Intrinsic Disorder and Salt-dependent Conformational Changes of the N-terminal TFIP11 Splicing Factor

Blinera Juniku[1,2], Julien Mignon[1], Alexia Genco[2], Anna Maria Obeid[2], Denis Mottet[2*], Antonio Monari[3], Catherine Michaux[1*]

[1]*Laboratory of Physical Chemistry of Biomolecules, Namur Research Institute for Life Sciences (NARILIS) and Namur Institute of Structured Matter (NISM), University of Namur, Rue de Bruxelles 61, B-5000 Namur, Belgium*

[2]*GIGA-Molecular Biology of Diseases, Molecular Analysis of Gene Expression (MAGE) Laboratory, University of Liege, B34, Avenue de l'Hôpital, B-4000 Liège, Belgium*

[3]*Université Paris Cité and CNRS, ITODYS, F-75006, Paris, France.*

***Corresponding authors**

E-mail: catherine.michaux@unamur.be

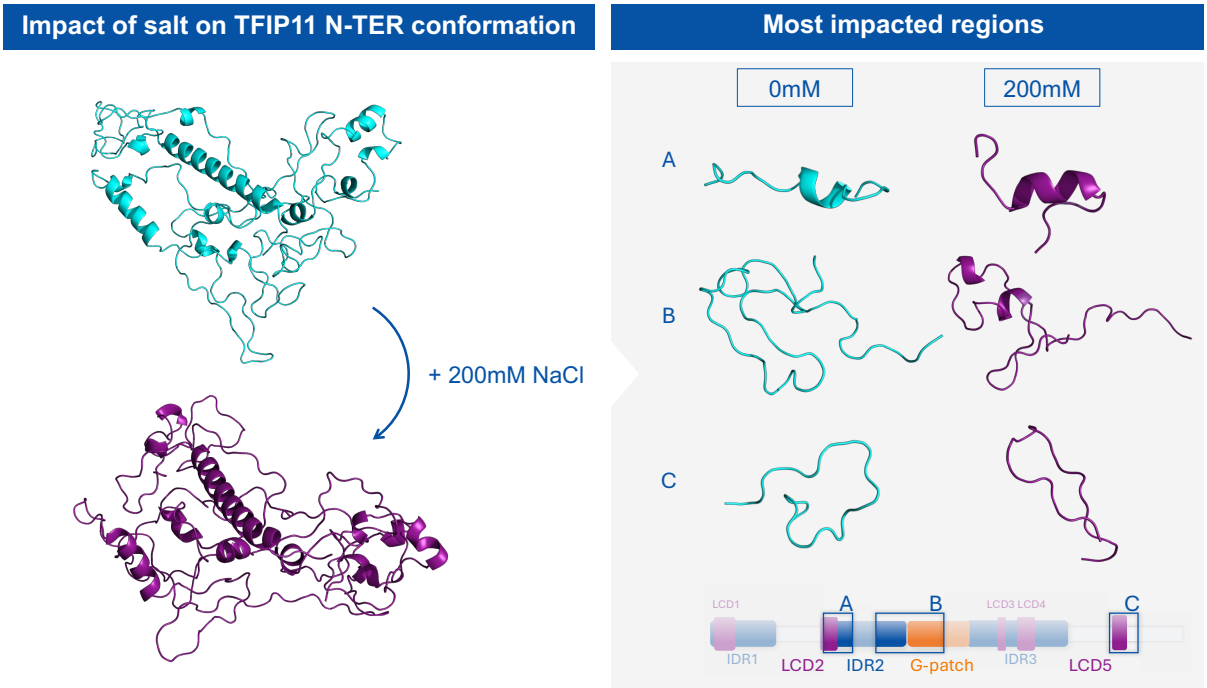
Phone: 0032472953989

E-mail: dmottet@uliege.be

Phone: 003243664422

Abstract

Tuftelin Interacting Protein 11 (TFIP11) was recently identified as a critical human spliceosome assembly regulator. Indeed, it is involved in many biological functions including the interaction with multiple spliceosome key proteins and its localisation in several membrane-less organelles. However, there is a lack of structural information on TFIP11, which limits the rationalisation of its biological role. TFIP11 has been predicted as a highly disordered protein, and specifically concerning its N-terminal (N-TER) region. Intrinsically disordered proteins (IDPs) lack a defined tertiary structure, existing as a dynamic conformational ensemble, which favours their role as hubs in protein-protein and protein-RNA interaction networks. Furthermore, IDPs are involved in liquid-liquid phase separation (LLPS) which drives the formation of subnuclear compartments. In this study, we have refined the disorder prediction of TFIP11 N-TER region and subsequently performed all-atom molecular dynamics (MD) simulations to assess its conformational flexibility and the interplay between their different N-TER domains. We further confirm that TFIP11 may be described as a polyampholyte IDP with a flexible conformation. Furthermore, since LLPS formation and IDP conformational changes are salt-dependent phenomena we have investigated by MD simulations the influence of salt concentration in shaping the conformational ensemble of the N-TER region of TFIP11. Increasing the salt concentration enhances the flexibility of the TFIP11 N-TER conformation, which presents a fuzzier conformational landscape, a more globular shape, and an unstructured arrangement that could favor LLPS segregation and protein-RNA interaction.



Keywords: Tuftelin Interacting Protein 11 – intrinsically disordered protein - molecular dynamics - LLPS

Introduction

Splicing of precursor messenger RNA (pre-mRNA) is a fundamental process in eukaryotic gene expression. The splicing of introns in pre-mRNA is carried out by the spliceosome, a dynamic macromolecular complex composed of five small nuclear RNA (snRNAs U1, U2, U4, U5, and U6) associated with more than 200 proteins. Due to its crucial role in RNA processing, spliceosome is a key cellular machinery in regulating gene expression, and its deregulation is associated with important debilitating diseases including cancers.[1–7]

Despite significant progresses in understanding the stepwise assembly of the spliceosome, the molecular mechanisms by which spliceosomal proteins interact together and mediate the ordered rearrangements within the spliceosome remain elusive. The multiple and dynamic protein-protein and protein-RNA interactions implicated in the assembly of the spliceosome machinery can be made possible by the large abundance of intrinsically disordered proteins (IDPs).

IDPs are characterized by the lack of stable secondary and three-dimensional structures. Due to their high conformational flexibility, IDPs are able to interact specifically, but transiently or weakly, with multiple protein partners, hence acting as hubs in protein-protein interaction (PPI) networks and molecular scaffolds that drive the formation of complex cellular machinery and membrane-less organelles (MLOs), such as Cajal bodies (CBs), nucleoli, or nuclear speckles, driven by liquid-liquid phase separation (LLPS).[8]

Recently, we demonstrated unrecognized functions for the G-Patch protein TFIP11 in the regulation of human spliceosome assembly and splicing efficiency.[9] We observed that TFIP11 is located in multiple membrane-less organelles (MLOs) in which the snRNPs biogenesis and maturation take place, such as nuclear speckles, Cajal bodies (CBs) and nucleoli. Interestingly, we found that TFIP11 interacts with multiple key U4/U6.U5 tri-snRNP-

specific proteins and we demonstrated that TFIP11 knockdown alters the assembly and stability of the U4/U6.U5 tri-snRNP complex, while blocking the spliceosome in a B-like conformation. Consistent with our data, several other studies reported that TFIP11 is recruited during B complex formation and stably integrated at this stage, suggesting that TFIP11 is a crucial actor during the assembly and activation of the spliceosome machinery.[10]

It is well known that structural information can provide key insights into protein function. A deeper characterisation of TFIP11 structure might therefore provide a better understanding of its function and mode of action. No structural data for TFIP11 are available so far. However, *in silico* bioinformatic prediction hypothesized the presence of three IDRs in the whole TFIP11 sequence.[9] The IDR#1 (residues 1-150) and IDR#2 (residues 175-250) surround the G-patch domain (residues 148–193). The IDR#3 (residues 720-770) overlaps the nuclear speckles targeting site (NSTS) at the C-terminal extremity. The G-patch domain constitutes a well-conserved glycine rich domain present in other eukaryotic RNA-processing proteins, such as PINX1[11], SON[12], GPKOW,[13] etc., and is involved in protein-protein and protein-nucleic acid interactions.[14,15] It is well-known that the G-patch domain of TFIP11 interacts with and activates the ATP-dependent RNA helicase DHX15.[16] Furthermore, the two predicted N-terminal IDRs in TFIP11 are also critical for interaction with coilin, the scaffold protein of CBs.[9]

In addition to IDRs, five low complexity domains (LCDs), LCD1 (residues 5–21), LCD2 (residues 82–96), LCD3 (residues 210–216), LCD4 (residues 226–240), and LCD5 (residues 291–306) are also predicted within the N-terminal region of TFIP11. LCDs are composed of a limited variety of amino acids and have distinct physico-chemical properties depending on the type of amino acid(s) each LCD is enriched with.[17] LCDs are mainly found in RNA- and DNA-binding proteins, such as the EWS[18] and TAR[19] proteins, where they

enforce gene regulation and functioning through the formation of dynamic complexes in MLOs such as CBs.[20] Indeed, the LCDs can drive the formation of extended biomolecular condensates and MLOs by favouring liquid-liquid phase separation (LLPS).[21,22] Together, these evidences strongly suggest that the N-Terminal region of TFIP11, including the two IDRs and the G-Patch domain, play a central role in promoting not only the interactions with spliceosomal proteins but also with CBs scaffold protein, such as coilin, and are therefore of crucial importance to TFIP11 function.

To get an additional insight on the N-terminal (N-TER) TFIP11 structure, its in-depth disorder-associated properties have been predicted. In addition, all-atom molecular dynamics (MD) simulations have been performed to assess TFIP11 conformational flexibility and the interplay between the different N-TER domains. Furthermore, given that LLPS formation and IDP conformational changes are salt-dependent and that *in vitro* experiments have been performed using different ionic strength conditions[23,24], the influence of salt concentration in shaping the conformational ensemble of TFIP11 N-TER has also been investigated by MD simulations.

The present study provides a better understanding of the TFIP11 intrinsic and environment-dependent structural behaviour, and in the long-term, may also offer hints on the molecular basis leading to LLPS and hence MLO formation. This in turn will allow a better understanding of the spliceosome assembly process, as well as the key role of TFIP11 IDPs in promoting the complex and finely tuned regulations of different biological processes.

