Biochimica et Biophysica Acta xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

Review

Enhanced sampling techniques in molecular dynamics simulations of biological systems☆

Rafael C. Bernardi [a], Marcelo C.R. Melo [b], Klaus Schulten [a,b,c,*]

[a] Beckman Institute, University of Illinois, Urbana, IL 61801, USA
[b] Center for Biophysics and Computational Biology, University of Illinois, Urbana, IL 61801, USA
[c] Department of Physics, University of Illinois, Urbana, IL 61801, USA

ARTICLE INFO

Article history:
Received 11 August 2014
Received in revised form 17 October 2014
Accepted 20 October 2014
Available online xxxx

Keywords:
Enhanced sampling
Molecular dynamics
Replica-exchange molecular dynamics
Metadynamics
Generalized simulated annealing
Cellulosome

ABSTRACT

Background: Molecular dynamics has emerged as an important research methodology covering systems to the level of millions of atoms. However, insufficient sampling often limits its application. The limitation is due to rough energy landscapes, with many local minima separated by high-energy barriers, which govern the biomolecular motion.
Scope of review: In the past few decades methods have been developed that address the sampling problem, such as replica-exchange molecular dynamics, metadynamics and simulated annealing. Here we present an overview over theses sampling methods in an attempt to shed light on which should be selected depending on the type of system property studied.
Major conclusions: Enhanced sampling methods have been employed for a broad range of biological systems and the choice of a suitable method is connected to biological and physical characteristics of the system, in particular system size. While metadynamics and replica-exchange molecular dynamics are the most adopted sampling methods to study biomolecular dynamics, simulated annealing is well suited to characterize very flexible systems. The use of annealing methods for a long time was restricted to simulation of small proteins; however, a variant of the method, generalized simulated annealing, can be employed at a relatively low computational cost to large macromolecular complexes.
General significance: Molecular dynamics trajectories frequently do not reach all relevant conformational substates, for example those connected with biological function, a problem that can be addressed by employing enhanced sampling algorithms. This article is part of a Special Issue entitled Recent developments of molecular dynamics.

1. Introduction

Computer simulations of biomolecular systems have grown rapidly over the past few decades, passing from simulating very small proteins in vacuum to simulating large protein complexes in a solvated environment [1,2]. All-atom molecular dynamics (MD) simulations, employing classical mechanics, allowed the study of a broad range of biological systems, from small molecules such as anesthetics [3] or small peptides [4,5], to very large protein complexes such as the ribosome [6] or virus capsids [7,8]. Hybrid classical/quantum MD simulations allowed the study of enzymatic activity [9] or polarizable molecules in biological membranes [10]. However, despite its success, MD simulations are still limited in two regards, inaccuracy of force fields and high computational cost. For example, one-microsecond simulation of a relatively small system (approximately 25,000 atoms) running on 24 processors requires months of computation to complete [11], and expensive petascale supercomputers must be employed to study larger systems [6,8,12–14]. Such limitations can lead to inadequate sampling of conformational states, which in turn limits the ability to analyze and reveal functional properties of the systems being examined. All relevant states of a system must be reached in simulations in order for its dynamics and function to be meaningfully characterized. In this review we will discuss some solutions to the sampling problem.

Molecular dynamics simulations have always been viewed as a general sampling method for the study of conformational changes of biomolecules [15]. However, biological molecules are known to have rough energy landscapes, with many local minima frequently separated by high-energy barriers [16], making it easy to fall into a non-functional state that is hard to jump out of in most conventional simulations. Recent studies have demonstrated indeed that, in long simulations, proteins can get trapped in non-relevant conformations without going back to the original relevant conformation [14,17]. In fact, many free

☆ This article is part of a Special Issue entitled Recent developments of molecular dynamics.

* Corresponding author at: Beckman Institute for Advanced Science and Technology room 3147-405 N. Matthews Ave, Urbana, IL 61801, USA. Tel./fax: +1 217 244 2212.
E-mail address: kschulte@ks.uiuc.edu (K. Schulten).

http://dx.doi.org/10.1016/j.bbagen.2014.10.019
0304-4165/

energy minima can trap molecules for a long time and in some sense slow down the sampling process, leading to a poor characterization of a protein's dynamic behavior [15,18].

Large conformational changes are often important for protein activity in biological systems, for example in case of catalysis large amplitude movements are frequently required [19,20] or in case of transport through membranes, channels and transporters have to undergo large conformational changes in the course of gating substrates [21]. Such complicated and time consuming processes are commonly beyond the ability of straightforward MD simulations and enhanced sampling algorithms are needed. In the past few decades several methods have been developed for this purpose, such as replica-exchange molecular dynamics (REMD), metadynamics and simulated annealing. Here we present a brief overview of these methods and also demonstrate a new application for the generalized simulated annealing (GSA) technique.

2. Replica-exchange

The limitations to handling "hardly-relaxing" systems in MD [22,23] has led Sugita and Okamoto [24] to develop an implementation of a new generalized-ensemble algorithm, the replica exchange molecular dynamics (REMD) method. This method employs independent parallel Monte Carlo random walks in several parallel simulations carried out at different temperatures. System states, defined by position of atoms are exchanged depending on temperature and energy differences between selected simulations (see Fig. 1). Due to the general occurrence of "hardly-relaxing" systems in biological systems, the replica exchange method was quickly adopted and implemented in MD codes. By using Monte Carlo weights to determine the probability of exchanging systems states, REMD assures that the probability of exchange is quickly determined from the system's characteristics, presenting an advantage over other generalized ensemble methods. This approach furnishes efficient free random walks on the "replica space", namely temperature and potential energy spaces.

