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

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.

© 2014 Elsevier B.V. All rights reserved.

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

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

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.

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].
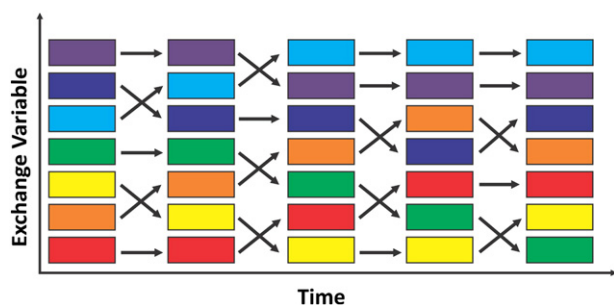
In the past decade REMD became a widely used method to enhance conformational sampling of MD simulations [4], and several variants of the traditional temperature dependent REMD (T-REMD) were implemented. While T-REMD has proven to be extremely useful in enhancing sampling in MD, it does not guarantee 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].
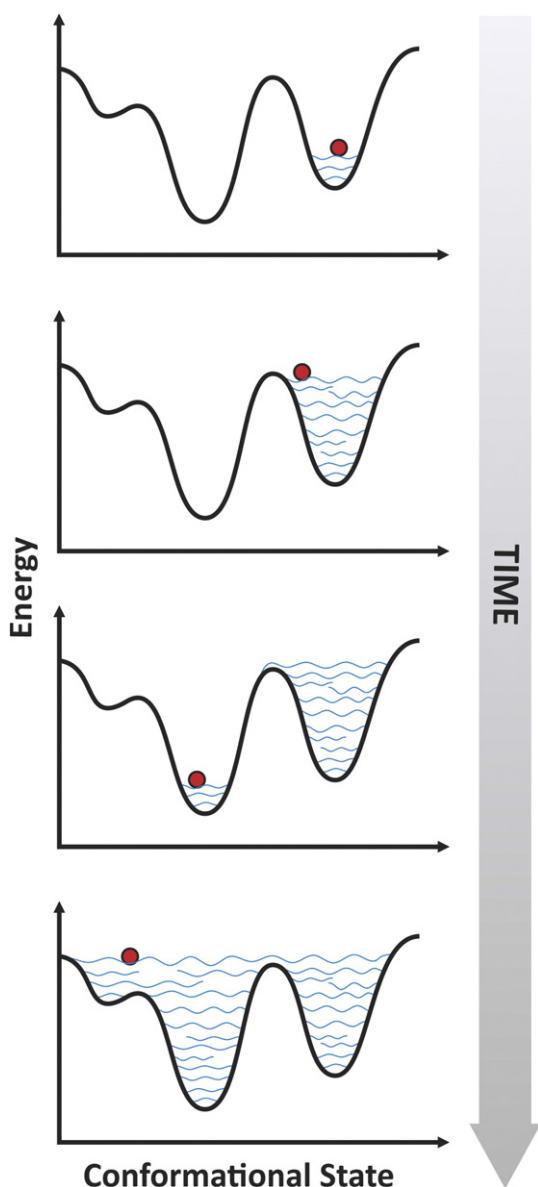
3. Metadynamics

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

Fig. 2. Illustration of the metadynamics method. Described as "filling the free energy wells with computational sand", the metadynamics method allows the search inside each energy well avoiding an oversampling of the same conformations. When the system reaches a point where the energy is higher than a barrier separating two minima the system goes to a state of lower energy in the new minimum, again searching many possible conformations.

same principles as adopted by metadynamics [45]. The implementation of metadynamics in MD codes, such as NAMD [41] and Gromacs [40], gave rise to a broad range of applications of the method, from solid-state physics to biology [56]. It was recently reported that metadynamics could be efficiently used to describe ligand–protein as well as protein–protein interactions [57].
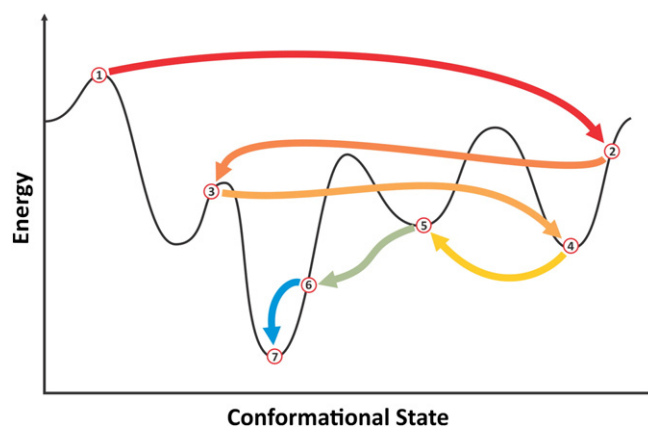
4. Simulated annealing

