Communications in Computational Chemistry
Commun. Comput. Chem. Vol. 8, No. 1, 25–36
doi: 10.4208/cicc.2025.273.01

Review

Advances in Ultra-Coarse-Grained Models for Large Biomolecules

Yuwei Zhang[1,2,*] and Fei Xia[3]

[1]Jiangsu Key Laboratory of New Power Batteries, Jiangsu Collaborative Innovation Centre of Biomedical Functional Materials, School of Chemistry and Materials Science, Nanjing Normal University, Wenyuan Road No. 1, Nanjing 210023, People's Republic of China;
[2]Key Laboratory of NSLSCS, Ministry of Education, Nanjing Normal University, Wenyuan Road No. 1, Nanjing 210023, People's Republic of China;
[3]School of Chemistry and Molecular Engineering, NYU-ECNU Center for Computational Chemistry at NYU Shanghai, East China Normal University, Shanghai 200062, People's Republic of China.

[*] Corresponding author: ywzhang@nnu.edu.cn.

Received 30 Sept. 2025; Accepted (in revised version) 19 Nov. 2025

Abstract: Recent advances in Ultra-Coarse-Graining (UCG) modeling for biological systems have improved both construction strategies and forcefield development. Empirical forcefields remain the primary choice for large systems due to their efficiency and ability to capture conformational dynamics, while non-restraining potentials and multiscale approaches have enhanced predictive capability for complex intermolecular interactions. Although bottom-up methods are currently limited by high parameterization costs, increasing physical interpretability and the growth of biomolecular trajectory databases are expected to make them more feasible and transferable in the future.

Key words: Ultra-coarse-grained models, proteins,bottom-up, coarse-graining, empirical coarse-grained models.

1. Introduction

Investigating the conformational dynamics of large biomolecules is essential for revealing the underlying mechanisms for critical macroscopic biological behaviors [1,2], including enzymatic catalysis [3-6], flexible docking [7-10], and the mechanical responses of cytoskeletal structures [11-14]. This investigation necessitates a molecular-level analysis of the conformational ensembles and their evolution across multiple spatial/temporal scales [15,16], grounded in the rigorous principles of physical chemistry and statistical mechanics [17-24]. A widely employed strategy is to perform extensive molecular dynamics (MD) simulations [25-29] under the assumption of ergodicity, wherein the appropriate choice of computational chemistry model is determined by the resolution required for the investigation of the specific properties. Quantum mechanical (QM) models offer accurate descriptions of electronic state transitions and chemical reactions of active sites [5,6,30,31]. All-atom (AA) models based on molecular mechanics, are well-suited for capturing localized conformational dynamics of large biomolecules [32]. Coarse-grained (CG) models [33-36] enable the efficient description of global conformational transitions and supramolecular assembly [37]. Given that the functional processes of large biomolecular systems often take place in micrometer and microsecond scales or even longer [38-41], their modeling and simulation presents significant challenges due to the prohibitive computational cost. Thus, CG models have gained growing interest and widespread adoption, as they offer a favorable compromise between computational efficiency and the ability to accurately capture essential structural and dynamic properties.

The principle of CG modeling lies in simplifying the representation of non-functional regions by reducing the system's degrees of freedom [42], thereby significantly boosting computational efficiency while aiming to preserve an accurate depiction of the structural and dynamic features within key functional domains. Therefore, the mapping strategies employed in CG models are largely guided by the structural characteristics of the target systems and the specific physical properties that need to be accurately captured. Given that biological systems are typically composed of a limited number of monomer types, such as amino acid residues or nucleobases, a widely adopted and generalizable CG modeling strategy, analogous to that employed in AA models, is to map these transferable monomers to CG resolution and use them as the fundamental building blocks for constructing large biomolecular systems. Such modeling strategies are designed to capture the geometric features of individual monomers and, to some extent, accurately reproduce the interactions between their chemical groups. Accurately predicting protein conformational changes typically requires high-resolution models that either preserve hydrogen-bond donor and acceptor atoms or employ finely tuned anisotropic (non-spherical) interaction potentials. Popular models, such as AMBER-UA [43], GBEMP [44,45], PACE [46,47], PRIMO [48,49] and UNRes [50], have been successfully applied to a variety of researches, including peptide folding and large-scale conformational transitions. CG models for simulating membrane or nucleic acid systems typically employ a resolution in which each bead represents two to three heavy atoms, such as Martini [51,52], SIRAH [53], oxDNA [54,55] and oxRNA [56]. This CG resolution enables the semi-quantitative description of essential interactions between key functional groups. Overall, the aforementioned high-resolution CG modeling strategies have been widely applied, and numerous well-organized review articles have been published [23,34-36,57-59]. Therefore, they will not be discussed in detail in this work.

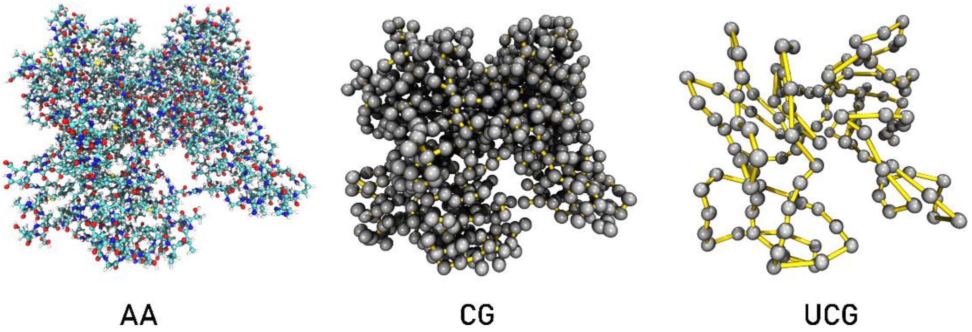

Figure 1. Structural representations of the G-actin monomer (PDB ID: 1ATN) at AA, high-resolution CG, and UCG resolutions.

However, due to computational limitations, CG models with resolutions higher than the residue level still face significant challenges in simulating complex biological processes involving multiple or large-scale proteins. To further enhance simulation efficiency and enable the description of large-scale conformational changes, Ultra-Coarse-Grained (UCG) [60-63] models, characterized by even higher levels of simplification, have been proposed in recent years. These models typically perform coarse-graining at the level of the entire biomacromolecule [64,65], employ customized mapping schemes designed to capture global molecular shape, secondary structures, and essential dynamic modes with minimal computational overhead. A comparison of UCG, higher-resolution CG, and AA representations is shown in **Figure 1**. Some UCG models developed for studying mechanical properties can represent an entire protein using fewer than ten particles, thereby significantly increasing the integration timestep and extending the accessible simulation timescales. UCG models have been successfully applied to describe global conformational changes in supramolecular assemblies or mechanical properties of macro-biomolecules. Carmichael demonstrated that a sidechain-free UCG model is possible to reproduce the folding dynamics of short peptides [66]. Voth and co-workers systematically investigated microscopic mechanisms of HIV-1 self-assembly and further discussed how capsid inhibitors modulate this process [64,67]. UCG models have also been employed to construct mesoscale representations of SARS-CoV-2 virus particle that preserve explicit conformational detail of the spike protein [65,68]. Bryer et al. introduced a division-flexible UCG framework that facilitates the modeling and simulation of the mesoscale mechanical properties of both HIV-1 and cofilin-2 [69]. Xia and co-workers employed UCG models to characterize the mesoscale torsional behavior of collagen and, respectively, the local conformational changes inside protein monomers under stretching process [11,38]. With the increasing demand for mechanistic insights at the molecular level in life sciences, the development of UCG models offers great potential for elucidating mesoscale functional dynamics of complex biomolecular systems.

