PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Phys. Chem. Chem. Phys.*, 2018, **20**, 8676

Characterization of the structural ensembles of p53 TAD2 by molecular dynamics simulations with different force fields†

Yanhua Ouyang, Likun Zhao and Zhuqing Zhang *

Intrinsically disordered regions (IDRs) or proteins (IDPs), which play crucial biological functions in essential biological processes of life, do not have well-defined secondary or tertiary structures when isolated in solution. The highly dynamic properties and conformational heterogeneity of IDPs make them challenging to study with traditional experimental techniques. As a powerful complementary tool for experiments, all-atom molecular dynamics simulation can obtain detailed conformational information on IDPs, but the limitation of force field accuracy is a challenge for reproducing IDP conformers. Here, we compared five empirical all-atom force fields AMBER03, AMBER99SB-ILDN, CHARMM27, OPLS-AA/L and CHARMM36m in modeling the conformational ensembles of wild-type peptide TAD2(41–62) from the human p53 tumor suppressor. Our results show that for the model peptide, the newest force field CHARMM36m produces more expanded coil ensemble followed by AMBER99SB-ILDN; CHARMM27 displays a predominant propensity for a helical structure; whereas OPLS-AA/L exhibits a apparent preference for a β-sheet structure and yields the most compact conformation. In the comparison of the simulated dimensions with theoretical prediction and the back-calculated chemical shifts with experimental measurements, AMBER99SB-ILDN gives a more consistent agreement than the other force fields. In addition, the region from residues 47 to 55, which commonly forms an amphipathic α-helix upon binding target proteins according to experimental observation, could form a helical structure with a different probability population in our simulations with different force fields. This implies that the binding process might be conducted by, or partly by "conformation selection" for this peptide. This work indicates that force field development for modeling general IDPs accurately has a long way to go, and more detailed experimental data of IDPs are also in demand.

Received 4th January 2018,
Accepted 5th March 2018

DOI: 10.1039/c8cp00067k

rsc.li/pccp

Introductions

Intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) in structured proteins are estimated to make up approximately one-third of proteins encoded in the human genome.[1–3] IDPs lack stable well-defined three-dimensional structures when isolated under physiological conditions, but play crucial biological functions in cellular processes including transcription and translation regulation, signaling transduction, protein phosphorylation and intracellular phase separation.[4] Many IDPs are associated with various human diseases ranging from cancer to neurodegenerative diseases.[3] The flexible feature of IDPs provides the advantage that many of them can bind with particular or various target molecules to perform their physiological functions. The binding process is deemed to be coupled folding and binding, by conformation selection or induced fitting or both through a "fly-casting" mechanism.[5] Compared with folded proteins, IDPs usually have more polar, charged and fewer hydrophobic amino acids in their sequences, as well as high conformational heterogeneity, which leads to the concept of a conformational ensemble proposed to understand their structure–function relations.

Characterizing the conformational ensembles of IDPs accurately is crucial for understanding their functions but quite challenging both experimentally and computationally. In experimental techniques, nuclear magnetic resonance (NMR), small-angle X-ray scattering (SAXS), as well as single-molecule fluorescence resonance energy transfer (smFRET) have been intensively applied in IDP studies, but the observables are usually ensemble-averaged over the interconverting conformations of IDPs, and their heterogeneous features are difficult to detect. In aspect of simulations, a variety of computational methods, in particular atom-level molecular dynamics (MD) simulations, have been increasingly applied to reproduce the conformational ensembles of IDPs due to the fact that atomic

---

College of Life Science, University of Chinese Academy of Sciences, Beijing, 100049, China. E-mail: zhuqingzhang@ucas.ac.cn

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8cp00067k

detailed information could be presented. However, atom-level MD simulations have two shortcomings: being time-consuming for sampling adequate conformations and limitation of force field accuracy. For the former, the rapid development of computing hardware facilities and sampling algorithms[6,7] have significantly alleviated the time scale limitation of simulations. For the latter, the classic force fields have been verified to be accurate enough in the application of folded proteins with a well-defined native structure.[8,9] However, it has been questioned whether they could be transferable to an IDP system since a number of reports have indicated that the simulated unstructured protein ensembles in an explicit solvent are usually much more compact than those in experimental observations.[10–15] In recent years, considerable attention has been focused on improving the force fields for IDPs.[12,16–22] The modifications include the optimization of backbone or sidechain torsion parameters, such as ff99IDPs[19] and ff14IDPSFF;[17] the optimization of water model, such as TIP4P-D;[23] or combining both strategies.[12] Recently, the residue-specific force field RSFF1[21] and its new version RSFF2+,[20,22] which includes statistical coil information based on protein data bank (PDB), have been reported to obtain a more rational prediction for IDPs. In addition, the newest developed force field termed CHARMM36m,[16] by refining the backbone CMAP potential derived from reweighting the calculation and a better description of specific salt bridge interactions, has been validated in both several model folded proteins and IDPs. In fact, whether or not these modified force fields are universal for general IDPs, as well as for folded proteins, is still unknown. Here we compared five force fields ranging from traditional force fields to the newest one by characterizing the conformational ensembles of peptide p53 TAD2 and evaluated these force fields' performance by comparing them with related theoretical predictions and experimental results.

Tumor suppressor p53 is a transcription factor, which has been found to regulate many processes related to DNA repair, cell cycle arrest, senescence, apoptosis, autophagy and metabolism.[24,25] The full-length p53 consists of an N-terminal transactivation domain (TAD), followed by a proline-rich region (PRR), the central DNA-binding domain (DBD), the tetramerization domain (TET), and the C terminal regulation domain (CT).[24] The free TAD is intrinsically disordered, and usually divided into two ill-defined subdomains, TAD1 (residues 1–40) and TAD2 (residues 40–61),[24] both of which have been found to bind various targets in their functional performance. Experimental measurements reveal that in bound complexes, an amphipathic helix forms frequently in both TAD1 and TAD2.[26–35] Furthermore, a recent study shows that TAD2 can form an extended string-like conformation with the human TFIIH subunit p62 PH domain.[36] The investigation on the structural features and dynamics of TAD1 and TAD2 is of great significance not only for revealing the functional mechanism of protein p53, but also for understanding of the general structure–function relationship of IDPs on account of their typicality. In the report of Mittal and coworkers, the NMR data of the p53 TAD1 fragment (residues 15–29) are consistent with those of full-length p53 TAD,[75] which means that the sub-domain interactions between TAD1 and TAD2 might not be significant. In this work, we used TAD2 of protein p53 as a model peptide in our atomic molecular simulations, and compared its target-free simulated conformational features with the theoretically predicted values, as well as the experimental NMR measurements of free p53 TAD (1–73)[69] and its bound conformations in complexes.

