



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 17 September 2014

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.

© 2014 Published by Elsevier B.V.

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.

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 $MM \rightarrow QM/MM$ 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 $MM \rightarrow QM$ 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 $MM \rightarrow QM$ 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:

$$E_{\text{rms}} = \sum_{i=1}^N \frac{1}{2} K_{\text{rms}} (r_i - \bar{r})^2 \quad (1)$$

$$r_i = \text{RMS}(i, i+1) \quad (2)$$

where RMSd is

$$\text{RMSd}(r_i, r_{\text{ref}}) = \sqrt{\frac{\sum_{i=1}^N (r_i - r_{\text{ref}})^2 m_i w_i}{\sum_{i=1}^N m_i w_i}} \quad (3)$$

$$\bar{r} = \sum_{i=1}^N \frac{r_i}{N} \quad (4)$$

where \bar{r} is the average step length over the entire path and N is the number of points along the pathway. 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.

$$E_{\text{angle}} = \sum_{i=1}^N E_{\text{angle}}^i = \sum_{i=1}^N \frac{1}{2} K_{\text{angle}} (\text{COSMAX} - \cos(\theta_i))^2; \quad \text{COSMAX} > \cos(\theta_i) \quad (5)$$

$$E_{\text{angle}}^i = 0; \quad \text{COSMAX} < \cos(\theta_i)$$

This term is controlled by two adjustable parameters; COSMAX and 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 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 mol^{−1} is a reasonable choice.

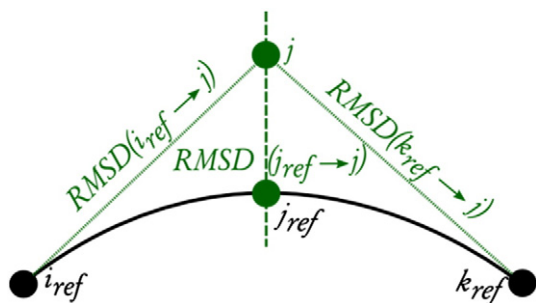


Fig. 1. Illustration of the restraints used to define an off-path simulation.

2.1. Classical, quantum, and hybrid QM/MM transition state searching

To complement the functionality of the replica path method CHARMM's Newton–Raphson minimizer has been extended [92,93]. Eigenvector following routines have been added in a flexible manner that allows efficient minimization of both harmonic and anharmonic degrees of freedom. Specifically, if a mode sufficiently satisfies the harmonic approximation then standard Newton–Raphson minimization will occur employing the energy (E), gradient of the energy (∇E) and Hessian of the energy ($\nabla^2 E$). However, if a mode is not sufficiently harmonic then a line search will be executed along the specified eigenvector(s).² During the line search procedure the aforementioned energy terms are employed along with two additional gradient calculations (∇E). ∇E in both the positive and negative directions of the eigenvector(s) are computed and combined with the three previous data points and fit to a third order polynomial ($f(x) = ax^3 + bx^2 + cx + d$) which is subsequently solved via a linear least squares fit procedure.

The default Newton–Raphson procedure, defined above, has been modified to change the direction of eigenvector(s) searching. Therefore, rather than minimizing all modes, which is typical, the new routines allow specific modes to be maximized thereby searching for a non-minimum stationary point. The order of the stationary point can be specified with transition states defined to have a stationary point of one. This new feature supports classical, QM, and QM/MM Hamiltonians [58,68].

2.2. Off-path simulation method

The off-path simulation (OPS) method is an extension of the replica path method that employs umbrella-like restraining potentials to compute potentials of mean force (PMFs). The basic premise of the OPS method is biased sampling of a pathway by allowing a simulation to propagate on planes orthogonal to a previously defined pathway. Using this information, data are accumulated and a biased PMF is computed at the conclusion of the simulation. As a usability feature, the statistical information can be saved and the simulation can be restarted if the PMF is determined to be not converged.

A prerequisite of this method is a previously computed reference path. It is not necessarily suggested that the reference path be a minimum energy pathway (MEP) as MEPs are likely to have a significant amount of curvature. This curvature can cause numerical inaccuracies in free energy procedures, therefore, a curvature correction has been built into the OPS method. For example, exploration of orthogonal path space can often lead to sampling regions of phase space that differ from the pathway of interest (i.e., the reaction coordinate). These differences can then result in either under- or overestimation of the free energy for moving along the chosen reaction coordinate. This is typically

referred to as “corner cutting”. Following this logic the nudged elastic band method is not recommended for generating reference paths. Rather, the RPATH method, which has extensive controls on the smoothness of the path is a better choice. Once a smooth path is obtained the procedure begins by assigning this as the reference path and creating a “simulation path”. In CHARMM this is accomplished by copying the path into the COMPArison coordinate set (i.e., COOR COPY COMP). At this point the RPATH command can be issued adding the keyword “OPTI”.

The OPS method consists of two major restraints that define the plane the points are allowed to move on during the simulation (Eq. (6)) and an additional penalty that specifies how far points are allowed to move away from the reference path.

$$E_{\text{rms}} = \sum_{i=1}^{N-2} \frac{1}{2} K_{\text{rms}} (r_{i_{\text{ref}}j} - r_{j_{\text{kref}}})^2 \quad (6)$$

where $i = j - 1 = k - 2$ and $r_{i_{\text{ref}}j}$ and $r_{j_{\text{kref}}}$ refer to the root mean square best-fit distances between point j of the current path and points i_{ref} and k_{ref} of the reference path. This effectively defines a plane bisecting points i and k of the reference path; allowing point j to move freely along this plane. The biasing term takes a similar form:

$$E_{\text{max}} = \sum_{i=1}^{N-2} \frac{1}{2} K_{\text{rms}} (r_{jj_{\text{ref}}} - \text{RMAX})^2 \quad (7)$$

where $r_{jj_{\text{ref}}}$ is the root mean square best-fit distance between points on the reference path and simulation path and RMAX is the user adjustable parameter (in Å). An illustration of the pathway restraints is presented in Fig. 1.

Once the pathway is defined and OPS parameters are set, the simulation begins collecting statistics that will be used to determine the PMF (i.e., work). This is done using the following approximation.

$$\text{Work} = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle d\lambda \approx \sum_{i=1}^{N-1} \langle \nabla \tilde{V}_i \rangle \Delta l_i^{\text{ref}} \quad (8)$$

where $\nabla \tilde{V}_i$ is

$$\nabla \tilde{V}_i = \nabla V_i \cdot \mathbf{R}_{\text{rot}} \quad (9)$$

the force on each simulation projected into the frame of the reference path and Δl_i^{ref} is

$$\Delta l_i^{\text{ref}} = \frac{1}{2} (r_{ij} + r_{jk}) * \left[\frac{r_{i,j} + r_{j,k}}{r_{i,k}} \right] \quad (10)$$

with all RMSd values in Eq. (10) referring to reference path distances, thus implicitly building in a curvature correction based on the fundamental triangle inequality theorem.

As with the replica path method, OPS is designed to work with MM, QM, or QM/MM potentials. This allows OPS to function in a highly flexible parallel/parallel manner. That is each QM/MM point can be run a different processor or group of processors with the constraint that the total number of processors be an integer multiple of the total number of pathway points (i.e., # of processors = # of replicas \times N; N = 1, 2, 4, 6, 8...). Currently, there are four quantum mechanical packages that support the QM and QM/MM off-path simulations method: Q-Chem [42,94], SCC-DFTB,³ [95,96,55], GAMESS-US [97,98], and GAMESS-UK [99,100].

² The criteria for performing this line search are determined from numerical estimates of each eigenvector's shape. Forward and backward perturbations are made along the eigenvector and energies and forces calculated. If the energies in both directions increase or the gradient of the perturbations are negative the mode is identified as “harmonic”.

³ As SCC-DFTB is not a parallel QM package the parallel/parallel functionality is not supported.

2.3. Distributed replica framework: REPDstr

Currently, the classical version of the off-path simulation method is implemented to run each point of the pathway sequentially, either in serial or parallel. In contrast, the QM or QM/MM version of OPS is fully parallelized; i.e., each point of the pathway executes on a different processor or group of processors. This difference in behavior is facilitated by the difference in time lengths of QM versus MM calculations; i.e., time of QM calculation \gg parallel communication of energy and forces.