While UCG models have demonstrated strong applicability and descriptive power for certain biomolecular systems, they still face significant challenges in predictive accuracy and transferability across different systems. Because each UCG particle typically represents multiple residues, usually corresponding to a unique particle type, UCG models must be customized according to their specific division. Unlike high-resolution CG models, they cannot be generated in a high-throughput way by taking the advantage of predefined standardized residues and sequence information. Significant progress has been made in developing general division algorithms for finding proper UCG representations and resolutions of bulk biomolecules that efficiently capture major conformational transitions. However, determining accurate and transferable UCG forcefields remains an open problem. The data requirements for accurately parameterizing UCG forcefields are prohibitively high, thereby limiting their application to relatively small systems containing only a few dozen residues. Empirical forcefields with lower data requirements for construction have been more widely adopted in UCG models. Although this simplification may limit their ability to accurately predict global conformational changes within individual proteins, such forcefields maintain strong predictive capability in simulating supramolecular behaviors involving multiple proteins.

This review summarizes recent advances in the UCG division algorithms, forcefield development. We also highlight the key challenges and opportunities in this field.

2. UCG mapping

Due to the highly customized and sub-residue resolution, UCG models cannot be constructed by simply assembling common monomeric units such as residues or nucleotides. The UCG representation of arbitrary biomolecules, conceptualized as a linear transformation from the fine-grained Cartesian coordinates, are often optimized on-the-fly using clustering techniques like residue sequence segmentation or Voronoi cells. To determine the optimal division, an error function [70] should be defined to quantify the accuracy of UCG representation, and a global optimization algorithm should be employed to find the global extremum. The definition of error function typically reflects the structural or dynamical similarity between the target UCG and AA model, aiming to assess whether molecular shapes and key conformational changes are properly projected in the low-dimensional phase space.

A class of UCG division methods aims to reproduce structural information, such as density maps [71] and topological networks [72], from high-resolution models, thereby maximizing the preservation of molecular shape information, as illustrated in the first-row workflow of **Figure 2**. These methods operate by first converting the all-atom (AA) model into a 3D density map, followed by clustering the density grids into a user-defined number of UCG beads using a vector quantization (VQ) [73] algorithm. The shape-based coarse-graining (SB-CG) method proposed by Arkhipov et al [74]. employs a neural network–inspired vector quantization (VQ) algorithm, enabling the accurate reproduction of AA mass density distributions. Building upon this framework, Bryer et al. introduced SB-CG2 [69], which incorporates an improved optimization strategy and a new quantitative metric to assess the similarity between AA and UCG density maps, thereby facilitating the optimization of UCG resolution. The SB-CG/SB-CG2 division algorithms have been successfully applied to the study of mesoscale bioprocesses, including membrane deformation [75], bacterial flagellar motion [76], mechanical properties of viral capsids [77], and actin filament stretching [69]. It is worth noting that, since density maps can also be obtained experimentally via electron microscopy (EM), such methods can be directly applied to construct UCG models from low-resolution EM densities in cases where AA models are not available. For example, the approach proposed by Ming et al [71]. enables UCG partitioning directly from EM density maps and accurately reproduces the corresponding vibrational modes. The convolution K-means coarse-graining (CK-CG) method developed by Zhang et al [78]. enables the coarse-graining of mesoscale macromolecules with uniform mass distribution by employing a convolution-based initialization of UCG positions. The ED-CG-EM method proposed by Zhang et al [79]. performs clustering based on vibrational modes extracted from EM density grids, yielding EM-based UCG divisions that can accurately reproduce the intrinsic dynamical behaviors. It is important to note that bonded information is typically undefined in EM-based UCG partitioning methods and they are therefore generally restricted to using elastic network models (ENMs) [80-82] for forcefield construction, which will be discussed in **Section 4**.

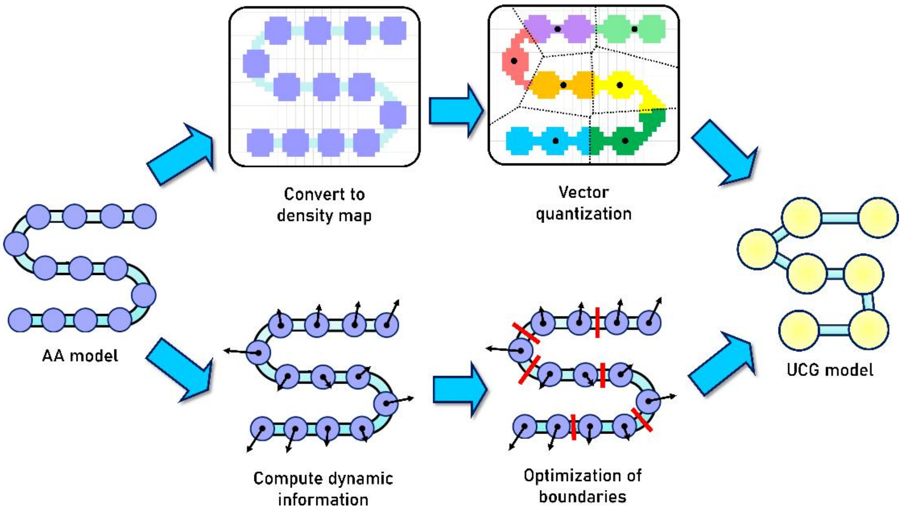

Figure 2. Schematic for the transformation of AA models into UCG representations.

Another class of UCG division methods aims to reproduce dynamical features from high-resolution models. These approaches can be understood as clustering algorithms operating in momentum space, as illustrated in the second-row workflow of **Figure 2**. They require not only high-resolution structural information but also detailed dynamical data, which can be obtained through MD simulations or normal mode analysis (NMA). Such methods are primarily applied to proteins, where the partitioning algorithm focuses on optimizing the placement of “boundaries” along the backbone so that each UCG bead captures the collective motion of its corresponding segment. Compared to purely structure-based partitioning approaches, these dynamics-informed methods incorporate molecular flexibility and are thus better suited for describing global conformational transitions in flexible molecules. Zhang et al [83,84]. proposed the essential dynamics coarse-graining (ED-CG) method, which constructs UCG mappings by grouping atoms that share similar essential vibrational modes into the same UCG bead. This strategy enables each UCG bead to capture the coherent motions of local structural domains. Subsequently, building upon the ED-CG framework, Wu et al [85]. proposed the K-means clustering coarse-graining (KMC-CG) method, which incorporates both structural and topological information. ED-CG has been widely employed in studying the conformational dynamics and NMA of large biomolecular assemblies, including actin filament contraction and extension [84,86], ribosomal motion patterns [87], conformational transitions of the SARS-CoV-2 spike protein [65,88], and the assembly of HIV-1 Gag proteins [89,90]. Li et al [91]. proposed the fluctuation maximization coarse-graining (FM-CG) method, which aims to maximize the pairwise fluctuations between UCG particles. This strategy enables the mapping to distinguish between flexible and rigid domains—for instance, α-helices and β-sheets—and to represent them at different resolutions [70,92]. As a result, FM-CG facilitates the development of UCG models capable of capturing large-scale conformational transitions, such as the open-to-closed motion of the adenylate kinase (AK) protein [63]. Additionally, several studies have explored the determination of optimal number of UCG beads. For example, Li et al [91,93]. developed a series of stepwise optimizers capable of efficiently generating coarse-grained representations across multiple resolutions in a high-throughput manner. Wu et al [94]. further demonstrated that the ultra-coarse-graining process can be statistically interpreted as a clustering problem, with ED-CG and FM-CG corresponding to minimizing within-cluster variation and maximizing between-cluster variation, respectively. And the optimal UCG representation can be systematically identified using statistical metrics such as the Calinski–Harabasz index [95].

Overall, current UCG division algorithms exhibit broad applicability across diverse biomolecular systems and are capable of efficiently generating UCG representations. However, the rapid and reliable determination of the optimal number [85,94,96] of UCG beads remains a significant challenge. This difficulty arises primarily because the ideal resolution is contingent on the specific properties of interest and the nature of the employed forcefield, for example, capturing large-scale functional conformational changes often requires more beads than simulating equilibrium fluctuations. Mathematically driven approaches have not fully addressed this issue. Integrating current UCG division strategies with forcefield fitting or experimental properties holds promise for substantially improving both the accuracy and utility of UCG division algorithms.

