Computational Methods for the Study of Enzymic Reaction Mechanisms. 1. Application to the Hydride Transfer Step in the Catalysis of Dihydrofolate Reductase

Peter L. Cummins, Stephen P. Greatbanks, Alistair P. Rendell,† and Jill E. Gready*

Computational Proteomics and Therapy Design Group, John Curtin School of Medical Research, Australian National University, P.O. Box 334, Canberra ACT 2601, Australia, and Department of Computer Science, Faculty of Engineering and Information Technology, Australian National University, P.O. Box 334, Canberra ACT 2601, Australia

Received: April 26, 2002

We report results and assessment from the use of several "state-of-the-art" computational methods to investigate the effect of the Asp27 ionization state on the hydride-ion transfer step in the enzymic reduction of folate and dihydrofolate (DHF) by DHFR from *Escherichia coli* (*E. coli*). The active forms of the DHFR complex are assumed protonated on the pterin/dihydropterin ring of the folate/dihydrofolate molecule prior to transfer of the hydride ion from the nicotinamide adenine dinucleotide phosphate (NADPH) cofactor. The calculations have been carried out for both protonated (neutral) and unprotonated (negatively charged) states of the conserved active-site Asp27 (*E. coli*) residue in the DHFR complexes. First, geometry optimizations at the semiempirical (PM3), density functional (DFT) and ab initio levels were performed on reaction-analogue clusters to obtain reactant, transition state (TS), and product complexes. The DFT and ab initio reaction energies obtained for the unprotonated Asp complexes were endothermic: exothermic reaction energies were obtained only for the protonated Asp complexes. In contrast, the PM3 method wrongly predicted endothermic reactions for both unprotonated and protonated Asp. Thus, for these simple analogue systems, the results suggest that reduction takes place only if the active-site Asp27 is protonated. Second, the effects of the remaining substrate, cofactor, protein and solvent interactions were studied using hybrid semiempirical (PM3) quantum mechanical and molecular mechanics (QM/MM) methods combined with molecular dynamics (MD) simulations. Within the MD simulation scheme, TS complexes were characterized using an efficient implementation of the coordinate-driving technique. Errors arising from the use of PM3 in the QM/MM calculations were estimated and corrected using the results of additional cluster calculations at the DFT and MP2 levels. Contrary to the initial cluster calculations, these corrected QM/MM calculations did not indicate unambiguously that reduction takes place only if the active-site Asp27 is protonated. For both ionized and neutral states of Asp27, analysis of the electrostatic interaction energies between the QM and MM regions in the MD simulations showed that the enzymic environment (MM region) plays a far greater role in stabilizing the final products than in stabilizing the TS complex. Activation free energies for the hydride-ion transfer ranged from 10 to 30 kcal/mol depending on the choice of QM region and ionization state of Asp27. For the larger QM region, the range of activation free energies (10−12 kcal/mol) is similar to an experimental value (13 kcal/mol), but in general quantitative agreement with experimental free energies was poor. Comparative estimates of reaction free energies suggest the catalytic advantage of a protonated Asp27 residue is small. Methodologically, the present work has identified issues relating to the composition and semiempirical (PM3) treatment of the QM region which need to be addressed in future studies.

Introduction

Hybrid quantum mechanics and molecular mechanics (QM/MM) methods have become the accepted approach for dealing with the complexities inherent in the computational study of enzymic reactions.[1−5] Use of QM/MM methods permits the computation of both energies and gradients for systems as large as proteins, thus opening the possibility of performing gradient minimization and/or molecular dynamics (MD) simulation studies. To incorporate rigorously the effects of the enzyme on the reaction center, it is highly desirable to apply the QM/MM methods within an MD simulation scheme. By performing MD simulations on reactant, transition state and product complexes, it is possible to describe both the spacial rearrangement of the QM and MM atoms, as well as their dynamical fluctuations, which are necessary to calculate activation and reaction *free* energies.

However, a major limitation of this QM/MM plus MD simulation approach is that it is currently only tractable for semiempirical QM methods, due to the prohibitive computational cost of conventional ab initio or density functional (DFT) methods. This can often be a troublesome deficiency as semiempirical QM methods are prone to unsystematic errors. When changes in bonding are involved, as in a chemical reaction, deficiencies in the semiempirical methods may become obvious. Consequently, in a semiempirical-QM/MM calculation a large part of the error is likely associated with the QM part of the calculation. It is, therefore, necessary to check the validity

* Corresponding author. E-mail: Jill.Gready@anu.edu.au.
† Department of Computer Science.

10.1021/jp021070q CCC: $22.00
Published on Web 08/31/2002

of the simulation results against more accurate ab initio QM results for the QM reaction center. Other reported hybrid ab initio QM/MM approaches[6] while providing a better description of the QM energy neglect the free energy changes associated with dynamical fluctuations of the QM subsystem. In the present work we have investigated the use of cluster calculations on a smaller analogue system as a way of assessing optimized geometries and energies calculated at semiempirical levels. By calculating ab initio QM-optimized geometries and energies for a relatively small cluster at different levels of theory, we can obtain an error estimate of the QM part in the semiempirical QM/MM calculation. Another potential problem with the QM/MM approach is the truncation of the QM region across covalent bonds, which can lead to artifacts in the properties of the QM system. In the present study we have also varied the size of the QM region in order to assess the reliabiltiy of the results.

We have applied the methods to dihydrofolate reductase (DHFR), an enzyme with a catalytic mechanism which has proven surprisingly difficult to define, and, hence, is a suitable challenge for application and assessment of the methods. DHFR catalyses the reduction of folate to dihydrofolate (DHF) and DHF to tetrahydrofolate (THF). These reductions 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

$$S + H^+ + NADPH \rightarrow SH_2 + NADP^+ \qquad (1)$$

The folate, DHF and NADPH molecules are illustrated in Figure 1. Also shown are the H-bond interactions between the substrates, water and active-site residues of DHFR.[7] Although, as noted, DHFR has been studied intensively,[8] many details of the catalytic mechanism remain unclear. It is not known, for example, how the substrates become protonated in the active site. Equally uncertain is the ionization state of the conserved acidic residues (Asp or Glu) in the active form of the complex and its specific role in the catalytic mechanism. The active site of DHFR contains a number of polar interactions which may influence the ionization behavior.[7] These interactions form an extensive H-bonded network involving the conserved Asp or Glu and active-site water molecules, as shown in Figure 1 for the *Escherichia coli* (*E. coli*) enzyme. Note that, as typical in the DHFR literature, the acidic (Asp27) residue in Figure 1 is depicted in the anionic form, so that one of the conserved waters, W206, is assumed to donate a H bond to both OD2 and O4 of the pterin or dihydropterin ring of the substrate. Note, however, this H-bonding orientation with W206 shown in Figure 1 will change if either OD2 or O4 is protonated. We recently reported results of both quantum chemical calculations and MD simulations within a QM/MM scheme for both the substrate and active-site residue protonations.[9−11] These studies demonstrated the importance of the conserved water molecule, corresponding to W206 in the *E. coli* DHFR complexes (Figure 1), and also suggested that the issue of OD2 protonation is likely to be important mechanistically.[11] Overall, these results tended to support a mechanism for both folate and DHF reductions in which the OD2 carboxyl oxygen is first protonated, followed by a direct protonation at N8 (folate) or N5 (DHF) to obtain the doubly protonated active cation complexes. Consequently, in the present study for the subsequent hydride-ion transfer step, we have modeled the enzymic reaction for both ionized and neutral states of Asp27.

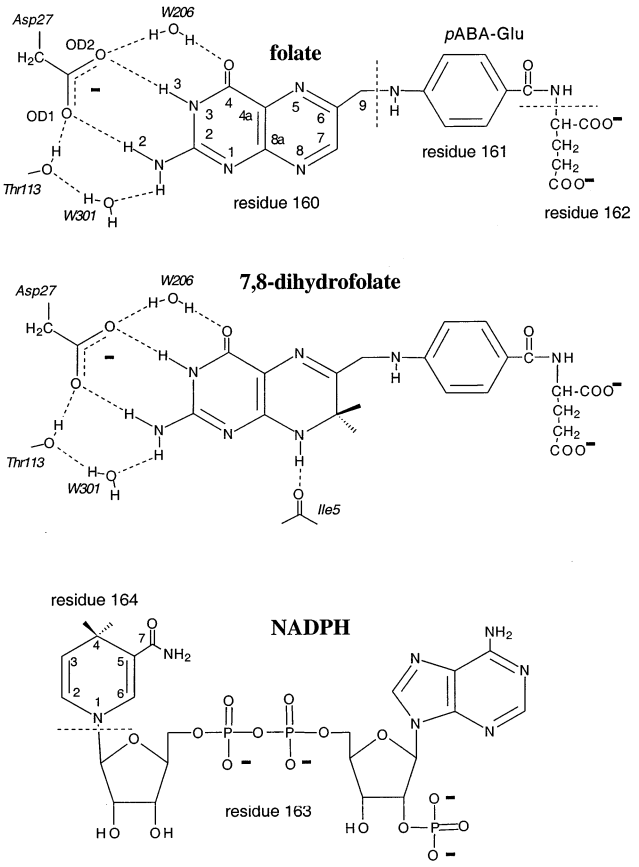

Figure 1. Structures of folate, 7,8-dihydrofolate (DHF) and NADPH cofactor. Substrates folate and DHF are shown H bonded to the carboxylate side chain of the conserved Asp27 residue in the active site of *E coli* DHFR. pABA = *p*-aminobenzoylglutamate side chain of the substrate molecules at C9. Division of structures into residues 160−165 used in the simulation analyses in Figures 5 and 6 are shown. Other possible H-bond interaction sites, including with water molecules W206 and W301, are indicated for folate and DHF on the bases of X-ray structures of *E. coli* DHFR complexes available in the Protein Data Bank.

