

Statistically Optimal Force Aggregation for Coarse-Graining Molecular Dynamics

Andreas Krämer,[†] Aleksander E. P. Durumeric,[†] Nicholas E. Charron, Yaoyi Chen, Cecilia Clementi,* and Frank Noé*



Cite This: *J. Phys. Chem. Lett.* 2023, 14, 3970–3979



Read Online

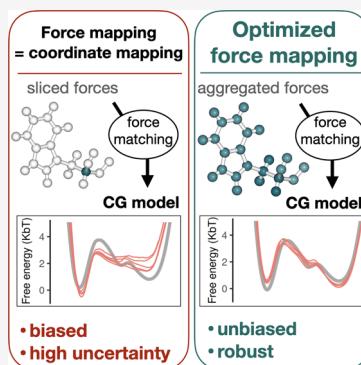
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Machine-learned coarse-grained (CG) models have the potential for simulating large molecular complexes beyond what is possible with atomistic molecular dynamics. However, training accurate CG models remains a challenge. A widely used methodology for learning bottom-up CG force fields maps forces from all-atom molecular dynamics to the CG representation and matches them with a CG force field on average. We show that there is flexibility in how to map all-atom forces to the CG representation and that the most commonly used mapping methods are statistically inefficient and potentially even incorrect in the presence of constraints in the all-atom simulation. We define an optimization statement for force mappings and demonstrate that substantially improved CG force fields can be learned from the same simulation data when using optimized force maps. The method is demonstrated on the miniproteins chignolin and tryptophan cage and published as open-source code.



Atomistic molecular dynamics (MD) simulations provide fundamental insight into physical phenomena by elucidating the behavior of individual atoms.^{1–3} While current simulations scale to millions of atoms and millisecond time scales, their application is constrained by an extremely large computational cost. One leading approach to investigate even larger systems for longer time periods is reducing the computational burden via coarse-graining, where molecular systems are simulated using fewer degrees of freedom than those associated with the atomistic positions.⁴ Particulate coarse-grained (CG) models typically define CG degrees of freedom (referred to as beads) as instantaneous averages of the positions of multiple atoms.^{5–10} Once the resolution (i.e., the definition of the CG degrees of freedom) is chosen, the central challenge is finding a force field that accurately represents the physical interactions that can be used to simulate the complex behavior of large molecular systems.

Bottom-up coarse-graining focuses on CG force fields that systematically approximate the CG behavior implied by a reference atomistic force field^{7,9,10} and has been recently used to parametrize machine-learned CG force fields based on deep neural networks.^{11–22} However, these applications require large amounts of MD data from the reference atomistic force field and, in the case of proteins, have often not quantitatively reproduced free energy surfaces of high-dimensional reference systems.^{11,13,15,16,19} These inaccuracies are often attributed to limited data, because the functional forms underpinning the force field are highly flexible.

There are multiple approaches to parametrizing bottom-up CG force fields.^{7,9,10,23} Unfortunately, many^{24–33} of these

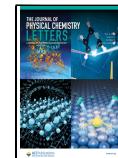
approaches require the repeated converged simulation of candidate CG force fields, creating a significant computational barrier to their application in complex systems.^{21,34} A leading approach circumventing repeated simulation is multiscale coarse-graining (i.e., “variational force matching”), where CG potentials are parametrized to directly approximate the effective mean force of an atomistic force field projected to the CG resolution.^{35–37} Noid et al.³⁶ showed that minimizing the mean-squared deviation between a CG candidate force field and suitably mapped atomistic forces yields the many body potential of mean force (PMF) and, in doing so, reproduces the reference configurational distribution at the appropriate resolution.

Numerous aspects of the coarse-graining procedure have been studied in depth; we refer readers to recent reviews for a comprehensive overview.^{9,10,23} For example, previous work has extensively studied the influence of the atom-to-bead mapping,^{14,31,38–47} functional form of the candidate potential,^{12,13,15,16,48–52} and other details of the fitting routine.^{20,33,34,53–57} However, to our knowledge, no work has directly and systematically investigated the influence of the mapping that projects fine-grained (FG) forces to the CG resolution. When the theoretical optimization statement

Received: February 16, 2023

Accepted: April 10, 2023

Published: April 20, 2023



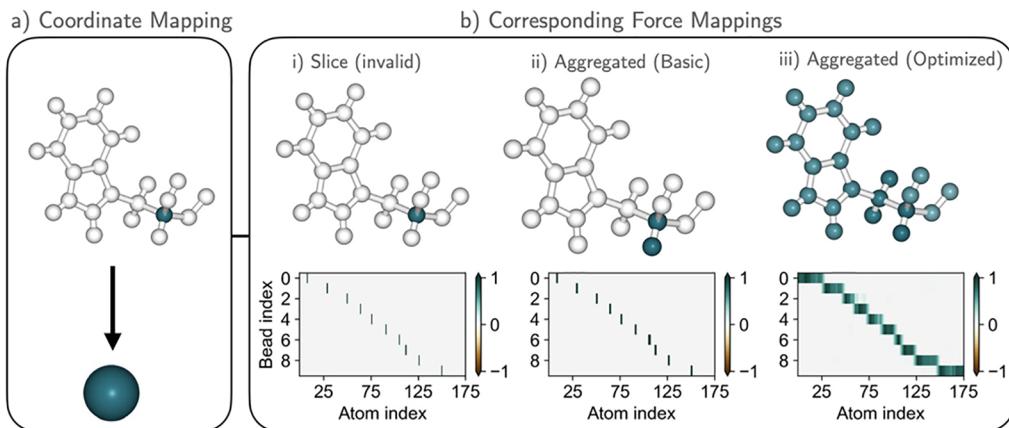


Figure 1. Different force mappings for the same “slice” coordinate mapping visualized using the TRP9 residue of the chignolin miniprotein. Non-zero mapping contributions are color-coded. The coordinate slice mapping preserves only C_{α} and is shown in panel a. Multiple possible force maps are presented in panel b: (i) applying the same slice mapping to forces is invalid as a result of a rigid bond between C_{α} and connected hydrogen and leads to an inaccurate CG force field; (ii) the “basic” aggregated force mapping, in which the forces of all holonomically constrained atoms contribute with equal weight to the mapped force; and (iii) the statistically optimal force mapping, in which all atoms can contribute with weights that are optimized to reduce the statistical uncertainty of the CG force.

defining force matching in the infinite sample limit is considered, this force mapping only affects a seemingly inconsequential constant offset to the variational statement determining the optimal force field.^{10,36} However, when learning force fields in practice, phase space averages are replaced by statistics calculated from MD trajectories to create tractable sample-based variational statements. When the force field being parametrized is not highly flexible, the distinction between phase space averages and trajectory statistics is often not important. In contrast, when using highly flexible modern machine-learned force field representations (e.g., neural networks), this distinction is critical. Parameterizing a machine-learned force field on a finite trajectory may lead to overfitting, in which a force field with optimal performance on a training trajectory may perform poorly on new configurations.^{13,21,58} More flexible potentials usually require more data for their optimization; with a fixed reference trajectory, this imposes an effective upper bound on the complexity of feasible force fields, limiting the application of flexible functional forms.