In analogy to the tempering process (annealing) in metallurgy, where a molten metal is gradually cooled until it reaches a crystalline structure (global minimum), the simulated annealing methods depend on an artificial temperature [58] that decreases during the simulation. For decades, several methods were derived from the stochastic "Monte Carlo" method [59], among them are classical simulated annealing (CSA) [60], fast simulated annealing (FSA) [61] and generalized simulated annealing (GSA) [58]. A similar approach is provided by the

genetic algorithms (GA), which generate solutions for search, optimization and machine learning problems, applying for the purpose techniques inspired by biological evolution [62]. Different attempts were made to quantitatively compare GA and the different versions of simulated annealing, concluding that the Monte Carlo derived methods are more efficient for biomolecular studies [63]. In particular, GSA has been shown to be useful when applied to global optimization problems that involve long range interactions, such as in atomic parameterization [64], gravitational systems [65] and conformational optimization of small molecules and peptides [66–72]. GSA has been shown to be more efficient than CSA and FSA as the number of dimensions in the phase space increases [73].

GSA efficiency emerges from the fact that it can explore the conformational space more homogeneously, maintaining the ability to make long jumps even at lower temperatures (see Fig. 3). The characteristics of GSA that makes it a useful method in molecular simulations stems from its basis in the non-additive statistical mechanics proposed by Tsallis in 1988 [74]. In Tsallis' statistical mechanics, the added entropy values of two systems A and B is not equal to the entropy of the compound system A + B. Instead, the long-range interactions are given an opportunity to add a special weight that depends on a free parameter q, which defines how much one system depends on the other due to the long-range interactions. When q goes to 1, the statistics takes the form of the additive entropies proposed by Boltzmann–Gibbs.

In order to apply GSA as a global optimization algorithm, the annealing procedure is encapsulated in three equations [58], one to describe the "temperature" decay, another for the random variable which drives the changes in the system at each iteration (also called the visitation function), and one to describe the acceptance probability. The temperature, in fact, does not represent the physical temperature of the system. Rather, it is only a parameter of the method which controls the random search. As it decreases, the probability of accepting a state with a higher cost (higher energy in an energy minimization, for example) will decrease, as will the changes in the system given by the random variable. Each equation has a free parameter, respectively, q_t, q_v and q_a, which have no direct physical interpretation. Initially only two were proposed and q_v would be equal q_a, but it was shown that using a dedicated parameter for the temperature decay gave better results in the case of protein structure prediction [67].

Fig. 3. Illustration of the simulated annealing method. Starting from a non-minimized structure and with a high temperature, the structure can randomly assume any conformation. A jump from conformation (1) to (2) is accepted if the simulated annealing acceptance criterion is achieved. Since the energy of conformation (2) is lower than the energy of the conformation (1), it is automatically accepted. Because the temperature decreases with each step, the modifications have a tendency to be progressively smaller. A conformation with a higher temperature can also be accepted as in the jump from (4) to (5). At the end the jumps are much smaller and there is a tendency to find a local minimum. Due to the high conformational energy at the beginning of the simulation, a large number of independent simulations should be capable of sampling all the possible stable conformations of the studied molecule.

The implementation of GSA in a broadly used MD code, namely NAMD, proved to be very powerful in case of folding studies. In fact, with GSA folding of small peptides is feasible even on desktop and small cluster computers [69]. A better interconnection between the GSA implementation (GSAFold algorithm) and NAMD is under development and will increase significantly the efficiency of the code when the potential energy of structures is calculated. These improvements allow one to apply GSA to the investigation of large conformational transitions in very large proteins, as demonstrated in the following section.

5. An example of a new application of Generalized Simulated Annealing

Studies by means of enhanced sampling algorithms involved either structure prediction of biomolecules or description of conformational changes that molecules go through as they interact with the environment [69,75–79]. Studying processes that occur on long timescales is also a common usage of sampling algorithms [80]. Macromolecules that exhibit structural diversity and do not have a stable folded conformation, but instead constantly adapt their shape to a changing environment, are very well suited for enhanced sampling methods. A good example of protein complexes with the latter characteristics are cellulosomes, macromolecular complexes specialized in cellulose degradation [81,82]. Analogous to a "Swiss army knife", cellulosomes can carry out a plethora of different catalytic and substrate binding activities that facilitate the degradation of plant cell wall material, a facility which is of great importance for the emerging biofuel industry [83–85]. Simulated annealing methods can help one to identify at a relatively low computational cost the conformations that a cellulosome can assume. As previously discussed, GSA has been shown to be more efficient than other simulated annealing methods [73]. In order to demonstrate for the case of cellulosomes the application of a powerful annealing method, we studied a fragment of the *Clostridium thermocellum* CipA scaffolding in complex with the SdbA type II cohesin module (CohII) [86,87]. The CipA fragment studied consisted of four components: (i) CipA's ninth type I cohesin (CohI$_9$) (ii) a small flexible linker (iii) an X-module (Xmod) and (iv) a type II dockerin (DocII).

