Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Machine learned coarse-grained protein force-fields: Are we there yet?

Aleksander E. P. Durumeric[1], Nicholas E. Charron[2,3,4], Clark Templeton[3], Félix Musil[3], Klara Bonneau[3], Aldo S. Pasos-Trejo[3], Yaoyi Chen[1], Atharva Kelkar[1], Frank Noé[5,1,3,6] and Cecilia Clementi[3,4,6,2]

Abstract

The successful recent application of machine learning methods to scientific problems includes the learning of flexible and accurate atomic-level force-fields for materials and biomolecules from quantum chemical data. In parallel, the machine learning of force-fields at coarser resolutions is rapidly gaining relevance as an efficient way to represent the higher-body interactions needed in coarse-grained force-fields to compensate for the omitted degrees of freedom. Coarse-grained models are important for the study of systems at time and length scales exceeding those of atomistic simulations. However, the development of transferable coarse-grained models via machine learning still presents significant challenges. Here, we discuss recent developments in this field and current efforts to address the remaining challenges.

Addresses

[1] Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 12, 14195, Berlin, Germany
[2] Department of Physics and Astronomy, Rice University, 6100 Main Street, Houston, 77005, Texas, USA
[3] Department of Physics, Freie Universität Berlin, Arnimallee 12, 14195, Berlin, Germany
[4] Center for Theoretical Biological Physics, Rice University, 6100 Main Street, Houston, 77005, Texas, USA
[5] Microsoft Research AI4Science, Karl-Liebknecht Str. 32, Berlin, 10178, Berlin, Germany
[6] Department of Chemistry, Rice University, 6100 Main Street, Houston, 77005, Texas, USA

Corresponding author: Clementi, Cecilia (c.clementi@fu-berlin.de)
(Templeton C.),
(Musil F.),
(Pasos-Trejo A.S.),
(Chen Y.),
(Noé F.)
(Clementi C.)

Current Opinion in Structural Biology 2023, **79**:102533

This review comes from a themed issue on **Theory and Simulation/Computational Methods (2023)**

Edited by **Turkan Haliloglu** and **Gregory A. Voth**

For complete overview of the section, please refer the article collection - Theory and Simulation/Computational Methods (2023)

Available online 31 January 2023

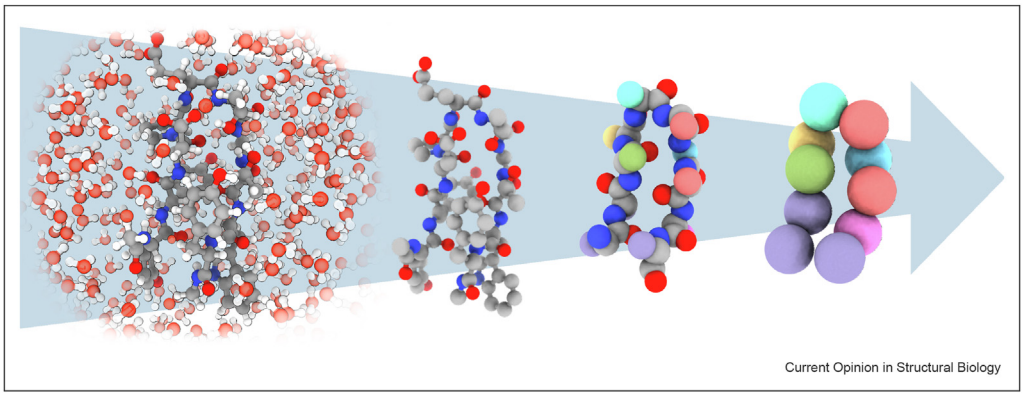
https://doi.org/10.1016/j.sbi.2023.102533

Introduction

The definition of simplified models is central to physical sciences; proteins are no exception [1,2]. Statistical mechanical approaches to describe protein folding and dynamics [3–5], as well as the analysis of long molecular dynamics (MD) trajectories [6–8], have demonstrated that slow processes in large biomolecular systems can be described by a reduced number of variables despite hundreds of thousands of atoms comprising the full system. In this spirit, many coarse-grained (CG) models have been proposed to study proteins through MD and energy minimization. These CG models have been used to investigate the principles underlying protein folding [9–12], intermolecular binding/interactions [13,14], protein-mediated membrane phenomena [15,16], and to make predictions about novel biological systems of immediate medical interest [17,18].

Despite their successes, CG models of proteins have not yet achieved the predictive performance of their atomistic counterparts. CG models are primarily designed by specifying their resolution, which defines the coarse degrees of freedom (referred to as "sites" or "beads," see Figure 1), and by their effective energy function, which dictates how these beads interact. Traditionally, the resolution is first chosen using either chemical intuition or through optimization designed to reproduce chosen properties (*e.g.*, the study by Giulini et al. [19]). The model's effective energy function is then parameterized to reproduce experimental or simulation data. The fundamental goal of the *transferable* CG models discussed in this article is to predict the conformational landscape of proteins not used for their parameterization, ideally using only the primary structure of the proteins of interest. Atomistic models have been able to explore the relevant landscape of small globular proteins [20–22]; however, it is still an open question as to whether there exists a resolution at which a chemically

Figure 1



Sequential reduction in resolution of a variant of the miniprotein Chignolin (CLN025) from a solvated all-atom representation containing many thousands of atoms, to an implicit solvent representation, to a heavy-backbone representation with C_β beads, and finally to a C_α CG representation containing 10 beads.

transferable CG model can quantitatively describe the configurational landscape of arbitrary proteins.