While difficulties with finite reference data are similarly exhibited with atomistic machine-learned force field development, the training data used when force matching at the CG resolution contain less information than their atomistic counterparts: energies are not available, and forces are noisy.²¹ The noise present in the forces may be an order of magnitude greater than the signal and can be viewed as a major factor in the high data requirements of machine-learned CG force fields.

The present work shows that designing the force mapping to reduce this noise improves trained CG force fields considerably. We leverage Ciccotti et al.,⁵⁹ showing that the mean force can be obtained via multiple distinct force mappings as long as they obey consistency requirements related to the configuration mapping and molecular constraints in the reference system (Figure 1). We formulate a variational statement that minimizes the noise of the mapped forces, significantly improving the signal-to-noise ratio of the force matching training objective. We also show that both high noise and constraint-inconsistent force mappings significantly degrade learned CG force fields. While these results apply to

all force-matched CG models, they are especially important for neural network CG potentials, which are sensitive to noise.⁶⁰ An open-source implementation of the proposed force mapping optimization is provided at <https://github.com/noegroup/aggforce>.

Force Matching with Constraints. Consider an atomistic system with atom positions $\mathbf{r} \in \mathbb{R}^{3n}$ and a potential energy function $V(\mathbf{r})$ in the canonical ensemble at temperature T . Atomistic holonomic constraints (e.g., rigid bond lengths) are incorporated as a system of equations, $\sigma(\mathbf{r}) = 0$.

We consider a linear mapping operator $\mathcal{M}: \mathbb{R}^{3n} \rightarrow \mathbb{R}^{3N}$, $\mathbf{r} \rightarrow \mathbf{R}$ that maps from fine to coarse configurational degrees of freedom. Under mild assumptions,^{36,59} this mapping induces the many body PMF $W: \mathbb{R}^{3N} \rightarrow \mathbb{R}$ through the principle of thermodynamic consistency

$$e^{-\beta W(\mathbf{R})} \propto \int e^{-\beta V(\mathbf{r})} \delta(\mathbf{R} - \mathcal{M}(\mathbf{r})) \delta(\sigma(\mathbf{r})) d\mathbf{r} \quad (1)$$

where $\beta = (k_B T)^{-1}$ and k_B is the Boltzmann constant. The integral in eq 1 represents a Boltzmann-weighted average over all FG configurations that correspond to a given CG configuration and obey the constraints. Computing this integral over FG states directly is not feasible for most systems of practical interest. Instead, W can be approximated by optimizing over candidate potentials $U(\mathbf{R}; \theta)$ with tunable parameters θ using variational principles, such as relative entropy minimization²⁸ or force matching.³⁶

In force matching, FG positions \mathbf{r} and forces $\mathbf{f} = -\nabla V(\mathbf{r})$ are recorded from an equilibrium simulation and mapped to the CG space to yield a training data set of instantaneous force-coordinate pairs $\{(\mathbf{R}, \mathcal{F})\}$. The optimization statement underlying force matching is found by minimizing the mean-squared deviation between model and training forces

$$\mathcal{L}_{FM}(\theta) = \langle I_{FM}(\mathbf{r}; \theta) \rangle_r = \langle \| -\nabla_{\mathbf{R}} U(\mathbf{R}; \theta) - \mathcal{F} \|_2^2 \rangle_r \quad (2)$$

where $\langle x \rangle_r := \int x p(\mathbf{r}) \delta(\sigma(\mathbf{r})) d\mathbf{r}$ denotes the thermodynamic average over the FG equilibrium distribution $p(\mathbf{r}) \propto e^{-\beta V(\mathbf{r})}$. As previously noted, in practice, force fields are produced by minimizing a sample-based approximation to eq 2 using

$\{(\mathbf{R}, \mathcal{F})\}$, possibly with regularization.^{13,61,62} Analogous to the configurational map \mathcal{M} , we need to define a force map that projects atomistic forces to the CG space in such a way that the mapped forces \mathcal{F} are an unbiased estimator of the mean force

$$\langle \mathcal{F} \rangle_{r|R} = -\nabla W(\mathbf{R}) \quad (3)$$

where we use the notation $\langle x \rangle_{r|R} := \langle x \delta(\mathbf{R} - \mathcal{M}(\mathbf{r})) \rangle_r / \langle \delta(\mathbf{R} - \mathcal{M}(\mathbf{r})) \rangle_r$ for conditional averages.

Defining Valid Force Mapping Operators. Ciccotti et al.⁵⁹ found the relation between the CG mean force, $-\nabla W(\mathbf{R})$, and the atomistic forces, $-\nabla V(\mathbf{r})$, by differentiating through the analytical expression of the many body PMF in eq 1. They showed that the (negative) mean force may be expressed as

$$\nabla W(\mathbf{R}) = \langle \underline{\mathcal{B}(\mathbf{r}) \nabla V(\mathbf{r})} - k_B T \operatorname{div} \mathcal{B}(\mathbf{r}) \rangle_{r|R} = -\mathcal{F}(\mathbf{r}) \quad (4)$$

where $\operatorname{div} \mathcal{B}(\mathbf{r}) = (\operatorname{div} B_1, \operatorname{div} B_2, \dots)^T$ denotes the divergence per CG coordinate. The (local) mapping $\mathcal{B}(\mathbf{r}) \in \mathbb{R}^{3N \times 3n}$ is a valid force projection if it obeys the following relations:

(i) orthogonality to the constraints

$$\mathcal{B}(\mathbf{r}) \nabla \sigma(\mathbf{r})^T = 0 \quad (5)$$

(ii) compatibility with the configurational mapping

$$\mathcal{B}(\mathbf{r}) \nabla \mathcal{M}(\mathbf{r})^T = \mathbf{I} \quad (6)$$

Condition i ensures that the mapped forces do not act against any atomistic constraints. This is important because rigid constraints do not transmit force information. Thus, the mapping operator \mathcal{B} must remove spurious (off-manifold) contributions to the force to not pollute the mean force computation. Condition ii ensures that the force mapping is consistent with the many body PMF induced by the configurational map.

Importantly, eqs 5 and 6 define a system of equations for each atomistic configuration \mathbf{r} that is usually highly underdetermined. This means that the force mapping operator is generally ambiguous for a fixed configurational mapping. It can even vary as a function of the FG coordinates \mathbf{r} . Previous work has not made full use of this flexibility. Instead, a common choice to meet condition ii is to define the force map as the pseudoinverse,^{14,15,18} i.e., $\mathcal{B} = (\nabla \mathcal{M} \nabla \mathcal{M}^T)^{-1} \nabla \mathcal{M}$. Alternatively, Noid et al.³⁶ defined a set of conditions to satisfy both i and ii in the case of specialized configurational and force mapping operators. They demand that all atoms that are involved in a constraint must contribute with the same force mapping coefficient. Furthermore, atoms must be configurationally uniquely associated with a single bead to have force contributions to that bead. These conditions restrict the design of \mathcal{B} considerably and do not have a solution for some configurational maps when molecular constraints are present (e.g., the slice mappings considered in this article).

To give an example of the actual flexibility of the force mapping operator, consider the setup underlying most of our computational experiments. Reference FG simulations are run with constrained covalent hydrogen bonds as is typical for biomolecular simulations.⁶³ For the configurational mappings, we use slice mappings, where bead positions are identical to the positions of selected individual heavy atoms (Figure 1a).

