



Transferable Implicit Solvation via Contrastive Learning of Graph Neural Networks

Justin Airas, Xinqiang Ding, and Bin Zhang*



Cite This: ACS Cent. Sci. 2023, 9, 2286–2297



Read Online

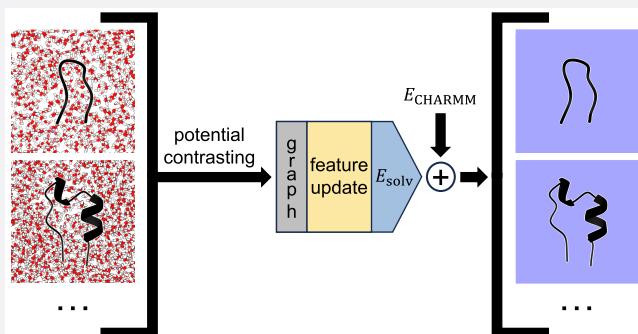
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Implicit solvent models are essential for molecular dynamics simulations of biomolecules, striking a balance between computational efficiency and biological realism. Efforts are underway to develop accurate and transferable implicit solvent models and coarse-grained (CG) force fields in general, guided by a bottom-up approach that matches the CG energy function with the potential of mean force (PMF) defined by the finer system. However, practical challenges arise due to the lack of analytical expressions for the PMF and algorithmic limitations in parameterizing CG force fields. To address these challenges, a machine learning-based approach is proposed, utilizing graph neural networks (GNNs) to represent the solvation free energy and potential contrasting for parameter optimization. We demonstrate the effectiveness of the approach by deriving a transferable GNN implicit solvent model using 600,000 atomistic configurations of six proteins obtained from explicit solvent simulations. The GNN model provides solvation free energy estimations much more accurately than state-of-the-art implicit solvent models, reproducing configurational distributions of explicit solvent simulations. We also demonstrate the reasonable transferability of the GNN model outside of the training data. Our study offers valuable insights for deriving systematically improvable implicit solvent models and CG force fields from a bottom-up perspective.



Downloaded via 43.206.209.194 on November 9, 2024 at 18:29:22 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

Coarse-grained (CG) force fields play a pivotal role in advancing molecular dynamics (MD) simulations of biomolecules by bridging the gap between computational efficiency and biological realism.^{1–38} In these simulations, biomolecules are represented with simplified models that group several atoms into a single interaction site, reducing the computational burden without sacrificing essential structural and dynamical information. This approach is crucial for studying large and complex biomolecular systems over biologically relevant timescales, which would be infeasible with atomistic simulations. CG force fields enable researchers to explore phenomena such as protein folding,^{1–5} protein aggregation,^{10–17,39–41} and large-scale conformational changes,^{42,43} providing insights into biological processes that are otherwise inaccessible. Understandably, there is great interest in developing methodologies for building accurate and transferable CG force fields.^{7,18–26}

Tremendous advancement in the theoretical framework provides guiding principles for developing accurate CG force fields from atomistic simulation data, i.e., the bottom-up approach.^{18–23} For example, one can consider the mapping \mathbf{M} that transforms the atomic Cartesian coordinates \mathbf{r} to the CG configuration \mathbf{R} . The CG energy function $W(\mathbf{R})$ can be explicitly defined as

$$\exp[-W(\mathbf{R})/k_B T] \int d\mathbf{r} \exp[-u(\mathbf{r})/k_B T] \times \delta(\mathbf{M}(\mathbf{r}) - \mathbf{R}) \quad (1)$$

where $u(\mathbf{r})$ is the potential of the assumed atomistic model. A CG force field that reproduces the potential of mean force (PMF), $W(\mathbf{R})$, is referred to as *thermodynamically consistent* with the atomistic force field.^{18–23} Being thermodynamically consistent indicates that statistical properties predicted by the CG force field will match the atomistic results.

However, despite the elegant theoretical framework, developing thermodynamically consistent CG force fields from the bottom up faces several challenges in practice. First, while pairwise potentials have found great success in explicit solvent atomistic models,^{44–46} it has long been known that many-body effects become prominent when the degrees of freedom are reduced.^{1,34,35,47–49} The PMF is inherently a many-body potential. Owing to the lack of analytical

Received: September 19, 2023

Revised: October 26, 2023

Accepted: October 31, 2023

Published: November 16, 2023



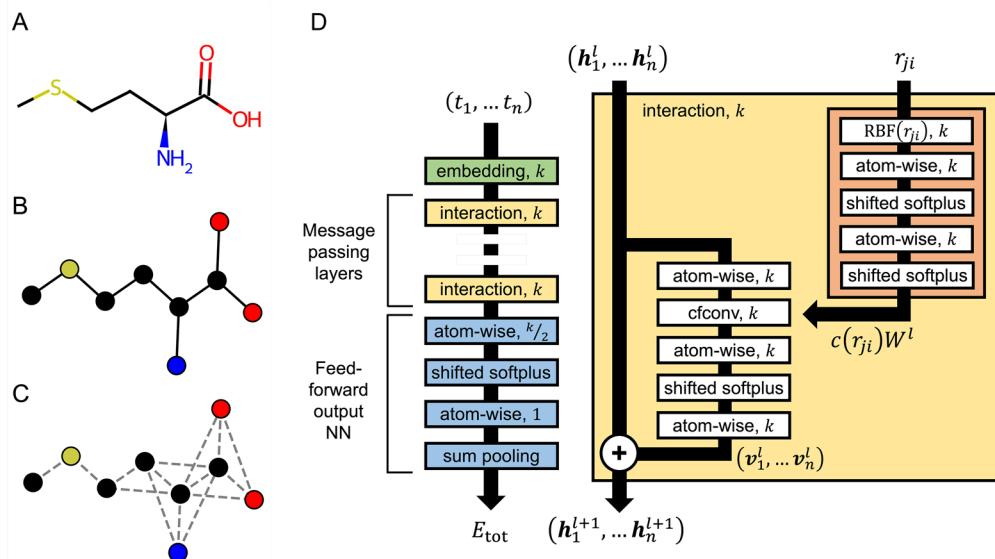


Figure 1. Graph neural networks allow convenient encoding of structural and chemical information of molecules for energy prediction. (A-C) Illustration of the mapping from a chemical representation of the methionine residue (A) to a graph with edges representing covalent bonds (B) and a fully connected graph of methionine with edges representing all atoms within $r_{\text{cut}} = 2.5 \text{ \AA}$ (C). Hydrogens are omitted from the graphs for clarity. (D) Illustration of the SchNet architecture.⁵⁵ The embedding layer (green) assigns k learnable parameters, $\mathbf{h}_i^0 = \{h_{i,1}^0, h_{i,2}^0, \dots, h_{i,k}^0\}$, for each atom i in the protein based on its atom type, t_i . The featurizations are updated by a series of message passing layers, i.e., interaction blocks (yellow), that account for the chemical environment surrounding individual atoms. See the Methods section for additional discussion of interaction blocks. After N_{IB} interaction blocks, the final featurizations, $(\mathbf{h}_1^{N_{\text{IB}}}, \dots, \mathbf{h}_n^{N_{\text{IB}}})$, are provided to the energy-predicting NN (blue), which computes the total potential energy for the protein as $E_{\text{tot}} = \sum_{i=1}^n E_i$.

expressions, how to best approximate $W(\mathbf{R})$ to achieve the desired accuracy is unclear. Second, despite significant methodological developments,^{22–25,47,50,51} parameterizing CG force fields with atomistic simulation data remains nontrivial. Existing algorithms often face challenges in achieving thermodynamic consistency due to error accumulation. Often, they can leverage only small data sets due to computational cost, limiting the transferability of the resulting force field.

We propose a machine learning (ML) approach to tackle both challenges. First, we employed graph neural networks (GNNs) to represent the CG force field. GNNs trained as neural network (NN) potentials use the Cartesian coordinates of the entire molecule to calculate energy. Crucially, their high expressibility enables them to parameterize complex functional relationships, capturing many-body effects. Second, we utilize a recently developed force field optimization method, potential contrasting,²⁵ to parameterize GNNs from atomistic simulation data. Potential contrasting optimizes the overlap between CG and atomistic configurational distributions, aiming to ensure thermodynamic consistency. Moreover, it is highly parallelizable and computationally efficient, allowing the use of large training sets that are essential for deriving transferable force fields.

We use the ML approach to develop transferable implicit solvent models based on 600,000 conformations from explicit solvent atomistic simulations of six proteins. An implicit solvent model is a high-resolution CG model that approximates the solvation free energy, i.e., $W(\mathbf{R})$, obtained by integrating the solvent degrees of freedom.^{52–54} We demonstrate that the SchNet architecture⁵⁵ provides a precise representation of the solvation free energy. When parameterized, the SchNet implicit solvent model, along with the gas-phase force field, can accurately reproduce the free energy

profiles from explicit solvent atomistic simulations. The SchNet model performs well for proteins outside the training set, indicating its transferability. We also introduce several methodological advancements necessary for parameterizing and simulating the GNN-based force field, such as a pretraining procedure that enhances the robustness and stability of MD simulations and a reweighting scheme that enables efficient free energy calculations. Our results will be of wide interest to researchers building accurate coarse-grained models. The methodological insights should also help advance the training of GNNs as transferable force fields.