A transferable CG protein model would have significant consequences. By employing special-purpose supercomputers [23] or distributed simulation combined with Markov State Models [24,21,25], the dynamics of small solvated proteins can be simulated over millisecond timescales [20]. However, biological phenomena routinely involve larger complexes and span longer timescales (seconds or more). CG models promise to reach such scales by reducing the computational cost via decreasing the number of degrees of freedom and increasing the effective simulation timestep. This increased efficiency would vastly improve the use of MD for both fundamental research and applications, for example, in protein design.

There has been a surge of interest in using machine learning (ML) methods for molecular simulation [26], including learning CG models from large amounts of data. In a sense, the development of ML CG models can be seen as an extension of ongoing research on the design of accurate atomistic force-fields from quantum mechanical calculations. In this area, ML has already produced highly accurate force-fields which have facilitated groundbreaking computational studies [27–29]. When combined with the field of bottom-up CG [30,31], these approaches provide a seemingly clear strategy to leverage ML to learn a CG force-field from existing atomistic MD trajectories. Indeed, thanks to the flexibility of ML algorithms, some frameworks developed for the atomistic resolution [32,33] have been transferred to the CG resolution [34–37].

Despite these advances, a completely bottom-up transferable CG model still does not exist for proteins or other biopolymers. This limited progress is due to multiple outstanding challenges, which together firmly differentiate the creation of ML bottom-up CG force-fields from their atomistic counterparts. We here discuss these difficulties and current efforts to overcome them.

Thermodynamic consistency: Why is it difficult?

Bottom-up coarse-graining typically models the following free energy surface (U) [30,26,31] referred to as the effective CG (free) energy:

$$U(\mathbf{R}) = -\beta^{-1}\ln \int \delta[\mathbf{R} - \mathcal{M}(\mathbf{r})]\exp[-\beta u(\mathbf{r})]\mathrm{d}\mathbf{r}, \quad (1)$$

where $\mathcal{M}$ maps all-atom configurations $\mathbf{r}\in\mathbb{R}^{3n}$ to their CG counterparts $\mathbf{R} = \mathcal{M}(\mathbf{r})\in\mathbb{R}^{3N}$, u is the reference all-atom energy, and β is the inverse temperature. Intuitively, $\mathcal{M}$ defines the CG resolution and U defines how particles at this resolution interact; the design of a bottom-up CG model then entails defining $\mathcal{M}$ and approximating U, and the two tasks are interdependent. A CG energy that, up to a constant, equals U is said to be *thermodynamically consistent* with the atomistic counterpart. Such a U produces free energy landscapes identical to the reference in any reaction coordinates that are a function of the CG coordinates. We note that the phrase *thermodynamically consistent* here does not refer to thermodynamic observables (*e.g.*, pressure), but instead considers the configurational distributions of the CG and atomistic force-fields. For information on thermodynamic properties in CG models we defer to recent articles [38–40,31].

Although the thermodynamically consistent CG energy is uniquely defined up to a constant by Eq. (1), the integral cannot be solved for non-trivial systems [30]. As

a result, multiple strategies [41,30,31] to approximate U have been proposed. Traditionally, the functional forms for CG (free) energy functions have been low body-order with physically motivated terms [30,31]. However, recent studies have employed higher body-order terms parameterized by neural networks with success [42–44,36,35]. Kernel methods have also been proposed (*e.g.*, the study by John et al. [34]) but have not been applied to proteins. While exceptions exist [42,45–47], for reasons of computational efficiency existing ML CG models [44,36,26] have been primarily based on the Multiscale Coarse-Graining [48] ("force-matching") variational framework, primarily due to the fact that it does not require the CG model to be simulated during its parameterization.

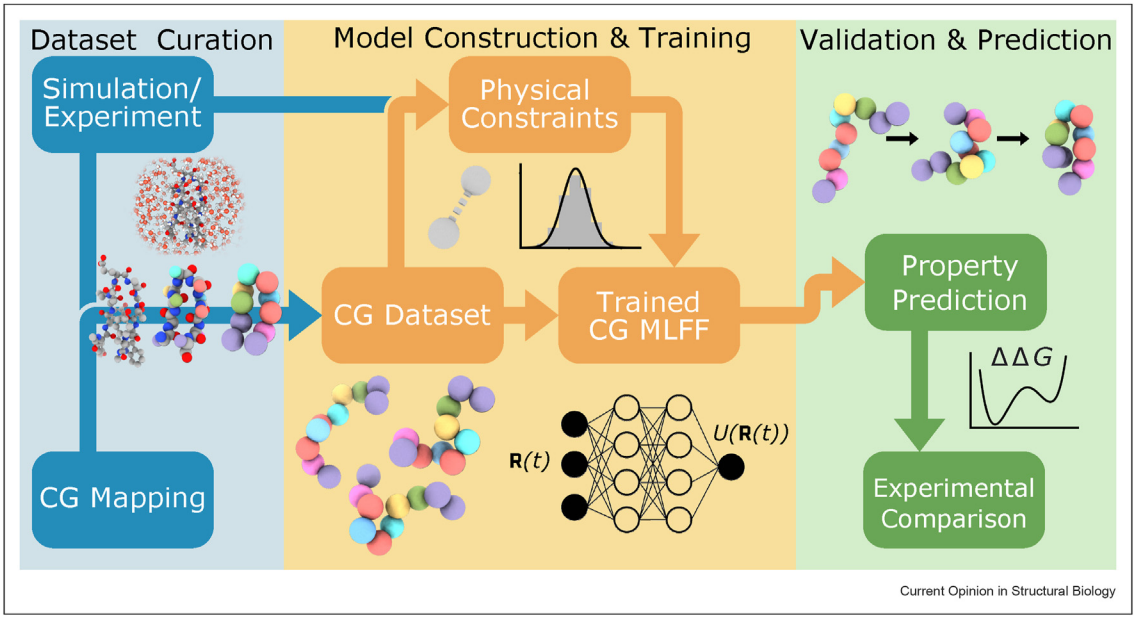
In principle, Eq. (1) suggests that once the CG resolution has been chosen the creation of the CG model should be straightforward. However, designing an accurate ML CG model is not trivial (see Figure 2). The choices of reference atomistic system (u), of the resolution ($\mathcal{M}$), and of the different terms of U all entail challenges unique to models designed at a CG resolution. These challenges are compounded by difficulties with validation, which are also present in the development of ML atomistic force-fields from quantum chemical data. We address these challenges in detail, beginning with the data used for training, continuing by discussing the design of U and subsequently the resolution of the CG model, and finishing by discussing validation and robustness. For brevity we only discuss algorithms which are applicable to ML CG force-fields; for a more comprehensive introduction to bottom-up coarse-graining we refer readers to recent perspective articles [30,31].

