JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

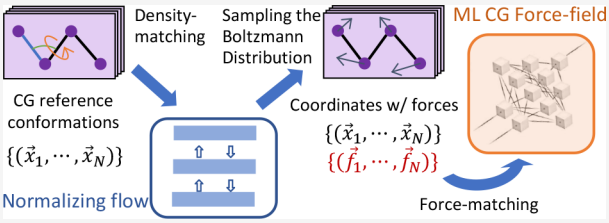
Flow-Matching: Efficient Coarse-Graining of Molecular Dynamics without Forces

Jonas Köhler,[▽] Yaoyi Chen,[▽] Andreas Krämer,[*,▽] Cecilia Clementi,[*] and Frank Noé[*]

Cite This: *J. Chem. Theory Comput.* 2023, 19, 942−952 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Coarse-grained (CG) molecular simulations have become a standard tool to study molecular processes on time and length scales inaccessible to all-atom simulations. Parametrizing CG force fields to match all-atom simulations has mainly relied on force-matching or relative entropy minimization, which require many samples from costly simulations with all-atom or CG resolutions, respectively. Here we present *flow-matching*, a new training method for CG force fields that combines the advantages of both methods by leveraging normalizing flows, a generative deep learning method. Flow-matching first trains a normalizing flow to represent the CG probability density, which is equivalent to minimizing the relative entropy without requiring iterative CG simulations. Subsequently, the flow generates samples and forces according to the learned distribution in order to train the desired CG free energy model via force-matching. Even without requiring forces from the all-atom simulations, flow-matching outperforms classical force-matching by an order of magnitude in terms of data efficiency and produces CG models that can capture the folding and unfolding transitions of small proteins.



1. INTRODUCTION

Molecular dynamics (MD) simulations have become a major computational tool to study biophysical processes on molecular scales. Presently, MD simulations at all-atom resolution can reach multiple microseconds for small- to medium-sized protein systems on retail hardware. By using special-purpose supercomputers[1,2] or combining distributed computing with Markov State Modeling[3,4] or enhanced sampling approaches, it is possible to probe millisecond time scales and sometimes beyond.[5,6]

Despite this progress, many biomolecular processes of interest exceed these time and length scales by orders of magnitude. Also, high-throughput simulations that would be needed, e.g., to screen protein sequences for high-affinity protein−protein interactions, cannot be efficiently done with all-atom MD.

A common approach to go to larger time and length scales or high-throughput simulations is coarse-grained (CG) molecular dynamics.[7−22] In "bottom-up" coarse-graining,[23] one defines a mapping from the all-atom representation to the CG model (e.g., by grouping sets of atoms to CG beads). The choice of mapping determines the resolution and has to suit the system as well as the scientific question, which is by itself a challenge.[13,16,24,25] Given that the CG mapping is chosen, a frequently used CG principle is known as thermodynamic consistency in the coarse-graining literature and as density matching in machine learning: the CG model should generate the same equilibrium distribution in the CG coordinates, as one would obtain from a fully converged all-atom simulation after applying the coarse-graining map to all simulation frames.[13] In principle, the requirement of thermodynamic consistency uniquely defines the free energy function in the CG coordinates, which is also known as the *potential of mean force* (PMF). Ideally, if this thermodynamically consistent PMF were known, it could be used to compute exactly any equilibrium property expressible as an ensemble average over the CG coordinates. Note that this definition does not guarantee that the CG model reproduces *all* thermodynamic observables, counterexamples being heat capacity, pressure, and entropy.[26−30] However, the PMF by definition involves high-dimensional integrals that cannot be estimated for nontrivial systems in practice. A pivotal challenge is to find a good approximation for the PMF with tractable functional forms to serve as the CG potential.[13]

Among the techniques for such bottom-up modeling,[13,15,31−34] two methods have been explicitly developed to approach thermodynamic consistency: variational force-matching (also known as multiscale coarse-graining)[32,33] and relative entropy minimization.[34] Force-matching (Figure 1a) is straightforward to implement but requires the forces on the CG particles mapped from all-atom sampling. Because these instantaneous forces depend on all degrees of freedom, they

Received: January 4, 2023
Published: January 20, 2023

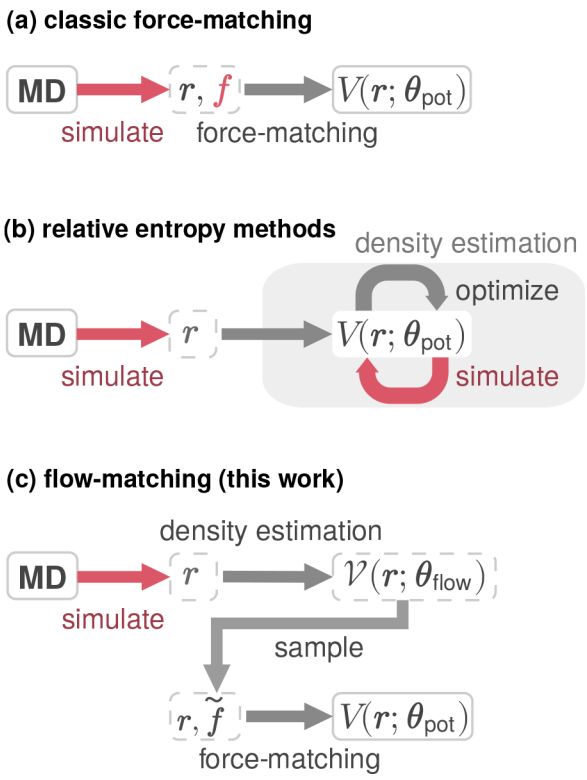


Figure 1. Overview of the flow-matching method. (a) Classical force-matching: parameters $\boldsymbol{\theta}_{\text{pot}}$ of a CG potential $V(\cdot; \boldsymbol{\theta}_{\text{pot}})$ optimized to minimize the mean-squared error of model forces with respect to projected atomistic forces $\boldsymbol{f}$ on the training configurations $\boldsymbol{r}$. (b) Relative entropy methods: simulations performed with CG potential to produce samples and enable evaluating (and minimizing) the relative entropy. (c) Present method: parameters $\boldsymbol{\theta}_{\text{flow}}$ of a normalizing flow first optimized to match the CG density from the ground-truth samples $\boldsymbol{r}$. This defines the flow-based potential $\mathcal{V}(\cdot; \boldsymbol{\theta}_{\text{flow}})$. The samples and forces from the flow are used to train a CG potential $V(\cdot; \boldsymbol{\theta}_{\text{pot}})$ via force-matching. Slow/inaccurate sampling steps are highlighted in red.

provide a very noisy signal that makes training the CG force field data inefficient. This approach has been connected with the blooming field of machine-learned potentials and led to several successes.[20−22] Relative entropy minimization (Figure 1b), the Inverse Monte Carlo method,[35] and Iterative Boltzmann Inversion[31] do not require forces to be recorded and are more data-efficient but require the CG model to be resimulated during the iterative training procedure, which can be extremely costly and even lead to failure in convergence. Reference 36 developed a hybrid approach combining force-matching and relative entropy methods in order to parametrize CG models where not all particles have force information available.

This manuscript presents a third alternative—the *flow-matching* method—which is shown to be more efficient. Our approach combines aspects of force-matching and relative entropy minimization with deep generative modeling. The centerpiece of this novel method is a *normalizing flow*,[37−39] a generative neural network that can approximate arbitrary probability distributions by transforming them into simple, easy-to-sample prior distributions. Once trained, normalizing flows can generate uncorrelated samples and compute normalized probability densities, energies, and forces, which makes them an exciting emerging tool for physical applications.[40−45,45,46] For example, Boltzmann generators[40] use flows that are trained on MD data and energies as one-shot importance samplers for molecular equilibrium distributions. Other types of generative neural networks have also been used for back-mapping of CG structures.[47,48]

Flow-matching applies normalizing flows to the coarse-graining problem. Like force-matching and relative entropy minimization, it starts from CG samples in equilibrium, which are usually created by mapping snapshots from an all-atom simulation to the CG space. In order to find a thermodynamically consistent CG potential, the method proceeds in two steps (Figure 1c). First, a normalizing flow is trained by density matching, such that it learns to sample directly from the target ensemble. Second, the CG equilibrium distribution that the flow has learned is taught to a CG force field by force-matching to coordinate−force pairs generated by the flow.

