Combining experiments and simulations to examine the temperature-dependent behaviour of a disordered protein

Francesco Pesce and Kresten Lindorff-Larsen*

Structural Biology and NMR Laboratory, The Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, Copenhagen, Denmark

E-mail: lindorff@bio.ku.dk

Abstract

Intrinsically disordered proteins are a class of proteins that lack stable folded conformations and instead adopt a range of conformations that determine their biochemical functions. The temperature-dependent behaviour of such disordered proteins is complex and can vary depending on the specific protein and environment. Here, we have used molecular dynamics simulations and previously published experimental data to investigate the temperature-dependent behaviour of Histatin 5, a 24-residue-long polypeptide. We examined the hypothesis that Histatin 5 undergoes a loss of polyproline II structure with increasing temperature, leading to more compact conformations. We found that the conformational ensembles generated by the simulations generally agree with small-angle X-ray scattering data for Histatin 5, but show some discrepancies with the hydrodynamic radius as probed by pulsed-field gradient nuclear magnetic resonance spectroscopy, and with the secondary structure information derived from circular dichroism. We attempted to reconcile these differences by reweighting the conformational ensembles against the scattering and NMR data. By doing so, we were

in part able to capture the temperature-dependent behaviour of Histatin 5 and to link the observed decrease in hydrodynamic radius with increasing temperature to a loss of polyproline II structure. We were, however, unable to achieve agreement with both the scattering and NMR data within experimental errors. We discuss different possibilities for this outcome including inaccuracies in the force field, differences in conditions of the NMR and scattering experiments, and issues related to the calculation of the hydrodynamic radius from conformational ensembles. Our study highlights the importance of integrating multiple types of experimental data when modelling conformational ensembles of disordered proteins and how environmental factors such as the temperature influence them.

Introduction

Intrinsically disordered proteins (IDPs) are a class of proteins characterized by the lack of stable folded conformations. Instead, IDPs may adopt a wide range of conformations, which help determine their biochemical functions.[1,2] This ensemble of conformations, although disordered, may show sequence-dependent populations of specific regions of the Ramachandran map,[3] and the formation of transient secondary structures is common.[4,5] Therefore, they are better described as statistical coils.[6] Biophysical measurements make it possible to get insights into the propensity of IDPs to form (transient) secondary structures, although the interpretation of ensemble-averaged data is often ambiguous.[7] Alternatively, molecular dynamics (MD) simulations enable direct access to the population of conformations of IDPs at atomistic resolution, but its utility may be limited by difficulties in sampling and by the accuracy of the force fields used.[8–12]

The complexity of the problem increases even more when considering that the behaviour of IDPs is often highly dependent on the environment.[13] One factor that may affect the behaviour of IDPs is the temperature.[14–23] The temperature-dependence of the conformational properties of IDPs may, however, be complex, as they depend on the combined effects of

many different physical effects including the hydrophobic effect, changes in local structure, and other entropic contributions, and may vary depending on the specific protein and the specific environment in which it is found.[24–26] Other than influencing the single chain behaviour, temperature is also known to influence how IDPs interact and their tendency to form high-order entities.[27,28] Therefore, understanding how temperature affects the single-chain behaviour, may provide new insights into how aggregates and biomolecular condensates can form as a consequence of particular single-chain behaviours.[29]

Experimentally, the temperature-dependent behaviour of the local structure of IDPs has previously been studied using for example circular dichroism (CD) spectroscopy,[4,5,30–32] where changes in the spectra, although difficult to interpret at atomistic resolution, have been suggested to be due to formation of α-helices and/or loss of polyproline II (PPII) structures with increasing temperature.[15,23,33,34] This change in propensity to form certain secondary structure elements has also been linked to changes in the overall dimension of IDPs.[5,14,15,23,35] For example, it has been suggested that the loss of PPII with increasing temperature is responsible for the conformational ensemble of certain IDPs to get more compact.[15,23] Even though this might sound counter intuitive, it could be explained by the fact that the PPII structure is a particularly extended secondary structure element, though other mechanisms are also likely to be important.[14,24,25]

The overall chain dimension of IDPs can be probed, for example, by small-angle X-ray scattering (SAXS) or pulsed-field gradient (PFG) NMR spectroscopy. The former yields structural information that are often linked to the average radius of gyration (R_{g}) of an IDP in solution,[36] while the latter yields the translational diffusion coefficient, which is in turn related to the average hydrodynamic radius (R_{h}) of an IDP in solution.[37] R_{g} and R_{h} are both measures that describe chain compaction, and are therefore correlated.[38,39] Nonetheless, SAXS and PFG NMR may sometimes be used to probe subtly different aspects of compaction.[38] Specifically, SAXS probes the ensemble averaged R_{g} as $\sqrt{(\sum_i (R_{\mathrm{g,i}}^2))}$ for each conformation i in the solution ensemble. Conversely, the translational diffusion coefficient

measured by PFG NMR, when converted into an R_h is averaged as $1/\sum_i(1/R_{\mathrm{h,i}})$. By consequence of this different averaging, compaction measured by SAXS may put greater weight on expanded conformations, while compaction measured by PFG NMR is more representative of compact conformations.[38] Both theory and numerical calculations show that the$R_\mathrm{g}/R_\mathrm{h}$ depends on compaction[38–40] and so a joint analysis of SAXS and PFG NMR measurements may provide additional information compared to analysing these individually.[38,41]

Here, we use the small protein Histatin 5 (hereafter Hst5) to study the relationship between local and global structure of a disordered protein and the effect of temperature on these properties. This 24 residue-long polypeptide has previously been characterized by a multitude of experiments.[23,42–47] CD spectra recorded at different temperatures between 283 K and 323 K suggested that a loss of PPII structure takes place with increasing temperature. Increasing the temperature also results in a decrease of the R_h measured by PFG NMR, while the R_g extracted from SAXS data appears not to be particularly sensitive to temperature. Previous attempts of computational characterization of the conformational ensembles of Hst5 at different temperatures did not completely succeed in explaining the experimental results,[23] but it was suggested that the drop in R_h with increasing temperature might be due to the temperature-dependent change in the propensity of forming PPII structures.

To provide a structural explanation of the experimental data, we extensively sample the conformational ensemble of Hst5 at different temperatures by running MD simulations enhanced with metadynamics.[48] We then compare the conformational ensembles to the available temperature-dependent experimental data, but do not observe substantial temperature-dependent changes in compaction and propensity to form PPII structures. Inspired by previous work that combined experiments and simulations to reconcile different experimental observations,[49–51] we use the Bayesian/Maximum-entropy reweighting approach[52–54] to refine the simulations against SAXS and R_h from PFG NMR.[41] We find that when we improve the agreement with the R_h we also observe a temperature-dependent change in the amount of PPII structure in qualitative agreement with the CD data, therefore providing

a molecular-level description of a link between chain compaction and propensity to form PPII-like structure. Conversely, when we combine the simulations with the SAXS data we do not observe any temperature-dependent change in compaction and PPII structure content. When we refine the ensembles simultaneously against the SAXS and NMR data, we were not able to achieve the same level of agreement as when we optimized against them individually. We discuss possible origins of this observation, including minor differences in the conditions of the experiments, force fields inaccuracies, and issues related to how R_{h} values are estimated from conformational ensembles of IDPs.

