Quantitative prediction of ensemble dynamics, shapes and contact propensities of intrinsically disordered proteins

Lei Yu[1] and Rafael Brüschweiler[1,2*]

[1]Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio, United States of America

[2]Department of Biological Chemistry and Pharmacology, The Ohio State University, Columbus, Ohio, United States of America

[*] Corresponding author

E-mail: bruschweiler.1@osu.edu (RB)

Short title: "Quantitative prediction of structural dynamics of IDPs"

ABSTRACT

Intrinsically disordered proteins (IDPs) are highly dynamic systems that play an important role in cell signaling processes and their misfunction often causes human disease. Proper understanding of IDP function not only requires the realistic characterization of their three-dimensional conformational ensembles at atomic-level resolution but also of the time scales of interconversion between their conformational substates. Large sets of experimental data are often used in combination with molecular modeling to restrain or bias models to improve agreement with experiment. It is shown here for the N-terminal transactivation domain of p53 (p53TAD) and Pup, which are two IDPs that fold upon binding to their targets, how the latest advancements in molecular dynamics (MD) simulations methodology produces native conformational ensembles by combining replica exchange with series of microsecond MD simulations. They closely reproduce experimental data at the global conformational ensemble level, in terms of the distribution properties of the radius of gyration tensor, and at the local level, in terms of NMR properties including ^{15}N spin relaxation, without the need for reweighting. Further inspection revealed that 10–20% of the individual MD trajectories display the formation of secondary structures not observed in the experimental NMR data. The IDP ensembles were analyzed by graph theory to identify dominant inter-residue contact clusters and characteristic amino-acid contact propensities. These findings indicate that modern MD force fields with residue-specific backbone potentials can produce highly realistic IDP ensembles sampling a hierarchy of nano- and picosecond time scales providing new insights into their biological function.

AUTHOR SUMMARY

Accurate prediction of the conformational ensemble dynamics sans bias is shown for intrinsically disordered proteins including the transactivation domain of p53.

INTRODUCTION

Intrinsically disordered proteins (IDPs) and protein regions (IDRs) are an integral part of the proteomes of many different organisms with more than 30% of all eukaryotic proteins possessing 40 or more consecutive disordered residues.(1, 2) While IDPs and IDRs in isolation do not adopt well-defined three-dimensional (3D) structures, they often play important biological roles in molecular recognition processes by interacting in specific ways with binding partners that are typically well-ordered.(3-5) For instance, the human oncoprotein protein p53 possesses the N-terminal transactivation domain (p53TAD) that binds to the N-terminal domain of human MDM2 protein adopting a stable α-helix.(6) Prokaryotic ubiquitin-like protein (Pup) is another IDP that is directly linked to protein degradation folding into an α-helix when binding to Mpa protein.(7) In addition to binding to their target protein(s), IDPs can also be involved in liquid-liquid phase separation (LLPS).(8-11) LLPS is the segregation of molecules in solution into a condensed phase and a dilute phase with high and low biomolecular concentrations. These membraneless droplet-like compartments formed by IDPs and other biomolecules are important for cellular function. Knowledge of the structural and dynamic propensities of IDPs both in isolation and in complex biological environments is essential for understanding these processes and their role in human diseases.

In order to relate IDP sequences to biological function, detailed knowledge of IDP conformational ensembles is needed. The description of conformational ensembles can range from local secondary structure populations to explicit ensembles in 3D space with atomic resolution.(12) Some of the earliest approaches generate random coil conformational ensembles that are subsequently refined against a host of experimental data reflecting both local and global structural features.(13-15) These approaches continue to be successfully applied through integrative modeling provided that a large amount of high quality experimental data is available for each system under investigation.(16, 17) Even when data from various complementary experimental techniques are being used, the amount of experimental information obtainable is still sparse when compared to the information needed to uniquely characterize large, highly heterogeneous structural ensembles that are the hallmark of IDPs. As a consequence, the amount of information that can be gained and that is not directly reflected in the experimental data used to refine the ensemble is restricted to robust descriptors ranging from coarse-grained to global

that can be compared with predictions by polymer theory under various assumptions.(16) In addition, site-specific interaction information, such as transient inter-residue contacts, can be obtained at medium to low resolution from paramagnetic relaxation enhancement (PRE) experiments by attaching electron spin labels to selected sites.(15, 17) Because empirical ensembles generated based on such data lack a time axis, they do not include dynamics time scales of IDPs associated with interconversion rates between substates and, hence, they do not inform about an essential part of the energy landscape.

From a theoretical and computational perspective, all-atom molecular dynamics (MD) simulations are an attractive alternative to empirical approaches for the generation of IDP conformational ensembles, including dynamic time scale information, for the comprehensive interpretation of experimental results.(18) However, for many years limitations in computer power precluded the generation of statistically well-converged results and MD force fields primarily developed for ordered proteins turned out to be unsatisfactory for applications to IDPs. With the continuing increase in computer power, the quality of sampling has reached a level that allows rigorous validation by quantitative comparison with a rich body of experimental data. In cases where discrepancies are observed between simulation and experiment, as is commonly the case, approaches have been developed that use restraining or reweighting that bias the original simulation to obtain results that agree better with experimental data.(19-26) When not only the conformational ensemble but also the underlying dynamics time scales are of interest, suitable rescaling of the MD time step or correlation times of the dominant motional modes can be applied to improve agreement with experiment.(27-30) Because these methods can often improve the unaltered simulations only within certain boundaries, they are best suited when the original predictions are fairly close to experimental data.(31) Although these methods rarely fail to produce better agreement, at least on average for those experimental parameters directly used as restraints or for reweighting, they naturally depend on large amounts of experimental data of good quality as input for each protein system studied. This amounts to a laborious experimental effort that needs to be repeated for each new protein system as the experimental data are protein-specific rendering them non-transferrable between systems.

An alternative and more principled approach is to improve the MD force fields themselves enabling them to increasingly accurately predict experimental data in a way that is

fully transferrable between protein systems, both ordered and disordered. This premise has led to a recent proliferation of protein force field developments(32-37) and new explicit water models(38-40) specifically geared toward the improved representation of disordered proteins. In a significant development, residue-specific force fields have been introduced.(41) These force fields use in addition coil library information from the Protein Data Bank (PDB) by incorporating the individual backbone φ,ψ propensities of each residue type.(41-47) Such residue-specific force fields, in combination with suitable water models, can provide an improved representation of disordered states while retaining the properties of ordered proteins. With respect to water models, TIP4P-D and closely related derivatives have been notably successful in preventing overly compact conformations by favoring more extended IDP structures showing improved agreement with experiment.(38)

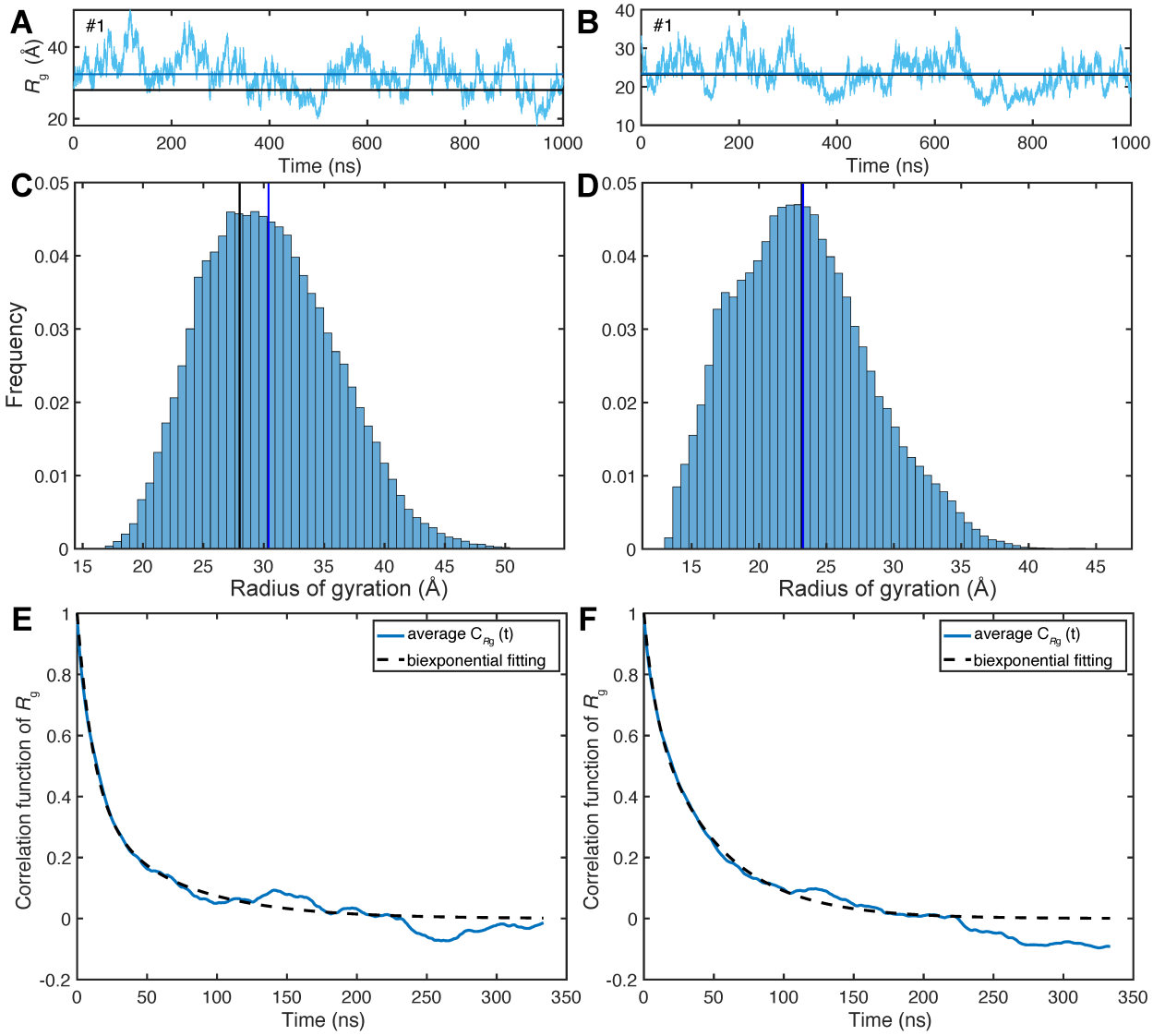
Besides global properties, such as the radii of gyration and asphericities, IDP ensembles and trajectories should also accurately reproduce local dihedral angle distributions and secondary structure propensities. Moreover, they should also replicate dynamic and kinetic IDP properties, such as librational motions and time scales of interconversion between conformational substates. Such information is important for understanding recognition events between IDPs and their binding targets, including IDP interactions with other disordered biomolecules, for example, during the formation of LLPS condensates. Experimental IDP dynamics information can be gained from fluorescence depolarization spectroscopy,(48) Förster resonance energy transfer (FRET),(16) and nuclear magnetic resonance (NMR) relaxation.(15) NMR ^{15}N longitudinal R_1 and transverse R_2 spin relaxation rates are exquisitely sensitive to the dynamics of disordered proteins and the underlying time scales.(49-51) R_2 relaxation rates, for example, have been linked to residual intramolecular interactions in chemically unfolded proteins.(51-53) ^{15}N R_1 and R_2 rates can be experimentally determined for each protein residue and therefore they are valuable for validating MD simulations with respect to amplitudes and time scales of IDP dynamics.(29, 54-56)

We recently developed the AMBER ff99SBnmr2 force field by modifying the backbone dihedral angle potentials of each amino-acid residue type to reproduce the φ,ψ dihedral angle distributions found in a random coil library.(57) The ff99SBnmr2 force field has been validated against experimental nuclear magnetic resonance (NMR) scalar ^3J-couplings of α-synuclein and

β-amyloid IDPs demonstrating that this force field accurately reproduces their sequence-dependent local backbone structural propensities.(58) The primary goal of this work is to learn whether state-of-the-art replica exchange and extended MD simulations of IDPs can also realistically reproduce NMR R_1, R_2 relaxation rates with their strong and unique dependence on motional time scales without the need of any additional corrections such as constraints or reweighting. Moreover, in-depth analysis of the MD trajectories generated yields a wealth of information about the radius of gyration tensor distribution and dominant dynamics modes allowing graph-theory based identification of specific inter-residue interaction propensities and residue clusters for the better understanding of IDP behavior.