Other computational studies on the NADPH-dependent hydride-ion transfer step of DHFR have been reported over many years.[12−18] A common feature of these early calculations, however, is that they were performed for analogue reactions in vacuo, and thus lack the explicit effects of the enzyme itself. More recent semiempirical QM/MM calculations of transition state structures for the reduction of DHF have explicitly included the enzyme.[19] Also, thus far, we have reported one MD simulation study employing QM/MM methods,[20] for the designed 8-methylpterin substrate, incorporating the enzyme explicitly. As in that previous study, the key feature of our present approach for the natural substrates, folate and dihydrofolate, is that the conformational flexibility is accounted for rigorously using MD, for both the QM and MM atoms, albeit at the expense of accuracy in the description of the interaction energies, i.e., the use of semiempirical QM methods. In the present work we have implemented an efficient coordinate driving approach for characterizing the transition state (TS) structures within the MD simulation scheme. We are then able to obtain an estimate of the TS on the free energy surface, and, hence, an activation free energy estimate for the enzymic reactions.

Methods

Cluster Calculations. All ab initio, DFT, and semiempirical PM3 cluster calculations were performed using the Gaussian

98 program (G98).[21] The effects of electron correlation were included using the B3LYP density functional,[22,23] as it usually yields results for many properties in close agreement with those obtained from MP2 calculations,[24−30] and is more efficient than conventional ab initio correlated methods for larger-scale calculations. The geometry optimizations on clusters were carried out at the PM3, HF/3-21G, B3LYP/3-21G, and B3LYP/6-31G** levels. However, as a cautionary note, there appear to be some instances, specifically for systems containing anionic species, and where field-response properties are calculated, where the DFT method breaks down.[31−34] Consequently, some single-point calculations at the MP2/6-31G** level were also performed in order to validate the B3LYP results. On the bases of the findings of our recent work,[9,10] we have used PM3 rather than AM1 as it better describes the H-bond interaction between Asp27 and the pterin ring of the substrates.

The primary objective of these cluster calculations was to compare ab initio, DFT and semiempirical optimized structures and energetics for the hydride-ion transfer reaction for ionized and neutral Asp27. Although not a complete model of the reaction in the enzymic system, it remains useful for investigating the effects of ionization of Asp27 and H bonding of the conserved water W206, on the relative energies of reactant, product and TS species. An initial set of coordinates for these reacting fragment clusters was obtained from a snapshot of the reactant DHF complex structure along an initial MD trajectory. The size of the ab initio cluster was chosen such that the calculations were tractable. It should be noted that the major computational cost for these calculations is the number of gradient evaluations which must be performed to optimize the structures along the reaction path. However, the cluster must include sufficient numbers of atoms to allow appropriate constraints to be applied. Ideally, the cluster should experience the flexibility of the true system, i.e., be no more or less flexible than in the enzyme. Less flexibility may prevent proper movement of participants in the reaction, while too much flexibility may allow for movements that would be prevented within the active site of the enzyme. Consequently, after optimization it might not be possible to dock the cluster back into the active site. The clusters used to represent the reactive fragments in the present study are shown in Figure 2. The constraints were imposed using fixed positions of the heavy (non-hydrogen) atoms of the Asp27 backbone, the ribose ring of the nicotinamide moiety, and the carbon atom of the 6-methyl group of the substrates. All hydrogen-atom positions were optimized.

Semiempirical methods were found unsuitable for initial investigation of the reaction complex, as the high gradients at the beginning of the optimization procedure lead to formation of unrealistic H bonds. Hence, the reactant, TS, and product clusters that were optimized at the ab initio level were used as starting structures for the PM3 geometry optimizations. For systems considered at the HF/3-21G level, no such problems inherent in the semiempirical methods were encountered. Semiempirical methods can, however, be used to speed up the generation of starting structures for the ab initio and DFT calculations. Specifically, as the first stage in a multistep optimization procedure, it was found useful to relax all hydrogens using a semiempirical calculation, while all heavy atoms in the cluster were frozen. An initial guess to the TS structure was then generated at the ab initio Hartree−Fock (HF) level using a "coordinate driving" technique.[35−37] The constraints imposed on the system, however, restrict the relative motions of the reacting fragments. Consequently, the only way to effect

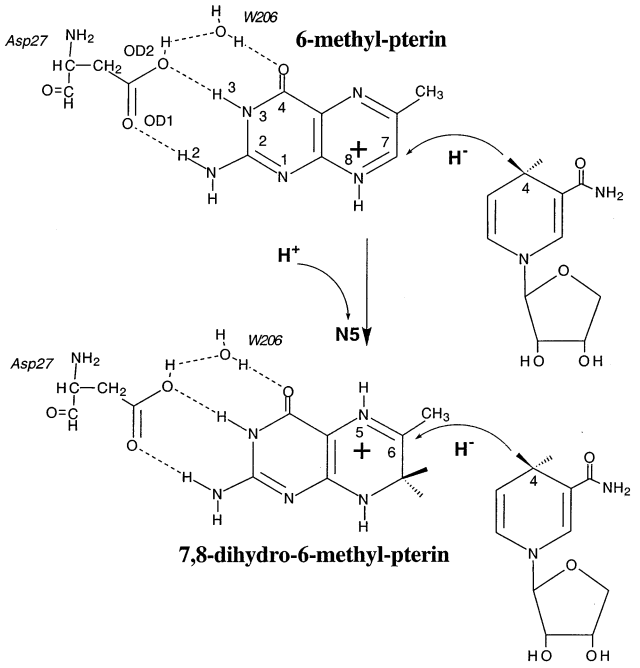

Figure 2. QM clusters used to optimize reactant, TS and product structures in the reduction of 6-methylpterin (folate analogue) and 7,8-dihydro-6-methylpterin (DHF analogue). The protonated form of Asp27 (OD2) is shown. Note that for the unprotonated OD2 form the H-bonding configuration is as shown in Figure 1.

the hydride-ion transfer is to push the hydride ion away from the carbon atom to which it is covalently bound, rather than bring the reacting fragments closer together along the lower-energy reaction path.

Calculations were first carried out for the ionized Asp27 cluster. In this case it was found more reliable to start with the product and move the hydride ion away from the product toward the reactant, as opposed to moving the hydride ion directly from reactants to products. When using the redundant internal coordinate optimizer with constraints in G98, it was necessary to move the hydride ion in small increments of 0.05 Å to ensure the integrity of successive geometry optimizations. If larger steps were employed, there was a tendency for the hydride ion to escape from the vicinity of the carbon atoms involved in the reaction [C4 on the nicotinamide ring and C6 (DHF) or C7 (folate) on the pterin ring]. Where calculations with a larger step size did converge, they required larger numbers of gradient evaluations. Thus, a larger number of smaller steps proved to be more efficient. Typically, an optimization between two successive points along the driven coordinate required several hundred gradient evaluations. Thus, while each SCF calculation can be performed quickly, the rate-limiting step becomes the number of gradient evaluations.

Near the TS, the reaction coordinate is no longer defined by the driven coordinate, because another covalent bond is either being formed or broken, i.e., the reaction coordinate is curved.[36,37] At some point beyond the TS, the reaction path becomes dominated by the bond that is being minimized, resulting in an abrupt decrease in the energy. Consequently, the maximum in the energy just before this sudden decrease is used as the initial guess, only, for the TS. At the structure corresponding to this maximum, a frequency calculation was performed and the derivatives at this maximum were used as the starting point for the TS search using the appropriate option (opt=ts) in G98. From this starting point the TS was obtained without complication. The TS was then characterized by another frequency calculation which generally yielded three imaginary frequencies,

of which one (in the region of 1900 cm^{-1}) corresponded to the hydride ion being transferred, while the remainder corresponded to hindered motions of atoms that were constrained in the optimization.

