QM-FE Calculations of Aliphatic Hydrogen Abstraction in Citrate Synthase and in Solution: Reproduction of the Effect of Enzyme Catalysis and Demonstration that an Enolate Rather than an Enol Is Formed

Oreola Donini,[†,§] Tom Darden,[‡] and Peter A. Kollman*[,†]

Contribution from the Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, CA, 94143-0446, and Laboratory of Structural Biology, NIEHS, National Institutes of Health, Research Triangle Park, North Carolina 27709

Received March 24, 2000. Revised Manuscript Received October 10, 2000

Abstract: Mechanistic enzymologists have long debated how enzymes catalyze the abstraction of an unactivated C−H group. Citrate synthase, due to its ability to catalyze this abstraction and its central role in the respiratory cycle, has been extensively studied both experimentally and theoretically. Despite this scrutiny, the question remains as to whether the initial aliphatic hydrogen abstraction step of the mechanism is stabilized by the formation of an enol-imidazolate intermediate through "short, strong" hydrogen bonds, as opposed to the more traditional enolate-imidazole complex. In an attempt to present a definitive answer to this question, quantum mechanical-free energy (QM-FE) calculations were performed for the formation of the enolate-imidazole complex from the reactants, as well as for the further formation of the enol-imidazolate system. These reactions were found to be extremely sensitive to the use of nonbonded cutoffs, and reliable results were only obtained with the use of particle mesh Ewald (PME) to treat the electrostatic interactions. Because of the length of these simulations, we also used a coarse-grained parallel approach to free energy calculations. The results indicate that the enolate-imidazole complex is the more stable one within the enzyme by approximately 13 kcal/mol. The calculated barrier to the formation of the enolate is in good quantitative agreement with the k_{cat} for this enzyme.

1.0. Introduction

The question of how enzymes achieve catalysis of reactions involving the abstraction of an unactivated C−H group has been of interest to mechanistic enzymologists. Concepts such as "low barrier hydrogen bonds"[1] and "pK_a matching"[2−4] have been proposed to rationalize the impressive level of enzyme catalysis achieved by enzymes such as triose phosphate isomerase (TIM)[5] and citrate synthase (CS) on their particular substrates. In this work we focus on the first catalytic step, the abstraction of the C−H proton from acetyl coenzyme A (AcCoA) by CS.

Citrate synthase is a central enzyme in the respiratory cycle and has been intensely studied. CS catalyzes the condensation of oxaloacetate (OAA) and acetyl-CoA (AcCoA) to form citrate and CoA. After binding the substrates, and undergoing a protein rearrangement, the first rate-limiting step of the reaction involves the abstraction of an aliphatic hydrogen from the AcCoA by aspartate (Scheme 1).[6] This reaction is extremely slow in solution and it is unclear how the enzyme catalyzes this step. Proposals vary from the simple abstraction of the hydrogen, stabilized by surrounding protein, to the formation of an "enol" by subsequent abstraction of a hydrogen from a neighboring neutral histidine to form an enol-imidazolate intermediate.[7,8] Previous computational studies have addressed this question by quantum mechanical/molecular mechanical (QM/MM) techniques that consider specific points on the proposed reaction path and compare the interaction energies.[9−11] These studies have suggested that the enol-imidazolate does not form and stabilization of the enolate occurs through interactions with the protein. These previous studies are not definitive, however, because they address interaction energies, and not free energies. We perform a free energy analysis of this enzyme, comparing the enolate and enol forms of the intermediate.

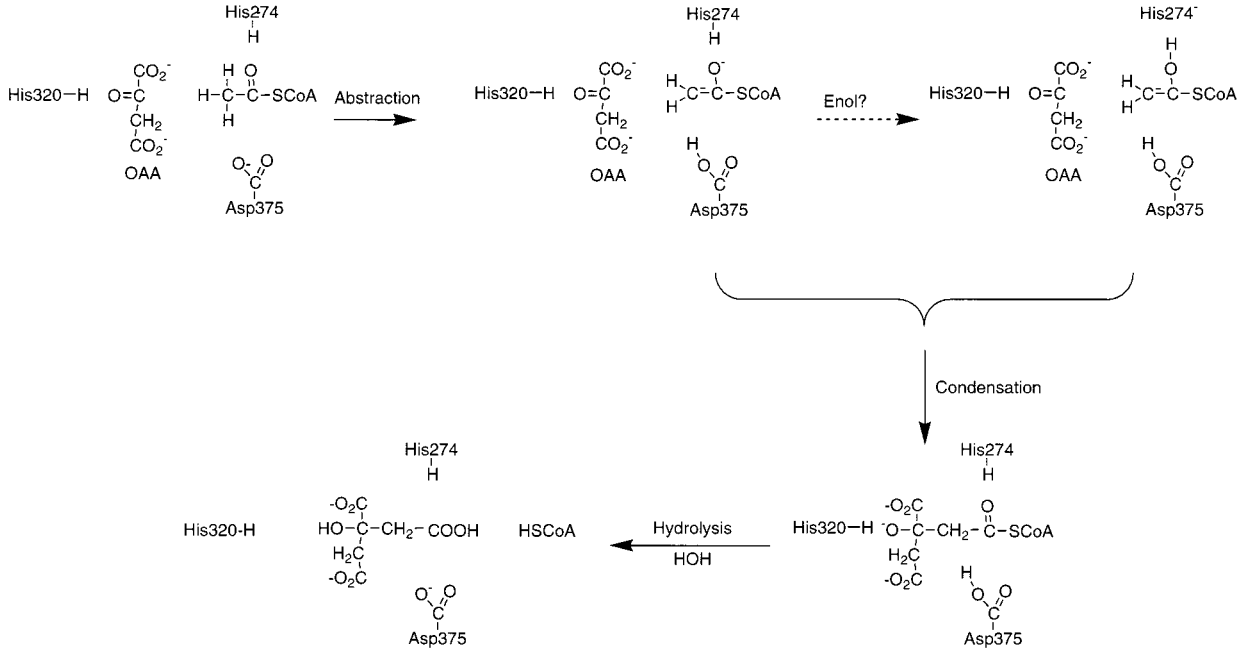
The computational analysis of enzyme reactions is difficult. Since enzymes are involved in catalyzing reactions, the bond-breaking and bond-making steps are of paramount importance. However, classical protein modeling is generally done with force fields, which intrinsically average the effects of electrons and cannot treat the bond-making and -breaking processes. There

* To whom correspondence should be addressed. Telephone: (415) 476-4637. Fax: (415) 476-0688. E-mail: pak@cgl.ucsf.edu.
† University of California at San Francisco.
‡ Laboratory of Structural Biology, National Institutes of Health.
§ Current address: Kinetek Pharmaceuticals, Vancouver, BC, Canada.

(1) Frey, P. A.; Whitt, S. A.; Tobin, J. B. *Science* **1994**, *264*, 1927−30.
(2) Gerlt, J. A.; Gassman, P. G. *Biochemistry* **1993**, *32(45)*, 11943−11952.
(3) Gerlt, J. A.; Kozarich, J. W. *J. Am. Chem. Soc.* **1991**, *113*, 9667−9669.
(4) Gerlt, J. A.; Gassman, P. G. *J. Am. Chem. Soc.* **1993**, *115*, 11552−11568.
(5) Alagona, G.; Ghio, C.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 9855−62.
(6) Beeckmans, S. *Int. J. Biochem.* **1984**, *16(4)*, 341−351.
(7) Remington, S. *Curr. Opin. Struct. Biol.* **1992**, *2*, 730−735.
(8) Kurz, L. C.; Drysdale, G. R.; Riley, M. C.; Evans, C. T.; Srere, P. A. *Biochemistry* **1992**, *31*, 7908−14.
(9) Mulholland, A. J.; Richards, W. G. *Proteins: Struct., Funct., Genet.* **1997**, *27*, 9−25.
(10) Mulholland, A. J.; Richards, W. G. *J. Phys. Chem. B* **1998**, *102*, 6635−6646.
(11) Mulholland, A. J.; Lyne, P. D.; Karplus, M. *J. Am. Chem. Soc.* **2000**, *122*, 534−535.