Cellulosomes are formed by interactions of dockerins and cohesins that are connected by flexible linkers made of proteins with just a few to more than a thousand amino acids. The many flexible linker regions allow for very complex structural dynamics. Some insight into cellulosome dynamics can be gained by determining possible conformations through the use of enhanced sampling methods. GSAFold [69], the implementation of GSA in NAMD introduced above, is employed here to explore conformations of the linker region. The structure used in the present simulation is from the Protein Data Bank [88] (PDB ID: 3KCP). The crystal structure was minimized and relaxed through 100 ps MD simulation using NAMD [41,89]. The resulting structure was employed then in a stochastic calculation with GSAFold where only the linker region was allowed to rotate (see Fig. 4A). Briefly stated, the random sampling of conformations was accomplished by generating perturbations to the peptide backbone and side chain. Each single covalent bond in the flexible linker region was allowed to rotate; none of the other chemical bonds in the CohI$_9$-Xmod-DocII:CohII fragment were rotated, maintaining their minimized and equilibrated structure.

Using the temperature, visitation and acceptance parameters presented in [69], we carried out 51,200 GSA simulations with 10,000 steps each, which required about 4.5 h on 51,200 Blue Waters Supercomputer processors. The results, presented in Fig. 4B, show that the system is most likely to be found in two conformations. As shown in Fig. 4B the crystal structure corresponds to the "center" of the most frequently visited conformations, demonstrating that the method finds a thermodynamically stable state. In order to visualize the two conformations mentioned, the center of mass of the CohI$_9$ module was plotted in Figs. 1B and C to depict the conformations visited by the complex. As one can observe, there is a tolerance to the positioning of the module
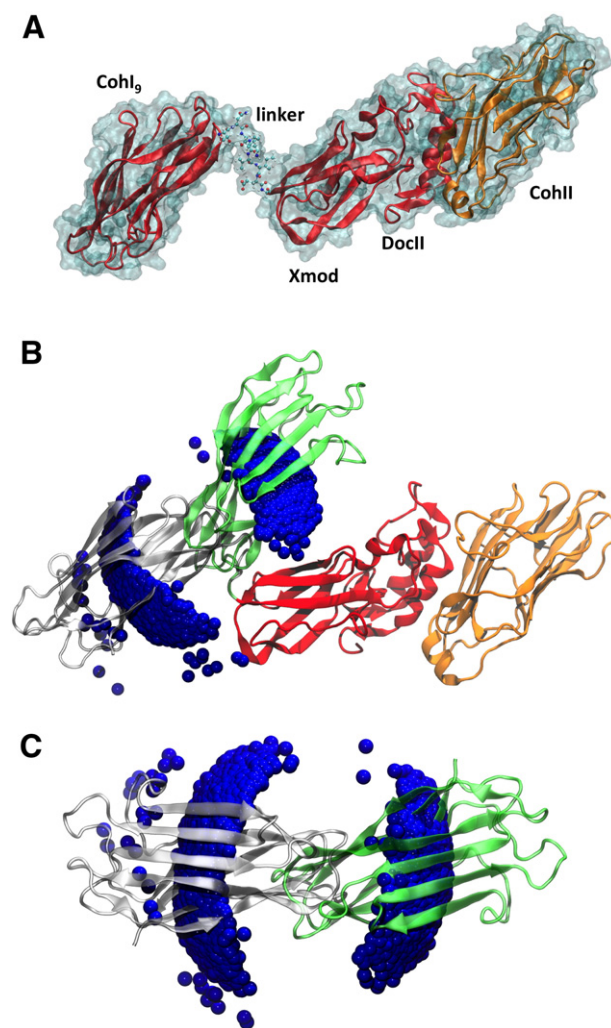
Fig. 4. Generalized simulated annealing study of a cellulosome fragment. A: *C. thermocellum* CipA scaffolding fragment in complex with the SdbA type II cohesin module (CohII). CipA fragment components: ninth type I cohesin (CohI9), small flexible linker, X-module (Xmod) and type II dockerin (DocII). The cartoon representation depicts the shape of the protein backbone while the flexible linking region is shown in atomic representation. This region includes THR150 and GLY157 and all six amino acids in between. Only the single covalent bonds from the eight residues in this region were allowed to change throughout the simulations, both for backbone and side chains. B: Center of Mass of CohI$_9$ module in all visited conformations, shown as a blue sphere. The center of mass determination was done using only the atoms in the CohI9 module, excluding the flexible X module. Both calculation and figure were created using VMD [90]. The analysis shows that the CohI9 module visited mainly two opposite positions. The arched-band positioning of the centers of mass show how much the module can fluctuate around each position. In gray, one can observe the "native" position of the CohI$_9$ module and the linker in the final minimized and equilibrated structure, initially taken from the crystal structure deposited in the PDB. The structure in gray was used to initiate all simulations and is not part of the results, it is, however, in the center of the region most visited by the independent simulations. In green, a second structure is shown that represents in an exemplary fashion the "alternate" position of the CohI$_9$ module and linker visiting the second conformational cluster. C: Detail of the two conformations of CohI$_9$ observed in B, and the arched bands showing the positions of centers of mass of final conformations from all 51,200 simulations.

