



## **Determination of Structural Ensembles of Proteins: Restraining vs Reweighting**

Ramya Rangan, Massimiliano Bonomi, Gabriella Heller, Andrea Cesari, Giovanni Bussi, Michele Vendruscolo

### **► To cite this version:**

Ramya Rangan, Massimiliano Bonomi, Gabriella Heller, Andrea Cesari, Giovanni Bussi, et al.. Determination of Structural Ensembles of Proteins: Restraining vs Reweighting. *Journal of Chemical Theory and Computation*, 2018, 14 (12), pp.6632-6641. <10.1021/acs.jctc.8b00738>. <hal-03136516>

**HAL Id: hal-03136516**

**<https://hal.science/hal-03136516v1>**

Submitted on 9 Feb 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Determination of Structural Ensembles of Proteins: Restraining vs Reweighting

Ramya Rangan<sup>†§</sup>, Massimiliano Bonomi<sup>\*††</sup>, Gabriella T. Heller<sup>†</sup>,  
Andrea Cesari<sup>‡</sup>, Giovanni Bussi<sup>‡</sup>, and Michele Vendruscolo<sup>\*†</sup>

<sup>†</sup>*Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK*

<sup>‡</sup>*Scuola Internazionale Superiore di Studi Avanzati (SISSA), 34136 Trieste, Italy*

## Abstract

The conformational fluctuations of proteins can be described by structural ensembles. To address the major challenge of determining these ensembles accurately, a wide range of strategies have recently been proposed to combine molecular dynamics simulations with experimental data. Quite generally, there are two ways of implementing this type of approach, either by applying structural restraints during a simulation, or by reweighting *a posteriori* the conformations from an *a priori* ensemble. It is not yet clear, however, whether these two approaches can offer ensembles of equivalent quality. The advantages of the reweighting method are that it can involve any type of starting simulation and that it enables the integration of experimental data after the simulations are run. A disadvantage, however, is that this procedure may be inaccurate when the *a priori* ensemble is of poor quality. Here, our goal is to systematically compare the restraining and reweighting approaches and to explore the conditions required for the reweighting ensembles to be accurate. Our results indicate that the reweighting approach is computationally efficient and can perform as well as the restraining approach when the *a priori* sampling is accurate. More generally, to enable an effective use of the reweighting approach by avoiding the pitfalls of poor sampling, we suggest metrics for the quality control of the reweighted ensembles.

## Introduction

It is increasingly recognized that proteins populate different conformations while performing their functions (Mittermaier, Anthony, and Lewis E. Kay. "New tools provide new insights in NMR studies of protein dynamics." *Science* 312.5771 (2006): 224-228).<sup>1-2</sup> For instance, membrane channels open and close to transport ions, enzymes twist to incorporate substrates and release products, and receptors adapt to bind signaling molecules.<sup>3-5</sup> Moreover, many proteins are intrinsically disordered, as they adopt large ensembles of heterogeneous conformations in their native states.<sup>6-8</sup> To understand the molecular basis of protein function, it is thus often critical to not only determine a single structure, but also to accurately characterize a set of conformational states along with their probabilities of occupation.<sup>9</sup> (Bonomi and Vendruscolo, Determination of protein structural ensembles using cryo-electron microscopy. *Curr. Op. Struct. Biol.* in press)

A major problem, however, is that experimental measurements cannot directly determine these structural ensembles (Fig. 1a, green curve) for a variety of technical reasons, including that in many cases these measurements yield data averaged over the ensembles themselves (Fig. 1a, black dashed line).<sup>9-10</sup> Molecular dynamics (MD) simulations provide a complementary tool for characterizing these ensembles, although one should consider that their accuracy is limited by the quality of the force fields, or ‘prior’ (Fig. 1a, red curve and red dashed line) (Bonomi and Vendruscolo, Determination of protein structural ensembles using cryo-electron microscopy. *Curr. Op. Struct. Biol.* in press). By incorporating simulations and experiments into integrative models, one can generate ensembles more consistent with experimental data while remaining close to the force fields (Fig. 1a, blue curve).

Various methods have emerged for integrating experimental information with MD simulations.<sup>9, 11-13</sup> (Bonomi and Vendruscolo, Determination of protein structural ensembles using cryo-electron microscopy. *Curr. Op. Struct. Biol.* in press) Maximum entropy approaches have been designed to minimally modify MD force fields to match experimental data.<sup>14-15</sup> To induce a better agreement with experimental data, some maximum entropy approaches known as ‘restraining’ methods include additional forces during the simulations.<sup>16-17</sup> Some of these approaches such as the metainference method include also error models that help prevent overfitting to noisy experimental data.<sup>18-19</sup>

As an alternative strategy, it is possible to first produce an ensemble from unrestrained simulations, and to change the statistical weights of the resulting configurations to obtain ensemble averages that better match experimental data. This ‘reweighting’ strategy has been recently implemented in some ensemble refinement methods<sup>20-22</sup>, providing important advantages. In particular, it can be computationally impractical to run simulations of multiple replicas, as required by some restraining methods.<sup>18</sup> In addition, restraining methods based on Lagrange multipliers require a careful choice of the learning parameters to converge quickly.<sup>19</sup> Furthermore, reweighting methods allow for the incorporation of additional experimental data after a simulation has finished. In addition, reweighting methods can also enable the use of more complex and computationally expensive forward models, which are used to approximate experimental data given a protein conformation. In particular, since restraining methods involve evaluating the forward model during a simulation, they are restricted to models that are computationally tractable with gradients that are reasonable to evaluate. However, the reweighting approach has limitations when the prior and the true ensembles overlap poorly (Fig. 1b).<sup>9, 23</sup>

Here, we compare the restraining and reweighting approaches, and we investigate the conditions in which the reweighting approach can improve an MD ensemble. To carry out a systematic analysis we applied a reweighting method on two systems, dialanine with interatomic distances for reweighting, and a peptide from the binding region of the oncogenic protein c-Myc (residues 402 to 412, referred to as c-Myc<sub>402-412</sub>) with NMR chemical shifts for reweighting.

## Methods

***Derivation of reweighting linear biasing potential.*** We use the restraining method of Cesari et al.<sup>19</sup> in a reweighting mode<sup>24</sup>, as it allows for the specification of an error model for the data. Importantly, in this approach a maximum entropy ensemble can be derived, while requiring that the reweighted ensemble matches the experimental data when including an error term. Here we discuss how the biasing potential for reweighting a conformational ensemble with experimental data arises from the maximum entropy principle, as introduced by Jaynes<sup>25</sup>.

A molecular dynamics simulation can be viewed as producing a distribution  $\mathbf{P}_0(\mathbf{x})$  of conformations  $\mathbf{x}$ . In a reweighting approach the goal is to reweight the ensemble  $\mathbf{P}_0(\mathbf{x})$  to a new en-

semble  $\mathbf{P}_1(\mathbf{x})$  that better matches the experimental data  $\mathbf{f}_i^{exp}$  for each experiment  $i$ , which is expected to result from an ensemble average over the underlying true ensemble. To determine the fit to experimental data, we use a forward model  $\mathbf{f}_i(\mathbf{x})$  which predicts the value of the observable  $i$  for each conformation  $\mathbf{x}$ . In the method that we implemented here, we require that  $\mathbf{P}_1(\mathbf{x})$  satisfies the ensemble-averaged experimental observables and minimizes the Kullback-Leibler (KL) divergence of  $\mathbf{P}_1(\mathbf{x})$  to  $\mathbf{P}_0(\mathbf{x})$ , causing a minimal shift of the prior to match the experimental data. This strategy leads to the following constrained optimization problem, where we maximize  $S$  when constraints  $\mathbf{c}_i$  are set to zero:

$$S = - \int P_1(x) \ln \left( \frac{P_1(x)}{P_0(x)} \right) dx$$

$$c_i = \int P_1(x) f_i(x) dx - f_i^{exp}$$

With Lagrange multipliers and normalization, we arrive at the ensemble  $\mathbf{P}_1(\mathbf{x})$  that shifts the ensemble  $\mathbf{P}_0(\mathbf{x})$  with a biasing potential linear in the experimental observables.<sup>14, 19, 24</sup>

$$P_1(x) = \frac{P_0(x) e^{-\sum_i \lambda_i f_i(x)}}{\int P_0(x) e^{-\sum_i \lambda_i f_i(x)} dx}$$