QM/MM Calculations. *Solvated Protein Model.* Initial coordinates for *E. coli* DHFR were obtained from the DHFR·FOL·$NADP^+$ complex in the Protein Data Bank (1rx2). In this structure the mobile M20 loop is in the closed conformation, which is thought to be required for the hydride-ion transfer to take place.[38] Most details of the MD simulation conditions follow closely from previous work.[39] The active-site region was solvated using a 26 Å radius shell of water molecules with a harmonic tethering-potential force constant of 0.005 kcal/mol/$Å^2$ applied to the water oxygen atom. Including crystallographic waters, there were 1219 molecules solvating the protein. In addition to the tethering potential, a half-harmonic restoring potential with a force constant of 0.6 kcal/mol/$Å^2$ was applied to prevent waters escaping from the dynamics zone, i.e., from moving outside the 26 Å solvation zone. All 1219 water molecules, but only those protein residues within a 16 Å radius of the active site, were included in the dynamics zone. Protein residues outside this radius were kept frozen at the crystallographic positions. The nonbonded interactions were treated as follows. All residues in the MM region were subject to a 20 Å cutoff for the electrostatic terms, except for interactions between charged residues for which no cutoffs were applied. Also, no cutoffs were applied to electrostatic interactions between the QM and MM regions. A 10 Å cutoff was applied to all van der Waals (vdW) interactions between residues. After energy minimization to relieve any bad contacts in the initial setup, MD simulations were carried out at a constant temperature of 300 K using a time step of 0.001 ps and with hydrogen masses set to 3 amu. After an initial equilibration phase requiring not more than 100 ps of MD, trajectories were typically run from 100 to 300 ps to allow sampling for the calculation of free energies. The force field of Cornell et al.[40] for the protein atoms and the TIP3P model[41] for water were used. The force field used for the nonbonded electrostatic and vdW interactions between the QM and MM regions has been described previously.[42,43] Hydrogen link atoms were used at the QM/MM interface. As in the cluster calculations, the PM3 method was used for the QM region. All MD simulations and QM/MM calculations were performed using Molecular Orbital Programs for Simulations (MOPS).[44]

QM/MM Partitioning. The partitioning of a system into a reaction zone (QM region) and environment (MM region) is entirely artificial. Indeed, the a priori definition of an "enzyme environment" has rather an obscure meaning. It is not always obvious how much of the system needs to be included in the QM region in order to describe adequately the reaction. There are not only boundary effects requiring addition of link atoms to consider. There are also differences in the description of interactions described by the QM/MM potential function, which has elements of the MM potential, as compared with the same interactions described by the purely QM semiempirical Hamiltonian approximation. Consequently variation in the results when the size and composition of the QM region are changed is likely. To validate any conclusions we can draw from the results it is necessary to gauge the magnitude of these variations by performing at least two sets of MD simulations using different QM/MM partitions. In the present study, one simulation was performed for the "small" QM region as shown in Figure 3 and one for an extended or "large" QM region as shown in Figure 4. From earlier simulation studies[20] on a substrate analogue bound to DHFR, we found that including the nicotinamide ring only of the NADPH cofactor in the QM region, with a hydrogen link atom, was, in fact, a better approximation for the whole cofactor than including larger fragments of it. Hence, the nicotinamide fragment of NADPH was included in both small and large QM partitions. The remaining fragments in the small QM region represent the minimum suggested by chemical intuition, and include the reactive pterin or dihydropterin ring of the substrate and the carboxylate side chain of Asp27 that forms a strong H-bond interaction with the substrate. The large QM region is an extension of the small region to include the important polar (H-bond) interactions in the active site (Figure 1). Also included in the large QM region are the pABA-amide side chain (-NH-Ph-CONH-) of the substrates and the backbone atoms of the Asp27 residue.

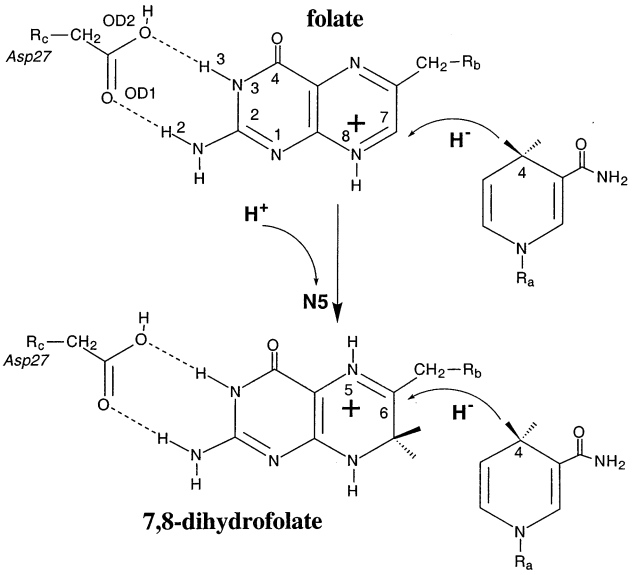

Figure 3. Small QM region in the hybrid QM/MM calculations. The protonated form of Asp27 (OD2) is shown. Note that for the unprotonated OD2 form the H-bonding configuration is as shown in Figure 1. R_a, R_b, and R_c are part of the MM region.

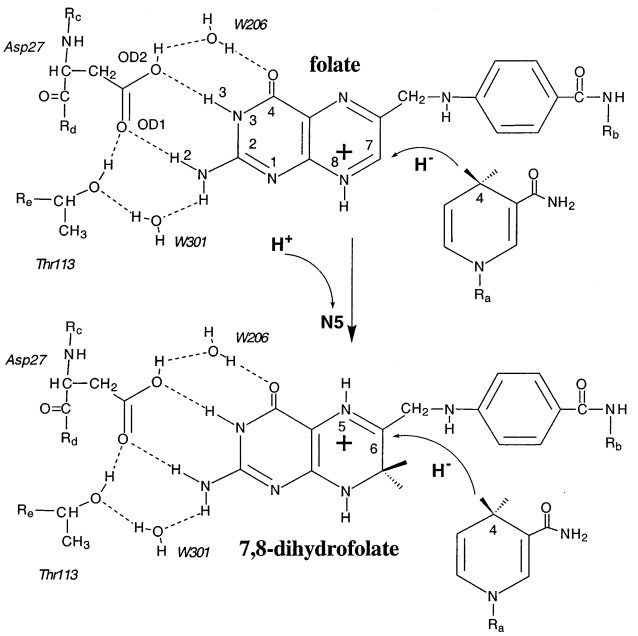

Figure 4. Large QM region in the hybrid QM/MM calculations. The protonated form of Asp27 (OD2) is shown. Note that for the unprotonated OD2 form the H-bonding configuration is as shown in Figure 1. R_a, R_b, R_c, R_d, and R_e are part of the MM region.

Characterization of Transition States. One option for calculating the free energy of activation would be to constrain the H4−C4 distance and C6−H4 or C7−H4 distance to generate a two-dimensional free energy surface using the free energy perturbation (FEP) approach. This would seem to represent the standard way of approaching the problem of chemical reactions in condensed phases.[20,45,46] However, such a free energy surface calculation is computationally demanding and will always contain statistical uncertainties in the ensemble averages, which would result in errors as the free-energy gradients are minimized. Consequently, it was deemed too expensive and impractical to refine the TS by calculating free energy differences between intermediate states on the reaction coordinate. In the present study, where a number of TS complexes must be characterized, we have implemented an approximate and more efficient coordinate driving method[35−37] within the MD simulation scheme. The efficiency of this approach derives from the fact that the vibrations of the bonds that are to be cleaved in the reaction are on a much shorter time scale than many of those associated with relaxation of the protein "environment" and, thus, can be sampled in a relatively short time. Thus, it should not be necessary to fully equilibrate and sample configurations corresponding to states along the reaction coordinate that are between the reactant and transition states.

This implementation is analogous to the coordinate driving method as applied to potential energy (PE) surfaces, except that an MD simulation instead of an energy minimization is carried out. Thus, for any fixed point on the chosen reaction coordinate, the system is relaxed using MD. The complications encountered in the optimizations of the cluster complexes (Figure 2) do not arise in the MD simulations, as there are no longer any artificial constraints imposed on atomic positions, other than those required to define the reaction coordinate. Thus, the reacting molecules are free in principle to move relative to one another along a minimum energy path as they would in the actual enzyme complex. The reaction coordinate was chosen to be the distance between the reactant hydride ion (bound to C4 of the NADPH cofactor) and the target carbon atom of the substrate (C6 for DHF and C7 for folate). The C6−H4 or C7−H4 distance constraint that defines the reaction coordinate was imposed during the MD simulation using the SHAKE algorithm.[47] Starting from the equilibrated reactant complex, the hydride ion (H4) and the target carbon of the substrate are brought together by reducing this distance R_{SHAKE} by finite amounts $\Delta R_{SHAKE} = -0.25$ Å. Short MD trajectories of 50 time steps were run at the points thus generated along this coordinate corresponding to the fixed C6−H4 or C7−H4 distances, primarily to relax the H4−C4 bond stretch. Closer to the TS ($R_{SHAKE} < 1.75$ Å) ΔR_{SHAKE} was reduced to −0.05 Å for increased precision. This procedure is illustrated schematically in Figure 5 which shows the typical fluctuations in the unconstrained H4−C4 distance as the constrained C6−H4 or C7−H4 distance is reduced. It was found to be very efficient, and quickly converges to a state where the relaxed H4−C4 bond dominates the reaction coordinate, i.e., just after the TS there is an abrupt change due to breakage of the H4−C4 bond and the formation of product-like species. This estimate of the TS complex is then subjected to further MD to ensure that, under proper equilibration conditions, the vibrations do not lead to H4−C4 bond cleavage. If the H4−C4 bond *does* break, the immediate previous point should be tested. This process of elimination can be repeated, if required. However, our experience with the method suggests that this is rarely necessary.

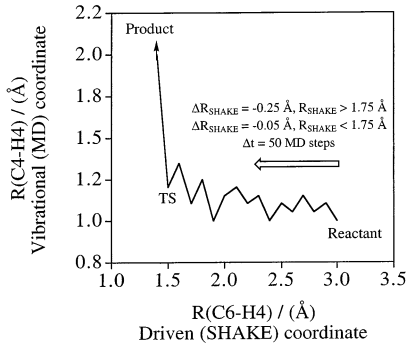

Figure 5. Scheme for application of the coordinate driving method within MD simulations. The driven coordinate, e.g. R(C6−H4) in the DHF → THF reaction, is fixed using SHAKE, initially at 3 Å (Reactant state). This distance, R_{SHAKE}, is then successively reduced by intervals of 0.25 Å, with short simulations of 50 MD time steps at each point, until $R_{SHAKE} = 1.75$ Å, at which point the interval is reduced to 0.05 Å.

