JCTC Journal of Chemical Theory and Computation

Calculation of a Complete Enzymic Reaction Surface: Reaction and Activation Free Energies for Hydride-Ion Transfer in Dihydrofolate Reductase

Peter L. Cummins, Ivan V. Rostov, and Jill E. Gready*

Computational Proteomics Group, John Curtin School of Medical Research, Australian National University, P.O. Box 334, Canberra ACT 2601, Australia

Received October 25, 2006

Abstract: We present a two-dimensional grid method for the calculation of complete free-energy surfaces for enzyme reactions using a hybrid quantum mechanical/molecular mechanical (QM/MM) potential within the semiempirical (PM3) QM approximation. This implementation is novel in that parallel processing with multiple trajectories (replica-exchange molecular dynamics simulations) is used to sample configuration space. The free energies at each grid point are computed using the thermodynamic integration formalism. From the free-energy surface, the minimum free-energy pathway for the reaction can be defined, and the computed activation and reaction energies can be compared with experimental values. We illustrate its use in a study of the hydride-transfer step in the reduction of dihydrofolate to tetrahydrofolate catalyzed by *Escherichia coli* dihydrofolate reductase with bound nicotinamide adenine dinucleotide phosphate cofactor. We investigated the effects of changing the QM region, ionization state of the conserved active-site Asp27 residue, and polarization contributions to the activation and reaction free energy. The results clearly show the necessity for including the complete substrate and cofactor molecules in the QM region, and the importance of the overall protein (MM) electrostatic environment in determining the free energy of the transition state (TS) and products relative to reactants. For the model with ionized Asp27, its inclusion in the QM region is essential. We found that the reported [Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *J. Mol. Biol.* **2003**, *327*, 549] stabilization of the TS due to polarization is an artifact that can be attributed to truncation of the electrostatic interactions between the QM and MM atoms. For neutral (protonated) Asp27, our calculated reaction free energy of -4 ± 2 kcal/mol agrees well with the experimental value of -4.4 kcal/mol, although the corresponding activation free-energy estimate is still high at 21 $\pm$ 2 kcal/mol compared with the experimental value of 13.4 kcal/mol. The results are less supportive of the ionized Asp27 model, which gives rise to a much higher activation barrier and favors the reverse reaction.

Introduction

Dihydrofolate reductase (DHFR) is a key enzyme[1] that catalyzes the reduction of folate to dihydrofolate (DHF) and DHF to tetrahydrofolate (THF), cycling the THF cofactors required for the synthesis of DNA and other precursors. It has been the subject of much theoretical investigation. Aspects of the mechanism that have been studied using computational methods include kinetic isotope effects,[2,3] transition-state (TS) geometry,[4] the ionization state of active-site residues,[5,6] the role of dynamical structural fluctuations in promoting hydride transfer,[7−9] and the effects of electronic polarization of the active site.[10−12] Methods employed vary

* Corresponding author e-mail: Jill.Gready@anu.edu.au.

10.1021/ct600313b CCC: $37.00
Published on Web 03/29/2007

from simple geometry optimization to the use of molecular dynamics (MD), or combinations of both. The ultimate goal of these studies is to understand how DHFR facilitates its reaction. While many general theories have been proposed[13−15] to rationalize the catalytic activity of enzymes, to test these theories using computational methods, and to provide a detailed picture of how a particular enzyme catalyzes a particular reaction, we must first have a sufficiently accurate description of the basic physical and chemical processes involved. Combined quantum mechanical/molecular mechanical (QM/MM) methods have become very popular as a basis for such a description. In the 30 years since the pioneering work of Warshel and Levitt,[16] QM/MM methods have undergone considerable development,[17−31] providing a computationally efficient way of including the effects of solvent and enzyme environments on a reaction center. The total energy of the system, as an explicit function of the reaction coordinate (r), is given by the sum of the QM and MM energies plus a QM/MM interaction term:

$$E(r) = E_{QM}(r) + E_{QM/MM}(r) + E_{MM} \quad (1)$$

As it must allow for the breaking and formation of chemical bonds, the reaction coordinate is contained within the QM region and, as indicated in eq 1, is an explicit function of E_{QM} and $E_{QM/MM}$ only.

However, in order to compute reaction free energies that are directly comparable with experimental quantities, it is necessary to perform appropriate statistical mechanical averaging[32,33] along the predefined reaction coordinate. While ab initio or density functional (DFT) QM methods can yield a more reliable description of reaction energies than the more efficient semiempirical PM3 or AM1 methods, their relatively high computational cost generally prohibits statistical mechanical calculations for enzymic systems.[29,31] Thus, despite the known limitations of semiempirical methods in terms of their overall chemical accuracy, the primary need to compute free-energy changes associated with chemical processes has led to their continued application in studies of enzymic reaction mechanisms, including that of DHFR.[6,7] They are particularly useful for problems where systematic trends or alternative mechanisms are of interest, rather than cases where absolute accuracy in computed reaction free energies is required.

Regardless of the level of QM used—ab initio, DFT, or semiempirical—a primary consideration when applying QM/MM methods is the choice of QM region. The positioning of the QM/MM boundary is arbitrary, except that it should not divide the reaction coordinate. Although chemical intuition and critical analysis of the active-site structure can be used to guide choices, ultimately, they require validation by performing a number of calculations with QM regions of varying size and composition.