Results and Discussion

Comparison of MD simulations and experiments

We sampled the conformational ensembles of Hst5 at 283 K, 293 K, 310 K and 323 K with MD simulations using the all-atom Amber ff99SB-disp (a99SB-disp) force field,[11] as this has previously been shown to be reliable in capturing the overall chain dimension of IDPs[11] and in reproducing NMR scalar couplings of a peptide with a high PPII structure propensity.[55] In order to improve the sampling of the overall dimensions and secondary structure preferences of Hst5, we used Parallel Bias Metadynamics[56] (PBMetaD; see Methods) simulations with two semi-independent replicas biasing the same free-energy surface (FES).[57] We ran these simulations until the sampling of the collective variables (CVs) that we biased with PBMetaD converged within a reasonable error (*i.e.* an average error on the estimated free-energy surface (FES) of the R_{g} lower than 1 kJ/mol; Fig. S1). We achieved this in all cases in a simulation time comprised between 2 and 2.5 μs per replica. We also verify that, along the simulations, both R_{g} and the backbone dihedrals diffuse in the CV space and do not get stuck in a specific state, supporting that the extent of biasing used in the simulations is sufficient for the system to escape local minima (Fig. S2 and S3). Furthermore, because of the well-tempered biasing[58] (see Methods), it would be ideal for the amount of bias dispensed

in the system, which decreases over time, to reach a quasi-static regime, to ensure a proper convergence of the simulation. This happens when all accessible regions of the FES have been explored. In our simulations, we observe that overall the bias deposition and free-energy surfaces become quasi-static over time (Fig. S4), with the exception when sampling frames whose R_g is outside the interval in which the bias is active (Fig. S4b). Therefore these few frames are not taken into account when calculating averages and errors (see Methods).

We then calculated SAXS intensities, R_g, R_h, and per-residue secondary structure propensities, and compared these to the experimental values. We find that all simulations—apart from that at 310 K—provide a χ_r^2 between the calculated and experimental SAXS data close to 1 (Fig. 1a), which, given the use of normalized error bars for SAXS intensities (see Methods), shows that the ensembles accurately fit the SAXS experiments. As previously reported,[23] we find that the R_g values derived from the experimental SAXS profile (with the Guinier approximation, using the ATSAS package[59]) show that the average dimension for Hst5 in solution at the different temperatures are very similar (Fig. 1b). Similarly, the conformational ensembles sampled by MD provide average R_g values that are all comparable when taking the errors into account (with a subtle trend of expansion with increasing temperature), but consistently smaller than those from the Guinier approximation. This observation is in line with the expectation that SAXS-derived R_g carries the contribution of the protein hydration layer,[60] while from the conformational ensembles the R_g is calculated on the coordinates of the protein atoms only.[61] For this reason we only rely on the χ_r^2 to SAXS data to determine whether a certain ensemble is accurate in terms of chain dimensions as described by SAXS data, since in calculating the SAXS we take into account both the protein hydration layer and the excluded volume effect (see Methods).

The calculated R_h values appear to be consistently smaller than the experimental values determined by PFG NMR at any of the temperatures (Fig. 1c). This observation might indicate that the a99SB-disp force field underestimates the R_h of Hst5. However, we note a number of complications when calculating the R_h from conformational ensembles of IDPs[62,63] as

well as issues related to how the PFG NMR measurements are used to estimate R_h.[63] Specifically, we recently showed[63] that the R_h calculated with HullRadSAS[64] provides the best overall agreement with experiments across a series of IDPs. Nonetheless, we note that this observation depends on the R_h value used for the reference compound (here 1,4-dioxane[63,65]) that is used to derive the R_h from the diffusion data.[62,63] Thus, earlier values of the R_h of dioxane[65] suggested that a different approach to calculate R_h was more accurate.[62] Finally, the comparison between averages of R_h values calculated for individual conformations and the translational diffusion coefficient rests on the assumption of treating R_h as an ensemble- and time-averaged quantity.[66] Together, these issues mean that it is difficult to determine the origin of the deviation between experiments and simulations.

The CD spectra of Hst5 show profiles often taken to represent PPII structure. The decrease in intensity of the PPII signature with increasing temperature suggests that this secondary structure element is less populated with increasing temperature.[23] Comparing CD spectra to simulations is difficult as previous attempts at calculating CD spectra from conformational ensembles of disordered peptides have shown large discrepancies between the experimental and calculated spectra.[55,67,68] Also, we are not aware of approaches to deconvolute CD spectra of disordered proteins that takes into account PPII-like signals and which can accurately extract the average content of PPII structure. For these reasons, we instead perform a qualitative comparison of our ensembles to the CD data of Hst5 by calculating the average content of PPII structure in our ensembles and examine its temperature dependence. We find that the ensembles do not show any substantial differences in PPII structure at the different temperatures, with changes of at most 0.6% (Fig. 1d). All together these observations suggest that the force field underestimates the R_h from PFG NMR and does not show changes in compaction and propensity to form PPII structure with temperature.

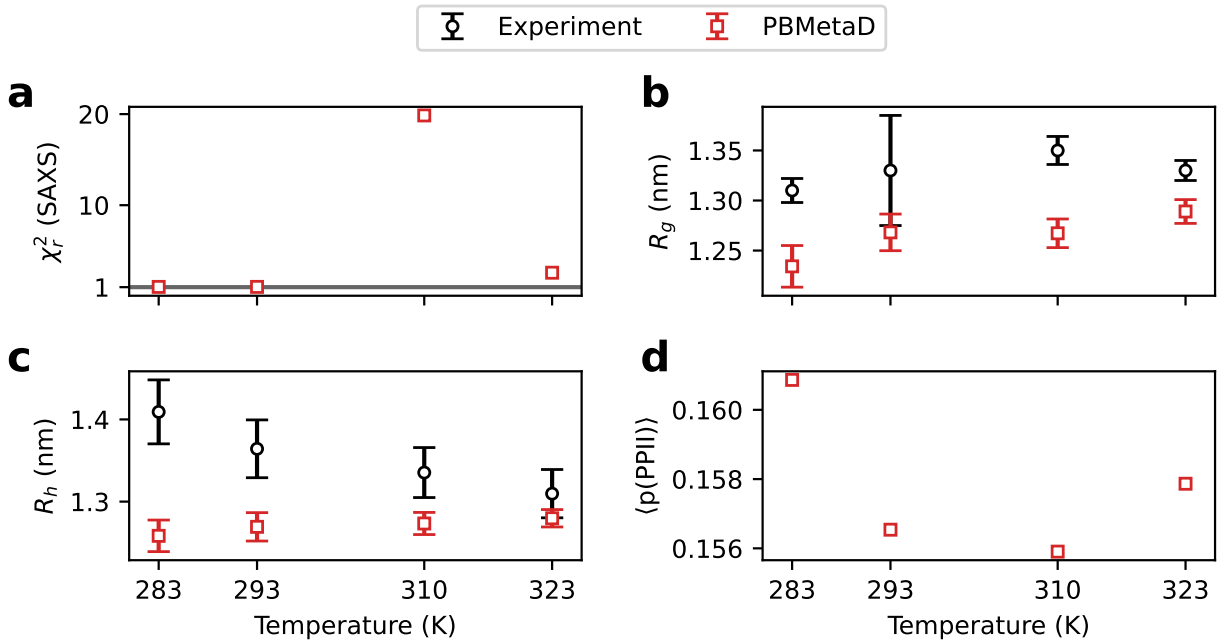

Figure 1: Average structural features calculated from MD simulations (red squares) at four different temperatures are compared to their experimental counterparts (black circles, when applicable). (a) χ^2 to SAXS data (grey horizontal line indicates χ_r^2=1, that denotes ensembles that correctly fit the experiments; see Methods), (b) R_{g}, (c) R_{h}, (d) PPII structure content. As discussed in the main text, the deviation in R_{g} (panel b), despite good agreement with the SAXS intensities (panel a), may likely in part be attributed to the contribution of the solvation layer to the R_{g} derived from Guinier fitting.[61]

Refinement of the simulations against SAXS experiments

To test the hypothesis that a decrease in chain dimension comes with a lowered propensity to form PPII structure, we improved the agreement between the ensembles and data that encode information about the protein dimension at different temperatures. We first used the iBME[69] approach to reweight the ensembles against SAXS data; in this way we construct conformational ensembles that are in better agreement with experiments, but with minimal perturbations to the ensembles generated by the force field alone. With very mild reweighting (Fig. S5), we achieved a good agreement with SAXS data for all simulations (Fig. 2a,b), but without any major change in R_{h} (Fig. 2c) and PPII structure propensity (Fig. 2d), and only the ensemble simulated at 310 K getting slightly more compact as seen by both R_{h} and R_{g} (Fig. 2c and S6). We note that the larger χ_r^2 before reweighting for the simulation at