Under the additional conditions that \mathcal{B} does not change as a function of configuration and contributions are the same along each spatial component, conditions i and ii are satisfied by

$$\mathcal{B}_{Ii} = \begin{cases} 1, & \text{for the one heavy atom } i \text{ that is identified with bead } I \\ 0, & \text{for heavy atoms } i \text{ that are identified with a different bead} \\ \mathcal{B}_{Ij}, & \text{for all hydrogens connected to heavy atom } j \\ \text{arbitrary } \in \mathbb{R}, & \text{for all other heavy atoms} \end{cases}$$

where we have used \mathcal{B}_{Ii} to denote the static contribution of atom i to CG bead I in \mathcal{B} (see the Supporting Information). The arbitrary coefficients of all heavy atoms that are not identified with or constrained to any CG bead imply considerable flexibility in choosing the force map, which we exploit for noise reduction.

Dual Variational Principle for Force Matching and Noise Reduction. As pointed out in previous work,¹³ the force residual in eq 2 can be decomposed into PMF error and noise. The PMF error represents the bias and variance as a result of the limited expressivity of the CG model and finite data, while the noise is associated with the dimensionality reduction and represents the inherently stochastic nature of the mapped training forces from the perspective of the CG model. When machine-learned force fields are optimized with force matching, the noise contribution can dominate the force residual,^{13,21} which leads to high variance and, thus, data inefficiency and a tendency to overfit.⁶⁰ The inherent flexibility in the choice of force mapping suggests that this situation can be improved by simply switching to a different force mapping scheme. We will therefore search for force maps that both satisfy the consistency relations in eqs 5 and 6 and reduce the noise in the gradient estimator associated with the force residual in eq 2. To this end, we first derive a new dual variational principle for force matching and noise reduction. We then use this insight to propose an efficient algorithm to produce forces that make for a more robust training objective.

To formalize the optimization of the force mapping, assume that we have a family of valid force mapping operators $\mathcal{B}(\mathbf{r}; \boldsymbol{\eta})$ that are parametrized by real vector $\boldsymbol{\eta}$. This means that $\mathcal{B}(\mathbf{r}; \boldsymbol{\eta})$ satisfies conditions i and ii for all choices of $\boldsymbol{\eta}$ (see the Supporting Information). Given such a parametrization of force maps, the force matching residual in eq 2 becomes a function of both the force map and the CG potential parameters. The integrand of the residual can be decomposed into three components

$$\begin{aligned} l_{FM}(\mathbf{r}; \boldsymbol{\theta}, \boldsymbol{\eta}) &= \| -\nabla_R U(\mathbf{R}; \boldsymbol{\theta}) - \mathcal{F}(\mathbf{r}; \boldsymbol{\eta}) \|_2^2 \\ &= \text{PMF error}(\mathbf{R}; \boldsymbol{\theta}) + \text{noise}(\mathbf{r}; \boldsymbol{\eta}) + \text{mixed term}(\mathbf{r}; \boldsymbol{\theta}, \boldsymbol{\eta}) \end{aligned} \quad (7)$$

similar to that by Wang et al.¹³ While Wang et al. use these terms to denote averages, we use them here in a pointwise sense and with a parametrized force map. The mixed term is mean-free (in the limit of infinite sampling),¹³ and the mapped force \mathcal{F} is defined as in eq 4. This decomposition is discussed in detail in the Supporting Information. Here, we summarize

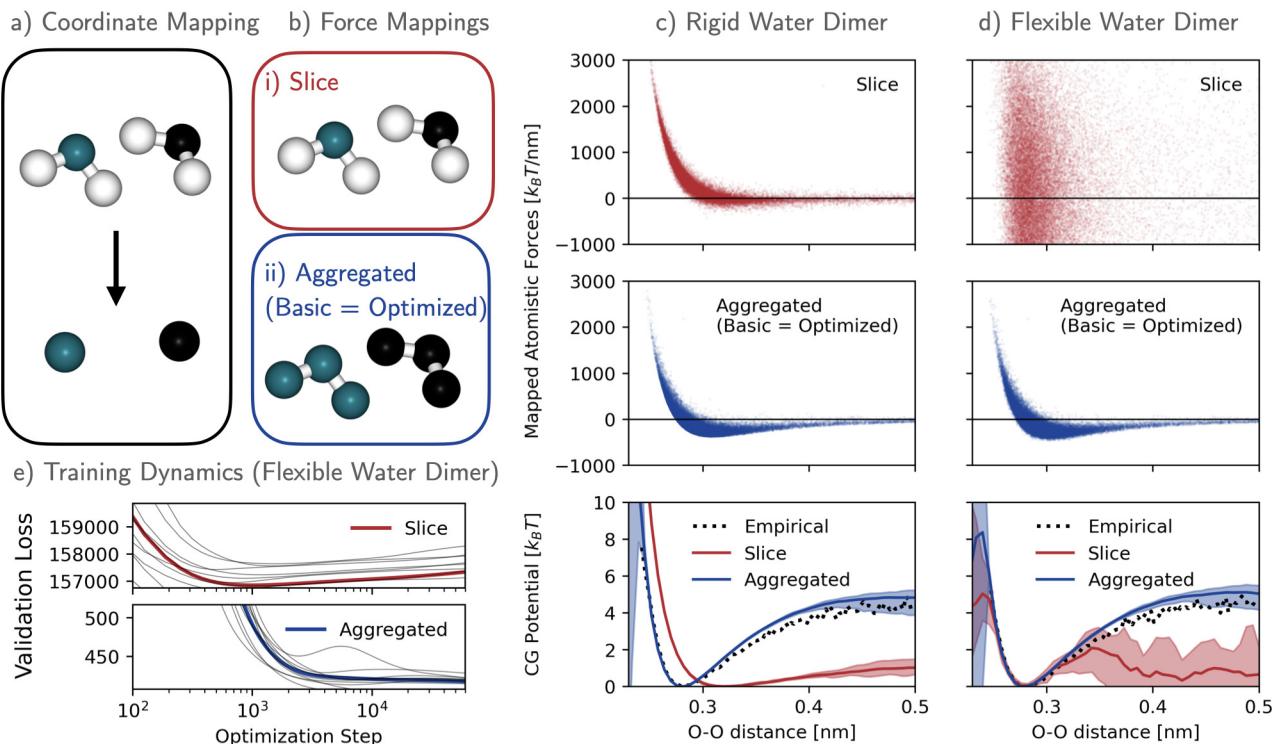


Figure 2. Coarse-graining of water dimers. (a) CG coordinates are defined by retaining only oxygens. (b) Two force mappings were investigated: a slice map and a map with equal weights for oxygens and hydrogens. Turquoise and black represent contributions to beads 1 and 2, respectively. (c) Results from constrained atomistic data: forces projected onto the oxygen–oxygen distance through a slice and aggregation mapping. Last row: CG potentials obtained from the projected force data compared to the empirical PMF. (d) Same as panel c for atomistic data without constraints. (e) Mean validation loss during training of the flexible water dimer. The shaded areas in panels c and d represent the values observed over 10 experiments. The gray lines in panel e represent individual experiments.

the most important implications. First, the PMF error does not depend upon η . Thus, for any valid force mapping scheme, minimizing the force matching loss $\langle I_{FM}(r; \theta, \eta) \rangle_r$ with respect to θ asymptotically yields the many body PMF (given a sufficiently powerful class of candidate potentials). Second, the noise term does not depend upon θ . Thus, for any guess of candidate potential, minimizing the force matching loss $\langle I_{FM}(r; \theta, \eta) \rangle_r$ with respect to η gives the same force map. A perfect, possibly nonlinear, zero-noise map would project each atomistic force exactly onto the mean force. Third, the mixed term controls the amount of noise on the parameter gradients. Improving the force map facilitates finding the CG potential and vice versa.