in both conformations that allows the cellulosome modules to adjust the large enzymatic domains attached to CohI$_9$. The results show that, even though very flexible, the linker led the CohI$_9$ module to two preferred positions, with about 2/3 occurrences of the "native" conformation as seen in the crystal structure, and 1/3 occurrences of the "alternate" conformation. The short flexible linker, one of the shortest in the *C. thermocellum* cellulosome, appears to have the function of keeping the system in two major conformations and not completely random ones as one might have expected.

Regarding the potential energy of the cellulosome system, each of the 51,200 simulations led to a structure that minimized locally the system's energy. The distribution of these final energies was found to fit a Gaussian distribution as shown in Fig. 5. The small difference in the average potential energy (about 1.2 kcal/mol) between the two conformations, larger than the magnitude of the thermal energy k_BT, suggests that it is essential for the proper functioning of the module to have two stable conformations, but also to switch between them.

6. Conclusions

The evolution in the development of REMD based methods have shown that these methods can be applied to a broad range of problems, from the smallest peptides to large molecular systems. However, REMD seems to be limited to systems that present energy landscapes which are not particularly rough. Metadynamics, on the other hand, is useful in cases where the local equilibration of intermediate steps of simulations is particularly difficult, allowing a broader exploration of the energy surface. However, despite the intelligent algorithm that prevents re-sampling, metadynamics is computationally expensive, restricting its usage to smaller systems, such as peptides or small proteins. Metadynamics usage in drug docking to protein and enzymes is certainly an important contribution of this method to cases where normal docking tools would easily fail, such as systems where big conformational changes are necessary or when the solvent plays an important role.

As we show here, simulated annealing methods, in particular GSA, have moved beyond their initial application to small molecule conformational optimization. GSA has emerged as a tool suitable for large macromolecular systems, which would require, in case those other sampling methods are employed, a prohibitive amount of computer time to comprehensively explore their conformational space. The fact that GSA does not employ full MD gives it high speed and parallelizability, though GSA is still hindered by its inability to explicitly resolve solvation. As we show here, GSAFold is a relatively cheap algorithm to study conformational distributions. Usage of other techniques, such as adaptive biasing force

simulation, could prove useful for a clearer picture of the kinetics of the process of moving from one conformation to another.

