Contrastive Learning of Coarse-Grained Force Fields

Xinqiang Ding and Bin Zhang*

Cite This: https://doi.org/10.1021/acs.jctc.2c00616

Read Online

ACCESS | 📊 Metrics & More | 📖 Article Recommendations | 🆂🆘 Supporting Information

ABSTRACT: Coarse-grained models have proven helpful for simulating complex systems over long time scales to provide molecular insights into various processes. Methodologies for systematic parametrization of the underlying energy function or force field that describes the interactions among different components of the system are of great interest for ensuring simulation accuracy. We present a new method, potential contrasting, to enable efficient learning of force fields that can accurately reproduce the conformational distribution produced with all-atom simulations. Potential contrasting generalizes the noise contrastive estimation method with umbrella sampling to better learn the complex energy landscape of molecular systems. When applied to the Trp-cage protein, we found that the technique produces force fields that thoroughly capture the thermodynamics of the folding process despite the use of only α-carbons in the coarse-grained model. We further showed that potential contrasting could be applied over large data sets that combine the conformational ensembles of many proteins to improve force field transferability. We anticipate potential contrasting as a powerful tool for building general-purpose coarse-grained force fields.
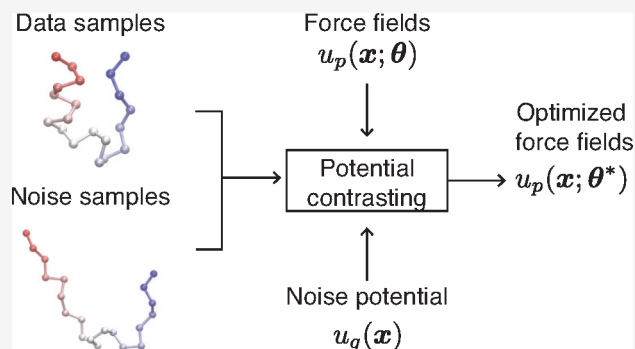
■ INTRODUCTION

Coarse-grained (CG) molecular dynamics simulations are computationally efficient and can simulate long time scale processes that are not accessible to all-atom simulations.[1−4] They are widely used for understanding dynamical processes in physics, chemistry, and biology.[5−17] The accuracy of these simulations depends on how well the force fields can describe the interactions among various components of the system under investigation. Therefore, algorithms and methodologies that can produce high-quality coarse-grained force fields (CGFFs) or CG potential energy are of key interest.

Several approaches have been introduced for systematically parametrizing CGFFs.[18,19] Top-down approaches often rely on a set of experimental structural or thermodynamic properties to fine-tune CGFFs and ensure the physical relevance of CG simulations.[3,5,18,20−24] On the other hand, bottom-up approaches learn CGFFs from an ensemble of atomistic configurations collected using simulations performed at finer resolution, typically with all-atom force fields.[25−28] From the configurational ensemble, various physical quantities and correlation functions can be computed to serve as targets for re-creation with CGFF.[22,23,29−40] In addition, CGFFs can also be optimized to enforce the statistical consistency between their corresponding Boltzmann distributions and the reference configurational distribution with variational methods.[41−44] The consistency is achieved when the CG potential energy matches the potential of the mean force dictated by the all-atom force field and the mapping from atomistic to CG configurations.

Existing variational methods optimize force field parameters by formulating and solving regression problems or maximizing the likelihood of observing the reference configurations. The force matching method[25,41] and the flow-matching method[45] belong to the former category and aim to minimize the difference between forces for CG coordinates calculated from the CGFF and target values estimated either from forces in all-atom simulations or from normalizing flows trained with conformations sampled in all-atom simulations. A perfect match in forces ensures that the CG energy function reproduces the potential of mean force. The generalized Yvon−Born−Green (gYBG) method[46] relates forces computed from the CGFF to structural correlation functions that can be directly estimated from configurations, allowing "force matching without forces". On the other hand, the relative entropy method,[42] or equivalently the maximum likelihood method,[34] directly optimizes the CG energy function by minimizing the relative entropy and maximizing the overlap between the CG Boltzmann distribution and the configurational distribution from all-atom simulations. The relative entropy is minimized when the CG energy function reproduces
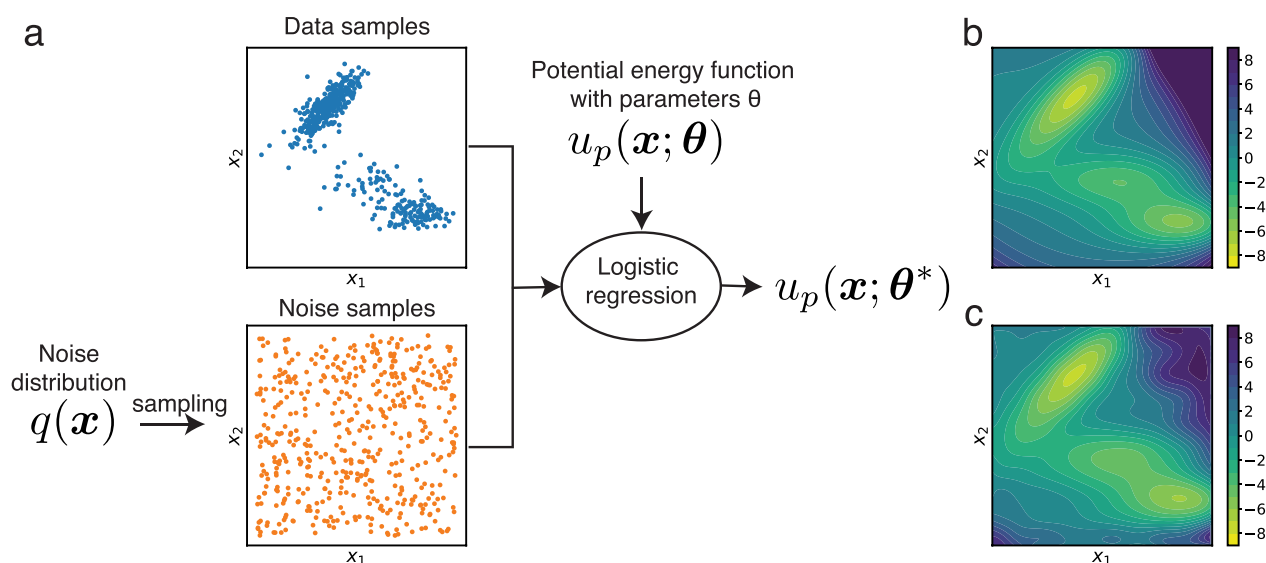
Figure 1. Noise contrastive learning accurately reproduces the Müller potential from sampled data. (a) Illustration of the noise contrastive estimation method. Representative data samples from Monte Carlo sampling of the Müller potential and noise samples from a noise distribution $q(\mathbf{x})$ are shown in blue and orange, respectively. The target data distribution is parametrized using a potential energy function, $u_p(\mathbf{x}; \boldsymbol{\theta})$, that is, $p(\mathbf{x}; \boldsymbol{\theta}) \propto \exp(-\beta u_p(\mathbf{x}; \boldsymbol{\theta}))$, where $\boldsymbol{\theta}$ is a set of parameters. $\boldsymbol{\theta}$ is optimized in a logistic regression to classify the data and noise samples. (b) Contour plot of the Müller potential. Energy is shown in units of $k_B T$. (c) Contour plot of the potential energy function $u_p(\mathbf{x}; \boldsymbol{\theta}^*)$ learned using noise contrastive estimation. Energy is shown in units of $k_B T$.

the potential of mean force and the CG Boltzmann distribution assigns high probabilities to configurations from all-atom simulations.

While existing force field parametrization methods have found great success in many applications, they are not without limitations. For example, the relative entropy method needs to run simulations to sample from trial CG potentials in every optimization step and can be computationally expensive. While the force matching method can learn the CG potential directly without iterative sampling, it often requires extra atomic force information, and the quality of the resulting potential can be sensitive to the accumulation of errors through the integration of the estimated force.