The difficulty in training CG force-fields

The principal challenge in bottom-up coarse-graining with machine-learned force-fields lies in finding a suitable ML formulation that directly or indirectly estimates the intractable integral described in Eq. (1). The situation is more difficult compared to learning atomistic potential energy surfaces from quantum mechanical data, where reference energies and forces are known: When learning a CG free energy, neither U nor its gradient for a given CG structure $\mathbf{R}$ are known because the integral Eq. (1) is intractable.

The most straightforward approach to Eq. (1) is to directly estimate the behavior of the probability density proportional to $\exp(-\beta U(\mathbf{R}))$ from simulation data. This requires an equilibrium sample of atomistic conformations $\mathbf{r}$, for example, obtained by MD simulations. After mapping them to the CG resolution, a ML model is then trained to approximate $U(\mathbf{R})$, by minimizing the Kullback–Leibler divergence between the CG and atomistic probability densities. This is called relative entropy minimization in the coarse-graining literature [49,50,47] and maximum likelihood estimation in the ML energy-based model community [51]. Similar

Figure 2



A pipeline for creating and using ML CG models from atomistic simulation data and experimental measurements. A chosen CG mapping (resolution) can reduce reference information into a CG dataset that can be used to train ML CG force-fields (FFs). This training can rely on both simulation and experimental observables in order to reduce the complexity of the learning task and respect physical constraints. A trained ML CG model can then be validated through CG MD and used for general property predictions.

approaches [42,45,46], which estimate and reduce the difference between a CG force-field and U or optimize selected observables, in turn expand on other approaches from the ML community (*e.g.*, the study by Gutmann et al. [52]).

The difficulty with most of these approaches is that the CG model must be periodically re-simulated during training in order to evaluate the equilibrium density generated by the current model of $U(\mathbf{R})$. While this is feasible for quickly equilibrating CG models, such as those of liquids [47], it is extremely challenging for models that exhibit rare events, such as realistic CG models of protein folding. This limitation is even more problematic for complex parameterizations of U (*e.g.*, neural networks) and significantly impedes simultaneously training over multiple molecules when creating transferable models. Approaches have attempted to reduce this burden by, for example, reweighting the density of previous iterations [50,45,53] or by modifying the sampling of the atomistic system [46]. However, these approaches have not yet reached the simplicity and applicability of approaches without iterative simulation.

The most common bottom-up approach for approximating U is force-matching [48], which fits a CG free energy such that its negative gradient matches projected instantaneous atomistic forces on average. Critically, this does not require simulations of the CG model during training, and was proposed for ML CG protein force-fields in the study by Wang et al. [44]. As many atomistic coordinates $\mathbf{r}$ map to the same CG coordinate $\mathbf{R}$, the instantaneous force is noisy, and the signal-to-noise ratio becomes smaller the more degrees of freedom are "CGed away"; thus CG force-matching requires more data as compared to atomistic force-fields. A second difficulty comes from the fact that U is obtained by implicitly integrating the mean force, and as a result, obtaining the free energy difference between minima depends on estimates of the forces along the transition path, where the uncertainties are the largest.

The recently proposed flow-matching method [54] combines relative entropy estimation and force-matching by employing generative deep learning: the CG density is estimated by a normalizing flow, a neural network that can generate one-shot samples of equilibrium CG conformations. This flow can then generate samples to train a downstream ML CG force-field by force-matching. The limitation of this approach still lies in finding flow architectures that can scale to large macromolecules.

The distribution of atomistic configurations is fundamental to the discussed algorithms. Rare events are important but infrequently sampled in the canonical distribution; directing atomistic sampling towards barriers and areas of "high uncertainty" may be beneficial. While ML models are more expressive than, for example, pair potentials, they require more data. For example, ML CG force-matching may use upwards of one million canonically distributed samples covering the configurational space for small proteins [36], in contrast to harmonic models parameterized using short trajectories in the folded state [55]. Modifying the distribution of samples may reduce data requirements [56], but it is unclear how such approaches scale with system complexity.

Concurrently, iterative methods may overcome their computational barriers if non-canonical sampling is used; expanding discriminative training may remove the need for repeated training simulations [46], and biasing potentials may promote diversity and produce more accurate parameters [57]. However, approaches that require data to be drawn from a modified distribution impede the use of preexisting atomistic trajectories. Nevertheless, these approaches will be critical to expanding current ML CG success to multi-domain proteins.

