Theor Chem Acc (2003) 109:140–148
DOI 10.1007/s00214-002-0421-3

Theoretical Chemistry Accounts

Regular article

Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase

H. Lee Woodcock[1,2], Milan Hodošček[3], Paul Sherwood[4], Yong S. Lee[5], Henry F. Schaefer III[2], Bernard R. Brooks[1]

[1] Laboratory of Biophysical Chemistry, NHLBI, National Institutes of Health, Bethesda, MD 20892, USA
[2] Center for Computational Quantum Chemistry, University of Georgia, Athens, GA 30602-2556, USA
[3] National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia
[4] CLRC Daresbury Lab, Warrington WA4 4AD, UK
[5] Center for Molecular Modeling, CIT, National Institutes of Health, Bethesda, MD 20892, USA

Received: 16 December 2001 / Accepted: 6 September 2002 / Published online: 8 April 2003

Abstract. A replica path method has been developed and extended for use in complex systems involving hybrid quantum/classical (quantum mechanical/molecular mechanical) coupled potentials. This method involves the definition of a reaction path via replication of a set of macromolecular atoms. An ''important'' subset of these replicated atoms is restrained with a penalty function based on weighted root-mean-square rotation/translation best-fit distances between adjacent ($i \pm 1$) and next adjacent ($i \pm 2$) pathway steps. An independent subset of the replicated atoms may be treated quantum mechanically using the computational engine Gamess-UK. This treatment can be performed in a highly parallel manner in which many dozens of processors can be efficiently employed. Computed forces may be projected onto a reference pathway and integrated to yield a potential of mean force (PMF). This PMF, which does not suffer from large errors associated with calculated potential-energy differences, is extremely advantageous. As an example, the QM/MM replica path method is applied to the study of the Claisen rearrangement of chorismate to prephenate which is catalyzed by the *Bacillus subtilis* isolated, chorismate mutase. Results of the QM/MM pathway minimizations yielded an activation enthalpy $\Delta H^{\dagger\dagger}$ of 14.9 kcal/mol and a reaction enthalpy of −19.5 kcal/mol at the B3LYP/6-31G(d) level of theory. The resultant pathway was compared and contrasted with one obtained using a forced transition approach based on a reaction coordinate constrained repeated walk procedure ($\Delta H^{\dagger\dagger} = 20.1$ kcal/mol, $\Delta H_{\text{rxn}} = -20.1$ kcal/mol, RHF/4-31G). The optimized replica path results compare favorably to the experimental activation enthalpy of 12.7 ± 0.4 kcal/mol.

Keywords: Quantum mechanical/molecular mechanical – Replica path method – Chorismate mutase

Contribution to the Proceedings of the Symposium on Combined QM/MM Methods at the 22nd National Meeting of the American Chemical Society, 2001.

Correspondence to: H.L. Woodcock
e-mail: hlwood@ccqc.uga.edu

1. Introduction

1.1 Reaction path methods

Reaction pathways yield much insight into the nature of chemical reactions. Thus, it is not surprising that many methods have been developed for the study of both large and small systems. The most popular of these methods, the intrinsic reaction coordinate method, makes use of internal coordinates to map a pathway connecting products to reactants through a known transition state [1]. Another popular method, which makes use of internal coordinates and an eigenvector following algorithm [2], is capable of optimizing pathways without *a priori* knowledge of the transition state [3]. This pathway determination method developed by Ayala and Schlegel is an extension of the self penalty walk (SPW) method of Elber and Karplus [4] and Elber and Czerminiski, [5, 6], who used energy and gradient information to minimize pathways.

Pathway optimization methods that replace the expensive analytic Hessian computation with a simple method, based on gradient information, are extremely desirable. Ayala and Schlegel's reaction pathway procedure is stated to be very efficient and to require as few as five points to locate a transition state and the pathway conjoining reactants to products [3]. However, typical eigenvector following methods become prohibitive as

the systems of interest increase in size (e.g., proteins). This is easily discernible as the generation of internal coordinates requires diagonalization of a Cartesian coordinate matrix which scales as N^2 in memory and N^3 in time (where N is the number of atoms). For larger, biochemical, systems (more than 10000 atoms) this quickly becomes impractical; therefore, the use of Cartesian coordinates, which do not experience the system size limitations of internal coordinates, is preferable for the study of macromolecular systems. Consequently, application of the internal-coordinate-based Ayala–Schelegel method is ill suited for the study of biological systems.

Recently the replica path method in CHARMM [7] has been extended to be used with QM/MM energy surfaces. This method has four major components:

1. Implementation of a path restraint penalty function that uses weighted root-mean-square (rms) best-fit coordinates of a smaller, replicated, subset of atoms along a discretized path.
2. A larger subset of the macromolecular system (which could be the entire system) is replicated so that the position of atoms may vary from step to step along the reaction path.
3. A subset of replicated atoms may be treated quantum mechanically in a QM/MM framework to allow the formation and/or breaking of chemical bonds.
4. Each replicate can be computed efficiently on a different cluster of processors, allowing the simultaneous use of vast computational resources.

Since the pathway defined by the replica path method is nondirectional, typical hysteresis problems encountered in determining the detailed structure of the pathway are avoided. Furthermore, it is possible to anneal to low-energy structures on the replica path surface using high-temperature molecular dynamics.

Another advantage gained in applying the replica path method is the ability to use forces in the energetic analysis of a minimized path without need for accurate energy calculations at each pathway point. When systems become large and many explicit water molecules are included, the potential energy of the entire system is no longer a desirable tool in computing reaction enthalpies or free energies. A water rotation far from the active site will perhaps perturb the potential energy but will have no significant effect on the forces of the active-site atoms. By focusing on the forces displacing the atoms used to define the pathway, the potential of mean force (PMF) in going from step i to step $i+1$ can be determined accurately without the use of the potential energy. Both PMF and direct potential-energy plots are presented for comparison.

