Towards accurate, force field independent conformational ensembles of intrinsically disordered proteins

Kaushik Borthakur,[1] Thomas R. Sisk,[1] Francesco P. Panei ,[2] Massimiliano Bonomi,[2] and Paul Robustelli*,[1]

[1] *Dartmouth College, Department of Chemistry, Hanover, NH, 03755, USA*

[2] *Institut Pasteur, Université Paris Cité, CNRS UMR 3528, Computational Structural Biology Unit, Paris, France*

* E-mail: paul.j.robustelli@dartmouth.edu

Abstract

Determining accurate atomic resolution conformational ensembles of intrinsically disordered proteins (IDPs) is extremely challenging. Molecular dynamics (MD) computer simulations provide atomically detailed conformational ensembles of IDPs, but their accuracy is highly dependent on the quality of the underlying physical models, or *force fields*, used. Integrative methods that combine experimental data with computational models offer a promising approach to address force field limitations and generate accurate conformational ensembles of IDPs, shedding light on their functional mechanisms. Here, we present a simple and robust maximum entropy reweighting procedure to refine atomic resolution conformational ensembles of IDPs with large experimental datasets consisting of several different types of data. We apply this approach to refine structural ensembles obtained from long timescale MD simulations and generate IDP ensembles with substantially improved agreement with a variety of nuclear magnetic resonance (NMR) and small-angle X-ray scattering (SAXS) measurements. We ask if reweighted IDP ensembles derived from MD simulations run with different force fields

converge to similar conformational distributions when extensive experimental datasets are used for refinement. We find that in favorable cases IDP ensembles derived from different force fields become highly similar after reweighting with experimental data. The maximum entropy reweighting procedure presented here enables the integration of atomic resolution MD simulations with extensive experimental datasets and can facilitate the elucidation of accurate, force field independent conformational ensembles of IDPs.

Introduction

Many proteins that perform important biological functions are completely or partially disordered under physiological conditions.[1,2] These so-called intrinsically disordered proteins (IDPs) lack a well-defined tertiary-structure in solution and instead populate a conformational ensemble of rapidly interconverting structures. Structurally characterizing the heterogeneous conformational ensembles adopted by IDPs can provide mechanistic insight into their physiological interactions and functions.[3,4] IDPs are implicated in many human diseases and are increasingly being pursued as drug targets.[5] Determining accurate conformational ensembles of IDPs and IDPs in complex with small molecules can provide valuable insight for the rational design of IDP inhibitors.[6–11]

Experimentally determining atomic resolution conformational ensembles of IDPs is extremely challenging. Most experimental techniques used to structurally characterize IDPs in solution, such as nuclear magnetic resonance (NMR) spectroscopy and small angle X-ray scattering (SAXS), report on conformational properties averaged over many molecules over long periods of time.[12,13] Such *ensemble-averaged* measurements can be consistent with a large number of conformational distributions. Typical experimental datasets used to characterize conformational ensembles of IDPs are also sparse, meaning that they report on a small subset of structural properties of IDPs. Additionally, many experimental data, such as NMR chemical shifts, are challenging to interpret and predict, as they are sensitive to a

combination of many structural properties.[14–18]

Molecular dynamics (MD) computer simulations are a powerful approach for determining atomic resolution conformational ensembles of IDPs *in silico*. In principle, long timescale all-atom MD simulations of an IDP driven by an accurate physical model, or *force field*, can provide atomically detailed structural descriptions of the conformational states populated in solution along with their equilibrium populations. In practice, however, MD simulations are limited by the accuracy of the force fields used to describe the interactions between atoms in molecules.[19–24] Recent improvements in molecular mechanics force fields and water models have dramatically improved the accuracy of MD simulations of IDPs as assessed by agreement with a large variety of experimental measurements. However, systematic and system-dependent discrepancies between simulations and experiments remain among the best performing force fields.[19–24]

Due to the challenges of determining conformational ensembles of IDPs from experimental or computational methods alone, integrative approaches, where experimental data are used to construct or refine computational models of IDP ensembles, have grown increasingly popular.[25–37] The maximum entropy principle[38,39] is the basis for a number of successful *reweighting*[25–33] and *biasing*[40–44] approaches to determine conformational ensembles of proteins. In the reweighting framework, after generating a conformational ensemble using a technique such as unbiased MD simulations, one seeks to introduce the minimal perturbation to the ensemble required to match a given set of experimental data. This is achieved by assigning a new statistical weight to each conformation of the ensemble, thus promoting or disfavoring regions of the conformational space and ultimately altering the size of the reweighted ensemble (the *effective ensemble size*).

The conformational ensembles of IDPs derived from integrative approaches are highly sensitive to the strengths of the restraints used to enforce agreement with experimental data. These strengths reflect the relative confidence in the experimental data compared to the prior simulation model (ie. the molecular mechanics force field) and should account

for errors and uncertainties in both experimental measurements and computational *forward models* used to predict experimental data from structures.[12] When integrating experimental data from several sources, researchers are often required to make subjective decisions about the relative importance of satisfying different classes of restraints.

In maximum entropy reweighting approaches, *regularization parameters* are often used to balance confidence in the prior simulation model against the satisfaction of each class of experimental restraints.[29–33] Choosing regularization parameters effectively requires researchers to decide what fraction of an initial ensemble they are willing to modify to satisfy a set of experimental data within a specified threshold. If many of the conformations in an initial ensemble are in poor agreement with experimental data, this choice can dramatically reduce the effective ensemble size of the reweighted ensemble. While several Bayesian and heuristic approaches have been proposed to infer optimal values of regularization parameters,[29–33,35,45] these approaches remain sensitive to estimates of the errors of forward models and experimental measurements, and/or subjective decisions about the desired level of agreement with unrestrained cross-validating data. This flexibility complicates the development of automated procedures for calculating structural ensembles of IDPs and objectively comparing the accuracy of IDP ensembles obtained by reweighting ensembles derived from different force fields.

Here, we introduce a maximum entropy reweighting procedure for combining several types of experimental data with a single adjustable parameter: the desired effective ensemble size of the final reweighted ensemble. This approach does not require manually tuning regularization parameters or the relative weights of experimental restraints. We use our approach to address the following question: if one uses all available experimental data to reweight IDP conformational ensembles obtained from long timescale MD simulations run with different state-of-the-art force fields, how similar are the resulting ensembles? In other words, when extensive experimental data is available, can we determine force field independent IDP ensembles? This question is particularly timely as IDP simulations are increasingly

being used to train machine learning approaches based on deep generative models to predict the conformational ensembles of IDPs.[46] These methods potentially offer an efficient alternative to MD for generating conformational ensembles in the same way that methods like AlphaFold3[47] provide accurate structural models of folded proteins. To be most effective, training and validating these approaches will require improved examples of "ground truth" conformational ensembles of IDPs.

To address the questions above, we reweighted conformational ensembles of five well studied IDPs obtained with unbiased long timescale MD simulations run with different force fields.[19] We considered five IDPs ranging in length from 40 to 140 residues: Aβ40,[48] drkN SH3,[49] ACTR,[50] PaaA2,[51] and α-synuclein.[52] We first demonstrate that our reweighting procedure simultaneously improves agreement with several types of NMR data and SAXS data while maintaining reasonable statistical sampling of the most populated conformational states observed in unbiased MD simulations. In all cases studied, we find that the ensembles of IDPs derived from different force fields become more similar to one another upon reweighting. In favorable cases, we observe that reweighted ensembles converge to very similar conformational distributions which may be considered a reasonable approximation of the true solution ensemble. We observe that MD ensembles of longer IDPs and IDPs with highly stable secondary structure elements derived from different force fields can have little-to-no overlap in conformational space before or after reweighting. Our analyses underscore the importance of using accurate force fields and achieving sufficient sampling in MD simulations of IDPs, even if reweighting methods are used to refine the resulting ensembles with experimental data. We anticipate that the reweighting protocol proposed here will provide a valuable tool for integrating MD simulations with extensive experimental datasets to improve the accuracy of atomic resolution conformational ensembles of IDPs and will facilitate the calculation of accurate, force field independent ensembles of IDPs.

Results

In this investigation we performed maximum entropy reweighting of IDP conformational ensembles obtained from long timescale MD simulations run with different state-of-the-art force fields and compared the resulting ensembles. We performed reweighting on 30μs unbiased MD simulations of five IDPs that were previously used to benchmark the ability of modern force fields to describe both IDPs and structured proteins:[19] Aβ40[48] (40 residues), drkN SH3[49] (59 residues), ACTR[50] (69 residues), The ParE2-associated antitoxin (PaaA2)[51] (70 residues), and α-synuclein[52] (140 residues). These IDPs span a range of secondary propensity: Aβ40 and α-synuclein contain little-to-no experimentally detectable residual secondary structure, ACTR and drkN SH3 contain regions of residual helical structure and PaaA2 contains two stable helical elements connected by a flexible linker.

We compared MD simulations run with three different protein force field and water model combinations: a99SB-*disp*[19] with a99SB-*disp* water, Charmm22*[53] with TIP3P water,[54] and Charmm36m[22] with TIP3P water. We henceforth refer to these protein force field and water model combinations as a99SB-*disp*, C22* and C36m, respectively. We selected these force field combinations as they were among the top performing force fields in a previous benchmark study considering both folded and disordered proteins.[19] We used previously reported experimental datasets containing different combinations of experimental NMR and SAXS data to reweight simulations of each protein. Backbone NMR chemical shifts[48–51,55] and residual dipolar couplings (RDCs)[50–52,56,57] were used as restraints for all proteins. Three-bond backbone ^3J scalar coupling constants were used as restraints for Aβ40[58] and α-synuclein[59] ensembles. Paramagnetic relaxation enhancements (PREs) were used as restraints for drkN SH3,[34] ACTR,[50] and α-synuclein[52,60–63] ensembles. SAXS data were used as restraints for drkN SH3,[64] ACTR,[65] PaaA2,[51] and α-synuclein[37] ensembles.

Developing a maximum entropy reweighting procedure with a single free parameter

We provide a detailed description of our reweighting procedure in the Theory section. Here, we summarize the key features of our approach and provide illustrative examples that highlight the rationale behind the choices made. We build upon the formalism of Bussi and coworkers,[26–28] where the weight of the restraint on each experimental data point i is determined by the value of an adjustable regularization parameter σ_i (Eq. 9). A key feature of our reweighting procedure is the use of effective ensemble sizes of reweighted ensembles to select the values of regularization parameters for reweighting. We quantify the effective ensemble size of reweighted ensemble using the Kish ratio[28,66] (Eq. 11). Briefly, the Kish ratio (K) is a measure of the fraction of frames of an original ensemble with significant statistical weights after reweighting. A reweighted ensemble where ∼90% of the frames have statistical weights close to zero will have a Kish ratio of $K \approx 0.10$. Conversely, if an unbiased MD ensemble contains 10,000 frames, a reweighted ensemble with a Kish ratio $K \approx 0.01$ would contain approximately 100 frames with statistical weights substantially larger than zero.

We initially attempted to reweight IDP ensembles using a cross-validation approach to determine the relative weights of restraints for different experimental data types. We performed reweighting using a single value of the regularization parameter σ for all experimental data points of a given type (ie. Cα chemical shifts, Cβ chemical shifts, SAXS intensities) and iteratively decreased the value of σ while monitoring: i) the root mean squared error (RMSE) between calculated and experimental data used as restraints; ii) the RMSE between calculated end experimental cross-validating data not used as restraints; iii) the Kish ratio of the reweighted ensemble. We sought to identify an optimal value of σ for each experimental data type corresponding to the point where further reducing σ caused agreement with unrestrained cross-validating experimental data to deteriorate. We found, however, that many experimental data types were highly correlated and the RMSE of several cross-validating data sets continued to improve as the Kish ratio approached zero (SI Figure 1).

We tested several Bayesian and heuristic approaches to infer the optimal values of regularization parameters for each type of experimental data by monitoring the agreement of restrained and unrestrained cross-validation data using empirical RMSE thresholds and L-curve or "elbow method" analyses.[29,31,45] In many instances, we observed that these approaches produced reweighted ensembles with extremely small effective ensemble sizes, where only hundreds of structures from an initial MD ensemble of nearly 30,000 structures maintained appreciable statistical weights after reweighting ($K < 0.01$). When reweighting ensembles derived from different force fields, we observed that these approaches produced reweighted ensembles with dramatic differences in their effective ensemble size, depending on the initial agreement of unbiased MD ensembles with experimental data. This complicated global comparisons of the accuracy and structural properties of reweighted ensembles derived from different force fields.

As we desire reweighted ensemble sizes that are large enough to provide reasonable statistics for the conformational properties of IDPs, such as the overall topology of backbone conformations and the relative positions of sidechains, we opted to utilize the value of the Kish ratio of reweighted ensembles to determine regularization parameters for each data type. To account for the errors of forward models used to calculate experimental data and the effects of calculating average quantities over a finite-size ensemble, we introduced an additional regularization parameter $\sigma_{i,MD}$ for each experimental data point i used for reweighting (Eq. 10). We calculated $\sigma_{i,MD}$ by applying the Flyvbjerg block analysis method[67] to the forward model predictions obtained from the unbiased MD trajectory. With this addition, we obtain one regularization parameter for each experimental data point $\sigma_i = \sqrt{\sigma_{i,MD}^2 + \sigma_{reg}^2}$, where σ_{reg} is the same for all data of a given data type (ie. $\sigma_{reg-C\alpha}$, $\sigma_{reg-C\beta}$, $\sigma_{reg-SAXS}$).

To determine an optimal minimum Kish ratio threshold for reweighting IDP ensembles, we examined how the conformational properties of reweighted ensembles changed as a function of σ_{reg} for each data type. We sought to identify a minimum Kish ratio threshold where the conformational properties of reweighted ensembles, such as secondary structure content,

radius of gyration (R_g) and populations of intramolecular contacts, remained relatively consistent as σ_{reg} was decreased and further improvements in the agreement between calculated and experimental data were largely achieved by increasing the sparsity of the reweighted ensemble. For each data type, as the weight of experimental restraints was increased (by decreasing σ_{reg}) we generally observed an initial regime where agreement with experimental data was achieved by changing the statistical weights of the structures belonging to the most populated regions of conformational space sampled in unbiased MD simulations. As Kish ratios approached smaller values (ie. $K < 0.10$), improved agreement with experimental data was often achieved by favoring a smaller number of structures from regions of conformational space with unfavorable free energies in unbiased MD simulations.