Computational Methodology

Sequence-based bioinformatics predictions

Disorder-associated properties have been predicted from the known amino acid sequence of the TFIP11 (Uniprot ID: Q9UBB9) N-TER involving residue 1 to 350. Overall, the per-residue percentage of intrinsic disorder (PPID) has been determined considering the average of 20 disorder predictors available online. This includes the Predictors of Natural Disordered Regions (PONDR) series of algorithms VL-XT, XL1-XT, CAN-XT, VL3, and VSL2[25–27]; Prediction of Intrinsically Unstructured Proteins (IUPred3) for short and long disordered segments[28]; Protein DisOrder prediction System (PrDOS)[29]; ESpritz algorithms respectively trained on X-ray, Disprot, and NMR datasets;[30] non-evolutionary and evolutionary-based Prediction of Order and Disorder by evaluation of NMR data (ODiNPred) [31]; deep-learning-based predictor metapredict (v2.2)[32]; NORSnet, Ucon, and MetaDisorder MD algorithms on the PredictProtein webserver[33]; NetSurfP-3.0[34]; putative function and linker-based Disorder Prediction using deep neural network (flDPnn)[35]; and finally DisoMine from Bio2Byte tools[36]. Disorder profiles were generated with the Rapid Intrinsic Disorder Analysis Online (RIDAO) tool[37], comprising PONDR and IUPred predictions. On the disorder plot, residues with a score above the 0.5 threshold are considered as disordered, while flexible segments typically exhibit scores between 0.2 and 0.5.

The disorder propensity and the conformational ensemble of TFIP11 N-TER was further examined by cumulative distribution function (CDF) and charge-hydropathy (CH) plots obtained from the PONDR server[38], as well as by the Das-Pappu phase diagram with the Classification of Intrinsically Disordered Ensemble Regions (CIDER)[39]. Namely, κ and Ω patterning parameters were also extracted from the CIDER analysis. The κ parameter describes the extent of charged amino acid mixing in a given sequence, while Ω describes the distribution of charged and proline residues with respect to all other residues. Presence and localisation of

molecular recognition features (MoRFs) along the TFIP11 N-TER sequence have also been computationally identified with the MoRFchibi tool.[40]

Molecular dynamics (MD) simulations

All atoms equilibrium MD simulations of the TFIP11 N-TER, in different environmental conditions, have been performed in three independent replicates (triplicates) using the GROMACS 2020.5 suite.[41] All the simulations have used the relaxed first-ranked AlphaFold2-generated model of TFIP11 as the initial structure (**Fig. S1**). The starting protein system was centred in a truncated octahedron water box including a 10 Å buffer from each edge and has been neutralised with the required minimal amount of Na^+ cations. In a second system a NaCl concentration of 200 mM has been enforced, using the same protocol. The initial systems have been obtained using the *tleap* module of AmberTools. The AMBER ff14SB[42] force field has been used to model the protein, while water has been described with the OPC model[43]. To better account for the TFIP11 disordered character, grid-based energy correction maps (CMAP) method, available for the amber ff14SB force fields, was applied to the protein intrinsically disordered regions (IDRs) and the flexible segments, i.e. residues 1-24, 36-44, 57-143, and 166-337[44]. Hydrogen mass repartitioning (HMR)[45], redistributing the mass of non-solvent hydrogen atoms has been consistently used to allow, in combination with Rattle and Shake, the use of a 4 fs time-step for the integration on the Newton equations of motion. HMR was enforced by modifying the initial topology with the *ParmEd* package available in Amber Tools. Prior to the MD, the system has been minimised for 50000 steps, and subsequently thermalised and equilibrated. MD has been propagated in the isothermal and isobaric (NPT) ensemble. Constant temperature (300 K) and pressure (1 bar) have been maintained using a modified Berendsen thermostat and the Parrinello-Rahman barostat, respectively. Long-range electrostatic interactions have been calculated using the Particle Mesh Ewald (PME)

summation[46]. Each MD replicate was propagated for 1 μs in periodic boundaries conditions (PBC) in all three dimensions, energies and trajectories being recorded every 20 ps. Root-mean square deviation (RMSD), RMSD distribution, root-mean square fluctuation (RMSF), and radius of gyration (R_g) data were calculated from the resulting trajectories, which have been visualised and analysed using PyMOL[47] and VMD[48].

Results and discussion

Prediction of sequence-based disorder-associated properties for TFIP11 N-TER

Given the complex behaviour of the N-TER domain of TFIP11 we aimed at refining the bioinformatic predictions assessing and comparing complementary descriptors. Based on the average of the 20 disorder predictors, the overall per-residue percentage of intrinsic disorder (PPID) globally amounts to 70% for the TFIP11 N-TER. The predicted PPID tendency is also corroborated by the content in order-promoting (OPRs) and disorder-promoting residues (DPRs). The latter takes into account the effects of short non-polar (glycine and alanine), charged (lysine, arginine, and glutamate), polar (serine and glutamine), and secondary structures breaking (proline) residues. As previously reported[9], full-length TFIP11 has a higher proportion of DPRs (50.2%) compared to OPRs (34.2%). Yet, this tendency is further amplified for its N-TER domain, in which the amount of DPRs increases up to 59.4%, while the amount of OPRs is concomitantly decreased to 25.7%.

By using three independent algorithms (MFDp2 v2.0, PrDOS, and DISOPRED v3.1), two intrinsically disordered regions (IDRs) have been previously identified in TFIP11 N-TER at positions 1-150 and 175-250.[9] We have refined the occurrence and location of IDRs with Rapid Intrinsic Disorder Analysis Online (RIDAO), generating and analysing the PONDR-FIT and IUPred predicted disorder profiles (**Fig. 1a**). Although some variability can be observed between the different predictors, even for the more disordered N-TER region, three main and

discontinued IDRs clearly stand out showing a disorder score well above 0.5. The three IDRs are located at the N-TER region and are separated by shorter flexible segments whose disorder score varies between 0.2 and 0.5. The first disordered region (IDR1) is comprised between residues 1 and 50 and is followed by two longer disordered regions comprising residues 85-145 (IDR2) and 190-265 (IDR3). Due to a much larger variability amongst the predictors for the residues comprised between the position 270 and 350, we restrain to assign this segment as an IDR. Nevertheless, scores close to the 0.5 threshold indicate that a large flexibility is still present in this region. Interestingly, the three IDRs have a high content of charged and hydrophilic residues (**Fig. S2**), a very common characteristic of IDP.[49]

TFIP11 disorder propensity was further investigated with the cumulative distribution function (CDF), as well as the charge-hydropathy (CH) plots (**Fig. 1b-c**).[50] In the case of CDF plot, IDPs should consistently remain below the defined boundary, while ordered proteins (ORDPs) should overcome the threshold.[51] The N-TER CDF profile is located slightly below the boundary, thus supporting its global disordered nature. Indeed, the presence of the tuftelin-binding domain, which should (might) be ordered, as well as partially folded sub-structures linking the N-TER IDRs may shift the plot close to the CDF boundary. The difference in disorder content is far more noticeable on the CH plot, discriminating IDPs from ORDPs on the basis of their absolute mean net charge $\langle R \rangle$ with respect to their mean scaled hydropathy $\langle H \rangle$.[52] The TFIP11 N-TER region displays both low mean net charge and hydropathy and is found above the boundary defined by the equation $\langle R \rangle = 2.785\langle H \rangle - 1.15$, i.e. clearly laying in the IDP area, strongly supporting the assignation of the N-TER domain as a mainly disordered one.

The disorder-related conformational state of TFIP11 was further examined with the Das-Pappu phase diagram, as well as the κ and Ω parameters defined by the Classification of Intrinsically Disordered Ensemble Regions (CIDER) (**Fig. 1d**). Notably, the Das-Pappu phase

diagram describes the conformational ensemble of a selected IDP from its fraction of positively against negatively charged residues.[53] The TFIP11 N-TER region is found within the R2 region of the diagram, i.e. the region which comprises most of IDPs (~40%). R2-associated IDPs are referred to as Janus sequences, reflecting their structural duality, namely the coexistence of ordered and disordered phases, also depending on their coupling with the environment. Indeed, R2-associated IDPs exist either as a collapsed ensemble (molten globule), an arrangement which is mostly driven by hydrophobic effect, or as extended conformations (random coil) driven by favourable protein chain solvation. Interestingly, the TFIP11 N-TER domain nearly reaches the R3 category (~30% of IDPs), comprising strong polyampholytes with coiled or hairpin conformations.[54]

The analysis of the patterning parameters informs on the distribution of positively and negatively charged amino acids along the sequence (κ), as well as on the alternation of charged and proline residues (Ω). On a scale from 0 to 1, the closest the indexes are to 1, the more segregated the charged and/or the proline residues in the protein sequence are, thus resulting in more collapsed and compact conformers.[53,55] TFIP11 N-TER has κ and Ω values of 0.24 and 0.14, respectively. κ and Ω values indicate the absence of extended patches of charged and proline residues, while their values remain sufficiently low to suggest that the N-TER region preferentially adopt more expanded conformations.

Molecular recognition features (MoRFs) are short disordered regions (10-70 residues) which can be found within a given protein sequence, and which are able to fold by binding to the interaction partners and, thus, enforce protein-protein interactions in complexes involving IDPs.[56] The number and the location of such segments in TFIP11 N-TER have been determined using the MoRFchibi tool. We have identified five MoRFs, which are consistently localised in the N-TER domain, and three of them are found in the IDRs. Indeed, while IDR1 contains two MoRFs, MoRF1 (residues 1-15) and MoRF2 (residues 25-31), MoRF5 (residues 235-249)

belongs to IDR3. Interestingly, two additional MoRFs, MoRF3 (residues 75-81) and MoRF4 (residues 175-180), are situated in the short flexible regions linking IDR1 to IDR2 and IDR2 to IDR3, respectively. These regions are highlighted within the sequence in **Fig. 2**.

All together, these analyses suggest the classification of TFIP11 as an IDP and reveal that its main disorder area lies in its N-TER domain, as particularly evidenced by the disorder-associated predictions. Instead, the presence of several MoRFs further supports the critical role of the N-TER region in TFIP11 functionality, probably in favouring protein-protein interactions in molecular complexes.