3. Bottom-up forcefields

Once an appropriate UCG mapping matrix is established, the next essential step is to determine the effective interactions between CG beads. A widely adopted strategy is to microscopically fit these interactions based on systematically accessible potential energy surface data obtained from high-resolution models. This approach, commonly referred to as the bottom-up methods [23,34,36,97,98], supports the construction of CG forcefields under arbitrary division matrix $\boldsymbol{\mathcal{M}}$, and is therefore methodologically generalizable across a wide range of systems. One of the most representative implementations of this strategy is the multiscale coarse-graining (MS-CG) method developed by Voth and co-workers [99-101]. The MS-CG method reflects a conceptually simple yet rigorous idea: when the CG forcefield is constructed to reproduce the AA free energy surface, a strict statistical mechanical correspondence exists between the CG potential $U^{CG}(\boldsymbol{R})$ and the partition function of the AA model, as shown in Eq. (1). Here, $U^{AA}(\boldsymbol{r})$ denotes the AA potential energy, $\boldsymbol{R}$ and $\boldsymbol{r}$ represent the coordinates in the CG and AA models, respectively. A group of other bottom-up approaches are theoretically equivalent to MS-CG, including Iterative Boltzmann Inversion (IBI) [102], the generalized Yvon–Born–Green (YBG) theory [103], and entropy minimization methods [104]. Several recent review articles have provided comprehensive overviews of their theoretical foundations, methodological advances, and practical applications [23,34,36,97,98]. Owing to their physically grounded treatment of the microscopic interactions, bottom-up approaches generally achieve high accuracy in predicting a wide range of thermo and dynamic properties. Most bottom-up methods for constructing UCG models are implemented in the OpenMSCG package [105] and are fully compatible with the LAMMPS simulation package [106].

$$U^{CG}_{bottom-up}(\boldsymbol{R}) = -\beta^{-1} ln\left[\int \delta(\boldsymbol{R} - \boldsymbol{\mathcal{M}}\boldsymbol{r}) e^{-\beta U^{AA}(\boldsymbol{r})} d\boldsymbol{r}\right] \qquad (1)$$

A key challenge in bottom-up forcefield parameterization lies in obtaining the partition function of the AA reference system. For complex biomolecular systems, computing the full AA partition function is practically infeasible. In practice, AA-MD simulations are typically used as an approximation to sample low-energy conformational ensembles. To ensure that the CG potential accurately captures the target conformational transitions, the AA-MD simulations must sufficiently explore not only local energy minima but also all transition pathways. However, the global conformational changes typically described by UCG models occur on spatial and temporal scales that far exceed the capabilities of AA simulations. As a result, pre-sampling is generally restricted to local equilibrium states, making it challenging to access intermediate or transition-state conformations without prior knowledge. Furthermore, the absence of comprehensive databases for AA conformational ensembles significantly limits the applicability of bottom-up approaches in high-throughput CG modeling. This transferability limitation becomes particularly critical for UCG models, where the cost of extensive pre-sampling is prohibitively high. Although recent efforts have focused on developing open-access conformational databases, these initiatives remain in their early stages and require sustained data accumulation. Currently, bottom-up methods are primarily employed to accurately parameterize native interactions in structurally stable regions, such as intra-molecular bonded interactions, or to model relatively simple systems, including the folding of short peptides.

$$U^{CG}_{bottom-up}(\boldsymbol{R}) = U^{prior}(\boldsymbol{R}) + U^{ml}(\boldsymbol{R}) \qquad (2)$$

One promising strategy to address the challenge of insufficient sampling involves enhancing the physical interpretability of the potential energy function, thereby improving its ability to generalize across unseen conformational space. This approach is conceptually aligned with the Δ-learning framework [107] commonly employed in all-atom machine-learned potentials (MLPs) [108,109]. In this scheme [110], a machine-learning-based correction potential $U^{ml}(\boldsymbol{R})$ is trained to match the difference between a prior potential $U^{prior}(\boldsymbol{R})$ and the target effective bottom-up CG potential $U^{CG}_{bottom-up}(\boldsymbol{R})$, as shown in Eq. (2). $U^{prior}(\boldsymbol{R})$ is typically represented by classical forcefields to ensure strong physical interpretability, while the correction potential $U^{ml}(\boldsymbol{R})$ is often modeled using polynomials or neural network potentials with rotational and translational invariance to accurately capture details of the potential energy landscape and account for many-body effects [111], such as CGnet [112], GDML [113], CGSchNet [114]. The underfitting problem is more pronounced in UCG models than in high-resolution CG models, therefore the construction of a physically interpretable and accurate $U^{prior}(\boldsymbol{R})$ is particularly critical. Unlike Δ-learning approaches based on low-level QM methods, $U^{prior}(\boldsymbol{R})$ is difficult to construct using predefined, systematically low-level models. In UCG modeling, it is typically derived by composing the effective potential into physical contributions such as repulsion, bonded interactions, and point charges. Therefore, it is critically important to develop potential-based matching or decomposition strategies tailored to specific physical contributions for UCG modeling.

A series of recently developed potential-based coarse-graining methods have been proposed to parameterize specific physical contributions in CG models even without MD simulation information, particularly for nonbonded interactions. These approaches facilitate the rapid and customizable generation of high-precision physical-based CG models, offering a promising strategy to address the challenges associated with defining $U^{prior}(\boldsymbol{R})$. UCG Among them, various electrostatic potential (ESP)-based matching methods have been introduced to estimate the point charges of CG beads [115]. These approaches can effectively reproduce the ESP distribution of the corresponding atomistic models. However, due to the intrinsic limitations of point charge representations, their accuracy remains insufficient for highly polarized UCG particles. Incorporating higher-order multipole effects may offer a promising direction for future improvements. For the Lennard-Jones (LJ) potential, Zhang et al [116]. developed the Lennard-Jones static potential matching (LJSPM) method, which can be considered an LJ-based analogue of ESP matching. This enables rapid parameterization of CG-level LJ interactions and allows accurate reproduction of pairwise LJ interactions between functional groups [9,117,118]. In addition, the method developed by Tillack et al [119]. enables efficient parameterization of the Gay-Berne ellipsoidal potential for UCG particles and has demonstrated excellent performance in systems such as benzene. Given the inherent anisotropy and irregular shapes of UCG particles, this approach represents a highly promising direction for future development.

Moreover, the inherently coarse representations of UCG particles tend to magnify certain physicochemical effects that are often negligible in fine-grained CG models, such as conformational entropy [120], multistate particle behavior [121]. Exploration of these effects represents a cutting-edge direction in current research and contributes significantly to our understanding of the fundamental of UCG interactions. Significant advancements have been made in the investigation of these effects. One important discovery is the multistate behavior of UCG particles, arising from substantial internal conformational fluctuations that are significantly more pronounced than those in high-level CG models. These fluctuations have a non-negligible influence on global conformational transitions and cannot be accurately captured by the mean-field approximation employed in Eq. (1). To address this limitation, Voth and collaborators developed a multistate UCG modeling framework that extends the MS-CG method by incorporating an Ehrenfest-like approach to explicitly account for the contributions of internal conformational dynamics within UCG particles. Notably, this strategy also offers an effective means of describing chemical reactions within a UCG framework. Furthermore, they identified that conformational entropy makes a substantial contribution to the effective potential. By explicitly removing this entropic component, the resulting potential exhibits markedly improved temperature transferability. As such entropic effects are expected to be even more pronounced in coarser CG models, developing efficient strategies for estimating these contributions is of critical importance.

Currently, the use of bottom-up approaches in UCG modeling has been relatively limited, primarily due to the challenges associated with insufficient conformational sampling. Nevertheless, with ongoing improvements in the physical interpretability of UCG models and the growing availability of MD trajectory databases, the cost of parameterizing UCG models is expected to decline. As a result, the development of user-friendly and transferable bottom-up forcefields is becoming achievable.