In the five selected force fields, AMBER03,[37] CHARMM27[38] and OPLS-AA/L[39] have been extensively applied in the study of globular proteins,[40] such as protein folding[41] or drug design;[42] AMBER99SB-ILDN has been reported to present a better prediction in a couple of IDP studies;[43–45] the newest developed CHARMM36m was reported to improve the accuracy both in conformational sampling for IDPs and in stability prediction for the studied model folded proteins.[16] Our work suggests that for the peptide p53 TAD2, AMBER99SB-ILDN generates conformations more in line with theoretical prediction, as CHARMM36m produces much more extended coil conformations and the other force fields exhibit more compact structures. On comparing the back-calculated chemical shifts with data from NMR experiments, it indicates that AMBER99SB-ILDN gives a more consistent result for p53 TAD2. This work will offer useful guidance for further improving the accuracy of force fields for IDPs.

Models and methods

In this work, five force fields—AMBER03,[37] CHARMM27[38] OPLS-AA/L,[39] AMBER99SB-ILDN[46] and CHARMM36m,[16] were selected for comparison. Except CHARMM36m, which was combined with the CHARMM-modified TIP3P water model, the classical TIP3P water model[47] was used in simulations with the other force fields. All the simulations were performed with GROMACS 5.0.2, and the structure schematics were presented by PYMOL. An extended conformation of human wild-type p53 TAD2 (^{41}DDLMLSPDDIEQWFTEDPGPDE^{62}), was built by the *tleap* program in the AmberTools15 package.[48] To get random conformations as initial structures for simulations, a short 4 ns MD simulation was performed at 600 K (with the AMBER99SB-ILDN force field and TIP3P water model). The initial structures for all the simulations were taken from this high temperature trajectory randomly. The periodic rhombic dodecahedral box was used for solvating p53 TAD2 with a box vector length of 7.5 nm, within which around 9700 water molecules were added, as well as 9 Na^+ to keep the simulated system neutralized. The Particle Mesh Ewald (PME)[49] was used for long-range electrostatic interactions with a real-space cutoff of 10 Å, and the same cutoff distance was adopted for van der Waals interactions. The bond lengths with hydrogen atoms were constrained by the LINCS algorithm.[50] For each simulation, after energy minimization by the steepest descent algorithm, a short *NVT* simulation for 200 ps and a *NPT* simulation for 1 ns were performed at 300 K for pre-equilibration. Then a 1020 ns *NVT* productive simulation was performed. The V-rescale algorithm[51] and Parrinello–Rahman barostat[52] were adopted to keep the temperature at 300 K and the pressure at 1 bar for the *NVT* and

NPT simulations. The time step in all the simulations was set as 2 fs, and simulated data were saved every 2 ps. For each force field, to obtain enough sampled conformations, three independent conventional molecular dynamics (CMD) simulations with different initial structures were performed, and the trajectory of the final 1 μs in each simulation was used for analysis.

In the present work, to confirm the sampling adequacy of conventional molecular dynamics (CMD) simulations, we performed a replica-exchange molecular dynamics (REMD) simulation for p53 TAD2 using the AMBER99SB-ILDN force field, with a total of 67 temperature replicas ranging from 270 K to 500 K. Adjacent replica exchanges were attempted every 2 ps and the mean exchange acceptance probability was above 0.2. After discarding the beginning 20 ns for each replica, the productive 200 ns trajectories were collected for analysis. The comparison of the conformation distribution between CMD (using the conformations sampled from all three independent trajectories) and REMD (using conformations sampled at 300 K) is shown in Fig. S1 (ESI†). The distribution of the radius of gyration (R_g), as well as the averaged secondary structure element analysis, indicate that the sampling of CMD is rational and adequate, in accordance with the energy landscape of the intrinsically disordered peptides being relatively flat, with multifunnel energy minima separated by lower-energy barriers.[53] In the present work, we used the sampled conformations from a total of 15 μs CMD simulations, 3 μs from three independent trajectories for each force field, for analysis.

Most of the analysis work was done using GROMACS software and in-house tools. Clustering was carried out by *gmx_cluster* based on the GROMOS algorithm[54] in GROMACS, with a backbone RMSD cutoff of 0.4 nm at a time interval of 30 ps for each system. The DSSP algorithm[55] was used to analyze secondary structure content. Contact order (CO)[56] was calculated for each saved conformation based on the definition in ref. 56 and residue contact was defined as the distance between any pair of heavy atoms in two amino acids within a cutoff of 0.45 nm separated by more than 3 residues. The end-to-end distance (R_{ee}) was defined as the distance between the first and the last C_α atoms. To compare with experimental NMR data, calculated chemical shifts and residual dipolar couplings (RDCs) were conducted with the SPARTA+[57] and PALES[58] programs respectively, using the coordinates saved every 500 ps from the simulations with each force field because the different time intervals 50 ps, 100 ps, 200 ps, and 1000 ps presented similar results.

Results and discussion

The dimensions of p53 TAD2 with different force fields

In order to characterize the average dimension of p53 TAD2 conformational ensembles generated by the five force fields, we firstly analyzed the distribution of the radius of gyration R_g and the end-to-end distance R_{ee}, as shown in Fig. 1A and B, respectively. This indicates that the simulated peptide dimensions depend strongly on the force field. Both figures show that CHARMM36m produces the most extended conformations, which is followed by AMBER99SB-ILDN, then CHARMM27. Ensembles simulated with AMBER03 and OPLS-AA/L exhibit much more compact structures than the others and both have two evident peaks located at R_g smaller than 1 nm. The two-dimensional probability distribution of conformational ensembles in Fig. 1C also indicates multiple populated regions for the AMBER03 and OPLS-AA/L force fields. Meanwhile, CHARMM36m and AMBER99SB-ILDN produce more heterogeneous conformations than the other three force fields due to much wider populated regions. The decoupling of R_g and R_{ee}, which was proposed to explain the inconsistency in results measured from small-angle X-ray scattering (SAXS) and single-molecule Forster resonance energy transfer (smFRET) respectively,[59,60] is observed in our simulations with AMBER03, AMBER99SB-ILDN and CHARMM27 for p53 TAD2. To compare the average separated sequence length between two contact residues, we plotted the conformational distribution using contact order (CO)[56] and R_g as order parameters, as shown in Fig. S2 (ESI†). It shows that a larger fraction of local contacts are observed in the simulations with CHARMM36m (around more than 50% of all sampled conformations are extended without any contact) with small CO values, while more long-range contacts are detected in simulations with OPLS-AA/L and AMBER03, consistent with the observations from R_{ee} measurements.