While this stepwise approach leans on the same underlying principles as previous coarse-graining methods, it avoids their key limitations. In contrast to force-matching (Figure 1a,c), it does not rely on atomistic reference forces. Although saving forces during the simulation is in principle straightforward to do, in most cases of already existing long simulations, forces have not been stored and are often nontrivial to recompute a posteriori. To bypass the need for force data, an alternative method has been previously proposed as the generalized Yvon−Born−Green theory,[49] which determines a CG force field (usually as a sum of basis functions) directly according to structural correlations. However, it is not clear whether this can be generalized to CG force fields based on neural networks.

Additionally, the flow can generate an indefinite number of "synthetic" configurations and forces, which do not carry noise from the atomistic environment. In contrast to relative entropy minimization[34] and Iterative Boltzmann Inversion,[31] flow-matching does not require repeated resimulation of the CG model during training, as the flow can generate independent samples that represent the thermodynamic equilibrium (Figure 1b,c). In practice, by removing the need for costly simulations during training, flow-matching makes coarse-graining by density estimation/relative entropy methods feasible for molecules with rare events, such as biomolecules. In contrast to force-matching, density estimation does not suffer from the noise problem due to the omitted degrees of freedom, and consequently, flow-matching is significantly more data-efficient.

Using the flow only as an intermediate offers complete freedom in choosing the functional form of the final CG force field. In particular, the candidate potential can incorporate the desired physical symmetries and asymptotics[20,21] as well as share parameters across chemical space.[22] Conversely, directly using a normalizing flow as the CG force field would not be a good idea, because transferable properties cannot be easily incorporated into invertible[37,39] or at least statistically reversible[50] neural networks, which are required by the flows. For example, transferability across molecular systems of different sizes and topologies requires parameter sharing and a transformation of random variables of different dimensionality—features not yet supported by existing normalizing flows. To this end, flow-matching combines the advantages of normalizing flows and energy-based models in a novel way. Flow-matching per se does not enable transferability. However, it helps toward this goal by allowing the training of neural network force fields in a data-efficient way, thus significantly reducing the burden of generating extensive training data.

https://doi.org/10.1021/acs.jctc.3c00016
J. Chem. Theory Comput. 2023, 19, 942−952

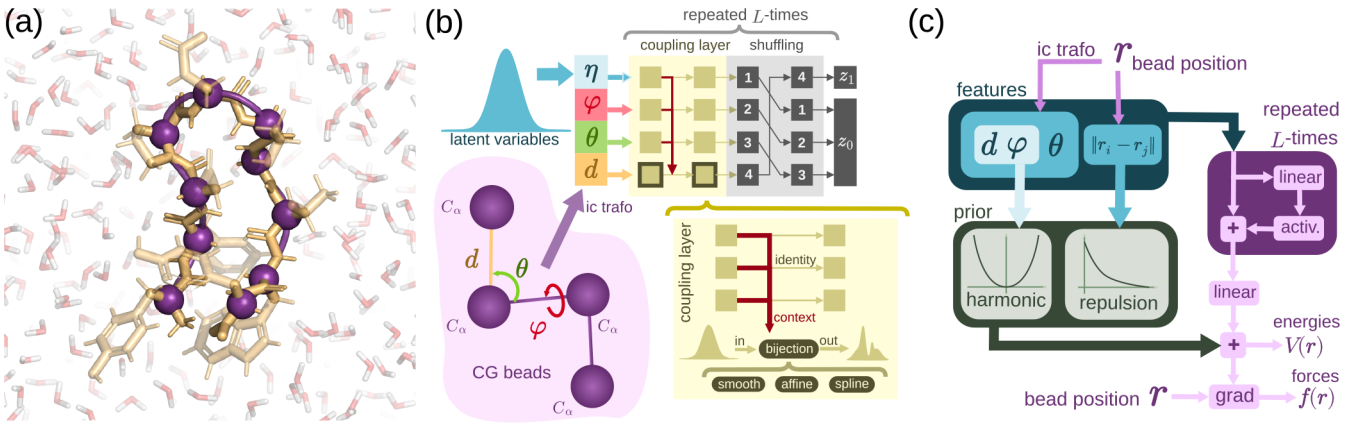


Figure 2. (a) Chignolin in explicit solvent. The magenta spheres show the CG beads at C_α resolution. (b) Normalizing flow architecture used in this work to model $\mathcal{V}(\cdot; \boldsymbol{\theta}_{\text{flow}})$. After transforming the CG beads into an internal coordinate (IC) representation made from bonds (d), angles (θ), and dihedral torsions (φ) a trainable stack of coupling layers transform them into uniform noise. See Figure S1 for a more detailed illustration of the flow architecture. (c) Modified *CGnet* architecture used in this work to model $V(\cdot; \boldsymbol{\theta}_{\text{pot}})$. "grad" stands for computing the gradient using automatic differentiation.

As a proof of concept, we apply the method to the coarse-graining of small protein systems. We show that accurate CG models can be fit to equilibrium trajectories without using atomistic forces or intermediate simulations. Even when forces are available, we find that flow-matching is much more data-efficient than force-matching and yields surprisingly accurate force fields on small data sets.

2. COARSE-GRAINING WITH FLOW-MATCHING

2.1. Coarse-Graining with Thermodynamic Consistency. We consider a molecular system with atomic coordinates $\boldsymbol{R} \in \mathbb{R}^{3N}$ in thermodynamic equilibrium following an equilibrium distribution

$$\mu(\boldsymbol{R}) \propto \exp(-u(\boldsymbol{R})) \tag{1}$$

where u is the reduced potential energy of the system, whose exact form depends on the choice of the ensemble, e.g., $u(\boldsymbol{R}) = U(\boldsymbol{R})/kT$ for the canonical ensemble with potential energy $U(\boldsymbol{R})$, temperature T, and Boltzmann constant k.

Coarse-graining considers a mapping Ξ: $\mathbb{R}^{3N} \to \mathbb{R}^{3n}$ that projects fine-grained states $\boldsymbol{R}$ onto a lower-dimensional representation $\boldsymbol{r}$. In the present work, we only consider linear and orthogonal maps, $\boldsymbol{r} = \Xi R$. For nonorthogonal or even nonlinear maps, the subsequent mathematical treatment must be generalized.[51,52] As an example, the conformational dynamics of a protein with N atoms can be projected onto a chosen set of beads by only considering the C_α-atoms in the backbone (Figure 2a). Coarse-graining with thermodynamic consistency aims at parametrizing a CG model which yields the same density over the CG coordinates as the marginal distribution from the original system; i.e.,

$$\nu(\boldsymbol{r}) = \int \mathrm{d}\boldsymbol{R}\, \mu(\boldsymbol{R}) \cdot \delta_{[\Xi \boldsymbol{R} = r]}(\boldsymbol{R}) \tag{2}$$

The CG model is often defined by a CG potential $V(\cdot; \boldsymbol{\theta}_{\text{pot}})$ with parameters: $\nu(\cdot; \boldsymbol{\theta}_{\text{pot}}) \propto \exp(-V(\cdot; \boldsymbol{\theta}_{\text{pot}}))$. Two conventional parametrization approaches will be introduced below. It is important to stress that designing a CG force-field by trying to optimize *thermodynamic* consistency does not imply that also the *dynamical* properties are well approximated.[19,53]

2.2. Variational Force-Matching. One option is to optimize a candidate potential $V(\cdot; \boldsymbol{\theta}_{\text{pot}})$ with the force information from the ground-truth potential u (Figure 1a). Given a set of fine-grained samples (e.g., MD trajectory) $\mathcal{D} = (\boldsymbol{R}_1, \ldots, \boldsymbol{R}_T)$ with corresponding forces $\boldsymbol{f}(\boldsymbol{R}) = -\nabla u(\boldsymbol{R})$, it is shown that the thermodynamically consistent CG potential (eq 2) can be approximated by the potential minimizing the *variational force-matching loss*[33]

$$\mathcal{L}(\boldsymbol{\theta}_{\text{pot}}) = \mathbb{E}_{\boldsymbol{R},\boldsymbol{f}\sim\mathcal{D}}[\|\Xi_f \boldsymbol{f} + \nabla_{\Xi \boldsymbol{R}} V(\Xi \boldsymbol{R}; \boldsymbol{\theta}_{\text{pot}})\|_2^2] \tag{3}$$

in which Ξ_f is a force mapping operator dependent on map Ξ. When infinite samples $\mathcal{D}$ and all functional forms for V are available, the minimization of the loss (eq 3) yields exactly the thermodynamically consistent potential defined by eq 2. Even with finite samples and restrictions on the $V(\cdot; \boldsymbol{\theta}_{\text{pot}})$, the result from the loss minimization still provides a variational approximation in practice. Because of their enhanced expressiveness, neural networks with physical inductive biases have been shown to be a useful model class for the parametrization of $V(\cdot; \boldsymbol{\theta}_{\text{pot}})$.[20,22]