RESULTS

Ensemble properties of radius of gyration tensor. The radius of gyration $R_g(t)$ is shown as a function of time for representative 1-μs MD trajectories of p53TAD and Pup in **Fig. 1A,B** (see also **Fig. S1**). The trajectories exhibit predominantly stationary stochastic behavior reflecting random expansion and contraction of the overall IDP size with the mean value (blue horizontal lines) in good agreement with the experimentally determined $\langle R_g \rangle$ (black line) or the predicted $\langle R_g \rangle$ from polymer theory (**Eq. 6**). The MD-distributions of R_g of all 10 MD trajectories are shown as histograms in **Fig. 1C,D**. The Flory exponent ν of the polymer scaling law was determined from the REMD ensembles at 298 K. Using ρ_0 = 1.927 Å, we obtain a value of ν = 0.601 for Pup, which closely matches the theoretical value ν_{theory} = 0.588 of a fully disordered, self-avoiding random coil.(59, 60) For p53TAD, the REMD $\langle R_g \rangle$ value of 28.1 Å is in almost perfect agreement with experiment(61) (28.0 Å) corresponding to ν = 0.624, which clearly exceeds ν_{theory}.



168

169 **Fig. 1.** Radius of gyration, R_g, properties of two IDPs p53TAD and Pup from microsecond MD
170 simulations. Time-dependence of $R_g(t)$ from representative 1-μs MD trajectories (cyan) of (A)
171 p53TAD and (B) Pup where the horizontal blue lines correspond to the mean R_g values
172 calculated from the trajectories and the black lines correspond to the experimentally determined
173 R_g for p53TAD and the predicted R_g according to polymer theory (**Eq. 6**) for Pup. R_g profiles for
174 all 10 1-μs trajectories of each protein are shown in Figure S1. Histograms of the $R_g(t)$
175 distributions over all 10 MD simulations are shown in Panels C, D (blue and black lines have the
176 same meaning as in Panels A, B). The standard deviation of R_g over all 10 MD trajectories is 5.4
177 Å for p53TAD and 5.0 Å for Pup. Offset-free time-correlation functions $C_{Rg}(t)$ of $R_g(t)$ averaged
178 over all 10 1-μs MD trajectories are shown for (E) p53TAD and (F) Pup. The dashed lines
179 belong to non-linear least squares fits of $C_{Rg}(t)$ by biexponential functions whereby the best fits
180 are obtained for p53TAD with τ_a = 12 ns (63% of total amplitude), τ_b = 62 ns (37%) and for Pup
181 with τ_a = 8 ns (29%), τ_b = 48 ns (71%).

The characteristic time scales of $R_g(t)$ fluctuations can be obtained from the time-correlation functions $C_{Rg}(t)$ (**Eq. 5**), which are well-converged over the course of the 1-μs trajectories (**Fig. 1E,F**). $C_{Rg}(t)$ of both proteins decay in good approximation biexponentially with reconfigurational correlation times $\tau_a \cong 10$ ns and $\tau_b \cong 55$ ns. The normalized variance of the $R_g(t)$ fluctuations, given by

$$\kappa^2_{Rg} = 1 - \left\langle R_g \right\rangle^2 \Big/ \left\langle R_g^2 \right\rangle \quad (1)$$

is almost the same for p53TAD (0.03) and Pup (0.04). The ensemble distribution of the gyration tensor $\boldsymbol{S}$ (**Eq. 2**) contains information about the deviation of individual MD snapshots from spherical shape, which can be directly compared with a random Gaussian chain serving as a perfect random coil (**Fig. 2**).(62) Both proteins show unimodal asphericity distributions (**Eq. 3**) with maxima around $A \cong 0.18$, which qualitatively differ from the Gaussian chain model (**Fig. 2C**) peaking at $A = 0$. Compared to p53TAD, Pup has a higher tendency to adopt a more spherical conformation. Another useful measure of the overall shape of individual snapshots is the prolateness P (**Eq. 4**). The distribution of P is bimodal for both proteins with the global maximum corresponding to prolate-shaped (cigar-like) structures ($P = 1$) and a second (local) maximum corresponding to disk-like structures ($P = -1$). The distribution of the prolateness of Pup is more balanced between positive and negative values with $\langle P \rangle = 0.2$ than for p53TAD, which has a higher tendency to adopt prolate-shaped conformers ($\langle P \rangle = 0.35$), whereas the Gaussian chain distribution ($\langle P \rangle = 0.3$) lies between the two IDP distributions. The distinct asphericity distribution and increased prolateness of p53TAD is at the origin of its increased $\langle R_g \rangle$ over the Gaussian random coil model.

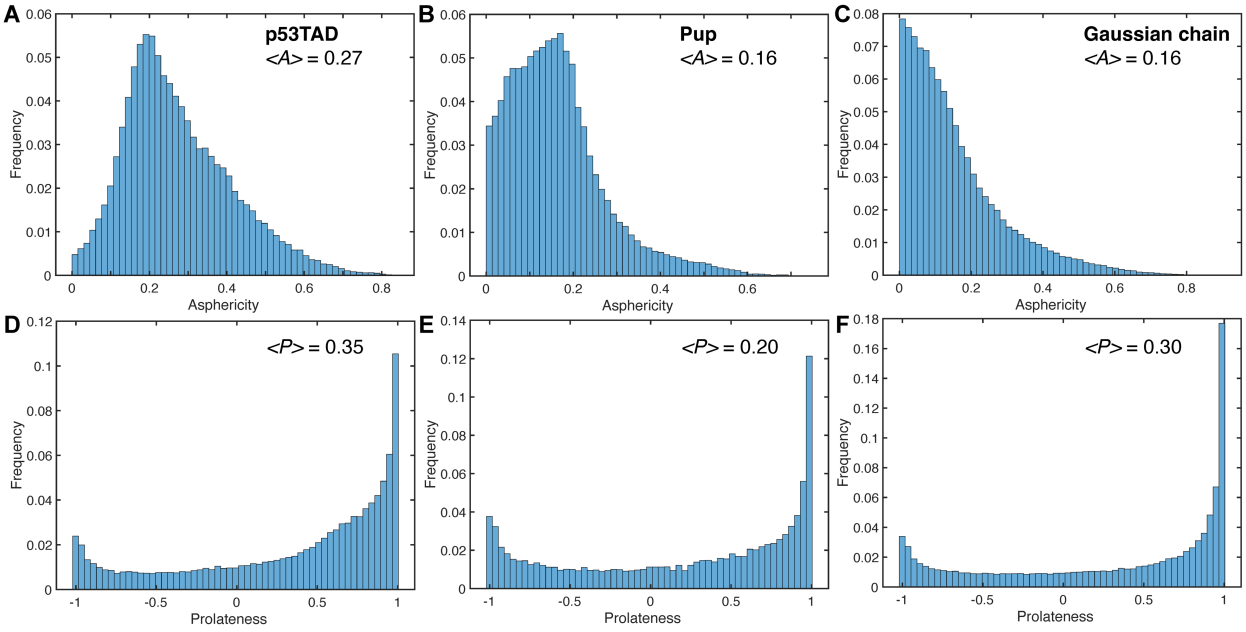


Fig. 2. Gyration tensor properties of IDP ensembles of p53TAD and Pup across 10 1-μs MD trajectories. The distributions of gyration tensor asphericities A are shown for (A) p53TAD and (B) Pup in comparison with (C) a Gaussian chain. The distributions of gyration tensor prolateness P are shown for (D) p53TAD and (E) Pup in comparison with a (F) Gaussian chain.

Validation against R_1, R_2 relaxation data. Experimental and computed ^{15}N R_1, R_2 relaxation rates are shown in **Fig. 3**. R_1 relaxation rates determined from simulations (**Eq. 7–12**) are in close agreement with experiment(63) evidenced by small RMSEs (0.10 s^{-1} for p53TAD and 0.12 s^{-1} for Pup) and Pearson correlation coefficients R of 0.78 for p53TAD and 0.86 for Pup (**Fig. 3A,B**). R_2 relaxation rates determined from the simulations are also in good agreement with experiment with correlation coefficients R of 0.88 for p53TAD and 0.70 for Pup and RMSEs of 0.84 s^{-1} for p53TAD and 0.81 s^{-1} for Pup and (**Fig. 3C,D)**. It can be seen that the simulations tend to underestimate R_1 and overestimate R_2 rates, although only slightly, in a manner that is notably uniform for the R_1 values of both proteins and for the R_2 values of p53TAD. The 10 N-terminal residues of p53TAD are very flexible with small R_2's, which closely follow the experiment. For Pup, differences in R_2 between MD and experiment display the same trend and are most pronounced for residues 30–48. The error bars of the computed relaxation rates, which represent the root-mean-square deviations over all 10 MD trajectories, are fairly uniform along the polypeptide chains and systematically larger for R_2 than for R_1, again with the exception of the

10 N-terminal residues of p53TAD. For both proteins, not all 10 1-μs MD trajectories individually reproduce the experimental data equally well. Either 1 (p53TAD) or 2 (Pup) trajectories have more compact average IDP structures, which quantitatively affect the agreement with experiment (**Fig. S2**).

Correlation times of backbone N-H bond vectors in both proteins fitted from the average correlation functions range from picoseconds to about 20 ns (**Fig. 3E,F**). Consistent with the finding for other IDPs,(55, 64) the dominant contribution to the time correlation functions stems from dynamics on the intermediate time scale around 1 ns reporting about backbone φ,ψ jumps. Fast dynamics on the time scale of 100 ps or faster report on local ^{15}N-^{1}H bond librations, similar to those observed in secondary structures of folded proteins,(65) and slower dynamics on the time scale between 3 and 20 ns reports on collective IDP chain motions. The presence of slower modes correlate with increased R_2 values most pronounced for residues 30–48 in Pup. This is consistent with relaxation theory (**Eq. 12**), which predicts that in solution transverse spin relaxation rates R_2 are in good approximation proportional to the effective overall correlation time experienced by the ^{15}N-^{1}H spin pairs.