RESULTS

GNNs Enable Transferable Learning of Solvation Free Energy. ML techniques have proven to be powerful for force field learning as they capture many-body effects.^{56–58} While early ML force fields relied on hand-picked descriptors,^{59,60} GNNs with only Cartesian coordinates and atom types as inputs have become increasingly popular for force field parameterization.^{61–78} Molecules in 3D space can be conveniently abstracted into a graph where nodes are atoms. Edges represent covalent and noncovalent interactions⁷⁹ (Figures 1A–C). Such a graph is naturally invariant to atom indexing, and the rotational and translational symmetry of energy functions can be encoded straightforwardly.

Following Chen et al.,⁷⁸ we adopt the SchNet architecture introduced by Schütt et al.⁵⁵ to represent the PMF of CG models. As depicted in Figure 1D, the network architecture can be broadly divided into three parts: an embedding layer that assigns the initial featurization to each atom, multiple message-passing layers that update the featurization, and, finally, a feed-forward NN that predicts an atomic contribution to the total system energy from the updated featurization. Featurization defines the characteristics associated with a specific atom type.

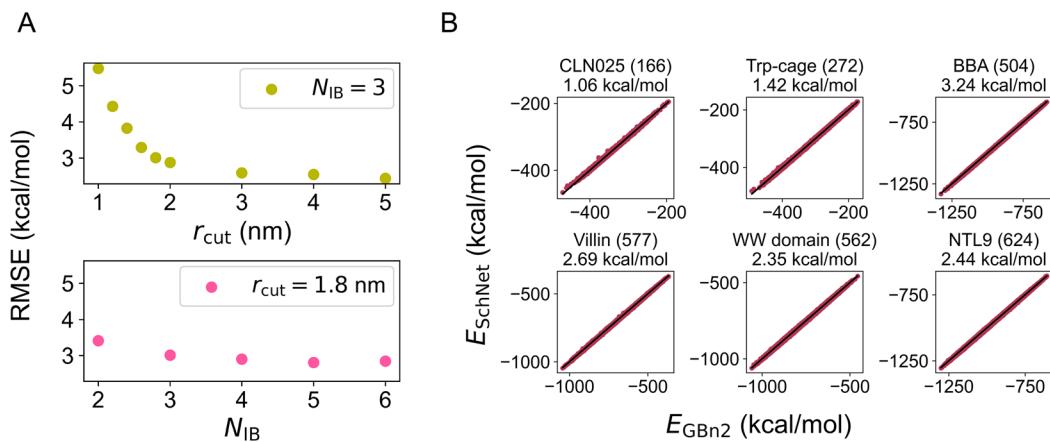


Figure 2. SchNet accurately reproduces the many-body solvation free energy for proteins of various sizes. (A) Average root-mean-squared error (RMSE) across all configurations and all proteins with different hyperparameters r_{cut} and N_{IB} in SchNet design. The hyperparameter indicated in the legend was held constant, while the hyperparameter indicated on the x-axis was varied. (B) Scatter plots detailing the agreement between the energy from the best performing SchNet model ($N_{\text{IB}} = 3$, $r_{\text{cut}} = 5 \text{ nm}$), E_{SchNet} , and the GBn2 solvation free energy, E_{GBn2} , for different configurations of each protein. The number of atoms in each protein is shown in parentheses. Average RMSEs are shown for each protein.

The message-passing step is typically performed iteratively, allowing nodes to gather information from increasingly distant neighbors, thus accounting for interactions among atoms. These layers are frequently referred to as interaction blocks. Feature updates explicitly make use of structural information inferred from Cartesian coordinates. Therefore, the SchNet model resembles a typical energy function that transforms the Cartesian coordinates of the entire molecule, along with knowledge about its chemical makeup, into a single value representing the energy. Furthermore, backward propagation can compute forces for MD simulations.

Most evaluations of SchNet and other GNNs focus on organic molecules with fewer than 21 atoms.^{55,80} Therefore, we need to assess the ability of SchNet to predict interaction energies for much larger protein molecules. We selected six proteins: chignolin CLN025 (CLN025), Trp-cage, BBA, Villin, the WW domain, and NTL9. The smallest protein in this data set has 166 atoms. For each protein, we collected 200,000 conformations from MD simulations (see Supporting Information and Tables S1–S3 for details) and computed the solvation free energy, E_{GBn2} , using the GB-neck2 (GBn2) implicit solvent model.⁸¹ Since E_{GBn2} is a many-body potential, its prediction is a challenging problem for SchNet.

We explored several SchNet architectures by focusing on adjusting two hyperparameters that directly impact the message passing layers: the cutoff distance r_{cut} and the number of interaction blocks N_{IB} . Here, r_{cut} defines the nearest neighbors connected with edges to a given atom and used to compute feature updates in each interaction block, while N_{IB} determines the number of iterations for feature updates. Increasing r_{cut} while keeping the number of interaction blocks constant at $N_{\text{IB}} = 3$ results in a continuously decreasing training root-mean-squared error (RMSE), as illustrated in Figure 2A. Next, after selecting a well-performing smaller $r_{\text{cut}} = 1.8 \text{ nm}$, we explored the effect of varying N_{IB} . Figure 2B suggests that N_{IB} has a less pronounced impact on training RMSE compared to r_{cut} . For the most accurate model ($N_{\text{IB}} = 3$, $r_{\text{cut}} = 5 \text{ nm}$), SchNet indeed performs well for all proteins. Higher RMSE values are observed for increasingly larger proteins, as shown in Figure 2B.

However, as noted in previous works,^{21,82} training RMSE alone does not directly indicate the accuracy of an ML force

field in simulations. Notably, the training RMSEs for the best- and worst-performing hyperparameter combinations differ by only approximately 2.5 kcal/mol. To further evaluate different SchNet architectures, we devised an efficient scheme for computing another important quantity in the simulations: free energy. Specifically, we computed the free energy profile as a function of the RMSD from the folded state using the SchNet model and a free energy perturbation method.⁸³ Instead of conducting computationally intensive simulations with the SchNet model, we estimated the free energy using configurations sampled from RMSD-biased umbrella sampling in the GBn2 implicit solvent. This scheme is detailed in the Methods section.

The estimated free energy profiles depicted in Figure 3 suggest that, despite a mere ~ 2.5 kcal/mol RMSE difference,

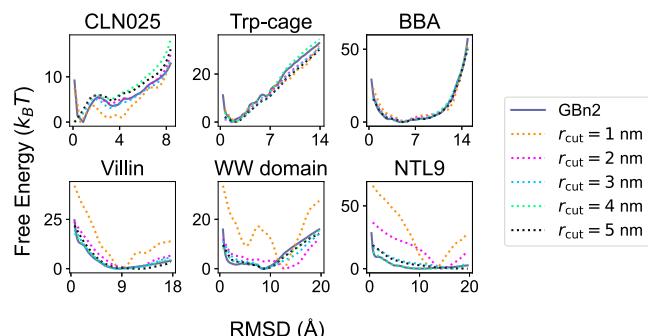


Figure 3. Free energy profiles for SchNet models trained to fit E_{GBn2} with different values of r_{cut} . The number of interaction blocks is held constant at $N_{\text{IB}} = 3$. The x axes in the plots correspond to the RMSD values from the folded structures. For reference, the free energy profiles computed from umbrella simulations with the implicit solvent model, GBn2, are also shown as solid lines. Results for the SchNet model were determined using a reweighting scheme as detailed in the Methods section.

the SchNet model with $r_{\text{cut}} = 3 \text{ nm}$ significantly outperforms the model with $r_{\text{cut}} = 1 \text{ nm}$ in replicating the results from the implicit GBn2 solvent model. Conversely, despite similar training RMSEs, the SchNet model with $r_{\text{cut}} = 5 \text{ nm}$ and $N_{\text{IB}} = 3$ proves to be significantly more accurate than the model with

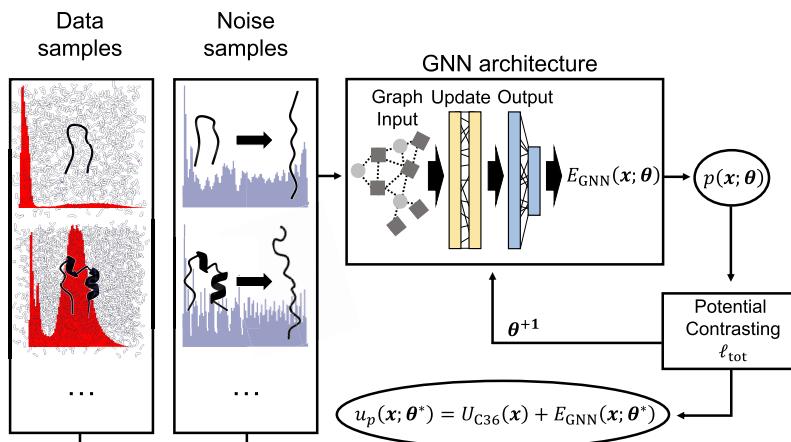


Figure 4. Illustration of the potential contrasting method for training SchNet to learn the solvation free energy from explicit solvent simulation data. For each protein, two ensembles of configurations are prepared to represent data and noise distributions. Data configurations are produced from unbiased long time scale explicit solvent simulations, while noise configurations are obtained from umbrella simulations using the implicit solvent model, GBn2. The Boltzmann distributions of the two ensembles are denoted as $p(x;\theta) \propto \exp(-\beta[u_p(x;\theta)])$ and $q(x) \propto \exp(-\beta[u_q(x)])$, respectively. $u_p(x;\theta) = U_{C36}(x) + E_{\text{SchNet}}(x;\theta)$, where U_{C36} is the gas-phase energy evaluated using the CHARMM force field^{91–93} and $E_{\text{SchNet}}(x;\theta)$ is the solvation free energy, defined by the SchNet model with parameters θ , for configuration x . $u_q(x)$ represents the potential energy for the generalized ensemble constructed by combining configurations from different umbrella windows (see Supporting Information for more details). The loss function in potential contrasting, ℓ_{tot} , is designed to maximize $p(x;\theta)$ over data samples and simultaneously minimize its values over noises. Thus, the optimal parameters θ^* will ensure that $p(x;\theta^*)$ closely matches the Boltzmann distribution from which the atomistic explicit solvent configurations were drawn from.

