BBAGEN-28036; No. of pages: 10; 4C: 3, 4, 6, 7

Biochimica et Biophysica Acta xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Efficiently computing pathway free energies: New approaches based on chain-of-replica and Non-Boltzmann Bennett reweighting schemes[☆]

Phillip S. Hudson [a,1], Justin K. White [a,1], Fiona L. Kearns [a], Milan Hodoscek [b], Stefan Boresch [c], H. Lee Woodcock [a,*]

[a] Department of Chemistry, University of South Florida, 4202 E. Fowler Ave., CHE205, Tampa, FL 33620-5250, USA
[b] Center for Molecular Modeling, National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia
[c] Department of Computational Biological Chemistry, Faculty of Chemistry, University of Vienna, Währingerstraße 17, A-1090 Vienna, Austria

ARTICLE INFO

Article history:
Received 14 July 2014
Received in revised form 9 September 2014
Accepted 10 September 2014
Available online xxxx

Keywords:
Reaction path
Free energy
QM/MM
Potential of mean force
Bennett's acceptance ratio
BAR
Reweighting
QM-Non-Boltzmann Bennett
QM-NBB

ABSTRACT

Background: Accurately modeling condensed phase processes is one of computation's most difficult challenges. Include the possibility that conformational dynamics may be coupled to chemical reactions, where multiscale (i.e., QM/MM) methods are needed, and this task becomes even more daunting.
Methods: Free energy simulations (i.e., molecular dynamics), multiscale modeling, and reweighting schemes.
Results: Herein, we present two new approaches for mitigating the aforementioned challenges. The first is a new chain-of-replica method (off-path simulations, OPS) for computing potentials of mean force (PMFs) along an easily defined reaction coordinate. This development is coupled with a new distributed, highly-parallel replica framework (REPDstr) within the CHARMM package. Validation of these new schemes is carried out on two processes that undergo conformational changes. First is the simple torsional rotation of butane, while a much more challenging glycosidic rotation (in vacuo and solvated) is the second. Additionally, a new approach that greatly improves (i.e., possibly an order of magnitude) the efficiency of computing QM/MM PMFs is introduced and compared to standard schemes. Our efforts are grounded in the recently developed method for efficiently computing QM-based free energies (i.e., QM-Non-Boltzmann Bennett, QM-NBB). Again, we validate this new technique by computing the QM/MM PMF of butane's torsional rotation.
Conclusions: The OPS–REPDstr method is a promising new approach that overcomes many limitations of standard pathway simulations in CHARMM. The combination of QM-NBB with pathway techniques is very promising as it offers significant advantages over current procedures.
General significance: Efficiently computing potentials of mean force is a major, unresolved, area of interest.
This article is part of a Special Issue entitled Recent developments of molecular dynamics.

1. Introduction

The determination of free energy differences, which has been stated to be the most important general quantity in physical chemistry, has fascinated computational scientists for many years. Work in this area falls into two major categories. First, it is of obvious interest to calculate the free energy difference between two states, such as solvation free energies or binding free energies. Such calculations rely on long known statistical mechanical identities, such as "free energy perturbation" (FEP) [1–4] or "thermodynamic integration" (TI) [5]. More recently, Bennett's acceptance ratio (BAR) method [6] has been rediscovered [7] and shown to be more efficient than either FEP or TI [8–10]. The second group of application is concerned with the determination of potentials-of-mean-force (PMFs), i.e., the free energy along a degree of freedom of interest, such as a torsional angle. Methods to compute PMFs include TI or the use of umbrella potentials [11,12], often in connection with the weighted histogram analysis method (WHAM) [13,14].

More recently, a vast array of techniques have been developed that combine chain-of-replica methods [15–17] with sampling approaches to compute free energy profiles [18–31]. In contrast to these methods, which have primarily been used to model processes via classical mechanics (i.e., MM), many alternative "string" type methods have been developed with the intention of modeling reactive processes via quantum mechanical (QM) or hybrid QM/MM Hamiltonians. This shift is not surprising as the improvement of QM methods has made possible the study of new areas of chemical space [32–37]. Currently, the majority of these new QM and QM/MM based reaction path methods have focused on obtaining minimum energy pathways (MEPs) [27,38–53]; however, more and more effort is being devoted to employing

---

[☆] This article is part of a Special Issue entitled Recent developments of molecular dynamics.
[*] Corresponding author.
E-mail address: hlw@usf.edu (H. Lee Woodcock).
[1] Authors contributed equally.

http://dx.doi.org/10.1016/j.bbagen.2014.09.016
0304-4165/© 2014 Published by Elsevier B.V.

