Modelling the structure and interactions of intrinsically disordered peptides with multiple-replica, metadynamics-based sampling methods and force-field combinations

Lunna Li[1,‡], Tommaso Casalini[2,‡], Paolo Arosio[2] and Matteo Salvalaglio[1*]

[1]*Thomas Young Centre and Department of Chemical Engineering,*

University College London, London, UK

[2]*Department of Chemistry and Applied Biosciences, ETH Zurich,*

Zurich, Switzerland

Intrinsically disordered proteins (IDPs) play a key role in many biological processes, including the formation of biomolecular condensates within cells. A detailed characterization of their configurational ensemble and structure-function paradigm is crucial for understanding their biological activity and for exploiting them as building blocks in material sciences. In this work, we incorporate bias-exchange metadynamics and parallel-tempering well-tempered metadynamics with CHARMM36m and CHARMM22* to explore the structural and thermodynamic characteristics of a short archetypal disordered sequence derived from a DEAD-box protein. The conformational landscapes emerging from our simulations are largely congruent across methods and forcefields. Nevertheless, differences in fine details emerge from varying forcefield/sampling method combinations. For this protein, our analysis identifies features that help to explain the low propensity of this sequence to undergo self-association *in vitro*, which can be common to all force-field/sampling method combinations. Overall, our work demonstrates the importance of using multiple force-field/enhanced sampling method combinations for accurate structural and thermodynamic information in the study of general disordered proteins.

INTRODUCTION

Intrinsically disordered proteins (IDPs) are abundantly present in nature and play important roles in carrying out diverse biological functions, including cellular signaling, cytoplasmic compartment and regulation of gene expression[1-8]. A class of IDPs, defined as low complexity domains (LCDs), have been recently discovered in association with the dynamic formation of open compartments in cells[4-8]. These compartments, which are defined as condensates, are formed by liquid-liquid phase separation (LLPS) of proteins and nucleic acids and could underlie important functions and dysfunctions in biology. In addition to their role in biology, these LCDs represent promising building blocks for the design of synthesizing organelles to encode novel biochemical functionalities in bioengineering in a controllable and programmable manner, either alone or conjugated to soluble globular domains[9-15].

IDPs can be characterized by the lack of stable well-defined native structures of folded proteins, and amino acid compositions biased towards a high fraction of charged and polar residues, secondary-structure disrupters such as proline and glycine, and a low amount of bulky hydrophobic amino acids[1,15-19]. It is believed the multivalent attractive interaction between side chains may give rise to favorable energetic gain that is responsible for counteracting the entropic loss during LLPS[15,20-21], and the phase behaviours of IDPs can be specifically encoded in their protein sequences[21-31]. Nevertheless, our understanding of IDPs and their key role in mediating phase separations of multicomponent structures into coacervate assemblies is still largely limited. In the literature, mean-field theories such as Flory-Huggins theory and a recently emerging sticker-and-spacer model provide highly generalizable frameworks for describing the thermodynamics of the phase behavior of associate polymers in

solution[32-33], with emerging applications to complex biological systems both experimentally and computationally[33-36]. The thermodynamic driving force towards phase separation is likely to be inherently determined by the conformational free energy (FE) landscape of individual IDPs (intramolecular factor) and their molecular interactions with other sequences in solution including their own replicates (intermolecular factor). Thus, uncovering protein dynamics and their structure-function relationship can be pivotal for establishing versatile and sensitive protocols to design, program and predict the bottom-up assembly of multifunctional bio-inspired protein-based materials based on IDPs.

Molecular dynamics (MD) simulations can in principle offer the opportunity of resolving structural, kinetic and thermodynamic characteristics of biological systems such as IDPs and IDP-rich bodies at the length and time scales that are not easily accessible experimentally[22,37-51]. These simulations complement state-of-the-art experimental methods such as NMR spectroscopy[52-53], single-molecule Förster resonance energy transfer (sm-FRET)[54-55], and small-angle X-ray and neutron scattering (SAXS and SANS)[56-57], which may exhibit challenges in measuring molecular motions at atomic resolution and conformational heterogeneity associated with structural disorder[37,41-42,49]. Nevertheless, conformational sampling in classical molecular dynamics (MD) simulations can be restrained to local minima of free energy surfaces (FES), and accessing the full and complex landscape of disordered proteins can be nontrivial with standard unbiased simulations. To mitigate the problem, various enhanced sampling schemes have been developed to help probing regions of the configurational space that could be rarely explored otherwise. Unbiased enhanced samplings are largely based upon the concept of parallel-tempering (PT)[50-51]. In temperature PT/replica-exchange algorithm, multiple replicas are simulated at different

temperatures, in which conformations are exchanged at regular intervals, based on the Metropolis acceptance criterion. The stochastic nature of the method ensures generation of Boltzmann-weighted ensemble from which thermodynamic averages can be straightforwardly obtained. However, temperature differences between neighboring replicas must be moderate in order to yield practically large acceptance rates. For very large biomolecular systems simulated in explicit solvent, the number of replicas needed increases as O ($f^{1/2}$) with f as the system degrees of freedom, so reaching high temperature ensembles requires challenging computational costs[51,58]. In Replica Exchange with Solute Tempering (REST2)[37,40,45,59-60], such issue can be overcome by scaling the protein intramolecular potential energy, in which case the acceptance probability depends only on the differences of intramolecular solute energy and intermolecular solute-water energy. Recently, Replica exchange with hybrid tempering (REHT)[41] was developed to enhance exploring complex free energy landscape with large free energy barriers in the replica-exchange framework. At the same time, metadynamics-based techniques are powerful tools to accelerate the exploration of rare events such as protein folding and conformational transitions[38-39,22,61-70]. These methods promote the exploration of free energy minima, by iteratively building a history-dependent bias potential as a sum of Gaussians defined as a function of a chosen set of collective variables (CVs). Boltzmann-weighted configuration ensemble can then be obtained via appropriate reweighting techniques[71-73]. Among the many versions of metadynamics-based methods, well-tempered metadynamics (MetaD-WTE) allows the Gaussian height to be decreased with time so that the bias potential smoothly converges to the exact FE surface in the long-time limit[66-67], while Metadynamics with Adaptive Gaussians can adapt the Gaussian width on the fly to the local features of FES in order to improve sampling efficiency[61].

Here, we critically compare the application combinations of two metadynamics-based, multiple-replica sampling methods and two protein force-fields to develop a systematic understanding of the conformational ensemble of an intrinsically disordered protein. The methods chosen are bias-exchange metadynamics (BEMD)[64-65] and parallel-tempering well-tempered metadynamics (PTMetaD-WTE)[38,46,66-68], which merge the advantages of PT method and metadynamics-based techniques. PTMetaD-WTE can significantly reduce the number of replicas required thanks to the potential-energy bias introduced in the MetaD-WTE part; at the same time, the high temperature replicas from the PT part may compensate for the limited number of CVs directly biased to explore a high-dimensional phase space. On the other hand, in the BEMD method, a larger set of structural CVs is separately biased, and the system is often simulated at one temperature. Overall, the efficiency of the CV-based metadynamics methods depends on a suitable set of collective variables, which usually requires *a priori* knowledge of topological, chemical, and physical properties of the protein of interest. In this context, we have selected structural CVs commonly used for exploring folded and disordered proteins[38,46,64].

It is well-known that the results of MD simulations can strongly depend strongly on the accuracy of the applied protein-water force-fields, which may lead to large discrepancies with various experimental measurements[42,45,49,74-82]. Examples have been observed with state-of-the-art Amber force-fields a99SB*-ILDN[83-84] with TIP3P[85], a99SB-ILDN with TIP4P-D[80], and the a03ws with empirically optimized solute-solvent dispersion interactions[81], CHARMM force-fields CHARMM22*[86] and CHARMM36m[87] with CHARMM-modified TIP3P[85,87-88]. For example, these force-fields have been demonstrated to calculate very different helical propensities from the experimental estimates of NMR data for various proteins of interests[74]. While the

recently developed a99SB-disp is optimally parameterized for both folded and disordered proteins with substantially improved accuracy[74], the requirement of four-point water model significantly increases computational costs. Here, we propose and demonstrate the use of two protein force-fields compatible with three-point CHARMM-modified TIP3P water, namely, CHARMM36m and CHARMM22*, in combination with our chosen sampling methods. In particular, CHARMM36m showed improved accuracy in generating polypeptide backbone conformational ensembles for intrinsically disordered proteins[77,87], despite the issue of over-compact structures and over-stabilized helices[74,89]. The "helix-coil-balanced" CHARMM22*, instead, is not only able to reproduce experimental native-state structure and proteins folding rate, but also shows good agreement with experimental secondary structure propensities and NMR chemical shifts for many disordered proteins[74,87,90].

In summary, we use BEMD and PTMetaD-WTE with CHARMM22* and CHARMM36m protein force-fields to effectively sample IDP conformational landscapes within a microsecond timescale. We apply this approach to elucidate the structural and thermodynamic properties of an archetypal IDP sequence derived from N-terminus of DEAD-box protein DHH1[91-92]. DHH1N is a 46 amino-acid sequence that contains a low fraction of hydrophobic residues. The peptide is enriched in polar residue such as Asparagine (Asn) and Threonine (Thr), negatively charged Aspartic acid (Asp), and positively charged large-sized Arginine (Arg) and Lysine (Lys), and with a few Prolines (Pro) distributed along the C-terminal half of the peptide. Experimental observations indicate that *in vitro* DHH1N does not undergo LLPS on its own at physiological pH, despite its composition is characteristic of low complexity domains involved in LLPS, and the fact that it participates in the formation of molecular adhesives to promote LLPS of chimera proteins[15]. In this work, we investigate a

relatively short IDP that does not undergo phase separation, demonstrating how information from multiple sampling algorithms and forcefields reveal a detailed and congruent multi-dimensional FES within microsecond timescales. This method represents a basis to next investigate the behavior of IDPs undergoing phase separation and connect structural properties with phase transition.

