

Exploring SCC-DFTB Paths for Mapping QM/MM Reaction Mechanisms<sup>†</sup>H. Lee Woodcock,<sup>\*,‡</sup> Milan Hodošček,<sup>§</sup> and Bernard R. Brooks<sup>‡</sup>

Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, and Center for Molecular Modeling, National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia

Received: February 19, 2007; In Final Form: May 12, 2007

A new first-order procedure for locating transition structures (TS) that employs hybrid quantum mechanical/molecular mechanical (QM/MM) potentials has been developed. This new technique (RPATH+RESO) combines the replica path method (RPATH) and standard reaction coordinate driving (RCD) techniques in an approach that both efficiently determines reaction barriers and successfully eliminates two key weaknesses of RCD calculations (i.e., hysteresis/discontinuities in the path and the sequential nature of the RCD procedure). In addition, we have extended CHARMM's QM/MM reaction pathway methods, the RPATH and nudged elastic band (NEB) methods, to incorporate SCC-DFTB wave functions. This newly added functionality has been applied to the chorismate mutase-catalyzed interconversion of chorismate to prephenate, which is a key step in the shikimate pathway of bacteria, fungi, and other higher plants. The RPATH+RESO barrier height ( $\Delta E^\ddagger = 5.7$  kcal/mol) is in good agreement with previous results from full-energy surface mapping studies (Zhang, X.; Zhang, X.; Bruce, T. C. *Biochemistry* 2005, 44, 10443–10448). Full reaction paths were independently mapped with RPATH and NEB methods and showed good agreement with the final transition state from the RPATH+RESO "gold standard" and previous high-level QM/MM transition states (Woodcock, H. L.; Hodošček, M.; Gilbert, T. B.; Gill, P. M. W.; Schaefer, H. F.; Brooks, B. R. *J. Comput. Chem.* 2007, 28, 1485–1502). The SCC-DFTB TS geometry most closely approximates the MP2/6-31+G(d) QM/MM result. However, the barrier height is underestimated and possibly points to an area for improvement in SCC-DFTB parametrization. In addition, the steepest descents (SD) minimizer for the NEB method was modified to uncouple the in-path and off-path degrees of freedom during the minimization, which significantly improved performance. The convergence behavior of the RPATH and NEB was examined for SCC-DFTB wave functions, and it was determined that, in general, both methods converge at about the same rate, although the techniques used for convergence may be different. For instance, RPATH can effectively use the adopted basis Newton–Raphson (ABNR) minimizer, where NEB seems to require a combination of SD and ABNR.

## 1. Introduction

Modeling and characterizing detailed mechanisms of chemical and biochemical (e.g., enzymatically catalyzed) processes is a rapidly expanding area of computational chemistry and biology. Due to the size and complexity of the systems of interest coupled with hardware and software constraints, computational modeling of biological processes has typically been carried out via classical methods (i.e., molecular mechanics, MM). Still, many approximations are needed to compute enzymatic processes at more rigorous levels of theory; for example, ab initio quantum mechanical (QM) methods, which scale as  $N^3$  or worse, where  $N$  is the number of basis functions, are still too expensive to completely describe biological systems.<sup>3</sup> Although much effort has been put into linear-scaling ab initio methodology,<sup>4–8</sup> the use of hybrid methods (e.g., quantum mechanical/molecular mechanical, QM/MM) is the direction in which the field has progressed most rapidly.<sup>9–13</sup>

Despite the advances that have been made in ab initio methodology and the widespread use of AM1,<sup>14</sup> PM3 (and later PM5),<sup>15–17</sup> and MNDO<sup>18–23</sup> methods, efforts to find more accurate and more efficient semiempirical QM methods (SEQM)

have continued. These have largely been driven by the three factors, (1) the need for sampling and characterization of dynamic properties such as free energy ( $\Delta G$ ),<sup>24–27</sup> (2) the desire to study larger and larger systems, which has particularly become important as computational biology has gained in prominence,<sup>28,29</sup> and (3) improved accuracy of SEQM, which has been deemed essential as the failures of SEQM theories have been well documented.<sup>11,30–32</sup> Highlights of the continued efforts to improve SEQM methodology include the following:

1. The addition of orthogonalization corrections into the neglect of diatomic differential overlap (NDDO). These reparameterizations have led to the development of the orthogonalization models (OM1 and OM2), which have been shown to yield better descriptions of conformational properties than typical NDDO approaches.<sup>33</sup>

2. The definition of the pairwise distance-directed Gaussian function (PDDG),<sup>34</sup> which was put forth as a three-pronged approach, (1) a single function should be added to PM3 and MNDO pairwise core repulsion interactions, (2) a reparameterization of the standard SEQM parameters, and (3) a modification to the computation of gaseous atom energy of formations. Results from the PDDG improvements are shown to compare favorably with experiments and density functional theory (DFT) calculations.

<sup>†</sup> Part of the "DFTB Special Section".

<sup>‡</sup> National Institutes of Health.

<sup>§</sup> National Institute of Chemistry.

3. Development of the self-consistent-charge density functional tight binding (SCC-DFTB) method as a second-order expansion of the Kohn–Sham energy function with respect to charge density variations.<sup>35</sup> This approach has been shown to improve transferability of parameters and, in general, to be more accurate than AM1 and PM3 for biological applications.<sup>36</sup> In addition, the cost of the computations are not significantly increased when compared to AM1 and PM3.

A major focus of the work presented here is the extension of CHARMM's currently supported reaction path mapping methods to function with the QM/MM capability of SCC-DFTB. In general, reaction mapping methods fall into three categories.

1. Eigenvector-following methods: The most popular of this class of methods is the intrinsic reaction coordinate (IRC) method.<sup>37,38</sup> The IRC method requires a known transition state and employs internal coordinates to map the steepest descents or minimum energy pathway (MEP) that connects reactants and products. However, these methods are typically not applied to high dimensional systems (i.e., proteins) due to the cost of coordinate transformations (scales as  $N^3$ ) and the ambiguous definition of internal coordinates in very large systems (1000 > degrees of freedom).

2. Forced transition or reaction coordinate driving method (RCD): This method involves predefining a chemically relevant reaction coordinate and constraining/restraining that coordinate as the reaction is gradually “forced” to occur. In a chemical sense, this is a logical way to approach a problem. However, there are three major drawbacks that prevent this from being a panacea. (a) The definition of a reaction coordinate can be problematic in many cases and impossible in some. The choice of reaction coordinate is extremely important! If chosen unwisely, the results of all further work can be significantly biased and, in many cases, worthless. (b) The portions of the system that are not directly involved in the reaction can experience hysteresis effects (i.e., discontinuities). For example, one can easily imagine the breaking of hydrogen bonds far away from the reactive center of an enzymatically catalyzed reaction. These changes would have a rather large effect on the total energy of the system but an insignificant effect on the actual reaction occurring. Currently, this problem is most often resolved by making multiple passes along the reaction coordinate of choice and iterating until self-consistency is reached; this is a rather inefficient process. (c) The uncoupled nature of this approach leaves much to be desired with respect to computational efficiency. Typically, a single point on the pathway is optimized until some convergence criteria is reached and then is propagated and used as the starting point of the next step in the path. This, by its very nature, is sequential and inefficient.