By performing simulations on the replica path energy restraint surface, the PMF of a reference, or average, path may be computed to obtain relative free-energy differences, $\Delta\Delta G$ of steps along that path. The technical details of this procedure will be presented in a subsequent paper. Whereas this analysis is warranted for classical potential-energy surfaces it becomes prohibitively expensive for the QM/MM surfaces presented in the present study. This is because many time-steps are required to obtain sufficiently accurate statistical sampling. However, as the availability of parallel computer time increases, obtaining QM/MM free-energy curves will thus become more practical. By repeating the calculation while varying the temperature, both entropy $\Delta S^{\dagger\dagger}$, and free energy, $\Delta G^{\dagger\dagger}$, can be computed via Eq. (1).

$$\Delta G^{\dagger\dagger} = \Delta H^{\dagger\dagger} - T\Delta S^{\dagger\dagger} \tag{1}$$

It is also practical to perform this analysis on a less expensive QM/MM surface with the assumption that the entropy contribution will be nearly independent of the QM method employed.

1.2 Claisen rearrangement of chorismate to prephenate

Development of the parallel QM/MM repliaca path method is also exciting owing to the many reactions which may not be accurately modeled by gas-phase QM methods, but which rather depend on specific environments only encountered in the active sites of rate-enhancing proteins. One such case is examined in the current work. The well-studied [8, 9] Claisen rearrangement of chorismate (A) to prephenate (C) through a transition state analogue (B) (overall scheme in Fig. 1) is used as a demonstration of the QM/MM replica path procedure. This is an ideal test case for many reasons. First, the experimental activation enthalpy (12.7 ± 4 kcal/mol) has been well established [10]. A second reason for the use of chorismate mutase as a test case is the availability of *Bacillus subtilis* X-ray structures, elucidated by Chook et al. in 1994 [11] and again in 1996 by Kast et al. [10]. Finally, since it is well known [10, 11, 12] that side chains do not explicitly participate in the Claisen rearrangement we will not have to consider the difficult problem of QM/MM barrier interactions, which may cause significant errors and are the focus of ongoing methodological development [13, 14, 15].

The chemical and biological significance of chorismate mutase is unquestioned. The 10^6-fold increase in rate that is observed in the active site relative to the chorismate rearrangement in solution is direct evidence of this [16]. This rate enhancement plays a key role in bacteria, fungi, and higher-order plant synthesis of vital aromatic compounds, including phenylalanine and tyrosine (shikimate acid pathway) [17].

Owing to the importance of this reaction, many previous studies, both QM and QM/MM, have been performed to gain insight into the possible development of new antibiotics, fungicides, and herbicides to combat various infections. One early QM study, by Wiest and Houk [18], employed model systems in an effort to ascertain the factors that influence the rate enhancement observed for the enzymatically catalyzed chorismate-to-prephenate rearrangement. The latter work also examined energetically important active-site side chains and attempted to determine how they influence the change in reaction rate.

Like the Wiest–Houk work, a 1995 study by Lyne et al. [19] also investigated the importance of active-site side chains in the chorismate-to-prephenate reaction. The work of Lyne et al. in contrast to the study of Wiest and Houk, examined the entire protein via QM/MM

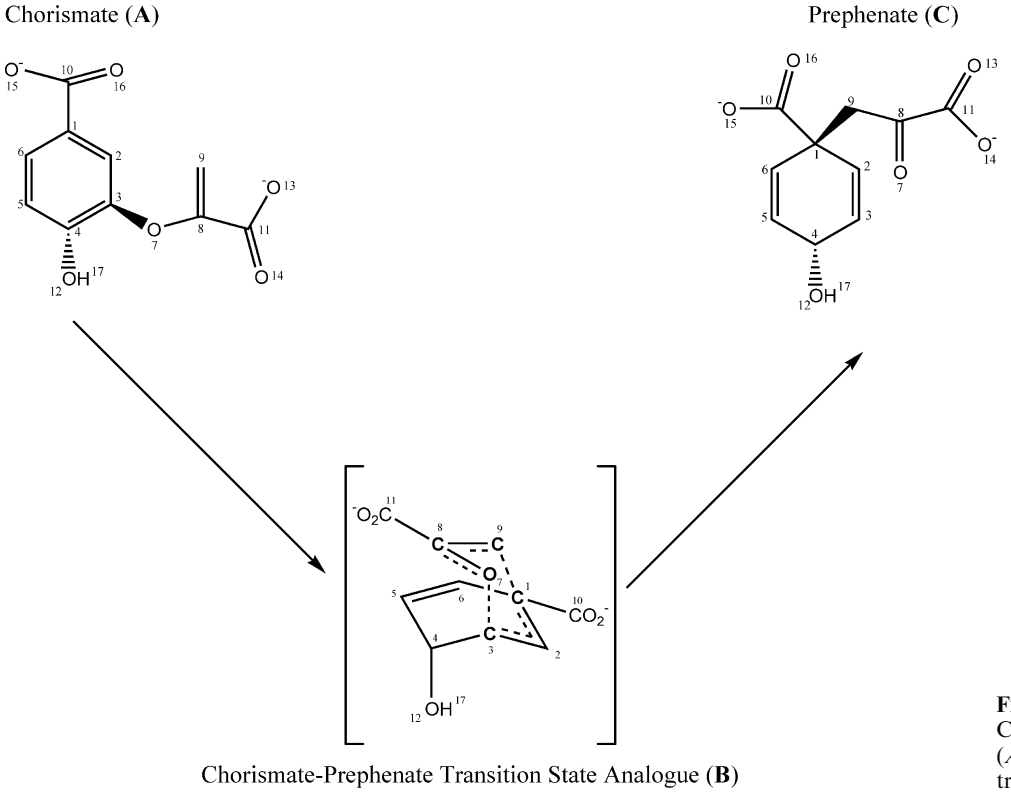

Fig. 1. Chorismate mutase catalyzed Claisen rearrangement of chorismate (*A*) to prephenate (*C*) through a transition-state analogue (*B*)