2.3. Density Estimation/Relative Entropy Method. Force-matching requires the mapped CG forces to be saved during fine-grained sampling, which is not common practice. Alternatively, one can directly learn a CG model via density estimation on the observed conformational space. Density estimation aims at minimizing the following objective

$$\mathcal{L}(\boldsymbol{\theta}_{\text{pot}}) = \mathbb{E}_{\boldsymbol{R}\sim\mathcal{D}}[-\log \nu(\Xi \boldsymbol{R}; \boldsymbol{\theta}_{\text{pot}})] \tag{4}$$

The minimum can be interpreted as the maximum-likelihood solution of an energy-based model trained on the projected samples $\Xi\mathcal{D} = (\Xi \boldsymbol{R}_1, \ldots \Xi \boldsymbol{R}_T)$. This approach can be related to the relative entropy method in molecular simulation[34] and is used for training an energy-based model in the field of machine learning.[54] Unfortunately, computing the gradients of eq 4 with respect to parameters generates a sampling problem. Computing the gradient contribution of the normalizing constant involves sampling from the model density ν, which means that the CG model needs to be periodically resampled during training (Figure 1b).

2.4. Flow-Based Density Estimation. We can avoid the sampling problem of eq 4 by using the density $\nu(\cdot; \boldsymbol{\theta}_{\text{flow}})$

https://doi.org/10.1021/acs.jctc.3c00016
J. Chem. Theory Comput. 2023, 19, 942−952

corresponding to a model that can be efficiently sampled, such as normalizing flows.[37−39] Flows are invertible neural networks $\Phi(\cdot; \boldsymbol{\theta}_{\text{flow}})$: $\mathbb{R}^n \to \mathbb{R}^n$ that transform an easy-to-sample reference distribution $q(\boldsymbol{z})$, e.g., a Gaussian or uniform density, into our target density. If we sample $\boldsymbol{z} \sim q(\boldsymbol{z})$ and transform it into $\boldsymbol{r} = \Phi(\boldsymbol{z}; \boldsymbol{\theta}_{\text{flow}})$, the resulting density is given by

$$p(\boldsymbol{r}; \boldsymbol{\theta}_{\text{flow}}) = q(\Phi^{-1}(\boldsymbol{r}; \boldsymbol{\theta}_{\text{flow}}))\cdot|J_{\Phi^{-1}}(\boldsymbol{r}; \boldsymbol{\theta}_{\text{flow}})| \tag{5}$$

Inserting eq 5 into eq 4, we get an efficient training objective. After training, the energy of the normalizing flow

$$\mathcal{V}(\boldsymbol{r}; \boldsymbol{\theta}_{\text{flow}}) = -\log p(\boldsymbol{r}; \boldsymbol{\theta}_{\text{flow}}) \tag{6}$$

approximates the CG PMF.

2.5. Variational Density Estimation. Direct density estimation with flow models suffers from the fact that the flow architecture is constrained in order to represent an invertible function, which compromises their representative power and training dynamics. As a solution, we consider relaxing the bijectivity constraint by introducing k additional variables and sampling a joint state $\boldsymbol{z} = (\boldsymbol{z}_0, \boldsymbol{z}_1) \in \mathbb{R}^{n+k}$ from a joint (Gaussian/uniform) reference density $q(\boldsymbol{z}_0, \boldsymbol{z}_1)$ (Figure 2b). Now we define Φ as an invertible coordinate transformation (e.g., a flow model) over those joint $n + k$ degrees of freedom. Similarly as before, we get the output density $p(\boldsymbol{r}, \boldsymbol{\eta}; \boldsymbol{\theta}_{\text{flow}})$ of a transformed pair $(\boldsymbol{r}, \boldsymbol{\eta}) = \Phi(\boldsymbol{z}_0, \boldsymbol{z}_1; \boldsymbol{\theta}_{\text{flow}})$. The marginal density over $\boldsymbol{r}$ of this augmented model cannot be computed efficiently. However, we can still optimize a variational bound of the likelihood: we first define a joint density $\nu(\boldsymbol{r}, \boldsymbol{\eta}) = \nu(\boldsymbol{r})\cdot\tilde{\nu}(\boldsymbol{\eta}|\boldsymbol{r})$ by introducing a Gaussian conditional density $\tilde{\nu}(\boldsymbol{\eta}|\boldsymbol{r})$ and then minimize

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}_{\text{flow}}) &= \mathbb{E}_{\boldsymbol{R}\sim\mathcal{D},\boldsymbol{\eta}\sim\tilde{\nu}(\boldsymbol{\eta}|\boldsymbol{r})}[-\log p(\Xi\boldsymbol{R}, \boldsymbol{\eta}; \boldsymbol{\theta}_{\text{flow}})] \\ &\geq \mathbb{E}_{\boldsymbol{R}\sim\mathcal{D}}[-\log p(\Xi\boldsymbol{R}; \boldsymbol{\theta}_{\text{flow}})]\end{aligned} \tag{7}$$

As shown in refs 55 and 56, normalizing flows with additional noise dimensions can alleviate limitations of invertible neural networks to transform a simple, unimodal, prior density to a complex, multimodal target density.[50,57,58] While the extra dimensions do not allow us to directly compute the density $p(\boldsymbol{r}; \boldsymbol{\theta}_{\text{flow}})$, and thus $\mathcal{V}(\boldsymbol{r}, \boldsymbol{\theta}_{\text{flow}})$ as well as the corresponding forces, we can still compute a joint energy model over CG coordinates and latent variables

$$\mathcal{V}(\boldsymbol{r}, \boldsymbol{\eta}; \boldsymbol{\theta}_{\text{flow}}) = -\log p(\boldsymbol{r}, \boldsymbol{\eta}; \boldsymbol{\theta}_{\text{flow}}) \tag{8}$$

which can be used to train an arbitrary model of the CG potential as follows.

2.6. Teacher−Student Force-Matching. Our idea is to teach the information about the distribution of the CG coordinates $\boldsymbol{r}$ contained in a trained latent-variable model $\mathcal{V}(\boldsymbol{r}, \boldsymbol{\eta}; \boldsymbol{\theta}_{\text{flow}})$ to a "student" CG potential $V(\boldsymbol{r}; \boldsymbol{\theta}_{\text{pot}})$ that does not suffer from the architectural constraints of flows (Figure 2c). We first draw samples $(\boldsymbol{r}, \boldsymbol{\eta})$ from our flow model and compute instantaneous forces over CG coordinates $\boldsymbol{r}$:

$$\tilde{\boldsymbol{f}}(\boldsymbol{r}, \boldsymbol{\eta}; \boldsymbol{\theta}_{\text{flow}}) = -\nabla_{\boldsymbol{r}}\mathcal{V}(\boldsymbol{r}, \boldsymbol{\eta}; \boldsymbol{\theta}_{\text{flow}}) \tag{9}$$

Any given $\boldsymbol{r}$ may correspond to different $\tilde{\boldsymbol{f}}$, but on average they give rise to the unbiased mean force:

$$\boldsymbol{f}(\boldsymbol{r}; \boldsymbol{\theta}_{\text{flow}}) = \mathbb{E}_{\boldsymbol{\eta}\sim p(\boldsymbol{\eta}|\boldsymbol{r};\boldsymbol{\theta}_{\text{flow}})}[\tilde{\boldsymbol{f}}(\boldsymbol{r}, \boldsymbol{\eta}; \boldsymbol{\theta}_{\text{flow}})] \tag{10}$$

This relation allows us to efficiently train an unconstrained $V(\boldsymbol{r}; \boldsymbol{\theta}_{\text{pot}})$ via the variational force-matching objective

$$\mathcal{L}(\boldsymbol{\theta}_{\text{pot}}) = \mathbb{E}_{(\boldsymbol{r},\boldsymbol{\eta})\sim p(\boldsymbol{\theta}_{\text{flow}})}[\|\tilde{\boldsymbol{f}}(\boldsymbol{r}, \boldsymbol{\eta}; \boldsymbol{\theta}_{\text{flow}}) + \nabla_{\boldsymbol{r}}V(\boldsymbol{r}; \boldsymbol{\theta}_{\text{pot}})\|_2^2] \tag{11}$$

As shown in the Supporting Information (SI), the gradients of eq 11 with respect to the $\boldsymbol{\theta}_{\text{pot}}$ provides an unbiased estimator that does not depend on $\boldsymbol{\theta}_{\text{flow}}$. The proposed approach resembles the conventional one for coarse-graining but with the difference being that it averages over fewer degrees $\boldsymbol{\eta}$ rather than a larger amount of (mainly solvent) degrees of freedom.