To find the Lagrange multipliers, we can enforce the constraints  $\mathbf{c}_i$ , finding the parameters  $\lambda_i$  such that the ensemble average  $\langle \mathbf{f}_i(\mathbf{x}) \rangle$  on the new ensemble equals the experimental value. As described by Cesari, et al.<sup>19</sup>, we can incorporate experimental and forward model errors into the linear biasing potential by modifying  $\mathbf{c}_i$  to include error terms  $\boldsymbol{\varepsilon}_i$  with a distribution  $\mathbf{P}(\boldsymbol{\varepsilon}_i)$ . Then, we must find  $\lambda_i$  with  $\langle \mathbf{f}_i(\mathbf{x}) \rangle + \langle \boldsymbol{\varepsilon}_i \rangle = \mathbf{f}_i^{exp}$  where  $\langle \boldsymbol{\varepsilon}_i \rangle$  is the average of the error terms  $\boldsymbol{\varepsilon}_i$  on distribution  $\mathbf{P}(\boldsymbol{\varepsilon}_i) e^{-\sum_i \lambda_i \varepsilon_i}$ .

The reweighting approach that we use here is closely related to the Bayesian ensemble refinement method.<sup>22</sup> In that method, an additional parameter is introduced to take into account the confidence in the prior distribution; this parameter enters as a global scaling factor in the errors for each data point. The errors used here can similarly be used to modulate both our confidence in the experimental data and our confidence in the original force field. Cesari, et al.<sup>24</sup> include further discussion on the relationship between the Bayesian ensemble refinement method and the reweighting approach presented in this paper.

**Evaluation of the error terms for the reweighting algorithm.** We explore two error models for the experimental data. First, we use a Gaussian error model with  $P(\boldsymbol{\varepsilon}_i) \propto \text{Norm}(\mathbf{0}, \boldsymbol{\sigma}_i)$ . This error model is especially useful when the variance of the experimental measurement and forward model is known. In this case, the error term becomes

$$\langle \boldsymbol{\varepsilon}_i \rangle = -\lambda_i \boldsymbol{\sigma}_i^2$$

We also try using an uninformative prior, Jeffreys prior, as an error model. We treat the observables as having Gaussian error as before, with distribution  $\text{Norm}(\mathbf{0}, \boldsymbol{\sigma}_i)$ . However, rather than specifying the variance  $\boldsymbol{\sigma}_i$  explicitly for each observable, we model the variance as arising from Jeffreys prior with  $P(\boldsymbol{\sigma}_i) \propto \frac{1}{\boldsymbol{\sigma}_i}$ . This prior is useful for situations in which we have no prior guess for the experimental error on the observables.

When using Jeffreys prior on the variance directly, the average error diverges. We therefore bound the prior to act on the domain  $\boldsymbol{\sigma}_{min}$  and  $\boldsymbol{\sigma}_{max}$ . We then have the following form for  $\boldsymbol{\varepsilon}_i$ , where  $Ei$  is the exponential integral function

$$\langle \boldsymbol{\varepsilon}_i \rangle = -\frac{2}{\lambda_i} \frac{e^{\frac{\lambda_i^2 \boldsymbol{\sigma}_{max}^2}{2}} - e^{\frac{\lambda_i^2 \boldsymbol{\sigma}_{min}^2}{2}}}{Ei\left(\frac{\lambda_i^2 \boldsymbol{\sigma}_{max}^2}{2}\right) - Ei\left(\frac{\lambda_i^2 \boldsymbol{\sigma}_{min}^2}{2}\right)}$$

We can use these closed form expressions for  $\langle \boldsymbol{\varepsilon}_i \rangle$  to find  $\lambda_i$  with  $\langle \mathbf{f}_i(\mathbf{x}) \rangle + \langle \boldsymbol{\varepsilon}_i \rangle = \mathbf{f}_i^{exp}$ .

**Reweighting as an optimization problem.** The problem of calculating the Lagrange multipliers to satisfy  $\langle \mathbf{f}_i(\mathbf{x}) \rangle + \langle \boldsymbol{\varepsilon}_i \rangle = \mathbf{f}_i^{exp}$  can be formulated as an optimization problem, which can be tackled with any available minimization method. As the objective function, we use the following function, which depends on all the Lagrange multipliers

$$\Gamma(\boldsymbol{\lambda}) = \ln \left( \sum_{\mathbf{x}} P_0(\mathbf{x}) \int d\boldsymbol{\varepsilon} P(\boldsymbol{\varepsilon}) e^{-\sum_i \lambda_i (\mathbf{f}_i(\mathbf{x}) + \boldsymbol{\varepsilon}_i)} \right) + \sum_i \lambda_i \mathbf{f}_i^{exp}$$

The gradient of  $\Gamma(\lambda)$  is  $\langle f_i(x) \rangle + \langle \varepsilon_i \rangle - f_i^{exp}$ , such that minimizing  $\Gamma(\lambda)$  leads to an optimal setting of the Lagrange multipliers. Furthermore, with analytical solutions for each error model, it is tractable to evaluate this objective function for both the Gaussian error model with known variance and the Gaussian error with a bounded uninformative Jeffreys prior.

This objective function behaves in a particularly convenient manner for optimization. It is strictly convex in the cases of uncorrelated observables without error, and it is also strictly convex in the case with Gaussian error even with correlated observables.<sup>19</sup> Furthermore, even when the experimental observables are correlated and an error is not added, all minima of the objective function satisfy  $\langle f_i(x) \rangle + \langle \varepsilon_i \rangle = f_i^{exp}$  and remain equally far from the prior  $P_0(x)$ .<sup>19</sup> All minima in this case are thus equally valid solutions for the reweighted ensemble. In the case of correlated observables without error, we are not guaranteed to be able to find a set of Lagrange multipliers that achieves this optimal value. In practice, this situation is essentially never encountered, as in all real systems one should model some amount of error, and thus the convexity in the error case leads the objective function to be convex in essentially all practical applications. Examples showing what would be the result of such a procedure when the trajectory is not consistent with the data are reported in Cesari, et al.<sup>24</sup>

With a convex objective function, it is not necessary to use techniques such as genetic algorithms and simulated annealing, which are meant to help the optimization procedure explore beyond the current local minimum. Instead, a gradient descent approach is guaranteed to approach a unique minimum. In this approach, we begin with a random guess for parameters  $\lambda_i$ , evaluate  $\langle f_i(x) \rangle + \langle \varepsilon_i \rangle$ , and update our guess for  $\lambda_i$  to  $\lambda_i - \eta(f_i^{exp} - \langle f_i(x) \rangle - \langle \varepsilon_i \rangle)$  for learning the parameter  $\eta$ . When the Lagrange multipliers converge, this iterative procedure is complete, yielding a reweighted ensemble  $P_1(x)$ .

Despite the guarantee of convexity, gradient descent methods can be very slow to reach a minimum. For instance, if the Hessian is not well conditioned, these procedures can oscillate around the optimal Lagrange multiplier setting, with step sizes too large to converge on the minimum. For this reason, we use the conjugate gradient approach as an alternative for determining the Lagrange multipliers.

We note that it is possible to pursue other minimization strategies to achieve further convergence, such as methods that require the calculation of the Hessian like Newton's method. Furthermore, techniques such as mini-batch gradient minimization<sup>26</sup> can aid computational efficiency by operating on randomly chosen smaller batches of the dataset in each iteration, taking in more data for tuning the Lagrange multipliers in further iterations as the parameters approach their optimal value. Depending on the particulars of the ensemble being reweighted and the number of available experimental observables, improving efficiency using these methods may be desirable.

***Molecular dynamics simulations of dialanine.*** As a first test system, we evaluated reweighting on dialanine. By systematically altering the free energy (FE) landscape of dialanine and generating data from these altered FE landscapes, we can assess the ability for reweighting to correctly shift the simulation ensembles towards the true one. Furthermore, we can readily generate long trajectories with this system to test the computational efficiency of reweighting.