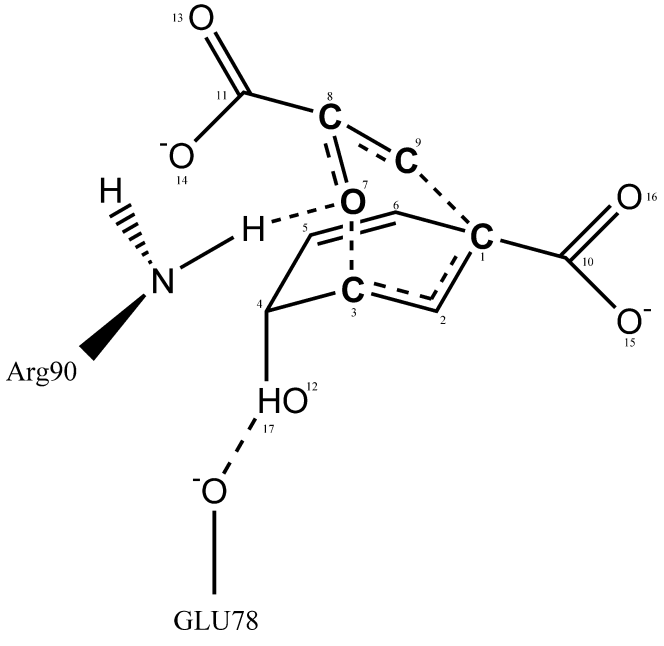

Fig. 2. Transition-state analogue interactions with GLU78 and ARG90 in the active site of chorismate mutase

methods. Lyne et al., used a semiempirical (AM1) [20] Hamiltonian and found, in agreement with Wiest and Houk, that Arg-90 and Arg-7 have important energetic effects in the stabilization of the chorismate–prephenate transition state. However, Lyne et al. also found that Glu-78 plays an important role in transition-state stabilization.

The most recent theoretical study of chorismate mutase was performed in 2001 by Marti et al. [25] and made use of density functional theory (DFT) (B3LYP/6-31G(d)) as well as QM/MM (AM1, CHARMM) [7, 20] computations. This work examined the preorganizational energetic effects of binding the pseudoequatorial conformation of chorismate into the active site of chorismate mutase. In addition, they computed the PMF of the Claisen rearrangement and applied an empirical DFT correction to predict the activation free energy ($\Delta G^{\dagger\dagger}$).

The experimental work on chorismate mutase has been just as vast as the theoretical work. The chorismate-to-prephenate rearrangement has been examined under many different experimental conditions, with the chorismate mutase protein being isolated from many different organisms. [10, 11, 22]. Recently, researchers carried out site-directed-mutagenesis studies in an effort to experimentally determine what computational research had previously investigated (i.e., energetic effects of active-site side chains) [23, 24]. These investigations yielded results that agree with theory as well as allow the accuracy of past and future QM and QM/MM studies to be evaluated.

The present work is undertaken as an initial investigation of the highly parallel QM/MM replica path method and is applied to the chorismate mutase catalyzed chorismate-to-prephenate rearrangement. We employed the QM/MM replica path method starting from the end points of a QM/MM 4-31G optimized pathway, obtained from an intensive 3 week repeated walk calculation. We utilized these end points to generate and optimize a simple linear interpolated guess

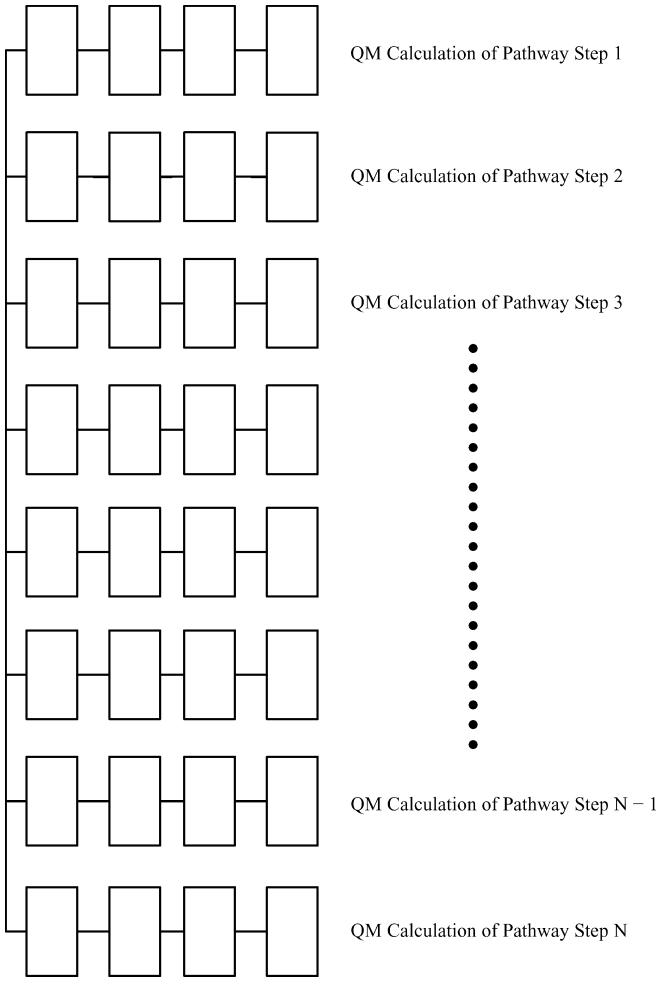

Fig. 3. Schematic of the processor allocation in the massively parallel quantum mechanical/molecular mechanical (QM/MM) replica path procedure. The example illustrates a 16-step pathway with clusters of four processors being allocated per QM computation (self-consistent field and analytic gradient). For the schematic, QM computations take place on horizontal processors, while intercluster communication only occurs when a QM pathway step is completed and CHARMM is ready to evaluate QM gradients and generate the next pathway minimization step

pathway that upon optimization yields a good reference path and allows the accuracy of the PMF analysis to be tested versus a more traditional energetic analysis. The optimized reference pathway can also be utilized in more elaborate calculations to obtain accurate transition structure(s).