10.1021/ja001043i CCC: $19.00
Published on Web 11/18/2000

Scheme 1: Overall Proposed Reaction Mechanism of Citrate Synthase[23]



have been a number of methods suggested to overcome these challenges; primary among them are the empirical valence bond[12] (EVB), QM/MM, and quantum mechanical-free energy (QM-FE) methodologies. In the QM/MM approach, the active-site residues are treated in a QM framework and are linked directly to the surrounding protein atoms, which are treated in a molecular mechanical (force field) framework. To date, these calculations for the CS system have been carried out for minimizations of specific points along the reaction path as defined by QM calculations of the isolated active site.[9−11] These calculations yield interaction energies (*but not free energies*) of various points along the reaction path. On the other hand, Jorgensen has applied a QM-FE method to study organic reactions in solution.[13,14] We have further developed the QM-FE approach for enzyme systems.[15−17] In ref 16, we introduced the term QM-FE to distinguish it from more fully coupled QM/MM approaches. The parameters for the active site are derived from QM calculations for the initial and final active site geometries. These parameters are then changed throughout the calculation from the "initial" to the "final" state, and the response of the remainder of the protein is used to calculate the resulting free energy. This approach has been used previously to yield quantitative free energies for both catechol *O*-methyl transferase (COMT)-[16] and trypsin-catalyzed[17] reactions and the corresponding solution reactions.

In the course of our investigations, it became apparent that this system was extremely sensitive to the nature of the nonbonded electrostatic treatment. This is not very surprising, given the highly charged nature of the substrates. Initial simulations involving a nonbonded cutoff and simulation of the enzyme in a solvent cap revealed significant variations in the free energy to changes in simulation protocol. To overcome this, more extensive calculations, involving the use of the particle mesh Ewald treatment for nonbonded electrostatics and solvation of the protein in a periodic box of water, were required to achieve convergence.

2.0. Methods

2.1. Modeling Software. All ab initio calculations were performed with Gaussian 94 or 98.[18]

All molecular mechanics calculations were performed with the AMBER suite of programs[19] and the parm94 parameter set.[20] A constant dielectric of 1 was used throughout. Minimizations were run with the conjugate gradient method and without any constraints on bond lengths. The dynamics were run with constraints on covalent bonds to hydrogen using the SHAKE[21] algorithm, a time step of 1 fs, and with separate coupling of solute and solvent to a heat bath (Berendsen coupling[22]) at a temperature of 300 K and a coupling frequency of 0.5 ps. The free energy calculations were done with the thermodynamic integration protocol (*Gibbs* module) using windows of fixed length ($\Delta\lambda = 0.02$).

(12) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; Wiley: New York, 1991

(13) Chandresekehar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 3049−59.

(14) Severance, D. L.; Jorgensen, W. L. In ACS Symposium Series 568; Cramer, C. J.; Truhlar, D. G.; Eds.; American Chemical Society: Washington, DC, 1994; pp 243−59.

(15) Kollman, P.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. *Acc. Chem. Res.* Manuscript accepted for publication.

(16) Kuhn, B.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, 122, 2586−2596.

(17) Stanton, R. V.; Perakyla, M.; Bakowies, D.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 3448−3457.

(18) 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 94/98*; Gaussian, Inc.: Pittsburgh, PA, 1998.

(19) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1−41.

(20) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(21) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1997**, *23*, 327−341.

(22) Berendsen H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−90.

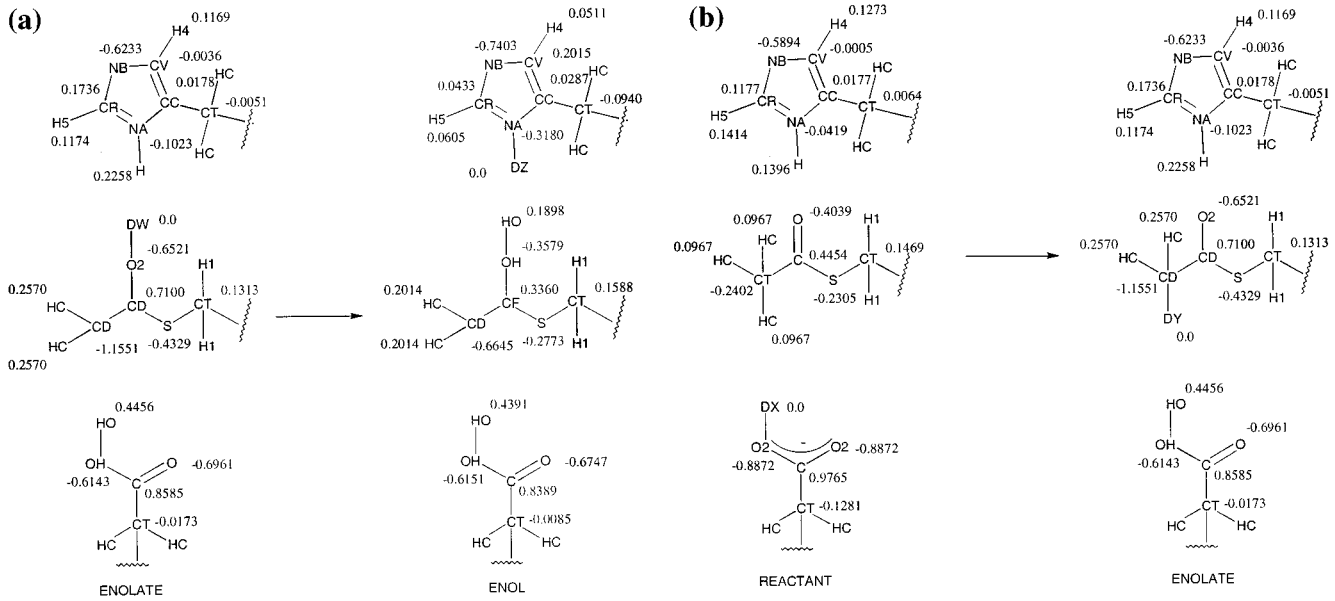


Figure 1. Atom types and partial charges used to describe the (a) reactant → enolate and (b) enolate → enol reactions. Dummy atoms are indicated by atom type D? and have zero van der Waals parameters, zero partial charge, an equilibrium bond length of 0.4 Å, and the angle and dihedral parameters of the corresponding "real" atom. Atom types CD and CF are derived explicitly to describe the enolate and enol geometries. All other atom types are from parm94. A complete listing of additional required geometrical parameters is available in the Supporting Information.

2.2. Choice of Model Structure. A number of crystal structures of citrate synthase complexed with various substrates are available. Since the hydrogen abstraction occurs in the closed form of the enzyme, such a protein structure was used. Due to the known sequence of pig (Sus Scrofa) citrate synthase, a pig crystal structure (2cts,[23] 2.0 Å) was chosen. There is also a variety of mutational and kinetic data concerning pig CS.[24−26] However, the substrate models (CMC and OAA) were taken from a chicken CS structure (5cts,[27] 1.9 Å) since they were the closest analogue to the reactive species (AcCoA and OAA). The AcCoA structure was then built from the CMC analogue (using the *xleap* module of AMBER).

2.3. Derivation of the ab Initio Model. The QM region of citrate synthase for this first hydrogen abstraction step included the side chains of his274 and asp375, and the thioacetate portion of AcCoA. These fragments were isolated from the crystal structure, and any broken covalent bonds were replaced by hydrogens.

The his, asp, and thioacetate monomers were individually geometry optimized at the Hartree−Fock level using a 6-31+G* basis set (HF/6-31+G*) with constraints. For the his and asp moieties, constraints were imposed by including the C_α, C, and N atoms of the residue backbone as dummy atoms. These atoms constituted the first three atoms in the *Z*-matrix input, and all bonds, angles, and dihedrals formed with these dummy atoms were fixed. A similar approach was taken with the thioacetate moiety where the dummy atoms included the next three heavy atoms in the composite AcCoA molecule.

After geometry optimization of the individual monomers, scans of the asp−thio and his−thio dimer surfaces were performed. The asp−thio OD2−methyl C distance was scanned with partial optimization of the methyl hydrogens. A minimum in the asp−thio distance was found at 3.2 Å; however, the well was broad with all geometries between 2.9 and 3.5 Å within 1 kcal/mol of the minimum. A similar scan of the his ND−thio carbonyl C distance was performed with a partial optimization of the histidine ND hydrogen and thioacetate carbonyl oxygen. A minimum in the ND−C distance was found at 3.9 Å (corresponding to an H−O distance of 2.13 Å); however, the potential surface was again quite broad with distances greater than 3.7 Å, yielding structures within 1 kcal/mol of the minimum. A short molecular dynamics simulation of the citrate synthase system was then performed to determine if these ab initio generated distances were feasible within the enzyme. The ab initio minimum energy distances were attained during the dynamics, although these distances were generally at the minimum end of the distance distributions.

The monomer positions were then built into the enzyme using the overlap with a congruent structure from the short exploratory MD simulation. The resulting "reactant" trimer was again isolated, and ab initio geometry optimization was performed on the complete trimer at the HF/6-31+G* level with the terminal methyl groups on the asp, thioacetate, and his moieties frozen to constrain the trimer to the enzymatic framework.

The enolate trimer was then formed by geometry optimization after transferring the appropriate hydrogen. The terminal methyls as well as the internal coordinates of the histidine ring were constrained. The reactant → enolate transition was also monitored by determining the energies of "intermediates" along the pathway. These were constructed by fixing the $H_{transfer}$−OD2 distance and allowing partial optimization of the resulting structures (i.e., the terminal methyls and histidine internal coordinates were frozen).

