J|A|C|S ARTICLES

Published on Web 04/28/2010

Effects of Point Mutation on Enzymatic Activity: Correlation between Protein Electronic Structure and Motion in Chorismate Mutase Reaction

Toyokazu Ishida

Research Institute for Computational Sciences (RICS), National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 2, 1-1-1 Umezono, Tsukuba 305-8568, Japan

Received January 27, 2010; E-mail: toyokazu.ishida@aist.go.jp

Abstract: Assignment of particular roles to catalytic residues is an important requirement in clearly understanding enzyme functions. Therefore, predicting the catalytic activities of mutant variants is a fundamental challenge in computational biochemistry. Although site-directed mutagenesis is widely used for studying enzymatic activities and other important classes of protein function, interpreting mutation experiments is usually difficult mainly due to side effects induced by point mutations. Because steric and, in many cases, electrostatic effects may affect the local, fine geometries conserved in wild-type proteins that are usually believed to be thermodynamically stable, simply reducing a loss in catalytic activity into clear elements is difficult. To address these important but difficult issues, we performed a systematic *ab initio* QM/MM computational analysis combined with MD-FEP simulations and all-electron QM calculations for the entire protein matrix. We selected chorismate mutase, one of the simplest and well-known enzymes, to discuss the details of mutational effects on the enzymatic reaction process. On the basis of the reliable free energy profiles of the wild-type enzyme and several mutant variants, we analyzed the effects of point mutations relative to electronic structure and protein dynamics. In general, changes in geometrical parameters introduced by a mutation were usually limited to the local mutational site. However, this local structural modification could affect the global protein dynamics through correlated motions of particular amino acid residues even far from the mutation site. Even for mutant reactions with low catalytic activity, transition state stabilization was observed as a result of conformational modifications and reorganization around the active site. As for the electrostatic effect created by the polar protein environment, the wild-type enzyme was most effectively designed to stabilize the transition state of the reactive substrate, and the effect of global polarization in the electronic structure was found to be a small catalytic element during the process. As electrostatic media for optimum catalysis, both wild-type and mutant variant proteins were generally robust against external electrostatic perturbations. Protein structures have a certain flexibility, which allows them to slightly modulate their conformations to maximize the transition state stabilization in response to the steric perturbations induced by mutations.

1. Introduction

Enzymes potentially possess tremendous catalytic abilities such as drastic rate enhancements, specific substrate recognition, and fine stereochemical control and thus are crucial in most biological processes.[1−3] In the enzymatic reactions, it is usually said that *enzymes bind to the transition state of their substrates more tightly than their enzyme−substrate (ES) complex*.[4] However, despite this widely accepted popular proposition, the mechanistic principles of enzymatic reactions are still unclear to date because most modern techniques still encounter difficulties in directly determining correlations between protein 3D structures and functions at atomistic resolution.[5,6] Among many experimental protein engineering techniques, site-directed mutagenesis is a powerful technique that is widely used to identify the specific roles of target amino acid residues.[1,2] Because point mutations mainly predict the effect of each amino acid residue by replacing the original residues with other residues, we are able to estimate the expected role of each residue during the catalytic event through kinetic measurements of enzyme activity. On the other hand, mutations usually induce other side effects to the protein tertiary structures to a certain extent, making it difficult to clearly distinguish the role of original catalytic residues and unpredictable catalytic behavior introduced by point mutations.[7,8] For example, in many enzymatic reactions, muta-

(1) Frey, P. A.; Hegeman, A. D. *Enzymatic Reaction Mechanism*; Oxford University Press: New York, 2007.
(2) Fersht, A. *Structure and Mechanism in Protein Science. A Guide to Enzyme Catalysis and Protein Folding*, 2nd ed.; W. H. Freeman and Company: New York, 1999.
(3) Jencks, W. P. *Catalysis in Chemistry and Enzymology*; Dover Publications: New York, 1987.
(4) Pauling, L. C. *Chem. Eng. News.* **1946**, *24*, 1375–1377.
(5) Special Issue on Principles of Enzyme Catalysis. *Chem. Rev.* **2006**, *106*, 8.
(6) Focus on Biological Catalysis. *Nat. Chem. Biol.* **2009**, *5*, 8.
(7) Taylor, S. V.; Kast, P.; Hilvert, D. *Angew. Chem., Int. Ed.* **2001**, *40*, 3310–3335.
(8) Toscano, M. D.; Woycechowsky, K. J.; Hilvert, D. *Angew. Chem., Int. Ed.* **2007**, *46*, 3212–3236.

