models, respectively. Upper triangles show residue-wise contact maps, while lower triangles display domain-wise contact maps, calculated by summing the contact probabilities and averaging over the total number of residues within each domain. Residue- and domain-wise contact maps share the same colorbar. (D). Averaged inter-molecular contact probability of each residue across the three models: EN in green, $G\bar{o}_{\text{intra}}$ in orange, and $G\bar{o}_{\text{intra}+\text{inter}}$ in dark green. Aromatic residues are indicated with arrows.

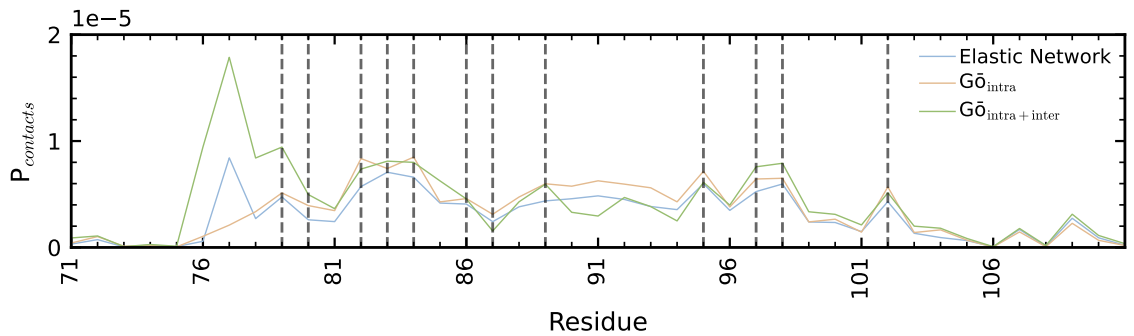


Figure 3.9: Zoomed-in view of residue-level averaged intermolecular contact probabilities within the NLS region. Charged residues are indicated by black dashed lines.

3 Regulating TDP-43 LLPS via structured domain stabilization and inter-chain bridging

Overall, the $G\bar{o}$ -based models not only preserved the multivalent interactions driven by disordered regions but also expanded the inter-chain interaction network to include structured domains. These results suggest that flexible and folded regions cooperate synergistically to stabilize the condensate while maintaining its dynamic internal organization, underscoring the importance of multivalent interactions across all domains in shaping condensate properties.

3.3.3 Density distribution within condensates

The equilibrium concentrations of TDP-43 in the dilute and dense phases were quantified under the $G\bar{o}_{\text{intra}}$ and $G\bar{o}_{\text{intra}+\text{inter}}$ models using the same method as for the Elastic Network model (Method 5.2.1) within the slab geometry.

At $\lambda_{1.06}$, the $G\bar{o}_{\text{intra}+\text{inter}}$ model exhibited more pronounced phase separation than the $G\bar{o}_{\text{intra}}$ model, as evidenced by a higher dense-phase concentration and a correspondingly lower dilute-phase concentration (Figure 3.10B and D). This enhanced stability of the dense phase was closely linked to differences in chain dynamics at the condensate interface.

In the $G\bar{o}_{\text{intra}}$ model, individual chains frequently exchanged between the dilute and dense phases (Figure 3.10A). This dynamic behavior was consistently observed across replicates at $\lambda_{1.06}$ (Figure 3.11) and contributed to a relatively elevated dilute-phase concentration (Figure 3.12). By contrast, in the $G\bar{o}_{\text{intra}+\text{inter}}$ model, chains approaching the condensate surface were often retained through attractive inter-chain interactions, particularly those mediated by N-terminal domains (NTDs) (Figure 3.13). This “surface capture” mechanism effectively impeded chain escape, thereby reducing the dilute-phase concentration and stabilizing the dense phase.

Together, these observations suggest that the inclusion of inter-chain $G\bar{o}$ interactions not only strengthens condensate cohesion but also alters interfacial dynamics, suppressing the exchange of chains between phases. This mechanistic insight highlights how subtle differences in interaction networks at the condensate boundary can have substantial consequences for macroscopic phase behavior.

Concentrations of coexisting dense and dilute phases were quantified across a range of λ values, with standard errors used to construct phase diagrams (Figure 3.14A). As λ increased, the dense-phase concentration systematically decreased, indicating that stronger protein–water interactions enhanced the solubility of TDP-43 in the aqueous phase and progressively suppressed LLPS. This trend highlights the delicate balance between solvation forces and intermolecular interactions in determining the propensity for phase separation.

Across the λ range examined, the $G\bar{o}_{\text{intra}}$ model consistently exhibited more pronounced phase separation compared to the Elastic Network baseline. Notably, at $\lambda_{1.06}$ and $\lambda_{1.07}$, condensates in the $G\bar{o}_{\text{intra}}$ model remained highly stable, despite the stronger pro-

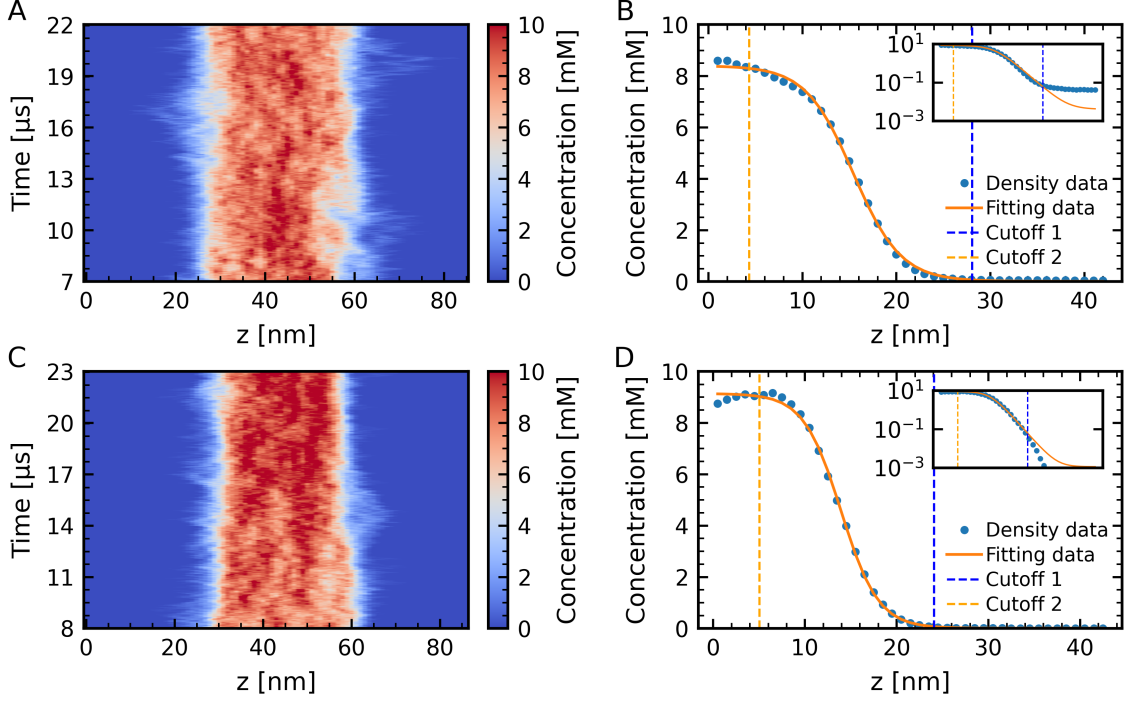


Figure 3.10: Condensate density analysis under G_{intra} and $G_{\text{intra+inter}}$ models at $\lambda_{1.06}$. (A, C). Time-dependent density histograms of the condensates for the G_{intra} and $G_{\text{intra+inter}}$ models, respectively. (B, D). Corresponding averaged density histograms, fitted to a hyperbolic tangent (tanh) function (Eq. 5.2) to extract dense- and dilute-phase concentrations.

tein–water interactions that would otherwise favor solubilization. This enhanced stability is likely attributable to the increased inter-chain contacts captured by the G_{intra} potential, which reinforce the internal network of the dense phase and resist the escape of chains into the dilute phase. In other words, attractive inter-chain interactions effectively counterbalance the solubilizing effect of protein–water interactions, preserving condensate integrity under conditions that partially destabilize LLPS.

Analysis of the full phase diagrams further revealed that the G_{intra} model shifted the coexistence curve outward, resulting in a higher critical point relative to the EN model. This outward shift indicates that condensates could remain stable under conditions where the baseline model would no longer support phase separation, suggesting that the inter-chain interaction network plays a key role in expanding the range of conditions compatible with stable condensates.

Taken together, these findings demonstrate that LLPS is governed by a competition between solubilizing forces and inter-chain cohesive interactions. Increasing protein–water

3 Regulating TDP-43 LLPS via structured domain stabilization and inter-chain bridging

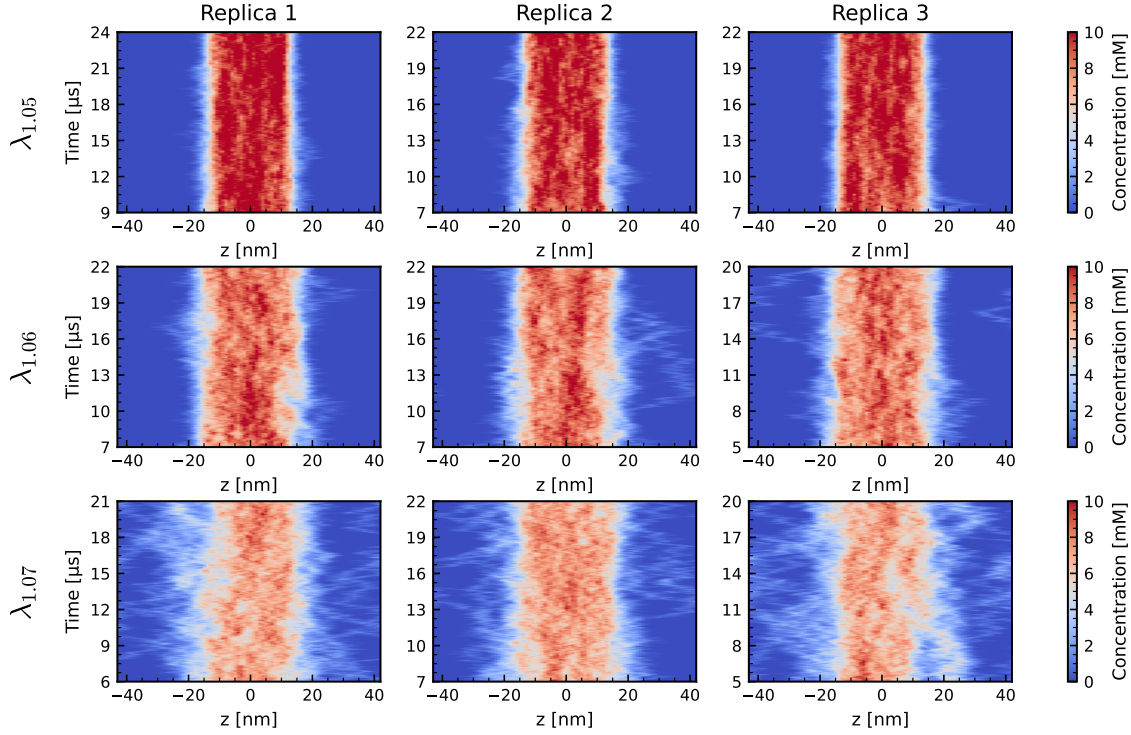


Figure 3.11: Time-dependent density histograms along the Z-axis across replicates of condensates for the $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.05}$, $\lambda_{1.06}$, and $\lambda_{1.07}$ in $G\bar{o}_{\text{intra}}$ model.

interaction strength counteracts phase separation by promoting solubility and weakening the driving forces that stabilize the dense phase, whereas enhanced inter-chain contacts act to preserve condensate structure and maintain phase integrity. These results provide mechanistic insight into how subtle modifications of interaction networks can fine-tune the stability and composition of biomolecular condensates.

3.3.4 Surface tension and shear viscosity

The surface tension of protein condensates is a critical biophysical parameter that governs droplet morphology, stability, and biological function [78]. It reflects the net balance between cohesive protein–protein interactions and solvation effects at the condensate boundary. Condensate interfacial tension was computed for the Elastic Network (EN) model and two distinct $G\bar{o}$ -based models to examine how inter-chain interactions influence surface properties.

As shown in Figure 3.14B, the surface tension decreased with increasing values of λ , indicating that stronger protein–water interactions weaken the cohesive forces stabilizing the dense phase. For a fixed solvent interaction strength, the different models yielded

3.3 Full-length TDP-43 condensation via LJ-enhanced models

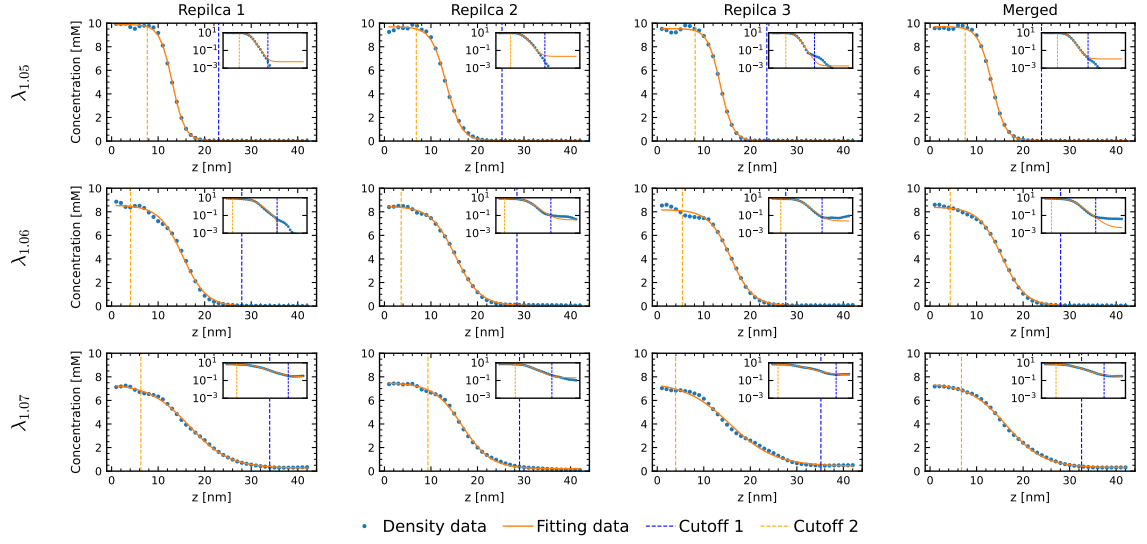


Figure 3.12: Averaged density histograms fitted to a hyperbolic tangent (tanh) function 5.2 to extract dense and dilute phase concentrations under the $G\bar{o}_{\text{intra}}$ model at $\lambda_{1.05}$, $\lambda_{1.06}$ and $\lambda_{1.07}$.