2 Computational methods

2.1 Computational procedures

All replica path QM/MM calculations were carried out with the CHARMM [7] molecular mechanics force field using the CHARMM27 [25] parameter set in conjunction with the Gamess-UK [26] QM package. End-point structures used in the generation of the linear interpolated pathway were obtained from the repeated walk optimized pathway. These repeated walk end points were initially taken from the X-ray structure [10] of *B. subtilis* and were minimized during the repeated walk procedure. The linear interpolated replica path was defined from the aforementioned end points, which only contained the crystal structure waters. Subsequently, solvation was increased via the addition of 2903 modified TIP3P [27, 28] waters in a rhombododecahedran followed by minimization of the entire system. For all chorismate mutase trimer calculations presented in this paper, only one active site was studied, while the remaining two were treated as nonreacting via the use of classical model potentials. Additionally, no long-range cutoffs were utilized and all electrostatic interactions were included.

The replica path optimization was performed using the adopted basis Newton–Rhaphson (ABNR) [7] method and minimized until the rms gradient (GRMS) was consistently less than 0.3 kcal/cal/Å and the total pathway energy change was less than 1.0 kcal/mol for at least ten consecutive steps.

This pathway was then used to determine the PMF and to perform energetic analyses to determine both the reaction enthalpy (ΔH_{rxn}) and the enthalpy of activation ($\Delta H^{\dagger\dagger}$). The DFT B3LYP, [29, 30], BLYP [30, 31], and BP86 [31, 32] methods with various basis sets were employed for the final energetic analysis of the optimized linear interpolated pathway.

2.2 Methodological details

Development of the replica path method has evolved over the past few years, like the work of Schlegel et al. [3], from the SPW method of Elber and coworkers [4, 5, 6]. The replica path method, however, has a significant advantage over previous methods; instead of making complete copies of the original system and modifying each of them to define a reaction pathway, like the SPW method, specific ''important'' parts of the biomolecular system may be chosen and replicated to define a more appropriate pathway. An example of this is a pathway that is chosen by selecting a buffer region around and including the active site of a catalytic enzyme and treating the remaining portion of the system as a molecular bath. This allows active-site reactions to be modeled with the emphasis placed on the appropriate region of interest. Furthermore, independent subsets of the replicated atoms may be treated quantum mechanically or used for defining the best-fit restraint penalty function.

In the replica path procedure, the initially created replicas are not assigned coordinates; therefore, each replica must be defined before use in generation of a full reaction pathway. There are three primary procedures for accomplishing this task. First, a simple linear interpolation of Cartesian coordinates may be performed to create a discrete reaction pathway. This is rather simplistic but is effective for many reactions. Second, the use of CHARMM's internal coordinate manipulation [33] can be employed to define more complicated pathways. Third, structures from prior pathway computations can be imported and the replica path procedure employed to refine previously optimized pathways. Using a defined pathway, the QM/MM replica path procedure can be initiated. Integral to the replica path procedure is CHARMM's BLOCK subroutine. The BLOCK algorithm has been modified to define specific regions of the biochemical system and control how these regions interact during the pathway minimization (QM replicated region, MM replicated region, MM nonreplicated region).

The replica path procedure, in CHARMM, is called via the RPATH keyword and accepts several variables that yield maximum flexibility to the user [34]. For the chorismate mutase test case, the following keyword arguments were used during the RPATH procedure: KRMS, KANGLE, RMAX, KMAX, COSMAX, MASS, WEIGHT, ROTATION, TRANSLATION.

– KRMS: force constant used to restrain distances between adjacent points along the reaction pathway via an energetic penalty term.

$$E_{\mathrm{rms}} = \sum_{i=1}^{N} \frac{1}{2} K_{\mathrm{rms}} (\mathrm{r}_i - \bar{\mathrm{r}})^2, \qquad (2)$$

where N is the number of points along the pathway,

$$r_i = \mathrm{RMSd}_{\text{best fit}}(i, i+1), \tag{3}$$

and

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

The optimal choice of KRMS should result in a path step variation no greater than 15–20% while not introducing vibrational frequencies in excess of 3000 cm. The value used in the current study was 250000 kcal/mol/Å^2.

- KANGLE: force constant used to restrain the angle between points i, $i \pm 1$, and $i \pm 2$ of the reaction pathway through an angle energetic penalty term. Implemented using the law of cosines.

$$E_{\text{angle}} = \sum_{i=1}^{N} \frac{1}{2} K_{\text{angle}} [\mathrm{COSMAX} - \cos(\Theta)_i]^2 \tag{5}$$

$$\mathrm{COSMAX} > \cos(\Theta)_i,$$

$$E_{\text{angle}} = 0 \qquad \mathrm{COSMAX} \leq \cos(\Theta)_i. \tag{6}$$

The angle Θ is the deviation from linearity. A straight path is defined to have Θ values of zero. The KANGLE parameter determines, in part, the rigidity of the pathway. Its value is chosen so as not to be so small that the path doubles back onto itself, but also not so large that the path is too rigid and misses passing through the transition structure(s). The value used in the current study was 200 kcal/mol.

- RMAX: the maximum distance points along a pathway are allowed to be separated, before the energetic penalty (E_{rms}) is activated increasing KRMS to KMAX. This additional restraint term is used to limit the maximum step length of reaction pathways. The value used for the current study was 0.06 Å.
- KMAX: force constant used to apply an energetic penalty (E_{rms}) during pathway minimization if points along the path move farther than some critical distance RMAX. The value used for the current study was 750000 kcal/mol/Å^2.
- COSMAX: magnitude of the cos Θ values below which angle restraints are caused to be applied via the E_{angle} energetic penalty. The value used in the current study was 0.975.
- ROTATION and TRANSLATION: turns on use of the RMSd best-fit procedure between adjacent pathway points, thus preventing rigid body rotations and translations from effecting the overall minimization. This prevents minimization problems by requiring the selected region to maintain the same general shape from point to point along the reaction pathway. Without the RMSd best-fit procedure, the active-site atoms would shift their positions through unwanted rotation and/or translation as a means of increasing the distance between adjacent path steps.
- MASS : specifies the use of mass weighting in the RMSd best-fit procedure used when the ROTATION and/or TRANSLATION keyword is invoked. The use of the MASS keyword is not appropriate for reactions involving the transfer of hydrogen atoms in that these atoms would need a more significant weighting factor in the RMSd best-fit procedure. The MASS keyword was employed in the current study.
- WEIGHT : specifies the use of values in the main weighting array during the RPATH minimization procedure.