Acknowledgements

References

[1] M. Karplus, J.A. McCammon, Molecular dynamics simulations of biomolecules, Nat. Struct. Biol. 9 (2002) 646–652.
[2] M. Orozco, A theoretical view of protein dynamics, Chem. Soc. Rev. 43 (2014) 5051–5066.
[3] R.C. Bernardi, D.E.B. Gomes, R. Gobato, C.A. Taft, A.T. Ota, P.G. Pascutti, Molecular dynamics study of biomembrane/local anesthetics interactions, Mol. Phys. 107 (2009) 1437–1443.
[4] N.-V. Buchete, G. Hummer, Peptide folding kinetics from replica exchange molecular dynamics, Phys. Rev. E. 77 (2008) 030902.
[5] Y. Liu, J. Strümpfer, P.L. Freddolino, M. Gruebele, K. Schulten, Structural characterization of λ-repressor folding from all-atom molecular dynamics simulations, J. Phys. Chem. Lett. 3 (2012) 1117–1123.
[6] S. Sothiselvam, B. Liu, W. Han, H. Ramu, D. Klepacki, G.C. Atkinson, et al., Macrolide antibiotics allosterically predispose the ribosome for translation arrest, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 9804–9809.
[7] X. Wang, F. Xu, J. Liu, B. Gao, Y. Liu, Y. Zhai, et al., Atomic model of rabbit hemorrhagic disease virus by cryo-electron microscopy and crystallography, PLoS Pathog. 9 (2013) e1003132.
[8] G. Zhao, J.R. Perilla, E.L. Yufenyuy, X. Meng, B. Chen, J. Ning, et al., Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics, Nature 497 (2013) 643–646.
[9] H.M. Senn, W. Thiel, QM/MM methods for biological systems, At. Approaches Mod, Biol. From Quantum Chem. to Mol. Simulations, 2682007. 173–290.
[10] R.C. Bernardi, P.G. Pascutti, Hybrid QM/MM molecular dynamics study of benzocaine in a membrane environment: how does a quantum mechanical treatment of both anesthetic and lipids affect their interaction, J. Chem. Theory Comput. 8 (2012) 2197–2203.
[11] J.D. Durrant, J.A. McCammon, Molecular dynamics simulations and drug discovery, BMC Biol. 9 (2011) 71.
[12] R. Schulz, B. Lindner, L. Petridis, J.C. Smith, Scaling of multimillion-atom biological molecular dynamics simulation on a petascale supercomputer, J. Chem. Theory Comput. 5 (2009) 2798–2808.
[13] K. Schulten, J.C. Phillips, L.V. Kalé, A. Bhatele, Biomolecular modeling in the era of petascale computing, in: D.A. Bader (Ed.), Petascale Comput. — Algorithms Appl, Chapman & Hall/CRC, Boca Raton, 2007, pp. 165–181.
[14] C. Bergonzo, N.M. Henriksen, D.R. Roe, J.M. Swails, A.E. Roitberg, T.E. Cheatham, Multidimensional replica exchange molecular dynamics yields a converged ensemble of an RNA tetranucleotide, J. Chem. Theory Comput. 10 (2014) 492–499.
[15] C. Chen, Y. Huang, Y. Xiao, Enhanced sampling of molecular dynamics simulation of peptides and proteins by double coupling to thermal bath, J. Biomol. Struct. Dyn. 31 (2013) 206–214.
[16] J.N. Onuchic, Z. Luthey-Schulten, P.G. Wolynes, Theory of protein folding: the energy landscape perspective, Annu. Rev. Phys. Chem. 48 (1997) 545–600.
[17] S. Marsili, G.F. Signorini, R. Chelli, M. Marchi, P. Procacci, Software news and updates ORAC: a molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level, J. Comput. Chem. 31 (2009) 1106–1116.
[18] C. Liao, J. Zhou, Replica-exchange molecular dynamics simulation of basic fibroblast growth factor adsorption on hydroxyapatite, J. Phys. Chem. B 118 (2014) 5843–5852.
[19] P.R. Batista, G. Pandey, P.G. Pascutti, P.M. Bisch, D. Perahia, C.H. Robert, Free energy profiles along consensus normal modes provide insight, J. Chem. Theory Comput. 7 (2011) 2348–2352.
[20] R.C. Bernardi, I. Cann, K. Schulten, Molecular dynamics study of enhanced Man5B enzymatic activity, Biotechnol. Biofuels 7 (2014) 1–8.
[21] F. Khalili-Araghi, J. Gumbart, P.-C. Wen, M. Sotomayor, E. Tajkhorshid, K. Schulten, Molecular dynamics simulations of membrane channels and transporters, Curr. Opin. Struct. Biol. 19 (2009) 128–137.
[22] K. Hukushima, Extended ensemble Monte Carlo approach to hardly relaxing problems, Comput. Phys. Commun. 147 (2002) 77–82.
[23] K. Hukushima, K. Nemoto, Exchange Monte Carlo method and application to spin glass simulations, J. Phys. Soc. Jpn. 65 (1996) 1604–1608.
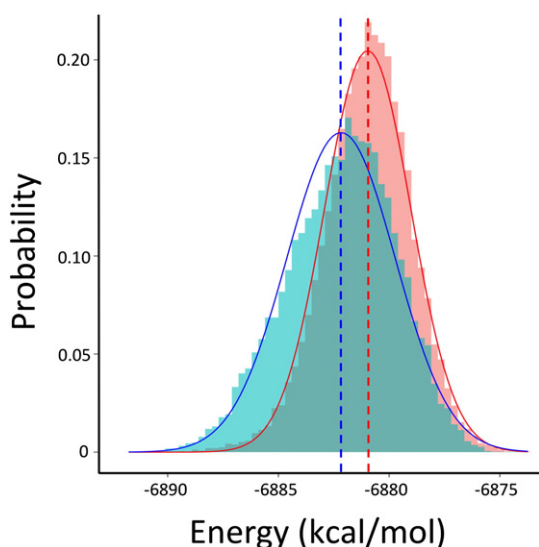
Fig. 5. Potential energy distribution of the cellulosome fragment after 10,000 steps of GSA simulations. The 51,200 different conformations resulted in 51,200 different energies that are coupled according to the two main clusters of conformations. The "native" cluster (represented in red) occurs in 34,936 structures, or 68% of the cases. The "alternate" cluster (represented in blue) occurs in 16,264 structures, or 32% of the cases. All structures were assigned to a cluster using the software package Weka [91]. The software package R [92] was used to calculate a normalized histogram of the potential energy of the structures in each cluster. A Gaussian distribution fits each of the distributions and the difference in average energy between the two conformations is about 1.2 kcal/mol.