Please cite this article as: P.S. Hudson, et al., Efficiently computing pathway free energies: New approaches based on chain-of-replica and Non-Boltzmann Bennett reweighting schemes, Biochim. Biophys. Acta (2014), http://dx.doi.org/10.1016/j.bbagen.2014.09.016

simulations and/or normal mode analysis to generate QM/MM potentials of mean force (PMFs) [54–68].

Although copious work has been done in this area there are still two fundamental, if not contradictory, requirements for both classes of approaches. First is the need for adequate sampling of all relevant conformational degrees of freedom. In this, the former class of methods (i.e., using MM potentials) has a significant advantage over QM based approaches. This is not based simply on efficiency, but also on the fact that numerous enhanced sampling methods have been developed specifically for classical simulations [69–77]. The other requirement is the accurate description of intra- and intermolecular interactions; here, the use of a QM-based Hamiltonian is desirable if not essential, in particular if chemical reactions are to be described. While satisfying these two requirements is extremely difficult, even for MM based approaches, they create a quandary when QM or QM/MM is used [64].

To circumvent this problem, the so-called *indirect* scheme [78–83] for computing QM/MM free energies between two states was developed by Gao and co-workers and Warshel and co-workers and subsequently generalized and extended by numerous others [56,57,59,66,84–86]. One particular extension (i.e., using "fixed" QM regions to avoid expensive QM/MM simulations) was proposed by Yang and co-workers as well as Ryde and co-workers. This technique has been used on numerous occasions to approximate QM/MM free energies as a function of reaction coordinate. The approach is carried out as follows: (1) The free energy difference between the MM and QM/MM descriptions of the initial state A is calculated by using a single-step FEP. (2) The atoms belonging to the "QM region" are assigned electrostatic potential derived point charges (ESP) and held fixed during the classical $A_{MM} \rightarrow B_{MM}$ step, followed by (3) a single step FEP for $B_{MM} \rightarrow B_{QM/MM}$. In its simplest form the FEP steps are based on MM simulations of *A* and *B*, i.e., no MD with a QM/MM Hamiltonian is carried out.

This approach has been successfully applied to numerous questions surrounding enzyme mechanisms [57,59,84,85,87]; however, recently serious doubts have been raised about both the accuracy and precision of results obtained from these *indirect* approaches if no QM/MM simulations are actually performed [86]. In that study Heimdal and Ryde proposed two new strategies to more accurately account for MM → QM/MM connections: (1) heavily optimizing a classical force field to improve overlap with the QM region during MM simulations and (2) performing the underlying sampling using semi-empirical QM/MM (SQM/MM) rather than MM. For both cases they determined that the "QM region" degrees of freedom were too dissimilar from either the optimized force field or the SQM methods employed.

One potential source of error in these approaches is the use of FEP in the MM → QM connection steps. The poor convergence of FEP is well known for the calculation of free energy differences between two states and has been well documented [4,9,88], with superior methods gaining popularity in such applications, in particular BAR [8] and generalizations thereof [89]. Recently, we have developed a robust new method for connecting MM and QM descriptions of a system, combining "unusual" biasing potentials, reweighting, and BAR [90]. This new technique, to which we refer to as QM-Non-Boltzmann Bennett (QM-NBB), was successfully applied to a series of solvation free energy calculations and showed vast improvements compared to FEP for connecting MM → QM levels of theory [90,91].

In the current paper we will focus on both aspects critical to computing QM/MM reaction free energies via the *indirect scheme*. Initially, we will introduce and validate a novel pathway sampling method; referred to as Off-path simulations (OPS). This new method combines the chain-of-states replica path framework (RPATh, see below for a review of this methodology) with an umbrella like approach for determining the PMF of a particular reaction in root mean square deviation (RMSD) phase space. Further, we will describe the implementation of a new replica framework (i.e., distributed replica or REPDstr) and validate this using OPS. In addition, we will detail a newly introduced classical, QM, and QM/MM eigenvector following procedure implemented in CHARMM and Q-Chem [92–94,42]. Finally, we present results that demonstrate the potential for greatly improved efficiency (as compared to standard FEP) for obtaining QM/MM reaction PMFs using a combination of the *indirect scheme* and our new QM-NBB method.

