JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Review

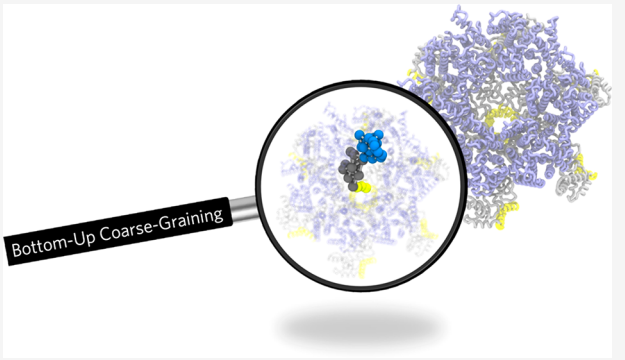
Bottom-up Coarse-Graining: Principles and Perspectives

Jaehyeok Jin, Alexander J. Pak, Aleksander E. P. Durumeric, Timothy D. Loose, and Gregory A. Voth*

Cite This: J. Chem. Theory Comput. 2022, 18, 5759−5791 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Large-scale computational molecular models provide scientists a means to investigate the effect of microscopic details on emergent mesoscopic behavior. Elucidating the relationship between variations on the molecular scale and macroscopic observable properties facilitates an understanding of the molecular interactions driving the properties of real world materials and complex systems (e.g., those found in biology, chemistry, and materials science). As a result, discovering an explicit, systematic connection between microscopic nature and emergent mesoscopic behavior is a fundamental goal for this type of investigation. The molecular forces critical to driving the behavior of complex heterogeneous systems are often unclear. More problematically, simulations of representative model systems are often prohibitively expensive from both spatial and temporal perspectives, impeding straightforward investigations over possible hypotheses characterizing molecular behavior. While the reduction in resolution of a study, such as moving from an atomistic simulation to that of the resolution of large coarse-grained (CG) groups of atoms, can partially ameliorate the cost of individual simulations, the relationship between the proposed microscopic details and this intermediate resolution is nontrivial and presents new obstacles to study. Small portions of these complex systems can be realistically simulated. Alone, these smaller simulations likely do not provide insight into collectively emergent behavior. However, by proposing that the driving forces in both smaller and larger systems (containing many related copies of the smaller system) have an explicit connection, systematic bottom-up CG techniques can be used to transfer CG hypotheses discovered using a smaller scale system to a larger system of primary interest. The proposed connection between different CG systems is prescribed by (i) the CG representation (mapping) and (ii) the functional form and parameters used to represent the CG energetics, which approximate potentials of mean force (PMFs). As a result, the design of CG methods that facilitate a variety of physically relevant representations, approximations, and force fields is critical to moving the frontier of systematic CG forward. Crucially, the proposed connection between the system used for parametrization and the system of interest is orthogonal to the optimization used to approximate the potential of mean force present in all systematic CG methods. The empirical efficacy of machine learning techniques on a variety of tasks provides strong motivation to consider these approaches for approximating the PMF and analyzing these approximations.



1. INTRODUCTION

Understanding how molecular phenomena translate into emergent mesoscopic and macroscopic behavior is a common theme throughout biology, chemistry, physics, materials science, and engineering. Experimental techniques have offered microscopic insights into systems from these fields. For example, ensemble-averaged atomic structures can be resolved at high-resolution using X-ray crystallography or cryo-electron microscopy.[1,2] Alternatively, fluorescence techniques[3,4] or nuclear magnetic resonance (NMR) spectroscopy[5,6] can provide dynamic information, albeit at lower spatial resolution. To complement these experimental approaches, theorists leverage classical molecular dynamics (MD) simulations to investigate dynamical phenomena at high spatial resolution, most commonly at the atomistic level.[7] However, within the space of MD simulation techniques, coarse-grained (CG) modeling and simulation are particularly attractive for the study of systems with hierarchical length and time scales such as biomolecular systems[8−14] (including UNRES,[15−18] OPEP,[13,19] PRIMO,[20] SIRAH,[21,22] MARTINI,[23−27] MS-CG,[28−34] and REM[35−41]).

By design, CG models are reduced representations of fine-grained (FG) atomistic resolution molecules, where CG sites represent groups of corresponding FG atoms through a process that can be called *mapping*. Effective interactions between the CG sites are *parametrized* to retain the essential aspects of the

Received: June 20, 2022
Published: September 7, 2022

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

system of interest under the chosen equations of motion. However, defining these essential aspects depends upon the scientific question at hand and the reference data available, variations in which have led to the development of the *top-down* and *bottom-up* approaches, which are the two general classes of CG models that we will discuss later. CG simulations have three primary benefits compared to FG simulations. First, these models enable simulations of larger systems at appropriate length scales by virtue of the reduced number of particles. Second, a larger integration time step can be used in CG simulations since the removal of highly fluctuating short wavelength atomistic degrees of freedom results in a smoother CG free energy surface that accelerates the sampling under Hamiltonian mechanics. Finally, the construction of useful CG models grants tacit insight into molecular features (i.e., from CG mappings) and energetics (i.e., from CG interactions and associated equations of motion) that are essential for understanding mesoscopic and macroscopic behavior. For these reasons, CG simulations provide perspectives that would otherwise be inaccessible from more detailed atomistic MD simulations, which has driven their continued use and development.

The top-down strategy is perhaps the more typical CG modeling approach. Often, the scientific question posed by top-down CG studies is to determine if a particular set of interactions is capable of reproducing specific macroscopic properties. For example, the original MARTINI CG model for lipids was parametrized to recapitulate partition coefficients.[23−27] Studies of self-assembling systems have also benefited from top-down approaches. Using simplified geometries and interactions, it has been possible to broadly explore how morphologies may be dictated by a few adjustable parameters.[42−45] However, as these approaches neglect the direct validation of microscopic details by design, it is unclear if the resultant CG models faithfully reproduce microscopic physics. For example, the original MARTINI model, by construction, lacks a rigorous CG mapping from atomistic degrees of freedom by design[46] and also may not reflect the underlying nature of atomistically mapped interactions onto the CG representation, such as the correct enthalpy−entropy decomposition for certain calculated potentials of mean force (PMF).[47,48]

Bottom-up approaches use the opposite strategy and attempt to reproduce microscopic (mapped atomistic) statistics. The underlying principles of most bottom-up approaches are that properties observed in reference simulations are to be captured by the correct CG equations of motion describing equilibrium and certain nonequilibrium processes. For example, static properties are to be reproduced by the effective CG interactions as determined by equilibrium statistical mechanical principles. The majority of bottom-up CG approaches aim to reproduce static correlations. One common strategy, which we refer to as *thermodynamic consistency*, is to systematically parametrize CG models such that the sampled distribution recapitulates the multidimensional configurational distribution of their FG counterparts when mapped to the CG phase space.[28,31−33] Under this criterion, the ideal effective CG Hamiltonians are the conditioned (or CG mapped) many-body PMFs expressed in the CG coordinates or configurations. Reproducing the many-body PMFs using an arbitrarily complex set of functions or "basis set",[34] however, is challenging and sometimes problematic, as even if it is computationally feasible to capture the properties of the simulation being analyzed for parametrization, the resulting potential must also describe the larger system of primary interest to the study at hand—an extrapolative task that becomes increasingly difficult as the basis set grows in complexity. Instead, bottom-up studies have explored if CG models that recapitulate reduced sets of microscopic statistics using similarly simplified basis functions are also capable of collectively recapitulating mesoscopic and macroscopic behavior.[9,49−51] Unlike static properties, dynamical processes are correlated with both temporal and spatial variables and thus can be difficult to represent at CG resolution. Extracting the time evolution of CG systems from the FG reference provides a strategy for rigorously integrating the many-body nature of time-dependent processes into CG models.[52−54]

Bottom-up CG models are created to generate samples that systematically approximate high-dimensional data produced by a reference model. This approach is fundamentally similar to those of contemporary methods in machine learning (ML).[55,56] Algorithms for high-dimensional regression have been applied as the building blocks for force fields that recapitulate high order correlations in CG mapped atomistic data.[57−60] Simultaneously, ML techniques focused on directly generating samples from high-dimensional distributions have provided approaches for quantifying the error in existing CG models and more efficient methods for generating atomistic configurations. These novel generative approaches have additionally shown promise in producing atomistic configurations from CG simulations (an approach often referred to as "backmapping").[61−69]

In this Review, we summarize recent advances in bottom-up CG modeling (with some, albeit abbreviated, historical context) and discuss promising future directions. In particular, we focus on how bottom-up CG models can be derived by establishing a systematic connection between the microscopic (atomistic) and the reduced descriptions. We first review fundamental concepts in CG models that have been proposed over the past two decades. We then discuss limitations and challenges in CG modeling in terms of consistency, representability, and transferability. We briefly survey recent scientific findings and breakthroughs that benefited from methodological advances, e.g., concepts from ML. We conclude with a brief overarching summary and future outlook toward the next generation of bottom-up CG modeling.

2. BASICS OF BOTTOM-UP COARSE-GRAINED MODELING

Two ingredients are necessary for any bottom-up CG modeling recipe. While intertwined, these processes are typically performed separately. First, one needs to define the CG mapping that formally defines the correspondence between the FG and reduced resolutions. Then, once a mapping is selected, the CG mechanics need to be defined in the desired CG phase space on the basis of the FG statistics mapped onto that space. Various techniques can be applied for parametrizing the interactions governing the CG equations of state. These two ingredients are not only essential to construct CG models but also to provide the theoretical basis to understand the challenges underlying bottom-up CG modeling, i.e., consistency, transferability, and representability. Below, we briefly describe these essential steps required for bottom-up CG modeling and discuss how they are related to the aforementioned challenges.

2-1. Coarse-Grained Mapping. The process of applying a CG mapping reduces the high-dimensional atomistic phase space to a low-dimensional CG phase space. Ideally, this should involve mapping over both configurational and momentum variables in phase space, but most molecular CG models do not

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

involve momentum in their equilibrium distribution consistency, and under this assumption, the majority of CG mappings can be readily applied to only the configurational variables. The field of chemistry has an established history of breaking complex molecules into moieties and functional groups in order to predict and understand atomistic behavior. This process provides an intuitive basis for designing CG configurational mappings. Even though there are numerous mapping schemes that one can take in order to preserve the desired behavior in CG models, the resultant CG interaction associated with the specific mapping should be designed based on the statistical mechanical principles given by the CG methodology, which will be reviewed in Section 2-2 below.

2-1.A. Based on Real Particles. Representing each chemical moiety as a group of CG sites is perhaps most easily realized by defining each CG site as a weighted average of the configurations of various atoms. For these real particle-based mappings, the CG mapping operator on configurational variables is expressed as a set of N linear functions $\mathbf{M}(\mathbf{r}):(\mathbf{M}_1(\mathbf{r}),...,\mathbf{M}_N(\mathbf{r}))$, where the FG coordinates can be mapped into CG site I following

$$\mathbf{M}_I(\mathbf{r}) = \sum_i c_{Ii}\mathbf{r}_i \tag{1}$$

One example is the center-of-mass mapping ($c_{Ii} \propto m_i$), enabling one to retain important molecular configurations and momenta at the reduced level. For force-based CG methodologies, the center-of-mass mapping provides thermodynamically consistent forces that act on the center-of-mass phase space variables in comparison to atomistic forces.[32] However, information beyond configurations may be lost using the center-of-mass mapping. Recent advances have suggested that one can alternatively perform center-of-charge mapping, which is a reweighted mass based on the partial charges of atoms within CG sites,[70] to better encode electrostatic information[71,72] for systems in which electrostatic interactions play a major role, e.g., ionic liquids.[73] It is also conceivable that the geometry of the system can be better conserved by performing center-of-geometry mapping as $c_{Ii} = 1/n_I$, where n_I denotes the number of FG particles involved in CG site I.

2-1.B. Based on Virtual Particles. Interactions that are centered on CG sites mapped from the FG variables alone can be a limitation in CG modeling. Particles that do not explicitly represent specific FG particles can be included as additional interaction centers, thereby introducing a general means to increase the expressivity of the desired model; we holistically refer to these particles as "virtual sites". Virtual sites have been used to impart subtle anisotropic projections of forces acting upon real sites. One prototypical example of this idea is the atomistic TIP4P water model.[74] Similar types of virtual sites have been used in the context of high-resolution CG models, notably for sterols and for aromatic hydrocarbons.[75,76] Overall, virtual sites can be thought of as relatively inexpensive augmentations to conventional real particle-based mappings.

Virtual sites in CG models have been increasingly utilized in recent years, most predominantly in top-down CG models.[77−79] Moreover, virtual CG sites were required to describe directional interactions at protein−protein interfaces that are responsible for viral self-assembly.[44,80−86] However, the use of virtual sites in bottom-up CG models has been limited due to a lack of systematic rules that describe effective virtual site interactions. One proposed approach is to use the so-called center-of-symmetry framework that maintains thermodynamic consistency while preserving the molecular asymmetry via virtual particles.[76] We note that the necessity of preserving molecular symmetry in CG mapping is still an open problem,[87] but the center-of-symmetry framework can effectively encode the missing quadrupole information into CG models (e.g., $\pi-\pi$ stacking from benzene rings), enhancing the fidelity of structural correlations and transferability. These early successes demonstrate the utility of virtual sites and motivate the need for additional efforts to determine systematic rules for virtual site mappings and effective interactions for complex molecular systems, e.g., polymers.[88] Another approach that shares a similar physical principle is to introduce virtual sites that help to represent the effects of explicit solvent in implicit solvent models.[89] These so-called "solvent-free" CG models can potentially provide an accurate and transferable CG modeling for biomolecules, e.g., amphiphilic assemblies of lipids.

2-1.C. Mesoscopic Mapping: Clustering. On larger mesoscopic scales, CG particles can instead be represented as supramolecular "blobs", and the CG mapping at this resolution becomes less clear than at the molecular level. If the target system is composed of bonded systems, one can still employ a linear center-of-mass mapping.[90−93] However, the same strategy cannot be applied for unbonded systems since the construction of CG blobs becomes a nonlinear and time-dependent procedure. As developed by Español and co-workers, who introduced the Voronoi cell representation,[94] suitable mesoscopic representations for unbonded systems must be obtained via alternate approaches such as clustering methods other than conventional center-of-mass and related mappings. One such example is the application of the k-means clustering algorithm to unbonded fluids,[95,96] and another option proposed by Praprotnik et al.[97,98] uses spatial tessellation at adaptive resolutions. By adjusting the center of each Voronoi cell based on its center-of-mass, these clustering methods allow for mesoscopic CG blobs to faithfully represent their FG counterparts. Yet, conformation-based clustering suffers from the nonanalytical nature of the mapping process, which makes the derivation of CG interactions impractical, and requires frequent reclustering over the simulation. This limits the development of highly CG models to study mesoscopic behavior.[99] Hence, several alternatives have been reported in the literature, including a spherical CG blob mapping by Ayton et al.[100] This latter work was also extended (mainly, but not completely, in a top-down manner) to treat biomolecular membranes, with and without bound proteins.[61,101−104]

In general, the nonlinear, nonanalytic, and iterative nature of the mesoscopic CG mapping is considered a major bottleneck in mesoscale modeling of unbonded molecules (i.e., liquids). Recently, a "dynamic mapping" scheme was developed by mapping velocities instead of configurations, which only requires the initial configuration from smoothed centroidal Voronoi tessellations.[105] To note, this mapping scheme is based on a Lagrangian description to track individual fluid particles, but a complementary Eulerian description can also be established in a similar vein.[106] In turn, this new approach allows stable propagation of the CG blobs over time, indicating its applicability to various fluids,[107] e.g., heterogeneous multiphase systems.

2-1.D. Backmapping. Due to information loss during the coarse-graining process, there is a seemingly inherent lower bound to the CG resolution when creating a model for a particular scientific question. While the CG resolution can be tuned to optimize the level of detail remaining as to only include the relevant parts of the system at hand (see Item E later), this

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

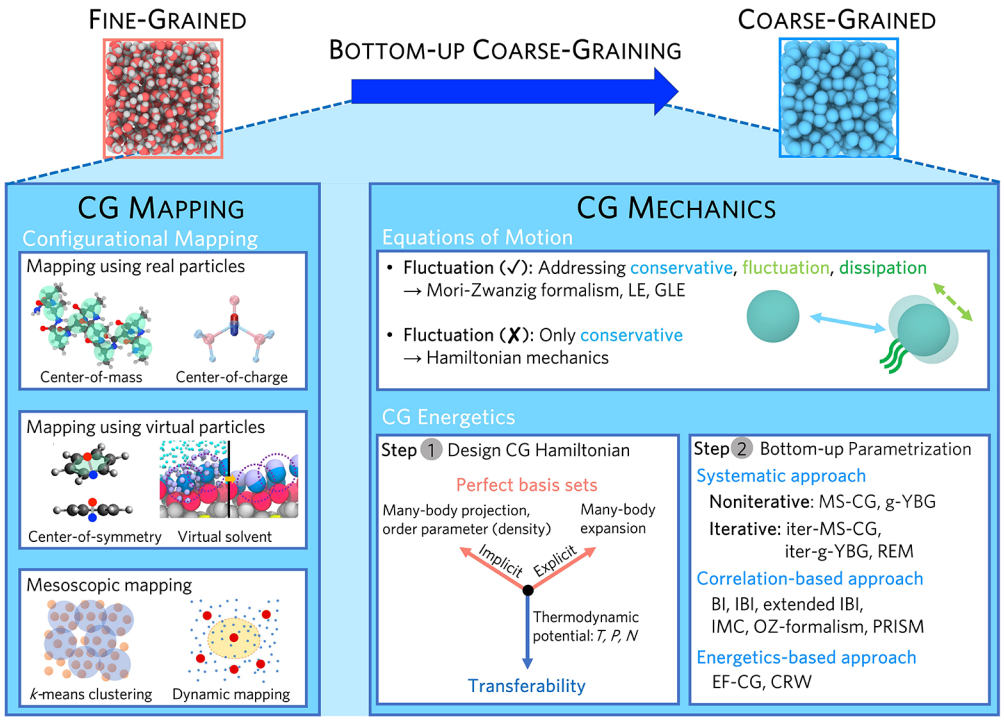


Figure 1. Broad summary of bottom-up CG modeling. Based on FG reference statistics, CG modeling is composed of two steps. (1) CG mapping (often performed on configurational variables) involves real or virtual CG particles at molecular resolution or includes mesoscopic mapping at coarser resolutions. (2) CG mechanics are defined by the specific consistency criteria and design principles that determine the CG equation of motion and CG interactions. CG equations of motions are generally chosen based on the target dynamical information, e.g., with or without fluctuation forces. CG interactions are determined by the designed CG Hamiltonian, which may suffer from an imperfect basis set and transferability issues. Bottom-up parametrization methodologies are then applied to yield effective CG interactions that optimally approximate the level of physics specified by the consistency criteria and design principles.

resolution may itself prove to be too computationally expensive. As a result, it is often desirable to recapture the full FG details by "backmapping" the CG configurations onto FG configurations.[108] In prior studies, backmapped structures from CG simulations have been used as starting points for FG simulations in order to improve FG sampling.[61−69] Generally, the procedure is performed in two steps. First, an initial model for the backmapped FG structure is predicted using geometric algorithms. Then, the structure is equilibrated using a short MD simulation. However, due to the degeneracy associated with the CG mapping, this process is highly nontrivial, and some recent advances have introduced ML techniques to perform backmapping (see Section 6). Early work also developed reasonably rigorous statistical mechanical methods for backmapping from the CG resolution to obtain the Boltzmann distribution in the FG variables (or close to it).[109−113]

2-1.E. Optimal Resolution. Along with CG mapping, the resolution of CG models impacts the resultant CG model phase space and the performance of CG models in reproducing the values of key observables. These factors are particularly important for modeling complex biomolecules, where grouping different atomistic entities becomes less clear. While there is no definitive answer to this problem, the optimal resolution for the desired CG models can be chosen to maximally reproduce the key observables from CG models or the loss of information from the coarse-graining process.[114]

Notably, essential dynamics coarse-graining (ED-CG) has been developed to systematically estimate the optimal partitioning scheme for large biomolecules by recapitulating key dynamics (or essential dynamics).[115] ED-CG can be employed in conjunction with principal component analysis (PCA) of atomistic trajectories[115] or CG models of large proteins[116] described by an heterogeneous elastic network model (HeteroENM).[117] Various determination protocols have also been suggested recently to determine the optimal CG representation, e.g., constrained minimization,[118] stepwise optimization,[119] and fluctuation maximization.[120]

On the other hand, an alternative approach can be achieved by minimizing the loss of information accompanied by the coarse-graining process. References 121 and 122 systematically investigated this concept by reproducing underlying fluctuations in terms of the CG model spectrum using a Gaussian network model, which can be further extended to elastic network models (ENMs).[123] As this information loss is intrinsically related to the mapping entropy from the coarse-graining process, ref 124 suggests a CG mapping optimization strategy by minimizing the mapping entropy with the aid of ML techniques[125] or enhanced sampling algorithms[126] (e.g., Wang−Landau[127,128]) in a way to preserve the maximum possible FG information. We emphasize that the mapping entropy introduced here is one of the most central quantities not only to determine the CG mapping but also to correctly understand the representability and transferability of CG models as will be discussed in Section 2-4.

Beyond a single resolution, CG models with multiple resolutions can be a way to capture different levels of detail in inhomogeneous systems. However, these so-called adaptive resolution models also require a concurrent coupling among multiple levels of resolution and changing the particles' resolution on the fly. One of the most commonly practiced frameworks, the adaptive resolution simulation (AdResS)

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

approach, allows for a smooth transition between different resolutions by introducing a hybrid transition region.[129] Then, the coupling between each resolution is designed based on the ensemble of interest and desired resolution ranging from quantum to hydrodynamics levels. For the molecular level, the original AdResS used a force coupling,[129] and then Hamiltonian AdResS (H-AdResS) was developed by designing the global Hamiltonian.[130−132] For different spatial levels, several AdResS-based schemes have been developed, such as for larger proteins,[133] continuum hydrodynamics,[134,135] systems with quantized nuclei,[136,137] open systems,[138,139] and the grand canonical ensemble.[140,141] One particularly nontrivial question would be to determine the thermodynamically consistent bottom-up interactions at the multiple resolutions. We will briefly discuss this in Section 2-4.

2-2. Coarse-Grained Mechanics. Once the mapping operator is decided, CG mechanics must be determined via (1) the equations of motion for the CG phase space variables and (2) the effective CG interactions from the chosen equation of motion in phase space, which will then enable CG simulations to be performed. Similar to the process of CG mapping, in the canonical (constant *NVT*) ensemble, the effective CG interactions are only dependent on the CG configurations. The essential building blocks for CG models are summarized in Figure 1.

2-2.A. Coarse-Grained Equations of Motion. CG equations of motion and energetics affect the fidelity of CG models in different ways. In order to correctly address dynamical properties, CG models should faithfully represent the friction and fluctuations observed in the reference FG system.[52−54] On the other hand, the equilibrium static properties, e.g., structural correlations, are not dependent on dynamical behavior. Correct recapitulation of static correlations is possible through only conservative interactions. While the latter can be propagated under Hamiltonian mechanics, the proper dynamics (time-dependent behavior) requires equations of motion accounting for nonconservative interactions, e.g., the generalized Langevin equation. In this section, we specifically discuss the performance of CG models in terms of static correlations. The related discussion on dynamical properties is presented in Section 5.

2-2.B. Coarse-Grained Energetics: Design Principles. In general, it is impractical to determine exact forms of the renormalized many-body CG interactions. Therefore, constructing approximate bottom-up CG models is performed in two sequential steps. The first step is to design the form of the CG Hamiltonian in terms of configurational variables. This is often done by adopting molecular mechanics functional forms similar to that of the atomistic description, such that the approximate CG interactions are written as a combination of analytical and tractable forms, including bond, angle, torsion, and pair nonbonded interactions. Various design principles for the CG Hamiltonian will be described in detail in Section 3. The second step is to determine the interaction parameters for the defined CG Hamiltonian, which is the focus of this subsection.

2-3. Coarse-Grained Force Fields. *2-3.A. Bottom-up Philosophies.* In order to address a myriad of chemical and biological systems, most CG methodologies provide a general principle to determine CG interaction parameters regardless of the CG Hamiltonian form. This is often manifested in the bottom-up manner by enforcing certain statistical mechanical principles to maintain the fundamental properties of the FG system. Depending on the microscopic target of interest (thermodynamic properties or static correlations), various bottom-up CG methodologies have been proposed in the field. In this subsection, we briefly survey some of the leading strategies for approximating CG energetics, and we will review naturally emerging issues in Section 3.

2-3.B. Based on Variational Principles. As noticed from the CG equations of motion, forces are central to equilibrium thermodynamics, and several methodologies have been designed based on the conservative forces. The idea of employing force-matching (without coarse-graining) originated from the early work of Ercolessi and Adams[142] and from Izvekov, Parrinello, Burnham, and Voth[143] to define molecular mechanics force fields on an *ad hoc* basis from *ab initio* calculations. The extension of force-matching to CG configurational space was established with the Multiscale Coarse-Graining (MS-CG) method developed first by Izvekov and Voth for biomolecular systems[28,30] and liquids.[29] This advance was accomplished through a recognition that force-matching could be carried out along with a resolution reduction (coarse-graining) and that this would be a variational route to determine the many-body PMF for the CG variables. Later, in a series of papers,[31−34,70,144−151] the MS-CG method was more fully developed and explored. By design, the MS-CG methodology determines the effective force field acting on CG site I, $\mathbf{F}_I(\mathbf{M}(\mathbf{r}^n))$, by minimizing the least-squared force residual $\chi^2[\mathbf{F}]$ between a target CG model and the FG counterpart. Here, $\mathbf{M}$ denotes a mapping operator that maps the FG configuration $\mathbf{r}^n$ to the CG configuration $\mathbf{R}^N$. These force differences are often expressed as a quadratic residual, and thus, a systematic determination is possible by variationally minimizing this force metric

$$\chi^2[\mathbf{F}] = \frac{1}{3N}\left\langle \sum_{I=1}^{N} |\mathbf{F}_I(\mathbf{M}(\mathbf{r}^n)) - \mathbf{f}_I(\mathbf{r}^n)|^2 \right\rangle \tag{2}$$

where $\mathbf{F}_I(\mathbf{M}(\mathbf{r}^n))$ is the unknown CG forces at the CG configuration, and $\mathbf{f}_I(\mathbf{r}^n)$ is the projected microscopic forces on the CG site I. The unknown CG forces can be linearly expressed through the CG force field parameters ϕ: $U_{\mathrm{CG}}(\mathbf{R}^N) = \sum_k \phi_k V_k(\mathbf{R}^N)$ via $\mathbf{F}_I(\mathbf{R}^N) = -\nabla_I U_{\mathrm{CG}}(\mathbf{R}^N)$. In general, a two-body (pairwise) approximation is often adopted to express the CG force field basis sets, and subsequent algorithmic advances[34,152] introduced spline interpolation to describe force field parameters as $\mathbf{F}_I(\mathbf{R}^N) = \sum_{J\neq I} \xi_2(R_{IJ})\hat{e}_{IJ} = \sum_{J\neq I}\sum_k c_k u_k(R_{IJ})\hat{e}_{IJ}$ which reduces the least-squares problem in eq 2 to an overdetermined system of linear equations,[153] resulting in the following matrix equation

$$\mathbf{F}\phi = \mathbf{f} \tag{3}$$

In eq 3, $\mathbf{F}$ is the force matrix calculated from the CG configurations and the CG force field parameters ϕ, and the column vector $\mathbf{f}$ represents the projected FG forces at the CG resolution. From a statistical mechanical perspective, the MS-CG methodology satisfies the thermodynamic consistency between the FG and CG phase spaces.[32,33] It is important to note that this approach to training a bottom-up CG force field from FG data can be considered an early example of "machine learning", which has become very popular in recent times (including for coarse-graining), albeit a deep neural network[59] was not utilized in that early MS-CG work of almost 20 years ago.