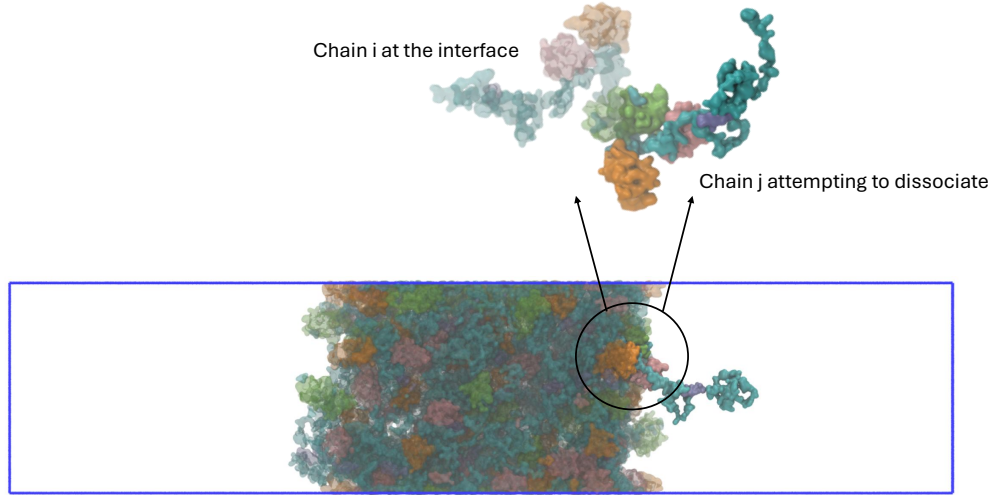


Figure 3.13: A chain trapped at the condensate interface in the $G\bar{o}_{\text{intra}+\text{inter}}$ model at $\lambda_{1.06}$.

3 Regulating TDP-43 LLPS via structured domain stabilization and inter-chain bridging

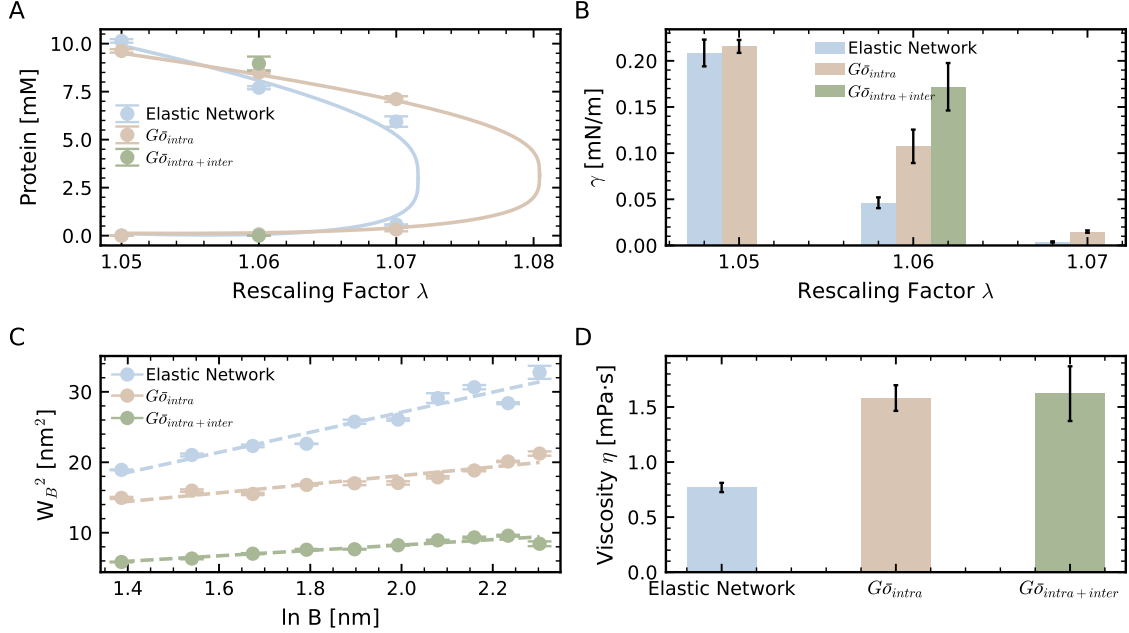


Figure 3.14: (A). Phase diagram of the Elastic Network (EN) and $G\bar{\sigma}_{intra}$ models across varied rescaling factors λ , with the dense- and dilute-phase concentrations of $G\bar{\sigma}_{intra+inter}$ shown at $\lambda_{1.06}$. (B). Surface tension measurements for each model. (C). Interfacial width of condensates across models at $\lambda_{1.06}$. (D). Viscosity estimated from end-to-end distance autocorrelations for each model at $\lambda_{1.06}$.

quantitatively distinct values for both surface tension and interfacial width (Figure 3.14B, C and 3.15). At $\lambda_{1.05}$, the EN and $G\bar{\sigma}_{intra}$ models produced comparable values of surface tension and interfacial width, suggesting a similar balance between cohesive inter-chain forces and solvation effects. In contrast, at $\lambda_{1.06}$, the EN model exhibited a lower surface tension and a broader interfacial width relative to both $G\bar{\sigma}_{intra}$ and $G\bar{\sigma}_{intra+inter}$ models, indicating that weaker inter-chain interactions allow greater interfacial fluctuations and reduce energetic penalties for residue exposure to solvent.

The two $G\bar{\sigma}$ -based models introduce additional attractive contacts between chains, providing stronger energetic stabilization within the dense phase. This increased cohesion produces a larger energetic penalty for exposing residues to solvent [79], resulting in a higher surface tension and a sharper, narrower interface. Strong inter-chain attractions effectively pull chains inward, restricting protrusion at the boundary, whereas weaker interactions permit chain segments to extend and fluctuate more freely, producing broader and more diffuse interfacial regions.

At $\lambda_{1.07}$, all models showed a dramatic reduction in surface tension, consistent with behavior approaching the critical point where dense and dilute phases become indistinguishable.

3.3 Full-length TDP-43 condensation via LJ-enhanced models

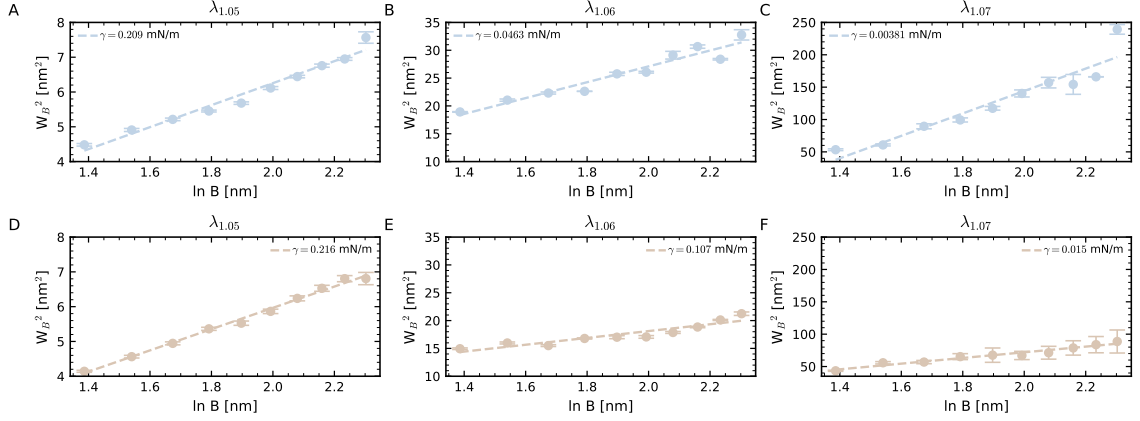


Figure 3.15: Interfacial width W^2 as a function of block size B . The interface region is divided into rectangular segments, and the interfacial tension γ is determined using $W^2(B) = W_0^2 + \frac{k_B T}{2\pi\gamma} \ln \frac{L}{B}$, where W_0 is the intrinsic interfacial width, k_B is the Boltzmann constant, T is the temperature, and L is the simulation box size.

Despite this general decrease, the G $\bar{o}$ -based models consistently exhibited higher surface tension and sharper interfaces compared to the EN model, reflecting the stabilizing effect of stronger inter-chain contacts. These results highlight how variations in intra- and inter-chain interactions tune the physical properties of condensate surfaces, modulating both interfacial width and surface tension, which in turn can influence droplet stability, morphology, and potentially their biological function.

The shear viscosity of the protein condensates was estimated from the variance of the chain end-to-end distance and its relaxation time using Eq. 5.8 under $\lambda_{1.06}$ across the three models. The end-to-end (ee) distance relaxation time, τ , was determined from the autocorrelation function of the ee distance, $C(t)$, which quantifies how the chain's end-to-end distance decorrelates over time:

$$C(t) = \frac{\langle \mathbf{R}_{ee}(t_0) \cdot \mathbf{R}_{ee}(t_0 + t) \rangle}{\langle \mathbf{R}_{ee}(t_0) \cdot \mathbf{R}_{ee}(t_0) \rangle}, \quad (3.3)$$

where $\mathbf{R}_{ee}(t)$ is the end-to-end distance of a polymer chain at time t , and $\langle \cdot \rangle$ denotes a time average over all starting times t_0 .

To account for multi-timescale relaxation behavior, $C(t)$ was fitted to a biexponential function:

$$C(t) = A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2}, \quad (3.4)$$

3 Regulating TDP-43 LLPS via structured domain stabilization and inter-chain bridging

where A_1 and A_2 are amplitudes, and τ_1 and τ_2 are characteristic relaxation times of the fast and slow components, respectively. The effective relaxation time τ was then extracted from the fitted parameters as the amplitude-weighted sum of the two relaxation times in the biexponential fit to the autocorrelation function, averaged over all protein chains in the system.

This approach enabled a quantitative estimation of the internal constraints imposed on polymer chains within the condensate that limit their conformational relaxation, yielding an effective measure of shear viscosity.

Condensates under the Elastic Network model, with fewer inter-chain contacts and faster chain reconfiguration, exhibited lower viscosity than either of the two $G\bar{o}$ -based models.

Between the two $G\bar{o}$ -based models, the viscosities were statistically similar in terms of their mean values and chain reconfiguration. However, the viscosity estimate for the $G\bar{o}_{\text{intra}+\text{inter}}$ model displayed a larger error bar than that of the $G\bar{o}_{\text{intra}}$ model, leading to increased uncertainty in the viscosity estimation.

Although the shear viscosity was inferred from end-to-end relaxation dynamics, it primarily reflects the extent to which chain conformational relaxation may be influenced by crowding from neighboring polymer chains and interactions with the surrounding solvent. Increased inter-chain coupling is expected to slow relaxation and potentially elevate viscosity, whereas reduced intermolecular constraints may allow more rapid reconfiguration and lower viscosity.

The larger error bar observed for the viscosity estimate of the $G\bar{o}_{\text{intra}+\text{inter}}$ model likely reflects greater variability in chain relaxation dynamics rather than a systematically higher or lower viscosity. While the mean viscosity is comparable to that of the $G\bar{o}_{\text{intra}}$ model, the stronger inter-chain interactions in the $G\bar{o}_{\text{intra}+\text{inter}}$ model may impose additional constraints on some chains. As a result, different chains could relax at different rates depending on their transient interaction networks, leading to a broader spread in end-to-end relaxation times. This variability contributes to increased statistical uncertainty in the viscosity estimate without necessarily affecting the average value.

The similar mean viscosities suggest a comparable overall resistance to flow at the condensate scale, while the enhanced inter-chain interactions in the $G\bar{o}_{\text{intra}+\text{inter}}$ model may lead to stronger and more variable constraints that fluctuate over time. The viscosity estimates are based on end-to-end relaxation, which can exhibit substantial fluctuations in relaxation times. If estimated carefully, viscosity could also be assessed using mean-squared displacement (MSD) analysis as an alternative approach.

3.4 Discussion

This study applied fewer restraints within the structured domains by using Lennard-Jones (LJ) potentials, allowing the folded regions of the proteins to remain flexible. This increased conformational freedom facilitated additional inter-chain interactions, particularly through N-terminal domain (NTD) contacts, which further strengthened the overall network of interactions within the condensate. As a result, the enhanced inter-chain interactions promoted protein phase separation under identical solvent conditions, emphasizing the importance of specific residue-residue contacts in stabilizing the dense phase.

Modifications in the internal inter-molecular interaction network also affected the properties of the condensate interface. Stronger inter-chain interactions, particularly those mediated by NTD–NTD contacts, contributed to a sharper and more cohesive interface by limiting chain protrusion and reducing fluctuations at the boundary. In contrast, the more weakly interacting EN regions allowed greater chain mobility near the interface, resulting in a broader and more diffuse interfacial region. Despite these differences at the boundary, the internal chain dynamics within the bulk of the condensate remained largely similar between the $G\bar{o}_{\text{intra}}$ and $G\bar{o}_{\text{intra}+\text{inter}}$ models. This suggests that the primary impact of enhanced inter-chain contacts is on interfacial cohesion, rather than on overall internal mobility within the condensate.

Overall, these results indicate that modifications in the internal interaction network, particularly through enhanced inter-chain contacts, can tune both the stability of the dense phase and the structural characteristics of the condensate interface, even when the internal dynamics of chains within the bulk remain largely unchanged.

The three semi-coarse-grained models reveal differences in the internal interaction networks of the condensates, although the overall patterns are similar. In the case of TDP-43, it has been shown that NTD dimerization plays an important role in modulating condensate behavior. For other proteins, the role of inter-chain interactions has not been clearly characterized, and further studies are needed to determine whether the models capture the key contacts accurately.

From a computational perspective, the $G\bar{o}_{\text{intra}}$ and $G\bar{o}_{\text{intra}+\text{inter}}$ models include additional virtual beads to implement the extra potentials, which increases the total number of beads in the system. As a result, using the same computational resources, simulations with the $G\bar{o}_{\text{intra}}$ and $G\bar{o}_{\text{intra}+\text{inter}}$ models require roughly two to three times longer than simulations with the EN model. Therefore, the choice of model largely depends on the specific scientific questions and computational constraints of the study.

4 Helical Structure within the TDP-43 C-Terminal Intrinsically Disordered Region

Within the C-terminal disordered region of TDP-43, a helical structure enhances the liquid–liquid phase separation (LLPS) of the LCD fragment. The helix spans residues 321 to 332 as a core and extends to residue 343 [51, 69, 80]. Trp334, located at the boundary of the helical core, plays a key role in mediating inter-chain interactions within the dynamic, liquid-like condensates. Notably, this same region can also adopt β -sheet conformations, which are associated with less dynamic, pathological aggregates [81].