In summary, the symmetry of the generalized force matching loss in eq 7 reflects two orthogonal approaches to approximate the mean force. The first approach (classic force matching) tries to find the force field that best explains the atomistic forces. The second approach (noise reduction) tries to find the mapping that minimizes the variance of the mapped forces. These approaches will benefit from each other when used together. In the following section, we exploit this concept by defining force maps that facilitate efficient optimization of the candidate potential.

Computationally Efficient Optimization of Linear Force Mappings. One way to use this variational principle is the joint optimization of the force residual over θ and η . However, such an approach requires significant effort, e.g., computing the expression in eq 4 at each joint optimization step. Instead, we construct a configuration independent (“linear”) force map,

which minimizes the average magnitude of the mapped forces; i.e., we find the optimal map parameters as

$$\eta_{opt} = \arg \min_{\eta} \langle \| \mathcal{F}(r; \eta) \|_2^2 \rangle_r \quad (8)$$

Note that this optimization term has previously been used to select optimal configurational maps¹⁴ but not optimal force maps. We algebraically show in the Supporting Information that the force mapping scheme obtained in this way reduces a bound on the variance of the parameter gradient. The gradient variance is crucial because neural networks are typically trained using stochastic gradient descent-based algorithms, which iteratively follow the parameter gradient estimated on small batches of training examples.^{64,65} The gradient estimated using a single batch can be viewed as a noisy estimate of the gradient that would be obtained using all of the training samples; this noise can slow the training convergence of neural networks. Significant effort has aimed at reducing the noise generated at each update by utilizing control variates generated from previous optimization iterations.^{65–69} However, these modified optimization approaches have had limited success when applied to neural networks, likely as a result of the speed at which optimization iterations diverge from the calculated variates.⁷⁰ Equation 8 may be viewed as utilizing control variates in the force-averaging procedure to minimize gradient noise. The control variates are the linear combination of various atomistic forces. Unlike existing modifications of stochastic gradient descent, these control variates incorporate information into the training data that would be lost when

using a basic, non-optimized force mapping and result in a considerable reduction in variance.

Furthermore, solving [eq 8](#) is computationally efficient⁷¹ and allows us to optimize the mapped forces before optimizing the CG potential. Consequently, the force optimization becomes a part of the data preparation pipeline, and we can perform force matching as usual but with more robust gradients.

The choice of force mapping can significantly affect the quality of the resulting CG force field. This is first demonstrated using a low-dimensional CG potential to model a water dimer, which allows us to visualize and discuss the issues caused by atomistic constraints. We then conclude by investigating the effect on high-dimensional CG neural network potentials trained to reproduce the folding behavior of a fast-folding variant of the miniprotein chignolin (CLN025) and tryptophan (Trp) cage, systems commonly used to benchmark machine-learned CG force fields.^{13,15,16,19} For both test cases, the [Supporting Information](#) contains detailed descriptions of the simulations, CG models, and training procedures.

Water Dimers Demonstrate the Importance of Force Mappings. The water dimer system ([Figure 2](#)) contains two TIP3P molecules⁷² interacting via Coulomb and Lennard-Jones interactions in a harmonic external potential. Two data sets are created by running MD simulations with and without rigid bond and angle constraints. In both simulations, the most favorable configuration is the dimer state, with an oxygen–oxygen distance slightly below 0.3 nm, although distances of up to 3 nm are also explored.

The configurational mapping and candidate CG force field basis were fixed: bead positions were identified with oxygen positions ([Figure 2a](#)), and the CG potential was defined as a linear combination of radial basis functions on the oxygen–oxygen distance. Two aspects of the coarse-graining task were varied: the force mapping ([Figure 2b](#)) and the training data (rigid versus flexible). We first focus on the rigid system to discuss the influence of atomistic constraints.

Rigid Water: Sliced Forces Are Invalid with Bond Constraints. Most biomolecular simulations constrain the fastest moving chemical bonds to enable time steps greater than 1 fs.⁶³ MD engines enforce these constraints by modifying particle positions and velocities at each time step but do not modify the forces. As a result, the reported forces contain off-manifold contributions, such as spurious radial forces acting along a rigid bond; these artifacts do not influence the atomistic distribution or dynamics but can pollute the force matching objective when not properly taken into account. The orthogonality condition in [eq 5](#) ensures that force mappings eliminate such spurious atomistic contributions to the mapped force. The simplest way to enforce this condition is by setting $\mathbf{B}_{ii} = \mathbf{B}_{jj}$ for any pair of constrained atoms i and j , such that forces felt by atoms connected to atoms preserved in the configurational map via constrained bonds always contribute equally to the mapped force. Throughout this work, we refer to force mappings that only include force contributions from configurationally preserved and their constraint-connected atoms as basic (aggregated) force mappings (cf. [Figure 1](#)). For the water dimer with constraints, [Figure 2b](#) shows the sliced and basic aggregated force mapping schemes. Slicing in panel i of [Figure 2b](#) violates the orthogonality condition, while basic aggregation produces valid force mapping for the configurational slice mapping in [Figure 2a](#).

Using invalid force mappings can have a detrimental effect on learning CG force fields. [Figure 2c](#) shows the mapped forces (using both mapping schemes) versus the bead-to-bead distance. Both force mappings reproduce the intermolecular repulsion at small distances. However, only the basic aggregated forces capture the hydrogen-bond-driven water–water attraction. This flaw is most salient after training CG potentials and evaluating them: potentials trained using basic aggregated forces match the empirical PMF computed from a histogram of the data. In contrast, potentials trained against the sliced forces are inaccurate: they express an overly weak attraction and overestimate the equilibrium distance. This example illustrates how force mappings that violate atomistic constraints can impede convergence to the many body PMF.

Flexible Water: Aggregated Forces Drive Data-Efficient Coarse-Graining. For the water dimer without constraints, both the slice and basic aggregated force mappings are consistent with the configurational map but do not both perform equally well. [Figure 2d](#) shows forces mapped to the bead-to-bead distance. The sliced forces are dominated by the noise produced by fluctuations in the intramolecular bonds and angles. In contrast, basic aggregation annihilates these contributions completely and greatly reduces the noise in the mapped forces, which is reflected by the magnitude of force matching loss in [Figure 2e](#). Notably, solving the minimization task ([eq 8](#)) yields the basic aggregation scheme as the optimal linear force mapping (up to a 10^{-3} numerical tolerance). This shows the noise reduction mechanism at work: aggregating the force over groups of adjacent atoms removes force fluctuations coming from the “stiff” local terms of the atomistic potential.

Improving the signal-to-noise ratio of the mapped forces helps train CG potentials on finite data sets. As shown in [Figure 2d](#), CG models trained on sliced forces only reproduce the mean force in regions where data are abundant, i.e., near the equilibrium distance. In contrast, models trained on basic aggregated forces yield a high-fidelity approximation to the many body PMF that agrees well with atomistic statistics. This result supports the idea that, even when slice force mappings are valid given underlying atomistic constraints, using noise-reducing force mappings improves the data efficiency of creating CG force fields.