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

In contrast to the force-based metric, Shell and co-workers have identified and implemented the information-theoretic relative entropy as a target metric.[35−41] Relative entropy is defined as the differences between the FG and CG probability distributions, given by the Kullback−Leibler divergence[154]

$$S_{\mathrm{rel}} = \int d\mathbf{r} p_{\mathrm{FG}}(\mathbf{r}^n) \ln \frac{p_{\mathrm{FG}}(\mathbf{r}^n)}{p_{\mathrm{CG}}(\mathbf{R}^N)} + \langle S_{\mathrm{map}} \rangle \tag{4}$$

where $\langle S_{\mathrm{map}} \rangle_{\mathrm{CG}}$ denotes the mapping entropy (introduced in Section 2-1) defined by a mapping operator, $S_{\mathrm{map}} = \ln \int \delta[\mathbf{M}(\mathbf{r}) - \mathbf{R}] d\mathbf{r}$. Thus, minimizing S_{rel} enforces minimizing the log difference between the FG and CG probability distributions. Based on this metric, relative entropy minimization (REM) can be expressed within the canonical ensemble as $S_{\mathrm{rel}} = \beta \langle U_{\mathrm{FG}} - U_{\mathrm{CG}} \rangle_{\mathrm{CG}} - (A_{\mathrm{FG}} - A_{\mathrm{CG}}) + \langle S_{\mathrm{map}} \rangle$, where U and A denote the internal energy and free energy, respectively. Then, a set of CG model parameters $\{\lambda_i\}$ can be variationally determined by minimizing the relative entropy differences between the FG and CG systems, resulting in the following conditions for local optimality[38]

$$\frac{\partial S_{\mathrm{rel}}}{\partial \lambda_i} = \beta \left\langle \frac{\partial U_{\mathrm{CG}}(\mathbf{R}^N)}{\partial \lambda_i} \right\rangle_{\mathrm{FG}} - \beta \left\langle \frac{\partial U_{\mathrm{CG}}(\mathbf{R}^N)}{\partial \lambda_i} \right\rangle_{\mathrm{CG}} = 0 \tag{5a}$$

$$\frac{\partial^2 S_{\mathrm{rel}}}{\partial \lambda_i^2} > 0 \tag{5b}$$

Due to the systematic nature of REM, several variations were reported to enhance the predictability of CG models.[155] The connection between MS-CG and REM has also been analyzed:[38,156] both will give the exact many-body CG variable PMF if a "perfect" basis set is used to describe the CG interactions, but the two approaches will differ in complementary ways if more approximate basis sets are used.

Even though the aforementioned approaches can variationally determine the effective CG interaction parameters, it is not immediately clear if the resultant CG models will reproduce a particular target atomistic correlation correctly given the approximate nature of the CG Hamiltonian as well as the basis set chosen to describe the CG interactions. Recent studies have, however, shed light on such connections. Noid et al. demonstrated that the MS-CG method determined from eq 2 satisfies the Yvon−Born−Green (YBG) hierarchical equation,[157] indicating that MS-CG with two-body interactions attempts to capture two-body and three-body structural correlations.[31] In a related fashion, Chaimovich and Shell showed that REM guarantees capturing any n-body statistical correlations that are explicitly represented in a corresponding n-body CG Hamiltonian.[38]

By establishing links between force-based models and structural correlations, explicit consideration of FG structural correlations can result in modified strategies for force field parametrization. Notably, Mullinax and Noid provided a practical link by developing the generalized YBG (g-YBG) framework to determine optimal interaction potentials for complex classical force fields by utilizing only structural correlation functions.[158,159] The g-YBG approach was readily applied to CG systems on the basis of the MS-CG framework, where the structural correlations were used instead of forces.[158−162] This observation is based on eq 3 that the normal equation form can be constructed by acting a transpose $\mathbf{F}^{\mathrm{T}}$ on the left-hand side. This produces the following equation

$$\mathbf{G}\phi := \mathbf{b} \tag{6}$$

where $\mathbf{G} := \mathbf{F}^{\mathrm{T}}\mathbf{F}$ and $\mathbf{b} := \mathbf{F}^{\mathrm{T}}\mathbf{f}$. In the g-YBG approach, b is expressed in terms of a set of structural correlation functions,[158−160] and $\mathbf{G}$ contains the ensemble average of cross-correlations between the CG degrees of freedom.

An iterative refinement to eq 6 can improve the reproduction of specified correlation functions. Namely, two iterative schemes are possible. The first approach is to update the G matrix iteratively by matching the CG forces at a given CG structure to the mapped FG forces. Cho and Chu applied this scheme to the MS-CG methodology,[163] and Rudzinski and Noid made similar extensions to the g-YBG framework.[164] Another iterative treatment is to recalculate the b column vector by matching the CG forces at a given FG structure to the CG forces at a given CG structure. The latter approach is equivalent to matching expectations of basis function derivatives, as demonstrated by Lu et al.[165] Another recent advance has resulted in a noniterative parametrization scheme, while still based on eq 6, in order to directly reproduce pair correlations by transforming the atomistic cross-correlations.[166] Lastly, the parametrization strategies discussed above can be improved by leveraging ML techniques. While general principles and examples of ML-based CG parametrization will be discussed in Section 6, we briefly note here that force-matching and relative entropy ideas can be translated into ML. For example, the Kullback−Leibler divergence in eq 4 can be extended to general f-divergence,[58] and the force-matching scheme in eq 2 can be utilized to train the CG free energy functional.[59] Beyond force-matching, the effective *flow* by combining eq 2 and 4 can also be trained to recapitulate the CG probability density.[167]

2-3.C. Based on Static Correlations. Alternatively, several bottom-up CG approaches have been specifically designed to capture target static correlations (as opposed to the many-body PMF of the CG variables like MS-CG and REM). One of the earliest attempts was to capture pair correlations, or radial distribution functions (RDFs), from FG systems[168,169] based on Henderson's uniqueness theorem,[170] which asserts that there is a unique pair potential that gives rise to a given RDF. Under dilute (low-density) conditions, one can ignore the many-body correlations in the system, and the effective pair interactions can be approximated based on the RDF or $g_{\mathrm{FG}}(R)$. This approach is known as the (direct) Boltzmann inversion[171]

$$U(R) = -k_{\mathrm{B}}T \ln g_{\mathrm{FG}}(R) \tag{7}$$

However, most condensed matter systems are not at the low-density limit, and thus, the model RDF from the CG trajectory using eq 7 often deviates from the reference RDF ($g_{\mathrm{FG}}(R)$). This deviation can be corrected using an iterative scheme, known as iterative Boltzmann inversion (IBI), where one can iteratively improve the fidelity of the CG models by updating the CG interactions according to

$$U_{k+1}(R) = U_k(R) - \alpha k_{\mathrm{B}}T \ln \frac{g_k(R)}{g_{\mathrm{FG}}(R)} \tag{8}$$

with a convergence rate α.[172] Even though the IBI approach has been applied to various chemical systems with its relatively simple update scheme, eq 8 is neither a variational approach nor strictly based on rigorous statistical mechanical principles, and thus, improving IBI models to resolve thermodynamic issues

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

cannot be done systematically. As a result, several IBI-based approaches require a heuristic, phenomenological theory to design the parametrization strategy. We note that ref 173 substantiates the well-posedness of IBI under certain conditions, but not all systems of interest fall into this case. Also, pair correlation-based CG methodologies suffer from numerical degeneracy in the RDF. Even though Henderson's theorem proves uniqueness, several practically constructed CG models (such as those parametrized via IBI and REM) can yield similar RDFs due to the sensitivity of pair distributions as well as numerical issues. The pair distribution itself intrinsically contains configurational degeneracy,[174] and recent work by Wang et al. systematically demonstrated that the RDF can be insensitive to pair interactions,[175] as substantiated by previously reported systems using various CG modeling methods.[176−178] Therefore, several extensions from IBI have been designed to surmount these issues, including the multistate generalization (multistate IBI) by McCabe and co-workers[179,180] as well as thermodynamic property-based IBI approaches such as Kirkwood−Buff IBI by van der Vegt and co-workers[181] and C−IBI by Junghans and Mukherji.[182]

Similarly, an approach named inverse Monte Carlo (IMC) by Lyubartsev and Laaksonen also targets the reference RDF and determines the CG interactions in an iterative manner.[183,184] Yet, these IMC iterations often suffer from computational overhead due to the sampling of all possible configurations.[185] In light of these issues, recent methodologies have been designed based on an integral equation approach that can approximate the many-body correlations via closure equations. For example, Guenza and co-workers have proposed an Ornstein−Zernike integral equation-based analytical approach to determine the effective CG interactions for polymeric systems.[186−188] It is also possible to utilize an inverted integral equation as an initial guess for IBI to improve parametrization efficiency.[189]

2-3.D. Based on Energetics. As opposed to target- or correlation-based approaches, one can also directly derive CG interactions from reference FG systems. Based on atomistic energetics, this class of approaches attempts to extract the reduced energetics, i.e., forces or free energies, directly from the atomistic energetics. For example, by borrowing from the force-matching philosophy, the effective force CG (EF-CG) approach computes the averaged forces acting on the groups of atoms by projecting onto corresponding radial vectors.[190] By directly extracting the force information from the atomistic simulations, the CG Hamiltonian is approximated as atomistically averaged forces, and EF-CG can be derived from an averaged description of the MS-CG methodology. Similar averaging schemes, as well as the EF-CG method, can be employed to capture structural correlations,[191−194] since these averaged interactions correctly account for the existing pair- and many-body correlations in FG systems. Similarly, in order to correctly capture the underlying FG energetics, a recent extension of the MS-CG method[195,196] (called "energy-matching") has shown that one could variationally determine the pair energy functions by minimizing the energy differences akin to eq 2. This allows for the pinpointing of energetic contributions from the many-body CG variable PMF, which is advantageous for better understanding of transferability.

A direct assessment of the PMF is also possible by explicitly computing effective pair potentials following the definition of the PMF in the low-density limit. In this case, by adopting a free energy perturbation approach, effective CG interactions can be computed via conditional reversible work (CRW).[197] This fragment-based CG approach can be advantageous in terms of transferability for different system conditions, yet, by design, it does not guarantee that the static correlations will be correctly addressed.[198−202]

2-4. Representability and Transferability in Coarse-Grained Modeling. *2-4.A. Accuracy in Coarse-Grained Modeling.* The accuracy of bottom-up CG models is undoubtedly important as it reflects the fidelity of CG models relative to the FG reference. Yet, appropriate metrics for accuracy are ambiguous at present. In this section, we define three separate yet related measures of bottom-up CG model fidelity: *consistency*, *representability*, and *transferability*. The use and importance of each of these measures are dependent upon the scientific question of interest. It is therefore worthwhile to discuss each of these metrics with the understanding that all three contribute to the overall accuracy of a bottom-up CG model. It should be noted that to an extent top-down CG models do not satisfy some or all of the properties discussed below, and it will generally mean that those top-down models are not consistent with statistical mechanics (meaning they do not provide a direct connection between the FG and CG worlds but are instead primarily models in the larger sense of the word).

2-4.B. Consistency. Consistency refers to the specific statistical mechanical principle used to generate a bottom-up CG model from a reference FG model. This consistency can serve as the theoretical basis to derive CG force fields or be utilized as the criterion to evaluate how the constructed CG models capture FG statistics in a consistent manner.

Since the emergence of CG modeling as a "field", various consistency conditions have been proposed. For example, the MS-CG methodology was initially built upon the consistency of excess free energies, asserting that the CG model should provide an excess free energy $A_{ex} = -k_B T \ln \left[\int d\mathbf{R}^N e^{-\beta U_{CG}(\mathbf{R}^N)}/V^N\right]$ that is identical to the FG excess free energy $a_{ex} = -k_B T \ln \left[\int d\mathbf{r}^n e^{-\beta u_{FG}(\mathbf{r}^n)}/V^n\right]$.[203] This condition is sufficient to derive the effective CG interaction form in configurational space. However, in order to consider the full phase space, the most commonly used criterion built upon equilibrium statistical mechanics is *thermodynamic consistency*,[31−33] indicating that the CG variables $p_{CG}(\mathbf{R}^N, \mathbf{P}^N)$ should have exactly identical probability distributions as compared to that of FG variables $p_{FG}(\mathbf{r}^n, \mathbf{p}^n)$ that are mapped to the specific CG phase variables via the mapping operator $\mathbf{M}:(\mathbf{r}^n, \mathbf{p}^n) \rightarrow (\mathbf{R}^N, \mathbf{P}^N)$. Both probability distributions follow Boltzmann sampling at equilibrium, and configurational and momentum variables can be separated as discussed earlier. Mathematically, this can be expressed as

$$p_{CG}(\mathbf{R}^N, \mathbf{P}^N) = \int d\mathbf{r}^n \delta(\mathbf{M}(\mathbf{r}^n) - \mathbf{R}^N) \int d\mathbf{p}^n \delta(\mathbf{M}(\mathbf{p}^n) - \mathbf{P}^N) \times p_{FG}(\mathbf{r}^n, \mathbf{p}^n) \quad (9)$$

where the mapping operators on configurational and momentum variables are folded into the delta functions, which are understood to be a product of delta functions in the expressions here, one for each FG to CG mapping, i.e., $\delta(\mathbf{M}(\mathbf{r}^n) - \mathbf{R}^N) := \prod_I^N (\delta(\mathbf{M}_I(\mathbf{r}^n) - \mathbf{R}_I))$ and $\delta(\mathbf{M}(\mathbf{p}^n) - \mathbf{P}^N) := \prod_I^N (\delta(\mathbf{M}_I(\mathbf{p}^n) - \mathbf{P}_I))$. This thermodynamic consistency can be reduced into configurational and momentum consistency relationships, respectively

$$p_{CG}(\mathbf{R}^N) = \int d\mathbf{r}^n \delta(\mathbf{M}(\mathbf{r}^n) - \mathbf{R}^N) p_{FG}(\mathbf{r}^n) \quad (10a)$$

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

$$p_{\mathrm{CG}}(\mathbf{P}^N) = \int d\mathbf{p}^n \delta(\mathbf{M}(\mathbf{p}^n) - \mathbf{P}^N) p_{\mathrm{FG}}(\mathbf{p}^n) \tag{10b}$$

The configurational consistency implies that the effective CG interaction potential should have the following form

$$U_{\mathrm{CG}}(\mathbf{R}^N) = -k_{\mathrm{B}}T \ln \int d\mathbf{r}^n \delta(\mathbf{M}(\mathbf{r}^n) - \mathbf{R}^N) \times \exp\left(-\frac{u_{\mathrm{FG}}(\mathbf{r}^n)}{k_{\mathrm{B}}T}\right) + (\text{const.}) \tag{11}$$

Eq 11 also demonstrates that the effective CG interaction that ensures configurational consistency is equivalent to the many-body PMF, which is a configuration-dependent free energy function. Notably, this condition serves as a design principle for CG force fields,[33] and a rigorous link between eq 11 and force-matching from eqs 2 and 3 was shown in ref 32.

Furthermore, the renormalized nature of the CG interactions shown in eq 11 points to the nontrivial challenge in designing a multiresolution CG model (one of the first systematic attempts can be found in ref 204). Especially, at the AdResS-level linking the FG and CG resolutions (assuming that we map to n_0 FG particles and N_0 CG particles), the overall renormalized interaction $U_{\mathrm{FG/CG}}(\mathbf{r}^n, \mathbf{R}^N)$ should be written as

$$U_{\mathrm{FG/CG}}(\mathbf{r}^n, \mathbf{R}^N) = -k_{\mathrm{B}}T \ln \int d\mathbf{r}^n \prod_i^{n_0} (\delta(\mathbf{M}_i(\mathbf{r}^n) - \mathbf{r}_i)) \times \prod_I^{N_0} (\delta(\mathbf{M}_I(\mathbf{r}^n) - \mathbf{R}_I)) \exp\left(-\frac{u_{\mathrm{FG}}(\mathbf{r}^n)}{k_{\mathrm{B}}T}\right) \tag{12}$$

where the constant term that is independent of configurational variables is omitted for simplicity. In eq 12, the delta functions containing the mapping function (to two different resolutions) enter inside the integrand of the many-dimensional integral. Therefore, the overall multiresolution interaction must be considered to be renormalized at both levels, FG and CG. However, this systematic thermodynamic connection is often missing in many current multiresolution models.[205−207] For example, the AdResS treats the overall potential as[130]

$$U_{\mathrm{AdResS}} = \sum_\alpha \{\lambda_\alpha V_\alpha^{\mathrm{FG}} + (1 - \lambda_\alpha) V_\alpha^{\mathrm{CG}}\} + V^{\mathrm{int}} \tag{13}$$

where coupling parameter λ_α varies from 0 to 1 to couple the FG and CG systems with their bulk (naked) interactions V_α^{FG} and V_α^{CG}, respectively. While this interpolative approach has shown to be able to capture various thermodynamic properties at equilibrium, the thermodynamic consistency principle of eq 12 suggests that there is no formal theory to support the idea that eq 12 can be written in terms of the "naked" FG potential as done in eq 13. As such, since the topic of adaptive and multiresolution modeling is a complex one,[204] it can still benefit from additional development based on rigorous statistical mechanical principles. In the remainder of this Review, we will therefore only focus on CG models with a single level of resolution.

While configurational consistency reveals the many-body nature of CG interactions, momentum consistency imposes a rigorous condition when designing CG mapping operators. Unlike configurational interactions, the momentum contribution to the system Hamiltonian is rather simple if the mapping functions are linear: $\prod_{i=1}^n \mathbf{P}_i^2/2m_i$ and $\prod_{I=1}^N \mathbf{P}_I^2/2M_I$ for the FG and CG systems, respectively. Moreover, this form of the momentum contribution, combined with eq 10b, asserts that no FG particles can be mapped to more than one CG site.[32]

2-4.C. Representability. The general connection between microscopic statistics and dynamical or thermodynamic observables (e.g., internal energy, entropy, pressure, temperature, etc.) has been rigorously derived through statistical mechanics. It is these relationships that allow computer simulations of FG models (and generally only FG models) to quantitatively predict experimental properties; in this paper, we holistically refer to these relationships as *representability*. In the case of dynamical representability, we will refer to how well the CG models reproduce FG dynamical properties, and this is often related to the choice of equations of motion used throughout the CG simulations. We specifically discuss dynamical representability in Section 5. Furthermore, in contrast to time-dependent properties and dynamical representability, the static picture of CG models can be assessed via structural representability. Structural representability is defined as how well the CG models reproduce FG structural correlations, which is highly dependent on the quality of the approximated conservative interactions at the CG level. This particular topic will be discussed in Section 3.

Our particular interest in this section is in representability issues that arise from thermodynamic inconsistencies observed in various CG systems, which are broadly referred to as the (thermodynamic) representability problem.[49−51] Perhaps the most common example is the trade-off between consistent recapitulation of pairwise correlations (via configurational consistency) and dramatic overestimation of pressure as computed from the virial theorem due to missing degrees of freedom in the CG model.[208,209] From a statistical mechanical perspective, the representability problem is rooted in the nature of the renormalized degrees of freedom. Due to the missing configurational and momentum variables in the CG phase space, the FG observables from the FG ensemble are not always equivalent in value to their CG counterparts if the latter are simply calculated using the same expressions as in the FG model.[49−51,121,195]

To demonstrate this perspective on representability, consider the fact that the ideal effective CG Hamiltonian in the CG configurational space, $U_{\mathrm{CG}}(\mathbf{R}^N)$, in the canonical ensemble under thermodynamic consistency is equivalent to the many-body PMF.[31−33] As a projection of the free-energy, U_{CG} has two important attributes. First, $U_{\mathrm{CG}}(\mathbf{R}^N)$ is clearly state point-dependent, i.e., $U_{\mathrm{CG}}(\mathbf{R}^N) = f(\mathbf{R}^N, V, T)$. Second, U_{CG} encodes both energetic (E_{CG}) and entropic (S_{CG}) contributions, i.e., $U_{\mathrm{CG}}(\mathbf{R}^N) = E_{\mathrm{CG}} - TS_{\mathrm{CG}}(\mathbf{R}^N)$;[148] the entropic contribution represents the entropy "lost" and folded into the CG interactions due to CG mapping, that is, the entropy associated with the FG degrees of freedom that map to the same CG configuration. Therefore, evaluating $\langle U_{\mathrm{CG}} \rangle$ would not give the FG internal energy. Instead, a reformulated expression $\langle U_{\mathrm{CG}} + TS_{\mathrm{CG}} \rangle$ could recover the FG internal energy; finding approximations for S_{CG} is an active area of research.[51,121,124]

The thermodynamic representability problem becomes more apparent for observable expressions explicitly mapped from the FG to CG ensembles.[49] For example, if the observable of interest, A, only depends upon configurational variables, this inconsistency can be mathematically formulated by examining the difference between $A_{\mathrm{FG}} = \langle A(\mathbf{r}^n) \rangle_{\mathbf{r}^n}$ and $A_{\mathrm{CG}} = \langle A_{\mathrm{CG}}(\mathbf{R}^N) \rangle_{\mathbf{R}^N}$. By introducing thermodynamic consistency into A_{FG}, it can be shown that observables at the CG resolution use equivalent expressions to that of the FG resolution only when the following definition is used:

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

$$A_{CG}(\mathbf{R}^N) = \frac{\int d\mathbf{R}^N e^{-\beta U_{CG}(\mathbf{R}^N)}}{\int d\mathbf{r}^n e^{-\beta u_{FG}(\mathbf{r}^n)}} \times \frac{\int d\mathbf{r}^n \delta(\mathbf{M}(\mathbf{r}^n) - \mathbf{R}^N) e^{-\beta u_{FG}(\mathbf{r}^n)} A(\mathbf{r}^n)}{e^{-\beta U_{CG}(\mathbf{R}^N)}} \quad (14)$$

However, as can be seen from the complex expression in eq 14, not all observables satisfy this criterion (eq 14), and this often results in $A_{FG} \neq A_{CG}$.[49] There has been considerable interest in determining the observable incompatibility from a rigorous statistical mechanical perspective.[210,211]

Pressure is an example of a thermodynamic observable that does not meet the criterion in the equation above. All FG configurational variables contribute to pressure according to the virial theorem,[157,212] rather than only the reduced CG configurational variables. As such, the evaluation of pressure in a CG model by simply using the virial expression as if the CG variables were the FG variables has essentially no connection to the pressure of the system at the actual FG level. Stated differently, the "pressure" in the CG model is rather meaningless unless interpreted as a part of a global model in which virtually everything is a model, not just the interactions between the CG particles. On the other hand, if the volume dependence of U_{CG} were known, it would be possible to construct a modified virial expression to compute the pressures using only CG variables that are consistent with the FG pressures.[49,146] One way to resolve this discrepancy is to determine a compatible observable expression using correct basis sets according to the thermodynamic properties of interest. A recent study on determining the correct observable expression for pressure using particle-wise decompositions suggested that such an approach can resolve the thermodynamic representability issue.[213] In related fashion, Lebold and Noid developed a dual-potential approach that combines both structure- and energy-based variational principles, resulting in a more faithful recapitulation of FG energetics at CG resolution.[195,196] We believe that this approach is generalizable and that finding appropriate expressions at CG resolution within the parametrized state point forms the basis of CG representability. Moreover, these approaches are expected to further reveal the fluctuations underlying fundamental thermodynamic quantities, e.g., heat capacity and isothermal compressibility, although, similar to CG potentials, they must be able to extrapolate to the system of primary interest.

2-4.D. Transferability. In CG modeling, the transferability issue naturally emerges from the differences between the FG and CG Hamiltonians and can be defined as a measure of how predictive or extrapolatory the CG models are to the statistics beyond the parametrized conditions. Since effective bottom-up CG interactions are free energy functions, unlike FG Hamiltonians, the CG interactions will vary at different state points (e.g., pressure, temperature, and composition). They are clearly defined as a function of the thermodynamic state point. For example, the constant *NVT* ensemble is generally chosen in practice. However, this fact does not necessarily mean that bottom-up CG models will have zero transferability. In this regard, imbuing transferability onto bottom-up CG models needs to be carried out in a manner consistent with the underlying thermodynamics. For example, based on the entropy-enthalpy decomposition approach introduced earlier,

$$U_{CG}(\mathbf{R}^N) = E_{CG}(\mathbf{R}^N) - TS_{CG}(\mathbf{R}^N) \quad (15)$$

the changes in entropic contributions to CG interactions $S_{CG}(\mathbf{R}^N)$ at different densities or temperatures should be correctly reflected while designing CG models. Most of these conditions are natural variables of free energy (temperature, pressure, volume) and system composition (bulk to mixtures with different ratios). Yet, these conditions are inextricably linked. For example, in mixture conditions, each molecular entity will experience differences in pressure and volume due to the presence of other molecules, resulting in different interaction profiles than that of bulk conditions. We will discuss recent advances in dealing with the transferability issue in Section 3.

2-5. Current Challenges in Bottom-up Coarse-Grained Modeling. Currently, major challenges in CG modeling originate from the approximate nature of CG models that aim to faithfully describe the complex many-body correlations and properties of atomistic systems. Here, we present some important challenges faced in these areas that have been and are being actively pursued by a number of researchers in the field.

- *Structural Representability*: How can one design CG models to capture higher-order structural correlations correctly? Most CG models suffer from this issue due to the use of relatively simple pairwise interactions.
- *Thermodynamic Representability*: While structural representability can be directly computed from CG simulations, correct representation of CG thermodynamic properties requires a systematic treatment. For example, how can one obtain comparable pressures, internal energies, or entropies with respect to the FG reference?
- *Dynamical Representability*: Since most bottom-up CG methodologies focus on configurational variables, the resultant CG dynamics is not guaranteed (or even likely) to be consistent with the FG reference. How can one overcome this inconsistency?
- *Transferability*: How can one design bottom-up CG models that can be applied to nonparameterized conditions? This issue is directly related to the applicability of CG models.
- *Machine Learning:* How can one benefit from emerging ML techniques to reduce the complexity underlying CG modeling?

With this in mind, we aim to address each of these issues in the remainder of this Review.

3. TOWARD MORE EXPRESSIVE COARSE-GRAINED BASIS SETS

3-1. Design Principles for Coarse-Grained Energetics. A bottom-up CG model that follows thermodynamic consistency requires that the effective CG Hamiltonian, $U_{CG}(\mathbf{R}^N)$, be exactly equivalent to the many-body PMF in terms of the CG variables.[31−33] However, it is computationally difficult to derive a many-body expression for $U_{CG}(\mathbf{R}^N)$, and many-body potentials are computationally expensive to use. Instead, low-dimensional basis sets are commonly used by adopting commonly used molecular mechanics functional forms, e.g., pairwise nonbonded interactions and bonded interactions (typically up to four-body terms, i.e., dihedrals and impropers).[214−219] In summary, $U_{CG}(\mathbf{R}^N)$ is often approximated as

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

$$U_{\mathrm{CG}}(\mathbf{R}^N) \approx \sum_{IJ} U_{\mathrm{nb}}^{(2)}(R_{IJ}) + \sum_{\mathrm{bonds}} U_b(d_{IJ}) + \sum_{\mathrm{angles}} U_\theta(\theta_{IJK}) + \sum_{\mathrm{dihedrals}} U_\psi(\psi_{IJKL}) \tag{16}$$

where $U_{\mathrm{nb}}^{(2)}(R_{IJ})$ is the two-body nonbonded potential that depends upon the distance R_{IJ} between CG sites I and J, $U_b(d_{IJ})$ is the two-body bonded potential, $U_\theta(\theta_{IJK})$ is the three-body angle potential, and $U_\psi(\psi_{IJKL})$ is the four-body dihedral or improper potential.

3-2. Current Challenges and Breakthroughs. By approximating the CG force field as eq 16, two types of errors are naturally introduced. The first error is due to the simplified nature of the interaction form (e.g., the pairwise approximation in nonbonded potential). The other error is caused by the inconsistency between FG force fields representing energetics while CG force fields are representing free energies (the CG PMFs). Even though eq 16 asserts that the CG PMF will be only a function of CG configurational variables, $U_{\mathrm{CG}}(\mathbf{R}^N)$ ignores explicit contributions from other thermodynamic variables (e.g., volume or temperature) as introduced in Section 2-4. It may be typical that the former negatively impacts the structural representability of CG models, whereas the latter limits the thermodynamic representability and transferability of CG models. A summary of these issues in CG force fields is presented in Figure 1.