310 K (and to a minor extent also that at 323 K) is likely due to a higher signal-to-noise ratio in the experimental SAXS profile at these temperatures compared to those at 283 K and 293 K (Fig. 2a). This difference arises from the use of more concentrated samples in the SAXS measurements at 310 K and 323 K.[23] This, and the fact that we do not observe major changes in the ensembles upon reweighting against SAXS, suggest that the ensembles from PBMetaD fit the experimental SAXS data well at all temperatures and differences in χ_r^2 are due to different magnitudes of the error bars among the experiments at different temperatures. Indeed the normalized average error bars (see Methods) are 20%, 14%, 3% and 8% for SAXS data collected respectively at 283 K, 293 K, 310 K and 323 K. This suggests that the differences in χ_r^2 observed are in part due to the different magnitude of the errors across the experiments, though issues related to convergence and/or force field inaccuracy might also be in play. As seen also from the R_{g} values extracted from the experimental SAXS profiles, the SAXS data do not suggest any substantial change in average chain dimension with temperature. Nonetheless, even though the SAXS data does not change substantially as a function of temperature, this does not mean that the temperature may not influence the average dimension of Hst5 in solution, as SAXS experiments only probe certain aspects of chain compaction.[38,49,50,62] We note also the difference between the R_{g} value calculated from the protein coordinates and from a Guinier analysis of SAXS data calculated from the same coordinates (Table S6), and that such differences have previously been shown to arise from a contribution from the solvent to the SAXS data.[61]

Refinement of the simulations against R_{h} from PFG NMR

We also reweighted the simulations against the R_{h} from PFG NMR independently from the SAXS data. With the level of reweighting that we applied (Fig. S7), the reweighted R_{h} values from the simulations at 293 K, 310 K and 323 K come within the error bars associated to the experimental values, while the reweighted R_{h} of the simulations at 283 K remains below the experimental value (Fig 3a). The sampling at 283 K is more restricted

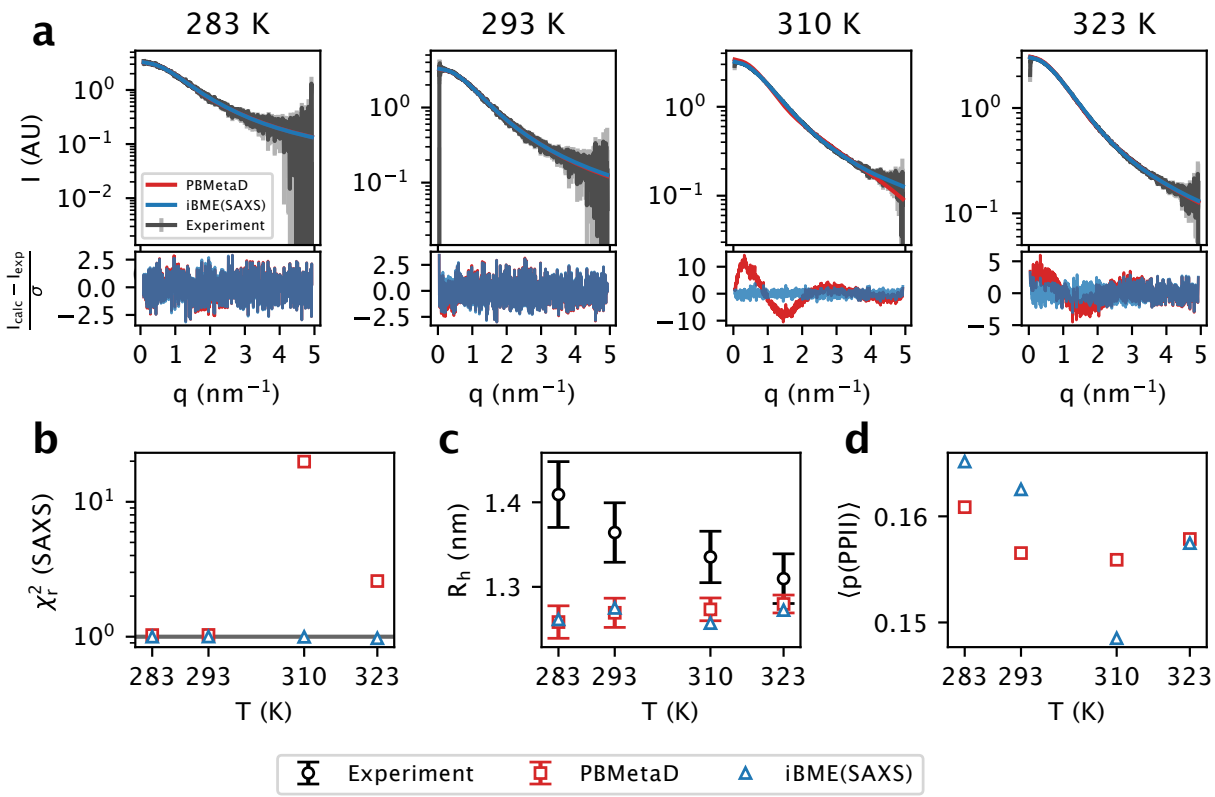

Figure 2: Refinement of the simulations against SAXS data. (a) SAXS profiles of prior (red) and reweighted (blue) ensembles compared with the experiment (grey). (b) χ^2_r calculated from the profiles in panel a. (c) Ensemble-averaged R_h calculated with either the prior or SAXS-refined weights. (d) Ensemble-averaged per-residue PPII structure population calculated with either the prior or SAXS-refined weights.

to compact conformations and lacks expanded conformations whose population could be increased by reweighting (Fig. S8). Indeed, by reweighting the simulation at 323 K against the experimental R_h recorded at the lower temperatures, in all cases (but in particular for the lowest temperature) we achieve a better agreement with the experiments (Fig. S9), because there are more expanded conformations that can be up-weighted to fit the experiments (Fig. S8). Remarkably, when improving the agreement with the experimental R_h, we also observe an overall increase in the amount of PPII structure, that follows the same trend suggested by CD data (Fig. 3b). Observing the relationship between conformational R_h and number of residues in PPII conformation, we see that, while conformations with only few residues involved in a PPII structure can be either very compact or very expanded,

conformations with many residues in a PPII structure are all expanded (Fig. 3c). Indeed, by reweighting the ensembles against the experimental R_h values (that likely increases the weight of conformations with larger R_h), we also observe that conformations with five to ten residues involved in a PPII structure are up-weighted, while conformations with no or little of this secondary structure element are down-weighted (Fig. 3d). The down-side arising from reweighting the simulations of Hst5 against R_h is that the agreement with SAXS data decreases (Fig. 3e), that also translates in the ensemble-averaged R_g becoming larger (Fig. S6).