As stated previously, if the ROTATION and/or TRANSLATION keyword is invoked in the RPATH procedure, unlike previous methods, an RMSd best-fit between adjacent points along the pathway is performed, at each step of the minimization, to prevent rigid body rotations and/or translations from affecting pathway forces. The use of the RMSd best-fit procedure is a major improvement over previous methods and reinforces the power of the replica path method. Also, the WEIGHT keyword allows users to place additional weight on the most important atoms or residues in specific reactions. In this manner arbitrary atomic restraint weights may be applied.

Once an optimized pathway is achieved, energetic analysis may be performed to yield an approximate reaction (ΔH_{rxn}) and activation enthalpy ($\Delta H^{\dagger\dagger}$) via two procedures. The first and simplest is to separate the replicated regions into N (N is the number of points along the pathway) different structures and perform a QM/MM energy calculation on each of these points. Secondly the PMF contribution for a specific pathway can be computed from the forces at each point along the reaction pathway via Eq. 7.

$$\mathrm{PMF}(N) = \sum_{i=3}^{N-1} \left\{ \frac{\nabla E(i-1) \cdot [\mathrm{rms}(i, i-1) - \mathrm{rms}(i-2, i-1)]}{4} \right\} + \left\{ \frac{\nabla E(i) \cdot [\mathrm{rms}(i+1, i) - \mathrm{rms}(i-1, i)]}{4} \right\} \tag{7}$$

where N is the number of points in the pathway, rms (i, j) is the weighted best-fit coordinates of structure i onto structure j as a reference, and ∇E is the negative force excluding the path restraints. This equation is the symmetric discretization of the integral $\nabla E \cdot \mathrm{d}r$ for a specific path.

Another extremely powerful tool that can be applied in the study of an RPATH minimized reaction pathway is ab initio molecular dynamics (MD). Using the replica path method to obtain a reference path, MD simulations may be performed and free energy ($\Delta G^{\dagger\dagger}$) information obtained via a PMF analysis. The entropy ($\Delta S^{\dagger\dagger}$) may then be determined via a simple rearrangement of Eq. (1). Details of this will be presented in a subsequent paper.

A novel feature of the new QM/MM replica path method is the highly parallel nature in which it is implemented. The QM/MM interface was constructed to allow each replica created to be passed to a different set of processors if the number of processors is equal to or an integer multiple of the number of points along the pathway. For those that do not have access to large Beowulf clusters [35], a secondary parallelization scheme was included. In the secondary implementation, successive gradient computations at each point along the pathway are carried out by the total number of processors allotted to the calculation. These two parallelization schemes offer maximum flexibility to all users regardless of available computational resources. Note that each set of processors maintains its own most recent QM wavefunction which facilitates convergence of the self-consistent-field procedure in subsequent minimization or MD iterations.

Another very beneficial aspect of this implementation is the choice of the quantum computational package, Gamess-UK [26]. Gamess-UK is effective as the QM engine owing to the additional parallelization that it offers. Application of Gamess-UK/CHARMM, QM/MM computations offer researchers the ability to efficiently use well over 100 processors (Fig. 5) for studies that were recently believed to be impossible. For a general description of QM/MM methods refer to recent review articles [36]. One final benefit in using Gamess-UK is the fully supported double link atom method of partitioning QM/MM interface bonds via Gaussian convolution of nearby classical MM charges.

3 Discussion and results

The QM/MM replica path method is initially applied to investigate the chorismate-to-prephenate Claisen rearrangement. This reaction, which is catalyzed by chorismate mutase, undergoes a 10^6-fold rate enhancement in the active site compared to the same reaction carried out in solution [11, 16, 37]. The results of the current work are summarized in Table 1 and are compared to previous studies, both theoretical and experimental.

This chorismate rearrangement has been studied by numerous researchers [10, 11, 18, 19, 21, 23, 24, 38]. Our initial estimate used the inefficient repeated pathway walk procedure to study the Clasien rearrangement catalyzed by the *B. subtilis* chorismate mutase. This procedure involved defining a reaction pathway and

Table 1. Quantum mechanical (QM)/molecular mechanical (MM) restricted Hartree–Fock (RHF)/4-31G replica path optimization of the chorismate-to-prephenate rearrangement catalyzed by chorismate mutase (*Bacillus subtilis*). Activation ($\Delta H^{\dagger\dagger}$) and reaction ($\Delta H_{rxn}$) enthalpies (kcal/mol) were computed, using various density functional theory (DFT) methods and basis sets, from the optimized pathway and are compared to experiment and other theoretical work when possible

Method of energetic analysis	Level of theory	Basis set	$\Delta H^{\dagger\dagger}$	ΔH_{rxn}
QM/MM repeated walk[a]	RHF	4-31G	20.1	−20.1
QM/MM replica path[b]	RHF	4-31G	33.4	−15.3
QM/MM replica path[b]	B3LYP	TZVP	14.6	−19.2
QM/MM replica path[b]	B3LYP	6-31G(d)	14.9	−19.5
QM/MM replica path[b]	B3LYP	6-31G(d,p)	14.8	−19.4
QM/MM replica path[b]	B3LYP	6-311G(d,p)	14.4	−20.0
QM/MM replica path[b]	B3LYP	6-311++G(d,p)	14.5	−19.3
QM/MM replica path[b]	BLYP	6-311++G(d,p)	9.5	−16.9
QM/MM replica path[b]	BP86	6-311G(d,p)	9.1	−20.1
QM/MM potential of mean force[b]	B3LYP	6-31G(d)	14.2	−20.6
Gas-phase QM[20]	AM1		42.30	
Gas-phase QM[20]	RHF	6-31G(d)	48.44	
Gas-phase QM[20]	MP2	6-31G(d)	19.42	
Gas-phase QM[c]	B3LYP	6-31G(d)	19.30	−23.26
QM/MM reaction coordinate[20]	AM1		17.80	
Experimental[11]			12.7 ± 0.4	

[a] Optimized via repeated walk procedure at the RHF/4-31G level of theory
[b] Optimized via replica path procedure at the RHF/4-31G level of theory
[c] Gas-phase DFT energy profile, computed from the RHF/4-31G optimized replica path

stepping along it performing a QM/MM minimization at each point. The reaction pathway, in this case, was defined using reaction coordinate distance constraints of the chorismate-to-prephenate rearrangement. This procedure is not only expensive, requiring thousands of QM/MM minimization steps, but also limits variation in the reaction pathway during minimization.