[24] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, Chem. Phys. Lett. 314 (1999) 141–151.

[25] Y. Meng, A.E. Roitberg, V. Gaines, Constant pH replica exchange molecular dynamics in biomolecules using a discrete protonation model, 2010. 1401–1412.

[26] C. Lockhart, D.K. Klimov, Alzheimer ' s A β 10–40 peptide binds and penetrates DMPC bilayer: an isobaric–isothermal replica exchange molecular dynamics study, J. Phys. Chem. B 118 (2014) 2638–2648.

[27] H. Nymeyer, How efficient is replica exchange molecular dynamics? An analytic approach, J. Chem. Theory Comput. 4 (2008) 626–636.

[28] D.A.C. Beck, G.W.N. White, V. Daggett, Exploring the energy landscape of protein folding using replica-exchange and conventional molecular dynamics simulations, J. Struct. Biol. 157 (2007) 514–523.

[29] S. Gnanakaran, Peptide folding simulations, Curr. Opin. Struct. Biol. 13 (2003) 168–174.

[30] G. Hummer, I.G. Kevrekidis, Coarse molecular dynamics of a peptide fragment: free energy, kinetics, and long-time dynamics computations, J. Chem. Phys. 118 (2003) 10762.

[31] P.H. Nguyen, G. Stock, E. Mittag, C.-K. Hu, M.S. Li, Free energy landscape and folding mechanism of a beta-hairpin in explicit water: a replica exchange molecular dynamics study, Proteins 61 (2005) 795–808.

[32] A. Liwo, C. Czaplewski, S. Ołdziej, H.A. Scheraga, Computational techniques for efficient conformational sampling of proteins, Curr. Opin. Struct. Biol. 18 (2008) 134–139.

[33] D.R. Roe, C. Bergonzo, T.E. Cheatham, Evaluation of enhanced sampling provided by accelerated molecular dynamics with Hamiltonian replica exchange methods, J. Phys. Chem. B 118 (2014) 3543–3552.

[34] K. Ostermeir, M. Zacharias, Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins, Biochim. Biophys. Acta 1834 (2013) 847–853.

[35] Y.M. Rhee, V.S. Pande, Multiplexed-replica exchange molecular dynamics method for protein folding simulation, Biophys. J. 84 (2003) 775–786.

[36] W. Jiang, M. Hodoscek, B. Roux, Computation of absolute hydration and binding free energy with Free Energy Perturbation Distributed Replica-Exchange Molecular Dynamics (FEP/REMD), J. Chem. Theory Comput. 5 (2009) 2583–2588.

[37] W. Jiang, J.C. Phillips, L. Huang, M. Fajer, Y. Meng, J.C. Gumbart, et al., Generalized scalable multiple copy algorithms for molecular dynamics simulations in NAMD, Comput. Phys. Commun. 185 (2014) 908–916.

[38] W. Jiang, B. Roux, Free energy perturbation Hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations, J. Chem. Theory Comput. 6 (2010) 2559–2565.

[39] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, et al., The Amber biomolecular simulation programs, J. Comput. Chem. 26 (2005) 1668–1688.

[40] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, et al., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29 (2013) 845–854.

[41] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, et al., Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781–1802.

[42] A. Laio, M. Parrinello, Escaping free-energy minima, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 12562–12566.

[43] T. Huber, A.E. Torda, W.F. Gunsteren, Local elevation: a method for improving the searching properties of molecular dynamics simulation, J. Comput. Aided Mol. Des. 8 (1994) 695–708.