3-2.A. Beyond Pairwise Basis Sets. Due to their pairwise approximate nature, CG models constructed from eq 16 are often unable to reproduce the many-body correlations from the FG reference systems. This problem is exacerbated by the isotropic nature of CG particles (sites) that is commonly adopted upon the CG mapping. By instead introducing virtual sites that are designed to capture such correlations, CG models can be improved while keeping the computational benefit of pairwise basis sets. For example, one can introduce virtual sites to represent complex chemical environments, e.g., the hydration layer surrounding lipid bilayers, where the virtual site interactions can be determined from a hybrid framework that combines structure-based methods and force-based variational principles.[89] These choices are often more favorable than introducing nonisotropic descriptions for CG particles, e.g., Gay−Berne interactions,[220] due to the complexity and computational cost of both the CG parametrization and CG simulation. Thus, despite some preliminary efforts in nonisotropic CG particle representation[221−227] and parametrization strategy,[228] we will focus on efforts to improve CG models using isotropic CG mapping representations in this section.

Alternatively, based on the many-body expansion,[229] an improvement can be achieved by introducing higher-order interaction terms in the CG Hamiltonian

$$U_{\mathrm{nb}}^{(n)}(\mathbf{R}^N) = \sum_I \sum_{J\neq I} U_{\mathrm{nb}}^{(2)}(R_{IJ}) + \sum_I \sum_{J\neq I} \sum_{K>J} U_{\mathrm{nb}}^{(3)}(\theta_{IJK}, R_{IJ}, R_{IK}) + \sum_{IJKL} U_{\mathrm{nb}}^{(4)}(\theta_1, \theta_2, \psi) + \dots \tag{17}$$

where $U_{\mathrm{nb}}^{(n)}(\cdot)$ denotes the n-body nonbonded interaction with configurational variables $(\cdot)$. For example, studies of bottom-up 1-site CG water have shown that pairwise interactions are incapable of recapitulating the local structure due to hydrogen bonds but can be recovered with the explicit addition of three-body interactions $U_{\mathrm{nb}}^{(3)}$.[147] The importance of three-body interactions in molecular systems[230−236] can be seen by applications of Stillinger−Weber interaction-based models,[237] including in the top-down CG water model, e.g., mW.[238]

In principle, the aforementioned CG methodologies can be readily applied to determine the interaction parameters for higher-order Hamiltonians, e.g., MS-CG[147,150] or IMC[239] for three-body Hamiltonians.[240] However, introducing many-body interactions inevitably reduces the efficiency gains from CG modeling. In contrast, recent developments have proposed two different generalized approaches to include many-body interactions at a reduced computational cost. The first approach[193,194] implicitly projects the many-body interaction (up to N-body) onto lower-order basis sets based on the conditional probability $p(\boldsymbol{O}_I^{(n)} \mid R_{IJ})$ of higher-order configurational variables $\boldsymbol{O}_I^{(n)}$

$$U_{\mathrm{nb}}^{(N)}(\mathbf{R}^N) = \sum_I \sum_{J\neq I} \left\{ U_{\mathrm{nb}}^{(2)}(R_{IJ}) + \sum_{n=3}^{N} C_n \int d\boldsymbol{O}_I^{(n)} p(\boldsymbol{O}_I^{(n)} | R_{IJ}) U_{\mathrm{nb}}^{(n)}(\boldsymbol{O}_I^{(n)}, R_{IJ}) \right\} \tag{18}$$

which can be interpreted as an extended Bogoliubov-Born−Green-Kirkwood-Yvon (BBGKY) hierarchy for the configurational interactions.[241−245] Recently, eq 18 was applied to water, where the three-body interactions are projected onto effective pairwise interactions, resulting in the Bottom-up Many-body Projected Water (BUMPer) model.[193,194] The importance of the many-body nature of interactions in water and aqueous systems has been also investigated at the atomistic level, e.g., by the MB-pol potential,[246−248] or ML approaches.[249] Notably, BUMPer also highlights the computational efficiency of such approaches while still recapitulating higher-order correlations.

The alternative approach can be realized by explicitly introducing higher-order order parameters (or collective variables) into the CG Hamiltonian. Explicitly evaluating arbitrary higher-order order parameters will often reduce simulation efficiency, but several order parameters that can be computed using pairwise statistics can account for many-body correlations with reduced computational cost. Inspired by many-body Dissipative Particle Dynamics (DPD),[250,251] there have been active efforts to combine local number density ρ-dependent interactions with conventional pairwise interactions in order to improve the CG Hamiltonian,[252,253] for example,

$$U_{\mathrm{nb}}^{(n)}(\mathbf{R}^N) = \sum_I \sum_{J\neq I} U_{\mathrm{nb}}^{(2)}(R_{IJ}) + \sum_I U_{\mathrm{nb}}^{(\rho)}(\rho_I) \tag{19}$$

Due to notable improvements in modeling implicit solvents[254−256] and material systems,[257−260] several papers have reported that utilizing the local density enhances the structural representability of CG models for phase separating or interfacial systems.[261−266] Taking a step further, one could incorporate variations in the local density in terms of a local density gradient $\sum_I U_{\mathrm{nb}}^{(\nabla\rho)}(\nabla_I\rho_I)$ into eq 19 to accurately describe inhomogeneous systems.[267]

Nevertheless, a direct advantage of eq 19 is that one can introduce any kind of order parameter that can be computed in a pairwise manner, other than local density ρ (e.g., particle orientations,[268,269] as well as order parameters related to liquid crystals[270] and glass transitions[271,272]) to accurately describe the corresponding correlations. However, the choice of order parameters used is often determined phenomenologically, and

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

thus, a systematic principle should instead be established. We will provide a systematic framework in Section 4. Finally, more complex, high-dimensional N-body order parameters are generally difficult to regress; ML techniques can help solve such nonlinear problems (see Section 6).

3-2.B. Transferable Coarse-Grained Force Fields. In the context of CG equations of motion, conjugate forces may additionally contribute to observables (on top of forces due to particle configurations). Since the effective CG interactions obeying thermodynamic consistency are free energy variables,[31−33] this point of view provides physical insights that introduce conjugate forces in terms of thermodynamic variables.

Early attempts to address temperature transferability were based on heuristics, such as temperature rescaling $U_{CG}(R, T) \approx U_{CG}(R, T_0) \times \sqrt{T/T_0}$ for polymers.[273,274] Recently, more thermodynamically consistent approaches have been pursued using entropy-enthalpy decomposition (eq 15) under the pairwise approximation: $U_{CG}(R) = E_{CG}(R) - TS_{CG}(R)$. By estimating the pairwise thermodynamic functionals ($E_{CG}(R)$, $S_{CG}(R)$) and extrapolating to nonparameterized temperatures, temperature transferability can be achieved in both constant NVT (Helmholtz free energy) and NPT (Gibbs free energy) conditions.[51,121,148,275] Even though eq 15 is a natural extension of the free energy to CG PMFs, the physical meaning of the pairwise thermodynamic functionals $E_{CG}(R)$, $S_{CG}(R)$ remains relatively unclear at present.[276] A recent report elucidated that these functionals may be deeply connected to the thermodynamic representability issue for entropy.[51,276] Alternatively, a more direct approach based on statistical mechanics that does not suffer from the ambiguity of pairwise thermodynamic functionals was developed by numerically transferring the phase space expectation value at different temperatures.[145] It should be noted that reweighting approaches often suffer from inefficient sampling and non-negligible numerical noise.[277,278] Notably, the dual approach by Lebold and Noid circumvented this limitation in reweighting without sampling other temperatures by employing the least-squares minimization to the energy quantity to obtain $E_{CG}(R)$, which is analogous to force-matching.[195,196] Recently, Pretti and Shell showed that the CG models constructed from microcanonical basis sets in conjunction with REM for capturing entropy functions can naturally provide temperature transferable CG models by recapitulating atomistic energy fluctuations.[41] Altogether, current findings and reports, regardless of methodological details, emphasize the role of entropy in temperature transferable CG models.

Temperature transferability is inevitably coupled with pressure transferability in the case of the constant NPT ensemble, as both thermodynamic variables affect the system volume.[146,279] Early improvements were based on rescaling CG interactions with respect to pressure but lacked theoretical rigor. Notably, one can introduce a volume-dependent conjugate force into the CG Hamiltonian, a strategy that dates back to the 1970s for liquid metals.[280] Das and Andersen showed that introducing volume-dependent interactions $U_V(V)$ to the CG Hamiltonian can correct the virial pressures in CG models[146]

$$H(\mathbf{R}^N, \mathbf{P}^N, V) := \sum_I^N \frac{\mathbf{P}_I^2}{2M_I} + U(\mathbf{R}^N) + U_V(V) \tag{20}$$

This interpretation has been further explored in refs 281−283 to impart transferable CG models while adequately addressing the representability issue. Other than explicitly relying on eq 15 or 20, it is also possible to correctly reflect the changes in CG interactions using order parameters that are coupled to system conditions. One notable example is the Ultra-Coarse-Graining (UCG) approach, which will be discussed in Section 4.

Transferability across composition is a more complicated issue, where both temperature and pressure transferability come into play. The ultimate goal would be to correctly address the reduced pressure in mixture conditions, which is a nonlinear process due to nonideal interactions, and to design the cross-interactions between different molecular moieties. While this direction has not been actively pursued due to its complexity, several preliminary reports have paved the way for developing so-called mixing rules for cross-interactions, indicating that having a correct description of the CG thermodynamic quantities is essential for achieving such transferability.[51,284] Another free energy-based direction would be to introduce a chemical potential-like term $\Delta\mu(R, N)$ as an analog to the chemical potential term in the Helmholtz free energy. Eventually, addressing the aforementioned transferability issues will elucidate how to achieve chemical transferability where the CG Hamiltonian can be determined *a priori* by grouping atoms into molecular building blocks (or functional groups) and sampling these groups at various configurations and state variables.[198,285−287] For example, an extended ensemble approach by Mullinax and Noid suggested the use of topology in composition and chemical transferability.[285] Combined with ML techniques, such a systematic treatment to predict CG interactions for new molecules is deemed possible by sufficiently sampling numerous small molecules.[287]

3-2.C. Holy Grail for a Bottom-up Coarse-Grained Force Field. An accurate CG Hamiltonian should be able to account for both many-body correlations and transferability issues. On the basis of the efforts described above, we argue that the ideal effective CG Hamiltonian may be expressed as

$$U_{CG}^*(\mathbf{R}^N, T, V, N) \approx U_{nb}^{(n)}(\mathbf{R}^N) + U_{temp}(\mathbf{R}^N, T) + U_{press}(\mathbf{R}^N, V) + U_{chem}(\mathbf{R}^N, N) \tag{21}$$

In eq 21, $U_{nb}^{(n)}(\mathbf{R}^N)$ accounts for the nonbonded interactions up to n-body, which can be explicitly cast based on eq 17 or implicitly described by eq 18. Alternatively, using an appropriate n-body order parameter O_I under the pairwise approximation, this term would be $U_{nb}^{(n)}(\mathbf{R}^N) \approx \sum_{IJ} U_{nb}^{(2)}(R_{IJ}) + \sum_I U_{nb}^{(n)}(O_I)$ and shares similar mesoscopic physics as the many-body DPD method.[250,251] The last three terms in eq 21 correspond to temperature-dependent, volume-dependent, and composition-dependent potentials through a conjugate interaction to thermodynamic observables. The first term $U_{nb}^{(n)}(\mathbf{R}^N)$ is independent of state point, while the latter three terms $U_{temp}(\mathbf{R}^N, T)$, $U_{press}(\mathbf{R}^N, V)$, and $U_{chem}(\mathbf{R}^N, N)$ are examples of state-dependent potentials. Both classes are examples of efforts to improve the expressivity of $U_{CG}^*(\mathbf{R}^N, T, V, N)$.

Alternatively, eq 21 can be interpreted as a projected Helmholtz free energy functional along pairwise basis sets. Note that $\Delta A(V, T, N) = -P\Delta V - S\Delta T + \sum_i \mu_i \Delta N_i$ and $\Delta A_{CG}(V, T, N) = \langle U_{CG}^*(\mathbf{R}^N, T, V, N)\rangle$. This interpretation is in line with the state-dependent potential derived from the free energy perspective.[288,289] Recently, some CG models have been developed based on the above principles and shown favorable results, such as the combination of density-dependent interactions with volume-dependent terms[290] or the UCG models in the mean-field ansatz.[275]

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

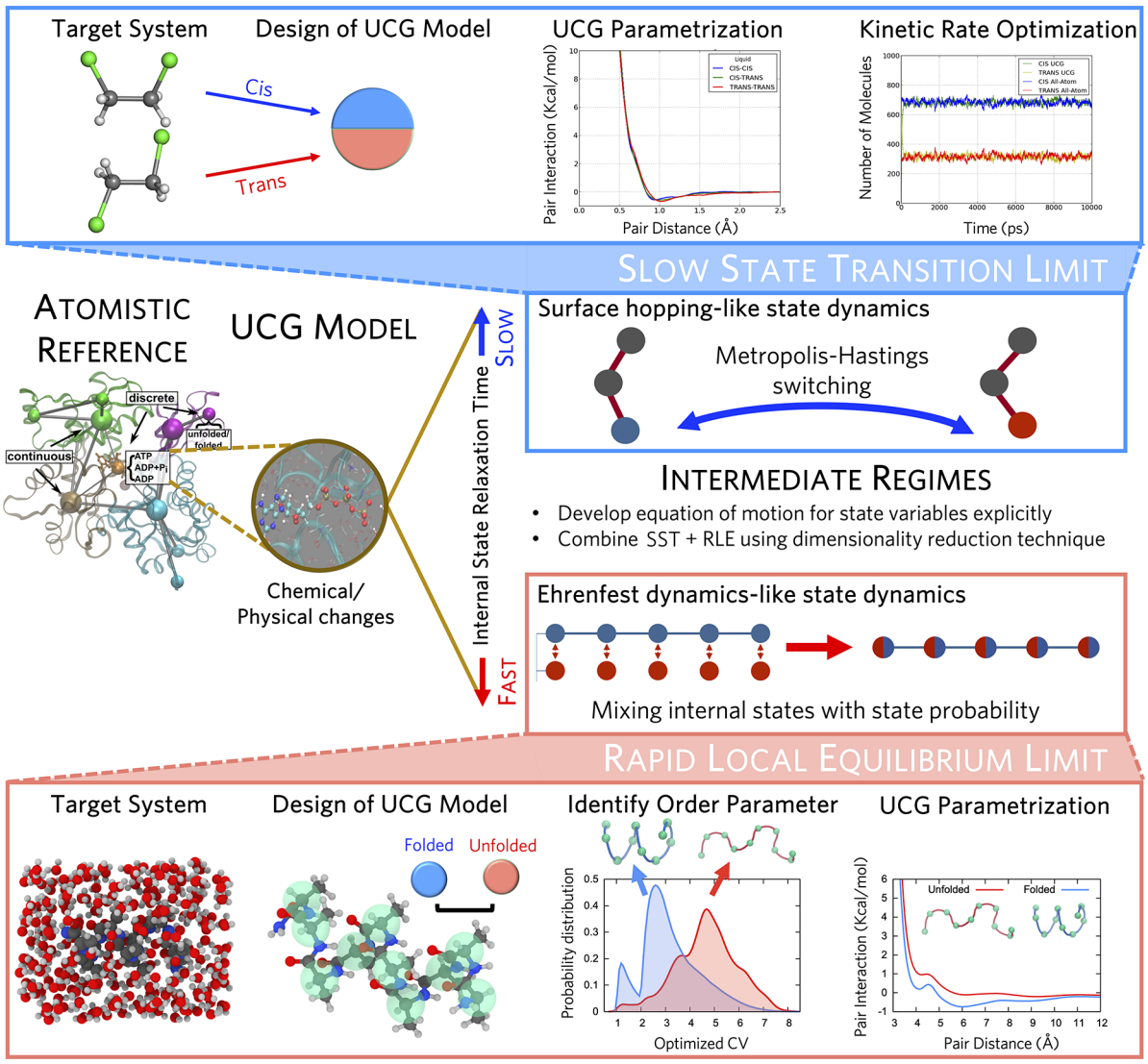


Figure 2. UCG models are designed to capture the chemical or physical changes "beneath" the CG resolution (illustrated upper middle left for ATP hydrolysis in F-actin). Practical design principles of UCG models are based on the relaxation time of internal state dynamics. (1) In the slowest limit (SST), UCG state dynamics can be treated as a kind of surface hopping. An example of this is the gauche- and anti- configurations from 1,2-dichloroethane (top panel). Based on the target system, distinct UCG states are identified, and the UCG models are built by parametrizing the state-wise interactions and optimizing the kinetic rates described by the Metropolis-Hastings algorithm. (2) The internal states at the fastest switching limit (RLE) can be thought to be in quasi-equilibrium, and the Ehrenfest dynamics idea can describe the internal states by mixing them with the state probability. The UCG models are then constructed by identifying the rapidly varying states with the corresponding order parameters. Then, bottom-up CG methodologies can be applied to determine the UCG state-wise interactions. As an example, we depict the solvated peptide here exhibiting folded and unfolded states determined by the optimal CV (bottom panel).

While eq 21 aims to directly determine conjugate force components in free energy expressions, an indirect approach based on the free energy is also possible by applying the perturbations to the conjugate variables of interest in the coarse-graining process.[284] By deliberately extending the simulation ensemble, matching a response of the system to perturbations has recently shown to sample a broad range of parameters. This wide parameter space with more abundant information allows for determining the most transferable CG interactions by maximizing an information metric, e.g., Fisher information.[291] In the future, it would be informative to rigorously elucidate the physical nature of these pairwise thermodynamic functionals and their relationship to CG systems at different thermodynamic state points and ensemble conditions.

3-3. Mini Outlook. As researchers continue to explore different avenues to increase the expressivity of U_{CG}, it is important to remain mindful of the trade-offs between model complexity and computational cost, the latter of which includes the cost of parametrization, implementation, and runtime. The development of effective CG interactions bears resemblance to "Jacob's Ladder" as seen in density functional approximations across quantum chemistry.[292] Each degree of complexity can be thought of as a new "rung" on the "ladder" that represents the field of CG modeling. We note, however, that climbing successive rungs does not necessarily guarantee improvement in the overall accuracy of a given CG model. For instance, it is possible that the increased complexity enhances the adherence to CG model consistency at the expense of CG thermodynamic representability and transferability.[156] To avoid these pitfalls, a

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

standardized method to derive, explore, and validate the benefit of increasingly complex basis sets is still needed. Our focus in the next two sections will be to discuss frameworks that can systematically increase the expressivity of CG force fields using theoretical and algorithmic procedures.

4. ULTRA-COARSE-GRAINING: MACHINERY TO GENERALIZE REPRESENTABILITY AND TRANSFERABILITY

4-1. Necessity of Generalized Coarse-Grained Framework. In Section 3, we introduced several challenging problems that are intrinsic to the parametrization nature of CG interactions. For different systems, conventional approaches introduced earlier have tackled these problems by altering the CG Hamiltonian form or parametrization strategy case-by-case, which limits general applicability. In order to attain a more generalized approach, the Ultra-Coarse-Grained (UCG) model and methodology were developed from the observation that these problems originate from the inability of conventional CG models to correctly address underlying chemical or physical changes in the reference system.[263,293,294] The idea underpinning the UCG approaches is to introduce internal quantum-like "states" into the CG sites, and thus, the CG Hamiltonian can effectively account for the driving forces associated with these discrete changes that are missing in conventional CG force field treatments by modulating the internal state interactions (called the "state-wise" interactions). The basic idea of a UCG model is to utilize a kind of isomorphism with quantum mechanics, i.e., in the latter case the system nuclei can evolve on multiple potential energy functions, which in turn depend on the quantum mechanical state space of the system and its underlying dynamics, and this in turn is tied back to the evolution of the dynamics of the nuclei. "Simple" nonadiabatic dynamics with surface hopping between electronic states[295,296] is one example of this behavior, albeit there can be other examples. The basic notion of the UCG approach is to utilize this quantum isomorphism to increase the expressivity of the CG model so that the influences of the processes that become implicit at the CG level are still included to a certain degree.

The importance of the underlying molecular nature beneath the CG resolution is pronounced in many systems, especially for highly coarsened representations. As indicated by refs 297 and 298 for ATP hydrolysis in actin protein, the missing important molecular details beneath the CG models can affect the free energy landscape and should be incorporated into the pertinent UCG models.[299,300]

4-2. Ultra-Coarse-Grained State Dynamics: Practical Realization. A UCG idea can be mathematically formulated by introducing internal state variables into CG interactions but in principle requires rigorous formulation and parametrization of the correct equation of motion for the state variables, which can significantly increase computational cost.[293] Since the chemical or physical changes in the molecular system have clearly separated relaxation times, a separation of time scales can be introduced. For example, physical changes such as bond breaking or formation happen much slower than the CG particle mass translations. On the other hand, chemical changes in electronic states or solvation will directly affect the CG system without any relaxation in state dynamics. Thus, instead of seeking a generalized Hamiltonian involving both CG configuration and state variable dynamics, two specific approximations at each time scale limit can be designed,[263,294] as described below. Figure 2 delineates a design principle for the UCG models in terms of state dynamics.

4-2.A. Slow State Transition Limit. Under the slow state transition (SST) limit, the internal states rarely change with the characteristic relaxation time due to conformational transitions that take place at time scales of nanoseconds or longer. A kinetic Metropolis−Hastings-like approach[294] can thus be utilized to approximate the instantaneous rate of switching from states α to β that are defined by the order parameter m (e.g., a dihedral angle ψ in a protein[299,300]).

$$K_{\alpha\to\beta} = k_{\alpha\to\beta}(m)\min\left[\frac{k_{\beta\to\alpha}(m)}{k_{\alpha\to\beta}(m)}\exp(-\beta(U_\beta - U_\alpha) - \epsilon_{\alpha\beta}),\ 1\right] \tag{22}$$

In eq 22, the prefactors $k_{\alpha\to\beta}$ and $k_{\beta\to\alpha}$ are determined from the FG simulations, experimental information, or they may be treated phenomenologically; and the model parameter $\epsilon_{\alpha\beta}$ is introduced to correctly capture the surface hopping-like dynamics from the given state-wise CG interactions U_α and U_β.

4-2.B. Rapid Local Equilibrium Limit. The opposite limit of the SST limit is when internal states undergo rapid state transitions or the so-called rapid local equilibrium (RLE). In this case, in an analogy to Ehrenfest dynamics, the effective UCG Hamiltonian can be expressed as a mixed interaction with respect to its state probabilities that follow quasi-equilibrium distributions. Ideally, one should consider the overall s^N configurations, where s is the number of internal states per CG particle, and N represents the number of CG particles. However, under the RLE limit, one can choose a local order parameter to decouple state correlations between different particles, resulting in the expression

$$U_{\mathrm{mix}}^{\mathrm{UCG}}(\mathbf{R}^{\mathrm{N}}) = \sum_{I,J}\sum_{s_I,s_J} p(s_I|\mathbf{R}^{\mathrm{N}})p(s_J|\mathbf{R}^{\mathrm{N}})U_{s_I,s_J}^{(2)}(R_{IJ}) \tag{23}$$

where $p(s_I|\mathbf{R}^{\mathrm{N}})$ denotes the probability of CG site I being s_I, and $U_{s_I,s_J}^{(2)}(R_{IJ})$ is the state-wise interaction between s_I and s_J.[263,301] Unlike the SST limit, the RLE limit explicitly accounts for the driving forces (the term $-\nabla p(s_I|\mathbf{R}^{\mathrm{N}})$ in eq 23), and thus, UCG models should be carefully designed using the correct order parameters to extract the nonuniform physical or chemical nature from the system. In this regard, various order parameters suggested from the previous studies can be utilized: external fields, relative particle positions, coordination numbers, and local number densities.[263] Notably, the UCG force field from eq 23 can be generalized to other CG Hamiltonians, such as the multiconfigurational CG (MCCG)[302] and conformational surface hopping methods.[303,304]

It must be noted that the UCG approach has already achieved notable success in treating realistic and highly complex biomolecular systems, an example of which is the assembly of the HIV-1 virus capsid from the more than 1,000 copies of its capsid (CA) protein component.[81,84,305]

4-3. Structural Representability. *4-3.A. Chemical Accuracy.* Even though most computationally efficient CG models are not able to describe changes in the chemical nature, a key attribute of UCG models in the SST limit, by design, is the ability to explicitly model chemical transformations while avoiding the use of many-body force fields by introducing distinct states that represent different chemical phenomena. This advance has enabled UCG models to capture previously missing conformational transitions between gauche- and anti-

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

conformations in 1,2-dichloroethane[294] and the characterization of ATP hydrolysis and phosphate-release reactions in actin filaments.[299,300] In a coarser description (lower resolution), UCG models based largely on the SST limit have been applied to capture the dynamic self-assembly behavior originating from many-protein HIV viral capsids.[81,84,305] As such, when conformational transitions in complex biomolecules rarely occur, the SST limit can effectively embed the finely detailed chemical nature into a reduced level of representation.

4-3.B. Many-Body Correlations. Introducing flexibility into CG interaction forms enables the UCG models to improve the description of the many-body correlations necessary for more accurate CG models. For the SST limit, chemical reactions due to complex many-body correlations that involve many-particle nonbonded and electrostatic interactions can be faithfully folded into the UCG models. Namely, the SST limit implicitly encodes the many-body correlations into distinct state-dependent pairwise UCG potentials, which have a strong similarity to a "polarizable" CG model.[306]

Other types of many-body correlations in terms of nonbonded structures can be practically impossible to address using only pairwise basis sets. We note, as before, that pairwise MS-CG models attempt to capture two-body and three-body correlations concurrently by satisfying the YBG equation,[31] but three-body correlations and even higher-order correlations are often not well-captured in complex CG systems, e.g., water.[147,165,193,194] Even though the UCG models at the RLE limit are built upon the pairwise basis sets, these state-wise interactions are linked via local order parameters. By choosing these local order parameters properly, the resultant UCG model can faithfully capture complex many-body correlations. Based on the observation that the local density is an N-body property, several studies have reported that the local density-based UCG models can readily reproduce the many-body phenomena of interest. For example, the solvophobic association of hydrophobic solutes due to many-body correlations between solute molecules was captured with the UCG formalism as well as pair correlations and clustering behavior.[263] The local density can also be an important order parameter for distinguishing different phases in heterogeneous systems. This inhomogeneity is pronounced in interfacial systems, where the conventional CG models failed to properly describe two or more phases at their interfaces. Notably, UCG interface models can differentiate distinct characteristics emergent in the system, e.g., liquid and vapor states in liquid/vapor interfaces with well-reproduced liquid "slab" density profiles and structural correlations.[265] Extending beyond interfacial UCG models, it was recently demonstrated that the single UCG model is able to encompass the structural correlations emergent from distinct bulk phases, resulting in multiphase CG models.[275]

Since the local density of CG moieties reflects the various chemical and physical natures (e.g., changes in structures and electrostatics that may affect coordination), it is conceivable that the local density can be utilized as a generalized order parameter to indirectly represent the nonuniform nature of a given system. For example, it has been shown that effects of FG hydrogen-bonding interactions can be faithfully captured in a local density-based UCG model by differentiating the donor and acceptor states,[307] with the hydrogen-bonding not being explicitly resolved at the CG level.

4-4. Transferability. The main advantage of UCG modeling not only lies in being able to reproduce important correlations faithfully described by the order parameters but also in its flexible interaction form. This is especially true in the RLE limit, where the UCG Hamiltonian inherits transferability by design; several distinct state-wise interactions are folded into a single Hamiltonian form, allowing for transferring to different state conditions. This flexibility is akin to Ehrenfest dynamics[308] or the empirical valence bond theory,[309] where a systematic connection between these different theories has been recently demonstrated.[310]

First, the flexibility of eq 23 enables one to directly employ the MS-CG variational principles to parametrize many state-wise interactions that are distinguished by the imposed order parameter. It has been seen that the parametrized UCG interactions are comparable to bulk MS-CG interactions, confirming the transferability of UCG models. For example, state-wise interactions between denser states in liquid/vapor interfaces are equivalent to bulk liquid interactions, and a similar conclusion holds for liquid/liquid interfaces where the UCG interactions are transferred to liquid mixture systems with different compositions.[265]

Alternatively, one can introduce the already determined MS-CG interactions into the UCG framework and utilize adequate order parameters to distinguish each interaction. For example, one can utilize the local density in order to mix bulk interactions at high temperatures and low temperatures, resulting in temperature transferable UCG models. A recent study showed that such an approach has a direct link to the energy-entropy decomposition under the mean-field ansatz and further demonstrated that one can possibly design phase transferable UCG models by combining liquid and gas phase MS-CG interactions.[275] Since number density directly responds to the system condition, local density-based UCG models can be good candidates for properly describing the conjugate forces exerted by the thermodynamic variables of the system. For more complex systems exhibiting various conformations, determining the correct order parameter would be the most important step toward transferability; this will potentially benefit from advances in ML techniques that will be described in Section 6.