Finally, the enol trimer was formed by transferring the hydrogen from the histidine of the final optimized enolate structure to the thioacetate carbonyl group. Again, the terminal methyls and the aspartate internal coordinates were fixed during the proceeding minimization.

Finally, to accurately determine the ab initio energy difference between the reactant, enolate, and enol species, single point calculations using the second-order Moller−Plesset perturbation approach to the correlation energy (MP2/6-31+G*) were performed on the final geometries.

2.4. Determination of Parameters. Charges. The partial charges of the reactant, enolate, and enol trimers were determined using the restrained electrostatic potential (RESP) protocol[28,29] on the HF/6-31+G* generated electrostatic potentials of the final optimized structures (Figure 1). The trimers were constrained to have a net charge

(23) Remington, S.; Wiegand, G.; Huber, R. *J. Mol. Biol.* **1982**, *158*, 111−152.
(24) Zhi, W.; Srere, P. A.; Evans, C. T. *Biochemistry* **1991**, *30*, 9281−9286.
(25) Evans, C. T.; Kurz, L. C.; Remington, S. J.; Srere, P. A. *Biochemistry* **1996**, *33*, 10661−10672.
(26) Kurz, L. C.; Nakra, T.; Stein, R.; Plungkhen, W.; Riley, M.; Hsu, F.; Drysdale, G. R. *Biochemistry* **1998**, *37*, 9724−9737.
(27) Karpusas, M.; Branchaud, B.; Remington, S. J. *Biochemistry* **1990**, *29*, 2213−19.
(28) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(29) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.

Table 1: Internal Constraints Used in FEP Simulations

reactant → Enolate			enolate → enol		
constraint	eqb. value	K_{eqb}	constraint	eqb. value	$K_{eqb.}$
Asp OD2−Thio C_{Me}	3.15 Å	100 kcal/mol Å	Asp OD2−Thio C_{Me}	3.15 Å	100 kcal/mol Å
Thio $O_{carbonyl}$−His HD1	2.10 Å	100 kcal/mol Å	Thio $O_{carbonyl}$−His ND1	3.1 Å (enolate)	100 kcal/mol Å
			Thio $O_{carbonyl}$−His ND1	2.9 Å (enol)	
His ND1−His HD1−Thio $O_{carbonyl}$−	175°	50 kcal/mol deg	His ND1−Thio $O_{carbonyl}$−Thio $C_{carbonyl}$	125° (enolate)	50 kcal/mol deg
			His ND1−Thio $O_{carbonyl}$−Thio $C_{carbonyl}$	115° (enol)	
Asp CG−Asp OD2−Thio C_{Me}−	145°	50 kcal/mol deg	Asp CG−Asp OD2−Thio C_{Me}	145°	50 kcal/mol deg
His ND1−Thio $C_{carbonyl}$ −Asp CG	125°	50 kcal/mol deg	His ND1−Thio $C_{carbonyl}$−Asp CG	125°	50 kcal/mol deg

of −1. All hydrogens attached to carbons forming a covalent bond to the protein (i.e., "link" atoms) were set to zero partial charge.

The oxaloacetate (OAA) charges were generated from a single point HF/6-31+G* calculation of the OAA structure as found in the crystal structure using the restrained electrostatic potential methodology (RESP) methodology.[28,29] The acetyl-CoA (AcCoA) charges were determined by breaking the group into two fragments. The first fragment incorporated the thioacetate moiety and all atoms up to the first methyl-phosphate oxygen bond. The second fragment included the methyl attached to the first phosphate oxygen and all other atoms. A single point HF/6-31+G* calculation yielded the electrostatic potentials of each fragment. The first fragment partial charges were fit using the RESP methodology *after* fixing the thioacetate partial charges to those found in the reactant trimer. The link hydrogens on the carbons at the interface between the thioacetate and the remainder of AcCoA and between fragments 1 and 2, were set to zero. The electrostatic potential of the second fragment was fitted *after* fixing the adenine moiety charges to those determined by Cornell et al.[20] and setting all link hydrogens to zero.

Parameters. Parameters were required for the thioacetate and methyl phosphate moieties of AcCoA, for oxaloacetate, and for the enolate and enol forms of the thioacetate. The parameters were derived from a mixture of information including ab initio generated torsional potentials (thioacetate: CT−S−C−CT, OAA: C−C−CT−C, O2−C−C−CT, enolate: H−C−C−S, C−C−S−CT), cumulated average bond length data,[30] crystal structures of similar moieties (pyruvate as a mimic for OAA[31]), and vibrational frequencies (thioacetate[32] and pyruvate for OAA[33]). In addition parameters for the enol were derived from existing AMBER parameters for hydroxyl groups.[20] All atom types and geometric parameters are provided in the Supporting Information.

2.5. Free Energy Simulations. Solution Reaction. The reactant and enolate trimers with link hydrogens were each embedded in a periodic box of water ∼45 × 33 × 31 Å^3. The net charge of the systems was −1.

Molecular dynamics calculations were carried out to equilibrate the system. Restraints were imposed in each case to keep the geometries close to the QM determined one (see Table 1). A nonbonded cutoff of 13 Å was used. Equilibration was achieved after minimization (2000 steps conjugate gradient), heating and equilibration (total 40 ps for the reactant → enolate simulation and 31 ps for the enolate → enol simulation). Equilibration was assessed by monitoring fluctuations in the temperature and total and potential energies, as well as the density of the periodic box.

Upon equilibration, a free energy calculation was performed from the reactant → enolate and from the enolate → enol. The free energy was calculated over 150 ps with $\Delta\lambda = 0.02$. At each value of λ, 1000 steps of equilibration and 2000 steps of data collection were performed. The free energy calculation *excluded* intra-perturbed group interactions since these are accounted for in the quantum mechanical (QM) energy previously determined.

Enzymatic Reaction. The following procedure was followed for both the isolated monomer and the model dimer system, and is described for the model dimer system only since it is the larger model system.

The reactant and enolate trimers were inserted into the 2cts protein crystal structure. The second monomer of the CS dimer was constructed and viewed. Residues 163−165 (including Arg164 to bind AcCoA phosphate) and 419−424 (including Arg421 to bind OAA) were then isolated from the "second monomer". These two strings of residues were capped with an acetyl group and an *N*-methyl group on the N- and C-terminals, respectively. The final protein system then included one complete monomer and the 163−165 and 419−424 "pieces" of the second monomer to form a complete active site. All crystallographic waters were retained. The resulting system was then placed in a periodic box of water (final box size ∼87 × 90 × 96 Å^3) and five Na^+ counterions were added to neutralize the system. The ions were positioned with the "addions" functionality of *tleap*, which defines an electrostatic grid and places the ions at the most negative points on that grid. In the reactant → enolate system, the ions were at least 20 Å distant from the active site methyl carbon in thioacetate while in the enolate → enol system they were at least 19 Å from the methyl carbon in thioacetate and 21 Å from His274 ND1. The total system is ∼61000 atoms. Restraints were imposed to keep the active site geometry close to the QM determined one (Table 1). Finally, in the dimer model system only, harmonic positional restraints ($K = 500.0$ kcal/molÅ^2) on the N- and C-terminals of the monomer and 163−165 and 419−424 "pieces" were used to anchor the dimer structure during the simulations. All histidines (except his274 in the active site) were assumed to be neutral, although there is some evidence[34] that, in the D375G enzyme, His274 is positively charged. With particle mesh Ewald, one would expect additional charges, as long as the entire system is neutral, to be reasonably accurately described. If anything, a positive His274 would further stabilize an enolate over an enol even more than found below.

Initial simulations of this system in a cap of water (as opposed to a periodic box) indicated an extreme sensitivity of the free energy results to the nonbonded cutoff and to the identity of the moving atoms (i.e., the residues included in the belly). Thus, these simulations were carried out using the particle mesh Ewald (PME) procedure implemented in the *Sander* and *Gibbs* modules of AMBER.[35] Within the *Gibbs* module, additional modifications to the AMBER5.0 release were required to *exclude* intra-perturbed group interactions when using the PME method. While direct intraperturbed group electrostatic interactions are excluded, one perturbed group interacting with the image of another perturbed group is not excluded; however, given the distances involved, this contribution should be negligible. The nonbonded cutoff for the van der Waals interaction and the switch between the direct and Fourier spaces of the PME procedure occurred at a cutoff of 9 Å.

Equilibration was a multistep procedure. Initially the water positions only were energy minimized for 4000 steps of conjugate gradient minimization. Then the total system was minimized for 4000 steps as well. Heating and equilibration of the waters *only* was then carried out for a total of 20 ps. Finally, heating and equilibration of the system as a whole was performed (41.5 ps for the reactant → enolate reaction and 37.5 ps for the enolate → enol reaction). Equilibration was again assessed by monitoring temperature, total and potential energies, and density fluctuations of the system.