3. Chain-of-replica methods:<sup>39–45</sup> These methods involve discretizing a (bio)chemical reaction by defining points (replicas) along the path as intermediate steps between the reactant and product. A target function (typically energy) for the entire pathway is then minimized, governed by various restraints/constraints (e.g., replica distance, angle). The result, once the pathway is fully relaxed, is a MEP or approximate MEP. The chain-of-replica class methods have two key benefits over reaction coordinate driving. The first is the elimination of bias in choosing a reaction coordinate. For example, the replica path method uses root-mean-square distance (rmsd) changes to define the distance between points. This allows for “global” pathway motion to define a reaction path rather than relying on a combination of distances. The second big advantage is formulation of these methods, which are easily translated to efficient parallel computations using modern Beowulf clusters and even

grid environments. In addition, when coupling these reaction path methods to QM/MM theory, one can make further use of parallel computing by taking advantage of the dual parallelism that exists, where the QM code runs in parallel while sitting on top of the parallel reaction path algorithms. A sub category of this class of methods could be deemed “interpolation methods”. An example of this is the conjugate peak refinement (CPR) technique, as implemented in the TRAVel and TREK modules of CHARMM.<sup>46,47</sup> The CPR method uses an interpolation-based scheme and refines previously guessed pathways via an iterative approach that is able to add points to the pathway to avoid high-energy regions. Although this method removes the predetermination of reaction coordinate, thereby overcoming one weakness of RCD, it does not fully take advantage of the parallel/parallel nature of previously mentioned chain-of-replica methods.

The following section (Methods) briefly reviews the RPATH and NEB methods. In addition, the details of a novel application of reaction coordinate driving (via distance restraints) and the RPATH method will be presented. Section 3 (Discussion) details the application of SCC-DFTB-based QM/MM RPATH and NEB calculations to map the chorismate mutase-catalyzed Claisen rearrangement of chorismate to prephenate. Results are compared to those from the novel RPATH/RCD calculations and analyzed in terms of transition-state geometry,  $\Delta E^\ddagger$ , and methodological convergence. Section 4 presents the Concluding Remarks.

## 2. Computational Methodology

All QM/MM calculations were carried out with the CHARMM (c34a1) program for macromolecular simulations.<sup>46</sup> The QM method used throughout the current work is the self-consistent-charge density functional tight binding method (SCC-DFTB),<sup>35,36</sup> unless otherwise noted. The chorismate mutase reactant and product structures used in the current work were taken from our previous study on this system, which initially employed the ICOM structure.<sup>39,48</sup> All initial chorismate mutase pathways consisted of 21 replicas and were generated via a linear interpolation of Cartesian coordinates to better test the convergence behavior of the RPATH and NEB methods. However, we have developed the CHARMM scripts necessary to generate segmented paths using a guess transition structure, which would likely lead to significant computational savings in most cases (scripts are available upon request). All pathway calculations were converged to at least 0.002 kcal/mol/Å in a total pathway root-mean-squared gradient (GRMS), projected out net translations and rotations, and employed both the MASS and WEIGHT keywords. All NEB calculations used  $K_{\text{rms}} = 1000.0$  kcal/mol/Å<sup>2</sup>,  $K_{\text{max}} = 0.0$  kcal/mol/Å<sup>2</sup>,  $K_{\text{angle}} = 0.0$  kcal/mol/Å<sup>2</sup>, and COSMAX = 0.96 radian. All full RPATH calculations used  $K_{\text{rms}} = 2000.0$  kcal/mol/Å<sup>2</sup>,  $K_{\text{max}} = 2000.0$  kcal/mol/Å<sup>2</sup>,  $r_{\text{max}} = 0.025$  Å,  $K_{\text{angle}} = 100.0$  kcal/mol/Å<sup>2</sup>, and COSMAX = 0.96 radian. We will now briefly review the theoretical basis of the RPATH and NEB methods and introduce the novel RPATH+RESD procedure.

**2.1. Replica Path Method.** The replica path method is an extension of the self-penalty walk method of Elber and co-workers.<sup>49–51</sup> However, the RPATH method has the advantage of allowing users to select only a subsection of atoms that define the pathway of interest.<sup>39</sup> Furthermore, users can treat the entire or a smaller portion of the subsection quantum mechanically with an array of ab initio packages.<sup>2,52–54</sup> For example, the substrate of chorismate mutase and a buffer region of 6 Å was chosen to define our path, with the chorismate/prephenate being defined as the QM region and the remainder

of the system treated as a molecular bath. The RPAth method consists of two major restraint terms. With the distance restraint, a force constant is used to restrain the distances between adjacent pathway points. This effectively keeps the path smooth and evenly spaced

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

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

$$\bar{r} = \sum_{i=1}^N \frac{r_i}{N} \quad (3)$$

where  $\bar{r}$  is the average step length and  $N$  is the number of points along the pathway. The best choice of  $K_{\text{rms}}$  is one that limits the variation of the pathway step to 10–15% and prevents vibrational frequencies of over 3000  $\text{cm}^{-1}$ . With angle restraints, a force constant is used to restrain the angle between adjacent and next adjacent pathway points ( $i, i \pm 1, i \pm 2$ )

$$E_{\text{angle}} = \sum_{i=1}^N \frac{1}{2} K_{\text{angle}} (\text{COSMAX} - \cos(\theta)_i)^2$$

$$\text{COSMAX} > \cos(\theta)_i$$

$$E_{\text{angle}} = 0 \quad \text{COSMAX} < \cos(\theta)_i \quad (4)$$

where  $\theta$  is the deviation from linearity. In general, we have found that choosing a  $K_{\text{angle}}$  value that is too large can significantly damage the performance of the ABNR minimizer, as high-frequency modes can be artificially introduced. We recommend a  $K_{\text{angle}}$  value in the range of 100–500 kcal/mol/Å<sup>2</sup>, depending on the flexibility and desired smoothness of the pathway. COSMAX should typically be in the 0.90–0.98 range, with 0.98 restraining the path to be rather stiff and 0.90 allowing for much more flexibility.