As will be shown in Results, the student model can mitigate flaws in the flow models, namely, samples that deviate from physics laws (e.g., containing steric clashes) and the ruggedness of the CG free energy surface. The student model is also regularized to entail a more robust CG potential than the direct force output of the flow for molecular dynamics simulation. In addition, the flexibility in choosing the functional form of the CG free energy allows built-in symmetries such as roto-translational energy invariance[20] and parameter sharing for obtaining a transferable force field.[22]

3. RESULTS

We now employ the flow-matching method to obtain CG molecular models of small proteins. To this end, we train flows on the CG coordinate samples extracted from all-atom simulation trajectories. Trained flow models can generate CG coordinates and accompanying forces, which in turn are used to train a neural CG potential via force-matching. For demonstration purposes, this work uses an improved version of the CGnet architecture[20] to represent the CG force field (Figure 2c; see also Methods in the SI). Therefore, these secondary CG models will be denoted as "Flow-CGnets".

3.1. Flow-Matching Learning of Accurate CG Force Fields. As a first example, we consider capped alanine, also known as alanine dipeptide, to demonstrate that flow-matching can learn accurate CG force fields and achieve much higher statistical efficiency than force-matching. As in previous work,[20,22] the CG mapping is defined as slicing out the coordinates of five backbone carbons and nitrogens (Figure 3a).

We quantify the accuracy of different methods based on equilibrium statistics from either direct sampling (for flows) or long simulation trajectories (for CGnets). We focus on the joint distributions of the ϕ and ψ dihedral angles along the backbone (i.e., Ramachandran plot, Figure 3b), which are the main degrees of freedom for this system.[59] The ground truth for comparison comes from all-atom MD simulation (2 μs in total; see Methods in the SI). As for baseline, we use CGnets trained with classical force-matching[20,33] employing forces stored during all-atom simulations. As illustrated by Figure 3b, the flow and Flow-CGnet can recover the reference distribution to a very good approximation when only 20000 reference all-atom conformations are used. In contrast, a normal CGnet cannot effectively model the dihedral free energy in this low data regime, even with the additionally available force information: The free energy minima are more or less located according to the ground truth (representative conformations from all-atom and two CGnet models illustrated and compared in SI Figure S6), but the dihedral distribution smears over the whole space. When increasing the amount of training data, also CGnet trained with force-matching can well approximate the free energy landscape (as reported in ref 20, where 8×10^5 configurations and forces were used) but never reaches the flow-matching accuracy for the available data set

https://doi.org/10.1021/acs.jctc.3c00016
J. Chem. Theory Comput. 2023, 19, 942−952

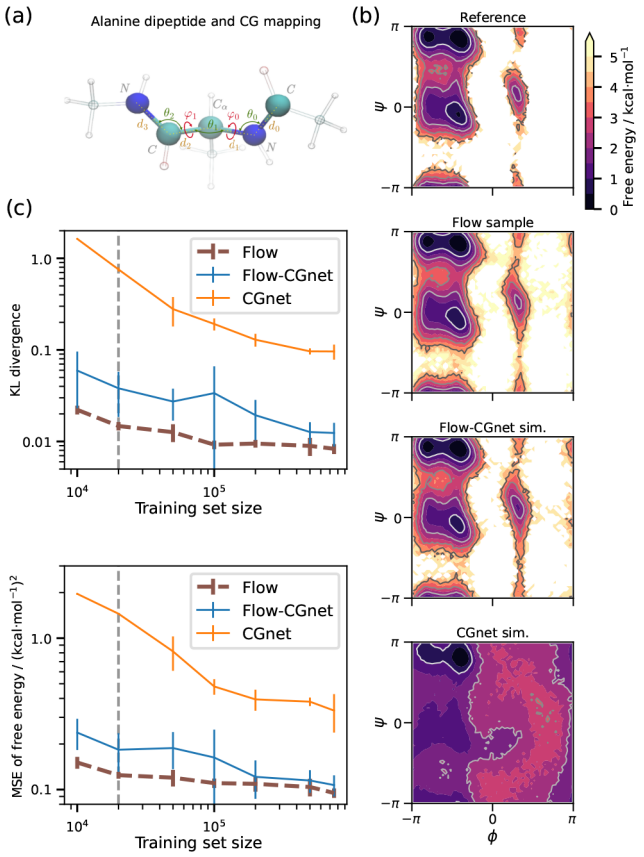


Figure 3. Application of flow-matching on capped alanine. (a) CG mappings used for the flow and CGnets, φ_0 and φ_1, represent main chain torsion angles ϕ and ψ, respectively. (b) Free energy profile of capped alanine projected on the ϕ/ψ plane (Ramachandran plot) for the all-atom ground truth from MD simulation (ground truth), for the flow model, for the Flow-CGnet, and for original CGnet model (baseline). The latter three were trained against only 20000 data points from the reference data (vertical gray dashed lines in panel c). (c) Model accuracy as a function of training set size for capped alanine. Shown metrics are estimated KL divergence and MSE between discrete free energies on the ϕ/ψ plane. Brown dashed curves correspond to the flow after MLE training, while solid lines show values for the *CGnets* trained on either the flow sample (blue) or the all-atom ground-truth sample (orange).

(Figure 3c). This comparison displays the advantage of the flow-matching method, which infers the boundary of free energy basins as well as relative weighting between different metastable states better than force-matching, especially for regions rarely covered by the training data, e.g., at transition states.

3.2. Flow-Matching Delivering Better Data Efficiency than Force-Matching. The better accuracy of Flow-CGnet models can be attributed to higher statistical efficiency. For illustration, we measure the effects of the training set size on the KL divergence and mean-squared error of torsional free energy, which are computed on a discrete histogram against the validation data distribution.[22] Concretely, we perform training with a varying number of samples in the training set for both flow-matching and baseline force-matching. Detailed training setup can be found in the SI.

It can be observed that the direct samples from the flow model rank first regarding both criteria (Figure 3c), which renders the knowledge transfer to a student Flow-CGnet model to be "lossy". Nevertheless, the secondary model provides a potential that is not only faster to evaluate but also numerically more stable for CG molecular dynamics. Despite that the flow model automatically provides a differentiable energy function, it is not fully accurate in regions with low Boltzmann probabilities: A simulation with flow potential often visits spurious states outside of the distribution and sometimes experiences numerical blow-ups on the boundary of training data distribution. This issue is solved by our two-stage training strategy, in which the CGnet can incorporate an additive, physics-inspired term (i.e., the prior energy) to set simulation-friendly free energy barriers and rule out outlier conformations.[20] Flow samples with unrealistically high force magnitude or located in unrealistic conformational regions can be filtered or reweighted before feeding to the CGnet training (see Methods in the SI). The remaining samples mostly lie in the high-probability region, thus bringing informative forces for force-matching training. As a result, the Flow-CGnet also benefits from the flow's efficiency: it achieves an equivalent performance of CGnet at full data set size even with the smallest tested input data amount (Figure 3c).

In SI Figure S2 we show a similar analysis of the data efficiency of CGnet, the flow, and Flow-CGnet for the miniprotein chignolin. As expected, the situation is even more extreme than for alanine dipeptide: The Flow-CGnet trained on only 2×10^4 data points is on par with the CGnet trained on all available 1.4×10^6 data points, resulting in a 70× data efficiency, which is expected to further increase for larger systems.