10.1021/ja100744h

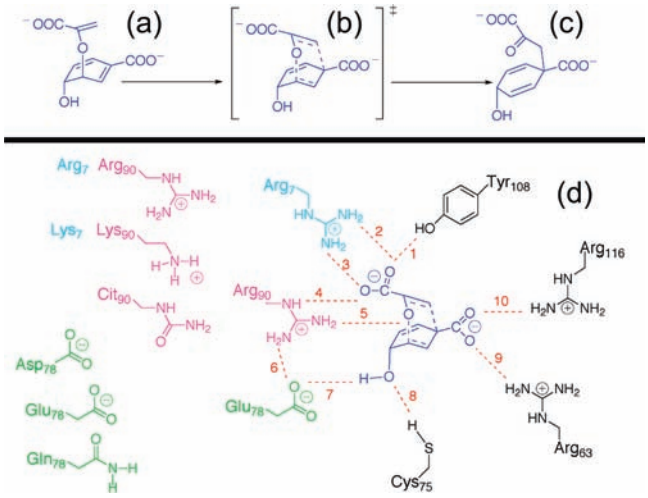

Figure 1. (Top) Reaction scheme of the Claisen rearrangement of chorismate, catalyzed by *Bacillus subtilis* chorismate mutase (BsCM). Through the putative transition state (TS) structure (b), the reactant chorismate (a) is converted into the product prephenate (c). (Bottom) Schematic illustration of the major hydrogen bonds at the wild-type BsCM active site (d). The Arabic numerals correspond to the 10 major hydrogen-bond pairs, and bond distances sampled from the QM/MM MD-FEP runs are summarized in Table 2. The three target amino acid residues in the computed mutations are Arg90 (purple), Glu78 (green), and Lys7 (light blue). The resulting mutant side chain structures appear on the left-hand side of the figure.

tions of catalytically crucial residues usually lower the rate enhancement of chemical reactions, although the affinities of target ligands increase. In addition, mutations that replace bulky, highly steric side chains by simpler ones and those that replace polar residues with nonpolar ones may alter the optimum alignment of the catalytic residues at the active site. The resulting catalytic efficiencies are largely affected by the perturbation of steric or electrostatic effects. In such cases, care must be taken when interpreting the observed rate constants during the kinetic measurements.

With the rapid progress of computational bioscience, computer-aided protein modeling and *design* has emerged as an important research area.[9−11] Designing effective enzymes and predicting their activity are two of main challenges in this research field. In these computational challenges, simulating and predicting the effects of mutations are important but minimum essential prerequisites. Chorismate mutase (CM) is now a particular class of target enzymes that has been practically tested through comprehensive mutational studies. CM enzymes catalyze the internal conversion of a chorismate into a prephenate through a Claisen rearrangement (Figure 1).[12] So far, several types of CM structures have already been determined through well-refined crystal structures,[13−15] and as a result of many additional kinetic measurements,[16−18] several catalytic residues have been suggested, along with their predicted roles.[19−22] Hilvert and co-workers have long been examining the nature of the CM catalytic mechanism as a template for general enzyme principles.[23−32] For example, in a recent paper investigating the relative tolerance against point mutation effects, they reported that activity differences between mutant variants were very much context-dependent and likely to stem from subtle changes in the fine structure of the active site.[32] Mayo and co-workers have investigated the computational enzyme design, which aims at mimicking CM reactivity inside a completely novel catalyst.[33,34] To study the effect of point mutations at active site residues, which are not directly hydrogen-bonded to the reactive substrate, they performed exhaustive mutagenesis studies for all possible substitutions and reported the importance of secondary amino acid residues in determining enzymatic activity.[34]

Simulations of enzymatic reactions are presently highly debated in computational and theoretical chemistry,[35−38] and CMs have been widely used as target molecules because of their characteristic features in the reactions. For example, there are no chemical bond formations between protein and substrate. Moreover, the catalytic reactions inside enzymes and the solution-phase counterpart have already been determined by several kinetic measurements.[12] Many computational studies have been reported so far;[39−48] however, most of them focus only on the wild-type enzyme reaction by introducing new

