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

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 authors: ywzhang@nnu.edu.cn; fxia@chem.ecnu.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.

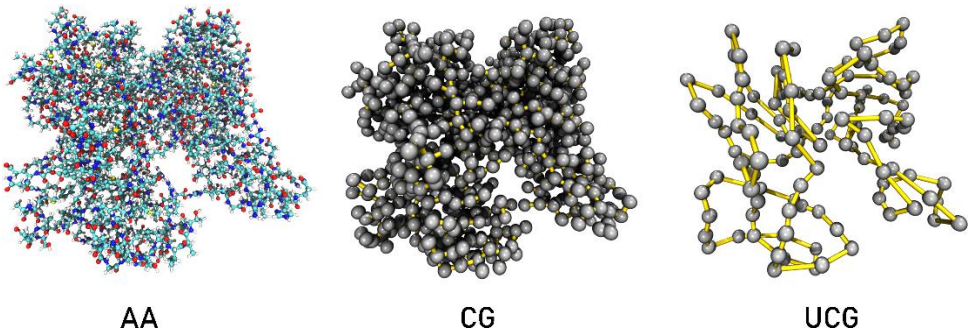

Figure 1. Structural representations of the G-actin monomer (PDB ID: 1ATN) at AA, high-resolution CG, and UCG resolutions.

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,