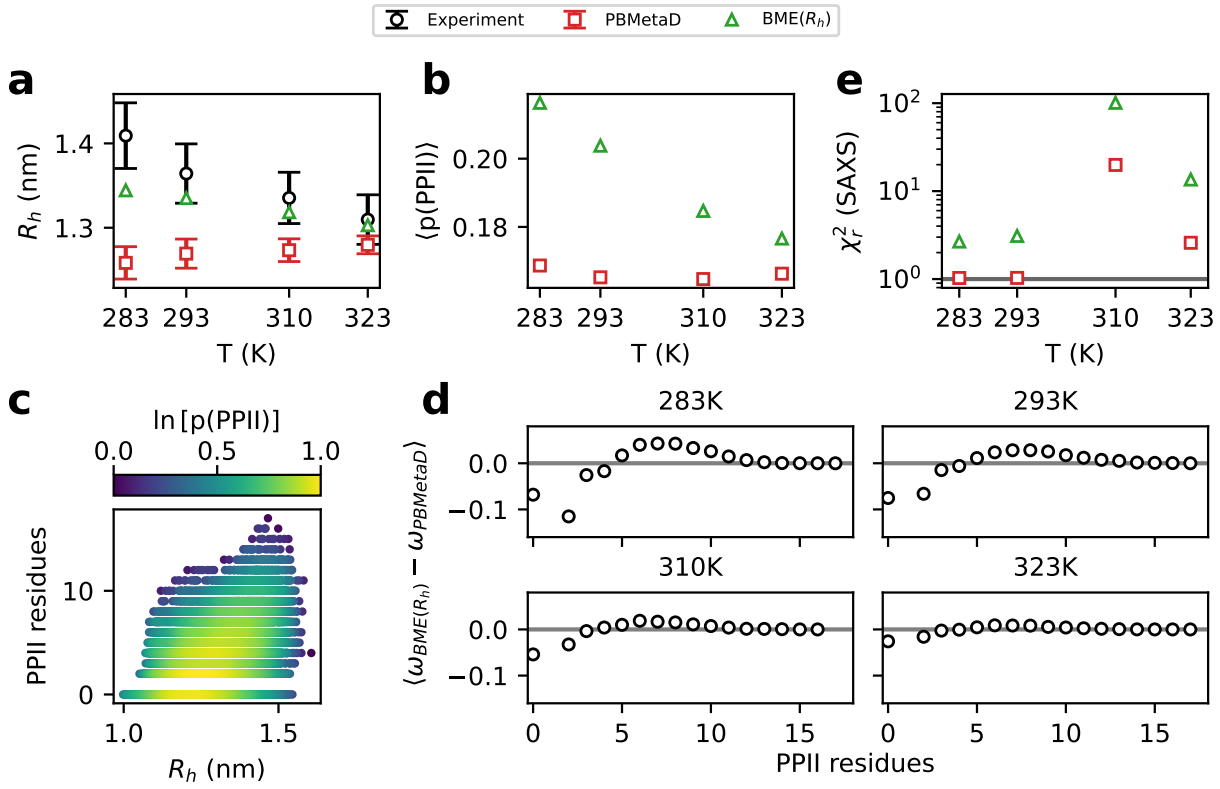

Figure 3: Refinement of the simulations against R_h from PFG NMR measurements. We show the (a) R_h, (b) average PPII structure propensity and (e) χ_r^2 to SAXS data for the expectation value from the experiment (black, when applicable), the ensemble average from the simulations (red) and the simulations reweighted against the experimental R_h (green). (c) Density plot showing the number of residues involved in a PPII structure as a function of the conformational R_h, together with their relative frequency. (d) Average difference between R_h-refined weights and prior weights, for frames with a certain number of residues involved in forming a PPII structure.

Simultaneous refinement of the simulations against experimental R_{h} and SAXS

As discussed above, SAXS and R_{h} can probe different aspects of the compaction of protein conformations; therefore refining against both experiments might in principle provide additional information compared to analysing them individually. We therefore refined the ensembles simultaneously against SAXS and R_{h} (see Methods). We note here one complication when refining simultaneously against different sources of experiments. Specifically, setting the appropriate balance between the force field and the experiments depends (among other things) on both the experimental errors and the errors associated with how we calculate observables from experiments.[52,54,70] Thus, when setting the balance between SAXS and the PFG NMR data we are assuming that the relative balance between these experiments is correct as the reweighting approach only tunes the balance between the simulations and the experiments via a single parameter.

With these caveats in mind, we find that the results of the combined reweighting (Fig. 4) are similar to that obtained when reweighting against SAXS alone (Fig. 2). The main reason for this is that in this combined reweighting the SAXS experiments overall get a substantially greater weight due to a combination of the larger number of data points and the associated errors. Specifically, the SAXS experiments are represented by a large number of intensity measurements while the experimental R_{h} is represented by a single data point (that itself is derived from a small number of points in the PFG NMR experiments). Therefore, if the reweighting cannot satisfy both experiments simultaneously, it will tend to optimize mostly on the experiment with the largest number of data points, that is the SAXS experiments, unless the other experiments have much smaller errors. To demonstrate this effect, we changed the balance of the reweighting by computationally expanding the number of data points for the R_{h} to the same number of data points available in the SAXS profile. When reweighting against this alternative dataset we find a solution that shows intermediate level agreement to both sets of experiments (Fig. S10).

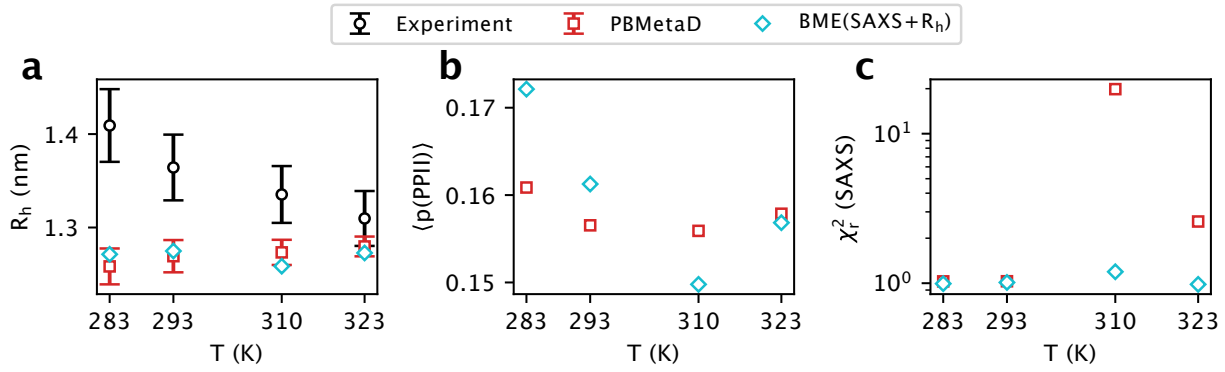

Figure 4: Refinement of the simulations against SAXS and R_h simultaneously. We show the (a) χ^2_r to SAXS data (b) R_h and (c) average PPII structure propensity from the experiment (black, when applicable), the ensemble average from the simulations (red) and the simulations reweighted against both SAXS and R_h (cyan).

Conclusions

Integrative modelling makes it possible to provide atomistic models of IDPs by combining information from experiments probing ensemble-averaged structural features with conformational models generated for example by MD simulations. Taken alone, many experiments only provide information about a specific structural feature, for example the average chain dimension, while being mostly insensitive to other properties of the conformational ensemble. Therefore, integrating simultaneously multiple sources of independent experimental information can provide a more realistic determination of the conformational ensemble of an IDP.[6,38,41,49,50,71]

We have here focused on Hst5 as a case study, in an attempt to explain the previously observed trend showing the R_h from PFG NMR decreasing with temperature, together with a drop of the PPII structure signal from CD spectra[23] and a relatively temperature-independent R_g from SAXS. While Hst5 is relatively simple in terms of its conformational preferences, its small size makes it useful because it is easier to sample by means of all-atom molecular simulations. This in turn makes it possible to study small differences across different conditions as well as potential coupling between local and global conformational preferences. We find that the simulations alone were mostly in good agreement with the

SAXS data, but did not appear to reproduce the R_h from PFG NMR and the loss of PPII structure with increasing temperature. We therefore applied a reweighting approach to refine the ensembles to achieve a better agreement with the experiments. While we could refine against the SAXS and NMR data separately, we were not able to achieve simultaneous agreement with the SAXS and R_h data within experimental errors. Refining against the CD data would have been interesting, but is hampered by lack of methods to calculate accurate CD spectra for disordered proteins; however, we find that refining against the R_h provided a picture in qualitative agreement with the previous interpretation of the CD data.