$r_{\text{cut}} = 1.8$ nm and $N_{\text{IB}} = 6$ (Figure S1). Free energy profiles were also computed for SchNet models with N_{IB} ranging from 2 to 6, but little difference is observed between these models (Figure S2). These estimated profiles thus offer a more precise assessment of model performance than errors in energy alone, showcasing the utility of our reweighting scheme for hyperparameter optimization. These results confirm that SchNet can perform quite effectively for proteins of up to 624 atoms.

Contrastive Learning of Solvation Free Energy from Explicit Solvent Simulations. Having demonstrated that SchNet can replicate the solvation free energy defined by the GBn2 model, our next objective is to further enhance its accuracy. Specifically, we aim to optimize SchNet to better replicate the conformational distribution obtained from explicit solvent atomistic simulations. Current implicit solvent models, including GBn2, rely on analytical expressions for the solvation free energy.^{52,84–89} While these models are computationally efficient, they often entail significant approximations, thereby limiting the achievable level of accuracy. Consequently, there is a compelling need to develop more accurate and systematically improvable representations of the solvation free energy. GNNs have recently emerged as alternatives to traditional implicit solvent functional forms.^{78,90}

We employed a method known as potential contrasting²⁵ to optimize SchNet using the configurational ensemble of the six aforementioned proteins gathered from explicit solvent simulations conducted by Lindorff-Larsen et al.⁹⁴ (refer to Figure 4 and Methods). Potential contrasting, recently introduced by us, serves as a means to optimize the parameters of a CG force field by extending the noise contrastive estimation method.⁹⁵ This approach is related to the adversarial-residual-coarse-graining method in that it formulates force field optimization as classification between data and noise samples.⁹⁶ Potential contrasting offers computational efficiency, high parallelizability, and the ability to leverage a substantial training data set. It is worth noting that, due to the

unavailability of the exact solvation free energy for each solute configuration, employing RMSE as a loss function is unfeasible. Additionally, while force matching has been used to train GNNs as implicit solvent models, pre-existing data can only be used if it is possible to extract the forces exerted on the solute atoms by the solvent.^{78,90} Without forces or exact solvation free energies available to fit, the training process transitions into an unsupervised learning problem; hence, our adoption of potential contrasting.

With the hyperparameters that were demonstrated to be effective for fitting SchNet to E_{GBn2} ($N_{\text{IB}} = 3$, $r_{\text{cut}} = 5$ nm), we observed that the SchNet model optimized by using potential contrasting surpasses GBn2 by a significant margin. As illustrated in Figure 5, the free energy profiles for the SchNet model, computed with the reweighting scheme, align with explicit solvent results. The lower accuracy for BBA, WW

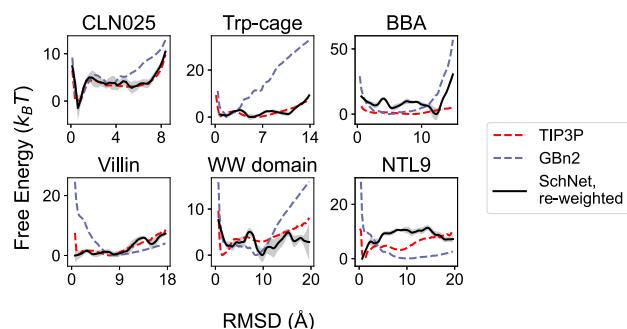


Figure 5. Comparison between the free energy profiles computed from explicit solvent simulations (TIP3P⁹⁷), implicit solvent simulations (GBn2), and the optimized SchNet model with hyperparameters $N_{\text{IB}} = 3$ and $r_{\text{cut}} = 5$ nm. Due to the sharpness of the minimum in the SchNet free energy profile for CLN025, the free energy profile was uniformly shifted downward by $1.5 k_B T$ for better visualization. Results for the SchNet model were determined using a reweighting scheme as detailed in the Methods section.

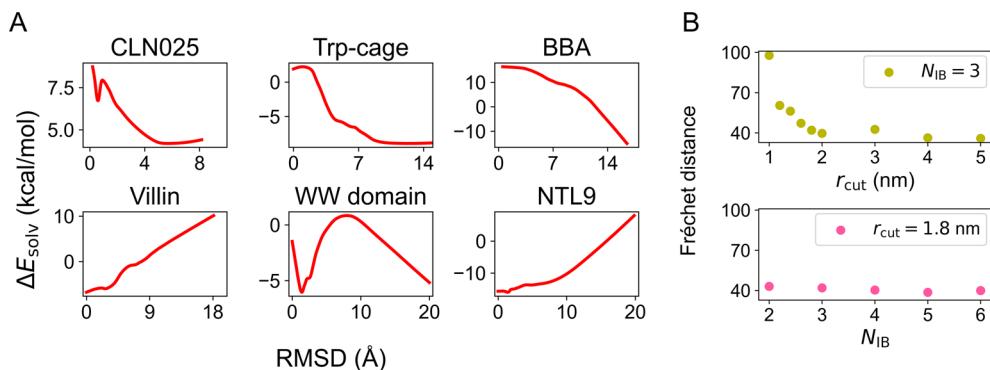


Figure 6. Solvation free energy learned from explicit solvent simulations by SchNet improves over that of the implicit solvent model, GBn2. (A) Average solvation free energy difference between the optimized SchNet and GBn2 models, i.e., $\Delta E_{\text{solv}} = E_{\text{SchNet}} - E_{\text{GBn2}}$ as a function of the RMSD to the folded structures. We averaged over the configurations from explicit solvent simulations to compute ΔE_{solv} at a given RMSD value and applied the LOWESS⁹⁸ algorithm to smooth the curves. (B) The performance of the SchNet models was optimized with potential contrasting at various hyperparameters. The hyperparameter indicated in the legend was held constant, while the hyperparameter indicated on the x -axis was varied. The performances are evaluated using the discrete Fréchet distance between atomistic and SchNet free energy profiles as a function of the RMSD across all proteins. The SchNet free energy profiles used to compute these distances are shown in Figures S6 to S8.

domain, and NTL9 arises mainly from limits to the flexibility of the SchNet architecture as discussed in the *Supporting Information* and Figures S3 to S4. Remarkably, we find that if we reduce the total number of training configurations from 200,000 to 20,000 configurations per protein, potential contrasting is still capable of training SchNet as an implicit solvent that is more accurate than GBn2. With the exception of NTL9, the free energy profiles generated from this data-limited model (Figure S5) are similar to those presented in Figure 5, supporting the data efficiency of potential contrasting.

Figure 6A illustrates that SchNet is effectively learning solvation free energies that, in comparison to GBn2, result in the destabilization of folded and compact states in favor of unfolded/extended states. This trend is notably reversed for Villin and NTL9, where GBn2 fails to capture the free energy minimum at the folded states (as shown in Figure 5). Given that it is a well-known characteristic of GB-based implicit solvents to generally overstabilize folded/compact states, these findings support the notion that SchNet is assimilating chemically meaningful information to accurately replicate explicit solvent configurational distributions.

Pretraining for Robust Parameter Optimization. We further explored the role of the hyperparameters, r_{cut} and N_{IB} , on the accuracy of models optimized by potential contrasting. To assess the performance of the SchNet implicit solvent models, we computed the discrete Fréchet distance⁹⁹ between their free energy profiles and those from the explicit solvent simulations. The discrete Fréchet distance is a widely used measure of similarity between two arbitrary curves.

For all of the hyperparameter combinations we examined, the sum of these distances across all proteins in the training set is presented in Figure 6B. For reference, the sum of the discrete Fréchet distances between the explicit solvent and GBn2 free energy profiles for all proteins is 130.98. This metric suggests that all SchNet models outperform GBn2 in reproducing explicit solvent simulations. Consistent with the conclusions drawn from training SchNet to fit E_{GBn2} , we observed that increasing r_{cut} leads to more accurate implicit solvent models. However, maintaining r_{cut} at 1.8 nm while varying the number of interaction blocks from $N_{IB} = 3$ does not significantly affect accuracy.

We observed that a pretraining procedure is essential for robust parameter optimization using potential contrasting. In this procedure, we initialize the SchNet parameters using those obtained from training the model to fit E_{GBn2} . Additionally, we keep the parameters within the interaction blocks constant during the training process with potential contrasting. Without the initial pretraining, the free energy profiles depicted in Figure S9 indicate that the resulting model performs poorly, failing to replicate the results from explicit solvent simulations for all six proteins. Similarly, as shown in Figure S10, not holding the parameters within the interaction blocks constant also results in a poor performance of the SchNet model.