It is clear that the repeated walk procedure, which is subject to hysteresis problems owing to its dependence on the total energy of the system, is limited by its very nature. These problems are only overcome by performing multiple back and forth iterations along the constrained pathway, thus adding expense to the procedure. In contrast, the replica path method, which is symmetric with respect to the path direction, only depends on the forces of each point along the pathway and thus avoids the energy-dependent hysteresis problems observed in the repeated walk procedure.

Nevertheless, using the repeated walk QM/MM procedure, the chorismate mutase catalyzed Clasien rearrangement was examined via a distance-constrained 21-step pathway. This investigation was performed, in a manner consistent with prior QM/MM studies [39], at the restarted Hartree–Fock (RHF)/4-31G level of theory for the quantum region (24 atoms) (Gamess-US) [40] and the MSI all-atom-parameter set [41] for the classically treated portion (in CHARMM). The reaction (ΔH_{rxn}) and activation ($\Delta H^{\dagger\dagger}$) enthalpies were computed to be −20.1 and 20.1 kcal/mol, respectively (Fig. 4). These results are of similar quality to that of other QM/MM studies of the chorismate mutase system. However, given the experimental activation energy of 12.7± 0.4 kcal/mol [11], it is clear that the RHF/4-31G level of theory is insufficient when quantitative QM/MM results are desired. The repeated walk procedure was performed before the replica path method was extended to treat QM/MM reactions and the former method therefore provides a useful comparison with the current work.

In the replica path study we began with the identical end points (reactant and product) used previously, and an additional 2903 water molecules [27, 28] were added to fill a large rhombododecahedron. The replicated region included the substrate and all protein residues or water molecules that had any atom within 6 Å of either the chorismate in the reactant or prephenate in the product. Atoms outside this region were not replicated and started from the same positions for all steps of the reaction pathway. This exclusion was performed to limit the total number of atoms replicated and to provide a more consistent potential energy for the pathway steps. Within the replicated region, only the chorismate/prephenate atoms were selected to be treated quantum mechanically. Note that through exclusion each replicate was generated in a consistent manner as to include the same number of amino acid and water molecules throughout the entire path. Mass-weighted restraints were applied to all QM atoms; however, the restraint weights were doubled for the five atoms that most directly participated in formation of the second ring in the chorismate/prephenate transition state (B, Fig. 1). A total of 21 replicates were then created and a reaction path defined via linear interpolation of Cartesian coordinates from the reactant replicate to the product replicate. Again the RHF/4-31G level of theory was employed to describe the quantum region (24 atoms) during the pathway minimization procedure; however, the updated CHARMM27 [25] parameter set was employed to treat the nonquantum region.

Next, 400 steps of ABNR minimization was performed on the linear replicated pathway. After 400 minimization steps the pathway was determined to be significantly optimized with a total pathway of 0.3 kcal/mol/Å. Pathway energetic analysis was then performed using the RHF/4-31G level of theory, which resulted in $\Delta H^{\dagger\dagger}$ and ΔH_{rxn} values of 33.4 and −15.3 kcal/mol, respectively (Table 1). It is clear that this level of theory, which lacks the vital dynamic electron correlation needed to accurately describe the transition state (QM) stabilization via interaction with the MM field, cannot be used to yield quantitative results. It should be noted that the current implementation of the replica path algorithm is

