ARTICLE IN PRESS

BBAGEN-27985; No. of pages: 12; 4C: 2, 3, 4, 5, 7, 9, 10

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

Recent advances in QM/MM free energy calculations using reference potentials[☆]

Fernanda Duarte[*], Beat A. Amrein, David Blaha-Nelson, Shina C.L. Kamerlin[*]

Science for Life Laboratory, Department of Cell and Molecular Biology (ICM), Uppsala University, BMC Box 596, S-751 24 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 24 May 2014
Received in revised form 6 July 2014
Accepted 7 July 2014
Available online xxxx

Keywords:
Multiscale modeling
QM/MM free energy calculation
Averaging potential
Reference potential
Mean field approximation

ABSTRACT

Background: Recent years have seen enormous progress in the development of methods for modeling (bio)molecular systems. This has allowed for the simulation of ever larger and more complex systems. However, as such complexity increases, the requirements needed for these models to be accurate and physically meaningful become more and more difficult to fulfill. The use of simplified models to describe complex biological systems has long been shown to be an effective way to overcome some of the limitations associated with this computational cost in a rational way.

Scope of review: Hybrid QM/MM approaches have rapidly become one of the most popular computational tools for studying chemical reactivity in biomolecular systems. However, the high cost involved in performing high-level QM calculations has limited the applicability of these approaches when calculating free energies of chemical processes. In this review, we present some of the advances in using reference potentials and mean field approximations to accelerate high-level QM/MM calculations. We present illustrative applications of these approaches and discuss challenges and future perspectives for the field.

Major conclusions: The use of physically-based simplifications has shown to effectively reduce the cost of high-level QM/MM calculations. In particular, lower-level reference potentials enable one to reduce the cost of expensive free energy calculations, thus expanding the scope of problems that can be addressed.

General significance: As was already demonstrated 40 years ago, the usage of simplified models still allows one to obtain cutting edge results with substantially reduced computational cost. This article is part of a Special Issue entitled Recent developments of molecular dynamics.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

1. Introduction

Recent years have seen enormous progress in the development of new methods for modeling molecular systems [1–6]. The introduction of massively parallelized computer architectures [7–9], together with the availability of more efficient algorithms, has allowed for a spectacular increase in terms of the size of the molecular systems studied [10,11] as well as the length of the simulations that can be performed [6,12]. For instance, the development of approximate methods for first-principles quantum chemistry [1,3,13] has made it possible to carry out electronic structure calculations for systems as big as millions of atoms [13–15]. Concurrently, the introduction of alternative hardware to perform molecular dynamics (MD) simulations [7–9] now allows one to run single trajectories of μs [16,17] and, more recently, even MD simulations of ms in length [12,18]. Despite the many advances in the field, some of these technologies have failed to make the expected impact on the scientific community, in part due to their still excessive computational cost. Additionally, in the case of recent developments in first-principles quantum chemistry [1,3,13], their more difficult implementation and limited accessibility compared to conventional codes remain bottlenecks. Finally, despite the fact that running ms trajectories is computationally impressive, the computational cost involved discourages running sufficient replicas in order to test for reproducibility.

To perform large scale free energy calculations on complex systems, two key requirements need to be fulfilled: 1) the approach should be sufficiently accurate to describe the system under study in a physically meaningful way and 2) sufficient resources are needed to adequately sample the configurational space to draw unambiguous conclusions that are not dependent on starting structure. Therefore, one is stuck in balancing the cost of the simulations with the resources available. The availability of more powerful computer resources has partially resolved these issues allowing simulations of larger and larger systems. Examples

---

Abbreviations: ABF, adaptive biasing force; CG, coarse-grained; EVB, empirical valence bond; FEP, free energy perturbation; LRA, linear response approximation; MD, molecular dynamics; PD, paradynamics; PMF, potential of mean force; QM/MM, quantum mechanics/molecular mechanics; QTCP, QM/MM thermodynamic cycle perturbation; US, umbrella sampling; REMD, replica exchange molecular dynamics; RS, reactant state; SCC-DFTB, self-consistent charge density functional tight binding; TS, transition state

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

[*] Corresponding authors.

E-mail addresses: fernanda.duarte@icm.uu.se (F. Duarte), kamerlin@icm.uu.se (S.C.L. Kamerlin).

http://dx.doi.org/10.1016/j.bbagen.2014.07.008

0304-4165/© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Please cite this article as: F. Duarte, et al., Recent advances in QM/MM free energy calculations using reference potentials, Biochim. Biophys. Acta (2014), http://dx.doi.org/10.1016/j.bbagen.2014.07.008

of this include the ribosome system of approximately 2.64 million atoms [11] or the complete shell of southern bean mosaic virus, which in explicit solvent led to a system that comprised more than 4.5 million atoms and a total simulation time of about 100 ns [10]. Even with contemporary computer power, as the complexity of the systems of interest increases, these two requirements become more and more difficult to fulfill, further pushing the need for custom codes and specialized hardware. An alternative solution, in order to extend the range of problems that can be addressed, is the introduction of some level of simplification to the models being used. The use of simplified models is particularly important in the case of QM/MM free energy calculations, where even current computational resources do not allow for both a full high-level QM description of the reacting system and enough sampling of the conformational space [19] to obtain reasonable convergent free energies. Often the only solution to this problem has been to move to cheaper (but less quantitatively precise) semi-empirical models [5,19–21].