In light of our observations, one conclusion is that while the force field used provides an overall good agreement with the SAXS data, it does not appear to capture the temperature-dependence shown by the experimental NMR and CD data. This force field, as other force fields shown to be relatively accurate for disordered proteins, has been tested to reproduce R_g extrapolated from SAXS data, and indeed it does a good job in reproducing that experimental information. When it comes to R_h, it seems that the force field overall underestimates the R_h, with the exception of that at 323 K. We note, however, that the observed disagreement between the simulations and the R_h values from PFG NMR may also be due to inaccuracies in the forward model for R_h[62,63] as well as remaining uncertainty about how to determine R_h experimentally.[63] The possibility to calculate accurately CD spectra from conformational ensembles would also aid in understanding the reliability of the forward model for the R_h. Given the observed correlation between R_h and PPII structure content, a quantitative comparison to CD spectra could provide direct information about the possibility that the force field underestimates the PPII structure content of Hst5, together with its R_h. In absence of such a model, we decided to rely on a qualitative analysis of the secondary structures. Based on the backbone dihedral angles, we can calculate the per-residue secondary structure for each frame of our simulations. Doing so, we did not find any temperature dependence on the amount of PPII structure in our simulations, but we could qualitatively reproduce the picture from CD spectra when reweighting the ensembles against the R_h from experiments.

Furthermore, minor differences in the experimental conditions used in the SAXS and PFG NMR experiments[23] might potentially lead to different behaviours of the ensembles in solution and hence some of the observed discrepancies. The Hst5 sample used for SAXS was in a buffer at pH 7.5 and ionic strength of about 150 mM, while the sample used for PFG NMR was at pH 7 and ionic strength of about 40 mM.[23] Nonetheless, differences in ionic strength can affect the extent of inter-chain interactions.[45] Furthermore, given that 7 out of the 24 residues of Hst5 are histidines, the difference in pH between the SAXS and NMR buffers can affect the protonation state of these residues and consequently affect the ensemble of conformations in solution.[44] To examine this, we performed coarse-grained simulations with the CALVADOS2 model[72,73] using the IDRLab implementation on Google Colab.[74] Although approximate, this model has previously been shown to capture some aspects of the ionic-strength dependency of chain compaction.[73] The CALVADOS2 simulations of Hst5, however, suggest at most a small effect of pH and ionic strength (Table S1), though a more detailed analysis would ideally require experimental measurements or simulations with a better description of the effects of ions and mixed protonation states.[75]

In summary, we have here shown an example of the importance of using multiple and diverse experiments to obtain detailed insights into the conformational ensembles of IDPs, but also highlight some of the problems associated with analysing a diverse set of data. Even with such experiments in hand, we have also shown that it is important to have reliable models to calculate experimental observables from computational models. The development of these models is nonetheless impaired by the fact that it is far from trivial to generate accurate conformational ensembles of IDPs that can be used to train these models.[76] Despite these caveats, we have used multiple experiments to shed further light on the structural basis of the temperature-dependent behaviour of Hst5. By generating conformational ensembles that reproduce the R_{h} from PFG NMR measurements, we support the hypothesis that there is a link between how temperature affects the amount of PPII structures and compaction.[15,23,77,78] Some issues remain, that are related to whether the force field em-

ployed here underestimates the R_h of Hst5 or not. Future development of forward models, especially for CD, can help in better characterizing the content of PPII structures in IDPs. Furthermore, the study of more IDPs (ideally bigger, although much more difficult to sample efficiently by all-atom MD simulations) that show similar features would be useful to examine whether the behaviour observed for Hst5 occurs in other IDPs.

Methods

Molecular Dynamics Simulations

We performed MD simulations of Hst5 in solution with GROMACS 2018.6.[79] We used the Amber99SB-disp force field[11] with the TIP4P-D water model.[80] The histidine residues in the peptide were modelled as their Nϵ tautomers. We solvated the peptide in a dodecahedric box (with image distance equal to 10 nm) adding Na^+ and Cl^- ions to neutralize the system and reach an ionic strength of 150 mM (same as used in the experiments). We ran a two-steps energy minimization, using first the steepest-descent algorithm with energy tolerance set at 200 kJ mol^{-1} nm^{-1} and than the conjugate-gradient algorithm with energy tolerance set at 100 kJ mol^{-1} nm^{-1}. Then, we equilibrated the system at the target temperature and pressure (1 bar) by running first a 200 ps MD simulation in NVT with position restraints on the protein atoms with the v-rescale thermostat,[81] then a 1 ns MD simulation in NPT adding the Berendsen barostat.[82] For production simulations, we used the Parrinello-Rahman barostat.[83] We used a 2 fs time step in all simulations and constrained all bonds involving hydrogen atoms with the LINCS algorithm.[84]

To accelerate the sampling of the conformational space of Hst5, we employed Parallel Bias Metadynamics (PBMetaD)[56] as enhanced sampling scheme. In PBMetaD a set of user-defined one-dimensional collective variables (CVs) is subject to history-dependent bias potentials that discourage the system from visiting regions of the CV space already explored. Biasing multiple one-dimensional CVs removes the limitation on the number of CVs that

affects the classical metadynamics,[48] where a single multidimensional space defined by the CVs is biased. Additionally, as in well-tempered metadynamics,[58] the extent of biasing decreases when the same region of a CV space is visited multiple times, so to have a smoother convergence than with classical metadynamics, and the free-energy surface (FES) as a function of the CVs can be estimated within a well-defined error.[85] In our PBMetaD setup we used 47 CVs, consisting in all 23 ϕ and all 23 ψ dihedral angles and the radius of gyration of the C_α atoms (R_{g,C_α}). We employ R_{g,C_α} as CV to bias because it is cheaper to compute on-the-fly, but in all further analyses we calculate the R_g for all the atoms. Gaussian bias was deposited every 400 ps. We set the initial height of the Gaussian bias to 0.3 kJ mol^{-1}. This height decreased during the simulations by means of the well-tempered feature, according to a bias factor of 24. We used an adaptive algorithm to define the width of the Gaussian bias potentials,[86] where, at each bias deposition, the width is based on the space covered by the CV in the last 2 ps. To enhance the sampling efficiency even further we used a double walker approach,[57] where two simulations run in parallel sharing the deposited bias. For the CV representing the R_{g,C_α}, we set an interval in the CV space between 0.6 and 2.5 nm outside which the biasing force of the PBMetaD does not act.[87] We did so because we observed a 'FES wall climbing effect': since the FES as a function of the R_{g,C_α} is shaped as a single basin, as bias accumulates, new areas in the CV space, corresponding to more compact and more extended high-energy conformations, gets available (Fig. S11). Since these areas are newly explored, bias is deposited with the initial height, while for the rest of the CV space the extent of biasing is already very low. The advantage of the well-tempered feature in PBMetaD is that when all the accessible conformations in CV space are explored, the biasing potential decreases at the point of becoming quasi-static and the simulation converges to a well definite FES. But if new areas of the CV space become continuously accessible and a large bias is deposited, the convergence is impaired. Setting the interval in the FES of the R_{g,C_α}, prevents the system from climbing the FES walls corresponding to high-energy conformations that are not relevant. Despite the use of the interval in the FES of the R_g,

we still observe few frames depositing a large amount of bias outside the interval (Fig. S4). This, however, is not a problem since our calculations are not based on those few frames that explore regions of the FES outside the interval range. To enable PBMetaD in GROMACS we used PLUMED 2.5.[88]

At the end of a simulation, we recovered the total bias (B) for each conformation x explored and used it to calculate the Boltzmann weights (ω) associated to each conformation according to $\omega_x = \exp(\frac{B_s}{k_b T})$, where k_b is the Boltzmann constant and T is the temperature. The FES for a CV is calculated from the weighted histogram (H) of the CV as $FES(CV) = -k_b T \ln H(CV)$. The error on the FES and ensemble averaged values is calculated by block averaging.[85,89] The package used for these calculations is available at `https://github.com/fpesceKU/BLOCKING`.