Free Energy Determination. The energy of the interacting QM and MM regions in the hybrid QM/MM calculations is given by the sum of the QM and MM energies and the QM/MM energy of interaction between these two regions, i.e.,

$$E = E_{QM} + E_{QM/MM} + E_{MM} \tag{2}$$

We also assume that the free energy can be similarly expressed as the sum of QM, QM/MM, and MM contributions. From previous studies,[20] we have found this to be an excellent approximation, and it allows us to make effective use of the ab initio cluster calculations by separating out the QM-region energy, E_{QM}. The free energy at some reference state λ is written as

$$\Delta G(\lambda) = \Delta G_{QM} + \Delta G_{QM/MM} + \Delta G_{MM} \tag{3}$$

In the linear-response approximation (LRA), the change in the free energy of the MM region, ΔG_{MM}, and the change due to its interaction with the QM region, $\Delta G_{QM/MM}$, is given as half the average QM/MM interaction energy.[48,49] Thus, if we also neglect entropy changes in the QM region (i.e., $\Delta G_{QM} = \langle E_{QM} \rangle$), the energy components in eq 2 averaged over the trajectory length provide all the information necessary to estimate the required free energy differences. Then eq 3 becomes approximately

$$\Delta G(\lambda) = \langle E_{QM} \rangle + \frac{1}{2}\langle E_{QM/MM} \rangle \tag{4}$$

Further, the QM/MM interaction energy can be expressed as the sum of electrostatic and vdW contributions, $\langle E_{QM/MM} \rangle = \langle E_{ele} \rangle + \langle E_{vdW} \rangle$, so that we can also distinguish between these different contributions to the free energy.

The free energy of activation and the free energy for the reaction, respectively, are then calculated from eq 4 as

$$\Delta G_a = \Delta G(\text{TS}) - \Delta G(\text{reactants}) \tag{5}$$

$$\Delta G_r = \Delta G(\text{products}) - \Delta G(\text{reactants}) \tag{6}$$

The LRA is found to be an excellent approximation for cases where the QM region is not allowed to be polarized by the charges of the MM atoms.[43] This treatment is also consistent with the use of a typical MM force field for the protein,[40] which also neglects explicit polarization terms in the nonbonded interactions.

TABLE 1: R(C7−H4) and R(C4−H4) Distances (Ångstroms) at the TS for the Folate Analogue Reduction (FOL → DHF), and R(C6−H4) and R(C4−H4) Distances at the TS for the Dihydrofolate Analogue Reduction (DHF → THF)

Asp27 (OD2)	water	method	FOL → DHF		DHF → THF	
			R_{C7-H4}	R_{C4-H4}	R_{C6-H4}	R_{C4-H4}
unprotonated	no	PM3	1.303	1.572	1.374	1.592
		HF/3-21G	1.359	1.482	1.452	1.535
		B3LYP/3-21G	1.315	1.531	1.442	1.468
		B3LYP/6-31G**	1.343	1.522	1.439	1.493
	yes	PM3	1.310	1.566	1.399	1.584
		HF/3-21G	1.370	1.470	1.459	1.525
		B3LYP/3-21G	1.291	1.599	1.451	1.458
		B3LYP/6-31G**	1.300	1.578	1.453	1.486
protonated	no	PM3	1.404	1.489	1.487	1.514
		HF/3-21G	1.418	1.407	1.504	1.482
		B3LYP/3-21G	1.400	1.414	1.510	1.407
		B3LYP/6-31G**	1.423	1.416	1.505	1.436
	yes	PM3	1.410	1.488	1.494	1.503
		HF/3-21G	1.395	1.400	1.490	1.489
		B3LYP/3-21G	1.370	1.419	1.486	1.414
		B3LYP/6-31G**	1.422	1.412	1.499	1.444

Correction Terms. As it is not yet possible to perform proper Boltzmann averages at the ab initio QM level, we require an estimate of the error incurred by using the PM3 model in the hybrid QM/MM calculations. To compensate for these errors in the semiempirical QM energies, correction terms were added to the free energies obtained using eqs 5 and 6 according to

$$\Delta\Delta G_a = \Delta E_a(\mathrm{DFT}) - \Delta E_a(\mathrm{PM3}) \quad (7)$$

$$\Delta\Delta G_r = \Delta E_r(\mathrm{DFT}) - \Delta E_r(\mathrm{PM3}) \quad (8)$$

where ΔE_a and ΔE_r are quantum mechanical energy differences calculated at the PM3 and DFT (B3LYP/6-31G**) levels using QM-region geometries from a single configuration chosen at random from the MD simulations. Thus, the clusters consisted of the QM region in the QM/MM calculations. Unlike the cluster calculations in Figure 2, these do not involve generating gradients and performing optimization procedures at the ab initio level. However, we do not perform the single-point calculations directly using the raw MD structures. Starting with these raw coordinates from the MD trajectories, we first minimize the gradients of the entire enzyme complex at the QM/MM level before performing the PM3 and DFT calculations and applying eqs 7 and 8.

Results

Cluster Calculations. The C4−H4 (NADPH analogue), C6−H4 (DHF analogue), and C7−H4 (folate analogue) distances in the TS structures are given in Table 1. Without exception, protonation of Asp27 increases the C6−H4 and C7−H4 distances and decreases the C4−H4 distance, relative to the unprotonated state. The explicit inclusion of the water molecule representing W206 has little impact on either bond distance. The TS complexes in which Asp27 is unprotonated are thus more like the product state. These trends are observed at all levels of theory. In some cases, the PM3 methods give distances in close quantitative agreement with the B3LYP values. The agreement between the C7−H4 distances is particularly good for folate reduction (FOL → DHF). The largest differences, of the order 0.1 Å, are obtained for the C4−H4 distances for DHF reduction (DHF → THF).

The activation energies (ΔE_a) and reaction energies (ΔE_r) for the optimized complexes are given in Table 2. There appears to be no general correlation between the activation energies and the ionization state of Asp27. When the Asp27 residue is ionized, the reaction for the transfer of the hydride ion is always predicted to be endothermic ($\Delta E_r > 0$). However, if Asp27 is protonated the reaction energies are usually exothermic. Exceptions are for the PM3 method and some of the ab initio and DFT methods for the DHF → THF analogue reaction. The presence of the water molecule has a lesser effect on the relative energies of reactants, TS, and products. Except for the FOL → DHF reaction with unprotonated OD2, the agreement between PM3 and ab initio/DFT energies for all other reactions is quite poor.

TABLE 2: Energies (kcal/mol) of TS and Products Relative to Reactants, i.e., ΔE_a = activation energy, ΔE_r = Heat of Reaction

Asp27 (OD2)	water	method	FOL → DHF		DHF → THF	
			ΔE_a	ΔE_r	ΔE_a	ΔE_r
unprotonated	no	PM3	43.6	27.7	48.0	13.9
		HF/3-21G//PM3	59.1	20.3	58.7	21.9
		HF/3-21G	60.6	20.2	61.0	19.1
		B3LYP/3-21G	32.2	21.1	30.0	12.6
		B3LYP/6-31G**	36.4	21.8	28.8	11.3
		MP2/6-31G**// B3LYP/6-31G**	38.9	24.3	25.8	17.4
		B3LYP/6-31G**//PM3	32.4	14.2	30.0	12.9
	yes	PM3	38.5	21.5	48.8	26.4
		HF/3-21G	58.0	10.7	69.1	22.0
		B3LYP/3-21G	32.7	16.1	45.4	15.5
		B3LYP/6-31G**	32.3	15.4	28.7	11.9
		MP2/6-31G**// B3LYP/6-31G**	39.6	16.2	27.7	15.2
		B3LYP/6-31G**//PM3	29.6	10.1	38.2	18.4
protonated	no	PM3	33.2	−13.3	48.8	20.9
		HF/3-21G	45.9	−4.7	51.6	−1.2
		B3LYP/3-21G	20.4	−2.3	23.6	−3.3
		B3LYP/6-31G**	20.9	−5.5	34.3	7.1
		MP2/6-31G**// B3LYP/6-31G**	21.0	−4.5	30.7	1.1
		B3LYP/6-31G**//PM3	20.1	−23.2	32.8	7.5
	yes	PM3	33.2	5.9	39.5	16.7
		HF/3-21G//PM3	41.6	−9.6	51.9	2.6
		HF/3-21G	42.5	−8.2	47.4	−1.7
		B3LYP/3-21G	19.7	−6.3	18.6	−4.0
		B3LYP/6-31G**	37.1	−15.8	42.4	−8.0
		MP2/6-31G**// B3LYP/6-31G**	40.6	−13.6	42.8	−13.5
		B3LYP/6-31G**//PM3	18.8	−11.4	24.0	−3.0

Results in Table 2 for two basis sets and using various different levels of theory, allow an assessment of the reliability of the computational protocols. We consider first the ab initio results. The differences between the HF/3-21G and B3LYP/3-21G results give an indication of the importance of electron correlation in our systems. These differences are much smaller for the heat of reaction compared with the activation energies, averaging 3.5 and 26.7 kcal/mol, respectively. In all cases, the B3LYP activation energies are much less than the HF results, in line with the expectation that electron correlation will be more important for transition states compared with ground states. The effect of basis set can be gauged by comparing the B3LYP/3-21G and B3LYP/6-31G** results. The average difference between the reaction and activation energies calculated with these two basis sets is 4.2 and 9.4 kcal/mol, respectively. In this case there is no systematic behavior: in some cases the