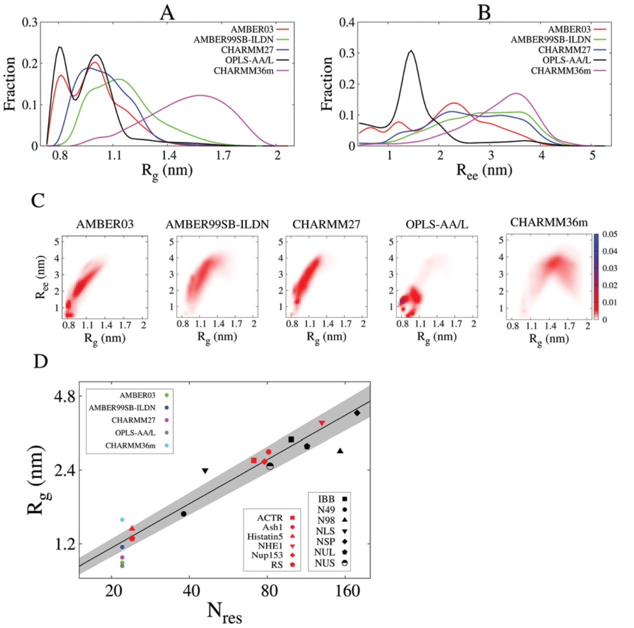


Fig. 1 Comparison of dimensions. (A) Distributions of radius of gyration R_g. (B) Distributions of end-to-end distance R_{ee}. (C) Two-dimensional probability distribution as a function of R_g and R_{ee} for the five force fields. (D) Measured R_g as a function of residue number N_{res}. The black filled and semi-filled points present data from SAXS experimental measurements published in ref. 59. The red filled points come from the experimental measured data of ACTR,[65] Ash1,[64] Histain5,[66] NHE1,[65] Nup153[63] and RS.[14] The green, blue, magenta, grey and cyan filled circle points denote the simulated average R_g with AMBER03, AMBER99SB-ILDN, CHARMM27, OPLS-AA/L and CHARMM36m, respectively. The black line and grey band show the power-law relationship[67] function $R_g = R_0 \times N^{\nu}$ with R_0 = 1.927 Å and ν = 0.598 ± 0.028.

Most IDPs have more polar and charged residues than folded proteins in their sequences,[61] therefore they are considered to possess more extended conformations.[62] An increasing number of SAXS experimental measurements have been reported for numerous peptides recently. Here, we collected the R_g values of a couple of peptides[14,59,63–66] from SAXS measurements and fitted them with the power-law relation $R_g \sim N^{\nu}$ from polymer model, where N denotes the residue number of peptide and ν represents the scaling factor. The black and red filled points in Fig. 1D indicate that, like unfolded proteins,[67] the sizes of various IDPs can also roughly be described by the self-avoiding random walk (SARW) model,[67] since the shaded grey region corresponding to the value of ν as 0.598 ± 0.028, covers most of the experimental data. The mean size, R_g, of the sampled conformations for the peptide p53 TAD2, is shown as a different colored circle filled point for each of the five force fields in Fig. 1D. For this peptide with 22 residues, the theoretically estimated R_g value is around 1.122–1.334 nm. Overall, Fig. 1D shows that the force field CHARMM36m probably produces conformations much more extended, while AMBER03, OPLS-AA/L and CHARMM27 might generate conformations too compact. Conformations sampled with AMBER99SB-ILDN might be more rational than those from other force fields referring to the blue filled circle point covered by the shaded gray region of the theoretical relationship in Fig. 1D. This implies that although a number of studies suggest the classical force fields usually sample more collapsed conformations,[11–14,68] for the peptide p53 TAD2, AMBER99SB-ILDN and the TIP3P water model might possibly produce rational dimensions, and this needs more experiments to validate it. In the meantime, the new modified CHARMM36m has been improved for several model IDPs,[16] but the results shown here indicate more efforts may still be needed for force field development for general IDPs.

Conformational ensemble analysis

To characterize the structural properties of p53 TAD2, we compared the secondary structure content analysis of the ensembles obtained by simulations with different force fields. Fig. S3 (ESI†) shows the secondary structure assessment for all trajectories with the DSSP algorithm.[55] For each force field, one independent trajectory may show distinct structural properties from another one, therefore, we combined all three trajectories for statistical analysis, since the sampled conformations from all of them could be comparable with those from REMD as validated by AMBER99SB-ILDN (see "Models and methods" section). Like the chain dimension, the secondary structure content also displays an apparent difference with the studied five force fields for the peptide p53 TAD2, as shown in Fig. 2. Firstly, it reveals that coil is the largest fraction in simulations with all five force fields, especially for CHARMM36m. The conformational ensemble using CHARMM27 has a much larger fraction of α-helix than the other force fields, reaching to almost 40%. Besides, in the ensemble simulated with OPLS-AA/L, more bend and β-sheet contents, and almost no α-helix are detected, although the standard deviation of β-sheet content displays a slightly larger value. Structures obtained by both AMBER03 and AMBER99SB-ILDN have a small fraction of α-helix, as well as a little larger fraction of turn and bend.

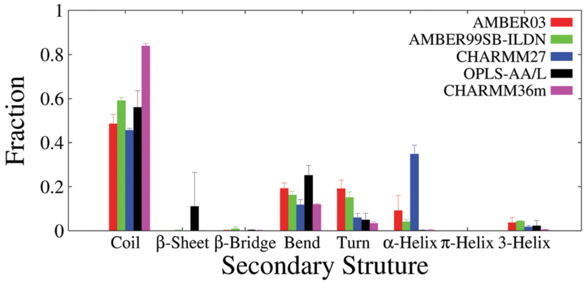


Fig. 2 The averaged fraction of residues assigned to different types of secondary structure elements based on the DSSP algorithm.[55] The standard deviation was obtained based on the average values of three independent trajectories for each force field. 3-Helix indicates 3_{10}-helix in this plot.