The reductions in DHFR take place via the transfer of a hydride ion from the nicotinamide adenine dinucleotide phosphate (NADPH) cofactor bound to the enzyme. The hydride-ion transfer also requires that a proton be transferred to the substrate (S), so that the overall reaction proceeds according to

$$S + H^+ + NADPH \rightarrow SH_2 + NADP^+ \quad (2)$$

The active site of *Escherichia coli* DHFR is shown in Figure 1. For this particular redox reaction, it is commonly believed that the substrate is protonated first, prior to hydride-ion transfer.[34] Protonation produces an increase in positive charge on the substrate, thereby activating it toward acceptance of the negatively charged hydride ion. Our recent computational studies[5,6] have provided some evidence that protonation of the conserved Asp/Glu that hydrogen bonds the substrate may also form part of the activation mechanism. The neutral (protonated) and ionized forms of Asp27 are depicted in Figure 1, H bonded to W206 and O4 of the substrate pterin ring. Despite its apparent importance in binding the substrate to DHFR, this residue has been omitted from the QM region in a large number of other QM/MM studies[2−4,7,9−12] of the hydride-ion transfer mechanism. Another QM/MM study has also suggested that electronic polarization of the QM region by the MM region is an essential feature of TS stabilization for the hydride-ion transfer in DHFR.[12] The choice of QM region is critical to an understanding of the origin of these proposed polarization effects as polarization is clearly a function of the QM/MM boundary; that is, the QM region is polarized by the partial charges in the MM region.

In the present study, we have used MD simulation methods to perform an a priori calculation of the free-energy change for the hydride-ion transfer in DHFR. In order to perform the necessarily large number of simulations efficiently, we use the semiempirical PM3-QM/MM approximation[6,27,35,36] for the force field in the MD and parallel processing capabilities of state-of-the-art computer clusters. From our simulations, we have determined the reaction coordinate corresponding to the minimum pathway on the free-energy surface for the hydride-ion transfer. Thus, we have a precise definition of the activation and reaction free energies for direct comparison with experimental results. This minimum free-energy path (MFEP) provides a convenient integration path for carrying out further investigative simulations to study the effects of changing the size and composition of the QM region, the ionization state of Asp27, and the polarization contribution to the free energy. Due to the approximations of the semiempirical QM method used in generating the MFEP, we also performed some higher-level ab initio and DFT QM/MM calculations along the MFEP to assist in validation.

Methods

Solvated Enzyme and QM/MM Model. Initial coordinates for *E. coli* DHFR were obtained from the DHFR·FOL·$NADP^+$ complex in the Protein Data Bank (1rx2).[37] In this structure, the mobile M20 loop is in the closed conformation, which is thought to be necessary for the hydride-ion transfer to take place. Hydrogens were added to this structure using Insight II[38] to form the DHFR·$DHFH^+$·NADPH complex. This complex, together with the 153 crystallographic water molecules, was solvated in a 36-Å-radius sphere containing 4962 additional water molecules. To maintain the system at constant density, the water molecules between 32 and 36 Å

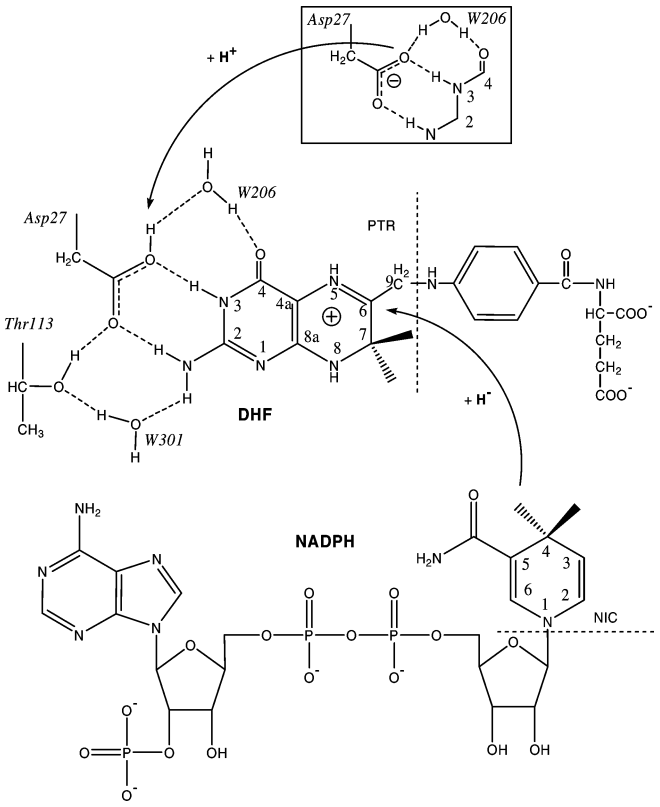

Figure 1. Active site in *E. coli* DHFR with N5-protonated 7,8-dihydrofolate ($DHFH^+$) substrate and NADPH cofactor. $DHFH^+$ is shown H-bonded to both the protonated carboxylate (carboxylic acid) and ionized (carboxylate) side chain of the conserved Asp27 residue. Other possible H-bond interaction sites, with Thr113 and water molecules W206 and W301, are shown as indicated from X-ray structures of *E. coli* DHFR complexes available in the Protein Data Bank.

were held fixed at their initial positions during the MD simulation to form a rigid outer shell of solvent, leaving a total of 3229 waters free to move. Where necessary, hydrogen (link) atoms were added to the QM region at the QM/MM boundaries. The force field of Cornell et al.[39] for the protein atoms and the TIP3P model[40] for water were used for the MM terms. The hybrid force field and parametrization procedure used for the nonbonded electrostatic and van der Waals (vdW) interactions between the QM and MM regions has been described in detail previously.[27,39,40] Briefly, the QM/MM term is given by

$$E_{\mathrm{QM/MM}} = tr(\mathbf{PV}) + E_{\mathrm{vdW}} \tag{3}$$

where **P** is the density matrix and **V** is the one-electron matrix for the interaction between the QM atoms and the partial charges (q) of the MM atoms. The matrix elements of **V** have the form

$$V_{\mu\nu} = \sum_k q_k(\mu\nu,ss) \tag{4}$$

where the summation is over MM atoms and ($\nu\mu,ss$) are the semiempirical approximations for the one-electron integrals. The electrostatic interactions between atoms in the QM and MM regions were not subject to any cutoff in the calculations of the MFEP. All residues in the MM region were subject to a 12 Å cutoff for the electrostatic terms. An 8 Å cutoff was applied to all vdW interactions between residues. In order to reduce the computational time, we also adopted a multiple time-step approach in which the electrostatic forces due to interactions beyond 8 Å were computed once every 10 time steps.