Furthermore, we have discovered that pretraining enables stable ML-MD simulations. These simulations combine forces derived from the gas-phase CHARMM force field with those from the SchNet solvation free energy for the dynamic evolution of the equation of motion. For instance, we successfully conducted hundreds of nanoseconds of simulation for smaller proteins, such as CLN025 and Trp-cage, using SchNet models optimized with pretraining. Larger proteins impose greater demands on SchNet hyperparameters, but stable ML-MD simulations can still be performed with larger values for N_{IB} . Conversely, SchNet models optimized without pretraining do not support long time scale simulations, regardless of the hyperparameter configuration.

It is noteworthy that optimization with or without pretraining yielded comparable values of the loss function, indicating that it is an inadequate indicator of SchNet model performance. Upon examining the predicted solvation free energies E_{SchNet} , it becomes evident that our pretraining scheme prevents SchNet from assigning unphysical energies (see Figures S11 and S12).

Our pretraining procedure is closely related to transfer learning, which involves transferring the knowledge a model has acquired from one task to another.¹⁰⁰ Transfer learning is a widely employed technique to enhance the accuracy of machine learning models across a diverse range of problems.^{101–103} By pretraining SchNet to fit E_{GBn2} for configurations of our training set proteins, the interaction blocks learn a set of parameters that optimally update the featurization for each atom. By holding these parameters constant while training with potential contrasting, we are effectively trans-

ferring the “knowledge” of how to optimally update atomic featurizations. Thus, training with potential contrasting serves to adjust the solvation free energy predictions by further optimizing only the parameters of the initial embedding layer and the energy-predicting NN.

SchNet Enables Highly Accurate ML-MD Simulations.

We conducted hybrid ML-MD simulations to further assess the quality of the SchNet models trained using potential contrasting. Such simulations have the potential to yield more accurate free energy estimations by directly sampling configurations using the SchNet solvation free energy. In our investigation, we focused on a computationally more efficient SchNet architecture featuring three interaction blocks ($N_{IB} = 3$) and a cutoff distance of $r_{cut} = 1.8$ nm to define nearest neighbors. The free energy profiles computed using our reweighting scheme indicate that this setup delivers accurate results, particularly for small proteins.

We performed 260 and 600 ns of RMSD-biased umbrella sampling simulations for CLN025 and Trp-cage, respectively. Each individual umbrella simulation represents 13 and 20 ns for CLN025 and Trp-cage, respectively. Details of these simulations can be found in the *Methods* section and the *Supporting Information*. Simulating larger systems with ML implicit solvents is computationally demanding and was not attempted. Additional details regarding the free energy calculation methods are provided in the *Supporting Information*.

We observed that the free energy profiles computed from ML-MD simulations (Figure 7A) closely aligned with the reweighted results (Figure S13) and exhibited strong agree-

ment with the explicit solvent results. Additionally, the free energy profiles computed as a function of the radius of gyration (Rg) of all backbone heavy atoms from ML-MD simulations matched those from explicit solvent simulations (Figure S14A). These simulation outcomes affirm the accuracy of the derived SchNet model and validate the reweighting scheme as a valuable tool for generating rapid and efficient approximations of free energy profiles.

ML-MD Simulations Show Transferability of SchNet.

We have demonstrated that a single SchNet model, trained using the potential contrasting method, accurately captures the conformational distribution of all six proteins in the training set (Figure 5). This remarkable achievement underscores the potential of GNNs as attractive alternatives for generating transferable representations of the solvation free energy. Next, we delve into assessing the transferability of the SchNet model by applying it to three proteins outside of the training set.

We first considered two variants of chignolin CLN025,¹⁰⁴ denoted as 1ua0 and T8P. 1ua0 contains glycine residues at both termini, in contrast to tyrosine in the CLN025 sequence.¹⁰⁵ T8P, on the other hand, features a threonine-to-proline mutation at position 8 of 1ua0, which discourages the native hairpin conformation.¹⁰⁶ These variants share 80% and 70% sequence identities, respectively, with CLN025. For both proteins, 260 ns umbrella sampling simulations were conducted with each individual simulation lasting for 13 ns.

We calculated free energy profiles through simulations using the SchNet, GBn2, and TIP3P explicit solvent. Detailed simulation information can be found in the *Methods* section and the *Supporting Information*. As depicted in Figure 7B, for the 1ua0 system, SchNet once again outperforms GBn2 in predicting the relative stability of native and unfolded states. However, for the T8P mutant, SchNet’s performance is slightly inferior to that of GBn2, although it still accurately predicts the impact of the mutation, namely, the destabilization of the β -hairpin structure observed in 1ua0. Free energy profiles as a function of Rg support these conclusions (Figure S14B).

We further evaluated the SchNet model for an intrinsically disordered peptide (IDP), which is drastically different from the fast-folding proteins in our training set. The IDP corresponds to a structurally active, 10-mer fragment of the C-Jun amino-terminal kinase-interacting protein 1 (JIP1), and has been studied extensively by Ojaghlu et al.¹⁰⁷ with explicit solvent simulations. It differs significantly from CLN025 with a sequence identity of 30%. For this protein, 600 ns umbrella sampling simulations were conducted with each individual simulation lasting for 30 ns.

Free energy profiles as a function of Rg (Figure 8A) suggest that similar to GBn2, the SchNet model overstabilizes the collapsed state. However, from the two-dimensional free energy profiles computed using the time-lagged independent component analysis (TICA) coordinates, we found that the SchNet model correctly captures the three basins of the compact states explored in explicit solvent simulations (Figure 8B-C). Therefore, SchNet successfully predicted the right conformations, although it provided inaccurate free energies. Increasing the training set size, especially with the addition of disordered proteins, could further improve its transferability.

CONCLUSIONS AND DISCUSSION

We have introduced a comprehensive strategy for parametrizing GNN implicit solvent models from explicit solvent simulation data with the potential contrasting method. Several

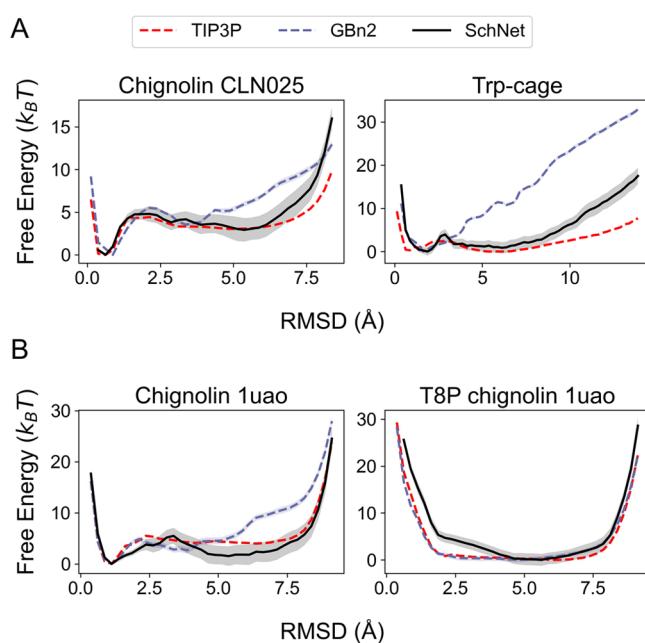


Figure 7. Comparison between the free energy profiles computed from explicit solvent simulations (TIP3P), implicit solvent simulations (GBn2), and the optimized SchNet model. Results for the SchNet model were determined using ML-MD umbrella simulations. (A) Free energy profiles for chignolin CLN025 and Trp-cage. TIP3P and GBn2 results are identical to those shown in Figure 5. (B) Free energy profiles for chignolin 1ua0 and a T8P mutant. TIP3P and GBn2 free energy profiles were computed from umbrella simulations. See the *Supporting Information* and Tables S4 to S5 for simulation details.

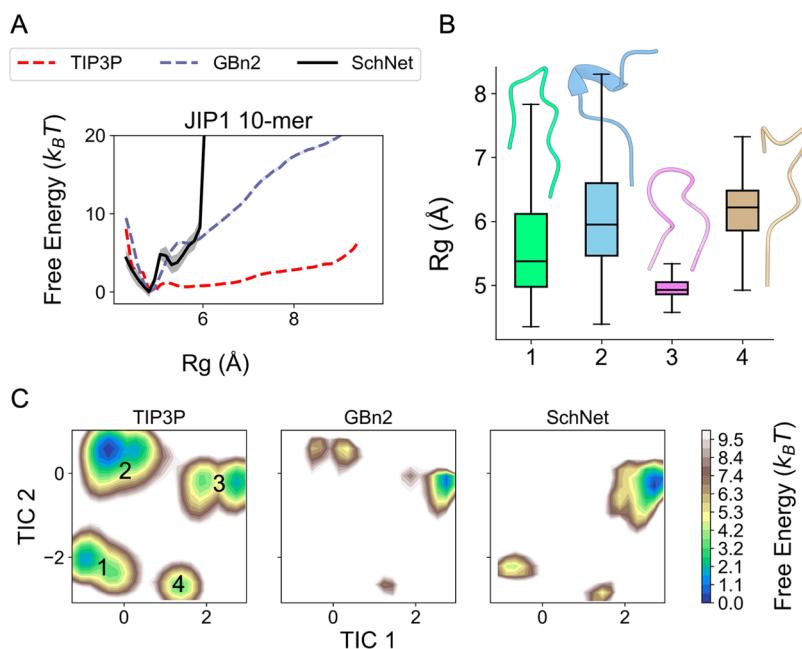


Figure 8. Comparing one- and two-dimensional free energy profiles between explicit solvent (denoted as TIP3P), GBn2 and the optimized SchNet model for the JIP1 10-mer. All explicit solvent results were generated by Ojaghlu et al.,¹⁰⁷ while results for GBn2 and the SchNet model were generated from umbrella simulations. (A) Free energy profiles with respect to R_g . (B) R_g box plots and representative structures for each basin in the two-dimensional free energy profiles shown in part C. (C) Two-dimensional free energy profiles generated through TICA dimensionality reduction.^{108–110} The procedure used by Ojaghlu et al.¹⁰⁷ was followed (see Supporting Information for more detail).

methodological advancements are crucial for enhancing the performance of the resulting models in describing the distribution of the entire configurational ensemble. These include a pretraining strategy for initializing parameters and a reweighting scheme that facilitates efficient evaluation of hyperparameter combinations. Our results demonstrate that optimized SchNet models can accurately replicate the free energy profiles obtained from explicit solvent simulations for all six proteins in the training set. Simultaneously achieving such results for all six proteins represents a significant achievement, showcasing the model's transferability. Furthermore, we have shown that the optimized SchNet model demonstrates reasonable transferability outside the training set.