4-5. Mini Outlook. In contrast to conventional CG methodologies, UCG theory provides a generalized framework to greatly enhance the fidelity of CG models, and the practical realization of UCG state dynamics imparts a physical presence of the missing degrees of freedom in terms of internal states at the reduced CG resolution. The UCG method has recently been shown to be successful for describing various systems ranging from liquids to biomolecules, demonstrating its applicability and versatility. We conclude this section by providing open problems for future developments.

- *Intermediate State Dynamics.* Various chemical and physical processes may still exhibit intermediate state dynamics that does not fall into either the SSL or RLE limit. One illustrative example would be proteins that slowly fold but have rapidly changing behaviors due to their interactions with the environment or solvents. Even though one direct direction for modeling intermediate state dynamics would be to incorporate equations of motion for state variables and to develop the corresponding UCG Hamiltonian, we propose a hybrid UCG model as a more efficient alternative by simultaneously accounting for both the SST and RLE limits. For the hybrid UCG model, defining two distinct UCG types that undergo different dynamical limits can benefit from existing dimensionality reduction methods. For example,

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

time-lagged independent component analysis (TICA) can be utilized to extract the most rapidly varying and the slowest varying order parameters.[311] With the recent success of such dimensionality reduction techniques in Markov State Models (MSM) for CG dynamics modeling,[312−317] the combination of the UCG methodology with TICA is expected to further extend the range of dynamical transitions explicitly included in CG modeling.

- *Toward Large Biomolecules.* Despite recent developments in UCG models having been mostly focused on relatively simple liquids, the UCG methodology in principle can be extended to larger spatial scales, and MS-CG-based UCG models can be also extended to much larger systems by correctly differentiating the chemical environments using local density parameters. For example, the numerous conformations in protein folding result in complex energy landscapes, e.g., dodecaalanine,[318] and UCG models that have internal states designed based on the local Cα density could be expected to accurately capture the folded and unfolded states. Similarly, the effect of solvents in a solvated system can be faithfully modulated by solvent density-based UCG models, e.g., lipid bilayers.

On the other hand, for more complex biomolecules, effective large-scale UCG models can be built upon coarser descriptions of state-wise CG interactions. Since the UCG methodology does not specify any of the interaction forms but rather provides a systematic framework for embedding internal state information into CG models, combining several CG interactions from different methodologies with the UCG theory can correctly account for the state dynamics. Recently, the UCG models based on the fluctuation maximization with harmonic interactions were developed for the glutamine-binding protein and lactoferrin and were able to correctly describe the protein conformational transitions.[319] Relatedly, a network-based UCG model was reported to effectively assess the mechanical properties of microtubules.[320]

5. DYNAMICS OF THE COARSE-GRAINED MODELS

5-1. Limitations and Challenges. While the thermodynamic properties of CG models differ from their FG counterparts, thermodynamic representability can be systematically improved based on the FG and CG observable expressions, e.g., pressure and energy. However, dynamical properties pose completely different problems in comparison to those from thermodynamic properties.[52−54] In this section, we elaborate on such difficulties arising from the dynamics of CG models and showcase recent advances in this area.

In order to assess the relevant dynamical variables in CG representation, the Mori−Zwanzig projection operator formalism can be applied to the microscopic Hamiltonian dynamics at the FG resolution.[321−324] The idea behind the Mori−Zwanzig formalism is to project the relevant dynamic variables that are left in the CG systems, resulting in a generalized Langevin equation (GLE) form of the equation of motion.[325] The following integro-differential equation thus faithfully describes the dynamics of the CG model[326−329]

$$\dot{\mathbf{P}}_I \approx -\frac{1}{\beta}\frac{\partial}{\partial \mathbf{R}_I}\ln w(\mathbf{R}^N) - \beta\sum_{J=1}^{N}\int_0^t ds\langle[\delta\mathbf{F}_I^Q(t-s)]\cdot[\delta\mathbf{F}_J^Q(0)]\rangle\mathbf{V}_J(s) + \delta\mathbf{F}_I^Q(t) \tag{24}$$

which is composed of particle conservative force, frictional, and stochastic (random) forces, respectively. The notations in eq 24 are consistent with ref 91, where $w(\mathbf{R}^N)$ is a normalized partition function of the microscopic (renormalized) configurations at $\mathbf{R}^N$. See ref 330 for the detailed discussion of the approximation made in eq 24. Therefore, performing the CG simulations under Hamiltonian mechanics using only conservative forces often results in an accelerated CG dynamics due to the missing friction. Note that friction and fluctuations are connected through the second fluctuation dissipation theorem.[91,331] These fast CG dynamics might be considered advantageous for performing CG simulations, allowing for simulations that span large temporal scales with relatively small time steps. Nevertheless, correct dynamical information from CG simulations is required to evaluate the dynamical properties when making contact with experimental kinetics.

From our perspective, two different approaches may elucidate the correct CG dynamics. The first approach would be to reconstruct the dissipation and fluctuation information from eq 24, such that the velocity correlations, diffusion features, and nonequilibrium properties of the system can be well reproduced. Due to complexity in parametrizing the correct friction and fluctuation terms, an alternative approach can be built upon establishing the correspondence between FG and CG dynamics by analyzing the fast CG dynamics when using Hamiltonian mechanics with the conservative forces along. In this section, we briefly review recent advances in both of these directions. More detailed perspectives for each approach can be found in refs 332 and 333 for molecular CG modeling at equilibrium and in ref 334 for out of equilibrium conditions. Furthermore, ref 335 provides a general review of CG modeling for both in and out of equilibrium conditions.

5-2. Incorporating Missing Friction. In order to faithfully represent the frictional and stochastic forces at the reference level, continued attention has been paid to parametrizing the frictional and stochastic forces from the Mori−Zwanzig equation of motion in eq 24.[54,91−93,105,107,327,336−342] Since the Mori−Zwanzig formalism cannot be directly employed in practical simulations due to its complexity and large computational cost, various approximations have been introduced in the literature to simplify the nature of the CG dynamics, resulting in various types of stochastic differential equations to describe the time propagation of the CG system. The simplest approach one can take is to parametrize the friction coefficient as described by the Langevin equation in which the time correlations and frictional kernels are omitted. For relatively simple CG systems, such an approach has been shown to recapitulate the correct diffusion behavior from the FG level.[54] Nevertheless, the Langevin equation is a rough approximation of the complete dynamical behavior of CG models.

A more accurate description of frictional and stochastic forces present in the CG equation of motion can be established in two steps. First, one needs to choose an appropriate stochastic differential equation as the equation of motion. Then, based on the chosen equation of motion, one needs to parametrize the friction kernels and associated stochastic forces in a "bottom-up"

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

manner. Typically, the Mori−Zwanzig equation of motion is approximated using a single-particle GLE or DPD-like equation of motion. A single-particle GLE assumes that there are no spatial correlations for the random forces acting on different CG particles. Such an approximation reduces eq 24 into a more tractable form that allows for developing various parametrization methods,[105,341,343−345] but missing pairwise nature in such approximation violates macroscopic physical principles by not conserving the momentum. Conservation of momentum is particularly important for reproducing the long-time tail of velocity autocorrelations resulting from the hydrodynamic effect.[346] This discrepancy can be correctly addressed by introducing a fluid mechanical description. For example, the smoothed particle hydrodynamics (SPH)[347−349] and the smoothed DPD[350,351] based on the discretized Navier−Stokes equation can resolve the momentum conservation issue by introducing pairwise frictional kernels to eq 24, yet most fluid mechanics-based approaches often suffer from the top-down nature. Thus, one should carefully choose the appropriate physical descriptions at the desired resolution to embed into the CG model in order to reduce eq 24 into an approximate stochastic differential equation.

Once the form of the CG equation of motion is chosen, the remaining step is to determine the friction kernels to be consistent with the FG reference. Here, we provide a brief discussion on bottom-up approaches for parametrizing these nonconservative interactions. For bonded systems, e.g., star polymers, a pairwise decomposition of instantaneous forces into parallel and perpendicular directions at the FG level is possible at the CG resolution. Practically, Hijon and Español introduced the so-called "constraint dynamics" technique to extract the pair decomposed forces under the Markovian DPD equation of motion.[92] This was further extended to the non-Markovian DPD regime by Yoshimoto,[338] and in recent years, Karniadakis and co-workers have established a systematic parametrization of GLE and DPD equations of motion for both Markovian and non-Markovian limits.[93,337,352−354] In this case, the friction kernels are readily obtained from the stochastic forces, allowing direct utilization of the Mori−Zwanzig formalism. However, pairwise decomposition of instantaneous forces is only feasible for bonded systems, and thus, this approach cannot be applied to unbonded systems.

Alternatively, the friction kernel can be constructed indirectly by inverting the time correlation functions from the FG reference.[93,337,338,353] In practice, this inverse approach matches the pairwise velocity autocorrelation function and force−velocity cross-correlation function, resulting in the Volterra integral equation.[344] Still, most of these approaches are limited to bonded systems. Notably, a recent breakthrough for constructing friction kernels of unbonded fluids was reported using the dynamic mapping approach.[107] By estimating the instantaneous forces on dynamic blobs based on the velocity Verlet algorithm,[355] conservative forces in the form of the many-body DPD interactions[250,251] were determined using the MS-CG principle. The deconvolution of correlation functions was then applied to derive a DPD-like equation of motion in Markovian and non-Markovian limits. This approach further establishes the bottom-up link between the microscopic origins of fluids and macroscopic physics. Such a bottom-up inference of frictional and stochastic interactions has not been widely investigated due to its complexity but remains a promising direction for future research. For example, using the REM framework, Español and Zuñiga designed a variational approach to infer drift and diffusion terms in the Fokker−Planck equation.[356] Another interesting extension of the MS-CG method was developed by Davtyan, Andersen, and Voth, where they introduced fictitious particles and coupled them with CG sites to effectively introduce a memory kernel under the GLE,[339,340] which shares a similar physical idea with the auxiliary model later developed by Karniadakis and co-workers.[353]

5-3. Understanding Accelerated Hamiltonian Dynamics. Alternatively, a computationally less expensive yet challenging direction would be to perform CG simulations under Hamiltonian mechanics and then elucidate how the accelerated CG diffusion is related to the FG (reference) dynamics through various rescaling approaches.

5-3.A. Time Rescaling. A naïve yet straightforward approach is to think of the CG time scales as uniformly accelerated time with respect to the physical time of the FG reference. This uniform time rescaling approach assumes that the frictional forces are not dependent on time, and thus, the Mori−Zwanzig projection operator can remove the configuration, momentum, and time dependence in the friction kernel.[327] This approach has been reported for polymer systems at different resolutions and chain lengths but with limited applicability due to the strong assumptions made.[357−364] We note that even though one could assume a uniform scaling ratio and estimate such value naïvely based on the FG and CG diffusion, there is no theoretical guarantee that such a factor exists[365] and may necessitate an explicit consideration of the Mori−Zwanzig formalism.[366−369] Also, the uniform scalar friction term obtained from the scaling factor itself is a many-body quantity (renormalized memory kernel) and differs by system conditions, e.g., thermodynamic state point, and hinders its applicability to other chemical systems.

5-3.B. Free Energy Landscape. Inspired by the energy-landscape theory,[370,371] e.g., protein folding,[372] the free energy landscape approach aims to address the dynamical properties underlying barrier-crossing dynamics by correctly representing the CG energy landscape. While the barrier-crossing dynamics is quite different from the microscopic dynamics of the system, recent advances have elucidated the structural-kinetic-thermodynamic relationships for helix−coil transitions of helix-forming peptides.[373] More importantly, such approaches have been shown to have a direct link to the MSMs, where the system dynamics is represented by transitions between microstates.[374−377] Notably, based on variational approaches to understanding the conformational dynamics and then follow-up work,[378,379] Nüske et al. have developed a spectral matching method that targets the dynamical propagator of CG systems, resulting in correct long-time dynamics.[380] Outside of the MSM-based framework, Rudzsinki and Bereau have developed the Bayesian dynamical reweighting scheme[381] to correctly recapitulate the kinetics of CG peptides.[382] These pioneering efforts in barrier-cross dynamics have highlighted how existing observed deficiencies in CG dynamics and kinetics may be due to not only incorrect CG equations of motion but also inaccuracies in the approximation of the conservative forces (i.e., the many-body CG variable PMF).

5-3.C. Excess Entropy Scaling. Another important recent advance in CG dynamics has been an attempt to understand accelerated CG dynamics using the excess entropy scaling relationship. First proposed by Rosenfeld,[383−385] the excess entropy scaling relationship is an empirical, semiquantitative relationship that links the dynamic property of the system D^* to

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

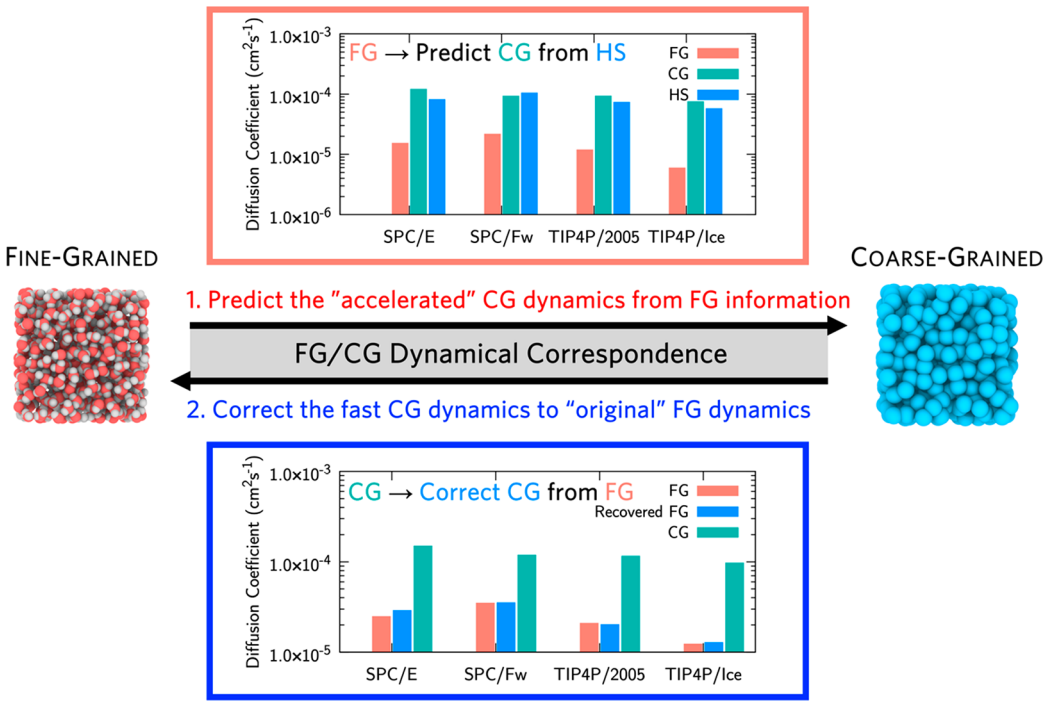


Figure 3. Summary of the excess entropy-based approach to achieve dynamical representability under Hamiltonian mechanics. The dynamical representability of CG models can be addressed by having a dynamical correspondence between the CG and FG models. In this case, without correct fluctuations, CG dynamics is spuriously accelerated compared to the reference FG dynamics (e.g., diffusion coefficients in this figure). Therefore, the ultimate goal in dynamical correspondence would be to address both directions across the FG and CG systems: (1) Predict the accelerated CG dynamics from the FG information and (2) correct the fast CG dynamics to match the original FG dynamics by observing the missing degrees of freedom upon the coarse-graining process. We depict the molecular liquids (water in this case) at the single-site CG resolution as an example, where the hard sphere (HS) mapping theory can achieve (1) and incorporating the missing rotational information can address (2) (recovered FG).

its molar excess entropy s_{ex}, which is the entropy difference between the system and ideal gas, such that

$$D^* = D_0 \exp(\alpha s_{ex}) \tag{25}$$

To date, only a handful of studies paid attention to the potential usefulness of eq 25 to assess CG dynamics in terms of entropy, e.g., the perspective from the REM,[39] and until recently, employing eq 25 for CG systems had not been extensively pursued. This is mainly because Rosenfeld scaling is not established from first-principle physics, limiting the applicability of this scaling relationship.[383,386−391] The empirical nature of this scaling relationship further exacerbates this limitation when applied to CG dynamics. First, it is not guaranteed that the FG and CG systems will obey the same scaling relationship, i.e., $\alpha_{FG} \stackrel{?}{=} \alpha_{CG}$. Also, the correspondence between D^*_{FG} and D^*_{CG} is still unclear because there is no physical explanation for the D_0 term, which is the "entropy-free" coefficient from eq 25.

Recent progress on the excess entropy scaling has addressed these aforementioned problems,[392−394] starting from the excess entropy difference between the FG and CG systems, known as the mapping entropy.[51] By computing the excess entropy based on earlier arguments from Karplus, Lazaridis, and Zielkiewicz,[395,396] ref 392 has confirmed that the universal scaling relationship will hold for the same molecular systems upon the coarse-graining process for fluids. In addition to unraveling the universality of the Rosenfeld scaling in CG models, it was recently demonstrated that D_0 at the single CG-site resolution can be physically understood from the hard sphere nature of CG models of liquids, resulting in the analytical form of D_0 determined by specific equations of state.[393] Such an approach requires an additional layer of coarsening of CG systems to describe them as dynamically consistent hard spheres, and it has been demonstrated that classical perturbation theory can determine the effective hard sphere by mapping the short-range repulsions (e.g., Barker−Henderson theory[397,398]) or long-wavelength fluctuations (e.g., Weeks−Chandler−Andersen theory[399−401] or fluctuation matching[393]). While the hard sphere treatment of CG systems determines the D_0^{CG}, the entropy-free diffusion coefficient for FG systems is usually larger than D_0^{CG}. This discrepancy can be understood from the degrees of freedom that are missing at the CG resolution.[394] By incorporating the missing rotations and vibrations back into the single CG-site resolution, the complete dynamic correspondence between the FG and CG systems can be recapitulated for liquids. In practice, then, one may (1) predict the accelerated CG diffusion under the Hamiltonian mechanics by estimating the D_0^{CG} or (2) recover the reference FG diffusion from the CG level by incorporating the missing diffusion into the translational CG diffusion. However, these recent findings have primarily focused on relatively simple CG systems of liquids, so considerable effort to extend such dynamic correspondence to nontrivial, complex CG systems, e.g., biomolecules, will be a challenge and should be pursued. A detailed description of the dynamical correspondence between the FG and CG systems using the excess entropy scaling is given in Figure 3.

5-4. Mini Outlook and Future Challenges. In order to correctly address dynamical representability in bottom-up CG models, recent studies have examined the dynamical properties of CG models using both GLE-like equations of motion and rescaled Hamiltonian mechanics. While the GLE description provides a rigorous statistical mechanical description of CG

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

dynamics, complexity and numerical stability are currently a bottleneck for application to complex molecular CG systems. In addition, similar to the conservative interactions, dissipative and stochastic forces present problems with transferability. The majority of attempts has only been tested on uniform single-component fluids, as multicomponent systems tend to face transferability issues. Also, in the same context as the transferability of the conservative forces (i.e., the many-body CG PMF), thermodynamic transferability of dynamic properties should be addressed to impart a high fidelity CG model. This particular area has not been extensively explored at the current stage, yet several preliminary directions have provided potential directions: dynamical rescaling,[366,367] energy renormalization,[402−406] and transfer learning using ML techniques.[407]

On the other hand, understanding accelerated CG dynamics produced using Hamiltonian mechanics can lessen complications from frictional interactions by introducing *ad hoc* physical scaling principles, e.g., excess entropy scaling. In this regard, rigorous physical scaling principles beyond the hard sphere description for explaining the excess entropy scaling are a promising area for future research. A few possible directions are based on the mode coupling theory[408,409] and transition state theory.[410] A grand goal in CG dynamics would be to correctly understand and faithfully reproduce transport phenomena under nonequilibrium conditions. Continuous development of a rigorous and practical bottom-up CG theory based on nonequilibrium statistical mechanics would therefore pave the way to a new era of CG modeling.

6. MACHINE LEARNING AND MOLECULAR COARSE-GRAINING

6-1. When Bottom-up CG Modeling Meets Machine Learning. ML is a subfield of artificial intelligence that uses algorithms to study and analyze data.[411] For example, in order to design an automated method that can determine the content of a picture by inspecting its pixels, e.g., whether the image contains a cat or a dog, two strategies could be considered. First, an algorithm could be explicitly programmed to analyze each pixel in the image to determine whether the picture is more similar to a cat or a dog—an admittedly difficult task. On the other hand, one could use ML procedures to extract patterns from a large set of images that are already labeled as containing a cat or dog. This ML-based approach would then use this labeled data to learn the connection between individual pixels and the overall content of the image, resulting in a new algorithm that is able to discern overall picture content. In other words, ML approaches can leverage high volumes of data to perform tasks that seem intractable when using other strategies. These techniques, however, naturally require a large amount of data to succeed and produce solutions whose quality and accuracy fundamentally depend on the quality of the data used for parametrization. Furthermore, the produced solutions may be opaque and extrapolate poorly to cases outside the data used for parametrization, creating a natural barrier to transferability. Nevertheless, the ability to create complex algorithms from data has revolutionized a number of areas such as computer vision and advertisement targeting.[412]

Since bottom-up CG approaches focus on learning patterns from data sets generated from FG simulations, it should come as no surprise that algorithms developed in ML have found use in molecular CG modeling. Given the similarities between CG and FG models, a natural extension is to employ atomistic ML techniques for CG systems. Our particular interest in this subsection, however, is to describe applications of ML to CG modeling with a focus on how these methods differ from ML used in the atomistic setting. For a thorough survey of atomistic (as well as some CG) ML methods, we refer readers to recent reviews.[60,413−418]

6-2. Machine Learned CG Force fields. *6-2.A. Machine Learning Design Principles.* In the language typical to ML, the MD-based CG and atomistic models discussed in this article fall under a class of methods referred to as energy-based models (EBMs).[419,420] These EBMs are algorithms that describe a distribution by specifying its probability up to an unknown normalizing constant (CG and atomistic force fields are naturally related to the "energy" term of EBMs). However, the data available for training and the need for scientific extrapolation create a specialized domain with its own approaches and difficulties: similar to the atomistic setting, the availability of a noisy estimate of the forces of the many-body PMFs provides an important avenue for parametrization not available in most EBM applications, and the need for physical transferability impedes the use of many EBM architectures. These two key details underpin the training and design of force fields. Despite these differences, various function representation approaches have been successfully adapted from ML to increase the flexibility and accuracy of atomistic force fields for over two decades,[411] and many of these same methods have now been transferred to the CG resolution. Recent reviews have detailed several ML-based force field approaches, including several kernel-based methods such as the Gaussian Approximation Potential by John and Csányi,[421] the method introduced by Scherer et al. that uses Gaussian process regression projected onto tabulated potentials,[422] and Gradient Domain ML-based methods as described by Chmiela et al.[423] and Wang et al.[424]

A growing number of CG neural network approaches have also been introduced for estimating the many-body CG variable PMFs, likely beginning with Lemke and Peter who developed a convolutional neural network-based approach to learn corrections to an existing CG force field through ideas connected with noise contrastive estimation and adversarial learning.[57,425,426] We also note that publications such as Schneider et al.[427] similarly proposed using neural networks to capture free energy surfaces but did not do so for a high-dimensional particle representation. The Deep CG Potential (DeePCG), introduced by Zhang et al.,[428] was the first to adapt more traditional ideas from the atomistic force field community,[429] followed by CGnet.[59] We note that while typically used with highly general feature sets, these same approaches could also be applied to custom high-dimensional order parameters as discussed in Section 3-2. Wang and Bombarelli subsequently introduced an autoencoder augmented approach.[424] Traditional autoencoders are commonly used in data compression as they temporarily reduce the feature space within the network.[430−432] This bottleneck feature allows the neural network to find a CG mapping operator before force-matching.[433] In addition, CGSchNet developed by Husic et al.[434] and the recent work by Ruza et al.[435] have introduced graph neural network-based methods to CG force field development. These approaches can be viewed as combining multilayer perceptron-based approaches (e.g., CGnet), which produce a CG force field using user-supplied featurization, with SchNet,[436] a graph neural network architecture originally designed to reproduce *ab initio* forces and energies, and a compatible simulation engine, TorchMD.[437] This use of graph neural networks requires no molecular featurization to be supplied by the user, as the

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

network learns its own features via its graph subnetwork. This also has the added benefit of ideally making CGSchNet transferable across system composition and accurately modeling the solvation environment for biomolecules in a novel manner with ISSNet.[438] We note that graph neural network architectures[439] and graph-based approaches[440] have additionally found widespread use in other molecular tasks such as the selection of CG mapping operators and automatic sampling of atomistic configurations.[441,442] The CGnet architecture has also been adapted to only consider many-body interactions up to a specified order. For example, it is shown that even 5-body interactions notably improve the quality of the resulting CG model when studying a small protein.[443]

6-2.B. Parametrization. When parametrizing an atomistic model to match *ab initio* approaches, the force field is often trained via regression to reproduce the connection between molecular configurations and the energies and forces present in the reference data set. While the initial data set may often be generated via MD using a reference energy function, this is not required; in fact, configurations outside the stable basins of the reference system are often critical for reproducing the barriers fundamental to chemical behavior.[416,417,444−448] These additional structures are either included through human intervention or actively added to the reference set through a variety of strategies, one of which is of particular importance to the current discussion when considering neural networks: query by committee (QBC).[417] In QBC, multiple neural networks are trained with varying initial optimization conditions on a given reference data set.[449] While each of these neural networks can typically reproduce energies on the reference data set, their predictions outside this domain may differ. This disagreement is then used as an engine for adding new structures to the reference data set: If a candidate structure results in disagreement among the various neural networks in the committee, a reference *ab initio* calculation is performed, and the configuration is used for parametrization. While QBC is an important technique, its construction emphasizes how neural networks trained on finite samples (especially those drawn from the reference Boltzmann distribution, as is often used for the initial *ab initio* data set) often exhibit inaccuracy outside the stable basins of the reference system—thereby highlighting the importance of complex sampling for high-dimensional force field parametrization.

While this ability to add particular configurations (along with energies and forces) to a data set has proven to be critical to the creation of high-dimensional atomistic force fields, this route has not typically been pursued in the creation of CG force fields. In order to understand the barriers to doing so, it is helpful to compare the settings in which CG force fields are parametrized in comparison to their atomistic counterparts. The atomistic configurational energy corresponds to the many-body PMF in terms of the CG variables, which is extremely difficult to evaluate in practice. However, for the force-based parametrization strategy given in Section 2-3, it is often straightforward to create a noisy version of the forces given by the many-body PMF using the forces in an atomistic trajectory. This noisy signal is compatible with techniques such as least-squares regression, partially reproducing the setting typical to atomistic modeling. However, it is unclear how to add a single CG configuration to a force-based reference data set unless conditional sampling is used, as forces from a single configuration have no clear connection to those of the many-body PMF—only the conditional mean over all such configurations weighted according to the Boltzmann distribution does. The modeler is thus often forced to either select the corresponding atomistic configuration from said conditional distribution by running constrained atomistic MD or to use such constrained MD to directly provide noiseless forces at arbitrary CG configurations, e.g., using blue moon sampling.[421,450] Doing so creates an avenue to use active sampling strategies such as those described in the previous paragraph.