Here we developed a new variational method called potential contrasting for learning CGFFs and applied it for multiscale coarse-graining of protein folding. Potential contrasting generalizes the noise contrastive estimation method[47−49] to formulate force field parametrization into a classification problem. Input for the method is a target ensemble of protein conformations from all-atom simulations, and no atomic force information is required. When applied to the peptide Trp-cage, we found that potential contrasting can produce force fields that accurately reproduce the all-atom conformational ensemble and capture the complex folding landscape. The method also revealed the importance of including many-body potentials in CG models to describe protein biophysics with a reduced degree of freedom and implicit solvation. In addition, we showed that potential contrasting is computationally efficient and trivially parallelizable, enabling the parametrization of force fields using large data sets collected from multiple proteins.

■ METHODS

Potential contrasting combines a machine learning method called noise contrastive estimation[49] (NCE) with molecular simulation techniques. In this section, we first introduce NCE using the Müller potential[50] as an example. Then we present

how the NCE method is generalized and used in potential contrasting to learn CGFFs for protein folding.

Noise Contrastive Estimation. The NCE method[49] learns a probabilistic model on observed data. It is especially useful for learning unnormalized statistical models where the probability density function is only specified up to a normalization constant. It is evident that NCE is connected to bottom-up force field optimization, which aims to parametrize an energy function or an unnormalized Boltzmann distribution from data produced by all-atom simulations.

Here we use the Müller potential as an example to show how NCE helps to learn energy functions. Given a set of data (Figure 1a) drawn from the Müller potential with Markov chain Monte Carlo sampling, NCE aims to approximate their probability distribution with $p(\mathbf{x}; \boldsymbol{\theta})$ defined as $\log p(\mathbf{x}; \boldsymbol{\theta}) = -\beta[u_p(\mathbf{x}; \boldsymbol{\theta}) - F_p]$, where $u_p(\mathbf{x}; \boldsymbol{\theta})$ is the potential energy parametrized with $\boldsymbol{\theta}$ and F_p is the free energy. To optimize the parameters $\boldsymbol{\theta}$, NCE performs a logistic regression to discriminate the N_p data samples $\{\mathbf{x}_p^i\}_{i=1}^{N_p}$ from N_q noise samples $\{\mathbf{x}_q^i\}_{i=1}^{N_q}$ (Figure 1a) that are drawn from a noise distribution $q(\mathbf{x})$. Specifically, we assign binary labels of $y = 1$ and $y = 0$ to data and noise samples, respectively. NCE parametrizes the energy function by maximizing the following averaged log-likelihood of labels:

$$l(\boldsymbol{\theta}, F_p) = \frac{1}{N_p}\left[\sum_{i=1}^{N_p} \log P(y = 1|\mathbf{x}_p^i) + \sum_{i=1}^{N_q} \log P(y = 0|\mathbf{x}_q^i)\right]$$

(1)

with

$$P(y = 1|\mathbf{x}) = \frac{p(\mathbf{x}; \boldsymbol{\theta})}{p(\mathbf{x}; \boldsymbol{\theta}) + \nu q(\mathbf{x})} \quad \text{and}$$

$$P(y = 0|\mathbf{x}) = \frac{\nu q(\mathbf{x})}{p(\mathbf{x}; \boldsymbol{\theta}) + \nu q(\mathbf{x})}$$

(2)

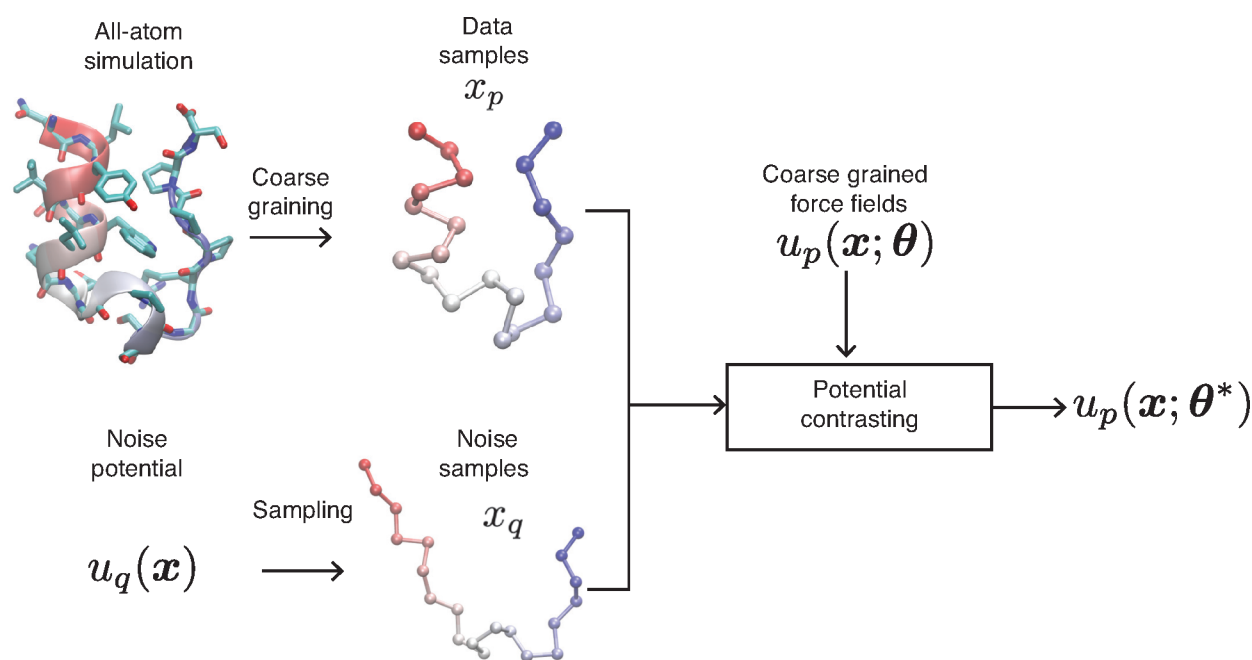where $\nu = P(y = 0)/P(y = 1) = N_q/N_p$.

Figure 2. Workflow of the potential contrasting method for learning coarse-grained force fields for the Trp-cage protein. The functional form of the potential energy function is chosen as $u_p(x; \theta)$, where θ is the set of parameters that need to be learned. The ensemble of conformations from all-atom simulations is converted into a coarse-grained ensemble using a predefined CG mapping and used as data samples. Here we map each amino acid into one coarse-grained particle at the C_α position. Based on the data samples, a noise potential $u_q(x)$ is designed and used to generate an ensemble of noise conformations and optimize the parameters θ with potential contrasting.

By definition, maximizing the above objective function forces the probability function $p(x; \theta)$ to assign high values to data samples (the first term) and low values to noise samples (the second term). In that regard, NCE is similar to the standard maximum likelihood estimation,[51] which assigns high probability on training data. Previous works[49] have proven that the solution θ^* for optimizing $l(\theta, F_p)$ behaves like the maximum likelihood estimator for large noise sample sizes and $p(x; \theta^*)$ converges to the true data distribution. The advantage of NCE over maximum likelihood estimation is that the free energy F_p is treated as a free parameter, and the optimization avoids the computationally expensive procedure for evaluating F_p rigorously. In addition, a nice property of $l(\theta, F_p)$ is that it is a concave function and has a unique maximum point if the potential energy function $u_p(x; \theta)$ is linear to θ.

Treating F_p as an independent variable, while being advantageous, also introduces a dependence of NCE's performance on the noise distribution because the noise sample size is always limited in practice. If $p(x; \theta)$ is a normalized density with conserved probability mass, as in the maximum likelihood optimization, increasing its value on data samples would implicitly decrease its value on regions outside the data. Such a balance of probability density is not guaranteed in NCE since $p(x; \theta)$ is not strictly normalized due to the approximate treatment of F_p. The use of a noise distribution remedies this issue by allowing an explicit probability minimization for the region covered by noise samples. While a comprehensive theory is still missing on designing optimal noise distributions,[52] we find that a useful guiding principle is to design the noise distribution such that it covers the phase space occupied by and surrounding the data samples. Without significant overlap between data and noise samples, the objective function, $l(\theta, F_p)$, can be trivially optimized by assigning high probability on data samples and low probability on noise samples without forcing $p(x; \theta)$ to capture the distributional structure within the data samples. In such cases, both terms in the objective function approach the constant zero, and the gradient on θ vanishes, hindering the optimization.