**2.2. Nudged Elastic Band Method.** Recently, a superlinear minimization scheme<sup>40</sup> for the nudged elastic band method (NEB)<sup>55–57</sup> was implemented in CHARMM. The NEB method, which determines minimum energy pathways (MEP), is fundamentally governed by the definition of the force acting on each replica

$$\vec{F}_i^0 = -\nabla\vec{V}(\vec{R}_i)|_{\perp} + \vec{F}_i^S \cdot \hat{\tau}_{\parallel} \hat{\tau}_{\parallel} \quad (5)$$

$$\vec{F}_i^S = k_{i+1}(\vec{R}_{i+1} + \vec{R}_i) - k_i(\vec{R}_i + \vec{R}_{i-1}) \quad (6)$$

where  $k$  is a spring constant,  $\hat{\tau}_{\parallel}$  is the pathway tangent vector, and

$$\nabla\vec{V}(\vec{R}_i)|_{\perp} = \nabla\vec{V}(\vec{R}_i) - \nabla\vec{V}(\vec{R}_i) \cdot \tau_{\parallel} \tau_{\parallel} \quad (7)$$

is defined as the projection of the perpendicular component of  $\nabla\vec{V}(\vec{R}_i)$  and the parallel component of the spring force. The newly introduced minimization algorithm is based upon CHARMM's ABNR minimizer. During the minimization, each ABNR step is performed self-consistently in a user-defined subspace. The new superlinear minimization scheme for NEB has been shown to be much more efficient than the standard method of quenched molecular dynamics minimization. In addition, the NEB method with improved minimizer is also able to take advantage of CHARMM's root-mean-squared (rms) best-fit pathway definitions and employ its flexible weighting options.

This gives the user the ability to account for mobile solvent effects while using a finite number of replicas. Since the NEB implementation is coupled to the replica code, CHARMM's parallel/parallel QM/MM pathway functionality can efficiently be employed to examine bond making and bond breaking processes.<sup>2,40</sup> Overall, application of the improved NEB force projection procedures, flexible pathway definitions (in rms best-fit space), and the superlinear minimization algorithm affords users a framework to aid in the examination of complicated transition processes in biological systems.

**2.3. Replica Path Method + RESDistance.** The RESDistance facility in CHARMM allows the general definition and restraint of a linear combination of distances. This restraint term has been added to the energy function to allow for searching of user-specified reaction coordinates (i.e., reaction coordinate driving, RCD)

$$E_{\text{RESD}} = \frac{1}{E_{\text{val}}} K_{\text{val}} \cdot D_{\text{ref}}^{E_{\text{val}}} \quad (8)$$

$$D_{\text{ref}} = (K_1 R_1^{I_{\text{val}}} + K_2 R_2^{I_{\text{val}}} + \dots + K_n R_n^{I_{\text{val}}}) - R_{\text{val}} \quad (9)$$

where  $E_{\text{val}}$  is the exponent used on the restraint function,  $K_{\text{val}}$  (set to 1500 kcal/mol/Å<sup>2</sup> for the current work) is the force constant used to restrain the distances,  $I_{\text{val}}$  is the exponent on the individual distances (all set to 1.0 for the current work), and  $K_1, K_2, \dots, K_n$  are the scale factors used to assign the linear combination. For example, in the current work, two distances were used to define a linear combination

$$\delta = d_{\text{C}_1-\text{C}_9} - d_{\text{C}_3-\text{O}_7} \quad (10)$$

resulting in a  $D_{\text{ref}}$  defined as

$$D_{\text{ref}} = ((1.0 \cdot d_{\text{C}_1-\text{C}_9}) + (-1.0 \cdot d_{\text{C}_3-\text{O}_7})) - R_{\text{val}} \quad (11)$$

Once an appropriate distance metric is determined, a three-step replica path is constructed, with point 1 being defined as the reactant, point 3 defined as the product, and point 2 being constructed as a guess at the transition structure (TS). Our current experience has shown us that Hammond's Postulate can be utilized to great effect when constructing the initial guess of the TS and can save significant time in the overall pathway minimization. The application of RPAth+RESD is straight forward and proceeds in six steps. These are as follows: (1) set all pathway restraints to be 0.0 (i.e.,  $K_{\text{rms}}$  and  $K_{\text{angle}}$ ) in the RPAth command, making sure to turn on rotation and translation projections; (2) set the initial  $\delta$  value for the guess transition structure; (3) invoke the QM/MM RPAth command as directed in the CHARMM documentation (i.e., replica.doc), taking note that this is a parallel calculation with a requirement being that the number of processors must equal to an integer multiple of the number of pathway points (vide infra) (4) once the initial three-point calculation has converged, print the energy and force of the restraint function toward the end of the output file. At this point, a new  $\delta$  value can be computed using the aforementioned information and Newton's Method

$$\delta_2 = \delta_1 - \frac{E(\delta_1)}{\partial E_1 / \partial \delta_1} \quad (12)$$

(5) adjust the  $\delta$  value and perform a new QM/MM RPAth minimization; and (6) once the second round of minimization is completed, compute a new  $\delta$  using the first-order Newton Method or use the following second-order Newton Method

**TABLE 1: QM/MM Activation ( $\Delta E^\ddagger$ ) Energies and Transition-State Geometrical Parameters (See Figure 2) for the Chorismate Mutase-Catalyzed Claisen Rearrangement<sup>a</sup>**

pathway method	minimization details	$\Delta E^\ddagger$	$d_{C_1-C_9}$	$d_{C_3-O_7}$
RPAth	1000 ABNR	5.8	2.42	1.81
RPAth	1000 SD	7.0	2.38	1.91
RPAth	500 SD + 500 ABNR	5.8	2.43	1.81
RPAth	500 SD + 500 SD	6.6	2.39	1.90
RPAth	750 ABNR	5.8	2.42	1.81
RPAth	750 SD	6.9	2.37	1.90
RPAth	250 SD + 500 ABNR	5.9	2.42	1.81
RPAth	500 ABNR	6.0	2.31	1.92
RPAth	500 SD	7.2	2.40	1.90
NEB	1000 ABNR	7.5	2.45	1.95
NEB	1000 SD	6.3	2.37	1.92
NEB	500 SD + 500 ABNR	6.1	2.33	1.88
NEB	500 SD + 500 CIMG	6.3	2.37	1.88
NEB	500 SD + 500 SD	6.3	2.32	1.88
NEB	750 ABNR	7.6	2.45	1.95
NEB	750 SD	6.3	2.42	1.83
NEB	250 SD + 500 ABNR	6.0	2.36	1.84
NEB	500 SD + 250 CIMG	6.5	2.38	1.89
NEB	500 ABNR	7.6	2.45	1.95
NEB	500 SD	6.7	2.44	1.86
NEB	250 SD + 250 CIMG	7.3	2.41	1.89
RPAth+RESd	SCC-DFTB/CHARMM	5.7	2.34	1.86
RPAth+RESd <sup>2</sup>	HF/6-31+G(d)/CHARMM	26.2	2.68	2.20
RPAth+RESd <sup>2</sup>	B3LYP/6-31+G(d)/CHARMM	9.0	2.72	2.13
RPAth+RESd <sup>2</sup>	MP2/6-31+G(d)/CHARMM	8.2	2.41	1.92
RCD <sup>1</sup>	SCC-DFTB QM/MM	6.1	2.32	1.90
experimental <sup>48</sup>		$12.7 \pm 0.4$		