4. Empirical forcefields

Currently, empirical forcefields are more widely adopted than bottom-up approaches for constructing biomolecular UCG models. These empirical forcefields have been extensively applied to investigations of various biological functions, including mechanical properties [69], protein conformational changes [63,122], and self-assembly [123]. The distinction between empirical and bottom-up forcefields lies in their matching philosophy: empirical approaches emphasize feasibility and computational efficiency rather than strict adherence to fundamental physical principles. For example, empirical forcefields are often developed by reproducing experimentally accessible properties, such as structural features, vibrational modes, and solvation enthalpies. In terms of potential energy functions, empirical forcefields prioritize reproducing macroscopic dynamical behavior over accurately capturing microscopic interactions. For instance, accurately describing the fluctuations around the equilibrium position between two protein domains inherently requires accounting for numerous microscopic interactions. However, empirical forcefields typically approximate such behavior using simple single-well potentials derived from crystallographic structures. Fitting high-dimensional potential energy surfaces based on low-dimensional macroscopic properties may result in deviations from the AA model, thereby limiting the model's predictive capability. Although some empirical high-resolution CG models have exhibited strong transferability and predictive performance, these successes are largely attributed to extensive, cross-system, data-driven parameter optimization. However, for highly customized and large-scale UCG models of complex biological systems, the availability of relevant data is often limited. Consequently, generating predictive empirical potentials remains a significant and unresolved challenge.

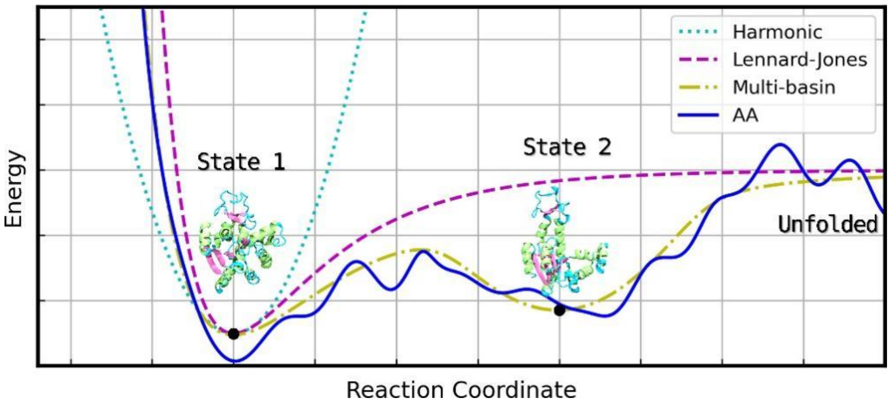

Figure 3. Schematic illustration of three types of empirical UCG potentials and their comparison with the AA model.

In many biological processes at the UCG scale, macromolecular domains or protein monomers often exhibit considerable structural stability. For example, during the self-assembly of the HIV capsid, the two major domains of the monomer, the N-terminal domain and the C-terminal domain, undergo minimal conformational changes. The conformational dynamics of such rigid proteins or domains can be effectively captured using simple restraining potentials. This approach enables a reliable description of conformational dynamics around a predefined reference structure at minimal modeling and simulation cost, a strategy commonly referred to as structure-based models. Elastic network models (ENMs) [124-126] represent one of the most widely used classes of restraining potentials. In ENMs, all pairwise interactions are modeled using computationally efficient harmonic oscillators, as illustrated by the light blue potential energy curve in **Figure 3**. They are particularly effective at constraining molecular structures and capturing vibrational modes [125] in the vicinity of equilibrium states. However, ENMs inherently suffer from the limitation that their potentials do not converge upon dissociation states, rendering them unsuitable for describing conformational states far from the reference structure. To overcome this limitation, ENMs are often integrated with more flexible bottom-up or empirical potentials, thereby enhancing the model's capacity to capture the dynamics of flexible functional regions such as linkers. ENMs have been extensively employed in UCG forcefields for analyzing vibrational modes between protein domains, constraining rigid domains, and reproducing elastic mechanical properties. Furthermore, ENMs play a pivotal role in UCG division methods, particularly in the rapid extraction of vibrational features from all-atom simulations or cryo-EM density data.

Other empirical UCG forcefields, particularly those designed to capture protein folding dynamics, require the potential energy to converge smoothly at large distances. A pragmatic approach involves the use of Gō-like potentials [127-129]—a class of restraining potentials that preserve the native bonding relations of protein chains. These potentials employ Lennard-Jones-type functions that smoothly converge in the unfolded state to describe inter-residue non-bonded attractions, which are critical for stabilizing folded conformations, as illustrated by the purple curve in **Figure 3**. Gō potentials have been widely applied to predict folding process of small to medium-sized proteins [130-133], including folding pathways and residue-level correlations, as well as for imposing constraints on backbone conformations. Gō-type potentials are implemented in the CafeMol MD simulation package developed by Takada and co-workers [134,135].

Many proteins undergo substantial conformational changes during their functional cycles, and structural biology studies have revealed that they often populate multiple stable intermediate states—for example, the opening and closing of the active site in adenylate kinase (AK). Accurately capturing such functionally relevant transitions at the single-protein level requires reproducing a global potential energy surface with multiple local minima, which is difficult to achieve using the aforementioned single-well restraining potentials. To address this issue, some UCG models employ empirical multi-well potentials to capture conformational transitions. Notably, unlike bottom-up approaches, these multi-well functions are not derived from the summation of microscopic interactions. Instead, they serve as phenomenological approximations that reproduce the experimentally observed multistate behavior of proteins, as illustrated in **Figure 3**. A common strategy for constructing such models involves assigning stable crystal structures as local minima of the multi-well potential [136], followed by the setting of their relative stabilities and transition parameters. Some approaches, based on the adiabatic approximation, analytically construct multi-well potentials by coupling single-well restraining potentials. Representative examples include the multi-state Gō model proposed by Takada and colleagues [137], and the double-well ENM developed by Voth and co-workers [138]. Another approach is to employ piecewise multi-state potentials, which offer improved computational efficiency and finer control over the microscopic energy landscape. Notable examples include the two-state model developed by Wang and collaborators [122,139,140], and the DW-UCG model proposed by Xia and collaborators [63]. These models provide an easy-to-construct and transferable framework for building UCG models capable of capturing conformational transitions. In some cases, they can even predict intermediate states that are closely associated with the reference structures.

In the context of intermolecular interactions, restraining potentials remain the most widely used approach. These potentials are typically formulated as simple single-well functions, such as harmonic or Lennard-Jones types, depending on whether dissociated states need to be represented, for example, in studies of mechanical properties or the transition dynamics between assembled and dissociated states. The construction of restraining potentials requires prior knowledge of equilibrium conformations, typically derived from crystal structures. Consequently, such potentials can only capture binding states observed in these reference conformations. This inability to account for the diversity of possible molecular associations significantly limits their predictive capabilities. To address this limitation, several non-restraining potentials with enhanced predictive capabilities beyond reference binding states have been developed by incorporating greater physical interpretability. For example, the potential used in SB-CG models can be efficiently estimated from properties such as the solvent-accessible surface area (SASA) and radius of gyration of the AA model, which encode key features like particle hydrophobicity and particle size. This allows for the rapid generation of intermolecular forcefields capable of stabilizing mesoscale and complex supramolecular structures, such as viral capsids, without relying on equilibrium conformations from crystal structures, and further enables the prediction of non-elastic deformations.

However, non-restraining potentials for modeling intermolecular interactions often rely on classical pairwise functional forms, which are inadequate for capturing solvation effects, many-body interactions, and the influence of non-spherical molecular UCG shapes. Overcoming these limitations can be achieved by the development of new potential functions that are both physically interpretable and capable of incorporating these complex effects, an effort that typically demands huge amounts of simulation and experimental data across different biomolecular systems. In the absence of such comprehensive data, a promising alternative is to adopt multiscale models, wherein key contact regions are described using high-resolution, bottom-up models. This approach provides an accurate, predictive, and transferable representation of critical intermolecular interactions. For example, in the open-to-closed conformational transition of ligand-free adenylate kinase (AK) shown in **Figure 4**, we found that the UA/UCG multiscale model retains more detailed features of the local basins on the free energy landscape compared to the purely UCG-based model proposed by Zhang et al [63]. These details stem from the physically grounded microscopic interactions between residues captured in the high-resolution regions, thereby enhancing both accuracy and predictability. This multiscale strategy has recently been applied to drug design and the simulation of membrane-associated systems.