To generate a converged ensemble for dialanine *in vacuo*, we ran an MD simulation with a bias potential from a well-tempered metadynamics simulation.<sup>27</sup> We used GROMACS 5.1.4<sup>28</sup> for all simulations, and we accelerated sampling for our first simulation with well-tempered metadynamics (WT-MetaD)<sup>27</sup> as implemented in PLUMED 2.3.1<sup>29</sup>. We used the Amber99SB-ILDN<sup>30</sup> force field, running *in vacuo*. For both simulations, we ran for 20 ns with a time step of 2 fs, generating ~200,000 frames. We used no van der Waals or Coulomb force cutoffs as our system was very small. We kept the temperature of the system at 300 K using temperature coupling with the v-rescale thermostat<sup>31</sup>.

For the initial WT-MetaD simulation, we used the backbone dihedrals  $\phi$  and  $\varphi$  as collective variables, depositing Gaussians of height 1.2 kJ/mol and width 0.35 every 500 time steps, using a bias factor of 8. These parameters were chosen to match those used by Bonomi, et al.<sup>32</sup> We used the resulting WT-MetaD potential to bias our second simulation, in which we calculated the free energy curve projected onto  $\phi$ , and we calculated the 36 distances between non-bonded heavy atoms as a dataset for reweighting. To generate shifted ensembles for testing reweighting, we added Gaussians of the form  $-ke^{-(\phi-1)^2}$  to the free energy projected onto  $\phi$  at  $\phi = 1$ , and we measured ensemble averages of the 36 non-bonded heavy atom distances on these shifted ensembles.



**Reweighting of the dialanine simulations.** To assess the ability of the reweighting procedure to improve incorrect ensembles, we reweighted the prior dialanine ensemble with synthetic data from the shifted ensembles, using 36 distances between non-bonded heavy atoms. To generate the reweighted ensemble, we initially used gradient descent minimization with step sizes of 0.5, initializing Lagrange multipliers to 0. The MetaD bias potential was included with the prior ensemble. For these initial tests, we incorporate no error into our dataset or in the reweighting algorithm, and we generated data from an ensemble shifted by a Gaussian of magnitude 15 kJ/mol.

We measured convergence levels by measuring the current data error, taking the L1 norm of the deviation between the data from the expected curve and the current 36 average distances on the reweighted ensemble. If the data from the true ensemble is given by  $f_i$ , with  $i$  ranging from 1 to 36, and the data from the current ensemble is  $f_i^*$ , then the data error is

$$\text{data error} = \sum_i |f_i - f_i^*|$$

We found that it was challenging to reach convergence with a gradient descent approach, with the data error reaching a plateau above  $3 * 10^{-4}$ . Even after scanning for alternative step sizes for gradient descent, this challenge remained. As discussed previously, gradient descent methods may fail to converge even in the case of a unique minimum, as the Lagrange multipliers may oscillate around an optimal setting.

Thus, we used a conjugate gradient descent strategy as an alternative minimization technique, finding that the data errors converged more quickly with this approach. We terminated the minimization procedure once the data error was below  $1.67 * 10^{-5}$ , finding that the magnitude of the deviation of Lagrange multipliers between iterations fell to less than 1% of their magnitude after this threshold was reached.

To evaluate the reweighted ensembles, we used two measures to compare the deviation between the prior ensemble and the true ensemble with the deviation between the reweighted ensemble and the true ensemble. Both measures were made after aligning all FE curves at the left of the two energy basins for backbone dihedral  $\phi$ , at  $\phi = -2.6$ .

First, we computed the RMSD FE error between two ensembles as follows. We projected the free energy onto the backbone dihedral  $\phi$ , binning the dihedral angles into 51 bins; this procedure yielded free energy values  $fe_j^i$  for each ensemble  $i$  and each  $\phi$  bin  $j$ . Then the RMSD FE error is defined as follows

$$RMSD\ FE\ error = \sqrt{\frac{\sum_j (fe_j^1 - fe_j^2)^2}{num\_bins}}$$

We also compared ensembles using a weighted RMSD FE error, which weights errors by the population in that  $\phi$  bin. Varying the temperature in the weighted RMSD FE error calculation can alter the influence of bin populations; indeed, at low temperature, only the free energy minima of the system would contribute to the weighted RMSD FE error, whereas in the large temperature limit the weighted RMSD FE becomes equivalent to the unweighted RMSD FE error. Note that the population in bin  $j$  of ensemble  $i$  is given by

$$w_j^i = e^{-fe_j^i/k_bT} / \sum_j e^{-fe_j^i/k_bT}$$

Then the weighted RMSD FE error can be computed as

$$weighted\ RMSD\ FE\ error = \sqrt{\sum_j w_j^i (fe_j^1 - fe_j^2)^2}$$

**Molecular dynamics simulations of c-Myc<sub>402-412</sub>.** We next examined the performance of reweighting on c-Myc<sub>402-412</sub> in the presence of the small molecule 10058-F4. This system provides a biologically relevant<sup>33</sup> and computationally tractable case for evaluating reweighting. Furthermore, NMR chemical shifts of the peptide in the presence 10058-F4 are available.<sup>34-35</sup>

Using GROMACS 5.1.4<sup>28</sup>, we simulated c-Myc<sub>402-412</sub> and 10058-F4 with three force field settings: Amber94<sup>36</sup> with the TIP3P<sup>37</sup> water model, Amber03w<sup>38</sup> with the TIP4P/2005<sup>39</sup> water model, and Amber03ws<sup>40</sup> with the TIP4P/2005<sup>39</sup> water model. For 10058-F4, we used the force field parameterization determined in previous work.<sup>35</sup> These parameters were meant to better

model the compound, particularly the potential energy function of the dihedral angle of the molecule, which was observed to be poor when using the GAFF force field parameters<sup>41</sup>.

We used the following steps to set up the systems for simulation: We placed N-terminally acetylated and C-terminally amidated c-Myc<sub>402-412</sub> in a 5.86 nm cubic box with one molecule of 10058-F4. We then solvated the system with water, yielding 6791 water molecules for the TIP4P/2005 and 6661 water molecules for the TIP3P. We added one sodium ion to each simulation to neutralize the net charge of the system. We then ran an energy minimization until the maximum force applied on the system per step was less than 1000 kJ/mol/nm. To generate velocities for the system, we ran a 500 ps NPT simulation with a Berendsen barostat<sup>42</sup> maintaining the pressure at 1 atm and generating temperatures at 300 K with v-rescale thermostat<sup>31</sup>. We constrained bonds with the LINCS constraint algorithm<sup>43</sup>.

To generate a set of varied configurations across the free energy landscape of c-Myc<sub>402-412</sub> for starting 128 separate trajectories, we ran an NVT simulation at 600 K for 2 ns to generate a 1000 frame trajectory. Again, we used the v-rescale thermostat in GROMACS for temperature coupling, generating velocities at 600 K. We chose 128 frames randomly from the resulting trajectory as a starting point for the production simulation replicas.

In the final production c-Myc<sub>402-412</sub> simulations using these starting configurations, we generated 5.0  $\mu$ s of aggregated simulation time across 128 simulation replicas at 300 K. We also ran a restraining simulation for each force field for 5.0  $\mu$ s of aggregated simulation time, with multiple replica restraints as implemented in metainference<sup>18</sup> applied to the 128 replicas, using the chemical shift data and the forward model discussed below (section c-Myc<sub>402-412</sub> Forward Model and Reweighting). To further enhance sampling for the unrestrained simulation, we also ran a replica exchange molecular dynamics (REMD)<sup>44</sup> simulation for each force field for 8.5  $\mu$ s of simulation time across 128 replicas, yielding 66.047 ns of simulation in 83,009 frames for the 300 K replica. For each simulation setting, we ran with 2 fs timesteps, using Particle Mesh Ewald calculations<sup>45</sup> for Coulomb forces with a cutoff of 0.9, and cutting off van der Waals forces at 0.9. We constrained bonds with the LINCS algorithm<sup>43</sup>. We generated velocities on each replica to match the temperature of that replica.