<sup>a</sup> All results are SCC-DFTB QM/MM-based unless otherwise noted. All energies are reported in kilocalories per mole, and all distances are in angstroms.

formula that takes into account information from both minimizations and appears to be generally faster at converging to the transition structure

$$\delta_3 = \frac{(\partial E_2 / \partial \delta_2) \cdot \delta_1 - (\partial E_1 / \partial \delta_1) \cdot \delta_2}{\partial E_2 / \partial \delta_2 - \partial E_1 / \partial \delta_1} \quad (13)$$

Given the fact that the restraint energy is fit to a quadratic function (eq 13), this approach should converge nearly quadratically. Although this method retains the weakness of predefining a reaction coordinate, there are two fundamental improvements, (1) hysteresis effects that are commonly encountered with the RCD method are eliminated due to the way RPAth handles interactions not explicitly defined as part of the pathway (see Section 2.1), and (2) the combination of RPAth+RESd eliminates the sequential nature of coordinate driving by utilizing the parallel implementation of the replica path method. If using SCC-DFTB, which is not parallelized, each point (i.e., QM/MM replica calculations) must be run on only one processor (e.g., 3 points = 3 processors), as compared to ab initio packages that can take advantage of the parallel/parallel nature of the replica path method and efficiently employ 24 or more processors to examine a three-step path (i.e.,  $X \cdot n$  processors can be used where  $X = 1, 2, 4, 8, \dots$  and  $n =$  is the number of pathway points, 3 in the case of RPAth+RESd).

### 3. Discussion

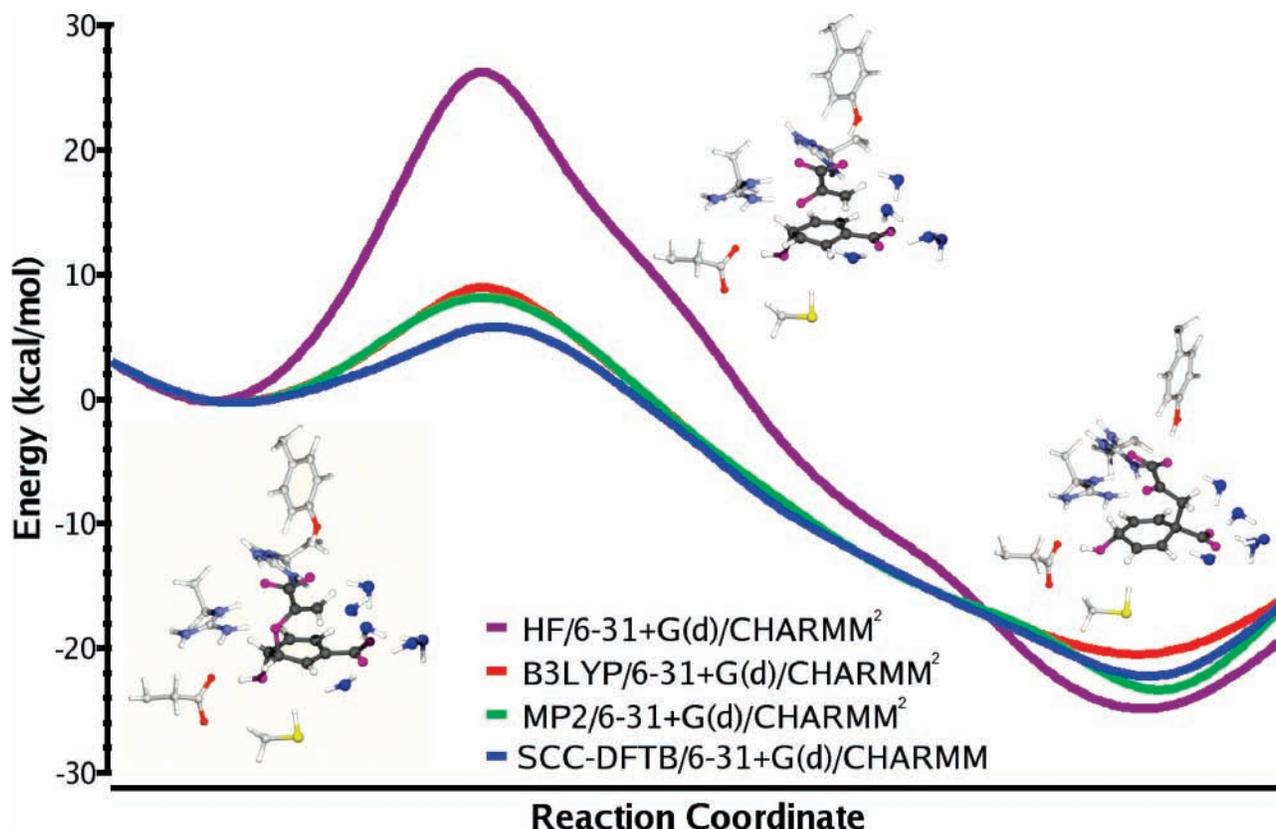
The chorismate mutase-catalyzed Claisen rearrangement has been the focus of numerous experimental and computational studies.<sup>48,58-64</sup> Scientists are interested in this pathway largely because of the role it plays in the shikimate pathway of bacteria, fungi, and other higher plants. This pathway is the first committed step in the biosynthesis of the aromatic amino acids phenylalanine and tyrosine and thus has been suggested as a possible antimicrobial and antifungal target.<sup>65,66</sup> However,

chorismate mutase is also of interest because it is one of the only known enzymes to catalyze a pericyclic rearrangement.<sup>67</sup>

A more recent point of emphasis has been the role this reaction plays in benchmarking QM/MM methodology. As a test case for the current reaction pathway mapping methods, we computed the activation and reaction energies of the *Bacillus subtilis*-catalyzed Claisen rearrangement of chorismate to prephenate. This was done with the RPAth+RESd method, which served as a ‘‘gold standard’’ for which to compare the full RPAth and NEB calculations. We also examined the geometries of the transition-state analogue (TSA), comparing these with the standard from the RPAth+RESd path and previously published results (Table 1).