Optimized Forces Improve Protein Models. The proposed force mappings produce significant improvements when coarse-graining proteins using high-dimensional force fields. Chignolin and Trp cage, miniproteins consisting of 10 and 20 residues, respectively, exhibit folding behavior and serve as computationally efficient systems for investigating CG force field design. Here, we model these proteins by only preserving the positions of their C_α ([Figure 3](#)) via the approach described by Husic et al.¹⁵ using sliced forces and two modified force mapping operators.

The reference atomistic simulations utilized constrained bonds to hydrogens; as a result, the sliced force approach, which only includes the forces present on C_α , is not a valid force mapping for either protein. To investigate valid force mappings, we considered two options. First, we tested the basic aggregation force mapping: forces for each CG site were defined as summing the forces of each C_α with its connected hydrogen(s). Second, we produced an optimized force mapping by solving [eq 8](#); this is referred to as the optimized mapping ([Figure 1](#)). Note that water was not considered when creating the optimized force mapping.

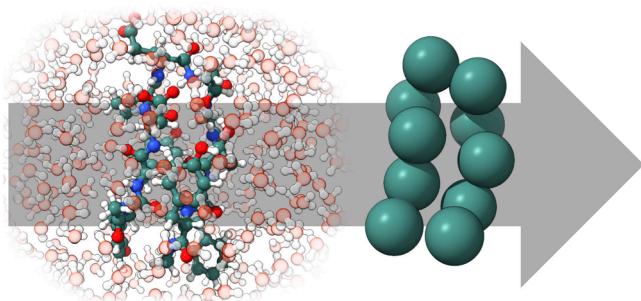


Figure 3. Visualization of the configurational CG mapping used to model chignolin. The solvated atomistic resolution used for the reference simulations is shown on the left, while the CG representation (which preserves only C_α) is shown on the right.

The resulting CG force fields were validated using MD and resulting free energy surfaces defined along slow coordinates produced via time-lagged independent component analysis (TICA)^{73–75} on the reference atomistic trajectories. These surfaces were compared to that of the reference atomistic trajectory in three ways. For all approaches, the statistics along the first two TIC components from the model and reference data were histogrammed. In the first approach, the difference in the free energy was squared and averaged across bins. For the second approach, the Jeffreys divergence (the arithmetic mean of the Kullback–Leibler divergence performed in both directions) was calculated between the two binned distributions. In the third approach, the Jensen–Shannon divergence was similarly calculated between the two binned distributions. Further details on calculating divergences may be found in the Supporting Information.

These measures of errors were calculated for models trained using various subsets of the atomistic data; these subsets were produced using two strategies. First, the effect of reduced data

set size was investigated by striding the atomistic data at a variety of values (see the Supporting Information). Second, for each stride, the atomistic data were equally partitioned into five sections, and five models were trained using different subsets of these sections in a strategy similar to cross-validation: each model was trained using a different 4/5 of the strided atomistic data. These approaches allow us to study the effect of the training set size while quantifying sensitivity to the particular data used.

The free energy surfaces of the CG models parametrized using large training sets are visualized in Figures 4 and 5, and the performance of these training procedures as a function of the training set size is visualized in Figure 6. Collectively, the sliced force models exhibit the worst accuracy; their erroneous behavior at a large sample size for Trp cage under the Jeffreys metric is due to spurious states between the folded and unfolded basins (Figure 5 and the Supporting Information). Similar artifacts are seen for large-data chignolin slice models, because the folded basin is slightly shifted (Figure 4 and the Supporting Information). The behavior in Figure 6 suggests that optimized forces increase efficiency by a factor of approximately 3 over basic forces, each of which avoid the errors produced by the sliced forces. Note that, as in the case of the water dimer, optimized forces result in significantly lower force residuals (Figures S1 and S7 of the Supporting Information). Evaluation of models trained using various force strategies on hold sets using a fixed force aggregation strategy (Table S3 of the Supporting Information) demonstrates that optimized force models result in lower force residuals; however, we note that force residuals and accurate free energy surfaces often have a complex relationship.^{76–78}

Similar to the case of the rigid water dimer, these results strongly suggest that training using invalid slice force mappings introduces artifacts. These errors appear to be resolved using force maps that satisfy the requirements outlined above.

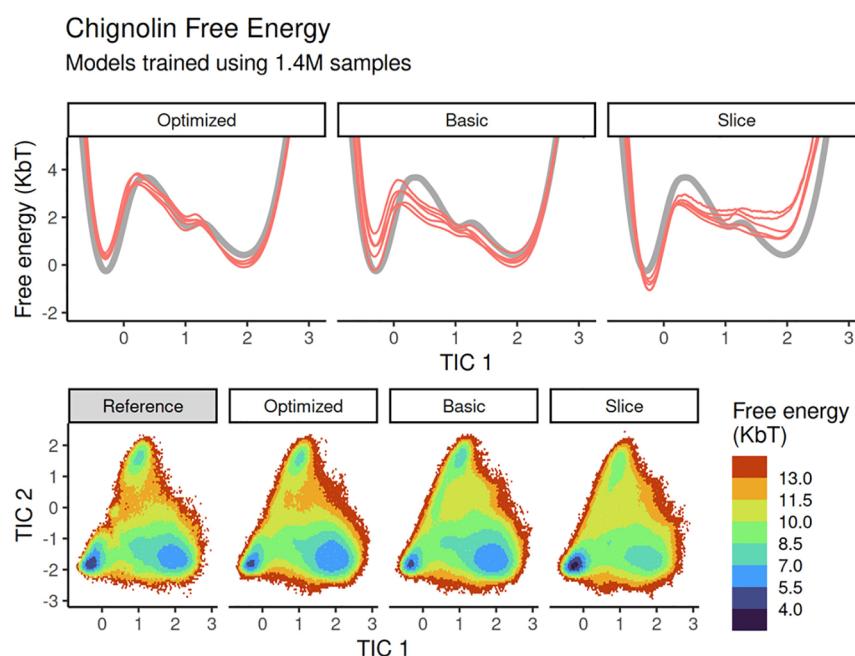


Figure 4. Free energy surfaces calculated for chignolin. The top row compares surfaces along the slowest TIC. Colored lines represent multiple force fields, each trained using a different subset of the reference trajectory. Gray lines indicate the free energy of the reference trajectory. The bottom row contains surfaces calculated for chignolin across the two slowest TICs using a single shared subset of the data. Each panel contains data generated using a different force mapping or the reference data for comparison.