This structural duality raises a fundamental mechanistic question, what are the respective contributions of secondary structure versus specific “sticky” residues in driving phase behavior? Specifically, does the transient α -helix primarily serve as a scaffold that optimally presents residues such as Trp334 for dynamic intermolecular interactions, or does the helical conformation itself encode a unique interaction interface that is critical for maintaining liquid-like properties? Resolving this balance is essential for understanding how the same sequence can give rise to either dynamic liquid condensates or static solid aggregates, and provides insight into the molecular determinants that govern condensate material properties.

4.1 Residue W334 intra-contacts within the helix and inter-contacts in the condensate

The single-chain helical structure was simulated at the all-atom level using two different initial configurations (Figure 4.1), each with three independent replicas. The simulations were performed using the same setup as that applied to the full-length protein described in Chapter 2.2.2.2.

The intra-residue contacts were defined using specific physical cutoff distances. Within this framework, Trp334 exhibited a high contact probability with hydrophobic residues Phe313, Leu330, and Met337, as well as with polar residues Gln327 and Gln331. Neighboring residues at positions $i \pm 2$ (i.e., 332 and 336) were omitted from certain contact analyses for clarity (Figure 4.2A and B); nevertheless, these residues contributed to stabilizing the helical fold through conventional backbone hydrogen bonding and local steric packing.

The interactions among hydrophobic residues follow the canonical i , $i + 3$, and $i + 4$

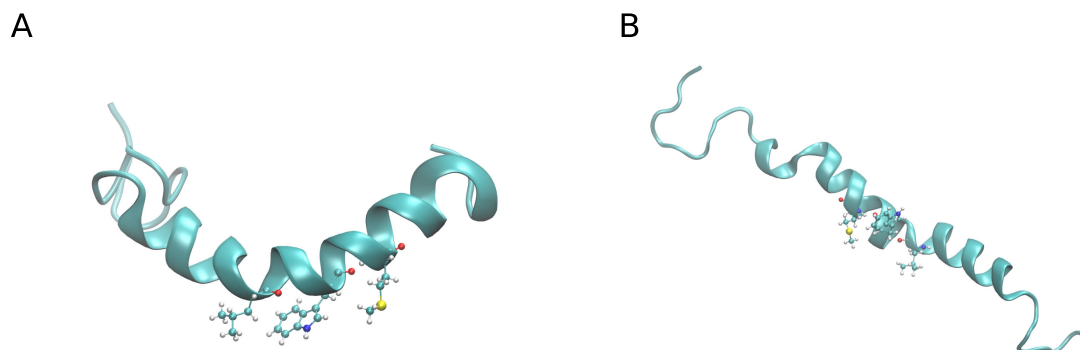


Figure 4.1: All-atom simulations of helix starting configurations. Initial structures derived from PDB entries 2N2C (A) and 2N2X (B), with residues Leu330, Trp334, and Met337 highlighted.

spacing characteristic of a helical structure, in which the carbonyl oxygen (O) at position i forms a hydrogen bond with the amide hydrogen (H) at position $i + 4$. Specifically, the carbonyl oxygen of Leu330 forms a hydrogen bond with the amide hydrogen of Trp334, corresponding to a characteristic $i + 4$ interaction, while the carbonyl oxygen of Trp334 forms a hydrogen bond with the amide hydrogen of Met337, corresponding to an $i + 3$ interaction (Figure 4.2C and D; bonds are indicated by black dashed lines). This hydrogen-bonding pattern is consistent with the canonical backbone interactions of an α -helix. Beyond this backbone framework, direct side-chain contacts, particularly among Leu330, Trp334, and Met337, contribute significantly to structural cohesion and further stabilize the helical fold (Figure 4.2C and D).

The Trp334–Phe313 ($i - 21$) contact represents a long-range π – π interaction that buries hydrophobic surface area, thereby shielding both residues from the polar solvent (Figure 4.3A).

Trp334 also forms contacts with nearby polar residues. The Trp334–Gln327 ($i - 7$) and Trp334–Gln331 ($i - 3$) pairs engage in amino– π interactions (Figure 4.3B and C). The Trp334–Gln327 interaction involves side-chain amino– π stacking and introduces a subtle local bend in the helix (Figure 4.3B). In contrast, the Trp334–Gln331 interaction is structurally compatible with the α -helical geometry and primarily serves to desolvate the hydrophobic face of Trp334 by shielding it from solvent exposure (Figure 4.3C).

In addition, Trp334 and Gln327 form a stabilizing hydrogen bond, while Met337 contributes hydrophobic packing interactions. Together, these contacts create a tightly packed three-residue interaction motif that reinforces local helical stability and reduces solvent exposure of Trp334 (Figure

4.1 Residue W334 intra-contacts within the helix and inter-contacts in the condensate

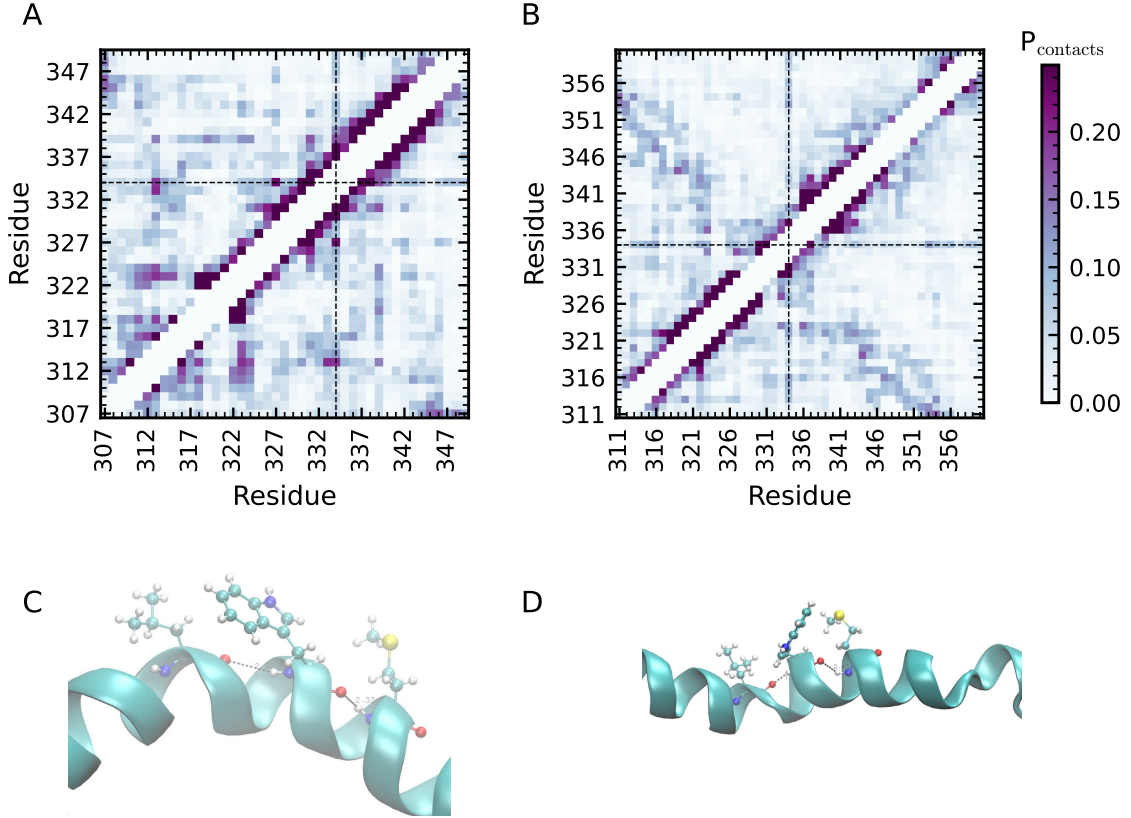


Figure 4.2: Intra-chain interactions within the helix in all-atom simulations. (A–B). Intra-chain contact probability heatmaps for 2N2C (A) and 2N2X (B), with residue Trp334 indicated by dashed lines. (C–D). Zoomed-in views illustrating the local packing and contacts among residues Leu330, Trp334, and Met337 within the helical structures of 2N2C and 2N2X, respectively. Dashed black lines indicate hydrogen bonds between residues Leu330–Trp334 and Trp334–Met337.

In single-helix W334A mutant simulations, the Trp334–Leu330, Trp334–Gln331, and Trp334–Met337 interactions were partially retained, whereas the Trp334–Gln327 contact was lost (Figure 4.4A and B). The helix propensity across residues within the helical core region was generally reduced in the mutant compared to the WT (Figure 4.4C and D). Importantly, the backbone $i \rightarrow i + 3$ and $i \rightarrow i + 4$ hydrogen bonds that define the helix were not completely lost, but were weakened relative to the WT (e.g., the 334–Leu330 and 334–Met337 interactions). As a consequence, the helix within the core region (residues 320–332) was modestly destabilized rather than fully unfolded, whereas the extended helical segment (up to residue 343) became disordered or unstructured (Figure 4.5A). Notably, the backbone helical network involving residues 330–334–337 was disrupted.

Replacement of Trp with Ala removed the bulky aromatic indole side chain that mediates

4 Helical Structure within the TDP-43 C-Terminal Intrinsically Disordered Region

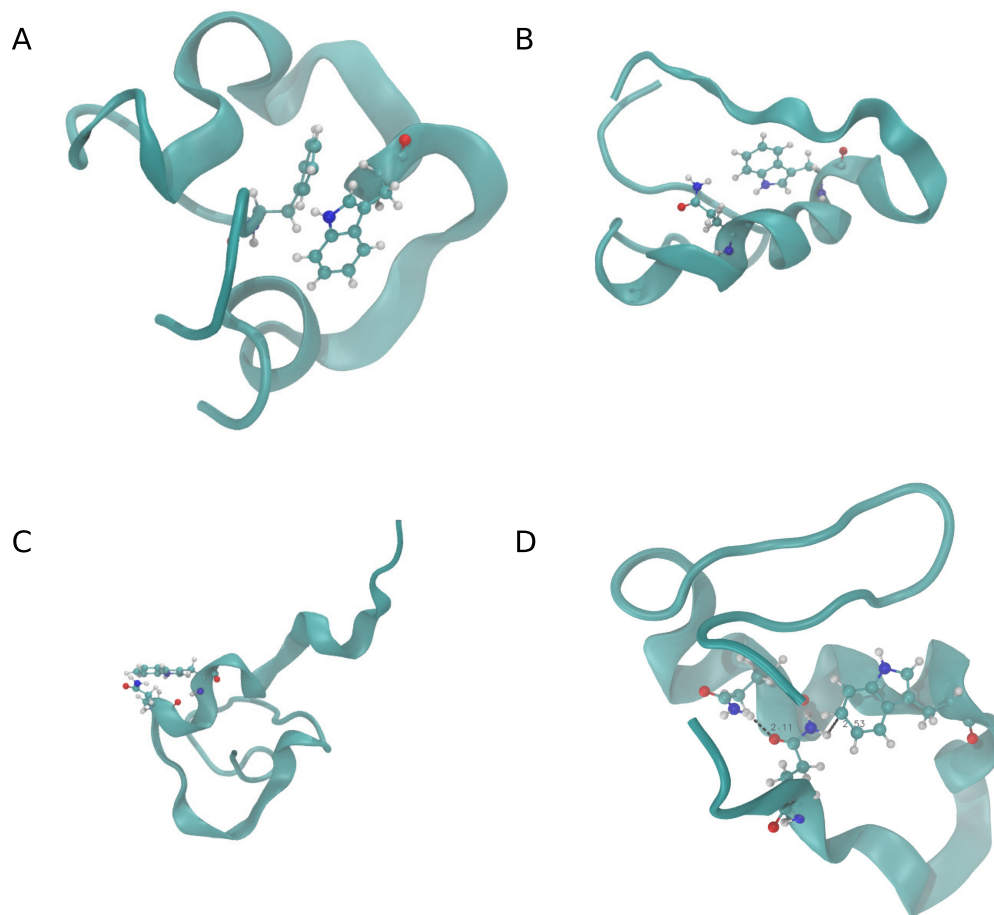


Figure 4.3: Representative structures illustrating side-chain contacts mediated by Trp334 within the helix. A. Representative snapshot of the Phe313–Trp334 side-chain contact within the helix, with the atoms of Phe313 and Trp334 shown. B. Side-chain interaction between residues Gln327 and Trp334. C. Side-chain interaction between residues Gln331 and Trp334. D. Representative configuration showing the combined side-chain contacts involving residues Gln327, Trp334, and Met337.

extensive side-chain packing and solvent shielding. As a result, the side-chain-driven hydrophobic face of the helix was disrupted, and the 334–337 side-chain contact was substantially weakened, indicating that helix stability in the WT is supported by Trp-mediated side-chain interactions. Although the 334–331 interaction persisted in the W334A mutant, it was maintained primarily through backbone hydrogen bonding and inward backbone packing rather than the extensive side-chain packing observed in the WT.

4.1 Residue W334 intra-contacts within the helix and inter-contacts in the condensate

Overall, the W334A mutation reduced side-chain-driven hydrophobic stabilization while preserving the backbone hydrogen-bond scaffold, resulting in a measurable loss of helical stability.