The pathway was defined to be the substrate (chorismate/prephenate) with a 6 Å buffer region. We next constructed the RPAth+RESd pathway by defining the reactants and products as an initial two-step pathway with no restraints. This allowed us to exhaustively minimize the endpoints and establish a reaction energy ( $\Delta E$ ) of approximately  $-22.0$  kcal/mol, which was in good agreement with both B3LYP and MP2 QM/MM results of  $-19.5$  and  $-23.1$  kcal/mol, respectively (Figure 1). The endpoints obtained from this calculation were used to generate all paths via linear interpolation of Cartesian coordinates and were fixed at their minimized geometries for all additional calculations. All fully mapped pathways consisted of the reaction being discretized into 21 steps (each QM/MM calculation was run on a separate processor of LoBoS, <http://www.lobos.nih.gov>), which is consistent with our previous study.<sup>39</sup>

We next began the RPAth+RESd procedure by adding an additional point (i.e., the transition structure guess) to our initial two-step pathway and using  $\delta = 0.59$  Å, the B3LYP/6-31+G(d)/CHARMM result previously determined.<sup>2</sup> The initial  $\delta$  value turned out to be nearly 20% larger than the final SCC-DFTB result and was a good test for convergence of the procedure.



**Figure 1.** Illustrations of reaction energetics computed with RPATH+RESD.

The first-order newton approximation was employed to generate a guess for  $\delta_2$  (eq 12), and the pathway was reoptimized. Using the restraint forces from the first two steps of the optimization procedure, we computed  $\delta_3$  from the second-order Newton approximation (eq 13). The final pathway was solved in four iterations, with each successive QM/MM calculation minimizing more quickly. The optimized path yielded  $\delta = 0.49$  Å ( $d_{C_1-C_9} = 2.34$  Å and  $d_{C_3-O_7} = 1.81$  Å) and a  $\Delta E^\ddagger$  of 5.7 kcal/mol. Although this procedure is iterative and still somewhat sequential, the overall computational time required to find the correct transition structure is reduced dramatically compared to performing a full RCD procedure.

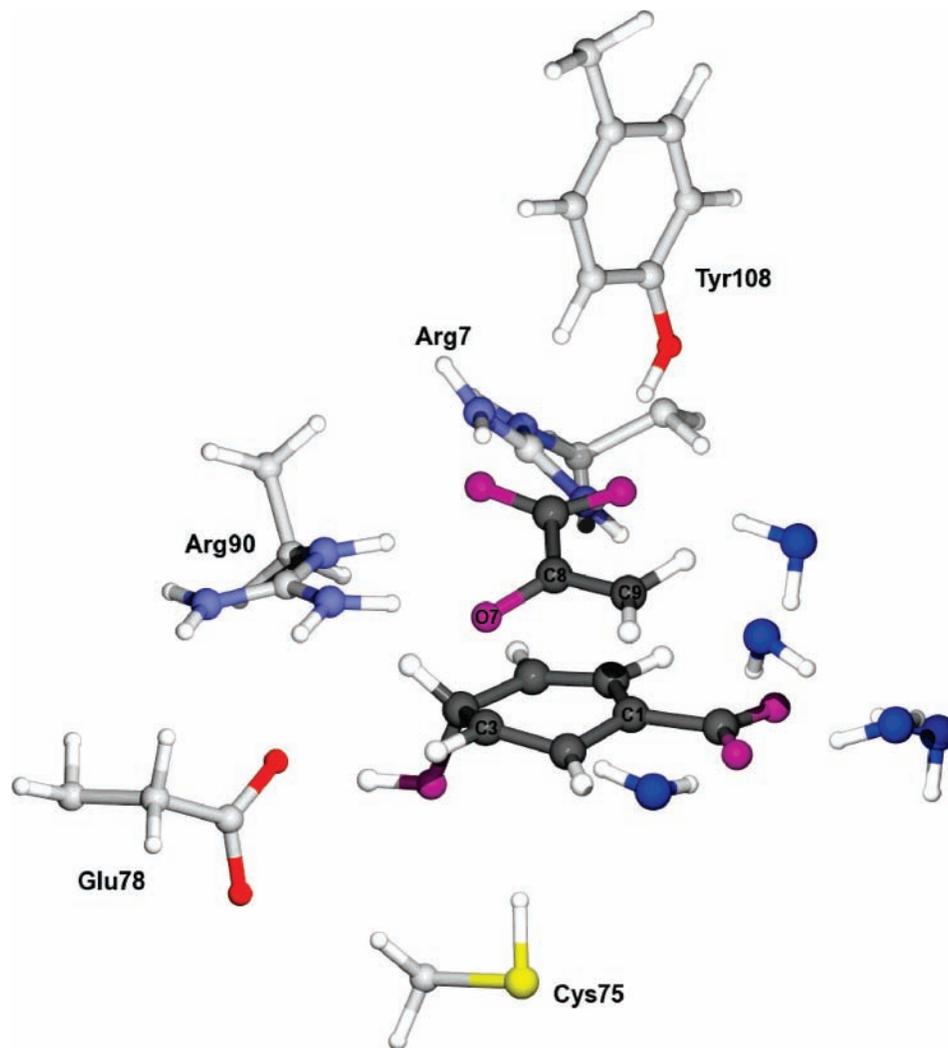
We also examined key geometrical features of the active site. This was done by analyzing four hydrogen-bonding distances that have been determined to be critical for the transition-state stabilization, one from Arg7 and three from Arg90 (Figure 2).<sup>68</sup> Using these four distances, we computed root-mean-square deviations (rmsd) between the SCC-DFTB-optimized transition structure and the HF, B3LYP, and MP2 QM/MM results. This resulted in variations of 0.06, 0.04, and 0.03 Å, respectively. Although these differences are small, it is clear that SCC-DFTB is doing a good job of reproducing the MP2/6-31+G(d) QM/MM active site geometry, with the HF/6-31+G(d) result being twice the rmsd.

Results were compared to a recent study by Zhang, Zhang, and Bruce, which mapped out the full SCC-DFTB QM/MM surface using the RCD method and arrived at a final  $\Delta E^\ddagger$  of 6.1 kcal/mol and  $d_{C_1-O_9} = 2.32$  Å and  $d_{C_3-O_7} = 1.90$  Å, which are in good agreement with the RPATH+RESD result.<sup>1</sup> The slight differences in the barrier between the Zhang et al. study and the current work can likely be attributed to small variations in the reactant structure ( $d_{C_1-O_9} = 2.80$  Å and  $d_{C_3-O_7} = 1.53$  Å). Full reactant and TS active site comparisons could not be

carried out with the geometrical parameters listed in their published work.