For a transferable ML CG model, more requirements for the training dataset arise. It is straightforward, and important, to simultaneously force-match a model using reference data from multiple proteins as evidence has shown that extended ensembles can act as regularization [58]. Previous pioneering work developing bottom-up transferable CG models used this approach, but fell short of unassisted folding and relied on artificially lowering simulation temperatures to stabilize states of interest [59]; we associate these inaccuracies to limitations of the force-field basis and training set. We anticipate that the proportion of structural motifs in the dataset plays an important role. In the ideal case, a general CG model of proteins would likely include globular, fibrous, and intrinsically disordered proteins in its training procedure. Such a transferable training setup naturally expands the amount of atomistic data available to train a given model; whether this will improve predictions on individual proteins remains to be seen.

Choice of the CG representation

In the design of an atomistic force-field, the Born-Oppenheimer approximation justifies the separation between electronic and nuclear degrees of freedom and provides the framework for effective nuclear potential energy surfaces. However, the separation of scales is less clear for CG models. Consequently, the selection of the CG resolution ($\mathcal{M}$ from Eq. (1)) is non-trivial and influences the free energy surface that must be learned. The fundamental questions in this area are which resolutions are "easy to learn" and which are conducive to creating transferable models. These points highlight the

challenges of validating ML CG models that are capable of extrapolating to unseen systems. For certain resolutions, it may be easier to learn an effective CG energy and extrapolate into unknown regions of phase space. On the other hand, certain resolutions may be conducive to accurate ML CG models but may be difficult to interpret.

Current successful CG ML protein applications [42,44,36,46,54] typically focus on a single site per residue (typically C_α); however, this appears to be mostly due to simplicity and not systematic validation. "Optimal" resolutions have been studied [60,61], but it is unclear how they impact ML CG models. Back-mapping, that is, reconstructing details from CG models, is a current area of investigation [62] and may alleviate interpretability constraints on the CG resolution. Views which link back-mapping with potential optimization can facilitate a joint optimization of the representation alongside the CG energy model [63]; however, these approaches do not yet in themselves search for transferable resolutions.

Functional form of the many-body effective CG (free) energy

In practice, training ML CG models via force-matching from equilibrium data requires a baseline (or "prior") potential to reduce catastrophically incorrect extrapolation into unphysical regions of phase space [44,36,64]. Ultimately, a good prior potential incorporates physical principles, reduces learning complexity, and allows for stable simulation. Similar to Δ-learning [65] for atomistic force-fields [66,67], the CG energy is usually decomposed into:

$$U(\mathbf{R};\boldsymbol{\theta}) = U_{prior}(\mathbf{R}) + U_{net}(\mathbf{R};\boldsymbol{\theta}) \tag{2}$$

where $U_{net}(\mathbf{R};\ \boldsymbol{\theta})$ is a trainable multibody potential expressed by an ML model with parameters $\boldsymbol{\theta}$ and $U_{prior}(\mathbf{R})$ is the prior energy.

Designing the priors is non-trivial as it depends on an interplay between the CG resolution, the ML architecture, and the training data. Poor choices of priors can significantly reduce the performance of an ML force-field [47,68]. Currently, the prevailing strategy involves proposing a prior inspired by the low body-order terms from classical force-fields, and then iteratively developing an ML CG model over both the prior terms and traditional hyperparameters [44,36,35,37,47]. Systematic strategies have yet to be developed to design prior energies that are transferable to different molecules.

While priors help enforce important physical asymptotic interactions, the ML model architecture itself should respect basic physical constraints. These include invariance with respect to permutations of particles of the same type, invariance to translations and rotations of the reference frame, and curl-free force predictions [69,44]. A way to allow the learnable energy U_{net} from Eq. (2) to be transferable [36] is to decompose it bead-wise such that:

$$U_{net}(\mathbf{R},\mathbf{a}) = \sum_i u_{net}(\mathbf{R}-\mathbf{R}_i, a_i), \tag{3}$$

where, $\mathbf{R}_i$ is the ith bead (with type a_i) in the configuration $\mathbf{R}$ so that $\mathbf{R}-\mathbf{R}_i$ are the relative displacements of all beads around bead $\mathbf{R}_i$, and u_{net} is the bead-wise contribution to the potential.

On top of these constraints, bottom-up coarse-graining involves additional architectural challenges. Coarse-graining a variety of different groups of atoms leads to a large number of CG bead types, for example, at least 20 types (one for each amino acid) for proteins at the C_α resolution. Furthermore, as mentioned above, training with noisy forces requires a large number of training configurations. As a result, the ML approach must accommodate large training sets (indicating that neural networks may be preferable over kernel methods) and should not scale with the number of bead types so that evaluation times do not increase when considering transferable models. This constraint favors the use of deep learning architectures like SchNet [33] over models based on fixed representation, for example, symmetry functions [70,71].

Validation and robustness

While atomistic ML force-field development has matured, there exists no appropriate set of best practices for probing stability and robustness. It is common to assess atomistic model accuracy with point wise metrics, such as the mean force error, over fixed test datasets [27]. However, without an understanding of how models extrapolate into data-poor regions, these metrics cannot be used as indicators of simulation stability or accuracy [72,64], as simulations may explore uncovered configurations. For ML CG models, even with the use of prior energy terms, force error does not guarantee a stable model [68].

Due to the difficulty in constructing comprehensive test sets, the robustness and accuracy of a trained ML model can only be ascertained through extensive sampling, for example, by using the model to run long MD simulations. Recent investigations into ML architectures have revealed the need for such metrics for both atomistic and CG ML models [64,68]. Unfortunately, obtaining a converged CG MD simulation can require several million force evaluations; for large systems and complex architectures this may present a computational bottleneck [73]. Validation difficulties impede hyperparameter optimization (*e.g.*, regularization strength, cutoff, or prior potential), as searches may become

prohibitively expensive. We note, however, that existing applications provide suitable initial choices of hyperparameters for select architectures and resolutions (see the study by Husic et al. [36]), but that the introduction of novel architectures naturally requires substantial effort for the initial hyperparameter search.