TABLE 3: Average QM, ele, and vdW Energies (kcal/mol) of Reactants, TS, and Products from MD Simulations with Small QM Region (Figure 3)

substrate	Asp27 (OD2)	species	$\langle E_{QM}\rangle$	$\langle E_{ele}\rangle$	$\langle E_{vdW}\rangle$
FOL	unprotonated	reactant	-38.3 ± 0.4	-58.5 ± 0.8	-45.0 ± 0.2
		TS	-5.0 ± 0.6	-51.2 ± 1.7	-47.4 ± 0.5
		product	1.5 ± 0.3	-137.9 ± 1.2	-44.7 ± 0.2
	protonated	reactant	76.3 ± 1.2	-291.5 ± 1.9	-38.7 ± 0.8
		TS	92.9 ± 0.5	-287.1 ± 1.4	-37.5 ± 0.6
		product	83.5 ± 1.5	-345.9 ± 1.1	-42.0 ± 0.6
DHF	unprotonated	reactant	-62.0 ± 0.7	5.6 ± 0.7	-53.7 ± 0.3
		TS	-16.0 ± 0.4	-12.9 ± 1.1	-53.8 ± 0.3
		product	-21.7 ± 1.3	-104.4 ± 3.8	-48.0 ± 0.7
	protonated	reactant	48.9 ± 1.3	-250.6 ± 1.0	-44.1 ± 0.5
		TS	85.6 ± 0.7	-252.2 ± 1.9	-44.8 ± 0.7
		product	69.4 ± 1.1	-348.0 ± 2.4	-41.1 ± 0.7

TABLE 4: Average QM, ele, and vdW Energies (kcal/mol) of Reactants, TS, and Products from MD Simulations with Large QM Region (Figure 4)

substrate	Asp27 (OD2)	species	$\langle E_{QM}\rangle$	$\langle E_{ele}\rangle$	$\langle E_{vdW}\rangle$
FOL	unprotonated	reactant	-227.8 ± 0.9	-95.5 ± 1.1	-92.7 ± 0.4
		TS	-204.7 ± 2.1	-108.3 ± 1.0	-91.2 ± 0.6
		product	-198.2 ± 2.4	-187.9 ± 0.3	-88.3 ± 0.1
	protonated	reactant	-107.2 ± 0.4	-317.9 ± 0.2	-91.5 ± 0.3
		TS	-89.5 ± 1.1	-314.0 ± 0.3	-93.5 ± 0.7
		product	-98.2 ± 1.0	-377.9 ± 1.8	-90.5 ± 0.8
DHF	unprotonated	reactant	-246.8 ± 1.7	-108.8 ± 1.4	-90.9 ± 1.2
		TS	-203.8 ± 0.8	-100.6 ± 0.3	-92.9 ± 0.5
		product	-207.9 ± 1.7	-170.3 ± 0.8	-86.9 ± 0.1
	protonated	reactant	-126.9 ± 0.2	-316.5 ± 1.1	-90.9 ± 0.1
		TS	-101.1 ± 1.4	-303.2 ± 0.4	-92.0 ± 0.2
		product	-108.6 ± 2.3	-361.8 ± 1.6	-90.8 ± 0.1

6-31G** results are less than those for the 3-21G basis, while in other cases they are greater. Also the effect of basis set enlargement differs greatly between the unprotonated and protonated systems: the former shows relatively little effect, while the latter shows a much larger effect.

As previously mentioned, in earlier work we found density functional methods to be unreliable when computing field-dependent properties of anionic systems.[31,33] Hence, to validate the B3LYP results, we computed the activation and reaction energies using MP2 theory and the 6-31G** basis set. Optimization of the geometry at this level is, however, not practical, so our results were computed using the B3LYP/6-31G** geometries. The average differences between the B3LYP and MP2 reaction and activation energies are 3.4 and 2.7 kcal/mol, respectively, with a maximum difference of 7.3 kcal/mol. Overall, we conclude that B3LYP performs reasonably well for these systems.

Considering the semiempirical PM3 results, the average differences between the PM3 and B3LYP/6-31G** reaction and activation energies are 12.1 and 10.8 kcal/mol, respectively. These differences reflect effects from different geometry as well as between the PM3 and B3LYP methods themselves. To discriminate these two components, we performed B3LYP calculations at the PM3 geometry. The average differences between B3LYP results at the PM3 geometry and PM3 results at the PM3 geometry are 11.8 and 13.5 kcal/mol for the reaction and activation energies, respectively. Conversely, the corresponding differences between the B3LYP//PM3 and B3LYP//B3LYP results are 6.1 and 7.1 kcal/mol. This indicates that the primary deficiency of the PM3 method is in calculating accurate energetics, although errors due to use of a PM3 geometry are still quite large. Consequently, there are significant overall differences between the B3LYP and PM3 potential energy surfaces.

Energies from MD simulation. The average energy components are given in Table 3 for the MD simulations with the small QM region, and in Table 4 for the simulations with the large QM region. The average energies in Tables 3 and 4 can be used to compute enthalpies of activation and reaction. The results are given in Table 5 for the QM-region enthalpies, and Table 6 for the enthalpy corresponding to the electrostatic component of the QM/MM interaction. We do not give results for the vdW component of the QM/MM interaction as the results in Tables 3 and 4 clearly show that there are only minor differences between reactant, TS, and product states. In general, the atomic charges in the MM region have a large effect on the electrostatic stabilization of the product relative to reactant species (Table 6). However, although still significant, these same types of electrostatic effects are much smaller for the energy

TABLE 5: PM3 QM-only Energy Component $\langle E_{QM}\rangle$ (kcal/mol) of TS and Products Relative to Reactants, i.e., $\Delta\langle E_{QM}\rangle_a$ = Component of Activation Enthalpy, $\Delta\langle E_{QM}\rangle_r$ = Component of Reaction Enthalpy

Asp27 (OD2)	QM region[a]	FOL → DHF		DHF → THF	
		$\Delta\langle E_{QM}\rangle_a$	$\Delta\langle E_{QM}\rangle_r$	$\Delta\langle E_{QM}\rangle_a$	$\Delta\langle E_{QM}\rangle_r$
unprotonated	small	33.3	39.8	46.0	40.3
	large	23.1	29.6	43.0	38.9
protonated	small	16.6	7.2	36.7	20.5
	large	17.7	9.0	25.8	18.3

[a] Small QM region in Figure 3, large QM region in Figure 4.

TABLE 6: PM3 Electrostatic-Only Energy Component $\langle E_{ele}\rangle$ (kcal/mol) of TS and Products Relative to Reactants, i.e., $\Delta\langle E_{ele}\rangle_a$ = Component of Activation Enthalpy, $\Delta\langle E_{ele}\rangle_r$ = Component of Reaction Enthalpy

Asp27 (OD2)	QM region[a]	FOL → DHF		DHF → THF	
		$\Delta\langle E_{ele}\rangle_a$	$\Delta\langle E_{ele}\rangle_r$	$\Delta\langle E_{ele}\rangle_a$	$\Delta\langle E_{ele}\rangle_{rr}$
unprotonated	small	7.3	−79.4	−7.3	−98.8
	large	−12.8	−92.4	8.2	−61.5
protonated	small	4.4	−54.4	−1.6	−97.4
	large	3.9	−60.0	13.3	−45.3

[a] Small QM region in Figure 3, large QM region in Figure 4.

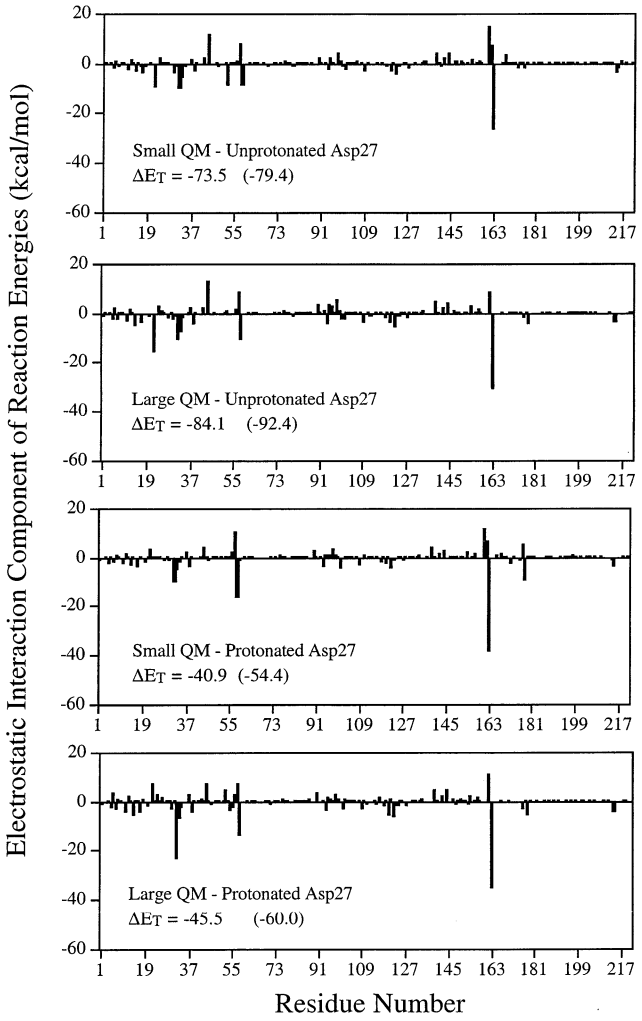

Figure 6. FOL → DHF reaction energies. Energy differences between products and reactants for the electrostatic interaction of the first 220 MM residues with the small (Figure 3) and large (Figure 4) QM regions, with unprotonated and protonated Asp27. Residue numbers above 220 correspond to the remaining water molecules. The total electrostatic interaction energy summed over all residues in the MM region is given by ΔE_T. The results in parentheses are the electrostatic components of the reaction enthalpies given in Table 6.