An illustrative example of this behavior is shown in Figure 1. Here, we use only two experimental backbone NMR chemical shifts of a single residue as restraints to reweight the unbiased a99SB-*disp* MD ensemble of PaaA2. We examine how reweighted distributions of the predicted values of these chemical shifts change as we increase the weight of these two restraints. We observe that as we decrease σ_{reg} from 16.0 to 2.0, we improve agreement with experimental data by increasing the statistical weights of conformations near the most populated free energy basins of the original unbiased MD ensemble. When we further increase the weight of experimental restraints by setting $\sigma_{reg} = 1.0$, the Kish ratio drops to $K = 0.003$, corresponding to an effective reweighted ensemble size of approximately one hundred structures from an initial ensemble of 29,976 structures. Notably, we observe that very few of the conformations that maintain non-negligible statistical weights in the reweighted ensemble are located within the free energy basins of the original unbiased MD ensemble (Figure 1). This dramatic reduction in the effective ensemble size and reshaping of the free energy surface results in an extremely marginal improvement in the agreement between the calculated and experimental values of these chemical shifts relative to the reweighted ensemble obtained with $\sigma_{reg} = 2.0$, which has a Kish ratio $K = 0.11$. This marginal improvement is likely within the noise of chemical shift predictor SPARTA+.[16]

We performed similar analyses for each experimental data type, monitoring how the free energy surfaces of reweighted ensembles and agreement with unrestrained cross-validating data varied with σ_{reg}. Based on these analyses, we elected to select regularization parameters using a Kish ratio threshold of $K = 0.10$. This means we iteratively decrease σ_{reg} for each individual data type until the resulting reweighted ensemble has a Kish ratio $K < 0.10$, and select the smallest value of σ_{reg} where $K \geq 0.10$ as the final σ_{reg} value for that data type. After establishing an initial σ_{reg} for each individual data type, we perform a global reweighting using all experimental data as restraints simultaneously. In order to ensure a suitable ensemble size for the final reweighted ensemble, we rescaled the regularization parameter of each data type by a global scaling factor $\sigma_{reg-Global}$. We iteratively decrease the value of $\sigma_{reg-Global}$ and select the minimum value of $\sigma_{reg-Global}$ that produces a reweighted ensemble with Kish ratio $K \geq 0.10$ as the final value for global reweighting using all experimental data as restraints. The full reweighting protocol is illustrated for the a99SB-*disp* MD ensemble of Aβ40 in Figure 2 and summarized in the Methods section. We note that the choice of an optimal Kish ratio threshold is ultimately a subjective decision that will depend on the initial ensemble size of unbiased ensembles, the desired level of agreement with experimental data and the the desired degree of statistical sampling of properties of interest in reweighted ensembles. We found that varying the Kish ratio threshold between $K = 0.05$ and $K = 0.25$ did not appreciably change the structural properties of reweighted IDP ensembles or any of the conclusions of this study.

Reweighting IDP ensembles improves agreement with extensive experimental data sets from NMR spectroscopy and SAXS

We sought to determine if our proposed reweighting protocol could simultaneously improve agreement with all data types in large experimental NMR and SAXS datasets for IDPs or if improving agreement with some experimental data types requires sacrificing agreement with others. We treat the NMR chemical shifts of each backbone nucleus ($C\alpha, C\beta, C, H\alpha, H, N$),

each type of ^3J scalar coupling constant (ie. $^3J_{HNH\alpha}$ and $^3J_{CC}$) and PREs measured from each nitroxide spin-label as separate experimental data types. The experimental datasets used for reweighting Aβ40, drkN SH3, ACTR, PaaA2 and α-synuclein ensembles therefore contain 7, 10, 11, 8, and 14 experimental data types, respectively. For the a99SB-*disp*, C36m and C22* MD ensembles of each protein we performed: i) reweighting using each experimental data type as the only restraint; ii) a global reweighting using all experimental data as restraints; and iii) a leave-one-out cross-validation reweighting using all but one experimental data type as restraints.

We present a comparison of the results of reweighting calculations performed on the a99SB-*disp* MD ensemble of Aβ40 in Figure 2. We illustrate the process of determining σ_{reg} for each experimental data type of Aβ40 in Figure 2A and the process of determining the global regularization parameter scaling factor $\sigma_{reg-Global}$ in Figure 2B. In both plots, we denote the Kish ratio threshold $K = 0.10$ with a dashed line. In Figure 2B we observe that the RMSE between calculated and experimental data continues to improve for all data types as $\sigma_{reg-Global}$ is decreased beneath the selected $K = 0.10$ Kish ratio threshold. With our approach, we choose to prioritize obtaining a reweighted ensemble with a desired minimum effective ensemble size over maximizing agreement with experimental data.

We compare the RMSE between calculated and unrestrained experimental data for each leave-one-out cross-validation reweighting calculation performed on the a99SB-*disp* Aβ40 MD ensemble in Figure 2C. To compare the relative accuracy of ensembles before and after reweighting we introduce an expression for the normalized RMSE ($\mathrm{RMSE_N}$), where we normalize the value of the RMSE calculated for each experimental data type by the corresponding value in the unbiased a99SB-*disp* MD ensemble ($\mathrm{RMSE_N} = \mathrm{RMSE}/\mathrm{RMSE}_{\mathrm{a99SB}-disp}$). We note that for this Aβ40 ensemble, with the exception of backbone RDCs, the agreement of unrestrained cross-validating data with experiment continues to improve as the Kish ratio decreases beneath the selected $K \geq 0.10$ threshold. With our approach, we also choose to prioritize obtaining a desired minimum effective ensemble size over maximizing agreement

with unrestrained cross-validating data.

We compare the agreement between calculated and experimental data for each reweighted Aβ40 ensemble in Figure 2D. We observe that when we use all available experimental NMR data ($C\alpha$, $C\beta$, $H\alpha$, H, and N backbone NMR chemical shifts, backbone $^3J_{HNH\alpha}$ coupling constants, and backbone RDCs) as restraints our reweighting protocol simultaneously improves agreement with all experimental data types (Figure 2D, "All Restraints"). For each experimental data type, we obtain a smaller RMSE between calculated and experimental data when we perform reweighting using only those data as restraints (Figure 2D, "Individual Restraints") compared to the RMSE obtained when performing reweighting using all available data as restraints. This demonstrates that satisfying multiple types of data requires trade offs in the agreement with each type of restraint and that our reweighting procedure naturally balances the weights of different types of restraints without requiring manual adjustment of regularization parameters or cross-validation thresholds.

We observe that the RMSE between calculated and experimental data for each data type withheld in leave-one-out cross validation reweighting calculations is very similar to the RMSE obtained when reweighting using all experimental data as restraints (Figure 2D, "Cross validation"). This suggests that using a Kish ratio threshold of $K \geq 0.10$ to determine regularization parameters for each data type results in minimal overfitting when all experimental data are used as restraints. We observe that the structural properties (secondary structure populations, intramolecular contact populations and R_g distribution) of the Aβ40 conformational ensemble obtained by reweighting with all experimental data and the structural properties of conformational ensembles obtained from leave-one-out cross-validation reweighting calculations were largely indistinguishable (data not shown). This suggests that we are in a data rich regime where all of the reweighted ensembles of Aβ40 derived from the unbiased a99SB-*disp* Aβ40 MD ensemble are extremely similar.

We present a comparison of the results of reweighting calculations performed on a99SB-*disp* MD ensembles of drkN SH3, ACTR, PaaA2 and α-synuclein in SI Figure 6, SI Figure

12, SI Figure 16, and SI Figure 19, respectively. We see similar global improvements in the agreement between calculated and experimental data in all proteins, illustrating the robustness of the reweighting protocol proposed here on a range of IDPs. We note one exception in the case of the protein ACTR, where the agreement between calculated and experimental SAXS data obtained using all experimental data as restraints is slightly (6%) worse than the agreement observed in the unbiased a99SB-*disp* MD ensemble of ACTR (SI Figure 12). We also observe that the RMSE between calculated and experimental SAXS data is 18% worse than the RMSE of the unbiased a99SB-*disp* MD ensemble when SAXS data is excluded as cross-validating data. This demonstrates some incompatibility in the agreement between experimental NMR data (backbone chemical shifts, backbone RDCs and PREs) and SAXS data in the conformations sampled in the unbiased a99SB-*disp* MD ensemble of ACTR.

Remarkably, only three of the fifty experimental data types withheld in leave-one-out cross-validation reweighting calculationsperformed on a99SB-*disp* MD ensembles of Aβ40, drkN SH3, ACTR, PaaA2 and α-synuclein were found to have worse agreement with experiment in reweighted ensembles than in unbiased MD ensembles: ACTR SAXS data, PRE data measured with a spin-label on residue 59 of drkN SH3 (drkN SH3 PRE-59), and α-synuclein backbone $^3J_{HNH\alpha}$ coupling constants (SI Figure 6, SI Figure 12, SI Figure 19). Similarly, only five data types withheld in leave-one-out cross-validation reweighting tests performed on C36m ensembles had worse agreement with experiment in reweighted ensembles than in unbiased MD ensembles (ACTR PRE-3, ACTR PRE-41, ACTR PRE-61, ACTR SAXS data and α-synuclein PRE-103) and only three data types withheld in leave-one-out cross-validation reweighting tests performed on C22* ensembles had worse agreement in reweighted ensembles (drkN SH3 PRE-59, ACTR PRE-41 and α-synuclein PRE-103). These results demonstrate that the reweighting protocol introduced in this investigation to refine IDP ensembles improves agreement with extensive experimental datasets from NMR spectroscopy and SAXS with minimal overfitting.

Comparing the accuracy of reweighted IDP ensembles derived from different force fields

We sought to compare the accuracy of reweighted ensembles of Aβ40, drkN SH3, ACTR, PaaA2 and α-synuclein derived from 30μs unbiased a99SB-*disp*, C22* and C36m MD simulations based on their agreement with experimental data. In order to compare the accuracy of the reweighted ensembles obtained from different force fields we define a global quality index for each system: the averaged normalized RMSE ($\overline{\mathrm{RMSE_N}}$). First, for each force field and data type i, we compute the agreement between calculated and experimental data in the reweighted ensemble ($\mathrm{RMSE_i}$). We then normalize each value of $\mathrm{RMSE_i}$ by the corresponding RMSE value in the unbiased a99SB-*disp* MD ensemble ($\mathrm{RMSE_{N,i}} = \mathrm{RMSE_i}/\mathrm{RMSE_{i,a99SB-}}_{disp}$) and finally average across all M data types to obtain $\overline{\mathrm{RMSE_N}} = 1/\mathrm{M} \sum_{\mathrm{i=1}}^{\mathrm{M}} \mathrm{RMSE_{N,i}}$.

We observe similar agreement between calculated and experimental data in the reweighted a99SB-*disp*, C22*, and C36m ensembles of ACTR (Figure 3), Aβ40 (SI Figure 2) and drkN SH3 (SI Figure 7). The $\overline{\mathrm{RMSE_N}}$ of the reweighted ensembles of ACTR are 0.84, 0.84, 0.85, for a99SB-*disp*, C22* and C36m, respectively (Figure 3C); the $\overline{\mathrm{RMSE_N}}$ of reweighted ensembles of Aβ40, are 0.74, 0.74, 0.66, for a99SB-*disp*, C22* and C36m, respectively (SI Figure 2C); and the $\overline{\mathrm{RMSE_N}}$ of reweighted ensembles of drkN SH3 are 0.80, 0.95, 0.91, for a99SB-*disp*, C22* and C36m, respectively (SI Figure 7C). All of the reweighted ensembles of these three IDPs are in exceptionally good agreement with experimental NMR and SAXS data, and are in substantially better agreement with experimental data than the most accurate unbiased MD ensembles of these proteins reported in previous benchmark studies.[19,21]

We observed substantially larger differences between calculated and experimental data in reweighted ensembles of PaaA2 (Figure 4) and α-synuclein (SI Figure 20). The $\overline{\mathrm{RMSE_N}}$ of reweighted ensembles of PaaA2 are 0.80, 1.39, 1.14, for a99SB-*disp*, C22* and C36m, respectively (Figure 4C), while the $\overline{\mathrm{RMSE_N}}$ of reweighted ensembles of α-synuclein are 0.79, 2.46, 1.97, for a99SB-*disp*, C22* and C36m, respectively (SI Figure 20C). Reweighted a99SB-*disp*, C36m and C22* ensembles of PaaA2 have a similar agreement with experimental NMR

chemical shift and RDC data (Figure 4C). The agreement between calculated and experimental data in reweighted C22* and C36m α-synuclein ensembles is, however, substantially worse than the agreement observed in the reweighted a99SB-*disp* ensemble (SI Figure 20C).

The agreement with experimental SAXS data is particularly poor in reweighted C22* and C36m ensembles of PaaA2 and α-synuclein. The C22* and C36m MD ensembles of these IDPs are substantially more compact than the a99SB-*disp* ensembles both before and after reweighting (Figure 4, Figure 5, SI Figure 20, SI Figure 22). Reweighting C22* and C36m ensembles of PaaA2 and α-synuclein only produces a marginal improvement in their agreement with the experimental SAXS data (Figure 4C, SI Figure 20). This results from the fact that unbiased C36m and C22* ensembles of PaaA2 and α-synuclein contain almost no conformations with a R_g as large as the experimental R_g determined from SAXS[19,37,51] (Figure 5, SI Figure 22). This places an upper limit on the agreement that can be obtained from reweighting regardless of the specified Kish ratio threshold. Reweighted C36m and C22* ensembles of α-synuclein also have substantially worse agreement with experimental NMR chemical shift, RDC, and PRE data than the reweighted a99SB-*disp* ensemble, due to an overestimation of β-sheet content and overly compact ensemble dimensions (SI Figures 20-23).

Comparing the structural properties of reweighted IDP ensembles derived from different force fields

We next sought to determine if the ensembles of Aβ40, drkN SH3, ACTR, PaaA2 and α-synuclein obtained by reweighting MD ensembles derived from different force fields converge to similar conformational distributions when extensive experimental datasets are used as restraints. To assess the similarity of unbiased and reweighted conformational ensembles of each protein, we compared the populations of α-helical and β-sheet secondary structure elements, the populations of intramolecular contacts, and the free energy surfaces of each ensemble as a function of R_g and the α-helical order parameter $S\alpha$[68] (See Methods). The

populations of α-helices in unbiased and reweighted ensembles of ACTR and PaaA2 are shown in Figure 3 and Figure 4, respectively. The populations of α-helices in unbiased and reweighted ensembles of Aβ40, drkN SH3 and α-synuclein are shown in SI Figure 2, SI Figure 7, and SI Figure 20, respectively. We observe highly consistent populations of moderately stable helices (populations of 20%-40%) in the reweighted ensembles of ACTR and drkN (Figure 3, SI Figure 7). In the case of PaaA2, we observe large differences in the locations and populations of helical elements in unbiased MD ensembles that become substantially more similar after reweighting (Figure 4). We note that there are strong signals for the helical elements of drkN SH3, ACTR and PaaA2 in experimental NMR chemical shift, RDC, and NOE data and that the populations of helical conformations observed in our reweighted ensembles are in excellent agreement with helical populations directly estimated from experimental data.[49–51,56]

The populations of β-sheets in unbiased and reweighted ensembles for Aβ40, drkN SH3, ACTR, PaaA2 and α-synuclein are shown in SI Figure 2, SI Figure 8, SI Figure 13, SI Figure 17, and SI Figure 21, respectively. We observe that all reweighted ensembles of Aβ40 converge to very similar secondary structure populations, with residues 15-20 and residues 30-35 forming β-sheets with populations of 20%-40% (SI Figure 2). We observe several highly populated β-sheets (populations of 40%-95%) in unbiased C36m and C22* MD ensembles of drkN SH3 and α-synuclein that produce notable discrepancies with experimental NMR data (SI Figure 7, SI Figure 8, SI Figure 20, SI Figure 21). The populations of these β-sheets are only partially decreased upon reweighting. The unbiased C36m MD ensemble of drkN SH3 contains β-sheets between by residues 1-10 and residues 20-30 with populations of 40%-80%. These populations are reduced to 20%-40% after reweighting with all experimental data. The reweighted C36m drkN SH3 ensemble, however, has substantially worse agreement with NMR chemical shifts and $^3J_{HNH\alpha}$ scalar couplings than the reweighted a99SB-disp drkN SH3 ensemble, which contains very little residual β-sheet structure. We observe similarly modest reductions in spurious populations of β-sheets in reweighted C22* and C36m ensembles of

α-synuclein (SI Figure 20, SI Figure 21).