To improve the performance of the classical OPS methodology, the fully parallel REPDstr module was developed and implemented in CHARMM. This module is a facility to group a specified number of independent CHARMM executions. Each group can either read the same input script or each can execute its own input script. In this way, one can run many different input scripts from a single parallel run. Similarly, output and all other files written to disk during the simulation can optionally be divided based on the replica numbers.

A major benefit of grouping parallel processes is that everything within the group works the same way as a standard CHARMM parallel job. There is no code change needed for intragroup communication or special splitting of the workload. The only addition of such division is that for some methods one needs to communicate data amongst specific groups. Most of the time this is just data swap between the neighboring groups and almost never requires global communication, i.e., that every processor in all groups has the same data. Therefore, the communication is significantly reduced, and parallel efficiency is almost perfect.

For CHARMM developers, it is easy to implement new methods which can take advantage of the two level parallel framework. Specifically for OPS we needed to add the communication of comparison coordinate sets from two neighboring replicas. This is performed only at the start of the simulation, since the reference path does not change during the course of simulation. The simulations of individual replicas are then independent of each other and parallel efficiency is close to 100%, unless replicas have significantly different numbers of atoms and load unbalance comes into play. Additionally, final communication was added at the end of the off-path simulation job to compute and print the path summary.

Recently we implemented a variety of methods using the REPDstr module, most of which are derivatives of standard replica exchange methods, such as Self Guided Langevin Dynamics Replica Exchange [77] or calculations of absolute free energy [101]. Additionally, the REPDstr module greatly facilitates improving parallel performance in CHARMM without having to make code modifications. For example, it is now possible to parallelize standard PERT (free energy perturbation) calculations in CHARMM [92]. All the user needs to do is specify the REPDstr command in a PERT input script and use IF statements to select which group of processors is attached to which λ state/simulation.

3. Results and discussion

In this paper, we are reporting the implementation and validation of a novel method for calculating PMFs within a chain-of-states (i.e., replica) framework. Further, this framework is compatible with standard restraint pathway determination procedures (i.e., reaction coordinate driving) and herein is combined with our new QM-NBB procedure and shown to greatly enhance convergence in *indirect* QM/MM pathway free energy simulations. To validate the implementation of the new chain-of-states framework, we calculate the simple conformational pathway of butane's torsional rotation in the gas phase. As part of this test case, both the original REPLica and new REPDstr frameworks are employed. Following the initial butane test case, the one-dimensional PMF is calculated for the rotation about the ϕ/ψ glycosidic dihedral angles of maltose in both gas phase and explicit solvent; results are compared to previously published data from Kuttel and Naidoo [102]. Finally, the model system butane is again used to introduce the

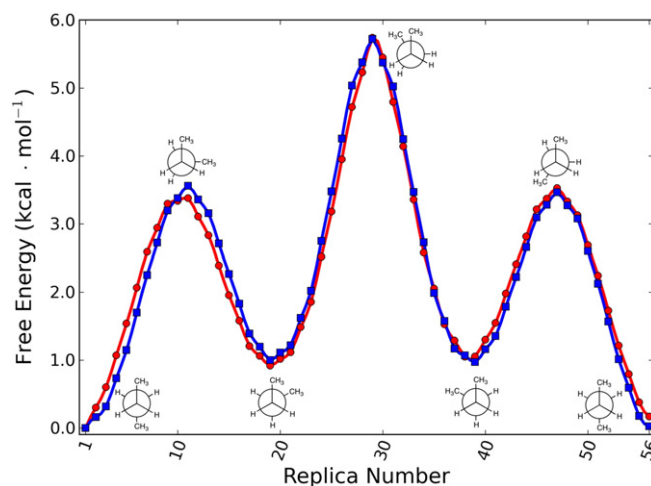


Fig. 2. Comparison of the OPS-REPL (red) and OPS-REPD (blue) methods. Each simulation was run for 10 ps at 300 K with an RMAX of 0.275.

application of QM-NBB to efficiently compute QM/MM reaction pathway free energies.

3.1. Torsional potential of butane

The torsional rotation of butane is an iconic illustration of steric effects on the conformational positions of a molecule. This potential is defined as a total of 4 unique conformations: anti, gauche (2,4), eclipsed-gauche (1,5), and eclipsed (3); see Fig. 2. The surface of the torsional potential is dominated by the van der Waals repulsion arising from varying levels of substituent interactions on the chain. The anti and gauche conformations correlate to minima along the potential while the eclipsed-gauche and eclipsed positions correspond to maxima. Standard harmonic limit free energies (using the vibrational analysis, VIBRan, module) were obtained using the CHARMM force field and are used as the “gold standard” with which to compare OPS results. A comparison of replica frameworks (i.e., OPS-REPL versus OPS-REPDstr) was performed to validate implementation of REPDstr.

All OPS simulations used force constants of $10,000 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for both KRMS and KMAX, a 1 fs time step, and a temperature of 300 K. The reaction was discretized into 58 points (i.e., 58 replicas) while the total simulation time and RMAX values were varied during testing of the OPS-REPL method. The range of simulation times was: 1 ps, 5 ps, 10 ps, 25 ps, 50 ps, 100 ps, 250 ps, and 1000 ps. The RMAX value varied from 0.050 to 0.450 in increments of 0.025 for each time length, thus producing a wide array of simulation conditions used in determining the overall success of the new methodology. A simulation lasting 10 ps was utilized for the OPS-REPL method to highlight the possible reduction in computational time. Table 1 contains a comparison of the values calculated for each of the minima and maxima along the energy surface that correspond to the conformations illustrated in Fig. 2.

This paper is meant to outline the implementation of the OPS method as well a new replica framework (i.e., REPDstr). To validate this implementation, the torsional potential of butane was also calculated with OPS-REPD. Results of these simulations were again compared to harmonic limit free energies and those generated via OPS-REPL (Table 1). The torsional energy surface was reproduced to within quantitative accuracy and differed only in the replica framework employed with OPS (Fig. 2). Thus, we conclude that both the OPS-REPL and OPS-REPD procedures function efficiently and accurately.

3.2. Double torsional potential of the disaccharide maltose

The majority of systems of interest are larger than a simple butane molecule, thus it is necessary for us to demonstrate that OPS can

Table 1

Computed potentials of mean force using all three methods at 300 K: Normal Mode Analysis (NMA), OPS–REPL, and OPS–REPD. Each of the OPS simulations was run for 10 ps with an RMAX of 0.275. All values are given in kcal mol^{−1}.

Methods	1	2	3	4	5
NMA	3.5	1.0	5.7	1.0	3.5
OPS–REPL	3.5	0.9	5.7	1.1	3.5
OPS–REPD	3.4	1.0	5.7	1.0	3.5

produce qualitative and quantitative results of interest for a larger system. To this end, the PMF for the rotation about the $\alpha(1 \rightarrow 4)$ -glycosidic linkage (Fig. 3) in a molecule of maltose was computed in vacuum and explicitly solvated in TIP3P waters.

In 2005, Kuttel and Naidoo [102] used an iterative adaptive umbrella sampling method [103–107] to compute the PMF along the ϕ, ψ dihedral angles of the $\alpha(1 \rightarrow 4)$ -glycosidic linkage of maltose in vacuum and water. Approximately 500 ns of simulation time was required to produce each of the final PMF surfaces with individual data collections in each simulation being 1 ns–10 ns. Populations of ϕ values were calculated for each ψ , and these relative populations were then used to calculate a one-dimensional PMF. Herein, we have used these results as a comparison set for the PMF values determined via OPS (Table 2). The maltose was built using CHARMM topology and parameters for the CSFF type force field [108], corresponding to the setup of Naidoo and coworkers. It is crucial to employ the same force fields in these simulations to allow for comparison of the computed PMFs.

As mentioned in the Methods section, one shortcoming of the REPLICA framework is the inability to efficiently replicate and parallelize simulations that employ explicit solvent, therefore the REPDstr infrastructure was developed and employed herein. The force constants (i.e., KRMS, KMAX) for all maltose simulations were set to 10,000 kcal mol^{−1} Å^{−2} and the pathway was discretized into 78 points (i.e., replicas).