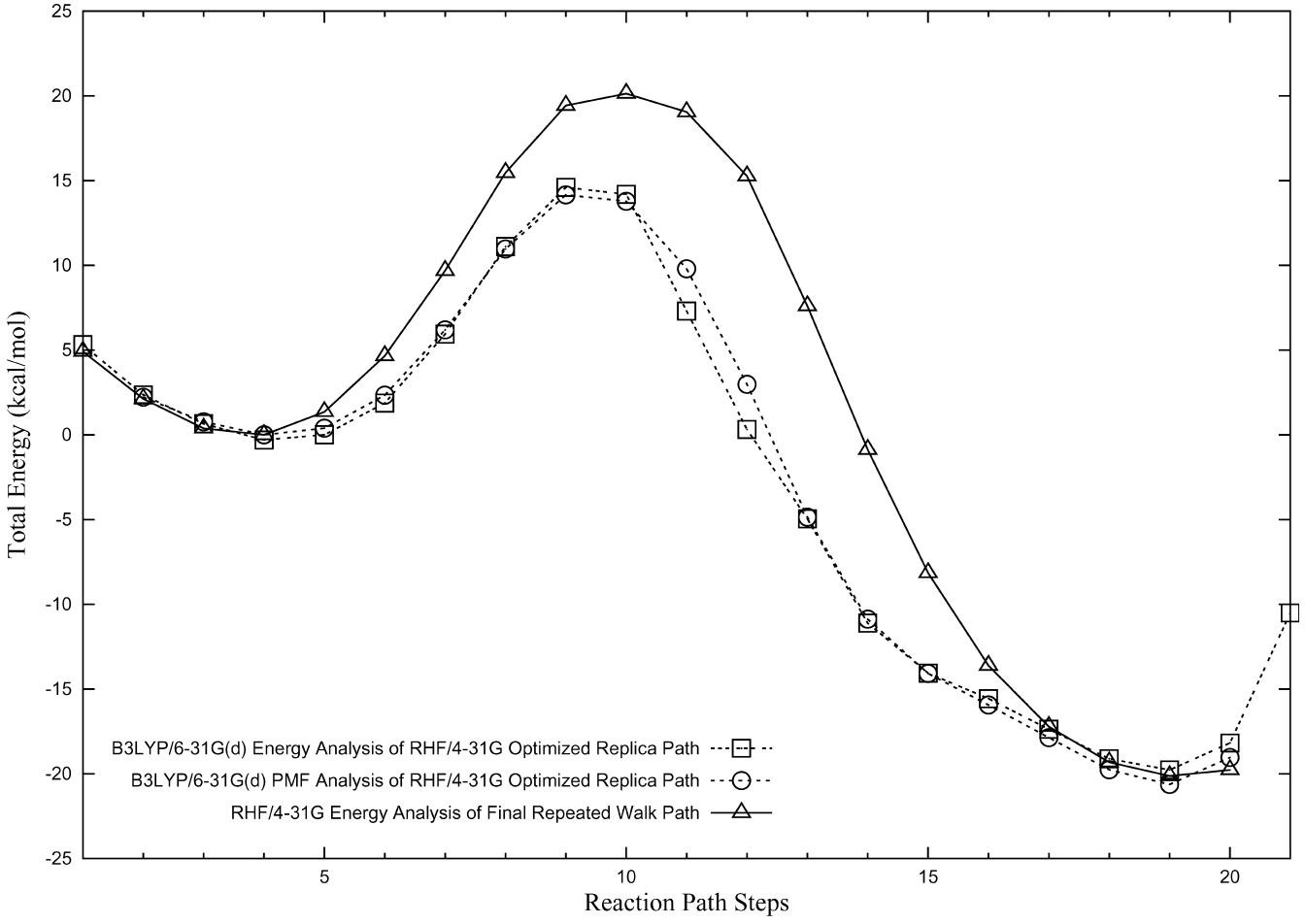

Fig. 4. Energetic profiles (kcal/mol) of chorismate-to-prephenate reaction pathways optimized with both the replica path method as well as the repeated walk procedure. Standard energetic as well as potential of mean force (*PMF*) analysis was performed on the optimized replica path at the B3LYP/6-31G(d) level of theory. The repeated walk pathway was evaluated at the RHF/4-31G level of theory. All energetic profiles were normalized to begin at 0.0 to facilitate comparison. ($\Delta H^{\dagger\dagger}$ [B3LYP/6-31G(d) PMF] = 14.2 kcal/mol, $\Delta H^{\dagger\dagger}$ [B3LYP/6-31G(d) energy analysis] = 14.9 kcal/mol, $\Delta H^{\dagger\dagger}$(RHF/4-31G repeated walk) = 20.1 kcal/mol)

not designed for precise transition state searching. However, given the reference path obtained in the current study further transition searches may be performed via alternative replica path procedures. We then applied, the assumed more accurate, B3LYP/6-31G(d) level of theory to determine energetics of our reference path. This resulted in $\Delta H^{\dagger\dagger}$ and ΔH_{rxn} of 14.9 and −19.5 kcal/mol for a traditional energetic analysis and 14.2 and −20.6 kcal/mol for the PMF analysis, respectively (Table 1, Figure 4). Here it must be stated that DFT is known to overstabilize transition states thus artificially lowering activation energies. This extends to the, force dependent, PMF procedure and care must be taken not to overinterpret the accuracy of DFT energetic results. Therefore, it is important to extend the current replica path scheme to include MP2 QM methods which should help us to evaluate this problem in the future.

Optimized structures for the replicated path were also obtained (Table 2). These structures show smooth transition of important geometrical parameters as the Claisen rearrangement procedes from product to reactant due to the replica path angle restraints. From these structures and their energetic profile additional information may be obtained about the transition-state analogue and how it may interact with various active-site side chains (Arg90 and Glu78).

4 Conclusions

In this work we demonstrate that the newly developed and implemented highly parallel hybrid QM/MM, replica path method is an efficient and effective tool for the examination of both biomolecular and chemical reactions. There are four features of this method that are highly beneficial in the study of large biochemical reaction pathways. First, a subset of atoms may be replicated and varied to assist in pathway descriptions. Second, the use of rmsd best-fit cartesian coordinate restraints may be applied to a subset of the replicated atoms in pathway minimizations to maintain similar size and shape of adjacent points. Third, an independent subset of replicated atoms can be selected and treated with *ab initio* or DFT levels of theory, allowing accurate pathway energetics and/or forces to be determined. Finally, the entire computational framework has been

Table 2. Variation of important bond lengths (Å) versus step number in the optimized chorismate-to-prephenate replicated reaction pathway. Refer to Fig. 1 for the numbering scheme

Step number	C_3–O_7	C_1–C_9	C_4–O_{12}	Arg90-O_7	GLU78-H_7
1	1.468	3.460	1.434	1.913	1.656
2	1.497	3.356	1.431	1.890	1.628
3	1.503	3.258	1.431	1.870	1.617
4	1.484	3.134	1.432	1.863	1.608
5	1.505	2.994	1.432	1.844	1.595
6	1.562	2.852	1.431	1.828	1.583
7	1.649	2.700	1.427	1.812	1.570
8	1.764	2.542	1.427	1.803	1.565
9	1.905	2.396	1.427	1.795	1.560
10	2.041	2.239	1.424	1.800	1.567
11	2.165	2.071	1.428	1.819	1.572
12	2.281	1.896	1.430	1.822	1.578
13	2.408	1.745	1.430	1.822	1.572
14	2.553	1.620	1.431	1.823	1.560
15	2.720	1.555	1.430	1.825	1.551
16	2.891	1.520	1.425	1.826	1.554
17	3.055	1.526	1.428	1.832	1.544
18	3.209	1.553	1.429	1.834	1.543
19	3.343	1.583	1.246	1.830	1.545
20	3.464	1.609	1.423	1.815	1.555
21	3.561	1.618	1.412	1.790	1.565

implemented in a highly parallel scheme that makes it possible to efficiently use well over 100 processors for QM/MM pathway studies.

The QM/MM replica path procedure was employed to study the chorismate mutase catalyzed rearrangement of chorismate to prephenate. The results were compared to previous theoretical and experimental investigations. Employing RHF and various DFT levels of theory we demonstrated that for the chorismate mutase test case, the replica path method is capable of yielding good approximate or refernce pathways that give qualitative insights into reaction mechanisms or can be further refined to give accurate transition state analogs and activation energies. This procedure can take as few a 3–4 days running in parallel on dual processor pentium III 866 MHz machines.