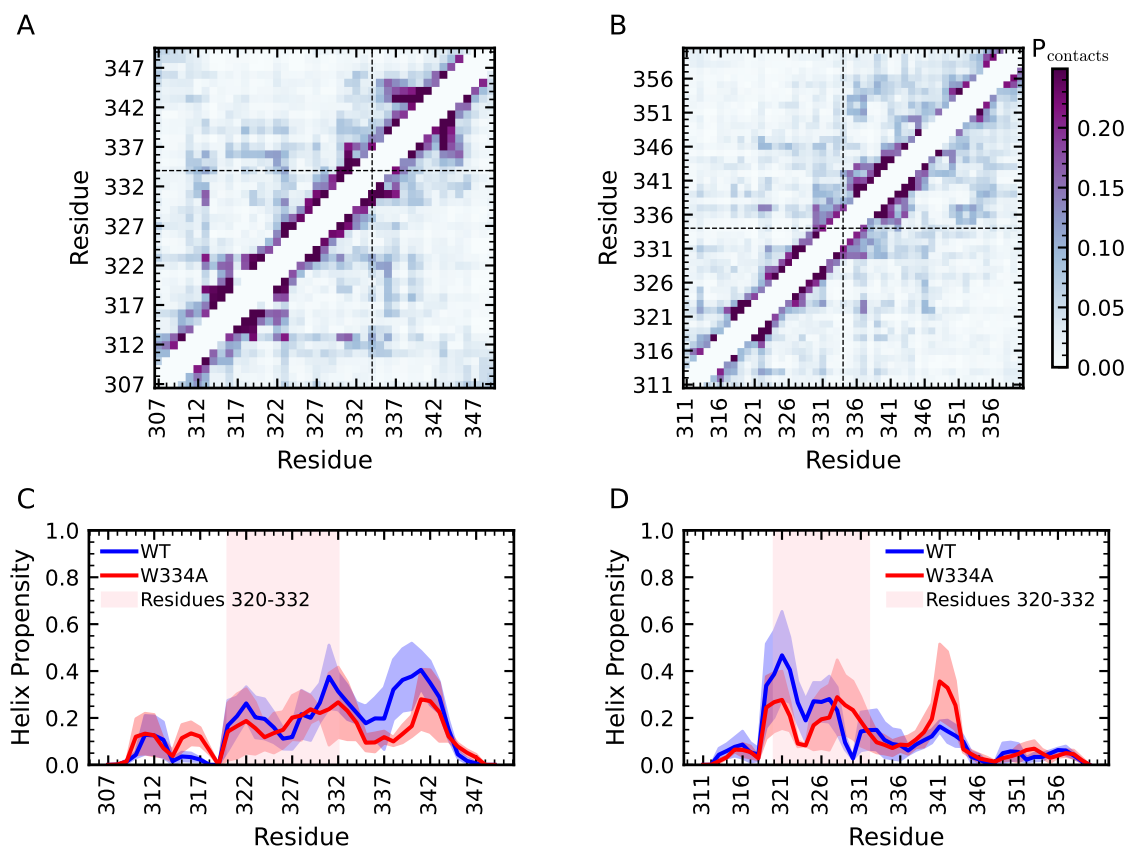


Figure 4.4: Intra-chain interactions and helix propensities of the W334A mutant in all-atom simulations. (A). Intra-chain contact probability heat map for the W334A mutant using PDB:2n2c as the starting structure. (B). Intra-chain contact probability heat map for the W334A mutant using PDB:2n2x as the starting structure. (C). Residue-resolved helix propensities for the W334A mutant starting from PDB:2n2c. (D). Residue-resolved helix propensities for the W334A mutant starting from PDB:2n2x.

Within the liquid-like condensate, Trp334 maintained high inter-chain contact activity, while helix preservation was enforced through harmonic-bond constraints in the adapted Martini 3 force field (Chapter 2.2.2). Trp334 formed strong inter-chain interactions with aromatic residues from neighboring chains, supporting its role as a key aromatic interaction node within the condensate environment (Figure 4.6A). In addition, Trp334 exhibited strong intramolecular contacts with residues 327–331, reinforcing the helical backbone hydrogen-bond scaffold, and the Trp334–Met337 interaction remained stable in the dense phase (Figure 4.6C).

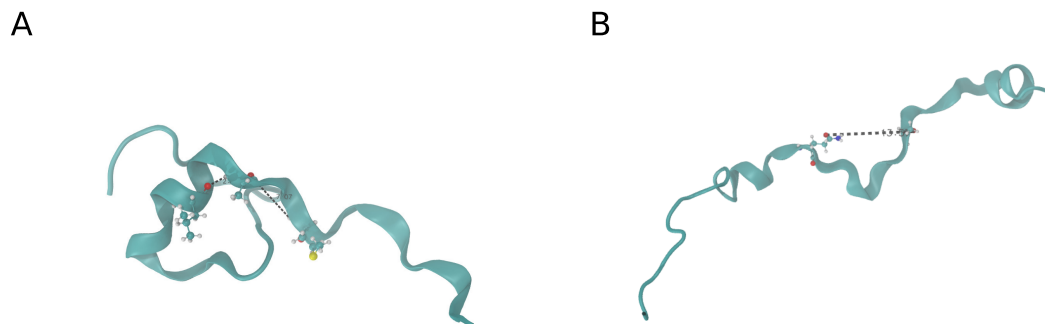


Figure 4.5: Representative snapshots illustrating the Leu330–Trp334–Met337 interaction motif and the Trp334–Met337 contact within the helix of the W334A mutant. (A). Leu330–Trp334–Met337 interaction motif; black dashed lines indicate inter-residue distances. (B). Trp334–Met337 contact within the helix; black dashed lines indicate inter-residue distances.

In contrast, the W3A condensate—in which all Trp residues in the disordered region were mutated to Ala (Chapter 2.3.1)—exhibited substantially reduced intra-chain interactions (Figure 4.6B). Although the 330–334–337 interaction motif could still be detected, its occurrence was strongly diminished compared to the WT condensate. This reduction in intramolecular contacts is likely due to the absence of the Trp334 side chain, the lack of additional helix-stabilizing harmonic restraints (Elastic Network), or a combination of both. Consequently, the relative contributions of sequence mutation versus enforced helix stabilization cannot be fully disentangled in the current setup. A WT condensate simulation without helix stabilization would therefore be required for a direct and controlled comparison, representing a limitation of the present study.

Consistent with these observations, the W3A condensate showed a dramatic loss of inter-chain contact activity at the position corresponding to residue 334 (Figure 4.6B). This result highlights a mechanistic ambiguity: while the presence of the helix has been shown to be critical for condensate formation, it remains unclear whether this requirement arises from the helical secondary structure itself or from the aromatic residues embedded within the helix that mediate inter-chain interactions. In the W3A system, disruption of aromatic side chains within the helical region coincides with reduced helix stability and diminished intermolecular contacts, making it difficult to disentangle the relative contributions of helix geometry versus Trp-mediated aromatic interactions to condensate cohesion. Addressing this question would require either a WT condensate simulation without helix stabilization

4.1 Residue W334 intra-contacts within the helix and inter-contacts in the condensate

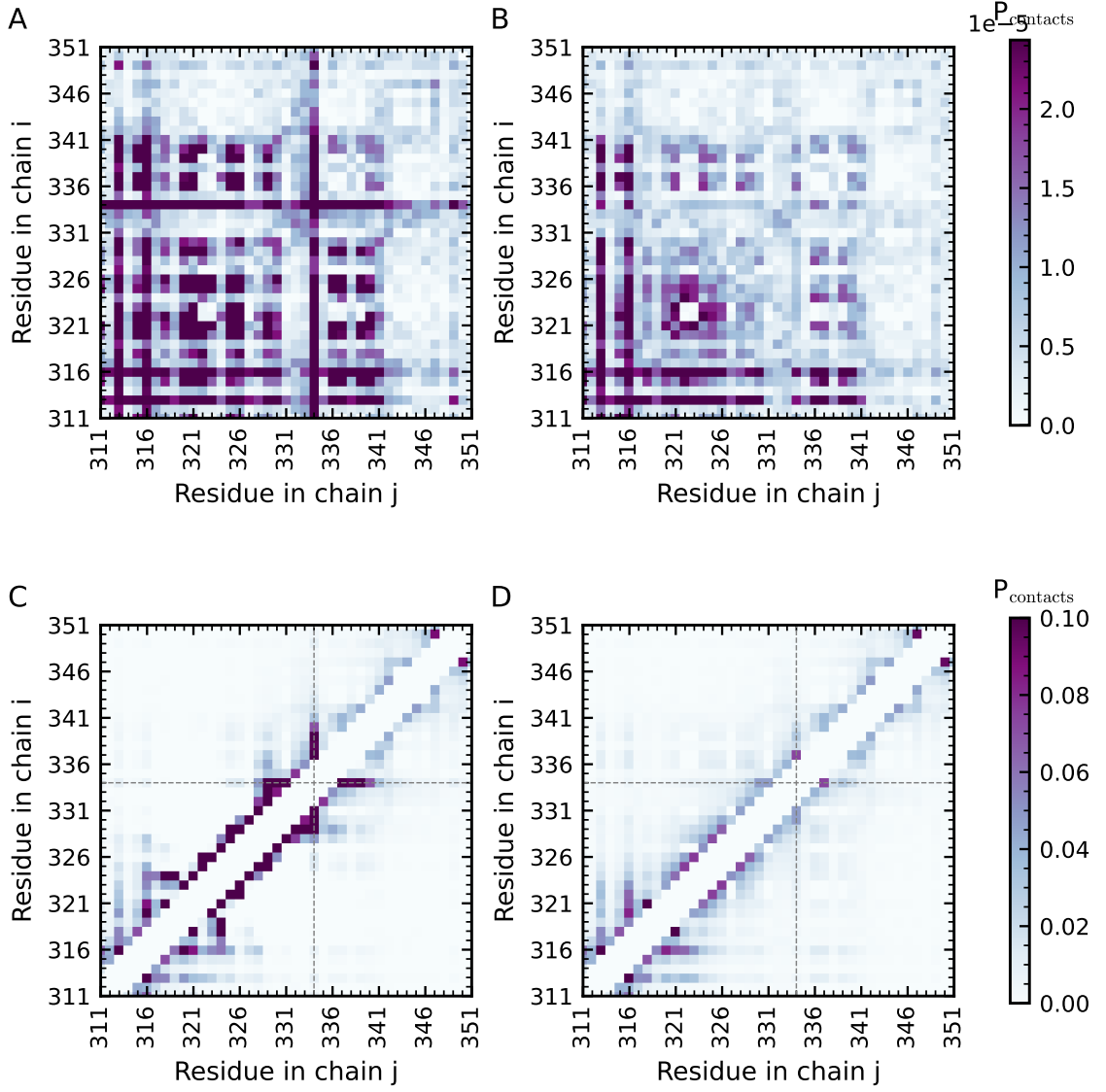


Figure 4.6: Contact heatmaps of WT and W3A condensates showing inter- and intra-chain interactions under the EN model at $\lambda_{1.06}$. (A–B). Inter-chain contact heatmaps for WT and W3A condensates, respectively. (C–D). Intra-chain contact heatmaps for WT and W3A condensates, respectively.

or a W3A mutant with enforced helix stabilization.

Together, these results indicate that Trp334 is essential not only for stabilizing the local helical structure but also for mediating the inter-chain aromatic and hydrophobic interactions that underpin condensate cohesion.

4.2 Residue Ser332 and Ser333 in helix

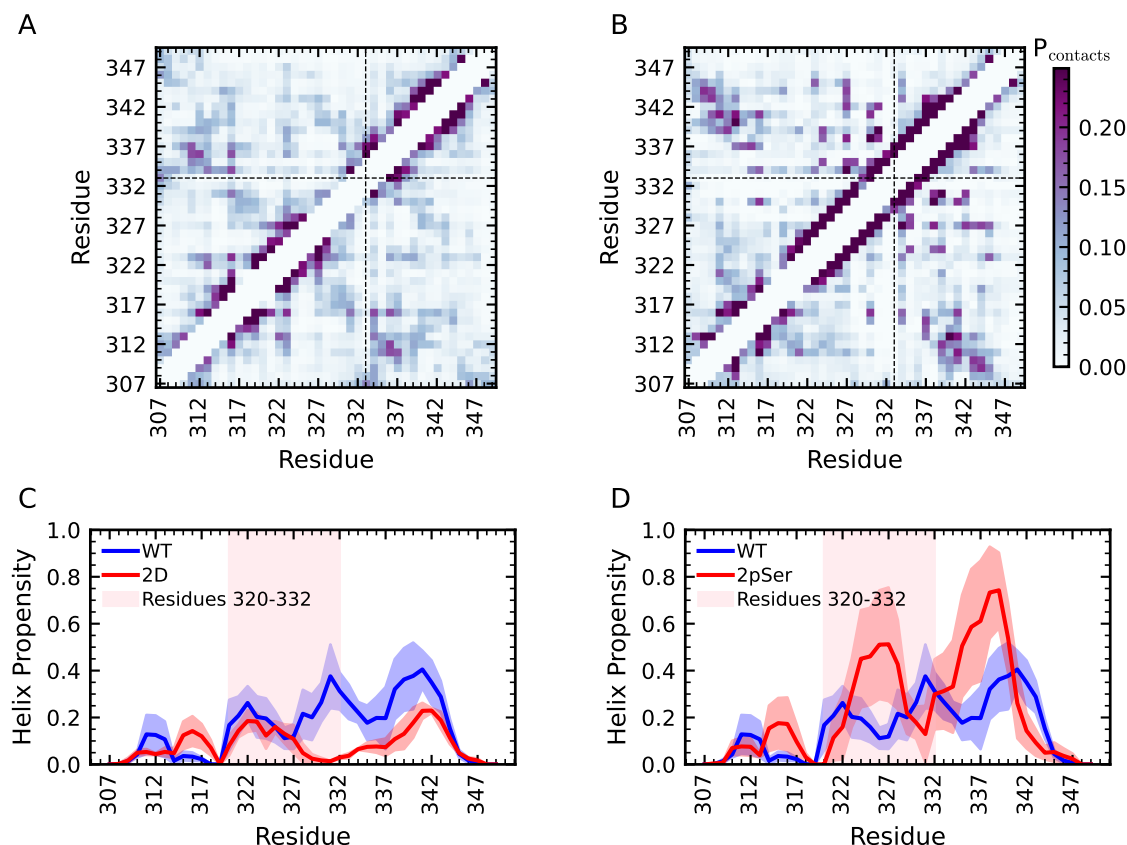


Figure 4.7: Phospho-mimicking or phosphorylated Ser residues at positions 332 and 333 and their effect on helix stability in all-atom simulations. (A–B). Intra-chain contact heatmaps for the phospho-mimicking and phosphorylated Ser residues, respectively. (C–D). Residue-resolved helix propensities for the phospho-mimicking and phosphorylated Ser residues, respectively.

Two serine residues lie at the edge of the core helical region, and serines in the C-terminal disordered region represent major phosphorylation sites that are key pathological hallmarks in neurodegenerative disease. Within the isolated helix, Ser332 and Ser333 form several stabilizing intra-chain contacts, including Ser333–Leu330, Ser333–Met336, Ser332–Ala329, and Ser332–Gly335, which reinforce the local helical scaffold (Figure 4.2A and B).

Within the condensate environment, the contribution of these serine residues to inter-chain interactions was minimal. Helix–helix contacts are primarily mediated by hydrophobic and aromatic residues, with Ser332 and Ser333 participating only weakly at the intermolecular level. Ser333 exhibited somewhat higher inter-chain interactions with nearby aromatic



Figure 4.8: Representative snapshots illustrating two phosphorylated Ser residues in all-atom simulations. All atoms of the phosphorylated Ser residues are shown.

residues (e.g., Phe313, Phe316, and Trp334), likely due to its proximity to Trp334. Overall, their primary role is intramolecular stabilization rather than supporting condensate cohesion (Figure 4.6A).