In examining the in vacuo and solvated PMFs of maltose, there are several features highlighted by Naidoo and coworkers. One such feature is the lowering of the PMF upon solvation. The general decrease and broadening of the energy wells is thought to arise from the reduction in enthalpic advantage from intramolecular hydrogen bonds, particularly that of the inter-monomer hydrogen bonds. This also increases the likelihood of forming the X, and Y conformations (Fig. 4) that cannot form inter-monomer hydrogen bonds due to the favorable interactions with the solvent. Another feature we have examined is how the solvation of maltose has affected the transition barriers between different states in the molecule, specifically, the transition barrier between the Anti, X and Y conformations. The results from Naidoo and coworkers show no significant change in relative transition energy from Anti \rightarrow X barrier but show a marked change in the Anti \rightarrow Y transition. Further, they determined that the difference between intramolecular and intermolecular hydrogen bonding from solvent is in the range of 1 kcal mol^{−1}. This difference was determined using QM and classical MD simulations and gives us an idea of the significance that each hydrogen bond plays in determining the potential of mean force.

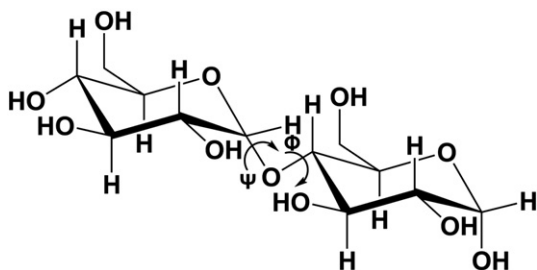


Fig. 3. Illustration of maltose molecule (gluc $\alpha(1 \rightarrow 4)$ gluc). The arrows in the image refer to the angles of interest in the calculation of the PMF.

Table 2

Table of potential of mean force from Naidoo and Kuttel and Naidoo [102] (top) and from the OPS–REPD simulations (bottom). The data represents the transition barriers between the SYN, Anti, X and Y conformations. All values are given in kcal mol^{−1}.

Transition	Vacuum	Solvated
<i>Naidoo and coworkers</i>		
$\Delta G(\text{syn} \rightarrow \text{anti})$	4	3
$\Delta G(\text{syn} \rightarrow \text{X})$	10	8
$\Delta G(\text{syn} \rightarrow \text{Y})$	15	12
$\Delta G(\text{anti} \rightarrow \text{X})$	6	6
$\Delta G(\text{anti} \rightarrow \text{Y})$	11	9
<i>OPS</i>		
$\Delta G(\text{syn} \rightarrow \text{anti})$	4.1	2.5
$\Delta G(\text{syn} \rightarrow \text{X})$	9.7	8.3
$\Delta G(\text{syn} \rightarrow \text{Y})$	15.1	11.7
$\Delta G(\text{anti} \rightarrow \text{X})$	5.6	5.8
$\Delta G(\text{anti} \rightarrow \text{Y})$	11.0	9.2

Results from the OPS–REPD PMF calculations are presented in Fig. 5. The values corresponding to the transition barriers between each of the major conformations mentioned previously are listed in Table 2. Previously published data is also presented in that table. Finally, it is clear from Table 2 that PMFs generated using OPS–REPD, in vacuo and solvent, agree well with previously published exhaustive sampling results. Additionally, there is a considerable reduction in the total simulation time required to obtain these results; however, additional testing needs to be performed to establish optimal simulation conditions (i.e., user adjustable parameters, time length, time step). Overall, the off-path simulation method appears to be a promising step forward for efficiently computing PMFs of conformational changes and ultimately reactions via QM/MM.

3.3. Accurate PMFs using a novel reweighting scheme: QM–NBB

As highlighted in the Introduction section, computing accurate free energies has two fundamental requirements: adequate sampling and accurate description of inter- and intramolecular interactions. To help overcome these, often, conflicting requirements, we recently developed a new reweighting scheme that allows more efficient connecting of low level simulations (e.g., MM, semi-empirical QM) to high level results (e.g., Hartree–Fock, DFT) [90]. The efficiency and accuracy of this new approach for calculating solvation free energy have been further demonstrated in the SAMPL4 competition [91]. Further, the potential that QM–NBB has for improving the computation of free energies as a function of reaction is vast.

As stated previously, FEP has traditionally been the method of choice for connecting MM \rightarrow QM while using the indirect QM/MM free energy scheme. However, if the potential energy surfaces of the MM force field are not representative of the ensemble that would exist at the QM level of theory then overlap between them will be poor and lead to convergence problems. This is clearly an issue for all methods that seek to connect two disparate ensembles; however, the QM–NBB procedure has been shown to be far more robust in this respect as compared to FEP [90]. Further, Pohorille et al. recently reported a detailed analysis of why two-sided methods, such as BAR and, thus, also QM–NBB, to compute free energy differences converge more rapidly than one-sided approaches, such as FEP. Therefore, we again take the simple butane torsion as our initial validation of coupling QM–NBB to QM/MM PMF generation.

The conformational changes in butane (i.e., torsion) were discretized into 10° increments (i.e., 37 points along the path).⁴ Gas phase MM and semi-empirical QM (i.e., SCC-DFTB with the 3OB parameters) [55,95,96,

⁴ Although the REPLICA framework was not used for the initial data collection, we subsequently confirmed that combining simple harmonic restraints (i.e., as demonstrated in Ref. [41]) with either the original RPATH method or OPS will yield equivalent results.

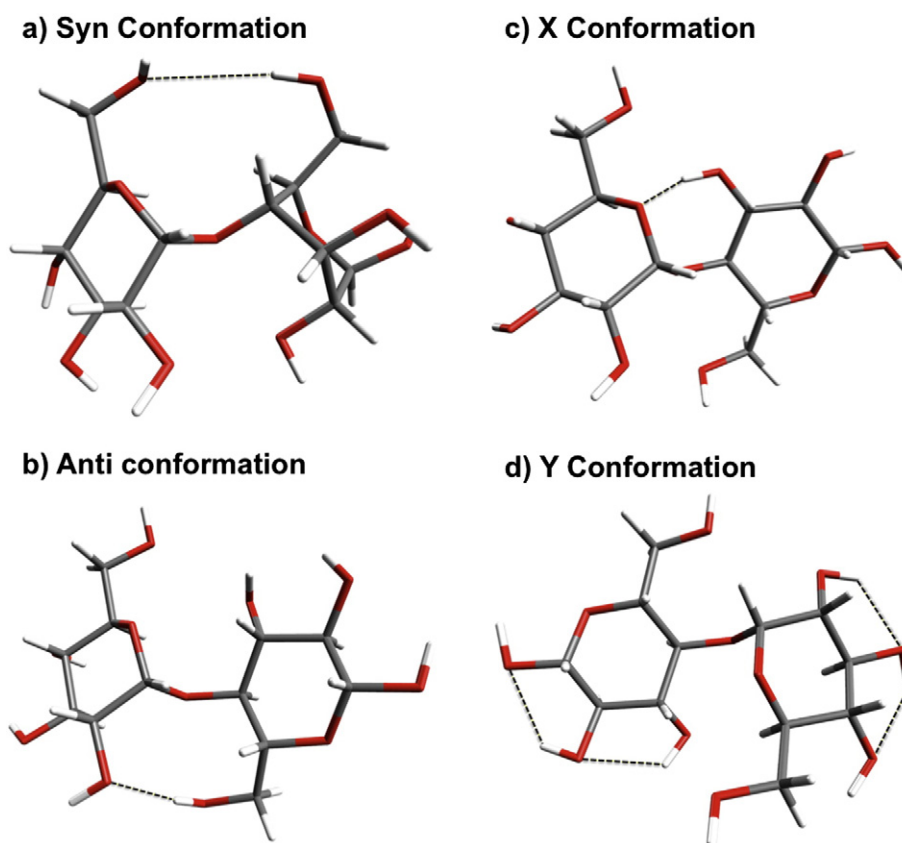


Fig. 4. Visual representations of the 4 most prominent conformations of the maltose molecule. These images display the 4 conformations for which Kuttel and Naidoo [102] based their method evaluation on.