Overall, empirical UCG potentials are predominantly implemented as restraining functions, which enable efficient characterization of conformational dynamics near equilibrium and facilitate the reproduction of multistate transitions observed between crystal structures. For modeling intermolecular interactions, some emerging non-restraining potentials incorporating hydrophobic/hydrophilic effects have shown potential in predicting non-elastic mechanical behaviors.

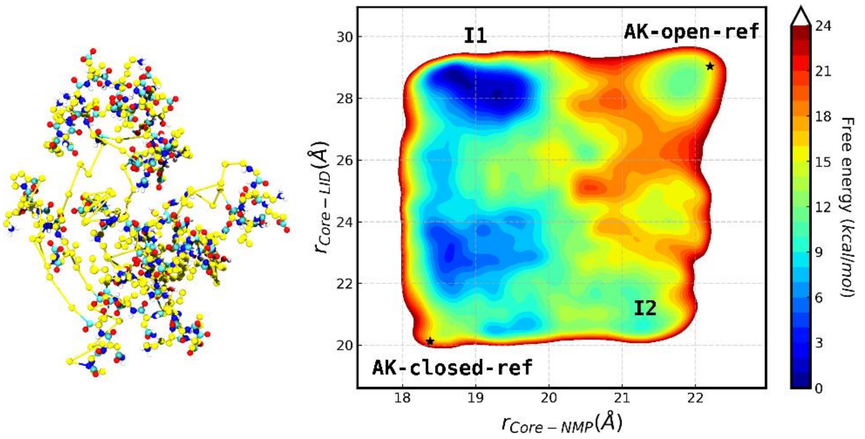

Figure 4. Structure of the UA/UCG AK model. UCG beads are colored in yellow, while UA regions are depicted in other colors. The UA potential is parameterized using the LJSPM method, and the UCG potential adopts the DW-UCG model. The reference closed and open conformations, as well as the two major intermediate states (I1 and I2), are labeled in the figure.

5. Conclusions

This review summarizes recent advances in biological UCG models from both application and methodological perspectives, with a particular focus on progress in mapping CG representations and forcefield development. For CG division, a range of generalized optimization algorithms have been established to customize mappings for a wide range of biomolecular systems based on high-resolution structural models or EM density maps, enabling partitions that faithfully capture the system's vibrational modes or density distributions. Nonetheless, due to the inherent heterogeneity in research objectives and system-specific characteristics, the a priori determination of an optimal resolution in the absence of parameterization remains a nontrivial challenge. From the perspective of forcefield development, empirical potentials remain the popular choice the modeling of large-scale biomolecular systems because of their high efficiency in parameterization and construction, as well as their capability in reproducing single-molecule dynamic behaviors like equilibrium fluctuations or complex multistate behavior. Moreover, recent advances in non-restraining empirical potentials, coupled with progress in multiscale modeling methodologies, have markedly expanded the predictive capability of UCG models in supramolecular assembling or docking contributed by intermolecular interactions. Meanwhile, bottom-up forcefield parameterization methods grounded in rigorous statistical mechanics have undergone substantial advancement over the past decades. Nevertheless, their application to large biomolecular systems remains constrained by the considerable computational cost of parameterization. Enhancing the physical interpretability of UCG potentials, such as through the explicit decomposition of electrostatic, Lennard-Jones, and bonded contributions, offers a promising avenue for reducing prior-data requirements for parametrization and improving transferability. Also, the continued accumulation of open-source biomolecular trajectory databases is expected to reduce the reliance on pre-sampling, thereby improving the feasibility and expanding the applicability of bottom-up methods.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grants No. 22403051), the open project foundation of Key Laboratory of NSLSCS, Ministry of Education.

References

[1] Klepeis J. L., Lindorff-Larsen K., Dror R. O. and Shaw D. E., Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.*, **19** (2009), 120–127.

[2] Wei G., Xi W., Nussinov R. and Ma B., Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell. *Chem. Rev.*, **116** (2016), 6516–6551.

[3] Petrovic D., Risso V. A., Kamerlin S. C. L. and Sanchez-Ruiz J. M., Conformational dynamics and enzyme evolution. *J. R. Soc. Interface*, **15** (2018), 20180330.

[4] Schwartz S. D., Protein Dynamics and Enzymatic Catalysis. *J. Phys. Chem. B*, **127** (2023), 2649–2660.

[5] Sokkar P., Boulanger E., Thiel W. and Sanchez-Garcia E., Hybrid Quantum Mechanics/Molecular Mechanics/Coarse Grained Modeling: A Triple-Resolution Approach for Biomolecular Systems. *J. Chem. Theory Comput.*, **11** (2015), 1809–1818.

[6] Senn H. M. and Thiel W., QM/MM studies of enzymes. *Curr. Opin. Chem. Biol.*, **11** (2007), 182–187.

[7] Paggi J. M., Pandit A. and Dror R. O., The Art and Science of Molecular Docking. *Annu. Rev. Biochem.*, **93** (2024), 389–410.

[8] Zhou Y., Jiang Y. and Chen S. J., RNA-ligand molecular docking: advances and challenges. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **12** (2022), e1571.

[9] Lu D., Luo D., Zhang Y. and Wang B., A Robust Induced Fit Docking Approach with the Combination of the Hybrid All-Atom/United-Atom/Coarse-Grained Model and Simulated Annealing. *J. Chem. Theory Comput.*, **20** (2024), 6414–6423.

[10] Souza P. C. T., Limongelli V., Wu S., Marrink S. J. and Monticelli L., Perspectives on High-Throughput Ligand/Protein Docking With Martini MD Simulations. *Front. Mol. Biosci.*, **8** (2021), 657222.

[11] Zha J., Zhang Y., Xia K., Grater F. and Xia F., Coarse-Grained Simulation of Mechanical Properties of Single Microtubules With Micrometer Length. *Front. Mol. Biosci.*, **7** (2020), 632122.

[12] Chu J. W. and Voth G. A., Coarse-grained modeling of the actin filament derived from atomistic-scale simulations. *Biophys. J.*, **90** (2006), 1572–1582.

[13] Tavakol M. and Vaughan T. J., A coarse-grained molecular dynamics investigation of the role of mineral arrangement on the mechanical properties of mineralized collagen fibrils. *J. R. Soc. Interface*, **20** (2023), 20220803.

[14] Vaziri A. and Gopinath A., Cell and biomolecular mechanics in silico. *Nat. Mater.*, **7** (2008), 15–23.

[15] Capone M., Romanelli M., Castaldo D., Parolin G., Bello A., Gil G. and Vanzan M., A Vision for the Future of Multiscale Modeling. *ACS Phys. Chem. Au*, **4** (2024), 202–225.

[16] Noe F., Tkatchenko A., Muller K. R. and Clementi C., Machine Learning for Molecular Simulation. *Annu. Rev. Phys. Chem.*, **71** (2020), 361–390.

[17] Zeng J., Weng J., Zhang Y., Xia F., Cui Q. and Xu X., Conformational Features of Ras: Key Hydrogen-Bonding Interactions of Gln61 in the Intermediate State during GTP Hydrolysis. *J. Phys. Chem. B*, **125** (2021), 8805–8813.

[18] Zha J. and Xia F., Developing Hybrid All-Atom and Ultra-Coarse-Grained Models to Investigate Taxol-Binding and Dynein Interactions on Microtubules. *J. Chem. Theory Comput.*, **19** (2023), 5621–5632.

[19] Invernizzi M., Valsson O. and Parrinello M., Coarse graining from variationally enhanced sampling applied to the Ginzburg-Landau model. *Proc. Natl. Acad. Sci. U. S. A.*, **114** (2017), 3370–3374.

[20] Kanada R., Terayama K., Tokuhisa A., Matsumoto S. and Okuno Y., Enhanced Conformational Sampling with an Adaptive Coarse-Grained Elastic Network Model Using Short-Time All-Atom Molecular Dynamics. *J. Chem. Theory Comput.*, **18** (2022), 2062–2074.