When used with our reweighting scheme, the parametrized SchNet model can offer immediate practical utility by delivering more precise solvation free energies than existing implicit solvent models. As an illustration, one can contemplate conducting simulations with GBn2 and subsequently applying the SchNet model to reweight the obtained configurations, thereby bringing more accurate estimations of statistical properties.

However, ML-MD simulations using our SchNet implicit solvent model are considerably more computationally intensive than those using the TIP3P explicit solvent. Table 1 compares the simulation speeds for CLN025, Trp-cage, and BBA in a vacuum, explicit solvent, GBn2, and our SchNet implicit solvent model. Simulations with the SchNet implicit solvent ($N_{IB} = 3$, $r_{cut} = 1.8$ nm) are approximately 10 to 15 times slower than those in explicit solvent. These sluggish simulation speeds hinder effective conformational sampling for large proteins. We encountered several technical challenges that impeded more efficient simulations with our SchNet implicit solvent model, which we expect to be resolved in the near future. First, we faced software dependency issues that prevented us from running simulations on the GPU. Second,

Table 1. Simulation Speed for CLN025, Trp-cage, and BBA with Different Solvent Models^a

System	CLN025	Trp-cage	BBA
Vacuum (ns/day)	103	80.3	89.1
GBn2 (ns/day)	51.2	43.1	51.8
TIP3P (ns/day)	32.4	35.0	22.1
SchNet (ns/day)	3.62	3.05	1.50 ^b

^aSimulations for each protein were run on an Intel Xeon Platinum 8260 (4, 12, and 48 cores for CLN025, Trp-cage, and BBA, respectively) processor, and all simulation parameters are shown in Table S6. ^bUnstable with 5 fs time step used here, a 0.5 fs time step is required for ns simulations.

our GNN implicit solvent implementation relies on slow communication between two software packages, PyTorch¹¹¹ and OpenMM.¹¹² The native integration of GNN models into MD simulation packages could circumvent this communication bottleneck and has the potential to significantly accelerate the simulations.

While our work primarily focused on parametrizing transferable implicit solvent models, the methodologies outlined here have broad applicability. They can be readily applied to develop lower-resolution CG models with fewer particles per amino acid. Utilizing GNNs as a flexible architecture for representing CG force fields means that accuracy does not have to be constrained by the approximations often required to derive analytical expressions for the PMF. Moreover, using potential contrasting enables robust, data-efficient parametrization of GNNs to replicate the conformational distribution observed in atomistic simulations. The combination of these two techniques has the potential to revolutionize coarse-graining practices, paving the way for the creation of transferable CG force fields that can rival the accuracy of explicit solvent atomistic simulations.

METHODS

Detailed Implementation of SchNet. As shown in Figure 1D, the SchNet architecture⁵⁵ consists of three parts: an embedding layer (with learnable parameters θ_0) for featurization, message-passing layers (with learnable parameters θ_F) for feature updates, and a feed-forward NN (with learnable parameters θ_E) for energy prediction. The embedding layer assigns a feature vector of k learnable parameters, h_i^0 , to the i th atom based on its atom type t_i . Each message-passing layer, i.e., the interaction block, further updates the feature vector of each atom by aggregating information from neighbors. Updates use continuous filters W learned from the distances between pairs of neighboring atoms. These filters are scaled by the distance between neighbors i and j , r_{ji} , using a Behler-style cosine cutoff function $c(r_{ji})$. The layers that compute W are shown in orange in Figure 1D. After N_{IB} total interaction blocks, the updated featurizations are inputted into the energy-predicting NN, which computes the total potential energy for the protein as a sum of individual contributions from each atom. Overall, the set of learnable parameters for SchNet is $\theta = \{\theta_0, \theta_F, \theta_E\}$. With the hyperparameters used to produce our optimal implicit solvent model (detailed in the Supporting Information and Table S7–S8), each atom type is assigned $k = 32$ learnable parameters in the embedding layer, each interaction block contains 5,248 learnable parameters, and the energy-predicting NN contains 545 learnable parameters.

We mostly followed the implementation of SchNet from the PyTorch Geometric package,¹¹³ with only one change. Specifically, instead of featurizing each atom based on the atomic numbers, we used the atom types defined by the CHARMM force field. For example, we apply numeric labels to the atom type strings (outside of SchNet) and then use the embedding layer to assign h_0 to the numeric labels. As such, encoding atom types does not require additional layers. Like atomic numbers and partial charges, CHARMM atom types are available *a priori*. Since these atom types are defined based on neighboring atoms, the intention of this change was to speed convergence toward optimal parameters θ^* .

GNN Optimization with Potential Contrasting. Potential contrasting builds upon the noise-contrastive estimation method⁹⁵ to parametrize force fields. While it shares similarities with the maximum likelihood method, instead of directly maximizing the probability defined by the Boltzmann distribution of the force field on training data (i.e., configurations produced from atomistic simulations), the potential contrasting method optimizes force field parameters to best differentiate data configurations from noise configurations. The use of noise configurations circumvents the need to evaluate the partition function of the force field, thus, significantly reducing the computational cost of optimization by eliminating the need for iterative sampling. The potential contrasting objective function, the loss function we minimize when parametrizing a force field, is detailed in the Supporting Information.

We generated noise configurations for each protein by conducting umbrella simulations employing the GBn2 implicit solvent model.⁸¹ These simulations involved applying biases on the RMSD from the folded structure, starting from 0.0 Å and incrementing by intervals of 0.5 Å up to their maximum RMSD, for each protein. Each simulation ran for a minimum duration of 155 ns, and configurations were sampled at intervals of 40, 50, or 200 ps during the simulations (see Table

S2). We merged the configurations obtained from different umbrella windows using free energy biases computed via MBAR.¹¹⁴ Further information on these umbrella simulations and the construction of the noise distribution is presented later in this section and the Supporting Information.

We conducted an initial pretraining phase before training SchNet with configurations from explicit solvent simulations using the potential contrasting method.²⁵ This pretraining served two primary purposes: first, to assess SchNet's ability to accurately predict the rather complex function representing the GBn2 solvation free energy, denoted as E_{GBn2} , for all the proteins in the training set; and second, to learn the optimal parameters θ_F^* within the message passing layers. Consequently, the framework for performing feature updates became fully optimized through pretraining. The set of parameters acquired during this pretraining phase is denoted as $\theta^{pre} = \{\theta_0^{pre}, \theta_F^{pre}, \theta_E^{pre}\}$. This pretraining process can be seen as imparting SchNet with the physical insights offered by the GBn2 implicit solvent model. As indicated in the text and illustrated in Figures S9, S10, and S12, in the absence of pretraining, potential contrasting, on its own, cannot ensure the satisfactory performance of the resulting SchNet models.

When training with potential contrasting, SchNet parameters were initialized as θ^{pre} . In the subsequent optimization steps, only θ_F^{pre} and θ_E^{pre} were permitted to change, while θ_0^{pre} remained fixed. This approach effectively transferred the “knowledge” acquired by the message-passing layers during pretraining.

Training concluded after 120 epochs, corresponding to the number of times the loss function has been evaluated over the entire training set, for pretraining, and after 30 epochs for potential contrasting. This termination criterion ensured that both loss functions reached a plateau. The precise expressions of the loss functions are available in the Supporting Information. Both pretraining and potential contrasting employed the Adam optimizer.¹¹⁵ Learning rates of 1×10^{-3} and 1×10^{-4} were used during pretraining and potential contrasting, respectively.

Reweighting Configurations to Estimate Free Energy Profiles.

Evaluating the accuracy of our SchNet implicit solvent model can be challenging, primarily because we lack precise values for the solvation free energy that the model aims to reproduce. Conducting extensive sampling through ML-MD simulations enables the creation of low-dimensional probability distributions of selected collective variables, which can then be compared to the explicit solvent simulation results. While this evaluation technique is reliable, it is often computationally demanding. To address this, we propose a straightforward reweighting scheme using the free energy perturbation method⁸³ to swiftly generate reasonably accurate estimates of our SchNet implicit solvent models' performance.

We introduce a weighting factor, denoted as $w(\mathbf{x}) = \frac{p(\mathbf{x}; \theta^*)}{q(\mathbf{x})} = \exp(-\beta[u_p(\mathbf{x}; \theta^*) - u_q(\mathbf{x})])$, for each noise configuration. Here, $u_q(\mathbf{x})$ represents the energy function of the noise distribution, while $u_p(\mathbf{x}; \theta^*)$ corresponds to the energy defined by the optimized SchNet. These weights are subsequently normalized so that $\max w(\mathbf{x}) = 1$. Using the noise configurations and their corresponding weighting factors, we construct a weighted histogram with L RMSD bins, where we choose L to ensure that the RMSD width of each bin is 0.25 Å.