[109,110] simulations were carried out using Langevin dynamics with a time step of 1 fs and time length of 2 ns. To ensure that simulations stayed close to their respective dihedral angle, harmonic restraints were employed with a force constant of 50 kcal mol⁻¹ rad⁻². Coordinates were saved every 10 steps and reevaluated at the Hartree–Fock (HF) level of theory using the 6-31G* basis set. This resulted in a total of 7.4 million (200 K per simulation) QM calculations to compute the complete PMF. Further, all frames in the MM trajectory were reevaluated using SCC-DFTB and vice versa.

Next, we computed the PMFs at the MM and SCC-DFTB levels, i.e., the levels of theory at which the simulations were carried out. We employed BAR for these calculations, and states 0 and 1 in the working equations (see below) refer to neighboring values of the restraining potentials.

$$\Delta G(0 \rightarrow 1) = k_B T \left(\ln \frac{\langle f(U_0 - U_1 + C) \rangle_1}{\langle f(U_1 - U_0 - C) \rangle_0} \right) + C \quad (11)$$

where $f(x)$ denotes the Fermi function $f(x) = \left(1 + \exp \left(\frac{x}{k_B T} \right) \right)^{-1}$ and

$$C = k_B T \ln \frac{Q_0 N_1}{Q_1 N_0}. \quad (12)$$

Here Q_0 and Q_1 are the canonical partition functions of the two states, and N_0 and N_1 are the number of data points used to compute the ensemble averages for states 0 and 1, respectively. Eq. (11) is iterated until the condition

$$\langle f(U_0 - U_1 + C) \rangle_1 = \langle f(U_1 - U_0 - C) \rangle_0 \quad (13)$$

is fulfilled. With C determined in this manner, one immediately obtains

$$\Delta G(0 \rightarrow 1) = -k_B T \ln \frac{N_1}{N_0} + C. \quad (14)$$

The individual free energy differences obtained in this manner were combined to yield the PMF as a function of the central torsional angle. While the use of BAR to compute a PMF is unusual, the resulting PMFs

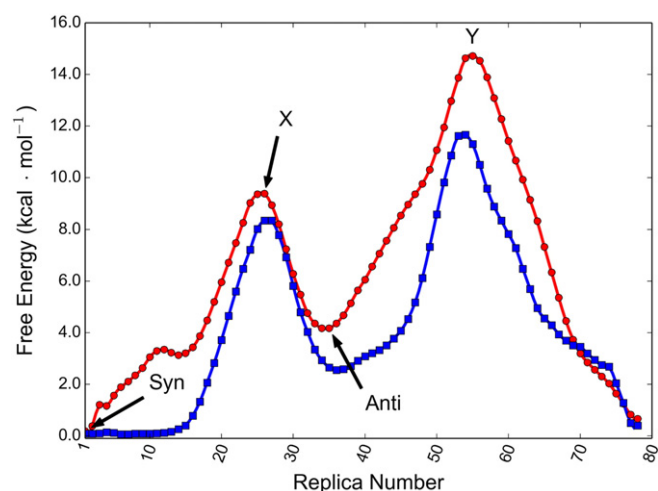


Fig. 5. Illustration of the simulations performed with maltose using OPS-REPD methodology in vacuum and explicit solvent. An RMAX value of 0.225 was run for 40 ps in vacuum (red curve) for the above results. The solvated system (blue curve) was run for 1000 ps with an RMAX of 0.225. This data represents a quantitative match to the previous published data used for comparison [102].

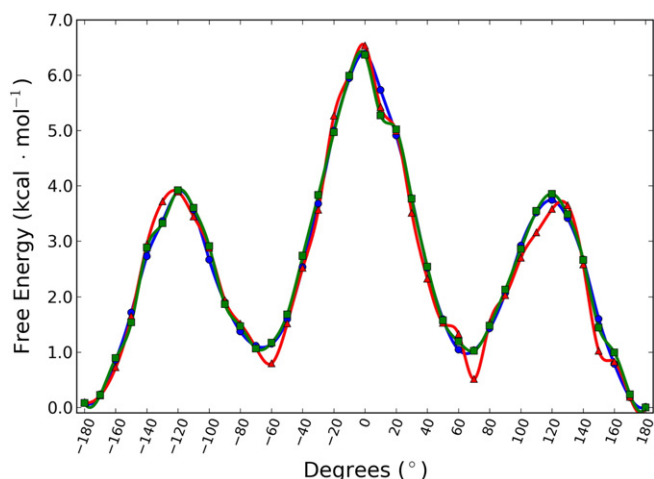


Fig. 6. Butane PMFs computed via standard FEP (i.e., with $\Delta G(\text{MM} \rightarrow \text{QM})$) using all 200 K data points (blue, 2 ns simulations) per simulation, 50 K data points (green, 500 ps simulations), and 10 K data points (red, 100 ps simulations).

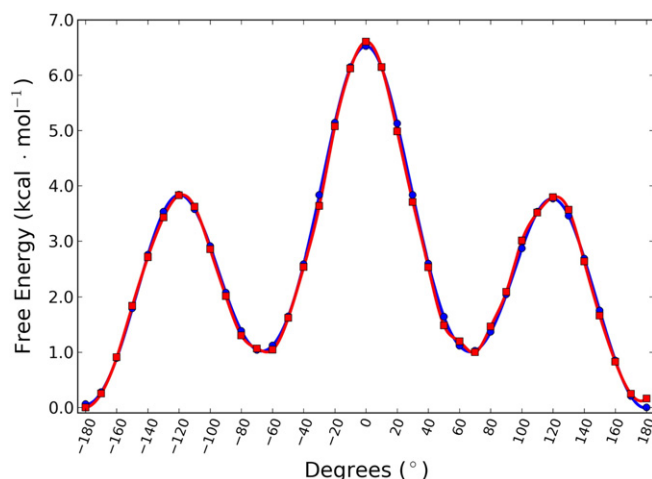


Fig. 7. Butane PMFs computed via standard FEP (i.e., with $\Delta G(\text{SCC-DFTB} \rightarrow \text{QM})$) using all 200 K data points (blue) and with QM-NBB (red) using only 10 K data points from MM simulations (i.e., 100 ps simulations) and 1 K data points from SCC-DFTB simulations (i.e., 10 ps simulations).

were highly precise, i.e., differences between the free energy at $\varphi = -180^\circ$ and $\varphi = +180^\circ$, which were not restrained to be identical, were $\leq 0.02 \text{ kcal mol}^{-1}$ in all cases. Further, our emphasis is on connecting a low level of theory (e.g., MM, SCC-DFTB) to a high level (e.g., QM). This was initially done using FEP as follows:

$$\Delta G(0 \rightarrow 1) = -k_B T \ln \langle \exp [-(U_1 - U_0)/k_B T] \rangle_0 \quad (15)$$

where U_0 is MM or SCC-DFTB and U_1 is QM. Care was taken (by employing suitable offsets) to avoid numerical overflow when calculating the exponential functions in Eq. (15).

Results from these calculations serve as a powerful illustration as to what can happen when there is poor overlap between ensembles you are trying to connect via FEP. Results showing the butane PMF generated from using all 200 K points per simulation are plotted in Fig. 6 along with a subset of points that were generated during the first 100 ps and 500 ps (i.e., time scales that are currently accessible if QM simulations were required). Just by visual inspection, several problems with these PMFs are evident. The first concerns the PMF generated from the full 200 K simulation points. There is a clear asymmetry about the global maximum; the point at 10° does not correspond to the one at -10° . This problem is only exacerbated when more limited data sets are considered (i.e., 100 ps, 500 ps). Further asymmetry can be observed when comparing the non-global maxima ($\pm 120^\circ$) and minima ($\pm 70^\circ$). For later reference we introduce two simplistic measures to quantify such asymmetries: the maximum absolute difference between $\text{PMF}(-\varphi)$ and $\text{PMF}(+\varphi)$, as well as the root mean squared deviation over all points of the PMF that should be symmetric. For the 500 ps data, these are $0.72 \text{ kcal mol}^{-1}$ and $0.048 \text{ kcal mol}^{-1}$, respectively; for the 100 ps data, the numbers are $0.62 \text{ kcal mol}^{-1}$ and $0.075 \text{ kcal mol}^{-1}$. The large discrepancy between the maximal absolute deviation and the root-mean-square criterion results from the fact that the asymmetries affect mostly individual points in the PMF; as is clearly discernible also in the figure.