The use of simplified models in biomolecular simulations dates back to the 1970s, when the pioneering work of Levitt and Warshel [22] introduced the use of coarse-grained (CG) methods to the study of protein-folding processes. These models allowed one to significantly reduce the number of degrees of freedom of the system under study and therefore run longer simulations that would otherwise have been impossible using the resources available at that time. The first example of this was on bovine pancreatic trypsin inhibitor, BPTI [22], a small (<100 amino acids) single polypeptide chain protein of known conformation at that time (Fig. 1). In that study, the protein side chains were represented by spheres with an effective potential (implicitly representing the average potential of the solvated side chains) and the main chain was represented by virtual bonds between the C_α's. Using such a model, it was possible to correctly fold this protein in about 1000 minimization cycles, as outlined in Ref. [22] and illustrated in Fig. 1.

Similar approaches have subsequently been used in a variety of processes, including DNA and RNA folding [23,24], assemblies of membrane proteins [25], and vesicle formation [26]. More recently, the idea of using a simplified model as a reference potential has been expanded to a wide range of chemical problems [27–31], long time-scale conformational dynamics of proteins [32], and other related processes [33,34].

Having addressed the issue of cost *vs.* accuracy of the calculations, the second problem is the need for extensive conformational sampling. In principle, one would expect that the evaluation of a standard unbiased trajectory would be sufficient to visit the different regions of the conformational space multiple times. However, this requires the unbiased trajectory to be extremely (and inefficiently) long, as the system under study will spend a large fraction of the time in regions of phase space that have already been visited. A number of enhanced and rare event sampling techniques have been developed in order to reduce this problem: umbrella sampling [36], thermodynamics integration [37], replica exchange molecular dynamics (REMD) [38], the adaptive biasing force (ABF) method [39], transition path sampling [40], accelerated MD [41], metadynamics (MTD) [42] and paradynamics [28], just to name a few examples (for further information on some of these approaches, we refer readers to *e.g.* Ref. [43]). When combined with simplified models, these techniques have been shown to be capable of overcoming some of the limitations associated with computational cost in rational ways.

Earlier works have already discussed the methodological aspects of QM/MM approaches in detail (*e.g.* [20,44]) and their applications in the modeling of complex biological systems (*e.g.* [5,45,46]). In the present review, we hope to complement these excellent works, now focusing on a particular case study, namely the use of simplified reference potentials for performing higher-level hybrid QM/MM calculations. We will first provide a generalized introduction to this concept, followed by a discussion of the state-of-the-art of QM/MM free energy calculations using reference potentials as well as illustrating some recent applications. We will then conclude with an overview and discussion of open challenges and future perspectives for the field.

2. Approximating the high-level behavior of complex systems with low-level models

When using a simplified model to study complex systems, it is essential to use a model that captures the physics of the full explicit

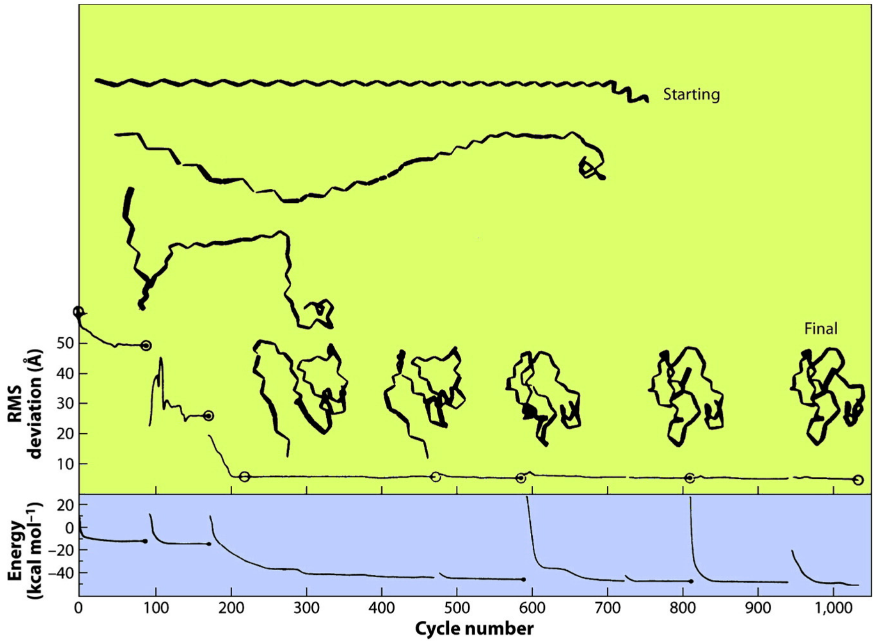

Fig. 1. Simulation of bovine pancreatic trypsin inhibitor (BPTI) folding from an extended starting conformation with the terminal helix. Each residue has only one degree of freedom, *i.e.* the torsional angle α between the four successive C_α atoms. In all cases, $\alpha = 180°$, except for residues 48 to 58 where $\alpha = 45°$. No other knowledge whatsoever about the native protein was used for the simulation. In order to avoid nonproductive changes in the protein conformation, after a cycle of minimization, thermal fluctuations were introduced under the condition that each mode has an average kinetic energy of $kT/2$ (where k is the Boltzmann constant and T is the absolute temperature).
Reprinted by permission from Macmillan Publishers Ltd: Nature [22], copyright 1975. Also adapted with permission from [35].

Please cite this article as: F. Duarte, et al., Recent advances in QM/MM free energy calculations using reference potentials, Biochim. Biophys. Acta (2014), http://dx.doi.org/10.1016/j.bbagen.2014.07.008

Download English Version:

https://daneshyari.com/en/article/10800049

Download Persian Version:

https://daneshyari.com/article/10800049

Daneshyari.com