However, often due to computational reasons, most studies instead resort to canonically distributed nonconstrained atomistic reference samples for parametrization (the use of non-Boltzmann distributed data sets to parametrize CG models has been performed recently[421,451] but does not seem to be commonplace). This lack of non-Boltzmann sampling, when combined with the similarity between the model architectures used for atomistic and CG force fields and the nature of active sampling strategies, seems to imply that current highly flexible CG potentials trained using forces may persistently face difficulties outside the stable basins of the system under study. The situation for CG models may indeed be sometimes worse than the atomistic case as implied by the previously mentioned results of Wang et al.,[443] where high order CG potential terms were required to reproduce atomistic results, implying that low capacity model representations may often be insufficient. Despite this, however, the preliminary success of the results discussed in earlier subsections provides hope for quantitatively accurate CG force fields using ML-based algorithms.

Additional strategies beyond force-matching have been developed (see Section 2) to parametrize CG models. These approaches generally aim to either reproduce a particular correlation (low-dimensional marginal distribution) or a high-dimensional distribution described by a reference atomistic trajectory mapped to the CG resolution.[168,169,452,453] While not the focus of this Review, we note that the direct inversion of radial distribution functions to pair potentials using neural networks has shown to be of repeated interest and represents a route to parametrization requiring no additional reference data once successful CG methods are established for similar systems. Since the forces present in atomistic trajectories are not referenced in these approaches, they are sometimes applicable to general EBMs when the model architectures are compatible. For example, the optimization procedure underpinning REM in the CG literature closely matches the maximum likelihood training of EBMs,[420,454] although, in the CG case, the supporting theory[35] has a stronger multiresolution focus. The use of classification to differentiate between data produced by a candidate CG force field and that of the atomistic reference has led to both additive updates reminiscent of noise contrastive estimation[57,425] as well as the adversarial strategies imitating those found in Generative Adversarial Networks (GANs).[58,426] Despite these initial connections, however, explicit cross-pollination between these fields remains sparse and is an area for future development.

6-3. Machine Learning-based Analysis of CG Models.

6-3.A. Recent Advances in Machine Learning. CG simulations, similar to their atomistic counterparts, create large amounts of data, and transforming this data into knowledge and understanding is of utmost importance and yet a difficult task. As mentioned in Section 6-1, a large number of ML-based techniques have been developed to understand the data produced by atomistic simulations and are generally also applicable at the CG resolution.[60,413−417] We refer interested readers to the aforementioned reviews. In this section, we instead focus on two novel applications of ML that focus on

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

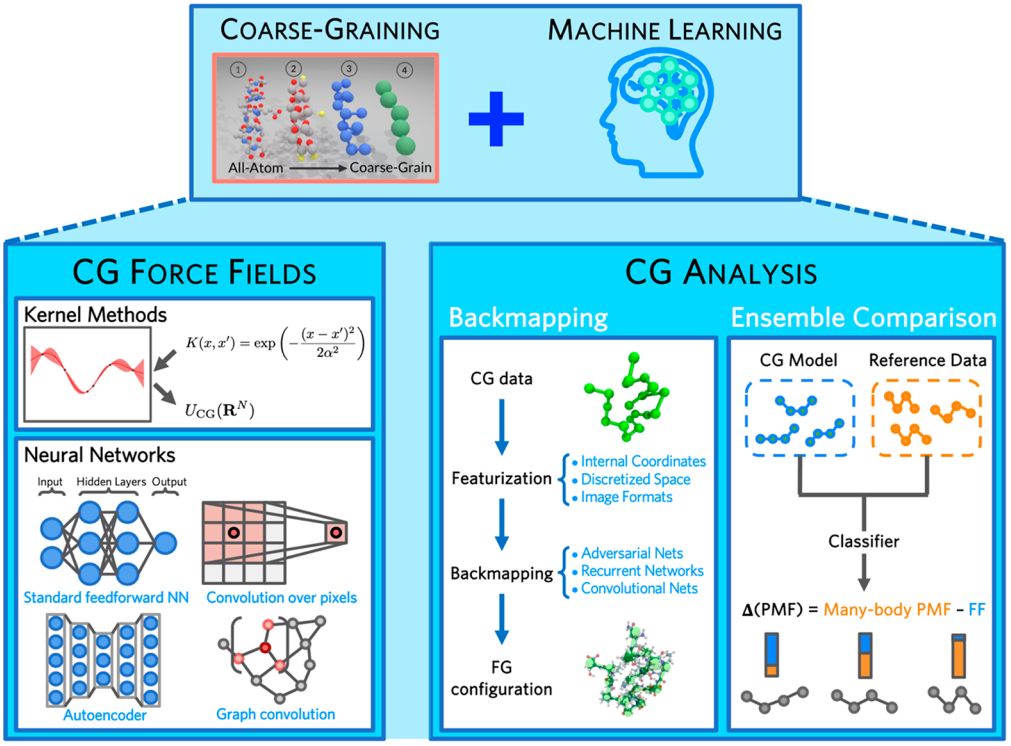


Figure 4. A summary of machine learning methods related to CG modeling. Kernel methods utilize a covariance function to specify a random process which can be used as a nonlinear estimator for CG potentials (left panel, top). Artificial neural networks (ANNs) can also be used to generate nonlinear force fields. A variety of ANN architectures have been developed with applications in CG modeling, such as autoencoders which compress the data stream before expanding it, which has natural connections to CG methods, and graph convolutions, which apply filters over graphs such as molecular topologies (left panel, bottom). Machine learning has also found uses in analyzing CG models and trajectories. Neural networks can be used to backmap CG configurations (right panel, left). Classifiers, especially interpretable ones, can be used to differentiate between configurations generated by different models (such as a CG model and the reference data it was parametrized from) and also provide specific information about how the ensembles are different (right panel, right).

problems common to CG simulation: the lack of atomistic detail and the inability to identify high-dimensional structural representability issues.

6-3.B. Sample Generation and Reconstruction. CG models are more efficient than atomistic simulations due to their reduced resolution, yet as posed earlier, this simplification inhibits understanding the molecular driving forces underpinning emergent behavior. Approaches that reintroduce atomistic details into CG configurations (aptly referred to as "backmapping" methods) provide strategies to take advantage of the efficiency of CG models while maintaining a clear atomistic picture. Several ML-based techniques have recently been developed to backmap CG data. Most of these approaches use generative models[455] or GANs in order to create samples at the atomistic resolution that are consistent with a given CG sample. GANs work by simultaneously training two neural networks which "compete" against each other.[426] One network (the "generator") is trained in the task of creating the backmapped configurations, while the other (the "adversary") is trained to classify structures as being generated either via atomistic molecular dynamics or the generator network. Over the course of training, the generator ideally improves to the point that a fully trained adversary can no longer distinguish between the two sources of samples. In this ideal case, the generator produces atomistic samples that are Boltzmann distributed conditioned on a given CG sample. In application, similar to ML-based CG force fields, the network could first be trained on a smaller system and then applied in a larger context, providing a way to atomistically interpret emergent behavior discovered in CG simulations.

There are currently multiple examples of GAN-based backmapping methods. One method introduced by Li et al. uses the Pix2Pix network architecture which was originally designed as a style transfer network for images.[456] By converting CG positions into a two-dimensional image, the method is able to perform backmapping with minimal additions to an already existing framework.[457] Another method by Stieffenhoffer, Wand, and Bereau utilizes a new network architecture designed specifically for the task of backmapping.[458] In this so-called deepBackmap approach, the network not only sees CG configurations but also has access to force field information, allowing higher energy structures to be directly penalized. The network also distinguishes between different chemical groups such as aromatic rings and backbone atoms and additionally builds the high-resolution structure one particle at a time, which allows information from previous atoms to inform the placement of later atoms. Recently, this method showed promising results for complicated systems such as polymer melts and can translate well to crystalline structures.[286] We note that such stochastic backmapping methods would also allow the force-based active learning strategies previously described to be performed at a minimal cost.

We also note that a number of ML methods other than GANs can be also employed to perform backmapping. To note, various backmapping methods have been developed by utilizing traditional supervised/unsupervised methods[459] (e.g., graph

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

methods and PCA) or advanced ML methodologies, including Bayesian inference,[460] Gaussian process regression,[461] and autoencoders.[433] In particular, the approach of directly generating molecular configurations without extensive MD has similarly been exploited by Boltzmann Generators, which are trained on small amounts of molecular dynamics data and the atomistic force field in an unconditional manner.[462] This technique, while not currently implemented for CG data, could similarly be adapted to the backmapping domain and would have the advantage of taking into account atomistic force field information in a rigorous manner.

6-3.C. Ensemble Comparison. Even though bottom-up CG models are rigorously built upon FG statistics, many CG models have structural representability issues. In addition, the high-dimensional parametrizations typical to bottom-up models can make it difficult to intuit how two approximate CG models differ beyond their qualitative or projected behavior. When parametrizing atomistic force fields, the per-configuration error in the energy or forces is known for each entry in the reference data set, and these known discrepancies can provide intuition on the regions of phase space well-described by the model at hand. Unfortunately, this type of analysis is not possible for bottom-up CG models, as neither the true many-body CG variable PMF nor its gradients are typically known (although, as noted previously, the gradients may be estimated by constrained simulation techniques). Instead, the validation of CG models is generally limited to comparing their performance using low-dimensional free energy surfaces (i.e., marginal distributions), leaving the differences in their full high-dimensional behavior relatively undescribed.

In this regard, recent work[463] has proposed a different strategy for capturing and describing the errors present in parametrized CG models. The output of a calibrated classifier trained to differentiate between configurations in the reference atomistic data set and those produced by a candidate CG model can be transformed into an estimate of the difference in the CG force field and the true many-body PMF at each configuration (and, if the classifier is differentiable, the difference in the forces at these configurations), variationally recapturing the information present when dealing with atomistic force fields. Furthermore, the applicability of classifiers in this domain opens the door to exploiting techniques from explainable ML in order to interpret the inferred configurational errors. These same ideas are also applicable to resolutions coarser than that of the CG force field and provide an avenue for understanding classification-based training techniques, such as those in Lemke and Peter's work as well as in adversarial-residual-CG (ARCG).[57,58]

We summarize the current advances in ML-based CG approaches in Figure 4.

6-4. Mini Outlook. Both bottom-up CG and ML use algorithms to discern patterns in data. This similarity has motivated the application of a variety of ML algorithms in CG frameworks, ranging from novel force field approaches to systematic methods for reintroducing atomistic detail. Collectively, this increased expressivity can produce atomistic explanations for previously uninvestigated phenomena. However, this increase in accuracy has come with a loss in model transparency (whether this trade-off is inevitable is a topic of debate).[464,465] In situations where extrapolation and physical intuition are critical, this trade-off can create problems. For example, while bottom-up ML CG models may recapitulate the ensemble it is parametrized to match, are the derived force field parameters transferable to the primary system of interest where emergent behavior is expected? Do backmapping approaches parametrized in the stable basins of a system extrapolate correctly to the transition regions? If a model is inaccurate in portions of phase space, can this be predicted and scientifically understood? As the parametrization of bottom-up methods is often based on reference data sets that inherently do not guarantee the performance of CG simulations, these questions are critical to the many ML-based approaches currently in development. These issues, of course, are not new to the bottom-up CG community (see Section 2) or the ML community.[466−471] Unfortunately, the nature of molecular models with high-dimensional parametrizations producing molecular configurations in a high-dimensional phase space makes these issues even more pertinent. Nevertheless, the initial successes mentioned in previous paragraphs, along with the successful creation of increasingly general purpose atomistic potentials,[472] motivate future development for all the applications described in this section. Once the transferability of an approach has been systematically established (ideally through a combination of transparency and application), further work can focus on computational efficiency and ease of use, ideally leading to a class of CG models that underpin a new wave of research devoted to previously insurmountable scientific problems. In general, one may also expect an explosion of new literature in the future in terms of ML methods applied to coarse-graining.

7. CONCLUDING REMARKS AND PERSPECTIVE

For the past two decades, bottom-up CG models have been developed by studying the microscopic origins underlying many macroscopic processes and have emerged as efficient, powerful, and multiscale computational tools in several fields of science. However, due to the enormous complexity of atomistic systems, statistical mechanics-driven CG modeling has primarily only been possible for relatively simple systems, e.g., liquids and small peptides, using various practical approximations from *ad hoc* design principles. To move toward true "multiscale" models, such bottom-up principles should maintain accuracy across different physical scales. Notably, some recent advances are pushing this limit forward to much smaller (quantum regime) and larger (meso- to macro-scopic) regimes. An extension of the MS-CG framework to the quantum regime,[473] described by quantum Boltzmann statistics, corroborates that such a multiscale treatment is possible in one extreme, while the other extreme toward macroscale CG modeling has been actively pursued via mesoscopic fluids.[105,107] Similarly, the Electronic Coarse-Graining (ECG) method[474,475] has been developed to target configurationally dependent electronic structure and applied to semiconductors[476] and optoelectronic materials.[477] These advances promise to significantly extend the spatiotemporal scale of systematically parametrized simulations, which currently encompass biological entities comprising millions to billions of atoms, e.g., the SARS-CoV-2 virion.[478] When bridging across distinctly different scales, transferability between these scales is critical to reconcile different emergent physics within a single unified model. While at an early stage, recent success in quantum mechanics/CG molecular mechanics (QM/CG-MM),[479,480] which encompasses both quantum and molecular regimes, implies that it is possible to design CG mappings, equations of motion, and energetics based on the scale of interest. In the other limit, a promising direction would be to incorporate classical field theory into the bottom-up CG framework.[481]

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

While future efforts should focus on extending the multiscale regime, an equal amount of attention might be spent developing new CG theories at the molecular level. Possible directions for new CG methodologies include the following: (1) stable and extensible nonlinear CG mappings for complex mesoscale systems, (2) fully expressive CG energetics, (3) a more complete description of UCG state dynamics, (4) a more complete dynamic representability between the FG and CG systems, and (5) advanced ML-based CG methodologies.

Lastly, we conclude this Review by emphasizing the necessity and importance of infrastructure for computer software and data set handling for the next generation of CG modeling. As models and derivation strategies increase in complexity, infrastructure for software and data becomes much more crucial. Since the naissance of modern computer simulations for molecules,[482] computer software has been inseparable from molecular models. Given the *ad hoc* nature of many existing CG models, CG modeling software should be amenable to new feature implementation and easy dissemination to facilitate both access and utility to the general scientific community. Currently, there are several options that satisfy these criteria with different features and capabilities according to their objectives: VOTCA[185] (BI, IBI; https://github.com/votca), MAGIC[483,484] (IMC; http://bitbucket.org/magic-su/magic-3), BOCS[485] (g-YBG, iter-YBG; https://github.com/noid-group/BOCS), openMSCG[486] (MS-CG, iter-MS-CG, MC-CG, REM; https://software.rcc.uchicago.edu/mscg/), and CGNet[59] (ML-based approaches; https://github.com/coarse-graining/cgnet). Continued development in both CG methodology and software necessitates standardized data sets that are used to validate existing or new CG methods. For example, in the free energy sampling community, alanine dipeptide serves as a standard example to validate new methodologies,[487] and ultralong atomistic MD trajectories for fast-folding proteins[488−490] using the ANTON supercomputer by the D. E. Shaw lab[491] have been extensively employed in kinetics studies such as MSMs. Even though CG modeling tackles a variety of chemical and biological systems across many different scales, there are relatively few data sets that are publicly available in the CG community. Therefore, by taking inspiration from the ML community, future efforts should also aim to achieve community agreement on which data sets are reliable to be shared and standardized in order to benchmark various CG methodologies. Altogether, we expect that a continuous exploration along these directions will push the frontiers of bottom-up CG modeling into the exploration and characterization of increasingly complex molecular systems.

AUTHOR INFORMATION

Corresponding Author

Gregory A. Voth − *Department of Chemistry, Chicago Center for Theoretical Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0002-3267-6748; Email: gavoth@uchicago.edu

Authors

Jaehyeok Jin − *Department of Chemistry, Chicago Center for Theoretical Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0003-3859-3545

Alexander J. Pak − *Department of Chemistry, Chicago Center for Theoretical Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, Illinois 60637, United States*

Aleksander E. P. Durumeric − *Department of Chemistry, Chicago Center for Theoretical Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, Illinois 60637, United States*

Timothy D. Loose − *Department of Chemistry, Chicago Center for Theoretical Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, Illinois 60637, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c00643

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation (NSF grant CHE-2102677) and the National Institutes of Health (NIH grant R01-GM063796). J.J. acknowledges a Kwanjeong Educational Fellowship during graduate studies and a Harper Dissertation Fellowship from the University of Chicago to finalize this Review. A.J.P. acknowledges support of the Ruth L. Kirschstein National Research Service Award Postdoctoral Fellowship by the National Institutes of Health (Grant No. F32-AI150477). The authors also thank Drs. James F. Dama and Yining Han for insightful discussions on this Review.

REFERENCES

(1) Shi, Y. A glimpse of structural biology through x-ray crystallography. *Cell* **2014**, *159*, 995−1014.

(2) Henderson, R. Overview and future of single particle electron cryomicroscopy. *Arch. Biochem. Biophys.* **2015**, *581*, 19−24.

(3) Shashkova, S.; Leake, M. C. Single-molecule fluorescence microscopy review: Shedding new light on old problems. *Biosci. Rep.* **2017**, *37*, BSR20170031.

(4) Sieben, C.; Douglass, K. M.; Guichard, P.; Manley, S. Super-resolution microscopy to decipher multi-molecular assemblies. *Curr. Opin. Struct. Biol.* **2018**, *49*, 169−176.

(5) Kleckner, I. R.; Foster, M. P. An introduction to nmr-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta Proteins Proteom.* **2011**, *1814*, 942−968.

(6) Kovermann, M.; Rogne, P.; Wolf-Watz, M. Protein dynamics and function from solution state nmr spectroscopy. *Q. Rev. Biophys.* **2016**, *49*, e6.

(7) Hollingsworth, S. A.; Dror, R. O. Molecular dynamics simulation for all. *Neuron* **2018**, *99* (6), 1129−1143.

(8) Saunders, M. G.; Voth, G. A. Coarse-graining methods for computational biology. *Annu. Rev. Biophys.* **2013**, *42*, 73−93.

(9) Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139* (9), 090901.

(10) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-grained protein models and their applications. *Chem. Rev.* **2016**, *116*, 7898−7936.

(11) Pak, A. J.; Voth, G. A. Advances in coarse-grained modeling of macromolecular complexes. *Curr. Opin. Struct. Biol.* **2018**, *52*, 119−126.

(12) Joshi, S. Y.; Deshmukh, S. A. A review of advancements in coarse-grained molecular dynamics simulations. *Mol. Simul.* **2021**, *47* (10−11), 786−803.

(13) Maupetit, J.; Tuffery, P.; Derreumaux, P. A coarse-grained protein force field for folding and structure prediction. *Proteins: Struct., Funct., Bioinf.* **2007**, *69* (2), 394−408.

(14) Latham, A. P.; Zhang, B. Unifying coarse-grained force fields for folded and disordered proteins. *Curr. Opin. Struct. Biol.* **2022**, *72*, 63−70.
(15) Liwo, A.; Ołdziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. A united-residue force field for off-lattice protein-structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J. Comput. Chem.* **1997**, *18* (7), 849−873.
(16) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. *J. Chem. Phys.* **2001**, *115* (5), 2323−2347.
(17) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Ołdziej, S.; Wachucik, K.; Scheraga, H. A. Modification and optimization of the united-residue (unres) potential energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins. *J. Phys. Chem. B* **2007**, *111* (1), 260−285.
(18) Liwo, A.; He, Y.; Scheraga, H. A. Coarse-grained force field: General folding theory. *Phys. Chem. Chem. Phys.* **2011**, *13* (38), 16890−16901.
(19) Derreumaux, P. From polypeptide sequences to structures using monte carlo simulations and an optimized potential. *J. Chem. Phys.* **1999**, *111* (5), 2301−2310.
(20) Kar, P.; Gopal, S. M.; Cheng, Y.-M.; Predeus, A.; Feig, M. Primo: A transferable coarse-grained force field for proteins. *J. Chem. Theory Comput.* **2013**, *9* (8), 3769−3788.
(21) Darré, L.; Machado, M. R.; Brandner, A. F.; González, H. C.; Ferreira, S.; Pantano, S. Sirah: A structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *J. Chem. Theory Comput.* **2015**, *11* (2), 723−739.
(22) Machado, M. R.; Barrera, E. E.; Klein, F.; Sóñora, M.; Silva, S.; Pantano, S. The sirah 2.0 force field: Altius, fortius, citius. *J. Chem. Theory Comput.* **2019**, *15* (4), 2719−2733.
(23) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The martini force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111* (27), 7812−7824.
(24) Marrink, S. J.; Tieleman, D. P. Perspective on the martini model. *Chem. Soc. Rev.* **2013**, *42* (16), 6801−6822.
(25) Arnarez, C. m.; Uusitalo, J. J.; Masman, M. F.; Ingólfsson, H. I.; de Jong, D. H.; Melo, M. N.; Periole, X.; de Vries, A. H.; Marrink, S. J. Dry martini, a coarse-grained force field for lipid membrane simulations with implicit solvent. *J. Chem. Theory Comput.* **2015**, *11* (1), 260−275.
(26) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A. Martini 3: A general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18* (4), 382−388.
(27) Alessandri, R.; Barnoud, J.; Gertsen, A. S.; Patmanidis, I.; de Vries, A. H.; Souza, P. C.; Marrink, S. J. Martini 3 coarse-grained force field: Small molecules. *Adv. Theory Simul.* **2022**, *5*, 2100391.
(28) Izvekov, S.; Voth, G. A. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **2005**, *109* (7), 2469−2473.
(29) Izvekov, S.; Voth, G. A. Multiscale coarse graining of liquid-state systems. *J. Chem. Phys.* **2005**, *123* (13), 134105.
(30) Izvekov, S.; Voth, G. A. Multiscale coarse-graining of mixed phospholipid/cholesterol bilayers. *J. Chem. Theory Comput.* **2006**, *2* (3), 637−648.
(31) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Voth, G. A. Multiscale coarse-graining and structural correlations: Connections to liquid-state theory. *J. Phys. Chem. B* **2007**, *111* (16), 4116−4127.
(32) 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* (24), 244114.
(33) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. The multiscale coarse-graining method. Ii. Numerical implementation for coarse-grained molecular models. *J. Chem. Phys.* **2008**, *128* (24), 244115.
(34) 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* (3), 954−965.
(35) Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **2008**, *129* (14), 144108.
(36) Chaimovich, A.; Shell, M. S. Anomalous waterlike behavior in spherically-symmetric water models optimized with the relative entropy. *Phys. Chem. Chem. Phys.* **2009**, *11* (12), 1901−1915.
(37) Chaimovich, A.; Shell, M. S. Relative entropy as a universal metric for multiscale errors. *Phys. Rev. E* **2010**, *81* (6), 060104.
(38) Chaimovich, A.; Shell, M. S. Coarse-graining errors and numerical optimization using a relative entropy framework. *J. Chem. Phys.* **2011**, *134* (9), 094112.
(39) Shell, M. S. Systematic coarse-graining of potential energy landscapes and dynamics in liquids. *J. Chem. Phys.* **2012**, *137* (8), 084503.
(40) Shell, M. S. Coarse-graining with the relative entropy. *Adv. Chem. Phys.* 2016; Vol. *161*, DOI: 10.1002/9781119290971.ch5.
(41) Pretti, E.; Shell, M. S. A microcanonical approach to temperature-transferable coarse-grained models using the relative entropy. *J. Chem. Phys.* **2021**, *155* (9), 094102.
(42) Nguyen, H. D.; Brooks, C. L. Generalized structural polymorphism in self-assembled viral particles. *Nano Lett.* **2008**, *8* (12), 4574−4581.
(43) Nguyen, H. D.; Reddy, V. S.; Brooks, C. L., III Invariant polymorphism in virus capsid assembly. *J. Am. Chem. Soc.* **2009**, *131* (7), 2606−2614.
(44) Perlmutter, J. D.; Qiao, C.; Hagan, M. F. Viral genome structures are optimal for capsid assembly. *eLife* **2013**, *2*, e00632.
(45) Mohajerani, F.; Hagan, M. F. The role of the encapsulated cargo in microcompartment assembly. *PLOS Comput. Biol.* **2018**, *14* (7), e1006351.
(46) Alessandri, R.; Souza, P. C. T.; Thallmair, S.; Melo, M. N.; De Vries, A. H.; Marrink, S. J. Pitfalls of the martini model. *J. Chem. Theory Comput.* **2019**, *15* (10), 5448−5460.
(47) Jarin, Z.; Newhouse, J.; Voth, G. A. Coarse-grained force fields from the perspective of statistical mechanics: Better understanding of the origins of a martini hangover. *J. Chem. Theory Comput.* **2021**, *17* (2), 1170−1180.
(48) Majumder, A.; Straub, J. E. Addressing the excessive aggregation of membrane proteins in the martini model. *J. Chem. Theory Comput.* **2021**, *17* (4), 2513−2521.
(49) Wagner, J. W.; Dama, J. F.; Durumeric, A. E. P.; Voth, G. A. On the representability problem and the physical meaning of coarse-grained models. *J. Chem. Phys.* **2016**, *145* (4), 044108.
(50) Dunn, N. J. H.; Foley, T. T.; Noid, W. G. Van der waals perspective on coarse-graining: Progress toward solving representability and transferability problems. *Acc. Chem. Res.* **2016**, *49* (12), 2832−2840.
(51) Jin, J.; Pak, A. J.; Voth, G. A. Understanding missing entropy in coarse-grained systems: Addressing issues of representability and transferability. *J. Phys. Chem. Lett.* **2019**, *10* (16), 4549−4557.
(52) Español, P. Statistical mechanics of coarse-graining. In *Novel methods in soft matter simulations*; Springer: 2004; pp 69−115, DOI: 10.1007/978-3-540-39895-0_3.
(53) Öttinger, H. C. *Beyond equilibrium thermodynamics*; John Wiley & Sons: 2005; DOI: 10.1002/0471727903.
(54) Izvekov, S.; Voth, G. A. Modeling real dynamics in the coarse-grained representation of condensed phase systems. *J. Chem. Phys.* **2006**, *125*, 151101.
(55) Ceriotti, M.; Clementi, C.; Anatole von Lilienfeld, O. Machine learning meets chemical physics. *J. Chem. Phys.* **2021**, *154* (16), 160401.
(56) Ye, H.; Xian, W.; Li, Y. Machine learning of coarse-grained models for organic molecules and polymers: Progress, opportunities, and challenges. *ACS Omega* **2021**, *6* (3), 1758−1772.
(57) 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* (12), 6213−6221.
(58) Durumeric, A. E. P.; Voth, G. A. Adversarial-residual-coarse-graining: Applying machine learning theory to systematic molecular coarse-graining. *J. Chem. Phys.* **2019**, *151* (12), 124110.
(59) 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* (5), 755−767.
(60) Noé, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine learning for molecular simulation. *Annu. Rev. Phys. Chem.* **2020**, *71*, 361−390.
(61) Lyman, E.; Cui, H.; Voth, G. A. Reconstructing protein remodeled membranes in molecular detail from mesoscopic models. *Phys. Chem. Chem. Phys.* **2011**, *13* (22), 10430−10436.
(62) Rzepiela, A. J.; Schäfer, L. V.; Goga, N.; Risselada, H. J.; De Vries, A. H.; Marrink, S. J. Reconstruction of atomistic details from coarse-grained structures. *J. Comput. Chem.* **2010**, *31* (6), 1333−1343.
(63) Ghanbari, A.; Bohm, M. C.; Muller-Plathe, F. A simple reverse mapping procedure for coarse-grained polymer models with rigid side groups. *Macromolecules* **2011**, *44* (13), 5520−5526.
(64) Wassenaar, T. A.; Pluhackova, K.; Böckmann, R. A.; Marrink, S. J.; Tieleman, D. P. Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models. *J. Chem. Theory Comput.* **2014**, *10* (2), 676−690.
(65) Lombardi, L. E.; Martí, M. A.; Capece, L. Cg2aa: Backmapping protein coarse-grained structures. *Bioinformatics* **2016**, *32* (8), 1235−1237.
(66) Machado, M. R.; Pantano, S. Sirah tools: Mapping, backmapping and visualization of coarse-grained models. *Bioinformatics* **2016**, *32* (10), 1568−1570.
(67) Badaczewska-Dawid, A. E.; Kolinski, A.; Kmiecik, S. Computational reconstruction of atomistic protein structures from coarse-grained models. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 162−176.
(68) Pezeshkian, W.; König, M.; Wassenaar, T. A.; Marrink, S. J. Backmapping triangulated surfaces to coarse-grained membrane models. *Nat. Commun.* **2020**, *11* (1), 2296.
(69) Vickery, O. N.; Stansfeld, P. J. Cg2at2: An enhanced fragment-based approach for serial multi-scale molecular dynamics simulations. *J. Chem. Theory Comput.* **2021**, *17* (10), 6472−6482.
(70) 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* (24), 243116.
(71) Wu, Z.; Cui, Q.; Yethiraj, A. A new coarse-grained model for water: The importance of electrostatic interactions. *J. Phys. Chem. B* **2010**, *114* (32), 10524−10529.
(72) Riniker, S.; van Gunsteren, W. F. A simple, efficient polarizable coarse-grained water model for molecular dynamics simulations. *J. Chem. Phys.* **2011**, *134* (8), 084110.
(73) Wang, Y.; Feng, S.; Voth, G. A. Transferable coarse-grained models for ionic liquids. *J. Chem. Theory Comput.* **2009**, *5* (4), 1091−1098.
(74) 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* (2), 926−935.
(75) Melo, M. N.; Ingólfsson, H. I.; Marrink, S. J. Parameters for martini sterols and hopanoids based on a virtual-site description. *J. Chem. Phys.* **2015**, *143* (24), 243152.
(76) Jin, J.; Han, Y.; Voth, G. A. Coarse-graining involving virtual sites: Centers of symmetry coarse-graining. *J. Chem. Phys.* **2019**, *150* (15), 154103.
(77) Rzepiela, A. J.; Louhivuori, M.; Peter, C.; Marrink, S. J. Hybrid simulations: Combining atomistic and coarse-grained force fields using virtual sites. *Phys. Chem. Chem. Phys.* **2011**, *13* (22), 10437−10448.
(78) Melo, M.; Ingólfsson, H. I.; Marrink, S. J. Parameters for martini sterols and hopanoids based on a virtual-site description. *J. Chem. Phys.* **2015**, *143* (24), 243152.
(79) Liu, Y.; De Vries, A. H.; Barnoud, J.; Pezeshkian, W.; Melcr, J.; Marrink, S. J. Dual resolution membrane simulations using virtual sites. *J. Phys. Chem. B* **2020**, *124* (19), 3944−3953.
(80) Ruiz-Herrero, T.; Hagan, M. F. Simulations show that virus assembly and budding are facilitated by membrane microdomains. *Biophys. J.* **2015**, *108* (3), 585−595.
(81) Grime, J. M. A.; Dama, J. F.; Ganser-Pornillos, B. K.; Woodward, C. L.; Jensen, G. J.; Yeager, M.; Voth, G. A. Coarse-grained simulation reveals key features of hiv-1 capsid self-assembly. *Nat. Commun.* **2016**, *7* (1), 11568.
(82) Perlmutter, J. D.; Hagan, M. F. The role of packaging sites in efficient and specific virus assembly. *J. Mol. Biol.* **2015**, *427* (15), 2451−2467.
(83) Pak, A. J.; Grime, J. M. A.; Sengupta, P.; Chen, A. K.; Durumeric, A. E. P.; Srivastava, A.; Yeager, M.; Briggs, J. A. G.; Lippincott-Schwartz, J.; Voth, G. A. Immature hiv-1 lattice assembly dynamics are regulated by scaffolding from nucleic acid and the plasma membrane. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114* (47), E10056−E10065.
(84) Pak, A. J.; Grime, J. M. A.; Yu, A.; Voth, G. A. Off-pathway assembly: A broad-spectrum mechanism of action for drugs that undermine controlled hiv-1 viral capsid formation. *J. Am. Chem. Soc.* **2019**, *141* (26), 10214−10224.
(85) Yu, A.; Skorupka, K. A.; Pak, A. J.; Ganser-Pornillos, B. K.; Pornillos, O.; Voth, G. A. Trim5α self-assembly and compartmentalization of the hiv-1 viral capsid. *Nat. Commun.* **2020**, *11* (1), 1307.
(86) Sawade, K.; Peter, C. Multiscale simulations of protein and membrane systems. *Curr. Opin. Struct. Biol.* **2022**, *72*, 203−208.
(87) Chakraborty, M.; Xu, J.; White, A. D. Is preservation of symmetry necessary for coarse-graining? *Phys. Chem. Chem. Phys.* **2020**, *22* (26), 14998−15005.
(88) Dhamankar, S.; Webb, M. A. Chemically specific coarse-graining of polymers: Methods and prospects. *J. Polym. Sci.* **2021**, *59* (22), 2613−2643.
(89) Pak, A. J.; Dannenhoffer-Lafage, T.; Madsen, J. J.; Voth, G. A. Systematic coarse-grained lipid force fields with semiexplicit solvation via virtual sites. *J. Chem. Theory Comput.* **2019**, *15* (3), 2087−2100.
(90) Schweizer, K. S. Microscopic theory of the dynamics of polymeric liquids: General formulation of a mode−mode-coupling approach. *J. Chem. Phys.* **1989**, *91* (9), 5802−5821.
(91) Kinjo, T.; Hyodo, S.-a. Equation of motion for coarse-grained simulation based on microscopic description. *Phys. Rev. E* **2007**, *75* (5), 051109.
(92) Hijón, C.; Español, P.; Vanden-Eijnden, E.; Delgado-Buscalioni, R. Mori−zwanzig formalism as a practical computational tool. *Faraday Discuss.* **2010**, *144*, 301−322.
(93) Li, Z.; Bian, X.; Caswell, B.; Karniadakis, G. E. Construction of dissipative particle dynamics models for complex fluids via the mori−zwanzig formulation. *Soft Matter* **2014**, *10* (43), 8659−8672.
(94) Español, P.; Serrano, M.; Zuñiga, I. Coarse-graining of a fluid and its relation with dissipative particle dynamics and smoothed particle dynamic. *Int. J. Mod. Phys. C* **1997**, *08* (04), 899−908.
(95) Hadley, K. R.; McCabe, C. On the investigation of coarse-grained models for water: Balancing computational efficiency and the retention of structural properties. *J. Phys. Chem. B* **2010**, *114* (13), 4590−4599.
(96) Izvekov, S.; Rice, B. M. Multi-scale coarse-graining of non-conservative interactions in molecular liquids. *J. Chem. Phys.* **2014**, *140* (10), 104104.
(97) Zavadlav, J.; Marrink, S. J.; Praprotnik, M. Adaptive resolution simulation of supramolecular water: The concurrent making, breaking, and remaking of water bundles. *J. Chem. Theory Comput.* **2016**, *12* (8), 4138−4145.
(98) Zavadlav, J.; Praprotnik, M. Adaptive resolution simulations coupling atomistic water to dissipative particle dynamics. *J. Chem. Phys.* **2017**, *147* (11), 114110.
(99) Li, M.; Lu, W.; Zhang, J. Z. Ultra-coarse-graining modeling of liquid water. *J. Chem. Phys.* **2021**, *154* (22), 224506.
(100) Ayton, G. S.; Tepper, H. L.; Mirijanian, D. T.; Voth, G. A. A new perspective on the coarse-grained dynamics of fluids. *J. Chem. Phys.* **2004**, *120* (9), 4074−4088.
(101) Ayton, G.; Voth, G. A. Bridging microscopic and mesoscopic simulations of lipid bilayers. *Biophys. J.* **2002**, *83* (6), 3357−3370.