We parametrized the potential energy function $u_p(x; \theta)$ using a two-dimensional cubic spline[53] with 169 spline coefficients. The noise distribution $q(x)$ was chosen as the uniform distribution, and 500 000 samples were generated from both data and noise distributions. We learned the parameters θ by maximizing the NCE objective function $l(\theta, F)$ (eq 1) using the L-BFGS algorithm[54] (In practice, we minimize the negative of the NCE objective function). As shown in Figure 1c, $u_p(x; \theta^*)$ closely matches the underlying Müller potential (Figure 1b), supporting the effectiveness of NCE for learning potential energy functions.

Potential Contrasting for Learning Force Fields. Although theoretically sound, the current formulation of NCE is not effective for learning molecular force fields in practice. Therefore, we developed a new method named potential contrasting by generalizing NCE and introducing a customized way of defining the noise distribution. We present details of the method with applications to protein molecules in mind (Figure 2), for which the development of CGFFs is of great significance but has been challenging.[2,20] However, potential contrasting is general and can be applied to other types of molecules.

Generalizing NCE to Unnormalized Noise Distributions. The current formulation of NCE[49] requires specifying noise distributions for which the normalized probability density can be easily determined. This requirement restricts the choice of noise distributions because the normalization constant is difficult to compute for many probabilistic functions, including

Boltzmann distributions defined by complex potentials. Here we propose that this requirement is unnecessary and generalize NCE to use noise distributions specified with a potential energy function $u_q(\mathbf{x})$. Specifically, we set $q(\mathbf{x}) = e^{-\beta[u_q(\mathbf{x})-F_q]}$, where F_q is the free energy. Similarly to F_p, we treat F_q as an extra parameter in the optimization instead of computing its value explicitly. As a result, the logistic regression objective function in eq 1 becomes

$$l(\boldsymbol{\theta}, F_p, F_q) = \frac{1}{N_p}\left[\sum_{i=1}^{N_p} \log \frac{1}{1 + \nu\, e^{-\beta[u_q(\mathbf{x}_p^i)-u_p(\mathbf{x}_p^i;\boldsymbol{\theta})+F_p-F_q]}} \right.$$
$$\left. + \sum_{i=1}^{N_q} \log \frac{1}{1 + \nu^{-1}\, e^{-\beta[u_p(\mathbf{x}_q^i;\boldsymbol{\theta})-u_q(\mathbf{x}_q^i)+F_q-F_p]}}\right]$$
(3)

Because the value of $l(\boldsymbol{\theta}, F_p, F_q)$ in eq 3 only depends on $\boldsymbol{\theta}$ and the difference between F_p and F_q, we merge the two free energies into one free parameter, $\Delta F = F_p - F_q$, that is,

$$l(\boldsymbol{\theta}, \Delta F) = \frac{1}{N_p}\left[\sum_{i=1}^{N_p} \log \frac{1}{1 + \nu\, e^{-\beta[u_q(\mathbf{x}_p^i)-u_p(\mathbf{x}_p^i;\boldsymbol{\theta})+\Delta F]}} \right.$$
$$\left. + \sum_{i=1}^{N_q} \log \frac{1}{1 + \nu^{-1}\, e^{-\beta[u_p(\mathbf{x}_q^i;\boldsymbol{\theta})-u_q(\mathbf{x}_q^i)-\Delta F]}}\right]$$
(4)

Potential contrasting uses $l(\theta, \Delta F)$ as the objective function and optimizes the parameters $\boldsymbol{\theta}$ by maximizing $l(\theta, \Delta F)$. $\theta*$ is used to represent optimized parameters.

Defining the Noise Distribution for Learning CGFFs of Protein Folding. As mentioned before, the performance of NCE depends critically on the noise distribution, which should produce samples with sufficient overlap with the training data. For low dimensional systems, the uniform distribution used in the Müller potential example is a feasible choice for the noise. For complex systems such as protein molecules, uniform distributions suffer the dimensionality curse to cover the relevant phase space. Our generalization to unnormalized Boltzmann distributions significantly broadens the choices of noise distributions to facilitate producing complex molecular structures that resemble data samples. We further propose an umbrella sampling procedure to design noise potential energy functions and enhance overlap between noise and data samples.

We design the noise potential energy function such that the noise samples contain both folded and unfolded structures to match the configurational ensemble from all-atom simulations. For a given protein, we start with an energy function that includes terms for bonds, angles, and dihedral angles defined as

$$u_{\text{bonded}}(\mathbf{x}) = \sum_{i=1}^{L-1} \frac{1}{2}k_i(b_i - b_i^\circ)^2 + \sum_{i=1}^{L-2} S_{\text{angle}}(a_i; \mathbf{c}_i^a)$$
$$+ \sum_{i=1}^{L-3} S_{\text{dihedral}}(d_i; \mathbf{c}_i^d)$$
(5)

where L is the number of residues in the protein and b_i, a_i, and d_i represent the ith bond, angle, and dihedral angle, respectively. A quadratic function is used for energies on bonds. k_i and b_i° are the force constant and the equilibrium value for the ith bond, respectively. Cubic spline functions,

S_{angle} and S_{dihedral}, are used for energies on angles and dihedral angles (Figure S1). $\mathbf{c}_i^a$ and $\mathbf{c}_i^d$ are spline coefficients for the ith angle and dihedral angle, respectively. Using data samples, we fit each bonded energy term in $u_{\text{bonded}}(\mathbf{x})$ independently such that it will reproduce the marginal distribution of the corresponding degree of freedom from the data samples. To generate both folded and unfolded structures for noise samples, we further carried out umbrella sampling simulations[55,56] with $u_{\text{bonded}}(\mathbf{x})$ by biasing the root-mean-squared-deviation (RMSD) from the folded structure toward different values.

We combined configurations sampled from all M umbrella simulations to form a generalized ensemble which is used as the noise ensemble. The probability distribution of the generalized ensemble can be described as $p_{\text{gm}}(\mathbf{x}) \propto \sum_{i=1}^{M} \exp(-\beta[u_i(\mathbf{x}) + v_i])$,[57−61] where $u_i(\mathbf{x})$ is the energy function used in the ith umbrella simulation that includes both $u_{\text{bonded}}(\mathbf{x})$ and the bias function on the RMSD. v_i is adjustable energy that needs to be fitted and added to the potential energy $u_i(\mathbf{x})$ so that the relative free energies of the M states match the relative populations of structures sampled from these states. Correspondingly, the noise potential function can be computed as $u_q(\mathbf{x}) = -\beta^{-1} \log \sum_{i=1}^{M} \exp(-\beta[u_i(\mathbf{x}) + v_i])$. More details on the procedure are included in the Supporting Information.

Extending Potential Contrasting to Multiple Proteins. With the developments outlined above, potential contrasting can be used to parametrize CG energy functions for a specific protein by optimizing $l(\theta, \Delta F)$ defined in eq 4 (Figure 2). It can be further generalized to learn CG potential functions with shared parameters. Suppose that we can produce data and noise samples for a collection of proteins, the objective function to ensure that the CGFF reproduces the target configurational distribution for each protein can be defined as

$$l_{\text{tot}}(\boldsymbol{\theta}, \{\Delta F_k\}_{k=1}^K) = \sum_{k=1}^{K} \frac{1}{N_p^k}$$
$$\left[\sum_{i=1}^{N_p^k} \log \frac{1}{1 + \nu_k\, e^{-\beta[u_q^k(\mathbf{x}_p^{ki})-u_p(\mathbf{x}_p^{ki};\boldsymbol{\theta})+\Delta F_k]}} \right.$$
$$\left. + \sum_{i=1}^{N_q^k} \log \frac{1}{1 + \nu_k^{-1}\, e^{-\beta[u_p(\mathbf{x}_q^{ki};\boldsymbol{\theta})-u_q^k(\mathbf{x}_q^{ki})-\Delta F_k]}}\right]$$
(6)

The above expression is a sum of potential contrasting objective functions (eq 4) introduced for each protein. $\{\mathbf{x}_p^{ki}: i = 1, ..., N_p^k\}$ and $\{\mathbf{x}_q^{ki}: i = 1, ..., N_q^k\}$ represent the data and noise samples for the kth protein, with N_p^k and N_q^k corresponding to the respective sample sizes, and $\nu_k = N_q^k/N_p^k$. While the same energy function $u_p(\mathbf{x}; \boldsymbol{\theta})$ with shared parameters $\boldsymbol{\theta}$ is used, different noise potential energy functions, $u_q^k(\mathbf{x}_q^k)$, can be introduced for individual proteins. The aggregated objective function maintains the property of being concave if the CG energy function is linear to $\boldsymbol{\theta}$. We note that the objective function can be generalized straightforwardly if the CGFF introduces protein-specific parameters, as detailed in the Supporting Information.