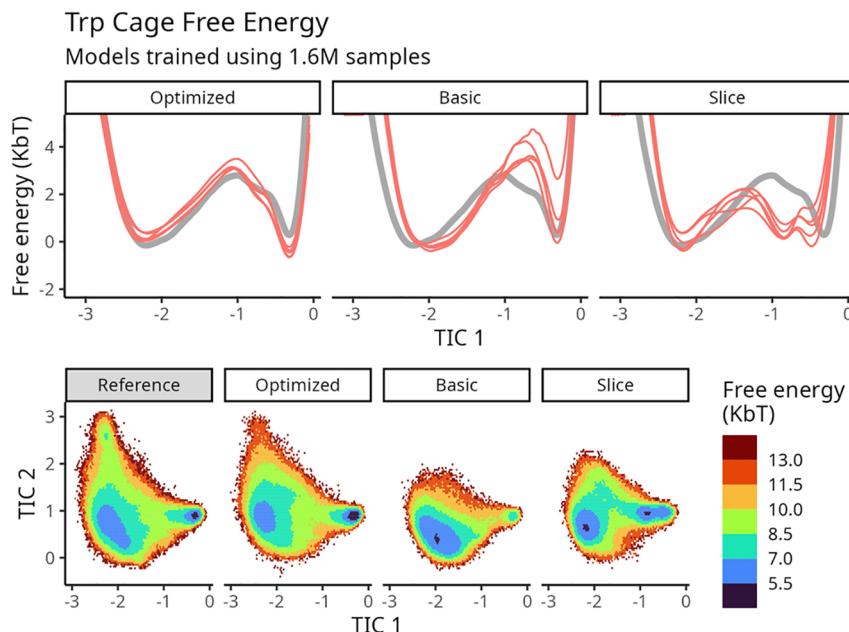


Figure 5. Free energy surfaces calculated for Trp cage. The top row compares surfaces along the slowest TIC. Colored lines represent multiple force fields, each trained using a different subset of the reference trajectory. Gray lines indicate the free energy of the reference trajectory. The bottom row contains surfaces calculated for Trp cage across the two slowest TICs using a single shared subset of the data. Each panel contains data generated using a different force mapping or the reference data for comparison.

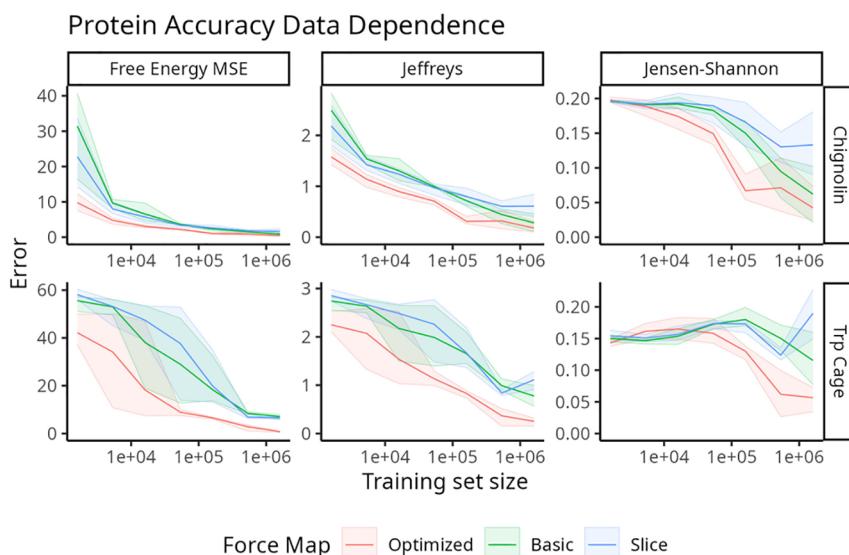


Figure 6. TIC1–TIC2 free energy error versus training size. Each column specifies an error measure: the mean squared error (MSE) of the free energies, the Jeffreys divergence, or the Jensen–Shannon divergence; each row specifies a protein; and each color represents a force mapping. Each force mapping and training size was investigated by training five models on subsets of the reference data (see the main text); the mean of the error is plotted as a line, while the maximum and minimum of errors correspond to the bounds of the ribbon.

Furthermore, using optimized forces results in a substantial improvement over basic aggregated forces, motivating their application also to systems that do not contain holonomic constraints. While large amounts of training data diminish the advantage of using optimized forces, there does not appear to be a downside to their application in all situations. Collectively, our results suggest that optimized forces result in less overfitting and lower model variance with regard to both the force residual and free energy surface.

It is important to note that, while the expressions in this paper apply to configurational maps that average positions (e.g., a center of mass mapping encompassing each amino

acid), these aggregated configurational maps may be less likely to exhibit the problems demonstrated for sliced configurational mappings. This is because the force mappings derived from such aggregation mappings using previously established rules³⁶ may satisfy i and ii in eqs 5 and 6 for typical constraints and incorporate a diverse set of atomistic forces. However, whether such mappings are appropriate for the application depends upon other aspects of force field preparation, such as the imposition of functional forms on bonded force field contributions. Similarly, we note that future comparisons between force matching results using different configurational mappings should be cognizant of the force mapping used and

that such force mappings should be reported to facilitate reproduction.

Conclusion. As machine-learned force fields become increasingly powerful, the present work paves the way for their more efficient optimization. We demonstrate that the selection of force mapping may significantly affect the resulting force field. The proposed optimized force mapping schemes reduce overfitting and increase accuracy, robustness, and data efficiency. The possibility to partly decouple force mapping coefficients from the configurational map may also elevate approaches to optimize configurational mappings alongside the CG potential.¹⁴ Future work may further exploit the presented variational principle using position-dependent force mappings and joint optimization of the force map and CG potential.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcllett.3c00444>.

Experimental details of simulations, coarse-grained models, and training procedure ([PDF](#))

■ AUTHOR INFORMATION

Corresponding Authors

Cecilia Clementi — Department of Physics, Freie Universität Berlin, 14195 Berlin, Germany; Center for Theoretical Biological Physics, Rice University, Houston, Texas 77251, United States; Department of Chemistry, Rice University, Houston, Texas 77005, United States; Department of Physics and Astronomy, 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

Andreas Krämer — Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany; orcid.org/0000-0002-7699-3083

Aleksander E. P. Durumeric — Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany

Nicholas E. Charron — Department of Physics and Astronomy, Rice University, Houston, Texas 77005, United States; Center for Theoretical Biological Physics, Rice University, Houston, Texas 77251, United States; Department of Physics, Freie Universität Berlin, 14195 Berlin, Germany

Yaoyi Chen — Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany; International Max Planck Research School for Biology and Computation (IMPRS-BAC), Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany; orcid.org/0000-0002-9299-2005

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jpcllett.3c00444>

Author Contributions

†Andreas Krämer and Aleksander E. P. Durumeric contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Clark Templeton, Félix Musil, Andrea Guljas, Iryna Zaporozhets, Atharva Kelkar, Klara Bonneau, David Rosenberger, and Brooke Husic for helpful discussions, additional experiments, and contributions to the code framework. The authors gratefully acknowledge funding from the European Commission (Grant ERC CoG 772230 “ScaleCell”), the International Max Planck Research School for Biology and Computation (IMPRS-BAC), the Federal Ministry of Education and Research (BMBF) [Berlin Institute for the Foundations of Learning and Data (BIFOLD)], the Berlin Mathematics Center MATH+ (AA1-6 and EF1-2), and the Deutsche Forschungsgemeinschaft (DFG, GRK DAEDALUS, SFB1114/A04 and B08). Cecilia Clementi 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).