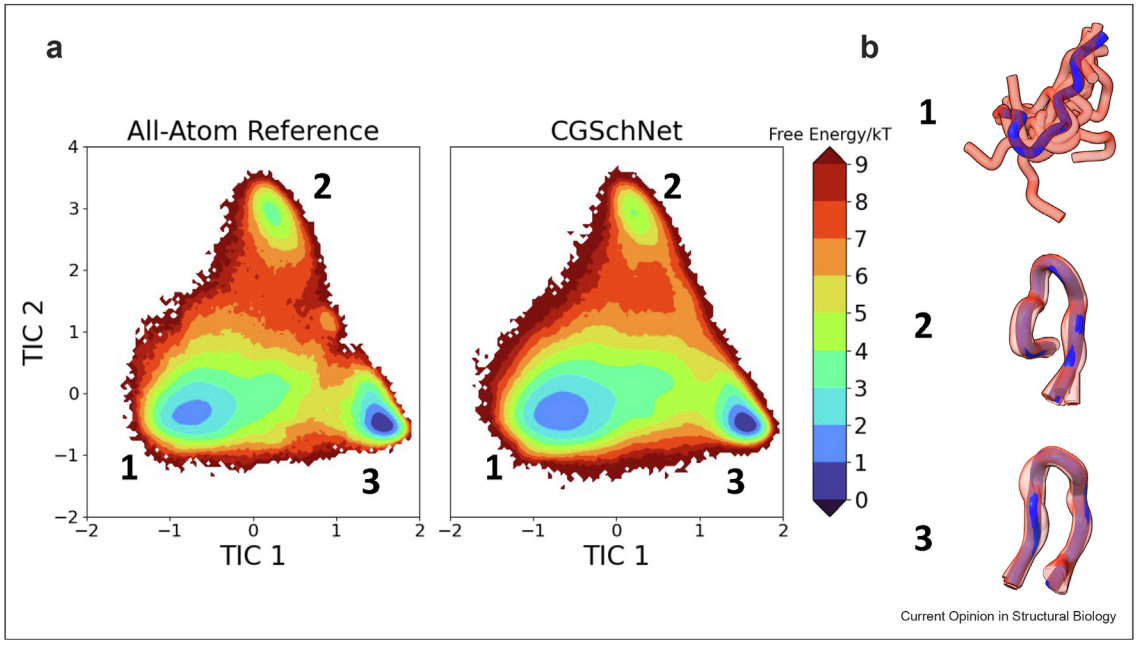
Even once MD has been used to characterize a ML CG model, validation still poses difficulties. When characterizing atomistic force-fields on selected configurations it is typically possible to compare the model's energy and force predictions to noiseless reference values; unfortunately, these are not available at the CG resolution (Eq. (1)). Instead, analysis typically projects CG configurations onto low-dimensional collective variables (*e.g.*, Figure 3). However, as ML CG models are now able to reproduce such collective variable surfaces, the need for more rigorous validation is emerging. Recent work [75] has proposed classification as an approach to generate energy-like errors for CG models and may provide an avenue for connecting atomistic and CG force-field validation.

A related challenge is presented by model uncertainty: How robust is a ML model to different training seeds or data partitioning strategies? For neural networks, these can be expensive questions to answer. However, recent advances have started to enable estimates of uncertainty [76,77]. A promising strategy involves estimating the uncertainty of predictions and minimizing it either before or during model deployment, either through iterative training or through "on-the-fly" frameworks [78] where data is added to the training set based on such estimates.

Conclusion

At the moment of writing, state-of-the-art ML CG models can quantitatively reproduce the behavior of small proteins, as shown in Figure 3 for Chignolin and in the study by Köhler et al. [54] for Trpcage, BBA, and Villin. Currently, the largest barrier to describe larger proteins is gathering sufficient training data. To what extent such an approach can be extended to define *transferable* CG models remains an open question. It may be possible only for a class of proteins, or at particular resolutions. Before the advent of ML methods, these questions remained challenging to answer, as thermodynamic consistency between an atomistic and a CG model (Eq. (1)) could only be approximately enforced; it was not clear whether the limitations of transferable models [59,79–81,39] were due to the limited expressivity of the CG energy and limited reference data or to more fundamental problems with transferability. Now, as ML CG models can quantitatively enforce thermodynamic consistency for single proteins (as shown in Figure 3), we have the tools to address these questions

Figure 3



State-of-the-art performance for a C_α CG ML model on the benchmark protein CLN025. a) Comparison of the CG free energy landscape of CLN025 (produced using MD) for a learned CG ML model with the corresponding free energy for the reference all-atom dataset projected onto slow degrees of freedom (TICA) [74]. b) Ensembles of structures sampled from the CG ML model MD simulation (in red) are superimposed onto all-atom reference structure counterparts (in blue). Basin 1 represents the unfolded state, basin 2 the misfolded state, and basin 3 the folded state.

and explore the trade-off between accuracy and transferability. Here, we have discussed the practical challenges towards this goal, but we remain optimistic that such a line of research can be pursued.