**Enhanced sampling of c-Myc<sub>402-412</sub>.** With PLUMED 2.3.1<sup>29</sup>, MetaD was used to encourage the system to sample more efficiently across the energy landscape. The nine collective variables that were chosen match those used previously.<sup>35</sup> In the description below, we include the abbreviations used for these collective variables.

As a first collective variable, we used the number of  $\phi$  and  $\varphi$  angles that matched the expected angles for a right  $\alpha$ -helix (*ahelixright*). Our second, third, and fourth collective variables were designed similarly, measuring  $\beta$ -sheet content (*betasheet*), left  $\alpha$ -helix content (*ahelixleft*), and polyproline helix content (*polypro*), respectively. These first four collective variables encouraged exploration of various distributions of secondary structure in the peptide. As a fifth collective variable, we used the radius of gyration (*rgyr*), or the RMSD distance between atoms in the peptide to its center of mass, encouraging the peptide to explore varying extended states. For the final collective variable for the peptide, we used the number of backbone hydrogen bonds (*ohbond*), counting the number of relevant hydrogen and electronegative atoms less than 2.5 Å apart. The final three collective variables measured the distance between sections of the peptide and the small molecule. The first of these measured the distance between c-Myc<sub>402-405</sub> and the small molecule (*fstdr*), the next between c-Myc<sub>406-408</sub> and the small molecule (*scddr*), and the final between c-Myc<sub>409-412</sub> and the small molecule (*trddr*). These distances ensured that the simulation explored the variety of potential interactions between 10058-F4 and the peptide.

To determine the desired Gaussian sizes for these collective variables, we ran a short 1 ns version of the production simulation, computing the collective variables' values across the resulting frames. Half the standard deviation of each collective variable after this short simulation was used as the sigma value for deposited Gaussians in MetaD. We deposited Gaussians with height 1.2 every 500 time-steps, running with a bias factor of 10.

We ran additional simulations that used REMD<sup>44</sup> to enhance the sampling of our c-Myc<sub>402-412</sub> system, with a low temperature of 300 K and a high temperature of 600 K, and 128 total replicas. The temperatures chosen for each replica were in a geometric series from 300 K to 600 K. Exchanges were attempted between neighboring replicas every 500 time-steps. Upon observing exchange frequencies during the simulation, we found that it was unnecessary to include additional replicas.

**Convergence of the *c-Myc*<sub>402-412</sub> REMD simulations.** We performed various checks to ensure that the REMD simulations were sufficiently converged before applying reweighting. First, we ensured that the full domain of each collective variable had been explored sufficiently during the simulation (Fig. S3).

As another check for convergence, we assessed whether the simulation had reached a regime in which sections of the simulation were uncorrelated with one another. To determine if given sections, or blocks, of the simulation are correlated, we estimated an observable from the simulation by measuring this value across blocks of simulation, and we viewed how the error of this estimate varies with the block size. For the observable, we used the free energy curves projected onto each collective variable. The average error reached by every collective variable is relatively low, and we observed that the block errors reach a plateau in all cases, indicating that the correlation between blocks is no longer present at larger block sizes (Fig. S4).

Additionally, we estimated whether sampling in coordinate space is complete by clustering the trajectory's configurations, checking whether blocks of the simulation have similar distributions of populations in these clusters, and determining if simulation blocks are uncorrelated with respect to cluster populations. We clustered the trajectory frames using the Gromos algorithm<sup>46</sup>, considering frames as neighbors if the backbone C $\alpha$  atoms were less than 2 Å RMSD away. With the Amber03ws force field, 335 clusters were generated; and with Amber94, 257 clusters were generated. For each block size, we computed the number of frames belonging to each cluster, weighted by the MetaD bias, and we computed the average error of these cluster sizes across blocks. We can see that as in the previous section, the block errors level off with increasing block size, indicating convergence (Fig. S5).

For the REMD simulations, additional checks are necessary to ensure convergence, as it may be the case then that the simulations never actually explore the full free energy landscape, but rather exchange configurations in separate free energy basins across replicas, giving the appearance of exploration on a single replica. To check for this potential artefact, we produced alternative demuxed trajectories that followed one starting configuration across temperatures as it was exchanged across replicas. For each collective variable, we plotted the average histogram across the 128 demuxed trajectories, along with the minimum and maximum bin population for each position across all the replicas.<sup>47</sup> We saw that the distributions of collective variables across demuxed trajectories is conserved, with the minimum and maximum value in each histogram bin

deviating minimally from the mean distribution (Fig. S6). Thus, individual demuxed trajectories are not contributing distinct sections of configuration space.

***Forward model and reweighting of the c-Myc<sub>402-412</sub> simulations.*** As a forward model to predict chemical shifts from a configuration of the system, we used CamShift<sup>48</sup>. We ran reweighting on the ensembles from Amber94 with TIP3P, Amber03w with TIP4P/2005, and Amber03ws with TIP4P/2005, using the 12 C $\alpha$  and C $\beta$  chemical shifts as experimental observables. The restraining simulations also used these 12 C $\alpha$  and C $\beta$  chemical shifts. For the determination of the Lagrange multipliers, we used conjugate gradient optimization. The MetaD bias potential from each simulation was used as its prior ensemble. The search for Lagrange multipliers was terminated once reaching a convergence threshold of  $2.18 * 10^{-5}$ . To avoid overflow errors with the exponents in the objective function and gradient, it was critical to shift the linear bias potential with an additive constant, which does not alter the correctness of the reweighted ensemble.

For both restraining and reweighting simulations, we used the Gaussian error model for the data, taking the typical error of the Camshift predictor<sup>48</sup> as an approximation of the standard deviation of the error. In particular, we used the reported error for CamShift on a 28-protein test set, computed as an RMSD between experimental and predicted chemical shifts (1.3 for C $\alpha$  shifts, 1.36 for C $\beta$  shifts, and 0.28 for H $\alpha$  shifts).<sup>48</sup>

***Details of the sampling experiment.*** To test reweighting in the setting of poor sampling, we generated ensembles of varying qualities by shifting and subsampling the simulation ensembles for dialanine and c-Myc<sub>402-412</sub>. Analysis of these trials is presented in Fig. 4, S8, and S9.

For dialanine, we began with the free energy curve generated from simulation, and we shifted this landscape with Gaussians at  $\phi = 1$  as discussed previously. We ran experiments with Gaussians of magnitude between 15 and 40 in increments of 5. For each ensemble, we subsampled the trajectory frames using weights computed from the simulation bias, thus generating samples that ergodically sampled from the initial simulation ensemble. We generated ensembles of sizes between 10,000 and 100,000 frames in increments of 10,000. For each sample size and shift size, we generated 5 trials. Then, we used the dataset of 36 distances as collected on the original simulation ensemble to reweight each prior, using for the error model Gaussian error

with fixed standard deviation 0.01. We measured the quality of the reweighted ensemble by computing the free energy error to the original simulation ensemble.

For c-Myc<sub>402-412</sub>, we again shifted the ensemble from simulation with Gaussians to generate priors of varying quality. In particular, we began with the free energy landscape projected onto the  $\beta$ -sheet content (number of dihedral angles with angles corresponding to  $\beta$ -sheet structure), and added Gaussians to this landscape of the form  $-ke^{-1.5(\beta_{sheet}-6)^2}$  with  $k$  ranging from 15 to 40 in increments of 5. Again, we generated ensembles of sizes between 10,000 and 100,000 in increments of 10,000 by ergodically sampling from the original simulation, and we conducted 5 trials for each shift size and ensemble size setting. We ran reweighting on these ensembles using the 12 C $\alpha$  and C $\beta$  chemical shifts, and we tested the quality of the resulting ensembles by measuring the match to the 9 H $\alpha$  chemical shifts not used for reweighting.

***Indicators for the robustness of the reweighted ensemble.*** We explored various metrics that can be easily computed to help assess the reliability of the reweighted ensemble when additional validation data are unavailable.