■ REFERENCES

- (1) Hollingsworth, S. A.; Dror, R. O. Molecular dynamics simulation for all. *Neuron* **2018**, *99*, 1129–1143.
- (2) Bottaro, S.; Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **2018**, *361*, 355–360.
- (3) Gartner, T. E., III; Jayaraman, A. Modeling and simulations of polymers: A roadmap. *Macromolecules* **2019**, *52*, 755–786.
- (4) Clementi, C. Coarse-grained models of protein folding: Toy models or predictive tools? *Curr. Opin. Struct. Biol.* **2008**, *18*, 10–15.
- (5) Baschnagel, J.; Binder, K.; Doruker, P.; Gusev, A. A.; Hahn, O.; Kremer, K.; Mattice, W. L.; Müller-Plathe, F.; Murat, M.; Paul, W.; Santos, S.; Suter, U. W.; Tries, V. Bridging the gap between atomistic and coarse-grained models of polymers: Status and perspectives. In *Viscoelasticity, Atomistic Models, Statistical Chemistry*; Abe, A., Albertsson, A.-C., Cantow, H.-J., Dušek, K., Edwards, S., Höcker, H., Joanny, J.-F., Kausch, H.-H., Kobayashi, T., Lee, K.-S., McGrath, J. E., Monnerie, L., Stupp, S. I., Suter, U. W., Thomas, E. L., Wegner, G., Young, R. J., Eds.; Springer: Berlin, Germany, 2000; Advances in Polymer Science, Vol. 152, pp 41–156, DOI: [10.1007/3-540-46778-5_2](https://doi.org/10.1007/3-540-46778-5_2).
- (6) Klein, M. L.; Shinoda, W. Large-scale molecular dynamics simulations of self-assembling systems. *Science* **2008**, *321*, 798–800.
- (7) Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, 090901.
- (8) Pak, A. J.; Voth, G. A. Advances in coarse-grained modeling of macromolecular complexes. *Curr. Opin. Struct. Biol.* **2018**, *52*, 119–126.
- (9) Dhamankar, S.; Webb, M. A. Chemically specific coarse-graining of polymers: Methods and prospects. *J. Polym. Sci.* **2021**, *59*, 2613–2643.
- (10) 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.
- (11) 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.

- (12) 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.
- (13) Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N. E.; de Fabritiis, G.; Noé, F.; Clementi, C. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Cent. Sci.* **2019**, *5*, 755–767.
- (14) Wang, W.; Gómez-Bombarelli, R. Coarse-graining auto-encoders for molecular dynamics. *npj Comput. Mater.* **2019**, *5*, 125.
- (15) 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.
- (16) Wang, J.; Charron, N.; Husic, B.; Olsson, S.; Noé, F.; Clementi, C. Multi-body effects in a coarse-grained protein force field. *J. Chem. Phys.* **2021**, *154*, 164113.
- (17) 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.
- (18) Chennakesavalu, S.; Toomer, D. J.; Rotskoff, G. M. Ensuring Thermodynamic Consistency with Invertible Coarse-graining. *J. Chem. Phys.* **2023**, *158*, 124126.
- (19) Majewski, M.; Pérez, A.; Thölke, P.; Doerr, S.; Charron, N. E.; Giorgino, T.; Husic, B. E.; Clementi, C.; Noé, F.; De Fabritiis, G. Machine Learning Coarse-Grained Potentials of Protein Thermodynamics. *arXiv.org, e-Print Arch., Quant. Biol.* **2022**, arXiv:2212.07492.
- (20) Ding, X.; Zhang, B. Contrastive Learning of Coarse-Grained Force Fields. *J. Chem. Theory Comput.* **2022**, *18*, 6334–6344.
- (21) Durumeric, A. E.; Charron, N. E.; Templeton, C.; Musil, F.; Bonneau, K.; Pasos-Trejo, A. S.; Chen, Y.; Kelkar, A.; Noé, F.; Clementi, C. Machine learned coarse-grained protein force-fields: Are we there yet? *Curr. Opin. Struct. Biol.* **2023**, *79*, 102533.
- (22) Yao, S.; Van, R.; Pan, X.; Park, J. H.; Mao, Y.; Pu, J.; Mei, Y.; Shao, Y. Machine learning based implicit solvent model for aqueous-solution alanine dipeptide molecular dynamics simulations. *RSC Adv.* **2023**, *13*, 4565–4577.
- (23) Joshi, S. Y.; Deshmukh, S. A. A review of advancements in coarse-grained molecular dynamics simulations. *Mol. Simul.* **2021**, *47*, 786–803.
- (24) Schommers, W. A pair potential for liquid rubidium from the pair correlation function. *Phys. Lett. A* **1973**, *43*, 157–158.
- (25) 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.
- (26) Müller-Plathe, F. Coarse-graining in polymer simulation: From the atomistic to the mesoscopic scale and back. *ChemPhysChem* **2002**, *3*, 754–769.
- (27) Tóth, G. Interactions from diffraction data: Historical and comprehensive overview of simulation assisted methods. *J. Phys.: Condens. Matter* **2007**, *19*, 335220.
- (28) Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **2008**, *129*, 144108.
- (29) Cho, H. M.; Chu, J.-W. Inversion of radial distribution functions to pair forces by solving the Yvon-Born-Green equation iteratively. *J. Chem. Phys.* **2009**, *131*, 134107.
- (30) Lu, L.; Dama, J. F.; Voth, G. A. Fitting coarse-grained distribution functions through an iterative force-matching method. *J. Chem. Phys.* **2013**, *139*, 121906.
- (31) Rudzinski, J. F.; Noid, W. G. Investigation of coarse-grained mappings via an iterative generalized Yvon-Born-Green method. *J. Phys. Chem. B* **2014**, *118*, 8295–8312.
- (32) Schöberl, M.; Zabaras, N.; Koutsourelakis, P.-S. Predictive coarse-graining. *J. Comput. Phys.* **2017**, *333*, 49–77.
- (33) Thaler, S.; Zavadlav, J. Learning neural network potentials from experimental data via Differentiable Trajectory Reweighting. *Nat. Commun.* **2021**, *12*, 6884.
- (34) Thaler, S.; Stupp, M.; Zavadlav, J. Deep coarse-grained potentials via relative entropy minimization. *J. Chem. Phys.* **2022**, *157*, 244103.
- (35) Izvekov, S.; Voth, G. A. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.
- (36) 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.
- (37) Lu, L.; Voth, G. A. The Multiscale Coarse-Graining Method. *Adv. Chem. Phys.* **2012**, *149*, 47–81.
- (38) Zhang, Z.; Lu, L.; Noid, W. G.; Krishna, V.; Pfaendtner, J.; Voth, G. A. A systematic methodology for defining coarse-grained sites in large biomolecules. *Biophys. J.* **2008**, *95*, 5073–5083.
- (39) Cao, Z.; Voth, G. A. The multiscale coarse-graining method. XI. Accurate interactions based on the centers of charge of coarse-grained sites. *J. Chem. Phys.* **2015**, *143*, 243116.
- (40) Foley, T. T.; Shell, M. S.; Noid, W. G. The impact of resolution upon entropy and information in coarse-grained models. *J. Chem. Phys.* **2015**, *143*, 243104.
- (41) Madsen, J. J.; Sinit斯基, A. V.; Li, J.; Voth, G. A. Highly coarse-grained representations of transmembrane proteins. *J. Chem. Theory Comput.* **2017**, *13*, 935–944.
- (42) Diggins, P., IV; Liu, C.; Deserno, M.; Potestio, R. Optimal coarse-grained site selection in elastic network models of biomolecules. *J. Chem. Theory Comput.* **2019**, *15*, 648–664.
- (43) Webb, M. A.; Delannoy, J.-Y.; de Pablo, J. J. Graph-Based Approach to Systematic Molecular Coarse-Graining. *J. Chem. Theory Comput.* **2019**, *15*, 1199–1208.
- (44) Giulini, M.; Menichetti, R.; Shell, M. S.; Potestio, R. An Information-Theory-Based Approach for Optimal Model Reduction of Biomolecules. *J. Chem. Theory Comput.* **2020**, *16*, 6795–6813.
- (45) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grunewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tielemans, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: A general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18*, 382–388.
- (46) Kidder, K. M.; Szukalo, R. J.; Noid, W. Energetic and entropic considerations for coarse-graining. *Eur. Phys. J. B* **2021**, *94*, 153.
- (47) Yang, W.; Templeton, C.; Rosenberger, D.; Bittracher, A.; Nüske, F.; Noé, F.; Clementi, C. Slicing and Dicing: Optimal Coarse-Grained Representation to Preserve Molecular Kinetics. *ACS Cent. Sci.* **2023**, *9*, 186–196.
- (48) Larini, L.; Lu, L.; Voth, G. A. The multiscale coarse-graining method. VI. Implementation of three-body coarse-grained potentials. *J. Chem. Phys.* **2010**, *132*, 164107.
- (49) Sanyal, T.; Shell, M. S. Coarse-grained models using local-density potentials optimized with the relative entropy: Application to implicit solvation. *J. Chem. Phys.* **2016**, *145*, 034109.
- (50) John, S.; Csányi, G. Many-body coarse-grained interactions using Gaussian approximation potentials. *J. Phys. Chem. B* **2017**, *121*, 10934–10949.
- (51) Scherer, C.; Andrienko, D. Understanding three-body contributions to coarse-grained force fields. *Phys. Chem. Chem. Phys.* **2018**, *20*, 22387–22394.
- (52) DeLyser, M. R.; Noid, W. Coarse-grained models for local density gradients. *J. Chem. Phys.* **2022**, *156*, 034106.
- (53) Dama, J. F.; Sinit斯基, A. V.; McCullagh, M.; Weare, J.; Roux, B.; Dinner, A. R.; Voth, G. A. The theory of ultra-coarse-graining. I. General principles. *J. Chem. Theory Comput.* **2013**, *9*, 2466–2480.
- (54) Sharp, M. E.; Vázquez, F. X.; Wagner, J. W.; Dannenhoffer-Lafage, T.; Voth, G. A. Multiconfigurational coarse-grained molecular dynamics. *J. Chem. Theory Comput.* **2019**, *15*, 3306–3315.
- (55) Rudzinski, J. F.; Bereau, T. Coarse-grained conformational surface hopping: Methodology and transferability. *J. Chem. Phys.* **2020**, *153*, 214110.