(30) Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen, A. G.; Taylor, R. *J. Chem. Soc., Perkin Trans. 2* **1987**, S1−S19.

(31) Kennard, O.; Watson, D. G.; Allen, F. H.; Isaacs, N. W.; Motherwell, W. D. S.; Pettersen, R. C.; Town, W. G. *Interatomic Distances 1960-1965: Organic and Organometallic Crystal Structures. Molecular Structures and Dimensions*; Utrecht: Cambridge, 1972; Vol. A1.

(32) El-Aasar, A. M.; Nash, C. P.; Ingraham, L. L. *Biochemistry* **1982**, *21*,: 1972−76.

(33) Keller, R. J. *The Sigma Library of FT-IR Spectra, first ed.*; Sigma Chemical Company: Missouri, 1986; *Chem. Abstr.* **1990**, *113*, 24−26).

(34) Gu, Z. T.; Drueckhammer, D. G.; Kurz, L.; Lui, K.; Martin, d. P.; McDermott, A. *Biochemistry* **1999**, *38*, 8022−8031.

(35) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−92.

Free energy determinations were carried out using the coarse-grained parallel approach. Specifically a starting structure for each value of λ was generated from a fast (~0.1 ps) free energy run generating a rapidly accessible manifold of structures.[36] These structures are then each used in parallel free energy (FE) calculations for each individual value of λ. Thus, each parallel node, for example, would be running a separate $\Delta\lambda$ window. This approach makes extending the simulation to check for convergence much easier, as it is necessary only to continue the dynamics in each independent $\Delta\lambda$ window as opposed to re-running the entire free energy calculation. The $dH/d\lambda$ values from each window are collected, and the integration is done using the trapezoidal rule. These simulations were run on a 4-processor Origin R10000 and a single $\Delta\lambda$ window of 2 ps required ~6.3 CPU h.

The use of the PME protocol does preclude the determination of free energy components during the FE determination. It is not possible to separate individual components of the electrostatic interaction and thus the electrostatic contribution to the free energy components of each residue cannot be assessed.

Finally, results are also presented for simulations on the isolated monomer using both PME (as described above) and a nonbonded cutoff. In the latter case, a solvent cap was centered on the CG of the active-site aspartate (radius: 24 Å), and only atoms within 14 Å of the active-site aspartate were allowed to move during the simulations. The phosphate "side chain" of the AcCoA group is modified to have a net −1 versus −2 charge because the larger charge induces SHAKE errors in interactions with neighboring atoms. A variety of nonbonded cutoffs were employed. Heating and equilibration were carried out as described for the periodic simulations. The free energy calculations were perfomed without requiring the coarse-grained parallel approach and used a nonbonded cutoff rather than the PME electrostatic treatment. These results are presented for comparison.

2.6. Cratic Free Energy Contribution. The cratic free energy contribution to the solvation free energy of the reference reaction must be included to calculate a reasonable reference reaction.[14] This contribution includes the desolvation and entropic cost of bringing the three components of the trimer together in solution. The desolvation cost is determined using the PCM model of G94. Single point polarized continuum model (PCM)[37] calculations were performed for the optimized monomers and reactant trimer. For each PCM calculation, a united atom Hartree−Fock (UAHF) approach was used with the 6-31+G* basis set and a pentakisdodecahedral cavity. The entropic cost for bringing together the substrates is semiquantitatively estimated using the *NMODE* module of AMBER. The monomers and reactant trimer are minimized with a distance dependent dielectric of $4r_{ij}$ (where r_{ij} is the nonbonded distance between atoms i and j) to a tolerance of 8×10^{-5} in the gradient rms using an adopted basis Newton−Raphson minimization. The resulting structures are then used to perform a normal-mode analysis from which the entropy is determined using classical equations (*NMODE* module of AMBER). The sum of the entropic and PCM differences between the reactant trimer and the constituent monomers then forms an estimate of the cratic free energy contribution.

2.7. Energetic Analysis of Free Energy Contributions. While the use of PME precludes the calculation of free energy contributions on a per-residue basis, it is possible to examine the interaction energy contributions to the free energy calculation. Thus, we analyze the interaction energies at the two endpoints ($\lambda = 1$ and $\lambda = 0$) of the free energy simulation.

The total interaction energy difference as a function of λ can be estimated as the difference in potential energies between the $\lambda = 1$ and $\lambda = 0$ states minus the internal perturbed group interactions in these same states (since this was excluded from the FE calculation and included in the QM calculation). The average potential energy is taken from the averages printed in the Gibbs output, using 3000 MD steps in the solution simulation, and 2000 in the enzyme. The perturbed group interactions with each other are calculated separately for the saved snapshots using three snapshots for the solution simulation and 20 for the enzyme using a nonbonded cutoff of 999 Å and the ANAL module of AMBER. The net difference is the change in interaction energy. Since the total free energy change over the simulation is known, subtraction of the net interaction energy from the total free energy then yields an estimate of the entropy. Finally, it is also possible to calculate the interaction energy of the $\lambda = 1$ and $\lambda = 0$ states with the first hydrogen-bonding solvation shell of the active site, as shown in Figure 3a, using a nonbonded cutoff of 999 Å. The "first solvation shell" of the solvent reaction was determined by selecting all waters within 3.0 Å of at least one active-site residue, for each snapshot. Subtraction of this "first shell" interaction energy from the net interaction energy illustrates the long-range effect of the protein or the solvent, respectively.

Table 2: Ab Initio Geometries of Trimer Complexes

trimer	HF/6-31+G*	MP2/6-31+G*
reactant ($R = 2.0$ Å)[a]	0[b]	0[b]
inter 1 ($R = 1.75$ Å)	1.8	0.7
inter 2 ($R = 1.48$ Å)	10.4	7.2
inter 3 ($R = 1.20$ Å)	26.7	19.5
enolate ($R = 0.95$ Å)	18.8	18.4
enol	30.1	26.7

[a] Single point calculations were performed on the geometries shown in the Figure 1 for the reactant, enolate, and enol. R represents the distance between the $H_{transfer}$ moving from thioacetate to aspartate and OD2 of the aspartate. [b] Energies are given in kcal/mol relative to the reactant trimer in each column.

3.0. Results

3.1. QM Calculations. The energy difference between reactant and enolate is 18.4 kcal/mol, while the formation of the enol costs an additional 8.3 kcal/mol (Table 2). The energy of various "intermediates" along the reactant → enolate pathway are also included. We are implicitly assuming that the energy difference between the reactant and enolate and enol intermediates approximates that of the true transition states. This assumption is reasonable, since only "Inter 3" is found to have a higher energy than the enolate intermediate used as an endpoint. Furthermore, at the MP2 level of calculation this energy difference is only 1 kcal/mol, which is within the error of the combined QM+FE calculations.

3.2. Free Energy Simulations.. The overall free energy of the reaction is calculated using the relationship[16]

$$\Delta G^* = \Delta E_{QM} + \Delta G_{int} \quad (1)$$

ΔE is calculated using MP2/6-31+G* single point energies for representative intermediates along the reaction path. ΔG_{int} represents the interaction free energy between the protein, solvent and counterions (or solvent for the reference calculation), and the active-site atoms, as the active-site atoms are mutated from the initial to final structures.

In CS, one of the primary questions that remains unresolved is the relative stability of the proposed enolate and enol intermediates of the first hydrogen abstraction step. Due to the complexity of the reaction mechanism,[38] which involves hydrogen abstraction, condensation ,and hydrolysis, it has been impossible to answer this question definitively with experimental information. Thus, we calculate the free energies required to "mutate" the reactants into the enolate and then into the enol. The resulting free energies indicate the relative stabilities of these two intermediates, and the likelihood of their formation. These calculations are carried out both within the protein, and within solvent as a reference reaction. The energetic results are summarized in Table 3 and Figure 2.

(36) DeBolt, S. E.; Pearlman, D. A.; Kollman, P. A. *J. Comput. Chem.* **1994**, *15*, 351−373.

(37) Tomasi, J.; Cammi, R.; Mennucci, B. *Int. J. Quantum Chem.* **1999**, *75*, 783−803.

(38) Kurz, L. C.; Drysdale, G.; Riley, M.; Tomar, M. A.; Chen, J.; Russell, R. J. M.; Danson, M. J. *Biochemistry* **2000**, *39*, 2283−2296.