As a basic first check, it is useful to ensure that the range of forward model predictions for an experimental observable from an ensemble frames actually overlaps with the observed experimental value. We term ‘domain failures’ cases in which this overlap does not occur. While none of the c-Myc<sub>402-412</sub> ensembles tested had a domain failure with the measured 12 C $\alpha$  and C $\beta$  chemical shifts, domain failures were observed for some of the larger ensemble shifts tested for the dialanine system did. These domain failures mapped to cases in which the reweighting procedure led to a worse match with the true dialanine ensemble (Fig. S8).

If further analysis of a system depends on some experimental observable, the error of this observable in the reweighted ensemble is a useful measure for whether it is reasonable to use the reweighted ensemble. We explored how the quality of the prior ensemble tracked with the error in an experimental observable: the experimentally measured 9 H $\alpha$  chemical shifts for this system. We computed the average H $\alpha$  chemical shift measurement in each of 10 blocks of the reweighted ensemble, and we computed the error of this measurement across blocks. In Fig. S9, we see that this block error increases when the prior quality is lower, either when the prior has a

smaller number of frames or deviates more from the true ensemble. This result supports the expectation that the reweighted ensemble should be less reliable with poor sampling.

We finally explored the Kish effective sample size<sup>49</sup> as a metric to assess whether a reweighted ensemble is reliable. After reweighting, say each frame of the simulation is given a weight  $w_i$ . Then the Kish effective sample size is computed as:

$$\text{Kish effective sample size} = \frac{(\sum w_i)^2}{\sum w_i^2}$$

Note that when the reweighting algorithm preferences a single frame over all the rest, the Kish effective sample size takes the value 1; when the reweighting algorithm preferences all frames equally, the Kish effective sample size is the size of the original ensemble. We normalize this metric by the original size of the ensemble. The log-normalized Kish effective sample size, termed the Kish score, then takes lower values when the reweighting algorithm leads to a small ensemble relative to the original ensemble; when this is the case, we expect further analysis of the ensemble to be less robust.

## Results

The dialanine simulation yielded approximately 200,000 frames, producing the FE landscape in Fig. 2a. Following previous work on this system<sup>32</sup>, we considered a modified ensemble with the FE shifted by an added Gaussian shift of 15 kJ/mol at  $\phi=1$ , altering the balance between the two FE minima (Fig. 2b, green curve). We reweighted the prior with synthetic data from this shifted ensemble, using 36 distances between non-bonded heavy atoms.

The reweighting algorithm converged in a computationally efficient manner, requiring just minutes on a personal computer to reweight the trajectory frames. The procedure shifted the prior FE closer to the true ensemble, reducing the average FE error from 6.70 kJ/mol to 3.87 kJ/mol. The remaining error can be interpreted by considering intrinsic features of the method. The reweighted ensemble converged to an ensemble different from the true one even as the match to the data reached RMSD below  $10^{-4}$  nm (Fig. 2d). We note, however, that the error was most reduced in highly populated regions of the true ensemble, where data were most in-



formative (Fig. 2c). Indeed, we can compute an FE error weighted by ensemble population (see SI); reweighting shifts the weighted FE error from 13.80 kJ/mol to 0.41 kJ/mol. This weighted FE error measurement highlights that the true ensemble and reweighted ensemble are quite similar, agreeing closely in highly populated regions. We recall that maximum entropy restraints can be shown to minimize the KL divergence from the posterior to the experimental ensemble,<sup>24, 50</sup> and that this metric by construction upweights highly populated conformations. The maximum entropy method did not find the true ensemble exactly because, in this case, a different ensemble closer to the prior exists that also matches the experimental data; indeed, the KL divergence from the reweighted to the prior ensemble (3.03 kJ/mol) is lower than the KL divergence from the true to the prior ensemble (3.10 kJ/mol). We would expect a restrained simulation using the same dataset to similarly not be able to recover the true ensemble in less populated regions of  $\phi$  space. When using a dataset that is also informative for less populated regions of  $\phi$  space, reweighting can bring the modelled ensemble closer to the true ensemble in high FE regions as well (Fig. S1). The procedure continues to improve the prior in tests with larger Gaussian shifts, and when the data are modeled with Gaussian error or with an uninformative Jeffreys prior<sup>51</sup> on the data variance (Fig. S2, Table S1).

For the c-Myc<sub>402-412</sub> system, we ran simulations with parallel bias metadynamics<sup>52</sup> (PBMETAD) using collective variables chosen based on previous work.<sup>35</sup> In order to test reweighting with various priors, we ran simulations with three force fields: Amber03ws<sup>40</sup>, Amber03w<sup>38</sup>, and Amber94<sup>36</sup>, using GROMACS<sup>28</sup> equipped with PLUMED<sup>29</sup>. We used 12 previously published C $\alpha$  and C $\beta$  chemical shifts as metainference restraints or for *a posteriori* reweighting.<sup>53</sup> To compare the use of data for restraining vs for reweighting with equivalent sampling, for each force field we ran a metadynamics metainference simulation<sup>18</sup> and an unrestrained PBMETAD simulation, both employing 128 replicas for 5  $\mu$ s of aggregated simulation time. Additionally, we explored reweighting replica exchange molecular dynamics (REMD)<sup>44</sup> simulations with 8.5  $\mu$ s of aggregated simulation time for each force field (see Figs. S3-6 for simulation convergence).

To evaluate the results of the restraining and reweighting simulations, we assessed the agreement to 9 H $\alpha$  peptide chemical shifts not used to generate the ensembles. For each ensemble, we computed the deviation of ensemble-averaged predicted H $\alpha$  chemical shifts from the experimentally measured value (Fig. 3). The Amber94 simulation with no experimental data yielded the worst match to the chemical shifts (Fig. 3, red bars). After reweighting, the average deviation from the H $\alpha$  chemical shifts reduced in all force fields to between 0.06 and 0.08 ppm (Fig. 3,

blue bars). The reweighting and restraining approaches (Fig. 3, green bars) performed comparably, with  $H\alpha$  chemical shift deviations within simulation statistical error. Reweighting the REMD ensemble (Fig. 3, orange bars) provided a similar match to the chemical shifts, indicating that, for this system, both sampling approaches allowed for successful reweighting.

We additionally compared the reweighted and the restrained ensembles on the 9 collective variables used for metadynamics enhanced sampling. For each simulation approach and prior, we computed the FE landscape projected onto these collective variables, which include features such as the secondary structure content of the peptide or the distance between the peptide and the small molecule (Fig. S7). The reweighting and restraining approaches produced a similar FE landscape for each collective variable. Whereas restraining and reweighting yielded a closer match to the experimental  $H\alpha$  chemical shifts when applied to the priors from all three force fields (Fig. 3), these approaches produced distinct collective variable distributions across force fields, remaining close to the corresponding prior ensembles (Fig. S7). For instance, the ensembles based on the Amber94 prior retain a high  $\alpha$ -helical content for the peptide compared to the ensembles based on the Amber03w and Amber03ws priors. It is possible that other experimental data sets can better correct this discrepancy between force fields.

Since reweighting converges within 10 s for each simulation, the procedure in this case is an efficient post-processing step for incorporating data into ensembles. When experimental data are collected after a simulation has completed, reweighting presents an opportunity to include new information without restarting simulations. In some cases, even if data are available before generating a prior ensemble, it may still be computationally prohibitive to evaluate restraints during a simulation. Here, however, including the  $C\alpha$  and  $C\beta$  chemical shifts as metainference restraints did not substantially increase simulation time when compared to unrestrained simulations (Table S2).

In this case, reweighting with experimental data improved the match of the ensembles to the  $H\alpha$  chemical shifts and produced ensembles that agreed with restrained simulations across various collective variables. More generally, however, we expect that the procedure might fail in cases of poor sampling. It is thus important to develop strategies for evaluating the quality of a reweighted ensemble when validation data are limited. To investigate cases with poor sampling, we created priors of varying qualities by shifting and subsampling the ensembles for dialanine and c-Myc<sub>402-412</sub> with Gaussian shifts of varying magnitudes (Fig. 4a,c). For dialanine, we re-

weighted these ensembles using synthetic data from the original ensemble (Fig. 4a, red curve), whereas for c-Myc<sub>402-412</sub> we began with the Amber03ws REMD ensemble (Fig. 4c, red curve) and reweighted using the 12 C $\alpha$  and C $\beta$  chemical shifts. Using these priors, we explored metrics that signal unreliable reweighting.