Replacing Ser with Asp, commonly used as a phosphomimetic (2D), or explicitly introducing phosphate groups at both Ser332 and Ser333 (2pSer) produced distinct effects on helix stability. The Ser→Asp substitutions modestly destabilized the helix (Figure 4.7C), whereas phosphorylation did not destabilize the structure; instead, it increased the stability of the isolated core helical region (Figure 4.7D). In the Asp mutant, contacts involving Ser333–Leu330, Ser333–Met336, Ser332–Ala329, and Ser332–Gly335 were substantially reduced and often undetectable (Figure 4.7A). In contrast, these interactions were enhanced in the phosphorylated system (Figure 4.7B).

Phosphorylated serine side chains preferentially projected outward from the helix, enabling new long-range intramolecular interactions (e.g., Phe313–Leu340). These interactions promoted a subtle bend within the helical fragment, suggesting that phosphorylation stabilized the helix through a mechanism distinct from local side-chain packing (Figure 4.8).

4 Helical Structure within the TDP-43 C-Terminal Intrinsically Disordered Region

The results described above were obtained from single-helix all-atom simulations, and how these effects translate to condensate formation remains unknown. Nevertheless, several aspects can be assessed once condensate simulations are performed to evaluate the contributions of helix stability and residue-specific interactions to phase separation. Notably, Asp carries a net charge of -1 , whereas a phosphate group carries a net charge of -2 . Introducing these additional negative charges may either dissolve the condensate or alter its compactness. If condensate dissolution is observed, it remains unclear whether this effect arises primarily from increased electrostatic repulsion or from changes in helix stability induced by the modifications. Disentangling these contributions will be essential for understanding how charge modulation and secondary-structure stability jointly regulate condensate behavior.

4.2.1 12pSer

To extend the study of serine phosphorylation within the low-complexity domain (LCD) (Chapter 2.3.2), twelve phosphate groups were incorporated into the full-length protein model. Building on earlier insights, preliminary investigations were conducted using both all-atom simulations and residue-level implicit coarse-grained approaches. These initial results motivated a more detailed analysis of phosphorylation-induced effects on protein conformation and phase behavior, which is presented in the following section.

Consistent with the single-chain WT TDP-43 simulations, all-atom simulations of the phosphorylated protein (mutant 12pSer) were performed using the AMBER99SB*-ILDN-q protein force field in combination with the TIP4P-D water model. Phosphoserine parameters were adopted from [82], with corrections to the oxygen parameters from [83]. Each simulation was run for 1 μ s, with three independent replicates.

Phosphate groups located distal to the C-terminal helix enhanced helix stability (Figure 4.9A and B). The helical propensity increased across nearly all residues, affecting not only the core helix region (residues 320–332) but also the extended helical segments. In contrast, intra-chain interactions differed only modestly from those of the WT protein (Figure 4.10A and B). Most contacts remained localized within folded regions, while long-range inter-domain contacts, particularly between the second linker and the LCD, were reduced upon phosphorylation.

Focusing on the LCD region (residues 261–414), the positions of phosphorylated serine residues are indicated by arrows at the top of both the WT and 12pSer mutant contact heatmaps (Figure 4.10C and D). Overall, long-range contacts within the LCD were reduced in the 12pSer mutant, with the strongest decreases observed at phosphorylated serine sites. In addition, contacts between the LCD and the C-terminal helix were diminished. This effect may result from the negatively charged phosphate groups recruiting counterions or water molecules, thereby screening intermolecular interactions and hindering residue–residue contacts.

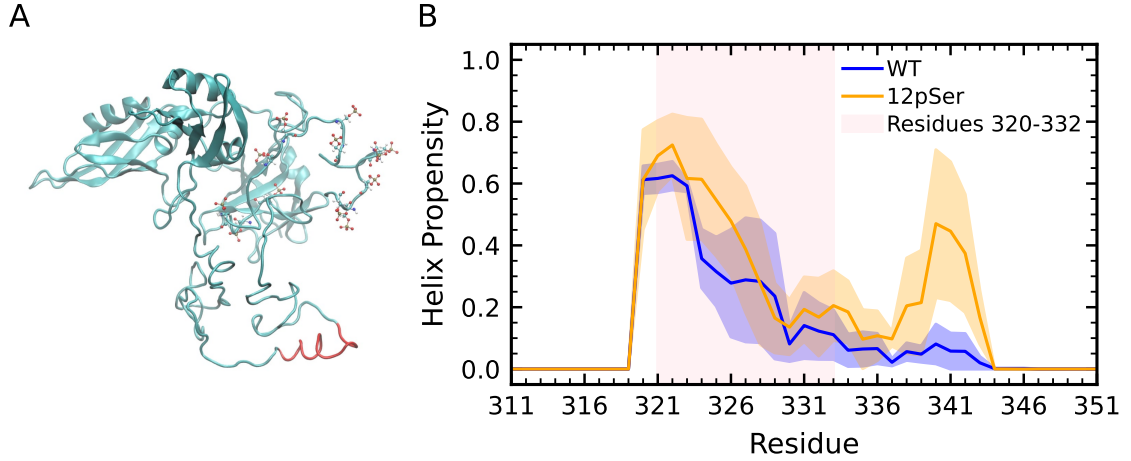


Figure 4.9: Helix stability of the 12 phosphorylated Ser residues (12pSer) in full-length TDP-43 AA simulations. (A). Representative snapshots illustrating 12pSer in all-atom simulations. All atoms of the phosphorylated Ser residues are shown, and the helical region is colored in red. (B). Residue-resolved helix propensities for the 12pSer mutant.

In condensate simulations, residue parameters were taken from the CALVADOS2 force field [84, 85]. All folded regions, including the helix, were modeled as rigid bodies, disallowing internal flexibility within regions possessing defined three-dimensional structures. Phosphorylation of twelve serine residues dissolved the condensate (Figure 4.11). At the same simulation temperature (290 K), condensate formation was observed for the WT system but not for the 12pSer mutant.

Analysis of averaged density histograms along the slab z-axis revealed clear separation into dense and dilute phases for the WT condensate, whereas the dense phase was difficult to distinguish in the 12pSer system. This result is consistent with previous findings by Silva et al. [68], who reported that phosphorylation of all twelve serine sites dissolved the LCD condensate. Notably, in the present study, phosphorylation was sufficient to dissolve condensates formed by the full-length protein. Overall, this behavior is consistent with the idea that increased negative charge favors solvated states, thereby weakening TDP-43 condensates.

4 Helical Structure within the TDP-43 C-Terminal Intrinsically Disordered Region

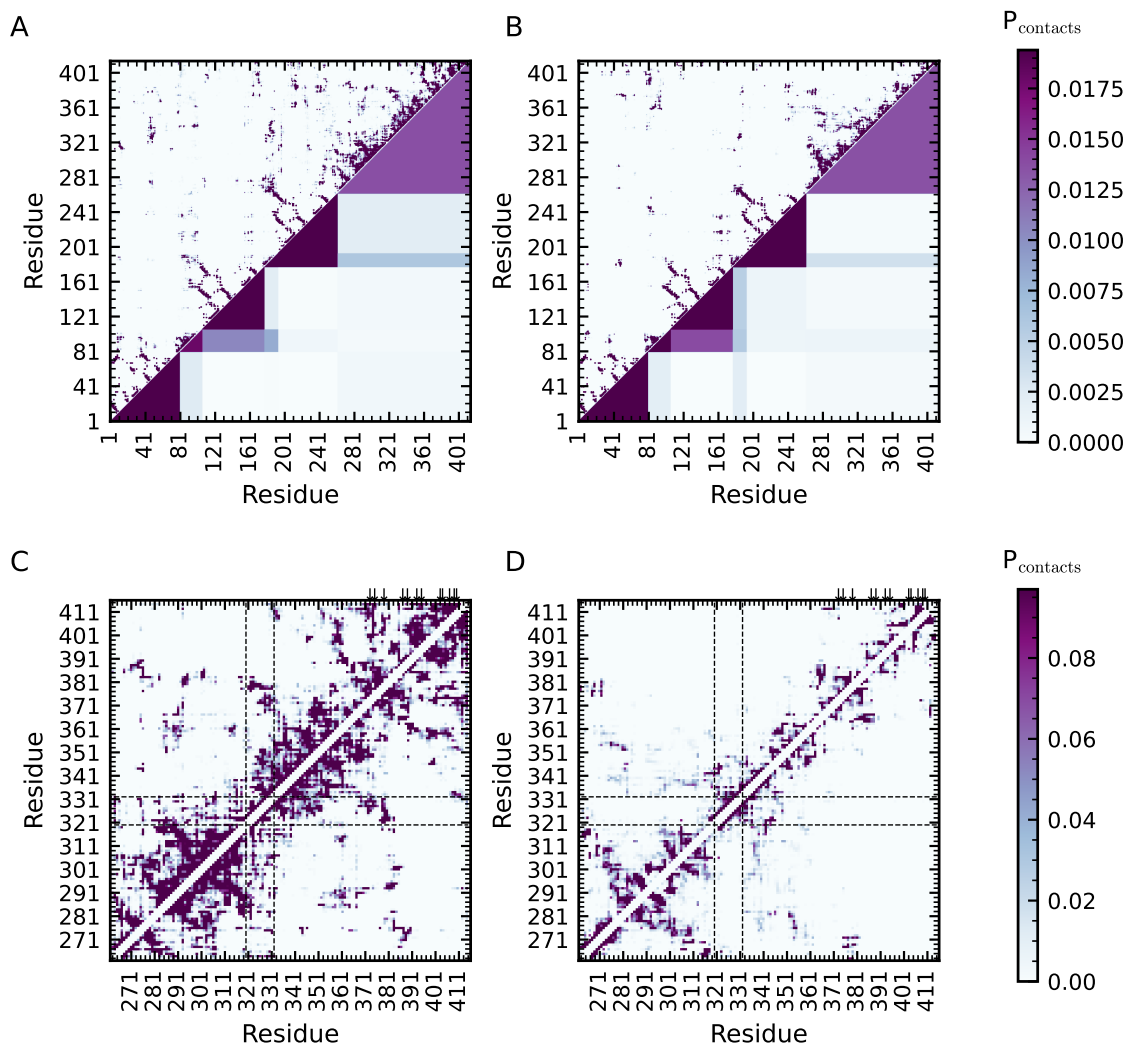


Figure 4.10: Intra-chain contact heatmaps of WT and 12pSer mutant in all-atom simulations. (A–B). Intra-chain contact heatmaps for WT and 12pSer mutant, respectively. (C–D). Zoomed-in intra-chain contact heatmaps for the LCD region of WT and 12pSer mutant, respectively.

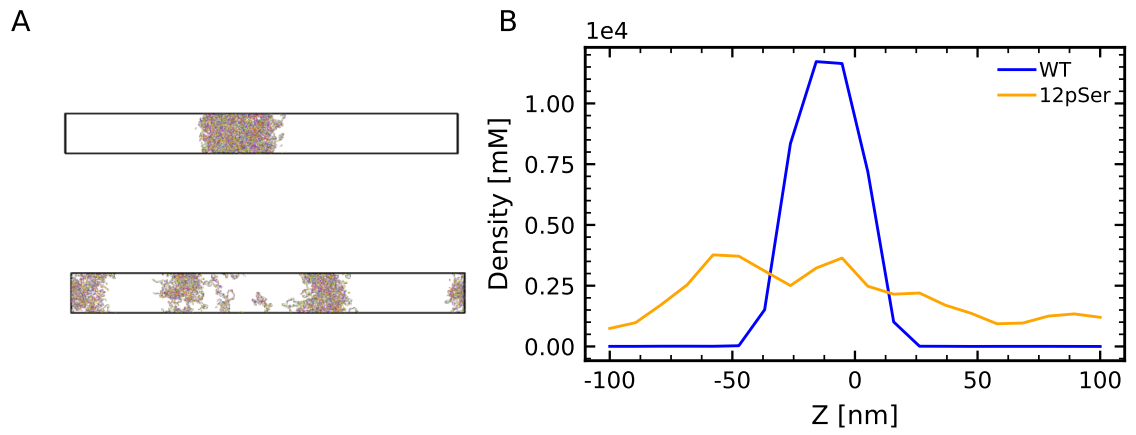


Figure 4.11: 12pSer condensates in implicit coarse-grained simulations. (A). Representative snapshots of the condensates: WT (upper) and 12pSer mutant (lower). (B). Averaged protein density along the z-axis in the slab geometry for WT and 12pSer condensates.

4.3 Discussion

In all-atom simulations, the aromatic mutant W334A modestly destabilized the helix, whereas phosphorylated serines, particularly in the 12pSer mutant, enhanced helix stability. How helix stability contributes to condensate formation, however, remains unclear. Although numerous phenomena have been observed, such as alterations in intra-chain contacts, side-chain orientations, and inter-chain interactions, it is still unknown whether the secondary structure itself or the outward projection of specific residue side chains primarily governs condensate stability.

The interplay between helix stability and residue-specific interactions is likely context-dependent. For example, a stable helix may serve as a scaffold that optimally positions key residues like Trp334 for intermolecular contacts, whereas destabilization could reduce the efficiency of these interactions. Conversely, outward-facing side chains may drive condensate cohesion even in partially disordered helices. Disentangling these contributions requires carefully designed control simulations, such as WT helices with enforced rigidity, mutants that preserve helical secondary structure but alter key side chains, or variants with partial phosphorylation.

Helix stability remains a central concern in TDP-43 condensation studies, motivating systematic simulation strategies to probe how structural and chemical modifications influence both local folding and macroscopic phase behavior. Such analyses are critical for understanding the mechanistic links between secondary structure, residue chemistry, and condensate formation.

In CG simulations, direct atomic-level resolution is often absent, making it challenging to assess secondary structure formation using traditional metrics. Therefore, specialized algorithms are required to quantify helix, β -sheet, and other structural motifs in CG models. Several approaches have been proposed, including mapping coarse-grained beads back to pseudo-atomic coordinates, analyzing backbone dihedral angle distributions, or computing local hydrogen-bonding patterns inferred from bead positions. Additionally, residue-level simulation models should be carefully validated to ensure that they can accurately capture helix and β -sheet propensities observed in all-atom simulations or experiments. Such validation is essential for interpreting CG simulations of phase separation, as the ability to maintain secondary structures can significantly influence intra-chain interactions, inter-chain contacts, and ultimately the stability and morphology of protein condensates.