Table 3: Free Energy Components

	reactant → enolate (kcal/mol)	enolate → enol (kcal/mol)
solution - *FEP*	17.9	6.5 kcal/mol
solution - *cratic*	9.4 (PCM) + 19.6 (ENT) = 29.0	N/A
enzymatic - monomer cut = 14/44[a]	−12.6 (150 ps)	−33.4 (150 ps)
enzymatic - monomer cut = 12/42[a]	−22.3 (50 ps)	N/A
enzymatic - monomer	−3.0 (100 ps), −3.0 (200 ps)[b]	+4.3 (100 ps)
PME		+4.3 (150 ps)[b]
enzymatic - dimer	−3.1 (100 ps)	+5.1 (100 ps)
PME		+5.3 (150 ps)[b]

[a] Phosphate charges on the AcCoA are scaled back to prevent SHAKE errors, no counterions are used and the *x*/*y* cut indicates a nonbonded cutoff of *y*Å for the perturbed groups and *x*Å for all other residues. [b] The MD simulation in each $\Delta\lambda$ window is continued for an additional 1 or 2 ps yielding a total simulation length of 150 or 200 ps, respectively.

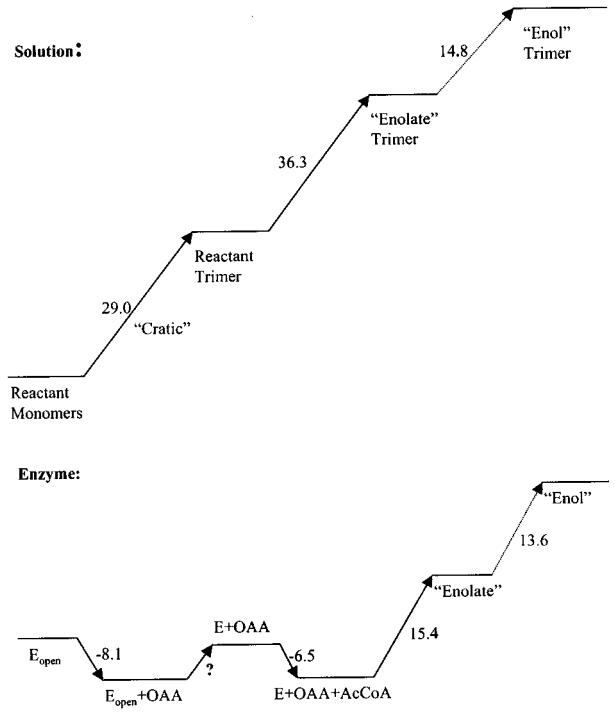


Figure 2. Reaction profiles of the (a) solution and (b) enzymatic reactions. These profiles are based on the assumption that the barrier between the initial and final intermediates is very similar to the barrier to the transition state. In the enzymatic profile, E_{open} indicates the open form of the enzyme, while E is the closed form. The step between E_{open} and E represents the structural rearrangement that occurs upon OAA binding, and is of unknown free energy. The OAA and AcCoA binding energies are derived from K_i values of OAA alone (K_d = 1.2 μM[8]) and AcCoA in the presence of OAA ($K_M = 1.6 \times 10^{-5}$ at pH 8.0[38]). Calculated barriers are derived from the MP2/6-31+G* QM energies and the PME dimer free energies.

The free energy barriers in the solvent calculation (Figure 2a) are found to be quite large. This is expected since the abstraction of an aliphatic hydrogen by a weak base such as aspartate is known to be extremely unfavorable. These barriers are further increased by the cratic energy term,[15] which we estimate to be 9 kcal/mol (PCM) and 19 kcal/mol (normal-mode analysis). As noted above, these two estimates come from calculation of the ab initio energy of the trimer compared to the three monomers including the continuum solvation free energy using the PCM model[18] (PCM contribution) and carrying out classical mechanical normal-mode analysis of the reactant trimer and the three monomers and calculating the $T\Delta S$ from the difference in the translational, rotationa, and vibrational entropies of trimer versus monomers. These estimates are crude, particularly since the use of the normal-mode analysis to estimate solute entropy is a very approximate model. Nonetheless, bringing charged molecules together can cost considerable free energy, and aligning three fragments was calculated to cost 9−11 kcal/mol in trypsin.[17] Even if we use this latter number, the total cratic free energy cost is still ∼20 kcal/mol.

The enzymatic free energy calculations were quite challenging. In particular, the highly charged nature of the active site (including the OAA and AcCoA substrates) resulted in an extreme sensitivity of the results to the simulation protocol. Initial simulations were carried out by solvating the active site in a solvent cap and allowing only a subset of the atoms around the active site to move. Various nonbonded cutoffs were then used to ensure convergence. Unfortunately, the free energy results varied with respect to the placement of the counterions, the subset of atoms allowed to move and the size of the nonbonded cutoff (e.g., Table 3, rows 4, 5). The enolate → enol reaction in particular was very cutoff-sensitive and could give rise to large, nonsensical free energy values (Table 3, row 4). Previous QM-FE calculations[16,17] have not shown this degree of sensitivity to cutoff, presumably because the substrates and the proteins were not so highly charged.

As a result of this sensitivity, the calculations were repeated using solvation in a periodic box, allowing all atoms to move and including all nonbonded interactions using the PME option in AMBER (see section 2.5.). The convergence of these results was checked in two ways. First, the calculations were carried out with 2 ps of dynamics per window of thermodynamic integration and then extended for a further 1 or 2 ps of dynamics per window (Table 3). This ensured that each window had indeed converged. Sensitivity to simulation starting conditions was further tested by re-running the calculations for both the monomer and the dimer. The dimer model included an additional two arginines and one glutamate, as well as a different placement of the counterions. The similarity of the resulting free energy differences (Table 3, rows 6, 7) confirms that with the inclusion of all nonbonded interactions, the results are stable.

The enzymatic reaction profile is shown in Figure 2b. The profile includes all reaction steps up to the postulated formation of the enol. In particular, the binding of the OAA and AcCoA substrates are included, and the relevant free energies are estimated from experimental information. It is difficult to ascertain the free energy cost that accompanies the structural rearrangement of CS from the open to closed forms upon binding OAA. This cost could potentially be included in the OAA K_d used to estimate the free energy of OAA binding, and this ambiguity is indicated by a question mark in Figure 2b. As is seen in the solvent reaction, the further formation of the enol from the enolate is even more costly and unlikely to occur. Significantly, while the enzyme stabilizes the formation of the enolate as compared to the solution reaction (FE_{enz} = −3.0 kcal/mol, FE_{solv} = 17.9 kcal/mol), the formation of the enol is not stabilized (FE_{enz} = 5.3 kcal/mol, FE_{solv} = 6.5 kcal/mol).

3.3. Enzymatic Structural Features. Analysis of the residues in close contact with the active-site residues (asp375, his274 and thioacetate) gives some insight into the structural features that stabilize the formation of the enolate. Comparison of the

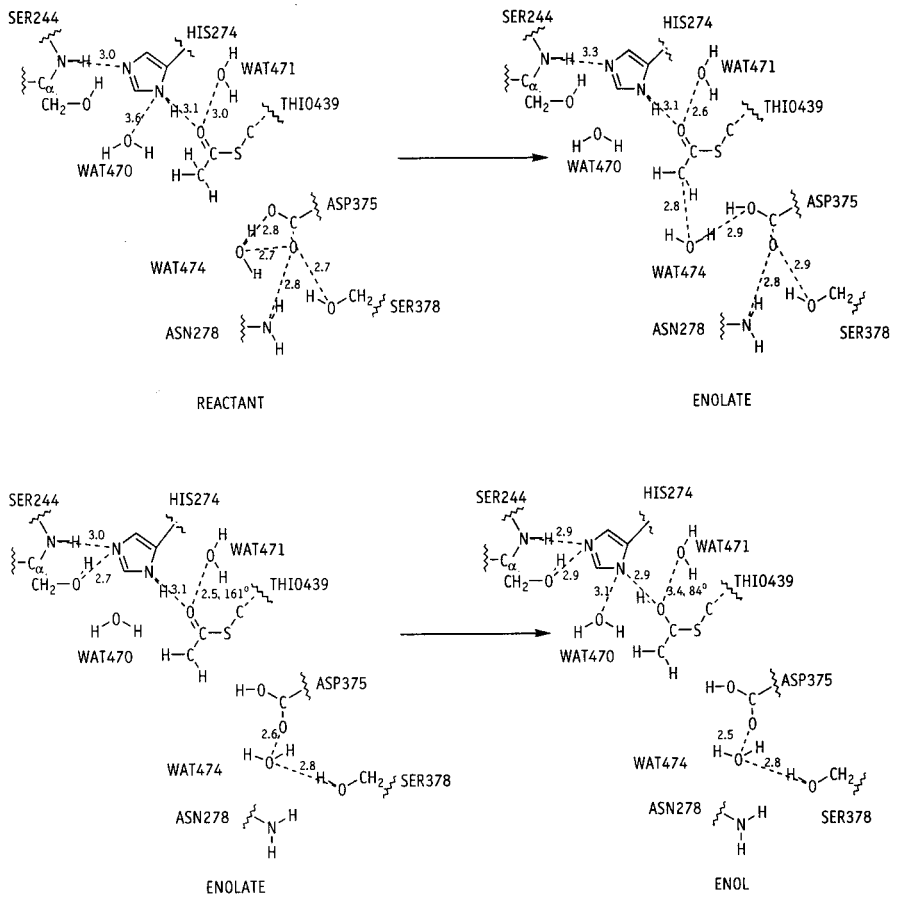


Figure 3. Representative snapshots of initial and final structures in the (a) reactant → enolate and (b) enolate → enol model dimer reactions. Residues that are observed to interact directly ("first solvation shell") with the active site (thio439, his274, asp375) are indicated. Note that the distances given represent a single snapshot only and do not indicate the degree of structural flexibility that occurs over the simulation. All distances are between heavy atoms only.