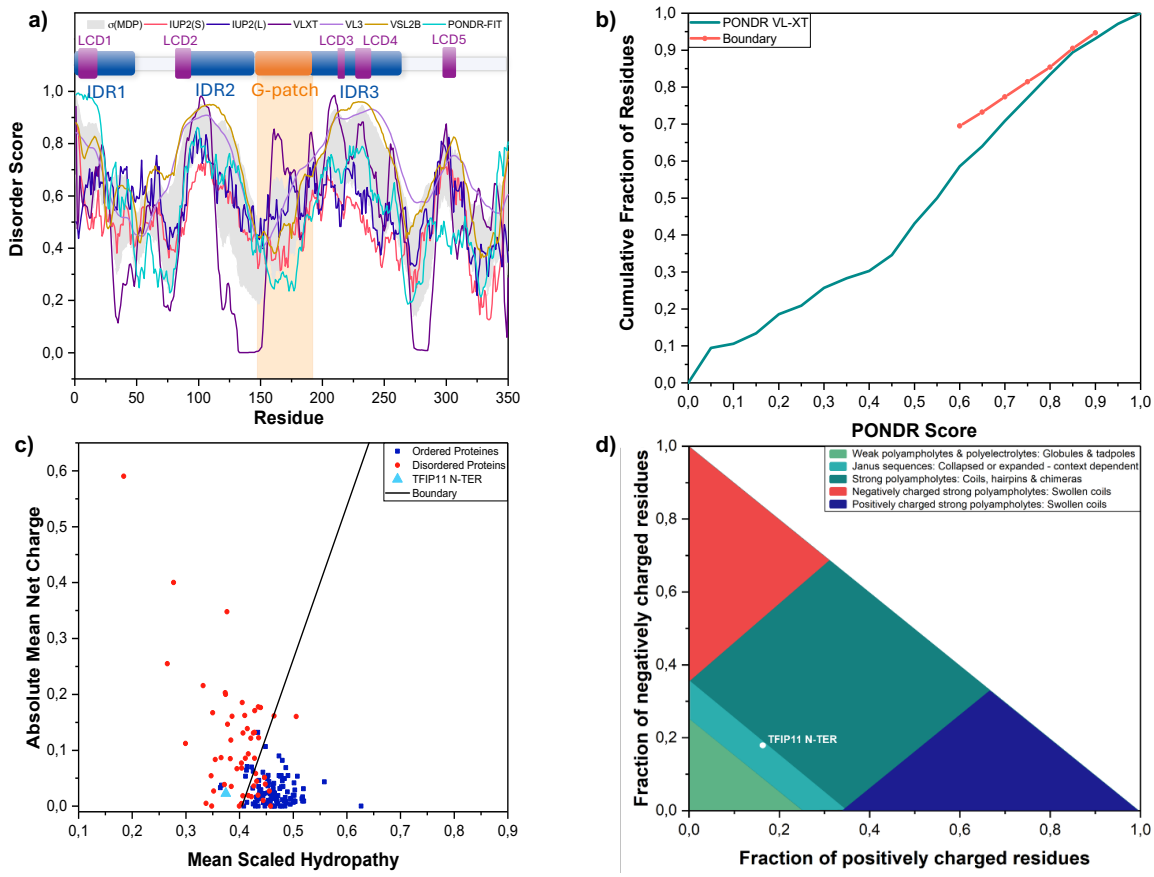


Fig. 1: *Predicted disorder-associated properties of TFIP11 N-TER. (A) PONDR and IUPred disorder profile, sequence and domains organisation of TFIP11 N-TER; MDP*

corresponds to the mean disorder profile. Every residue displaying a score above the 0.5 threshold is considered disordered, (B) CDF, (C) CH plots and (D) Das-Pappu phase diagram.

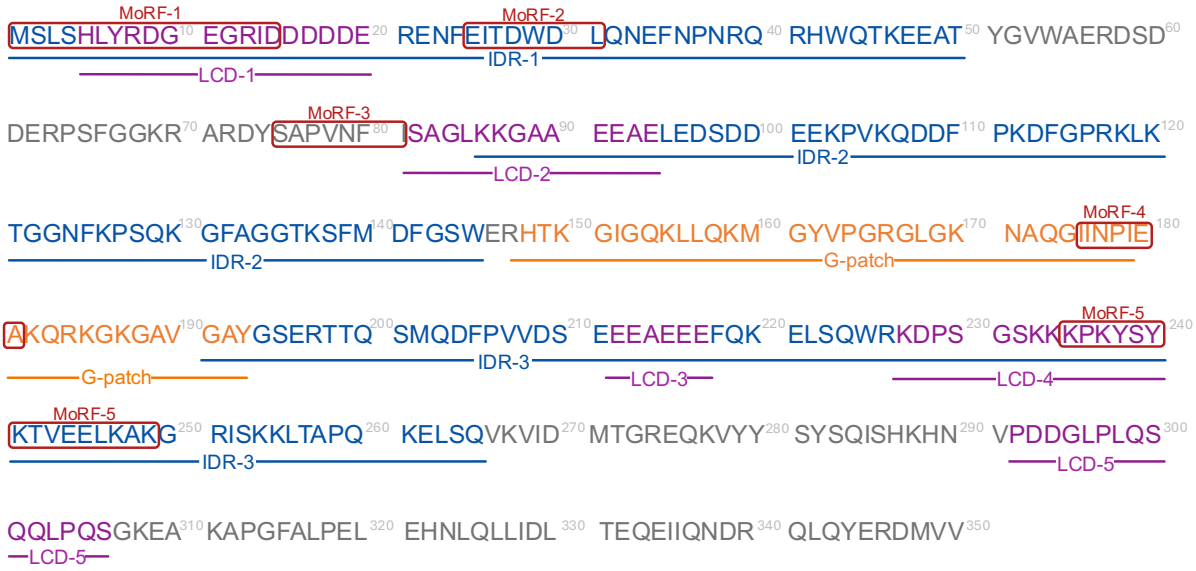


Fig. 2: *TFIP11 N-TER sequence with domain organisation (1-350 amino acid residues) comprising IDRs, LCDs, G-patch, and MoRFs.*

Molecular dynamics simulations of TFIP11 N-TER

Since LLPS formation and IDP conformational changes are complex and salt-dependent phenomena[23], we have investigated by all-atom MD simulations the intrinsic structural properties and the influence of salt concentration in shaping the conformational ensemble of the N-TER region of TFIP11.

First, we compare the stability of the TFIP11 N-TER conformational ensembles at 0 and 200 mM NaCl concentration, by calculating the RMSD for the protein backbone over the entire TFIP11 N-TER sequence for both conditions (**Fig. 3a; Fig. S3**). During the simulation, the overall fold of the protein in 0 mM NaCl remains stable despite its intrinsic disorder. However, after the first 400 ns of simulation, a slight increase in RMSD value is observed for the protein at 200 mM NaCl, suggesting that the presence of salt allows TFIP11 N-TER to

explore a slightly different conformational space during the simulation. In addition, the higher RMSF values for both conditions are unsurprisingly located in the IDRs and LCDs, which is consistent with their flexible properties (**Fig. 3b; Fig. S4**).

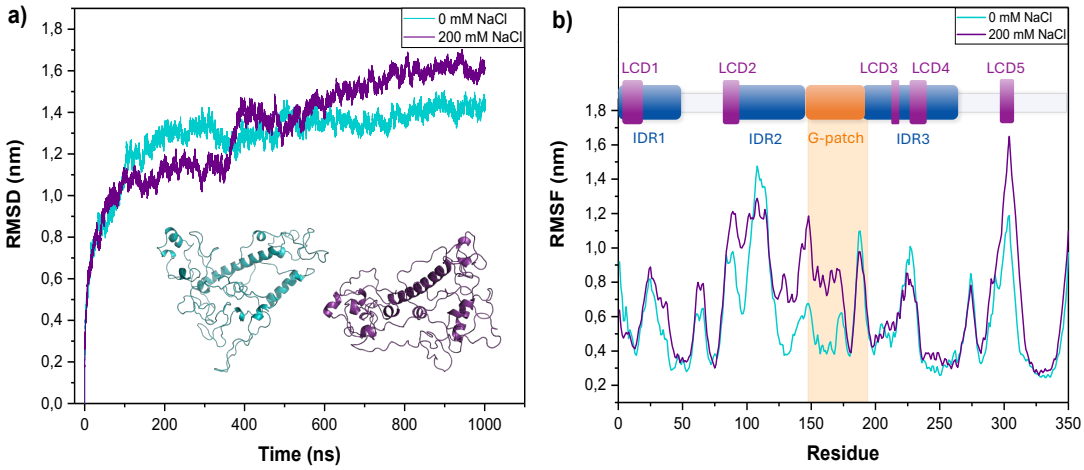


Fig. 3: *RMSD and RMSF profile for TFIP11 N-TER. a) The root mean square deviation (RMSD) of the backbone atoms from the equilibrated conformation (0 ns) is presented as a function of time. The RMSD time profiles for 0 mM (blue) and 200 mM (violet) NaCl are shown. b) Root mean square fluctuation (RMSF) values of atomic positions computed for the backbone atoms are shown as a function of residue number. The RMSF values for 0 mM (blue) and 200 mM (violet) NaCl are shown.*

Interestingly, an important difference for the two salt concentrations is observed within LCD2, IDR2, G-patch, and LCD5, which present almost systematically higher RMSF values in the 200 mM NaCl condition. These domains are comprised within the three regions (A, B, C) with the highest variation of RMSF (ΔRMSF) (**Fig. 4a**). The variation of RMSF is particularly high within the region B and C containing IDR2, G-patch, and LCD5, indicating a large backbone flexibility in these regions. Interestingly, G-patch proteins are known to interact with helicases via their eponymous glycine-rich motifs. In the peculiar case of TFIP11, its G-patch domain interacts with the RNA helicase DHX15.[57] From a structural point of view, and

given their sequence similarity, it is suggested that the G-patch domain of TFIP11 could act in a similar way as the one of the ribosome biogenesis factor NKRF.[58] Indeed, as determined in the crystal structure of the human helicase DHX15/Prp43 in complex with the G-patch domain of NKRF, the latter acts like a flexible arm linking dynamic portions of DHX15 and tethering the two mobile parts of the protein together. The NKRF G-patch motif is mostly unstructured and flexible, apart from a short N-terminal α-helix, which is in line with the structure of the B region in TFIP11 (**Fig. 4b**), and thus again confirms the role of structural flexibility in regulating TFP11 biological function. In addition, the structural changes shown in **Fig. 4b** highlight the environmental sensitivity of the conformation of the TFIP11 G-patch, IDRs, and LCDs.