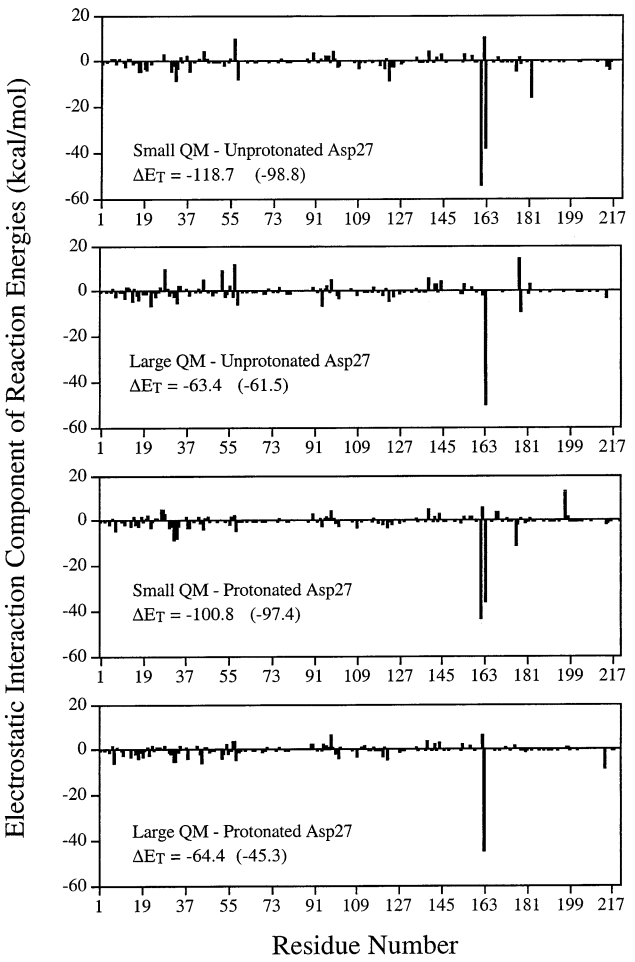

Figure 7. DHF → THF reaction energies. Energy differences between products and reactants for the electrostatic interaction of the first 220 MM residues with the small (Figure 3) and large (Figure 4) QM regions, with unprotonated and protonated Asp27. Residue numbers above 220 correspond to the remaining water molecules. The total electrostatic interaction energy summed over all residues in the MM region is given by ΔE_T. The results in parentheses are the electrostatic components of the reaction enthalpies given in Table 6.

difference between the TS and reactants, i.e., the activation enthalpy. For the DHF → THF reaction, the electrostatic component of the reaction enthalpy is significantly larger for the small-QM region than for the large-QM region.

To gain insight into the origin of these large differences in the electrostatic component of the reaction enthalpies arising from interactions of the small and large QM regions with the MM region, we partitioned the total electrostatic component of the QM/MM interaction into contributions from individual residues. The results are shown in Figure 6 for the FOL → DHF reaction, and in Figure 7 for the DHF → THF reaction. These residue interaction energies were calculated using structures obtained by gradient minimization of a single reactant and product configuration randomly chosen from the MD trajectories. The total electrostatic energy differences between product and reactant states summed over all residues (ΔE_T) are also indicated on the figures. These total reaction energies are similar to the corresponding electrostatic components of the reaction enthalpies in Table 6, i.e., the MD averaged energies, which are given for reference in parentheses in the figures. This analysis shows that the negatively charged end of NADPH (residue 163 in the QM/MM calculations) makes the single most significant contribution to stabilization of the products, except in the case of the DHF → THF reaction. In the DHF → THF reaction, the interaction of the −NH−Ph-CONH− (residue 161) fragment in pABA-Glu (residues 161 and 162) with the small QM region gives rise to the largest (50 kcal/mol) stabilizing effect. However, this 50 kcal/mol stabilization due to residue 161 is not reflected in the QM contributions to the reaction enthalpies for the large QM region given in Table 5. These QM reaction enthalpies are quite similar for both the small and large QM regions. The largest difference of 10 kcal/mol is for the FOL → DHF reaction (unprotonated OD2), while the remaining differences are only 2 kcal/mol. In contrast, the effect of residue 161 is smaller for the FOL → DHF reaction and destabilizes the products (Figure 6). The remaining residues belonging to the DHFR molecule (numbers less than 160) and solvent water molecules (numbers greater than 165) collectively make a smaller but still significant contribution. This is readily seen by subtracting the sum of contributions for residues 161, 162, and 163 from the total (ΔE_T) values. Without exception, these contributions, as similarly for the NADPH cofactor, always stabilize the product state.

Free Energy Predictions. The estimated activation free energies (ΔG_a) and reaction free energies (ΔG_r) are given in Tables 7 and 8. The results in Table 8 are the Table 7 results

TABLE 7: Free Energies (kcal/mol) of TS and Products Relative to Reactants, i.e., ΔG_a = Activation Free Energy, ΔG_r = Free Energy of Reaction

Asp27 (OD2)	QM region[a]	FOL → DHF		DHF → THF	
		ΔG_a	ΔG_r	ΔG_a	ΔG_r
unprotonated	small	35.8	0.2	42.3	−6.3
	large	17.4	−14.4	46.1	10.2
protonated	small	19.2	−21.7	35.5	−26.7
	large	18.6	−20.5	31.9	−4.3

[a] Small QM region in Figure 3, large QM region in Figure 4.

TABLE 8: Corrected PM3 Free Energies (kcal/mol) of TS and Products Relative to Reactants, i.e., ΔG_a = Activation Free Energy, ΔG_r = Free Energy of Reaction

Asp27 (OD2)	QM region[a]	FOL → DHF		DHF → THF	
		ΔG_a	ΔG_r	ΔG_a	ΔG_r
unprotonated	small	29.0	−8.4	16.9	−16.2
	large	12.4	−12.9	11.4	−20.1
protonated	small	12.7	−29.1	28.5	−41.5
	large	10.8	−29.6	10.2	−24.0

[a] Small QM region in Figure 3, large QM region in Figure 4.

after correction for errors in the PM3 method according to eqs 7 and 8. The uncorrected PM3 activation free energies in Table 7 range from 17.4 up to 46.1 kcal/mol, compared with 24 and 32 kcal/mol obtained from previous AM1 semiempirical QM/MM calculations for the reduction of DHF[19] and the folate analogue, 8-methylpterin,[20] respectively. Note that the value of 24 kcal/mol[19] was obtained using gradient minimization methods and not by MD simulation. For the FOL → DHF reaction with unprotonated Asp27, the PM3 (Table 7) and corrected PM3 (Table 8) free energies are quite similar. The corrections for the same reaction with protonated Asp27 become larger (∼10 kcal/mol), while for the DHF → THF reactions, the correction reduces the activation barriers by 20−30 kcal/mol and the reactions are also predicted to be more exothermic by 20−30 kcal/mol. The corrected PM3 results suggest that the catalytic difference between active-site complexes with protonated or unprotonated Asp27 is small, with activation barriers in the 10−12 kcal/mol range.

Transition State Geometries. The probabilities of finding the C4−H4 distance at a certain value for a given value of the C7−H4 (FOL → DHF) or C6−H4 (DHF → THF) reaction coordinate, fixed near either the reactant state or TS, are shown in Figure 8. As a further small reduction in the indicated value (range 1.25 to 1.45 Å) of the reaction coordinate fixed near the TS would result in cleavage of the C4−H4 bond, the best estimate of the C4−H4 distance at the transition point (indicated by "TS" in Figure 8) is given by the largest amplitude vibration of the C4−H4 bond as determined by the probability distribution. Thus, the C4−H4 distances at the TS are estimated to be approximately 1.40 and 1.35 Å for folate (unprotonated and protonated Asp27, respectively), and 1.50 and 1.40 Å for DHF (unprotonated and protonated Asp27, respectively). The results show significant differences between the TS geometries obtained for protonated Asp27 and those obtained for unprotonated Asp27. These differences are qualitatively the same as those found in the cluster calculations (Table 1), i.e., unprotonated Asp27 produces longer C4−H4 distances and shorter C6−H4 (DHF) and C7−H4 (folate) distances at the TS, compared with protonated Asp27. The QM/MM partitioning has a minimal effect, as both small and large QM regions predict very similar TS geometries.