Minimum Free-Energy Path. We used a two-dimensional (2D) grid-based method in order to determine the MFEP. This type of approach has been used extensively to calculate adiabatic and free-energy surfaces for chemical reactions in both aqueous[41−43] and enzymic[17,44,45] environments. More recently, potential of mean force calculations in 2D have been carried out in order to generate free-energy surfaces for investigating concerted reactions.[46,47] Thus, grid methods would seem a natural choice when two parameters are required for a complete description of a reaction pathway in complex molecular environments such as enzymes.

The MFEP was determined using a QM region (Asp27·$DHFH^+$·NADPH) consisting of the N5-protonated substrate and cofactor molecules and the carboxylic acid ($-CH_2-COOH$) side chain of Asp27 (see Figure 1). In eq 1, we wrote the general reaction coordinate as $r = (r_1, r_2)$, where the C6−H4 and C4−H4 bond distances involving the hydride ion (H4) to be transferred are given by r_1 and r_2, respectively. In order to locate reactant, product, and transition states on the reaction free-energy surface, we first define a set of curvilinear reaction paths $\{x\}$ with $x \geq 1$ for the path parameter and $0 \leq \lambda \leq 1$ for the coupling parameter which drives the system between reactant and product states.[48] The $\lambda = 0$ and $\lambda = 1$ states correspond to reactant-side and product-side states, respectively. The equations for these paths are given (in Å) by

$$r_1(x,\lambda) = 3.4 - 2.4[1 - (1 - \lambda)^x] \tag{5a}$$

and

$$r_2(x,\lambda) = 1 + 2.4\lambda^x \tag{5b}$$

with values of $\{x\} = \{1.3, 1.5, 2.0, 2.5, 3.0, 4.0\}$; that is, a total of six paths were used to construct the free-energy surface. The λ-coupling parameter was divided into 21 discrete values using $\Delta\lambda = 0.05$. Thus, we obtain a grid of $6 \times 21 = 126$ points (x,λ); independent MD simulations were performed at these points for the free-energy calculation.

Reaction Path Simulations. The configurational sampling in these simulations was performed using the replica exchange molecular dynamics (REMD) method.[49,50] The REMD simulations consisted of independent trajectories (replicas) that were generated simultaneously (i.e., in parallel) on a Compaq AlphaServer and SGI Altix 3700Bx2 cluster. The trajectories were assigned temperatures in the range 290−320 K with unique initial velocity distributions. The trajectories were run using a time step of 0.001 ps, with hydrogen masses set to 2 amu. Temperature exchanges were attempted every 0.02 ps. As the temperature range used is quite narrow (290−320 K), we perform an average over the temperature to obtain a single value for the free energy. We note that the efficiency of REMD over conventional MD

has been questioned recently,[51] but this is not of concern here. In this study, we needed to exploit the advantages of parallel methods to reduce the wall time (i.e., the real time it takes to complete a calculation), as opposed to CPU time, to a reasonable time frame; this can be achieved by using either the REMD formalism or parallel computation of conventional MD trajectories.

To evaluate the required free energies, we used the thermodynamic integration (TI) formalism.[32,33] The free-energy change between 0 and a given value of λ along path x is exactly

$$\Delta G(x,\lambda) = \int_{a=0}^{a=\lambda} \left\langle \frac{\partial E[r(x,\alpha)]}{\partial \alpha} \right\rangle_a \mathrm{d}\alpha \tag{6}$$

where, in the present calculations, we have approximated the integrand in eq 6 by the sum over a finite number of time intervals

$$\left\langle \frac{\partial E[r(x,\alpha)]}{\partial \alpha} \right\rangle_\alpha \cong \frac{1}{N_{\mathrm{time}}} \sum_{\mathrm{i}=1}^{N_{\mathrm{time}}} \langle E_{\mathrm{i}}(x,\alpha)\rangle \tag{7}$$

with the subaverages over the number of processors in the REMD given by

$$\langle E_{\mathrm{i}}(x,\alpha)\rangle = \frac{1}{N_{\mathrm{proc}}} \sum_{\mathrm{j}=1}^{N_{\mathrm{proc}}} \frac{\partial E_{\mathrm{ij}}[r(x,\alpha)]}{\partial \alpha} \tag{8}$$

Following Ho et al.,[43] the SHAKE algorithm[52] was used to fix distances r given by eq 5 at the 126 (x,λ) grid points. However, as we use TI to compute the Hamiltonian derivatives with respect to the λ coupling parameter, our approach to constraint distance contributions is formally the same as that described by Pearlman.[53] The average value in eq 6 was computed as the average of the set of five moving averages with a time interval of $\Delta t = 1$ ps, for a total of 12 time intervals (12 ps), that is, $N_{\mathrm{time}} = 12$ in eq 7. The total number of derivatives that are computed is given by the product of the fixed number of time intervals, which numbered 12, and the number of processors (N_{proc}). In the present simulations, we used sampling parameters of $N_{\mathrm{proc}} = 12$ (total of 144 derivatives per λ) and $N_{\mathrm{proc}} = 48$ (total of 576 derivatives per λ) for the calculation of the MFEP. For each value of λ, the REMD was run prior to the start of sampling for at least 100 ps to ensure adequate equilibration. All REMD simulations were performed using Molecular Orbital Programs for Simulations (MOPS).[54]

To compute the necessary energy derivatives at the PM3-QM/MM level in eq 8, we first write the total system energy given by eq 1 in the form

$$\mathrm{E} = tr[\mathbf{PH}] + \epsilon \tag{9}$$

where **H** is the QM Hamiltonian matrix, **P** is the density matrix, and ϵ are the remaining terms in the expression for the energy which do not depend explicitly on **P**. We then make use of the Hellmann−Feynman theorem to obtain

$$\frac{\partial E}{\partial \alpha} = tr\left[\mathbf{P}\frac{\partial \mathbf{H}}{\partial \alpha}\right] + \frac{\partial \epsilon}{\partial \alpha} \tag{10}$$