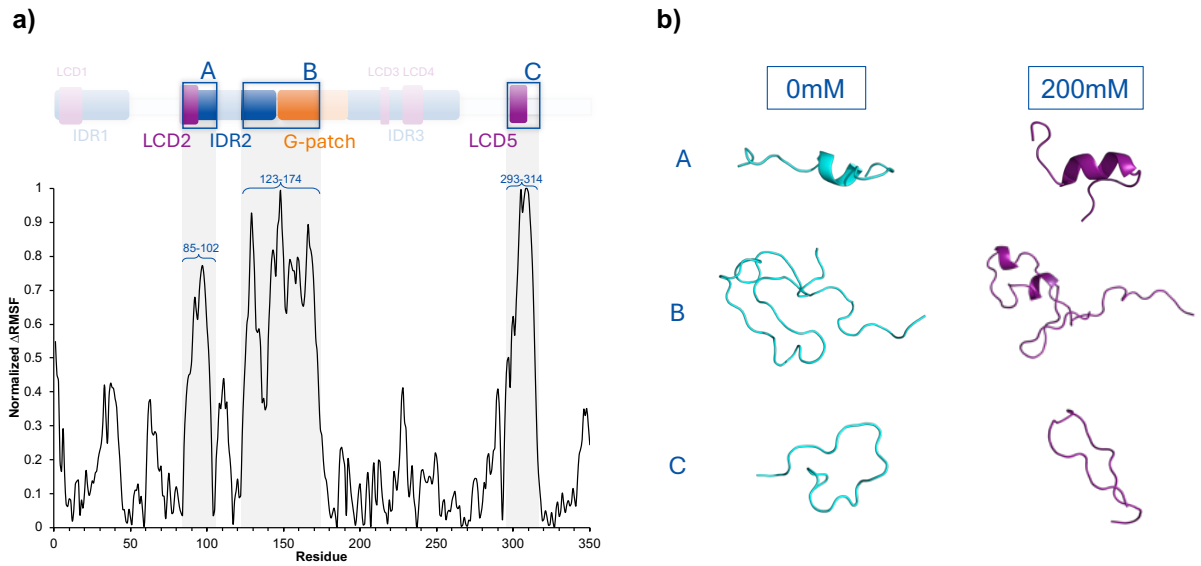


Fig. 4: *a) Per residue normalised variation of RMSD between 0 and 200 mM, with the most impacted regions highlighted in gray and shown within the sequence domains. b) 3D structure of the TFIP11 N-TER regions impacted by salt concentration at 0 mM (blue) and 200 mM (violet) NaCl are shown.*

A further indication of the stability of the protein structure may be obtained by analysing the evolution of the radius of gyration (R_g) (**Fig. 5a; Fig. S5**). It is shown that the

increase in ionic strength does not correlate directly with significant changes in the R_g values. However, at 200 mM NaCl, the R_g shows a slight increase and variation going from an average value of 2.71 nm to 2.76 nm, coherently with the variations in RMSD and RMSF. The average R_g values of TFIP11 N-TER are slightly higher than the ones for ordered protein sharing the same length, usually comprised around 2.0 ± 0.3 nm, further supporting its disordered features and its more extended conformation.[59]

Superimposed snapshots, extracted from the MD simulations for the two salt conditions (**Fig. 5b**), permit to visually appreciate the shape and conformational changes of the protein induced by the increase in the ionic strength, which, despite the modest effects on R_g, appear as substantial.

The intensity of the effects of salt concentration in shaping IDP conformation depends on the charge repartition within the protein, as well as on the salt concentration gradient.[24,60] Generally, at low salt concentration, the conformational changes are primarily driven by the possibility to maximise the screening of unfavourable electrostatic interactions. Indeed, as observed for TFIP11 N-TER, polyampholyte IDPs, have been shown to undergo conformation expansion, which may lead to an increase of 5 Å in the R_g when going from 0 to 1 M salt concentration.[61] Interestingly, such expanded conformation allows homotypic IDP-IDP interaction that further drive LLPS formation. This is the case of the tau protein, a polyampholyte IDP involved in Alzheimer disease, the conformation of which extends within LLPS droplets.[62] The weak, but significant, salt concentration-dependent increase in the R_g of TFIP11 N-TER is consistent with these observations. This leads to the hypothesis that TFIP11 could form LLPS under certain conditions, a characteristic which may be related to its biological functions inside MLOs.

To better assess the conformational variability along the MD simulations, the RMSD distribution of TFIP11 N-TER in 0 and 200 mM NaCl conditions are also compared (**Fig. 5b**). Both systems display a broad distribution, which is coherent with the inherent flexibility of the protein. At 200 mM NaCl, the distribution is globally shifted towards higher values indicating that the structure deviates further from its starting conformation. These structural changes are highlighted by superimposing representative snapshots corresponding to the first 10 clusters in the two conditions of 0 mM and 200mM NaCl (**Fig. 5b**). The protein at 200 mM NaCl has a fuzzier conformational landscape, a more globular shape, as well as an overall unstructured conformation as compared to the system at 0 mM NaCl. Furthermore, the 200 mM NaCl system also appears to present more accessible regions that might be involved in homotypic interactions, and hence should be more favourable to develop PPI networks.

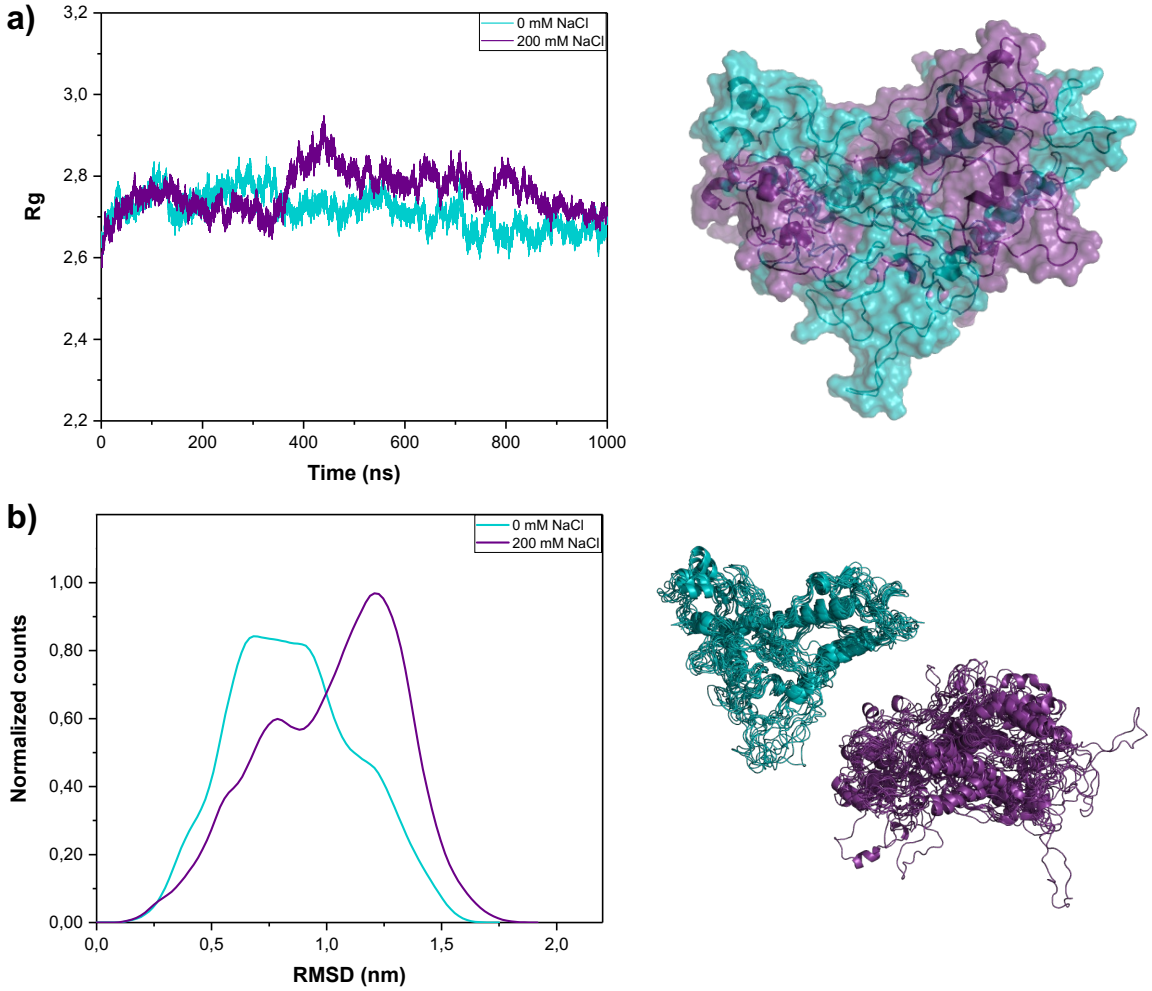


Fig. 5: a) *Radius of gyration (R_g) versus time plot during the 1000 ns molecular dynamics simulation for TFIP11 N-TER. The R_g values for 0 mM (blue) and 200 mM (violet) NaCl are shown. b) Backbone root-mean-square deviation (RMSD) distribution of TFIP11 N-TER.The first 10 clusters are superimposed for the system at 0mM (blue) and 200 mM (violet).*

To obtain a deeper understanding of the structural changes induced by salt concentration, the timeline analysis of the secondary structure content of the TFIP11 N-TER was performed (**Fig. 6**). Some structural features in the C-terminal end of the N-TER region (residues 209-225, 243-249, and 318-345), such as the α-helices, appears highly stable and persistent all along the MD simulation, and rather independent on salt conditions. This

observation is also coherent with the low flexibility observed for these regions in the RMSF analysis, and by the fact that these motifs have been correctly predicted by AlphaFold (**Fig. S1**). Interestingly, this secondary structure stability also involves structural motifs, once more in the form of α-helices, that are embedded in the IDR1 and IDR3 regions. Yet, many regions (residues 3-7, 28-35, 46-51, 89-95, 154-160, 261-265, and 282-286) contain transient helices that vary between α- and 3_{10} conformations, highlighting the dynamic nature of the TFIP11 N-TER. However, another stable structural motif involves the extended ß-sheet downstream of IDR3 (residues 268-270 and 277-279), which varies only slightly with salt condition.

Yet, a significant increase in random coil content at higher ionic strength is observed in every LCD (except for the shortest LCD3), the regions containing IDR2, IDR3, and the G-patch, which is also accompanied by a more global modification in the secondary structure composition. In the case of LCDs, turns are the most commonly disrupted structural motifs in favour of random coil, with the greatest change observed within LCD1. This LCD1 domain differs from the other similar sub-domains due to its high composition of consecutive negatively charged residues (HLYRDGEGRIDDDDDE), which could be preferentially involved in electrostatic interaction with the environmental salt. The tertiary structures corresponding to the first cluster show that the modification of turns into random coil participates to the expansion of the system in 200mM NaCl, specifically underlying the increased accessibility of LCDs and the G-patch. This is coherent with previous observations reported in the literature concerning the role played by LCDs in driving LLPS.[63] Coherently with the observations from the RMSF analysis, the G-Patch domain also appears highly modified by the increased salt concentration. In particular, we may observe a strong decrease of turns content in favour of extended conformations. Interestingly, the formation of a α-helical motif at 200 mM NaCl between the two first conserved glycine residues in the G-patch has also been observed in NKRF, highlighting the sequence and conformational similarity between

the two proteins.[64] It is also important to underline that high ionic strength conditions appears to favour the presence of random coil arrangement in the IDR2 domain, which is also one of the protein regions most affected by the difference in salt concentration. These data suggest that IDR2 and the G-patch are interdependently involved in conformational changes.