To determine if the persistence of spurious β-sheet populations is the result of our selected Kish ratio threshold of $K = 0.10$ we computed reweighted C22* and C36m ensembles of drkN SH3 and α-synuclein using a Kish ratio threshold of $K = 0.01$. We observed similar β-sheet populations in these reweighted ensembles, suggesting that the persistence of these β-sheets is likely the result of the relatively weak effect of residual β-sheet structure on experimental secondary NMR chemical shifts[36,69] and the properties of empirical structure based chemical predictors. Previous work has shown that experimental secondary NMR chemical shifts and structure-based backbone NMR chemical shift predictions from algorithms such as SHIFTX+[17] and SPARTA+[16] are inherently less discriminative of conformations in the β-sheet region of backbone ϕ/ψ Ramachandran space than conformations in the α-helical region.[36,69]

We observe that IDP ensembles derived from MD simulations run with different force fields converge to very similar populations of residual helical structure after reweighting, but IDP ensembles with identical populations of α-helices can have large differences in the lengths of helical elements observed in individual conformations. An IDP ensemble containing ten consecutive residues with 50% α-helical populations might be constituted of only structures containing a contiguous ten-residue helix or structures with no helical content. Alternatively, this ensemble could contain conformations with short helical elements where the populations of helical turns are relatively uncorrelated from neighboring residues. To obtain a more detailed measure of the similarity of the individual conformations within unbiased and reweighted IDP ensembles, we compared the free energy surfaces of the unbiased and reweighted ensembles of each protein as a function of R_g and the α-helical order parameter $S\alpha$. $S\alpha$ is a measure of the number of seven residue fragments in a structure that resemble an ideal helix[68] (See Methods). The free energy surfaces of unbiased and reweighted ensembles of PaaA2 are shown as a function of R_g and $S\alpha$ in Figure 5. The free energy surfaces of unbiased and reweighted ensembles of Aβ40, drkN SH3, ACTR and α-synuclein are shown

as a function of R_g and $S\alpha$ in SI Figure 3, SI Figure 9, SI Figure 14 and SI Figure 22, respectively.

We observe dramatic differences in the free energy surfaces of the unbiased a99SB-*disp*, C36m and C22* ensembles of PaaA2 both before and after reweighting (Figure 5). The reweighted a99SB-*disp*, C22* and C36m PaaA2 ensembles have similar per-residue populations of helical conformations but have very little overlap when projected using R_g and $S\alpha$ descriptors. For comparison, we plot the R_g and $S\alpha$ values of 50 conformations contained in a protein ensemble database (PED)[70] ensemble of PaaA2 calculated directly from NMR and SAXS restraints (PED00013)[51] as white dots in each free energy surface in Figure 5. We observe that the unbiased a99SB-*disp* PaaA2 MD ensemble has two broad free energy minima in the R_g and $S\alpha$ projection. One of these minima overlaps very well with the conformations in PED00013 ensemble, and the depth of this minimum is substantially increased upon reweighting. The overly compact unbiased C22* and C36m MD ensembles of PaaA2 have relatively little overlap with the experimental PaaA2 ensemble in this projection. Upon reweighting, the overlap with more compact conformations in the experimental ensemble increases, but the reweighted C22* and C36m ensembles are substantially more similar to the unbiased MD ensembles from which they were derived than to the experimentally restrained PED ensemble or the reweighted a99SB-*disp* PaaA2 ensemble. Reweighted PaaA2 ensembles clearly do not converge to similar conformational distributions across different force fields.

Free energy surfaces of unbiased and reweighted drkN SH3 ensembles are compared to an experimental ensemble (PED00427) in SI Figure 9. All three unbiased MD ensembles of drkN SH3 have substantial overlap with the experimental ensemble in the R_g and $S\alpha$ projection, and the free energy surfaces of reweighted ensembles are relatively similar. In contrast to PaaA2, reweighted drkN SH3 ensembles appear to converge to similar conformational distributions with similar lengths of helical elements and a similar coupling between helix formation and chain compaction. Free energy surfaces of reweighted and unbiased ensembles of Aβ40, ACTR and α-synuclein are shown as a function of R_g and $S\alpha$ in SI Figure 3, SI

Figure 14 and SI Figure 22, respectively. Reweighted ensembles of Aβ40, which have minimal helical content, are fairly similar in this projection. Unbiased ensembles of ACTR have fairly large differences in this projection, and become more similar after reweighting. Unbiased and reweighted α-synuclein ensembles have very little overlap in these projections.

We examine the similarity of the populations of intramolecular contacts, or *contact maps*, of unbiased and reweighted ensembles of Aβ40, drkN SH3, ACTR, PaaA2 and α-synuclein in SI Figure 4, SI Figure 10, SI Figure 15, SI Figure 18 and SI Figure 23, respectively. Unbiased Aβ40 ensembles have a similar pattern of intramolecular contacts, defined by contacts between residues 15-20 and residues 30-35, with notable differences in populations. The populations of these contacts, which are driven by transient β-sheet formation, become more similar after reweighting (SI Figure 4). The populations of intramolecular contacts of drkN SH3 and ACTR are quite similar in unbiased MD ensembles of C36m and C22*, while the more extended unbiased a99SB-*disp* MD ensembles contain fewer intramolecular contacts. Upon reweighting ensembles of drkN SH3 and ACTR with experimental datasets that contain PREs and SAXS data, the contact maps become more similar, and the contact maps of the reweighted ACTR ensembles are particularly similar (SI Figure 15).

The reweighted C36m and C22* ensembles of drkN SH3 contain several intramolecular contacts resulting from β-sheet structure not observed in the reweighted a99SB-*disp* ensemble (SI Figure 10). The overly compact unbiased C36m and C22* MD ensembles of PaaA2 and α-synuclein contain a large number of highly populated intramolecular contacts relative to the more realistically expanded unbiased a99SB-*disp* MD ensembles (SI Figure 18, SI Figure 23). The contacts observed in C36m and C22* PaaA2 and α-synuclein ensembles remain highly populated after reweighting, as there are too few extended conformations in the unbiased MD ensembles to reduce the populations of these contacts.

Quantifying the similarity of reweighted IDP ensembles derived from different force fields

While comparing IDP ensembles using selected reaction coordinates of interest (such as R_g and S_α) can be both intuitive and informative, we sought to obtain a more objective and quantitative description of the similarity of unbiased and reweighted ensembles of Aβ40, drkN SH3, ACTR, PaaA2 and α-synuclein. To do so, we utilized the energy landscape visualization method (ELViM) for dimensionality reduction.[71–73] The ELViM approach uses the distances between $C\alpha$ carbons in each conformation of a protein ensemble as an input, calculates a dissimilarity matrix between all pairs of conformations based on differences in the populations of $C\alpha$-$C\alpha$ contacts and projects the information contained in the high-dimensional dissimilarity matrix onto a low-dimensional latent space. The ELViM method is conceptually similar to using t-stochastic nearest neighbor embedding (t-SNE) to cluster conformations of IDPs.[8,74] Projecting IDP ensembles onto a low-dimensional latent space derived directly from atomic coordinates provides a more objective comparison of the similarity of ensembles than examining the similarity of a small number of subjectively selected structural descriptors.

We applied the ELViM approach to compare the similarity of the conformational ensembles sampled in the unbiased a99SB-*disp*, C36m and C22* MD simulations of Aβ40, drkN SH3, ACTR, PaaA2 and α-synuclein. For each protein, we concatenated the three unbiased MD ensembles into a single merged ensemble, computed the dissimilarity matrix between all conformations and used the ELViM algorithm to project the conformations of the merged ensemble onto a two-dimensional (2D) latent space (See Methods). We then used kernel density estimates of each unbiased and reweighted ensemble projected on to the ELViM latent space to compare the ensembles. A comparison of the projections of the unbiased and reweighted ensembles of ACTR and PaaA2 on their ELViM latent spaces are shown in Figure 6 and Figure 7, respectively. To obtain more insight into the nature of the ELViM latent spaces, we project the R_g and S_α values of each conformation of the merged

ensembles on their respective ELViM latent space and present snapshots of conformations from selected regions of ELViM density projections in unbiased and reweighted ensembles. A comparison of the projections of the unbiased and reweighted ensembles of Aβ40, drkN SH3 and α-synuclein on their respective ELViM latent space are shown in SI Figure 5, SI Figure 11 and SI Figure 24, respectively.

We display overlays of the ELViM latent space embeddings of all ensembles of Aβ40, drkN SH3, ACTR, PaaA2 and α-synuclein in Figure 8A. To provide a quantitative measure of the similarity of the unbiased and reweighted ensembles of each protein in the ELViM latent space we define a *density overlap* metric S (Eq. 13), which is analogous to the overlap integral used to quantify the overlap of electronic wave functions in quantum mechanics (See Methods). We first construct a kernel density of each ensemble in the ELViM latent space. We compare the similarity of two ensembles by computing the overlap integral S of their kernel densities, normalized such that the value of S ranges from [0,1] (Eq. 14). Two densities with no overlapping points will have an overlap integral value of S=0, while two identical densities will have an overlap integral value of S=1. We display the values of the overlap integrals between the reweighted and unbiased ensembles of each protein, expressed as an overlap percentage ($S \cdot 100\%$) in Figure 8B. Values in the blue triangles reflect the overlap of unbiased MD ensembles derived from different force fields, values in red triangles reflect the overlap of reweighted ensembles derived from different force fields, and the diagonal elements reflect the overlap of the reweighted ensemble derived from each force field with the unbiased MD ensemble from which it was derived.

The ELViM projections of ACTR and Aβ40 and ensembles shown in Figure 6, Figure 8 and SI Figure 5 illustrate that the unbiased a99SB-*disp*, C36m and C22* MD simulations of these proteins sample the same regions of conformational space with different probabilities, and that these probabilities are adjusted to produce highly similar conformational distributions after reweighting with experimental data. The reweighted ensembles of ACTR and Aβ40 share similar high density regions in the ELViM latent space, which is consistent with

the similarity of the intramolecular contact maps and secondary structure propensities of the reweighted ensembles of these proteins (SI Figures 2-4, SI Figures 13-15). Unbiased drkN SH3 ensembles initially have relatively little overlap in the ELViM latent space, but their overlap is substantially increased upon reweighting (Figure 8, SI Figure 11).

The ELViM projections of PaaA2 ensembles shown in Figure 7 and Figure 8 illustrate that there is very little overlap of unbiased MD ensembles derived from different force fields in the space of $C\alpha$ contacts. This is reflected by overlap percentages of less than 2% between unbiased ensembles (Figure 8B). As a result, while reweighting marginally increases the overlap of the ensembles, reweighted PaaA2 ensembles are substantially more similar to the unbiased MD ensembles from which they were derived than to one another. This results from the fact that the stable helical elements in PaaA2 ensembles can pack in distinct discrete orientations in the overly collapsed C22* and C36m ensembles, and the unbiased MD ensembles do not sample the same packing orientations. The unbiased and reweighted ensembles of α-synuclein derived from different force fields are as similarly disjoint as the PaaA2 ensembles in the ELViM latent space, with almost no overlap before or after reweighting (Figure 8, SI Figure 24).

Discussion

In this investigation we have proposed a simple and robust maximum entropy reweighting approach to refine atomic resolution conformational ensembles of IDPs using large datasets consisting of several different types of experimental data. Our proposed reweighting procedure contains a single free parameter, the desired effective ensemble size of the reweighted ensemble, and naturally balances the weights of restraints for different types of experimental data. This approach does not require any *a priori* knowledge of the accuracy of an initial IDP ensemble, the magnitude of errors of experimental measurements, the correlation of experimental observables or the accuracy of forward models for predicting experimental

data. We demonstrate, through extensive cross-validation, that the proposed reweighting approach simultaneously improves agreement with several types of experimental NMR data and SAXS data with minimal overfitting while maintaining a desired degree of sampling of the most populated regions of conformational space in unbiased MD ensembles.

The reweighting approach presented here builds upon several successful maximum entropy and Bayesian approaches for reweighting conformational ensembles of flexible molecules with experimental data.[26–32] Refining conformational ensembles of flexible molecules with maximum entropy reweighting poses two inherent challenges: i) determining how to balance the weights of experimental restraints against one's confidence in the initial conformational ensemble, and ii) determining the relative weights of different types of experimental restraints. Several heuristic and theoretical approaches have been proposed for determining or inferring the weights of experimental restraints when reweighting ensembles of flexible molecules.[29–33,35,45] When considering experimental datasets with a small number of experimental data types cross-validation or L-curve/elbow methods provide a practical and direct approach. When performing reweighting using a few sets of experimental data it is reasonable and expedient to determine the weight of each set of restraints by identifying weight values where the agreement of calculated cross-validating data with experiment begin to deteriorate.

When reweighting conformational ensembles with data from many different experiments this process becomes less straightforward and one must attempt to normalize the relative RMSE between predicted and experimental values based on estimates of experimental measurement errors and the errors of forward models used to predict experimental data. This is frequently done through the use of χ^2 values obtained by normalizing RMSE calculations with various error models. While it is possible to propose reasonable estimates of experimental measurement and forward model prediction errors for different types of data, these quantities are inherently uncertain and these estimates introduce a degree of subjectivity into reweighting protocols. Additionally, in cases where many experimental data are highly

correlated, such as NMR data that report on the distribution of backbone dihedral angles of IDPs (ie. chemical shfits, scalar coupling constants, RDCs), cross-validation may not be a practical approach to determine the relative weights of restraints.

We set out to develop a reweighting protocol to refine IDP conformational ensembles obtained from all-atom MD simulations that is minimally biased and maximally automated. We have proposed a heuristic approach for determining the relative weights of experimental restraints that directly builds upon previous work that considers the Kish ratios of reweighted ensembles.[28] This approach is motivated by several practical considerations. One motivation is a desire to objectively compare the accuracy of IDP ensembles obtained from reweighting unbiased MD simulations run with force fields with different levels of accuracy. If a reweighting procedure contains adjustable parameters to determine the relative importance of different types of experimental data in a reweighting objective function, it is possible to adjust the relative "scores" or likelihoods of reweighted ensembles derived from different force fields by adjusting these parameters.

A second motivation for using the effective ensemble size of reweighted ensembles to choose the relative weights of experimental restraints is a desire to directly compare the structural properties and similarity of reweighted ensembles derived from different force fields. If we choose to use RMSE (or χ^2) thresholds to determine the weights of experimental restraints, we observe that reweighted ensembles derived from different force fields can achieve similar RMSEs with experimental data with dramatically different effective ensemble sizes. Requiring an overly compact C22* or C36m ensemble to agree with SAXS data within a specified tolerance could require reducing the ensemble to a small handful of tens of structures, while a more realistically extended a99SB-*disp* ensemble might contain 1,000s or 10,000s of structures after reweighting. Comparing the structural properties and structural overlap of reweighted ensembles with orders-of-magnitude differences in effective ensemble sizes is less meaningful than comparing ensembles with similar effective ensemble sizes.