To further investigate the specific location of every secondary structure element in the sequence, the secondary structure propensity for each amino acid was statistically calculated based on the simulations with each studied force field. Fig. 3A reveals that the helical structures are mainly located in the region from residue 47 to 56 in ensembles sampled with CHARMM27, AMBER03 and AMBER99SB-ILDN. In fact, this region has been reported to form an α-helix in many complex structures when p53 TAD2 binds with other partners.[26,29,31,32] The population of the preexisting bound conformations in free-state ensemble is essential for determining the binding mechanism of IDPs. A more detailed discussion will be presented in the next section. The comparison of other secondary structure propensity with studied force fields for each residue is shown in Fig. S4 (ESI†), in which the β-sheet is only shown in simulations sampled with OPLS-AA/L apparently, and located in the regions of residues 43–47 and 52–56; the turn regions detected in ensembles with both AMBER03 and AMBER99SB-ILDN are located around residues from 42 to 45 and from 47 to 56, partly consistent with experimental measurements;[69] the coil occupies almost all residues in structures sampled with CHARMM36m. Although an evident distinction is revealed in the secondary structure propensity for specific residues with the five force fields, the Root Mean Square Fluctuation (RMSF) shown in Fig. 3B (taking the averaged conformation as a reference structure, which was obtained from no less than 90% of the whole sampled conformations, by clustering based on backbone RMSD with a cut off around 7 Å) demonstrates that for most force fields, the residue region from 45–58 shows a relatively lower mobility compared with N- and C-termini, and those residues with CHARMM27 present the lowest RMSF values, which means the structures that they formed in simulations are quite stable.

In a previous study[70] on β-hairpin folding of the nuclear factor erythroid 2-related factor 2 (Nrf2), a transcription factor regulating the expression of genes responsive to oxidative stress, it was observed that CHARMM27 and OPLS-AA/L were heavily biased towards α-helical and bend conformations, respectively. Additionally, earlier MD simulations of peptides including Aβ and short polyalanine, also revealed that AMBER03

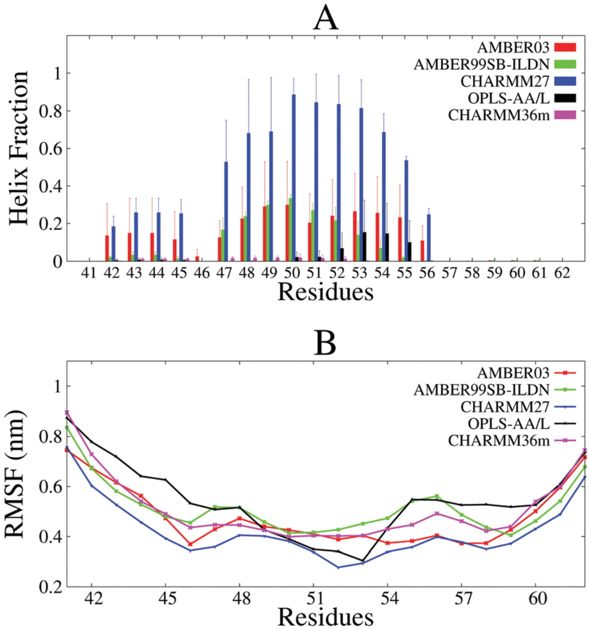


Fig. 3 Characteristic of each residue of p53 TAD2 in simulations. (A) Helix occupancy of each residue based on the DSSP algorithm.[55] The standard deviation was obtained based on the average values of three independent trajectories for each force field. Helix here includes α-helix, π-helix and 3_{10}-helix. (B) Root mean square fluctuation (RMSF) of each residue in simulations with five force fields, taking the averaged conformation of no less than 90% of the whole samples (obtained by clustering with the backbone RMSD cutoff at around 7 Å) as a reference structure.

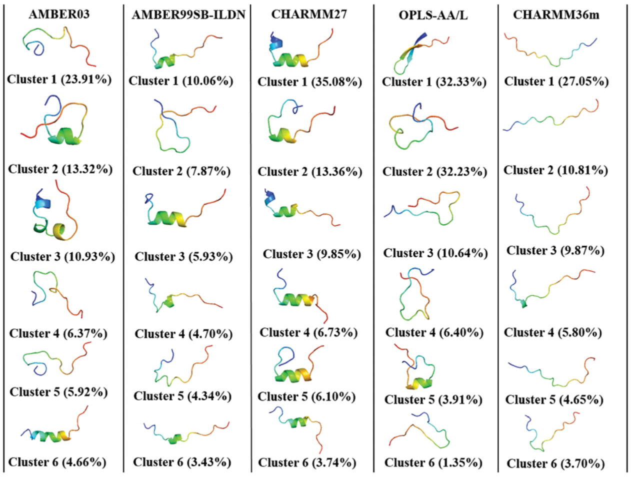


Fig. 4 Representative cartoon schematics for the top populated six clusters from simulations with five force fields, along with the percentage occupancy of the corresponding cluster. The cutoff of backbone RMSD is 0.4 nm for all force fields. The structures are colored from blue to red gradually for the peptide from N-termini to C-termini.

and CHARMM27 (CHARMM22 + CMAP) tended to over-stabilize helical structures,[43,71,72] and OPLS-AA/L yielded the most compact structures in modeling the intrinsically disordered Aβ peptide.[73] Moreover, in another investigation using several intrinsically disordered peptides as model systems, Rauscher *et al.* systematically compared couples of force fields, and concluded that force field was a stronger determinant of secondary structure content than peptide sequences.[14] In this work, using p53 TAD2 as a model peptide, our simulations reveal similar trends, which indicates that OPLS-AA/L favors β-sheet structures, CHARMM27 tends to over-stabilize α-helix, AMBER03 and AMBER99SB-ILDN also prefer to generate α-helix but with a smaller propensity than CHARMM27; the new improved CHARMM36M produces more coil conformations. To determine whether or not these are general features of these force fields, more investigations on different type of IDPs and folded proteins may need to be performed.