■ RESULTS

Potential contrasting is a general method for force field parametrization. We focus on its application to protein folding
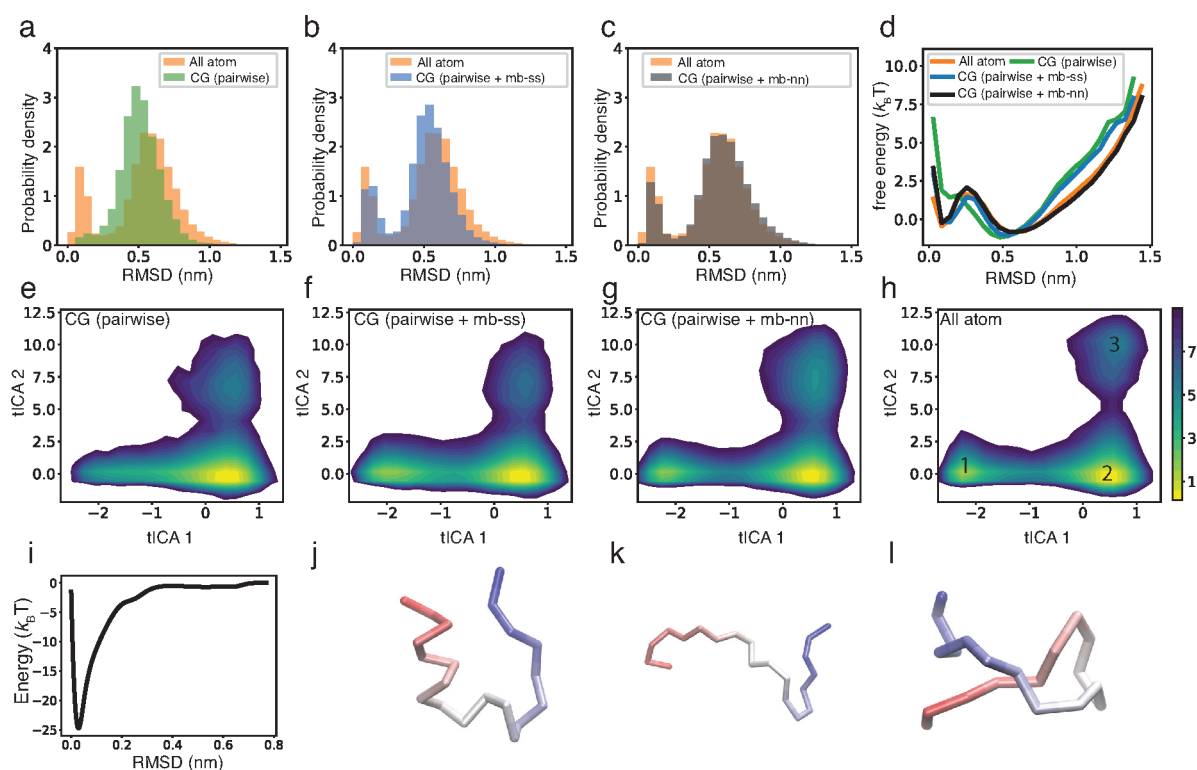
Figure 3. Parameterizing CGFFs for the Trp-cage protein using potential contrasting and all-atom simulations. (a−c) Distributions of RMSD with respect to the folded structure for conformations sampled from the all-atom simulation (orange) and CG simulations with learned CG potentials that differ in the representation of the nonbonded interactions (eqs 7−9). (d) Free energy profiles along the RMSD with respect to the folded structure for conformations sampled from the all-atom simulation and CG simulations with the three different learned potentials. (e−h) Free energy surfaces over the first two tICA coordinates for the all-atom simulation (h) and CG simulations with the three different learned potentials. The metastable states in panel h are labeled as 1, 2, and 3, with the corresponding representative structures shown in parts j, k, and l. (i) The many-body potential $u_{ss}^{mb}(\mathbf{x}; \varphi^*)$ as a function of the RMSD with respect to the folded α-helix structure.

and show that it can be used to optimize CGFFs for a specific protein and a collection of proteins. Given a sufficiently flexible functional form, the force field produced by potential contrasting can accurately reproduce the configurational distribution of all-atom simulations. We demonstrate its efficiency by optimizing over 12 proteins to derive CG potential functions with shared parameters. We provide potential contrasting's source code via the Github repository https://github.com/ZhangGroup-MITChemistry/PCCG.

Coarse-Grained Force Field for the Trp-Cage Protein. We applied potential contrasting to learn CGFFs for a 20 amino acid long peptide, Trp-cage. As detailed in the Methods section, potential contrasting parametrizes the force field by maximizing its effectiveness in differentiating data samples from noise samples. We use as data samples a total of $N_p = 1\,044\,000$ conformations from a 208 μs long molecular dynamics simulation with explicit solvents performed in ref 62. This fully atomistic simulation captures multiple folding and unfolding events for the peptide. We generated $N_q = 1\,044\,000$ noise samples (Figure S2) that include both folded and disordered configurations and computed the noise potential $u_q(\mathbf{x})$ using the umbrella sampling procedure described in the Methods section. In the following, we use potential contrasting to learn three CGFFs with different flexibility and complexity by optimizing eq 4 with the limited-memory BFGS (L-BFGS) method.[54] For simplicity, we only use C_α atoms to represent protein conformations and define energies, but potential contrasting can be easily generalized to more refined structural models.

CGFF with Bonded Terms and Pairwise Nonbonded Interactions. We first learned a CGFF, $u_p^{pair}(\mathbf{x}; \boldsymbol{\theta})$, that includes bonded terms and pairwise nonbonded terms defined as

$$u_p^{pair}(\mathbf{x}; \boldsymbol{\theta}) = u_{bond}(\mathbf{x}) + u_{angle}(\mathbf{x}) + u_{dihedral}(\mathbf{x}) + u_{elec}(\mathbf{x})$$
$$+ u_{nb}(\boldsymbol{x}) = \sum_{i=1}^{L-1} \frac{1}{2} k_i (b_i - b_i^\circ)^2 + \sum_{i=1}^{L-2} S_{angle}(a_i; \mathbf{c}_i^a)$$
$$+ \sum_{i=1}^{L-3} S_{dihedral}(d_i; \mathbf{c}_i^d) + \sum_{i=1}^{L-4} \sum_{j=i+4}^{L} \frac{q_i q_j}{4\pi\epsilon r_{ij}} \exp(-r_{ij}/\lambda_D)$$
$$+ \sum_{i=1}^{L-4} \sum_{j=i+4}^{L} S_{nb}(r_{ij}; \mathbf{c}_{ij})$$

$$(7)$$

The bond, angle, and dihedral terms are similarly defined as in eq 5. Nonbonded terms include electrostatics, $u_{elec}(\mathbf{x})$, and a short-range interaction energy term, $u_{nb}(\mathbf{x})$, both of which act between pairs of CG particles that are separated by four or more bonds. The electrostatic interaction is modeled using the Debye−Hückel theory, where q_i is the net charge of the ith residue, λ_D is the Debye screening length, and r_{ij} is the distance between residues i and j. The short-range nonbonded interaction energy is defined with cubic spline functions, $S_{nb}(r_{ij}; \mathbf{c}_{ij})$ and $\mathbf{c}_{ij}$ are spline basis coefficients (Figure S1). Because bond energies are much stronger than others, the parameters b_i° and k_i were directly fitted based on the mean and the variance of the ith bond's distribution in the data samples. Therefore, the parameter $\boldsymbol{\theta}$ only includes spline basis

coefficients, that is, $\theta = \{\mathbf{c}_i^a, \mathbf{c}_i^d, \mathbf{c}_{ij}\}$. To prevent overfitting, regularization terms on the potential energy $u_p(\mathbf{x}; \theta)$ are added in the optimization to control their smoothness. Details on regularization terms are included in the Supporting Information. Since the energy function depends on the parameters θ linearly, potential contrasting is guaranteed to produce a unique solution θ^*.