In contrast, the PMF generated using all 7.4 million saved frames from the SCC-DFTB simulations (i.e., $\Delta G(\text{SCC-DFTB} \rightarrow \text{QM})$ obtained using standard FEP) displayed none of the asymmetry observed in the $\text{MM} \rightarrow \text{QM}$ PMF (see Fig. 7). Again, this behavior highlights the importance of using a level of theory for conformational searching (e.g., fixed QM regions, MM, SCC-DFTB) that adequately describes the ensemble that would be generated if searching was done at the high level of theory (e.g., QM). This is especially problematic when using FEP to connect levels of theory.

To help mitigate the limitations of the *indirect* scheme, we have recently developed the QM-NBB approach. This new method shows significantly better convergence when computing the free energy between two disparate levels of theory as compared to FEP [90,91].

$$\Delta G(0 \rightarrow 1) = k_B T \ln \left(\frac{\langle f(U_0 - U_1 + C) \exp(\beta V_1^b) \rangle_{1,b} \langle \exp(\beta V_0^b) \rangle_{0,b}}{\langle f(U_1 - U_0 - C) \exp(\beta V_0^b) \rangle_{0,b} \langle \exp(\beta V_1^b) \rangle_{1,b}} \right) + C. \quad (16)$$

The notation follows that of Eq. (11); however, the additional subscript b indicates that the ensemble averages were obtained in the presence of an “unusual” biasing potential (Eq. (17)).

$$V^b = U^{\text{low}} - U^{\text{high}} \quad (17)$$

To use Eq. (16) it is necessary to evaluate three quantities for each frame of the trajectories: U_0 , U_1 and V_0^b for state 0 and U_0 , U_1 and V_1^b for state 1. Thereby, U_0 and U_1 are evaluated at the respective high level of theory; the fact that sampling was carried out at a low level is taken care of by Eq. (17). In regular BAR, each forward/backward energy difference contributes equally to the sums $\langle \rangle$ in the numerator and denominator of Eq. (11). In NBB, on the other hand, the summands have different weights $\exp(\beta V^b)$. In case of poor overlap the sums in Eq. (16) become dominated by just a few terms having the most positive V^b . One very attractive feature of both standard BAR and QM-NBB is the fact that no requirement exists that $N_1 = N_0$ (Eq. (14)); i.e., the simulations at different levels of theory can be of different lengths. Here we apply the QM-NBB methodology taking particular advantage of this feature and using the following definitions: $\text{MM} = \text{state 0}$; $\text{SCC-DFTB} = \text{state 1}$; $V_0^b = 0$; and $V_1^b = U^{\text{SCC-DFTB}} - U^{\text{QM}}$. Further, U_0 in Eq. (16) is just the MM energy, whereas U_1 is the QM (HF/6-31G*) energy for coordinates sampled at the MM and SCC-DFTB levels, respectively.

In practice, this entails using MM and QM potential energies obtained from saved trajectories of MM and SCC-DFTB simulations; $\langle U_0, U_1 \rangle_0$ and $\langle U_0, U_1 \rangle_1$ in Eq. (16). The forward and backward free energy differences between MM and QM are obtained from the real MM simulation and a “QM” simulation in the presence of the biasing potential $V_1^b = U^{\text{SCC-DFTB}} - U^{\text{QM}}$. In general, if an “intermediate” level of theory is used that more closely overlaps with QM then significant improvement in convergence should be realized.

PMFs obtained using QM-NBB and the aforementioned energies are shown in Fig. 7. Here, the reference PMF is obtained using all 200 K frames from each simulation in combination with SCC-DFTB \rightarrow QM free energies determined using standard FEP. Further, the QM-NBB PMF is computed using data obtained from a 100 ps MM simulation and a 10 ps SCC-DFTB simulation (i.e., 10 K–1 K in Fig. 7). This is roughly equivalent in computational effort to the FEP results generated using only 10 K MM data points; however, the result is remarkably improved. For example, the symmetry of the PMF is almost perfectly fulfilled and closely matches the reference. In terms of our quantifiers, the maximal absolute deviation from perfect symmetry was 0.11 kcal mol⁻¹, and the root-mean-square-deviation was 0.016 kcal mol⁻¹.

Again, it should be noted that this approach allows different time length simulations at the different levels of theory to be employed (i.e., low level = MM, intermediate level = SCC-DFTB). It is therefore easy to envision this approach offering significant improvement over current techniques (i.e., FEP) used with the *indirect* QM/MM scheme. For example, if SCC-DFTB is treated as the “low level” of theory then a simulation at a very cheap QM level of theory (e.g., HF) could be used for “intermediate” data and ultimately reweighted to a more accurate QM Hamiltonian (e.g., M06-2X, B97X-V, RIMP2) [111–114].

4. Conclusion

Computing free energies as a function of reaction coordinate is extremely challenging. Herein, we presented two new approaches that have the potential to greatly increase the accuracy and efficiency of this task. First, the off-path simulation method (OPS) has several attractive features. These include both a straightforward definition of reaction coordinate (i.e., based on weighted RMSD or a subset of atoms) and coupling to a new distributed replica (i.e., OPS-REPDstr) framework that offers both efficiency and usability improvements over the standard infrastructure (i.e., OPS-REPLICA).

This new approach and framework were validated by computing the PMF of rotating butane around its central torsion. Additionally, the much more challenging PMF of the disaccharide maltose's glycosidic rotation was also carried out both in vacuo and explicit solvation. The latter case was compared to a previously published adaptive umbrella sampling study and found to reproduce both the qualitative and quantitative results while significantly reducing the overall simulation time required. This application of the OPS-REPD procedure suggests that this new highly parallelized infrastructure overcomes the previous limitations of the OPS-REPL framework, which is highlighted by the maltose explicit solvent simulations.

In addition to the new OPS method and REPDstr framework, we have also implemented transition state searching functionality into CHARMM's Newton–Raphson minimizer. This procedure uses eigenvector following and allows the user to specify which order saddle point is of specific interest; this will be 1 in cases where the transition state is the ultimate target. As most potential energy surfaces can be rather complex it is suggested that an initial structural perturbation, in the direction of the desired saddle point, be performed prior to beginning the optimization. This can be done easily with coordinate fixing or restraining potentials in CHARMM or other software.

Finally, the recently developed QM-NBB scheme for computing high level (i.e., QM or QM/MM) free energy differences at affordable costs (i.e., MM or semi-empirical QM) has been incorporated into the so-called *indirect* QM/MM scheme for computing reaction free energies. In this procedure, an underlying PMF is generated at lower level of theory and then connected to higher level results via standard FEP. However, this connection has been shown to be problematic if the lower and higher level potential energy surfaces are too different or if fixed QM regions are employed and entropic effects play a significant role. To help circumvent these problems, connecting levels of theory via QM-NBB, rather than FEP, is suggested. Specifically, we have shown that even for a simple molecule like butane connecting MM \rightarrow HF/6-31G* via

FEP already encounters non-ideal behavior. However, augmenting the data with those obtained at an “intermediate” level of theory, (e.g., SCC-DFTB) that overlaps more closely with the high level data, significantly improves results while simultaneously reducing the number of QM calculations necessary by up to an order of magnitude; with a factor of 2–5 almost guaranteed. Although butane is a rather simplistic first example, it is clear that combining QM-NBB with reaction path techniques offers significant advantages over current procedures.

5. Acknowledgements

HLW3 would like to acknowledge NIH (1K22HL088341-01A1), DOE (DE-SC0011297TDD), and the University of South Florida (start-up) for the funding. Computations were performed at the USF Research Computing Center (NSF Grant No. CHE-0722887) and XSEDE (MCB120133), and both centers are greatly appreciated. Further, this research was supported in part by the Intramural Research Program of the NIH, National Heart, Lung, and Blood Institute. The NIH funding for the LoBoS (<http://www.lobos.nih.gov>) and Biowulf (<http://biowulf.nih.gov>) super computing systems is also acknowledged and appreciated. Finally, we would like to thank the following people for their helpful discussions: Dr. Bernard Brooks, Dr. Gerhard König, and Prof. Michael Shirts.