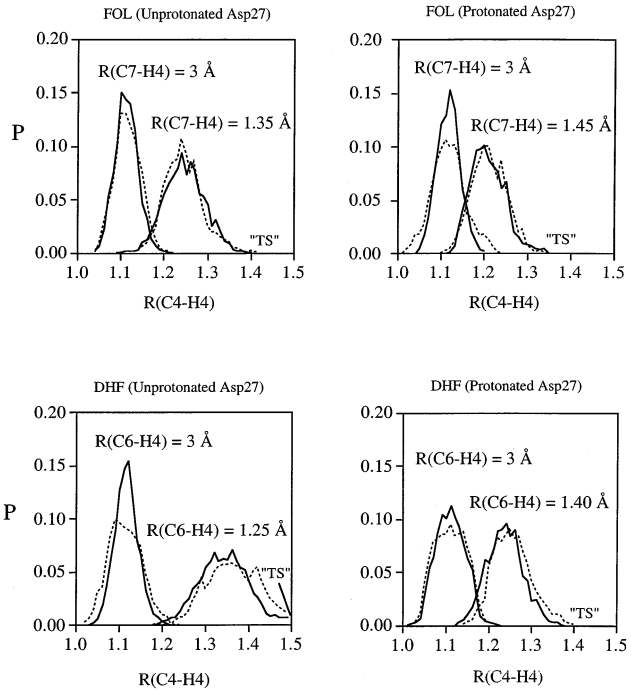

Figure 8. Probability distribution function (4 ps averages) of the C4−H4 distance (Å) in vicinity of reactants (C6−H4 = C7−H4 = 3 Å) and near the TS (C6−H4 = C7−H4 = 1.25−1.45 Å), as indicated. The dashed and solid lines are the results for the small QM region (Figure 3) and large QM region (Figure 4) simulations, respectively. "TS" indicates the C4−H4 distance at the transition state.

Discussion

On the basis of our more recent work,[9−11] we have refined our enzyme and computational models, and there are a number of differences between our current and previous[20] reaction studies. Previously, we assumed, as usually the case in the DHFR literature, that the active site Asp or Glu is ionized. The 8-methylpterin calculations also differ from the present calculations in that 8-methylpterin lacks the negatively charged side chain of folate, the source of DHFR was avian, Glu30 (analogous to Asp27) was partitioned to the MM region, and the AM1 method was used. Despite these differences, the 35.8 kcal/mol (uncorrected) value for the activation energy for folate reduction (small QM region, Table 7) with ionized Asp27, is close to the 30 kcal/mol value we obtained for reduction of the 8-methylpterin analogue with ionized Glu30. In contrast, the (uncorrected) result obtained for the large QM region is much smaller (17.4 kcal/mol), and similar to the activation energies of 19.2 (small QM) and 18.6 (large QM) obtained with protonated Asp27. The PM3 (i.e., again uncorrected) activation free energies (Table 7) for the DHF reduction are higher than the corresponding ones for folate reductions. The values of 42.3 and 46.1 kcal/mol (small and large QM, respectively) for DHF reduction with ionized Asp27 are considerably higher than the 25 kcal/mol enthalpic activation energy reported by Castillo et al.[19] (using AM1 methods). Castillo et al.[19] used geometry optimization methods to characterize the TS, whereas we have used averages obtained by MD simulation.

The QM/MM studies of the DHFR reaction have been conducted using the semiempirical methods, PM3 and AM1, for reasons of computational efficiency: currently, this is the only available option for performing MD simulations to obtain free energies. However, it is apparent from the all-QM cluster calculations that the PM3 method is inadequate for describing the relative energies of reactant, product and TS species for the

DHFR reaction. Hence, correction of the QM components using DFT or ab initio calculations is essential to obtain meaningful estimates of the free energies in the enzyme complex from the QM/MM simulations (Table 8). However, as eqs 7 and 8 neglect differences in the optimized geometries, i.e., do not properly take into account errors in the PM3 potential energy surface, the reliability of these corrections is difficult to predict. The results of the cluster calculations in Tables 1 and 2 suggest quite considerable errors. As these deviations are too large to be considered as perturbations, we conclude that eqs 7 and 8 do not offer a reliable method for correcting the semiempirical results. Nevertheless, in other ways there is qualitative agreement between the PM3 and DFT methods. They exhibit the same trend in TS structural changes with respect to ionization state of Asp27, i.e., for both methods the unprotonated complexes produce the more reactant-like TS, as measured by the key distances (Table 1) involving the hydride-ion moiety. The QM/MM calculations also exhibit this behavior (Figure 7). Furthermore, in the QM/MM calculations, the size of the QM region does not have a significant impact on the TS structure.

Another factor to consider is the reliability of the cluster calculations. To date, high level calculations have been restricted to reaction-analogue clusters, as shown in Figure 1. At best, these are crude approximations to the constraints and perturbations imposed by the enzyme. Hence, interpretation of the results from the cluster calculations must also be approached with some caution. In particular, in the cluster calculations the use of fixed-atom positions restrains reactants from coming closer together, which may result in any or all of the C4−H4, C6−H4 and C7−H7 distances in the TS being too long. The PM3 C4−H4 distance of 1.59 Å (Table 2) is in fact much larger than the analogous AM1 values of 1.40 (vacuo) and 1.37 Å (enzyme) obtained by Castillo et al.[19] Also, the MD-averaged values of the C4−H4 distances (Figure 7) are smaller than the corresponding PM3 values obtained in the cluster calculations (Table 1). A consequence of these constraints, which it is necessary to impose in order to obtain optimized structures, is that activation energies are likely to be overestimated.

An interesting feature of the present calculations is that the conserved water molecule, W206, appears to play little role in the actual hydride-ion reaction. There is only a small effect on the relative energies of reactants, TS and products from the cluster calculations (Table 2). Thus, the significance of the conserved water molecule, W206, may be limited to the activation step, namely in facilitating substrate protonation and in providing some additional stabilization for the OD2 protonated form of Asp27.[11]

As the ionization state, or states, of Asp27 appear critical to an understanding of the DHFR mechanism, a key mechanistic question is whether protonation of this residue offers any catalytic advantage in the hydride-ion transfer step? On this issue, the results from the small-cluster calculations disagree with those from the hybrid semiempirical QM/MM simulations. Whereas the cluster calculations (i.e. in vacuo) suggest that protonation of Asp27 lowers the activation barrier and reaction energies, this is not clear from the results of the QM/MM calculations for reaction in the enzyme. The corrected PM3 estimates (eq 7 or 8) of the activation free energies are in fact very similar for both protonated and unprotonated Asp27 (Table 8). Another consideration for this reaction arises because substrate and cofactor molecules are both highly charged and conjugated systems (Figure 1). Consequently, charge rearrangement and subsequent changes in π-bonding that occur on imposing a QM/MM boundary are likely to be considerable.[20] The present analysis of the QM/MM interactions (Figures 6 and 7) also shows a large contribution from these fragments to the overall reaction energetics. Thus, we concude that it is necessary to describe the whole of the reacting molecular fragments accurately.

Including an accurate treatment of the electronic structure of the complete cofactor and substrate molecules in modeling of the DHFR reaction presents a computational challenge greater than that for most other enzyme reactions for which high-level simulations have been attempted. Preliminary calculations we have carried out indicate other major deficiencies in the force-field description of the reaction center using the PM3 method applied to larger QM regions that include the whole of the NADPH molecule. In particular, the charged 2′-phosphate group is remarkably unstable in the MD simulations, while other nearby atoms form unexpected intramolecular H-bond interactions. However, an encouraging result of other recent work is that our QM/MM description is expected to perform quite well in representing the remaining bulk of the enzyme environment.[5,42,43]

Finally, we discuss our results to see what insights they offer in understanding experimental results. Experimental study of the chemical mechanism of folate and DHF reduction by DHFR has proven surprisingly intransigent. The step for folate reduction to DHF has been little studied experimentally: observed rates are very slow at neutral pH, quite possibly due to limitations in protonating this less basic substrate in the active site. Consequently, values for activation and reaction free energies are not available. Note, however, that for our designed 8-methylpterin substrate, protonation is enhanced and product release is not rate limiting, thus making it more suitable to direct study of the catalytic mechanism.[50,51] For the DHF reaction step, detailed pre-steady-state kinetic experiments for DHFR from *E. coli* [52] and other species (e.g., mouse,[53] *Lactobacillus casei*[54]) have established that product dissociation is the rate-limiting step under steady-state conditions (12.5 s^{-1}), with the chemical hydride-ion transfer step being rapid (950 s^{-1}) and essentially irreversible (back rate 0.6 s^{-1}; $K_{eq} = 1200$). Using these rates, activation and reaction free energies of 13.4 and −4.4 kcal mol, respectively, can be calculated. Similar values can be extracted from the analogous experiments for mouse (12.1 and −2.7 kcal mol)[53] and *L. casei* (13.9 and −3.4 kcal mol)[54] DHFRs. Although some of our estimates for the activation free energies (10−12 kcal/mol) compare well with experiment, the higher values appear too high for an enzyme-catalyzed reaction.[19] Also, our estimates for the DHF to THF reaction (−20−−24 kcal/mol) suggest much higher stabilization of the product complex compared with the experimental value of −4.4 kcal/mol.

Apart from the specific methodological problems regarding the PM3 approximation and choice of QM region we have already discussed, there are two other sources of error that should be mentioned. One is quantum dynamical effects, which have been neglected in the present studies, and are likely to further reduce activation energies for the hydride-ion transfer by up to 5 kcal/mol.[55,56] The other source of error is less predictable and relates to the adequacy of configuration space sampling in the MD. Experimental (NMR)[57] and simulation (MD)[58] studies for *E. coli* DHFR complexes implicate dynamical motions in promoting catalysis for the DHF to THF reaction. The MD studies[58] (nanosecond time scale) identified correlated dynamical motions in the reactive DHFR·DHF·NADPH complex but not in the product complex, DHFR·THF·$NADP^+$. The NMR studies[57] on complexes proposed to represent different steps in the catalytic cycle showed significantly different

dynamical behavior, with that mimicking the reactive complex exhibiting fluctuations on the microsecond to millisecond time scale. There may, therefore, be energetically important conformational differences between reactant, TS, and product complexes that have not been sampled in the present MD simulations spanning less than 1 ns. One possibility relates to the apparent overstabilization of our THF product complex. Conformational changes in the Met20 loop and coupled structures proposed to assist product dissociation, take place on a longer time scale (i.e., microsecond to millisecond) than sampled by MD simulation.[58] Our THF product complex may, thus, represent a more tightly bound predissociative conformational state of the enzyme.