We considered the role of ARG 63 in the catalytic process. We specifically examined if the reaction was facilitated by direct hydrogen bonding of ARG 63 to the substrate. ARG 63 has been suggested to play an important role<sup>1,27,68</sup> or an ambiguous role in catalysis.<sup>69</sup> Note that in published crystal structures (1COM,2CHT),<sup>48,70</sup> ARG 63 is bound to the substrate in only  $\sim 1/3$  of the resolved active sites. In the remaining, it is solvent exposed. Ignoring crystal packing effects, this ratio of active site conformations corresponds to an approximately 1 kcal/mol free-energy penalty for burying the ARG 63 in the active site. Modeling of ARG 63 bound to the substrate is illustrated in Figure 3. The bound state was modeled in addition to the unbound ARG 63 state (Figure 2), where the residue was solvent exposed. Mapping both reaction paths with the RPATH+RESD resulted in nearly identical barrier heights, with  $\Delta E^\ddagger$  of the bound ARG 63 state being 6.0 kcal/mol as compared to 5.7 kcal/mol in the unbound case, both of which are in agreement with the Zhang et al. study of 6.1 kcal/mol. Although hydrogen bonding ARG 63 directly to the substrate is enthalpically favored (here by roughly 12 kcal/mol), there is little evidence that this residue plays an important catalytic role since all three states (reactant, transition state, and product) are similarly effected. Without additional evidence that there is a free-energy penalty for ARG 63 to be solvent exposed during the reaction, we cannot conclude that ARG 63 plays an important catalytic role.

We employed the adopted basis Newton–Raphson (ABNR), the steepest descents (SD), and the NEB climbing image (CIMG) minimizers that are available in CHARMM to systematically explore the chorismate mutase-catalyzed reaction. For all RPATH and NEB full pathway calculations, we applied standard mass weights to the defined path (the substrate and



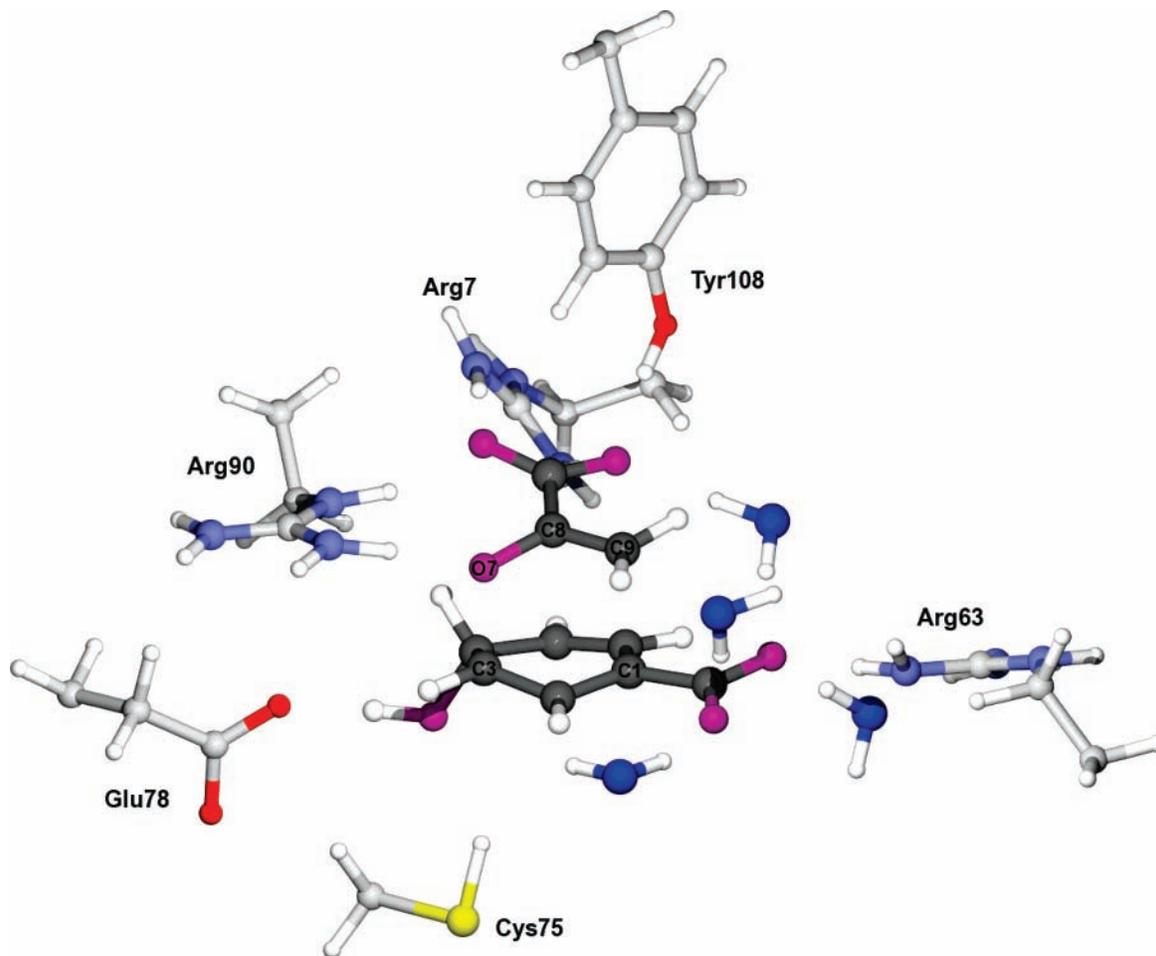
**Figure 2.** Active site amino acid residues of chorismate mutase showing the transition structure for the interconversion of chorismate to prephenate. Labeled atoms were double-mass-weighted in all RPATH and NEB calculations.

buffer region) and double mass weights to the five atoms participating most critically in the chemical reaction (C1, C9, C3, O7, and C8; Figure 2). The RPATH method was employed with the ABNR and SD minimizers and combinations of the two. For the RPATH method, we observed very little benefit from using the SD minimizer. For example, the 250 SD + 500 ABNR minimization yielded nearly identical results, in terms of barrier and geometry, as those from employing ABNR alone for 750 steps. Although these results show SD minimizations as being unnecessary, it is likely that at least a few steps will help to remove bad contacts and speed overall convergence. This could be especially useful if the initial guess path is poorly constructed. Using the same four distance metrics as done previously, we looked at the rmsd's of the RPATH results. Minimization employing the RPATH method for 1000 steps with ABNR yielded rmsd's of 0.01 Å with respect to the RPATH+RESD "standard" and 0.07, 0.04, and 0.03 Å with respect to HF, B3LYP, and MP2 QM/MM.