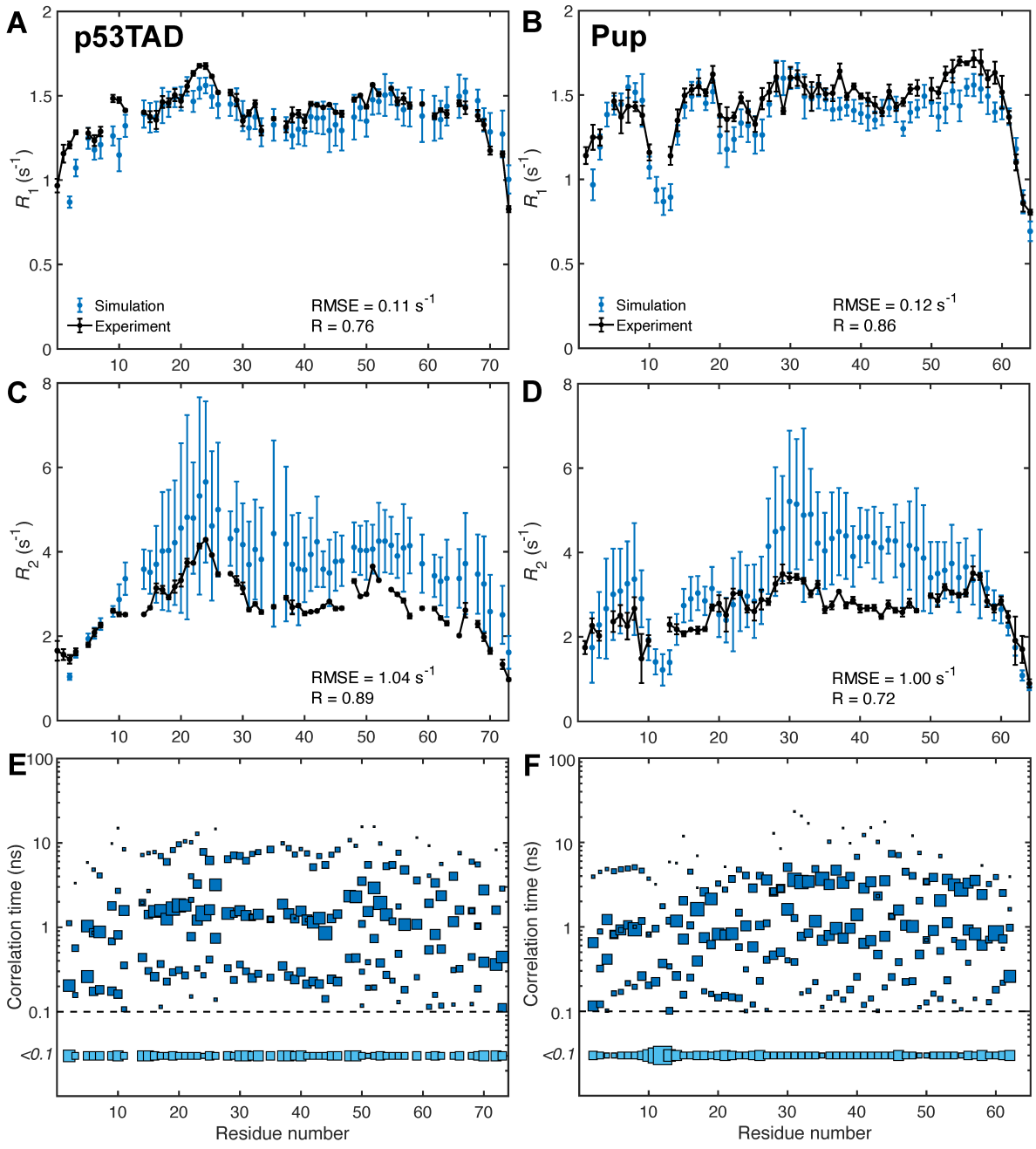


Fig. 3. Back-calculated R_1, R_2 NMR ^{15}N-spin relaxation rates in comparison with experiment along with underlying motional time scale distributions. R_1, R_2 rates calculated from average correlation functions are plotted in blue with error bars representing standard deviations across individual MD trajectories. Correlation time distribution of individual ^{15}N-^{1}H bonds of IDPs extracted from correlation functions for (E) p53TAD and (F) Pup where the size of the blue squares are proportional to the associated motional amplitudes A_i. The squares at the bottom indicate the aggregate of dynamics contributions with correlation times faster than 100 ps. Dominant dynamics time scales range from about 100 ps to about 10 ns depending on the residue, with the exception of Thr12 in Pup which exhibits dominant dynamics time scales faster than 100 ps.

Increased transverse NMR spin relaxation is indicative of the presence of collective segmental motions in IDPs, which are modulated by the formation of transient secondary structures and inter-residue side-chain interactions. To examine these relationships, instantaneous secondary structures and average contact maps were determined from the MD trajectories (**Fig. 4**). A contact is defined in an MD snapshot when the nearest distance between atoms from two different residues is smaller than 4 Å (uninformative first-neighbor (*i*,*i*+1) and second-neighbor (*i*,*i*+2) contacts between residues were excluded (white band along diagonal in **Fig. 4A,B**)). The most frequent contacts are relatively short range, but contacts over larger distances occur for p53TAD and even more frequently for Pup. Some contacts are linked to the transient formation of short secondary structures, α-helices and β-strands (**Fig. 4C,D**), whereas other regions display frequent contacts largely independent of secondary structure propensity often involving arginine residues, such as Arg65 of p53TAD and Arg28/29 and Arg56 of Pup. **Fig. 4C,D** also shows that selected trajectories possess regions with well above-average secondary structure propensities, such as trajectories #4 of p53TAD and trajectories #5 and #7 of Pup, which are the same trajectories that contribute to the lengthening of R_2 along parts of the polypeptide sequences mentioned above. Due to their atypical (outlier) nature, not representative of the other trajectories, they were not included in the following residue-cluster analysis. For p53TAD, regions that tend to form α-helices do not form β-strands and vice versa (except for trajectory #4). For Pup, on the other hand, a number of regions exist in its N-terminal half that can transiently switch between these two types of local secondary structures.

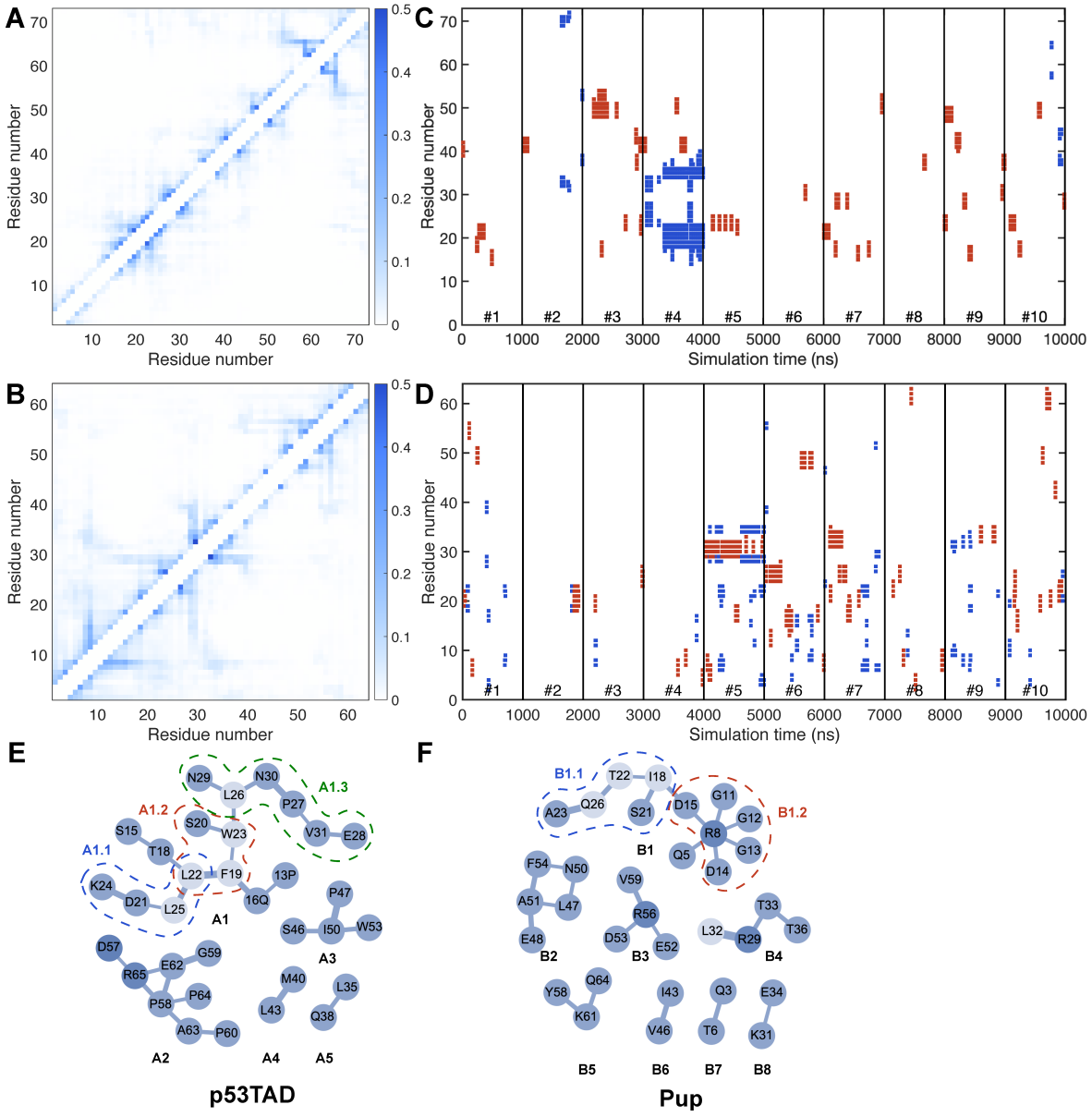


Fig. 4. Average IDP contact maps and time-dependent secondary structure formation of each residue. (A, B) Pairwise contact occupancies were determined from MD simulations (without outlier trajectories, **Fig. S2, Tables S4, S5**) for (A) p53TAD and (B) Pup. Darker/lighter shades of blue denote contacts that are more frequently/rarely formed according to legend (vertical bar). Self-contacts, first-neighbor contacts (between residues i,i+1), and second-neighbor contacts (between residues i,i+2) are not shown since they are present in most snapshots. (C, D) secondary structure of each residue in MD simulations are predicted using the DSSP algorithm with α-helices shown in red and β-strands in blue. (E, F) In the residue clusters at the bottom, pairwise contacts with occupancies > 0.2 are depicted as an edge connecting two nodes (residues) with edge widths proportional to the pairwise contact occupancies. Labels A1–A5 denote dominant clusters in p53TAD and B1–B8 in Pup. Examples of transiently formed subclusters are indicated by dashed lines (A1.1, A1.2, and A1.3 in p53TAD and B1.1 and B1.2 in Pup).