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

(102) Ayton, G. S.; Voth, G. A. Mesoscopic lateral diffusion in lipid bilayers. *Biophys. J.* **2004**, *87* (5), 3299−3311.
(103) Ayton, G. S.; McWhirter, J. L.; Voth, G. A. A second generation mesoscopic lipid bilayer model: Connections to field-theory descriptions of membranes and nonlocal hydrodynamics. *J. Chem. Phys.* **2006**, *124* (6), 064906.
(104) Ayton, G. S.; Lyman, E.; Krishna, V.; Swenson, R. D.; Mim, C.; Unger, V. M.; Voth, G. A. New insights into bar domain-induced membrane remodeling. *Biophys. J.* **2009**, *97* (6), 1616−1625.
(105) Han, Y.; Dama, J. F.; Voth, G. A. Mesoscopic coarse-grained representations of fluids rigorously derived from atomistic models. *J. Chem. Phys.* **2018**, *149* (4), 044104.
(106) Luo, S.; Thachuk, M. Conservative potentials for a lattice-mapped coarse-grained scheme. *J. Phys. Chem. A* **2021**, *125* (29), 6486−6497.
(107) Han, Y.; Jin, J.; Voth, G. A. Constructing many-body dissipative particle dynamics models of fluids from bottom-up coarse-graining. *J. Chem. Phys.* **2021**, *154* (8), 084122.
(108) Peter, C.; Kremer, K. Multiscale simulation of soft matter systems−from the atomistic to the coarse-grained level and back. *Soft Matter* **2009**, *5* (22), 4357−4366.
(109) Lyman, E.; Ytreberg, F. M.; Zuckerman, D. M. Resolution exchange simulation. *Phys. Rev. Lett.* **2006**, *96* (2), 028105.
(110) Lyman, E.; Zuckerman, D. M. Resolution exchange simulation with incremental coarsening. *J. Chem. Theory Comput.* **2006**, *2* (3), 656−666.
(111) Zuckerman, D. M.; Lyman, E. A second look at canonical sampling of biomolecules using replica exchange simulation. *J. Chem. Theory Comput.* **2006**, *2* (4), 1200−1202.
(112) Liu, P.; Voth, G. A. Smart resolution replica exchange: An efficient algorithm for exploring complex energy landscapes. *J. Chem. Phys.* **2007**, *126* (4), 045106.
(113) Liu, P.; Shi, Q.; Lyman, E.; Voth, G. A. Reconstructing atomistic detail for coarse-grained models with resolution exchange. *J. Chem. Phys.* **2008**, *129* (11), 114103.
(114) Giulini, M.; Rigoli, M.; Mattiotti, G.; Menichetti, R.; Tarenzi, T.; Fiorentini, R.; Potestio, R. From system modeling to system analysis: The impact of resolution level and resolution distribution in the computer-aided investigation of biomolecules. *Front. Mol. Biosci.* **2021**, *8*, 676976.
(115) 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* (11), 5073−5083.
(116) Zhang, Z.; Pfaendtner, J.; Grafmüller, A.; Voth, G. A. Defining coarse-grained representations of large biomolecules and biomolecular complexes from elastic network models. *Biophys. J.* **2009**, *97* (8), 2327−2337.
(117) Lyman, E.; Pfaendtner, J.; Voth, G. A. Systematic multiscale parameterization of heterogeneous elastic network models of proteins. *Biophys. J.* **2008**, *95* (9), 4183−4192.
(118) Sinitskiy, A. V.; Saunders, M. G.; Voth, G. A. Optimal number of coarse-grained sites in different components of large biomolecular complexes. *J. Phys. Chem. B* **2012**, *116* (29), 8363−8374.
(119) Li, M.; Zhang, J. Z.; Xia, F. A new algorithm for construction of coarse-grained sites of large biomolecules. *J. Comput. Chem.* **2016**, *37* (9), 795−804.
(120) Li, M.; Zhang, J. Z.; Xia, F. Constructing optimal coarse-grained sites of huge biomolecules by fluctuation maximization. *J. Chem. Theory Comput.* **2016**, *12*, 2091−2100.
(121) 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* (24), 243104.
(122) Foley, T. T.; Kidder, K. M.; Shell, M. S.; Noid, W. G. Exploring the landscape of model representations. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117* (39), 24061−24068.
(123) Koehl, P.; Poitevin, F.; Navaza, R.; Delarue, M. The renormalization group and its applications to generating coarse-grained models of large biological molecular systems. *J. Chem. Theory Comput.* **2017**, *13* (3), 1424−1438.
(124) 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* (11), 6795−6813.
(125) Errica, F.; Giulini, M.; Bacciu, D.; Menichetti, R.; Micheli, A.; Potestio, R. A deep graph network−enhanced sampling approach to efficiently explore the space of reduced representations of proteins. *Front. Mol. Biosci.* **2021**, *8*, 637396.
(126) Menichetti, R.; Giulini, M.; Potestio, R. A journey through mapping space: Characterising the statistical and metric properties of reduced representations of macromolecules. *Eur. Phys. J. B* **2021**, *94* (10), 204.
(127) Wang, F.; Landau, D. Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. *Phys. Rev. E* **2001**, *64* (5), 056101.
(128) Wang, F.; Landau, D. P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **2001**, *86* (10), 2050.
(129) Praprotnik, M.; Delle Site, L.; Kremer, K. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *J. Chem. Phys.* **2005**, *123* (22), 224106.
(130) Potestio, R.; Fritsch, S.; Espanol, P.; Delgado-Buscalioni, R.; Kremer, K.; Everaers, R.; Donadio, D. Hamiltonian adaptive resolution simulation for molecular liquids. *Phys. Rev. Lett.* **2013**, *110* (10), 108301.
(131) Kreis, K.; Donadio, D.; Kremer, K.; Potestio, R. A unified framework for force-based and energy-based adaptive resolution simulations. *Europhys. Lett.* **2014**, *108* (3), 30007.
(132) Español, P.; Delgado-Buscalioni, R.; Everaers, R.; Potestio, R.; Donadio, D.; Kremer, K. Statistical mechanics of hamiltonian adaptive resolution simulations. *J. Chem. Phys.* **2015**, *142* (6), 064115.
(133) Fogarty, A. C.; Potestio, R.; Kremer, K. Adaptive resolution simulation of a biomolecule and its hydration shell: Structural and dynamical properties. *J. Chem. Phys.* **2015**, *142* (19), 195101.
(134) Delgado-Buscalioni, R.; Kremer, K.; Praprotnik, M. Concurrent triple-scale simulation of molecular liquids. *J. Chem. Phys.* **2008**, *128* (11), 114110.
(135) Delgado-Buscalioni, R.; Kremer, K.; Praprotnik, M. Coupling atomistic and continuum hydrodynamics through a mesoscopic model: Application to liquid water. *J. Chem. Phys.* **2009**, *131* (24), 244107.
(136) Poma, A.; Delle Site, L. Classical to path-integral adaptive resolution in molecular simulation: Towards a smooth quantum-classical coupling. *Phys. Rev. Lett.* **2010**, *104* (25), 250201.
(137) Agarwal, A.; Delle Site, L. Path integral molecular dynamics within the grand canonical-like adaptive resolution technique: Simulation of liquid water. *J. Chem. Phys.* **2015**, *143* (9), 094102.
(138) Delgado-Buscalioni, R.; Sablić, J.; Praprotnik, M. Open boundary molecular dynamics. *Eur. Phys. J.: Spec. Top.* **2015**, *224* (12), 2331−2349.
(139) Wang, H.; Agarwal, A. Adaptive resolution simulation in equilibrium and beyond. *Eur. Phys. J.: Spec. Top.* **2015**, *224* (12), 2269−2287.
(140) Wang, H.; Hartmann, C.; Schütte, C.; Delle Site, L. Grand-canonical-like molecular-dynamics simulations by using an adaptive-resolution technique. *Phys. Rev. X* **2013**, *3* (1), 011018.
(141) Sablić, J.; Praprotnik, M.; Delgado-Buscalioni, R. Open boundary molecular dynamics of sheared star-polymer melts. *Soft Matter* **2016**, *12* (8), 2416−2439.
(142) Ercolessi, F.; Adams, J. B. Interatomic potentials from first-principles calculations: The force-matching method. *Europhys. Lett.* **1994**, *26* (8), 583.
(143) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching. *J. Chem. Phys.* **2004**, *120* (23), 10896−10913.
(144) Das, A.; Andersen, H. C. The multiscale coarse-graining method. Iii. A test of pairwise additivity of the coarse-grained potential and of new basis functions for the variational calculation. *J. Chem. Phys.* **2009**, *131* (3), 034102.

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

(145) Krishna, V.; Noid, W. G.; Voth, G. A. The multiscale coarse-graining method. Iv. Transferring coarse-grained potentials between temperatures. *J. Chem. Phys.* **2009**, *131* (2), 024103.
(146) Das, A.; Andersen, H. C. The multiscale coarse-graining method. V. Isothermal-isobaric ensemble. *J. Chem. Phys.* **2010**, *132* (16), 164106.
(147) 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* (16), 164107.
(148) Lu, L.; Voth, G. A. The multiscale coarse-graining method. Vii. Free energy decomposition of coarse-grained effective potentials. *J. Chem. Phys.* **2011**, *134* (22), 224107.
(149) Das, A.; Andersen, H. C. The multiscale coarse-graining method. Viii. Multiresolution hierarchical basis functions and basis function selection in the construction of coarse-grained force fields. *J. Chem. Phys.* **2012**, *136* (19), 194113.
(150) Das, A.; Andersen, H. C. The multiscale coarse-graining method. Ix. A general method for construction of three body coarse-grained force fields. *J. Chem. Phys.* **2012**, *136* (19), 194114.
(151) Das, A.; Lu, L.; Andersen, H. C.; Voth, G. A. The multiscale coarse-graining method. X. Improved algorithms for constructing coarse-grained potentials for molecular systems. *J. Chem. Phys.* **2012**, *136* (19), 194115.
(152) Lu, L.; Voth, G. A. Systematic coarse-graining of a multicomponent lipid bilayer. *J. Phys. Chem. B* **2009**, *113* (5), 1501−1510.
(153) Press, W. H.; William, H.; Teukolsky, S. A.; Saul, A.; Vetterling, W. T.; Flannery, B. P. *Numerical recipes: The art of scientific computing*, 3rd ed.; Cambridge University Press: New York, NY, USA, 2007.
(154) Kullback, S.; Leibler, R. A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22* (1), 79−86.
(155) Schöberl, M.; Zabaras, N.; Koutsourelakis, P. S. Predictive coarse-graining. *J. Comput. Phys.* **2017**, *333*, 49−77.
(156) Rudzinski, J. F.; Noid, W. G. Coarse-graining entropy, forces, and structures. *J. Chem. Phys.* **2011**, *135* (21), 214101.
(157) Hansen, J.-P.; McDonald, I. R. *Theory of simple liquids*; Elsevier: 2006; pp 17−19.
(158) Mullinax, J. W.; Noid, W. G. Generalized yvon-born-green theory for molecular systems. *Phys. Rev. Lett.* **2009**, *103* (19), 198104.
(159) Mullinax, J. W.; Noid, W. G. A generalized-yvon− born− green theory for determining coarse-grained interaction potentials. *J. Phys. Chem. C* **2010**, *114* (12), 5661−5674.
(160) Mullinax, J. W.; Noid, W. G. Reference state for the generalized yvon−born−green theory: Application for coarse-grained model of hydrophobic hydration. *J. Chem. Phys.* **2010**, *133* (12), 124107.
(161) Mullinax, J. W.; Noid, W. G. Recovering physical potentials from a model protein databank. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (46), 19867−19872.
(162) Ellis, C. R.; Rudzinski, J. F.; Noid, W. G. Generalized-yvon−born−green model of toluene. *Macromol. Theory Simul.* **2011**, *20* (7), 478−495.
(163) 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* (13), 134107.
(164) 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* (28), 8295−8312.
(165) Lu, L.; Dama, J. F.; Voth, G. A. Fitting coarse-grained distribution functions through an iterative force-matching method. *J. Chem. Phys.* **2013**, *139* (12), 121906.
(166) Wörner, S. J.; Bereau, T.; Kremer, K.; Rudzinski, J. F. Direct route to reproducing pair distribution functions with coarse-grained models via transformed atomistic cross correlations. *J. Chem. Phys.* **2019**, *151* (24), 244110.
(167) Köhler, J.; Chen, Y.; Krämer, A.; Clementi, C.; Noé, F. Force-matching coarse-graining without forces. 2022, arXiv:2203.11167. *arXiv preprint*. https://arxiv.org/abs/2203.11167 (accessed 2022-08-16).
(168) Tóth, G.; Király, N.; Vrabecz, A. Pair potentials from diffraction data on liquids: A neural network solution. *J. Chem. Phys.* **2005**, *123* (17), 174109.
(169) Tóth, G. Interactions from diffraction data: Historical and comprehensive overview of simulation assisted methods. *J. Phys.: Condens. Matter* **2007**, *19* (33), 335220.
(170) Henderson, R. L. A uniqueness theorem for fluid pair correlation functions. *Phys. Lett. A* **1974**, *49* (3), 197−198.
(171) Tschöp, W.; Kremer, K.; Batoulis, J.; Bürger, T.; Hahn, O. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polym.* **1998**, *49* (2−3), 61−74.
(172) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24* (13), 1624−1636.
(173) Hanke, M. Well-posedness of the iterative boltzmann inversion. *J. Stat. Phys.* **2018**, *170* (3), 536−553.
(174) Stillinger, F. H.; Torquato, S. Structural degeneracy in pair distance distributions. *J. Chem. Phys.* **2019**, *150* (20), 204125.
(175) Wang, H.; Stillinger, F. H.; Torquato, S. Sensitivity of pair statistics on pair potentials in many-body systems. *J. Chem. Phys.* **2020**, *153* (12), 124106.
(176) Fu, C.-C.; Kulkarni, P. M.; Shell, M. S.; Leal, L. G. A test of systematic coarse-graining of molecular dynamics simulations: Thermodynamic properties. *J. Chem. Phys.* **2012**, *137* (16), 164106.
(177) Potestio, R. Is henderson's theorem practically useful. *JUnQ* **2013**, *3*, 13−15.
(178) Khot, A.; Shiring, S. B.; Savoie, B. M. Evidence of information limitations in coarse-grained models. *J. Chem. Phys.* **2019**, *151* (24), 244105.
(179) Moore, T. C.; Iacovella, C. R.; McCabe, C. Derivation of coarse-grained potentials via multistate iterative boltzmann inversion. *J. Chem. Phys.* **2014**, *140* (22), 224104.
(180) Moore, T. C.; Iacovella, C. R.; McCabe, C. Development of a coarse-grained water forcefield via multistate iterative boltzmann inversion. In *Foundations of molecular modeling and simulation*; Springer: 2016; pp 37−52, DOI: 10.1007/978-981-10-1128-3_3.
(181) Ganguly, P.; Mukherji, D.; Junghans, C.; van der Vegt, N. F. Kirkwood−buff coarse-grained force fields for aqueous solutions. *J. Chem. Theory Comput.* **2012**, *8* (5), 1802−1807.
(182) de Oliveira, T. E.; Netz, P. A.; Kremer, K.; Junghans, C.; Mukherji, D. C−ibi: Targeting cumulative coordination within an iterative protocol to derive coarse-grained models of (multi-component) complex fluids. *J. Chem. Phys.* **2016**, *144* (17), 174106.
(183) Lyubartsev, A. P.; Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse monte carlo approach. *Phys. Rev. E* **1995**, *52* (4), 3730.
(184) Lyubartsev, A. P.; Laaksonen, A. Osmotic and activity coefficients from effective potentials for hydrated ions. *Phys. Rev. E* **1997**, *55* (5), 5689.
(185) Ruhle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. Versatile object-oriented toolkit for coarse-graining applications. *J. Chem. Theory Comput.* **2009**, *5* (12), 3211−3223.
(186) Sambriski, E. J.; Yatsenko, G.; Nemirovskaya, M. A.; Guenza, M. G. Analytical coarse-grained description for polymer melts. *J. Chem. Phys.* **2006**, *125* (23), 234902.
(187) Sambriski, E. J.; Guenza, M. G. Theoretical coarse-graining approach to bridge length scales in diblock copolymer liquids. *Phys. Rev. E* **2007**, *76* (5), 051801.
(188) Clark, A. J.; McCarty, J.; Guenza, M. G. Effective potentials for representing polymers in melts as chains of interacting soft particles. *J. Chem. Phys.* **2013**, *139* (12), 124906.
(189) Wang, Q.; Keffer, D. J.; Nicholson, D. M. A coarse-grained model for polyethylene glycol polymer. *J. Chem. Phys.* **2011**, *135* (21), 214903.
(190) Wang, Y.; Noid, W. G.; Liu, P.; Voth, G. A. Effective force coarse-graining. *Phys. Chem. Chem. Phys.* **2009**, *11* (12), 2002−2015.
(191) Müller, E. A.; Gelb, L. D. Molecular modeling of fluid-phase equilibria using an isotropic multipolar potential. *Ind. Eng. Chem. Res.* **2003**, *42* (17), 4123−4131.