[21] Fenwick R. B., Esteban-Martin S. and Salvatella X., Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *Eur. Biophys. J.*, **40** (2011), 1339–1355.

[22] Childers M. C. and Daggett V., Validating Molecular Dynamics Simulations against Experimental Observables in Light of Underlying Conformational Ensembles. *J. Phys. Chem. B*, **122** (2018), 6673–6689.

[23] Noid W. G., Szukalo R. J., Kidder K. M. and Lesniewski M. C., Rigorous Progress in Coarse-Graining. *Annu. Rev. Phys. Chem.*, **75** (2024), 21–45.

[24] Lu X., Chen J. and Huang J., The continuous evolution of biomolecular force fields. *Structure*, **33** (2025), 1138–1149.

[25] Case D. A., Cheatham T. E., 3rd, Darden T., Gohlke H., Luo R., Merz K. M., Jr., Onufriev A., Simmerling C., Wang B. and Woods R. J., The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26** (2005), 1668–1688.

[26] Brooks B. R., Brooks C. L., 3rd, Mackerell A. D., Jr., Nilsson L., Petrella R. J., Roux B., Won Y., Archontis G., Bartels C., Boresch S., et al., CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, **30** (2009), 1545–1614.

[27] Van Der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A. E. and Berendsen H. J. G., GROMACS: fast, flexible, and free. *J. Comput. Chem.*, **26** (2005), 1701–1718.

[28] Hollingsworth S. A. and Dror R. O., Molecular Dynamics Simulation for All. *Neuron*, **99** (2018), 1129–1143.

[29] Ciccotti G., Dellago C., Ferrario M., Hernández E. R. and Tuckerman M. E., Molecular simulations: past, present, and future (a Topical Issue in EPJB). *Eur. Phys. J. B*, **95** (2022), 3.

[30] Cui Q., Pal T. and Xie L., Biomolecular QM/MM Simulations: What Are Some of the "Burning Issues"? *J. Phys. Chem. B*, **125** (2021), 689–702.

[31] Kubar T., Elstner M. and Cui Q., Hybrid Quantum Mechanical/Molecular Mechanical Methods For Studying

Energy Transduction in Biomolecular Machines. *Annu. Rev. Biophys.*, **52** (2023), 525–551.

[32] Karplus M., Molecular dynamics simulations of biomolecules. *Acc. Chem. Res.*, **35** (2002), 321–323.

[33] Trylska J., Coarse-grained models to study dynamics of nanoscale biomolecules and their applications to the ribosome. *J. Phys. Condens. Matter*, **22** (2010), 453101.

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

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

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

[37] van Teijlingen A., Smith M. C. and Tuttle T., Short Peptide Self-Assembly in the Martini Coarse-Grain Force Field Family. *Acc. Chem. Res.*, **56** (2023), 644–654.

[38] Zhu R., Zhang Y., Zhu T. and Xia F., Coarse-Grained Simulation of Persistence Length and Twisting Dynamics of Micrometer-Scale Microtubules. *J. Phys. Chem. Lett.*, **16** (2025), 7300–7306.

[39] Bartocci A., Grazzi A., Awad N., Corringer P. J., Souza P. C. T. and Cecchini M., A millisecond coarse-grained simulation approach to decipher allosteric cannabinoid binding at the glycine receptor alpha1. *Nat. Commun.*, **15** (2024), 9040.

[40] Souza P. C. T., Thallmair S., Conflitti P., Ramirez-Palacios C., Alessandri R., Raniolo S., Limongelli V. and Marrink S. J., Protein-ligand binding with the coarse-grained Martini model. *Nat. Commun.*, **11** (2020), 3714.

[41] Jung J., Tan C. and Sugita Y., GENESIS CGDYN: large-scale coarse-grained MD simulation with dynamic load balancing for heterogeneous biomolecular systems. *Nat. Commun.*, **15** (2024), 3370.

[42] Giulini M., Rigoli M., Mattiotti G., Menichetti R., Tarenzi T., Fiorentini R. and 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.*, **8** (2021), 676976.

[43] Weiner S. J., Kollman P. A., Case D. A., Singh U. C., Ghio C., Alagona G., Profeta S., Jr. and Weiner P., A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.*, **106** (1984), 765–784.

[44] Shen H., Li Y., Ren P., Zhang D. and Li G., An Anisotropic Coarse-Grained Model for Proteins Based On Gay-Berne and Electric Multipole Potentials. *J. Chem. Theory Comput.*, **10** (2014), 731–750.

[45] Shen H., Li Y., Xu P., Li X., Chu H., Zhang D. and Li G., An anisotropic coarse-grained model based on Gay-Berne and electric multipole potentials and its application to simulate a DMPC bilayer in an implicit solvent model. *J. Comput. Chem.*, **36** (2015), 1103–1113.

[46] Han W., Wan C. K., Jiang F. and Wu Y. D., PACE Force Field for Protein Simulations. 1. Full Parameterization of Version 1 and Verification. *J. Chem. Theory Comput.*, **6** (2010), 3373–3389.

[47] Han W., Wan C. K., Jiang F. and Wu Y. D., PACE Force Field for Protein Simulations. 2. Folding Simulations of Peptides. *J. Chem. Theory Comput.*, **6** (2010), 3390–3402.

[48] Kar P., Gopal S. M., Cheng Y. M., Predeus A. and Feig M., PRIMO: A Transferable Coarse-grained Force Field for Proteins. *J. Chem. Theory Comput.*, **9** (2013), 3769–3788.

[49] Kar P., Gopal S. M., Cheng Y. M., Panahi A. and Feig M., Transferring the PRIMO Coarse-Grained Force Field to the Membrane Environment: Simulations of Membrane Proteins and Helix-Helix Association. *J. Chem. Theory Comput.*, **10** (2014), 3459–3472.

[50] Liwo A., Ołdziej S., Pincus M. R., Wawak R. J., Rackovsky S. and 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), 849–873.

[51] Marrink S. J., de Vries A. H. and Mark A. E., Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B*, **108** (2004), 750–760.

[52] Monticelli L., Kandasamy S. K., Periole X., Larson R. G., Tieleman D. P. and Marrink S. J., The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.*, **4** (2008), 819–834.

[53] Darre L., Machado M. R., Brandner A. F., Gonzalez H. C., Ferreira S. and Pantano S., SIRAH: a structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *J. Chem. Theory Comput.*, **11** (2015), 723–739.

[54] Ouldridge T. E., Louis A. A. and Doye J. P., Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.*, **134** (2011), 085101.

[55] Snodin B. E., Randisi F., Mosayebi M., Sulc P., Schreck J. S., Romano F., Ouldridge T. E., Tsukanov R., Nir E., Louis A. A. and Doye J. P., Introducing improved structural properties and salt dependence into a coarse-grained model of DNA. *J. Chem. Phys.*, **142** (2015), 234901.

[56] Sulc P., Romano F., Ouldridge T. E., Doye J. P. and Louis A. A., A nucleotide-level coarse-grained model of RNA. *J. Chem. Phys.*, **140** (2014), 235102.

[57] Marrink S. J., Monticelli L., Melo M. N., Alessandri R., Tieleman D. P. and Souza P. C. T., Two decades of Martini: Better beads, broader scope. *WIREs Comput. Mol. Sci.*, **13** (2022), e1620.

[58] Snodin B. E. K., Schreck J. S., Romano F., Louis A. A. and Doye J. P. K., Coarse-grained modelling of the structural properties of DNA origami. *Nucleic Acids Res.*, **47** (2019), 1585–1597.

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

[60] Dama J. F., Sinitskiy A. V., McCullagh M., Weare J., Roux B., Dinner A. R. and Voth G. A., The Theory of Ultra-Coarse-Graining. 1. General Principles. *J. Chem. Theory Comput.*, **9** (2013), 2466–2480.

[61] Davtyan A., Dama J. F., Sinitskiy A. V. and Voth G. A., The Theory of Ultra-Coarse-Graining. 2. Numerical Implementation. *J. Chem. Theory Comput.*, **10** (2014), 5265–5275.