Inter-residue contact propensities. Different residues along the polypeptide chain display different tendencies to form contacts with other residues. **Fig. 5A,B** shows the average number of contacts per snapshot for each residue, which was calculated as the total number of contacts formed by a residue divided by the total number of MD snapshots. To better visualize the different behaviors, the residues were divided into four distinct groups: the majority of residues that form 0.5–1.5 contacts per snapshot (colored in black), residues that form an unusually small number of contacts (< 0.5) (colored in blue), residues that form a moderately large number of contacts (1.5–2) (colored in yellow), and residues that form a relatively large number of contacts (> 2) are colored in red. For Pup, there are three distinct regions that form the largest numbers of contacts (red) comprising residues (1) Lys7, Arg8, (2) Arg28, Arg29, and (3) Arg56. They perfectly align with the three centers of **Fig. 3** with elevated R_2 values, namely (1) Arg8, (2) Arg29, and (3) Arg56. For p53TAD, the residue that forms the largest number of contacts is Arg65, which is surrounded by residues with a number of contacts below average between 0.5 and 1.0. This rationalizes why R_2 of Arg65 shows a local maximum that is still lower than R_2 in other regions of p53TAD, such as residues 19–26 forming a residue cluster with an intermediate number of contacts. Notably, the 11 N-terminal residues of p53TAD display a lower-than-average amount of contacts, which is consistent with low R_2 values observed across all 10 individual MD trajectories. When the same type of contact analysis is performed with side-chain atoms only, a similar behavior is observed with only a small, systematic reduction in contacts (**Fig. S3**) reflecting that the majority of medium- to long-range inter-residue contacts are made by side-chain atoms.

We also grouped the number of contacts per snapshot formed by each residue according to residue type and normalized them by the number of residues of the same type. The resulting value for each amino acid residue type present in p53TAD and Pup reflects their inherent contact propensity (**Fig. 5C,D**). These profiles display the following trends: positively charged residues arginine and lysine are on average most prone to form contacts, followed by hydrophobic residues isoleucine and leucine as well as aromatic residues tryptophan and phenylalanine. Negatively charged residues aspartate and glutamate, however, are least disposed to form contacts. This may be also a consequence that both IDPs are overall negatively charged (-14e for p53TAD and -12e for Pup). When acidic residues outnumber basic residues, the former tend to repulse each other, thereby increasing R_g, while the latter have more options to interact with an

acidic residue than vice versa leading to an increase of the contact propensity of basic over acidic residues.

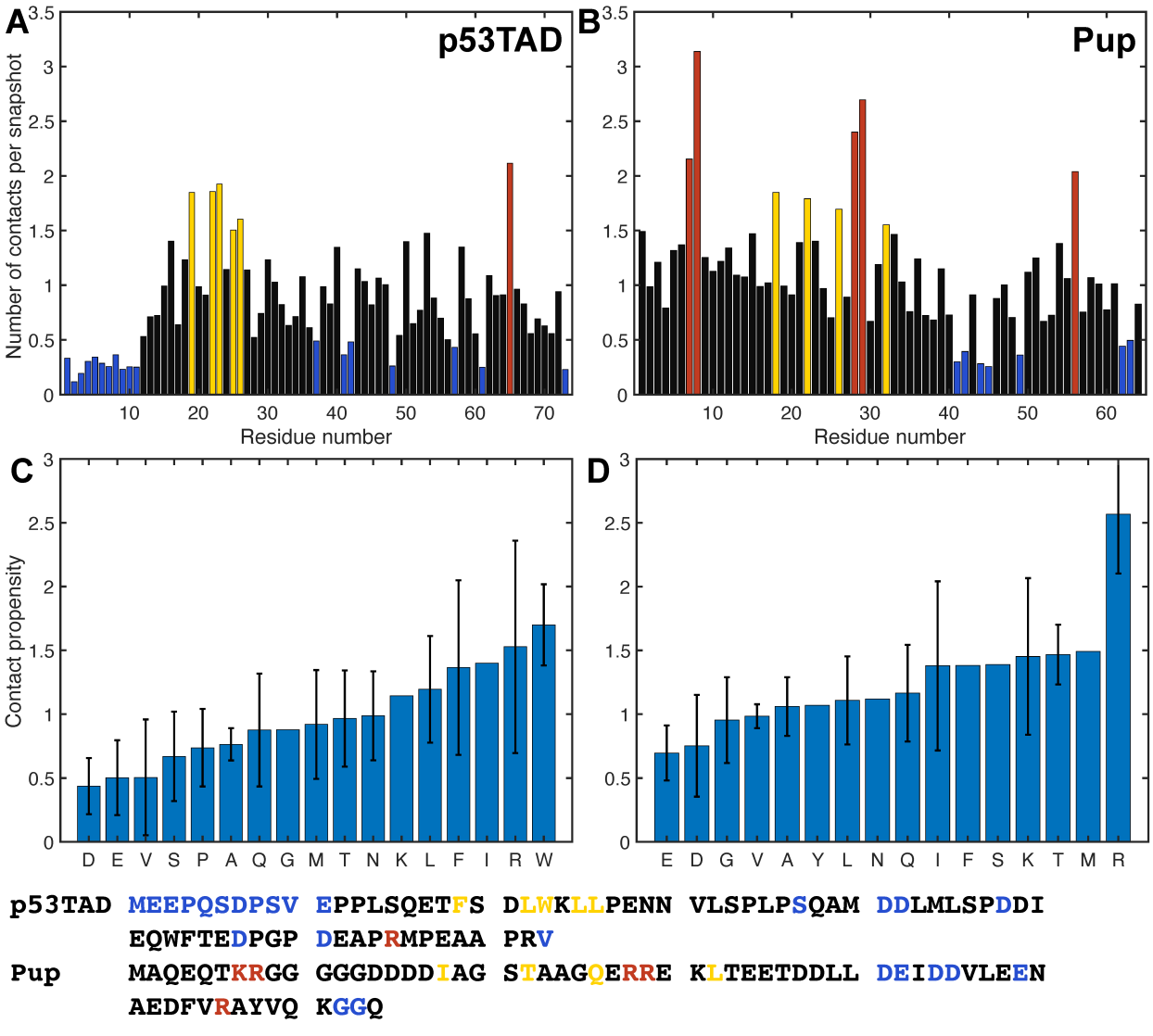


Fig. 5. Number of close contacts formed by each residue during MD simulations of p53TAD and Pup (without outliers) along with average residue-type specific contact propensities. For each residue, the number of contacts was normalized by the number of snapshots for (A) p53TAD and (B) Pup. Residues with their number of contacts per snapshot below 0.5 are depicted in blue, 0.5–1.5 in black, 1.5–2 in yellow, and above 2 in red. Primary sequences of p53TAD and Pup are given at the bottom and colored as in Panels A, B. Average contact propensities according to amino-acid residue type, which is the number of contacts per snapshot averaged over all residues of the same type, are shown for (C) p53TAD, (D) Pup. Error bars correspond to the standard deviations among different residues of the same type.

Contact analysis by graph theory. To investigate the nature of some of the most frequent pairwise contacts in these IDPs, the MD snapshots were analyzed by graph theory where each snapshot is represented as an undirected graph with each residue corresponding to an edge and an inter-residue contact corresponds to an edge connecting the two residues (nodes). The resulting graphs were then analyzed in terms of clusters, which are disconnected graph components that do not have any edges to nodes outside of the cluster. On average 6.0 clusters per snapshot are found for p53TAD and 5.4 clusters for Pup. The probabilities of a cluster to have a given size are represented for both IDPs by the histograms of cluster sizes (**Fig. 6A**), which reveal that clusters consisting of 2 nodes are most abundantly present (around 40%) in both p53TAD and Pup. Moreover, the cluster size probability decreases rapidly with increasing size. For instance, the fraction of clusters with 10 or more nodes (residues) is only 2–3%. Despite their sequence independence and different lengths, the two IDPs have strikingly similar cluster size distributions. The number of edges grows on average linearly with the number nodes (straight solid line), which is much slower than the quadratic behavior of complete graphs (dashed line, **Fig. 6B**). In fact, most of the clusters formed during MD simulations are sparse graphs with a relatively small average edge-to-node ratio of 1.54, which is indicative of tree-like graphs consisting mostly of linear branches with few cross-links. **Fig. 6** also depicts residue clusters (on the right) where pairwise contacts with occupancies > 0.2 are depicted as an edge connecting two nodes (residues) with edge widths proportional to the pairwise contact occupancies.

The graph-theoretical representation of the transient interaction network uncovers the relationship between R_2 profiles and transient contact formation and the types of interactions that are prevalent in IDP structures. For p53TAD, the three centers in the sequence with an elevated experimental R_2 profile are (1) Lys24, (2) Glu51, and (3) Met66, and they are involved in or are sequentially adjacent to clusters A1, A3, and A2, respectively. Electrostatic interactions are important for residue cluster formation in p53TAD, in particular in cluster A2 featuring the pairwise contacts Lys65–Asp57 and Arg65–Glu62. The largest elevation of R_2, however, is the result of the largest interaction network A1. Hydrophobic and aromatic residues Phe19, Leu22, Trp23, Leu25 and Leu26 belong to a p53TAD segment that displays increased helical propensity(66, 67) (secondary structure propensities determined from chemical shifts are shown in **Fig. S5**) and which undergoes distinct loop closure dynamics.(68) In particular, residues

Phe19, Trp23, and Leu26 form the hydrophobic triad that is crucial for the binding of p53TAD to MDM2.(67) Similar to cluster A1, the smaller cluster A3 centered around Ile50 is also driven by hydrophobic interactions.

The regions of Pup with elevated R_2 values (**Fig. 3D**) around Arg8, Ile18, Thr22, Arg29, Arg56 are all involved in clusters B1, B4, or B3 (**Fig. 4E,F**). Separate clusters can involve sequentially adjacent residues, such as clusters B2 and B3 or clusters B3 and B5 and thereby mediate cooperative behavior. The most dominant inter-residue interaction in Pup is of electrostatic nature resulting in the transient formation of salt bridges involving residue pairs in cluster B1.2 (Arg8–Asp14, Arg8–Asp15) and cluster B3 (Arg56–Asp53, Arg56-Glu52). Many of these residues appear to play the role of hubs promoting enhanced interactions also with other residues as visualized by the graphs in **Fig. 4E,F**.