(9) Vizcarra, C. L.; Mayo, S. L. *Curr. Opin. Chem. Biol.* **2005**, *9*, 622–626.
(10) Lippow, S. M.; Tidor, B. *Curr. Opin. Biotechnol.* **2007**, *18*, 305–311.
(11) Roca, M.; Vardi-Kilshtain, A.; Warshel, A. *Biochemistry* **2009**, *48*, 3046–3056.
(12) Ganem, B. *Angew. Chem., Int. Ed.* **1996**, *35*, 936–945.
(13) Chook, Y. M.; Ke, H.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8600–8603.
(14) Xue, Y.; Lipscomb, W. N.; Graf, R.; Schnappauf, G.; Braus, G. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10814–10818.
(15) Lee, A. Y.; Karplus, P. A.; Ganem, B.; Clardy, J. *J. Am. Chem. Soc.* **1995**, *117*, 3627–3628.
(16) Andrews, P. R.; Smith, G. D.; Young, I. G. *Biochemistry* **1973**, *12*, 3492–3498.
(17) (a) Copley, S. D.; Knowles, J. R. *J. Am. Chem. Soc.* **1987**, *109*, 5008–5013. (b) Guilford, W. J.; Copley, S. D.; Knowles, J. R. *J. Am. Chem. Soc.* **1987**, *109*, 5013–5019.
(18) Kast, P.; Asif-Ullah, M.; Hilvert, D. *Tetrahedron Lett.* **1996**, *37*, 2691–2694.
(19) Chook, Y. M.; Gray, J. V.; Ke, H.; Lipscomb, W. N. *J. Mol. Biol.* **1994**, *240*, 476–500.
(20) Haynes, M. R.; Stura, E. A.; Hilvert, D.; Wilson, I. A. *Science* **1994**, *263*, 646–652.
(21) Lee, A. Y.; Stewart, J. D.; Clardy, J.; Ganem, B. *Chem. Biol.* **1995**, *2*, 195–203.
(22) Cload, S. T.; Liu, D. R.; Pastor, R. M.; Schults, P. G. *J. Am. Chem. Soc.* **1996**, *118*, 1787–1788.
(23) Kast, P.; Asif-Ullah, M.; Jiang, N.; Hilvert, D. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5043–5048.
(24) Kast, P.; Hartgerink, J. D.; Asif-Ullah, M.; Hilvert, D. *J. Am. Chem. Soc.* **1996**, *118*, 3069–3070.
(25) Gustin, D. J.; Mattei, P.; Kast, P.; Wiest, O.; Lee, L.; Cleland, W. W.; Hilvert, D. *J. Am. Chem. Soc.* **1999**, *121*, 1756–1757.
(26) Kast, P.; Grisostomi, C.; Chen, I. A.; Li, S.; Krengel, U.; Xue, Y.; Hilvert, D. *J. Biol. Chem.* **2000**, *275*, 36832–36838.
(27) Gamper, M.; Hilvert, D.; Kast, P. *Biochemistry* **2000**, *39*, 14087–14094.
(28) Kienhöfer, A.; Kast, P.; Hilvert, D. *J. Am. Chem. Soc.* **2003**, *125*, 3206–3207.
(29) Vamvaca, K.; Vögeli, B.; Kast, P.; Pervushin, K.; Hilvert, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12860–12864.
(30) Eletsky, A.; Kienhöfer, A.; Hilvert, D.; Pervushin, K. *Biochemistry* **2005**, *44*, 6788–6799.
(31) Besenmatter, W.; Kast, P.; Hilvert, D. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 500–506.
(32) Woycchowsky, K. J.; Choutko, A.; Vamvaca, K.; Hilvert, D. *Biochemistry* **2008**, *47*, 13489–13496.
(33) Lassila, J. K.; Privett, H. K.; Allen, B. D.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16710–16715.
(34) Lassila, J. K.; Jennifer, R. K.; Kast, P.; Mayo, S. L. *Biochemistry* **2007**, *46*, 6883–6891.
(35) Special Issue on Molecular Dynamics Simulations of Biomolecules. *Acc. Chem. Res.* **2002**, *35*.
(36) Special Issue on Protein Simulations. *Adv. Prot. Chem.* **2003**, *66*.
(37) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 3188–3209.
(38) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, H. M. *Chem. Rev.* **2006**, *106*, 3210–3235.
(39) Lyne, P. D.; Mulholland, A. J.; Richards, W. G. *J. Am. Chem. Soc.* **1995**, *117*, 11345–11350.