initial equilibrated starting structure in the model dimer simulation to that obtained after perturbation into the enolate (Figure 3a), as well as the initial equilibrated structure of the enolate to that of the enol (Figure 3b) indicates that the primary residues of interest are Ser244, Ser378, Asn278, and three waters (470, 471, and 474 in our numbering scheme). The initial starting structure has a hydrogen-bonding network between asp375 and the neighboring serine and asparagine, as well as a water molecule. His274 appears to interact with both the backbone NH and side chain OH of ser244 (OH interaction >3.0 Å) on the NE side of the ring, while the ND side has a distant interaction with a water and a close hydrogen bond with the thioacetate (which is constrained throughout the dynamics: Table 2). After the free energy perturbation, the resulting enolate maintains the same interactions about the his274, while the hydrogen bonding around asp375 is variable. In this case, the water is bridging between the enolate and asp375, while the other oxygen of aspartate is stabilized by asn278 and ser378. Interestingly, in the monomer only calculation wat474 is missing, and the free energy difference is not changed significantly as a result. In contrast the water in the equilibrated starting structure for the enolate → enol transition has a similar interaction pattern for his274 (note that the ser244 OH−his274 NE interaction is highly variable with an average and standard deviation of 3.4 ± 0.4 Å over 20 ps of equilibration). In this case, wat474 has inserted between asp375 and the neighboring ser378. The water remains in this position throughout the free energy calculation. The variability of this water position in different simulations indicates that it is not tightly bound. On the other hand, the water that hydrogen bonds to the thioacetate carbonyl (wat471) is always found in that position in any enolate or starting structure. Wat471 is equivalent to wat585 in the previous QM/MM calculations.[9−11] In the enol structures, this water becomes more loosely bound, and the angle of the hydrogen bond is not as favorable (Figure 3b). Upon extension of the enolate → enol simulation, the wat471−thioacetate carbonyl interaction becomes more favorable again ($O_{HOH}-O_{thio\ carbonyl} \approx 2.9$ Å). Lengthening the MD simulation in each $\Delta\lambda$ window for the monomer only reactant → enolate, and the monomer and dimer enolate → enol simulations did not yield any other qualitatively different structural features around the active site. The free energy values of these simulations also did not change significantly (Table 3).

3.4. Comparison of Enzyme and Solution Reactions. The enzyme and solution reactions can be compared on both a structural and energetic level, in an attempt to hypothesize how the enzyme stabilizes the formation of the enolate. Figure 4 illustrates the first hydrogen-bonding solvation shell that is found in the reactant → enolate transition. There are a number of species hydrogen bonding to the polar his and asp moieties, as is expected. It is interesting to note that, in comparison to Figure 3, the number of hydrogen bonds to the enolate asp and thioacetate moieties is significantly greater.

Energetically, the interaction of the active-site residues with the all of the surroundings, with the first solvation shell, and with the surrounding minus the first solvation shell, are given in Table 4. In addition, the energy for the individual active-site residues interacting with the solvation shell is also given. It is

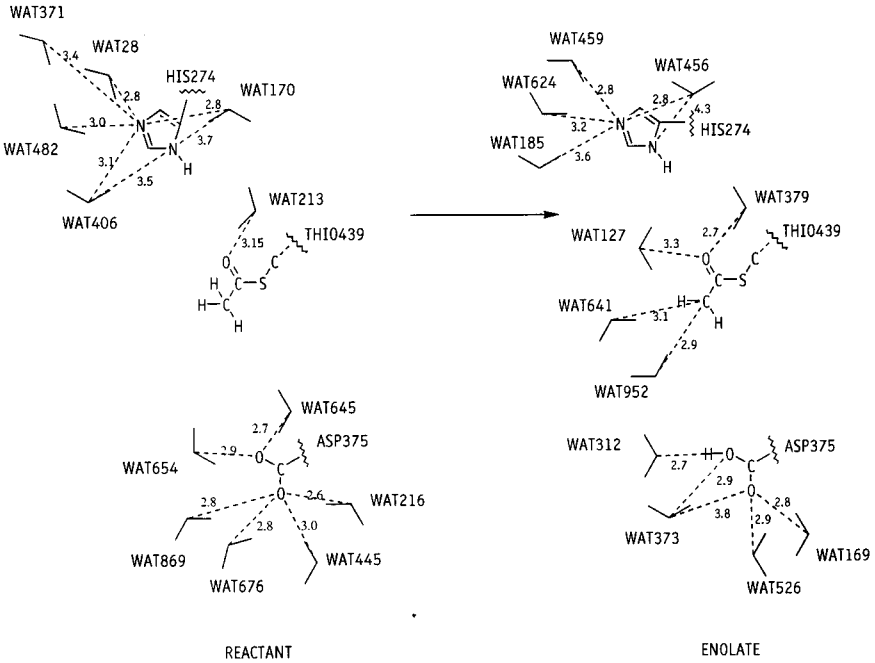


Figure 4. Representative snapshots of the initial and final structures in the reactant → enolate solution reaction. Residues that are observed to interact directly with the active site ("first solvation shell") are indicated. Note that the distances given represent a single snapshot only and do not indicate the degree of structural rearrangement that occurs over the simulation. All distances are between heavy atoms only.

Table 4: Free Energy Contributions[a]

	enzyme (kcal/mol)	solution (kcal/mol)
calculated free energy	−3.0	17.9
ΔE_{inte} (all)	−9.95 ± 160.5	−21.4 ± 44.6
$-T\Delta S$	6.95 ± 160.5	39.3 ± 44.6
ΔE_{inte} (first shell)	24.6 ± 5.8	5.1 ± 16.4
ΔE_{inte} (rest)	−34.6 ± 160.6	−28.8 ± 47.5
ΔE_{inte} (asp, first shell)	54.8 ± 4.3	54.1 ± 9.3
ΔE_{inte} (his, first shell)	0.6 ± 2.0	1.5 ± 6.8
ΔE_{inte} (thio, first shell)	−30.8 ± 3.0	−51.2 ± 12.3

[a] All differences are for $E_{product} - E_{reactant}$; therefore any negative contributions favor the product (enolate) while positive contributions favor the reactant. The calculated free energy is taken from Table 3, and interaction energy is calculated as described in section 2.7. The entropy is calculated $\Delta G_{FE} = \Delta E_{inte,\ all} - T\Delta S$. The interaction energy of the first shell is calculated as described in section 2.7., while the interaction energy with the rest of the protein is calculated from, $\Delta E_{inte,rest} = \Delta E_{inte,\ all} - \Delta E_{inte,\ first\ shell}$. Individual interaction energies for each active-site residue with its first shell are also given and are calculated individually as described in section 2.7. All values are given as the average ± standard deviations. Standard deviations for the total interaction energy are quite large since they represent differences between two large numbers (the total potential energies for each state).

apparent that the thioacetate in the enolate is *less* stabilized in the enzyme than in the solution reaction.

4.0. Discussion

4.1. Comparison to ΔG^*. Unfortunately there is no unequivocal experimental information regarding the role of specific residues in the stabilization of the enolate or the existence of an enolate as opposed to an enol. However, the overall k_{cat} of the reaction is 90−125 s^{-1} [24−26], which is equivalent to a $\Delta G^{\ddagger}$ of 14.6−14.8 kcal/mol. In some studies[38] hydrogen abstraction is believed to be the rate-determining step. Our value of 15.4 kcal/mol for the reactant → enolate step is reasonable in this respect. The further formation of the enol from the enolate is clearly disfavored in this context, as it would increase the cost of the reaction by ∼13 kcal/mol (Figure 2).