Calculation of experimental observables

We compare the ensembles from our simulations to SAXS, PFG NMR and CD data. We calculate SAXS data with Pepsi-SAXS,[90] following the protocol we previously described:[69] the intensity of the forward scattering and constant background are fit to the experimental SAXS profile by least square fitting; the values for the density of the hydration layer and the excluded volume are scanned on a two-dimensional grid and the combination of those that best fit the data upon reweighting is selected (Table S2).

We calculate the R_h using the HullRadSAS software[64] for reasons recently described.[62,63] We calculate the per-residue secondary structure with DSSP-PPII[91] where—if two or more consecutive residues that are not assigned to any secondary structure by the DSSP algorithm[92] have backbone dihedrals correspondent to those typical of PPII structure—those residues are assigned as PPII structure. For each residue we then report the propensity for PPII structure formation as a weighted sum of the frames where a residue forms PPII structures, using either weights from PBMetaD or BME (see below).

Experimental data

All experimental data used in this study were previously measured and published by Jephthah et al.[23] We corrected the published R_h values based on a recent estimate of the reference R_h value for 1,4-dioxane (2.27 Å ± 0.04 Å[63] compared to the previously used 2.12 Å[65]). Furthermore, we rescaled the error bars for the SAXS intensities by a factor estimated through the Bayesian indirect Fourier transform.[93]

Ensemble reweighting

We use the Bayesian/Maximum entropy (BME) approach[53] to reweight simulations in order to improve the agreement with experimental data. This is achieved by minimizing the functional $\mathcal{L}$:[52,54]

$$\mathcal{L}(\omega_1 \cdots \omega_n) = \frac{m}{2}\chi_\mathrm{r}^2(\omega_1 \cdots \omega_n) - \theta S_\mathrm{rel}(\omega_1 \cdots \omega_n) \tag{1}$$

where m is the number of experimental data points, $(\omega_1 \cdots \omega_n)$ are the weights for each simulation frame, χ_r^2 measures the agreement between the calculated and experimental observable. The relative entropy, $S_\mathrm{rel} = -\sum_{i=1}^{N} \omega_i \log\left(\frac{\omega_i}{\omega_i^0}\right)$, indicates how much the refined weights (*i.e.* the posterior distribution, ω) diverge from the initial weights (*i.e.* the prior distribution, ω^0), θ is a free-parameter that balances the extent of minimizing the χ_r^2 vs. maximizing S_rel. By taking the exponential of S_rel ($\phi_\mathrm{eff} = \exp(S_\mathrm{rel})$), it is possible to gain insights into the extent of the reweighting. Generally, a low ϕ_eff is not desirable, as it means that a low number of frames from the simulation contributes to the calculated weighted averages. Therefore different values for θ are explored in the reweighting procedure, and that corresponding to the χ_r^2 reaching plateau with the highest ϕ_eff is chosen to generate the reweighted ensemble.[53] This helps to ensure a posterior distribution with a good agreement with the experimental data while retaining as much information as possible from the prior distribution. The prior distribution, ω^0 is estimated as the weights from PBMetaD.

To reweight simulations against R_h values from PFG NMR we use the standard im-

plementation of BME. For SAXS data we use the iterative BME (iBME).[69] iBME couples ensemble reweighting with least square fitting of scale factor (s) and constant background (cst) to the calculated SAXS profile. Therefore here the χ^2_{r} in the $\mathcal{L}$ functional takes the following shape:

$$\chi^2_{\mathrm{r}}(\omega_1 \cdots \omega_n, s, cst) = \frac{1}{m} \sum_{q}^{m} \frac{[(s \sum_j^n \omega_j I_q(x_j) + cst) - I_q^{\mathrm{EXP}}]^2}{\sigma^2_{\mathrm{BIFT},q}} \tag{2}$$

where $I_q(x_j)$ is the calculated SAXS intensity at scattering angle q for the frame x_j, I_q^{EXP} is the experimental SAXS intensity at scattering angle q. $\sigma_{\mathrm{BIFT},q}$ is the rescaled error[93] of the experimental intensity at scattering angle q.

In reweighting simultaneously against R_{h} and SAXS we also use the standard BME, using calculated SAXS profiles with scale factor and offset optimized from previous iBME runs.

Data and code availability

MD simulations are available at https://doi.org/10.17894/UCPH.4D0D9A09-B40B-4A3A-93EF-D6C249F49A90. PLUMED inputs to enable PBMetaD are available on PLUMED-NEST[94] at https://www.plumed-nest.org/eggs/23/008/. Scripts used to analyze the simulations are available at `https://github.com/KULL-Centre/_2023_Pesce_Hst5`.

Acknowledgement

We thank Marie Skepö and Birthe B. Kragelund for sharing the SAXS, NMR and CD data and for discussions on disordered proteins, and Simone Orioli for discussions about the simulations. This research was funded by the Lundbeck Foundation BRAINSTRUC initiative in structural biology (R155-2015-2666, lundbeckfonden.com). We acknowledge access to computational resources from the ROBUST Resource for Biomolecular Simulations (supported by the Novo Nordisk Foundation grant no. NF18OC0032608) and from the

Biocomputing Core Facility at the Department of Biology, University of Copenhagen.

References

(1) Dyson, H. J.; Wright, P. E. Intrinsically unstructured proteins and their functions. *Nature reviews Molecular cell biology* **2005**, *6*, 197–208.

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

(3) Bernadó, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W.; Blackledge, M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proceedings of the National Academy of Sciences* **2005**, *102*, 17002–17007.

(4) Uversky, V. N.; Li, J.; Fink, A. L. Evidence for a partially folded intermediate in α-synuclein fibril formation. *Journal of Biological Chemistry* **2001**, *276*, 10737–10744.

(5) Jarvet, J.; Damberg, P.; Danielsson, J.; Johansson, I.; Eriksson, L.; Gräslund, A. A left-handed 3_1 helical conformation in the Alzheimer Aβ (12–28) peptide. *Febs Letters* **2003**, *555*, 371–374.

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

(7) Schramm, A.; Bignon, C.; Brocca, S.; Grandori, R.; Santambrogio, C.; Longhi, S. An arsenal of methods for the experimental characterization of intrinsically disordered proteins–How to choose and combine them? *Archives of biochemistry and biophysics* **2019**, *676*, 108055.

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

(9) Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; De Groot, B. L.; 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*, 5513–5524.

(10) Henriques, J.; Cragnell, C.; Skepo, M. Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment. *Journal of chemical theory and computation* **2015**, *11*, 3420–3431.

(11) Robustelli, P.; Piana, S.; 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*, E4758–E4766.

(12) Wang, W. Recent advances in atomic molecular dynamics simulation of intrinsically disordered proteins. *Physical Chemistry Chemical Physics* **2021**, *23*, 777–784.

(13) Uversky, V. N. Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *The protein journal* **2009**, *28*, 305–325.

(14) Nettels, D.; Müller-Späth, S.; Küster, F.; Hofmann, H.; Haenni, D.; Rüegger, S.; Reymond, L.; Hoffmann, A.; Kubelka, J.; Heinz, B., et al. Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proceedings of the National Academy of Sciences* **2009**, *106*, 20740–20745.

(15) Kjaergaard, M.; Nørholm, A.-B.; Hendus-Altenburger, R.; Pedersen, S. F.; Poulsen, F. M.; Kragelund, B. B. Temperature-dependent structural changes in in-

trinsically disordered proteins: Formation of α–helices or loss of polyproline II? *Protein Science* **2010**, *19*, 1555–1564.

(16) Best, R. B.; Mittal, J. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *The journal of physical chemistry B* **2010**, *114*, 14916–14923.

(17) Ciasca, G.; Campi, G.; Battisti, A.; Rea, G.; Rodio, M.; Papi, M.; Pernot, P.; Tenenbaum, A.; Bianconi, A. Continuous thermal collapse of the intrinsically disordered protein tau is driven by its entropic flexible domain. *Langmuir* **2012**, *28*, 13405–13410.