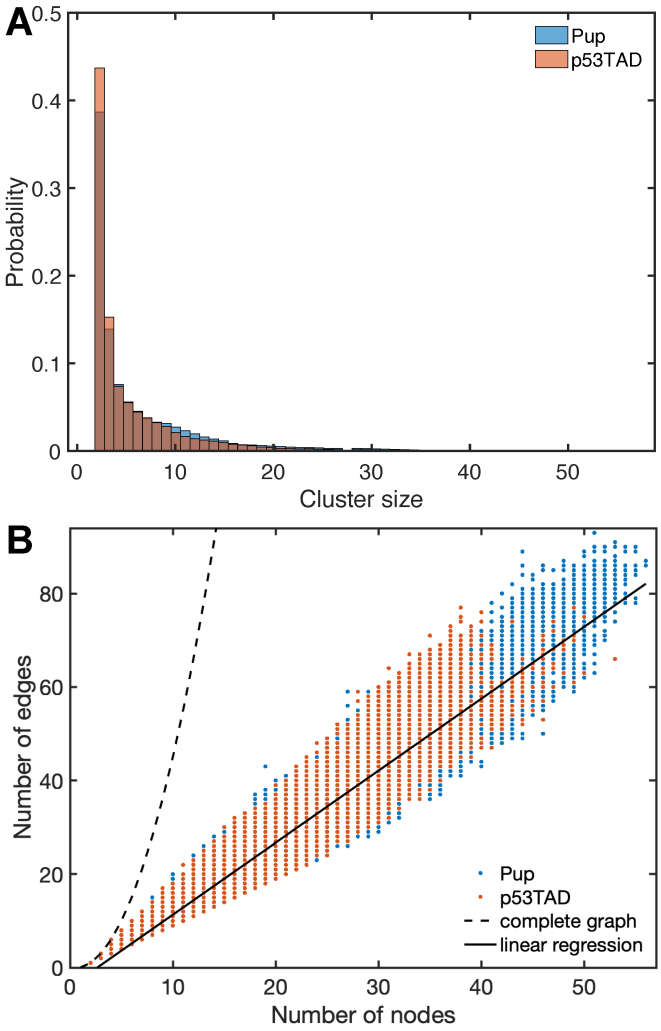


Fig. 6. Graph theoretical analysis of inter-residual interactions and transient interaction networks of p53TAD and Pup. (A) Clusters consisting of 2 nodes (residues) dominate in the MD structures of p53TAD and Pup (without outlier trajectories), followed by clusters of size 3, etc. (B) The majority of the unique clusters are sparse graphs, with their number of edges much smaller than the number of edges in complete graphs growing with $N(N\text{-}1)/2$ where N is the number of nodes. The average edge-to-node ratio is 1.54 (slope indicated by solid black line), indicating predominantly tree-like graphs that sometimes have a few additional edges (cross-linked branches).

DISCUSSION

Disordered proteins play a prominent role in many regulatory processes using their unique malleability to interact with their targets. Details of conformational substates of IDPs and how they are shaped by the complex interplay of inter-residue interaction networks are currently poorly understood both experimentally and computationally. In this work, we showed how the latest advances in MD force fields and computational protocols allow the nearly quantitative prediction of the complex behavior of the two IDPs p53TAD and Pup, including their dynamics time scales from site-resolved NMR spin relaxation. Both proteins have been characterized by a host of experimental techniques, including X-ray crystallography,(69, 70) NMR,(7, 63, 67, 71-73) small-angle X-ray scattering (SAXS),(61, 74) FRET,(75, 76) and fluorescence correlation spectroscopy.(68)

The global dimensions of IDPs can be experimentally characterized by SAXS providing information about their radius of gyration R_g for direct comparison with MD ensembles. For Pup, $\langle R_g \rangle$ from the 10 1-μs MD simulations follows the power law of **Eq. 6** with a Flory exponent ν = 0.601, which closely mirrors the behaviour of a self-avoiding random coil (ν = 0.598). By contrast, p53TAD is more expanded with ν = 0.624, which is consistent with previous experimental results reported for this protein.(61) Such behaviour could be the result of stronger repulsive intra-residual forces caused by a slightly higher negative net charge (-14e of p53TAD vs. -12e of Pup) and a high percentage of prolines (18% in p53TAD vs. none in Pup) known to increase extendedness.(77) The relatively high ν values of both proteins suggest that their interactions with water solvent are highly favorable preventing the hydrophobic collapse of their polypeptide chains.

The 10 1-μs MD trajectories allow extensive sampling of the radius of gyration over time and extract characteristic time scales from its autocorrelation function (**Fig. 1**). For both proteins, the time-correlation function follows in good approximation a biexponential decay with correlation times around 10 and 55 ns. Global distance fluctuations can be studied experimentally by nanosecond fluorescence correlation spectroscopy (nsFCS), which found for 8 M urea denatured ubiquitin global reconfiguration times τ_r in the range of 50–90 ns.(16) A nsFCS study of α-synuclein, which is about twice as long in sequence as the IDPs studied here, identified two reconfigurational correlation times of τ_{r1} = 23 ns and τ_{r2} = 136 ns.(30) These

correlation times are within a factor 2–3 of those found in the current study, although it should be kept in mind that they report about a donor/acceptor pair, i.e. S42C/T92C in the case of α-synuclein, rather than about R_g.

Heteronuclear ^{15}N relaxation offers a complementary view of IDP dynamics. Longitudinal R_1 and transverse R_2 relaxation rates are caused by local spin interactions, namely the magnetic dipole-dipole coupling and chemical shielding anisotropy, and they reflect reorientational dynamics amplitudes and timescales due to local conformational fluctuations as well as longer-range reorientational motional modes of the order of an IDP's persistence length and beyond. Model-free analysis is not applicable to IDP relaxation data due to the absence of a well-conserved global rotational diffusion tensor as reference frame.(27) Instead, a residue-by-residue interpretation can applied where the correlation function of each site is described as a multiexponential function of the type of **Eq. 8** with 6 exponential dynamics modes.(28, 50, 55, 64, 78) The hierarchy of dynamics modes depicted in **Fig. 3** shows a broad distribution of time scales including rapid librational motions (< 100 ps) and dominant low nanosecond motions, which sample the different local energy basins of backbone φ,ψ dihedral angles. The slowest modes with time scales in the range of 3–20 ns represent predominantly collective segmental reorientational motions. A similar hierarchy of time scales has been observed by fluorescence depolarization kinetics measurements of α-synuclein.(48) These collective motions involve medium to longer-range interactions between residues that can be elucidated by graph theoretical analysis of the MD trajectories described here. For Pup, many of these slower motional modes have correlation times around 3–4 ns whereas for p53TAD they are on average twice as large. For both proteins the three distinct bands of time scales are pervasive across their polypeptide sequence (**Fig. 3E,F**).

MD methodology has made great strides in recent years to toward an increasingly realistic representation of disordered proteins.(26) Besides experimental scattering data, quantitative NMR has played a key role for the independent validation of MD ensembles. Because NMR spin relaxation parameters fully quantitatively reflect IDP dynamics at atomic-level resolution both in terms of motional amplitudes and time scales, their accurate reproduction by MD has been an important but also very challenging task. A recent comparison of commonly used MD force fields that do not use residue-specific backbone potentials showed for several

IDPs significant force-field dependences with the best results obtained when the analysis was restricted to average correlation functions of chunks of 10-ns subtrajectories.(56) The need to exclude slower time-scale motions, which are prominent in both experimental data and simulations (see for example **Fig. 3**), may reflect the lack of convergence due to limited sampling. Beneficial for all simulations was the improvement of the TIP4P-D water model over TIP3P preventing overly collapsed IDP ensembles, which is consistent with other computational studies.(38, 57) Because of the observed discrepancies between experiments and MD simulations, some studies applied *post factum* adjustments to the MD simulations in order to improve agreement, which include uniform or selective scaling of the MD time scale or correlation times(27-30) or the reweighting of sub-trajectories.(64) Here, we chose a different approach: rather than relying on *post factum* modifications, we use the residue-specific ff99SBnmr2 force field, which was specifically designed for the improved representation of IDPs without the need of any corrections.(57, 58) A correction-free MD approach has recently been reported for the intrinsically disordered SH4UD protein with the Amber ff03ws force field, which does not use residue-type independent backbone dihedral angle potentials, and no time-scale dependent data, such as NMR spin relaxation, were used for validation.(79) NMR chemical shifts were back-calculated using SHIFTX2,(80) which, besides 3D structural information, makes extensive use of protein sequence data. Here, we back-calculated NMR chemical shifts using PPM(81) (**Fig. S4**), which only uses the physical parametrization of chemical shifts with respect to 3D protein structure of each snapshot, achieving very good agreement with experiment.(73)

The close correspondence observed between experimental and computed ^{15}N relaxation R_1 and R_2 relaxation rates for both IDPs studied here (**Fig. 3**), without the need for *post factum* corrections, attests to the accuracy and robustness of the computational protocol used. It applies REMD for the generation of conformational ensembles belonging to different temperatures from which 10 representative structures at 300 K were randomly selected as starting structures for 1-μs MD trajectories whereby all simulations made use of the ff99SBnmr2 force field and the TIP4P-D water model. MD-derived longitudinal ^{15}N R_1 follow the shapes of the experimental R_1 profiles with a small tendency to underestimate the experimental ^{15}N R_1 rates by 4–6% whereas ^{15}N R_2 relaxation rates overestimate the experimental values on average by 26% for Pup and 34% for p53TAD. This level of agreement is significantly better than for previously reported comparisons of this type.

Few individual trajectories (10–20%) show systematically larger differences with respect to experiment than the rest. For the proteins studied here, they are trajectories #4 of p53TAD and #5 and #7 of Pup (**Fig. S2, Table S4**). These trajectories are characterized by the persistent formation of secondary structure (#4 of p53TAD and #5 of Pup) (**Fig. 4C,D**) or by a collapsed overall geometry with a reduced $<R_g>$ compared to the other trajectories (trajectory #7 of Pup) (**Table S5**). At the individual trajectory level, these outlier trajectories are in poorer agreement with experimental data and their removal from the set of 10 trajectories during the back-calculation of relaxation rates further improves the agreement with experiment (**Fig. S3**). From such diagnostic analysis it follows that these outlier trajectories are either overrepresented in the original simulations or the result of simulation artifacts, for example, caused by inaccuracies of the underlying force field. Removal of individual trajectories based on comparison with experiment should be applied with great care and be reserved primarily for diagnostic purposes, such as the analysis of shortcomings of the simulations. While *post factum* trajectory selection or reweighting can provide better agreement with experiment, it is generally unclear whether the altered ensembles are in fact consistent with an alternative, physics-based force field, thereby complicating the physical interpretation of such ensembles.

Although it is difficult to identify individual force field terms responsible for the IDP behaviour observed in the outlier trajectories, these results can nonetheless provide useful input to guide future force field improvements. With more computer power, it will be possible to gain better statistics by generating a larger number of trajectories for the improved sampling of conformational space allowing the more rigorous assessment of the underlying force field, the water model, and other aspects of the computational methods used. Conversely, such insights may allow the further improvement of force fields and methods for applications also to other proteins. In fact, the ff99SBnmr1 force field, which is the parent force field of ff99SBnmr2, was developed and optimized using this strategy by the systematic reweighting of MD snapshots based on many trial force fields using experimental NMR data of intact proteins.(82)

The good agreement of the MD simulation with experimental observables both motivates and justifies the analysis of other protein properties observed in the MD trajectories that are difficult to measure. This includes the analysis of transient inter-residue interactions. The molecular driving forces of these interactions are fundamentally similar to those of ordered proteins although average hydration properties may differ.(79) In contrast to ordered proteins,

inter-residue interactions between non-sequential amino acids are short-lived. Therefore, the time-averaged interaction maps (**Fig. 4A,B**) offer only partial insights as they conceal the compositions and distributions of instantaneous interaction clusters. In fact, the relatively large network reflected by the average contact map contrasts the much smaller size of graphs that exist at any given time, which attests to the very heterogeneous and transient nature of instantaneous contact clusters. The highest occupancy of pairwise contacts found is around 0.5, which mostly belong to (*i*,*i*+3) contacts. For a list of the most frequent pairwise contacts, **see Tables S2, S3**.