Experimentally, the active-site residues are found to be tightly coupled, and small mutations, especially at his320, cause differences in multiple steps of the reaction.[25] Further work[38] suggests that both condensation and subsequent hydrolysis of the citoryl thieoester intermediate are rate-limiting. Given the uncertainty in our calculated ΔG^* for the initial proton-transfer step, all we can say is that it is of the same order of magnitude as that of the observed free energy, but we need to study the further steps in the reaction for a definite conclusion. This information indicates that while the hydrogen abstraction step is believed to be the rate-determining step, the other reactions (i.e., condensation and hydrolysis) are likely to have reaction barriers of a similar overall magnitude.

4.2. Previous Computational Studies. QM calculations on both the asp−thio and asp−thio−his triad in both reactant and enolate forms were reported by Mulholland and Richards.[10] These authors did unrestrained optimizations, whereas we restricted ours to conform to the active site. For the asp−thio diad, they found a 7.7 kcal/mol difference between reactant and enolate at the AM1 level and 17.0 kcal/mol at the MP2/6-31+G*. They studied the triad only at the AM1 level, finding an energy difference of 2.4 kcal/mol. If we assume that the difference between AM1 and MP2/6-31+G* for the diad can be extrapolated to the triad, their energy difference between reactant and enolate is estimated to be 2.4 + 9.3 = 11.7 kcal/mol. Our hydrogen bond distances are based on the reactant geometry; thus, stimulated by a reviewer's comments, we optimized only the two H-bonded distances in the enolate, keeping the remaining geometrical parameters as they were. This led to a decrease in the enolate energy of 1.8 kcal/mol and a shortening of the $O_{thio}-N_{his}$ distance from 3.1 to 2.97 Å (compared to 2.8 Å in ref 10). This leads to a ΔE of 17.4 kcal/mol for enolate compared to that for reactant, larger than the estimation based on the calculations in ref 10 of 11.7 kcal/mol. The difference in these values certainly corresponds to a reasonable amount of enzyme strain on the relative QM energies, which could be easily compensated by differential interactions with other enzyme residues.

More extensive[9] semiempirical QM/MM calculations have also been performed on this enzyme. In this case, it is more

difficult to compare the energies with those fround here, since these QM calculation are carried out at the AM1 level and the MM results are interaction energies as compared to our free energies. Nonetheless the qualitative results of these QM/MM calculations do indicate that the enolate has more favorable interaction energies than the enol, as found here.

Recently, Mulholland et al,[11] have extended their earlier work by applying ab initio methods within QM/MM at both the Hartree−Fock and MP2 levels (6-31G* basis sets), using energy minimization on reactant, enolate, and enol complex as well as two structures intermediate between these three. It is strange that the quantum mechanical energies are actually lower, by 4 kcal/mol for the enolate with MP2 (higher by 2 kcal/mol with Hartree−Fock) than for the reactant, very different from our results, where the enolate is a seemingly more reasonable 16−18 kcal/mol higher in energy than the reactant. Furthermore, they find that the enzyme environment net *destabilizes* the enolate relative to the reactant, presumably due to the anionic groups of the substrates. Nonetheless, as has been found previously by Mulholland et al,[9,10] and by us, the enol form is always less stable than the enolate.

4.3. Advantages and Limitations of the QM-FE Methodology. The EVB, QM/MM, and QM-FE methodologies are useful in the study of enzymes, as they allow both a realistic accounting of the bond-making and -breaking processes and a consideration of environmental effects with molecular mechanics/dynamics. The ability to calculate free energies is the major advantages of this method over the QM/MM approach. Ways to include free energy in QM/MM approaches have been described by Benzion et al,[39] Gao et al,[40] and Yang et al. [41,42] Benzien et al.[39] use an EVB potential as a reference potential in a QM (ab initio)/MM approach and calculate the free energy between target and reference potential. Gao and co-workers use a potential of mean force (PMF) approach to "drive" the system along a designated reaction coordinate. An approach developed by Yang[41,42] is closest in spirit to the one presented here; QM/MM calculations are carried out in an efficient iterative optimization approach, and then subsequent classical free energy calculations are carried out along this path. Each of these approaches can be usefully applied to interesting systems, as can QM-FE.

While recent advances in QM/MM calculations have allowed higher level ab initio calculations (HF/6-31G*) to be included, they are still time-intensive enough to preclude the calculation of dynamics, and hence free energy changes. Also, QM/MM methods inevitably require many more energy evaluations and gradient evaluations than required in the QM-FE used here and thus cannot go to as high an ab initio level with the available computational resources as one can with QM-FE.

One normally does not include solute entropy in QM-FE methods, but one can estimate this from the standard gas-phase QM calculations, as was done in ref 15. It is highly unlikely that these will be very different in enzyme than in solution or even in the gas phase. Although the charge distribution used for the free energy calculations can be considered "gas phase", in most reactions the polarization effects are modest. In fact, our use of 6-31G* RESP charges for the QM atoms in the free energy calculations makes them well balanced with the charges of the classical region; the use of this basis set model for the charges includes polarization effects implicitly since this basis set overestimates dipole moments by ∼10−20%.

A limitation of any methodology, which was particularly evident with the citrate synthase system, is the dependency of the results on the treatment of the long-range electrostatic interaction. Due to the charged nature of the active site, this enzyme is a particular challenge to model. We were able to address this challenge by using periodic boundary conditions and particle mesh Ewald (PME) treatment of long-range electrostatics in our free energy subroutine. The problems in the calculations of ref 11, where the enzyme environment net destabilizes the quantum mechanical atoms, have been noted above. Modifications of the standard AMBER5 release of *Gibbs* were required for this purpose (section 2.5.). These simulation conditions do increase the computational cost of a given free energy simulation, and we made them more efficient by using coarse-grained parallelization.[36]

Sham and Warshel[43] have shown that an alternative approach, using a spherical hydration model, calculates ion and ion-pair solvation independent of system size. Periodicity artifacts have been a concern for Ewald simulations, but it appears that these should be small if large enough boxes are used and the solvent has a high dielectric constant.[44,45]

Free energy simulations are usually carried out over relatively short simulation times, and this can give rise to sampling errors. The convergence of our results was checked through the use of different simulation conditions (i.e., monomer vs model dimer) and extension of the free energy simulations. In both cases, the numerical free energy results were *much* more stable than observed previously in the absence of PME and periodic boundary conditions. The instability that was observed in this system emphasizes the need to check for convergence with respect to simulation conditions in any free energy calculation.

4.4. Mechanistic Issues. The difficulty in abstracting C−H protons has been noted by mechanistic enzymologists for some time, and the fact that enzymes can do this with facility has spawned special concepts such as "low-barrier hydrogen bonds" and "pK_a matching". The interpretation of the mechanism of action of triose phosphate isomerase (TIM) by Komives et al.[46] as involving the protonation of an enolate by a nearby histidine was supported by calculations by Bash et al.,[47] but not supported by calculations of Alagona et al.,[5] who suggested that the enolate$^-$··His was more stable than the enol··His$^-$.

This is analogous to the situation in CS, where all previous studies[9−11] and this work support the greater stability of enolate$^-$··His. In CS, the enolate$^-$ has two strong hydrogen bonds to the anionic oxygen: one from His and the other from a water molecule (WAT471). Moving the proton from His to enolate leaves the enol hydrogen forming a H-bond to the His$^-$ and a water (470) forming a weaker H-bond to the N$_\delta$ of His$^-$.

More generally, this question is reminiscent of the argument in the serine proteases, where it was long debated whether His57, after receiving a proton from Ser195, transferred one to Asp102. Very strong evidence that His57 remains protonated during catalysis was provided by the neutron diffraction studies of

(39) Benzion, J.; Muller, R. P.; Florian, J.; Warshel, A. *J. Phys. Chem. B.* **1998**, *102*, 2293−2301.
(40) Alhambra, C.; Wu, L.; Zhang, Z. Y.; Gao, J. L. *J. Am. Chem. Soc.* **1998**, *120*, 3858−3566.
(41) Zhang, Y. K.; Liu, H. Y; Yang, W. T. *J. Chem. Phys.* **2000**, *112*, 3483−3492.
(42) Lui, H. Y., Zhang, Y. K., Yang, W. T. *J. Am. Chem. Soc.* **2000**, *122*, 6560−6570.
(43) Sham, Y. Y.; Warshel, A. *J. Chem. Phys.* **1998**, *109*, 7940−7944.
(44) Smith, P. E.; Pettitt, B. M. *J. Chem. Phys.* **1996**, *105*, 4289−4293.
(45) Weger, W.; Hunenberger, P. H.; McCammon, J. A. *J. Phys. Chem. B.* **2000**, *104*, 3668−3675.
(46) Komives, E. A.; Chang, L. C.; Lolis, E.; Lilton, R. F.; Petsko, G. A.; Knowles, J. R. *Biochemistry* **1991**, *30*, 3011−3019.
(47) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826−5832.

