Computational Chemistry | Research Article

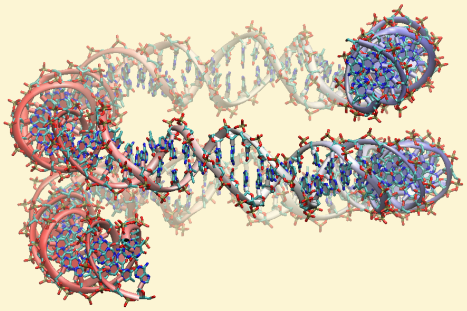
A Generalized Algorithm for Coarse-Graining Molecular Dynamics Simulations

Author
Forrest Bicker[†]

Supervisors
Dr. Sharon Loverde [§]
Phu Tang [§,‡]

[†]Hunter College High School, Manhattan, New York 10128, United States
[§]Department of Chemistry, College of Staten Island, Staten Island, New York 10314, United States
[‡]Ph.D. Program in Biochemistry, The Graduate Center of the City University of New York, New York 10314, United States

ABSTRACT: Atomistic molecular dynamics simulations are used to model the thermodynamic properties of macromolecular structures—but the high computational expense associated with these models limits the size and time-scale of systems that can be practically simulated. Coarse-graining techniques mitigate this issue by modeling atomic structures as low-resolution simplifications, but existing algorithms are inflexible to the customization of coarse-grain mappings, which inhibits the ability to switch between different coarse-grain models or develop new ones. In this paper, we generalize the process of coarse-graining to eliminate algorithmic dependence on specific mappings and offer an adaptable approach to model reduction. We demonstrate our algorithm performs this task in the theoretical minimum number of computations possible to ensure maximal efficiency. Additionally, we use machine learning to develop a method for accurately parametrizing our coarse-grained output. Finally, we implement and publish the algorithm under a MIT open-access software license alongside documentation contained in this article to facilitate widespread access and ease of use.



1. INTRODUCTION

The advancement of computational technology over recent decades has made possible the purely digital simulation of physical interactions between atomic structures. So-called molecular dynamics (MD) simulations make it possible to model the thermodynamic properties and temporal trajectories of small molecular structures over short time intervals.[1] MD is a valuable tool with diverse applications for examining interatomic interactions in a resolution unachievable through conventional methods. For example, simulation can be used to predict protein subunit quaternary conformations or observe translocation of potential drug compounds across cell membranes. However, these simulations carry with them a high computational expense: a simulation of meaningful size will typically require days of parallel CPU time to calculate nanoseconds of interactions.[2] This issue is compounded by the fact that van der Waals interactions require calculations with complexities that scale quadratically.[3] For this reason, the prospect of using all-atomistic MD simulations to analyze large macromolecular structures on biological time scales is still infeasible with modern computing resources.[4]

To circumvent this limitation, coarse-grain (CG) models have been developed which simplify complex atomic structures by clustering groups of atoms into collectivized beads.[5] This process shifts away from calculating minute, local motions of individual atoms and instead focuses resources on capturing the macroscopic dynamics of the molecule as a whole.[6] By defining CG mappings in such a that they eliminate localized low-impact calculations while remaining closely accurate to the true chemophysical nature of their target molecules, we are able to produce CG simulations closely resembling all-atomistic ones while greatly reducing computational expense of simulations.[7]

Numerous CG mappings for various molecules such as amino acids and DNA nucleotides have been developed, with each mapping offering different tradeoffs between computational efficiency and behavioral proximity to the atomistic model under varied conditions or differing molecular conformations.[8,9] However, available CG programs lack the functionality to accommodate for the existing variety of mappings.[10] This inflexibility it makes the process of interchanging mappings for different use-cases difficult and complex. Additionally, the inability to easily configure new mappings inhibits the ability to design and test new ones. Thus, the need for a generalized algorithm capable of coarse-graining atomistic models using any CG mapping arises.

2. THE ALGORITHM