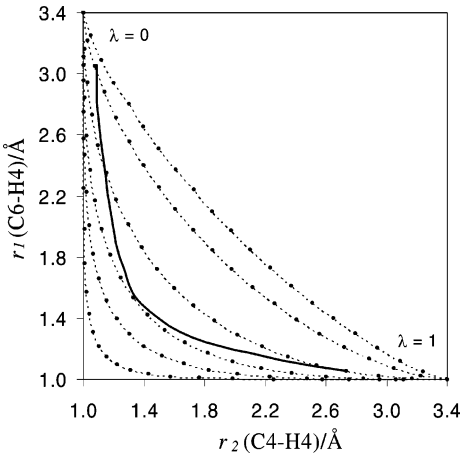

Figure 2. Reaction paths (dashed lines) used in the calculation of the minimum free-energy path (MFEP) (solid line) as a function of the C6−H4 and C4−H4 distances (Å). The MFEP was determined for λ ranging from 0.10 to 0.85 using a sampling parameter of $N_{\mathrm{proc}} = 48$. The dots (filled circles) indicate the points at which the MD simulations were performed for paths corresponding to $\{x\} = 1.3$, 1.5, 2.0, 2.5, 3.0, and 4.0 (1.3 gives the path with smallest curvature, 4.0 the path with highest curvature).

The derivatives of **H** and ϵ were conveniently obtained by finite difference ($\Delta\alpha = \pm 0.001$), using the density matrix **P** computed at α; at the semiempirical level, performing analytic evaluation of these derivatives does not provide any advantage in terms of precision or efficiency. Calculation of the energy derivatives (on the basis of configurational sampling at the PM3-QM/MM level along the MFEP) were also carried out using the ONIOM (QM/MM) method with electronic embedding[28,30] as implemented in the Gaussian 03 (G03) program.[55] Derivatives were obtained at the Hartree−Fock (HF) and DFT (B3LYP) levels using the 6-31G* basis and the formula

$$\frac{\partial E}{\partial \alpha} = \frac{\partial r_1}{\partial \alpha}\frac{\partial E}{\partial r_1} + \frac{\partial r_2}{\partial \alpha}\frac{\partial E}{\partial r_2} \tag{11}$$

where the derivatives of the energy E with respect to r_1 and r_2 were computed analytically (using G03) at the configurations generated in the PM3-QM/MM level simulations.

Results and Discussion

Minimum Free-Energy Path. The reaction paths (eq 5) are plotted in Figure 2. The (x,λ)-grid points at which the individual MD simulations were carried out are indicated by dots. The free-energy surface computed using $N_{\mathrm{proc}} = 12$ based on these grid points is illustrated in Figure 3. The isoenergy contours (2D projections in the x,λ plane) clearly show the reactant ($\lambda \sim 0.1$) and product ($\lambda \sim 0.85$) potential wells. The solid line in Figure 2 represents the MFEP obtained using a sampling parameter of $N_{\mathrm{proc}} = 48$. N_{proc} corresponds to the number of computer processors and is equivalent to the number of independent trajectories at each point (x,λ) in the free-energy calculation. This MFEP was found by a least-squares fitting of the path $\{x\}$ free-energy curves (dots in Figure 2) to a low-order (quadratic or cubic)

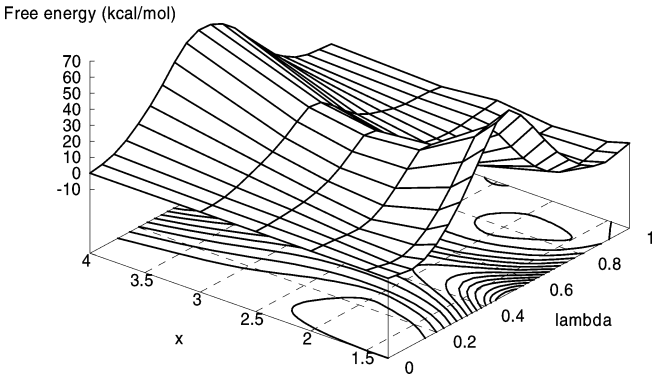

Figure 3. 3D and isoenergy contour representation of the PM3 free-energy surface. Free energy (kcal/mol) plotted as function of reaction path $\{x\}$ and λ coupling parameter (λ) for the hydride transfer to $DHFH^+$ in DHFR.

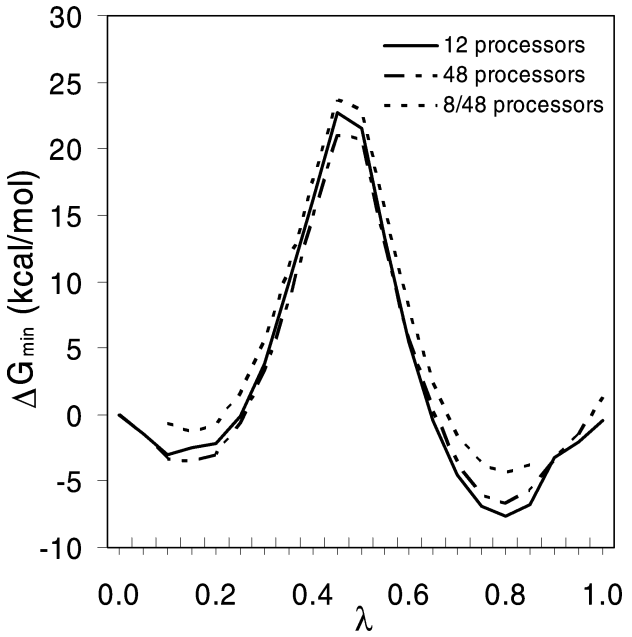

Figure 4. Change in free energy ΔG_{min} relative to $\lambda = 0$ for $N_{proc} = 12$ and $N_{proc} = 48$. The curve labeled 8/48 is not a ΔG_{min} (see text) but was calculated from independent simulations using $N_{proc} = 8$ at the points (x_{min}, λ) along the minimum free-energy path generated using $N_{proc} = 48$ (see Figure 3).

polynomial for each value of λ. We were unable to fit the free-energy curves for $\lambda < 0.1$ and $\lambda > 0.85$ values due to the fact that paths close to the end points ($\lambda = 0$ and $\lambda = 1$) rapidly converge (see Figure 2), forcing the actual small free-energy differences to be comparable with statistical sampling errors in the simulations.