RESULTS AND DISCUSSION

PTMetaD-WTE. To obtain a general understanding of the configurational ensemble of DHH1N using PTMetaD-WTE, we first calculated one-dimensional free energy (1D-FE) profiles as a function of individual collective variables at 300 K for both CHARMM36m and CHARMM22* (Fig. 1); in addition, results from the unbiased CHARMM22* simulations with nine different initial structures are also included for comparison. The selection of PLUMED-defined collective variables are the number of $C_\alpha - C_\alpha$ contacts and hydrophobic $C_\gamma - C_\gamma$ contacts, the number of backbone H-bonds, α-RMSD (proportional to α-helical content), antiparallel-β-RMSD and parallel-β-RMSD (proportional to β-sheet content), radius of gyration R_g, asphericity b and the relative shape anisotropy κ^2 which measures the conformational deviation from a perfectly spherical structure [Supplementary Note 1][93-94]. Generally, these collective variables enable us to understand protein structures in terms of size, shape, compactness and structural order.
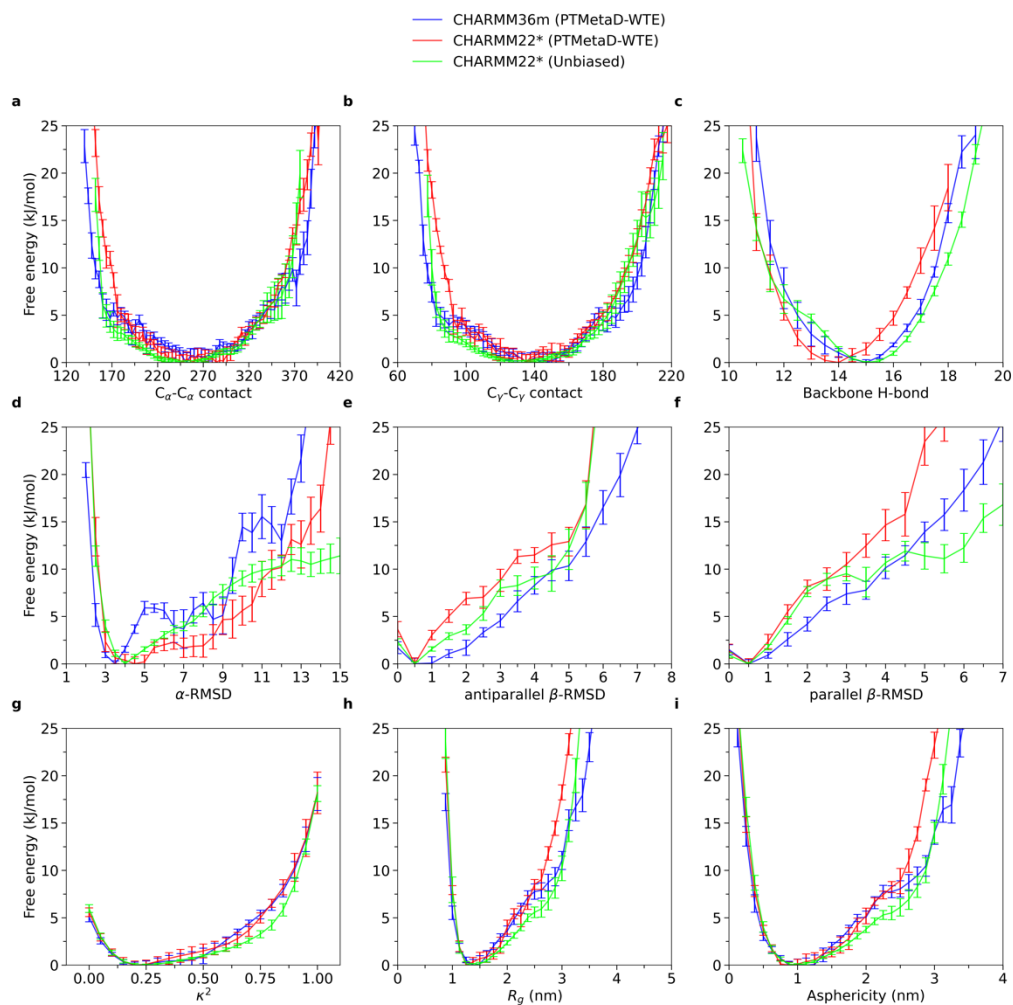
Figure 1 One-dimensional FE profiles of DHH1N for CHARMM36m PTMetaD-WTE, CHARMM22* PTMetaD-WTE and CHARMM22* unbiased simulations. Error bars are calculated from Tiwary reweighting scheme[72,97] and block averages[95-97].

Notably, the configurational ensemble projected in most CVs displays a single FE minimum upon convergence, corresponding to a monomodal probability density distribution (Supplementary Note 2, Figs. S1, S2, S3). An exception is represented by the α-RMSD distribution obtained with CHARMM36m, which shows a FE profile characterized by multiple local minima. A similar picture emerges, to a lesser extent, from the α-RMSD probability density obtained from CHARMM22* PTMetaD-WTE, which also exhibits shallow local minima, despite a smoother FE profile compared to CHARMM36m. To investigate the possibilities of multiple, metastable conformational states emerging from PTMetaD-WTE simulations, we investigated conformational degeneracies associated with individual CVs by constructing two-dimensional free energy surfaces (2D-FES) for pairs of CVs that are less correlated with each other (Supplementary Note 2, Fig. S4). Importantly, both force-fields map rugged weakly-funneled landscapes, with a single, large global basin at compact R_g and low α-RMSD value (Fig. 2). The local minima in the 1D-FE profile CHARMM36m (Fig. 1d) are shown more clearly in the 2D-FES (Figs. 2a), highlighting the existence of metastable basins associated to different conformational states separated by apparent free energy barriers higher than $k_B T$. Overall, CHARMM22* displays a conformational ensemble similar to that of CHARMM36m.
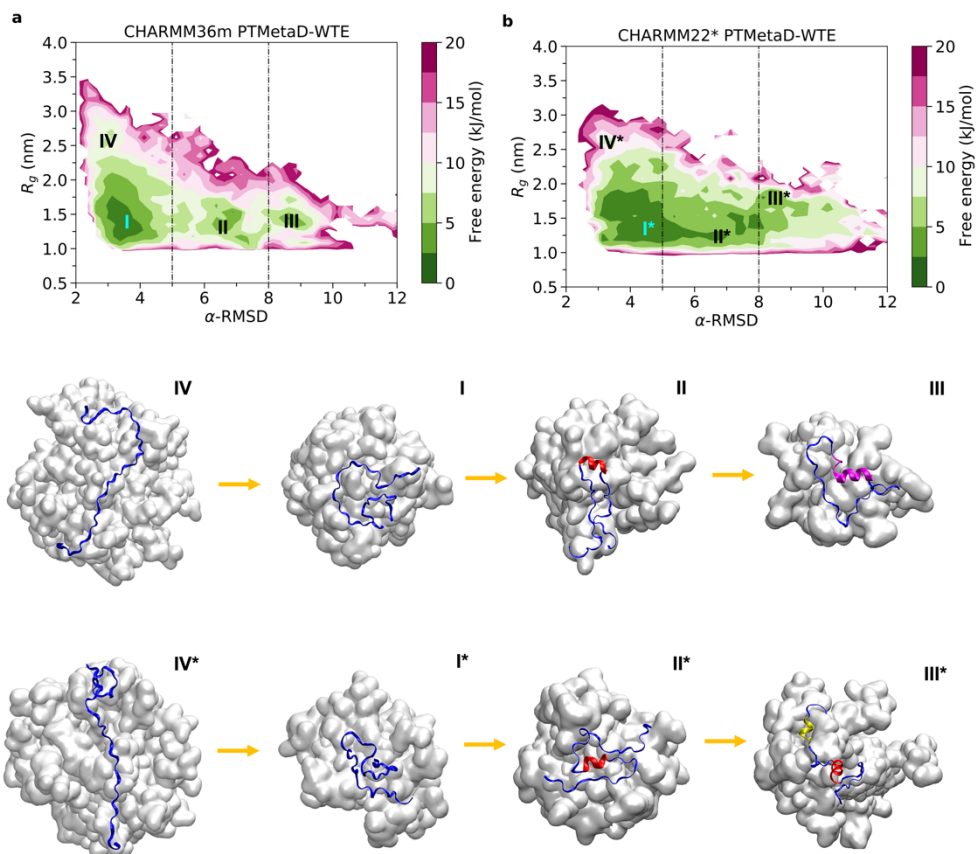
Figure 2 Two-dimensional free energy surfaces of DHH1N with PTMetaD-WTE on α-RMSD and R_g for CHARMM36m (a) and CHARMM22* (b) with the black dashed lines indicating cluster 1 (α-RMSD $\leq$ 5), cluster 2 ($5 < \alpha$-RMSD $\leq$ 8) and cluster 3 (α-RMSD $>$ 8). Conformations above 20 kJ/mol are not shown in the figure. Purple, red and yellow parts of the protein highlight the dynamic helical regions of Asn5-to-Asp18, Asp18-Asp26 and Thr36-Thr42. The grey clouds represent the overlap of trajectories in the same annotated state.

To probe DHH1N structures corresponding to different basins, we clustered the configurations sampled through PTMetaD-WTE into three groups. Cluster 1 contains all conformations with low α-RMSD values (≤ 5), cluster 2 for all structures showing intermediate α-RMSD values (> 5 and ≤ 8), and eventually cluster 3 for all frames showing high α-RMSD values (> 8). Because the collective variable α-RMSD correlates with the α-helix content by definition[93], conformations in cluster 3 may feature the highest amount of α-helices, while structures in cluster 1 are associated with the largest disorder. The global minima of CHARMM36m and CHARMM22* are both enclosed within cluster 1, which comprises approximately 70 % and 50 % of the equilibrium conformational population, respectively. Such evidence strongly indicates that DHH1N is intrinsically disordered regardless of the force-field choice, consistent with experimental observation[15,91-92]. While the two force-fields explore similar conformational phase space, no significant free energy barriers are displayed between clusters of CHARMM22*, suggesting that the force-field may describe DHH1N with a higher conformational flexibility, and with lower free energy costs for converting between compact and extended conformations. In contrast, the disordered global minimum of CHARMM36m is separated from the partially folded local minima by apparent FE barriers of approximately 5-10 kJ/mol, implying that the nucleation of α-helical domains is activated process for DHH1N simulated with CHARMM36m.