Within each of the n_l configurations present in the l th bin, we compute the free energy, denoted as $F_l = -\ln\left(\frac{\sum_{k=1}^n w(x_k)}{N_{\text{total}}}\right)$.

To ensure that the set of free energies for all L bins, $\{F_l\}_{l=1}^L$, aligns with a physically meaningful scale, we uniformly shift it to ensure that $\min\{F_l\}_{l=1}^L = 0$ in units of $k_B T$. Recognizing that physically meaningful free energy should exhibit smoothness, we further apply the LOWESS⁹⁸ smoothing scheme to $\{F_l\}_{l=1}^L$. This process yields a reasonably accurate estimate of the free energy as a function of RMSD.

Additionally, we compute standard deviations for each reweighted free energy by employing the block bootstrapping¹¹⁶ method on the noise configurations and their corresponding energies. The block size was set as the number of configurations collected from 2 ns of simulation. Ten bootstrapping samples were generated to calculate ten estimates, and their average and standard deviation are used as the free energy profile and the error bars in the figures.

Molecular Dynamics Simulation Details. RMSD-biased umbrella simulations were conducted for all proteins, with RMSD computed based on the heavy atoms in the peptide backbone in relation to the folded structure of each protein. These simulations were carried out using either the GBn2 or the SchNet implicit solvent model, with either the CHARMM^{91–93} or AMBER¹¹⁷ gas-phase force field. The Langevin middle integrator was used for all simulations.¹¹⁸ All umbrella simulations were performed with the OpenMM package¹¹² (version 8.0 beta). For simulations employing the SchNet model, the OpenMM-Torch package¹¹⁹ (version 1.0 beta) was also utilized. Further simulation details can be found in the Supporting Information.

Inputs were generated using CHARMM-GUI.^{44,120,121} Trajectory processing and the calculation of RMSD and Rg were accomplished using the MDTraj¹²² and pytraj¹²³ packages. Additional analyses were performed with the aid of the statsmodels^{98,124} (version 0.14.0) and scipy^{125,126} packages. All structure visualization was performed using ChimeraX.¹²⁷

Configurations generated from the umbrella simulations were utilized to calculate the free energy as a function of the RMSD or Rg. We employed the MBAR¹¹⁴ and UWHAM¹²⁸ methods for these calculations, utilizing the FastMBAR package.¹²⁹ To assess the reliability of our results, we determined standard deviations for the free energy profiles through the block bootstrapping technique.¹¹⁶ Following the same procedure used for the reweighted free energy profiles, the block size was set as the number of configurations collected from 2 ns of simulation, and ten bootstrap samples were created to compute average and standard deviation of the free energy profile. For more detailed information regarding the free energy calculations, refer to the Supporting Information.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscentsci.3c01160>.

Simulation methodology, data selection, machine learning methodology, noise potential construction, free energy calculation details, further discussion of the results displayed in Figure 5, and additional figures (PDF)

AUTHOR INFORMATION

Corresponding Author

Bin Zhang – Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, United States; orcid.org/0000-0002-3685-7503; Email: binz@mit.edu

Authors

Justin Arias – Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, United States; orcid.org/0000-0001-5045-9926

Xinqiang Ding – Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, United States; orcid.org/0000-0002-4598-8732

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acscentsci.3c01160>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (Grant R35GM133580). We acknowledge the Massachusetts Institute of Technology (MIT) Supercloud and Lincoln Laboratory Supercomputing Center for providing high-performance computing resources. We thank D. E. Shaw Research for sharing their atomistic simulation data. J.A. was the recipient of the MIT Dean of Science Fellowship and the National Science Foundation Graduate Research Fellowship.

REFERENCES

- (1) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing. *J. Phys. Chem. B* **2012**, *116*, 8494–8503.
- (2) Nielsen, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L. Coarse grain models and the computer simulation of soft materials. *J. Phys.: Condens. Matter* **2004**, *16*, R481.
- (3) Hyeon, C.; Thirumalai, D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* **2011**, *2*, 487.
- (4) Darré, L.; Machado, M. R.; Brandner, A. F.; González, H. C.; Ferreira, S.; Pantano, S. SIRAH: A Structurally Unbiased Coarse-Grained Force Field for Proteins with Aqueous Solvation and Long-Range Electrostatics. *J. Chem. Theory Comput.* **2015**, *11*, 723–739.
- (5) Machado, M. R.; Barrera, E. E.; Klein, F.; Sónora, M.; Silva, S.; Pantano, S. The SIRAH 2.0 Force Field: Altius, Fortius, Citius. *J. Chem. Theory Comput.* **2019**, *15*, 2719–2733.
- (6) Latham, A. P.; Zhang, B. Improving Coarse-Grained Protein Force Fields with Small-Angle X-ray Scattering Data. *J. Phys. Chem. B* **2019**, *123*, 1026–1034.
- (7) Latham, A. P.; Zhang, B. Consistent Force Field Captures Homologue-Resolved HP1 Phase Separation. *J. Chem. Theory Comput.* **2021**, *17*, 3134–3144.
- (8) Latham, A. P.; Zhang, B. Unifying coarse-grained force fields for folded and disordered proteins. *Curr. Opin. Struct. Biol.* **2022**, *72*, 63–70.
- (9) Lin, X.; Zhang, B. Explicit Ion Modeling Predicts Physicochemical Interactions for Chromatin Organization. *bioRxiv*, May 18, 2023. DOI: [10.1101/2023.05.16.541030](https://doi.org/10.1101/2023.05.16.541030) (accessed Aug. 11, 2023)
- (10) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLOS Comput. Biol.* **2018**, *14*, No. e1005941.