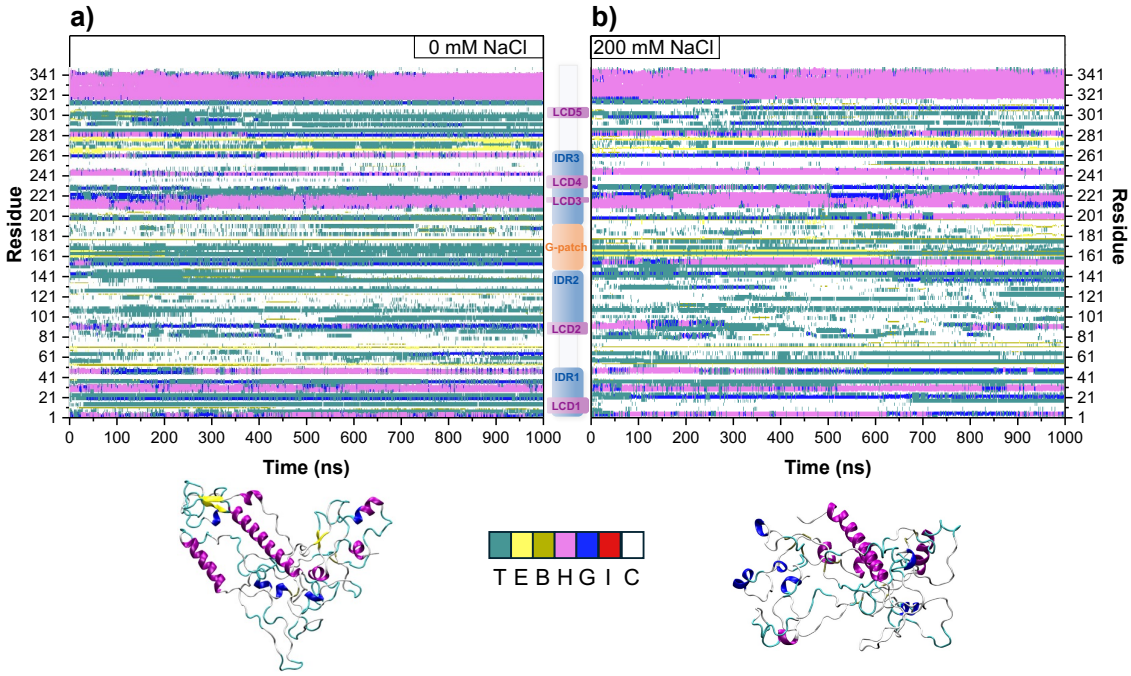


Fig. 6: *Secondary structure timeline analysis as computed by the timeline plugin contained in VMD. In the graphic, the extended β-sheet components (E) and isolated bridges(B) are represented in yellow and dark yellow, respectively; turns (T) are in teal; degrees of helix are α-helix (H) in pink, 3-10 helix (G) in blue, and π-helix (I) in red; random coil (C) is in white. 3D structures of the first cluster for both salt conditions at (a) 0 mM and (b) 200 mM NaCl are represented with the aforementioned colour-coded secondary structures.*

IDPs are prone to post translational modifications (PTMs) which can regulate their functions, for example, interaction with other protein partners, protein folding, etc. and consequently, change protein functions in various biological processes. PTMs of RNA-binding

proteins (RBPs) can directly enhance or reduce the interactions with other proteins and/or RNA components, contributing to the formation of MLOs. TFIP11 is known to undergo extensive PTMs. In yeast, a lysine (K) residue at position 67 within the G-patch domain of Ntr1/Spp382 is important for the binding of TFIP11 to RNA.[65] This K67 residue is equivalent to K residue at position 155 in the G-patch of human TFIP11 and is highly conserved across species (**Fig S6**), suggesting a crucial importance of this residue for structure and function of TFIP11. In addition, this specific K155 is reported by the PhosphoSitePlus[66] database as well as in the litterature[67] as an acetylated residue, suggesting an acetylation-dependent regulation of TFIP11 function. Interestingly, Tannukit S., *et al*.[68] report that one tyrosine residue (Y162) located in the G-Patch domain is phosphorylated. This conserved Y162 residue is required for binding to nucleic acid.[69] The presence of this phospho-Y162 in the G-patch, which is predicted to contain two α-helices with four out of the six glycine residues located within an intervening loop, is consistent with the observation that phospho-Tyr is more often observed at ordered interfaces (i.e. structures predicted to be ordered even in the unbound state).[70,71] In addition, proteome-wide analysis of arginine monomethylation reveal that the arginine residue at position 166 (R166) located in TFIP11 G-Patch domain is mono-methylated and sensitive to PRMT1/4/5 inhibition.[72] This R166 residue is adjacent to phospho-Y162 residue, which is consistent with Larsen S. C., *et al*. study[72] reporting that arginine methylation sites regulated by PRMT1/4/5 are often found to be adjacent to phosphorylation sites. Moreover, this R166 is located between two neighbouring glycine residues (GRG), a preferential site for the PRMT5 enzyme[73], a protein methyltransferase which has been detected in B spliceosomal complex. Therefore, it is tempting to speculate that PTMs in functional domains of TFIP11, such as the G-patch, could be the molecular basis for binding to spliceosomal proteins and/or RNA substrates, and consequently contribute to the structural arrangement and activation of the spliceosome complex.

Interestingly, Y162, K155, and R166 are located within the B region (**Fig. 4**; **Fig. S7**), presenting a high variation of RMSF, indicating the large impact of salt concentration in this region containing both IDR2 and the G-patch. By looking closer to the position of these residues in the MD structures, we can estimate their impact on salt-induced conformational changes. In **Fig. 7**, we compare the conformational environment and the interactions of Y162, K155, and R166 with other residues depending on salt concentration. For K155, the presence of salt modifies the interactions of such positively charged residue, which previously interacted with two negatively charged residues, D34 and E47, to finally become totally accessible to the solvent, thus maximising electrostatic interactions with ions in solution. This gain in accessibility may favour interactions with RNA and/or facilitate the addition of PTM, such as acetylation, to dynamically regulate RNA binding. The environment of R166 is also changed by the presence of salt. By switching from a π-cation interaction with residue W145 to an electrostatic interaction with residue D18, R166 participates in bringing the G-patch and LCD1 (two types of domains found in RNA-processing proteins) closer together. Although residues K155 and R166 are found close to Y162, the latter is not significantly affected by the presence of salt, probably because its polar nature makes it less sensitive to salinity than charged residues. Nevertheless, addition of a negative charge to Y162 through phosphorylation could significantly change the conformation of the G-patch and thus impact its function.

Another PTM frequently found in TFIP11 is the phosphorylation of multiple serine residues such as S59, S98, and S210 (**Fig. S7**) (see PhosphositePlus database for all references for each residue). Therefore, we decided to compare the variation in structure and interactions of such residues according to salt concentration (**Fig. S7**). S98 is located in the high ΔRMSF region A (**Fig. 4a**) in IDR2. Its interaction with K103, a charged residue within IDR2, is favoured by the presence of salt. A similar effect is observed for S59. At 200 mM NaCl, an electrostatic interaction with R70 is also promoted. The salt-induced proximity of these serine

residues to positively charged residues shows that the charge of the system locally influences intramolecular interactions. This suggests that positively charged R or K residues, known to be involved in interactions with nucleic acids[74], could be attracted by a local negative charge introduced by the addition of a phosphate group.[75] This feature is frequently found in phosphorylated splicing factors such as SF1[76] and ASF/SF2 (or SRPK1) and is known as "arginine claw" in reference to the cluster of arginine residues around the added phosphate group.

We have also investigated Short Linear Motifs (SLiMs) that refer to linear protein interaction sites containing ligand motifs mediating protein-protein interactions and post-translational modification motifs and are directly recognized and targeted for PTM by regulatory enzymes.[77] The number and the location of the predicted SLiMs in TFIP11 N-TER have been determined using the Eukaryotic Linear Motif (ELM) resource prediction tool.[78] The predicted SLiMs are shown in table (**Fig. S8**). Only those contained in the flexible regions A, B and C are listed. We focus on SLiMs involved in the regulation of LLPS. For example, the SLiM (residues 155-159) located within the G-patch is predicted as a Ubiquitin carboxyl-terminal hydrolase 7 (USP7 CTD) domain binding motif. This type of hydrolases is responsible for the deubiquitination of target proteins. This is interesting given that K155 is also predicted by PhosphositePlus as being ubiquitinylated (as well as acetylated). A USB7 MATH domain binding motif is also predicted in LCD5 within region C (residues 297-301). Ubiquitination is known to modulate LLPS for many proteins such as Tau[79], UBQLN2[80] and HECT E3[81]. Depending on the protein, ubiquitination can promote or inhibit the formation of LLPS. Another interesting motif is the SLiM (residues 303-309) found in the C region comprising LCD5 and predicted as a phosphorylation site for casein kinase 2 (CK2). This kinase plays a key role in IDP folding.[82] For example, the phosphorylation of FRMP by CK2 increases the density of negative charge in its IDRs and increases electrostatic interactions leading to

LLPS.[83] Studying the effect of phosphorylation by CK2 and other enzymes involved in the post-translational modification of TFIP11 will provide a better understanding of the mechanisms regulating the LLPS phenomenon within the cell.

Further structural and functional studies on post-translationally modified and repositioned residues will allow to identify their role in the multiple biological functions of TFIP11.

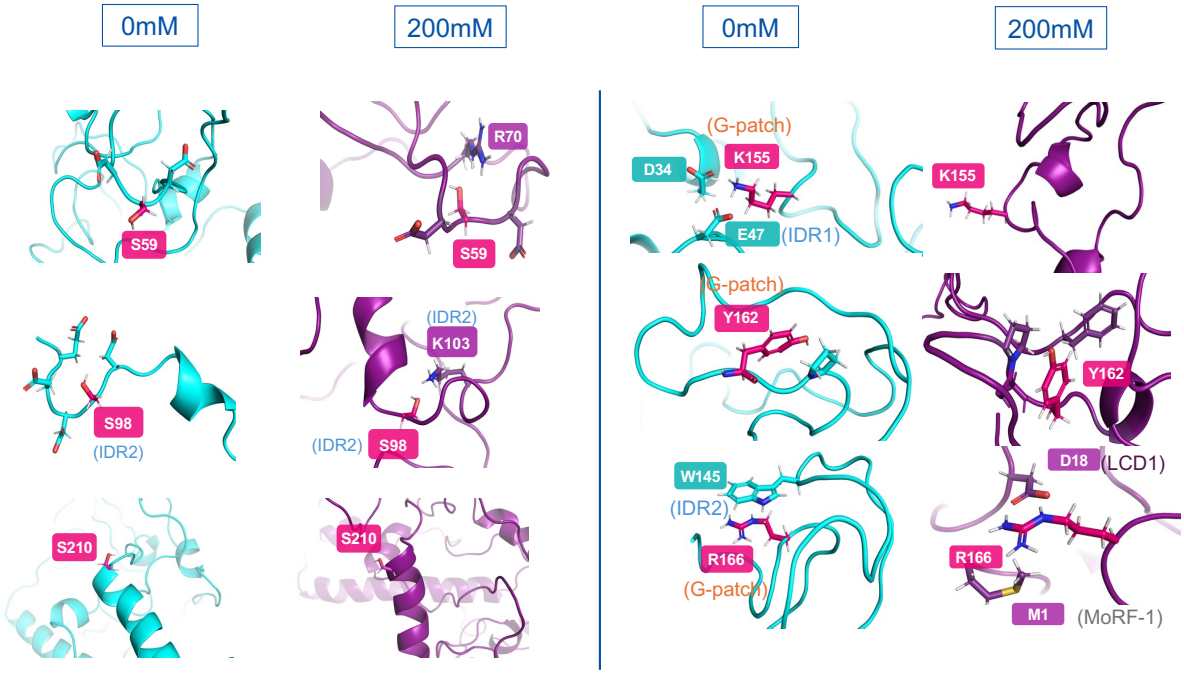


Figure 7: *PTM residues S59, S98, S210, K155, Y162, and R166 with their modified environment depending on salt concentration.*

Conclusions

In the present study, disorder predictors have demonstrated that TFIP11 N-TER is a mainly disordered protein that enters the category of polyampholyte IDP. According to the Das-Pappu phase diagram, this also implies a structural duality, namely the coexistence of ordered and disordered phases, the predominance of which may depend on their coupling with the environment. *In silico* predictions have also allowed to refine the position of three IDRs

and identified several MoRFs, further supporting the critical disorder-associated role of the N-TER region in TFIP11 functionality.