(192) Honda, K. A coarse-grained potential for liquid water by spherically projected hydrogen-bonding interactions. *J. Mol. Liq.* **2020**, *300*, 112246.
(193) 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* (4), 044104.
(194) Jin, J.; Pak, A. J.; Han, Y.; Voth, G. A. A new one-site coarse-grained model for water: Bottom-up many-body projected water (bumper). Ii. Temperature transferability and structural properties at low temperature. *J. Chem. Phys.* **2021**, *154* (4), 044105.
(195) Lebold, K. M.; Noid, W. G. Dual-potential approach for coarse-grained implicit solvent models with accurate, internally consistent energetics and predictive transferability. *J. Chem. Phys.* **2019**, *151* (16), 164113.
(196) Lebold, K. M.; Noid, W. G. Dual approach for effective potentials that accurately model structure and energetics. *J. Chem. Phys.* **2019**, *150* (23), 234107.
(197) Brini, E.; Marcon, V.; van der Vegt, N. F. A. Conditional reversible work method for molecular coarse graining applications. *Phys. Chem. Chem. Phys.* **2011**, *13* (22), 10468−10474.
(198) Brini, E.; Van der Vegt, N. Chemically transferable coarse-grained potentials from conditional reversible work calculations. *J. Chem. Phys.* **2012**, *137* (15), 154113.
(199) Deichmann, G.; Marcon, V.; van der Vegt, N. F. A. Bottom-up derivation of conservative and dissipative interactions for coarse-grained molecular liquids with the conditional reversible work method. *J. Chem. Phys.* **2014**, *141* (22), 224109.
(200) Deichmann, G.; van der Vegt, N. F. A. Conditional reversible work coarse-grained models of molecular liquids with coulomb electrostatics−a proof of concept study on weakly polar organic molecules. *J. Chem. Theory Comput.* **2017**, *13* (12), 6158−6166.
(201) Deichmann, G.; Dallavalle, M.; Rosenberger, D.; van der Vegt, N. F. A. Phase equilibria modeling with systematically coarse-grained models—a comparative study on state point transferability. *J. Phys. Chem. B* **2019**, *123* (2), 504−515.
(202) Deichmann, G.; van der Vegt, N. F. A. Conditional reversible work coarse-grained models with explicit electrostatics—an application to butylmethylimidazolium ionic liquids. *J. Chem. Theory Comput.* **2019**, *15* (2), 1187−1198.
(203) Ayton, G. S.; Noid, W. G.; Voth, G. A. Systematic coarse graining of biomolecular and soft-matter systems. *MRS Bull.* **2007**, *32* (11), 929−934.
(204) Shi, Q.; Izvekov, S.; Voth, G. A. Mixed atomistic and coarse-grained molecular dynamics: Simulation of a membrane-bound ion channel. *J. Phys. Chem. B* **2006**, *110* (31), 15045−15048.
(205) Praprotnik, M.; Site, L. D.; Kremer, K. Multiscale simulation of soft matter: From scale bridging to adaptive resolution. *Annu. Rev. Phys. Chem.* **2008**, *59*, 545−571.
(206) Zavadlav, J.; Bevc, S.; Praprotnik, M. Adaptive resolution simulations of biomolecular systems. *Eur. Biophys. J.* **2017**, *46* (8), 821−835.
(207) Cortes-Huerto, R.; Praprotnik, M.; Kremer, K.; Delle Site, L. From adaptive resolution to molecular dynamics of open systems. *Eur. Phys. J. B* **2021**, *94* (9), 189.
(208) Johnson, M. E.; Head-Gordon, T.; Louis, A. A. Representability problems for coarse-grained water potentials. *J. Chem. Phys.* **2007**, *126* (14), 144509.
(209) Lyubartsev, A.; Mirzoev, A.; Chen, L.; Laaksonen, A. Systematic coarse-graining of molecular models by the newton inversion method. *Faraday Discuss.* **2010**, *144* (0), 43−56.
(210) Jin, J.; Voth, G. A. Theory of thermodynamic entropy in coarse-grained models. I. Correct entropy correspondence. 2022, *Submitted*.
(211) Jin, J.; Han, Y.; Voth, G. A. Theory of thermodynamic entropy in coarse-grained models. Ii. Full representability relationship. 2022, *Submitted*.
(212) McQuarrie, D. A. *Statistical mechanics*; Sterling Publishing Company: 2000; pp 254−276.
(213) Dannenhoffer-Lafage, T.; Wagner, J. W.; Durumeric, A. E. P.; Voth, G. A. Compatible observable decompositions for coarse-grained representations of real molecular systems. *J. Chem. Phys.* **2019**, *151* (13), 134115.
(214) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular mechanics. The mm3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111* (23), 8551−8566.
(215) Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III. Dreiding: A generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94* (26), 8897−8909.
(216) Rappé, A. K.; Casewit, C. J.; Colwell, K.; Goddard, W. A., III; Skiff, W. M. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024−10035.
(217) Scott, W. R.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. The gromos biomolecular simulation program package. *J. Phys. Chem. A* **1999**, *103* (19), 3596−3607.
(218) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26* (16), 1668−1688.
(219) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S. Charmm: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30* (10), 1545−1614.
(220) Gay, J. G.; Berne, B. J. Modification of the overlap potential to mimic a linear site−site potential. *J. Chem. Phys.* **1981**, *74* (6), 3316−3319.
(221) Babadi, M.; Everaers, R.; Ejtehadi, M. R. Coarse-grained interaction potentials for anisotropic molecules. *J. Chem. Phys.* **2006**, *124* (17), 174708.
(222) Golubkov, P. A.; Ren, P. Generalized coarse-grained model based on point multipole and gay-berne potentials. *J. Chem. Phys.* **2006**, *125* (6), 064103.
(223) Golubkov, P. A.; Wu, J. C.; Ren, P. A transferable coarse-grained model for hydrogen-bonding liquids. *Phys. Chem. Chem. Phys.* **2008**, *10* (15), 2050−2057.
(224) Wu, J.; Zhen, X.; Shen, H.; Li, G.; Ren, P. Gay-berne and electrostatic multipole based coarse-grain potential in implicit solvent. *J. Chem. Phys.* **2011**, *135* (15), 155104.
(225) Shen, H.; Li, Y.; Ren, P.; Zhang, D.; Li, G. Anisotropic coarse-grained model for proteins based on gay−berne and electric multipole potentials. *J. Chem. Theory Comput.* **2014**, *10* (2), 731−750.
(226) Tanis, I.; Rousseau, B.; Soulard, L.; Lemarchand, C. A. Assessment of an anisotropic coarse-grained model for cis-1, 4-polybutadiene: A bottom-up approach. *Soft Matter* **2021**, *17* (3), 621−636.
(227) Cohen, A. E.; Jackson, N. E.; De Pablo, J. J. Anisotropic coarse-grained model for conjugated polymers: Investigations into solution morphologies. *Macromolecules* **2021**, *54* (8), 3780−3789.
(228) Nguyen, H. T.; Huang, D. M. Systematic bottom-up molecular coarse-graining via force matching using anisotropic particles. *J. Chem. Phys.* **2022**, *156* (18), 184118.
(229) Hankins, D.; Moskowitz, J. W.; Stillinger, F. H. Water molecule interactions. *J. Chem. Phys.* **1970**, *53* (12), 4544−4554.
(230) Moore, E. B.; Molinero, V. Structural transformation in supercooled water controls the crystallization rate of ice. *Nature (London)* **2011**, *479* (7374), 506.
(231) Moore, E. B.; Molinero, V. Is it cubic? Ice crystallization from deeply supercooled water. *Phys. Chem. Chem. Phys.* **2011**, *13* (44), 20008−20016.
(232) Li, T.; Donadio, D.; Russo, G.; Galli, G. Homogeneous ice nucleation from supercooled water. *Phys. Chem. Chem. Phys.* **2011**, *13* (44), 19807−19813.
(233) Jabes, B. S.; Nayar, D.; Dhabal, D.; Molinero, V.; Chakravarty, C. Water and other tetrahedral liquids: Order, anomalies and solvation. *J. Phys.: Condens.* **2012**, *24* (28), 284116.

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

(234) Bullock, G.; Molinero, V. Low-density liquid water is the mother of ice: On the relation between mesostructure, thermodynamics and ice crystallization in solutions. *Faraday Discuss.* **2013**, *167*, 371−388.
(235) Holten, V.; Limmer, D. T.; Molinero, V.; Anisimov, M. A. Nature of the anomalies in the supercooled liquid state of the mw model of water. *J. Chem. Phys.* **2013**, *138* (17), 174501.
(236) Russo, J.; Romano, F.; Tanaka, H. New metastable form of ice and its role in the homogeneous crystallization of water. *Nat. Mater.* **2014**, *13* (7), 733.
(237) Stillinger, F. H.; Weber, T. A. Computer simulation of local order in condensed phases of silicon. *Phys. Rev. B* **1985**, *31* (8), 5262.
(238) Molinero, V.; Moore, E. B. Water modeled as an intermediate element between carbon and silicon. *J. Phys. Chem. B* **2009**, *113* (13), 4008−4016.
(239) Russ, C.; Brunner, M.; Bechinger, C.; von Grünberg, H.-H. Three-body forces at work: Three-body potentials derived from triplet correlations in colloidal suspensions. *Europhys. Lett.* **2005**, *69* (3), 468.
(240) Scherer, C.; Andrienko, D. Understanding three-body contributions to coarse-grained force fields. *Phys. Chem. Chem. Phys.* **2018**, *20* (34), 22387−22394.
(241) Yvon, J. *La théorie statistique des fluides et l'equation d'etat*; Hermann & cie: 1935; Vol. *203*.
(242) Kirkwood, J. G. The statistical mechanical theory of transport processes i. General theory. *J. Chem. Phys.* **1946**, *14* (3), 180−201.
(243) Born, M.; Green, H. S. A general kinetic theory of liquids i. The molecular distribution functions. *Proc. R. Soc. London, Ser. A* **1946**, *188* (1012), 10−18.
(244) Bogolubov, N. Kinetic equations. *Zh. Eksp. Teor. Fiz.* **1946**, *16* (8), 691−702.
(245) Kirkwood, J. G.; Buff, F. P.; Green, M. S. The statistical mechanical theory of transport processes. Iii. The coefficients of shear and bulk viscosity of liquids. *J. Chem. Phys.* **1949**, *17* (10), 988−994.
(246) Babin, V.; Leforestier, C.; Paesani, F. Development of a "first principles" water potential with flexible monomers: Dimer potential energy surface, vrt spectrum, and second virial coefficient. *J. Chem. Theory Comput.* **2013**, *9* (12), 5395−5403.
(247) Babin, V.; Medders, G. R.; Paesani, F. Development of a "first principles" water potential with flexible monomers. Ii: Trimer potential energy surface, third virial coefficient, and small clusters. *J. Chem. Theory Comput.* **2014**, *10* (4), 1599−1607.
(248) Medders, G. R.; Babin, V.; Paesani, F. Development of a "first-principles" water potential with flexible monomers. Iii. Liquid phase properties. *J. Chem. Theory Comput.* **2014**, *10* (8), 2906−2910.
(249) Moradzadeh, A.; Aluru, N. Many-body neural network-based force field for structure-based coarse-graining of water. *J. Phys. Chem. A* **2022**, *126*, 2031.
(250) Pagonabarraga, I.; Frenkel, D. Dissipative particle dynamics for interacting systems. *J. Chem. Phys.* **2001**, *115* (11), 5015−5026.
(251) Warren, P. B. Vapor-liquid coexistence in many-body dissipative particle dynamics. *Phys. Rev. E* **2003**, *68* (6), 066702.
(252) DeLyser, M. R.; Noid, W. G. Analysis of local density potentials. *J. Chem. Phys.* **2019**, *151* (22), 224106.
(253) Vanya, P.; Elliott, J. A. Definitions of local density in density-dependent potentials for mixtures. *Phys. Rev. E* **2020**, *102* (1), 013312.
(254) Allen, E. C.; Rutledge, G. C. A novel algorithm for creating coarse-grained, density dependent implicit solvent models. *J. Chem. Phys.* **2008**, *128* (15), 154115.
(255) Allen, E. C.; Rutledge, G. C. Evaluating the transferability of coarse-grained, density-dependent implicit solvent models to mixtures and chains. *J. Chem. Phys.* **2009**, *130* (3), 034904.
(256) Allen, E. C.; Rutledge, G. C. Coarse-grained, density dependent implicit solvent model reliably reproduces behavior of a model surfactant system. *J. Chem. Phys.* **2009**, *130* (20), 204903.
(257) Izvekov, S.; Chung, P. W.; Rice, B. M. The multiscale coarse-graining method: Assessing its accuracy and introducing density dependent coarse-grain potentials. *J. Chem. Phys.* **2010**, *133* (6), 064109.
(258) Izvekov, S.; Chung, P. W.; Rice, B. M. Particle-based multiscale coarse graining with density-dependent potentials: Application to molecular crystals (hexahydro-1, 3, 5-trinitro-s-triazine). *J. Chem. Phys.* **2011**, *135* (4), 044112.
(259) Moore, J. D.; Barnes, B. C.; Izvekov, S.; Lísal, M.; Sellers, M. S.; Taylor, D. E.; Brennan, J. K. A coarse-grain force field for rdx: Density dependent and energy conserving. *J. Chem. Phys.* **2016**, *144* (10), 104501.
(260) Izvekov, S.; Rice, B. M. Bottom-up coarse-grain modeling of plasticity and nanoscale shear bands in α-rdx. *J. Chem. Phys.* **2021**, *155* (6), 064503.
(261) 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* (3), 034109.
(262) Wagner, J. W.; Dannenhoffer-Lafage, T.; Jin, J.; Voth, G. A. Extending the range and physical accuracy of coarse-grained models: Order parameter dependent interactions. *J. Chem. Phys.* **2017**, *147* (4), 044113.
(263) Dama, J. F.; Jin, J.; Voth, G. A. The theory of ultra-coarse-graining. 3. Coarse-grained sites with rapid local equilibrium of internal states. *J. Chem. Theory Comput.* **2017**, *13* (3), 1010−1022.
(264) Sanyal, T.; Shell, M. S. Transferable coarse-grained models of liquid−liquid equilibrium using local density potentials optimized with the relative entropy. *J. Phys. Chem. B* **2018**, *122* (21), 5678−5693.
(265) Jin, J.; Voth, G. A. Ultra-coarse-grained models allow for an accurate and transferable treatment of interfacial systems. *J. Chem. Theory Comput.* **2018**, *14* (4), 2180−2197.
(266) Berressem, F.; Scherer, C.; Andrienko, D.; Nikoubashman, A. Ultra-coarse-graining of homopolymers in inhomogeneous systems. *J. Phys.: Condens. Matter* **2021**, *33* (25), 254002.
(267) DeLyser, M.; Noid, W. G. Coarse-grained models for local density gradients. *J. Chem. Phys.* **2022**, *156*, 034106.
(268) Whitelam, S.; Rogers, C.; Pasqua, A.; Paavola, C.; Trent, J.; Geissler, P. L. The impact of conformational fluctuations on self-assembly: Cooperative aggregation of archaeal chaperonin proteins. *Nano Lett.* **2009**, *9* (1), 292−297.
(269) Ilie, I. M.; den Otter, W. K.; Briels, W. J. A coarse grained protein model with internal degrees of freedom. Application to α-synuclein aggregation. *J. Chem. Phys.* **2016**, *144* (8), 085103.
(270) De Gennes, P.-G.; Prost, J. *The physics of liquid crystals*; Oxford University Press: 1993; Vol. *83*.
(271) Leocmach, M.; Russo, J.; Tanaka, H. Importance of many-body correlations in glass transition: An example from polydisperse hard spheres. *J. Chem. Phys.* **2013**, *138* (12), 12A536.
(272) Xia, C.; Li, J.; Cao, Y.; Kou, B.; Xiao, X.; Fezzaa, K.; Xiao, T.; Wang, Y. The structural origin of the hard-sphere glass transition in granular packing. *Nat. Commun.* **2015**, *6* (1), 8409.
(273) Qian, H.-J.; Carbone, P.; Chen, X.; Karimi-Varzaneh, H. A.; Liew, C. C.; Müller-Plathe, F. Temperature-transferable coarse-grained potentials for ethylbenzene, polystyrene, and their mixtures. *Macromolecules* **2008**, *41* (24), 9919−9929.
(274) Farah, K.; Fogarty, A. C.; Böhm, M. C.; Müller-Plathe, F. Temperature dependence of coarse-grained potentials for liquid hexane. *Phys. Chem. Chem. Phys.* **2011**, *13* (7), 2894−2902.
(275) Jin, J.; Yu, A.; Voth, G. A. Temperature and phase transferable bottom-up coarse-grained models. *J. Chem. Theory Comput.* **2020**, *16* (11), 6823−6842.
(276) Kidder, K. M.; Szukalo, R. J.; Noid, W. G. Energetic and entropic considerations for coarse-graining. *Eur. Phys. J. B* **2021**, *94* (7), 153.
(277) Wagner, J. W.; Dama, J. F.; Voth, G. A. Predicting the sensitivity of multiscale coarse-grained models to their underlying fine-grained model parameters. *J. Chem. Theory Comput.* **2015**, *11* (8), 3547−3560.
(278) Monroe, J. I.; Hatch, H. W.; Mahynski, N. A.; Shell, M. S.; Shen, V. K. Extrapolation and interpolation strategies for efficiently estimating structural observables as a function of temperature and density. *J. Chem. Phys.* **2020**, *153* (14), 144101.

https://doi.org/10.1021/acs.jctc.2c00643
J. Chem. Theory Comput. 2022, 18, 5759−5791

(279) Agrawal, V.; Peralta, P.; Li, Y.; Oswald, J. A pressure-transferable coarse-grained potential for modeling the shock hugoniot of polyethylene. *J. Chem. Phys.* **2016**, *145* (10), 104903.
(280) Ashcroft, N. W.; Stroud, D. Theory of the thermodynamics of simple liquid metals. In *Solid state physics*; Elsevier: 1978; Vol. *33*, pp 1−81, DOI: 10.1016/S0081-1947(08)60468-3.
(281) Dunn, N. J. H.; Noid, W. G. Bottom-up coarse-grained models that accurately describe the structure, pressure, and compressibility of molecular liquids. *J. Chem. Phys.* **2015**, *143* (24), 243148.
(282) Dunn, N. J. H.; Noid, W. G. Bottom-up coarse-grained models with predictive accuracy and transferability for both structural and thermodynamic properties of heptane-toluene mixtures. *J. Chem. Phys.* **2016**, *144* (20), 204124.
(283) Rosenberger, D.; van der Vegt, N. F. A. Addressing the temperature transferability of structure based coarse graining models. *Phys. Chem. Chem. Phys.* **2018**, *20* (9), 6617−6628.
(284) Shen, K.; Sherck, N.; Nguyen, M.; Yoo, B.; Köhler, S.; Speros, J.; Delaney, K. T.; Fredrickson, G. H.; Shell, M. S. Learning composition-transferable coarse-grained models: Designing external potential ensembles to maximize thermodynamic information. *J. Chem. Phys.* **2020**, *153* (15), 154116.
(285) Mullinax, J. W.; Noid, W. G. Extended ensemble approach for deriving transferable coarse-grained potentials. *J. Chem. Phys.* **2009**, *131* (10), 104110.
(286) Stieffenhofer, M.; Bereau, T.; Wand, M. Adversarial reverse mapping of condensed-phase molecular structures: Chemical transferability. *APL Mater.* **2021**, *9*, 031107.
(287) Kanekal, K. H.; Rudzinski, J. F.; Bereau, T. Broad chemical transferability in structure-based coarse-graining. 2022, arXiv:2203.07487. *arXiv preprint*. https://arxiv.org/abs/2203.07487 (accessed 2022-08-16).
(288) Stillinger, F. H.; Sakai, H.; Torquato, S. Statistical mechanical models with effective potentials: Definitions, applications, and thermodynamic consequences. *J. Chem. Phys.* **2002**, *117* (1), 288−296.
(289) D'Adamo, G.; Pelissetto, A.; Pierleoni, C. Predicting the thermodynamics by using state-dependent interactions. *J. Chem. Phys.* **2013**, *138* (23), 234107.
(290) DeLyser, M. R.; Noid, W. G. Extending pressure-matching to inhomogeneous systems via local-density potentials. *J. Chem. Phys.* **2017**, *147* (13), 134111.
(291) Wasserman, L. *All of statistics: A concise course in statistical inference*; Springer Science & Business Media: 2013; DOI: 10.1007/978-0-387-21736-9.
(292) Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits. *J. Chem. Phys.* **2005**, *123* (6), 062201.
(293) Dama, J. F.; Sinitskiy, A. V.; McCullagh, M.; Weare, J.; Roux, B.; Dinner, A. R.; Voth, G. A. The theory of ultra-coarse-graining. 1. General principles. *J. Chem. Theory Comput.* **2013**, *9* (5), 2466−2480.
(294) Davtyan, A.; Dama, J. F.; Sinitskiy, A. V.; Voth, G. A. The theory of ultra-coarse-graining. 2. Numerical implementation. *J. Chem. Theory Comput.* **2014**, *10* (12), 5265−5275.
(295) Tully, J. C.; Preston, R. K. Trajectory surface hopping approach to nonadiabatic molecular collisions: The reaction of h+ with d2. *J. Chem. Phys.* **1971**, *55* (2), 562−572.
(296) Tully, J. C. Molecular dynamics with electronic transitions. *J. Chem. Phys.* **1990**, *93* (2), 1061−1071.
(297) McCullagh, M.; Saunders, M. G.; Voth, G. A. Unraveling the mystery of atp hydrolysis in actin filaments. *J. Am. Chem. Soc.* **2014**, *136* (37), 13053−13058.
(298) Sun, R.; Sode, O.; Dama, J. F.; Voth, G. A. Simulating protein mediated hydrolysis of atp and other nucleoside triphosphates by combining qm/mm molecular dynamics with advances in metadynamics. *J. Chem. Theory Comput.* **2017**, *13* (5), 2332−2341.
(299) Katkar, H. H.; Davtyan, A.; Durumeric, A. E. P.; Hocky, G. M.; Schramm, A. C.; Enrique, M.; Voth, G. A. Insights into the cooperative nature of atp hydrolysis in actin filaments. *Biophys. J.* **2018**, *115* (8), 1589−1602.
(300) Mani, S.; Katkar, H. H.; Voth, G. A. Compressive and tensile deformations alter atp hydrolysis and phosphate release rates in actin filaments. *J. Chem. Theory Comput.* **2021**, *17* (3), 1900−1913.
(301) Dama, J. F.; Jin, J.; Voth, G. A. Correction to the theory of ultra-coarse-graining. 3. Coarse-grained sites with rapid local equilibrium of internal states. *J. Chem. Theory Comput.* **2018**, *14* (4), 2288−2288.
(302) 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* (5), 3306−3315.
(303) Bereau, T.; Rudzinski, J. F. Accurate structure-based coarse graining leads to consistent barrier-crossing dynamics. *Phys. Rev. Lett.* **2018**, *121* (25), 256002.
(304) Rudzinski, J. F.; Bereau, T. Coarse-grained conformational surface hopping: Methodology and transferability. *J. Chem. Phys.* **2020**, *153* (21), 214110.
(305) Gupta, M.; Pak, A. J.; Voth, G. A. Critical Mechanistic Role of Inositol Hexakisphosphate (IP6) in HIV-1 Viral Capsid Assembly. *bioRxiv* **2022**, DOI: 10.1101/2022.05.03.490470v3.
(306) Lafond, P. G.; Izvekov, S. Multiscale coarse-graining with effective polarizabilities: A fully bottom-up approach. *J. Chem. Theory Comput.* **2018**, *14* (4), 1873−1886.
(307) Jin, J.; Han, Y.; Voth, G. A. Ultra-coarse-grained liquid state models with implicit hydrogen bonding. *J. Chem. Theory Comput.* **2018**, *14* (12), 6159−6174.
(308) Schatz, G. C.; Ratner, M. A. *Quantum mechanics in chemistry*; Courier Corporation: 2002.
(309) Warshel, A.; Weiss, R. M. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J. Am. Chem. Soc.* **1980**, *102* (20), 6218−6226.
(310) Jin, J.; Voth, G. A. Statistical mechanical design principles for coarse-grained interactions across different conformational surfaces. 2022, Submitted.
(311) 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* (1), 015102.
(312) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *J. Chem. Phys.* **2007**, *126* (15), 155102.
(313) Pan, A. C.; Roux, B. Building markov state models along pathways to determine free energies and rates of transitions. *J. Chem. Phys.* **2008**, *129* (6), 064107.
(314) Rains, E. K.; Andersen, H. C. A bayesian method for construction of markov models to describe dynamics on various time-scales. *J. Chem. Phys.* **2010**, *133* (14), 144113.
(315) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything you wanted to know about markov state models but were afraid to ask. *Methods* **2010**, *52* (1), 99−105.
(316) Bowman, G. R. Improved coarse-graining of markov state models via explicit consideration of statistical uncertainty. *J. Chem. Phys.* **2012**, *137* (13), 134111.
(317) He, Z.; Paul, F.; Roux, B. A critical perspective on markov state model treatments of protein−protein association using coarse-grained simulations. *J. Chem. Phys.* **2021**, *154* (8), 084101.
(318) Rudzinski, J. F.; Noid, W. G. Bottom-up coarse-graining of peptide ensembles and helix-coil transitions. *J. Chem. Theory Comput.* **2015**, *11*, 1278−1291.
(319) Zhang, Y.; Cao, Z.; Zhang, J. Z.; Xia, F. Double-well ultra-coarse-grained model to describe protein conformational transitions. *J. Chem. Theory Comput.* **2020**, *16* (10), 6678−6689.
(320) Zha, J.; Zhang, Y.; Xia, K.; Gräter, F.; Xia, F. Coarse-grained simulation of mechanical properties of single microtubules with micrometer length. *Front. Mol. Biosci.* **2021**, *7*, 632122.
(321) Zwanzig, R. Ensemble method in the theory of irreversibility. *J. Chem. Phys.* **1960**, *33* (5), 1338−1341.
(322) Zwanzig, R. Memory effects in irreversible thermodynamics. *Phys. Rev.* **1961**, *124* (4), 983.
(323) Zwanzig, R. On the identity of three generalized master equations. *Physica* **1964**, *30* (6), 1109−1123.

(324) Mori, H. Transport, collective motion, and brownian motion. *Prog. Theor. Phys.* **1965**, *33* (3), 423−455.
(325) Phillies, G. D. J. Projection operators and the mori-zwanzig formalism. In *Elementary lectures in statistical mechanics*; Springer: 2000; pp 347−364, DOI: 10.1007/978-1-4612-1264-5_32.
(326) Kawai, S.; Komatsuzaki, T. Derivation of the generalized langevin equation in nonstationary environments. *J. Chem. Phys.* **2011**, *134* (11), 114523.
(327) Izvekov, S. Microscopic derivation of particle-based coarse-grained dynamics. *J. Chem. Phys.* **2013**, *138* (13), 134106.
(328) Izvekov, S. Microscopic derivation of particle-based coarse-grained dynamics: Exact expression for memory function. *J. Chem. Phys.* **2017**, *146* (12), 124109.
(329) Meyer, H.; Voigtmann, T.; Schilling, T. On the dynamics of reaction coordinates in classical, time-dependent, many-body processes. *J. Chem. Phys.* **2019**, *150* (17), 174118.
(330) Glatzel, F.; Schilling, T. The interplay between memory and potentials of mean force: A discussion on the structure of equations of motion for coarse grained observables. *Europhys. Lett.* **2021**, *136*, 36001.
(331) Kubo, R. The fluctuation-dissipation theorem. *Rep. Prog. Phys.* **1966**, *29* (1), 255.
(332) Rudzinski, J. F. Recent progress towards chemically-specific coarse-grained simulation models with consistent dynamical properties. *Computation* **2019**, *7* (3), 42.
(333) Klippenstein, V.; Tripathy, M.; Jung, G.; Schmid, F.; van der Vegt, N. F. A. Introducing memory in coarse-grained molecular simulations. *J. Phys. Chem. B* **2021**, *125* (19), 4931−4954.
(334) Jung, G. Non-markovian systems out of equilibrium: Exact results for two routes of coarse graining. *J. Phys.: Condens. Matter* **2022**, *34* (20), 204004.
(335) Schilling, T. Coarse-grained modelling out of equilibrium. *Phys. Rep.* **2022**, *972*, 1−45.
(336) Gao, L.; Fang, W. Semi-bottom-up coarse graining of water based on microscopic simulations. *J. Chem. Phys.* **2011**, *135* (18), 184101.
(337) Li, Z.; Bian, X.; Li, X.; Karniadakis, G. E. Incorporation of memory effects in coarse-grained modeling via the mori-zwanzig formalism. *J. Chem. Phys.* **2015**, *143* (24), 243128.
(338) Yoshimoto, Y.; Kinefuchi, I.; Mima, T.; Fukushima, A.; Tokumasu, T.; Takagi, S. Bottom-up construction of interaction models of non-markovian dissipative particle dynamics. *Phys. Rev. E* **2013**, *88* (4), 043305.
(339) Davtyan, A.; Dama, J. F.; Voth, G. A.; Andersen, H. C. Dynamic force matching: A method for constructing dynamical coarse-grained models with realistic time dependence. *J. Chem. Phys.* **2015**, *142* (15), 154104.
(340) Davtyan, A.; Voth, G. A.; Andersen, H. C. Dynamic force matching: Construction of dynamic coarse-grained models with realistic short time dynamics and accurate long time dynamics. *J. Chem. Phys.* **2016**, *145* (22), 224107.
(341) Jung, G.; Hanke, M.; Schmid, F. Iterative reconstruction of memory kernels. *J. Chem. Theory Comput.* **2017**, *13* (6), 2481−2488.
(342) Jung, G.; Hanke, M.; Schmid, F. Generalized langevin dynamics: Construction and numerical integration of non-markovian particle-based models. *Soft Matter* **2018**, *14* (46), 9368−9382.
(343) Darve, E.; Solomon, J.; Kia, A. Computing generalized langevin equations and generalized fokker−planck equations. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (27), 10884−10889.
(344) Shin, H. K.; Kim, C.; Talkner, P.; Lee, E. K. Brownian motion from molecular dynamics. *Chem. Phys.* **2010**, *375* (2−3), 316−326.
(345) Bockius, N.; Shea, J.; Jung, G.; Schmid, F.; Hanke, M. Model reduction techniques for the computation of extended markov parameterizations for generalized langevin equations. *J. Phys.: Condens. Matter* **2021**, *33* (21), 214003.
(346) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Oxford University Press: 2017; DOI: 10.1093/oso/9780198803195.001.0001.
(347) Gingold, R. A.; Monaghan, J. J. Smoothed particle hydrodynamics: Theory and application to non-spherical stars. *Mon. Not. R. Astron. Soc.* **1977**, *181* (3), 375−389.
(348) Monaghan, J. J. Smoothed particle hydrodynamics. *Annu. Rev. Astron. Astrophys.* **1992**, *30* (1), 543−574.
(349) Monaghan, J. J. Smoothed particle hydrodynamics. *Rep. Prog. Phys.* **2005**, *68* (8), 1703.
(350) Español, P.; Revenga, M. Smoothed dissipative particle dynamics. *Phys. Rev. E* **2003**, *67* (2), 026705.
(351) Ellero, M.; Español, P. Everything you always wanted to know about sdpd★(★ but were afraid to ask). *Appl. Math. Mech.* **2018**, *39* (1), 103−124.
(352) Lei, H.; Caswell, B.; Karniadakis, G. E. Direct construction of mesoscopic models from microscopic simulations. *Phys. Rev. E* **2010**, *81* (2), 026704.
(353) Li, Z.; Lee, H. S.; Darve, E.; Karniadakis, G. E. Computing the non-markovian coarse-grained interactions derived from the mori−zwanzig formalism in molecular systems: Application to polymer melts. *J. Chem. Phys.* **2017**, *146* (1), 014104.
(354) Lei, H.; Yang, X.; Li, Z.; Karniadakis, G. E. Systematic parameter inference in stochastic mesoscopic modeling. *J. Comput. Phys.* **2017**, *330*, 571−593.
(355) Verlet, L. Computer″ experiments″ on classical fluids. I. Thermodynamical properties of lennard-jones molecules. *Phys. Rev.* **1967**, *159* (1), 98.
(356) Español, P.; Zúñiga, I. Obtaining fully dynamic coarse-grained models from md. *Phys. Chem. Chem. Phys.* **2011**, *13* (22), 10538−10545.
(357) Harmandaris, V. A.; Adhikari, N. P.; van der Vegt, N. F. A.; Kremer, K. Hierarchical modeling of polystyrene: From atomistic to coarse-grained simulations. *Macromolecules* **2006**, *39* (19), 6708−6719.
(358) Harmandaris, V. A.; Adhikari, N. P.; van der Vegt, N. F. A.; Kremer, K.; Mann, B. A.; Voelkel, R.; Weiss, H.; Liew, C. Ethylbenzene diffusion in polystyrene: United atom atomistic/coarse grained simulations and experiments. *Macromolecules* **2007**, *40* (19), 7026−7035.
(359) Harmandaris, V. A.; Reith, D.; Van der Vegt, N. F. A.; Kremer, K. Comparison between coarse-graining models for polymer systems: Two mapping schemes for polystyrene. *Macromol. Chem. Phys.* **2007**, *208* (19−20), 2109−2120.
(360) Harmandaris, V. A.; Kremer, K. Predicting polymer dynamics at multiple length and time scales. *Soft Matter* **2009**, *5* (20), 3920−3926.
(361) Harmandaris, V. A.; Kremer, K. Dynamics of polystyrene melts through hierarchical multiscale simulations. *Macromolecules* **2009**, *42* (3), 791−802.
(362) Harmandaris, V. A.; Floudas, G.; Kremer, K. Temperature and pressure dependence of polystyrene dynamics through molecular dynamics simulations and experiments. *Macromolecules* **2011**, *44* (2), 393−402.
(363) Fritz, D.; Harmandaris, V. A.; Kremer, K.; van der Vegt, N. F. A. Coarse-grained polymer melts based on isolated atomistic chains: Simulation of polystyrene of different tacticities. *Macromolecules* **2009**, *42* (19), 7579−7588.
(364) Fritz, D.; Koschke, K.; Harmandaris, V. A.; van der Vegt, N. F. A.; Kremer, K. Multiscale modeling of soft matter: Scaling of dynamics. *Phys. Chem. Chem. Phys.* **2011**, *13* (22), 10412−10420.
(365) Agrawal, V.; Holzworth, K.; Nantasetphong, W.; Amirkhizi, A. V.; Oswald, J.; Nemat-Nasser, S. Prediction of viscoelastic properties with coarse-grained molecular dynamics and experimental validation for a benchmark polyurea system. *J. Polym. Sci., Part B: Polym. Phys.* **2016**, *54* (8), 797−810.
(366) Lyubimov, I. Y.; McCarty, J.; Clark, A.; Guenza, M. G. Analytical rescaling of polymer dynamics from mesoscale simulations. *J. Chem. Phys.* **2010**, *132* (22), 224903.
(367) Lyubimov, I. Y.; Guenza, M. G. First-principle approach to rescale the dynamics of simulated coarse-grained macromolecular liquids. *Phys. Rev. E* **2011**, *84* (3), 031801.