computational methods.[49−51] A brief history of these computational CM studies has recently been summarized by Senn and Thiel.[52] Only a few studies have actually examined the reaction profiles of mutant enzymes.[53,54] The main focus of this article is to analyze and clarify the effect of point mutations on enzymatic activity through a detailed comparison between the wild-type enzyme and several active-site mutant variants. The reaction profile of the wild-type enzyme has previously been extensively investigated through systematic hybrid computational modeling procedure, ranging from all-atom molecular mechanics/molecular dynamics (MM-MD) simulations to *ab initio* quantum mechanics (QM)/MM computations, and all-electron QM analysis, to probe the polar protein environment.[55] This previous paper clearly demonstrated that (1) the major catalytic role of the protein environment is decomposed into a few critical residues around the binding pocket including Arg7, Glu78, and Arg90, and (2) *ab initio* QM/MM treatments reasonably predict the essential role of these residues because the polarization effect of protein is minor. Based on these previous reports, this article further discusses the role of the polar protein environment as an optimum electrostatic medium for catalysis by comparing the catalytic activities of mutant variants, in particular, the above-mentioned three residues. In the present study, we analyze the effect of mutations from two different viewpoints. The first is the dynamic factor that affects the (collective) enzyme motions by examining the correlation movement of amino acid residues inside the protein along the reaction coordinate. The other is the response of the protein electronic structure upon introduction of point mutations into the electrostatic medium by evaluating the strength of electrostatic interaction energies between all of the amino acid residue pairs, which are determined using all-electron QM calculations for geometries selected along the reaction coordinate. Through these detailed investigations, we discuss the effect of point mutations on the catalytic activity of *Bacillus subtilis* chorismate mutase (BsCM).

2. Methods and Computational Details

All calculations used *ab initio* QM/MM modeling techniques reported earlier.[55−58] Free energy profiles along selected reaction coordinates were evaluated by combining QM/MM with MD-free energy perturbation (FEP) simulations, and protein electronic structures were calculated by combining of QM/MM with all-electron QM computations. We followed the systematic modeling strategy proposed in a previous publication.[55] To estimate the reaction free energy profile, we first determined the reaction path using QM/MM structural optimizations along the selected reaction coordinate. In this QM/MM modeling process, we prepared and checked two types of QM models. In the small-QM-region model, the QM region was limited to the reactive substrate. On the other hand, in the large-QM-region model, several key surrounding residues were added to this small-QM-region model. The small-QM-region model was mainly used to estimate the reaction free energies and account for fluctuations in the surrounding protein environment, while the large-QM-region model was used in all-electron QM analyses to evaluate protein electronic structures. All QM/MM and MM modeling and simulation calculations used an original program developed in-house. The *ab initio* QM/MM program was developed from the HONDO package,[59] and MM modeling and simulation routines were added to the *ab initio* molecular orbital (MO) calculation component. In all QM/MM calculations, we used the so-called electrostatic embedding QM/MM partition scheme.[52,60−67] The interactions between QM and MM regions were defined as

$$E_{\mathrm{QM/MM}} = E^{\mathrm{elec}}_{\mathrm{QM/MM}} + E^{\mathrm{vdw}}_{\mathrm{QM/MM}} + E^{\mathrm{strain}}_{\mathrm{QM/MM}} \qquad (1)$$

where the first term was evaluated using electronic structure calculations, and the other two terms were approximated using a classical force field. The AMBER parameter set (parm.96) was used for the force field calculations.[68,69]

2.1. *Ab Initio* QM/MM Structural Modeling of Mutant Enzymes. The initial model structure to evaluate the wild-type reaction path was chosen from previous simulation results.[55] Enzyme coordinates were based on the X-ray crystal structure of BsCM bound to the endo-oxabicyclic transition state analogue (TSA) determined at a 2.2 Å resolution (PDB code 2CHT).[13] Hydrogen atoms were added to the ES complex in the standard method by assuming a standard protonation state for all of the polar residues under physiological pH condition. After selecting the reaction coordinate, we followed the reaction path by constrained QM/MM optimizations. Here, the reaction coordinate was defined as a linear combination of two bond lengths, which correspond to