The conformation clustering analysis shown in Fig. 4 presents more visualized structural details for the simulated ensembles of the peptide p53 TAD2. Using the backbone RMSD cutoff value of 0.4 nm, the total occupancies of the top six clusters for AMBER99SB-ILDN, AMBER03, CHARMM27, OPLS-AA/L and CHARMM36m are 65.10%, 36.33%, 74.85%, 86.86% and 61.87%, respectively. Consistent with the secondary structure analysis (Fig. 2), clusters obtained with CHARMM27 possess an obviously large fraction of helical structure. Whereas, those with OPLS-AA/L are distinctively occupied by a large number of β-sheet conformations, both AMBER03 and AMBER99SB-ILDN force fields produce a medial population of helical structure, and simulations with CHARMM36m predominantly exhibit expanded coil structures. From the whole point of view based on the occupancy of each cluster in Fig. 4, AMBER99SB-ILDN produces more heterogeneous conformations than the other force fields. The plotted Ramachandran map based on free energy profiles in Fig. S5 (ESI†) gives more conformational details. This indicates that except for the structural features mentioned above, conformations sampled with the newly developed force field CHARMM36m present a low population of left-handed α-helix (defined as $30^\circ < \varphi < 100^\circ$ and $7^\circ < \psi < 67^\circ$) as was reported,[16] meanwhile exhibiting much more polyproline II conformations ($90^\circ < \varphi < -20^\circ$ and $50^\circ < \psi < 180^\circ$), a left-handed helix that is found extensively distributed in disordered proteins/peptides.[74]

Comparison with experimental measurements

NMR chemical shifts measured on different nuclei are sensitive to chemical environment, which is determined by the special structure of proteins, and thus may provide independent information on the conformational ensembles. In this work, we calculated the chemical shifts H_α and H^N of p53 TAD2 based on our simulated trajectories, by using the widely applied software SPARTA+, and compared them with experimental NMR monitored values[69] to assess the five force fields' performance. According to the report of Mittal and coworkers, the NMR data of the p53 TAD fragment (residues 15–29) are consistent with those of full-length p53 TAD.[75] Here, we also use the available experimental data of free p53 TAD (1–73)[69] for comparison, assuming that the fragment TAD2 that we studied also has the same NMR chemical shifts. The simulated values from five studied force fields were plotted against experimental data as shown in Fig. 5. The predicted H_α chemical shifts of the sampled ensembles agree well with experimental ones, weakly depending on force field. However, the deviations between the calculated and experimental H^N chemical shifts are much

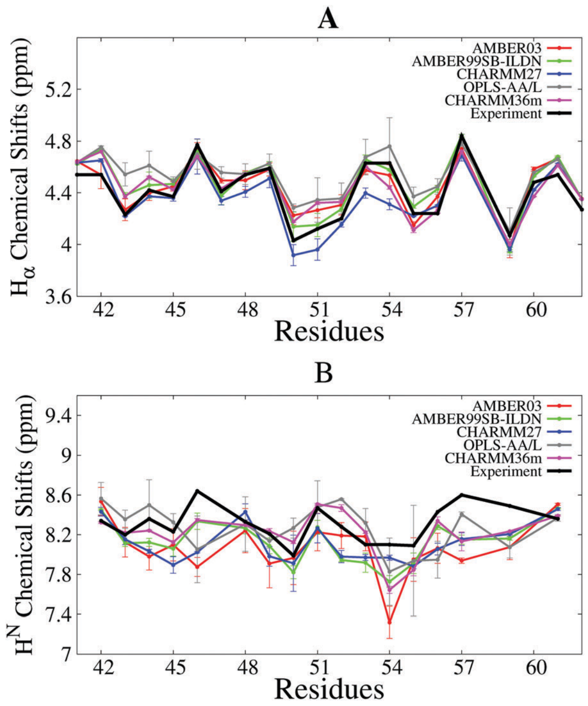


Fig. 5 Calculated (A) H_α and (B) H^N chemical shifts based on sampled conformations and experimental measured data.[69] For each residue, the standard deviation was obtained based on the average values of three independent trajectories for each force field.

larger than H_α and the differences among force fields are distinct. The standard deviations of H^N chemical shifts calculated from the averaged values of three independent trajectories are also a little larger than those of H_α. To quantify to what extent the simulation results can match experimental measurements, we calculated the RMS difference and Pearson's Correlation Coefficient (PCC) between simulated data and experimental ones, which are shown in Table 1. By evaluating both the H_α and H^N chemical shifts, it shows, apart from CHARMM36m giving a slightly smaller RMS difference for chemical shifts of H^N, generally, AMBER99SB-ILDN presents a better agreement with NMR measurements, showing a relatively lower RMS difference and a slightly higher Pearson's Correlation Coefficient.

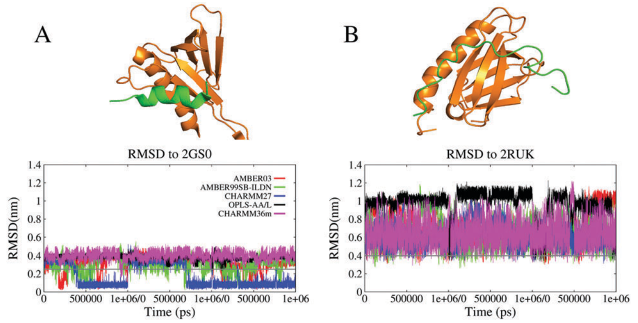
We also attempted to calculate the Residual Dipolar Couplings (RDCs), which provide long-range interaction information and are highly sensitive to solution conditions such as medium alignment, to evaluate the studied force fields by comparing with experimental data. RDCs were back-calculated using both stPales and bestFit modules in PALES program.[58] We found that the predicted RDCs using both modules based on simulated ensembles with all force fields are in poor line with experimental RDCs[25] (Fig. S6, ESI†), with RMS differences larger than 5 and PCC values reaching almost random for all force fields. Earlier studies for unstructured protein systems also reported a similar poor correlation between calculated and experimental RDCs.[10,14] Therefore, efforts for developing more accurate predictors for RDCs are needed.