To study more accurately the structural motifs emerging from the extensive sampling of DHH1N conformational ensemble, we also evaluate key secondary-structure content for CHARMM36m and CHARMM22* PTMetaD-WTE by means of the DSSP algorithm[98]. While DHH1N is predominantly disordered, as demonstrated by the high number of loops/irregular elements, bends and turns for both CHARMM36m and CHARMM22*, three domains of CHARMM36m display a relatively high

propensity to nucleate α-helices, located approximately along the sequences of Asn6-to-Asp16, Asp18-to-Asn26 and Thr36-to-Thr42 (Fig. 3c). Interestingly, the propensity of α-helices along the N-terminal Asn6-to-Asp16 is comparatively low compared with those of the other two regions, which becomes negligible in the case of CHARMM22* PTMetaD-WTE (Fig. 3f) and unbiased simulation (Fig. 3i), suggesting that DHH1N simulated under CHARMM22* demonstrates negligible probability to form α-helices in the Asn-rich N-terminal domain. Moreover, the sequence between Pro28 and Pro34, which separates the α-helical domains of Asp18-to-Asn26 from Thr36-to-Thr42, shows no α-helix propensity for both CHARMM36m and CHARMM22* (Figs. 3c, f, i). Such observation could be associated with the incompatibility of Pro with α-helix formation due to its rigid-ring structure and absence of an H atom on the peptide-bond N necessary for further H-bonding[99-100]. This feature is well captured by both force-fields. To study individual residue-residue interactions in detail, we constructed $C_\alpha - C_\alpha$ contact-maps for residue-pairs beyond the next nearest-neighbors, using a threshold distance of 0.5 nm (Figs. 3a, d, g). By focusing on residue-residue contacts with ~ 5 % frequencies, our analysis reveals a sparse contact-map with only a few key contacts along the map diagonal for all of CHARMM36m/CHARMM22* PTMetaD-WTE/unbiased simulations. The adjacent contacts along the diagonal in the domains of Asp18-to-Asn26 and Thr36-to-Thr42 coincide approximately with the two helical regions for both CHARMM36m and CHARMM22* from DSSP analysis (Figs. 3c, f, i).
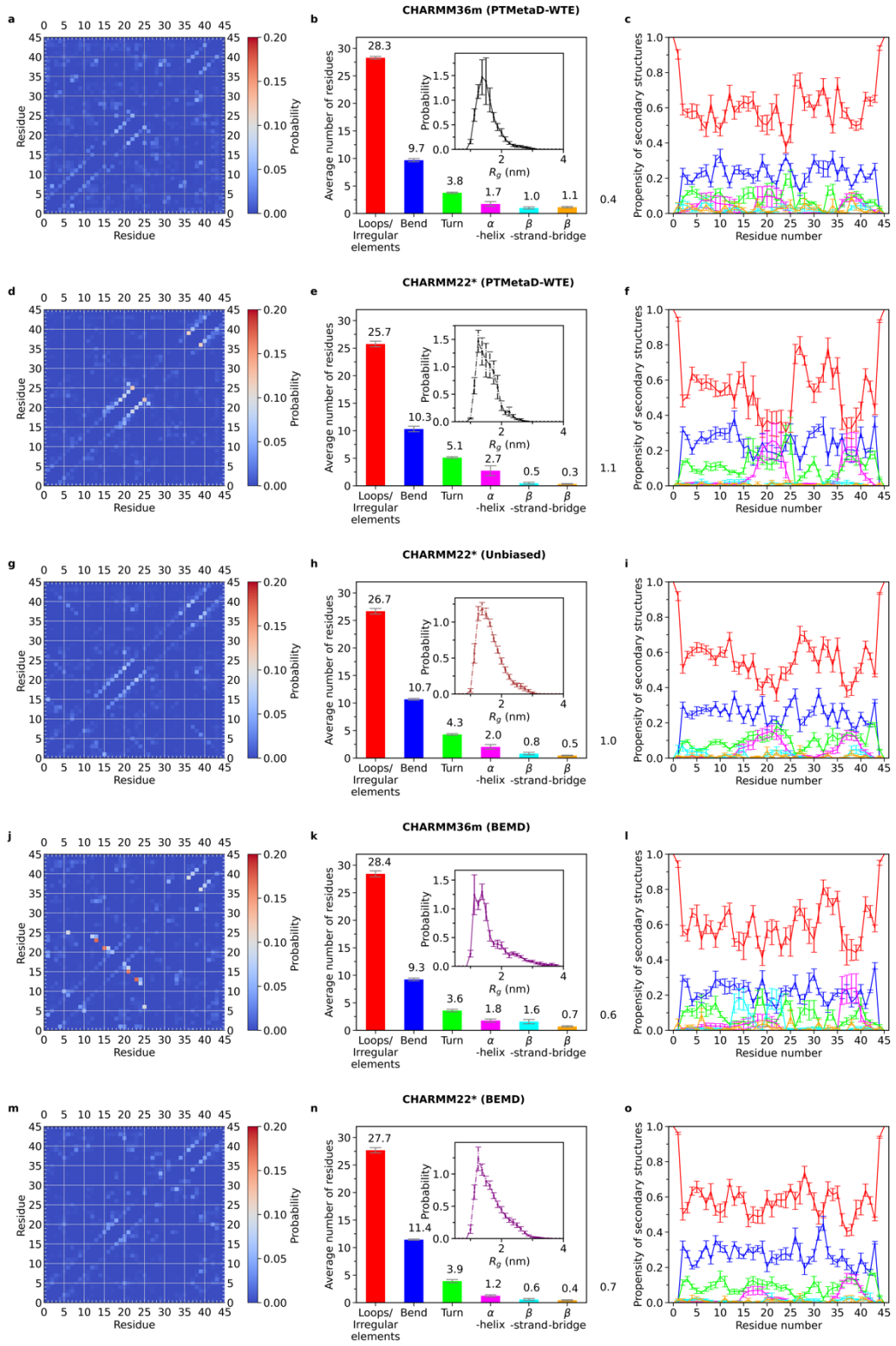
b CHARMM36m (PTMetaD-WTE)

e CHARMM22* (PTMetaD-WTE)

h CHARMM22* (Unbiased)

k CHARMM36m (BEMD)

n CHARMM22* (BEMD)

14

Figure 3 Contact-maps and secondary-structure analysis of DHH1N for CHARMM36m PTMetaD-WTE (a), (b) and (c), CHARMM22* PTMetaD-WTE (d), (e) and (f), and CHARMM22* unbiased (g), (h) and (i), CHARMM36m BEMD (i), (k) and (j) and finally CHARMM22* BEMD (m), (n) and (o). The insets of (b), (e), (h), (k) and (n) correspond to R_g distributions. The secondary structure assignment is based on the DSSP analysis codes[98]. Error bars are calculated from Tiwary reweighting scheme[72,97] and block averages[95-97].

Bias-exchange metadynamics. To complement the insights obtained with PTMetaD-WTE, we adopted BEMD method to further explore the conformational ensemble of DHH1N. We constructed 1D-FE profiles for CHARMM36m and CHARMM22* at 300 K and compare them with the results from PTMetaD-WTE. Similarly, the configurational ensemble projected in most CVs also displays a single FE minimum (Fig. 4), corresponding to a monomodal probability density distribution upon convergence (Supplementary Note 3, Figs. S5, S6). All our simulations show that DHH1N contain a very low amount of β-sheet (parallel/anti-parallel β-RMSD) and α-helical (α-RMSD) motifs, featuring the intrinsically disordered nature of the protein (Figs. 4d, e and f). However, BEMD tends to explore wider ranges of $C_\alpha - C_\alpha$ and $C_\gamma - C_\gamma$ contacts than PTMetaD-WTE (Figs. 4a, b) [Supplementary Note 4, Figs. S7a, b], with the effect being more prominent for CHARMM36m. For CHARMM36m, BEMD agrees well with PTMetaD-WTE on producing α-RMSD FE profiles characterized by multiple local minima (Figs. 4d) [Supplementary Note 4, Fig. S7d], despite that some of the local minima are located at different α-RMSD values. In addition, the incapability of DHH1N to undergo LLPS may also be reflected in its relative-shape-anisotropy κ^2. The broad distribution of κ^2 in all four simulations (Supplementary Note 4, Fig. S7g) suggests that multiple DHH1N molecules may dynamically adopt different molecular shapes because their conformations are not separated by large free energy barriers.
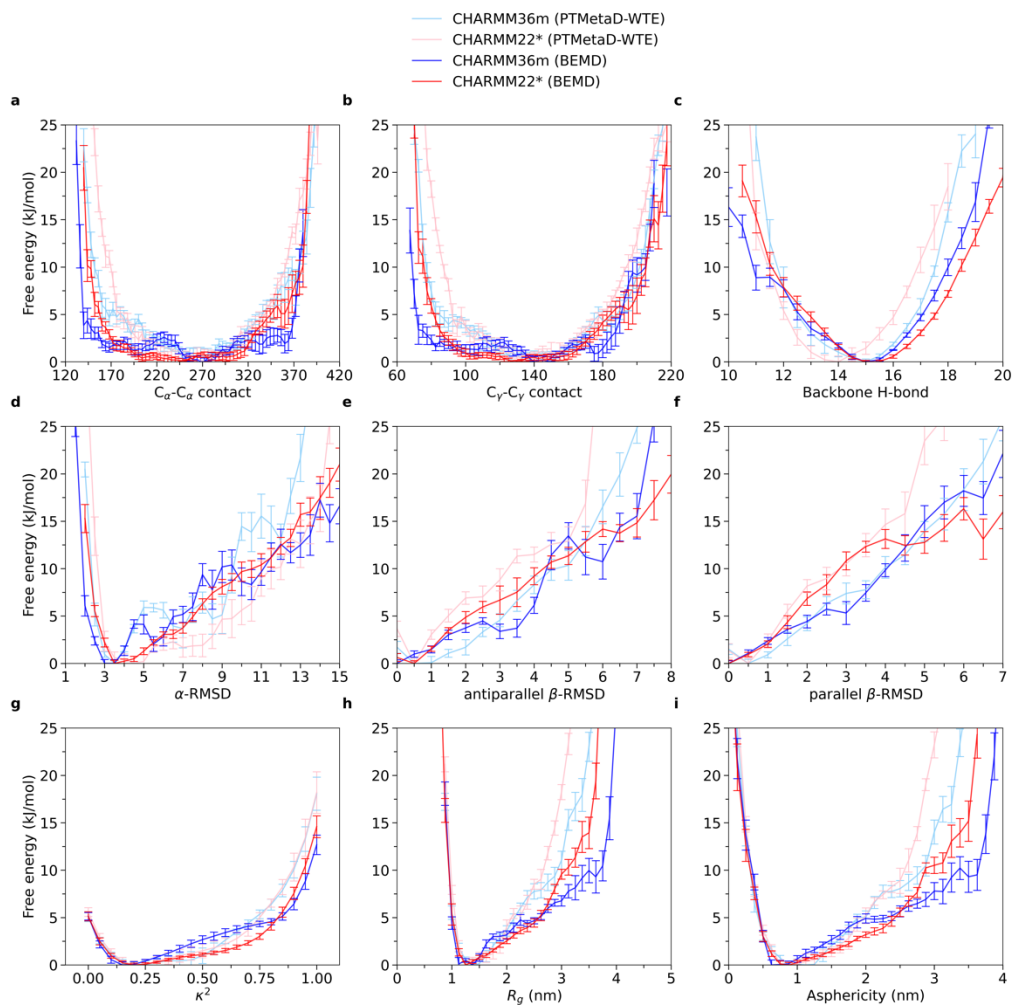
16