Kossiakoff et al.[48] Furthermore, there is no mechanistic advantage for $His57^+$ or the histidines stabilizing the enolates to give up their protons since, in the case of His57, it must deliver it to the amide N to promote peptide bond cleavage. There is also no catalytic advantage for CS or TIM to become an enol, since the enolates are better nucleophiles to accept a proton (from Ser195 in TIM) or attack oxaloacetate in CS, the latter point as previously noted by Mulholland and Richards.[49]

Structural Features. In the enzyme reaction, it is apparent that the water binding to the thioacetate (wat471) is quite tightly bound in the starting and enolate structures and more loosely bound in the enol. The binding of this water to the starting structure has also been found in previous studies.[10,11] A similar water is also apparent in the solvent reaction (Figure 4, WAT213, WAT379). Interestingly, the formation of the negatively charged thioacetate is not accompanied by many hydrogen bonds to the new negative carbon center in the enzyme, while in the solution reaction, two waters hydrogen bond solely with this negative center. This has implications for the enzyme mechanism since the next step of the reaction involves a condensation with this negative center and direct stabilization of this negative carbon center in the enzyme would inhibit the condensation.

Structurally, it appears that the starting structure asp375 is well solvated by the neighboring ser378, asn278, and wat474 (Figure 3) in the enzyme. Upon abstraction of the hydrogen, wat474 becomes more loosely bound, as do asn278 and ser378. The hydrogen-bonding network can either occur directly, or through wat474.

Energetic Features. Unfortunately, the use of PME precludes the use of free energy component analysis[50] to analyze the processes, reactant → enolate → enol. Still, we thought analysis of the interaction energy components was informative even though the subtraction of large numbers to yield small differences is always plagued by large standard deviations. We use this approach to gain some insight into the features of the enzyme that allow the reaction to be favorable in the enzyme compared to its energetics in solution. It is apparent from our estimates of the total interaction energy that there is a large entropic penalty in the solution reaction (39 kcal/mol, Table 4). The enzyme appears to stabilize both the reactant and the enolate intermediate with relatively little rearrangement (~7 kcal/mol penalty to form the product, Table 4, Figure 3a). This ties in nicely with Warshel's point[12] that enzymes can stabilize transition states better than solution because of pre-organized dipoles. Figure 4 illustrates the significant rearrangement that occurs going from reactant to enolate in solution.

A more detailed energetic analysis of the interactions with each residue of the active site indicate that the thioacetate part of the enolate is actually *less* stabilized in the enzyme than in solution ($\Delta E_{\text{thio, shell}}$ (enzyme) = −30.8 kcal/mol, $\Delta E_{\text{thio, shell}}$ (solution) = −51.2 kcal/mol, Table 4*)*. Since the overall free energy for the enzyme simulation is negative, this again suggests that the enzyme does not have to reorient its dipoles to achieve effective interaction with the thioacetate in contrast to what must happen in solution.[12]

A combination of structural and energetic analyses allows us to present an explanation for the success of the enzymatic reaction without involving "special concepts". In particular, the pre-organization of the enzyme to interact with both the reactant and enolate favors the enzyme reaction.[12] Interestingly, this pre-organization in the first solvation shell is not arranged to stabilize the enolate intermediate relatively more as compared to the solution reaction (Table 4, row 5). Longer range interactions do however, stabilize the enolate relative to the reactant (Table 4, row 6). This probably occurs because of the multistep nature of the reaction which requires the thioacetate to behave as a nucleophile in the proceeding condensation. Bernasconi[51] has noted that abstraction of a proton from carbon acids might be slower in solution because of asynchonization of the various elements, for example, solvation. This could also play a kinetic role; in the work presented here we have focused on how relatively unfavorable energetically C−H proton abstraction is intrinsically and how CS helps lower the energetic barrier.

4.5. Applications of QM-FE. This is the third application of QM-FE to enzyme systems. In the case of trypsin, we used previous studies that had established that the step from initial Michaelis complex to tetrahedral intermediate for acylation to be rate-limiting and studied just the relative free energies of two structures, with very good agreement with experimental free energies of activation for both enzyme and solution reactions. In the case of COMT,[16] we did a more complete study of methyl transfer and found an excellent reproduction of the $\Delta\Delta G^*$ (solution vs enzyme) as well as a quite good representation of ΔG^* (enzyme). CS provides another level of challenge, in that the large electrostatics in the active site make the reliable calculation of ΔG_{int} difficult. We have implemented PME to calculate ΔG_{int} and have shown the methodology to be robust and, together with ΔE_{QM}, to lead to a ΔG^* in encouragingly good agreement with experiment. Also, this is the first application of QM/FE to address a *mechanistic* question (the mechanisms of trypsin and COMT are well accepted) and our calculations allow us to rule out the enol as an intermediate in the CS pathway. Previous calculations[9−11] have also argued against the enol, but we feel our methodology allows the most definitive analysis to date and this is supported by the fact that our calculated ΔG^* for formation of the enolate is in surprisingly good quantitative agreement with experiment (calculated 15−16 kcal/mol, experimental 14.6−14.8 kcal/mol) and the most accurate value calculated to date. We expect the transition state for enolate formation is likely to be slightly higher in free energy than the enolate itself. On the other hand, higher levels of ab initio theory and consideration of proton tunneling would be expected to lower the calculated barrier; although, we do not know by how much.

The citrate synthase mechanism is actually a multistep one which involves hydrogen abstraction, condensation, and hydrolysis. A plethora of experimental information exists which documents the effects of mutations around the active site on the subsequent steps of the reaction (condensation and hydrolysis). The tightly coupled nature of the active site[52] suggests that the barriers to these reactions are comparable to the hydrogen abstraction step. For example, the initial enolate anion, once formed, might attack the OAA with little or no barrier, which could rationalize this tight coupling noted above. Interestingly, the reactant system has a thioacetate C−OAA carbonyl carbon distance of 3.6 Å, and a $C_{\text{thio}}-C_{\text{OAA}}-O_{\text{OAA}}$ angle of 66°, while the enolate structure after the free energy perturbation has a distance of 3.4 Å and an angle of 87°, favoring attack of the thioacetate at the carbonyl OAA center. Since we have identified

(48) Kossiakoff, A. A.; Spencer, S. A. *Biochemistry* **1981**, *20*, 6462−74.

(49) Mulholland, A. J.; Richards, W. G. *J. Mol. Struct.* (*THEOCHEM*) **1998**, *427*, 175−184.

(50) Perakyla, M.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 3436−3444.

(51) Bernasconi, C. F. *Acc. Chem. Res.* **1992**, *15*, 9−16.

(52) Schwarz, B.; Drueckhammer, D. G.; Usher, R. C.; Remington, S. J. *Biochemistry* **1995**, *34*, 15459−15466.

a stable methodology with which to assess the free energy changes in this system, we can proceed to quantify the actual barriers to these subsequent steps, as well as attempt to reproduce the experimental mutagenic information on this system.[24−26] By analyzing the resulting simulations on the mutated systems (i.e., by re-running the enzymatic free energy calculation in the field of the mutated residue), we should be able to dissect the particular interactions that are affected by a given mutation.

It is interesting that many of the enzymes which abstract C−H bonds (CS, triose phosphate isomerase, mandelate racemase, ketosteroid isomerase) have a significant number of charged groups nearby. Longer-range charge stabilization could be a strategy to stabilize both the catalytic base in its anionic form and to stabilize the nearby enolate by a comparable amount. In solution, water molecules much more strongly solvate the more charge localized CO_2^- than the delocalized enolate, leading to a much higher free energy barrier for C−H proton abstraction.

5.0. Conclusions

In conclusion, we have demonstrated the usefulness of the QM-FE (quantum mechanics-free energy) methodology in simulating the energetics of enzyme systems. For the first time, an enzyme involving a C−H bond abstraction has been simulated with this approach. Robust simulations required the use of PME in the free energy calculations, which we implemented in a coarse-grained parallel mode.

By calculating the chemically relevant free energy changes involved in the formation of both the enolate and the enol upon hydrogen abstraction in the citrate synthase mechanism, we have shown that it is the enolate and not the enol that is favored in the protein environment. The free energy barrier to the enolate formation is 15−16 kcal/mol, which is in good agreement with the observed free energy barrier of 14−15 kcal/mol. The calculated free energy barrier in solution was very high, further supporting the idea that enzymes are able to stabilize transition states with electrostatics (pre-aligned dipoles) and in orienting the reaction group in the active site in such a way that no "special concepts" need be invoked to explain enzyme catalysis.

Acknowledgment. O.D. gratefully acknowledges an NSERC PDF for salary support. P.A.K. acknowledges Grant GM-29072 from NIH. This research was partially supported by the National Computational Science Alliance utilizing the NCSA SGI/CRAY Origin2000 and the Maui High Performance Computing Center.

Supporting Information Available: A listing of geometric parameters used for the active-site residues (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA001043I