We carried out molecular dynamics simulations (see the Supporting Information for details) with the learned CGFF $u_p^{pair}(\mathbf{x}; \theta^*)$ to evaluate the resulting structural ensemble. Similar to that from the all-atom simulation, the distribution of RMSD with respect to the folded structure for CGFF is bimodal (Figure 3a). Therefore, the learned CG potential function $u_p^{pair}(\mathbf{x}; \theta^*)$ captures both folded and unfolded structures. However, a significant discrepancy exists between the two distributions. The CG simulation produced fewer folded structures, and the two maxima of the corresponding RMSD distribution do not exactly match that of the all-atom result. The discrepancy is clearer if we convert the RMSD distribution histogram into free energy surfaces (Figure 3d). Deviations can also be seen when comparing the free energy surface over the first two components of the time-independent component analysis[63−65] (tICA), which describe the slowest processes observed in the simulation. The all-atom surface has three metastable states: one folded state (state 1), one unfolded state (state 2), and one misfolded state (state 3) (Figure 3h,j,k,l). The first tICA coordinate separates the folded state from unfolded states and thus can be viewed as the protein's folding coordinate. On the other hand, tICA 2 captures the conformational transition between unfolded structures and a misfolded, metastable configuration. Although the CG simulation samples all three metastable states (Figure 3e), it produces a smaller population of the folded state and does not capture the cooperative transitions between folded and unfolded structures (Figure S3).

As mentioned in the Methods section, well-defined noise distributions are essential for the performance of potential contrasting. It is conceivable that the noise distribution $u_q(\mathbf{x})$ based on the umbrella sampling of RMSD does not provide sufficient coverage of the phase space given the degeneracy of the collective variable. As an alternative, we introduced a new noise distribution based on two-dimensional umbrella simulations using RMSD and tICA2 as collective variables. Because tICA2 is relatively orthogonal to tICA1 and RMSD, explicit biases on tICA2 in the two-dimensional umbrella sampling shall expand the conformational space explored by the noise distribution (Figure S4). However, optimizing $u_p^{pair}(\mathbf{x}; \theta^*)$ with the new noise distribution does not improve the force field's performance substantially (Figure S4), suggesting that the subpar results seen in Figure 3a are not due to the quality of the noise distribution.

Adding Many-Body Interactions Parametrized Using Neural Networks. The discrepancy between the CG and all-atom simulations could also be caused by the pairwise potential being too restrictive and cannot capture many-body interactions that might arise due to coarse-graining. Next, we learned a more flexible energy function that includes an extra term parametrized using a feed-forward neural network with parameters ϕ, that is,

$$u_p^{nn}(\mathbf{x}; \theta, \phi) = u_p^{pair}(\mathbf{x}; \theta) + u_{mb}^{nn}(\mathbf{x}; \phi) \tag{8}$$

The additional energy term, $u_{mb}^{nn}(\mathbf{x}; \phi)$, takes angles, dihedral angles, and pairwise distances as inputs and thus is invariant to translations and rotations (Figure S5). It can represent complex interactions involving multiple residues because the neural network is fully connected to couple different degrees of freedom.[66]

A CG simulation performed with the learned potential function $u_p^{nn}(\mathbf{x}; \theta^*, \phi^*)$ now indeed matches the all-atom results well. The maxima of the RMSD distribution are much better placed (Figure 3c), suggesting that the CG simulation accurately predicts the folded structure. Importantly, the CG simulation reproduces the relative population of the folded structure and the unfolded ensemble and the free energy barrier between them (Figure 3d and S3). Similarly, the free energy surface of the first two tICA coordinates (Figure 3g,h) and the distributions of pairwise distances (Figure S6) agree well with the all-atom ones. Therefore, despite only using only α-carbons, the CGFF captures the complex folding landscape of the peptide determined from atomistic explicit solvent simulations.

Adding Secondary Structure Inspired Many-Body Potentials. Although parametrizing the many-body energy term using a neural network improves the accuracy of the resulting force field, it has a few disadvantages. For instance, the potential function $u_p^{nn}(\mathbf{x}; \theta, \phi)$ is not linear to ϕ, and optimizations may produce multiple solutions with comparable performance when varying the initial conditions (see Figure S7). Moreover, it can be difficult to interpret the many-body energy in simple physical terms. To avoid these issues, we learned a CG potential function with a secondary structure based many-body energy term. Secondary structure biases are frequently incorporated into coarse-grained models as fragment memories for improved quality of structural predictions.[15,22,67] They help account for cooperative effects arising from water molecules involving many residues that are challenging to describe with pairwise potentials. Specifically, the secondary structure based many-body energy term is defined as

$$u_p^{ss}(\mathbf{x}; \theta, \phi) = u_p^{pair}(\mathbf{x}; \theta) + u_{mb}^{ss}(\mathbf{x}; \phi) \tag{9}$$

It is parametrized using cubic spline functions as $u_{mb}^{ss}(\mathbf{x}; \phi) = S_{ss}(\text{rmsd\_ss}(\mathbf{x}, \mathbf{x}_o); \mathbf{c}^{ss})$. Here rmsd_ss($\mathbf{x}, \mathbf{x}_o$) is the RMSD calculated on the α-helix (residue 3 to residue 15) between a given structure $\mathbf{x}$ and the folded structure $\mathbf{x}_o$. The function $S_{ss}(\text{rmsd}, \mathbf{c}^{ss})$ is parametrized using a cubic spline and the parameter ϕ includes all the spline basis coefficients $\mathbf{c}^{ss}$. This design of the energy function in eq 9 further ensures linear dependence on parameters and a unique solution for force field optimization.