Table 1 The Pearson Correlations Coefficient (PCC) and RMS difference between the calculated H_α and H^N chemical shifts and experimentally measured data[69]

	H_α chemical shifts		H^N chemical shifts	
Force field	RMS difference	PCC	RMS difference	PCC
AMBER03	0.095	0.912	0.372	0.195
AMBER99SB-ILDN	0.094	0.936	0.235	0.713
CHARMM27	0.121	0.885	0.278	0.445
OPLS-AA/L	0.151	0.906	0.265	0.150
CHARMM36m	0.111	0.870	0.209	0.522

Many studies have revealed that a lot of IDPs can interact with other proteins and fold a specific structure in the bound complex, and they may form transient local secondary structures in target-free state although mainly intrinsically disordered. The pre-existing transient secondary structure might determine the association mechanism when IDPs bind to their targeted partners to perform biological functions, but it is quite challenging to detect this experimentally. For full length p53 TAD, it has been proposed that target-free p53 TAD might form two nascent turns in TAD2 (residue 40–44 and 48–53).[69] Meanwhile, the continuity of sequential d_{NN} NOEs has been detected in two regions, near residues 39–45 and 48–55 of p53 TAD2, which implies that these two regions have a helical propensity in the free-state. Furthermore, a couple of investigations revealed that p53 TAD2 can form a stable amphipathic α-helix structure on binding to its targets.[26,29,31–33] One exception has been reported that TAD2 binds to human TFIIH subunit p62 PH domain through a different conformation—an extended string-like structure.[36] In our simulations with AMBER03, AMBER99SB-ILDN and CHARMM27 force fields, to more or less extent, the sampled conformations exhibit helix or turn conformations, as discussed above.

To investigate whether p53 TAD2 has probably already formed the target-binding conformation in its free state, we adopted two different TAD2 complex structures observed in experiments, with the PDB ID codes 2GS0[32] and 2RUK[36] (top panels of Fig. 6), as reference structures, to analyze RMSD based on C_α from the simulations with the five studied force fields. The RMSD values with time evolution are shown in the middle panels in Fig. 6. In the complex of yeast tfb1 PH domain (PDB ID 2GS0), residues 47–55 of p53 TAD2 form a short α-helix. The bottom panel of Fig. 6A presents that the total occupancies of the simulated conformations with RMSD < 2.5 Å are 13.25%, 21.13%, 64.09%, 0.02% and 0.49% for the force field AMBER03, AMBER99SB-ILDN, CHARMM27, OPLS-AA/L and CHARMM36m, respectively. In fact, the clusters shown in Fig. 4 and secondary structure propensity analysis shown in Fig. 3 and Fig. S4 (ESI†), have exhibited that the α-helix formed in the simulations with force field AMBER03, AMBER99SB-ILDN and CHARMM27 mainly locates in this region. AMBER99SB-ILDN, which may be the more rational force field for p53 TAD2 based on the predicted dimensions mentioned above, as well as the comparison of chemical shifts in this study, produces conformations with around 21% of pre-existing α-helix, which implies that p53 TAD2 might adopt a conformational selection mechanism to bind its target, or maybe with a combination of conformational selection and induced fit. On the other hand, when aligned to the extended string-like conformation in complex



Force fields	Percentage
AMBER03	13.25%
AMBER99SB-ILDN	21.13%
CHARMM27	64.09%
OPLS-AA/L	0.02%
CHARMM36m	0.49%

Force fields	Percentage
AMBER03	0.74%
AMBER99SB-ILDN	1.12%
CHARMM27	0.37%
OPLS-AA/L	0.15%
CHARMM36m	0.45%

Fig. 6 Comparison of sampled conformations with those in complexes. (A) Top: Cartoon structure of p53 TAD2 (green)/yeast tfb1 PH domain (orange) complex (PDB code 2GS0). Middle: Time evolution of C_α RMSD (residues 47–55) of p53 TAD2 referenced to the bound structure. Bottom: The percentage occupancy of conformations with RMSD < 2.5 Å for simulations with each force field. (B) Top: Cartoon structure of the p53 phosphorylated TAD2 (green)/p62 PH domain (orange) complex (PDB code 2RUK). Middle: Time evolution of C_α RMSD (residues 41–62) of p53 TAD2 referenced to the bound-state structure. Bottom: The percentage occupancy of conformations with RMSD < 4 Å for simulations with each force field. In the two middle panels, "1e + 06/0" in the *x*-axis means the end of the former 1 μs trajectory and the beginning of the latter trajectory.

with PDB ID 2RUK (top panel of Fig. 6B), nearly all RMSD values are higher than 3 Å. Taking RMSD < 4 Å, the total occupancies for each force field are also very low as shown in the bottom panel of Fig. 6B. Only AMBER99SB-ILDN produces a fraction of conformations larger than 1%. Although CHARMM36m force field could sample more extended conformations, the occupancy indicates that the sampled structures are not much like those in the complex. Therefore, for this situation, the binding process might be conducted by binding induced folding.

Taken together, by comparing the simulations with reported experimental data for p53 TAD2, we found that among the five force fields studied here, simulations with AMBER99SB-ILDN are more in line with experimental results. The target-free sampling simulations could provide a possible binding mechanism, while more association kinetics investigations both in experiments and in theoretical simulations are required to uncover the mechanism of this basic process.

Conclusions

Intrinsically disordered proteins have increasingly been realized to play a crucial role in many essential biological processes, ranging from molecular signaling to the formation of membraneless organelles. Obtaining accurate descriptions of IDPs by means of MD simulations is significant but quite challenging due to the limitation of force field accuracy. Here, we used five atomistic force fields to perform molecular dynamics simulations to characterize the conformational features of p53 TAD2. Our results demonstrate that the simulated structural features of p53 TAD2 are sensitive to force field: CHARMM36m produces a most expanded coil ensemble; CHARMM27 tends to over-stabilize helical structure; OPLS-AA/L exhibits a strong preference for β-sheet structure and yielded the most compact conformations. According to the theoretical power-law relationship of R_g *versus* peptide length,[67] for p53 TAD2, most force fields studied here exhibit rather collapsed conformations except the newest CHARMM36m and AMBER99SB-ILDN, the latter presenting a structural dimension more close to the theoretically predicted value. The clustering analysis indicates that AMBER99SB-ILDN could also generate more heterogeneous conformations than the other force fields. Furthermore, the comparison of the calculated chemical shifts with experimental NMR data implies that the conformations sampled with AMBER99SB-ILDN could provide a slightly more accurate prediction than those of the other force fields. On average, among the five investigated force fields, AMBER99SB-ILDN might be a more suitable one for simulating the peptide p53 TAD2, although earlier reports stated that AMBER99SB-ILDN with TIP3P model yielded overly collapsed IDP structures for other peptides.[68,76] This suggests that the force field might perform differently depending on the peptide sequences. Computational investigations of more sequences and more experimental studies are necessary for evaluating the performance of a force field.