An additional motivation for using the effective ensemble size of reweighted ensembles

to determine the relative weights of restraints is our desire to adequately represent the most populated regions of conformational space sampled in unbiased MD simulations in reweighted IDP ensembles. One application of our research efforts to determine accurate atomic resolution ensembles of IDPs is to obtain mechanistic insight into their binding mechanisms with small molecule drugs.[6–11] These dynamic and heterogeneous binding mechanisms are mediated by statistical distributions of backbone conformations and positions of sidechain pharmacophores. As such, IDP ensembles with a small number of conformations that sparsely sample backbone topologies and the relative positions of sidechain atoms are unlikely to be informative for predicting small molecule binding sites or understanding subtle shifts in the conformational ensembles of IDPs that occur upon small molecule binding.[6–8,10,11]

We note a conceptual similarity of the reweighting procedure proposed here with the concept of gentle ensemble refinement in the recently published work of Köfinger and Hummer.[32] In this work, the authors propose an elegant approach to balance the weight of experimental restraints in maximum entropy reweighting against confidence in an initial prior model. They relate the KL divergence between the weights of an initial and reweighted ensemble to an energy uncertainty of the unbiased ensemble, and propose to select the optimal value of restraint weights by relating the expected force field accuracy in the space of experimental observables to the expected energy variance upon reweighting. While motivated by theoretical considerations, this approach is similar in spirit and in practice to the heuristic approach proposed here, where we specify an acceptable limit to changes in the statistical weights of conformations in unbiased ensembles *a priori*.

We draw attention to the small effective ensemble sizes obtained with our proposed reweighting procedure if we utilize agreement with restrained or unrestrained data as the primary criterion to select the strength of the experimental restraints. We observe that the RMSE between calculated and experimental data of both restrained and unrestrained cross-validating data largely improve as the weight of experimental restraints is increased until ensemble sizes become very small (Kish ratio $K < 0.01$), corresponding to effective ensemble

sizes of a few hundred structures. We note that several approaches to calculate IDP conformational ensembles with a minimal set of structures, sometimes referred to as *maximum parsimony*[12] approaches, have previously found that ensembles of ∼50-500 structures are required to best reproduce NMR and SAXS data of IDPs.[34–37,50,51,61] It appears that the IDP ensembles obtained by reweighting long timscale MD simulations using cross-validation approaches produce ensembles with a similar size to ensembles calculated by minimal ensemble approaches. We observe that the reweighted ensemble of IDPs obtained in this investigation have similar structural properties to conformational ensembles calculated from NMR and SAXS data using minimal ensemble or maximum parsimony approaches. We find that α-helical populations of reweighted drkN SH3 ensembles obtained in this work agree well with the helical populations of drkN SH3 ensembles derived from NMR and SAXS data with approaches such as ENSEMBLE,[34] X-EISD[35] and IDPConformerGenerator.[75] We also observe that α-helical populations in PaaA2 ensembles obtained from our reweighting approach agree reasonably well with populations obtained in a previously reported PaaA2 ensemble[51] derived from NMR and SAXS data using the EOM approach.[76]

We find that our proposed reweigthing procedure only marginally reduces the populations of potentially spurious β-sheets in reweighted ensembles of drkN SH3 and α-synuclein relative to those calculated by X-EISD[35] and ASTEROIDS.[37] We note that these approaches, and other ensemble calculations approaches that generate conformations by statistically sampling backbone dihedral angles, may not sample any hydrogen-bonded β-sheet conformations in initial conformational pools used for ensemble selection, but rather aim to identify elevated sampling of β-sheet Ramachandran space relative to statistical coil models. Therefore, populations of β-sheets observed in reweighted MD ensembles may not be directly comparable to β-propensities identified in these ensemble calculation approaches. Nevertheless, we caution that the presence of substantially populated β-sheets in reweighted MD ensembles produced by our reweighting procedure does not necessarily suggest strong experimental evidence for residual β-sheet populations. We encourage a direct inspection of agreement with

experimental data for these β-sheet elements. We emphasize that an important difference between ensembles obtained from reweighting long timescale MD simulations and ensemble selection approaches such as ASTEROIDS,[36,37,61] EOM[76] or IDPConformerGenerator[75] is that the positions of side chains in reweighted MD ensembles are governed by the physics of the underlying force fields, compared to a stochastic sampling of side chain rotamer libraries to place side chains and avoid steric clashes. The positions of sidechains in reweighted MD ensembles may therefore have more physical meaning.

The maximum entropy reweighting procedure developed in this investigation enables us to address the following question: are modern force fields sufficiently accurate that MD simulations with adequate sampling will converge to the similar underlying conformational distribution when extensive experimental datasets are used for reweighting? This question can be alternatively viewed as: with sufficient experimental data, does the problem of determining structural ensembles of IDPs with modern force fields and maximum entropy reweighting methods become well-defined[12]? In the case of Aβ40 and ACTR, two shorter IDPs (40 and 69 residues, respectively) with moderate populations of residual secondary structure elements, we find substantial overlap in the unbiased a99SB-*disp*, C36m and C22* MD ensembles from both traditional structural descriptors and from a low dimensional projection of the ensembles onto a latent space defined by the positions of Cα carbons. Upon reweighting, we find that the ensembles become substantially more similar to one another and have similar levels of agreement with experimental data. These encouraging results suggest that accessible conformational space of these proteins is well sampled in 30μs-long MD simulations run with different force fields, and that available experimental NMR and SAXS datasets provide sufficient information to correct the simulated ensembles into similar conformational distributions. While this is not evidence that these reweighted ensembles are perfectly faithful representations of the *true* solution ensembles of these proteins, it is encouraging that each force field does not produce a unique reweighted ensemble with equally good agreement with experimental restraints. These results suggest that with modern force

fields and sufficient NMR and SAXS data, the challenge of determining structural ensembles of IDPs with fewer than 70 residues may now be becoming well-defined, in that consensus descriptions of solution ensembles are beginning to emerge.

In contrast, the reweighted ensembles of PaaA2 and α-synuclein produced in this work have extremely little overlap. The unbiased MD ensembles of these proteins appear to sample almost entirely disjoint regions of conformational space. The lack of overlap of unbiased ensembles is likely the result of a combination of force field inaccuracies and insufficient sampling in unbiased 30μs MD simulations. While reweighting improves some properties of the overly compact C22* and C36m ensembles of PaaA2 and α-synuclein, such as the distributions of backbone dihedral angles, we are clearly far from identifying a consensus description of the solution ensembles of these proteins. Based on the superior agreement with experimental data, the reweighted a99SB-*disp* ensembles of PaaA2 and α-synuclein are likely substantially more representative of the true solution ensembles of these proteins than the reweighted C22* and C36m ensembles. In the case of drkN SH3, the unbiased ensembles appear relatively similar aside from substantial populations of β-sheets in the C36m and C22* ensembles. Reweighting reduces the populations of these β-sheets and substantially increases the overlap of the conformational ensembles, but agreement with experimental data indicate that the reweighted a99SB-*disp* drkN SH3 ensemble is the most accurate representation of this protein in solution.

Accurate, force field independent conformational ensembles of IDPs, like the ones determined here for Aβ40 and ACTR, are extremely valuable for a variety of reasons. First, these reweighted ensembles can be used as potential target distributions when assessing the accuracy of coarse-grained representations of IDPs,[77,78] machine learning methods for predicting atomic resolution conformational ensembles,[46,79] and ensemble properties of IDPs.[3,80,81] Additionaly, these reweighted ensembles can serve as training data for machine learning approaches to generate IDP conformational ensembles at reduced computational cost compared to standard MD simulations. Force field independent ensembles could complement the in-

formation encoded in the co-evolution of amino-acid sequences and the structures deposited in the Protein Data Bank,[82] on which methods like AlphaFold3[47] were trained, but which provide limited information about the conformational heterogeneity of dynamic and disordered proteins. A major challenge for these efforts will be generating a sufficient number of high-quality conformational ensembles to train IDP models that generalize well to unseen data, especially when considering the vast conformational space potential of IDPs.

In conclusion, we have proposed and validated a robust approach for integrating MD simulations of IDPs with extensive experimental sets to improve the accuracy of atomic resolution conformational ensembles of IDPs. We have provided an in depth comparison of the reweighted IDP ensembles derived from different force fields and have shown that in favorable cases these ensembles converge to very similar conformational distributions after reweighting. We have also shown that for several proteins, IDP ensembles obtained from reweighting atomic resolution MD simulations have very similar properties to IDP ensembles calculated directly from experimental data with minimal ensemble methods. These results demonstrate substantial progress in the field of IDP ensemble modeling, and suggest that the field may be maturing from the realm of assessing the accuracy of disparate computational models to the realm of atomic resolution integrative structural biology. We anticipate that the maximum entropy reweighting protocol proposed here will provide a valuable tool for calculating accurate atomic resolution IDP ensembles, assessing the accuracy of future force field development efforts, scrutinizing IDP ensembles obtained from artificial intelligence and machine learning methods and ultimately providing high-quality data to train novel, efficient, and accurate deep learning models for IDP conformational ensemble generation.

Theory

In the following sections we provide a brief overview of the theory of the maximum entropy reweighting and the treatment of different sources error with regularization parameters. We

focus on the formalism proposed by Bussi et al.[26–28] which is the basis of our approach. We build upon this formalism by proposing a heuristic approach to determine regularization parameters in an automated fashion, using the Kish ratio (Eq. 11).

Overview of maximum entropy reweighting

A MD simulation generates a prior distribution $P_0(x)$ of conformations x. In initial implementations,[83–85] maximum entropy reweighting approaches were designed to determine the distribution $P_1(x)$ which is the closest to $P_0(x)$ that fits *exactly* each experimental data point f_i^{exp}. To measure the statistical distance between the two distributions $P_0(x)$ and $P_1(x)$, one can use the Kullback–Leibler divergence, defined as:

$$D_{KL}[P_1|P_0] = \int P_1(x) \ln \frac{P_1(x)}{P_0(x)} dx \tag{1}$$

To determine $P_1(x)$, the closest distribution to $P_0(x)$ that fits the ensemble-averaged data, the KL divergence is then minimized subject to the following two constraints:

$$\int P_1(x) f_i(x) dx - f_i^{exp} = 0 \text{ for } i = 0, 1, 2, ..., M \tag{2}$$

$$\int P_1(x) dx = 1 \tag{3}$$

where $f_i(x)$ is the forward model needed to predict the value of observable i for each conformation x. In Eq. 2, each experimental observation constrains the ensemble average of the corresponding forward model $f_i(x)$ computed over the distribution $P_1(x)$ to be exactly equal to f_i^{exp}. The second constraint in Eq. 3 ensures that $P_1(x)$ is normalized.

The problem of finding the distribution $P_1(x)$ that is the best representation of the state of knowledge of a system after observing a set of experimental data can be cast in the framework of information theory. In this framework, amongst all the possible distributions compatible with the data, the one that has the highest degree of uncertainty should be

selected in order to ensure minimal bias and that the data has been used as conservatively as possible. This uncertainty is expressed by the Shannon entropy[38] of the random variable x with distribution $P_1(x)$ with respect to the reference distribution $P_0(x)$:

$$S[P_1|P_0] = -\int P_1(x) \ln \frac{P_1(x)}{P_0(x)} dx = -D_{KL}[P_1|P_0] \tag{4}$$

Maximizing the Shannon entropy in Eq. 4 to determine $P_1(x)$ is thus equivalent to minimizing the KL divergence in Eq. 1, hence the name of maximum entropy reweighting.

The solution for the minimization of the KL divergence subject to the constraints in Eq. 2 and 3 can be achieved by using Lagrange multipliers. We first define the following Lagrange function:

$$\mathcal{L} = D_{KL}[P_1|P_0] - \sum_{i=1}^{M} \lambda_i \left[\int P_1(x) f_i(x) dx - f_i^{exp} \right] - \mu \left[\int P_1(x) dx - 1 \right] \tag{5}$$

where λ_i and μ are Lagrange multipliers. There is one Lagrange multiplier λ_i for each of the of the M experimental data points being used as constraints. To find stationary points of the Lagrangian, we set $\partial\mathcal{L}/\partial P_1 = 0$ and neglect the normalization factor. This leads to the following form for the desired distribution $P_1(x)$:

$$P_1(x) \propto e^{-\sum_{i=1}^{M} \lambda_i f_i(x)} P_0(x) \tag{6}$$

Approaches such as gradient descent or appropriate quasi-Newton methods can be utilized to calculate the values of the Lagrange multipliers, which ultimately yield to a new set of weights, one for each conformation of the ensemble:

$$w(x) \propto e^{-\sum_{i=1}^{M} \lambda_i f_i(x)} \tag{7}$$

Introducing different source of errors

By fitting exactly the experimental observations, the formalism introduced above neglects the fact that experimental data as well as forward models can be affected by random and systematic errors and therefore data should not be matched *exactly* to avoid overfitting to noise. In order to alleviate the problem of data overfitting, maximum entropy reweighting approaches can be modified to account for different sources of errors as well as a limited confidence in the prior distribution. Different approaches have been proposed over the years.[27,30–33] In the formalism of Bussi et al.,[27] errors are modeled by modifying the constraints of Eq. 2 by introducing an auxiliary variable for each data point ϵ_i, which represents the difference between experimental and predicted values. The new constraints are hence defined as follows:

$$\int P_1(x)\,(f_i(x) + \epsilon_i)\,dx - f_i^{exp} = 0 \text{ for } i = 0, 1, 2, ..., M \tag{8}$$

Errors can be modeled by choosing a proper prior distribution function for the variables ϵ_i. We choose a Gaussian prior with fixed standard deviation σ_i for the ith observable.

$$P_0(\epsilon_i) \propto \prod_{i=1}^{M} e^{-\left(\frac{\epsilon_i^2}{2\sigma_i^2}\right)} \tag{9}$$

The value of σ_i corresponds to the level of confidence in the ith experimental data point and ultimately determines how well the reweighted ensemble will fit the data. $\sigma_i = \infty$ implies no confidence in the data and results in a reweighted ensemble that is identical to the unbiased simulation. $\sigma_i = 0$ implies complete confidence in the data and results in a reweighted ensembles in which the ith data point is exactly matched, as in Eq. 2. The selection of appropriate values of the regularization parameters σ_i for each experimental data point is therefore essential to balance one's confidence in a prior model (ie. the accuracy of a force field), error estimates of experimental measurements and error estimates of the accuracy of forward models for predicting experimental data. Ultimately, the properties of the resulting conformational ensembles will be highly dependent on the relative magnitudes

of the regularization parameters selected for each experimental data point.

Different strategies have been developed to identify the optimal values of σ_i. Typically, when performing reweighting using multiple types of experimental data, one value of σ_i is assigned to each class of experimental data, such as SAXS intensities or NMR chemical shifts, RDCs or scalar couplings. A grid search is then performed, usually independently for each class of experimental data, and the optimal value of σ_i is identified for each class to maximize the agreement of the reweighted ensemble with independent experimental data not used for reweighting.[29] A final reweighting step using all the available experimental data is then performed.