The MFEP using the QM region for Asp27·$DHFH^+$·NADPH was obtained for two values of the sampling parameter, $N_{proc} = 12$ and $N_{proc} = 48$, which we will refer to as MFEP/12 and MFEP/48, respectively. We denote as x_{min} the value of x at which the free energy (as computed by the fitted polynomial) is a minimum value, given by ΔG_{min}, for a given value of the coupling parameter, λ. The free energy, ΔG_{min}, is shown in Figure 4, together with the free energy obtained from a subsequent simulation with $N_{proc} = 8$, calculated by following the MFEP/48 path. Note that the simulations with $N_{proc} = 8$ were carried out only for values

Table 1. Reaction Path Parameters for Reactant, Transition State, and Product Complexes

parameter[a]	reactant	TS	product
λ	0.1	0.475	0.8
x_{min} ($N_{proc} = 12$)	1.48	2.415	2.05
x_{min} ($N_{proc} = 48$)	1.45	2.45	2.07
R(C6−H4)/Å	3.06	1.50	1.09
R(C4−H4)/Å	1.09	1.40	2.52

[a] For PM3-QM/MM model simulations with Asp27(neutral)·$DHFH^+$·NADPH in QM region.

of λ ranging from 0.10 to 0.85. For the $N_{proc} = 12$ and 48 simulations used to determine the MFEP/12 and MFEP/48 paths, we averaged the free energy over x for all points, including $\lambda < 0.10$ and $\lambda > 0.85$. Note that the displacement between the ΔG_{min} curves for $N_{proc} = 8$ and 48 is due to the fact that the calculations start from different λ values, that is, $\lambda = 0.10$ and $\lambda = 0$, respectively.

Figure 4 reveals that there are no major differences in the results between different levels of sampling in the MD simulations. The free-energy change obtained by integration between $\lambda = 0$ and $\lambda = 1$, $<\Delta G(1)>$, where the average $<>$ is over paths$\{x\}$, was calculated to be -0.47 ± 2.35 kcal/mol for $N_{proc} = 12$ and 1.21 ± 2.16 kcal/mol for $N_{proc} = 48$. Note that, theoretically, the standard deviation in $<\Delta G(1)>$ should be zero (see Figure 2) as the reaction paths start and finish at common points. Thus, the nonzero standard deviation of $\sim \pm 2$ kcal/mol serves as a reliable estimate of the final error bars associated with incomplete sampling.

From the MFEP defined by coordinates (x_{min},λ), we can determine values for x and λ that are appropriate for reactants, TS, and products; these values are given in Table 1. The C6−H4 and C4−H4 distances in angstrom units, also given in Table 1, were obtained by substituting these values of x_{min} and λ into eq 5 for the curvilinear reaction paths shown in Figure 2. These distances were found to vary insignificantly (within ± 0.01 Å) between the two levels ($N_{proc} = 12$ and 48) of sampling.

QM Regions (with Protonated Asp27). The free-energy curves for different choices of the QM region are shown in Figures 5 and 6. These results are for protonated Asp27 only, with no cutoffs applied to electrostatic interactions between the QM and MM regions. They were obtained from simulations using $N_{proc} = 8$ and MFEP/48. In Figure 5, the complete $DHFH^+$ substrate and NADPH cofactor molecules are included in the QM region. The other enzyme fragments (W206, W301, and Thr113) that could be included (see Figure 1) interact with these reactant molecules exclusively via nonbonded interactions. It can be seen that the free-energy changes along the MFEP vary insignificantly (i.e., within the estimated statistical sampling error of ± 2 kcal/mol) among these choices of QM region. This suggests that the PM3-QM/MM force-field description we have used can adequately account for these types of nonbonded interactions. In contrast, Figure 6 shows that using smaller pterin (PTR) and nicotinamide (NIC) fragments (see Figure 1) of the substrate and cofactor in the QM region leads to much larger deviations from the reference MFEP. These results demonstrate that, at least for the neutral (protonated) state of Asp27, the minimum QM region need contain only the complete

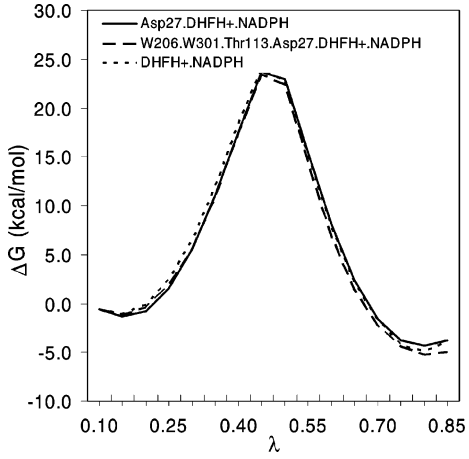

Figure 5. Free-energy change ($N_{proc} = 8$ and MFEP/48) relative to reactant state at $\lambda = 0.1$ for QM regions consisting of the substrate and cofactor molecules ($DHFH^+$·NADPH), and with additional active-site species forming H bonds with $DHFH^+$·NADPH (see Figure 1). Asp27 is neutral (protonated).