Figure 4 Comparison between the one-dimensional FE profiles of DHH1N from PTMetaD-WTE and BEMD simulations. Error bars are calculated from Tiwary reweighting scheme[72,97] and block averages[95-97].

To further examine the conformational states of DHH1N, we construct 2D-FES for CHARMM36m and CHARMM22* BEMD and compare them with the 2D-FES from PTMetaD-WTE. Similarly, every simulation reveals a global minimum at low α-RMSD in the disordered cluster 1, accompanied by multiple shallow local minima, indicating a large number of substates as a result of conformational heterogeneity (Fig. 5). Cluster 1 of CHARMM36m and CHARMM22* BEMD comprises approximately 70 % of the equilibrium conformational population, similar to the result of CHARMM36m PTMetaD-WTE, while cluster 1 population of CHARMM22* PTMetaD-WTE is much lower due to relatively shorter simulation time (Supplementary Note 4, Table S3). The largest disparities between the 2D-FES are primarily due to different choices of force-fields. In detail, the CHARMM36m results show that the nucleation of local α-helical domains is an activated process for both PTMetaD-WTE and BEMD, as suggested by the presence of relatively large free energy barriers (Figs. 5a, c, e, g, i and k), despite that the distribution of local minima differ slightly between the two sampling methods. Conversely, the assembly/disassembly of the partially ordered domains in CHARMM22* is associated with relatively small free energy costs, as indicated by a reduced number of scattered, local minima and FE barriers for both sampling methods (Figs. 5b, d, f, h, j and l). It is challenging to further compare the details of the scattered local minima because the weakly-funneled and rugged nature of FE landscape means that statistical errors could be of similar magnitude.
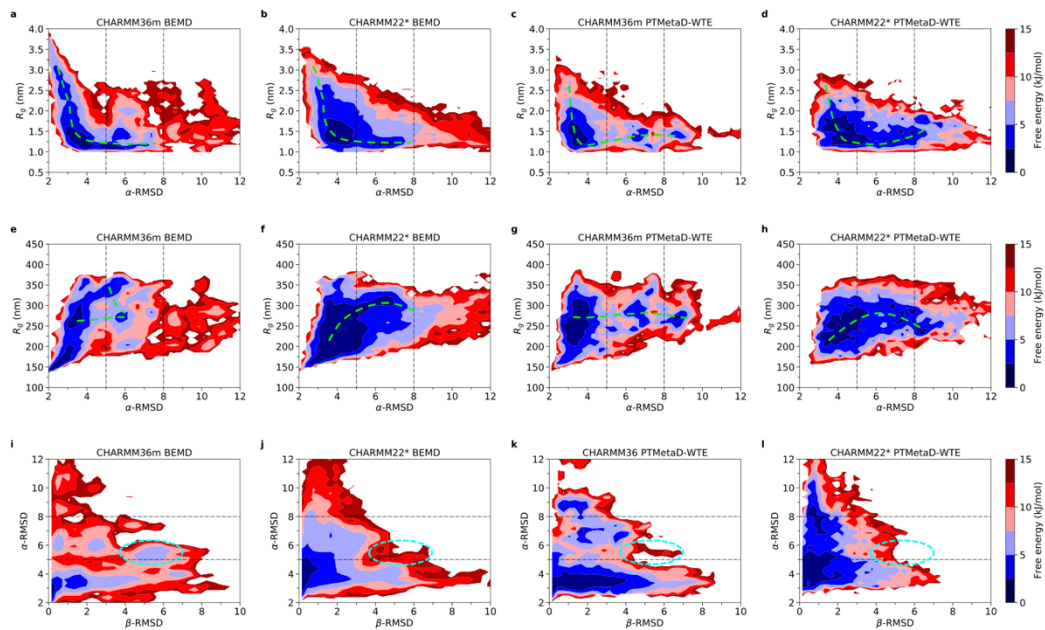
Figure 5 Two-dimensional free energy surfaces on α-RMSD and R_g for CHARMM36m BEMD (a), CHARMM22* BEMD (b), CHARMM36m PTMetaD-WTE (c) and CHARMM22* PTMetaD-WTE (d). Two-dimensional free energy surfaces on α-RMSD and the number of $C_\alpha - C_\alpha$ contact for CHARMM36m BEMD (e), CHARMM22* BEMD (f), CHARMM36m PTMetaD-WTE (g) and CHARMM22* PTMetaD-WTE (h). Two-dimensional free energy surfaces on β-RMSD and α-RMSD for CHARMM36m BEMD (i), CHARMM22* BEMD (j), CHARMM36m PTMetaD-WTE (k) and CHARMM22* PTMetaD-WTE (l). The black dashed lines indicate cluster 1 (α-RMSD $\leq$ 5), 2 ($5 < \alpha$-RMSD $\leq$ 5) and 3 (α-RMSD $>$ 8) as defined in Fig. 2. The cyan dashed lines indicate the regions primarily occupied by the subgroup of CHARMM36m BEMD with concurring Ala24 and Ser16-Lys22 contacts, which is less explored in all the other simulations.

Based upon the 2D free energy analysis, we proceeded to contact-maps and secondary-structures analysis for CHARMM36m and CHARMM22* of BEMD, in comparison with results from CHARMM36m and CHARMM22* of PTMetaD-WTE (Fig. 3). By focusing on residue-residue contacts with ~ 5 % frequencies, our analysis again reveals a sparse contact-map in which key contacts are mostly along the diagonal regions for all simulations. In respect of the off-diagonal regions, the four simulations show a lack of common residue-contacts. One exception is the Asn14-Ala24/Ser16-Lys22 domain of CHARMM36m BEMD. The contact-map of CHARMM36m BEMD displays prominent Asn14-Ala24 and Ser16-Lys22 contacts in this region, which is noticeably different from the other simulations within microsecond simulation length (Fig. 3i). At the same time, CHARMM36m BEMD explores a relatively lower number of α-helical structures, and a comparatively higher amount of β-strands in the domain of Asp18-to-Asn26, compared to the other method/force-field combinations (Fig. 3l). To investigate possible relationship between the two features, we extracted structures with concurring Asn14-Ala24 and Ser16-Lys22 contacts from the CHARMM36m BEMD configurational ensemble and found that these residue-contacts are correlated with the presence of β-strands within Asn14-to-Asn26 (Supplementary Note 5, Figs. S8, S9). The subgroup occupies ~ 18 % of the CHARMM36m BEMD population and spans both cluster 1 and 2 (Supplementary Note 5, Fig. S9b) [Fig. 5]. The relevant conformations, highlighted by the cyan circles in Fig. 5i, are less explored in the other force-field/method combinations (Figs. 5j, k and l).

We also constructed $C_\alpha - C_\alpha$ contact-maps and secondary-structure analysis for clusters 1, 2 and 3 of all simulations (Fig. 6). For CHARMM36m BEMD, the concurring Asn14-Ala24 and Ser16-Lys22 contacts, which have been proven to be highly associated with the β-strands observed along Asn14-to-Asn26, emerge primarily

in the less-ordered cluster 1 and 2 (Figs. 6j, k). In all cases, adjacent diagonal residue-contacts are focused within the two highest-order clusters (2 and 3), reflecting the presence of locally-ordered domains for the generally-disordered DHH1N. In contrast, the four force-field/method combinations share very few off-diagonal contacts, especially in the largest and most-disordered cluster 1, which contains about 70 % of overall population in most cases. A potential region of common contacts can be approximately between Pro28-Thr32 and Pro34-Leu40 (Fig. 6). The first sequence is capped by Pro28 and Pro34 and contains two adjacent units of positively-charged and bulky Lys29-Lys30), while the second sequence contains two adjacent units of negatively-charged Asp37-Asp38. Thus, electrostatic interactions may contribute to forming these residue-contacts, and it is also possible that participating in α-helical motifs reduces the likelihood of Pro34-Leu40 to interact with Pro28-Thr32 in cluster 2 and 3.