Determining regularization parameters

In our proposed maximum entropy approach, we adopt the above formalism of Bussi et al. with the following modifications. We assign one regularization parameter σ_i to each data point. To optimize the grid search over the space of σ_i, we decompose this parameter into two contributions:

$$\sigma_i = \sqrt{\sigma_{i,MD}^2 + \sigma_{reg}^2}. \tag{10}$$

where σ_{reg} is the regularization parameter, one per data type, that describes both experimental and forward model errors of this specific type of experiment. The $\sigma_{i,MD}$ parameter, one per data point, represents both the statistical errors of forward models to predict experimental data and statistical errors that result from calculating average quantities over a finite-size ensemble. The $\sigma_{i,MD}$ term can be calculated by using the Flyvbjerg block analysis[67] technique to account for the correlation between adjacent frames in the MD simulation or the standard error of the mean for uncorrelated data. We use a single free parameter, the Kish ratio (Eq. 11), to determine the optimal values of σ_{reg} for each experimental data type. This parameter will be described in the next section.

Kish Ratio

As described in the previous sections, maximum entropy reweighting algorithms change the statistical weights of each conformation of the input ensemble in a minimal way to improve the agreement with the available experimental data. In a conformational ensemble derived from an unbiased MD simulation with N frames, each conformation will have an equal statistical weight of $1/N$: $\boldsymbol{w_{MD}} = \{1/N, 1/N, \ldots, 1/N\}$. Upon reweighting, the weights of all frames are modified to $\boldsymbol{w} = \{w_0, w_1, ..., w_N\}$ as defined in Eq. 7. If the initial unbiased MD ensemble is already in excellent agreement with all the experimental data used for reweighting, it may be possible to satisfy these data with very small perturbations to the weights of each frame. In this instance, the weight of most frames will still be close to $1/N$. If instead a large fraction of frames of the unbiased MD ensemble are in poor agreement with the experimental data, in order to satisfy the data within a desired threshold it may be necessary to reduce the statistical weights of a significant number of frames to values close to 0, effectively discarding these frames from the reweighted ensemble.

To quantify the overall change in weights occurred upon reweighting, we use the (normalized) Kish effective sample size,[28,66] or *Kish ratio* K:

$$K = \frac{1}{N}\left(\sum_{i=1}^{N} w_i\right)^2 / \sum_{i=1}^{N} w_i^2 \tag{11}$$

where w_i is value of the statistical weight of the ith frame. The Kish ratio is a measure of the dimension of the reweighted ensemble expressed in terms of the fraction of frames of the original ensemble with significant statistical weight after reweighting. A reweighted ensemble where ∼50% of the frames have statistical weights close to zero will have a Kish ratio $K \approx 0.50$. Conversely, if an unbiased MD simulation has 10,000 frames a reweighted ensemble with a Kish ratio $K = 0.01$ would have approximately 100 frames (1% of frames) with statistical weights substantially larger than 0. By definition, a conformational ensemble derived from an unbiased MD simulation where all N weights are equal has a Kish ratio is

$K = 1.0$.

As discussed in the previous section, the σ_i parameters express our confidence in the experimental data. As the values of σ_i are changed during the optimization of the regularization parameters, the weights assigned to each conformation might be substantially altered to fit the experimental data within a desired threshold. Depending on the quality of the original MD ensemble, this might lead to dramatically reducing the Kish ratio of the reweighted ensemble, thus compromising the statistical accuracy of any average property calculated over the reweighted ensemble. Reweighted ensembles with small Kish ratios may effectively discard all structures representative of free energy basins in an initial MD simulation, and satisfy experimental restraints with a small set of rarely populated conformations (Figure 1). It is therefore desirable to ensure that the reweighted ensemble preserves a sufficiently large Kish ratio in order to be able to make predictions of the properties of IDPs, such as the relative positions and orientations of sidechain residues, with a desired statistical accuracy.

Methods

A maximum entropy reweighting protocol with a single free parameter

Our approach to combine multiple types of experimental data to reweight IDP ensembles uses only one free parameter: the Kish ratio of the reweighted ensemble (Eq. 11). Our protocol proceeds as follows:

1. We calculate $\sigma_{i,MD}$ for each data point from the unbiased MD simulation using a Fylvberg[67] blocking analysis.

2. For each experimental data type, we perform reweighting at a wide range of values of σ_{reg} and monitor the Kish ratio of the reweighted ensembles. For each data type, we choose a value of σ_{reg} based on a selected Kish ratio threshold. Here, we select the

minimum value of σ_{reg} that produces an ensemble with Kish ratio $K \geq 0.10$ for each data type. This procedure establishes the relative values of σ_{reg} for each data type (ie. $\sigma_{reg-C\alpha}$, $\sigma_{reg-J_{HNH\alpha}}$, $\sigma_{reg-SAXS}$, etc.).

3. A global reweighting is performed using all experimental data as restraints. In this step, we determine a global scaling factor $\sigma_{reg-Global}$ for the values of σ_{reg} calculated for each data type in step 2. We perform reweighting at a range of values of $\sigma_{reg-Global}$ and monitor the Kish ratio of the reweighted ensembles. We select the minimum value of $\sigma_{reg-Global}$ that produces a final reweighted ensembles with Kish ratio $K \geq 0.10$.

Calculation of experimental observations

All 30 μs MD simulations analyzed here were run on the Anton supercomputer[86] and contained 29976 frames, with a spacing of ∼1ns per frame. Backbone scalar coupling constants were calculated using previously determined Karplus equations.[87–89] RDCs were calculated with PALES[90] using a local alignment window of 15 residues.[56] Backbone chemical shifts were calculated using SPARTA+[16] with MDTraj.[91] The SAXS profiles of each frame were calculated using Pepsi-SAXS[92] following a previously described protocol[93] setting adjustable solvent parameters to $\delta\rho = 3.34\,\mathrm{e/nm}^3$ and $r_0 = 1.68$ $\mathring{A}$. The SAXS profiles were scaled by dividing the intensities by the intensities in the first scattering angle in order to normalize the datasets and set the initial intensity to 1. PREs were calculated using distances between $C\alpha$ atoms as described previously[20]

α-helical order parameter $S\alpha$

$S\alpha$ describes the similarity of each seven-residue fragment of a protein to an ideal helix.[68] $S\alpha$ is calculated as follows:

$$S_\alpha = \sum_i^N \frac{1 - \left(\frac{\mathrm{RMSD}\alpha_i}{r_0}\right)^8}{1 - \left(\frac{\mathrm{RMSD}\alpha_i}{r_0}\right)^{12}} \tag{12}$$

where RMSDα_i is the root mean square deviation (RMSD) between a seven-residue fragment of a protein (spanning from residue i to residue $i+6$) and an ideal seven residue helix. Eq. 12 functions as switching function which outputs values ranging from 0 (indicating not helical) to 1 (indicating perfectly helical) for each seven-residue segment. The threshold for this function is adjusted by the parameter r_0, which we set to 0.80 Å. Setting the parameter $r_0 = 0.80$ Å means that a seven-residue segment with RMSD$\alpha_i > 2.5$ Å effectively contributes a value 0 to the $S\alpha$ sum, while a seven-residue segment with RMSD$\alpha_i < 0.5$ Å contributes a value of 1.0 to the $S\alpha$ sum. The value of $S\alpha$ for a protein conformation can therefore be interpreted as a proxy for the number of seven-residue fragments that resemble an ideal helix.

ELViM ensemble comparisons

We provide a quantitative measure of the structural similarity of conformational ensembles by collectively analyzing a set of molecular dynamics simulations of the same protein by using the energy landscape visualization method (ELViM)[71–73] for dimensionality reduction and computing the overlap of latent space densities. For each protein, we concatenated all unbiased trajectories into a single merged ensemble. We used ELViM to compute a dissimilarity matrix for this merged ensemble with the hyper-parameters σ_0 and ϵ set to 1 and 0.15, respectively, and projected the conformations of the merged ensemble onto a 2D latent space. Individual kernel densities were constructed for each simulation dataset (each unbiased and reweighted ensemble of each protein) and sampled over the global extrema in each dimension of the ELViM latent space using an 80 by 80 grid. Gaussian kernel densities were estimated using the *SciPy*[94] Python package. The bandwidth of kernels were determined using Scott's rule[95] for both weighted and unbiased kernel density estimates. For a pair of kernel densities, D_1 and D_2, we evaluate the overlap integral, S:

$$S = N_1 N_2 \int \int D_1(x_1, x_2) D_2(x_1, x_2) dx_1 dx_2 \tag{13}$$

where N_1 and N_2 are normalization constants for the distributions D_1 and D_2 respectively, chosen such that the integrand is defined in the range : [0, 1]. For a kernel density D_i the normalization constant is defined as :

$$N_i = \frac{1}{\sqrt{\int\int D_i(x_1, x_2) D_i(x_1, x_2) dx_1 dx_2}} \tag{14}$$

If evaluations of kernel density estimators are compiled into vectors $\boldsymbol{D_1}$ and $\boldsymbol{D_2}$, this computation is equivalent the normalized dot product.

$$S = \frac{\langle \boldsymbol{D}_1 | \boldsymbol{D}_2 \rangle}{\|\boldsymbol{D}_1\| \|\boldsymbol{D}_2\|} \tag{15}$$

Using the definitions in Eq. 12-14, S is a positive value from [0,1]. Two kernel densities with no overlapping points have an overlap of value of $S = 0$, while two identical kernel densities have an overlap of $S = 1$. The values of overlap integrals S can be converted to a percentage by multiplying $S \cdot 100\%$. We computed the overlap integral of the ELViM projections of unbiased and reweighted ensembles derived from simulations of each protein with each force field, and compare the overlap of the ELViM projections for all pairs of unbiased and reweighted ensembles for each protein (Figure 8).

Acknowledgement

P.R., K.B. and T.R.S acknowledge the support of NIH award R35GM142750. M.B. and F.P.P acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-20-CE45-0002. The authors thank Vincent Schnapka and Vitor Leite for useful discussions.

Figures

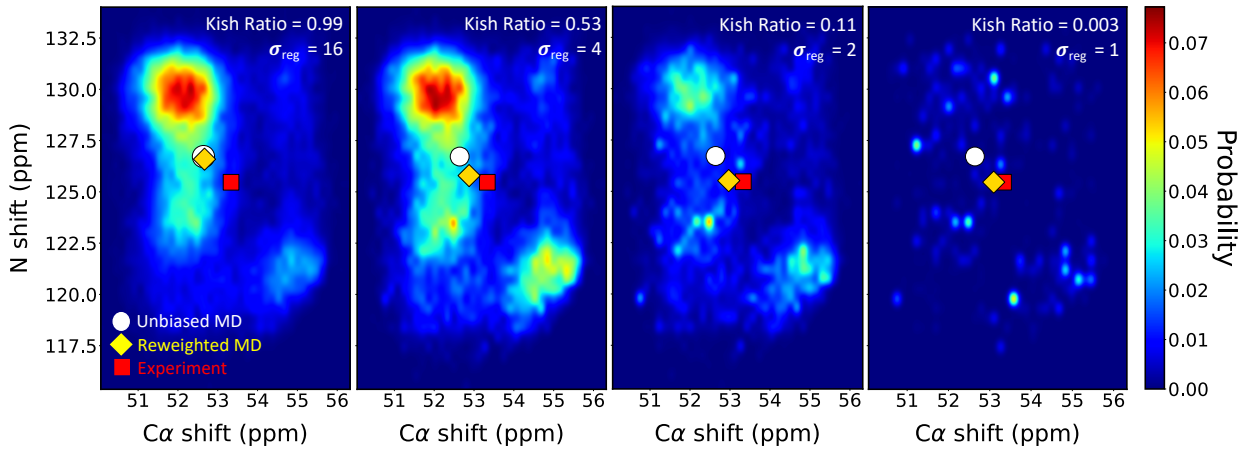

Figure 1: Reweighted IDP ensembles become less representative of the most populated conformational states sampled in unbiased MD simulations as the weight of experimental restraints is increased. We performed maximum entropy reweighting of a 30μs unbiased a99SB-*disp* MD ensemble of the ParE2 associated antitoxin protein (PaaA2) using the backbone $C\alpha$ and N NMR chemical shifts of residue ALA38 as restraints. We compare the distribution of the predicted values of these two chemical shifts in each reweighted ensemble as a function of the value of regularization parameter σ_{reg} of these restraints. Smaller values of σ_{reg} correspond to larger weights of experimental restraints. The calculated values of the restrained chemical shifts in each reweighted ensemble are shown as yellow diamonds. The calculated values of the chemical shifts in the unbiased MD ensemble are shown as a white circle and the experimental values of the chemical shifts are shown as a red square.

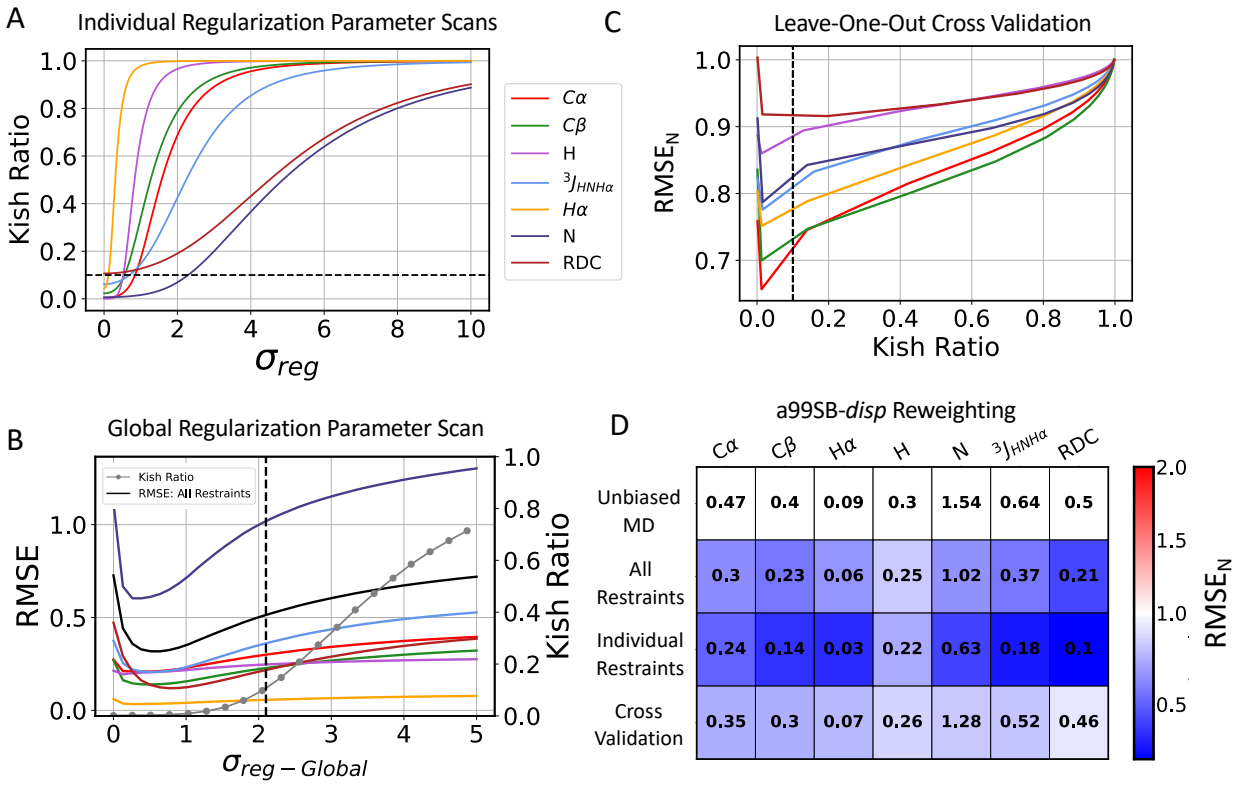

Figure 2: Maximum entropy reweighting of an a99SB-*disp* MD ensemble of Aβ40. A) The effective ensemble size of reweighted ensembles obtained using individual experimental data types as restraints as a function of the value of the regularization parameter σ_{reg}. **B)** Root mean square error (RMSE) between calculated and experimental data in reweighted ensembles as a function of the value of the global regularization parameter scaling factor ($\sigma_{reg-Global}$). The Kish ratio of each reweighted ensemble is indicated by gray dots. **C)** The normalized RMSE (RMSE_N) of unrestrained data in leave-one-out cross-validation tests where all other experimental data types are used as restraints for reweighting. RMSE_N values of each data type are normalized by the RMSE observed in the unbiased a99SB-*disp* Aβ40 MD ensemble. Black dotted lines indicate the Kish ratio threshold $K = 0.10$ in each plot. **D)** RMSE of experimental data in reweighted ensembles obtained by reweighting the a99SB-*disp* MD ensemble with different combinations of restraints. RMSEs are displayed for the unbiased MD ensemble ("Unbiased MD"), a reweighted ensemble obtained using all experimental data as restraints ("All Restraints"), reweighted ensembles obtained using each individual data type as the only restraint ("Individual Restraints") and reweighted ensembles obtained using all but one data type as restraints ("Cross Validation"). All reweighted ensembles were selected using a Kish ratio threshold of $K = 0.10$. Each square is colored to reflect the value of RMSE_N relative to the unbiased a99SB-*disp* MD ensemble.