MD simulations have provided information on the TFIP11 N-TER conformational space and how it is impacted by salt concentration. According to the RMSD distribution, TFIP11 N-TER displays a broad conformational ensemble, highlighting the coexistence of different conformations, coherently with the disorder predictions. Furthermore, changes in the shape and conformational space of the protein, induced by the increase in the ionic strength, have been observed, emphasizing a more flexible and extended conformation. The salt-induced fuzzier conformational landscape, more globular shape, and unstructured conformation observed are characteristic of polyampholyte IDPs.

The regions which are mostly impacted by the increase in salt concentration are those mostly composed of charged and hydrophilic residues, i.e. LCD2, IDR2, G-patch, and LCD5; coherently with their intrinsic flexible nature. According to secondary structure timeline analysis, LCDs show a tendency to lose their specific secondary structure by increasing the random coil content in high salinity conditions. Three other regions comprising IDR2, IDR3, and the G-patch follow the same pattern. However, they are characterised by a more global modification in the secondary structure composition. These conformational changes also correlate with a larger LCD and G-patch accessibility, which could enforce the LLPS segregation and protein-RNA interaction. Indeed, the expanded conformation and more accessible domains (especially LCDs) within polyampholyte IDPs are known to allow homotypic IDP-IDP interactions that further drive LLPS formation, and may, thus, be strongly related to the biological role of TFIP11. IDPs involved in multiple biological functions like TFIP11 are regulated by PTMs. The presence of salt modifies the environment and the interactions of the residues implied in PTMs such as K155, R166, S59 and S98. We observe

that K155 gains in accessibility which may favour interactions with RNA and/or facilitate the addition of PTM, such as acetylation, to dynamically regulate RNA binding. The R166 residue participates in bringing the G-patch and LCD1 (two types of domain found in RNA-processing proteins) closer together. The presence of salt also favours the interactions of serine residues S98 and S59 with the positively charged residues K103 and R70 respectively. This suggests that positively charged R or K residues, known to be involved in interactions with nucleic acids, could be attracted by a local negative charge introduced by the addition of a phosphate group and form "arginine claw", a frequently found feature in splicing factor.

Since LLPS formation and protein-RNA interactions depends on many factors, *i.e.* pH, temperature, RNA, salt concentration, and post-translational modifications, further studies considering these parameters will allow to better understand their impact on the structure and the biological functions of TFIP11 N-TER.

Funding and acknowledgements

B. J. and J. M. thank the Belgian National Fund for Scientific Research (F.R.S.-FNRS) for their FRIA (Fund for Research training in Industry and Agriculture) Doctoral grant and Research Fellow fellowship, respectively. A.G and A.M.O. are PhD students supported by University of Liège (Fonds Spéciaux Recherche). The authors are appreciative to the PTCI high-performance computing resource of the University of Namur. The present research benefited from computational resources provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the FNRS under grant n°2.5020.11 and by the Walloon Region, and made available on Lucia, the Tier-1 supercomputer of the Walloon Region, infrastructure funded by the Walloon Region under the grant agreement n°1910247. C. M. and D. M. also thank the FNRS for their Senior Research Associate position. A. M. thanks ANR and CGI for their financial support of this work through Labex SEAM ANR 11 LABEX 086,

ANR 11 IDEX 05 02. The support of the IdEx "Université Paris 2019" ANR-18-IDEX-0001 is also acknowledged. B. J. thanks Younes Bourenane Cherif for his precious help in designing the figures.

References

(1) Ivanova, O. M.; Anufrieva, K. S.; Kazakova, A. N.; Malyants, I. K.; Shnaider, P. V.; Lukina, M. M.; Shender, V. O. Non-Canonical Functions of Spliceosome Components in Cancer Progression. *Cell Death Dis* **2023**, *14* (2), 77. https://doi.org/10.1038/s41419-022-05470-9.

(2) Yang, H.; Beutler, B.; Zhang, D. Emerging Roles of Spliceosome in Cancer and Immunity. *Protein Cell* **2022**, *13* (8), 559–579. https://doi.org/10.1007/s13238-021-00856-5.

(3) Hsu, T. Y.-T.; Simon, L. M.; Neill, N. J.; Marcotte, R.; Sayad, A.; Bland, C. S.; Echeverria, G. V.; Sun, T.; Kurley, S. J.; Tyagi, S.; Karlin, K. L.; Dominguez-Vidaña, R.; Hartman, J. D.; Renwick, A.; Scorsone, K.; Bernardi, R. J.; Skinner, S. O.; Jain, A.; Orellana, M.; Lagisetti, C.; Golding, I.; Jung, S. Y.; Neilson, J. R.; Zhang, X. H.-F.; Cooper, T. A.; Webb, T. R.; Neel, B. G.; Shaw, C. A.; Westbrook, T. F. The Spliceosome Is a Therapeutic Vulnerability in MYC-Driven Cancer. *Nature* **2015**, *525* (7569), 384–388. https://doi.org/10.1038/nature14985.

(4) Dvinge, H.; Guenthoer, J.; Porter, P. L.; Bradley, R. K. RNA Components of the Spliceosome Regulate Tissue- and Cancer-Specific Alternative Splicing. *Genome Res* **2019**, *29* (10), 1591–1604. https://doi.org/10.1101/gr.246678.118.

(5) An, J.; Luo, Z.; An, W.; Cao, D.; Ma, J.; Liu, Z. Identification of Spliceosome Components Pivotal to Breast Cancer Survival. *RNA Biol* **2021**, *18* (6), 833–842. https://doi.org/10.1080/15476286.2020.1822636.

(6) Oltean, S.; Bates, D. O. Hallmarks of Alternative Splicing in Cancer. *Oncogene* **2014**, *33* (46), 5311–5318. https://doi.org/10.1038/onc.2013.533.

(7) Coltri, P. P.; dos Santos, M. G. P.; da Silva, G. H. G. Splicing and Cancer: Challenges and Opportunities. *WIREs RNA* **2019**, *10* (3). https://doi.org/10.1002/wrna.1527.

(8) Uversky, V. N. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Front Phys* **2019**, *7*, 10. https://doi.org/10.3389/fphy.2019.00010.

(9) Duchemin, A.; O’Grady, T.; Hanache, S.; Mereau, A.; Thiry, M.; Wacheul, L.; Michaux, C.; Perpète, E.; Hervouet, E.; Peixoto, P.; Ernst, F. G. M.; Audic, Y.; Dequiedt, F.; Lafontaine, D. L. J.; Mottet, D. DHX15-Independent Roles for TFIP11 in U6 SnRNA Modification, U4/U6.U5 Tri-SnRNP Assembly and Pre-MRNA Splicing Fidelity. *Nat Commun* **2021**, *12* (1), 1–20. https://doi.org/10.1038/s41467-021-26932-2.

(10) Deckert, J.; Hartmuth, K.; Boehringer, D.; Behzadnia, N.; Will, C. L.; Kastner, B.; Stark, H.; Urlaub, H.; Lührmann, R. Protein Composition and Electron Microscopy Structure of Affinity-

Purified Human Spliceosomal B Complexes Isolated under Physiological Conditions. *Mol Cell Biol* **2006**, *26* (14), 5528–5543. https://doi.org/10.1128/MCB.00582-06.

(11) Mouffok, S.; Capeyrou, R.; Belhabich-Baumas, K.; Joret, C.; Henras, A. K.; Humbert, O.; Henry, Y. The G-Patch Activators Pfa1 and PINX1 Exhibit Different Modes of Interaction with the Prp43 RNA Helicase. *RNA Biol* **2021**, *18* (4), 510–522. https://doi.org/10.1080/15476286.2020.1818458.

(12) Ahn, E. E.-Y.; Higashi, T.; Yan, M.; Matsuura, S.; Hickey, C. J.; Lo, M.-C.; Shia, W.-J.; DeKelver, R. C.; Zhang, D.-E. SON Protein Regulates GATA-2 through Transcriptional Control of the MicroRNA 23a∼27a∼24-2 Cluster*. *Journal of Biological Chemistry* **2013**, *288* (8), 5381–5388. https://doi.org/10.1074/jbc.M112.447227.

(13) Aksaas, A. K.; Larsen, A. C. V; Rogne, M.; Rosendal, K.; Kvissel, A.-K.; Skålhegg, B. S. G-Patch Domain and KOW Motifs-Containing Protein, GPKOW; a Nuclear RNA-Binding Protein Regulated by Protein Kinase A. *J Mol Signal* **2011**, *6*, 10. https://doi.org/10.1186/1750-2187-6-10.

(14) Robert-Paganin, J.; Réty, S.; Leulliot, N. Regulation of DEAH/RHA Helicases by G-Patch Proteins. *Biomed Res Int* **2015**, *2015*, 1–9. https://doi.org/10.1155/2015/931857.

(15) Bohnsack, K. E.; Ficner, R.; Bohnsack, M. T.; Jonas, S. Regulation of DEAH-Box RNA Helicases by G-Patch Proteins. *Biol Chem* **2021**, *402* (5), 561–579. https://doi.org/10.1515/hsz-2020-0338.

(16) Wen, X.; Tannukit, S.; Paine, M. L. TFIP11 Interacts with MDEAH9, an RNA Helicase Involved in Spliceosome Disassembly. *Int J Mol Sci* **2008**, *9* (11), 2105–2113. https://doi.org/10.3390/ijms9112105.