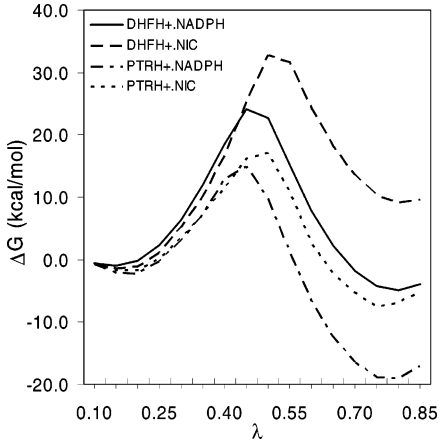

Figure 6. Free-energy change ($N_{proc} = 8$ and MFEP/48) relative to reactant state at $\lambda = 0.1$ for QM regions consisting of substrate and cofactor molecules ($DHFH^+$·NADPH), or combinations of nicotinamide-ring (NIC), and pterin-ring ($PTRH^+$) fragments (see Figure 1). Asp27 is neutral (protonated).

substrate and cofactor molecules. Note that many previous studies on DHFR include smaller fragments of these two species.[2−13] However, as demonstrated by the results in Figure 6, these choices represent very different reactions in terms of the free-energy change relative to the reactant state, and, although we do not compute them in the present study, represent somewhat different MFEPs. Hence, on this basis alone, we may also expect to find differences in the geometry of the TS, depending on the choice of QM region.[4−6]

Ionization State of Asp27. The results in Figures 7 and 8 were obtained from simulations using $N_{proc} = 8$ and MFEP/48, without cutoffs applied to electrostatic interactions between the QM and MM regions. As shown in Figure 7, if Asp27 is included in the MM region, the ionized state of Asp27 presents only a small perturbation on the neutral reference state. However, if ionized Asp27 is included in the QM region, a very different free-energy curve is obtained. The model with the ionized state gives rise to an endothermic reaction with a higher activation barrier, consistent with our

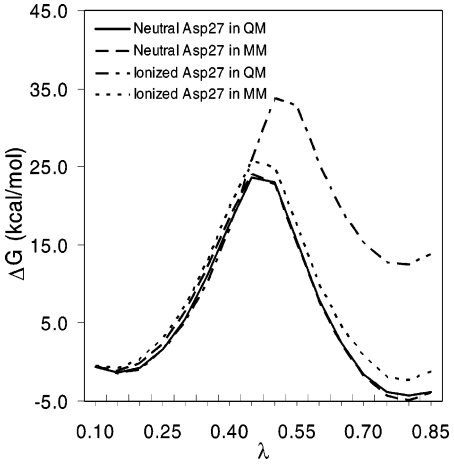

Figure 7. Free-energy change ($N_{proc} = 8$ and MFEP/48) relative to reactant state at $\lambda = 0.1$ for neutral and ionized states of Asp27 with QM regions Asp27·$DHFH^+$·NADPH (Asp27 in QM region) and $DHFH^+$·NADPH (Asp27 in MM region).

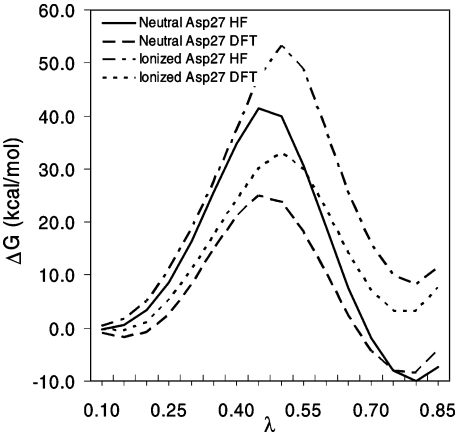

Figure 8. Free-energy change ($N_{proc} = 8$ and MFEP/48) relative to reactant state at $\lambda = 0.1$ calculated at the HF and DFT QM/MM levels for neutral and ionized states of Asp27 with QM region Asp27·$DHFH^+$·NADPH.

earlier QM/MM free energy simulations for the 8-methylpterin substrate.[56] As shown in Figure 8, the results are qualitatively similar if ab initio HF and DFT QM/MM methods are used to compute the free energies (based on the configurational sampling at the PM3-QM/MM level); that is, the reaction tends to be endothermic with a higher activation free energy. We have noted similar behavior between protonated and ionized Asp27 in previous geometry optimization studies using simplified model systems for the hydride transfer reaction.[6] The PM3, HF, and DFT methods predict very similar reaction free energies. As expected, HF overestimates the activation free energy relative to DFT, due to a lack of electron correlation in the former. The free-energy curve for PM3 is quite similar to that obtained for DFT.

From the PM3 data for neutral Asp (Figure 7), we estimate the free energy of reaction to be -4 ± 2 kcal/mol, compared with the experimental value of -4.4 kcal/mol.[57] The corresponding activation free energy of 24 ± 2 kcal/mol is considerably higher than the experimental value of 13.4 kcal/

Table 2. Polarization Contribution to the Free Energy (kcal/mol) at the Transition State, TS, at $\lambda = 0.475$, and Products at $\lambda = 0.80$, Relative to the Reactant State at $\lambda = 0.1$

system	method	TS ($\lambda = 0.475$)	products ($\lambda = 0.80$)
neutral Asp27 in QM	PM3	5.5 (−9.9)[a]	1.9 (−14.1)
	HF	4.5	6.1
	DFT	2.9	4.8
in MM	PM3	7.1 (−9.3)	5.5 (−10.7)
ionized Asp27 in QM	PM3	2.5 (−9.0)	4.5 (−14.1)
	HF	7.9	7.3
	DFT	9.1	9.7
in MM	PM3	8.7 (−11.1)	7.2 (−8.6)

[a] Free energies in parentheses were calculated using a 12 Å radius for the neglect of nonbonded interactions between atoms in the QM and MM regions.

mol.[57] Note that in the present simulations all atomic motions have been treated classically. If we include the 3 kcal/mol correction term for the effects of quantized vibrations,[2] we estimate a value of 21 ± 2 kcal/mol for the activation free energy. The DFT calculations yield a similar result for the activation free energy of ca. 25 kcal/mol (Figure 8). This result is unexpected given that in vacuo calculations on fragment complexes[6,7] show that PM3 consistently produces higher energy barriers compared with single-point DFT calculations carried out at the corresponding PM3 optimized geometries. The results in Table 2 show that this discrepancy cannot be rationalized in terms of polarization of the reaction center by the enzyme, as the difference in TS polarization free energy between PM3 and DFT amounts to only 2.6 kcal/mol (neutral Asp27). However, we also found from our fragment study[6] that optimization of the geometries at the DFT level can give barrier energies quite different from those for single-point calculations at PM3 or some other lower level of geometry optimization. This analysis suggests there may be some important differences between the semiempirical and DFT energy surfaces.