2. Methods

The replica path (RPATh) method has been developed as an extension to the self-penalty walk method of Elber and coworkers [15–17]. The RPATh method is built on top of topology replication functionality (i.e., the REPLica module of CHARMM). This allows N number of replicas to be created with the user having complete control as to the subset of atoms to include in these newly replicated structural elements. Energy penalty functions (i.e., restraints) are used to define the pathway, ensure equidistance of replicas, and prevent pathway reversal [39,41,42]. The definition of the pathway itself is one of the more attractive features of the RPATh method as smaller portions of the system can be selected and used to define the pathway of interest. Additionally, the entire or subsections of the pathway can subsequently be treated quantum mechanically (e.g., SCC-DFTB, ab initio, DFT). For example, in previous work a 6 Å region surrounding the substrate has been chosen as the pathway with the substrate itself being treated quantum mechanically [39,41,42]. As noted previously the RPATh method is comprised of two major restraints. The first of these is a distance restraint that consists of an energetic penalty which is applied when the distances between adjacent points deviate away from the average of all pathway distances. This restraint helps to maintain a smooth evenly spaced pathway and takes the following form:

$$\mathrm{E}_{\mathrm{rms}} = \sum_{i=1}^{N} \frac{1}{2} \mathrm{K}_{\mathrm{rms}} (\mathrm{r}_i - \bar{\mathrm{r}})^2 \tag{1}$$

$$\mathrm{r}_i = \mathrm{RMS}(\mathrm{i}, \mathrm{i}+1) \tag{2}$$

where RMSd is

$$\mathrm{RMSd}(\mathrm{r_i}, \mathrm{r_{ref}}) = \sqrt{\frac{\sum_{i=1}^{N} (\mathrm{r_i}, \mathrm{r_{ref}}) \mathrm{m_i w_i}}{\sum_{i=1}^{N} \mathrm{m_i w_i}}} \tag{3}$$

$$\bar{\mathrm{r}} = \sum_{i=1}^{N} \frac{\mathrm{r}_i}{N} \tag{4}$$

where $\bar{\mathrm{r}}$ is the average step length over the entire path and N is the number of points along the pathway. $\mathrm{K_{rms}}$ is a user defined parameter with suggested values that result in pathway-point variation being no larger than 10–15%.

The second major restraint term controls the angle between adjacent and next adjacent pathway points $(i, i+1, i+2)$ with θ defined as the deviation from linearity.

$$\begin{aligned} \mathrm{E}_{\mathrm{angle}} = \sum_{i=1}^{N} \mathrm{E}^{i}_{\mathrm{angle}} &= \sum_{i=1}^{N} \frac{1}{2} \mathrm{K}_{\mathrm{angle}} (\mathrm{COSMAX} - \cos\,(\theta)_i)^2; \quad \mathrm{COSMAX} > \cos\,(\theta)_i \\ \mathrm{E}^{i}_{\mathrm{angle}} &= 0; \quad \mathrm{COSMAX} < \cos\,(\theta)_i \end{aligned} \tag{5}$$

This term is controlled by two adjustable parameters; COSMAX and $\mathrm{K_{angle}}$. The recommended value of COSMAX is: 0.8–1.0. A COSMAX = 1.0 corresponds to a linear pathway and a value of 0.8 allows each angle to bend up to ~30° before a penalty is applied. The $\mathrm{K_{angle}}$ parameter should be set depending on the desired smoothness and flexibility of the pathway. For example, a value in the range 100–500 kcal $\mathrm{mol^{-1}}$ is a reasonable choice.

Please cite this article as: P.S. Hudson, et al., Efficiently computing pathway free energies: New approaches based on chain-of-replica and Non-Boltzmann Bennett reweighting schemes, Biochim. Biophys. Acta (2014), http://dx.doi.org/10.1016/j.bbagen.2014.09.016