5 Methods

5.1 Molecular Dynamics Simulations

5.1.1 Atomistic simulations

Atomistic molecular dynamics simulations were performed using the GROMACS 2024 [86–92] software package with the AMBER99SB*-ILDN-q [93–96] protein force field and the TIP4P-D [97] water model. All systems were neutralized to a physiological salt concentration of 150 mM NaCl. Periodic boundary conditions (PBC) were applied in all three spatial dimensions. Nonbonded interactions were computed using a 12 Å cutoff for both electrostatic and van der Waals (VDW) interactions, and long-range electrostatics were evaluated using the Particle Mesh Ewald (PME) method.

Equilibration was performed in the NPT ensemble using a 2 fs timestep, allowing the system volume to relax prior to the production phase. During equilibration, the temperature was maintained at 300 K using the Berendsen thermostat [98], and the pressure was controlled at 1 bar with the Berendsen barostat [98].

Production simulations were subsequently carried out in the NPT ensemble with a 2 fs integration timestep. Temperature and pressure were maintained at 300 K and 1 bar using the Bussi–Donadio–Parrinello velocity-rescaling thermostat [99] and the Parrinello–Rahman barostat [100], respectively.

5.1.2 Explicit coarse-grained simulations

Explicit coarse-grained simulations were performed using the Martini 3 force field, or an adapted variant with rescaled protein–water interaction strengths, within the GROMACS 2024 software framework, and subjected to PBC in all directions. All Martini-based coarse-grained systems were solvated in water and neutralized with 150 mM NaCl. Verlet cutoff scheme were applied as neighbor-searching algorithm within a cutoff radius of 11 Å for Nonbonded interactions (Both Lennard–Jones (VDW) and Coulomb interactions). Electrostatics were handled using the Reaction-Field (RF) approach with a dielectric constant of 15.

System equilibration was conducted under NPT conditions using a 10 fs timestep, with temperature maintained at 300 K via the Berendsen thermostat and pressure controlled at 1 bar with the Berendsen barostat.

5 Methods

Production simulations were performed using a 20 fs timestep. Temperature and pressure during production were maintained at 300 K and 1 bar, respectively, using the same velocity-rescaling thermostat and Parrinello–Rahman barostat as atomistic simulations.

5.1.2.1 Elastic Network model

Multi-domain proteins contain both structured regions and intrinsically disordered regions (IDRs). The structured domains were stabilized using an Elastic Network Model (ENM) [101], in which harmonic bonds are applied between coarse-grained beads to preserve the native fold.

Protein inputs (all-atom level) were coarse-grained by applying martinize2 in vermouth 0.14.1 [102] Python script. Secondary structures within the folded regions were assigned with the STRIDE algorithm [103] in the STRIDE Web interface¹. Elastic network restraints were applied using a force constant of 700 kJ mol⁻¹ nm⁻², with bead pair distance included between 0 nm and 0.85 nm. The Elastic restraints (rubber bands), dihedral and angle potential within linker regions and IDRs between structured domains were all removed as [104].

The adapted Martini 3 force field was obtained by rescaling the protein–water interaction strength [52, 104] through modification of the Lennard–Jones (LJ) well depth, ε . Specifically, the LJ ε parameter between all protein and water beads was scaled by a factor λ_{PW} . The Lennard–Jones potential is given by

$$V_{\text{LJ}}(r) = 4 \lambda_{\text{PW}} \varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right], \quad (5.1)$$

where ε is the interaction well depth, σ is the distance at which the LJ potential equals zero, and λ_{PW} is the scaling factor applied to protein–water interactions. The Python script used for protein–water interaction rescaling is available online².

5.1.2.2 Intra Gō-type potential model

Following a Gō-like approach [105], we stabilized native contacts in structured domains by creating virtual backbone beads and applying additional Lennard–Jones (LJ) potentials between relevant bead pairs. This section outlines the specific methodology for implementing LJ potentials for intra-chain contacts within folded domains.

Protein input structures were coarse-grained using martinize2 as described in the Elastic Network Model section. For simulation systems containing multiple chains, each

¹<https://webclu.bio.wzw.tum.de/cgi-bin/stride/stridecgi.py>

²https://github.com/KULL-Centre/papers/tree/main/2021/Martini-Thomasen-et-al/PW_rescaling_example

chain requires a unique identifier for distinguishing. To define intra-chain native contacts, single-chain contact maps were obtained from a web server³. The Python script `create_goVirt.py`⁴ was used to implement Gō-like bonds. The strength of the Go model structural bias was set to the default value of 9.414 kJ/mol. Additional statements were added to the Martini 3 force field to incorporate the created Gō-like bonds. Similar to the Elastic Network Model approach, Gō-bonds, dihedral, and angle potentials within linker regions and intrinsically disordered regions (IDRs) between structured domains were removed. For multi-chain systems (e.g., 60 chains), the .tpr file generated in GROMACS for equilibration and production requires 128 GB of memory allocation.

Application of the adapted Martini 3 force field follows the same procedure as for the Elastic Network Model.

The Gō-like model was implemented using the Vermouth 0.15.0 library, which provides a more advanced and streamlined setup procedure⁵.

5.1.2.3 Intra and inter Gō-type potential model

For multi-chain systems, inter-chain contacts can be stabilized with additional Lennard-Jones (LJ) potentials alongside the intra-Gō potential. We adapted the approach by Korshunova and Bruiniks [106] for multi-domain proteins involving intrinsically disordered proteins (IDPs), as implemented by L. Baltz⁶.

Similar to both the Elastic Network and intra Gō-like models, atom-level protein structures were first coarse-grained using martinize2. The additional inter- and intra-chain LJ potentials were implemented using the adapted Python script `sorted_goVirt.py`⁷. This script can exclude LJ potentials for linker regions and IDRs upon request. Additionally, dihedral and angle potentials within these flexible regions between structured domains have to manually remove. This approach can also be utilized for systems requiring only the Intra Gō-type potential model, which offers the advantage of reduced memory allocation compared to multi-chain implementations.

5.1.2.4 Multi-chain simulation setup

Multiple full-length TDP-43 chains were simulated in both cubic and slab geometries. Protein chains were initially placed at random positions within a cubic simulation box, which was subsequently solvated with water and neutralizing ions. Systems in the cubic geometry were simulated using the adapted Martini 3 force field with a rescaling factor of

³<http://pomalab.ippt.pan.pl/GoContactMap/>

⁴https://github.com/Martini-Force-Field-Initiative/GoMartini/blob/main/GoMartini/create_goVirt.py

⁵<https://cgmartini.nl/docs/tutorials/Martini3/ProteinsI/Tut1.html>

⁶<https://github.com/luhtzia/multichain-martini-go>

⁷<https://github.com/luhtzia/multichain-martini-go/tree/master>

5 Methods

$\lambda_{1.06}$ during energy minimization, equilibration, and production runs. In the slab geometry, multiple full-length TDP-43, its variants (12D, W3A, F6W3A) and LCD fragments were simulated in the slab geometry with different starting configurations. The production run duration and initial configuration in the slab geometry for WT protein system are provided in Table S1.

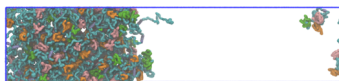
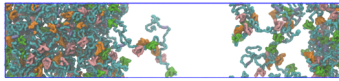
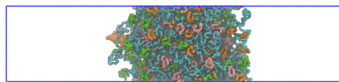
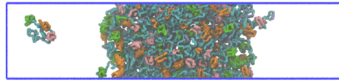
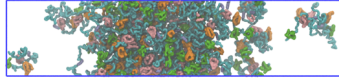
Model Name	Initial Configuration		Simulation Time
Martini3 Elastic	Replica 1		pw1.05: 25.24 μ s pw1.06: 26.94 μ s pw1.07: 48.30 μ s
	Replica 2		pw1.05: 19.55 μ s pw1.06: 27.15 μ s pw1.07: 27.81 μ s
	Replica 3		pw1.05: 31.80 μ s pw1.06: 31.28 μ s pw1.07: 27.30 μ s
	Replica 1		pw1.05: 24.03 μ s pw1.06: 22.46 μ s pw1.07: 21.41 μ s
	Replica 2		pw1.05: 21.97 μ s pw1.06: 24.46 μ s pw1.07: 22.45 μ s
	Replica 3		pw1.05: 22.63 μ s pw1.06: 19.96 μ s pw1.07: 20.17 μ s

Table 5.1: Initial configurations and simulation time lengths for each of the three replicas of the Martini3 Elastic Network and $G\bar{o}_{\text{intra}}$ models.

5.1.3 Implicit coarse-grained simulations

Following the simulation protocol described at ⁸ [107], simulations were performed using HOOMD-blue [108]. The system setup, force field parameters, and treatment of phosphorylated residues were implemented according to the repository guidelines, ensuring that the effects of phosphorylation on protein structure and condensate formation could be accurately modeled.

⁸https://github.com/ezipppo/hoomd3_phosphorylation

5.2 Molecular Dynamics Simulation Analysis

The quantitative analysis of LLPS was performed on simulations employing a slab geometry.

5.2.1 Density and phase diagram calculations

The system was aligned by translating the protein chains such that their collective center of mass (COM) was centered at $z = 0$. The spatial distribution of protein along the z -axis was computed and stored as a histogram for each trajectory frame. The time-averaged density profile was then symmetrized by averaging the density from both slab halves ($z < 0$ and $z > 0$), producing a single symmetrized profile for $z \geq 0$. This symmetrized profile was fit to a hyperbolic tangent function to characterize the interface:

$$\rho(z) = \frac{\rho_{\text{dil}} + \rho_{\text{den}}}{2} + \frac{\rho_{\text{den}} - \rho_{\text{dil}}}{2} \cdot \tanh\left(\frac{z - z_0}{w}\right) \quad (5.2)$$

where z_0 defines the midpoint of the interface and w parameterizes the interface width. The coexisting phase concentrations, ρ_{dil} and ρ_{den} , were determined directly from the averaged histogram by calculating the mean density in regions defined by the interface cutoffs. To estimate statistical uncertainties, the trajectory was divided into consecutive blocks. The average concentration within each block was computed, and the standard error of the mean (SEM) was calculated from these block-averaged values.

Phase diagrams were constructed by plotting the coexisting phase concentrations, ρ_{dil} and ρ_{den} , against the interaction rescaling factor λ . The critical point (λ_c) was obtained by

$$\Delta\rho = \rho_{\text{den}} - \rho_{\text{dil}} = B\left(1 - \frac{\lambda}{\lambda_c}\right)^\beta \quad (5.3)$$

B is a fitting parameter representing the critical amplitude, and assume the systems belong to the 3d Ising universality class, with a critical exponent $\beta = 0.325$.

The critical density ρ_c was estimated using the law of the rectilinear diameter:

$$\rho_d = \frac{\rho_{\text{den}} + \rho_{\text{dil}}}{2} = \rho_c + C(\lambda_c - \lambda) \quad (5.4)$$

is applied with a positive fitting constant C [109, 110].

5.2.2 Interfacial tension calculations

Estimation of condensate surface tension is based on Capillary wave theory with block analysis method in the slab geometry [111, 112] to describe the squared condensate interfacial thickness grows as logarithm of the block size.

5 Methods

$$W^2(B) = W_0^2 + \left(\frac{K_B \cdot T}{2\pi \cdot \gamma} \right) \cdot \left(\ln \frac{B}{L} \right) \quad (5.5)$$

the average of the interfacial widths measured from all blocks of a given size B , W_0 is an intrinsic width, $K_B T$ is the thermal energy (product of the Boltzmann constant K_B and temperature T), γ gives the surface tension, L is the system size of x or y dimension, B is the block size of x or y dimension.

In the block analysis method with slab geometry, the interface regions present along the x and y axis and are divided into rectangular blocks with dimensions as $Lx/nB \times Ly/nB \times Lz$, nB is grid number as an integer, ranging from 4 to 10. Within the segments, droplet interfacial width W values are obtained from a tanh fit to concentration profiles.

$$\rho(z) = \frac{\rho_{\text{den}} + \rho_{\text{dil}}}{2} + \frac{\rho_{\text{den}} - \rho_{\text{dil}}}{2} \cdot \tanh \left(\frac{Z - Z_0}{W_0} \right) \quad (5.6)$$

z_0 is the local interface position within a specific block, W_0 represents the intrinsic width.

5.2.3 Viscosity estimation

The viscosity within the condensate followed the approach by Benayad et al[113]. The variation of the end to end distances r and of the end to end distance from this one can arrive an estimate of (conformational) diffusion coefficient [113, 114] of the end-to-end dynamics of the TDP-43 chains

$$D = \frac{\text{var}(r)}{\tau} \quad (5.7)$$

By comparing the D from the condensates and dilute solution D_0 and assuming that their ratio scales with η we can arrive at an estimate of the shear viscosity

$$\eta = \frac{D_0}{D} \cdot \eta_w \quad (5.8)$$

The viscosity of Martini3 water $\eta_w = 0.66 \text{ mPa} \cdot \text{s}$ at 300 K was taken from Prass et al [115].