How can the greater data efficiency of Flow-CGnet compared to force-matching be explained? While the accuracy of the flow to approximate the Boltzmann distribution depends on the number of conformations used to train it, it achieves a very good approximation with relatively few observed conformations compared to force-matching (Figure S3). Although its free energy surface is not necessarily well behaved in all local details, the flow can generate abundant samples and forces from the learned distribution, thereby cheaply reducing the error of the trained Flow-CGnet to a similar level as the intrinsic approximation error of the flow (Figure S3). Additionally, the augmentation channels in the flow model are much fewer in number and have simpler distribution than the internal degrees of freedom in the all-atom system, and therefore the flow's sample forces have much less noise than instantaneous forces stored in all-atom simulations and better represent the CG mean force. In this sense, when a proper sample filtering scheme and regularizations on the CGnet models are adopted, the flow can become superior to a limited set of all-atom data in terms of the number of samples as well as the signal-to-noise ratio of forces it feeds to the secondary CGnet. The performance in this test case suggests Flow-CGnets may extend the application of neural CG potentials to more complex macromolecular systems, where usually only a limited amount of conformations and no forces are available.

3.3. Flow-Matching of Fast-Folding Proteins. The flow-matching method is applied to molecular trajectories of four small proteins from ref 5, namely, chignolin, tryptophan cage (trpcage), the α/β protein BBA (bba), and the villin headpiece (villin) that consist of 10, 20, 28, and 35 amino acids, respectively (see ref 5 for simulation details). These small proteins can be modeled by a flow that operates fully in internal coordinates. As for other fast-folding proteins in ref 5, some only have a marginally stable state that closely resembles

https://doi.org/10.1021/acs.jctc.3c00016
J. Chem. Theory Comput. 2023, 19, 942−952

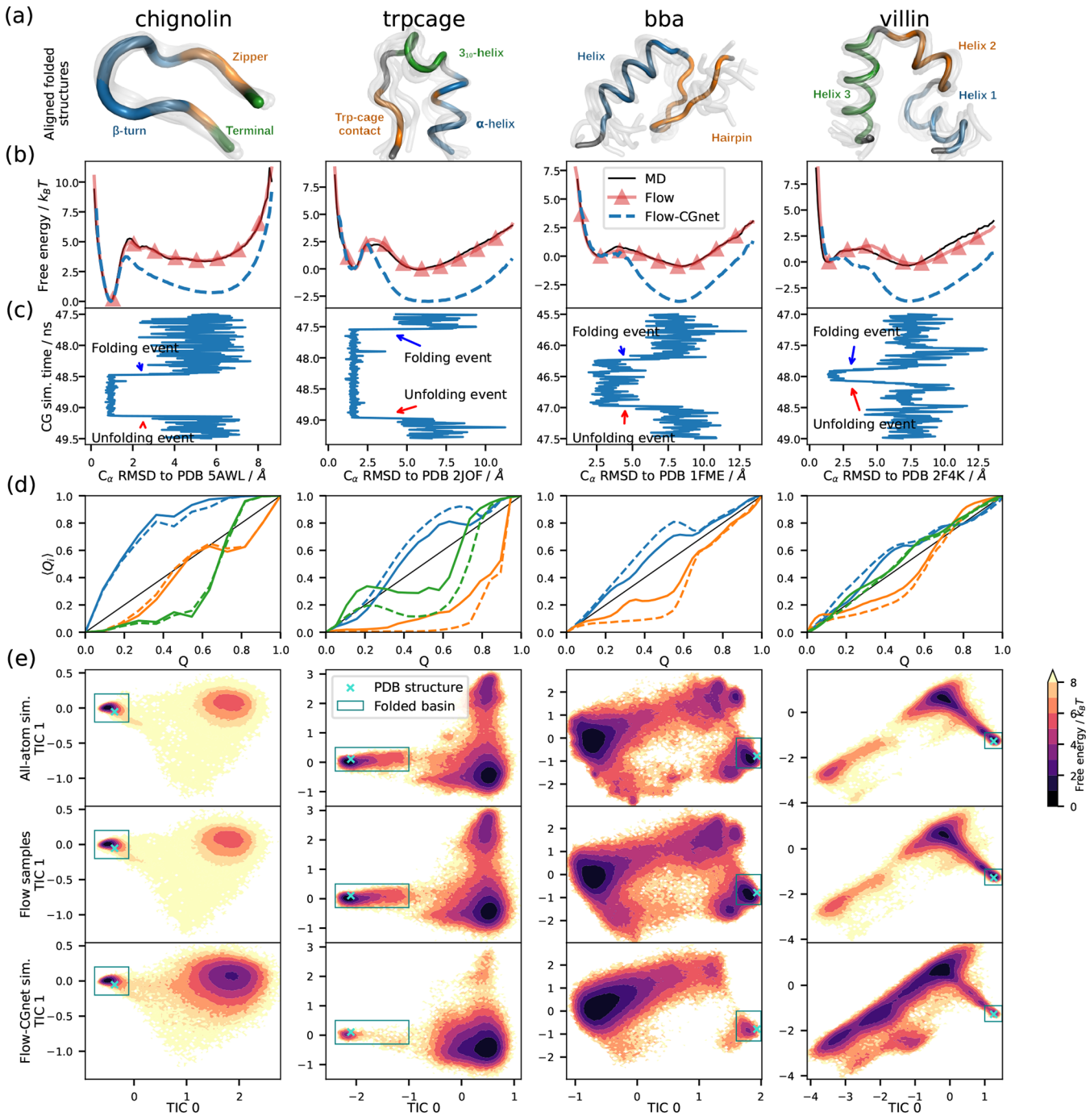


Figure 4. Flow-matching results for four fast-folding proteins. From top to bottom: (a) 10 exemplary folded samples from CG simulation (shown in half-transparent gray color) superposed on the experimental structure (colored segments correspond to important elements in the folding process); (b) free energy curve over RMSD for the MD, flow, and Flow-CGnet samples with PDB structure as the reference; (c) RMSD time series excerpted from CG simulation showing folding and unfolding events; (d) average fractions of native contacts $\langle Q_i \rangle$ in different segments of the protein formed at each stage of the folding process (identified by the fraction of all native contacts formed, Q; segments are determined mainly according to secondary structures and are highlighted with the corresponding color in panel a); (e) free energy landscapes of all-atom MD, flow, and Flow-CGnet model over TICs (turquoise crosses and teal rectangles denoting experimental structures and folded state according to MD trajectories, respectively).

the PDB structure throughout the all-atom trajectories, e.g., BBL; for some fast folders, we can acquire reasonably good flow models but the folded state cannot be stabilized by the subsequent Flow-CGnet models, e.g., wwdomain and homeodomain; for the rest, the internal-coordinate-based flow model cannot effectively capture the full free energy surface (see Discussion section on scalability). Each trajectory corresponds to at least 100 μs of all-atom MD. Note that the trajectories do not contain atomistic forces, so force-matching is not an option for parametrizing a CG force field based on these data. Relative entropy minimization is difficult because it would require iteratively resampling the CG model during training, introducing excessive computational cost.

https://doi.org/10.1021/acs.jctc.3c00016
J. Chem. Theory Comput. 2023, 19, 942−952

All four proteins are coarse grained using one bead per residue placed upon the C_α (see Figure 2a). First, normalizing flows are trained for each protein using likelihood maximization on the C_α coordinates. Second, synthetic position/force pairs are generated by the flow, of which the outliers are filtered and reweighted according to the extent they exceed the force magnitude boundary and violate the minimum pairwise distances, respectively. Last, the protein-specific CGnets are optimized via force-matching on the processed flow samples. The final CGnets are simulated using Langevin dynamics with parallel tempering to produce equilibrium samples from the CG model. The trajectories from the replica at the same temperature as the all-atom simulation are used for the analyses below. In order to show that folding and unfolding events occur without enhanced sampling strategies, we also performed pure Langevin dynamics simulations with learned Flow-CGnet models. The details on the procedure of training and simulation as well as hyperparameter choices can be found in Methods in the SI.