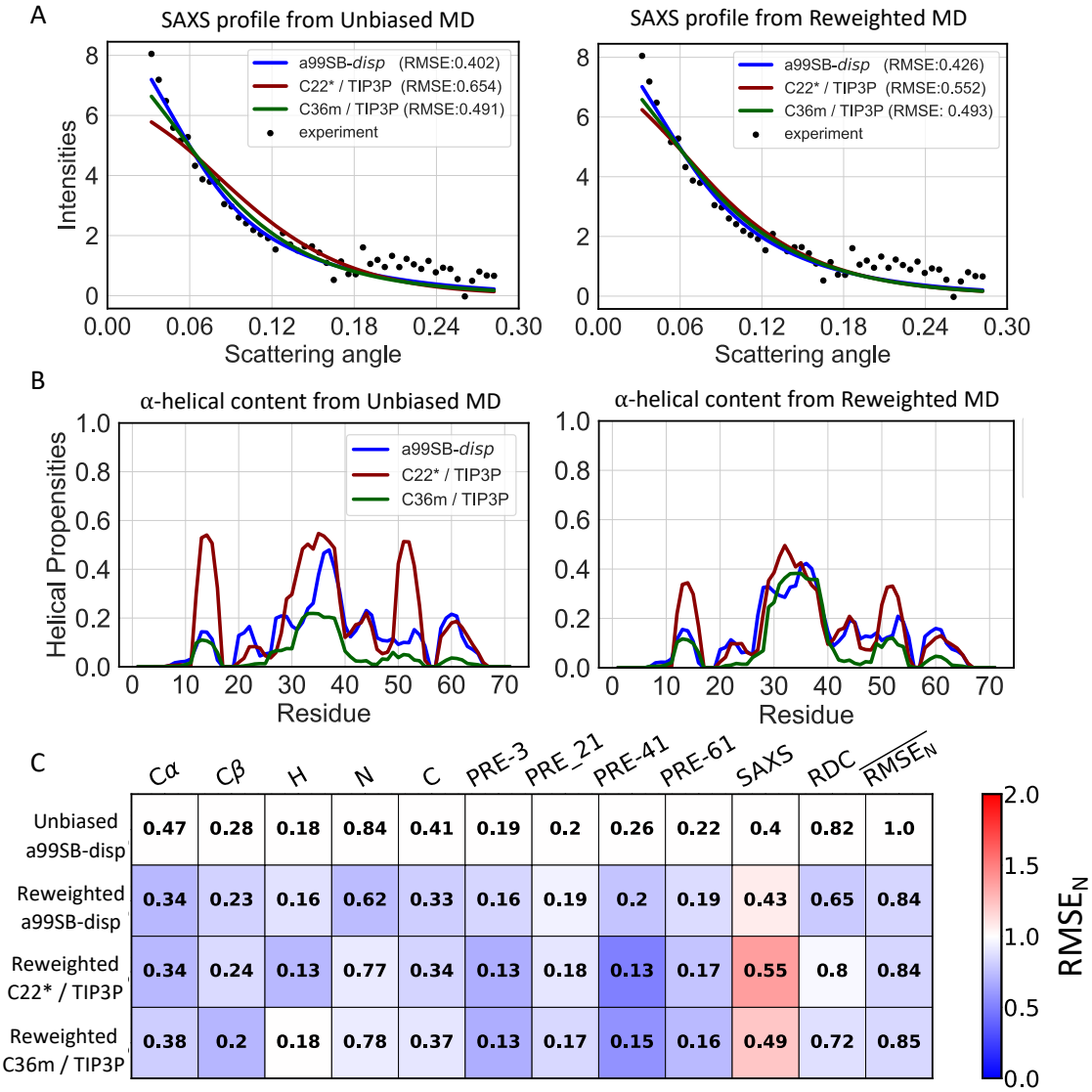

Figure 3: Comparison of unbiased and reweighted MD ensembles of ACTR. A) Comparison of SAXS profiles calculated from unbiased MD ensembles and reweighted MD ensembles of ACTR. **B)** Populations of α-helical conformations in unbiased MD ensembles and reweighted MD ensembles of ACTR. **C)** Comparison of the RMSE between calculated and experimental data in reweighted ACTR ensembles derived from different force fields. Reweighted ensembles were calculated using all experimental data as restraints with a Kish ratio threshold of K=0.10. Each square is colored to reflect the value of $\mathrm{RMSE_N}$ relative to the unbiased a99SB-*disp* MD ensemble.

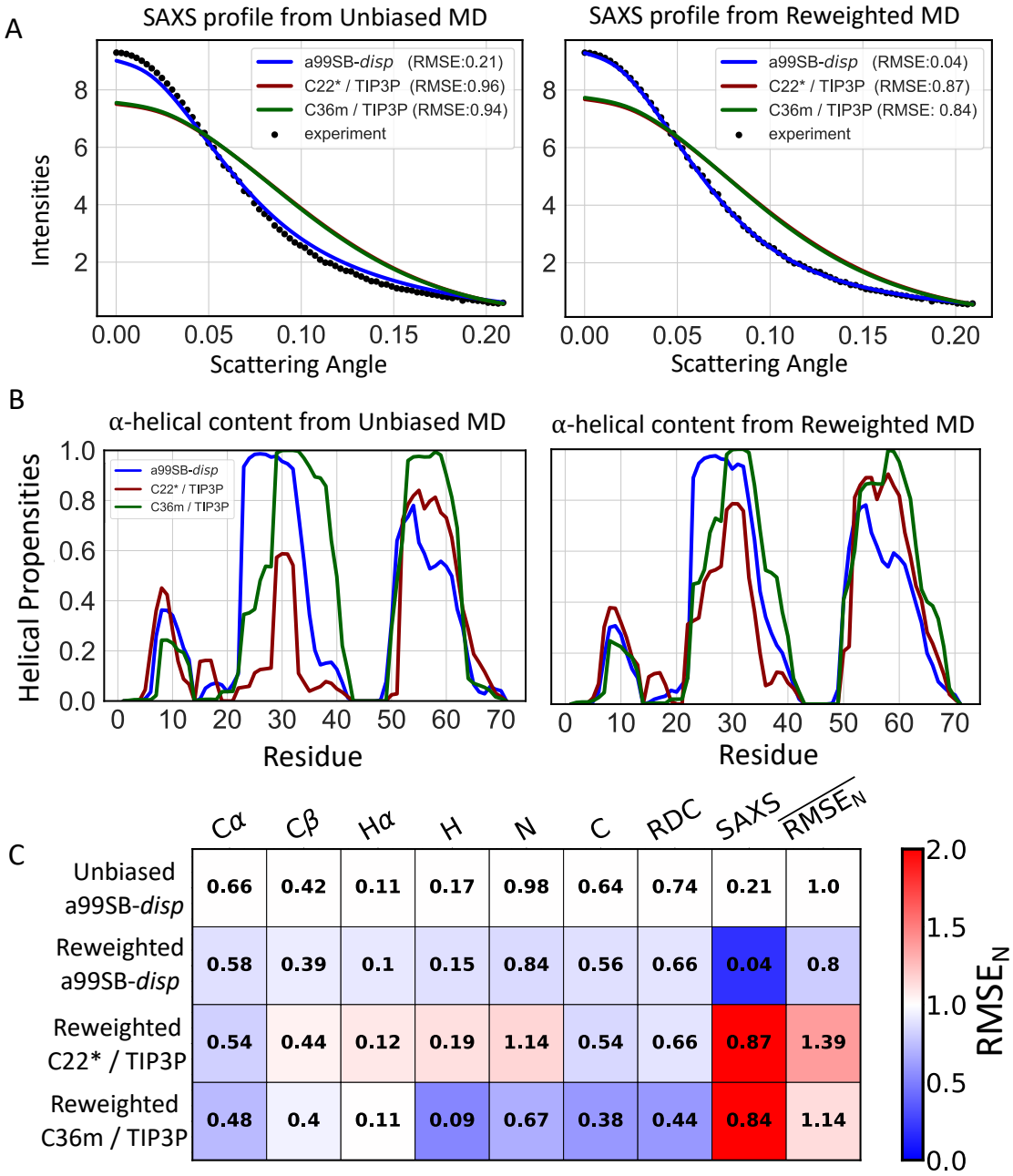

Figure 4: Comparison of unbiased and reweighted MD ensembles of PaaA2. A) Comparison of SAXS profiles calculated from unbiased MD ensembles and reweighted MD ensembles of PaaA2. **B)** Populations of α-helical conformations in unbiased MD ensembles and reweighted MD ensembles of PaaA2. **C)** Comparison of the RMSE between calculated and experimental data in reweighted PaaA2 ensembles derived from different force fields. Reweighted ensembles were calculated using all experimental data as restraints with a Kish ratio threshold of K=0.10. Each square is colored to reflect the value of $\mathrm{RMSE_N}$ relative to the unbiased a99SB-*disp* MD ensemble.

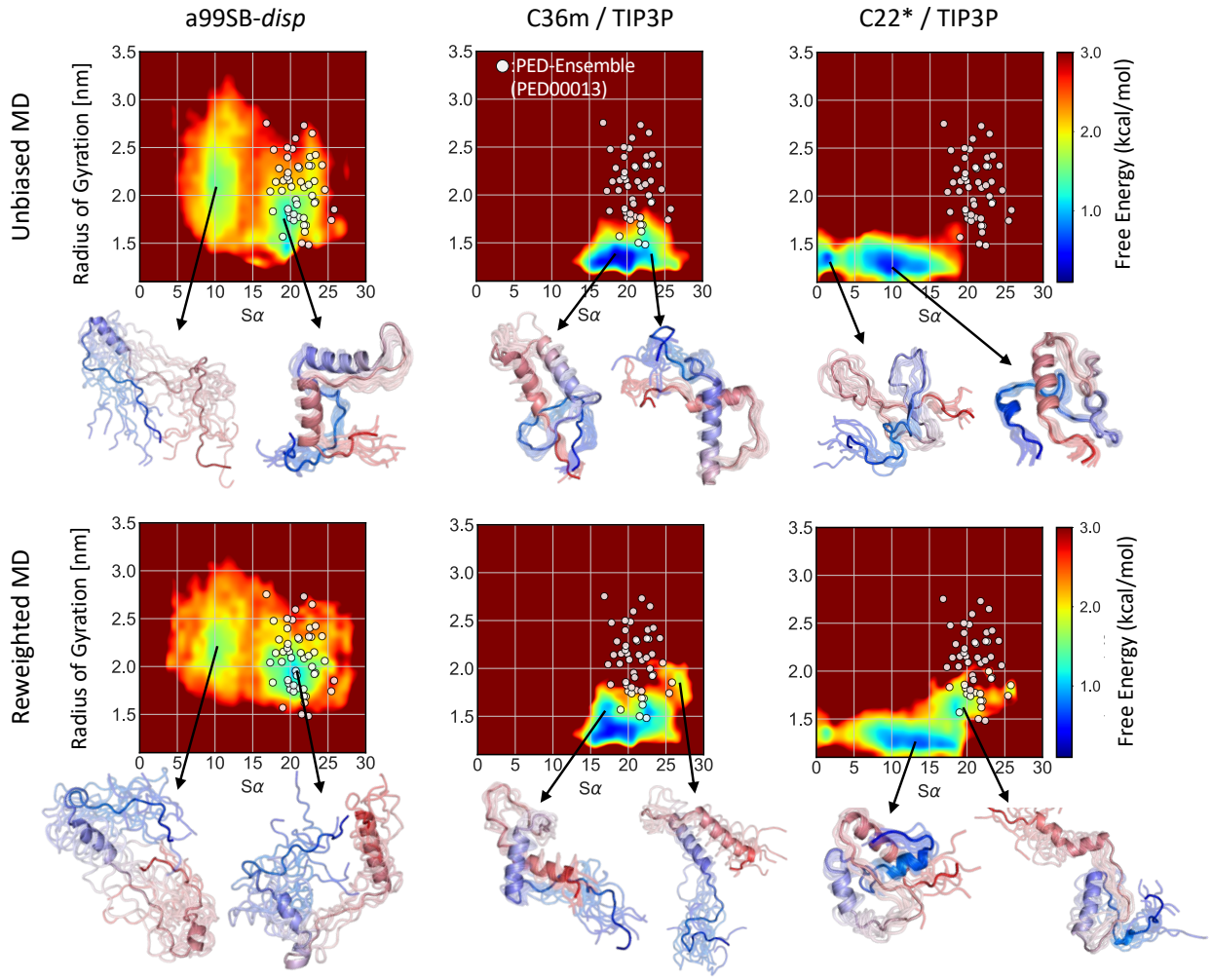

Figure 5: Comparison of free energy surfaces of unbiased and reweighted MD ensembles of PaaA2. Free energy surfaces of unbiased MD ensembles (top row) and reweighted ensembles (bottom row) of PaaA2 are shown as a function of the radius of gyration (R_g) and α-helical order parameter S_α of each conformation. The white dots represent the R_g and S_α values of structures from a PaaA2 protein ensemble database (PED) ensemble calculated from experimental NMR and SAXS data.[51] Representative structures are shown for selected free energy basins of each ensemble. Structures are colored with a blue-to-red gradient from the N-terminus to the C-terminus.