Snapshot by snapshot analysis revealed the dominance of small cluster sizes over larger ones (**Fig. 6**). For both p53TAD and Pup, clusters with 2 or 3 residues make up more than 50% of all clusters and clusters with more than 10 residues have notably low occurrence, although their formation could be functionally relevant during molecular recognition events. Because clusters consisting of residue pairs dominate intra-residual interactions in both IDPs, further analysis of the interaction network was performed based on pairwise contacts. Contact maps were generated for p53TAD and Pup averaged over all MD trajectories and pairwise contacts that have occupancies larger than 0.2 visualized as separate graphs (**Fig. 4E,F**). Instantaneous clusters can belong to such larger graphs as exemplified by clusters A1.1, A1.2, A1.3 for p53TAD and clusters B1.1 and B1.2 for Pup (**Fig. 4E,F**). The dominant clusters are characterized by a mix of hydrogen bonds, salt bridges (e.g., involving Arg65 in cluster A2, Arg8 in star-like cluster B1.2, and Arg56 in cluster B3), hydrophobic and aromatic interactions (e.g., Phe19, Leu22/25/26, and Trp23 in cluster A1). These are consistent with the driving forces attributed to liquid-liquid phase separation, namely intermolecular contacts among aromatic residues,(83-85) electrostatic interactions,(86-88) and hydrophobic interactions.(89)

The majority of clusters are linear graphs with few circular sub-graphs leading to the linear relationship between the number of nodes and number of edges (**Fig. 6B**). Acidic residues tend to have low cluster participation whereas arginine residues have the highest participation in both proteins (**Fig. 5A,B**). This difference in cluster participation between cationic and anionic residues is also evident in **Fig. 5C,D**. Among the neutral amino acids, those with larger side-chains are more prone to interactions with non-neighboring residues due to their intrinsically larger distance range. In fact, Pro, Val, Ser, Ala, Gly have the lowest interaction propensities among neutral residues and among pairs of chemically similar residues, such as Gln vs. Asn and Leu vs. Val, the larger residue (Gln, Leu) dominates the smaller one (Asn, Val).

A primary biological function of p53TAD is to negatively regulate p53 by interacting with the ubiquitin ligases MDM2 and MDMX for the degradation of p53. This interaction is one of the earliest and best studied interactions between an IDP and a folded protein both by experiment(67-69) and computation.(90) In order to better understand the molecular recognition mechanism underlying the formation of this complex, a realistic and accurate description of the free state of p53TAD is of central importance. For MD studies, the choice of the protocol, especially of the force field and water model, is consequential. A recent unbiased REMD study of free p53TAD reported the detailed comparison using five different MD force fields all without residue-specific backbone potentials. Based on 1-μs long replicas major differences were revealed in terms of the structural propensities among them and also with respect to experimental data.(91) An even longer simulation of residues 10–39 of p53TAD for a total length of 1.4 ms analyzed by Markov state models identified substantial populations of β-sheets across the sequence,(92) a behavior that is at variance with the above mentioned REMD ensembles(91) as well as with experimental solution NMR data.(67) These together with many other studies show that force fields need to be chosen following extensive testing to ensure that long trajectories, generated with considerable computational effort, offer the most realistic biophysical insights about these highly complex, heterogeneous systems.

In addition to forming transient intramolecular contacts, IDPs can also dynamically interact with other IDPs driving the formation of liquid-liquid phase separation. With a rapidly increasing body of experimental data on LLPS condensates,(9, 10, 93) all-atom MD simulations have an important role to play for a mechanistic understanding of emerging phase separation properties. Since the molecular driving forces of LLPS are the same as for intramolecular IDP interactions,(94) such as those described here, the optimal accuracy of force fields along with adequate sampling schemes of the heterogeneous condensate environment will be key for the quantitative interpretation of experimental data, allowing the prediction of condensate formation and eventually may open the way for new interventional approaches to actively reprogram condensates and their properties.

Although a possible role of Pup in LLPS is not known, LLPS involving full-length p53 has been documented and p53TAD has been implicated in both phase separation and oncogenic amyloid aggregation.(76, 95) Multivalent electrostatic interactions between the N-terminal

584 domain, p53TAD, and the C-terminal domain were identified as critical for LLPS, which were
585 shown to be positively modulated through molecular crowding and negatively modulated by the
586 addition of DNA and ATP molecules and post-translational modification. It was suggested that
587 compartmentalization of p53 into the droplets suppresses its transcriptional regulatory function,
588 while its release from droplets under cellular stress can activate p53.(76) These findings point to
589 the need for the comprehensive characterization of these intermolecular interactions at residue-
590 and atomic-level resolution. The agreement with experiment reported here clearly suggests that
591 MD methodology has reached a level of accuracy allowing it to make critical contributions
592 toward this goal.

593 The results of our study further advance the long-held premise of MD simulations to
594 realistically describe IDP ensembles on their native dynamics time scales toward the better
595 understanding of their biophysical properties and biological function. Both IDPs chosen in this
596 study, p53TAD and Pup, undergo folding upon binding to their protein targets and it will be
597 interesting to see how the protocol will perform for IDPs that do not fold when interacting with
598 other proteins. For both p53TAD and Pup, the use of REMD allows the adequate sampling of
599 conformational space for the generation of a representative set of initial structures that are then
600 subjected to long, continuous MD simulations. The close agreement found for the extendedness
601 of the simulated IDPs with experiment and polymer theory suggests an appropriate balance
602 between the ff99SBnmr2 force field and the TIP4P-D water model at the global scale. It
603 favorably complements the authentic IDP behavior achieved by this protocol on the local scale in
604 terms of its compliance at the individual residue level with coil libraries, scalar couplings, and
605 chemical shifts. In addition to the realistic modeling of ensemble properties, our protocol also
606 reproduces motional amplitudes and time scales encoded in quantitative NMR spin relaxation
607 data with near experimental accuracy suggesting that the dominant minima of the free energy
608 surface together with their many low-lying transition states are realistically captured by this
609 comprehensive computational framework. These results prompted a more detailed analysis of
610 short-lived inter-residue interactions, which was achieved by graph theory revealing
611 characteristic inter-residue contact patterns and the extraction of residue-type specific interaction
612 propensities. The realistic IDP conformational dynamics model achieved by the protocol
613 described here advances our increasingly mechanistic and predictive understanding of IDPs

along with their interactions and binding properties with ordered and disordered molecular targets ranging from regulatory pathways to emerging LLPS phenomena.

METHODS

Molecular dynamics simulations. Fully extended structures of p53TAD and Pup were prepared using the LEaP program in AmberTools16.(96) After equilibration, they were used to run replica-exchange MD (REMD) simulations for the sampling of conformational space (36 replicas for each IDP covering a temperature range from 298–353 K for p53TAD and 298–365 K for Pup, see Supplementary Material) with each replica being 1 μs of length. Exchange was attempted every 10 ps and the exchange probability was about 0.3. For each IDP, 10 structures were randomly selected from the room-temperature (298 K) REMD ensemble and used as initial structures to run free MD simulations for 1 μs in the NPT ensemble at 300 K and 1 atm. The protein force field and water model used in all simulations were AMBER ff99SBnmr2 and TIP4P-D.

All MD simulations were performed using the GROMACS 2020.2 package.(97) The integration time step was set to 2 fs with all bond lengths containing hydrogen atoms constrained by the LINCS algorithm. Na^+ or Cl^- ions were added to neutralize the total charge of the system. A 10 Å cutoff was used for all van der Waals and electrostatic interactions. Particle-mesh Ewald summation with a grid spacing of 1.2 Å was used to calculate long-range electrostatic interactions. A cubic simulation box extending 8 Å from the protein surface in all three dimensions was used. Energy minimization was performed using the steepest descent algorithm for 50,000 steps. The system was simulated for 100 ps at constant temperature and constant volume with all protein heavy atoms positionally fixed. The pressure was then coupled to 1 atm and the system was simulated for another 100 ps. The final production run of 1 μs length was performed in the NPT ensemble at 300 K and 1 atm. For simulation details, see **Table S1**.

Radius of gyration tensor calculations and derived quantities. In order to map the global shape of p53TAD and Pup conformers, radius of gyration tensors were computed as 3×3

matrices ***S*** from each snapshot of the room-temperature REMD ensemble and the free MD simulations as follows:(98)

$$S_{\alpha\beta} = \frac{1}{2N^2}\sum_{i,j=1}^{N}(r_\alpha^{(i)} - r_\alpha^{(j)})(r_\beta^{(i)} - r_\beta^{(j)}) \quad (2)$$

where $r_{\alpha(\beta)}^{(i)}$ is cartesian coordinate α (β) (= *x*, *y*, *z*) of atom *i* in the coordinate system that has its origin in the center of mass of the molecule. Diagonalization of ***S*** yields three non-negative eigenvalues $0 \le \lambda_1 \le \lambda_2 \le \lambda_3$ from which the *radius of gyration* R_g is obtained, $R_g = (\lambda_1 + \lambda_2 + \lambda_3)^{1/2}$, the *asphericity A*,(98, 99)

$$A = \frac{(\lambda_3 - \lambda_2)^2 + (\lambda_3 - \lambda_1)^2 + (\lambda_2 - \lambda_1)^2}{2(\lambda_3 + \lambda_2 + \lambda_1)^2} \quad (3)$$

and the *prolateness P,*(100)

$$P = \frac{(2\lambda_3 - \lambda_2 - \lambda_1)(2\lambda_2 - \lambda_3 - \lambda_1)(2\lambda_1 - \lambda_3 - \lambda_2)}{2(\lambda_3^2 + \lambda_2^2 + \lambda_1^2 - \lambda_3\lambda_2 - \lambda_3\lambda_1 - \lambda_2\lambda_1)^{3/2}} \quad (4)$$

The asphericity measures the degree to which the three axis lengths of the ellipsoid of inertia (eigenvalues) are equal, whereas the prolateness *P* indicates whether the largest or smallest axis length is closer to the middle axis length. *P* takes values between -1 and 1, quantifying the transition from oblate to prolate shapes. Normalized time-correlation functions of $R_g(t)$, made offset-free, were computed according to

$$C_{Rg}(t) = \langle(R_g(\tau) - \langle R_g\rangle)(R_g(t+\tau) - \langle R_g\rangle)\rangle_\tau / \langle(R_g(\tau) - \langle R_g\rangle)^2\rangle_\tau \quad (5)$$

as an average over all 1-μs MD trajectories.

According to polymer theory, for an unfolded polymer the ensemble-averaged R_g scales with the number of residues *N* as(62, 77)

$$\langle R_g\rangle = \rho_0 N^\nu \quad (6)$$

where ρ_0 is a constant reflecting the average size of a residue and the Flory exponent ν determines the overall compactness of the polymer serving as a reference.