(17) Cascarina, S. M.; Elder, M. R.; Ross, E. D. Atypical Structural Tendencies among Low-Complexity Domains in the Protein Data Bank Proteome. *PLoS Comput Biol* **2020**, *16* (1). https://doi.org/10.1371/journal.pcbi.1007487.

(18) Selig, E. E.; Bhura, R.; White, M. R.; Akula, S.; Hoffman, R. D.; Tovar, C. N.; Xu, X.; Booth, R. E.; Libich, D. S. Biochemical and Biophysical Characterization of the Nucleic Acid Binding Properties of the RNA/DNA Binding Protein EWS. *Biopolymers* **2023**, *114* (5). https://doi.org/10.1002/bip.23536.

(19) Lye, Y. S.; Chen, Y. TAR DNA-binding Protein 43 Oligomers in Physiology and Pathology. *IUBMB Life* **2022**, *74* (8), 794–811. https://doi.org/10.1002/iub.2603.

(20) Lee, J.; Cho, H.; Kwon, I. Phase Separation of Low-Complexity Domains in Cellular Function and Disease. *Exp Mol Med* **2022**, *54* (9), 1412–1422. https://doi.org/10.1038/s12276-022-00857-2.

(21) Fernández-Alvarez, A. J.; Thomas, M. G.; Giudice, J.; Boccaccio, G. L. Active Regulation Mechanisms of LLPS and MLOs Biogenesis. In *Droplets of Life*; Elsevier, 2023; pp 337–373. https://doi.org/10.1016/B978-0-12-823967-4.00005-1.

(22) Sehgal, P. B.; Westley, J.; Lerea, K. M.; DiSenso-Browne, S.; Etlinger, J. D. Biomolecular Condensates in Cell Biology and Virology: Phase-Separated Membraneless Organelles (MLOs). *Anal Biochem* **2020**, *597*, 113691. https://doi.org/10.1016/j.ab.2020.113691.

(23) Wohl, S.; Jakubowski, M.; Zheng, W. Salt-Dependent Conformational Changes of Intrinsically Disordered Proteins. *J Phys Chem Lett* **2021**, *12* (28), 6684–6691. https://doi.org/10.1021/acs.jpclett.1c01607.

(24) Maity, H.; Baidya, L.; Reddy, G. Salt-Induced Transitions in the Conformational Ensembles of Intrinsically Disordered Proteins. *Journal of Physical Chemistry B* **2022**, *126* (32), 5959–5971. https://doi.org/10.1021/acs.jpcb.2c03476.

(25) Williams, R. M.; Obradovi, Z.; Mathura, V.; Braun, W.; Garner, E. C.; Takayama, S.; Brown, C. J.; Dunker, A. K. The Protein Non-Folding Problem : Amino Acid Determinants of Intrinsic Order and Disorder. *Pac Symp Biocomp.* **2001**, 89–100.

(26) Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C. J.; Dunker, A. K.; Obradovic, Z. Optimizing Long Intrinsic Disorder Predictors With Protein Evolutionary Information. *J Bioinform Comput Biol* **2005**, *3* (1), 35–60. https://doi.org/10.1142/s0219720005000886.

(27) Garner; Romero; Dunker; Brown; Obradovic. Predicting Binding Regions within Disordered Proteins. *Genome Inform Ser Workshop Genome Inform* **1999**, *10*, 41–50.

(28) Erdos, G.; Pajkos, M.; Dosztányi, Z. IUPred3: Prediction of Protein Disorder Enhanced with Unambiguous Experimental Annotation and Visualization of Evolutionary Conservation. *Nucleic Acids Res* **2021**, *49*, 297–303. https://doi.org/10.1093/nar/gkab408.

(29) Ishida, T.; Kinoshita, K. PrDOS: Prediction of Disordered Protein Regions from Amino Acid Sequence. *Nucleic Acids Res* **2007**, *35*, 460–464. https://doi.org/10.1093/nar/gkm363.

(30) Walsh, I.; Martin, A. J. M.; Di domenico, T.; Tosatto, S. C. E. Espritz: Accurate and Fast Prediction of Protein Disorder. *Bioinformatics* **2012**, *28* (4), 503–509. https://doi.org/10.1093/bioinformatics/btr682.

(31) Dass, R.; Mulder, F. A. A.; Nielsen, J. T. ODiNPred: Comprehensive Prediction of Protein Order and Disorder. *Sci Rep* **2020**, *10* (1). https://doi.org/10.1038/s41598-020-71716-1.

(32) Emenecker, R. J.; Griffith, D.; Holehouse, A. S. Metapredict: A Fast, Accurate, and Easy-to-Use Predictor of Consensus Disorder and Structure. *Biophys J* **2021**, *120* (20), 4312–4319. https://doi.org/10.1016/j.bpj.2021.08.039.

(33) Bernhofer, M.; Dallago, C.; Karl, T.; Satagopam, V.; Heinzinger, M.; Littmann, M.; Olenyi, T.; Qiu, J.; Schütze, K.; Yachdav, G.; Ashkenazy, H.; Ben-Tal, N.; Bromberg, Y.; Goldberg, T.; Kajan, L.; O'Donoghue, S.; Sander, C.; Schafferhans, A.; Schlessinger, A.; Vriend, G.; Mirdita, M.; Gawron, P.; Gu, W.; Jarosz, Y.; Trefois, C.; Steinegger, M.; Schneider, R.; Rost, B. PredictProtein - Predicting Protein Structure and Function for 29 Years. *Nucleic Acids Res* **2021**, *49* (1), 535–540. https://doi.org/10.1093/nar/gkab354.

(34) Høie, M. H.; Kiehl, E. N.; Petersen, B.; Nielsen, M.; Winther, O.; Nielsen, H.; Hallgren, J.; Marcatili, P. NetSurfP-3.0: Accurate and Fast Prediction of Protein Structural Features by Protein Language Models and Deep Learning. *Nucleic Acids Res* **2022**, *50*, 510–515. https://doi.org/10.1093/nar/gkac439.

(35) Hu, G.; Katuwawala, A.; Wang, K.; Wu, Z.; Ghadermarzi, S.; Gao, J.; Kurgan, L. FlDPnn: Accurate Intrinsic Disorder Prediction with Putative Propensities of Disorder Functions. *Nat Commun* **2021**, *12* (1). https://doi.org/10.1038/s41467-021-24773-7.

(36) Orlando, G.; Raimondi, D.; Codicè, F.; Tabaro, F.; Vranken, W. Prediction of Disordered Regions in Proteins with Recurrent Neural Networks and Protein Dynamics. *J Mol Biol* **2022**, *434* (12). https://doi.org/10.1016/j.jmb.2022.167579.

(37) Dayhoff, G. W.; Uversky, V. N. Rapid Prediction and Analysis of Protein Intrinsic Disorder. *Protein Science* **2022**, *31* (12), 61a. https://doi.org/10.1002/pro.4496.

(38) Xue, B.; Dunbrack, R. L.; Williams, R. W.; Dunker, A. K.; Uversky, V. N. PONDR-FIT: A Meta-Predictor of Intrinsically Disordered Amino Acids. *Biochim Biophys Acta Proteins Proteom* **2010**, *1804* (4), 996–1010. https://doi.org/10.1016/j.bbapap.2010.01.011.

(39) Holehouse, A. S.; Das, R. K.; Ahad, J. N.; Richardson, M. O. G.; Pappu, R. V. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophys J* **2017**, *112* (1), 16–21. https://doi.org/10.1016/j.bpj.2016.11.3200.

(40) Malhis, N.; Jacobson, M.; Gsponer, J. MoRFchibi SYSTEM: Software Tools for the Identification of MoRFs in Protein Sequences. *Nucleic Acids Res* **2016**, *44* (W1), W488–W493. https://doi.org/10.1093/nar/gkw409.

(41) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

(42) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J Chem Theory Comput* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(43) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J Phys Chem Lett* **2014**, *5* (21), 3863–3871. https://doi.org/10.1021/jz501780a.

(44) Mu, J.; Liu, H.; Zhang, J.; Luo, R.; Chen, H.-F. Recent Force Field Strategies for Intrinsically Disordered Proteins. *J Chem Inf Model* **2021**, *61* (3), 1037–1047. https://doi.org/10.1021/acs.jcim.0c01175.

(45) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput* **2015**, *11* (4), 1864–1874. https://doi.org/10.1021/ct5010406.

(46) Petersen, H. G. Accuracy and Efficiency of the Particle Mesh Ewald Method. *J Chem Phys* **1995**, *103* (9), 3668–3679. https://doi.org/10.1063/1.470043.

(47) Yuan, S.; Chan, H. C. S.; Hu, Z. Using PyMOL as a Platform for Computational Drug Design. *WIREs Computational Molecular Science* **2017**, *7* (2). https://doi.org/10.1002/wcms.1298.

(48) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J Mol Graph* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

(49) Hansen, J. C.; Lu, X.; Ross, E. D.; Woody, R. W. Intrinsic Protein Disorder, Amino Acid Composition, and Histone Terminal Domains. *Journal of Biological Chemistry* **2006**, *281* (4), 1853–1856. https://doi.org/10.1074/jbc.R500022200.

(50) Xue, B.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N. CDF It All: Consensus Prediction of Intrinsically Disordered Proteins Based on Various Cumulative Distribution Functions. *FEBS Lett* **2009**, *583* (9), 1469–1474. https://doi.org/10.1016/j.febslet.2009.03.070.

(51) Xue, B.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N. CDF It All: Consensus Prediction of Intrinsically Disordered Proteins Based on Various Cumulative Distribution Functions. *FEBS Lett* **2009**, *583* (9), 1469–1474. https://doi.org/10.1016/j.febslet.2009.03.070.

(52) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why Are "natively Unfolded" Proteins Unstructured under Physiologic Conditions? *Proteins: Structure, Function and Genetics* **2000**, *41* (3), 415–427. https://doi.org/10.1002/1097-0134(20001115)41:3<415::AID-PROT130>3.0.CO;2-7.

(53) Das, R. K.; Pappu, R. V. Conformations of Intrinsically Disordered Proteins Are Influenced by Linear Sequence Distributions of Oppositely Charged Residues. *Proc Natl Acad Sci U S A* **2013**, *110* (33), 13392–13397. https://doi.org/10.1073/pnas.1304749110.

(54) Das, R. K.; Ruff, K. M.; Pappu, R. V. Relating Sequence Encoded Information to Form and Function of Intrinsically Disordered Proteins. *Curr Opin Struct Biol* **2015**, *32*, 102–112. https://doi.org/10.1016/j.sbi.2015.03.008.

(55) Martin, E. W.; Holehouse, A. S.; Grace, C. R.; Hughes, A.; Pappu, R. V.; Mittag, T. Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation. *J Am Chem Soc* **2016**, *138* (47), 15323–15335. https://doi.org/10.1021/jacs.6b10272.