In contrast to the SD behavior in the RPATH test case, we observed significant benefit from employing the SD minimizer when using the NEB method. The original NEB SD minimizer had both the parallel and perpendicular forces coupled and minimized both simultaneously. This, however, could present a problem if one degree of freedom is stiff (e.g., in path forces or distance between replicas) and is thus minimized very quickly and the other is more floppy (e.g., off-path forces like angular

dependence). For the current work, we modified the NEB SD routines in CHARMM to uncouple these degrees of freedom and minimize each independently. This resulted in significant improvements in minimizing a NEB path with SD. For example, comparing the 750 ABNR and 750 SD minimizations, it is clear that the SD result is closer to the RPATH+RESD result and to the previously published work of Bruice and co-workers. For example, the NEB paths minimized with 1000 and 750 steps of SD both yielded rmsd's of 0.02 Å compared to the "gold standard"; however, the NEB path that was minimized with a combination of 500 SD steps followed by 500 ABNR steps resulted in a slightly lower rmsd and a  $\Delta E^\ddagger$  in better agreement with RPATH+RESD. In addition, we examined the NEB rmsd's compared to ab initio QM/MM results and again found that the 500 SD + 500 ABNR is in similar agreement to the RPATH results, rmsd = 0.06, 0.04, and 0.03 for HF, B3LYP, and MP2 QM/MM, respectively.

At first, this may appear to be a failing of NEB with ABNR, but it is more likely that the ABNR procedure, which involves numerous projections and approximated Hessian matrices in a subspace, is adversely effected by these approximations and possibly the noise associated with more approximate semi-empirical levels of theory. Clearly, this needs to be investigated further employing both semiempirical and ab initio QM levels



**Figure 3.** Active site amino acid residues of chorismate mutase showing the transition structure (with ARG 63 bound) for the interconversion of chorismate to prephenate. Labeled atoms were double-mass-weighted in all RPATH and NEB calculations.

of theory. However, this is an ongoing area of research, and the current work will serve as a guide and benchmark for future studies.

Another interesting result is that the stationary point at the chorismate mutase barrier appears to be relatively flat. For example, changes in either the RPATH- or NEB-minimized transition states of 0.1 Å in either or both key bonds (the one being made, and/or the one being broken) has a negligible effect on  $\Delta E^\ddagger$ . This is not an artifact of the methods being used to examine the reaction paths, as two-dimensional energy surfaces, mapped out at the SCC-DFTB level, confirm these results.<sup>1</sup>

In general, both NEB and RPATH appear to converge at nearly the same rate; however, the path to convergence appears to be different. In the case of RPATH, the ABNR minimizer alone appears to give very good results and is essentially converged after 750 steps of minimization. On the other hand, the NEB method appears to be a bit more fickle than RPATH, a behavior that has been noted in the past when using the ABNR minimizer,<sup>71</sup> but nonetheless appears to converge when using a combination of SD and ABNR minimizations. Also of interest is the fact that the NEB ABNR minimizations are nearly double the cost of the RPATH minimizations (even at the same level of theory). This is due to the projections that occur in the NEB and ABNR routines to improve pathway description and improve the target function of the NEB. This performance issue is, of course, not the case when using NEB to perform ab initio QM/MM pathway calculations, as the cost of the projections, which are a larger part of the total computational time when running fast semiempirical levels of theory, is dwarfed by the

ab initio QM portion of the job. Overall, the pathway mapping results are somewhat unexpected, as tight binding DFT mirrored the MP2 QM/MM results more closely than B3LYP, which is actually a basis for the development of SCC-DFTB. More detailed benchmarking studies of SCC-DFTB in comparison to MP2 and pure, gradient-corrected, and meta DFT functionals are clearly needed and will be addressed in the future.

#### 4. Concluding Remarks

The current work introduces a rapid, new first-order procedure for locating transition states within a hybrid quantum mechanical/molecular mechanical framework and the extension of CHARMM's replica path and nudged elastic band methods to employ SCC-DFTB wave functions in pure QM and hybrid QM/MM reaction path calculations. The new procedure (RPATH+RESD), which combines RPATH with the reaction coordinate driving technique via CHARMM's RESDistance facility, helps to eliminate two key weaknesses of the standard RCD approach. Using the RPATH+RESD technique, we examined the Claisen rearrangement catalyzed by chorismate mutase and were able to locate the transition state very efficiently and compare this to results from previous studies. We found that the barrier computed at the SCC-DFTB QM/MM RPATH+RESD level of theory agreed well with a full-energy surface mapped out by Bruce and co-workers.

In addition, we compared the SCC-DFTB geometrical parameters with those computed at higher levels of QM/MM theory (i.e., HF, B3LYP, and MP2) and found that SCC-DFTB

most closely mimics the MP2 results. This analysis consisted of comparing  $C_1-C_9$  and  $C_3-O_7$  geometrical parameters and performing rmsd calculations on four key active site distances that are known to stabilize the transition state. It was also determined that the energy surface at the transition states is quite flat, as a 0.1 Å change in the two key distances (bond being made and bond being broken) resulted in virtually no change in the reaction barrier.

After determining the “gold standard” SCC-DFTB transition state, computed with the RPATH+RESD procedure, we mapped the full pathway using both the RPATH and the NEB methods, again employing SCC-DFTB QM/MM. Running RPATH and NEB calculations with multiple minimization algorithms allowed for a detailed methodological comparison and gave us the opportunity to examine the most effective strategies for studying reaction paths. In particular, we examined the convergence of RPATH and NEB reaction barriers to the gold standard result and noted that both methods converge at approximately the same rate but may require different approaches to reach that convergence. For example, the RPATH method is able to reproduce the RPATH+RESD result by employing the adopted basis Newton–Raphson minimizer alone for as little as 750 steps. On the other hand, the NEB method appears to need a combination of the steepest descents minimizer and ABNR for optimal performance. For example, performing a NEB minimization with 250 steps of SD followed by 500 steps of ABNR resulted in very good agreement with the gold standard result. It should be noted that the SD minimizer used in the current work was modified to uncouple the in-path and off-path degrees of freedom when being applied to NEB pathways, which significantly enhanced the performance.

This study not only describes a new method for rapidly locating transition states in a QM/MM scheme but also lays the groundwork for further study and improvements to be made to reaction pathway minimization techniques, an area that has already attracted much attention. All methods introduced here were implemented in CHARMM version c34a1 and will be available in upcoming releases.

**Acknowledgment.** We thank the following for helpful insight and fruitful discussions: Qiang Cui, University of Wisconsin; Damian Moran, Macquarie University; and Paul Sherwood and Christine Bailey, Daresbury Laboratory. This research was supported by the Intramural Research Program of the NIH, NHLBI. Use of the LoBoS ([www.lobos.nih.gov](http://www.lobos.nih.gov)) and Biowulf ([biowulf.nih.gov](http://biowulf.nih.gov)) supercomputers is also acknowledged and appreciated.