We present the following three indications to identify unreliable reweighting. First, for each data point, we can determine if the range of predicted values from the prior does not overlap with the experimental data. These domain failures are indicative of poor reweighting: in the case of dialanine, reweighting worsened the ensemble in 96.7% of trials with domain failures (Fig. S8). Second, we can compute the error of an observable of interest across blocks of the reweighted ensemble to assess the procedure (Fig. S9). Third, we can compute the Kish effective sample size<sup>49</sup> using reweighting weights to evaluate ensemble reliability (details in SI). This metric indicates the extent to which reweighting shifts the prior; it is equal to 1 when reweighting favors a single ensemble frame, and it is equal to the size of the prior ensemble if reweighting does not change the prior. In tests with dialanine (Fig. 4b) and c-Myc<sub>402-412</sub> (Fig. 4d), lower log normalized Kish effective sample sizes (Kish scores) correspond to cases in which reweighting performs inconsistently. When the Kish score is below -8 in these cases, reweighting is sharply less reliable, and it is preferable to use a more accurate prior.

## Conclusions

Taken together, our results indicate that reweighting approaches can efficiently improve unrestrained ensembles and match the quality of restrained ensembles when force fields and sampling are accurate. To enable the correct use of reweighting when validation data are insufficient, we discussed metrics that can be used for quality control. We anticipate that inventive combinations of restraining and reweighting approaches will enable the future development of increasingly accurate methods for integrating experiments and simulations, leading to improved mechanistic understanding of protein function through the determination of structural ensembles.

## Associated Content

The Supporting Information is available free of charge on the ACS Publications website at DOI:

## Corresponding Author

\*massimiliano.bonomi@pasteur.fr , \*mv245@cam.ac.uk

## Present Addresses

§Stanford University, Stanford, CA 94305, USA

+Structural Bioinformatics Unit, Institut Pasteur, CNRS UMR 3528, 75015 Paris, France

## Funding Sources

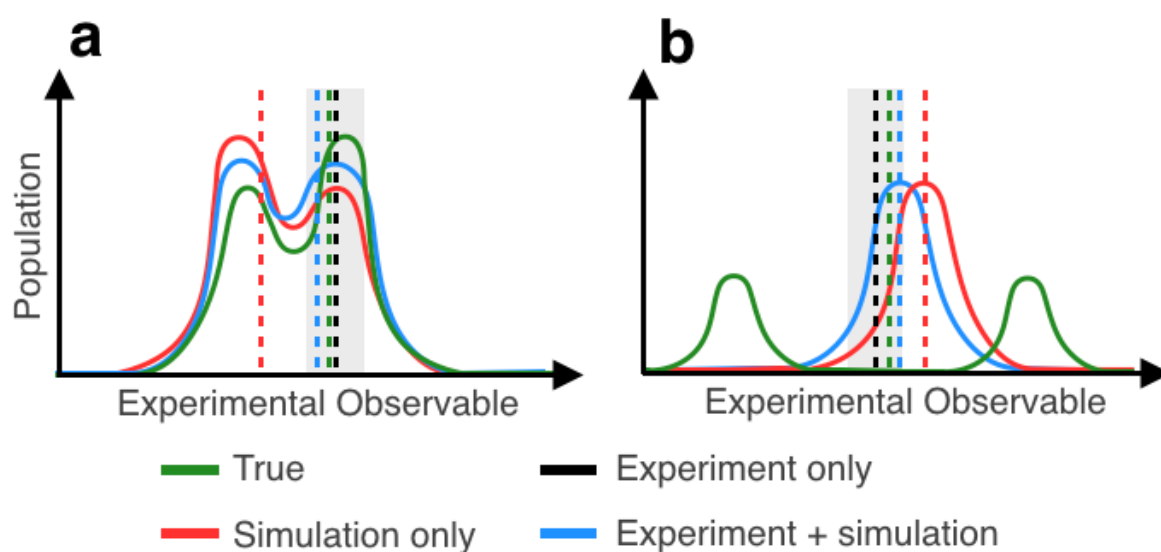
The authors acknowledge the Harvard-Cambridge Fiske Scholarship (RR) and the Gates Cambridge Scholarship (GTH) for support.

## References

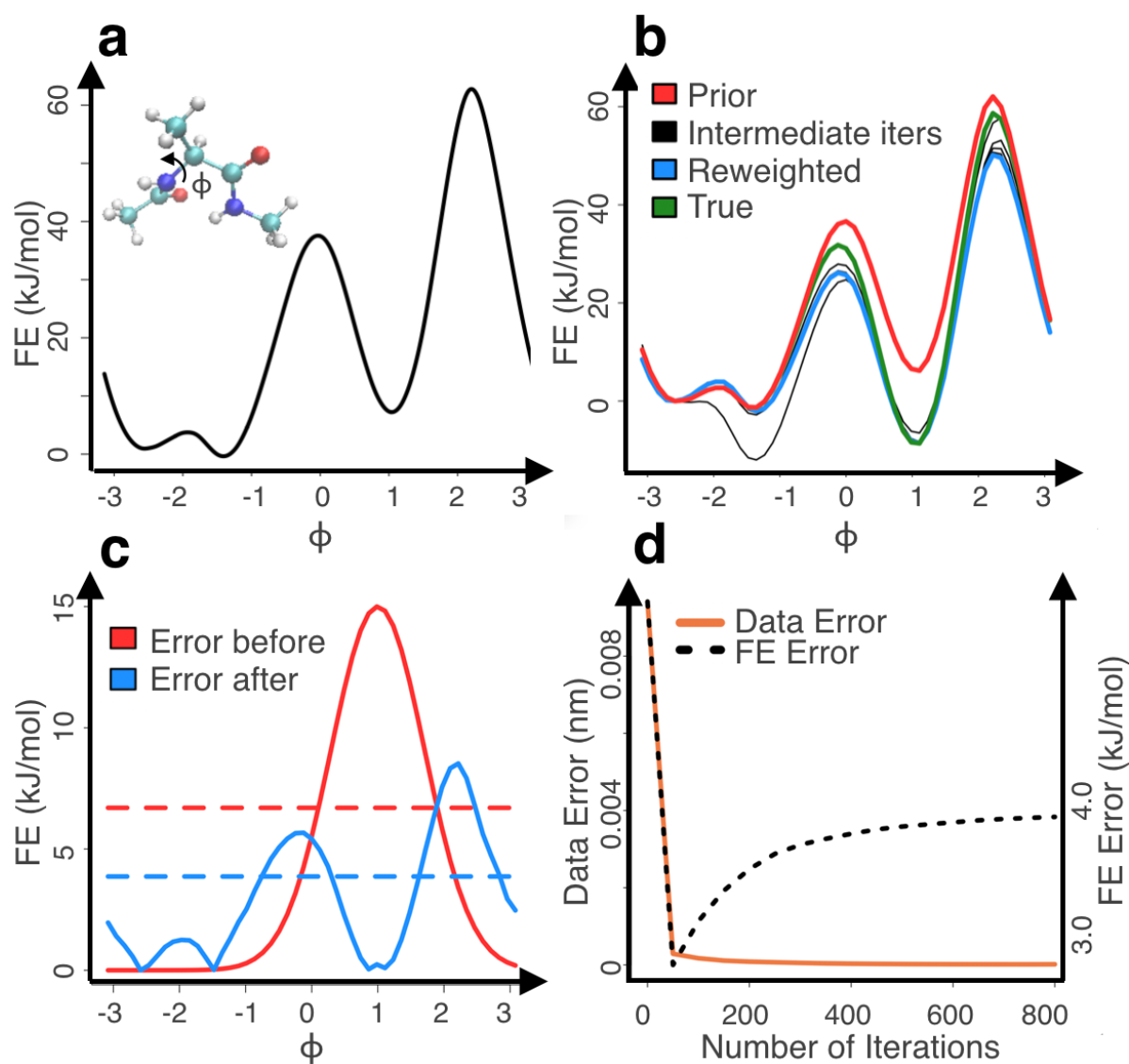
1. Henzler-Wildman, K.; Kern, D., Dynamic personalities of proteins. *Nature* **2007**, *450*, 964-972.
2. van den Bedem, H.; Fraser, J. S., Integrative, dynamic structural biology at atomic resolution - it's about time. *Nat. Methods* **2015**, *12*, 307-318.
3. Agarwal, P. K., Enzymes: An integrated view of structure, dynamics and function. *Microb. Cell Fact.* **2006**, *5* (2).
4. Khalili-Araghi, F.; Gumbart, J.; Wen, P. C.; Sotomayor, M.; Tajkhorshid, E.; Schulten, K., Molecular dynamics simulations of membrane channels and transporters. *Curr. Opin. Struct. Biol.* **2009**, *19* (2), 128-137.
5. Latorraca, N. R.; Venkatakrisnan, A. J.; Dror, R. O., GPCR Dynamics: Structures in Motion. *Chem Rev* **2017**, *117* (1), 139-155.
6. Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V. N., Introducing Protein Intrinsic Disorder. *Chem. Rev.* **2014**, *114*, 6561-6588.
7. Sormanni, P.; Piovesan, D.; Heller, G. T.; Bonomi, M.; Kukic, P.; Camilloni, C.; Fuxreiter, M.; Dosztanyi, Z.; Pappu, R. V.; Babu, M. M.; Longhi, S.; Tompa, P.; Dunker, A. K.; Uversky, V. N.; Tosatto, S. C. E.; Vendruscolo, M., Simultaneous quantification of protein order and disorder. *Nat. Chem. Biol.* **2017**, *13*, 339-342.
8. Tompa, P., Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* **2012**, *37* (12), 509-516.
9. Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M., Principles of protein structural ensemble determination. *Curr. Opin. Struct. Biol.* **2017**, *42*, 106-116.
10. Schneidman-Duhovny, D.; Pellarin, R.; Sali, A., Uncertainty in integrative structural modeling. *Curr. Opin. Struct. Biol.* **2014**, *28*, 96-104.
11. Allison, J. R., Using simulation to interpret experimental data in terms of protein conformational ensembles. *Curr. Opin. Struct. Biol.* **2017**, *43*, 79-87.
12. Fisher, C. K.; Stultz, C. M., Constructing ensembles for intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2011**, *3*, 426-431.