Polarization Free Energy. The polarization free energy arises from distortion of the continuous electron charge distribution of the QM region in the electrostatic field of the discrete point charges in the MM region. The polarization contribution to the free-energy change along the reaction coordinate may thus be defined as

$$\Delta\Delta G_{\mathrm{pol}} = \Delta G(\mathbf{P}) - \Delta G(\mathbf{P}_0) \qquad (12)$$

where $\mathbf{P}$ is the density matrix of the QM region obtained from a calculation in which the charge field of the MM region polarizes the QM region and $\mathbf{P}_0$ is obtained when this charge field is switched off in the calculation of the density matrix. Note that $\Delta G(\mathbf{P})$ and $\Delta G(\mathbf{P}_0)$ are computed from separate simulations. The polarization free energies at the TS ($\lambda = 0.475$) and product state ($\lambda = 0.8$), relative to the reactant state ($\lambda = 0.1$), are shown in Table 2 and were obtained from simulations using $N_{\mathrm{proc}} = 8$ and MFEP/48. The results obtained in the present study using a 12 Å radius as the cutoff for the neglect of interactions between the QM and MM regions (Table 2) are very similar to the 9 kcal/mol stabilization of the TS relative to reactants that was obtained by Garcia-Viloca et al.[12] using the same cutoff. However, on removing the cutoff for the QM/MM interaction, we found that polarization by the whole enzyme actually *destabilizes* the TS relative to reactants. The PM3-QM/MM polarization free energies are in broad agreement with the corresponding results obtained using the higher-level HF and DFT QM/MM methods. Thus, we may conclude, at least for this system, that the computationally less demanding semiempirical PM3 method can account for the bulk of the polarization effect and produce reliable estimates of the relative polarization free energies. Note also that the polarization free-energy differences between neutral and ionized Asp27, and the effect on the polarization free energy of moving Asp27 between QM and MM regions, are also not very large. But note again that the effect on the *total* free-energy differences due to these changes can be substantially greater.

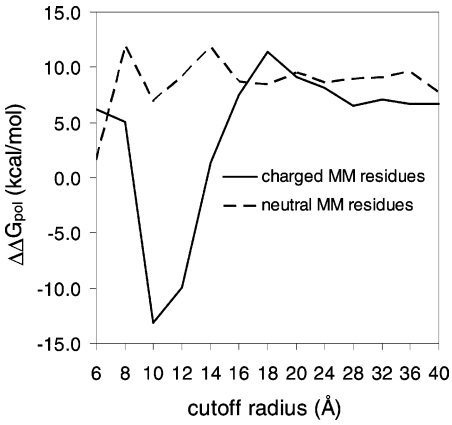

Figure 9. Polarization contribution (PM3 level) to the activation free energy as a function of cutoff for neglect of QM/MM interactions for both charged and neutral residues (see text) in the MM region, with Asp27·DHFH$^+$·NADPH (neutral Asp27) as the QM region.

The large increase in the magnitude of the polarization contribution obtained using a 12 Å cutoff can be explained largely in terms of the dipoles formed by ion pairs in the enzyme being split as a result of the application of the (residue-based) cutoff to the QM/MM interactions. The anomalous charge effect produced by the split dipole can be eliminated simply by neutralizing all acidic and basic residues in the MM region of the enzyme. However, rather than physically adding or deleting protons on the enzyme in order to gauge the magnitude of the split dipole effect, we have simply added a constant partial charge to each of the atoms in the ionized residues to enforce electrical neutrality. This was done only for the calculation of the free energy and not for the calculation of forces in the MD, so it does not otherwise affect the MD trajectory. The polarization component as a function of cutoff distance is shown Figure 9. In the calculation of the free energy using the normal residue charge state, the dip in the polarization contribution at around 12 Å is evident. However, the fluctuations with respect to cutoff distances are greatly reduced when all MM residues are made neutral.

Conclusions

In summary, using the semiempirical PM3-QM/MM method together with the TI formalism,[32,33,53] we mapped a sufficiently large part of the free-energy surface to allow us to obtain a precise MFEP for the hydride-ion transfer catalyzed by DHFR. For an accurate representation of the free-energy change along the reaction path, we found that it was necessary to include the whole substrate and cofactor molecules in the QM region of the active-site complex. For a model where the conserved active-site Asp27 residue is considered to be ionized, it also must be included in the QM region. Our calculations also show that electronic polarization by the enzyme actually *increases* the free energy of activation for the hydride-ion transfer and does not stabilize products either. Furthermore, we found that this polarization contribution was very sensitive to the neglect of electrostatic interactions between QM and MM atoms beyond a cutoff radius.

The MEFP obtained for the neutral state of Asp27 was found to be more consistent with the experimentally determined activation and reaction free energies[57] than that for the ionized state. This result adds further support to our findings in previous studies[5,6] that protonation of Asp27 is an essential factor in the activation mechanism. However, as complementary DFT calculations along the MEFP showed some discrepancies which may be due to the accuracy of the semiempirical representation of the free-energy surface, we are working on further methodological refinements which will allow a more definitive assessment of the Asp27 protonation mechanism.

Acknowledgment. We acknowledge support from Australian Research Council (ARC) Grants DP0346292 and DP0665816, the ANU (Australian National University) IAS (Institute of Advanced Studies) block grant, and APAC (Australian Partnership for Advanced Computing) National Facility and ANU Supercomputer Facility computer-time grants.