Bibliography

- [1] Yiming Tang, Santu Bera, Yifei Yao, Jiyuan Zeng, Zenghui Lao, Xuewei Dong, Ehud Gazit, and Guanghong Wei. Prediction and characterization of liquid-liquid phase separation of minimalistic peptides. *Cell Reports Physical Science*, 2(9), 2021.
- [2] Youhna M Ayala, Paola Zago, Andrea D’Ambrogio, Ya-Fei Xu, Leonard Petrucelli, Emanuele Buratti, and Francisco E Baralle. Structural determinants of the cellular localization and shuttling of tdp-43. *Journal of cell science*, 121(22):3778–3785, 2008.
- [3] Myungjin Jo, Shinrye Lee, Yu-Mi Jeon, Seyeon Kim, Younghwi Kwon, and Hyung-Jun Kim. The role of tdp-43 propagation in neurodegenerative diseases: integrating insights from clinical and experimental studies. *Experimental & molecular medicine*, 52(10):1652–1662, 2020.
- [4] Jannat Ahmed, Q Jane Wang, Oluwaseyi Balogun, Ning Ren, Roger England, and Frances Lockwood. Molecular dynamics modeling of thermal conductivity of several hydrocarbon base oils. *Tribology Letters*, 71(2):70, 2023.
- [5] T. Duve, L. Wang, L. Borges-Araújo, S. J. Marrink, P. C. T. Souza, and S. Thallmair. Martini 3 protein models — a practical introduction to different structure bias models and their comparison. *bioRxiv*, 2025. doi: 10.1101/2025.03.17.643608.
- [6] Gregory L Dignon, Wenwei Zheng, Young C Kim, Robert B Best, and Jeetain Mittal. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS computational biology*, 14(1):e1005941, 2018.
- [7] Edmund B Wilson. The structure of protoplasm. *Science*, 10(237):33–45, 1899.
- [8] Clifford P Brangwynne, Christian R Eckmann, David S Courson, Agata Rybarska, Carsten Hoege, Jöbin Gharakhani, Frank Jülicher, and Anthony A Hyman. Germline p granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, 324(5935):1729–1732, 2009.
- [9] Chen Wang, Yongjia Duan, Gang Duan, Qiangqiang Wang, Kai Zhang, Xue Deng, Beituo Qian, Jinge Gu, Zhiwei Ma, Shuang Zhang, et al. Stress induces dynamic, cytotoxicity-antagonizing tdp-43 nuclear bodies via paraspeckle lncrna neat1-mediated liquid-liquid phase separation. *Molecular cell*, 79(3):443–458, 2020.
- [10] Avinash Patel, Hyun O Lee, Louise Jawerth, Shovamayee Maharana, Marcus Jahnel, Marco Y Hein, Stoyno Stoynov, Julia Mahamid, Shambaditya Saha, Titus M

Bibliography

- Franzmann, et al. A liquid-to-solid phase transition of the als protein fus accelerated by disease mutation. *Cell*, 162(5):1066–1077, 2015.
- [11] SH Ou, Foon Wu, David Harrich, Leon F García-Martínez, and Richard B Gaynor. Cloning and characterization of a novel cellular protein, tdp-43, that binds to human immunodeficiency virus type 1 tar dna sequence motifs. *Journal of virology*, 69(6): 3584–3596, 1995.
- [12] Matthew J Winton, Lionel M Igaz, Margaret M Wong, Linda K Kwong, John Q Trojanowski, and Virginia M-Y Lee. Disturbance of nuclear and cytoplasmic tar dna-binding protein (tdp-43) induces disease-like redistribution, sequestration, and aggregate formation. *Journal of Biological Chemistry*, 283(19):13302–13309, 2008.
- [13] EM Tank, C Figueroa-Romero, LM Hinder, K Bedi, HC Archbold, X Li, K Weskamp, N Safren, X Paez-Colasante, C Pacut, et al. Abnormal rna stability in amyotrophic lateral sclerosis. *nat commun* 9: 2845, 2018.
- [14] Emanuele Buratti, Thilo Dörk, Elisabetta Zuccato, Franco Pagani, Maurizio Romano, and Francisco E Baralle. Nuclear factor tdp-43 and sr proteins promote in vitro and in vivo cftr exon 9 skipping. *The EMBO journal*, 2001.
- [15] Shovamayee Maharana, Jie Wang, Dimitrios K Papadopoulos, Doris Richter, Andrey Pozniakovsky, Ina Poser, Marc Bickle, Sandra Rizk, Jordina Guillén-Boixet, Titus M Franzmann, et al. Rna buffers the phase separation behavior of prion-like rna binding proteins. *Science*, 360(6391):918–921, 2018.
- [16] Claudia Colombrita, Eleonora Zennaro, Claudia Fallini, Markus Weber, Andreas Sommacal, Emanuele Buratti, Vincenzo Silani, and Antonia Ratti. Tdp-43 is recruited to stress granules in conditions of oxidative insult. *Journal of neurochemistry*, 111(4):1051–1061, 2009.
- [17] Paul Anderson and Nancy Kedersha. visibly stressed: the role of eif2, tia-1, and stress granules in protein translation. *Cell stress & chaperones*, 7(2):213, 2002.
- [18] Paul Anderson and Nancy Kedersha. Stress granules: the tao of rna triage. *Trends in biochemical sciences*, 33(3):141–150, 2008.
- [19] Karli K McDonald, Anaïs Aulas, Laurie Destroismaisons, Sarah Pickles, Evghenia Beleac, William Camu, Guy A Rouleau, and Christine Vande Velde. Tar dna-binding protein 43 (tdp-43) regulates stress granule dynamics via differential regulation of g3bp and tia-1. *Human molecular genetics*, 20(7):1400–1410, 2011.
- [20] Anaïs Aulas, Stéphanie Stabile, and Christine Vande Velde. Endogenous tdp-43, but not fus, contributes to stress granule assembly via g3bp. *Molecular neurodegeneration*, 7(1):1, 2012.
- [21] Yousra Khalfallah, Rachel Kuta, Camille Grasmuck, Alexandre Prat, Heather D Durham, and Christine Vande Velde. Tdp-43 regulation of stress granule dynamics in neurodegenerative disease-relevant cell types. *Scientific reports*, 8(1):7551, 2018.

- [22] J Paul Taylor, John Hardy, and Kenneth H Fischbeck. Toxic proteins in neurodegenerative disease. *science*, 296(5575):1991–1995, 2002.
- [23] Caroline Vance, Boris Rogelj, Tibor Hortobágyi, Kurt J De Vos, Agnes Lumi Nishimura, Jemeen Sreedharan, Xun Hu, Bradley Smith, Deborah Ruddy, Paul Wright, et al. Mutations in *fus*, an rna processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science*, 323(5918):1208–1211, 2009.
- [24] TJ Kwiatkowski Jr, DA Bosco, AL Leclerc, E Tamrazian, CR Vanderburg, C Russ, A Davis, J Gilchrist, EJ Kasarskis, T Munsat, et al. Mutations in the *fus/tls* gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science*, 323(5918):1205–1208, 2009.
- [25] Manuela Neumann, Deepak M Sampathu, Linda K Kwong, Adam C Truax, Matthew C Micsenyi, Thomas T Chou, Jennifer Bruce, Theresa Schuck, Murray Grossman, Christopher M Clark, et al. Ubiquitinated *tdp-43* in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science*, 314(5796):130–133, 2006.
- [26] Jemeen Sreedharan, Ian P Blair, Vineeta B Tripathi, Xun Hu, Caroline Vance, Boris Rogelj, Steven Ackerley, Jennifer C Durnall, Kelly L Williams, Emanuele Buratti, et al. *Tdp-43* mutations in familial and sporadic amyotrophic lateral sclerosis. *Science*, 319(5870):1668–1672, 2008.
- [27] Brian S Johnson, David Snead, Jonathan J Lee, J Michael McCaffery, James Shorter, and Aaron D Gitler. *Tdp-43* is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *Journal of Biological Chemistry*, 284(30):20329–20339, 2009.
- [28] Takashi Nonaka, Fuyuki Kametani, Tetsuaki Arai, Haruhiko Akiyama, and Masato Hasegawa. Truncation and pathogenic mutations facilitate the formation of intracellular aggregates of *tdp-43*. *Human molecular genetics*, 18(18):3353–3364, 2009.
- [29] Weirui Guo, Yanbo Chen, Xiaohong Zhou, Amar Kar, Payal Ray, Xiaoping Chen, Elizabeth J Rao, Mengxue Yang, Haihong Ye, Li Zhu, et al. An *als*-associated mutation affecting *tdp-43* enhances protein aggregation, fibril formation and neurotoxicity. *Nature structural & molecular biology*, 18(7):822–830, 2011.
- [30] Chia-Sui Sun, Cindy Yu-Hsiang Wang, Bryan Po-Wen Chen, Ruei-Yu He, Gerard Chun-Hao Liu, Chih-Hsien Wang, Wenlung Chen, Yijuang Chern, and Joseph Jen-Tse Huang. The influence of pathological mutations and proline substitutions in *tdp-43* glycine-rich peptides on its amyloid properties and cellular toxicity. *PLoS One*, 9(8):e103644, 2014.
- [31] Jacob R Mann, Amanda M Gleixner, Jocelyn C Mauna, Edward Gomes, Michael R DeChellis-Marks, Patrick G Needham, Katie E Copley, Bryan Hurtle, Bede Portz, Noah J Pyles, et al. Rna binding antagonizes neurotoxic phase transitions of *tdp-43*. *Neuron*, 102(2):321–338, 2019.

Bibliography

- [32] W Michael Babinchak, Raza Haider, Benjamin K Dumm, Prottusha Sarkar, Krystyna Surewicz, Jin-Kyu Choi, and Witold K Surewicz. The role of liquid–liquid phase separation in aggregation of the tdp-43 low-complexity domain. *Journal of Biological Chemistry*, 294(16):6306–6317, 2019.
- [33] Hiroshi Tsuji, Takashi Nonaka, Makiko Yamashita, Masami Masuda-Suzukake, Fuyuki Kametani, Haruhiko Akiyama, David MA Mann, Akira Tamaoka, and Masato Hasegawa. Epitope mapping of antibodies against tdp-43 and detection of protease-resistant fragments of pathological tdp-43 in amyotrophic lateral sclerosis and frontotemporal lobar degeneration. *Biochemical and biophysical research communications*, 417(1):116–121, 2012.
- [34] Shangxi Xiao, Teresa Sanelli, Helen Chiang, Yulong Sun, Avijit Chakrabartty, Julia Keith, Ekaterina Rogaeva, Lorne Zinman, and Janice Robertson. Low molecular weight species of tdp-43 generated by abnormal splicing form inclusions in amyotrophic lateral sclerosis and result in motor neuron death. *Acta neuropathologica*, 130(1):49–61, 2015.
- [35] Yong-Jie Zhang, Ya-Fei Xu, Casey Cook, Tania F Gendron, Paul Roettges, Christopher D Link, Wen-Lang Lin, Jimei Tong, Monica Castanedes-Casey, Peter Ash, et al. Aberrant cleavage of tdp-43 enhances aggregation and cellular toxicity. *Proceedings of the National Academy of Sciences*, 106(18):7607–7612, 2009.
- [36] Masato Hasegawa, Tetsuaki Arai, Takashi Nonaka, Fuyuki Kametani, Mari Yoshida, Yoshio Hashizume, Thomas G Beach, Emanuele Buratti, Francisco Baralle, Mitsuya Morita, et al. Phosphorylated tdp-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 64(1):60–70, 2008.
- [37] Takashi Nonaka, Genjiro Suzuki, Yoshinori Tanaka, Fuyuki Kametani, Shinobu Hirai, Haruo Okado, Tomoyuki Miyashita, Minoru Saitoe, Haruhiko Akiyama, Hisao Masai, et al. Phosphorylation of tar dna-binding protein of 43 kda (tdp-43) by truncated casein kinase 1 δ triggers mislocalization and accumulation of tdp-43. *Journal of Biological Chemistry*, 291(11):5473–5483, 2016.
- [38] Wm G Hoover. *Molecular dynamics*. Springer, 1986.
- [39] Joshua E Barnes. Evolution of compact groups and the formation of elliptical galaxies. *Nature*, 338(6211):123–126, 1989.
- [40] Luca Monticelli, Senthil K Kandasamy, Xavier Periole, Ronald G Larson, D Peter Tieleman, and Siewert-Jan Marrink. The martini coarse-grained force field: extension to proteins. *Journal of chemical theory and computation*, 4(5):819–834, 2008.
- [41] Djurre H De Jong, Gurpreet Singh, WF Drew Bennett, Clement Arnarez, Tsjerk A Wassenaar, Lars V Schafer, Xavier Periole, D Peter Tieleman, and Siewert J Marrink. Improved parameters for the martini coarse-grained protein force field. *Journal of chemical theory and computation*, 9(1):687–697, 2013.

- [42] Riccardo Alessandri, Fabian Grünewald, and Siewert J Marrink. The martini model in materials science. *Advanced Materials*, 33(24):2008635, 2021.
- [43] Cezary Czaplewski, Agnieszka Karczyńska, Adam K Sieradzan, and Adam Liwo. Unres server for physics-based coarse-grained simulations and prediction of protein structure, dynamics and thermodynamics. *Nucleic Acids Research*, 46(W1):W304–W309, 2018.
- [44] Siewert J Marrink, Alex H De Vries, and Alan E Mark. Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B*, 108(2):750–760, 2004.
- [45] Paulo CT Souza, Riccardo Alessandri, Jonathan Barnoud, Sebastian Thallmair, Ignacio Faustino, Fabian Grünewald, Ilias Patmanidis, Haleh Abdizadeh, Bart MH Bruininks, Tsjerk A Wassenaar, et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature methods*, 18(4):382–388, 2021.
- [46] Ligu Wang, Christopher Brasnett, Luís Borges-Araújo, Paulo CT Souza, and Siewert J Marrink. Martini3-idp: improved martini 3 force field for disordered proteins. *Nature Communications*, 16(1):1–14, 2025.
- [47] Yumeng Zhang, Xiaorong Liu, and Jianhan Chen. Toward accurate coarse-grained simulations of disordered proteins and their dynamic interactions. *Journal of chemical information and modeling*, 62(18):4523–4536, 2022.
- [48] Yumeng Zhang, Shanlong Li, Xiping Gong, and Jianhan Chen. Toward accurate simulation of coupling between protein secondary structure and phase separation. *Journal of the American Chemical Society*, 146(1):342–357, 2023.
- [49] Giulio Tesei, Anna Ida Trolle, Nicolas Jonsson, Johannes Betz, Frederik E Knudsen, Francesco Pesce, Kristoffer E Johansson, and Kresten Lindorff-Larsen. Conformational ensembles of the human intrinsically disordered proteome. *Nature*, 626(8000):897–904, 2024.
- [50] Fan Cao, Sören von Bülow, Giulio Tesei, and Kresten Lindorff-Larsen. A coarse-grained model for disordered and multi-domain proteins. *Protein Science*, 33(11):e5172, 2024.
- [51] Alexander E Conicella, Gregory L Dignon, Gül H Zerze, Hermann Broder Schmidt, Alexandra M D’Ordine, Young C Kim, Rajat Rohatgi, Yuna M Ayala, Jeetain Mittal, and Nicolas L Fawzi. Tdp-43 α -helical structure tunes liquid–liquid phase separation and function. *Proceedings of the National Academy of Sciences*, 117(11):5883–5894, 2020.
- [52] F Emil Thomasen, Francesco Pesce, Mette Ahrensback Roesgaard, Giulio Tesei, and Kresten Lindorff-Larsen. Improving martini 3 for disordered and multidomain proteins. *Journal of chemical theory and computation*, 18(4):2033–2041, 2022.
- [53] Liberty François-Moutal, Samantha Perez-Miller, David D Scott, Victor G Miranda,