Overall, CHARMM36m BEMD shows a tendency to sample more β-sheet-displaying conformations than the other force-field/method combinations. As shown earlier, a large fraction of the relevant structures corresponds to forming antiparallel β-sheets along Asp18-to-Asn26 and α-helices along Thr36-to-Thr42, with concurring Asn14-Ala24 and Ser16-Lys22 residue contacts (Supplementary Note 3, Fig. S7; Supplementary Note 4, Figs. S8, S9). Such trend is not observed for CHARMM22* simulated under the BEMD method and CHARMM36m simulated under PTMetaD-WTE. One hypothesis could be that the feature could be specifically related with using the backbone-optimized force-field in combination with the sampling method. In the literature, it is also reported that BEMD explores a larger portions of phase space than simple temperature PT simulations[42]. Our evidence also shows that BEMD is able to sample a slightly larger conformational ensemble than PTMetaD-WTE (Figs. S7a and

b). Thus, it is possible that the observed discrepancy can be due to the fact that both α-content (α-RMSD) and β-content (β-RMSD) were used as collective variables in BEMD, while the bias potential of PTMetaD-WTE (a function of $C_\alpha - C_\alpha$ and $C_\gamma - C_\gamma$ contacts) can only implicitly explore along the two secondary-structure reaction coordinates. Nevertheless, the difference may be relatively trivial in presence of the dominating disordered nature of DHH1N – over 70 % of the total population is contained in the low β-content (β-RMSD $\leq$ 4) region for both CHARMM36m BEMD and PTWTE. Furthermore, all simulations provide a similar insight into a possible nucleation pathway for the partially disordered regions of DHH1N (schematically represented by the green dashed lines in Fig. 5). While the 2D-FES cannot reveal an accurate multi-dimensional pathway for the full conformational transition kinetics, our evidence suggests that the dynamic disorder $\rightarrow$ α transition in the local domains of Asp18-to-Asn26 and Thr36-to-Thr42 is likely to occur within a relatively compact DHH1N ($R_g \sim 1.5$ nm) with moderate sphericity (Supplementary Note 4, Figs. S7g, h and i).
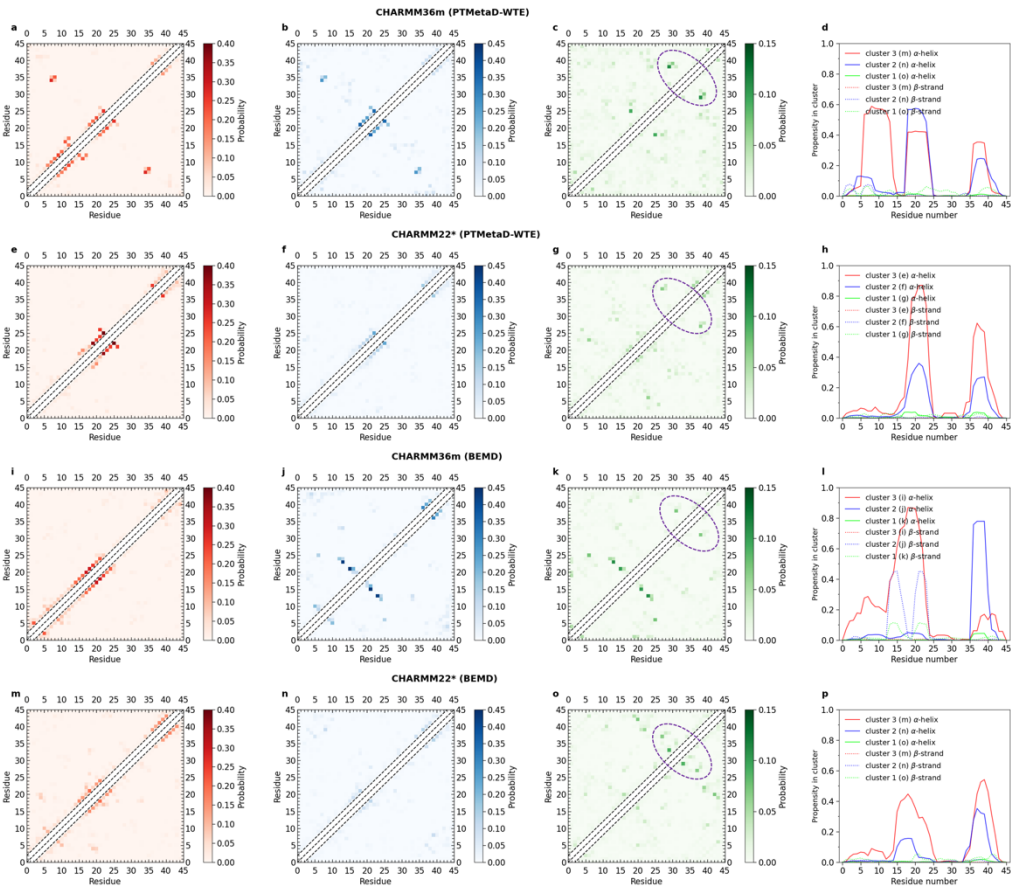
Figure 6 Contact-map and α-helical analysis of DHH1N of clusters 1, 2 and 3 for CHARMM36m PTMetaD-WTE (a)-(d), CHARMM22* PTMetaD-WTE (e)-(h), CHARMM36m BEMD (i)-(l) and CHARMM22* BEMD (m)-(p). The probabilities and propensities are computed relative to the overall population of every cluster. The purple dashed lines highlight possible regions of residues-contacts (approximately between Pro28-Thr32 and Pro34-Leu40) that are shared by all simulations.

23

CONCLUSIONS

In this work, we have examined the use of different protein force field/sampling method combinations to obtain a consensus picture for the structural and thermodynamic features of the disordered sequence DHH1N at ambient temperature and physiological pH. Despite different fine details, we show that the conformational landscapes emerging from different forcefield/sampling method combinations is largely congruent, The power of modern computers has increased dramatically, but many explicit-solvent protein simulations are still limited to microsecond timescale. As a result of the designs and advantages of various protein force-fields and sampling methods, different portions of high-dimensional phase space may be preferably explored within simulation time, despite the high possibility that all results should converge in the long-time limit. Relying on the results from a single force-field/sampling method simulation may restrain our understanding of the protein. For example, the temperature-dependence of certain protein properties can be easily recovered by analysing the sampling obtained from PTMetaD-WTE, while BEMD simulations explore more collective variables and possibly higher portions of phase space within limited time. Hence, in order to develop a comprehensive picture, it may be beneficial to include multiple FF/sampling methods combinations in the study of IDPs. Here, we observe discrepancies in some properties from separate simulations, but all results show good agreement in multi-dimensional FES, general features of residue-residue contact-map and secondary-structure analysis, which are consistent with the disordered nature of the protein.

Our evidence shows that DHH1N has an average R_g between 1.54 and 1.64 nm (Table I), which is comparable with that of the other phase-separating IDPs that are

much longer in length[22]. Such feature indicates a relatively low level of compactness. In the literature, it is reported that the phase-separating behavior of disordered proteins can be generally associated with their single-chain compactness, and the sequence-determinant of the compaction of disordered proteins is not only related to the overall protein charge, but also affected by the organization of charged and aromatic residues along the length[23,31,101-105]. At physiological pH, DHH1N is electrostatically neutral, but its charged residues, i.e., negatively-charged Asp and positively-charged Lys, Arg, are mostly focused within the C-terminal portion of the sequence from Asp16 onwards. However, DHH1N does not contain high-probability off-diagonal residue-contacts with that are shared by all force-field/method combinations, implying an absence of salt bridges or favorable electrostatic attractions between oppositely-charged patches of DHH1N, which could be key factors in inducing complex coacervation of many phase-separating IDPs[20,22]. Scarce presence of aromatic amino acids in DHH1N also means that the protein is not like to form condensates stabilized by cation-π, and π-π interactions.

The comprehensive approach implemented in this work, which we have initially applied to analyse a relatively short IDP that does not undergo phase separation *in vitro*, represents a prospecting platform to investigate the conformational ensemble and thermodynamic driving force of other IDPs that exhibit liquid-liquid phase separation under a broad range of solution conditions. Such analysis will provide crucial information at the atomistic level which will be key to unravel biological phase separation in general and assist the design of new building blocks for advanced protein-based materials and microreactors.

Table I Average radius of gyration for DHH1N. Error are calculated from block analysis.

	CHARMM36m PTMetaD-WTE	CHARMM22* PTMetaD-WTE	CHARMM22* BEMD	CHARMM36m BEMD
$\langle R_g \rangle$ (nm)	1.55(0.03)	1.54(0.02)	1.66(0.04)	1.65(0.04)

METHODS

System preparation and equilibration. Nine initial guess configurations for folded protein structure were obtained from i-TASSER[106] and Robetta webservers[107] (Supplementary Note 6). Force-field parameters were assigned by means of *pdb2gmx* command implemented in GROMACS 2019.3 software[108]; protonation states were assigned assuming physiological pH 7. The overall charge of the protein was equal to 0 and no counterions or salts were added to assure electroneutrality. CHARMM22*[86] and CHARMM36m[87] force-fields were chosen for protein, while CHARMM-modified TIP3P model[85,87-88] was used for water. The initial protein structure was placed in the center of the simulation box and subsequently solvated through *editconf* and *solvate* commands implemented in GROMACS 2019.3 software. All simulations were conducted using GROMACS 2019.3 patched with PLUMED 2.5.2[109]. Broadly, system equilibration was carried out according to the following protocol. First, system minimization was performed using steepest descent algorithm, using a tolerance value of 1000 kJ mol^{-1} nm^{-1} for the force. Temperature was raised to and kept at the target value for 50 ps in *NVT* ensemble (*N*: number of particles; *V*: volume; *T*: temperature), by means of V-rescale algorithm[110]. Solvent density was subsequently equilibrated in *NpT* ensemble (*N*: number of particles; *p*: pressure; *T*: temperature), adopting Parrinello-Rahman barostat [111] to keep the pressure to the target value of 1 atm. Neighbor list was updated every 10 simulation steps using a Verlet cutoff scheme[112]. Electrostatic long-range interactions were computed by Particle Mesh Ewald[113] using a cutoff value equal to 1.0 and 1.2 nm for CHARMM22* and CHARMM36m force-fields, respectively; the same cutoff values were employed for Van der Waals (VdW) interactions. In more detail, for CHARMM36m we adopted a force-switch scheme for VdW interactions as suggested by GROMACS, setting *rvdw-switch* equal to 1.0 nm[114].

