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