(40) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *J. Am. Chem. Soc.* **2004**, *126*, 311–319.
(41) Lee, Y. S.; Worthington, S. E.; Krauss, M.; Brooks, B. R. *J. Phys. Chem. B* **2002**, *106*, 12059–12065.
(42) Szefczyk, B.; Mulholland, A. J.; Ranaghan, K. E.; Sokalski, W. A. *J. Am. Chem. Soc.* **2004**, *126*, 16148–16159.
(43) Crespo, A.; Martí, M. A.; Estrin, D. A.; Roitberg, A. E. *J. Am. Chem. Soc.* **2005**, *127*, 6940–6941.
(44) Strajbl, M.; Shurki, A.; Kato, M.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 10228–10237.
(45) (a) Hur, S.; Bruice, T. C. *J. Am. Chem. Soc.* **2003**, *125*, 5964–5972. (b) Zhang, X.; Zhang, X.; Bruice, T. C. *Biochemistry* **2005**, *44*, 10443–10448.
(46) Guimarães, C. R. W.; Repasky, M. P.; Chandrasekhar, J.; Tirado-Rieves, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2003**, *125*, 6892–6899.
(47) Guo, H.; Cui, Q.; Lipscomb, W. N.; Karplus, M. *Angew. Chem., Int. Ed.* **2003**, *42*, 1508–1511.
(48) Ranaghan, K. E.; Mulholland, A. J. *Chem. Commun.* **2004**, 1238–1239.
(49) Woodcock, H. L.; Hodoscek, M.; Sherwood, P.; Lee, Y. S.; Schaefer, H. F., III.; Brooks, B. R. *Theor. Chem. Acc.* **2003**, *109*, 140–148.
(50) Crehuet, R.; Field, M. J. *J. Phys. Chem. B* **2007**, *111*, 5708–5718.
(51) Claeyssens, J.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schüetz, M.; Thiel, S.; Thiel, W.; Werner, H. J. *Angew. Chem., Int. Ed.* **2006**, *45*, 6856–6859.
(52) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.
(53) Ishida, T.; Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. B* **2006**, *110*, 1457–1463.
(54) Guimarães, C. R. W.; Udier-Blagović, M.; Tubert-Brohman, I.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2005**, *1*, 617–625.
(55) Ishida, T. *J. Chem. Phys.* **2008**, *129*, 125105.
(56) (a) Ishida, T.; Kato, S. *J. Am. Chem. Soc.* **2003**, *125*, 12035–12048. (b) Ishida, T.; Kato, S. *J. Am. Chem. Soc.* **2004**, *126*, 7111–7118.
(57) Ishida, T. *Biochemistry* **2006**, *45*, 5413–5420.
(58) Ishida, T. *J. Phys. Chem. B* **2010**, *114*, 3950–3964.
(59) Dupuis, M.; Marquez, A.; Davidson, E. R. *HONDO*; Quantum Chemistry Program Exchange (QCPE), Indiana University: Bloomington, IN 47405.
(60) Monard, G.; Merz, K. M., Jr. *Acc. Chem. Res.* **1999**, *32*, 904–911.
(61) Field, M. J. *J. Comput. Chem.* **2002**, *23*, 48–58.
(62) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467–505.
(63) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425–443.
(64) Ryde, U. *Curr. Opin. Chem. Biol.* **2003**, *7*, 136–142.
(65) Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389–427.
(66) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; Konig, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458–6469.
(67) Hu, H.; Yang, W. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573–601.
(68) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(69) Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. In *Computer Simulation of Biomolecular Systems*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; Kluwer Academic Publishers: Leiden, 1997; Vol. 3.

C−O bond breaking and C−C bond formation. To optimize the QM/MM reaction path, we used the second-order Møller−Plesset perturbation method (MP2) with the 6-31(+)G** basis set in which diffuse functions were added to the two carboxylic groups only. As clearly demonstrated in previous studies, the electron correlation is a crucial factor in this type of chemical reaction, and MP2 usually overestimates the correlation energy.[51,55] To correct for this overestimation, we performed MP4 calculations using the MP2 optimized geometries along the reaction coordinate. The QM/MM computations were thus conducted at the MP4/6-31(+)G**//MP2/6-31(+)G** level using the small-QM-region model. As noted previously, more refined geometries are usually required to analyze protein electronic structures by all-electron QM calculations.[55] Therefore, we performed additional QM/MM optimizations for geometries selected along the reaction coordinate using the large-QM-region models. The geometry parameters were fully optimized for the entire protein complex, except for the reaction coordinate. On the basis of our previous calculation results, the side chain atoms of Arg7, Glu78, and Arg90 were added to the QM region in the large-QM-region models. Boundaries between QM/MM regions (C_β-C_γ) were saturated using dummy hydrogen atoms that were allowed to move freely during the QM/MM optimizations.