LINCS algorithm[115] was employed to constrain all covalent bonds involving hydrogen atoms, which allowed a time step of 2 fs to propagate system dynamics via Leap-Frog algorithm. All simulations were performed adopting periodic boundary conditions (PBC). Further details of simulations are included in Supplementary Note 7 and 8.

Parallel-tempering well-tempered metadynamics. One of the initial guess configurations was first equilibrated in NpT ensemble for approximately 5 ns using V-rescale algorithm[110] and Parrinello-Rahman barostat[111] to keep the temperature and pressure at ambient conditions. After this, the equilibrated structure was used as the starting configuration for PTMetaD-WTE[38,46,66-68] simulations for both CHARMM36m[87] and CHARMM22*[86]. In the first stage, PTMetaD-WTE was implemented biasing only the potential energy (PE) collective variable (CV) of each temperature replica, simulating at T = 300, 308, 317, 326, 335, 345, 354, 364, 374, 385, 396, 407, 418, 430, 442, 455, 467, 481, 494, 508, 522, 537, 552 and 568 K for 20 ns per replica. The accumulated bias potential was subsequently used as a static PE bias potential in the second stage, where alpha carbon $C_\alpha - C_\alpha$ and gamma carbon hydrophobic $C_\gamma - C_\gamma$ contacts were both biased according to the well-tempered metadynamics algorithm. The definition and input parameters of the collective variables are included in Supplementary Note 1. The bias-factor for PE CV and the two conformational CVs were both 12, with initial Gaussian widths of 2000 kJ mol^{-1} and 1.0 at Gaussian heights of 1.2 kJ mol^{-1} respectively. A total of 1000 ns of data per temperature replica were used to reconstruct the free energy surface of DHH1N at T = 300 K. A similar protocol was applied to CHARMM22* PTMetaD-WTE to obtain approximately 600 ns of data. The PLUMED input files required to reproduce the results of PTMetaD-WTE are available on PLUMED-NEST (www.plumed-nest.org), the public repository of the PLUMED consortium[116].

Bias-exchange metadynamics. BEMD[64-65] simulations were carried out adopting seven collective variables (the number of $C_\alpha - C_\alpha$ contacts, $C_\gamma - C_\gamma$ hydrophobic contacts and backbone H-bonds, dihedral correlation, α-RMSD, antiparallel β-RMSD and parallel β-RMSD, as defined in Supplementary Note 1) and eight replicas, with one collective variable per replica plus the unbiased replica. One of the initial guess configurations was first equilibrated in *NpT* ensemble at 1 atm and 300 K for about 5 ns per replica in *NpT* ensemble according to the discussed protocol (*vide supra*); BEMD simulations were performed in *NVT* ensemble at 300 K, collecting 1000 ns of data for each replica; exchange of conformations between two randomly selected replicas were periodically attempted every 10000 simulation steps. CV were biased according to ordinary metadynamics scheme, adding the bias potential every 2500 simulation steps using a height value equal to 0.3 kJ mol^{-1}. After 160 ns, the system explored a wide region for each collective variable and we introduced loose lower and upper boundaries to improve convergence (Supplementary Note 7). The PLUMED input files required to reproduce the results of BEMD are available on PLUMED-NEST (www.plumed-nest.org), the public repository of the PLUMED consortium[116].

Unbiased simulations. Unbiased simulations were performed adopting only CHARMM22*[86] force-field. Nine of the input guess structures were solvated and equilibrated at 1 atm and 300 K according to the discussed protocol (*vide supra*); 1000 ns MD simulations were subsequently performed in *NVT* ensemble at 300 K for each system. The nine set of data were concatenated for data analysis.

Analysis of simulation data: All data analysis and error estimation are included in Supplementary Note 9.

AUTHOR INFORMATION

*Correspondence and requests for materials should be addressed to (e-mail: m.salvalaglio@ucl.ac.uk).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

‡ L. Li and T. Casalini contributed equally to the manuscript as co-first authors.

COMPETING INTEREST

The authors declare no competing interests.

FUNDING SOURCES

REFERENCES:

1. Wright, P. E. and Dyson, H. J. "Intrinsically disordered proteins in cellular signalling and regulation." *Nature reviews Molecular cell biology* **2015**, *16(1)*, 18-29.

2. Shin, Y. and Brangwynne, C. P. "Liquid phase condensation in cell physiology and disease." *Science* **2017**, *357*, 6357.

3. Feric, M., Vaidya, N., Harmon, T. S., Mitrea, D. M., Zhu, L., Richardson, T. M., Kriwacki, R. W. and Pappu, R.V., Brangwynne, C. P. "Coexisting liquid phases underlie nucleolar subcompartments." *Cell* **2016,** *165(7)*, 1686-1697.

4. Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A. P., Kim, H. J., Mittag, T. and Taylor, J. P. "Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization." *Cell* **2015**, *163(1)*, 123-133.

5. Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J. and Grishin, N. V., "Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels." *Cell* **2012**, *149(4)*, 753-767.

6. Lu, J., Cao, Q., Hughes, M. P., Sawaya, M. R., Boyer, D. R., Cascio, D. and Eisenberg, D. S. "CryoEM structure of the low-complexity domain of hnRNPA2 and its conversion to pathogenic amyloid." *Nature communications* **2020**, *11(1)*, 1-11.

7. Murthy, A. C., Dignon, G. L., Kan, Y., Zerze, G. H., Parekh, S. H., Mittal, J. and Fawzi, N. L. "Molecular interactions underlying liquid− liquid phase separation of the FUS low-complexity domain." *Nature structural & molecular biology* **2019**, *26 (7)*, 637-648.

8. Sysoev, V. O., Kato, M., Sutherland, L., Hu, R., McKnight, S. L. and Murray, D. T. "Dynamic structural order of a low-complexity domain facilitates assembly of intermediate filaments*." Proceedings of the National Academy of Sciences* **2020**, *117(38)*, 23510-23518.

9. Schuster, B. S., Reed, E. H., Parthasarathy, R., Jahnke, C. N., Caldwell, R. M., Bermudez, J. G., Ramage, H., Good, M. C. and Hammer, D. A. "Controllable protein phase separation and modular recruitment to form responsive membraneless organelles." *Nature communications* **2018**, *9(1)*, 1-12.

10. Knowles, T. P. and Mezzenga, R. "Amyloid fibrils as building blocks for natural and artificial functional materials." *Advanced Materials* **2016**, *28(31)*, 6546-6561.

11. Wei, G., Su, Z., Reynolds, N. P., Arosio, P., Hamley, I. W., Gazit, E. and Mezzenga, R. "Self-assembling peptide and protein amyloids: from structure to tailored function in nanotechnology." *Chemical Society Reviews* **2017**, *46(15)*: 4661-4708.

12. Pashuck, E. T., Duchet, B. J., Hansel, C. S., Maynard, S. A., Chow, L. W. and Stevens, M. M. "Controlled sub-nanometer epitope spacing in a three-dimensional self-assembled peptide hydrogel." *ACS nano* **2016**, *10(12)*, 11096-11104.

13. Bracha, D., Walls, M. T. and Brangwynne, C. P. "Probing and engineering liquid-phase organelles." *Nature biotechnology* **2019**, *37(12)*, 1435-1445.

14. Altenburg, W. J., Yewdall, N. A., Vervoort, D. F., Van Stevendaal, M. H., Mason, A. F. and van Hest, J. C. "Programmed spatial organization of biomacromolecules into discrete, coacervate-based protocells." Nature communications 11.1 (2020): 1-10.

15. Faltova, L., Küffner, A. M., Hondele, M., Weis, K. and Arosio, P. "Multifunctional protein materials and microreactors using low complexity domains as molecular adhesives." *ACS nano* **2018**, 12(10),9991-9999.

16. Papoian, G. A. "Proteins with weakly funneled energy landscapes challenge the classical structure–function paradigm." *Proceedings of the National Academy of Sciences* **2008**, *105(38)*, 14237-14238.

17. Malinovska, L., Kroschwald, S. and Alberti, S. "Protein disorder, prion propensities, and self-organizing macromolecular collectives." *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2013**, *1834(5)*, 918-931.

18. Uversky, V. N. and Dunker, A. K. "Understanding protein non-folding." *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 2010, *1804(6)*, 1231-1264.

19. Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T. and Kim, P. M. "Classification of intrinsically disordered regions and proteins." *Chemical reviews* **2014**, *114(13)*, 6589-6631.

20. Hazra, M. K. and Levy, Y. "Biophysics of Phase Separation of Disordered Proteins Is Governed by Balance between Short-And Long-Range Interactions." *The Journal of Physical Chemistry B* **2021**, *125(9)*, 2202-2211.

21. Schuster, B. S., Dignon, G. L., Tang, W. S., Kelley, F. M., Ranganath, A. K., Jahnke, C. N., Simpkins, A. G., Regy, R. M., Hammer, D. A., Good, M. C. and Mittal, J. "Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior." *Proceedings of the National Academy of Sciences* **2020**, *117(21)*, 11421-11431.

22. Paloni, M., Bailly, R., Ciandrini, L. and Barducci, A. "Unraveling Molecular Interactions in Liquid–Liquid Phase Separation of Disordered Proteins by Atomistic Simulations." *The Journal of Physical Chemistry B* **2020***, 124(41)*, 9009-9016.

23. Lin, Y. H. and Chan, H. S. "Phase separation and single-chain compactness of charged disordered proteins are strongly correlated." *Biophysical Journal* **2017**, *112(10)*, 2043-2046.

24. Chu, W. T. and Wang, J. "Thermodynamic and sequential characteristics of phase separation and droplet formation for an intrinsically disordered region/protein ensemble." *PLoS computational biology* **2021**, 17(3), e1008672.