The CG simulation results using the learned potential $u_p^{ss}(\mathbf{x}; \theta^*, \phi^*)$ are shown in Figures 3b,d,f, and S3. Although the many-body energy term is restricted within the α-helix, the CG simulation correctly reproduces the relative populations of folded and unfolded states and the free energy barrier. Its performance is almost as good as the potential with a neural network based many-body term defined over the whole protein. The learned many-body potential function $u_{mb}^{ss}(\mathbf{x}; \phi^*)$ along the α-helix RMSD is shown in Figure 3i. It has a deep well near 0 nm and quickly approaches zero when the RMSD is larger than 0.3 nm. Therefore, the potential only plays a significant role in stabilizing the folded structure when
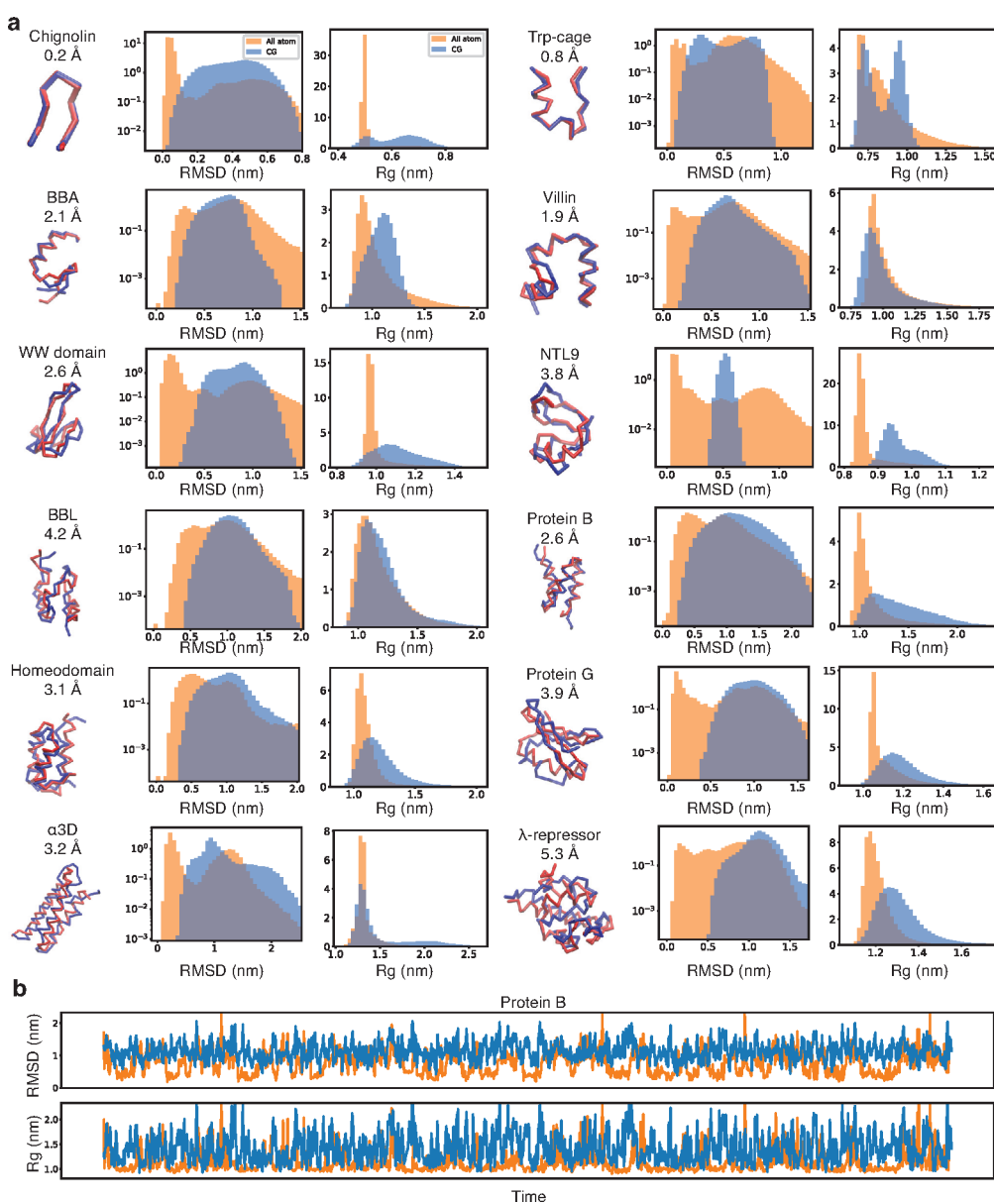
Figure 4. Comparison between results from all-atom simulations[62] and CG simulations performed using the force field with shared parameters. (a) For each of the 12 proteins, we show the folded structure (red) from the all-atom simulation, the structure (blue) from the CG simulation that has the lowest RMSD with respect to the folded structure, and the C_α-RMSD (over all residues) between the two structures. The two plots on the right of structures are distributions of RMSD to the folded structure and distributions of R_g (radius of gyration) for conformations sampled from all-atom simulations (orange) and CG simulations (blue). (b) Trajectories of R_g and RMSD with respect to the folded structure for the all-atom simulation (orange) and the CG simulation (blue) of Protein B. Although the data from all-atom simulations and CG simulations are plotted in the same figure, their time scales are different. Similar plots for other proteins are included in the Supporting Information.

the α-helix is already close to the native state. Its impact is minimal when the α-helix adopts unfolded configurations.

Efficient Optimization of CG Force Fields with Data from Multiple Proteins. The above results suggest that potential contrasting is a powerful tool to parametrize flexible CGFFs for specific proteins and capture their complex folding landscapes. However, to develop general-purpose force fields, optimizing the parameters against a diverse set of proteins is essential to improve their transferability. Next, we show that potential contrasting achieves high computational efficiency to enable parameter optimization using all-atom simulation data of 12 fast-folding proteins.[62]

We designed the energy function to allow sharing of parameters across proteins. The force field for the kth protein is defined using eq 9 as

$$u_p^k(\mathbf{x}_k) = u_{bond}(\mathbf{x}_k) + u_{angle}(\mathbf{x}_k) + u_{dihedral}(\mathbf{x}_k) + u_{nb}(\mathbf{x}_k)$$
$$+ u_{elec}(\mathbf{x}_k) + u_{mb}^{ss}(\mathbf{x}_k; \phi_k) \qquad (10)$$

As proof of principle, we only shared parameters for pairwise nonbonded interactions and allowed protein-specific parameters for bonded and many-body terms. The pairwise non-bonded interaction potential is now defined as

$$u_{nb}(\boldsymbol{x}) = \sum_{i=1}^{L-4} \sum_{j=i+4}^{L} S_{nb}(r_{ij}; \mathbf{c}_{IJ}^{nb}) \tag{11}$$

While $S_{nb}(r; \mathbf{c}_{IJ}^{nb})$ has the same functional form as that in eq 7, its parameters now only depend on residue types I and J. Because $\mathbf{c}_{IJ}^{nb}$ are made to depend on residue types alone, they are shared among proteins. Our choice of limiting the force field's shared parameters is due to the well-known challenges of predicting secondary structures in CG models.[2] While potential contrasting allows efficient optimization of CGFFs with all parameters being shared across proteins, the accuracy of the resulting CGFF may be poor. Allowing protein-specific potentials alleviates the challenges in describing secondary structures using CG models with only one particle per residue.

Both shared and protein-specific parameters were learned by optimizing the aggregated objective function defined in eq 6. For each of the 12 proteins, we used evenly spaced 250 000 conformations from the corresponding all-atom simulation as data samples. Using the umbrella sampling procedure described in the Methods section, we generated the same number of noise samples and computed the noise potentials $u_q^k(\mathbf{x})$. Because the energy function $u_p^k(\mathbf{x})$ is linear to all parameters, optimizing the aggregated objective function (eq 6) converges to a unique solution. In addition, because the aggregated objective function is a weighted sum of objective functions for individual proteins, its computing and optimization can be easily parallelized among proteins. Using 12 Nvidia Volta V100 GPUs with each assigned to calculate the potential contrasting objective function of one protein, we can optimize the aggregated objective function (Figure S8) and learn all parameters in 30 min.

CG simulations using the learned potential functions are compared to all-atom simulations (Figures 4, S9, S10, and S11) in terms of the radius of gyration (R_g) and the RMSD from the folded structures. Structures close to the native state are sampled in the CG simulations for all proteins (Figure 4). The lowest RMSD for configurations sampled in CG simulations range from 0.2 to 5.3 Å and are less than 4 Å for 10 out of 12 proteins. The short-range nonbonded interaction potentials between pairs of amino acids are shown in Figures 5 and S12. We parametrize these potentials
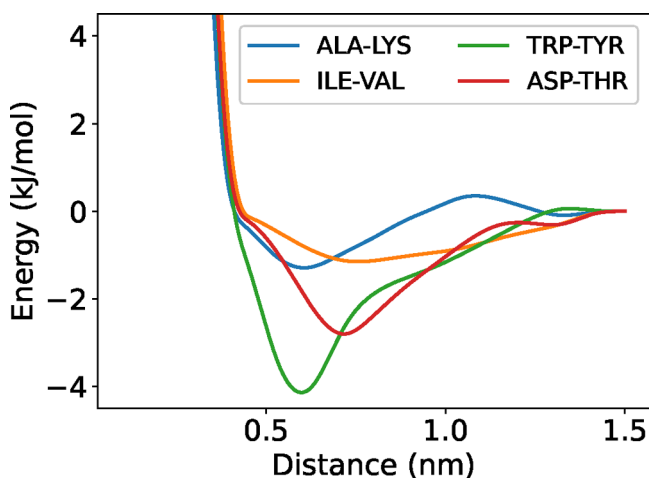

Figure 5. Learned contact potential energy functions between representative pairs of amino acids. Similar plots for other pairs of amino acids are included in the Supporting Information.

using cubic splines and do not restrict them to specific mathematical expressions. Nevertheless, many converge to functions with multiple minima, potentially arising from the formation of direct and water-mediated contacts between amino acids in all-atom simulations.[22,68−70]

Further analyses revealed that the CGFF captures the collapse transition of proteins with both expanded (large R_g) and compact (small R_g) configurations (Figure S11). The free energy surfaces along the tICA coordinates support the metastability of the different configurations (Figure S13). However, we found that, for most proteins, the native states do not appear as the global minimum in the sampled free energy profiles (Figure S10). Even for Trp-cage, the agreement between CG and all-atom simulations is not as good as those shown in Figure 3b. The CGFF's accuracy is impacted by the sharing of parameters across proteins and our use of a pairwise potential for the nonbonded interactions. Further improving the force field accuracy would require introducing transferable many-body potentials for nonbonded interactions.