References

- [1] R.W. Zwanzig, High-temperature equation of state by a perturbation method. I. Nonpolar gases, *J. Chem. Phys.* 22 (8) (1954) 1420–1426.
- [2] R.W. Zwanzig, High-temperature equation of state by a perturbation method. II. Polar gases, *J. Chem. Phys.* 23 (10) (1955) 1915–1922.
- [3] W.L. Jorgensen, L.L. Thomas, Perspective on free-energy perturbation calculations for chemical equilibria, *J. Chem. Theory Comput.* 4 (6) (2008) 869–876.
- [4] A. Pohorille, C. Jarzynski, C. Chipot, Good practices in free-energy calculations, *J. Phys. Chem. B* 114 (32) (2010) 10235–10253.
- [5] J.G. Kirkwood, Statistical mechanics of fluid mixtures, *J. Chem. Phys.* 3 (5) (1935) 300–313.
- [6] C.H. Bennett, Efficient estimation of free energy differences from Monte Carlo data, *J. Comput. Phys.* 22 (2) (1976) 245–268.
- [7] M. Shirts, E. Bair, G. Hooker, V. Pande, Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods, *Phys. Rev. Lett.* 91 (14) (2003) 140601.
- [8] M.R. Shirts, V.S. Pande, Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration, *J. Chem. Phys.* 122 (14) (2005) 144107.
- [9] S. Bruckner, S. Boresch, Efficiency of alchemical free energy simulations. I. A practical comparison of the exponential formula, thermodynamic integration, and Bennett's acceptance ratio method, *J. Comput. Chem.* 32 (7) (2011) 1303–1319.
- [10] S. Bruckner, S. Boresch, Efficiency of alchemical free energy simulations. II. Improvements for thermodynamic integration, *J. Comput. Chem.* 32 (7) (2011) 1320–1333.
- [11] G. Torrie, J. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling, *J. Comput. Phys.* 23(2), 187–199.
- [12] B. Roux, The calculation of the potential of mean force using computer simulations, *Comput. Phys. Commun.* 91 (1–3) (1995) 275–282.
- [13] A. Ferrenberg, R. Swendsen, Optimized Monte Carlo data analysis, *Phys. Rev. Lett.* 63 (12) (1989) 1195–1198.
- [14] S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen, P.A. Kollman, The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, *J. Comput. Chem.* 13 (8) (1992) 1011–1021.
- [15] R. Czereminski, R. Elber, Reaction path study of conformational transitions in flexible systems: applications to peptides, *J. Chem. Phys.* 92 (9) (1990) 5580–5601.
- [16] R. Czereminski, R. Elber, Self-avoiding walk between two fixed points as a tool to calculate reaction paths in large molecular systems, *Int. J. Quantum Chem.* 38 (S24) (1990) 167–185.
- [17] R. Elber, M. Karplus, A method for determining reaction paths in large molecules: application to myoglobin, *Chem. Phys. Lett.* 139 (5) (1987) 375–380.
- [18] E. W. Ren, E. Vanden-Eijnden, Simplified and improved string method for computing the minimum energy paths in barrier-crossing events, *J. Chem. Phys.* 126 (16) (2007) 164103.
- [19] W. Ren, E. Vanden-Eijnden, P. Maragakis, E. W., Transition pathways in complex systems: application of the finite-temperature string method to the alanine dipeptide, *J. Chem. Phys.* 123 (13) (2005) 134109.
- [20] L. Maragliano, A. Fischer, E. Vanden-Eijnden, G. Ciccotti, String method in collective variables: minimum free energy paths and isocommittor surfaces, *J. Chem. Phys.* 125 (2) (2006) 24106.
- [21] W. Quapp, Finding the transition state without initial guess: the growing string method for Newton trajectory to isomerization and enantiomerization reaction of alanine dipeptide and poly(15)alanine, *J. Comput. Chem.* 28 (11) (2007) 1834–1847.