25. Lin, Y. H., Forman-Kay, J. D. and Chan, H. S. "Theories for sequence-dependent phase behaviors of biomolecular condensates." *Biochemistry* **2018**, 57(17), 2499-2508.

26. Das, R. K. and Pappu, R. V. "Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues." *Proceedings of the National Academy of Sciences* **2013**, *110(33)*, 13392-13397.

27. Wang, J., Choi, J. M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D. and Poser, I. "A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins." *Cell* **2018**, *174(3)*, 688-699.

28. Martin, E. W., Holehouse, A. S., Peran, I., Farag, M., Incicco, J. J., Bremer, A., Grace, C. R., Soranno, A., Pappu, R. V. and Mittag, T. "Valence and patterning of aromatic residues determine the phase behavior of prion-like domains." Science **2020**, 367 (6478), 694-699.

29. Gabryelczyk, B., Cai, H., Shi, X., Sun, Y., Swinkels, P. J., Salentinig, S., Pervushin, K. and Miserez, A. "Hydrogen bond guidance and aromatic stacking drive liquid-liquid phase separation of intrinsically disordered histidine-rich peptides." *Nature communications* **2019**, *10(1)*, 1-12.

30. Lin, Y., Currie, S. L. and Rosen, M. K. "Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs." *Journal of Biological Chemistry* **2017**, *292(46)*, 19110-19120.

31. McCarty, J., Delaney, K. T., Danielsen, S. P., Fredrickson, G. H. and Shea, J. E. "Complete phase diagram for liquid–liquid phase separation of intrinsically disordered proteins." *The journal of physical chemistry letters* **2019**, *10(8)*, 1644-1652.

32. Martin, E.W. and Mittag, T. "Relationship of sequence and phase separation in protein low-complexity regions." *Biochemistry* **2018**, *57(17)*, 2478-2487.

33. Choi, J. M., Holehouse, A. S. and Pappu, R. V. "Physical principles underlying the complex biology of intracellular phase transitions*." Annual Review of Biophysics* **2020**, *49*, 107-133.

34. Yang, Y., Jones, H. B., Dao, T. P. and Castañeda, C. A. "Single amino acid substitutions in stickers, but not spacers, substantially alter UBQLN2 phase transitions and dense phase material properties." *The Journal of Physical Chemistry B* **2019**, *123(17)*, 3618-3629.

35. Quiroz, F. G. and Chilkoti, A. "Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers." *Nature materials* **2015**, *14(11)*, 1164-1171.

36. Choi, J. M., Dar, F. and Pappu, R. V. "LASSI: A lattice model for simulating phase transitions of multivalent proteins." *PLoS computational biology* **2019** 15(10), e1007028.

37. Shrestha, U. R., Juneja, P., Zhang, Q., Gurumoorthy, V., Borreguero, J. M., Urban, V., Cheng, X., Pingali, S. V., Smith, J. C., O'Neill, H. M. and Petridis, L. "Generation of the configurational ensemble of an intrinsically disordered protein from unbiased molecular dynamics simulation." *Proceedings of the National Academy of Sciences* **2019**, *116(41)*: 20446-20452.

38. Zerze, G. H., Miller, C. M., Granata, D. and Mittal, J. "Free energy surface of an intrinsically disordered protein: comparison between temperature replica exchange molecular dynamics and bias-exchange metadynamics." *Journal of chemical theory and computation* **2015**,*11(6)*, 2776-2782

39. Liu, N., Guo, Y., Ning, S. and Duan, M. "Phosphorylation regulates the binding of intrinsically disordered proteins via a flexible conformation selection mechanism." *Communications Chemistry* **2020**, *3(1)*, 1-9

40. Shrestha, U. R., Smith, J. C. and Petridis, L. "Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations." *Communications biology* **2021**, *4(1)*, 1-8.

41. Appadurai, R., Nagesh, J. and Srivastava, A. "High resolution ensemble description of metamorphic and intrinsically disordered proteins using an efficient hybrid parallel tempering scheme." *Nature communications* **2021**, *12(1)*, 1-11.

42. Bari, K. J. and Prakashchand, D. D. "Fundamental Challenges and Outlook in Simulating Liquid–Liquid Phase Separation of Intrinsically Disordered Proteins." *The Journal of Physical Chemistry Letters* **2021**, *12(6)*, 1644-1656.

43. Herrera-Nieto, P., Pérez, A. and De Fabritiis, G. "Characterization of partially ordered states in the intrinsically disordered N-terminal domain of p53 using millisecond molecular dynamics simulations." *Scientific reports* **2020**, *10(1)*, 1-8.

44. Pietrek, L. M., Stelzl, L. S. and Hummer, G. "Hierarchical ensembles of intrinsically disordered proteins at atomic resolution in molecular dynamics simulations." *Journal of chemical theory and computation* **2019**, *16(1)*, 725-737.

45. Das, P., Matysiak, S. and Mittal, J. "Looking at the disordered proteins through the computational microscope." *ACS central science* **2018**, *4(5)*, 534-542.

46. Deighan, M., Bonomi, M. and Pfaendtner, J. "Efficient simulation of explicitly solvated proteins in the well-tempered ensemble." *Journal of chemical theory and computation* **2012**, *8(7)*, 2189-2192.

47. Barducci, A., Bonomi, M., Prakash, M. K. and Parrinello, M. "Free-energy landscape of protein oligomerization from atomistic simulations." *Proceedings of the National Academy of Sciences* **2013**, *110(49)*, E4708-E4713.

48. Shea, J. E., Best, R. B. and Mittal, J. "Physics-based computational and theoretical approaches to intrinsically disordered proteins." *Current Opinion in Structural Biology* **2021**, *67*, 219-225.

49. Chong, S. H., Chatterjee, P. and Ham, S. "Computer simulations of intrinsically disordered proteins." *Annual review of physical chemistry* **2017**, *68*, 117-134.

50. Sugita, Y. and Okamoto, Y. "Replica-exchange molecular dynamics method for protein folding." *Chemical physics letters* **1999**, *314(1-2)*, 141-151.

51. Fukunishi, H., Watanabe, O. and Takada, S. "On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction*." The Journal of chemical physics* **2002**, *116(20)*, 9058-9067.

52. Jensen, M. R., Zweckstetter, M., Huang, J. R. and Blackledge, M. "Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy." *Chemical reviews* **2014**, *114(13)*, 6632-6660.

53. Robustelli, P., Stafford, K. A. and Palmer III, A. G. "Interpreting protein structural dynamics from NMR chemical shifts." *Journal of the American Chemical Society* **2012**, *134(14)*, 6365-6374.

54. Fuertes, G., Banterle, N., Ruff, K. M., Chowdhury, A., Mercadante, D., Koehler, C., Kachala, M., Girona, G. E., Milles, S., Mishra, A. and Onck, P. R. "Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements." *Proceedings of the National Academy of Sciences* **2017**, *114(31)*, E6342-E6351.

55. Schuler, B., Soranno, A., Hofmann, H. and Nettels, D. "Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins." *Annual Review of Biophysics* 2016, *45*, 207-231.

56. Riback, J. A., Bowman, M. A., Zmyslowski, A. M., Knoverek, C. R., Jumper, J. M., Hinshaw, J. R., Kaye, E. B., Freed, K. F., Clark, P. L. and Sosnick, T. R. "Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water." *Science* **2017**, *358(6360)*, 238-241.

57. Johansen, D., Jeffries, C. M., Hammouda, B., Trewhella, J. and Goldenberg, D. P. "Effects of macromolecular crowding on an intrinsically disordered protein characterized by small-angle neutron scattering with contrast matching." *Biophysical journal* 2011, *100(4)*, 1120-1128.

58. Li, L., Belcher, A. M. and Loke, D. K. "Simulating selective binding of a biological template to a nanoscale architecture: a core concept of a clamp-based

binding-pocket-favored N-terminal-domain assembly." *Nanoscale* **2020**, 12(47), 24214-24227.

59. Liu, P., Kim, B., Friesner, R. A. and Berne, B. J. "Replica exchange with solute tempering: A method for sampling biological systems in explicit water." *Proceedings of the National Academy of Sciences* **2005**, *102(39)*, 13749-13754.

60. Wang, L., Friesner, R. A. and Berne, B. J. "Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2)." *The Journal of Physical Chemistry B* **2011**, *115(30)*, 9431-9438.

61. Branduardi, D., Bussi, G. and Parrinello, M. "Metadynamics with adaptive Gaussians." *Journal of chemical theory and computation* **2012**, *8(7)*, 2247-2254.

62. Laio, A. and Parrinello, M. "Escaping free-energy minima." *Proceedings of the National Academy of Sciences* **2002**, *99(20)*, 12562-12566.

63. Bussi, G. and Laio, A. "Using metadynamics to explore complex free-energy landscapes." *Nature Reviews Physics* **2020**, *2(4)*, 200-212.

64. Piana, S. and Laio, A. "A bias-exchange approach to protein folding." *The journal of physical chemistry B* **2007**, *111(17)*, 4553-4559.

65. Granata, D., Baftizadeh, F., Habchi, J., Galvagnion, C., De Simone, A., Camilloni, C., Laio, A. and Vendruscolo, M. "The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments." *Scientific reports* **2015**, *5(1)*, 1-15.

66. Bonomi, M. and Parrinello, M. "Enhanced sampling in the well-tempered ensemble." *Physical review letters* **2010**, *104(19)*, 190601.

67. Barducci, A., Bussi, G. and Parrinello, M. "Well-tempered metadynamics: a smoothly converging and tunable free-energy method." *Physical review letter*s **2008**, *100(2)*, 020603.

68. Barducci, A., Bonomi, M. and Parrinello, M. "Linking well-tempered metadynamics simulations with experiments." *Biophysical journal* **2010**, *98(9)* , L44-L46

69. Pfaendtner, J. and Bonomi, M. "Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics." *Journal of chemical theory and computation* **2015**, *11(11)*, 5062-5067.