[44] H. Grubmüller, Predicting slow structural transitions in macromolecular systems: conformational flooding, Phys. Rev. E. 52 (1995) 2893–2906.

[45] E. Darve, A. Pohorille, Calculating free energies using average force, J. Chem. Phys. 115 (2001) 9169.

[46] G. Bussi, F.L. Gervasio, A. Laio, M. Parrinello, Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics, J. Am. Chem. Soc. 128 (2006) 13435–13441.

[47] F.L. Gervasio, A. Laio, M. Parrinello, Flexible docking in solution using metadynamics, J. Am. Chem. Soc. 127 (2005) 2600–2607.

[48] R. Martoňák, A. Laio, M. Parrinello, Predicting crystal structures: the Parrinello–Rahman method revisited, Phys. Rev. Lett. 90 (2003) 075503.

[49] R. Martonak, A. Laio, M. Bernasconi, C. Ceriani, P. Raiteri, M. Parrinello, Simulation of structural phase transitions by metadynamics, Z. Krist. 220 (2005) 489–498.

[50] V. Spiwok, P. Lipovová, B. Králová, Metadynamics in essential coordinates: free energy simulation of conformational changes, J. Phys. Chem. B 111 (2007) 3073–3076.

[51] O. Engkvist, G. Karlström, A method to calculate the probability distribution for systems with large energy barriers, Chem. Phys. 213 (1996) 63–76.

[52] J. Hénin, C. Chipot, Overcoming free energy barriers using unconstrained molecular dynamics simulations, J. Chem. Phys. 121 (2004) 2904–2914.

[53] J. Comer, B. Roux, C. Chipot, Achieving ergodic sampling using replica-exchange free-energy calculations, Mol. Simul. 40 (2013) 218–228.

[54] J. Hénin, E. Tajkhorshid, K. Schulten, C. Chipot, Diffusion of glycerol through Escherichia coli aquaglyceroporin GlpF, Biophys. J. 94 (2008) 832–839.

[55] A. Voter, Hyperdynamics: accelerated molecular dynamics of infrequent events, Phys. Rev. Lett. 78 (1997) 3908–3911.

[56] A. Barducci, M. Bonomi, M. Parrinello, Metadynamics, Wiley Interdiscip. Rev. Comput. Mol. Sci. 1 (2011) 826–843.

[57] V. Limongelli, M. Bonomi, M. Parrinello, Funnel metadynamics as accurate binding free-energy method, 110 (2013) 6358–6363.

[58] C. Tsallis, D.A. Stariolo, Generalized simulated annealing, Phys. A Stat. Theor. Phys. 233 (1996) 395–406.

[59] N. Metropolis, S. Ulam, The Monte Carlo method, J. Am. Stat. Assoc. 44 (1949) 335–341.

[60] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, Science 220 (1983) 671–680.

[61] H. Szu, R. Hartley, Fast simulated annealing, Phys. Lett. A 122 (1987) 157–162.

[62] H. Zang, S. Zhang, K. Hapeshi, A review of nature-inspired algorithms, J. Bionic Eng. 7 (2010) S232–S237.

[63] K. Hamacher, W. Wenzel, Scaling behavior of stochastic minimization algorithms in a perfect funnel landscape, Phys. Rev. E. 59 (1999) 938–941.

[64] M.D. De Andrade, M.A.C. Nascimento, K.C. Mundim, A.M.C. Sobrinho, L.A.C. Malbouisson, Atomic basis sets optimization using the generalized simulated annealing approach: new basis sets for the first row elements, Int. J. Quantum Chem. 108 (2008) 2486–2498.

[65] K. Mundim, T. Lemaire, A. Bassrei, Optimization of non-linear gravity models through generalized simulated annealing, Phys. A Stat. Mech. Appl. 252 (1998) 405–416.

[66] K.C. Mundim, C. Tsallis, Geometry optimization and conformational analysis through generalized simulated annealing, 58 (1995) 373–381.

[67] F.P. Agostini, D.D.O. Soares-Pinto, M.A. Moret, C. Osthoff, P.G. Pascutti, Generalized simulated annealing applied to protein folding studies, J. Comput. Chem. 27 (2006) 1142–1155.

[68] S.S. da Rocha Pita, T.V.A. Fernandes, E.R. Caffarena, P.G. Pascutti, Studies of molecular docking between fibroblast growth factor and heparin using generalized simulated annealing, Int. J. Quantum Chem. 108 (2008) 2608–2614.