(56) Morris, O. M.; Torpey, J. H.; Isaacson, R. L. Intrinsically Disordered Proteins: Modes of Binding with Emphasis on Disordered Domains. *Open Biol* **2021**, *11* (10). https://doi.org/10.1098/rsob.210222.

(57) Tannukit, S.; Crabb, T. L.; Hertel, K. J.; Wen, X.; Jans, D. A.; Paine, M. L. Identification of a Novel Nuclear Localization Signal and Speckle-Targeting Sequence of Tuftelin-Interacting Protein 11, a Splicing Factor Involved in Spliceosome Disassembly. *Biochem Biophys Res Commun* **2009**, *390* (3), 1044–1050. https://doi.org/10.1016/j.bbrc.2009.10.111.

(58) Studer, M. K.; Ivanović, L.; Weber, M. E.; Marti, S.; Jonas, S. Structural Basis for DEAH-Helicase Activation by G-Patch Proteins. *Proceedings of the National Academy of Sciences* **2020**, *117* (13), 7159–7170. https://doi.org/10.1073/pnas.1913880117.

(59) Ivankov, D. N.; Bogatyreva, N. S.; Lobanov, M. Y.; Galzitskaya, O. V. Coupling between Properties of the Protein Shape and the Rate of Protein Folding. *PLoS One* **2009**, *4* (8). https://doi.org/10.1371/journal.pone.0006476.

(60) Bianchi, G.; Longhi, S.; Grandori, R.; Brocca, S. Relevance of Electrostatic Charges in Compactness, Aggregation, and Phase Separation of Intrinsically Disordered Proteins. *Int J Mol Sci* **2020**, *21* (17), 6208. https://doi.org/10.3390/ijms21176208.

(61) Maity, H.; Baidya, L.; Reddy, G. Salt-Induced Transitions in the Conformational Ensembles of Intrinsically Disordered Proteins. *Journal of Physical Chemistry B* **2022**, *126* (32), 5959–5971. https://doi.org/10.1021/acs.jpcb.2c03476.

(62) Majumdar, A.; Dogra, P.; Maity, S.; Mukhopadhyay, S. Liquid-Liquid Phase Separation Is Driven by Large-Scale Conformational Unwinding and Fluctuations of Intrinsically Disordered

Protein Molecules. *Journal of Physical Chemistry Letters* **2019**, *10* (14), 3929–3936. https://doi.org/10.1021/acs.jpclett.9b01731.

(63) Dong, X.; Bera, S.; Qiao, Q.; Tang, Y.; Lao, Z.; Luo, Y.; Gazit, E.; Wei, G. Liquid-Liquid Phase Separation of Tau Protein Is Encoded at the Monomeric Level. *Journal of Physical Chemistry Letters* **2021**, *12* (10), 2576–2586. https://doi.org/10.1021/acs.jpclett.1c00208.

(64) Studer, M. K.; Ivanović, L.; Weber, M. E.; Marti, S.; Jonas, S. Structural Basis for DEAH-Helicase Activation by G-Patch Proteins. *Proceedings of the National Academy of Sciences* **2020**, *117* (13), 7159–7170. https://doi.org/10.1073/pnas.1913880117.

(65) Christian, H.; Hofele, R. V.; Urlaub, H.; Ficner, R. Insights into the Activation of the Helicase Prp43 by Biochemical Studies and Structural Mass Spectrometry. *Nucleic Acids Res* **2014**, *42* (2), 1162–1179. https://doi.org/10.1093/nar/gkt985.

(66) Hornbeck, P. V.; Kornhauser, J. M.; Tkachev, S.; Zhang, B.; Skrzypek, E.; Murray, B.; Latham, V.; Sullivan, M. PhosphoSitePlus: A Comprehensive Resource for Investigating the Structure and Function of Experimentally Determined Post-Translational Modifications in Man and Mouse. *Nucleic Acids Res* **2012**, *40* (D1), D261–D270. https://doi.org/10.1093/nar/gkr1122.

(67) Lundby, A.; Lage, K.; Weinert, B. T.; Bekker-Jensen, D. B.; Secher, A.; Skovgaard, T.; Kelstrup, C. D.; Dmytriyev, A.; Choudhary, C.; Lundby, C.; Olsen, J. V. Proteomic Analysis of Lysine Acetylation Sites in Rat Tissues Reveals Organ Specificity and Subcellular Patterns. *Cell Rep* **2012**, *2* (2), 419–431. https://doi.org/10.1016/j.celrep.2012.07.006.

(68) Tannukit, S.; Crabb, T. L.; Hertel, K. J.; Wen, X.; Jans, D. A.; Paine, M. L. Identification of a Novel Nuclear Localization Signal and Speckle-Targeting Sequence of Tuftelin-Interacting Protein 11, a Splicing Factor Involved in Spliceosome Disassembly. *Biochem Biophys Res Commun* **2009**, *390* (3), 1044–1050. https://doi.org/10.1016/j.bbrc.2009.10.111.

(69) Švec, M.; Bauerová, H.; Pichová, I.; Konvalinka, J.; Stříšovský, K. Proteinases of Betaretroviruses Bind Single-stranded Nucleic Acids through a Novel Interaction Module, the G-patch. *FEBS Lett* **2004**, *576* (1–2), 271–276. https://doi.org/10.1016/j.febslet.2004.09.010.

(70) Nishi, H.; Fong, J. H.; Chang, C.; Teichmann, S. A.; Panchenko, A. R. Regulation of Protein–Protein Binding by Coupling between Phosphorylation and Intrinsic Disorder: Analysis of Human Protein Complexes. *Mol Biosyst* **2013**, *9* (7), 1620. https://doi.org/10.1039/c3mb25514j.

(71) Guo, A.; Gu, H.; Zhou, J.; Mulhern, D.; Wang, Y.; Lee, K. A.; Yang, V.; Aguiar, M.; Kornhauser, J.; Jia, X.; Ren, J.; Beausoleil, S. A.; Silva, J. C.; Vemulapalli, V.; Bedford, M. T.; Comb, M. J. Immunoaffinity Enrichment and Mass Spectrometry Analysis of Protein Methylation. *Molecular & Cellular Proteomics* **2014**, *13* (1), 372–387. https://doi.org/10.1074/mcp.O113.027870.

(72) Larsen, S. C.; Sylvestersen, K. B.; Mund, A.; Lyon, D.; Mullari, M.; Madsen, M. V.; Daniel, J. A.; Jensen, L. J.; Nielsen, M. L. Proteome-Wide Analysis of Arginine Monomethylation Reveals Widespread Occurrence in Human Cells. *Sci Signal* **2016**, *9* (443). https://doi.org/10.1126/scisignal.aaf7329.

(73) Musiani, D.; Bok, J.; Massignani, E.; Wu, L.; Tabaglio, T.; Ippolito, M. R.; Cuomo, A.; Ozbek, U.; Zorgati, H.; Ghoshdastider, U.; Robinson, R. C.; Guccione, E.; Bonaldi, T. Proteomics Profiling

of Arginine Methylation Defines PRMT5 Substrate Specificity. *Sci Signal* **2019**, *12* (575). https://doi.org/10.1126/scisignal.aat8388.

(74) Nadassy, K.; Wodak, S. J.; Janin, J. Structural Features of Protein−Nucleic Acid Recognition Sites. *Biochemistry* **1999**, *38* (7), 1999–2017. https://doi.org/10.1021/bi982362d.

(75) Thapar, R. Structural Basis for Regulation of RNA-Binding Proteins by Phosphorylation. *ACS Chem Biol* **2015**, *10* (3), 652–666. https://doi.org/10.1021/cb500860x.

(76) Chatrikhi, R.; Wang, W.; Gupta, A.; Loerch, S.; Maucuer, A.; Kielkopf, C. L. SF1 Phosphorylation Enhances Specific Binding to U2AF 65 and Reduces Binding to 3'-Splice-Site RNA. *Biophys J* **2016**, *111* (12), 2570–2586. https://doi.org/10.1016/j.bpj.2016.11.007.

(77) Van Roey, K.; Uyar, B.; Weatheritt, R. J.; Dinkel, H.; Seiler, M.; Budd, A.; Gibson, T. J.; Davey, N. E. Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chem Rev* **2014**, *114* (13), 6733–6778. https://doi.org/10.1021/cr400585q.

(78) Kumar, M.; Michael, S.; Alvarado-Valverde, J.; Mészáros, B.; Sámano-Sánchez, H.; Zeke, A.; Dobson, L.; Lazar, T.; Örd, M.; Nagpal, A.; Farahi, N.; Käser, M.; Kraleti, R.; Davey, N. E.; Pancsa, R.; Chemes, L. B.; Gibson, T. J. The Eukaryotic Linear Motif Resource: 2022 Release. *Nucleic Acids Res* **2022**, *50* (D1), D497–D508. https://doi.org/10.1093/nar/gkab975.

(79) Parolini, F.; Tira, R.; Barracchia, C. G.; Munari, F.; Capaldi, S.; D'Onofrio, M.; Assfalg, M. Ubiquitination of Alzheimer's-Related Tau Protein Affects Liquid-Liquid Phase Separation in a Site- and Cofactor-Dependent Manner. *Int J Biol Macromol* **2022**, *201*, 173–181. https://doi.org/10.1016/j.ijbiomac.2021.12.191.

(80) Dao, T. P.; Kolaitis, R.-M.; Kim, H. J.; O'Donovan, K.; Martyniak, B.; Colicino, E.; Hehnly, H.; Taylor, J. P.; Castañeda, C. A. Ubiquitin Modulates Liquid-Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions. *Mol Cell* **2018**, *69* (6), 965-978.e6. https://doi.org/10.1016/j.molcel.2018.02.004.

(81) Li, J.; Zhu, K.; Gu, A.; Zhang, Y.; Huang, S.; Hu, R.; Hu, W.; Lei, Q.-Y.; Wen, W. Feedback Regulation of Ubiquitination and Phase Separation of HECT E3 Ligases. *Proceedings of the National Academy of Sciences* **2023**, *120* (33). https://doi.org/10.1073/pnas.2302478120.

(82) Zhang, L.; Zhang, D.; Liu, D.; Li, Y.; Li, H.; Xie, Y.; Wang, Z.; Hansen, B. O.; Olsson, S. Conserved Eukaryotic Kinase CK2 Chaperone Intrinsically Disordered Protein Interactions. *Appl Environ Microbiol* **2020**, *86* (2). https://doi.org/10.1128/AEM.02191-19.

(83) Tsang, B.; Arsenault, J.; Vernon, R. M.; Lin, H.; Sonenberg, N.; Wang, L.-Y.; Bah, A.; Forman-Kay, J. D. Phosphoregulated FMRP Phase Separation Models Activity-Dependent Translation through Bidirectional Control of MRNA Granule Formation. *Proceedings of the National Academy of Sciences* **2019**, *116* (10), 4218–4227. https://doi.org/10.1073/pnas.1814385116.