## References and Notes

- Zhang, X.; Zhang, X.; Bruice, T. C. *Biochemistry* **2005**, *44*, 10443–10448.
- Woodcock, H. L.; Hodošček, M.; Gilbert, T. B.; Gill, P. M. W.; Schaefer, H. F.; Brooks, B. R. *J. Comput. Chem.* **2007**, *28*, 1485–1502.
- Head-Gordon, M. *J. Phys. Chem.* **1996**, *100*, 13213–13225.
- Jung, Y.; Sodt, A.; Gill, P. M. W.; Head-Gordon, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6692–6697.
- White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1994**, *230*, 8–16.
- White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1996**, *253*, 268–278.
- Strain, M. C.; Scuseria, G. E.; Frisch, M. J. *Science* **1996**, *271*, 51–53.
- Lee, T. S.; Lewis, J. P.; Yang, W. T. *Comput. Mater. Sci.* **1998**, *12*, 259–277.
- Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.
- Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718–730.
- Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.
- Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
- Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357–19363.
- Dewar, M. J. S.; Storch, D. M. *J. Am. Chem. Soc.* **1985**, *107*, 3898–3902.
- Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209–220.
- Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221–264.
- Stewart, J. J. P. *MOPAC 2002*; CAChe Group: Portland, 2002.
- Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.
- Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4907–4917.
- Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1992**, *81*, 391–404.
- Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1996**, *93*, 315.
- Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616–626.
- Khandogin, J.; Gregersen, B. A.; Thiel, W.; York, D. M. *J. Phys. Chem. B* **2005**, *109*, 9799–9809.
- Li, G. H.; Zhang, X. D.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 8643–8653.
- Reddy, M. R.; Singh, U. C.; Erion, M. D. *J. Am. Chem. Soc.* **2004**, *126*, 6224–6225.
- Guimaraes, C. R. W.; Repasky, M. P.; Chandrasekhar, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2003**, *125*, 6892–6899.
- Marti, S.; Moliner, V.; Tunon, I.; Williams, I. H. *Org. Biomol. Chem.* **2003**, *1*, 483–487.
- Ermolaeva, M. D.; van der Vaart, A.; Merz, K. M. *J. Phys. Chem. A* **1999**, *103*, 1868–1875.
- Dixon, S. L.; Merz, K. M. *J. Chem. Phys.* **1996**, *104*, 6643–6649.
- Halgren, T. A.; Lipscomb, W. N. *Chem. Phys. Lett.* **1977**, *49*, 225–232.
- Loferer, M. J.; Loeffler, H. H.; Liedl, K. R. *J. Comput. Chem.* **2003**, *24*, 1240–1249.
- Lyne, P. D.; Hodošček, M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 3462–3471.
- Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495–506.
- Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601–1622.
- Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.
- Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569–585.
- Fukui, K. *J. Phys. Chem.* **1970**, *74*, 4161–4165.
- Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154–2161.
- Woodcock, H. L.; Hodošček, M.; Sherwood, P.; Lee, Y. S.; Schaefer, H. F.; Brooks, B. R. *Theor. Chem. Acc.* **2003**, *109*, 140–148.
- Chu, J. W.; Trout, B. L.; Brooks, B. R. *J. Chem. Phys.* **2003**, *119*, 12708–12717.
- Cisneros, G. A.; Liu, H. Y.; Lu, Z. Y.; Yang, W. T. *J. Chem. Phys.* **2005**, *122*, 114502–114508.
- Liu, H. Y.; Lu, Z. Y.; Cisneros, G. A.; Yang, W. T. *J. Chem. Phys.* **2004**, *121*, 697–706.
- Xie, L.; Liu, H. Y.; Yang, W. T. *J. Chem. Phys.* **2004**, *120*, 8039–8052.
- Khavrutskii, I. V.; Arora, K.; Brooks, C. L. *J. Chem. Phys.* **2006**, *125*, 174108–174114.
- Khavrutskii, I. V.; Byrd, R. H.; Brooks, C. L. *J. Chem. Phys.* **2006**, *124*, 194903–194916.
- Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- Fischer, S.; Karplus, M. *Chem. Phys. Lett.* **1992**, *194*, 252–261.
- Chook, Y. M.; Gray, J. V.; Ke, H. M.; Lipscomb, W. N. *J. Mol. Biol.* **1994**, *240*, 476–500.
- Czerminski, R.; Elber, R. *Int. J. Quantum Chem. Symp.* **1990**, 167–186.
- Czerminski, R.; Elber, R. *J. Chem. Phys.* **1990**, *92*, 5580–5601.
- Elber, R.; Karplus, M. *Chem. Phys. Lett.* **1987**, *139*, 375–380.
- Guest, M. F.; Bush, I. J.; Van Dam, H. J. J.; Sherwood, P.; Thomas, J. M. H.; Van Lenthe, J. H.; Havenith, R. W. A.; Kendrick, J. *Mol. Phys.* **2005**, *103*, 719–747.
- Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- Shao, Y.; Fusti-Molnar, L.; Jung, Y.; Kusmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O’Neill, D. P.; Distasio, R. A.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Daschel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair,

- N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, D. C.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.
- (55) Mills, G.; Jonsson, H.; Schenter, G. *Surf. Sci.* **1995**, *324*, 305–337.
- (56) Henkelman, G.; Jonsson, H. *J. Chem. Phys.* **2000**, *113*, 9978–9985.
- (57) Henkelman, G.; Uberuaga, B.; Jonsson, H. *J. Chem. Phys.* **2000**, *113*, 9901–9904.
- (58) Lee, A. Y.; Stewart, J. D.; Clardy, J.; Ganem, B. *Chem. Biol.* **1995**, *2*, 195–203.
- (59) Kast, P.; Asif-Ullah, M.; Hilvert, D. *Tetrahedron Lett.* **1996**, *37*, 2691–2694.
- (60) Ganem, B. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 937–945.
- (61) Galopin, C. C.; Zhang, S.; Wilson, D. B.; Ganem, B. *Tetrahedron Lett.* **1996**, *37*, 8675–8678.
- (62) Warshel, A. *Ann. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425–443.
- (63) Mulholland, A. J. *Drug Discovery Today* **2005**, *10*, 1393–1402.
- (64) Marti, S.; Roca, M.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J. *Chem. Soc. Rev.* **2004**, *33*, 98–107.
- (65) Haslam, E. *Shikimic Acid: Metabolism and Metabolites*; Wiley: New York, 1993.
- (66) Andrews, P. R.; Smith, G. D.; Young, I. G. *Biochemistry* **1973**, *12*, 3492–3498.
- (67) Knaggs, A. R. *Nat. Prod. Rep.* **2001**, *18*, 334–355.
- (68) Lee, Y. S.; Worthington, S. E.; Krauss, M.; Brooks, B. R. *J. Phys. Chem. B* **2002**, *106*, 12059–12065.
- (69) Ranaghan, K. E.; Ridder, L.; Szeferczyk, B.; Sokalski, W. A.; Hermann, J. C.; Mulholland, A. J. *Mol. Phys.* **2003**, *101*, 2695–2714.
- (70) Chook, Y. M.; Ke, H.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *90*, 8600–8603.
- (71) Bailey, C.; Sherwood, P. **2006**, personal communication.