13. Gaalswyk, K.; Muniyat, M. I.; MacCallum, J., The emerging role of physical modeling in the future of structure determination. *Curr. Opin. Struct. Biol.* **2018**, *49*, 145-153.
14. Pitera, J.; Chodera, J., On the use of experimental observations to bias simulated ensembles. *J. Chem. Theory Comput.* **2012**, *8*, 3445-3451.
15. Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K., Combining experiments and simulations using the maximum entropy principle. *PLOS Comput. Biol.* **2014**, *10* (2), e1003406.
16. Best, R. B.; Vendruscolo, M., Determination of protein structures consistent with NMR order parameters. *J. Am. Chem. Soc.* **2004**, *126*, 8090-8091.
17. Cavalli, A.; Camilloni, C.; Vendruscolo, M., Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem. Phys.* **2013**, *138*, 169903.
18. Bonomi, M.; Camilloni, C.; Cavalli, A.; Vendruscolo, M., Metainference: A Bayesian inference method for heterogeneous systems. *Sci. Adv.* **2016**, *2*.
19. Cesari, A.; Gil-Ley, A.; Bussi, G., Combining simulations and solution experiments as a paradigm for RNA force field refinement. *J. Chem. Theory Comput.* **2016**, *12*, 6192-6200.
20. Beauchamp, K.; Pande, V.; Das, R., Bayesian energy landscape tilting: towards concordant models of molecular ensembles. *Biophys. J.* **2014**, *106* (6), 1381-1390.
21. Leung, H.; Bignucolo, O.; Aregger, R.; Dames, S.; Mazur, A.; Berneche, S.; Grzesiek, S., A rigorous and efficient method to reweight very large conformational ensembles using average experimental data and to determine their relative information content. *J. Chem. Theory Comput.* **2016**, *12* (1), 383-394.
22. Hummer, G.; Kofinger, J., Bayesian ensemble refinement by replica simulations and reweighting. *J. Chem. Phys.* **2015**, *143* (24), 243150.
23. Ceriotti, M.; Brain, G. A. R.; Riordan, O.; Manolopoulos, D. E., The inefficiency of re-weighted sampling and the curse of system size in high-order path integration. *Proc. Royal Soc. A* **2011**, *468* (2137), 2-17.
24. Cesari, A.; Reißer, S.; Bussi, G., Using the Maximum Entropy Principle to Combine Simulations and Solution Experiments. *Computation* **2018**, *6* (1).
25. Jaynes, E. T., Information Theory and Statistical Mechanics. *Physical Review* **1957**, *106* (4), 620-630.
26. Ruder, S., An overview of gradient descent optimization algorithms. *CoRR* **2016**.
27. Barducci, A.; Bussi, G.; Parrinello, M., Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603.
28. Abraham, M.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J.; Hess, B.; Lindahl, E., Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19-25.
29. Tribello, G.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G., Plumed2: New feathers for an old bird. *Computer Physics Communications* **2014**, *185* (2), 604-613.
30. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side-chain torsion potentials for the Amber ff99sb protein force field. *Proteins* **2010**, *78* (8), 1950-1958.
31. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J Chem Phys* **2007**, *126* (1), 014101.
32. Bonomi, M.; Camilloni, C.; Vendruscolo, M., Metadynamic metainference: Enhanced sampling of the metainference ensemble using metadynamics. *Sci. Rep.* **2016**, *6*, 31232.
33. Follis, A.; Hammoudeh, D.; Wang, H.; Prochownik, E.; Metallo, S., Structural rationale for the coupled binding and unfolding of the c-Myc oncoprotein by small molecules. *Chemistry & Biology* **2008**, *15* (11), 1149-1155.

34. Hammoudeh, D.; Follis, A.; Prochownik, E.; Metallo, S., Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *J. Am. Chem. Soc.* **2009**, *131*, 7390-7401.
35. Heller, G. T.; Aprile, F.; Bonomi, M.; Camilloni, C.; Simone, A.; Vendruscolo, M., Sequence specificity in the entropy-driven binding of a small molecule and a disordered peptide. *J. Mol. Biol.* **2017**, *429* (18), 2772-2779.
36. Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179-5197.
37. Jorgensen, W.; Chandrasekhar, J.; Madura, J., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.
38. Best, R. B.; Mittal, J., Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B* **2010**, *114* (46), 14916-14923.
39. Abascal, J.; Vega, C., A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*.
40. Best, R. B.; Zheng, W.; Mittal, J., Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* **2014**, *10* (11), 5113-5124.
41. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J Comput Chem* **2004**, *25* (9), 1157-74.
42. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, *81* (8), 3684-3690.
43. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem* **1998**, *18* (12).
44. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314* (1-2), 141-151.
45. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98* (12), 10089-10092.
46. Daura, X.; Gademann, K.; Juan, B.; Seebach, D.; van Gunsteren, W.; Mark, A., Peptide folding: when simulation meets experiment. *Angew. Chem.* **1999**, *38* (1-2), 236-240.
47. Henriksen, N. M.; Roe, D. R.; Cheatham, T. E., 3rd, Reliable oligonucleotide conformational ensemble generation in explicit solvent for force field assessment using reservoir replica exchange molecular dynamics simulations. *J Phys Chem B* **2013**, *117* (15), 4014-27.
48. Kohlhoff, K.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M., Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *J. Am. Chem. Soc.* **2009**, *131*, 13894-13895.
49. Kish, L., *Survey Sampling*. Wiley: New York, 1965.
50. Dannenhoffer-Lafage, T.; White, A. D.; Voth, G. A., A Direct Method for Incorporating Experimental Data into Multiscale Coarse-Grained Models. *J. Chem. Theory Comput.* **2016**, *12* (5), 2144-53.
51. Sivia, D.; Skilling, J., *Data analysis: A Bayesian tutorial*. Oxford University Press: Oxford, 2006.
52. Pfandtner, J.; Bonomi, M., Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics. *J. Chem. Theory Comput.* **2015**, *11* (11), 5062-5067.
53. Bonomi, M.; Camilloni, C., Integrative structural and dynamical biology with PLUMED-ISDB. *Bioinformatics* **2017**, *33* (24), 3999-4000.