[69] M.C.R. Melo, R.C. Bernardi, T.V. a Fernandes, P.G. Pascutti, GSAFold: a new application of GSA to protein structure prediction, Proteins 80 (2012) 2305–2310.

[70] M.C.R. Melo, R.C. Bernardi, P.G. Pascutti, Large scale structure sampling for protein fold prediction using the generalized simulated annealing, Biophys. J. 104 (2013) 228.

[71] M.A. Moret, P.M. Bisch, K.C. Mundim, P.G. Pascutti, New stochastic strategy to analyze helix folding, Biophys. J. 82 (2002) 1123–1132.

[72] M.A. Moret, P.G. Pascutti, P.M. Bisch, K.C. Mundim, Stochastic molecular optimization using generalized simulated annealing, J. Comput. Chem. 19 (1998) 647–657.

[73] Y. Xiang, X.G. Gong, Efficiency of generalized simulated annealing, Phys. Rev. E. 62 (2000) 4473–4476.

[74] C. Tsallis, Possible generalization of Boltzmann–Gibbs statistics, J. Stat. Phys. 52 (1988) 479–487.

[75] G. Helles, A comparative study of the reported performance of ab initio protein structure prediction algorithms, J. R. Soc. Interface 5 (2008) 387–396.

[76] L.-H. Hung, S.-C. Ngan, T. Liu, R. Samudrala, PROTINFO: new algorithms for enhanced protein structure predictions, Nucleic Acids Res. 33 (2005) W77–W80.

[77] B. Jayaram, K. Bhushan, S.R. Shenoy, P. Narang, S. Bose, P. Agrawal, et al., Bhageerath: an energy based web enabled computer software suite for limiting the search space of tertiary structures of small globular proteins, Nucleic Acids Res. 34 (2006) 6195–6204.

[78] X. Zhang, T. Wang, H. Luo, J.Y. Yang, Y. Deng, J. Tang, et al., 3D protein structure prediction with genetic tabu search algorithm, BMC Syst. Biol. 4 (Suppl. 1) (2010) S6.

[79] A. Pavlopoulou, I. Michalopoulos, State-of-the-art bioinformatics protein structure prediction tools (Review), Int. J. Mol. Med. 28 (2011) 295–310.

[80] L.C.T. Pierce, R. Salomon-Ferrer, C. Augusto, F. de Oliveira, J.A. McCammon, R.C. Walker, Routine access to millisecond time scale events with accelerated molecular dynamics, J. Chem. Theory Comput. 8 (2012) 2997–3002.

[81] E. a Bayer, L.J. Shimon, Y. Shoham, R. Lamed, Cellulosomes—structure and ultrastructure, J. Struct. Biol. 124 (1998) 221–234.

[82] Y.J. Bomble, G.T. Beckham, J.F. Matthews, M.R. Nimlos, M.E. Himmel, M.F. Crowley, Modeling the self-assembly of the cellulosome enzyme complex, J. Biol. Chem. 286 (2011) 5614–5623.

[83] S.-Y. Ding, Q. Xu, M. Crowley, Y. Zeng, M. Nimlos, R. Lamed, et al., A biophysical perspective on the cellulosome: new opportunities for biomass conversion, Curr. Opin. Biotechnol. 19 (2008) 218–227.

[84] R.H. Doi, A. Kosugi, Cellulosomes: plant-cell-wall-degrading enzyme complexes, Nat. Rev. Microbiol. 2 (2004) 541–551.

[85] C.M.G.A. Fontes, H.J. Gilbert, Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates, Annu. Rev. Biochem. 79 (2010) 655–681.

[86] J.J. Adams, M.A. Currie, S. Ali, E.A. Bayer, Z. Jia, S.P. Smith, Insights into higher-order organization of the cellulosome revealed by a dissect-and-build approach: crystal structure of interacting Clostridium thermocellum multimodular components, J. Mol. Biol. 396 (2010) 833–839.

[87] M.A. Currie, J.J. Adams, F. Faucher, E.A. Bayer, Z. Jia, S.P. Smith, Scaffoldin conformation and dynamics revealed by a ternary complex from the Clostridium thermocellum cellulosome, J. Biol. Chem. 287 (2012) 26953–26961.

[88] H.M. Berman, The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242.

[89] L. Kale, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, et al., NAMD2: greater scalability for parallel molecular dynamics, J. Comput. Phys. 151 (1998) 283–312.

[90] W. Humphrey, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.

[91] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software, ACM SIGKDD Explor. Newsl. 11 (2009) 10.

[92] R Core Team, R: a language and environment for statistical, computing, 2014.