- [22] E. W., W. Ren, E. Vanden-Eijnden, String method for the study of rare events, *Phys. Rev. B* 66 (5) (2002) 052301.
- [23] A.C. Pan, D. Sezer, B. Roux, Finding transition pathways using the string method with swarms of trajectories, *J. Phys. Chem. B* 112 (11) (2008) 3432–3440.
- [24] A.C. Pan, B. Roux, Building Markov state models along pathways to determine free energies and rates of transitions, *J. Chem. Phys.* 129 (6) (2008) 064107.
- [25] E. Vanden-Eijnden, Some recent techniques for free energy calculations, *J. Comput. Chem.* 30 (11) (2009) 1737–1747.
- [26] E. W., W. Ren, E. Vanden-Eijnden, Finite temperature string method for the study of rare events, *J. Phys. Chem. B* 109 (14) (2005) 6688–6693.
- [27] E.F. Koslover, D.J. Wales, Comparison of double-ended transition state search methods, *J. Chem. Phys.* 127 (13) (2007) 134102.
- [28] J.M. Carr, S.A. Trygubenko, D.J. Wales, Finding pathways between distant local minima, *J. Chem. Phys.* 122 (23) (2005) 234903.
- [29] D.J. Wales, Energy landscapes: calculating pathways and rates, *Int. Rev. Phys. Chem.* 25 (1–2) (2006) 237–282.
- [30] S.A. Trygubenko, D.J. Wales, Analysis of cooperativity and localization for atomic rearrangements, *J. Chem. Phys.* 121 (14) (2004) 6689–6697.
- [31] L. Cao, C. Lv, W. Yang, Hidden conformation events in DNA base extrusions: a generalized-ensemble path optimization and equilibrium simulation study, *J. Chem. Theory Comput.* 9 (8) (2013) 3756–3768.
- [32] D. Jacquemin, B. Mennucci, C. Adamo, Excited-state calculations with TD-DFT: from benchmarks to simulations in complex environments, *Phys. Chem. Chem. Phys.* 13 (38) (2011) 16987–16998.
- [33] M.E. Casida, M. Huix-Rotllant, Progress in time-dependent density-functional theory, *Annu. Rev. Phys. Chem.* 63 (1) (2012) 287–323.
- [34] A.J. Cohen, P. Mori-Sánchez, W. Yang, Challenges for density functional theory, *Chem. Rev.* 112 (1) (2012) 289–320.
- [35] C.J. Cramer, D.G. Truhlar, Density functional theory for transition metals and transition metal chemistry, *Phys. Chem. Chem. Phys.* 11 (46) (2009) 10757–10816.
- [36] S. Grimme, Density functional theory with London dispersion corrections, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1 (2) (2011) 211–228.
- [37] W. Hieringer, A. Görling, Reply to Comment on 'Failure of time-dependent density functional methods for excitations in spatially separated systems' by Andreas Dreuw and Martin Head-Gordon, *Chem. Phys. Lett.* 426 (1–3) (2006) 234–236.
- [38] G. Henkelman, H. Jonsson, Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points, *J. Chem. Phys.* 113 (22) (2000) 9978.
- [39] H. Lee Woodcock, M. Hodoseck, P. Sherwood, Y.S. Lee, H.F. Schaefer III, B.R. Brooks, Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase, *Theor. Chem. Accounts Theory Comput. Model. (Theor. Chim. Acta)* 109 (3) (2003) 140–148.
- [40] J.-W. Chu, B.L. Trout, B.R. Brooks, A super-linear minimization scheme for the nudged elastic band method, *J. Chem. Phys.* 119 (24) (2003) 12708.
- [41] H.L. Woodcock, M. Hodoseck, B.R. Brooks, Exploring SCC-DFTB paths for mapping QM/MM reaction mechanisms, *J. Phys. Chem. A* 111 (26) (2007) 5720–5728.
- [42] H.L. Woodcock, M. Hodoseck, A.T.B. Gilbert, P.M.W. Gill, H.F. Schaefer, B.R. Brooks, Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations, *J. Comput. Chem.* 28 (9) (2007) 1485–1502.
- [43] S.A. Trygubenko, D.J. Wales, A doubly nudged elastic band method for finding transition states, *J. Chem. Phys.* 120 (5) (2004) 2082–2094.
- [44] B. Peters, A. Heyden, A.T. Bell, A. Chakraborty, A growing string method for determining transition states: comparison to the nudged elastic band and string methods, *J. Chem. Phys.* 120 (17) (2004) 7877–7886.
- [45] A. Goodrow, A.T. Bell, M. Head-Gordon, Development and application of a hybrid method involving interpolation and ab initio calculations for the determination of transition states, *J. Chem. Phys.* 129 (17) (2008) 174109.
- [46] D. Sheppard, R. Terrell, G. Henkelman, Optimization methods for finding minimum energy paths, *J. Chem. Phys.* 128 (13) (2008) 134106.
- [47] A. Goodrow, A.T. Bell, M. Head-Gordon, Transition state-finding strategies for use with the growing string method, *J. Chem. Phys.* 130 (24) (2009) 244108.
- [48] A. Goodrow, A.T. Bell, M. Head-Gordon, A strategy for obtaining a more accurate transition state estimate using the growing string method, *Chem. Phys. Lett.* 484 (4–6) (2010) 392–398.
- [49] W. Quapp, J.M. Bofill, A comment to the nudged elastic band method, *J. Comput. Chem.* 31 (13) (2010) 2526–2531.
- [50] J. Klimeš, D.R. Bowler, A. Michaelides, A critical assessment of theoretical methods for finding reaction pathways and transition states of surface processes, *J. Phys. Condens. Matter* 22 (7) (2010) 074203.
- [51] W. Quapp, J.M. Bofill, Reply to the comment by Sheppard and Henkelman on the nudged elastic band method, *J. Comput. Chem.* 32 (8) (2011) 1772–1773.
- [52] C. Shang, Z.-P. Liu, Constrained Brody dimer method with bias potential for exploring potential energy surface of multistep reaction process, *J. Chem. Theory Comput.* 8 (7) (2012) 2215–2222.
- [53] G.A. Cisneros, H. Liu, Z. Lu, W. Yang, Reaction path determination for quantum mechanical/molecular mechanical modeling of enzyme reactions by combining first order and second order "chain-of-replicas" methods, *J. Chem. Phys.* 122 (11) (2005) 114502.
- [54] H. Hu, W. Yang, Dual-topology/dual-coordinate free-energy simulation using QM/MM force field, *J. Chem. Phys.* 123 (4) (2005) 041102.
- [55] D. Riccardi, P. Schaefer, Y. Yang, H. Yu, N. Ghosh, X. Prat-Resina, P. König, G. Li, D. Xu, H. Guo, M. Elstner, Q. Cui, Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes, *J. Phys. Chem. B* 110 (13) (2006) 6458–6469.
- [56] H. Li, W. Yang, Sampling enhancement for the quantum mechanical potential based molecular dynamics simulations: a general algorithm and its extension for free energy calculation on rugged energy surface, *J. Chem. Phys.* 126 (11) (2007) 114104.
- [57] H. Hu, Z. Lu, W. Yang, QM/MM minimum free energy path: methodology and application to triosephosphate isomerase, *J. Chem. Theory Comput.* 3 (2) (2007) 390–406.
- [58] H.L. Woodcock, W. Zheng, A. Ghysels, Y. Shao, J. Kong, B.R. Brooks, Vibrational subsystem analysis: a method for probing free energies and correlations in the harmonic limit, *J. Chem. Phys.* 129 (21) (2008) 214109.
- [59] H. Hu, Z. Lu, J.M. Parks, S.K. Burger, W. Yang, Quantum mechanics/molecular mechanics minimum free-energy path for accurate reaction energetics in solution and enzymes: sequential sampling and optimization on the potential of mean force surface, *J. Chem. Phys.* 128 (3) (2008) 034105.
- [60] H. Hu, W. Yang, Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods, *Annu. Rev. Phys. Chem.* 59 (1) (2008) 573–601.
- [61] J.M. Parks, H. Hu, J. Rudolph, W. Yang, Mechanism of Cdc25B phosphatase with the small molecule substrate p-nitrophenyl phosphate from QM/MM-MFEP calculations, *J. Phys. Chem. B* 113 (15) (2009) 5217–5224.
- [62] M. Chen, W. Yang, On-the-path random walk to efficiently optimize minimum free energy path, *J. Comput. Chem.* 30 (11) (2009) 1649–1653.
- [63] O. Acevedo, W.L. Jorgensen, Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions, *Acc. Chem. Res.* 43 (1) (2010) 142–151.
- [64] W. Yang, Q. Cui, D. Min, H. Li, QM/MM alchemical free energy simulations: challenges and recent developments, *Ann. Rep. Comput. Chem.*, vol. 6, Elsevier, 2010.
- [65] D. Min, L. Zheng, M. Chen, C. Lv, W. Yang, Practically efficient QM/MM alchemical free energy simulations: the orthogonal space random walk strategy, *J. Chem. Theory Comput.* 6 (8) (2010) 2253–2266.
- [66] H. Hu, W. Yang, Elucidating solvent contributions to solution reactions with ab initio QM/MM methods, *J. Phys. Chem. B* 114 (8) (2010) 2755–2759.
- [67] D. Min, M. Chen, L. Zheng, Y. Jin, M.A. Schwartz, Q.-X.A. Sang, W. Yang, Enhancing QM/MM molecular dynamics sampling in explicit environments via an orthogonal-space-random-walk-based strategy, *J. Phys. Chem. B* 115 (14) (2011) 3924–3935.
- [68] A. Ghysels, H.L. Woodcock, J.D. Larkin, B.T. Miller, Y. Shao, J. Kong, D.V. Neck, V.V. Speybroeck, M. Waroquier, B.R. Brooks, Efficient calculation of QM/MM frequencies with the Mobile Block Hessian, *J. Chem. Theory Comput.* 7 (2) (2011) 496–514.
- [69] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, *Chem. Phys. Lett.* 314 (1999) 141–151.
- [70] A. Mitsuake, Y. Okamoto, Replica-exchange extensions of simulated tempering method, *J. Chem. Phys.* 121 (6) (2004) 2491–2504.
- [71] H. Kamberaj, A. van der Vaart, Multiple scaling replica exchange for the conformational sampling of biomolecules in explicit water, *J. Chem. Phys.* 127 (23) (2007) 234102.
- [72] A.E. Roitberg, A. Okur, C. Simmerling, Coupling of replica exchange simulations to a non-Boltzmann structure reservoir, *J. Phys. Chem. B* 111 (10) (2007) 2415–2418.
- [73] L. Zheng, M. Chen, W. Yang, Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems, *Proc. Natl. Acad. Sci. U. S. A.* 105 (51) (2008) 20227–20232.
- [74] S. Lee, M. Chen, W. Yang, N.G.J. Richards, Sampling long time scale protein motions: OSRW simulation of active site loop conformational free energies in formyl-CoA: oxalate CoA transferase, *J. Am. Chem. Soc.* 132 (21) (2010) 7252–7253.
- [75] X. Wu, B.R. Brooks, Toward canonical ensemble distribution from self-guided Langevin dynamics simulation, *J. Chem. Phys.* 134 (13) (2011) 134108.
- [76] G. König, B.T. Miller, S. Boresch, X. Wu, B.R. Brooks, Enhanced sampling in free energy calculations: combining SGLD with the Bennett's acceptance ratio and enveloping distribution sampling methods, *J. Chem. Theory Comput.* 8 (10) (2012) 3650–3662.
- [77] X. Wu, M. Hodoseck, B.R. Brooks, Replica exchanging self-guided Langevin dynamics for efficient and accurate conformational sampling, *J. Chem. Phys.* 137 (4) (2012) 044106.
- [78] J. Gao, X. Xia, A priori evaluation of aqueous polarization effects through Monte Carlo QM–MM simulations, *Science* 258 (5082) (1992) 631–635.
- [79] V. Luzhkov, A. Warshel, Microscopic models for quantum mechanical calculations of chemical processes in solutions: LD/AMPAC and SCAAS/AMPAC calculations of solvation energies, *J. Comput. Chem.* 13 (2) (1992) 199–213.
- [80] J. Gao, F.J. Luque, M. Orozco, Induced dipole moment and atomic charges based on average electrostatic potentials in aqueous solution, *J. Chem. Phys.* 98 (4) (1993) 2975.
- [81] P. Kollman, Free energy calculations: applications to chemical and biochemical phenomena, *Chem. Rev.* 93 (7) (1993) 2395–2417.
- [82] T. Wesolowski, A. Warshel, Ab initio free energy perturbation calculations of solvation free energy using the frozen density functional approach, *J. Phys. Chem.* 98 (20) (1994) 5183–5187.
- [83] J. Gao, M. Freindorf, Hybrid ab initio QM/MM simulation of N-methylacetamide in aqueous solution, *J. Phys. Chem. A* 101 (17) (1997) 3182–3188.
- [84] T.H. Rod, U. Ryde, Accurate QM/MM free energy calculations of enzyme reactions: methylation by catechol O-methyltransferase, *J. Chem. Theory Comput.* 1 (6) (2005) 1240–1251.
- [85] T. Rod, U. Ryde, Quantum mechanical free energy barrier for an enzymatic reaction, *Phys. Rev. Lett.* 94 (13) (2005) 138302.
- [86] J. Heimdal, U. Ryde, Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations, *Phys. Chem. Chem. Phys.* 14 (36) (2012) 12592–12604.