3.4. Flow-CGnets Recovering Native Structures. Figure 4 compares protein folding between the atomistic and CG simulations. All CG models recover the folded PDB structures up to 2.5 Å RMSD, which is of similar quality as the reference all-atom simulations. Figure 4a shows representative structures from the CG simulations superposed with the experimental crystal structures, and Figure S7 provides a detailed comparison between the CG and atomistic structural ensembles corresponding to the different minima in the free energy landscape of the four proteins studied, demonstrating excellent agreement. The free energy plots over the RMSD (Figure 4b) indicate that the CG conformational distribution matches the projected all-atom trajectory for the folded basin: The free energy valleys with the lowest RMSD values are centered around almost the same RMSD value and have nearly indistinguishable widths between the CG and MD densities, which indicates that all CG models accurately represent the flexibility of their respective folded states.

3.5. Flow-CGnets Matching Folding Thermodynamics Qualitatively. Moving into the unfolded region (RMSD $\geq$ 5 Å in Figure 4b), the match between atomistic and CG free energies deteriorates. While all CG models exhibit the characteristic folding free energy barrier, the height of this barrier and the folded/unfolded ratio differ between the MD and CG data. Generally, the folded states are less stable in the CG model. While the flow differs by less than $\approx 1kT$ from the all-atom result, the Flow-CGnet underestimates the folding free energy by up to $3kT$.

Nevertheless, frequent transitions between folded and unfolded configurations were observed in 50 ns simulation runs without parallel tempering (Figure 4c). This assures that the models still keep the two states kinetically connected.

3.6. Flow-CGnets Reproducing the Folding Mechanisms. Figure 4d illustrates the sequence of formation of the protein structure elements during folding, for the all-atom model and the corresponding Flow-CGnet model of the four proteins studied. The average fractions of native contacts $\langle Q_i \rangle$ formed in different segments of the protein along the folding process[60] are reported and show that the order of formation of the different secondary structure elements is recovered by Flow-CGnet to a good approximation.

3.7. Flow-CGnets Well Approximating Folding Free Energy Landscape. Figure 4e shows the joint densities over the first two TICA coordinates;[61−63] see SI Section D.5. These reaction coordinates visualize the slowest processes in the MD simulation, which correspond to folding and unfolding; see the SI for details. The Flow-CGnet densities resemble the atomistic densities, showing that the global patterns in the folding process are captured. The match deteriorates with increasing sequence length: for chignolin the Flow-CGnet recovers the shape of the distribution well, for trpcage and bba some minor metastable states are missing, and for villin some regions that are sparsely populated in the MD data are overstabilized.

4. DISCUSSION

4.1. Training Data Requirements. Flow-matching does not require the forces to be saved with the simulation data and is thus more readily applicable than force-matching. We have also shown that matching the empirical distribution benefits data efficiency. A drawback is that flow-matching requires the underlying all-atom data to come from an equilibrated ensemble. However, this does not need to be achieved by long simulation trajectories: reweighting from biased ensembles, such as replica-exchange simulations, or reweighting of short trajectories via Markov state models[3,4] is possible.

There are also theoretical developments in generalizing the force-matching method for non-equilibrium cases, such as ref 64. In such situations (but generally whenever atomistic force information is available), it might be beneficial to train the flow by combining density estimation with force-matching. Such a mixed loss can be especially efficient when using flows with continuous forces.[65]

4.2. Architectural Choices for Neural Networks. The teacher neural network needs to (i) be trainable via (approximate) likelihood on sampling data, (ii) permit efficient sampling, and (iii) allow us to compute the instantaneous forces (eq 9). We found that smooth mixture flows[65] on the internal coordinates are able to reproduce the CG conformational distribution very accurately. Other latent variable models, including different normalizing flow architectures as well as variational autoencoders[66] and their generalizations,[50,67] could be used as well. Examples of other generative networks used in coarse-graining applications can be found in refs 25, 47, and 48.

The student neural network is trained to represent the CG free energy. While here we used a modified version of the simple CGnet method,[20] this network could be replaced by more advanced neural network architectures, such as SchNet,[68] other graph neural networks,[22,69−75] or other machine learning methods.[76] In principle, flow-matching can be combined with any trainable CG model, either based on neural networks or fixed functional forms with adjustable parameters.

4.3. Scalability to Larger Molecules. We observed that the CG model quality deteriorated and eventually became unusable for larger proteins. This is because the present normalizing flows are built on a global internal coordinate representation. As the length of the peptide chain grows, the target potential energy becomes extremely sensitive with respect to these internal coordinates. For example, a tiny rotation of one torsion can easily cause steric clashes in a different part of the molecule. This may lead to, for example, a sigificant decrease of effective size of the training set after repulsion-based reweighting (see Figure S4). Other work[40−42,45,46] has also found suitable flow architectures for small molecules, proteins, and even explicitly solvated systems but did not report whether they could produce quantitatively

https://doi.org/10.1021/acs.jctc.3c00016
J. Chem. Theory Comput. 2023, 19, 942−952

matching forces. One possibility to scale to large molecules is to employ coupling flows with equivariant neural networks operating in Cartesian space while still informed by internal coordinates, but further work is needed in order to find suitable architectures that can sample low-energy structures and have the relevant physics built in.

5. CONCLUSIONS

We have developed a two-stage approach to bottom-up coarse-graining that addresses two major problems with classical approaches, namely, data availability and efficiency. The *flow-matching* method produces thermodynamically consistent CG models without relying on either all-atom ground-truth forces or subsequent CG simulations. The key ingredient of our method is a generative deep neural network that is introduced into the optimization workflow. Compared with classical force-matching, flow-matching combined with CGnet captures the global thermodynamics of small peptides much more accurately than CGnet models trained with force-matching. Interestingly, this was even the case, when only a fraction (<10%) of the data was used during training. The main factor determining the data efficiency of flow-matching with respect to force-matching is the ratio of the number of atoms versus the number of CG particles. For the examples described in this work—where the all-atom systems are solvated macromolecules and the CG models retain only a few solute atoms—this ratio is very large, and the instantaneous all-atom forces projected on the CG coordinates are very noisy.

Applications to four small proteins yielded CG potentials that were able to fold and unfold the proteins via the same pathways as all-atom MD. Biopolymers such as proteins are an especially interesting candidate for our proposed method, because they can be extremely difficult to sample, which makes the speedup obtained by a CG force field more practically attractive. Furthermore, bottom-up coarse-graining in the present manner is applicable to many other molecular systems, including other polymers, liquids, and materials. Thereby, the present work opens a new and efficient path to reach near-atomistic accuracy on scales not amenable to atomistic simulations.

The two-step machine learning architecture consisting of a teacher and a student model gives rise to an interesting strategy for training transferable CG potentials: One may train separate system-specific teacher networks (e.g., flows) and then train a shared CG force field to obtain a transferable molecular model across the chemical space represented by the training data. Again, biopolymers are particularly interesting candidates for transferable CG force fields, as they usually consist of relatively few chemical building blocks which simplifies the parametrization of a force field that can generalize across all sequences. We envisage that flow-matching will be an important contribution to the development of transferable CG force fields and thereby help us to access time and length scales currently inaccessible to accurate molecular models.

■ ASSOCIATED CONTENT

Data Availability Statement

All CG samples from flow, Flow-CGnet, and conventional CGnet models involved in the analyses in the main text have been deposited on the Zenodo platform with https://zenodo.org/record/7092157#.Y8gV_nbMITo. The accompanying code for flow-matching and application examples are available at https://github.com/noegroup/flowm.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.3c00016.