Conclusion

Using both high level QM (ab initio and DFT) cluster calculations and hybrid semiempirical (PM3) QM/MM calculations we have studied the effects of active-site protonation on the hydride ion transfer reaction from the NADPH cofactor to the substrates folate and dihydrofolate in the presence of the enzyme, dihydrofolate reductase. The ab initio and DFT calculations on analogue clusters suggest that reduction takes place when Asp27 is protonated. However, the same conclusions could not be drawn from the more realistic QM/MM plus MD model simulation results. In general, consistency with experimental free energies was poor. Consequently, the present study has identified several methodological deficiencies that need to be addressed. In particular, as free energy predictions were sensitive to the choice of QM region, future studies should include all atoms of both substrate and cofactor moieties, at least, in the QM region. Our preliminary investigations suggest that accurate QM calculations on structures of this size present a considerable challenge for computational methods. The other main area of concern is the accuracy of the QM/MM force fields used in the MD simulations, and the methods used to "correct" the PM3 energies.To address this issue we are currently assessing several approaches, including published methods (specific reaction parameters[59] and system specific parameters[60]), to obtain more accurate potential energy surfaces and reaction energetics.

Acknowledgment. We gratefully acknowledge the Australian National University Supercomputer Facility and Sun Microsystems for generous grants of computing resources, and the ANU-Fujitsu Area 3 contract for support (S.P.G.).

References and Notes

(1) Gao, J. L. *Rev. Comput. Chem.* **1996**, *7*, 119.
(2) Amara, P.; Field, M. J. In *Encyclopaedia of Computational Chemistry*; von Rogue Schleyer, P., Ed.; John Wiley: New York, 1998; p 431.
(3) Friesner, R. A.; Beachy, M. D. *Curr. Opin. Struct. Biol.* **1998**, *8*, 257.
(4) Mordasini, T. Z.; Thiel, W. *Chimia* **1998**, *52*, 288.
(5) Luque, F. J.; Reuter, N.; Cartier, A.; Ruiz-Lopez, M. F. *J. Phys. Chem. A* **2000**, *104*, 10923.
(6) Liu, H.; Zhang, Y.; Yang, W. T. *J. Am. Chem. Soc.* **2000**, *122*, 6560.
(7) Bystroff, C.; Oatley, S. J.; Kraut, J. *Biochemistry* **1990**, *29*, 3263.
(8) Blakely, R. L. In *Advances in Enzymology*; Meister, A., Ed.; John Wiley: New York, 1995; Vol. 70, p 23.
(9) Cummins, P. L.; Gready, J. E. *J. Phys. Chem. B* **2000**, *104*, 4503.
(10) Cummins, P. L.; Gready, J. E. *J. Mol. Graph. Model.* **2000**, *18*, 42.
(11) Cummins, P. L.; Gready, J. E. *J. Am. Chem. Soc.* **2001**, *123*, 3418.
(12) Donkersloot, M. C. A.; Buck, H. M. *J. Am. Chem. Soc.* **1981**, *103*, 6549.
(13) Brounts, R. H. A. M.; Buck, H. M. *J. Am. Chem. Soc.* **1983**, *105*, 1284.
(14) Krechl, J.; Kuthan, J. *J. Mol. Struct.*: *THEOCHEM* **1988**, *170*, 239.
(15) Tapia, O.; Candenas, R.; Andres, J.; Colonna-Cesari, F. *J. Am. Chem. Soc.* **1988**, *110*, 4046.
(16) Cummins, P. L.; Gready, J. E. *J. Comput. Chem.* **1990**, *11*, 791.
(17) Andres, J.; Safont, V. S.; Martins, J. B. L.; Beltran, A.; Moliner V. *J. Mol. Struct.*: *THEOCHEM* **1995**, *330*, 411.
(18) Andres, J.; Moliner, V.; Safont, V. S.; Domingo, L. R.; Picher, M. T.; Krechl, J. *Bioorg. Chem.* **1996**, *24*, 10.
(19) Castillo, R.; Andres, J.; Moliner, V. *J. Am. Chem. Soc.* **1999**, *121*, 12140.
(20) Cummins, P. L.; Gready, J. E. *J. Comput. Chem.* **1998**, *19*, 977.
(21) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, Revision A.6; Gaussian, Inc.: Pittsburgh, PA, 1998.
(22) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(23) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(24) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: Weinheim, 2000.
(25) Miani, A.; Cane, E.; Palmieri, P.; Trombetti, A.; Handy, N. C. *J. Chem. Phys.* **2000**, *112*, 248.
(26) Cohen, A. J.; Handy, N. C. *Chem. Phys. Lett.* **2000**, *316*, 160.
(27) McAllister, M. A. *Can. J. Chem.* **1997**, *75*, 1195.
(28) McAllister, M. A. *J. Mol. Struct.*: *THEOCHEM* **1998**, *427*, 39.
(29) Pan, Y. P.; McAllister, M. A. *J. Mol. Struct.*: *THEOCHEM* **1998**, *427*, 221.
(30) Lozynski, M.; Rusinskaroszak, D.; Mack, H. G. *J. Phys. Chem.* **1998**, *102*, 2899.
(31) Greatbanks, S. P,.; Gready, J. E.; Limaye, A. C.; Rendell, A. P. *Proteins*: *Struct., Funct., Genet.* **1999**, *37*, 157.
(32) Domene, C.; Fowler, P. W.; Jemmer, P.; Madden, P. *Chem. Phys. Lett.* **1999**, *299*, 51.
(33) Greatbanks, S. P.; Gready, J. E.; Limaye, A. C.; Rendell, A. P. *J. Comput. Chem.* **2000**, *21*, 788.
(34) Champagne, B.; Perpete, E. A.; Jacquemin, D.; van Gisbergen, S. J. A.; Baerends, E. J.; Soubra-Ghaoui, C.; Robins, K. A.; Kirtman, B. *J. Phys. Chem. A.* **2000**, *104*, 4755.
(35) Rothman, M. J.; Lohr, L. L. *Chem. Phys. Lett.* **1980**, *70*, 405.
(36) Burket, U.; Allinger, N. L. *J. Comput. Chem.* **1982**, *3*, 40.
(37) Schlegel, H. B. *Adv. Chem. Phys.* **1987**, *67*, 249.
(38) Sawaya, M. R.; Kraut, J. *Biochemistry* **1997**, *36*, 586.
(39) Cummins, P. L.; Gready, J. E. *J. Comput. Chem.* **1996**, *17*, 1598.
(40) 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.
(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(42) Cummins, P. L.; Gready, J. E. *J. Comput. Chem.* **1997**, *18*, 1496.
(43) Cummins, P. L.; Gready, J. E. *J. Comput. Chem.* **1999**, *20*, 1028.
(44) Cummins, P. L. *Molecular Orbital Programs for Simulations* (*MOPS*) **1996**.
(45) Gao, J.; Xia, X. *J. Am. Chem. Soc.* **1993**, *115*, 9667.
(46) Ho, L. L.; MacKerell, A. D.; Bash, P. A. *J. Phys. Chem.* **1996**, *100*, 4466.
(47) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.
(48) Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. *Protein Eng.* **1992**, *5*, 215.
(49) Åqvist, J.; Medina, C.; Samuelsson, J.-E. *Protein Eng.* **1994**, *7*, 385.
(50) Ivery, M. T. G.; Gready, J. E. *Biochemistry* **1995**, *34*, 3724.
(51) Jeong, S.-S.; Gready, J. E. *Biochemistry* **1995**, *34*, 3734.
(52) Fiercke, C. A.; Johnson, K. A.; Benkovic, S. J. *Biochemistry* **1987**, *26*, 4085.
(53) Thillet, J.; Adams, J. A.; Benkovic, S. J. *Biochemistry* **1990**, *29*, 5195.
(54) Andrews, J.; Fiercke, C. A.; Birdsall, B.; Ostler, G.; Feeney, J.; Roberts, G. C. K.; Benkovic, S. J. *Biochemistry* **1989**, *28*, 5743.
(55) Kong, Y. S.; Warshel, A. *J. Am. Chem. Soc.* **1995**, *117*, 6234.
(56) Hwang, J. K.; Chu, Z. T.; Yadav, A.; Warshel, A. *J. Phys. Chem.* **1991**, *95*, 8445.
(57) Osborne, M. J.; Schnell, J.; Benkovic, S. J.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2001**, *40*, 9846.
(58) Radkiewicz, J. L.; Brooks, C. L., III *J. Am. Chem. Soc.* **2000**, *122*, 225.
(59) Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *233*, 231.
(60) Bash, P. A.; Ho, L. L.; MacKerell, A. D.; Levine, D.; Hallstrom, P. *Proc. Nat. Acad. Sci. U.S.A.* **1996**, *93*, 3698.