Even if a transferable bottom-up ML CG model can be defined, eventually, the success of a computational model relies on its comparison to experiments. Bottom-up CG models rely on the reference atomistic models and necessarily inherit their inaccuracies and flaws. With the improvements in atomistic force-fields, we expect CG models to also become more accurate. However, even small inconsistencies between the CG and atomistic models may compound into a significant deviation from experimental data. We believe that ultimately bottom-up ML CG will need to be merged with top-down models for their useful and predictive applications.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Molecular visualizations for the graphical abstract and all in-text figures were produced using UCSF ChimeraX, a software developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco [82]. The free energy surfaces in Figure 3 were produced using the Python package Matplotlib [83]. We acknowledge the financial support of Deutsche Forschungsgemeinschaft DFG (SFB/TRR 186, Project A12; SFB 1114, Projects B03 and A04; SFB 1078, Project C7; and RTG 2433, Project Q04 and Q05), the National Science Foundation NSF (CHE-1900374 and PHY-2019745), the NLM Training Program in Biomedical Informatics and Data Science (Grant No. 5T15LM007093-27), and the Einstein Foundation Berlin (Project 0420815101). We also thank the members of the Clementi and Noé groups for their helpful discussions.
F.M. acknowledges support from the SNSF under the Postdoc.Mobility fellowship P500PT_203124.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Levitt M, Warshel A: **Computer simulation of protein folding**. *Nature* 1975, **253**:694–698.

2. Clementi C: **Coarse-grained models of protein folding: toy models or predictive tools?** *Curr Opin Struct Biol* 2008, **18**:10–15.

3. Bryngelson JD, Wolynes PG: **Spin glasses and the statistical mechanics of protein folding**. *Proc Natl Acad Sci USA* 1987, **84**:7524–7528.

4. Onuchic JN, Luthey-Schulten Z, Wolynes PG: **Theory of Protein Folding: the energy landscape perspective**. *Annu Rev Phys Chem* 1997, **48**:545–600.

5. Dill KA, Bromberg S, Yue K, Chan HS, Ftebig KM, Yee DP, Thomas PD: **Principles of protein folding — a perspective from simple exact models**. *Protein Sci* 1995, **4**:561–602, https://doi.org/10.1002/pro.5560040401.

6. Best RB, Hummer G: **Reaction coordinates and rates from transition paths**. *Proc Natl Acad Sci USA* 2005, **102**:6732–6737.

7. Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC: **Automatic discovery of metastable states for the construction of markov models of macromolecular conformational dynamics**. *J Chem Phys* 2007, **126**, 155101.

8. Noé F, Clementi C: **Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods**. *Curr Opin Struct Biol* 2017, **43**:141–147.

9. Clementi C, Nymeyer H, Onuchic JN: **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.

10. Liwo A, Khalili M, Scheraga HA: **Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains**. *Proc Natl Acad Sci USA* 2005, **102**:2362–2367.

11. Davtyan A, Schafer NP, Zheng W, Clementi C, Wolynes PG, Papoian GA: **Awsem-md: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing**. *J Phys Chem B* 2012, **116**:8494–8503, https://doi.org/10.1021/jp212541y. pMID: 22545654.

12. Bereau T, Deserno M: **Generic coarse-grained model for protein folding and aggregation**. *J Chem Phys* 2009, **130**, 06B621.

13. Souza PCT, Thallmair S, Conflitti P, Ramírez-Palacios C, Alessandri R, Raniolo S, Limongelli V, Marrink SJ: **Protein–ligand binding with the coarse-grained martini model**. *Nat Commun* 2020, **11**:3714, https://doi.org/10.1038/s41467-020-17437-5.

14. Roel-Touris J, Don CG, Honorato RV, Rodrigues JPGLM, Bonvin AMJJ: **Less is more: coarse-grained integrative modeling of large biomolecular assemblies with HADDOCK**. *J Chem Theor Comput* 2019, **15**:6358–6367, https://doi.org/10.1021/acs.jctc.9b00310.

15. Louhivuori M, Risselada HJ, van der Giessen E, Marrink SJ: **Release of content through mechano-sensitive gates in pressurized liposomes**. *Proc Natl Acad Sci USA* 2010, **107**:19856–19860.

16. Davies KM, Anselmi C, Wittig I, Faraldo-Gómez JD, Kühlbrandt W: **Structure of the yeast F1Fo-ATP synthase dimer and its role in shaping the mitochondrial cristae**. *Proc Natl Acad Sci USA* 2012, **109**:13602–13607.

17. Zheng W, Tsai M-Y, Chen M, Wolynes PG: **Exploring the aggregation free energy landscape of the amyloid-β protein (1–40)**. *Proc Natl Acad Sci USA* 2016, **113**:11835–11840.

18. Pak AJ, Yu A, Ke Z, Briggs JAG, Voth GA: **Cooperative multivalent receptor binding promotes exposure of the SARS-CoV-2 fusion machinery core**. *Nat Commun* 2022, **13**.

19. Giulini M, Menichetti R, Shell MS, Potestio R: **An information-theory-based approach for optimal model reduction of biomolecules**. *J Chem Theor Comput* 2020, **16**:6795–6813.

20. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE: **How fast-folding proteins fold**. *Science* 2011, **334**:517–520.

21. 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.

22. Bottaro S, Lindorff-Larsen K: **Biophysical experiments and biomolecular simulations: a perfect match?** *Science* 2018, **361**:355–360.

23. Shaw DE, Grossman J, Bank JA, Batson B, Butts JA, Chao JC, Deneroff MM, Dror RO, Even A, Fenton CH, *et al.*: **Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer**. In *SC'14: proceedings of the international conference for high*

performance computing, networking, storage and analysis. IEEE; 2014:41–53.

24. Prinz J-H, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C, Noé F: **Markov models of molecular kinetics: generation and validation**. *J Chem Phys* 2011, **134**, 174105.

25. Husic BE, Pande VS: **Markov state models: from an art to a science**. *J Am Chem Soc* 2018, **140**:2386–2396.

26. * Noé F, Tkatchenko A, Müller K-R, Clementi C: **Machine learning for molecular simulation**. *Annu Rev Phys Chem* 2020, **71**: 361–390.
A review of the application of machine learning on different aspects of molecular simulation.