(18) Wuttke, R.; Hofmann, H.; Nettels, D.; Borgia, M. B.; Mittal, J.; Best, R. B.; Schuler, B. Temperature-dependent solvation modulates the dimensions of disordered proteins. *Proceedings of the National Academy of Sciences* **2014**, *111*, 5213–5218.

(19) Zerze, G. H.; Best, R. B.; Mittal, J. Sequence-and temperature-dependent properties of unfolded and disordered proteins from atomistic simulations. *The Journal of Physical Chemistry B* **2015**, *119*, 14622–14630.

(20) Battisti, A.; Ciasca, G.; Grottesi, A.; Tenenbaum, A. Thermal compaction of the intrinsically disordered protein tau: entropic, structural, and hydrophobic factors. *Physical Chemistry Chemical Physics* **2017**, *19*, 8435–8446.

(21) Graziano, G. Comment on "Thermal compaction of the intrinsically disordered protein tau: entropic, structural, and hydrophobic factors" by A. Battisti, G. Ciasca, A. Grottesi and A. Tenenbaum, Phys. Chem. Chem. Phys., 2017, 19, 8435. *Physical Chemistry Chemical Physics* **2018**, *20*, 690–693.

(22) Hicks, A.; Zhou, H.-X. Temperature-induced collapse of a disordered peptide observed by three sampling methods in molecular dynamics simulations. *The Journal of Chemical Physics* **2018**, *149*, 072313.

(23) Jephthah, S.; Staby, L.; Kragelund, B.; Skepo, M. Temperature dependence of intrinsically disordered proteins in simulations: What are we missing? *Journal of chemical theory and computation* **2019**, *15*, 2672–2683.

(24) Holehouse, A. S.; Pappu, R. V. Collapse transitions of proteins and the interplay among backbone, sidechain, and solvent interactions. *Annual Review of biophysics* **2018**, *47*, 19–39.

(25) Best, R. B. Emerging consensus on the collapse of unfolded and intrinsically disordered proteins in water. *Current opinion in structural biology* **2020**, *60*, 27–38.

(26) Cubuk, J.; Stuchell-Brereton, M. D.; Soranno, A. The biophysics of disordered proteins from the point of view of single-molecule fluorescence spectroscopy. *Essays in Biochemistry* **2022**, *66*, 875–890.

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

(28) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Mittal, J. Temperature-controlled liquid–liquid phase separation of disordered proteins. *ACS central science* **2019**, *5*, 821–830.

(29) Mompeán, M.; Oroz, J.; Laurents, D. V. Do polyproline II helix associations modulate biomolecular condensates? *FEBS Open bio* **2021**, *11*, 2390–2399.

(30) Gast, K.; Zirwer, D.; Damaschun, G. Are there temperature-dependent structural transitions in the" intrinsically unstructured" protein prothymosin α? *European Biophysics Journal* **2003**, *31*, 586–594.

(31) Dawson, R.; Müller, L.; Dehner, A.; Klein, C.; Kessler, H.; Buchner, J. The N-terminal domain of p53 is natively unfolded. *Journal of molecular biology* **2003**, *332*, 1131–1141.

(32) Malm, J.; Jonsson, M.; Frohm, B.; Linse, S. Structural properties of semenogelin I. *The FEBS journal* **2007**, *274*, 4503–4510.

(33) Sánchez-Puig, N.; Veprintsev, D. B.; Fersht, A. R. Human full-length Securin is a natively unfolded protein. *Protein Science* **2005**, *14*, 1410–1418.

(34) Jeganathan, S.; Von Bergen, M.; Mandelkow, E.-M.; Mandelkow, E. The natively unfolded character of tau and its aggregation to Alzheimer-like paired helical filaments. *Biochemistry* **2008**, *47*, 10526–10539.

(35) Gast, K.; Zirwer, D.; Damaschun, H.; Hahn, U.; Müller-Frohne, M.; Wirth, M.; Damaschun, G. Ribonuclease T1 has different dimensions in the thermally and chemically denatured states: a dynamic light scattering study. *FEBS letters* **1997**, *403*, 245–248.

(36) Mertens, H. D.; Svergun, D. I. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of Structural Biology* **2010**, *172*, 128–141, New Trends in Protein Expression.

(37) Leeb, S.; Danielsson, J. In *Intrinsically Disordered Proteins: Methods and Protocols*; Kragelund, B. B., Skriver, K., Eds.; Springer US: New York, NY, 2020; pp 285–302.

(38) Choy, W.-Y.; Mulder, F. A.; Crowhurst, K. A.; Muhandiram, D.; Millett, I. S.; Doniach, S.; Forman-Kay, J. D.; Kay, L. E. Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. *Journal of molecular biology* **2002**, *316*, 101–112.

(39) Nygaard, M.; Kragelund, B. B.; Papaleo, E.; Lindorff-Larsen, K. An efficient method for estimating the hydrodynamic radius of disordered protein conformations. *Biophysical journal* **2017**, *113*, 550–557.

(40) Burchard, W.; Schmidt, M.; Stockmayer, W. Information on polydispersity and branching from combined quasi-elastic and intergrated scattering. *Macromolecules* **1980**, *13*, 1265–1272.

(41) Ahmed, M. C.; Crehuet, R.; Lindorff-Larsen, K. *Intrinsically Disordered Proteins*; Springer, 2020; pp 429–445.

(42) Brewer, D.; Hunter, H.; Lajoie, G. NMR studies of the antimicrobial salivary peptides histatin 3 and histatin 5 in aqueous and nonaqueous solutions. *Biochemistry and cell biology* **1998**, *76*, 247–256.

(43) Raj, P. A.; Marcus, E.; Sukumaran, D. K. Structure of human salivary histatin 5 in aqueous and nonaqueous solutions. *Biopolymers: Original Research on Biomolecules* **1998**, *45*, 51–67.

(44) Kurut, A.; Henriques, J.; Forsman, J.; Skepö, M.; Lund, M. Role of histidine for charge regulation of unstructured peptides at interfaces and in bulk. *Proteins: Structure, Function, and Bioinformatics* **2014**, *82*, 657–667.

(45) Cragnell, C.; Durand, D.; Cabane, B.; Skepö, M. Coarse-grained modeling of the intrinsically disordered protein Histatin 5 in solution: Monte Carlo simulations in combination with SAXS. *Proteins: Structure, Function, and Bioinformatics* **2016**, *84*, 777–791.

(46) Jephthah, S.; Henriques, J.; Cragnell, C.; Puri, S.; Edgerton, M.; Skepö, M. Structural characterization of histatin 5–spermidine conjugates: A combined experimental and theoretical study. *Journal of chemical information and modeling* **2017**, *57*, 1330–1341.

(47) McCaslin, T. G.; Pagba, C. V.; Yohannan, J.; Barry, B. A. Specific metallo-protein interactions and antimicrobial activity in Histatin-5, an intrinsically disordered salivary peptide. *Scientific reports* **2019**, *9*, 1–14.

(48) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* **2002**, *99*, 12562–12566.

(49) Borgia, A.; Zheng, W.; Buholzer, K.; Borgia, M. B.; Schüler, A.; Hofmann, H.; Soranno, A.; Nettels, D.; Gast, K.; Grishaev, A., et al. Consistent view of polypeptide

chain expansion in chemical denaturants from multiple experimental methods. *Journal of the American Chemical Society* **2016**, *138*, 11714–11726.

(50) Fuertes, G.; Banterle, N.; Ruff, K. M.; Chowdhury, A.; Mercadante, D.; Koehler, C.; Kachala, M.; Estrada Girona, G.; Milles, S.; Mishra, A., et al. 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*, E6342–E6351.

(51) Bengtsen, T.; Holm, V. L.; Kjølbye, L. R.; Midtgaard, S. R.; Johansen, N. T.; Tesei, G.; Bottaro, S.; Schiøtt, B.; Arleth, L.; Lindorff-Larsen, K. Structure and dynamics of a nanodisc by integrating NMR, SAXS and SANS experiments with molecular dynamics simulations. *Elife* **2020**, *9*, e56518.