[62] Zhang Y., Cao Z. and Xia F., Construction of ultra-coarse-grained model of protein with a Gō-like potential. *Chem. Phys. Lett.*, **681** (2017), 1–6.

[63] Zhang Y., Cao Z., Zhang J. Z. and Xia F., Double-Well Ultra-Coarse-Grained Model to Describe Protein Conformational Transitions. *J. Chem. Theory Comput.*, **16** (2020), 6678–6689.

[64] Grime J. M. A., Dama J. F., Ganser-Pornillos B. K., Woodward C. L., Jensen G. J., Yeager M. and Voth G. A., Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly. *Nat. Commun.*, **7** (2016), 11568.

[65] Yu A., Pak A. J., He P., Monje-Galvan V., Casalino L., Gaieb Z., Dommer A. C., Amaro R. E. and Voth G. A., A multiscale coarse-grained model of the SARS-CoV-2 virion. *Biophys. J.*, **120** (2021), 1097–1104.

[66] Carmichael S. P. and Shell M. S., A new multiscale algorithm and its application to coarse-grained peptide models for self-assembly. *J. Phys. Chem. B*, **116** (2012), 8383–8393.

[67] Pak A. J., Grime J. M. A., Yu A. and 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.*, **141** (2019), 10214–10224.

[68] Li F., Zhang Y., Xia F. and Xu X., Development of multiscale ultra-coarse-grained models for the SARS-CoV-2 virion from cryo-electron microscopy data. *Phys. Chem. Chem. Phys.*, **25** (2023), 12882–12890.

[69] Bryer A. J., Rey J. S. and Perilla J. R., Performance efficient macromolecular mechanics via sub-nanometer shape based coarse graining. *Nat. Commun.*, **14** (2023), 2014.

[70] Zhang Y., Cao Z., Zhang J. Z. and Xia F., Performance Comparison of Systematic Methods for Rigorous Definition of Coarse-Grained Sites of Large Biomolecules. *J. Chem. Inf. Model.*, **57** (2017), 214–222.

[71] Ming D., Kong Y., Lambert M. A., Huang Z. and Ma J., How to describe protein motion without amino acid sequence and atomic coordinates. *Proc. Natl. Acad. Sci. U. S. A.*, **99** (2002), 8620–8625.

[72] Martinetz T. and Schulten K., Topology Representing Networks. *Neural Networks*, **7** (1994), 507–522.

[73] Wriggers W., Milligan R. A., Schulten K. and McCammon J. A., Self-organizing Neural Networks Bridge the Biomolecular Resolution Gap. *J. Mol. Biol.*, **284** (1998), 1247–1254.

[74] Arkhipov A., Freddolino L. and Schulten K., Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure*, **14** (2006), 1767–1777.

[75] Arkhipov A., Yin Y. and Schulten K., Four-scale description of membrane sculpting by BAR domains. *Biophys. J.*, **95** (2008), 2806–2821.

[76] Arkhipov A., Freddolino L., Imada K., Namba K. and Schulten K., Coarse-grained molecular dynamics simulations of a rotating bacterial flagellum. *Biophys. J.*, **91** (2006), 4589–4597.

[77] Arkhipov A., Roos W. H., Wuite G. J. and Schulten K., Elucidating the mechanism behind irreversible deformation of viral capsids. *Biophys. J.*, **97** (2009), 2061–2069.

[78] Zhang Y., Xia K., Cao Z., Grater F. and Xia F., A new method for the construction of coarse-grained models of large biomolecules from low-resolution cryo-electron microscopy data. *Phys. Chem. Chem. Phys.*, **21** (2019), 9720–9727.

[79] Zhang Z. and Voth G. A., Coarse-Grained Representations of Large Biomolecular Complexes from Low-Resolution Structural Data. *J. Chem. Theory Comput.*, **6** (2010), 2990–3002.

[80] Tirion M. M., Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.*, **77** (1996), 1905–1908.

[81] Bahar I., Atilgan A. R. and Erman B., Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding & Design*, **2** (1997), 173–181.

[82] Haliloglu T., Bahar I. and Erman B., Gaussian dynamics of folded proteins. *Phys. Rev. Lett.*, **79** (1997), 3090–3093.

[83] Zhang Z., Lu L., Noid W. G., Krishna V., Pfaendtner J. and Voth G. A., A systematic methodology for defining coarse-grained sites in large biomolecules. *Biophys. J.*, **95** (2008), 5073–5083.

[84] Zhang Z., Pfaendtner J., Grafmuller A. and Voth G. A., Defining coarse-grained representations of large biomolecules and biomolecular complexes from elastic network models. *Biophys. J.*, **97** (2009), 2327–2337.

[85] Wu J., Xue W. and Voth G. A., K-Means Clustering Coarse-Graining (KMC-CG): A Next Generation Methodology for Determining Optimal Coarse-Grained Mappings of Large Biomolecules. *J. Chem. Theory Comput.*, **19** (2023), 8987–8997.

[86] Fan J., Saunders M. G. and Voth G. A., Coarse-graining provides insights on the essential nature of heterogeneity in actin filaments. *Biophys. J.*, **103** (2012), 1334–1342.

[87] Zhang Z., Sanbonmatsu K. Y. and Voth G. A., Key intermolecular interactions in the E. coli 70S ribosome revealed by coarse-grained analysis. *J. Am. Chem. Soc.*, **133** (2011), 16828–16838.

[88] Pak A. J., Yu A., Ke Z., Briggs J. A. G. and Voth G. A., Cooperative multivalent receptor binding promotes exposure of the SARS-CoV-2 fusion machinery core. *Nat. Commun.*, **13** (2022), 1002.

[89] Pak A. J., Gupta M., Yeager M. and Voth G. A., Inositol Hexakisphosphate (IP6) Accelerates Immature HIV-1 Gag Protein Assembly toward Kinetically Trapped Morphologies. *J. Am. Chem. Soc.*, **144** (2022), 10417–10428.

[90] Banerjee P. and Voth G. A., Conformational transitions of the HIV-1 Gag polyprotein upon multimerization and gRNA binding. *Biophys. J.*, **123** (2024), 42–56.

[91] Li M., Zhang J. Z. and Xia F., Constructing Optimal Coarse-Grained Sites of Huge Biomolecules by Fluctuation Maximization. *J. Chem. Theory Comput.*, **12** (2016), 2091–2100.

[92] Zhou J., Liu S., Zhang Y., Ma Y. and Liu J., Modulation of Phase Separation Morphology by Configuration Engineering in Bulk Heterojunction Organic Solar Cells. *Sol. RRL*, **8** (2024), 2400074.

[93] Li M., Zhang J. Z. and Xia F., A new algorithm for construction of coarse-grained sites of large biomolecules. *J. Comput. Chem.*, **37** (2016), 795–804.

[94] Wu Z., Zhang Y., Zhang J. Z., Xia K. and Xia F., Determining Optimal Coarse-Grained Representation for Biomolecules Using Internal Cluster Validation Indexes. *J. Comput. Chem.*, **41** (2020), 14–20.

[95] Calinski T. and Harabasz J., A dendrite method for cluster analysis. *Commun. Stat.- Theory Methods*, **3** (1974), 1–27.

[96] Sinitskiy A. V., Saunders M. G. and Voth G. A., Optimal number of coarse-grained sites in different components of

large biomolecular complexes. *J. Phys. Chem. B*, **116** (2012), 8363–8374.

[97] Durumeric A. E. P., Charron N. E., Templeton C., Musil F., Bonneau K., Pasos-Trejo A. S., Chen Y., Kelkar A., Noe F. and Clementi C., Machine learned coarse-grained protein force-fields: Are we there yet? *Curr. Opin. Struct. Biol.*, **79** (2023), 102533.

[98] Noid W. G., Perspective: Advances, Challenges, and Insight for Predictive Coarse-Grained Models. *J. Phys. Chem. B*, **127** (2023), 4174–4207.

[99] Izvekov S. and Voth G. A., A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B*, **109** (2005), 2469–2473.

[100] Izvekov S. and Voth G. A., Multiscale coarse graining of liquid-state systems. *J. Chem. Phys.*, **123** (2005), 134105.