From the initial application of REMD to the penta-peptide met-enkephalin [24], to more recent efforts where improved forms of the method were developed, such as constant pH replica exchange [25] or isobaric–isothermal REMD to study Alzheimer's peptides [26], REMD has gained ground and proven to be effective under a broad range of contexts. Applications of REMD have been shown to agree with MD results and as long as there is a positive activation energy for folding, REMD was shown to be more efficient than MD [27]. The method allowed the study of free energy landscape and folding mechanism of several peptides and proteins [28–31]. The effectiveness of REMD was demonstrated to be strongly dependent on the activation enthalpy and its efficiency in describing actual proteins was seen to depend sensitively on the choice of maximum temperature. Choosing the maximum temperature too high can result in REMD becoming significantly less efficient than conventional MD [27,32]. Nymeyer suggested that a good strategy is to choose the maximum temperature slightly above the temperature at which the enthalpy for folding vanishes [27].

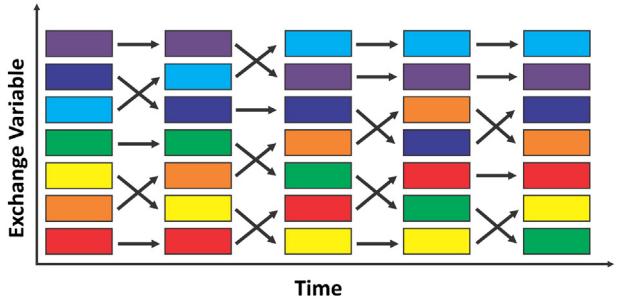

Fig. 1. Illustration of the replica exchange molecular dynamics (REMD) method. A set of non-interacting replicas runs at different values of an exchange variable, usually temperature (T-REMD). At specific intervals, replicas at neighboring values for the exchange variable are swapped based on a Monte Carlo acceptance criterion. In an efficient run, all trajectories will experience changing of the exchange variable value. At each value for the exchange variable, the trajectories will be discontinuous, but follow a proper Boltzmann distribution for the specific value being exchanged.

In the past decade REMD became a widely used method to enhance conformational sampling of MD simulations [4], and several variants of the traditional temperature dependent REMD (T-REMD) were implemented. While T-REMD has proven to be extremely useful in enhancing sampling in MD, it does not guarantee convergence and a better convergence was obtained with reservoir REMD (R-REMD) [33]. A more general form of REMD involves exchanges between different Hamiltonians (H-REMD) providing an enhanced sampling in dimensions other than temperature [33,34]. Methods employing multiple replicas, such as the multiplexed-REMD (M-REMD) for each temperature level were also developed, showing a more appropriate sampling in shorter simulation time than H-REMD and T-REMD [35]. Due to the large number of replicas, M-REMD scales to a large number of computer processors and the convergence was shown to be obtained in a short simulation time of around 50 ns. However, due to the large number of replicas, the total computational cost of M-REMD is prohibitive for most studies. Allowing the exchange of the thermodynamic coupling parameter λ also improved the search for alternative conformations of modeled biomolecules [36]. The exchanges along the thermodynamic coupling λ were shown to help distribute the side chain rotamers of a protein in different states [37]. A combination of a dual free energy perturbation (FEP) and λ-REMD was also proposed to be useful for calculation of the absolute binding free energy of p-xylene to a mutant of lysozyme [38]. Different implementations of REMD are common in some of the most popular MD packages, such as Amber [39], Gromacs [40] and NAMD [41], and, as described here, different types of REMD can be very effectively employed for sampling conformational changes. Although mostly used to sample different conformational states, REMD is not limited to such sampling and has been successfully employed for other problems such as the study of protein protonation states in case of constant pH REMD [25].

3. Metadynamics

To improve sampling of a system where ergodicity is hindered by the form of the system's energy landscape, Parrinello's group suggested a new algorithm called metadynamics that inserts memory in the sampling [42]. Local elevation [43] and conformational flooding [44] were the first methods to insert memory in an enhanced sampling for biomolecules. By discouraging that previously visited states be resampled, these and newer methods, like metadynamics, allow one to direct computational resources to a broader exploration of the free-energy landscape. Darve and Phorrille [45] described the method as "filling the free energy wells with computational sand" (see Fig. 2), suggesting that metadynamics can search through the entire free energy landscape, a well desired characteristic to study biological problems such as protein folding [46], molecular docking [47], phase transitions [48,49] and conformational changes [50].

Another advantage of the metadynamics method is that it does not depend on a very accurate description of the potential energy surface being explored. Since previous states are only *discouraged* to be resampled, misevaluated conformations can be re-calculated and error will tend to "even out"[42]. Metadynamics does depend on a low dimensionality of the system in order to produce an accurate description of the free energy surface, therefore using a small set of collective coordinates is essential. Such characteristics allow this method to be quickly used to provide qualitative information about the overall topology of the free energy surface being examined.

Parrinello's method was not the only one proposed to enhance the crossing of energy barriers; indeed, similar methods were created such as coarse molecular dynamics [30] or the method of Engkvist and Karlström [51]. Directed dynamics such as adaptive biasing force (ABF) [52–54] and hyperdynamics [55] were also derived from the

Please cite this article as: R.C. Bernardi, et al., Enhanced sampling techniques in molecular dynamics simulations of biological systems, Biochim. Biophys. Acta (2014), http://dx.doi.org/10.1016/j.bbagen.2014.10.019