(52) Hummer, G.; Köfinger, J. Bayesian ensemble refinement by replica simulations and reweighting. *The Journal of Chemical Physics* **2015**, *143*, 243150.

(53) Bottaro, S.; Bengtsen, T.; Lindorff-Larsen, K. In *Structural Bioinformatics: Methods and Protocols*; Gáspári, Z., Ed.; Springer US: New York, NY, 2020; pp 219–240.

(54) Orioli, S.; Larsen, A. H.; Bottaro, S.; Lindorff-Larsen, K. In *Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly*; Strodel, B., Barz, B., Eds.; Progress in Molecular Biology and Translational Science; Academic Press, 2020; Vol. 170; pp 123–176.

(55) Jephthah, S.; Pesce, F.; Lindorff-Larsen, K.; Skepo, M. Force field effects in simulations of flexible peptides with varying Polyproline II propensity. *Journal of chemical theory and computation* **2021**, *17*, 6634–6646.

(56) Pfaendtner, J.; Bonomi, M. Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics. *Journal of Chemical Theory and Computation* **2015**, *11*, 5062–5067, PMID: 26574304.

(57) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *The Journal of Physical Chemistry B* **2006**, *110*, 3533–3539, PMID: 16494409.

(58) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(59) Manalastas-Cantos, K.; Konarev, P. V.; Hajizadeh, N. R.; Kikhney, A. G.; Petoukhov, M. V.; Molodenskiy, D. S.; Panjkovich, A.; Mertens, H. D.; Gruzinov, A.; Borges, C., et al. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *Journal of Applied Crystallography* **2021**, *54*, 343–355.

(60) Svergun, D.; Richard, S.; Koch, M.; Sayers, Z.; Kuprin, S.; Zaccai, G. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proceedings of the National Academy of Sciences* **1998**, *95*, 2267–2272.

(61) Henriques, J.; Arleth, L.; Lindorff-Larsen, K.; Skepö, M. On the calculation of SAXS profiles of folded and intrinsically disordered proteins from computer simulations. *Journal of molecular biology* **2018**, *430*, 2521–2539.

(62) Pesce, F.; Newcombe, E. A.; Seiffert, P.; Tranchant, E. E.; Olsen, J. G.; Grace, C. R.; Kragelund, B. B.; Lindorff-Larsen, K. Assessment of models for calculating the hydrodynamic radius of intrinsically disordered proteins. *Biophysical Journal* **2023**, *122*, 310–321.

(63) Tranchant, E. E.; Pesce, F.; Jacobsen, N. L.; Fernandes, C. B.; Kragelund, B. B.; Lindorff-Larsen, K. Revisiting the use of dioxane as a reference compound for determination of the hydrodynamic radius of proteins by pulsed field gradient NMR spectroscopy. *bioRxiv* **2023**, 2023–06.

(64) Fleming, P. J.; Correia, J. J.; Fleming, K. G. Revisiting macromolecular hydration with HullRadSAS. *European Biophysics Journal* **2023**, 1–10.

(65) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques. *Biochemistry* **1999**, *38*, 16424–16431, PMID: 10600103.

(66) Yamamoto, E.; Akimoto, T.; Mitsutake, A.; Metzler, R. Universal relation between instantaneous diffusivity and radius of gyration of proteins in aqueous solution. *Physical review letters* **2021**, *126*, 128101.

(67) Fagerberg, E.; Månsson, L. K.; Lenton, S.; Skepö, M. The effects of chain length on the structural properties of intrinsically disordered proteins in concentrated solutions. *The Journal of Physical Chemistry B* **2020**, *124*, 11843–11853.

(68) Gopal, S. M.; Wingbermühle, S.; Schnatwinkel, J.; Juber, S.; Herrmann, C.; Schäfer, L. V. Conformational preferences of an intrinsically disordered protein domain: A case study for modern force fields. *The Journal of Physical Chemistry B* **2020**, *125*, 24–35.

(69) Pesce, F.; Lindorff-Larsen, K. Refining conformational ensembles of flexible proteins against small-angle x-ray scattering data. *Biophysical Journal* **2021**, *120*, 5124–5135.

(70) Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Combining experiments and simulations using the maximum entropy principle. *PLoS computational biology* **2014**, *10*, e1003406.

(71) Gomes, G.-N. W.; Krzeminski, M.; Namini, A.; Martin, E. W.; Mittag, T.; Head-Gordon, T.; Forman-Kay, J. D.; Gradinaru, C. C. Conformational ensembles of an intrinsically disordered protein consistent with NMR, SAXS, and single-molecule FRET. *Journal of the American Chemical Society* **2020**, *142*, 15697–15710.

(72) Tesei, G.; Schulze, T. K.; Crehuet, R.; Lindorff-Larsen, K. Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of

single-chain properties. *Proceedings of the National Academy of Sciences* **2021**, *118*, e2111696118.

(73) Tesei, G.; Lindorff-Larsen, K. Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range. *Open Research Europe* **2022**, *2*, 94.

(74) Tesei, G.; Trolle, A. I.; Jonsson, N.; Betz, J.; Pesce, F.; Johansson, K. E.; Lindorff-Larsen, K. Conformational ensembles of the human intrinsically disordered proteome: Bridging chain compaction with function and sequence conservation. *bioRxiv* **2023**, 2023–05.

(75) Fossat, M. J.; Posey, A. E.; Pappu, R. V. Quantifying charge state heterogeneity for proteins with multiple ionizable residues. *Biophysical Journal* **2021**, *120*, 5438–5453.

(76) Lindorff-Larsen, K.; Kragelund, B. B. On the potential of machine learning to examine the relationship between sequence, structure, dynamics and function of intrinsically disordered proteins. *Journal of Molecular Biology* **2021**, *433*, 167196.

(77) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. Polyproline II structure in a sequence of seven alanine residues. *Proceedings of the National Academy of Sciences* **2002**, *99*, 9190–9195.

(78) Shi, Z.; Chen, K.; Liu, Z.; Kallenbach, N. R. Conformation of the backbone in unfolded proteins. *Chemical reviews* **2006**, *106*, 1877–1897.

(79) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(80) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions

Strongly Influence Simulated Structural Properties of Disordered Protein States. *The Journal of Physical Chemistry B* **2015**, *119*, 5113–5123, PMID: 25764013.

(81) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007**, *126*, 014101.

(82) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.

(83) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, *52*, 7182–7190.

(84) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997**, *18*, 1463–1472.

(85) Bussi, G.; Tribello, G. A. In *Biomolecular Simulations: Methods and Protocols*; Bonomi, M., Camilloni, C., Eds.; Springer New York: New York, NY, 2019; pp 529–578.

(86) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *Journal of Chemical Theory and Computation* **2012**, *8*, 2247–2254, PMID: 26588957.

(87) Baftizadeh, F.; Cossio, P.; Pietrucci, F.; Laio, A. Protein Folding and Ligand-Enzyme Binding from Bias-Exchange Metadynamics Simulations. *Current Physical Chemistry* **2012**, *2*, 79–91.

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

(89) Flyvbjerg, H.; Petersen, H. G. Error estimates on averages of correlated data. *The Journal of Chemical Physics* **1989**, *91*, 461–466.

(90) Grudinin, S.; Garkavenko, M.; Kazennov, A. *Pepsi-SAXS*: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles. *Acta Crystallographica Section D* **2017**, *73*, 449–464.

(91) Mansiaux, Y.; Joseph, A. P.; Gelly, J.-C.; de Brevern, A. G. Assignment of PolyProline II conformation and analysis of sequence–structure relationship. *PloS one* **2011**, *6*, e18401.

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

(93) Larsen, A. H.; Pedersen, M. C. Experimental noise in small-angle scattering can be assessed using the Bayesian indirect Fourier transformation. *Journal of Applied Crystallography* **2021**, *54*, 1281–1289.

(94) Promoting transparency and reproducibility in enhanced molecular simulations. *Nature methods* **2019**, *16*, 670–673.