70. Laio, A., Rodriguez-Fortea, A., Gervasio, F.L., Ceccarelli, M. and Parrinello, M. "Assessing the accuracy of metadynamics." *The journal of physical chemistry B* **2005**, *109(14)*, 6714-6721.

71. Schäfer, T. M. and Settanni, G. "Data reweighting in metadynamics simulations." *Journal of chemical theory and computation* **2020**,*16(4)*, 2042-2052.

72. Tiwary, P. and Parrinello, M. "A time-independent free energy estimator for metadynamics." *The Journal of Physical Chemistry B* **2015**,119(3), 736-742.

73. Bonomi, M., Barducci, A. and Parrinello, M. "Reconstructing the equilibrium Boltzmann distribution from well‐tempered metadynamics." *Journal of computational chemistry* **2009**,30(11), 1615-1621.

74. Robustelli, P., Piana, S. and Shaw, D. E. "Developing a molecular dynamics force field for both folded and disordered protein states." *Proceedings of the National Academy of Sciences* **2018**, *115(21)*, E4758-E4766.

75. Palazzesi, F., Prakash, M. K., Bonomi, M. and Barducci, A. "Accuracy of current all-atom force-fields in modeling protein disordered states." *Journal of chemical theory and computation* **2015**, *11(1)*, 2-7.

76. Rauscher, S., Gapsys, V., Gajda, M. J., Zweckstetter, M., de Groot, B. L. and Grubmüller, H. "Structural ensembles of intrinsically disordered proteins depend

strongly on force field: a comparison to experiment." *Journal of chemical theory and computation* **2015**, *11(11)*, 5513-5524.

77. Yu, L., Li, D.W. and Brüschweiler, R. "Systematic Differences between Current Molecular Dynamics Force Fields To Represent Local Properties of Intrinsically Disordered Proteins." *The Journal of Physical Chemistry B* **2021**, *125(3)*, 798-804.

78. Huang, J. and MacKerell Jr, A.D. "Force field development and simulations of intrinsically disordered proteins." *Current opinion in structural biology* **2018**, *48*, 40-48.

79. Zerze, G. H., Zheng, W., Best, R. B. and Mittal, J. "Evolution of all-atom protein force fields to improve local and global properties." *The journal of physical chemistry letters* **2019**, *10(9)*, 2227-2234.

80. Piana, S., Donchev, A.G., Robustelli, P. and Shaw, D. E. "Water dispersion interactions strongly influence simulated structural properties of disordered protein states." *The journal of physical chemistry B* **2015**, *119(16)*, 5113-5123.

81. Best, R. B., Zheng, W. and 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)*, 5113-5124.

82. Nerenberg, P. S., Jo, B., So, C., Tripathy, A. and Head-Gordon, T. "Optimizing solute–water van der waals interactions to reproduce solvation free energies." *The Journal of Physical Chemistry B* **2012**, *116(15)*, 4524-4534.

83. Best, R. B. and Hummer, G. "Optimized molecular dynamics force fields applied to the helix−coil transition of polypeptides." *The journal of physical chemistry B* **2009**, *113(26)*, 9004-9015.

84. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O. and Shaw, D. E. "Improved side-chain torsion potentials for the Amber ff99SB protein force field." *Proteins: Structure, Function, and Bioinformatics* **2010**, *78(8)*, 1950-1958.

85. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. and Klein, M. L. "Comparison of simple potential functions for simulating liquid water." *The Journal of chemical physics* **1983**, *79(2)*, 926-935.

86. Piana, S., Lindorff-Larsen, K. and Shaw, D. E. "How robust are protein folding simulations with respect to force field parameterization?." *Biophysical journal* **2011**, *100(9)*, L47-L49.

87. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H. and MacKerell, A. D. "CHARMM36: An improved force field for folded and intrinsically disordered proteins." *Biophysical Journal* **2017**, *112(3)*, 175a-176a.

88. MacKerell Jr, A. D., Bashford, D., Bellott, M. L. D. R., Dunbrack Jr, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S. and Joseph-McCarthy, D. "All-atom empirical potential for molecular modeling and dynamics studies of proteins." *The journal of physical chemistry B* **1998**, *102(18)*, 3586-3616.

89. Pineda, L. I. G., Milko, L. N. and He, Y. "Performance of CHARMM36m with modified water model in simulating intrinsically disordered proteins: a case study." *Biophysics Reports* **2020**, *6(2)*, 80-87.

90. Carballo-Pacheco, M. and Strodel, B. "Comparison of force fields for Alzheimer's A: A case study for intrinsically disordered proteins." *Protein science* **2017**, *26(2)*, 174-185.

91. Hondele, M., Sachdev, R., Heinrich, S., Wang, J., Vallotton, P., Fontoura, B. M. and Weis, K. "DEAD-box ATPases are global regulators of phase-separated organelles." *Nature* **2019**, *573(7772)*, 144-148.

92. Linsenmeier, M., Kopp, M. R., Grigolato, F., Emmanoulidis, L., Liu, D., Zürcher, D., Hondele, M., Weis, K., Capasso Palmiero, U. and Arosio, P. "Dynamics of synthetic membraneless organelles in microfluidic droplets." *Angewandte Chemie* **2019**, *131(41)*, 14631-14636.

93. Pietrucci, F. and Laio, A."A collective variable for the efficient exploration of protein beta-sheet structures: application to SH3 and GB1." *Journal of Chemical Theory and Computation* **2009**, *5(9)*, 2197-2201.

94. Vymětal, J. and Vondrášek, J. "Gyration-and inertia-tensor-based collective coordinates for metadynamics. Application on the conformational behavior of polyalanine peptides and Trp-cage folding." *The Journal of Physical Chemistry A* **2011**, *115(41)*, 11455-11465.

95. Frenkel, D. and Smit, B. *Understanding molecular simulation: from algorithms to applications*. Vol. 1. Elsevier, **2001**.

96. Flyvbjerg, H. and Petersen, H.G. "Error estimates on averages of correlated data." *The Journal of Chemical Physics* **1989**, *91(1)*, 461-466.

97. Bussi, G. and Tribello, G.A. "Analyzing and biasing simulations with PLUMED." Biomolecular Simulations. Humana, New York, NY, 2019. 529-578.

98. Kabsch, W. and Sander, C. "Dictionary of protein secondary structure: pattern recognition of hydrogen‐bonded and geometrical features." *Biopolymers: Original Research on Biomolecules* **1983**, *22(12)*, 2577-2637.

99. Nilsson, I., SaÈaÈf, A., Whitley, P., Gafvelin, G., Waller, C. and von Heijne, G. "Proline-induced disruption of a transmembrane α-helix in its natural environment." *Journal of molecular biology* **1998**, *284(4)*, 1165-1175.

100. Woolfson, D. N. and Williams, D. H. "The influence of proline residues on α-helical structure." *FEBS letters* **1990**, 277(1-2), 185-188.

101. Bianchi, G., Longhi, S., Grandori, R. and Brocca, S. "Relevance of electrostatic charges in compactness, aggregation, and phase separation of intrinsically disordered proteins." *International Journal of Molecular Sciences* 2020, 21(17), 6208.

102. Martin, E.W., Holehouse, A.S., Peran, I., Farag, M., Incicco, J. J., Bremer, A., Grace, C. R., Soranno, A., Pappu, R. V. and Mittag, T. "Valence and patterning of aromatic residues determine the phase behavior of prion-like domains." Science **2020**, *367(6478)*, 694-699.

103. Pak, C. W., Kosno, M., Holehouse, A. S., Padrick, S. B., Mittal, A., Ali, R., Yunus, A. A., Liu, D. R., Pappu, R. V. and Rosen, M. K. "Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein." *Molecular cell* **2016**, *63(1)*, 72-85.

104. Vernon, R. M., Chong, P. A., Tsang, B., Kim, T. H., Bah, A., Farber, P., Lin, H. and Forman-Kay, J. D. "Pi-Pi contacts are an overlooked protein feature relevant to phase separation." *elife* **2018**, *7*, e31486.

105. Das, S., Lin, Y. H., Vernon, R. M., Forman-Kay, J. D. and Chan, H. S. "Comparative roles of charge, π, and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins." *Proceedings of the National Academy of Sciences*, **2020**, *117(46)*, 28795-28805.

106.   Yang, J. and Zhang, Y. "I-TASSER server: new development for protein structure and function predictions." *Nucleic acids research* **2015**, *43(W1)*, W174-W181.

107.   Song, Y., DiMaio, F., Wang, R. Y. R., Kim, D., Miles, C., Brunette, T. J., Thompson, J. and Baker, D. "High-resolution comparative modeling with RosettaCM." *Structure* **2013**, *21(10)*, 1735-1742.

108.   Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B. and Lindahl, E. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers." *SoftwareX* **2015**, *1*,19-25.

109.   Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. and Bussi, G. "PLUMED 2: New feathers for an old bird." *Computer Physics Communications* **2014**, *185(2)*, 604-613.

110.   Bussi, G., Donadio, D. and Parrinello, M. "Canonical sampling through velocity rescaling." *The Journal of chemical physics* **2007**, *126(1)*, 014101.

111.   Parrinello, M. and Rahman, A. "Polymorphic transitions in single crystals: A new molecular dynamics method." *Journal of Applied physics* **1981**, *52(12)*, 7182-7190.

112.   Páll, S. and Hess, B. "A flexible algorithm for calculating pair interactions on SIMD architectures." *Computer Physics Communications* **2013**, *184(12)*, 2641-2650.

113.   Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. and Pedersen, L. G. "A smooth particle mesh Ewald method." *The Journal of chemical physics* **1995**, *103(19)*, 8577-8593.

114.   https://manual.gromacs.org/2019/user-guide/force-fields.html

115. Hess, B. "P-LINCS: A parallel linear constraint solver for molecular simulation." *Journal of chemical theory and computation* **2008**, *4(1)*, 116-122.

116. Bonomi, M. "Promoting transparency and reproducibility in enhanced molecular simulations." *Nature methods* **2019**, *16(8)*, 670-673.