Bibliography

- Niloufar Mollasalehi, and May Khanna. Structural insights into tdp-43 and effects of post-translational modifications. *Frontiers in molecular neuroscience*, 12:301, 2019.
- [54] Miguel Mompeán, Valentina Romano, David Pantoja-Uceda, Cristiana Stuani, Francisco E Baralle, Emanuele Buratti, and Douglas V Laurents. The tdp-43 n-terminal domain structure at high resolution. *The FEBS journal*, 283(7):1242–1260, 2016.
- [55] Youhna M Ayala, Sergio Pantano, Andrea D’Ambrogio, Emanuele Buratti, Antonia Brindisi, Caterina Marchetti, Maurizio Romano, and Francisco E Baralle. Human, drosophila, and c. elegans tdp43: nucleic acid binding properties and splicing regulatory function. *Journal of molecular biology*, 348(3):575–588, 2005.
- [56] Peter J Lukavsky, Dalia Daujotyte, James R Tollervey, Jernej Ule, Cristiana Stuani, Emanuele Buratti, Francisco E Baralle, Fred F Damberger, and Frédéric HT Allain. Molecular basis of ug-rich rna recognition by the human splicing factor tdp-43. *Nature structural & molecular biology*, 20(12):1443–1449, 2013.
- [57] Haina Qin, Liang-Zhong Lim, Yuanyuan Wei, and Jianxing Song. Tdp-43 n terminus encodes a novel ubiquitin-like fold and its unfolded form in equilibrium that can be shifted by binding to ssdna. *Proceedings of the National Academy of Sciences*, 111(52):18619–18624, 2014.
- [58] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Židek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.
- [59] Lobanov MIu, NS Bogatyreva, and OV Galzitskaia. Radius of gyration is indicator of compactness of protein structure. *Molekuliarnaia biologii*, 42(4):701–706, 2008.
- [60] James A Austin, Gareth SA Wright, Seiji Watanabe, J Günter Grossmann, Svetlana V Antonyuk, Koji Yamanaka, and S Samar Hasnain. Disease causing mutants of tdp-43 nucleic acid binding domains are resistant to aggregation and have increased stability and half-life. *Proceedings of the National Academy of Sciences*, 111(11):4309–4314, 2014.
- [61] Gareth SA Wright, Tatiana F Watanabe, Kangsa Ampornnanai, Steven S Plotkin, Neil R Cashman, Svetlana V Antonyuk, and S Samar Hasnain. Purification and structural characterization of aggregation-prone human tdp-43 involved in neurodegenerative diseases. *Isience*, 23(6), 2020.
- [62] G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comput. Biol.*, 14:e1005941, 2018.
- [63] Martina Hallegger, Anob M Chakrabarti, Flora CY Lee, Bo Lim Lee, Aram G Amalietti, Hana M Odeh, Katie E Copley, Jack D Rubien, Bede Portz, Klara Kuret, et al. Tdp-43 condensation properties specify its rna-binding and regulatory repertoire. *Cell*, 184(18):4680–4696, 2021.

- [64] Justin P Gallivan and Dennis A Dougherty. Cation- π interactions in structural biology. *Proceedings of the National Academy of Sciences*, 96(17):9459–9464, 1999.
- [65] Rivka Calinsky and Yaakov Levy. Aromatic residues in proteins: Re-evaluating the geometry and energetics of π - π , cation- π , and ch- π interactions. *The Journal of Physical Chemistry B*, 128(36):8687–8700, 2024.
- [66] Hermann Broder Schmidt, Ariana Barreau, and Rajat Rohatgi. Phase separation-deficient tdp43 remains functional in splicing. *Nature Communications*, 10(1):4890, 2019. doi: 10.1038/s41467-019-12740-2. URL <https://doi.org/10.1038/s41467-019-12740-2>.
- [67] Paul J Flory. *Principles of polymer chemistry*. Cornell university press, 1953.
- [68] Lara A Gruijs da Silva, Francesca Simonetti, Saskia Hutten, Henrick Riemenschneider, Erin L Sternburg, Lisa M Pietrek, Jakob Gebel, Volker Dötsch, Dieter Edbauer, Gerhard Hummer, et al. Disease-linked tdp-43 hyperphosphorylation suppresses tdp-43 condensation and aggregation. *The EMBO journal*, 41(8):e108443, 2022.
- [69] Azamat Rizuan, Jayakrishna Shenoy, Priyesh Mohanty, Patricia MS Dos Passos, José F Mercado Ortiz, Leanna Bai, Renjith Viswanathan, Szu-Huan Wang, Victoria Johnson, Lohany D Mamede, et al. Structural details of helix-mediated tdp-43 c-terminal domain multimerization. *bioRxiv*, 2024.
- [70] Yuki Inukai, Takashi Nonaka, Tetsuaki Arai, Mari Yoshida, Yoshio Hashizume, Thomas G Beach, Emanuele Buratti, Francisco E Baralle, Haruhiko Akiyama, Shin-ichi Hisanaga, et al. Abnormal phosphorylation of ser409/410 of tdp-43 in ftld-u and als. *FEBS letters*, 582(19):2899–2904, 2008.
- [71] Manuela Neumann, Linda K Kwong, Edward B Lee, Elisabeth Kremmer, Andrew Flatley, Yan Xu, Mark S Forman, Dirk Troost, Hans A Kretzschmar, John Q Trojanowski, et al. Phosphorylation of s409/410 of tdp-43 is a consistent feature in all sporadic and familial forms of tdp-43 proteinopathies. *Acta neuropathologica*, 117(2):137–149, 2009.
- [72] Alexander Jussupow and Ville R. I. Kaila. Effective molecular dynamics from neural network-based structure prediction models. *Journal of Chemical Theory and Computation*, 19(7):1965–1975, 2023. doi: 10.1021/acs.jctc.2c01027. URL <https://doi.org/10.1021/acs.jctc.2c01027>. PMID: 36961997.
- [73] Lukas S Stelzl, Despoina AI Mavridou, Emmanuel Saridakis, Diego Gonzalez, Andrew J Baldwin, Stuart J Ferguson, Mark SP Sansom, and Christina Redfield. Local frustration determines loop opening during the catalytic cycle of an oxidoreductase. *eLife*, 9:e54661, jun 2020. ISSN 2050-084X. doi: 10.7554/eLife.54661. URL <https://doi.org/10.7554/eLife.54661>.
- [74] Adolfo B Poma, Marek Cieplak, and Panagiotis E Theodorakis. Combining the martini and structure-based coarse-grained approaches for the molecular dynamics

Bibliography

- studies of conformational transitions in proteins. *Journal of Chemical Theory and Computation*, 13(3):1366–1374, 2017.
- [75] Tariq Afroz, Eva-Maria Hock, Patrick Ernst, Chiara Foglieni, Melanie Jambeau, Larissa AB Gilhespy, Florent Laferriere, Zuzanna Maniecka, Andreas Plückthun, Peer Mittl, et al. Functional and dynamic polymerization of the als-linked protein tdp-43 antagonizes its pathologic aggregation. *Nature communications*, 8(1):45, 2017.
- [76] Wan-Ping Huang, Brittany CS Ellis, Rachel E Hodgson, Anna Sanchez Avila, Vedanth Kumar, Jessica Rayment, Tobias Moll, and Tatyana A Shelkvnikova. Stress-induced tdp-43 nuclear condensation causes splicing loss of function and stmn2 depletion. *Cell Reports*, 43(7), 2024.
- [77] Robert B Best, Gerhard Hummer, and William A Eaton. Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences*, 110(44):17874–17879, 2013.
- [78] Andrew GT Pyo, Yaojun Zhang, and Ned S Wingreen. Proximity to criticality predicts surface properties of biomolecular condensates. *Proceedings of the National Academy of Sciences*, 120(23):e2220014120, 2023.
- [79] Eugene Helfand and Yukiko Tagami. Theory of the interface between immiscible polymers. ii. *The Journal of chemical physics*, 56(7):3592–3601, 1972.
- [80] Alexander E Conicella, Gül H Zerze, Jeetain Mittal, and Nicolas L Fawzi. Als mutations disrupt phase separation mediated by α -helical structure in the tdp-43 low-complexity c-terminal domain. *Structure*, 24(9):1537–1549, 2016.
- [81] Diana Arseni, Masato Hasegawa, Alexey G Murzin, Fuyuki Kametani, Makoto Arai, Mari Yoshida, and Benjamin Ryskeldi-Falcon. Structure of pathological tdp-43 filaments from als with ftld. *Nature*, 601(7891):139–143, 2022.
- [82] Nadine Homeyer, Anselm HC Horn, Harald Lanig, and Heinrich Sticht. Amber force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *Journal of molecular modeling*, 12(3):281–289, 2006.
- [83] T Steinbrecher, J Latzer, and DA Case. Revised amber parameters for bioorganic phosphates. *Journal of chemical theory and computation*, 8(11):4405–4412, 2012.
- [84] Giulio Tesei, Thea K Schulze, Ramon Crehuet, and Kresten Lindorff-Larsen. Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proceedings of the National Academy of Sciences*, 118(44):e2111696118, 2021.
- [85] Giulio Tesei and Kresten Lindorff-Larsen. Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range. *Open Research Europe*, 2:94, 2023.

- [86] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015. doi: 10.1016/j.softx.2015.06.001.
- [87] S. Páll, M. J. Abraham, C. Kutzner, B. Hess, and E. Lindahl. Tackling exascale software challenges in molecular dynamics simulations with gromacs. In S. Markidis and E. Laure, editors, *Solving Software Challenges for Exascale*, volume 8759 of *Lecture Notes in Computer Science*, pages 3–27. Springer, 2015. doi: 10.1007/978-3-319-15976-8_1.
- [88] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, 2013. doi: 10.1093/bioinformatics/btt055.
- [89] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008. doi: 10.1021/ct700301q.
- [90] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. Gromacs: Fast, flexible and free. *Journal of Computational Chemistry*, 26(16):1701–1719, 2005. doi: 10.1002/jcc.20291.
- [91] E. Lindahl, B. Hess, and D. van der Spoel. Gromacs 3.0: A package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, 7(8):306–317, 2001. doi: 10.1007/s008940100045.
- [92] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56, 1995. doi: 10.1016/0010-4655(95)00042-E.
- [93] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.
- [94] Robert B Best and Gerhard Hummer. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *The journal of physical chemistry B*, 113(26):9004–9015, 2009.
- [95] K Lindorff-Larsen, S Piana, K Palmo, P Maragakis, JL Klepeis, and RO Dror. De shaw improved side-chain torsion potentials for the amber ff99sb protein force field., 2010, 78. DOI: <https://doi.org/10.1002/prot.22711>:1950–1958.
- [96] Robert B Best, David de Sancho, and Jeetain Mittal. Residue-specific α -helix propensities from molecular simulation. *Biophysical Journal*, 102(6):1462–1467, 2012.

Bibliography

- [97] Stefano Piana, Alexander G Donchev, Paul Robustelli, and David E Shaw. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *The journal of physical chemistry B*, 119(16):5113–5123, 2015.
- [98] Herman JC Berendsen, JPM van Postma, Wilfred F Van Gunsteren, ARHJ DiNola, and Jan R Haak. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81(8):3684–3690, 1984.
- [99] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1), 2007.
- [100] Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190, 1981.
- [101] Xavier Periole, Marco Cavalli, Siewert-Jan Marrink, and Marco A Ceruso. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *Journal of chemical theory and computation*, 5(9):2531–2543, 2009.
- [102] Peter C Kroon, Fabian Grünewald, Jonathan Barnoud, Marco van Tilburg, Paulo CT Souza, Tsjerk A Wassenaar, and Siewert-Jan Marrink. Martinize2 and vermouth: Unified framework for topology generation. *arXiv preprint arXiv:2212.01191*, 2022.
- [103] Dmitriy Frishman and Patrick Argos. Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics*, 23(4):566–579, 1995.
- [104] F Emil Thomasen, Tórrur Skaalum, Ashutosh Kumar, Sriraksha Srinivasan, Stefano Vanni, and Kresten Lindorff-Larsen. Rescaling protein-protein interactions improves martini 3 for flexible proteins in solution. *Nature Communications*, 15(1):6645, 2024.
- [105] Adolfo B Poma, Marek Cieplak, and Panagiotis E Theodorakis. Combining the martini and structure-based coarse-grained approaches for the molecular dynamics studies of conformational transitions in proteins. *Journal of Chemical Theory and Computation*, 13(3):1366–1374, 2017.
- [106] Ksenia Korshunova, Julius Kiuru, Juho Liekkinen, Giray Enkavi, Ilpo Vattulainen, and Bart MH Bruininks. Martini 3 oligomers: A scalable approach for multimers and fibrils in gromacs. *Journal of Chemical Theory and Computation*, 20(17):7635–7645, 2024.
- [107] Emanuele Zippo, Dorothee Dormann, Thomas Speck, and Lukas S Stelzl. Molecular simulations of enzymatic phosphorylation of disordered proteins and their condensates. *Nature Communications*, 16(1):4649, 2025.
- [108] J. A. Anderson, J. Glaser, and S. C. Glotzer. Hoomd-blue: A python package for high-performance molecular dynamics and hard particle monte carlo simulations.

- Computational Materials Science*, 173:109363, 2020. doi: 10.1016/j.commatsci.2019.109363.
- [109] Jiarul Midya, Sergei A Egorov, Kurt Binder, and Arash Nikoubashman. Phase behavior of flexible and semiflexible polymers in solvents of varying quality. *The Journal of chemical physics*, 151(3), 2019.
 - [110] D Janka Bauer, Lukas S Stelzl, and Arash Nikoubashman. Single-chain and condensed-state behavior of hnrnpa1 from molecular simulations. *The Journal of Chemical Physics*, 157(15), 2022.
 - [111] A Werner, F Schmid, M Müller, and K Binder. “intrinsic” profiles and capillary waves at homopolymer interfaces: A monte carlo study. *Physical Review E*, 59(1): 728, 1999.
 - [112] Arya Changiarath, Jasper J Michels, Rosa Herrera Rodriguez, Sonya M Hanson, Friederike Schmid, Jan Padeken, and Lukas S Stelzl. Conditions for the co-existence of promoter and gene-body condensates. *bioRxiv*, pages 2024–03, 2024.
 - [113] Zakarya Benayad, Sören von Bülow, Lukas S. Stelzl, and Gerhard Hummer. Simulation of FUS Protein Condensates with an Adapted Coarse-Grained Model. *J. Chem. Theory Comput.*, 17(1):525–537, jan 2021. ISSN 1549-9618. doi: 10.1021/acs.jctc.0c01064. URL <https://pubs.acs.org/doi/10.1021/acs.jctc.0c01064>.
 - [114] Gerhard Hummer. Position-dependent diffusion coefficients and free energies from bayesian analysis of equilibrium and replica molecular dynamics simulations. *New Journal of Physics*, 7(1):34, jan 2005. doi: 10.1088/1367-2630/7/1/034. URL <https://doi.org/10.1088/1367-2630/7/1/034>.
 - [115] Tobias M. Prass, Patrick Garidel, Michaela Blech, and Lars V. Schäfer. Predicting the dynamic viscosity of high-concentration antibody solutions with a chemically specific coarse-grained model. *bioRxiv*, 2025. doi: 10.1101/2025.09.17.676805. URL <https://www.biorxiv.org/content/early/2025/09/19/2025.09.17.676805>.