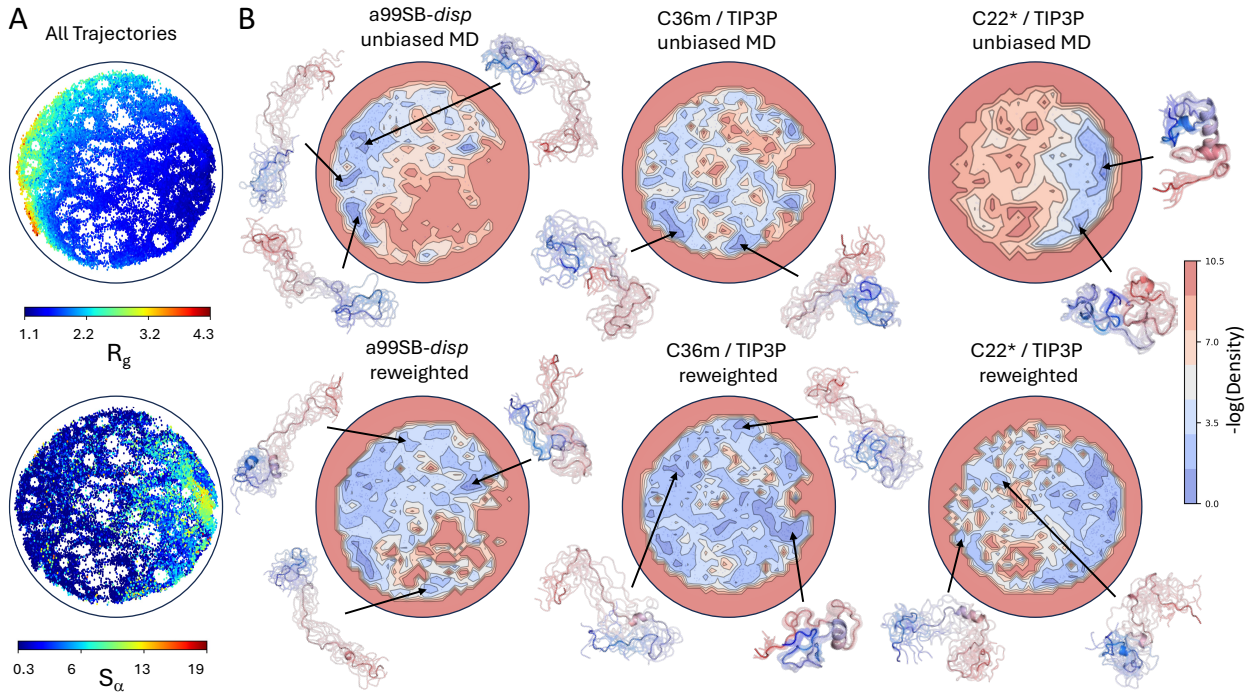

Figure 6: Comparison of energy landscape visualization method (ELViM) projections of unbiased and reweighted MD ensembles of ACTR. A 2D ELViM embedding was calculated using all conformations from unbiased a99SB-*disp*, C22* and C36m MD ensembles of ACTR. **A)** The values of the radius of gyration (R_g) and the α-helical order parameter S_α of each conformation in the unbiased MD ensembles of ACTR projected onto the ELViM latent space. **B)** Comparison of the density of ELViM projections of unbiased and reweighted MD ensembles of ACTR. Representative structures are shown for regions of high density in ELViM projections of each ensemble. Structures are colored with a blue-to-red gradient from the N-terminus to the C-terminus.

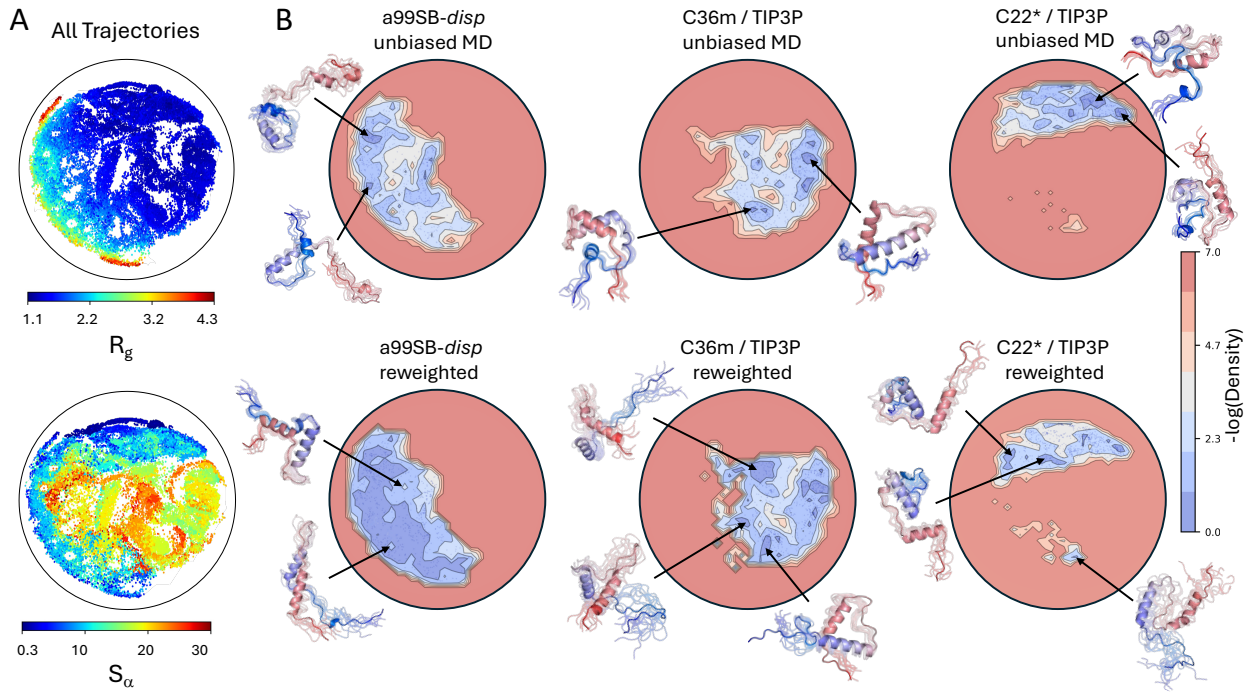

Figure 7: Comparison of energy landscape visualization method (ELViM) projections of unbiased and reweighted MD ensembles of PaaA2. A 2D ELViM embedding was calculated using all conformations from unbiased a99SB-*disp*, C22* and C36m MD ensembles of PaaA2. **A)** The values of the radius of gyration (R_g) and the α-helical order parameter S_α of each conformation in the unbiased MD ensembles of PaaA2 projected onto the ELViM latent space. **B)** Comparison of the density of ELViM projections of unbiased and reweighted MD ensembles of PaaA2. Representative structures are shown for regions of high density in ELViM projections of each ensemble. Structures are colored with a blue-to-red gradient from the N-terminus to the C-terminus.

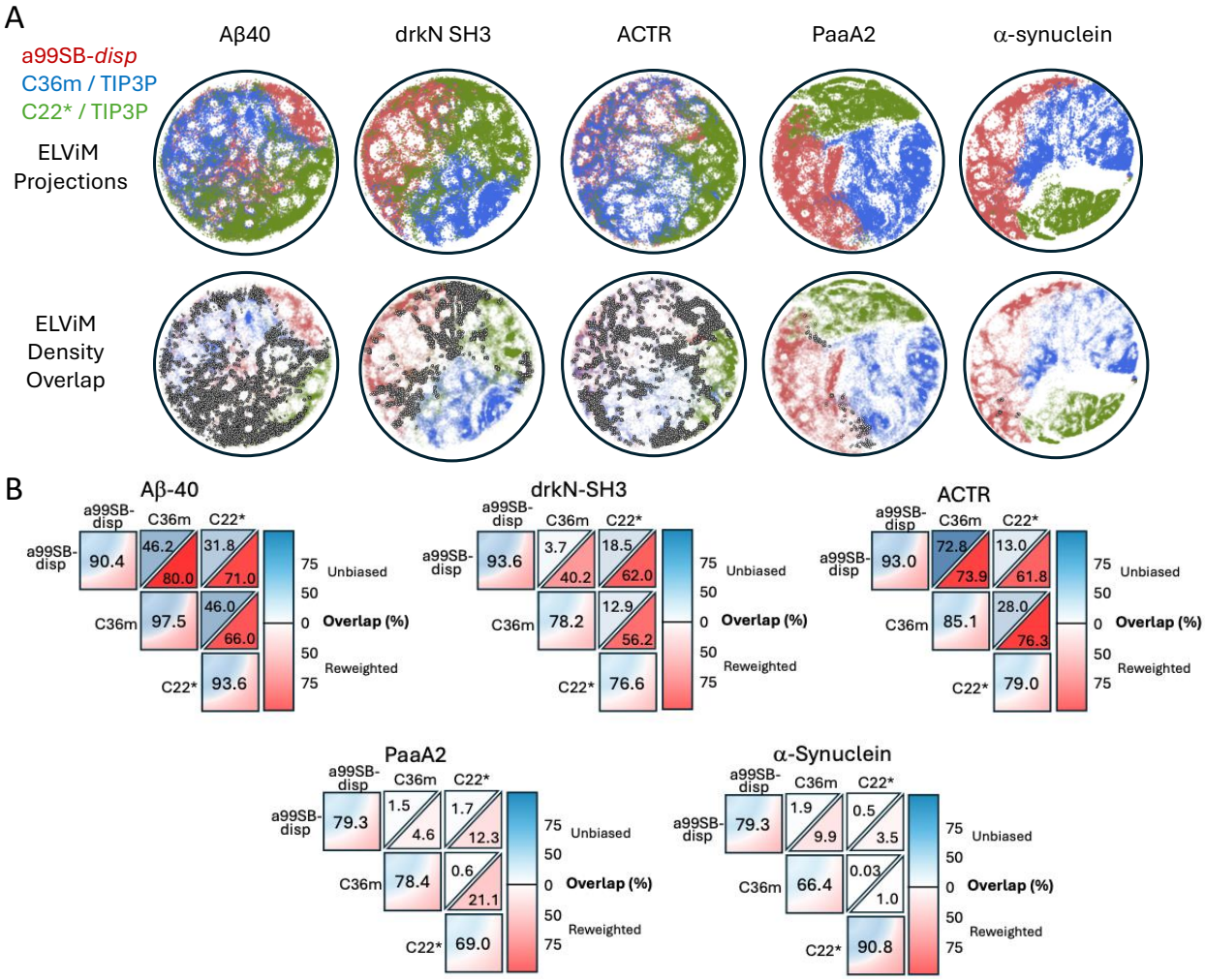

Figure 8: Comparison of the overlap of energy landscape visualization method (ELViM) projections of unbiased and reweighted MD ensembles. A) 2D ELViM embeddings were calculated for Aβ40, drkN SH3, ACTR, PaaA2 and α-synuclein. ELViM embeddings were calculated for each protein using all conformations from unbiased a99SB-*disp*, C22* and C36m MD ensembles. Projections of unbiased a99SB-*disp*, C22* and C36m MD ensembles on ELViM latent space are shown in the top row. In the bottom row, each point is colored with an opacity proportional to its statistical weight after reweighting with all available experimental data. More transparent points corresponding to structures lower statistical weights. Histograms were estimated for each projection using 60 bins in each dimension. Data points belonging to histogram bins that contain at least one data point from each force field are colored dark gray. B) Overlap of ELViM latent space kernel densities of unbiased and reweighted IDP ensembles. Values in the blue triangles reflect the ELViM density overlap of unbiased MD ensembles derived from different force fields, values in red triangles reflect the ELViM density overlap of reweighted ensembles derived from different force fields, and the diagonal elements reflect the ELViM density overlap of the reweighted ensemble derived from each force field with its parent unbiased MD ensemble.

Code Availability

All code used to calculate experimental data from MD trajectories, perform reweighting, analyze ensembles, and compare energy landscape visualization method (ELViM) projections[71–73] is freely available from the GitHub repository `https://github.com/paulrobustelli/Borthakur_MaxEnt_IDPs_2024/`. The ELViM code was adapted from the GitHub repository `https://github.com/VLeiteGroup/ELViM/`.

Data Availability

All experimental data used to reweight MD trajectories, the predicted values of experimental data calculated from each trajectory, and values of the structural descriptors used to compare ensembles are freely available from the GitHub repository (`https://github.com/paulrobustelli/Borthakur_MaxEnt_IDPs_2024/`). All previously reported[19] MD trajectories analyzed in this work are available for non-commercial use by request from D.E. Shaw Research (Trajectories@DEShawResearch.com)

References

(1) Holehouse, A. S.; Kragelund, B. B. The molecular basis for cellular function of intrinsically disordered protein regions. *Nature Reviews Molecular Cell Biology* **2024**, *25*, 187–211.

(2) Banerjee, P. R.; Holehouse, A. S.; Kriwacki, R.; Robustelli, P.; Jiang, H.; Sobolevsky, A. I.; Hurley, J. M.; Mendell, J. T. Dissecting the biophysics and biology of intrinsically disordered proteins. *Trends in biochemical sciences*

(3) Tesei, G.; Trolle, A. I.; Jonsson, N.; Betz, J.; Knudsen, F. E.; Pesce, F.; Johansson, K. E.; Lindorff-Larsen, K. Conformational ensembles of the human intrinsically disordered proteome. *Nature* **2024**, *626*, 897–904.

(4) Sisk, T. R.; Robustelli, P. Folding-upon-binding pathways of an intrinsically disordered protein from a deep Markov state model. *Proceedings of the National Academy of Sciences* **2024**, *121*, e2313360121.

(5) Biesaga, M.; Frigolé-Vivas, M.; Salvatella, X. Intrinsically disordered proteins and biomolecular condensates as drug targets. *Current Opinion in Chemical Biology* **2021**, *62*, 90–100.

(6) Zhu, J.; Salvatella, X.; Robustelli, P. Small molecules targeting the disordered transactivation domain of the androgen receptor induce the formation of collapsed helical states. *Nature Communications* **2022**, *13*, 6390.

(7) Robustelli, P.; Ibanez-de Opakua, A.; Campbell-Bezat, C.; Giordanetto, F.; Becker, S.; Zweckstetter, M.; Pan, A. C.; Shaw, D. E. Molecular basis of small-molecule binding to α-synuclein. *Journal of the American Chemical Society* **2022**, *144*, 2501–2510.

(8) Appadurai, R.; Koneru, J. K.; Bonomi, M.; Robustelli, P.; Srivastava, A. Clustering heterogeneous conformational ensembles of intrinsically disordered proteins with

t-distributed stochastic neighbor embedding. *Journal of chemical theory and computation* **2023**, *19*, 4711–4727.

(9) Basu, S.; Martínez-Cristóbal, P.; Frigolé-Vivas, M.; Pesarrodona, M.; Lewis, M.; Szulc, E.; Bañuelos, C. A.; Sánchez-Zarzalejo, C.; Bielskutė, S.; Zhu, J., et al. Rational optimization of a transcription factor activation domain inhibitor. *Nature Structural & Molecular Biology* **2023**, *30*, 1958–1969.

(10) Heller, G. T.; Aprile, F. A.; Michaels, T. C.; Limbocker, R.; Perni, M.; Ruggeri, F. S.; Mannini, B.; Löhr, T.; Bonomi, M.; Camilloni, C., et al. Small-molecule sequestration of amyloid-β as a drug discovery strategy for Alzheimer's disease. *Science advances* **2020**, *6*, eabb5924.

(11) Löhr, T.; Kohlhoff, K.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. A small molecule stabilizes the disordered native state of the Alzheimer's Aβ Peptide. *ACS Chemical Neuroscience* **2022**, *13*, 1738.