Concise summary of the methods applied for the training, validation and sampling of all coarse grained models; comprehensive theoretical derivations (Section B); all necessary details of the coarse grained flow and CGnet model training and validation (Section C) and of the sample analyses and comparisons (Section D) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Andreas Krämer − *Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany;* orcid.org/0000-0002-7699-3083; Email: andreas.kraemer@fu-berlin.de

Cecilia Clementi − *Department of Physics, Freie Universität Berlin, 14195 Berlin, Germany; Center for Theoretical Biological Physics, Department of Physics, and Department of Chemistry, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0001-9221-2358; Email: cecilia.clementi@fu-berlin.de

Frank Noé − *Microsoft Research AI4Science, 10178 Berlin, Germany; Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany; Department of Physics, Freie Universität Berlin, 14195 Berlin, Germany; Department of Chemistry, Rice University, Houston, Texas 77005, United States*; Email: frank.noe@fu-berlin.de

Authors

Jonas Köhler − *Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany*

Yaoyi Chen − *Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany;* orcid.org/0000-0002-9299-2005

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.3c00016

Author Contributions

[▽]J.K., Y.C., and A.K. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Aleksander E. P. Durumeric, Nicholas E. Charron, Brooke E. Husic, Klara Bonneau, Manuel Dibak, Leon Klein, Michele Invernizzi, and Leon Sixt for insightful discussions. We gratefully acknowledge funding from the European Commission (Grant No. ERC CoG 772230 "ScaleCell"), the International Max Planck Research School for Biology and Computation (IMPRS−BAC), the BMBF (Berlin Institute for Learning and Data, BIFOLD), the Berlin Mathematics center MATH+ (AA1-6, EF1-2), and the Deutsche Forschungsgemeinschaft DFG (GRK DAEDALUS, SFB1114/A04 and B08). C.C. acknowledges funding from the Deutsche Forschungsgemeinschaft DFG (SFB/TRR 186, Project A12; SFB 1114, Projects B03 and A04; SFB 1078, Project C7; and RTG 2433, Project Q05), the National Science Foundation (CHE-1900374, and PHY-2019745), and the Einstein

Foundation Berlin (Project 0420815101). The 3D molecular structures are visualized with PyMOL.[77]

REFERENCES

(1) Shaw, D. E.; Grossman, J. P.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Deneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Greskamp, B.; Ho, C. R.; Ierardi, D. J.; Iserovich, L.; Kuskin, J. S.; Larson, R. H.; Layman, T.; Lee, L. S.; Lerer, A. K.; Li, C.; Killebrew, D.; Mackenzie, K. M.; Mok, S. Y. H.; Moraes, M. A.; Mueller, R.; Nociolo, L. J.; Peticolas, J. L.; Quan, T.; Ramot, D.; Salmon, J. K.; Scarpazza, D. P.; Ben Schafer, U.; Siddique, N.; Snyder, C. W.; Spengler, J.; Tang, P. T. P.; Theobald, M.; Toma, H.; Towles, B.; Vitale, B.; Wang, S. C.; Young, C.Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. *SC'14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*; IEEE, 2014; pp 41−53. DOI: 10.1109/SC.2014.9.

(2) Shaw, D. E.; Adams, P. J.; Azaria, A.; Bank, J. A.; Batson, B.; Bell, A.; Bergdorf, M.; Bhatt, J.; Adam Butts, J.; Correi, T.; Dirks, R. M.; Dror, R. O.; Eastwoo, M. P.; Edwards, B.; Even, A.; Feldmann, P.; Fenn, M.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Gorlatova, M.; Greskamp, B.; Grossman, J. P.; Gullingsrud, J.; Harper, A.; Hasenplaugh, W.; Heily, M.; Heshmat, B. C.; Hunt, J.; Ierardi, D. J.; Iserovich, L.; Jackson, B. L.; Johnson, N. P.; Kirk, M. M.; Klepeis, J. L.; Kuskin, J. S.; Mackenzie, K. M.; Mader, R. J.; McGowen, R.; McLaughlin, A.; Moraes, M. A.; Nasr, M. H.; Nociolo, L. J.; O'Donnell, L.; Parker, A.; Peticolas, J. L.; Pocina, G.; Predescu, C.; Quan, T.; Salmon, J. K.; Schwink, C.; Shim, K. S.; Siddique, N.; Spengler, J.; Szalay, T.; Tabladillo, R.; Tartler, R.; Taube, A. G.; Theobald, M.; Towles, B.; Vick, W.; Wang, S. C.; Wazlowski, M.; Weingarten, M. J.; Williams, J. M.; Yuh, K. A.Anton 3: Twenty Microseconds of Molecular Dynamics Simulation before Lunch. *SC21: International Conference for High Performance Computing, Networking, Storage and Analysis*, Nov. 14−19, 2021, St. Louis, MO, USA; IEEE, 2021.

(3) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134*, 174105.

(4) Husic, B. E.; Pande, V. S. Markov state models: From an art to a science. *J. Am. Chem. Soc.* **2018**, *140*, 2386−2396.

(5) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517−520.

(6) Plattner, N.; Doerr, S.; De Fabritiis, G.; Noé, F. Complete protein−protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nat. Chem.* **2017**, *9*, 1005−1011.

(7) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **2000**, *298*, 937−953.

(8) Clementi, C. Coarse-grained models of protein folding: Toy-models or predictive tools? *Curr. Opin. Struct. Biol.* **2008**, *18*, 10−15.

(9) Matysiak, S.; Clementi, C. Optimal combination of theory and experiment for the characterization of the protein folding landscape of S6: How far can a minimalist model go? *J. Mol. Biol.* **2004**, *343*, 235−248.

(10) Matysiak, S.; Clementi, C. Minimalist Protein Model as a Diagnostic Tool for Misfolding and Aggregation. *J. Mol. Biol.* **2006**, *363*, 297−308.

(11) Das, P.; Matysiak, S.; Clementi, C. Balancing energy and entropy: A minimalist model for the characterization of protein folding landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10141−10146.

(12) Saunders, M. G.; Voth, G. A. Coarse-Graining Methods for Computational Biology. *Annu. Rev. Bioph. Biom.* **2013**, *42*, 73−93.

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

(14) Ingólfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. The power of coarse graining in biomolecular simulations. *WIREs Comput. Mol. Sci.* **2014**, *4*, 225−248.

(15) 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.

(16) Pak, A. J.; Voth, G. A. Advances in coarse-grained modeling of macromolecular complexes. *Curr. Opin. Struc. Biol.* **2018**, *52*, 119.

(17) Chen, J.; Chen, J.; Pinamonti, G.; Clementi, C. Learning effective molecular models from experimental observables. *J. Chem. Theory Comput.* **2018**, *14*, 3849−3858.

(18) Singh, N.; Li, W. Recent Advances in Coarse-Grained Models for Biomolecules and Their Applications. *Int. J. Mol. Sci.* **2019**, *20*, 3774.

(19) Nüske, F.; Boninsegna, L.; Clementi, C. Coarse-graining molecular systems by spectral matching. *J. Chem. Phys.* **2019**, *151*, 044116.

(20) 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 central science* **2019**, *5*, 755−767.

(21) Wang, J.; Chmiela, S.; Müller, K.-R.; Noé, F.; Clementi, C. Ensemble learning of coarse-grained molecular dynamics force fields with a kernel approach. *J. Chem. Phys.* **2020**, *152*, 194106.

(22) 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.

(23) Jin, J.; Pak, A. J.; Durumeric, A. E.; Loose, T. D.; Voth, G. A. Bottom-up Coarse-Graining: Principles and Perspectives. *J. Chem. Theory Comput.* **2022**, *18*, 5759−5791.

(24) Boninsegna, L.; Banisch, R.; Clementi, C. A data-driven perspective on the hierarchical assembly of molecular structures. *J. Chem. Theory Comput.* **2018**, *14*, 453−460.

(25) Wang, W.; Gómez-Bombarelli, R. Coarse-graining auto-encoders for molecular dynamics. *npj Comput. Mater.* **2019**, *5*, 125.

(26) Wagner, J. W.; Dama, J. F.; Durumeric, A. E.; Voth, G. A. On the representability problem and the physical meaning of coarse-grained models. *J. Chem. Phys.* **2016**, *145*, 044108.

(27) Dunn, N. J.; Foley, T. T.; Noid, W. G. Van der Waals perspective on coarse-graining: Progress toward solving representability and transferability problems. *Acc. Chem. Res.* **2016**, *49*, 2832−2840.

(28) Jin, J.; Pak, A. J.; Voth, G. A. Understanding missing entropy in coarse-grained systems: Addressing issues of representability and transferability. *J. Phys. Chem. Lett.* **2019**, *10*, 4549−4557.

(29) Dannenhoffer-Lafage, T.; Wagner, J. W.; Durumeric, A. E.; Voth, G. A. Compatible observable decompositions for coarse-grained representations of real molecular systems. *J. Chem. Phys.* **2019**, *151*, 134115.