27. Unke OT, Chmiela S, Sauceda HE, Gastegger M, Poltavsky I, Schütt KT, Tkatchenko A, Müller K-R: **Machine learning force fields**. *Chem Rev* 2021, **121**:10142–10186.

28. Kapil V, Schran C, Zen A, Chen J, Pickard CJ, Michaelides A: **The first-principles phase diagram of monolayer nanoconfined water**. *Nature* 2022, **609**:512–516.

29. Gigli L, Veit M, Kotiuga M, Pizzi G, Marzari N, Ceriotti M: **Thermodynamics and dielectric response of BaTiO3 by data-driven modeling**. *NPJ Comput Mater* 2022, **8**:1–17.

30. Noid WG: **Perspective: coarse-grained models for biomolecular systems**. *J Chem Phys* 2013, **139**, 090901.

31. * Jin J, Pak AJ, Durumeric AE, Loose TD, Voth GA: **Bottom-up coarse-graining: principles and perspectives**. *J Chem Theor Comput* 2022, **18**:5759–5791.
A review of the theory, implementation, and open challenges for bottom-up coarse-graining.

32. Bartók AP, Payne MC, Kondor R, Csányi G: **Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons**. *Phys Rev Lett* 2010, **104**, 136403.

33. Schütt KT, Sauceda HE, Kindermans PJ, Tkatchenko A, Müller KR: **SchNet - a deep learning architecture for molecules and materials**. *J Chem Phys* 2018, **148**, 241722.

34. John ST, Csányi G: **Many-body coarse-grained interactions using Gaussian approximation potentials**. *J Phys Chem B* 2017, **121**:10934–10949.

35. 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.

36. * Husic BE, Charron NE, 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.
A graph neural network is used to model the effective energy of an ML CG model. The network architecture makes the model in principle transferable across chemical space.

37. Chen Y, Krämer A, Charron NE, Husic BE, Clementi C, Noé F: **Machine learning implicit solvation for molecular dynamics**. *J Chem Phys* 2021, **155**, 084101.

38. Wagner JW, Dama JF, Durumeric AE, Voth GA: **On the representability problem and the physical meaning of coarse-grained models**. *J Chem Phys* 2016, **145**, 044108.

39. Dunn NJ, Foley TT, Noid WG: **Van der Waals perspective on coarse-graining: progress toward solving representability and transferability problems**. *Acc Chem Res* 2016, **49**: 2832–2840.

40. ** Jin J, Pak AJ, Voth GA: **Understanding missing entropy in coarse-grained systems: addressing issues of representability and transferability**. *J Phys Chem Lett* 2019, **10**:4549–4557.
A comprehensive discussion on the challenges of representability and transferability in coarse-grained systems.

41. Tóth G: **Interactions from diffraction data: historical and comprehensive overview of simulation assisted methods**. *J Phys Condens Matter* 2007, **19**, 335220.

42. Lemke T, Peter C: **Neural network based prediction of conformational free energies - a new route toward coarse-grained simulation models**. *J Chem Theor Comput* 2017, **13**:6213–6221.

43. Zhang L, Han J, Wang H, Car R, DeePCG WE: **Constructing coarse-grained models via deep neural networks**. *J Chem Phys* 2018, **149**, 034101.

44. ** Wang J, Olsson S, Wehmeyer C, Pérez A, Charron NE, de Fabritiis G, Noé F, Clementi C: **Machine learning of coarse-grained molecular dynamics force fields**. *ACS Cent Sci* 2019.
Introduction of machine learning into protein coarse-graining, by training a neural network with a physical prior potential with variational force-matching.

45. * Thaler S, Zavadlav J: **Learning neural network potentials from experimental data via Differentiable Trajectory Reweighting**. *Nat Commun* 2021, **12**:6884.
Coarse-graining with machine learning methods; in particular, the parameters of a neural network are tuned to reproduce selected observables using iterative simulations.

46. * Ding X, Zhang B: **Contrastive learning of coarse-grained force fields**. *J Chem Theor Comput* 2022, **18**:6334–6344.
An alternative approach for bottom-up coarse-graining of molecular systems, that does not require atomistic forces nor expensive sampling iterations.

47. Thaler S, Stupp M, Zavadlav J: *Deep coarse-grained potentials via relative entropy minimization*. 2022. arXiv e-prints, arXiv: 2208.10330.

48. Noid WG, Chu J-W, Ayton GS, Krishna V, Izvekov S, Voth GA, Das A, Andersen HC: **The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models**. *J Chem Phys* 2008, **128**, 244114.

49. Shell MS: **The relative entropy is fundamental to multiscale and inverse thermodynamic problems**. *J Chem Phys* 2008, **129**, 144108.

50. Carmichael SP, Shell MS: **A new multiscale algorithm and its application to coarse-grained peptide models for self-assembly**. *J Phys Chem B* 2012, **116**:8383–8393.

51. Hinton GE: **Training products of experts by minimizing contrastive divergence**. *Neural Comput* 2002, **14**:1771–1800.

52. Gutmann M, Hyvärinen A: **Noise-contrastive estimation: a new estimation principle for unnormalized statistical models**. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings; 2010:297–304.

53. Wieder M, Fass J, Chodera JD: **Teaching free energy calculations to learn from experimental data**. *bioRxiv* 2021, **24**, 457513. 2021.

54. ** Köhler J, Chen Y, Krämer A, Clementi C, Noé F: *Flow-matching – efficient coarse-graining molecular dynamics without forces*. 2022, 11167. arXiv:2203.
Data efficient method for learning coarse-grained force-fields by employing generative deep neural networks. It is shown to work well on a set of fast-folding proteins.