References

(1) Blakely, R. L. In *Advances in Enzymology and Related Areas in Molecular Biology*; Meister, A., Eds.; John Wiley: New York, 1995; Vol. 70, pp 23−102.

(2) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *Biochem.* **2003**, *42*, 13558−13575.

(3) Hammes-Schiffer, S. *Curr. Opin. Struct. Biol.* **2004**, *14*, 192−201.

(4) Castillo, R.; Andres, J.; Moliner, V. *J. Am. Chem. Soc.* **1999**, *121*, 12140−12147.

(5) Cummins, P. L.; Gready, J. E. *J. Am. Chem. Soc.* **2001**, *123*, 3418−3428.

(6) Cummins, P. L.; Greatbanks, S. P.; Rendell, A. P.; Gready, J. E. *J. Phys. Chem. B* **2002**, *106*, 9934−9944.

(7) Thorpe, I. F.; Brooks, C. L., III. *J. Phys. Chem. B* **2003**, 107, 14042−14051.

(8) Thorpe, I. F.; Brooks, C. L., III. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 444−457.

(9) Thorpe, I. F.; Brooks, C. L., III. *J. Am. Chem. Soc.* **2005**, *127*, 12997−13006.

(10) Greatbanks, S. P.; Gready, J. E.; Limaye, A. C.; Rendell. A. P. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 157−165.

(11) Greatbanks, S. P.; Gready, J. E.; Limaye, A. C.; Rendell. A. P. *J. Comput. Chem.* **2000**, *21*, 788−811.

(12) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *J. Mol. Biol.* **2003**, *327*, 549−560.

(13) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186−195.

(14) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425−443.

(15) Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. *Acc. Chem. Res.* **2001**, *34*, 72−79.

(16) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.

(17) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718−730.

(18) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.

(19) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.

(20) Gao, J. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Eds.; VHC Publishers: New York, 1996; Vol. 7, pp 119−185.

(21) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580−10594.

(22) Friesner, R. A.; Beachy, M. D. *Curr. Opin. Struct. Biol.* **1998**, *8*, 257−262.

(23) Merz, K. M., Jr.; Stanton, R. V. In *Encyclopedia of Computational Chemistry*; Wiley: New York, 1998; pp 2330−2343.

(24) Tomasi, J. In *Encyclopedia of Computational Chemistry*; Wiley: New York, 1998; pp 2343−2350.

(25) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *THEOCHEM* **1999**, *462*, 1−21.

(26) Murphy, R. B.; Phillipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *21*, 1442−1457.

(27) Luque, F. J.; Reuter, N.; Cartier, A.; Ruiz-Lopez, M. F. *J. Phys. Chem. A* **2000**, *104*, 10923−10931.

(28) Vreven, T.; Morokuma, K. *Theor. Chem. Acc.* **2003**, *109*, 125−132.

(29) Lu, Z.; Yang, W. *J. Chem. Phys.* **2004**, *121*, 89−100.

(30) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A., Jr.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815−826.

(31) Rosta, E.; Klahn, M.; Warshel, A. *J. Phys. Chem. B* **2006**, *110*, 2934−2941.

(32) Mezei, M.; Beveridge, D. L. *Ann. N.Y. Acad. Sci.* **1986**, *482*, 1−23.

(33) Straatsma, T. P. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VHC Publishers: New York, 1996; Vol. 9, pp 81−127.

(34) Huennekens, F. M.; Scrimgeour, K. G. In *Pteridine Chemistry*; Pfleiderer, W., Taylor, E. C., Eds.; Pergamon Press: New York, 1964; pp 355−376.

(35) Cummins, P. L.; Gready, J. E. *J. Comput. Chem.* **1997**, *18*, 1496−1512.

(36) Cummins, P. L.; Gready, J. E. *J. Comput. Chem.* **1999**, *20*, 1028−1038.

(37) Sawaya, M. R.; Kraut, J. *Biochemistry* **1997**, *36*, 586−603.

(38) *Insight II*, release 2000.1; Accelrys Inc.: San Diego, CA, 2002.

(39) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(40) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(41) Gao, J.; Xia, X. *J. Am. Chem. Soc.* **1993**, *115*, 9667−9675.

(42) Bash, P. A.; Ho, L. L.; MacKerell, A. D., Jr.; Levine, D.; Hallstrom, P. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 3698−3703.

(43) Ho, L. L.; MacKerell, A. D., Jr.; Bash, P. A. *J. Phys. Chem.* **1996**, *100*, 4466−4475.

(44) Strajbl, M.; Florian, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 4471−4484.

(45) Cui, Q.; Karplus, M. *J. Phys. Chem. B* **2002**, *106*, 1768−1798.

(46) Poulsen, T. D.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2003**, *107*, 9567−9578.

(47) Rajamani, R.; Naidoo, K. J.; Gao, J. *J. Comput. Chem*. **2003**, *24*, 1775−1781.

(48) Gready, J. E.; Rostov, I.; Cummins, P. L. In *Modelling Molecular Structure and Reactivity in Biological Systems*; Naidoo, K. J., Hann, M., Gao, J., Field, M., Brady, J., Eds.; Royal Society of Chemistry: London, 2006; pp 101−118.

(49) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(50) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96−123.

(51) Zuckerman, D. M.; Lyman, E. *J. Chem. Theory Comput.* **2006**, *2*, 1200−1202.

(52) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311−1327.

(53) Pearlman, D. A. *J. Chem. Phys*. **1993**, *98*, 8946−8957.

(54) Cummins, P. L. *Molecular Orbital Programs for Simulations (MOPS)*; Australian National University: Canberra, Australia, 1996.

(55) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian Inc.: Wallingford, CT, 2004.

(56) Cummins, P. L.; Gready, J. E. *J. Comput. Chem*. **1998**, *19*, 977−988.

(57) Fierke, C. A.; Johnson, K. A.; Benkovic, S. J. *Biochemistry* **1987**, *26*, 4085−4092.

CT600313B