(12) Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. Principles of protein structural ensemble determination. *Current opinion in structural biology* **2017**, *42*, 106–116.

(13) Bottaro, S.; Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **2018**, *361*, 355–360.

(14) Kohlhoff, K. J.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M. Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *Journal of the American Chemical Society* **2009**, *131*, 13894–13895.

(15) Robustelli, P.; Stafford, K. A.; Palmer III, A. G. Interpreting protein structural dynamics from NMR chemical shifts. *Journal of the American Chemical Society* **2012**, *134*, 6365–6374.

(16) Shen, Y.; Bax, A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *Journal of biomolecular NMR* **2010**, *48*, 13–22.

(17) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. SHIFTX2: significantly improved protein chemical shift prediction. *Journal of biomolecular NMR* **2011**, *50*, 43–57.

(18) Cavender, C. E.; Case, D. A.; Chen, J. C.-H.; Chong, L. T.; Keedy, D. A.; Lindorff-Larsen, K.; Mobley, D. L.; Ollila, O.; Oostenbrink, C.; Robustelli, P., et al. Structure-based experimental datasets for benchmarking of protein simulation force fields. *arXiv preprint arXiv:2303.11056* **2023**,

(19) 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.

(20) 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.

(21) Piana, S.; Robustelli, P.; Tan, D.; Chen, S.; Shaw, D. E. Development of a force field for the simulation of single-chain proteins and protein–protein complexes. *Journal of chemical theory and computation* **2020**, *16*, 2494–2507.

(22) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods* **2017**, *14*, 71–73.

(23) Best, R. B.; Zheng, W.; Mittal, J. Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association. *Journal of chemical theory and computation* **2014**, *10*, 5113–5124.

(24) Tian, C.; Kasavajhala, K.; Belfon, K. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q., et al. ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *Journal of chemical theory and computation* **2019**, *16*, 528–552.

(25) Olsson, S.; Frellsen, J.; Boomsma, W.; Mardia, K. V.; Hamelryck, T. Inference of structure ensembles of flexible biomolecules from sparse, averaged data. *PloS one* **2013**, *8*, e79439.

(26) Cesari, A.; Gil-Ley, A.; Bussi, G. Combining simulations and solution experiments as a paradigm for RNA force field refinement. *Journal of chemical theory and computation* **2016**, *12*, 6192–6200.

(27) Cesari, A.; Reißer, S.; Bussi, G. Using the maximum entropy principle to combine simulations and solution experiments. *Computation* **2018**, *6*, 15.

(28) Rangan, R.; Bonomi, M.; Heller, G. T.; Cesari, A.; Bussi, G.; Vendruscolo, M. Determination of structural ensembles of proteins: restraining vs reweighting. *Journal of chemical theory and computation* **2018**, *14*, 6632–6641.

(29) Bottaro, S.; Bussi, G.; Kennedy, S. D.; Turner, D. H.; Lindorff-Larsen, K. Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. *Science advances* **2018**, *4*, eaar8521.

(30) Bottaro, S.; Bengtsen, T.; Lindorff-Larsen, K. Integrating molecular simulation and experimental data: a Bayesian/maximum entropy reweighting approach. *Structural bioinformatics: methods and protocols* **2020**, 219–240.

(31) Hummer, G.; Köfinger, J. Bayesian ensemble refinement by replica simulations and reweighting. *The Journal of chemical physics* **2015**, *143*.

(32) Kofinger, J.; Stelzl, L. S.; Reuter, K.; Allande, C.; Reichel, K.; Hummer, G. Efficient ensemble refinement by reweighting. *Journal of chemical theory and computation* **2019**, *15*, 3390–3401.

(33) Köfinger, J.; Hummer, G. Encoding prior knowledge in ensemble refinement. *The Journal of Chemical Physics* **2024**, *160*.

(34) Marsh, J. A.; Forman-Kay, J. D. Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. *Journal of molecular biology* **2009**, *391*, 359–374.

(35) Lincoff, J.; Haghighatlari, M.; Krzeminski, M.; Teixeira, J. M.; Gomes, G.-N. W.; Gradinaru, C. C.; Forman-Kay, J. D.; Head-Gordon, T. Extended experimental inferential structure determination method in determining the structural ensembles of disordered protein states. *Communications chemistry* **2020**, *3*, 74.

(36) Ozenne, V.; Schneider, R.; Yao, M.; Huang, J.-r.; Salmon, L.; Zweckstetter, M.; Jensen, M. R.; Blackledge, M. Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution. *Journal of the American Chemical Society* **2012**, *134*, 15138–15148.

(37) Schwalbe, M.; Ozenne, V.; Bibow, S.; Jaremko, M.; Jaremko, L.; Gajda, M.; Jensen, M. R.; Biernat, J.; Becker, S.; Mandelkow, E., et al. Predictive atomic resolution descriptions of intrinsically disordered hTau40 and α-synuclein in solution from NMR and small angle scattering. *Structure* **2014**, *22*, 238–249.

(38) Shannon, C. E. A mathematical theory of communication. *The Bell system technical journal* **1948**, *27*, 379–423.

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

(40) Pitera, J. W.; Chodera, J. D. On the use of experimental observations to bias simulated ensembles. *Journal of chemical theory and computation* **2012**, *8*, 3445–3451.

(41) Cavalli, A.; Camilloni, C.; Vendruscolo, M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *The Journal of chemical physics* **2013**, *138*.

(42) White, A. D.; Voth, G. A. Efficient and minimal method to bias molecular simulations with experimental data. *Journal of chemical theory and computation* **2014**, *10*, 3023–3030.

(43) Bonomi, M.; Camilloni, C.; Cavalli, A.; Vendruscolo, M. Metainference: A Bayesian inference method for heterogeneous systems. *Science advances* **2016**, *2*, e1501177.

(44) Bonomi, M.; Camilloni, C.; Vendruscolo, M. Metadynamic metainference: Enhanced sampling of the metainference ensemble using metadynamics. *Scientific reports* **2016**, *6*, 31232.

(45) Crehuet, R.; Buigues, P. J.; Salvatella, X.; Lindorff-Larsen, K. Bayesian-maximum-entropy reweighting of IDP ensembles based on NMR chemical shifts. *Entropy* **2019**, *21*, 898.

(46) Janson, G.; Feig, M. Transferable deep generative modeling of intrinsically disordered protein conformations. *PLOS Computational Biology* **2024**, *20*, e1012144.

(47) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J., et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500.

(48) Hou, L.; Shao, H.; Zhang, Y.; Li, H.; Menon, N. K.; Neuhaus, E. B.; Brewer, J. M.; Byeon, I.-J. L.; Ray, D. G.; Vitek, M. P., et al. Solution NMR studies of the Aβ (1- 40)

and Aβ (1- 42) peptides establish that the Met35 oxidation state affects the mechanism of amyloid formation. *Journal of the American Chemical Society* **2004**, *126*, 1992–2005.

(49) Zhang, O.; Forman-Kay, J. D. Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer. *Biochemistry* **1995**, *34*, 6784–6794.

(50) Iesmantavicius, V.; Jensen, M. R.; Ozenne, V.; Blackledge, M.; Poulsen, F. M.; Kjaergaard, M. Modulation of the Intrinsic helix propensity of an intrinsically disordered protein reveals long-range helix–helix interactions. *Journal of the American Chemical Society* **2013**, *135*, 10155–10163.

(51) Sterckx, Y. G.; Volkov, A. N.; Vranken, W. F.; Kragelj, J.; Jensen, M. R.; Buts, L.; Garcia-Pino, A.; Jové, T.; Van Melderen, L.; Blackledge, M., et al. Small-angle X-ray scattering-and nuclear magnetic resonance-derived conformational ensemble of the highly flexible antitoxin PaaA2. *Structure* **2014**, *22*, 854–865.

(52) Bertoncini, C. W.; Jung, Y.-S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. Release of long-range tertiary interactions potentiates aggregation of natively unstructured α-synuclein. *Proceedings of the National Academy of Sciences* **2005**, *102*, 1430–1435.

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

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

(55) Rao, J. N.; Kim, Y. E.; Park, L. S.; Ulmer, T. S. Effect of pseudorepeat rearrangement on α-synuclein misfolding, vesicle binding, and micelle binding. *Journal of molecular biology* **2009**, *390*, 516–529.

(56) Marsh, J. A.; Baker, J. M.; Tollinger, M.; Forman-Kay, J. D. Calculation of residual dipolar couplings from disordered state ensembles using local alignment. *Journal of the American Chemical Society* **2008**, *130*, 7804–7805.

(57) Yan, Y.; McCallum, S. A.; Wang, C. M35 oxidation induces Aβ40-like structural and dynamical changes in Aβ42. *Journal of the American Chemical Society* **2008**, *130*, 5394–5395.

(58) Roche, J.; Shen, Y.; Lee, J. H.; Ying, J.; Bax, A. Monomeric Aβ1–40 and Aβ1–42 peptides in solution adopt very similar Ramachandran map distributions that closely resemble random coil. *Biochemistry* **2016**, *55*, 762–775.

(59) Lee, J. H.; Li, F.; Grishaev, A.; Bax, A. Quantitative residue-specific protein backbone torsion angle dynamics from concerted measurement of 3 J couplings. *Journal of the American Chemical Society* **2015**, *137*, 1432–1435.

(60) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *Journal of the American Chemical Society* **2005**, *127*, 476–477.

(61) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. NMR characterization of long-range order in intrinsically disordered proteins. *Journal of the American Chemical Society* **2010**, *132*, 8407–8418.

(62) Sung, Y.-h.; Eliezer, D. Residual structure, backbone dynamics, and interactions within the synuclein family. *Journal of molecular biology* **2007**, *372*, 689–707.

(63) Esteban-Martín, S.; Silvestre-Ryan, J.; Bertoncini, C. W.; Salvatella, X. Identification of fibril-like tertiary contacts in soluble monomeric α-synuclein. *Biophysical journal* **2013**, *105*, 1192–1198.

(64) 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.

(65) Kjaergaard, M.; Nørholm, A.-B.; Hendus-Altenburger, R.; Pedersen, S. F.; Poulsen, F. M.; Kragelund, B. B. Temperature-dependent structural changes in intrinsically disordered proteins: Formation of α–helices or loss of polyproline II? *Protein Science* **2010**, *19*, 1555–1564.

(66) Kish, L. Survey sampling. 1965.

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

(68) Pietrucci, F.; 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*, 2197–2201.

(69) Wood, G. P.; Rothlisberger, U. Secondary structure assignment of amyloid-β peptide using chemical shifts. *Journal of chemical theory and computation* **2011**, *7*, 1552–1563.

(70) Ghafouri, H.; Lazar, T.; Del Conte, A.; Tenorio Ku, L. G.; Tompa, P.; Tosatto, S. C.; Monzon, A. M. PED in 2024: improving the community deposition of structural ensembles for intrinsically disordered proteins. *Nucleic acids research* **2024**, *52*, D536–D544.

(71) Oliveira Jr, A. B.; Yang, H.; Whitford, P. C.; Leite, V. B. Distinguishing biomolecular pathways and metastable states. *Journal of chemical theory and computation* **2019**, *15*, 6482–6490.

(72) Viegas, R. G.; Martins, I. B.; Leite, V. B. Understanding the Energy Landscape of

Intrinsically Disordered Protein Ensembles. *Journal of Chemical Information and Modeling* **2024**, *64*, 4149–4157.

(73) Viegas, R. G.; Martins, I. B.; Sanches, M. N.; Oliveira Junior, A. B.; Camargo, J. B. d.; Paulovich, F. V.; Leite, V. B. ELViM: Exploring Biomolecular Energy Landscapes through Multidimensional Visualization. *Journal of Chemical Information and Modeling* **2024**, *64*, 3443–3450.

(74) van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **2008**, *9*, 2579–2605.

(75) Teixeira, J. M.; Liu, Z. H.; Namini, A.; Li, J.; Vernon, R. M.; Krzeminski, M.; Shamandy, A. A.; Zhang, O.; Haghighatlari, M.; Yu, L., et al. IDPConformerGenerator: a flexible software suite for sampling the conformational space of disordered protein states. *The Journal of Physical Chemistry A* **2022**, *126*, 5985–6003.

(76) Bernadó, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society* **2007**, *129*, 5656–5664.

(77) 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.

(78) Joseph, J. A.; Reinhardt, A.; Aguirre, A.; Chew, P. Y.; Russell, K. O.; Espinosa, J. R.; Garaizar, A.; Collepardo-Guevara, R. Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nature Computational Science* **2021**, *1*, 732–743.

(79) Janson, G.; Valdes-Garcia, G.; Heo, L.; Feig, M. Direct generation of protein conformational ensembles via machine learning. *Nature Communications* **2023**, *14*, 774.

(80) Lotthammer, J. M.; Ginell, G. M.; Griffith, D.; Emenecker, R.; Holehouse, A. S. Direct prediction of intrinsically disordered protein conformational properties from sequence. *Biophysical Journal* **2024**, *123*, 43a.

(81) Pesce, F.; Bremer, A.; Tesei, G.; Hopkins, J. B.; Grace, C. R.; Mittag, T.; Lindorff-Larsen, K. Design of intrinsically disordered protein variants with diverse structural properties. *Science Advances* **2024**, *10*, eadm9926.

(82) Burley, S. K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chao, H.; Chen, L.; Craig, P. A.; Crichlow, G. V.; Dalenberg, K.; Duarte, J. M., et al. RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic acids research* **2023**, *51*, D488–D508.

(83) Pitera, J. W.; Chodera, J. D. On the Use of Experimental Observations to Bias Simulated Ensembles. *Journal of Chemical Theory and Compututation* **2012**, *8*, 3445—3451.

(84) Roux, B.; Jonathan, W. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *Journal of Chemical Physics* **2013**, *138*, 084107.

(85) Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Combining Experiments and Simulations Using the Maximum Entropy Principle. *PLOS Computational Biology* **2014**, *10*, e1003406.

(86) Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J., et al. Millisecond-scale molecular dynamics simulations on Anton. Proceedings of the conference on high performance computing networking, storage and analysis. 2009; pp 1–11.

(87) Karplus, M. Contact electron-spin coupling of nuclear magnetic moments. *The Journal of chemical physics* **1959**, *30*, 11–15.

(88) Li, F.; Lee, J. H.; Grishaev, A.; Ying, J.; Bax, A. High Accuracy of Karplus Equations for Relating Three-Bond J Couplings to Protein Backbone Torsion Angles. *ChemPhysChem* **2015**, *16*, 572–578.

(89) Vögeli, B.; Ying, J.; Grishaev, A.; Bax, A. Limits on variations in protein backbone dynamics from precise measurements of scalar couplings. *Journal of the American Chemical Society* **2007**, *129*, 9377–9385.

(90) Zweckstetter, M.; Bax, A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *Journal of the American Chemical Society* **2000**, *122*, 3791–3792.

(91) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophysical journal* **2015**, *109*, 1528–1532.

(92) 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: Structural Biology* **2017**, *73*, 449–464.

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

(94) Developers, S. SciPy: Open Source Scientific Tools for Python. 2023; `https://www.scipy.org/`.

(95) Scott, D. *Multivariate Density Estimation: Theory, Practice, and Visualization*; John Wiley & Sons: New York, Chichester, 1992.