55. Lyman E, Pfaendtner J, Voth GA: **Systematic multiscale parameterization of heterogeneous elastic network models of proteins**. *Biophys J* 2008, **95**:4183–4192.

56. Podryabinkin EV, Tikhonov EV, Shapeev AV, Oganov AR: **Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning**. *Phys Rev B* 2019, **99**, 64114.

57. Shen K, Sherck N, Nguyen M, Yoo B, Köhler S, Speros J, Delaney KT, Fredrickson GH, Shell MS: **Learning composition-transferable coarse-grained models: designing external potential ensembles to maximize thermodynamic information**. *J Chem Phys* 2020, **153**, 154116.

58. Kanekal KH, Rudzinski JF, Bereau T: **Broad chemical transferability in structure-based coarse-graining**. *J Chem Phys* 2022, **157**, 104102.

59. Hills RD, Lu L, Voth GA: **Multiscale coarse-graining of the protein energy landscape**. *PLoS Comput Biol* 2010, **6**, e1000827.

60. Wang W, Gómez-Bombarelli R: **Coarse-graining auto-encoders for molecular dynamics**. *npj Comput. Mater.* 2019, **5**:125.

61. * Foley TT, Kidder KM, Shell MS, Noid W: **Exploring the landscape of model representations**. *Proc Natl Acad Sci USA* 2020, **117**:24061–24068.

The effect of different choices of resolution and mapping on coarse-graining model systems

62. Wang W, Xu M, Cai C, Miller BK, Smidt T, Wang Y, Tang J, Gómez-Bombarelli R: *Generative coarse-graining of molecular conformations*. 2022, 12176. arXiv:2201.

63. Chennakesavalu S, Toomer DJ, Rotskoff GM: *Ensuring thermodynamic consistency with invertible coarse-graining*. 2022. arXiv preprint arXiv:2210.07882.

64. * 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*. 2022. arXiv e-prints, arXiv:2210.07237.

The authors show that the force error is not a good metric to evaluate the performance of machine-learned force-fields and introduce a new benchmark suite for ML MD simulation.

65. Ramakrishnan R, Dral PO, Rupp M, von Lilienfeld OA: **Big data meets quantum chemistry approximations: the δ-machine learning approach**. *J Chem Theor Comput* 2015, **11**: 2087–2096.

66. Dolgirev PE, Kruglov IA, Oganov AR: **Machine learning scheme for fast extraction of chemically interpretable interatomic potentials**. *AIP Adv* 2016, **6**, 085318.

67. Deringer VL, Csányi G: **Machine learning based interatomic potential for amorphous carbon**. *Phys Rev B* 2017, **95**, 094203.

68. Ricci E, Giannakopoulos G, Karkaletsis V, Theodorou DN, Vergadou N: **Developing machine-learned potentials for coarse-grained molecular simulations: challenges and pitfalls**. In *Proceedings of the 12th hellenic conference on artificial intelligence*; 2022:1–6.

69. * Musil F, Grisafi A, Bartók AP, Ortner C, Csányi G, Ceriotti M: **Physics-inspired structural representations for molecules and materials**. *Chem Rev* 2021, **121**:9759–9815.

An extensive review of different representations and their properties for the modeling of molecules and materials with machine learning methods.

70. Behler J, Parrinello M: **Generalized neural-network representation of high-dimensional potential-energy surfaces**. *Phys Rev Lett* 2007, **98**.

71. Smith JS, Isayev O, Roitberg AE: **ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost**. *Chem Sci* 2017, **8**:3192–3203.

72. Stocker S, Gasteiger J, Becker F, Günnemann S, Margraf JT: **How robust are modern graph neural network potentials in long and hot molecular dynamics simulations?** *Chem* Apr 2022.

73. Unke OT, Stöhr M, Ganscha S, Unterthiner T, Maennel H, Kashubin S, Ahlin D, Gastegger M, Sandonas LM, Tkatchenko A, Müller K-R: *Accurate machine learned quantum-mechanical force fields for biomolecular simulations*. 2022. arXiv e-prints, arXiv: 2205.08306.

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. Durumeric AEP, Voth GA: *Explaining classifiers to understand coarse-grained models*. 2021. arXiv preprint arXiv:2109.07337.

76. Gal Y, Ghahramani Z: **Dropout as a bayesian approximation: representing model uncertainty in deep learning**. In *Proceedings of the 33rd international conference on international conference on machine learning*, vol. 48; 2016:1050–1059.

77. Lakshminarayanan B, Pritzel A, Blundell C: **Simple and scalable predictive uncertainty estimation using deep ensembles**. In *Proceedings of the 31st international conference on neural information processing systems*; 2017:6405–6416.

78. Vandermause J, Torrisi SB, Batzner S, Xie Y, Sun L, Kolpak AM, Kozinsky B: **On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events**. *NPJ Comput. Mater.* 2020, **6**:1–11.

79. Sanyal T, Mittal J, Shell MS: **A hybrid, bottom-up, structurally accurate, Gō-like coarse-grained protein model**. *J Chem Phys* 2019, **151**, 044111.

80. Potter TD, Tasche J, Wilson MR: **Assessing the transferability of common top-down and bottom-up coarse-grained molecular models for molecular mixtures**. *Phys Chem Chem Phys* 2019, **21**:1912–1927.

81. Rosenberger D, Van Der Vegt NF: **Addressing the temperature transferability of structure based coarse graining models**. *Phys Chem Chem Phys* 2018, **20**:6617–6628.

82. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE: **UCSF ChimeraX: meeting modern challenges in visualization and analysis**. *Protein Sci* 2018, **27**: 14–25.

83. Hunter JD: **Matplotlib: a 2D graphics environment**. *Comput Sci Eng* 2007, **9**:90–95.