CONCLUSION AND DISCUSSION

By generalizing noise contrastive estimation with unnormalized noise distributions, we developed a new method, potential contrasting, for learning force fields from reference molecular configurations. Potential contrasting combines the advantages of existing variational methods such as force matching and relative entropy minimization. As with the force matching method, it is computationally efficient and does not need sampling during force field optimization. Like the relative entropy method, potential contrasting does not require force information. We note that the gYBG method shares similar properties as potential contrasting and does not require iterative sampling of trial CG potentials or atomic force information. However, as Rudzinski and Noid pointed out,[71] the method can be sensitive to the accumulation of errors arising from the estimation of correlation functions and their gradients. Whether these challenges prevent the gYBG method from deriving transferable force fields for complex biomolecules is unclear. We showed that potential contrasting is effective and succeeds in producing CG energy functions that accurately reproduce configurational distributions obtained from all-atom simulations. In addition, potential contrasting can be trivially parallelized for efficient learning of CGFFs with shared parameters using simulation data of multiple systems.

With its efficacy and efficiency, potential contrasting is well-positioned to systematically learn transferable CGFFs based on all-atom force fields. To ensure the accuracy of CGFFs, energy functions that can account for many-body effects must be used. Indeed, the less-than-perfect results shown in Figure 4 arise precisely from the pairwise approximations used in the force field rather than a failure of potential contrasting. As we have demonstrated for Trp-cage, provided with a flexible enough energy function, potential contrasting is fully capable of producing CG models that reproduce all-atom configurational distributions with remarkable accuracy. Furthermore, unlike the neural network based potential with pairwise distances as inputs, the energy function needs to be agnostic with respect to protein-specific features for transferability. The recently proposed CGSchNet based on graph neural networks could potentially offer the capacity for capturing many-body effects while maintaining transferability.[72]

Although we focused in this study on using potential contrasting to learn CGFFs, the method is general. It can be

applied to learning various types of force fields. For instance, potential contrasting can be readily applied to parametrize implicit solvent models using all-atom simulations with explicit water molecules. With further development, it could also be used to improve existing all-atom force fields by incorporating information from quantum mechanical calculations or experimental data. Such applications and development will be investigated in future studies.

Using unnormalized noise distributions produced with umbrella sampling is essential for parametrizing accurate CGFFs. Unnormalized noise distributions defined with molecular energy functions allow the generation of noise samples that resemble the configurations produced from all-atom simulations. Therefore, a significant overlap in the phase space between noise and data samples can be achieved. Such overlap can be difficult to ensure with arbitrary noise distributions since all-atom simulations only sample limited regions of phase space with low energy. We note that, upon training molecular simulation data, probabilistic models parametrized with normalizing flows[73−75] have been shown to produce realistic and stable molecular conforma-tions.[45,76−80] These models have indeed been proposed to serve as noise distributions for contrastive learning to guarantee overlap with data samples.[49,76] However, we found that using flow-based models as noise distributions produced CGFFs with subpar quality. Similar findings have been reached in other recent studies as well.[52] By optimizing the overlap with data samples, flow-based models may hinder the minimization of probability for regions outside the data. Further research is needed to design optimal noise distributions in NCE.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00616.

> Detailed procedure for generating noise samples for learning CGFF of protein folding, learning CG potential functions with both shared and protein-specific param-eters, cubic splines for flexible potential energy para-metrization, parameter optimization and regularization, molecular dynamics simulations with the CGFFs, many-body interactions parametrized using neural network, B-spline basis for the cubic splines used to parametrize potential energies, distributions of RMSD with respect to the folded structure of the Trp-cage protein for the noise samples generated using one-dimensional umbrella sampling, trajectories of RMSD with respect to the folded structure of the Trp-cage protein for conforma-tions from the all-atom simulation and the CG simulations, performance of force fields optimized using a noise distribution constructed with two-dimen-sional umbrella simulations, neural network used for parametrizing the many-body energy term, distributions of distances between pairs of residues of the Trp-cage protein for both the all-atom simulation and the CG simulation, optimizations started from different initial values of neural network based potentials produce force fields with comparable quality, convergence of the aggregated loss function with weight decay during the optimization, trajectories of radius of gyration (R_g) and RMSD with respect to the folded structure from the all-

atom simulation and the CG simulation of all 12 proteins, free energy profiles along the RMSD with respect to the folded structure for conformations sampled from the all-atom simulation and CG simulations of the 12 proteins, free energy profiles along R_g with respect to the folded structure for conformations sampled from the all-atom simulation and CG simulations of the 12 proteins, learned transferable short-range nonbonded potential energy functions between pairs of amino acids, free energy surfaces over the first two tICA coordinates for the CG simulations of the 12 proteins using the learned force field with shared parameters, and setup used in umbrella sampling for generating noise samples (PDF)

AUTHOR INFORMATION

Corresponding Author

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

Author

Xinqiang Ding − *Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c00616

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Saunders, M. G.; Voth, G. A. Coarse-Graining Methods for Computational Biology. *Annual Review of Biophysics* **2013**, *42*, 73−93.

(2) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116*, 7898−7936.

(3) 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−R512.

(4) de Pablo, J. J. Coarse-Grained Simulations of Macromolecules: From DNA to Nanocomposites. *Annu. Rev. Phys. Chem.* **2011**, *62*, 555−574.

(5) Souza, P. C. T.; et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18*, 382−388.

(6) 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 Computational Biology* **2018**, *14*, No. e1005941.

(7) Best, R. B. Computational and theoretical advances in studies of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2017**, *42*, 147−154.

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

(9) Hyeon, C.; Thirumalai, D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* **2011**, *2*, 487.

(10) Denesyuk, N. A.; Thirumalai, D. Coarse-Grained Model for Predicting RNA Folding Thermodynamics. *J. Phys. Chem. B* **2013**, *117*, 4901−4911.

(11) Farr, S. E.; Woods, E. J.; Joseph, J. A.; Garaizar, A.; Collepardo-Guevara, R. Nucleosome plasticity is a critical element of chromatin liquid−liquid phase separation and multivalent nucleosome interactions. *Nat. Commun.* **2021**, *12*, 2883.

(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) Latham, A. P.; Zhang, B. Consistent Force Field Captures Homologue-Resolved HP1 Phase Separation. *J. Chem. Theory Comput.* **2021**, *17*, 3134−3144.

(16) 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.

(17) 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.

(18) Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, 090901.

(19) Gkeka, P.; et al. Machine Learning Force Fields and Coarse-Grained Variables in Molecular Dynamics: Application to Materials and Biological Systems. *J. Chem. Theory Comput.* **2020**, *16*, 4757−4775.

(20) Latham, A. P.; Zhang, B. Unifying coarse-grained force fields for folded and disordered proteins. *Curr. Opin. Struct. Biol.* **2022**, *72*, 63−70.

(21) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. A Coarse Grain Model for Phospholipid Simulations. *J. Phys. Chem. B* **2001**, *105*, 4464−4470.

(22) 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.

(23) 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.

(24) 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.

(25) Ercolessi, F.; Adams, J. B. Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *Europhysics Letters (EPL)* **1994**, *26*, 583−588.

(26) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching. *J. Chem. Phys.* **2004**, *120*, 10896−10913.

(27) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. DeePCG: Constructing coarse-grained models via deep neural networks. *J. Chem. Phys.* **2018**, *149*, 034101.

(28) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120*, 143001.

(29) Tschöp, W.; Kremer, K.; Batoulis, J.; Bürger, T.; Hahn, O. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polym.* **1998**, *49*, 61−74.

(30) Louis, A. A.; Bolhuis, P. G.; Hansen, J. P.; Meijer, E. J. Can Polymer Coils Be Modeled as "Soft Colloids. *Phys. Rev. Lett.* **2000**, *85*, 2522−2525.