(30) Lebold, K. M.; Noid, W. G. Dual-potential approach for coarse-grained implicit solvent models with accurate, internally consistent energetics and predictive transferability. *J. Chem. Phys.* **2019**, *151*, 164113.

(31) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24*, 1624−1636.

(32) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109*, 2469−2473.

(33) 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.

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

(35) Lyubartsev, A. P.; Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E* **1995**, *52*, 3730−3737.
(36) Pak, A. J.; Dannenhoffer-Lafage, T.; Madsen, J. J.; Voth, G. A. Systematic Coarse-Grained Lipid Force Fields with Semiexplicit Solvation via Virtual Sites. *J. Chem. Theory Comput.* **2019**, *15*, 2087−2100.
(37) Tabak, E. G.; Vanden-Eijnden, E.; et al. Density estimation by dual ascent of the log-likelihood. *Communications in Mathematical Sciences* **2010**, *8*, 217−233.
(38) Rezende, D.; Mohamed, S.Variational inference with normalizing flows. *ICML '15: Proceedings of the 32nd International Conference on Machine Learning*, Jul. 6−11, 2015, Lille, France; JMLR, 2015; Vol. *37*, pp 1530−1538.
(39) Papamakarios, G.; Nalisnick, E.; Rezende, D. J.; Mohamed, S.; Lakshminarayanan, B. Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.* **2021**, *22*, 1−64.
(40) 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.
(41) Gabrié, M.; Rotskoff, G. M.; Vanden-Eijnden, E. Adaptive Monte Carlo augmented with normalizing flows. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, e2109420119.
(42) Li, S. H.; Dong, C. X.; Zhang, L.; Wang, L. Neural Canonical Transformation with Symplectic Flows. *Phys. Rev. X* **2020**, *10*, 021020.
(43) Nicoli, K. A.; Nakajima, S.; Strodthoff, N.; Samek, W.; Müller, K. R.; Kessel, P. Asymptotically unbiased estimation of physical observables with neural samplers. *Phys. Rev. E* **2020**, *101*, 023304.
(44) Liu, Q.; Xu, J.; Jiang, R.; Wong, W. H. Density estimation using deep generative neural networks. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, e2101344118.
(45) Ding, X.; Zhang, B. DeepBAR: A Fast and Exact Method for Binding Free Energy Computation. *J. Phys. Chem. Lett.* **2021**, *12*, 2509−2515.
(46) Wirnsberger, P.; Ballard, A.; Papamakarios, G.; Abercrombie, S.; Racanière, S.; Pritzel, A.; Blundell, C.; et al. Targeted free energy estimation via learned mappings. *J. Chem. Phys.* **2020**, *153*, 144112−144112.
(47) Wang, W.; Xu, M.; Cai, C.; Miller, B. K.; Smidt, T. E.; Wang, Y.; Tang, J.; Gómez-Bombarelli, R.Generative Coarse-Graining of Molecular Conformations. *ICML '22: 39th International Conference on Machine Learning*, Jul. 17−23, 2022, Baltimore, MD, USA; OpenReview.net, 2022; pp 23213−23236.
(48) Stieffenhofer, M.; Wand, M.; Bereau, T. Adversarial reverse mapping of equilibrated condensed-phase molecular structures. *Mach. Learn. Sci. Technol.* **2020**, *1*, 045014.
(49) Mullinax, J. W.; Noid, W. G. Generalized Yvon-Born-Green Theory for Molecular Systems. *Phys. Rev. Lett.* **2009**, *103*, 198104.
(50) Wu, H.; Köhler, J.; Noe, F.Stochastic Normalizing Flows. *Advances in Neural Information Processing Systems 33*; Neural Information Processing Systems Foundation (NeurIPS), 2020; pp 5933−5944.
(51) Ciccotti, G.; Lelievre, T.; Vanden-Eijnden, E. Projection of diffusions on submanifolds: Application to mean force computation. *Commun. Pure Appl. Math.* **2008**, *61*, 371−408.
(52) Kalligiannaki, E.; Harmandaris, V.; Katsoulakis, M. A.; Plecháč, P. The geometry of generalized force matching and related information metrics in coarse-graining of molecular systems. *J. Chem. Phys.* **2015**, *143*, 084105.
(53) Davtyan, A.; Voth, G. A.; Andersen, H. C. Dynamic force matching: Construction of dynamic coarse-grained models with realistic short time dynamics and accurate long time dynamics. *J. Chem. Phys.* **2016**, *145*, 224107.
(54) LeCun, Y.; Chopra, S.; Hadsell, R.; Ranzato, M.; Huang, F.*Predicting Structured Data*; The MIT Press, 2007.
(55) Huang, C.-W.; Dinh, L.; Courville, A.Augmented normalizing flows: Bridging the gap between generative flows and latent variable models. *arXiv Preprint (Machine Learning)*, 2020. arXiv:2002.07101. https://arxiv.org/abs/2002.07101.
(56) Chen, J.; Lu, C.; Chenli, B.; Zhu, J.; Tian, T.Vflow: More expressive generative flows with variational data augmentation. *Proceedings of the 37th International Conference on Machine Learning*; JMLR, 2020; Vol. *119*, pp 1660−1669.
(57) Cornish, R.; Caterini, A.; Deligiannidis, G.; Doucet, A.Relaxing bijectivity constraints with continuously indexed normalising flows. *Proceedings of the 37th International Conference on Machine Learning*; JMLR, 2020; Vol. *110*, pp 2133−2143.
(58) Brofos, J.; Brubaker, M. A.; Lederman, R. R.Manifold Density Estimation via Generalized Dequantization. *ICML Workshop on Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models*, 2021.
(59) Tobias, D. J.; Brooks, C. L., III Conformational equilibrium in the alanine dipeptide in the gas phase and aqueous solution: A comparison of theoretical results. *J. Phys. Chem.* **1992**, *96*, 3864−3870.
(60) Clementi, C.; Garcıa, A. E.; Onuchic, J. N. Interplay Among Tertiary Contacts, Secondary Structure Formation and Side-chain Packing in the Protein Folding Mechanism: All-atom Representation Study of Protein L. *J. Mol. Biol.* **2003**, *326*, 933−954.
(61) 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.
(62) 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.
(63) 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.
(64) Harmandaris, V.; Kalligiannaki, E.; Katsoulakis, M.; Plecháč, P. Path-space variational inference for non-equilibrium coarse-grained systems. *J. Comput. Phys.* **2016**, *314*, 355−383.
(65) Köhler, J.; Krämer, A.; Noé, F.Smooth Normalizing Flows. *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, 2021.
(66) Kingma, D. P.; Welling, M.Auto-Encoding Variational Bayes. *2nd International Conference on Learning Representations, ICLR 2014*, Banff, AB, Canada, Apr. 14−16, 2014; Conference Track Proceedings, 2014.
(67) Nielsen, D.; Jaini, P.; Hoogeboom, E.; Winther, O.; Welling, M.Survae flows: Surjections to bridge the gap between vaes and flows. *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems*; 2020; pp 12685−12696.
(68) 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.
(69) Kovács, D. P.; Oord, C. V. D.; Kucera, J.; Allen, A. E.; Cole, D. J.; Ortner, C.; Csányi, G. Linear Atomic Cluster Expansion Force Fields for Organic Molecules: Beyond RMSE. *J. Chem. Theory Comput.* **2021**, *17*, 7696−7711.
(70) 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 Preprint (Machine Learning)*, 2018. arXiv:1802.08219. https://arxiv.org/abs/1802.08219.
(71) Klicpera, J.; Groß, J.; Günnemann, S.Directional Message Passing for Molecular Graphs. *International Conference on Learning Representations*; ICLR, 2020.
(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. *arXiv Preprint (Computational Physics)*, 2021. arXiv:2101.03164. https://arxiv.org/abs/2101.03164
(73) 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.

(74) 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.
(75) 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.
(76) John, S. T.; Csányi, G. Many-Body Coarse-Grained Interactions Using Gaussian Approximation Potentials. *J. Phys. Chem. B* **2017**, *121*, 10934−10949.
(77) Schrödinger. *PyMOL Molecular Graphics System*, Ver. 1.8; 2015.

https://doi.org/10.1021/acs.jctc.3c00016
J. Chem. Theory Comput. 2023, 19, 942−952