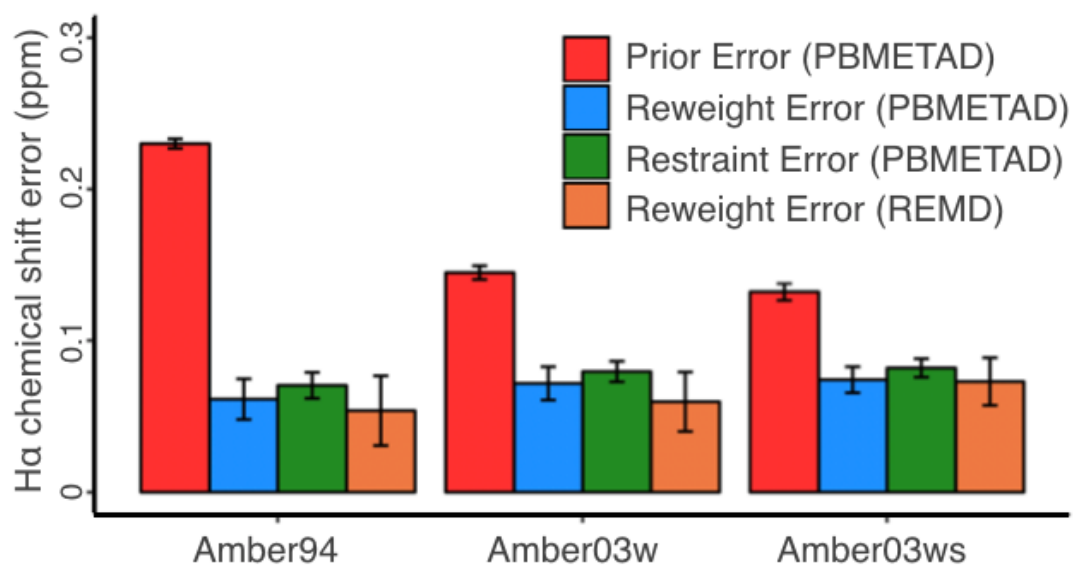


**Figure 1. Schematic illustration of different approaches for determining structural ensembles.** (a) The black dashed line indicates an experimental readout and the colored dashed lines ensemble averages, corresponding to different cases: the true ensemble (green curve), the prior ensemble (red curve) and the ensemble of the simulations combined with experimental data by reweighting or by restraining (blue curve); the shaded region indicates the experimental error. (b) When the prior (red curve) and true (green curve) ensembles have a poor overlap, reweighting (blue curve) does not help much.

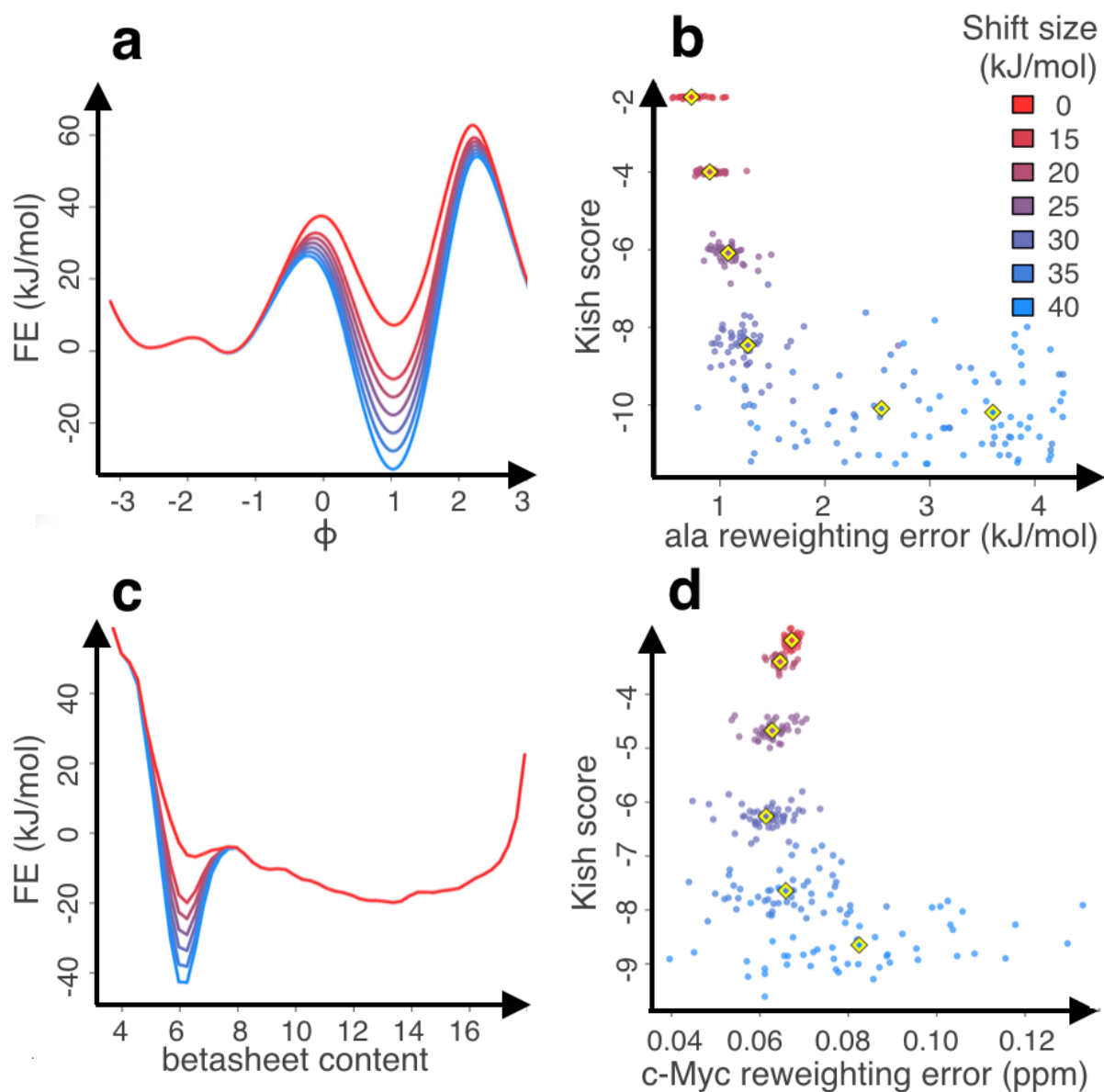


**Figure 2. Reweighting of the free energy landscape of dialanine.** (a) Dialanine structure and its free energy (FE) projected onto the dihedral  $\phi$ . (b) Reweighted FE using data from the true ensemble and conformations from the prior. Curves are shifted to be equal at  $\phi = -2.6$ . (c) Absolute deviation between the prior and true ensemble (red), and between the reweighted and true ensemble (blue). Dashed lines indicate RMSD errors over  $\phi$ . (d) RMSD FE error between reweighted and true ensemble, and convergence to data over iterations of minimization. When converging reweighting with an error model for the data (Table S1), the FE error resembles the minimum error obtained at 50 iterations with no data error model.





**Figure 3. Comparison of the performance of reweighting and restraining methods on c-Myc402-412.** Bars depict the average deviation between predicted and experimental H $\alpha$  chemical shifts. Error bars indicate the standard deviation of these errors, calculated with block analysis (SI).



**Figure 4. Kish effective sample size as a measure for reliable reweighting.** (a,b) Dialanine. (c,d) c-Myc402-412. (a,c) Ensembles from simulation (red), and shifted priors. (b,d) The Kish score is the log normalized Kish effective sample size. The dialanine reweighting error is the FE difference between the reweighted and true ensembles. The c-Myc402-412 reweighting error is the error of H $\alpha$  chemical shifts after reweighting; diamonds indicate shift size averages.

## ToC Graphic