(31) Savelyev, A.; Papoian, G. A. Molecular Renormalization Group Coarse-Graining of Polymer Chains: Application to Double-Stranded DNA. *Biophys. J.* **2009**, *96*, 4044−4052.

(32) Akkermans, R. L. C.; Briels, W. J. A structure-based coarse-grained model for polymer melts. *J. Chem. Phys.* **2001**, *114*, 1020−1031.

(33) Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. Versatile Object-Oriented Toolkit for Coarse-Graining Applications. *J. Chem. Theory Comput.* **2009**, *5*, 3211−3223.

(34) Noid, W. G. In *Systematic Methods for Structurally Consistent Coarse-Grained Models BT - Biomolecular Simulations: Methods and Protocols*; Monticelli, L., Salonen, E., Eds.; Humana Press: Totowa, NJ, 2013; pp 487−531.

(35) Tóth, G. Effective potentials from complex simulations: a potential-matching algorithm and remarks on coarse-grained potentials. *J. Phys.: Condens. Matter* **2007**, *19*, 335222.

(36) Li, W.; Takada, S. Characterizing Protein Energy Landscape by Self-Learning Multiscale Simulations: Application to a Designed β-Hairpin. *Biophys. J.* **2010**, *99*, 3029−3037.

(37) Hansen, J.-P.; Addison, C. I.; Louis, A. A. Polymer solutions: from hard monomers to soft polymers. *J. Phys.: Condens. Matter* **2005**, *17*, S3185−S3193.

(38) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. *J. Chem. Phys.* **2001**, *115*, 2323−2347.

(39) Clark, A. J.; McCarty, J.; Lyubimov, I. Y.; Guenza, M. G. Thermodynamic Consistency in Variable-Level Coarse Graining of Polymeric Liquids. *Phys. Rev. Lett.* **2012**, *109*, 168301.

(40) Mirzoev, A.; Lyubartsev, A. P. MagiC: Software Package for Multiscale Modeling. *J. Chem. Theory Comput.* **2013**, *9*, 1512−1520.

(41) Izvekov, S.; Voth, G. A. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **2005**, *109*, 2469−2473.

(42) Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **2008**, *129*, 144108.

(43) 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.

(44) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. *J. Chem. Phys.* **2008**, *128*, 244115.

(45) Köhler, J.; Chen, Y.; Krämer, A.; Clementi, C.; Noé, F. Force-matching Coarse-Graining without Forces. *arXiv preprint* 2022, arXiv:2203.11167; https://arxiv.org/abs/2203.11167.

(46) Mullinax, J. W.; Noid, W. G. Generalized Yvon-Born-Green Theory for Molecular Systems. *Phys. Rev. Lett.* **2009**, *103*, 198104.

(47) Garrido, L.; Juste, A. On the determination of probability density functions by using Neural Networks. *Comput. Phys. Commun.* **1998**, *115*, 25−31.

(48) Lemke, T.; Peter, C. Neural Network Based Prediction of Conformational Free Energies - A New Route toward Coarse-Grained Simulation Models. *J. Chem. Theory Comput.* **2017**, *13*, 6213−6221.

(49) 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*. Chia Laguna Resort, Sardinia, Italy, MLResearchPress: 2010; pp 297−304.

(50) Müller, K.; Brown, L. D. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theoretica chimica acta* **1979**, *53*, 75−93.

(51) Myung, I. J. Tutorial on maximum likelihood estimation. *J. Math. Psychol.* **2003**, *47*, 90.

(52) Chehab, O.; Gramfort, A.; Hyvarinen, A. The Optimal Noise in Noise-Contrastive Learning Is Not What You Think. *arXiv preprint* 2022, arXiv:2203.01110; https://arxiv.org/abs/2203.01110.

(53) Hastie, T.; Tibshirani, R.; Friedman, J. H.; Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction*; Springer, 2009; Vol. 2.

(54) Zhu, C.; Byrd, R. H.; Lu, P.; Nocedal, J. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software* 1997, 23, 550−560.

(55) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* 1977, 23, 187−199.

(56) Tiwary, P.; van de Walle, A. In *A Review of Enhanced Sampling Approaches for Accelerated Molecular Dynamics BT - Multiscale Materials Modeling for Nanomechanics*; Weinberger, C. R., Tucker, G. J., Eds.; Springer International Publishing: Cham, 2016; pp 195−221.

(57) Ferrenberg, A. M.; Swendsen, R. H. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* 1989, 63, 1195−1198.

(58) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 1992, 13, 1011−1021.

(59) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* 2008, 129, 124105.

(60) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* 1976, 22, 245−268.

(61) Ding, X.; Vilseck, J. Z.; Brooks, C. L. Fast Solver for Large Scale Multistate Bennett Acceptance Ratio Equations. *J. Chem. Theory Comput.* 2019, 15, 799−802.

(62) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* 2011, 334, 517−520.

(63) Molgedey, L.; Schuster, H. G. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.* 1994, 72, 3634−3637.

(64) Naritomi, Y.; Fuchigami, S. Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: The case of domain motions. *J. Chem. Phys.* 2011, 134, 065101.

(65) 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.

(66) Wang, Y.; Lamim Ribeiro, J. M.; Tiwary, P. Machine learning approaches for analyzing and enhancing molecular dynamics simulations. *Curr. Opin. Struct. Biol.* 2020, 61, 139−145.

(67) Rohl, C. A.; Strauss, C. E. M.; Misura, K. M. S.; Baker, D. *Numerical Computer Methods, Part D*; Academic Press, 2004; Vol. 383; pp 66−93.

(68) Cheung, M. S.; García, A. E.; Onuchic, J. N. Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. U. S. A.* 2002, 99, 685−690.

(69) Karanicolas, J.; Brooks, C. L. Improved Gō-like Models Demonstrate the Robustness of Protein Folding Mechanisms Towards Non-native Interactions. *J. Mol. Biol.* 2003, 334, 309−325.

(70) Jernigan, R. L.; Bahar, I. Structure-derived potentials and protein simulations. *Curr. Opin. Struct. Biol.* 1996, 6, 195−209.

(71) Rudzinski, J. F.; Noid, W. G. A generalized-Yvon-Born-Green method for coarse-grained modeling. *European Physical Journal Special Topics* 2015, 224, 2193−2216.

(72) Husic, B. E.; Charron, N. E.; Lemm, D.; Wang, J.; Pérez, A.; Majewski, M.; Krämer, A.; Chen, Y.; Olsson, S.; de Fabritiis, G.; Noé, F.; Clementi, C. Coarse graining molecular dynamics with graph neural networks. *J. Chem. Phys.* 2020, 153, 194101.

(73) Rezende, D. J.; Mohamed, S. Variational Inference with Normalizing Flows. *Proceedings of the 32nd International Conference on Machine Learning*; MLResearchPress: 2015; pp 1530−1538.

(74) Dinh, L.; Sohl-Dickstein, J.; Bengio, S. Density estimation using Real NVP. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*; 2016.

(75) Papamakarios, G.; Nalisnick, E.; Rezende, D. J.; Mohamed, S.; Lakshminarayanan, B. Normalizing Flows for Probabilistic Modeling and Inference. *arXiv*, 2019, https://arxiv.org/abs/1912.02762.

(76) Gao, R.; Nijkamp, E.; Kingma, D. P.; Xu, Z.; Dai, A. M.; Wu, Y. N. Flow contrastive estimation of energy-based models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 2020, 7518−7528.

(77) Noé, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* 2019, 365, No. eaaw1147.

(78) Ding, X.; Zhang, B. Computing Absolute Free Energy with Deep Generative Models. *J. Phys. Chem. B* 2020, 124, 10166−10172.

(79) Ding, X.; Zhang, B. DeepBAR: A Fast and Exact Method for Binding Free Energy Computation. *J. Phys. Chem. Lett.* 2021, 12, 2509−2515.

(80) Wirnsberger, P.; Ballard, A. J.; Papamakarios, G.; Abercrombie, S.; Racanière, S.; Pritzel, A.; Rezende, D. J.; Blundell, C. Targeted free energy estimation via learned mappings. *J. Chem. Phys.* 2020, 153, 144112.