As a force field is defined by potential energy functions and the corresponding parameters to describe bonded and non-bonded interactions between particles in a modeled system, the differing performance of the five contemporary force fields for protein in this study, should arise from the description forms and parameters for various interactions. In fact, the stability of a protein force field, is supposed to be transferable between folded protein and IDPs because of the same type of interactions.[77] In recent protein force field improvements, particularly for IDPs, optimizations of backbone or side chain torsion parameters[17,19] have been conducted to balance the propensity of sampled secondary structure. Meanwhile, including CHARMM36m, the rational water model and water–protein interactions have increasingly been recognized to be of significance.[12,16,23,77] Besides, more elaborate models, such as the atomic polarizable effect,[78] as well as many-body interactions which have been suggested to be non-negligible in protein folding,[79] may improve the prediction accuracy. Furthermore, not only the sampled conformations, but also intra-chain dynamics should also be considered.[80] Further studies on improving the force field by taking these fundamental elements into account would provide more insight into understanding the structure–function relationship of IDPs.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank Prof. Jizhong Lou for his helpful discussion, and Miss Keqi Xiang for her assistance in initial model building.

This work was supported by the National Science Foundation of China (No. 21633001, 31200548).

References

1 A. Bhowmick, D. H. Brookes, S. R. Yost, H. J. Dyson, J. D. Forman-Kay, D. Gunter, M. Head-Gordon, G. L. Hura, V. S. Pande, D. E. Wemmer, P. E. Wright and T. Head-Gordon, *J. Am. Chem. Soc.*, 2016, **138**, 9730–9742.

2 V. Csizmok, A. V. Follis, R. W. Kriwacki and J. D. Forman-Kay, *Chem. Rev.*, 2016, **116**, 6424–6462.

3 R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright and M. M. Babu, *Chem. Rev.*, 2014, **114**, 6589–6631.

4 H. J. Dyson and P. E. Wright, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 197–208.

5 Y. Huang and Z. Liu, *J. Mol. Biol.*, 2009, **393**, 1143–1159.

6 G. H. Zerze, C. M. Miller, D. Granata and J. Mittal, *J. Chem. Theory Comput.*, 2015, **11**, 2776–2782.

7 R. Zhou, *Methods Mol. Biol.*, 2007, **350**, 205–223.

8 K. Lindorff-Larsen, S. Piana, R. O. Dror and D. E. Shaw, *Science*, 2011, **334**, 517–520.

9 D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan and W. Wriggers, *Science*, 2010, **330**, 341–346.

10 F. Palazzesi, M. K. Prakash, M. Bonomi and A. Barducci, *J. Chem. Theory Comput.*, 2015, **11**, 2–7.

11 S. Piana, J. L. Klepeis and D. E. Shaw, *Curr. Opin. Struct. Biol.*, 2014, **24**, 98–105.

12 R. B. Best, W. Zheng and J. Mittal, *J. Chem. Theory Comput.*, 2014, **10**, 5113–5124.

13 J. Henriques, C. Cragnell and M. Skepo, *J. Chem. Theory Comput.*, 2015, **11**, 3420–3431.

14 S. Rauscher, V. Gapsys, M. J. Gajda, M. Zweckstetter, B. L. de Groot and H. Grubmuller, *J. Chem. Theory Comput.*, 2015, **11**, 5513–5524.

15 J. Hu, T. Chen, M. Wang, H. S. Chan and Z. Zhang, *Phys. Chem. Chem. Phys.*, 2017, **19**, 13629–13639.

16 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmuller and A. D. MacKerell Jr., *Nat. Methods*, 2017, **14**(1), 71–73.

17 D. Song, R. Luo and H. F. Chen, *J. Chem. Inf. Model.*, 2017, **57**, 1166–1178.

18 D. W. Li and R. Bruschweiler, *J. Chem. Theory Comput.*, 2011, **7**, 1773–1782.

19 W. Ye, D. Ji, W. Wang, R. Luo and H. F. Chen, *J. Chem. Inf. Model.*, 2015, **55**, 1021–1029.

20 C.-Y. Zhou, F. Jiang and Y.-D. Wu, *J. Phys. Chem. B*, 2015, **119**, 1035–1047.

21 F. Jiang, C.-Y. Zhou and Y.-D. Wu, *J. Phys. Chem. B*, 2014, **118**, 6983–6998.

22 H.-N. Wu, F. Jiang and Y.-D. Wu, *J. Phys. Chem. Lett.*, 2017, **8**, 3199–3205.

23 S. Piana, A. G. Donchev, P. Robustelli and D. E. Shaw, *J. Phys. Chem. B*, 2015, **119**, 5113–5123.

24 A. C. Joerger and A. R. Fersht, *Annu. Rev. Biochem.*, 2008, **77**, 557–582.

25 M. Wells, H. Tidow, T. J. Rutherford, P. Markwick, M. R. Jensen, E. Mylonas, D. I. Svergun, M. Blackledge and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 5762–5767.

26 J. P. Rowell, K. L. Simpson, K. Stott, M. Watson and J. O. Thomas, *Structure*, 2012, **20**, 2014–2024.

27 S. Rajagopalan, A. Andreeva, D. P. Teufel, S. M. Freund and A. R. Fersht, *J. Biol. Chem.*, 2009, **284**, 21728–21737.

28 C. W. Lee, M. Arai, M. A. Martinez-Yamout, H. J. Dyson and P. E. Wright, *Biochemistry*, 2009, **48**, 2115–2124.

29 E. Bochkareva, L. Kaustov, A. Ayed, G. S. Yi, Y. Lu, A. Pineda-Lucena, J. C. Liao, A. L. Okorokov, J. Milner, C. H. Arrowsmith and A. Bochkarev, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 15412–15417.

30 H. Feng, L. M. Jenkins, S. R. Durell, R. Hayashi, S. J. Mazur, S. Cherry, J. E. Tropea, M. Miller, A. Wlodawer, E. Appella and Y. Bai, *Structure*, 2009, **17**, 202–210.

31 C. W. Lee, M. A. Martinez-Yamout, H. J. Dyson and P. E. Wright, *Biochemistry*, 2010, **49**, 9964–9971.

32 P. Di Lello, L. M. Jenkins, T. N. Jones, B. D. Nguyen, T. Hara, H. Yamaguchi, J. D. Dikeakos, E. Appella, P. Legault and J. G. Omichinski, *Mol. Cell*, 2006, **22**, 731–740.