- (56) Jin, J.; Han, Y.; Pak, A. J.; Voth, G. A. A new one-site coarse-grained model for water: Bottom-up many-body projected water (BUMPer). I. General theory and model. *J. Chem. Phys.* **2021**, *154*, 044104.
- (57) Sahrmann, P. G.; Loose, T. D.; Durumeric, A. E. P.; Voth, G. A. Utilizing Machine Learning to Greatly Expand the Range and Accuracy of Bottom-Up Coarse-Grained Models through Virtual Particles. *J. Chem. Theory Comput.* **2023**, DOI: 10.1021/acs.jctc.2c01183.
- (58) Mohri, M.; Rostamizadeh, A.; Talwalkar, A. *Foundations of Machine Learning*; MIT Press: Cambridge, MA, 2018.
- (59) Ciccotti, G.; Kapral, R.; Vanden-Eijnden, E. Blue Moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics. *ChemPhysChem* **2005**, *6*, 1809–1814.
- (60) Köhler, J.; Chen, Y.; Krämer, A.; Clementi, C.; Noé, F. Flow-Matching: Efficient Coarse-Graining of Molecular Dynamics without Forces. *J. Chem. Theory Comput.* **2023**, *19*, 942–952.
- (61) Liu, P.; Shi, Q.; Daumé, H., III; Voth, G. A. A Bayesian statistics approach to multiscale coarse graining. *J. Chem. Phys.* **2008**, *129*, 214114.
- (62) Lu, L.; Izvekov, S.; Das, A.; Andersen, H. C.; Voth, G. A. Efficient, regularized, and scalable algorithms for multiscale coarse-graining. *J. Chem. Theory Comput.* **2010**, *6*, 954–965.
- (63) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.
- (64) *Neural Networks: Tricks of the Trade*; Montavon, G., Orr, G., Müller, K.-R., Eds.; Springer: Berlin, Germany, 2012; Vol. 7700, DOI: 10.1007/978-3-642-35289-8.
- (65) Bottou, L.; Curtis, F. E.; Nocedal, J. Optimization methods for large-scale machine learning. *SIAM Rev.* **2018**, *60*, 223–311.
- (66) Johnson, R.; Zhang, T. Accelerating stochastic gradient descent using predictive variance reduction. *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013 (NIPS 2013)*; Lake Tahoe, NV, Dec 5–10, 2013.
- (67) Defazio, A.; Bach, F.; Lacoste-Julien, S. SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives *Advances in Neural Information Processing Systems 27: 28th Annual Conference on Neural Information Processing Systems 2014 (NIPS 2014)*; Montreal, Quebec, Canada, Dec 8–13, 2014.
- (68) Schmidt, M.; Le Roux, N.; Bach, F. Minimizing finite sums with the stochastic average gradient. *Math. Program.* **2017**, *162*, 83–112.
- (69) Nguyen, L. M.; Liu, J.; Scheinberg, K.; Takáč, M. SARAH: A novel method for machine learning problems using stochastic recursive gradient. *Proceedings of the 34th International Conference on Machine Learning*; Sydney, New South Wales, Australia, Aug 6–11, 2017; pp 2613–2621.
- (70) Defazio, A.; Bottou, L. On the ineffectiveness of variance reduced optimization for deep learning. *Advances in Neural Information Processing Systems 32: 33rd Annual Conference on Neural Information Processing Systems 2019 (NeurIPS 2019)*; Vancouver, British Columbia, Canada, Dec 8–14, 2019.
- (71) Stellato, B.; Banjac, G.; Goulart, P.; Bemporad, A.; Boyd, S. OSQP: An operator splitting solver for quadratic programs. *Math. Program. Comput.* **2020**, *12*, 637–672.
- (72) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (73) 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.
- (74) 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.
- (75) 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.
- (76) Stocker, S.; Gasteiger, J.; Becker, F.; Günemann, S.; Margraf, J. T. How robust are modern graph neural network potentials in long and hot molecular dynamics simulations? *Mach. Learn.: Sci. Technol.* **2022**, *3*, 045010.
- (77) Fu, X.; Wu, Z.; Wang, W.; Xie, T.; Keten, S.; Gomez-Bombarelli, R.; Jaakkola, T. Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations. *arXiv.org, e-Print Arch., Phys.* **2022**, arXiv:2210.07237.
- (78) Ricci, E.; Giannakopoulos, G.; Karkaletsis, V.; Theodorou, D. N.; Vergadou, N. Developing Machine-Learned Potentials for Coarse-Grained Molecular Simulations: Challenges and Pitfalls. *Proceedings of the 12th Hellenic Conference on Artificial Intelligence (SETN 2022)*; Corfu, Greece, Sept 7–9, 2022; pp 1–6, DOI: 10.1145/3549737.3549793.