Back-calculation of R_1, R_2 relaxation rates. For IDPs, the normalized time-autocorrelation function $C(t)$ of the lattice part of the spin-relaxation active magnetic dipole-dipole interaction cannot be factorized into an overall tumbling part and an internal dynamics part. Rather, we compute the full $C(t)$ directly from an MD trajectory using the second-order Legendre polynomial:

$$C(t) = \frac{1}{2}\langle 3[\boldsymbol{e}(\tau)\boldsymbol{e}(\tau + t)]^2 - 1\rangle \quad (7)$$

where $\boldsymbol{e}(t)$ is the unit vector defining the ^{15}N–^{1}H bond orientation whereby snapshots were *not* aligned with respect to a reference snapshot. The angular brackets indicate averaging from time $\tau = 0$ to $T_{MD} - t$, where T_{MD} is the total trajectory length. The calculation of $C(t)$ was efficiently performed by the fast Fourier transform (FFT) using the Wiener–Khinchin theorem. For acceptable statistical convergence, the analysis of $C(t)$ was limited to its initial portion from t = 0 - T_{MD} /3. Next, a multiexpoential decay function was fitted to $C(t)$:(101)

$$C(t) = \sum_{i=1}^{6} A_i e^{-t/\tau_i} \quad (8)$$

where A_i and τ_i are the best fitting parameters subject to the conditions:

$$\sum_{i=1}^{6} A_i = 1 \quad A_i \geq 0, \tau_i \geq 0 \quad (9)$$

The spectral density function $J(\omega)$ can be then analytically obtained via Fourier transformation of $C(t)$:

$$J(\omega) = 2\int_0^\infty C(t)\cos(t)\mathrm{d}t = \sum_{i=1}^{6} \frac{2A_i\tau_i}{1+(\omega\tau_i)^2} \quad (10)$$

NMR spin relaxation parameters R_1 and R_2 were then computed using the standard expressions:(102-105)

$$R_1 = d_{00}[3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N)] + c_{00}\omega_N^2 J(\omega_N) \quad (11)$$

$$R_2 = \frac{1}{2}d_{00}[4J(0) + 3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{1}{6}c_{00}\omega_N^2[4J(0) + 3J(\omega_N)] \quad (12)$$

where $d_{00} = \frac{1}{20}(\frac{\mu_0}{4\pi})^2(\frac{h}{2\pi})^2\gamma_H^2\gamma_N^2\langle r_{NH}^{-3}\rangle^2$ and $c_{00} = \frac{1}{15}\Delta\sigma^2$. μ_0 is the permeability of vacuum, h is Plank's constant, γ_H and γ_N are the gyromagnetic ratios of ^{1}H and ^{15}N, and r_{NH} = 1.02 Å is the backbone N-H bond length. The ^{15}N chemical shift anisotropy was set to $\Delta\sigma$ = -160 ppm.

Analysis of inter-residue contacts and residue clusters by graph theory. Contact analysis was performed on all snapshots of the MD simulations of both p53TAD and Pup. A contact is considered formed when the nearest distance between atoms from two different residues is smaller than 4 Å. First-neighbor contacts (between residues *i*,*i*+1), and second-neighbor contacts (between residues *i*,*i*+2) were excluded since they are present for most residues. For each residue in p53TAD and Pup, the total number of contacts formed by a particular residue is determined and normalized by the number of MD snapshots. Each snapshot was converted to a graph where residues are represented as nodes and contacts between two residues are represented as edges between them. The initial graph was then decomposed into a maximal number of disconnected graph components called *clusters*, i.e. there is no edge between any node in the cluster and any node outside the cluster. The size of a cluster corresponds to the number of its nodes.

ACKNOWLEDGEMENTS

We thank Dr. Da-Wei Li for helping with the graph theoretical analysis and Dr. Mouzhe Xie for providing the experimental NMR relaxation data. MD and REMD simulations were performed at the Ohio Supercomputer Center. The authors declare that they have no competing interests. Molecular dynamics input files (topology, parameters, and initial coordinates) and generated trajectories as well as the codes used to produce the results and analyses presented in this manuscript are available from https://doi.org/10.5061/dryad.v9s4mw6z2.

SUPPORTING INFORMATION

Fig. S1. Radius of gyration of the IDPs p53TAD and Pup in 10 1-μs MD trajectories each at 300 K with starting structures randomly chosen from replica exchange simulations.

REFERENCES

1. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Bio. 2015;16(1):18-29.
2. Habchi J, Tompa P, Longhi S, Uversky VN. Introducing Protein Intrinsic Disorder. Chem Rev. 2014;114(13):6561-88.
3. Tompa P. Unstructural biology coming of age. Curr Opin Struct Biol. 2011;21(3):419-25.
4. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. Annu Rev Biophys. 2008;37:215-46.
5. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of Intrinsically Disordered Regions and Proteins. Chem Rev. 2014;114(13):6589-631.
6. Romer L, Klein C, Dehner A, Kessler H, Buchner J. p53 - A natural cancer killer: Structural insights and therapeutic concepts. Angew Chem Int Edit. 2006;45(39):6440-60.
7. Chen X, Solomon WC, Kang Y, Cerda-Maira F, Darwin KH, Walters KJ. Prokaryotic Ubiquitin-Like Protein Pup Is Intrinsically Disordered. J Mol Bio. 2009;392(1):208-17.
8. Alberti S, Gladfelter A, Mittag T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell. 2019;176(3):419-34.
9. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. Nat Rev Mol Cell Bio. 2017;18(5):285-98.
10. Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease. Science. 2017;357(6357).
11. Uversky VN. Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. Curr Opin Struct Biol. 2017;44:18-30.
12. Sormanni P, Piovesan D, Heller GT, Bonomi M, Kukic P, Camilloni C, et al. Simultaneous quantification of protein order and disorder. Nat Chem Biol. 2017;13(4):339-42.
13. Ozenne V, Bauer F, Salmon L, Huang JR, Jensen MR, Segard S, et al. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics. 2012;28(11):1463-70.
14. Marsh JA, Forman-Kay JD. Ensemble modeling of protein disordered states: Experimental restraint contributions and validation. Proteins. 2012;80(2):556-72.
15. Jensen MR, Zweckstetter M, Huang JR, Backledge M. Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. Chem Rev. 2014;114(13):6632-60.
16. Aznauryan M, Delgado L, Soranno A, Nettels D, Huang JR, Labhardt AM, et al. Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. Proc Natl Acad Sci USA. 2016;113(37):E5389-E98.
17. Gomes GNW, Krzeminski M, Namini A, Martin EW, Mittag T, Head-Gordon T, et al. Conformational Ensembles of an Intrinsically Disordered Protein Consistent with NMR, SAXS, and Single-Molecule FRET. J Am Chem Soc. 2020;142(37):15697-710.
18. Wang WN. Recent advances in atomic molecular dynamics simulation of intrinsically disordered proteins. Phys Chem Chem Phys. 2021;23(2):777-84.
19. Pitera JW, Chodera JD. On the Use of Experimental Observations to Bias Simulated Ensembles. Journal of chemical theory and computation. 2012;8(10):3445-51.
20. Roux B, Weare J. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. The Journal of chemical physics. 2013;138(8).

21. Cavalli A, Camilloni C, Vendruscolo M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. The Journal of chemical physics. 2013;138(9).
22. Camilloni C, Vendruscolo M. Statistical Mechanics of the Denatured State of a Protein Using Replica-Averaged Metadynamics. J Am Chem Soc. 2014;136(25):8982-91.
23. Hummer G, Koefinger J. Bayesian ensemble refinement by replica simulations and reweighting. The Journal of chemical physics. 2015;143(24).
24. Salvi N, Abyzov A, Blackledge M. Multi-Timescale Dynamics in Intrinsically Disordered Proteins from NMR Relaxation and Molecular Simulation. J Phys Chem Lett. 2016;7(13):2483-9.
25. Cheng P, Peng JH, Zhang ZY. SAXS-Oriented Ensemble Refinement of Flexible Biomolecules. Biophys J. 2017;112(7):1295-301.
26. Best RB. Computational and theoretical advances in studies of intrinsically disordered proteins. Curr Opin Struct Biol. 2017;42:147-54.
27. Prompers JJ, Brüschweiler R. General Framework for Studying the Dynamics of Folded and Nonfolded Proteins by NMR Relaxation Spectroscopy and MD Simulation. J Am Chem Soc. 2002;124(16):4522-34.
28. Xue Y, Skrynnikov NR. Motion of a Disordered Polypeptide Chain as Studied by Paramagnetic Relaxation Enhancements, N-15 Relaxation, and Molecular Dynamics Simulations: How Fast Is Segmental Diffusion in Denatured Ubiquitin? J Am Chem Soc. 2011;133(37):14614-28.
29. Rezaei-Ghaleh N, Parigi G, Zweckstetter M. Reorientational Dynamics of Amyloid-beta from NMR Spin Relaxation and Molecular Simulation. J Phys Chem Lett. 2019;10(12):3369-75.
30. Rezaei-Ghaleh N, Parigi G, Soranno A, Holla A, Becker S, Schuler B, et al. Local and Global Dynamics in Intrinsically Disordered Synuclein. Angew Chem Int Edit. 2018;57(46):15262-6.
31. Ahmed MC, Skaanning LK, Jussupow A, Newcombe EA, Kragelund BB, Camilloni C, et al. Refinement of alpha-Synuclein Ensembles Against SAXS Data: Comparison of Force Fields and Methods. Front Mol Biosci. 2021;8.
32. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. Journal of chemical theory and computation. 2015;11(8):3696-713.
33. Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Migues AN, et al. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. Journal of chemical theory and computation. 2020;16(1):528-52.
34. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017;14(1):71-3.
35. Robustelli P, Piana S, Shaw DE. Developing a molecular dynamics force field for both folded and disordered protein states. Proc Natl Acad Sci USA. 2018;115(21):E4758-E66.
36. Best RB, Zheng W, Mittal J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. Journal of chemical theory and computation. 2014;10(11):5113-24.
37. Nerenberg PS, Head-Gordon T. Optimizing Protein-Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. Journal of chemical theory and computation. 2011;7(4):1220-30.
38. Piana S, Donchev AG, Robustelli P, Shaw DE. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. J Phys Chem B. 2015;119(16):5113-23.