33 N. Raj and L. D. Attardi, *Cold Spring Harbor Perspect. Med.*, 2017, **7**, a026047.

34 B. Shan, D. W. Li, L. Bruschweiler-Li and R. Bruschweiler, *J. Biol. Chem.*, 2012, **287**, 30376–30384.

35 S. W. Chi, S. H. Lee, D. H. Kim, M. J. Ahn, J. S. Kim, J. Y. Woo, T. Torizawa, M. Kainosho and K. H. Han, *J. Biol. Chem.*, 2005, **280**, 38795–38802.

36 M. Okuda and Y. Nishimura, *J. Am. Chem. Soc.*, 2014, **136**, 14143–14152.

37 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999–2012.

38 A. D. Mackerell, Jr., M. Feig and C. L. Brooks, 3rd, *J. Comput. Chem.*, 2004, **25**, 1400–1415.

39 G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.

40 K. A. Beauchamp, Y. S. Lin, R. Das and V. S. Pande, *J. Chem. Theory Comput.*, 2012, **8**, 1409–1414.

41 Y. Qi, Y. Huang, H. Liang, Z. Liu and L. Lai, *Biophys. J.*, 2010, **98**, 321–329.

42 T. T. Nguyen, B. K. Mai and M. S. Li, *J. Chem. Inf. Model.*, 2011, **51**, 2266–2276.

43 A. K. Somavarapu and K. P. Kepp, *ChemPhysChem*, 2015, **16**, 3278–3289.

44 G. Chillemi, P. Davidovich, M. D'Abramo, T. Mametnabiev, A. V. Garabadzhiu, A. Desideri and G. Melino, *Cell Cycle*, 2013, **12**, 3098–3108.

45 E. Fadda and M. G. Nixon, *Phys. Chem. Chem. Phys.*, 2017, **19**, 21287–21296.

46 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950–1958.

47 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

48 D. A. Case, T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.

49 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

50 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.

51 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

52 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

53 Y. Chebaro, A. J. Ballard, D. Chakraborty and D. J. Wales, *Sci. Rep.*, 2015, **5**, 10386.

54 X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark, *Angew. Chem., Int. Ed.*, 1999, **38**, 236–240.

55 W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.

56 K. W. Plaxco, K. T. Simons and D. Baker, *J. Mol. Biol.*, 1998, **277**, 985–994.

57 Y. Shen and A. Bax, *J. Biomol. NMR*, 2010, **48**, 13–22.

58 M. Zweckstetter and A. Bax, *J. Am. Chem. Soc.*, 2000, **122**, 3791–3792.

59 G. Fuertes, N. Banterle, K. M. Ruff, A. Chowdhury, D. Mercadante, C. Koehler, M. Kachala, G. Estrada Girona, S. Milles, A. Mishra, P. R. Onck, F. Grater, S. Esteban-Martin, R. V. Pappu, D. I. Svergun and E. A. Lemke, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, E6342–E6351.

60 J. Song, G.-N. Gomes, T. Shi, C. C. Gradinaru and H. S. Chan, *Biophys. J.*, 2017, **113**, 1012–1024.

61 V. N. Uversky, J. R. Gillespie and A. L. Fink, *Proteins*, 2000, **41**, 415–427.

62 R. K. Das and R. V. Pappu, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 13392–13397.

63 D. Mercadante, S. Milles, G. Fuertes, D. I. Svergun, E. A. Lemke and F. Graeter, *J. Phys. Chem. B*, 2015, **119**, 7975–7984.

64 E. W. Martin, A. S. Holehouse, C. R. Grace, A. Hughes, R. V. Pappu and T. Mittag, *J. Am. Chem. Soc.*, 2016, **138**, 15323–15335.

65 M. Kjaergaard, A.-B. Norholm, R. Hendus-Altenburger, S. F. Pedersen, F. M. Poulsen and B. B. Kragelund, *Protein Sci.*, 2010, **19**, 1555–1564.

66 C. Cragnell, D. Durand, B. Cabane and M. Skepo, *Proteins*, 2016, **84**, 777–791.

67 J. E. Kohn, I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. S. Dothager, S. Seifert, P. Thiyagarajan, T. R. Sosnick, M. Z. Hasan, V. S. Pande, I. Ruczinski, S. Doniach and K. W. Plaxco, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 12491–12496.

68 J. Henriques and M. Skepo, *J. Chem. Theory Comput.*, 2016, **12**, 3407–3415.

69 H. Lee, K. H. Mok, R. Muhandiram, K. H. Park, J. E. Suk, D. H. Kim, J. Chang, Y. C. Sung, K. Y. Choi and K. H. Han, *J. Biol. Chem.*, 2000, **275**, 29426–29432.

70 E. A. Cino, W. Y. Choy and M. Karttunen, *J. Chem. Theory Comput.*, 2012, **8**, 2725–2740.

71 R. B. Best, N. V. Buchete and G. Hummer, *Biophys. J.*, 2008, **95**, L07–L09.

72 S. R. Gerben, J. A. Lemkul, A. M. Brown and D. R. Bevan, *J. Biomol. Struct. Dyn.*, 2014, **32**, 1817–1832.

73 D. J. Rosenman, C. Wang and A. E. Garcia, *J. Phys. Chem. B*, 2016, **120**, 259–277.

74 A. A. Adzhubei, M. J. E. Sternberg and A. A. Makarov, *J. Mol. Biol.*, 2013, **425**, 2100–2132.

75 J. Mittal, T. H. Yoo, G. Georgiou and T. M. Truskett, *J. Phys. Chem. B*, 2013, **117**, 118–124.

76 M. D. Smith, J. S. Rao, E. Segelken and L. Cruz, *J. Chem. Inf. Model.*, 2015, **55**, 2587–2595.

77 J. Huang and A. D. Mackerell Jr., *Curr. Opin. Struct. Biol.*, 2018, **48**, 40–48.

78 P. E. Lopes, J. Huang, J. Shim, Y. Luo, H. Li, B. Roux and A. D. Mackerell Jr., *J. Chem. Theory Comput.*, 2013, **9**, 5430–5449.

79 Z. Zhang, Y. Ouyang and T. Chen, *Phys. Chem. Chem. Phys.*, 2017, **18**, 31304–31311.

80 D. Mercadante, J. A. Wagner, I. V. Aramburu, E. A. Lemke and F. Graeter, *J. Chem. Theory Comput.*, 2017, **13**, 3964–3974.