Although DFT does offer some vital static and dynamic electron correlation, which are no doubt important when trying to accurately describe interactions in the active site of proteins, care must be taken when transition states are examined. This is due to a known problem with delocalized bonding patterns and gradient corrected functionals in which subsequent mistreatment of DFT exchange energy occurs.

For systems where larger portions of the protein must be replicated to accurately describe reaction pathways, the potential energy examined uncritically becomes a meaningless quantity. For example, if a distant water molecule rotates during pathway minimization breaking a hydrogen bond the potential energy may vary by as much as 3–5 kcal/mol. This renders a traditional energetic analysis suspect. However, a PMF analysis, where only forces are employed, allows useful results to be obtained. This is due to the fact that forces generated by overall energetic changes that occur far from the active site may have little effect on the reaction of interest (i.e. breaking a hydrogen bond or rotating a water molecule 10–15 Å from the active site of a protein may significantly change the total energy while having little effect on the reaction being catalyzed). Thus, a PMF analysis will on average yield more meaningful results than the previously employed energy analysis, as illustrated in the current work.

Although, this is an initial investigation, we are encouraged by the ability to efficiently generate good QM/MM reference pathways via the replica path method. These reference or approximate pathways are ideally suited for further studies with other new methods we are currently developing which can more accurately determine transition state analog structures and thus activation and reaction enthalpies and free energies. Therefore, we feel that additional studies applying the replica path method will yield similarly encouraging results and are thus being initiated.

The repeated pathway walk (forced transition) procedure has long been used as a valid method for the optimization of reaction pathways. However, the replica path method coupled with the PMF analysis clearly exhibits the potential to resolve weaknesses in current methods and perhaps supercede them as the primary tool used to examine reaction pathways and mechanisms.

Acknowledgements. The authors thank Eric Billings, Xiongwu Wu, and Stephen Bogusz for helpful discussions and related work. The authors also show grateful appreciation to The National Institutes of Health and The National Science Foundation for support of the current research.

References

1. Fukui K (1981) Acc Chem Res 14: 363
2. Gonazles C, Schlegel HB (1990) J Chem Phys 94: 5523
3. Ayala PY, Schlegel HB (1997) J Chem Phys 107: 375
4. Elber R, Karplus M (1987) Chem Phys Lett 139: 375
5. Czerminski R, Elber R (1990) Int J Quantum Chem Symp 24: 167
6. Czerminski R, Elber R (1990) J Phys Chem 92: 5580
7. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4: 187
8. Ganem B (1996) Angew Chem Int Ed Engl 35: 937
9. Lee AY, Stewart JD, Clardy J, Ganem B (1995) Chem Biol 2: 195
10. Kast P, Asif-Ullah M, Hilvert D (1996) Tetrahedron Lett 37: 2691
11. Chook YM, Gray JV, Ke H, Lipscomb WN (1990) J Mol Biol 29: 476
12. Gray JV, Eren D, Knowles JR (1990) J Am Chem 29: 8872
13. Eurenius KP, Chatfield DC, Hodoscek M, Brooks BR (1996) Int J Quantum Chem 60: 1189
14. Zhang Y, Lee T, Yang W (1999) J Chem Phys 110: 46
15. Murphy RB, Philipp DM, Friesner RA (2000) J Comput Chem 21: 1442
16. Andrews PR, Smith GD, Young IG (1973) Biochemistry 12: 3492
17. Haslam E (1993) Shikimic acid: metabolism and metabolites. Wiley, New York
18. Wiest O, Houk KN (1995) J Am Chem Soc 117: 11628
19. Lyne PD, Mulholland AJ, Richards WG (1995) J Am Chem Soc 117: 11345
20. Dewar MJS, Zoebisch EG, Healy EA, Stewart JJP (1985) J Am Chem Soc 107: 3902

21. Marti S, Andres J, Moliner V, Silla E, Tunon I, Bertran J, Field MJ (2001) J Am Chem Soc 123: 1709
22. Gorich H (1978) Biochemistry 17: 3700
23. Cload ST, Liu DR, Pastor RM, Schultz PG (1996) J Am Chem Soc 118: 1787
24. Kast P, Asif-Ullah M, Jiang N, Hilvert D (1996) Proc Natl Acad Sci 93: 5043
25. MacKerell AD et al. (1998) J Phys Chem B 102: 3586
26. Guest MF, van Lenthe JH, Kendrick J, Sherwood P (GAMESS-UK is a package of ab initio programs) with contributions from, Amos RD, Buenker RJ, van Dam H, Dupuis M, Handy NC, Hillier IH, Knowles PJ, Bonacic-Koutecky V, von Niessen W, Harrison RJ, Rendell AP, Saunders VR, Schoffel K, Stone AJ, Tozer D
27. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79: 926
28. Durell SR, Brooks BR, Ben-Naim A (1994) J Phys Chem 98: 2198
29. Becke AD (1993) J Chem Phys 98: 5648
30. Lee C, Yang W, Parr RG (1988) Phys Rev B 37: 785
31. Becke AD (1988) Phys Rev A 38: 3098
32. Perdew JP (1986) Phys Rev B 33: 7046
33. Intcor.doc http://www.lobos.nih.gov/Charmm
34. Replica.doc http://www.lobos.nih.gov/Charmm
35. Beowulf http://www.beowulf.org
36. (a) Gao J; (b) Merz KM, Stanton RV (1998) The encyclopedia of computational chemistry, Wiley, New York
37. Das D, Eurenius KP, Sherwood P, Chatfield DC, Hodoscek M, Brooks BR (2002) J Chem Phys 117(23): 10534
38. Gao J (1994) J Am Chem Soc 116: 1563
39. Khanjin NA, Snyder JP, Menger FM (1999) J Am Chem Soc 121: 11831
40. Lee YS, Hodoscek M, Brroks BR, Kador PF (1998) Biophys Chem 70: 203
41. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S, Windus TL, Dupuis M, Montgomery JA (1993) J Comput Chem 14: 1347
42. MSI (1992) MSI parameter file for CHARMM, version 22. MSI, San Diego, CA
43. Gritsenko OV, Ensing B, Schipper PRT, Baerends EJ (2000) J Phys Chem A 104: 8558