39. Izadi S, Anandakrishnan R, Onufriev AV. Building Water Models: A Different Approach. J Phys Chem Lett. 2014;5(21):3863-71.
40. Mu J, Pan Z, Chen H-F. Balanced Solvent Model for Intrinsically Disordered and Ordered Proteins. J Chem Inf Model. 2021;61(10):5141-51.
41. Kang W, Jiang F, Wu YD. How to strike a conformational balance in protein force fields for molecular dynamics simulations? Wires Comput Mol Sci. 2021.
42. Song D, Luo R, Chen H-F. The IDP-Specific Force Field *ff14IDPSFF* Improves the Conformer Sampling of Intrinsically Disordered Proteins. J Chem Inf Model. 2017;57(5):1166-78.
43. Liu H, Song D, Lu H, Luo R, Chen H-F. Intrinsically disordered protein-specific force field CHARMM36IDPSFF. Chem Biol Drug Des. 2018;92(4):1722-35.
44. Song D, Liu H, Luo R, Chen H-F. Environment-Specific Force Field for Intrinsically Disordered and Ordered Proteins. J Chem Inf Model. 2020;60(4):2257-67.
45. Jiang F, Zhou C-Y, Wu Y-D. Residue-Specific Force Field Based on the Protein Coil Library. RSFF1: Modification of OPLS-AA/L. J Phys Chem B. 2014;118(25):6983-98.
46. Zhou C-Y, Jiang F, Wu Y-D. Residue-Specific Force Field Based on Protein Coil Library. RSFF2: Modification of AMBER ff99SB. J Phys Chem B. 2015;119(3):1035-47.
47. Kang W, Jiang F, Wu Y-D. Universal Implementation of a Residue-Specific Force Field Based on CMAP Potentials and Free Energy Decomposition. Journal of chemical theory and computation. 2018;14(8):4474-86.
48. Das D, Arora L, Mukhopadhyay S. Fluorescence Depolarization Kinetics Captures Short-Range Backbone Dihedral Rotations and Long-Range Correlated Dynamics of an Intrinsically Disordered Protein. J Phys Chem B. 2021;125(34):9708-18.
49. Alexandrescu AT, Shortle D. Backbone Dynamics of a Highly Disordered 131 Residue Fragment of Staphylococcal Nuclease. J Mol Biol. 1994;242(4):527-46.
50. Brutscher B, Brüschweiler R, Ernst RR. Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by 1H, 13C, and 15N nuclear magnetic resonance spectroscopy. Biochemistry. 1997;36(42):13043-53.
51. Schwalbe H, Fiebig KM, Buck M, Jones JA, Grimshaw SB, Spencer A, et al. Structural and Dynamical Properties of a Denatured Protein. Heteronuclear 3D NMR Experiments and Theoretical Simulations of Lysozyme in 8 M Urea. Biochemistry. 1997;36(29):8977-91.
52. Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, et al. Long-Range Interactions Within a Nonnative Protein. Science. 2002;295(5560):1719-22.
53. Schwarzinger S, Wright PE, Dyson HJ. Molecular hinges in protein folding: The urea-denatured state of apomyoglobin. Biochemistry. 2002;41(42):12681-6.
54. Salvi N, Abyzov A, Blackledge M. Analytical Description of NMR Relaxation Highlights Correlated Dynamics in Intrinsically Disordered Proteins. Angew Chem Int Edit. 2017;56(45):14020-4.
55. Kampf K, Izmailov SA, Rabdano SO, Groves AT, Podkorytov IS, Skrynnikov NR. What Drives ^{15}N Spin Relaxation in Disordered Proteins? Combined NMR/MD Study of the H4 Histone Tail. Biophys J. 2018;115(12):2348-67.
56. Zapletal V, Mladek A, Melkova K, Lousa P, Nomilner E, Jasenakova Z, et al. Choice of Force Field for Proteins Containing Structured and Intrinsically Disordered Regions. Biophys J. 2020;118(7):1621-33.
57. Yu L, Li D-W, Bruschweiler R. Balanced Amino-Acid-Specific Molecular Dynamics Force Field for the Realistic Simulation of Both Folded and Disordered Proteins. Journal of chemical theory and computation. 2020;16(2):1311-8.

58. Yu L, Li D-W, Brüschweiler R. Systematic Differences between Current Molecular Dynamics Force Fields To Represent Local Properties of Intrinsically Disordered Proteins. J Phys Chem B. 2021;125(3):798-804.
59. Le Guillou JC, Zinn-Justin J. Critical Exponents for the *n*-Vector Model in Three Dimensions from Field Theory. Phys Rev Lett. 1977;39(2):95-8.
60. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, et al. Random-coil behavior and the dimensions of chemically unfolded proteins. Proc Natl Acad Sci USA. 2004;101(34):12491-6.
61. Daughdrill GW, Kashtanov S, Stancik A, Hill SE, Helms G, Muschol M, et al. Understanding the structural ensembles of a highly extended disordered protein. Mol Biosyst. 2012;8(1):308-19.
62. Flory PJ. Principles of polymer chemistry. Ithaca,: Cornell University Press; 1953. 672 p.
63. Xie M, Li D-W, Yuan J, Hansen AL, Brüschweiler R. Quantitative binding behavior of intrinsically disordered proteins to nanoparticle surfaces at individual residue level. Chem-Eur J. 2018;24(64):16997-7001.
64. Adamski W, Salvi N, Maurin D, Magnat J, Milles S, Jensen MR, et al. A Unified Description of Intrinsically Disordered Protein Dynamics under Physiological Conditions Using NMR Spectroscopy. J Am Chem Soc. 2019;141(44):17817-29.
65. Lienin SF, Bremi T, Brutscher B, Brüschweiler R, Ernst RR. Anisotropic intramolecular backbone dynamics of ubiquitin characterized by NMR relaxation and MD computer simulation. J Am Chem Soc. 1998;120(38):9870-9.
66. Wong TS, Rajagopalan S, Freund SM, Rutherford TJ, Andreeva A, Townsley FM, et al. Biophysical characterizations of human mitochondrial transcription factor A and its binding to tumor suppressor p53. Nucleic Acids Res. 2009;37(20):6765-83.
67. Shan B, Li DW, Bruschweiler-Li L, Brüschweiler R. Competitive Binding between Dynamic p53 Transactivation Subdomains to Human MDM2 Protein: implications for regulating the p53·MDM2/MDMX interaction. J Biol Chem. 2012;287(36):30376-84.
68. Lum JK, Neuweiler H, Fersht AR. Long-Range Modulation of Chain Motions within the Intrinsically Disordered Transactivation Domain of Tumor Suppressor p53. J Am Chem Soc. 2012;134(3):1617-22.
69. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, et al. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science. 1996;274(5289):948-53.
70. Wang T, Darwin KH, Li H. Binding-induced folding of prokaryotic ubiquitin-like protein on the *Mycobacterium* proteasomal ATPase targets substrates for degradation. Nat Struct Mol Biol. 2010;17(11):1352-7.
71. Lee H, Mok KH, Muhandiram R, Park K-H, Suk J-E, Kim D-H, et al. Local Structural Elements in the Mostly Unstructured Transcriptional Activation Domain of Human p53. J Biol Chem. 2000;275(38):29426-32.
72. Lowry DF, Stancik A, Shrestha RM, Daughdrill GW. Modeling the accessible conformations of the intrinsically unstructured transactivation domain of p53. Proteins. 2008;71(2):587-98.
73. Xie M, Hansen AL, Yuan J, Brüschweiler R. Residue-Specific Interactions of an Intrinsically Disordered Protein with Silica Nanoparticles and their Quantitative Prediction. J Phys Chem C. 2016;120(42):24463-8.
74. Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, et al. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc Natl Acad Sci USA. 2008;105(15):5762-7.

75. Huang F, Rajagopalan S, Settanni G, Marsh RJ, Armoogum DA, Nicolaou N, et al. Multiple conformations of full-length p53 detected with single-molecule fluorescence resonance energy transfer. Proc Natl Acad Sci USA. 2009;106(49):20758-63.

76. Kamagata K, Kanbayashi S, Honda M, Itoh Y, Takahashi H, Kameda T, et al. Liquid-like droplet formation by tumor suppressor p53 induced by multivalent electrostatic interactions between two disordered domains. Sci Rep. 2020;10(1).

77. Marsh JA, Forman-Kay JD. Sequence Determinants of Compaction in Intrinsically Disordered Proteins. Biophys J. 2010;98(10):2383-90.

78. Prompers JJ, Scheurer C, Brüschweiler R. Characterization of NMR relaxation-active motions of a partially folded A-state analogue of ubiquitin. J Mol Biol. 2001;305(5):1085-97.

79. Shrestha UR, Juneja P, Zhang Q, Gurumoorthy V, Borreguero JM, Urban V, et al. Generation of the configurational ensemble of an intrinsically disordered protein from unbiased molecular dynamics simulation. Proc Natl Acad Sci USA. 2019;116(41):20446-52.

80. Xu XP, Case DA. Automated prediction of ^{15}N, $^{13}C^{alpha}$, $^{13}C^{beta}$ and $^{13}C'$ chemical shifts in proteins using a density functional database. J Biomol Nmr. 2001;21(4):321-33.

81. Li D-W, Brüschweiler R. PPM: a side-chain and backbone chemical shift predictor for the assessment of protein conformational ensembles. J Biomol Nmr. 2012;54(3):257-65.

82. Li D-W, Brüschweiler R. NMR-based protein potentials. Angew Chem Int Edit. 2010;49(38):6778-80.

83. Brangwynne CP, Tompa P, Pappu RV. Polymer physics of intracellular phase transitions. Nat Phys. 2015;11(11):899-904.

84. Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowietz A, et al. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. Mol Cell. 2015;57(5):936-47.

85. Vernon RM, Forman-Kay JD. First-generation predictors of biological protein phase separation. Curr Opin Struct Biol. 2019;58:88-96.

86. Elbaum-Garfinkle S, Kim Y, Szczepaniak K, Chen CCH, Eckmann CR, Myong S, et al. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. Proc Natl Acad Sci USA. 2015;112(23):7189-94.

87. Tsang B, Arsenault J, Vernon RM, Lin H, Sonenberg N, Wang LY, et al. Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation. Proc Natl Acad Sci USA. 2019;116(10):4218-27.

88. Pak CW, Kosno M, Holehouse AS, Padrick SB, Mittal A, Ali R, et al. Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. Mol Cell. 2016;63(1):72-85.

89. Reichheld SE, Muiznieks LD, Keeley FW, Sharpe S. Direct observation of structure and dynamics during phase separation of an elastomeric protein. Proc Natl Acad Sci USA. 2017;114(22):E4408-E15.

90. Demir O, Barros EP, Offutt TL, Rosenfeld M, Amaro RE. An integrated view of p53 dynamics, function, and reactivation. Curr Opin Struct Biol. 2021;67:187-94.

91. Liu X, Chen J. Residual Structures and Transient Long-Range Interactions of p53 Transactivation Domain: Assessment of Explicit Solvent Protein Force Fields. Journal of chemical theory and computation. 2019;15(8):4708-20.

92. Herrera-Nieto P, Perez A, De Fabritiis G. Characterization of partially ordered states in the intrinsically disordered N-terminal domain of p53 using millisecond molecular dynamics simulations. Sci Rep. 2020;10(1).

93. Alberti S, Dormann D. Liquid-Liquid Phase Separation in Disease. Annu Rev Genet. 2019;53:171-94.
94. Dignon GL, Best RB, Mittal J. Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties. Annu Rev Phys Chem. 2020;71:53-75.
95. Petronilho EC, Pedrote MM, Marques MA, Passos YM, Mota MF, Jakobus B, et al. Phase separation of p53 precedes aggregation and is affected by oncogenic mutations and ligands. Chem Sci. 2021;12(21):7334-49.
96. Case DA, Betz RM, Cerutti DS, Cheatham III TE, Darden TA, Duke RE, et al. AMBER 2016. University of California, San Francisco. 2016.
97. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1-2:19-25.
98. Rawdon EJ, Kern JC, Piatek M, Plunkett P, Stasiak A, Millett KC. Effect of Knotting on the Shape of Polymers. Macromolecules. 2008;41(21):8281-7.
99. Rudnick J, Gaspari G. The aspherity of random walks. J Phys A. 1986;19(4):L191-L3.
100. Diehl HW, Eisenriegler E. Universal shape ratios for open and closed random walks: exact results for all *d*. J Phys A. 1989;22(3):L87-L91.
101. Bremi T, Brüschweiler R, Ernst RR. A Protocol for the Interpretation of Side-Chain Dynamics Based on NMR Relaxation: Application to Phenylalanines in Antamanide. J Am Chem Soc. 1997;119(18):4272-84.
102. Wangsness RK, Bloch F. The dynamical theory of nuclear induction. Phys Rev. 1953;89(4):728-39.
103. Bloch F. Dynamical theory of nuclear induction. II. Phys Rev. 1956;102(1):104-35.
104. Redfield AG. On the theory of relaxation processes. Ibm J Res Dev. 1957;1(1):19-31.
105. Abragam A. The Principles of Nuclear Magnetism: Clarendon Press; 1961.