- (11) Joseph, J. A.; Reinhardt, A.; Aguirre, A.; Chew, P. Y.; Russell, K. O.; Espinosa, J. R.; Garaizar, A.; Colleopardi-Guevara, R. Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nat. Comput. Sci.* **2021**, *1*, 732–743.
- (12) Grime, J. M. A.; Dama, J. F.; Ganser-Pornillos, B. K.; Woodward, C. L.; Jensen, G. J.; Yeager, M.; Voth, G. A. Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly. *Nat. Commun.* **2016**, *7*, 11568.
- (13) Mansbach, R. A.; Ferguson, A. L. Coarse-Grained Molecular Simulation of the Hierarchical Self-Assembly of π -Conjugated Optoelectronic Peptides. *J. Phys. Chem. B* **2017**, *121*, 1684–1706.
- (14) Shmilovich, K.; Mansbach, R. A.; Sidky, H.; Dunne, O. E.; Panda, S. S.; Tovar, J. D.; Ferguson, A. L. Discovery of Self-Assembling π -Conjugated Peptides by Active Learning-Directed Coarse-Grained Molecular Simulation. *J. Phys. Chem. B* **2020**, *124*, 3873–3891.
- (15) 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. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2111696118.
- (16) Wu, H.; Wolynes, P. G.; Papoian, G. A. AWSEM-IDP: A Coarse-Grained Force Field for Intrinsically Disordered Proteins. *J. Phys. Chem. B* **2018**, *122*, 11115–11125.
- (17) Jin, J.; Yu, A.; Voth, G. A. Temperature and Phase Transferable Bottom-up Coarse-Grained Models. *J. Chem. Theory Comput.* **2020**, *16*, 6823–6842.
- (18) Jin, J.; Pak, A. J.; Durumeric, A. E. P.; Loose, T. D.; Voth, G. A. Bottom-up Coarse-Graining: Principles and Perspectives. *J. Chem. Theory Comput.* **2022**, *18*, 5759–5791.
- (19) Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, 090901.
- (20) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* **2008**, *128*, 244114.
- (21) Durumeric, A. E. P.; Charron, N. E.; Templeton, C.; Musil, F.; Bonneau, K.; Pasos-Trejo, A. S.; Chen, Y.; Kelkar, A.; Noé, F.; Clementi, C. Machine learned coarse-grained protein force-fields: Are we there yet? *Curr. Opin. Struct. Biol.* **2023**, *79*, 102533.
- (22) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.
- (23) Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **2008**, *129*, 144108.
- (24) Köhler, J.; Chen, Y.; Krämer, A.; Clementi, C.; Noé, F. Flow-Matching: Efficient Coarse-Graining of Molecular Dynamics without Forces. *J. Chem. Theory Comput.* **2023**, *19*, 942–952.
- (25) Ding, X.; Zhang, B. Contrastive Learning of Coarse-Grained Force Fields. *J. Chem. Theory Comput.* **2022**, *18*, 6334–6344.
- (26) Yang, W.; Templeton, C.; Rosenberger, D.; Bittracher, A.; Nüske, F.; Noé, F.; Clementi, C. Slicing and Dicing: Optimal Coarse-Grained Representation to Preserve Molecular Kinetics. *ACS Cent. Sci.* **2023**, *9*, 186–196.
- (27) de Pablo, J. J. Coarse-Grained Simulations of Macromolecules: From DNA to Nanocomposites. *Annu. Rev. Phys. Chem.* **2011**, *62*, 555–574.
- (28) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18*, 382–388.
- (29) Best, R. B. Computational and theoretical advances in studies of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2017**, *42*, 147–154.
- (30) Saunders, M. G.; Voth, G. A. Coarse-Graining Methods for Computational Biology. *Annu. Rev. Biophys.* **2013**, *42*, 73–93.
- (31) Jarin, Z.; Newhouse, J.; Voth, G. A. Coarse-Grained Force Fields from the Perspective of Statistical Mechanics: Better Understanding of the Origins of a MARTINI Hangover. *J. Chem. Theory Comput.* **2021**, *17*, 1170–1180.
- (32) Yu, A.; Pak, A. J.; He, P.; Monje-Galvan, V.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Amaro, R. E.; Voth, G. A. A multiscale coarse-grained model of the SARS-CoV-2 virion. *Biophys.* **2021**, *120*, 1097–1104.
- (33) Denesyuk, N. A.; Thirumalai, D. Coarse-Grained Model for Predicting RNA Folding Thermodynamics. *J. Phys. Chem. B* **2013**, *117*, 4901–4911.
- (34) Wang, J.; Charron, N.; Husic, B.; Olsson, S.; Noé, F.; Clementi, C. Multi-body effects in a coarse-grained protein force field. *J. Chem. Phys.* **2021**, *154*, 164113.
- (35) Zaporozhets, I.; Clementi, C. Multibody Terms in Protein Coarse-Grained Models: A Top-Down Perspective. *J. Phys. Chem. B* **2023**, *127*, 6920–6927.
- (36) Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N. E.; de Fabritiis, G.; Noé, F.; Clementi, C. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Cent. Sci.* **2019**, *5*, 755–767.
- (37) Majewski, M.; Pérez, A.; Thölke, P.; Doerr, S.; Charron, N. E.; Giorgino, T.; Husic, B. E.; Clementi, C.; Noé, F.; De Fabritiis, G. Machine learning coarse-grained potentials of protein thermodynamics. *Nat. Commun.* **2023**, *14*, 5739.
- (38) Loose, T. D.; Sahrmann, P. G.; Qu, T. S.; Voth, G. A. Coarse-Graining with Equivariant Neural Networks: A Path Towards Accurate and Data-Efficient Models. *arXiv*, 2023, arXiv:2309.09290 (accessed Sept. 17, 2023).
- (39) Latham, A. P.; Zhang, B. On the stability and layered organization of protein-DNA condensates. *Biophys. J.* **2022**, *121*, 1727–1737.
- (40) Liu, S.; Wang, C.; Latham, A. P.; Ding, X.; Zhang, B. OpenABC enables flexible, simplified, and efficient GPU accelerated simulations of biomolecular condensates. *PLoS Comput. Biol.* **2023**, *19*, No. e1011442.
- (41) Latham, A. P.; Zhu, L.; Sharon, D. A.; Ye, S.; Willard, A. P.; Zhang, X.; Zhang, B. Frustrated Microphase Separation Produces Interfacial Environment within Biological Condensates. *bioRxiv*, Apr. 2, 2023. DOI: 10.1101/2023.03.30.534967 (accessed Aug. 14, 2023).
- (42) Lin, X.; Leicher, R.; Liu, S.; Zhang, B. Cooperative DNA looping by PRC2 complexes. *Nucleic Acids Res.* **2021**, *49*, 6238–6248.
- (43) Liu, S.; Lin, X.; Zhang, B. Chromatin fiber breaks into clutches under tension and crowding. *Nucleic Acids Res.* **2022**, *50*, 9738–9747.
- (44) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (45) Ponder, J. W.; Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.* **2003**, *66*, 27–85.
- (46) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 3499–3509.
- (47) Noid, W. G. In *Systematic Methods for Structurally Consistent Coarse-Grained Models BT - Biomolecular Simulations: Methods and Protocols*; Monticelli, L., Salonen, E., Eds.; Methods Mol. Biol.; Humana Press: Totowa, NJ, 2013; pp 487–531.
- (48) Kar, P.; Feig, M. Recent advances in transferable coarse-grained modeling of proteins. *Adv. Protein Chem. Struct. Biol.* **2014**, *96*, 143–180.
- (49) Liwo, A.; Baranowski, M.; Czaplewski, C.; Golaś, E.; He, Y.; Jagieła, D.; Krupa, P.; Maciejczyk, M.; Makowski, M.; Mozolewska, M. A.; et al. A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *J. Mol. Model.* **2014**, *20*, 2306.
- (50) Ercolessi, F.; Adams, J. B. Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *EPL* **1994**, *26*, 583.
- (51) Mullinax, J. W.; Noid, W. G. Generalized Yvon-Born-Green Theory for Molecular Systems. *Phys. Rev. Lett.* **2009**, *103*, 198104.

- (52) Onufriev, A. V.; Case, D. A. Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.* **2019**, *48*, 275–296.
- (53) Onufriev, A. V.; Izadi, S. Water models for biomolecular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, No. e1347.
- (54) Akkermans, R. L. C. Solvation Free Energy of Regular and Azeotropic Molecular Mixtures. *J. Phys. Chem. B* **2017**, *121*, 1675–1683.
- (55) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet – A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148*, 241722.
- (56) Stocker, S.; Gasteiger, J.; Becker, F.; Günemann, S.; Margraf, J. T. How robust are modern graph neural network potentials in long and hot molecular dynamics simulations? *Mach. Learn.: Sci. Technol.* **2022**, *3*, 045010.
- (57) Gastegger, M.; Marquetand, P. In *Molecular Dynamics with Neural Network Potentials BT - Machine Learning Meets Quantum Physics*; Schütt, K. T., Chmiela, S., von Lilienfeld, O. A., Tkatchenko, A., Tsuda, K., Müller, K.-R., Eds.; Lecture Notes in Physics; Springer: Cham, 2020; pp 233–252.
- (58) Ricci, E.; Giannakopoulos, G.; Karkaletsis, V.; Theodorou, D. N.; Vergadou, N. *Developing Machine-Learned Potentials for Coarse-Grained Molecular Simulations: Challenges and Pitfalls*. In *Proceedings of the 12th Hellenic Conference on Artificial Intelligence*; ACM: New York, NY, 2022; pp 1–6.
- (59) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
- (60) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8*, 3192–3203.
- (61) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678–3693.
- (62) Gasteiger, J.; Groß, J.; Günemann, S. Directional Message Passing for Molecular Graphs. *arXiv*, Apr. 5, 2022. DOI: [10.48550/arXiv.2003.03123](https://doi.org/10.48550/arXiv.2003.03123) (accessed July 13, 2023).
- (63) Liu, Y.; Wang, L.; Liu, M.; Zhang, X.; Oztekin, B.; Ji, S. Spherical Message Passing for 3D Graph Networks. *arXiv*, Nov. 24, 2022. DOI: [10.48550/arXiv.2102.05013](https://doi.org/10.48550/arXiv.2102.05013) (accessed Oct. 16, 2023).
- (64) Thomas, N.; Smidt, T.; Kearnes, S.; Yang, L.; Li, L.; Kohlhoff, K.; Riley, P. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. *arXiv*, May 18, 2018. DOI: [10.48550/arXiv.1802.08219](https://doi.org/10.48550/arXiv.1802.08219) (accessed July 13, 2023).
- (65) Anderson, B.; Hy, T.-S.; Kondor, R. Cormorant: Covariant Molecular Neural Networks. *arXiv*, Nov. 25, 2019. DOI: [10.48550/arXiv.1906.04015](https://doi.org/10.48550/arXiv.1906.04015) (accessed July 13, 2023).
- (66) Köhler, J.; Klein, L.; Noé, F. Equivariant Flows: Exact Likelihood Generative Learning for Symmetric Densities. *arXiv*, Oct. 26, 2020. DOI: [10.48550/arXiv.2006.02425](https://doi.org/10.48550/arXiv.2006.02425) (accessed Oct. 16, 2023).
- (67) Fuchs, F. B.; Worrall, D. E.; Fischer, V.; Welling, M. SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks. *arXiv*, Nov. 24, 2020. DOI: [10.48550/arXiv.2006.10503](https://doi.org/10.48550/arXiv.2006.10503) (accessed July 13, 2023).
- (68) Gasteiger, J.; Becker, F.; Günemann, S. GemNet: Universal Directional Graph Neural Networks for Molecules. *arXiv*, Apr. 5, 2022. DOI: [10.48550/arXiv.2106.08903](https://doi.org/10.48550/arXiv.2106.08903) (accessed July 13, 2023).
- (69) Schütt, K. T.; Unke, O. T.; Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. *arXiv*, June 7, 2021. DOI: [10.48550/arXiv.2102.03150](https://doi.org/10.48550/arXiv.2102.03150) (accessed July 13, 2023).
- (70) Jing, B.; Eismann, S.; Suriana, P.; Townshend, R. J. L.; Dror, R. Learning from Protein Structure with Geometric Vector Perceptrons. *arXiv*, May 16, 2021. DOI: [10.48550/arXiv.2009.01411](https://doi.org/10.48550/arXiv.2009.01411) (accessed July 13, 2023).
- (71) Satorras, V. G.; Hoogeboom, E.; Welling, M. E(n) Equivariant Graph Neural Networks. *arXiv*, Feb. 16, 2022. DOI: [10.48550/arXiv.2102.09844](https://doi.org/10.48550/arXiv.2102.09844) (accessed July 19, 2023).
- (72) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E. (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **2022**, *13*, 2453.
- (73) Huang, W.; Han, J.; Rong, Y.; Xu, T.; Sun, F.; Huang, J. Equivariant Graph Mechanics Networks with Constraints. *arXiv*, Mar. 12, 2022. DOI: [10.48550/arXiv.2203.06442](https://doi.org/10.48550/arXiv.2203.06442) (accessed July 19, 2023).
- (74) Brandstetter, J.; Hesselink, R.; van der Pol, E.; Bekkers, E. J.; Welling, M. Geometric and Physical Quantities Improve E(3) Equivariant Message Passing. *arXiv*, Mar. 26, 2022. DOI: [10.48550/arXiv.2110.02905](https://doi.org/10.48550/arXiv.2110.02905) (accessed July 13, 2023).
- (75) Thölke, P.; de Fabritiis, G. TorchMD-NET: Equivariant Transformers for Neural Network based Molecular Potentials. *arXiv*, Apr. 23, 2022. DOI: [10.48550/arXiv.2202.02541](https://doi.org/10.48550/arXiv.2202.02541) (accessed July 13, 2023).
- (76) Wang, Y.; Chodera, J. D. Spatial Attention Kinetic Networks with E(n)-Equivariance. *arXiv*, Jan. 24, 2023. DOI: [10.48550/arXiv.2301.08893](https://doi.org/10.48550/arXiv.2301.08893) (accessed July 13, 2023).
- (77) Husic, B. E.; Charron, N. E.; Lemm, D.; Wang, J.; Pérez, A.; Majewski, M.; Krämer, A.; Chen, Y.; Olsson, S.; de Fabritiis, G.; et al. Coarse graining molecular dynamics with graph neural networks. *J. Chem. Phys.* **2020**, *153*, 194101.
- (78) Chen, Y.; Krämer, A.; Charron, N. E.; Husic, B. E.; Clementi, C.; Noé, F. Machine learning implicit solvation for molecular dynamics. *J. Chem. Phys.* **2021**, *155*, 084101.
- (79) Han, J.; Rong, Y.; Xu, T.; Huang, W. Geometrically Equivariant Graph Neural Networks: A Survey. *arXiv*, Feb. 22, 2022. DOI: [10.48550/arXiv.2202.07230](https://doi.org/10.48550/arXiv.2202.07230) (accessed July 13, 2023).
- (80) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **2017**, *3*, No. e1603015.
- (81) Nguyen, H.; Roe, D. R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 2020–2034.
- (82) Wellawatte, G. P.; Hocky, G. M.; White, A. D. Neural potentials of proteins extrapolate beyond training data. *J. Chem. Phys.* **2023**, *159*, 085103.
- (83) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (84) Anandakrishnan, R.; Izadi, S.; Onufriev, A. V. Why Computed Protein Folding Landscapes Are Sensitive to the Water Model. *J. Chem. Theory Comput.* **2019**, *15*, 625–636.
- (85) Zhou, R.; Berne, B. J. Can a continuum solvent model reproduce the free energy landscape of a β -hairpin folding in water? *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777–12782.
- (86) Roe, D. R.; Okur, A.; Wickstrom, L.; Hornak, V.; Simmerling, C. Secondary Structure Bias in Generalized Born Solvent Models: Comparison of Conformational Ensembles and Free Energy of Solvent Polarization from Explicit and Implicit Solvation. *J. Phys. Chem. B* **2007**, *111*, 1846–1857.
- (87) Shao, Q.; Zhu, W. How Well Can Implicit Solvent Simulations Explore Folding Pathways? A Quantitative Analysis of α -Helix Bundle Proteins. *J. Chem. Theory Comput.* **2017**, *13*, 6177–6190.
- (88) Nymeyer, H.; García, A. E. Simulation of the folding equilibrium of α -helical peptides: A comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13934–13939.
- (89) Robinson, M. K.; Monroe, J. I.; Shell, M. S. Are AMBER Force Fields and Implicit Solvation Models Additive? A Folding Study with a Balanced Peptide Test Set. *J. Chem. Theory Comput.* **2016**, *12*, 5631–5642.
- (90) Katzberger, P.; Riniker, S. Implicit solvent approach based on generalized Born and transferable graph neural networks for molecular dynamics simulations. *J. Chem. Phys.* **2023**, *158*, 204101.
- (91) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. J. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the

- Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (92) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71–73.
- (93) Croitoru, A.; Park, S.-J.; Kumar, A.; Lee, J.; Im, W.; MacKerell, A. D.; Aleksandrov, A. Additive CHARMM36 Force Field for Nonstandard Amino Acids. *J. Chem. Theory Comput.* **2021**, *17*, 3554–3570.
- (94) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517–520.
- (95) Gutmann, M.; Hyvärinen, A. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics; Volume 9 of JMLR: W&CP 9*. Chia Laguna Resort, Sardinia, Italy, 2010; pp 297–304.
- (96) Durumeric, A. E. P.; Voth, G. A. Adversarial-residual-coarse-graining: Applying machine learning theory to systematic molecular coarse-graining. *J. Chem. Phys.* **2019**, *151*, 124110.
- (97) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (98) Cleveland, W. S. Robust Locally Weighted Regression and Smoothing Scatterplots. *J. Am. Stat. Assoc.* **1979**, *74*, 829–836.
- (99) Eiter, T.; Mannila, H. *Computing Discrete Frechet Distance*; Technische Universität Wien: Wien, Austria, 1994 (accessed Aug. 31, 2023).
- (100) Hosna, A.; Merry, E.; Gyalmo, J.; Alom, Z.; Aung, Z.; Azim, M. A. Transfer learning: a friendly introduction. *J. Big Data* **2022**, *9*, 102.
- (101) Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*; Miami, FL, June 20–25, 2009; pp 248–255.
- (102) Open AI. GPT-4 Technical Report. *arXiv*, Mar. 27, 2023. DOI: [10.48550/arXiv.2303.08774](https://doi.org/10.48550/arXiv.2303.08774) (accessed July 11, 2023)
- (103) Bubeck, S.; Chandrasekaran, V.; Eldan, R.; Gehrke, J.; Horvitz, E.; Kamar, E.; Lee, P.; Lee, Y. T.; Li, Y.; Lundberg, S. et al. Sparks of Artificial General Intelligence: Early experiments with GPT-4. *arXiv*, Apr. 13, 2023. DOI: [10.48550/arXiv.2303.12712](https://doi.org/10.48550/arXiv.2303.12712) (accessed July 11, 2023).
- (104) Honda, S.; Yamasaki, K.; Sawada, Y.; Morii, H. 10 Residue Folded Peptide Designed by Segment Statistics. *Structure* **2004**, *12*, 1507–1518.
- (105) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (106) Maruyama, Y.; Koroku, S.; Imai, M.; Takeuchi, K.; Mitsutake, A. Mutation-induced change in chignolin stability from π -turn to α -turn. *RSC Adv* **2020**, *10*, 22797–22808.
- (107) Ojaghlu, N.; Airas, J.; McRae, L. M.; Taylor, C. A.; Miller, B. R.; Parish, C. A. Understanding the Structure and Apo Dynamics of the Functionally Active JIP1 Fragment. *J. Chem. Inf. Model.* **2021**, *61*, 324–334.
- (108) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; de Fabritiis, G.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **2013**, *139*, 015102.
- (109) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* **2013**, *9*, 2000–2009.
- (110) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525–5542.
- (111) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv* 2019, arXiv:1912.01703, accessed on 2023–06–12.
- (112) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simonett, A. C.; Harrigan, M. P.; Stern, C. D.; et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Comput. Biol.* **2017**, *13*, e1005659.
- (113) Fey, M.; Lenssen, J. E. Fast Graph Representation Learning with PyTorch Geometric. *arXiv*, Apr. 25, 2019. DOI: [10.48550/arXiv.1903.02428](https://doi.org/10.48550/arXiv.1903.02428) (accessed June 12, 2023).
- (114) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
- (115) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv*, Jan. 30, 2017. DOI: [10.48550/arXiv.1412.6980](https://doi.org/10.48550/arXiv.1412.6980) (accessed June 12, 2023).
- (116) Kunsch, H. R. The Jackknife and the Bootstrap for General Stationary Observations. *Ann. Stat.* **1989**, *17*, 1217–1241.
- (117) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (118) Zhang, Z.; Liu, X.; Yan, K.; Tuckerman, M. E.; Liu, J. Unified Efficient Thermostat Scheme for the Canonical Ensemble with Holonomic or Isokinetic Constraints via Molecular Dynamics. *J. Phys. Chem. A* **2019**, *123*, 6056–6079.
- (119) Eastman, P. OpenMM PyTorch Plugin (2023); <https://github.com/openmm/openmm-torch> (accessed Apr. 18, 2023).
- (120) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.
- (121) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405–413.
- (122) 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. *Biophys. J.* **2015**, *109*, 1528–1532.
- (123) Roe, D. R.; Cheatham, T. E. I. PTraj and CPPtraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.
- (124) Seabold, S.; Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python in Science Conference (SciPy 2010)*; Austin, TX, June 28–July 3, 2010; pp 92–96.
- (125) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272.
- (126) Scott, D. W. *Multivariate Density Estimation: Theory, Practice, and Visualization*; John Wiley & Sons, 2015.
- (127) Petersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30*, 70–82.
- (128) Tan, Z.; Gallicchio, E.; Lapelosa, M.; Levy, R. M. Theory of binless multi-state free energy estimation with applications to protein-ligand binding. *J. Chem. Phys.* **2012**, *136*, 144102.
- (129) Ding, X.; Vilseck, J. Z.; Brooks, C. L. A Fast Solver for Large Scale Multistate Bennett Acceptance Ratio Equations. *J. Chem. Theory Comput.* **2019**, *15*, 799–802.