- [87] M. Kaukonen, P. Söderhjelm, J. Heimdal, U. Ryde, Proton transfer at metal sites in proteins studied by quantum mechanical free-energy perturbations, *J. Chem. Theory Comput.* 4 (6) (2008) 985–1001.
- [88] N. Lu, D.A. Kofke, T.B. Woolf, Improving the efficiency and reliability of free energy perturbation calculations using overlap sampling methods, *J. Comput. Chem.* 25 (1) (2004) 28–39.
- [89] M.R. Shirts, J.D. Chodera, Statistically optimal analysis of samples from multiple equilibrium states, *J. Chem. Phys.* 129 (12) (2008) 124105.
- [90] G. König, P.S. Hudson, S. Boresch, H.L. Woodcock, Multiscale free energy simulations: an efficient method for connecting classical MD simulations to QM or QM/MM free energies using non-Boltzmann Bennett reweighting schemes, *J. Chem. Theory Comput.* 10 (4) (2014) 1406–1419.
- [91] G. König, F.C. Pickard, Y. Mei, B.R. Brooks, Predicting hydration free energies with a hybrid QM/MM approach: an evaluation of implicit and explicit solvation models in SAMPL4, *J. Comput. Aided Mol. Des.* 28 (3) (2014) 245–257.
- [92] B.R. Brooks, C.L. Brooks, A.D. Mackerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: the biomolecular simulation program, *J. Comput. Chem.* 30 (10) (2009) 1545–1614.
- [93] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem.* 4 (2) (1983) 187–217.
- [94] Y. Shao, L.F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S.T. Brown, A.T.B. Gilbert, L.V. Slipchenko, S.V. Levchenko, D.P. O'Neill, R.A. DiStasio, R.C. Lochan, T. Wang, G.J.O. Beran, N.A. Besley, J.M. Herbert, C.Y. Lin, T. Van Voorhis, S.H. Chien, A. Sodt, R.P. Steele, V.A. Rassolov, P.E. Maslen, P.P. Korambath, R.D. Adamson, B. Austin, J. Baker, E.F.C. Byrd, H. Dachsel, R.J. Doerksen, A. Dreuw, B.D. Dunietz, A.D. Dutoi, T.R. Furlani, S.R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R.Z. Khaliullin, P. Klunzinger, A.M. Lee, M.S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E.I. Proynov, P.A. Pieniazek, Y.M. Rhee, J. Ritchie, E. Rosta, C.D. Sherrill, A.C. Simmonett, J.E. Subotnik, H.L. Woodcock, W. Zhang, A.T. Bell, A.K. Chakraborty, D.M. Chipman, F.J. Keil, A. Warshel, W.J. Hehre, H.F. Schaefer, J. Kong, A.I. Krylov, P.M.W. Gill, M. Head-Gordon, Advances in methods and algorithms in a modern quantum chemistry program package, *Phys. Chem. Chem. Phys.* 8 (27) (2006) 3172–3191.
- [95] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, G. Seifert, Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties, *Phys. Rev. B* 58 (11) (1998) 7260–7268.
- [96] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, M. Karplus, A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method, *J. Phys. Chem. B* 105 (2) (2001) 569–585.
- [97] M.W. Schmidt, K.K. Baldridge, J.A. Boat, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis, J.A. Montgomery, General atomic and molecular electronic structure system, *J. Comput. Chem.* 14 (11) (1993) 1347–1363.
- [98] D.C. Chatfield, B.R. Brooks, HIV-1 protease cleavage mechanism elucidated with molecular dynamics simulation, *J. Am. Chem. Soc.* 117 (20) (1995) 5561–5572.
- [99] M.F. Guest, I.J. Bush, H.J.J. Van Dam, P. Sherwood, J.M.H. Thomas, J.H. Van Lenthe, R.W.A. Havenith, J. Kendrick, The GAMESS-UK electronic structure package: algorithms, developments and applications, *Mol. Phys.* 103 (6–8) (2005) 719–747.
- [100] D. Das, K.P. Eurenium, E.M. Billings, P. Sherwood, D.C. Chatfield, M. Hodoscek, B.R. Brooks, Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method, *J. Chem. Phys.* 117 (23) (2002) 10534.
- [101] 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 (10) (2009) 2583–2588.
- [102] M.M. Kuttel, K.J. Naidoo, Free energy surfaces for the $\alpha(1 \rightarrow 4)$ -glycosidic linkage: implications for polysaccharide solution structure and dynamics, *J. Phys. Chem. B* 109 (15) (2005) 7468–7474.
- [103] C. Bartels, M. Karplus, Multidimensional adaptive umbrella sampling: applications to main chain and side chain peptide conformations, *J. Comput. Chem.* 18 (12) (1997) 1450–1462.
- [104] C. Bartels, M. Karplus, Probability distributions for complex systems: adaptive umbrella sampling of the potential energy, *J. Phys. Chem. B* 102 (5) (1998) 865–880.
- [105] R.K. Schmidt, B. Teo, J.W. Brady, Use of umbrella sampling in the calculation of the potential of mean force for maltose in vacuum from molecular dynamics simulations, *J. Phys. Chem.* 99 (29) (1995) 11339–11343.
- [106] E.M. Boczek, C.L. Brooks, Constant-temperature free energy surfaces for physical and chemical processes, *J. Phys. Chem.* 97 (17) (1993) 4509–4513.
- [107] R.W.W. Hooft, B.P. van Eijck, J. Kroon, An adaptive umbrella sampling procedure in conformational analysis using molecular dynamics and its application to glycol, *J. Chem. Phys.* 97 (9) (1992) 6690.
- [108] M. Kuttel, J.W. Brady, K.J. Naidoo, Carbohydrate solution simulations: producing a force field with experimentally consistent primary alcohol rotational frequencies and populations, *J. Comput. Chem.* 23 (13) (2002) 1236–1243.
- [109] M. Gaus, A. Goez, M. Elstner, Parametrization and benchmark of DFTB3 for organic molecules, *J. Chem. Theory Comput.* 9 (1) (2013) 338–354.
- [110] M. Gaus, X. Lu, M. Elstner, Q. Cui, Parameterization of DFTB3/3OB for sulfur and phosphorus for chemical and biological applications, *J. Chem. Theory Comput.* 10 (4) (2014) 1518–1537.
- [111] Y. Zhao, D.G. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other function, *Theor. Chem. Accounts* 120 (1–3) (2007) 215–241.
- [112] N. Mardirossian, M. Head-Gordon, ω B97X-V: a 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy, *Phys. Chem. Chem. Phys.* 16 (21) (2014) 9904–9924.
- [113] R.A. DiStasio, R.P. Steele, Y.M. Rhee, Y. Shao, M. Head-Gordon, An improved algorithm for analytical gradient evaluation in resolution-of-the-identity second-order Møller–Plesset perturbation theory: application to alanine tetrapeptide conformational analysis, *J. Comput. Chem.* 28 (5) (2007) 839–856.
- [114] M. Feyereisen, G. Fitzgerald, A. Komornicki, Use of approximate integrals in ab initio theory. An application in MP2 energy calculations, *Chem. Phys. Lett.* 208 (5–6) (1993) 359–363.