[101] Noid W. G., Chu J. W., Ayton G. S., Krishna V., Izvekov S., Voth G. A., Das A. and Andersen H. C., The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.*, **128** (2008), 244114.

[102] Reith D., Putz M. and Muller-Plathe F., Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.*, **24** (2003), 1624–1636.

[103] Mullinax J. W. and Noid W. G., Generalized Yvon-Born-Green theory for molecular systems. *Phys. Rev. Lett.*, **103** (2009), 198104.

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

[105] Peng Y., Pak A. J., Durumeric A. E. P., Sahrmann P. G., Mani S., Jin J., Loose T. D., Beiter J. and Voth G. A., OpenMSCG: A Software Tool for Bottom-Up Coarse-Graining. *J. Phys. Chem. B*, **127** (2023), 8537–8550.

[106] Thompson A. P., Aktulga H. M., Berger R., Bolintineanu D. S., Brown W. M., Crozier P. S., in 't Veld P. J., Kohlmeyer A., Moore S. G., Nguyen T. D., et al., LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.*, **271** (2022), 108171.

[107] Ramakrishnan R., Dral P. O., Rupp M. and von Lilienfeld O. A., Big Data Meets Quantum Chemistry Approximations: The Delta-Machine Learning Approach. *J. Chem. Theory Comput.*, **11** (2015), 2087–2096.

[108] Kulichenko M., Nebgen B., Lubbers N., Smith J. S., Barros K., Allen A. E. A., Habib A., Shinkle E., Fedik N., Li Y. W., et al., Data Generation for Machine Learning Interatomic Potentials and Beyond. *Chem. Rev.*, **124** (2024), 13681–13714.

[109] Kocer E., Ko T. W. and Behler J., Neural Network Potentials: A Concise Overview of Methods. *Annu. Rev. Phys. Chem.*, **73** (2022), 163–186.

[110] Husic B. E., Charron N. E., Lemm D., Wang J., Perez A., Majewski M., Kramer A., Chen Y., Olsson S., de Fabritiis G., et al., Coarse graining molecular dynamics with graph neural networks. *J. Chem. Phys.*, **153** (2020), 194101.

[111] Larini L., Lu L. and Voth G. A., The multiscale coarse-graining method. VI. Implementation of three-body coarse-grained potentials. *J. Chem. Phys.*, **132** (2010), 164107.

[112] Wang J., Olsson S., Wehmeyer C., Perez A., Charron N. E., de Fabritiis G., Noe F. and Clementi C., Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Cent. Sci.*, **5** (2019), 755–767.

[113] Wang J., Chmiela S., Muller K. R., Noe F. and Clementi C., Ensemble learning of coarse-grained molecular dynamics force fields with a kernel approach. *J. Chem. Phys.*, **152** (2020), 194106.

[114] Charron N. E., Bonneau K., Pasos-Trejo A. S., Guljas A., Chen Y., Musil F., Venturin J., Gusew D., Zaporozhets I., Kramer A., et al., Navigating protein landscapes with a machine-learned transferable coarse-grained model. *Nat. Chem.*, **17** (2025), 1284–1292.

[115] McCullagh P., Lake P. T. and McCullagh M., Deriving Coarse-Grained Charges from All-Atom Systems: An Analytic Solution. *J. Chem. Theory Comput.*, **12** (2016), 4390–4399.

[116] Zhang Y., Wang Y., Xia F., Cao Z. and Xu X., Accurate and Efficient Estimation of Lennard-Jones Interactions for Coarse-Grained Particles via a Potential Matching Method. *J. Chem. Theory Comput.*, **18** (2022), 4879–4890.

[117] Zhu Q., Zhang Y. and Xia F., Effective Volume Correction for Lennard-Jones Static Potential Matching on Coarse-Graining Small Molecules. *ChemRxiv*, (2024).

[118] Yin Y. D., Zhang Y. W., Song X. T., Hu J., Chen Y. H., Lai W. C., Li Y. F. and Gu Z. Y., Native globular ferritin nanopore sensor. *Nat. Commun.*, **16** (2025), 5268.

[119] Tillack A. F., Johnson L. E., Eichinger B. E. and Robinson B. H., Systematic Generation of Anisotropic Coarse-Grained Lennard-Jones Potentials and Their Application to Ordered Soft Matter. *J. Chem. Theory Comput.*, **12** (2016), 4362–4374.

[120] Jin J., Pak A. J. and Voth G. A., Understanding Missing Entropy in Coarse-Grained Systems: Addressing Issues of Representability and Transferability. *J. Phys. Chem. Lett.*, **10** (2019), 4549–4557.

[121] Dama J. F., Jin J. and Voth G. A., The Theory of Ultra-Coarse-Graining. 3. Coarse-Grained Sites with Rapid Local Equilibrium of Internal States. *J. Chem. Theory Comput.*, **13** (2017), 1010–1022.

[122] Lu Q. and Wang J., Single molecule conformational dynamics of adenylate kinase: energy landscape, structural correlations, and transition state ensembles. *J. Am. Chem. Soc.*, **130** (2008), 4772–4783.

[123] Sahrmann P. G. and Voth G. A., Entropy-based methods for formulating bottom-up ultra-coarse-grained models. *J. Chem. Phys.*, **162** (2025), 044102.

[124] Atilgan A. R., Durell S. R., Jernigan R. L., Demirel M. C., Keskin O. and Bahar I., Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.*, **80** (2001), 505–515.

[125] Xia F. and Lu L., Multiscale Coarse-Graining via Normal Mode Analysis. *J. Chem. Theory Comput.*, **8** (2012), 4797–4806.

[126] Xia F., Tong D. and Lu L., Robust Heterogeneous Anisotropic Elastic Network Model Precisely Reproduces the Experimental B-factors of Biomolecules. *J. Chem. Theory Comput.*, **9** (2013), 3704–3714.

[127] Taketomi H., Ueda Y. and Go N., Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.*, **7** (1975), 445–459.

[128] Ueda Y., Taketomi H. and Gō N., Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme. *Biopolymers*, **17** (1978), 1531–1548.

[129] Takada S., Go model revisited. *Biophys Physicobiol.*, **16** (2019), 248–255.

[130] Leopold P. E., Montal M. and Onuchic J. N., Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. U. S. A.*, **89** (1992), 8721–8725.

[131] Onuchic J. N., Luthey-Schulten Z. and Wolynes P. G., Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.*, **48** (1997), 545–600.

[132] Clementi C., Jennings P. A. and Onuchic J. N., How native-state topology affects the folding of dihydrofolate reductase and interleukin-1beta. *Proc. Natl. Acad. Sci. U. S. A.*, **97** (2000), 5871–5876.

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

[134] Murata Y., Niina T. and Takada S., OpenCafeMol: A coarse-grained biomolecular simulator on GPU with its application to vesicle fusion. *Biophys. J.*, (2025), 2025. 02. 20. 639390.

[135] Kenzaki H., Koga N., Hori N., Kanada R., Li W., Okazaki K., Yao X. Q. and Takada S., CafeMol: A Coarse-Grained Biomolecular Simulator for Simulating Proteins at Work. *J. Chem. Theory Comput.*, **7** (2011), 1979–1989.

[136] Okazaki K., Koga N., Takada S., Onuchic J. N. and Wolynes P. G., Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.*, **103** (2006), 11844–11849.

[137] Okazaki K. and Takada S., Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. U. S. A.*, **105** (2008), 11182–11187.

[138] Chu J. W. and Voth G. A., Coarse-grained free energy functions for studying protein conformational changes: a double-well network model. *Biophys. J.*, **93** (2007), 3860–3871.

[139] Lu Q. and Wang J., Kinetics and statistical distributions of single-molecule conformational dynamics. *J. Phys. Chem. B*, **113** (2009), 1517–1521.

[140] Lai Z. Z., Lu Q. and Wang J., Exploring the thermodynamic landscape, kinetics, and structural evolution of a protein conformational transition with a microscopic double-well model. *J. Phys. Chem. B*, **115** (2011), 4147–4159.