(368) Lyubimov, I. Y.; Guenza, M. G. Theoretical reconstruction of realistic dynamics of highly coarse-grained cis-1, 4-polybutadiene melts. *J. Chem. Phys.* **2013**, *138* (12), 12A546.
(369) McCarty, J.; Clark, A. J.; Copperman, J.; Guenza, M. G. An analytical coarse-graining method which preserves the free energy, structural correlations, and thermodynamic state of polymer melts from the atomistic to the mesoscale. *J. Chem. Phys.* **2014**, *140* (20), 204913.
(370) Stillinger, F. H.; Weber, T. A. Hidden structure in liquids. *Phys. Rev. A* **1982**, *25* (2), 978.
(371) Stillinger, F. H. A topographic view of supercooled liquids and glass formation. *Science* **1995**, *267* (5206), 1935−1939.
(372) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* **1997**, *48* (1), 545−600.
(373) Rudzinski, J. F.; Bereau, T. The role of conformational entropy in the determination of structural-kinetic relationships for helix-coil transitions. *Computation* **2018**, *6* (1), 21.
(374) Bowman, G. R.; Pande, V. S.; Noé, F. *An introduction to markov state models and their application to long timescale molecular simulation*; Springer Science & Business Media: 2013; Vol. *797*, DOI: 10.1007/978-94-007-7606-7.
(375) Klippenstein, S. J.; Pande, V. S.; Truhlar, D. G. Chemical kinetics and mechanisms of complex systems: A perspective on recent theoretical advances. *J. Am. Chem. Soc.* **2014**, *136* (2), 528−546.
(376) Klus, S.; Nüske, F.; Koltai, P.; Wu, H.; Kevrekidis, I.; Schütte, C.; Noé, F. Data-driven model reduction and transfer operator approximation. *J. Nonlinear Sci.* **2018**, *28* (3), 985−1010.
(377) Husic, B. E.; Pande, V. S. Markov state models: From an art to a science. *J. Am. Chem. Soc.* **2018**, *140* (7), 2386−2396.
(378) Noé, F.; Nuske, F. A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Model. Simul.* **2013**, *11* (2), 635−655.
(379) Nuske, F.; Keller, B. G.; Pérez-Hernández, G.; Mey, A. S. J. S.; Noé, F. Variational approach to molecular kinetics. *J. Chem. Theory Comput.* **2014**, *10* (4), 1739−1752.
(380) Nüske, F.; Boninsegna, L.; Clementi, C. Coarse-graining molecular systems by spectral matching. *J. Chem. Phys.* **2019**, *151* (4), 044116.
(381) Rudzinski, J. F.; Kremer, K.; Bereau, T. Communication: Consistent interpretation of molecular simulation kinetics using markov state models biased with external information. *J. Chem. Phys.* **2016**, *144*, 051102.
(382) Rudzinski, J. F.; Bereau, T. Concurrent parametrization against static and kinetic information leads to more robust coarse-grained force fields. *Eur. Phys. J.: Spec. Top.* **2016**, *225* (8), 1373−1389.
(383) Rosenfeld, Y. Relation between the transport coefficients and the internal entropy of simple systems. *Phys. Rev. A* **1977**, *15* (6), 2545.
(384) Rosenfeld, Y. Comments on the transport coefficients of dense hard core systems. *Chem. Phys. Lett.* **1977**, *48* (3), 467−468.
(385) Rosenfeld, Y. A quasi-universal scaling law for atomic transport in simple fluids. *J. Phys.: Condens. Matter* **1999**, *11* (28), 5415.
(386) Li, G. X.; Liu, C. S.; Zhu, Z. G. Excess entropy scaling for transport coefficients: Diffusion and viscosity in liquid metals. *J. Non-Cryst. Solids* **2005**, *351* (10−11), 946−950.
(387) Mittal, J.; Errington, J. R.; Truskett, T. M. Thermodynamics predicts how confinement modifies the dynamics of the equilibrium hard-sphere fluid. *Phys. Rev. Lett.* **2006**, *96* (17), 177804.
(388) Mittal, J.; Errington, J. R.; Truskett, T. M. Relationships between self-diffusivity, packing fraction, and excess entropy in simple bulk and confined fluids. *J. Phys. Chem. B* **2007**, *111* (34), 10054−10063.
(389) Malvaldi, M.; Chiappe, C. Excess entropy scaling of diffusion in room-temperature ionic liquids. *J. Chem. Phys.* **2010**, *132* (24), 244502.
(390) Voyiatzis, E.; Müller-Plathe, F.; Böhm, M. C. Do transport properties of entangled linear polymers scale with excess entropy? *Macromolecules* **2013**, *46* (21), 8710−8723.
(391) Jakse, N.; Pasturel, A. Excess entropy scaling law for diffusivity in liquid metals. *Sci. Rep.* **2016**, *6*, 20689.
(392) Jin, J.; Schweizer, K. S.; Voth, G. A. Understanding dynamics in coarse-grained models: I. Universal excess entropy scaling relationship. 2022, arXiv:2208.00078. *arXiv preprint*. https://arxiv.org/abs/2208.00078 (accessed 2022-08-16).
(393) Jin, J.; Schweizer, K. S.; Voth, G. A. Understanding dynamics in coarse-grained models: II. Coarse-grained diffusion modeled using hard sphere theory. 2022, arXiv:2208.01257. *arXiv preprint*. https://arxiv.org/abs/2208.01257 (accessed 2022-08-16).
(394) Jin, J.; Lee, E. K.; Voth, G. A. Understanding dynamics in coarse-grained models: III. Roles of rotational motion and translation-rotation coupling in coarse-grained dynamics. 2022, Submitted.
(395) Wang, L.; Abel, R.; Friesner, R. A.; Berne, B. J. Thermodynamic properties of liquid water: An application of a nonparametric approach to computing the entropy of a neat fluid. *J. Chem. Theory Comput.* **2009**, *5* (6), 1462−1473.
(396) Giuffré, E.; Prestipino, S.; Saija, F.; Saitta, A. M.; Giaquinta, P. V. Entropy from correlations in tip4p water. *J. Chem. Theory Comput.* **2010**, *6* (3), 625−636.
(397) Barker, J. A.; Henderson, D. Perturbation theory and equation of state for fluids: The square-well potential. *J. Chem. Phys.* **1967**, *47* (8), 2856−2861.
(398) Barker, J. A.; Henderson, D. Perturbation theory and equation of state for fluids. Ii. A successful theory of liquids. *J. Chem. Phys.* **1967**, *47* (11), 4714−4721.
(399) Weeks, J. D.; Chandler, D.; Andersen, H. C. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.* **1971**, *54* (12), 5237−5247.
(400) Andersen, H. C.; Weeks, J. D.; Chandler, D. Relationship between the hard-sphere fluid and fluids with realistic repulsive forces. *Phys. Rev. A* **1971**, *4* (4), 1597.
(401) Weeks, J. D.; Chandler, D.; Andersen, H. C. Perturbation theory of the thermodynamic properties of simple liquids. *J. Chem. Phys.* **1971**, *55* (11), 5422−5423.
(402) Xia, W.; Song, J.; Jeong, C.; Hsu, D. D.; Phelan, F. R., Jr; Douglas, J. F.; Keten, S. Energy-renormalization for achieving temperature transferable coarse-graining of polymer dynamics. *Macromolecules* **2017**, *50* (21), 8787−8796.
(403) Xia, W.; Song, J.; Hansoge, N. K.; Phelan, F. R., Jr; Keten, S.; Douglas, J. F. Energy renormalization for coarse-graining the dynamics of a model glass-forming liquid. *J. Phys. Chem. B* **2018**, *122* (6), 2040−2045.
(404) Xia, W.; Hansoge, N. K.; Xu, W.-S.; Phelan, F. R.; Keten, S.; Douglas, J. F. Energy renormalization for coarse-graining polymers having different segmental structures. *Sci. Adv.* **2019**, *5* (4), eaav4683.
(405) Dunbar, M.; Keten, S. Energy renormalization for coarse-graining a biomimetic copolymer, poly (catechol-styrene). *Macromolecules* **2020**, *53* (21), 9397−9405.
(406) Giuntoli, A.; Hansoge, N. K.; van Beek, A.; Meng, Z.; Chen, W.; Keten, S. Systematic coarse-graining of epoxy resins with machine learning-informed energy renormalization. *npj Comput. Mater.* **2021**, *7* (1), 168.
(407) Ma, Z.; Wang, S.; Kim, M.; Liu, K.; Chen, C.-L.; Pan, W. Transfer learning of memory kernels for transferable coarse-graining of polymer dynamics. *Soft Matter* **2021**, *17*, 5864.
(408) Samanta, A.; Ali, S. M.; Ghosh, S. K. Universal scaling laws of diffusion in a binary fluid mixture. *Phys. Rev. Lett.* **2001**, *87* (24), 245901.
(409) Samanta, A.; Ali, S. M.; Ghosh, S. K. New universal scaling laws of diffusion and kolmogorov-sinai entropy in simple liquids. *Phys. Rev. Lett.* **2004**, *92* (14), 145901.
(410) Acharya, S.; Bagchi, B. Study of entropy−diffusion relation in deterministic hamiltonian systems through microscopic analysis. *J. Chem. Phys.* **2020**, *153*, 184701.
(411) Hastie, T.; Tibshirani, R.; Friedman, J. *The elements of statistical learning: Data mining, inference, and prediction*; Springer Science & Business Media: 2009; DOI: 10.1007/978-0-387-84858-7.
(412) Takács, G.; Tikk, D. Alternating least squares for personalized ranking. In *Proceedings of the sixth ACM conference on Recommender systems*; 2012; pp 83−90, DOI: 10.1145/2365952.2365972.

(413) Ceriotti, M. Unsupervised machine learning in atomistic simulations, between predictions and understanding. *J. Chem. Phys.* **2019**, *150* (15), 150901.
(414) Gkeka, P.; Stoltz, G.; Barati Farimani, A.; Belkacemi, Z.; Ceriotti, M.; Chodera, J. D.; Dinner, A. R.; Ferguson, A. L.; Maillet, J.-B.; Minoux, H. Machine learning force fields and coarse-grained variables in molecular dynamics: Application to materials and biological systems. *J. Chem. Theory Comput.* **2020**, *16* (8), 4757−4775.
(415) Wang, Y.; Ribeiro, J. M. L.; Tiwary, P. Machine learning approaches for analyzing and enhancing molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2020**, *61*, 139−145.
(416) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine learning force fields. *Chem. Rev.* **2021**, *121* (16), 10142−10186.
(417) Behler, J. Four generations of high-dimensional neural network potentials. *Chem. Rev.* **2021**, *121* (16), 10037−10072.
(418) Clark, A. E.; Adams, H.; Hernandez, R.; Krylov, A. I.; Niklasson, A. M.; Sarupria, S.; Wang, Y.; Wild, S. M.; Yang, Q. The middle science: Traversing scale in complex many-body systems. *ACS Cent. Sci.* **2021**, *7* (8), 1271−1287.
(419) LeCun, Y.; Chopra, S.; Hadsell, R.; Ranzato, M.; Huang, F. *A tutorial on energy-based learning*; MIT Press: 2006; Vol. *1*.
(420) Song, Y.; Kingma, D. P. How to train your energy-based models. 2021, arXiv:2101.03288. *arXiv preprint*. https://arxiv.org/abs/2101.03288 (accessed 2022-08-16).
(421) John, S. T.; Csányi, G. Many-body coarse-grained interactions using gaussian approximation potentials. *J. Phys. Chem. B* **2017**, *121* (48), 10934−10949.
(422) Scherer, C.; Scheid, R.; Andrienko, D.; Bereau, T. Kernel-based machine learning for efficient simulations of molecular liquids. *J. Chem. Theory Comput.* **2020**, *16* (5), 3194−3204.
(423) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **2017**, *3* (5), e1603015.
(424) Wang, J.; Chmiela, S.; Müller, K.-R.; Noé, F.; Clementi, C. Ensemble learning of coarse-grained molecular dynamics force fields with a kernel approach. *J. Chem. Phys.* **2020**, *152* (19), 194106.
(425) Gutmann, M.; Hyvärinen, A. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of 13th International Conference on Artificial Intelligence and Statistics, JMLR Work. Conf. Proc.*: 2010; pp 297−304.
(426) Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63* (11), 139−144.
(427) Schneider, E.; Dai, L.; Topper, R. Q.; Drechsel-Grau, C.; Tuckerman, M. E. Stochastic neural network approach for learning high-dimensional free energy surfaces. *Phys. Rev. Lett.* **2017**, *119* (15), 150601.
(428) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deepcg: Constructing coarse-grained models via deep neural networks. *J. Chem. Phys.* **2018**, *149* (3), 034101.
(429) Zhang, L.; Han, J.; Wang, H.; Car, R.; Weinan, E. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **2018**, *120* (14), 143001.
(430) Baldi, P. Autoencoders, unsupervised learning, and deep architectures. In *Proceedings of ICML workshop on unsupervised and transfer learning, JMLR Work. Conf. Proc.*; 2012; pp 37−49.
(431) Doersch, C. Tutorial on variational autoencoders. 2016, arXiv:1606.05908. *arXiv preprint* https://arxiv.org/abs/1606.05908 (accessed 2022-08-16).
(432) Theis, L.; Shi, W.; Cunningham, A.; Huszár, F. Lossy image compression with compressive autoencoders. 2017, arXiv preprint. *arXiv:1703.00395*. https://arxiv.org/abs/1703.00395 (accessed 2022-08-16).
(433) Wang, W.; Gómez-Bombarelli, R. Coarse-graining auto-encoders for molecular dynamics. *npj Comput. Mater.* **2019**, *5* (1), 125.
(434) Husic, B. E.; Charron, N. E.; Lemm, D.; Wang, J.; Pérez, A.; Majewski, M.; Krämer, A.; Chen, Y.; Olsson, S.; de Fabritiis, G. Coarse graining molecular dynamics with graph neural networks. *J. Chem. Phys.* **2020**, *153* (19), 194101.
(435) Ruza, J.; Wang, W.; Schwalbe-Koda, D.; Axelrod, S.; Harris, W. H.; Gómez-Bombarelli, R. Temperature-transferable coarse-graining of ionic liquids with dual graph convolutional neural networks. *J. Chem. Phys.* **2020**, *153* (16), 164501.
(436) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. Schnet−a deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148* (24), 241722.
(437) Doerr, S.; Majewski, M.; Pérez, A.; Krämer, A.; Clementi, C.; Noe, F.; Giorgino, T.; De Fabritiis, G. Torchmd: A deep learning framework for molecular simulations. *J. Chem. Theory Comput.* **2021**, *17* (4), 2355−2363.
(438) 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* (8), 084101.
(439) Li, Z.; Wellawatte, G. P.; Chakraborty, M.; Gandhi, H. A.; Xu, C.; White, A. D. Graph neural network based coarse-grained mapping prediction. *Chem. Sci.* **2020**, *11* (35), 9524−9531.
(440) Webb, M. A.; Delannoy, J.-Y.; De Pablo, J. J. Graph-based approach to systematic molecular coarse-graining. *J. Chem. Theory Comput.* **2019**, *15* (2), 1199−1208.
(441) Empereur-Mot, C.; Pesce, L.; Doni, G.; Bochicchio, D.; Capelli, R.; Perego, C.; Pavan, G. M. Swarm-cg: Automatic parametrization of bonded terms in martini-based coarse-grained models of simple to complex molecules via fuzzy self-tuning particle swarm optimization. *ACS omega* **2020**, *5* (50), 32823−32843.
(442) Empereur-Mot, C.; Capelli, R.; Perrone, M.; Caruso, C.; Doni, G.; Pavan, G. M. Automatic multi-objective optimization of coarse-grained lipid force fields using swarmcg. *J. Chem. Phys.* **2022**, *156* (2), 024801.
(443) 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* (16), 164113.
(444) Li, Z.; Kermode, J. R.; De Vita, A. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* **2015**, *114* (9), 096405.
(445) Podryabinkin, E. V.; Shapeev, A. V. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.* **2017**, *140*, 171−180.
(446) Miwa, K.; Ohno, H. Interatomic potential construction with self-learning and adaptive database. *Phys. Rev. Mater.* **2017**, *1* (5), 053801.
(447) Gubaev, K.; Podryabinkin, E. V.; Shapeev, A. V. Machine learning of molecular properties: Locality and active learning. *J. Chem. Phys.* **2018**, *148* (24), 241727.
(448) Zhang, L.; Lin, D.-Y.; Wang, H.; Car, R.; Weinan, E. Active learning of uniformly accurate interatomic potentials for materials simulation. *Phys. Rev. Mater.* **2019**, *3* (2), 023804.
(449) Seung, H. S.; Opper, M.; Sompolinsky, H. Query by committee. In *Proceedings of the fifth annual workshop on Computational learning theory*; 1992; pp 287−294, DOI: 10.1145/130385.130417.
(450) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.* **1989**, *156* (5), 472−477.
(451) Bause, M.; Bereau, T. Reweighting non-equilibrium steady-state dynamics along collective variables. *J. Chem. Phys.* **2021**, *154*, 134105.
(452) Moradzadeh, A.; Aluru, N. R. Transfer-learning-based coarse-graining method for simple fluids: Toward deep inverse liquid-state theory. *J. Phys. Chem. Lett.* **2019**, *10* (6), 1242−1250.
(453) Jeong, J.; Moradzadeh, A.; Aluru, N. Extended deepilst for various thermodynamic states and applications in coarse-graining. *J. Phys. Chem. A* **2022**, *126* (9), 1562−1570.
(454) Younes, L. On the convergence of markovian stochastic algorithms with rapidly decreasing ergodicity rates. *Stochastics* **1999**, *65* (3−4), 177−228.
(455) Wang, W.; Xu, M.; Cai, C.; Miller, B. K.; Smidt, T.; Wang, Y.; Tang, J.; Gómez-Bombarelli, R. Generative coarse-graining of

molecular conformations. 2022, arXiv preprint. *arXiv:2201.12176*. https://arxiv.org/abs/2201.12176 (accessed 2022-08-16).
(456) Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A. A. Image-to-image translation with conditional adversarial networks. In *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*; 2017; pp 1125−1134, DOI: 10.1109/CVPR.2017.632.
(457) Li, W.; Burkhart, C.; Polińska, P.; Harmandaris, V.; Doxastakis, M. Backmapping coarse-grained macromolecules: An efficient and versatile machine learning approach. *J. Chem. Phys.* **2020**, *153* (4), 041101.
(458) Stieffenhofer, M.; Wand, M.; Bereau, T. Adversarial reverse mapping of equilibrated condensed-phase molecular structures. *Mach. Learn.: Sci. Technol.* **2020**, *1* (4), 045014.
(459) Louison, K. A.; Dryden, I. L.; Laughton, C. A. Glimps: A machine learning approach to resolution transformation for multiscale modeling. *J. Chem. Theory Comput.* **2021**, *17* (12), 7930−7937.
(460) Peng, J.; Yuan, C.; Ma, R.; Zhang, Z. Backmapping from multiresolution coarse-grained models to atomic structures of large biomolecules by restrained molecular dynamics simulations using bayesian inference. *J. Chem. Theory Comput.* **2019**, *15* (5), 3344−3353.
(461) An, Y.; Deshmukh, S. A. Machine learning approach for accurate backmapping of coarse-grained models to all-atom models. *Chem. Commun.* **2020**, *56* (65), 9312−9315.
(462) Noé, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **2019**, *365* (6457), aaw1147 DOI: 10.1126/science.aaw1147.
(463) Durumeric, A. E. P.; Voth, G. A. Explaining classifiers to understand coarse-grained models. 2021, arXiv:2109.07337. *arXiv preprint*. https://arxiv.org/abs/2109.07337 (accessed 2022-08-16).
(464) Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1* (5), 206−215.
(465) Allen, A. E.; Dusson, G.; Ortner, C.; Csányi, G. Atomic permutationally invariant polynomials for fitting molecular force fields. *Mach. Learn. Sci. Technol.* **2021**, *2* (2), 025017.
(466) Kodratoff, Y. The comprehensibility manifesto. *KDD Nugget Newsletter*; 1994; Vol. *94* (9).
(467) Rüping, S. Learning interpretable models. PhD Thesis, Univ. Dortmund, 2006; DOI: 10.17877/DE290R-8863.
(468) Molnar, C. *Interpretable machine learning*; Lulu.com, 2020 (accessed 2022-09-03).
(469) Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Inf. Fusion* **2020**, *58*, 82−115.
(470) Molnar, C.; Casalicchio, G.; Bischl, B. Interpretable machine learning−a brief history, state-of-the-art and challenges. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*; Springer: 2020; pp 417−431, DOI: 10.1007/978-3-030-65965-3_28.
(471) Rudin, C.; Chen, C.; Chen, Z.; Huang, H.; Semenova, L.; Zhong, C. Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statist. Surv.* **2022**, *16*, 1−85.
(472) Bartók, A. P.; Kermode, J.; Bernstein, N.; Csányi, G. Machine learning a general-purpose interatomic potential for silicon. *Phys. Rev. X* **2018**, *8* (4), 041048.
(473) Han, Y.; Jin, J.; Wagner, J. W.; Voth, G. A. Quantum theory of multiscale coarse-graining. *J. Chem. Phys.* **2018**, *148* (10), 102335.
(474) Jackson, N. E.; Bowen, A. S.; Antony, L. W.; Webb, M. A.; Vishwanath, V.; de Pablo, J. J. Electronic structure at coarse-grained resolutions from supervised machine learning. *Sci. Adv.* **2019**, *5* (3), eaav1190.
(475) Sivaraman, G.; Jackson, N. E. Coarse-grained density functional theory predictions via deep kernel learning. *J. Chem. Theory Comput.* **2022**, *18* (2), 1129−1141.
(476) Jackson, N. E. Coarse-graining organic semiconductors: The path to multiscale design. *J. Phys. Chem. B* **2021**, *125* (2), 485−496.
(477) Jackson, N. E.; Bowen, A. S.; De Pablo, J. J. Efficient multiscale optoelectronic prediction for conjugated polymers. *Macromolecules* **2020**, *53* (1), 482−490.
(478) Yu, A.; Pak, A. J.; He, P.; Monje-Galvan, V.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Amaro, R. E.; Voth, G. A. A multiscale coarse-grained model of the sars-cov-2 virion. *Biophys. J.* **2021**, *120* (6), 1097−1104.
(479) Sinitskiy, A. V.; Voth, G. A. Quantum mechanics/coarse-grained molecular mechanics (qm/cg-mm). *J. Chem. Phys.* **2018**, *148* (1), 014102.
(480) Mironenko, A. V.; Voth, G. A. Density functional theory-based quantum mechanics/coarse-grained molecular mechanics: Theory and implementation. *J. Chem. Theory Comput.* **2020**, *16* (10), 6329−6342.
(481) Sherck, N.; Shen, K.; Nguyen, M.; Yoo, B.; Köhler, S.; Speros, J. C.; Delaney, K. T.; Shell, M. S.; Fredrickson, G. H. Molecularly informed field theories from bottom-up coarse-graining. *ACS Macro Lett.* **2021**, *10*, 576−583.
(482) Rahman, A. Correlations in the motion of atoms in liquid argon. *Phys. Rev.* **1964**, *136* (2A), A405.
(483) Mirzoev, A.; Lyubartsev, A. P. Magic: Software package for multiscale modeling. *J. Chem. Theory Comput.* **2013**, *9* (3), 1512−1520.
(484) Mirzoev, A.; Nordenskiöld, L.; Lyubartsev, A. Magic v. 3: An integrated software package for systematic structure-based coarse-graining. *Comput. Phys. Commun.* **2019**, *237*, 263−273.
(485) Dunn, N. J. H.; Lebold, K. M.; DeLyser, M. R.; Rudzinski, J. F.; Noid, W. G. Bocs: Bottom-up open-source coarse-graining software. *J. Phys. Chem. B* **2018**, *122* (13), 3363−3377.
(486) Peng, Y.; Pak, A. J.; Durumeric, A. E. P.; Mani, S.; Jin, J.; Loose, T.; Sahrmann, P. G.; Beiter, J.; Voth, G. A. Openmscg: A software tool for multiscale coarse-graining with high performance and reproducibility. 2022, *Submitted*.
(487) Brooks, C.; Case, D. A. Simulations of peptide conformational dynamics and thermodynamics. *Chem. Rev.* **1993**, *93* (7), 2487−2502.
(488) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. Atomic-level characterization of the structural dynamics of proteins. *Science* **2010**, *330* (6002), 341−346.
(489) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334* (6055), 517−520.
(490) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **2011**, *100* (9), L47−L49.
(491) Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J. Millisecond-scale molecular dynamics simulations on anton. In *Proceedings of the conference on high performance computing networking, storage and analysis*, 2009; pp 1−11, DOI: 10.1145/1654059.1654099.