Next, we prepared the reaction path models of five mutant systems from the wild-type reaction path model. In this modeling process, we assumed that each mutant reaction followed the same reaction profile where the intramolecular isomerization proceeds in a concerted manner through two scissile bonds, without covalent bond formation with the amino acid residues inside the catalytic pocket. Transition state (TS) structures were determined for each mutant reaction for which pathways were then evaluated in the same manner as for wild-type reaction. The original side chain positions were replaced by the variant side chains with maximum overlap between heavy atom coordinates in the mutation residues. Similar to wild-type modeling, all mutant structures were optimized using *ab initio* QM/MM calculations. Note that all stationary structures were validated by vibrational frequency calculations.

2.2. Free Energy Simulations Using *ab Initio* QM/MM Calculations Combined with MD-FEP. Free energy calculations were conducted following the same strategy developed in the previous papers.[55–58] The initial solvation structure used for the wild-type simulations was modified by point mutations for the target amino acid residues to prepare the initial solvated mutant structures. We obtained reasonable solvation structures for each reaction system after several hundred picoseconds of MD runs. Each ES complex was solvated in a sphere of TIP3P water molecules with a 40 Å radius centered on the center of mass of the complex.[70] A spherical boundary condition was used to maintain the solvation structure by adding a weak harmonic constrain potential to the surface boundary of the solvation sphere. Each system contained about ~9000 water molecules and the ES complex. The Nosé-Hoover-chain (NHC) method was used in all MD simulations to generate the NVT ensemble, and the system temperature was maintained at 303 K by attaching the five chains of thermostat with the thermostat mass corresponding to $\tau = 0.5$ ps.[71,72] The reversible reference system propagation (rRESPA) algorithm extended to the non-Hamiltonian NHC system was also used.[73,74] Long-range nonbonded forces, mainly electrostatic and van der Waals interactions, were integrated in a long time scale (2.0 fs), while short-range bonding forces, including bond, bend, torsion, and improper terms, were integrated in a short time scale (0.25 fs). No cutoff was introduced for nonbonding interactions in all MD simulations.

Total free energy profiles for each mutant reaction were evaluated along the predefined reaction coordinate using *ab initio* QM/MM methods combined with constrained MD-FEP computations as[55–58]

$$F \simeq G_{QM}(\boldsymbol{R}_{QM}^{min}) - k_B T \ln \int e^{-\beta[E_{QM/MM}(\boldsymbol{R}_{QM}^{min},\boldsymbol{R}_{MM})+E_{MM}(\boldsymbol{R}_{MM})]}\, d\boldsymbol{R}_{MM} \qquad (2)$$

where the first term represents the vibrational free energy in the QM region and the second term is the free energy contribution from the MM region including surrounding protein and solvent. The constrained MD technique was applied to estimate the second term of eq 2. The internal geometries and the electron density of QM region, which were determined through QM/MM calculations for the wild-type simulations, were employed for mutant variants. By gradually changing these constrained parameters, the reactant geometry was transformed into the product geometry over the free energy barrier. During this adiabatic transformation process, free energy differences were calculated using the same concept of FEP[75] as follows:

$$\begin{aligned}\Delta F_i &= -k_B T \ln\langle e^{-\beta[E_{QM/MM}(\boldsymbol{q}_{QM}^{min,i+1},\boldsymbol{R}_{QM}^{min,i+1})-E_{QM/MM}(\boldsymbol{q}_{QM}^{min,i},\boldsymbol{R}_{QM}^{min,i})]}\rangle_i \\ &= k_B T \ln\langle e^{\beta[E_{QM/MM}(\boldsymbol{q}_{QM}^{min,i+1},\boldsymbol{R}_{QM}^{min,i+1})-E_{QM/MM}(\boldsymbol{q}_{QM}^{min,i},\boldsymbol{R}_{QM}^{min,i})]}\rangle_{i+1}\end{aligned} \qquad (3)$$

where $\mathbf{R}_{QM}^{min,i}$ and $\mathbf{q}_{QM}^{min,i}$ represent the internal coordinates of QM region and the partial atomic charges of QM atoms at the *i*th optimized point, respectively.

We applied the small-QM-region model to all MD-FEP calculations to reasonably sample the fluctuations between substrate and surrounding amino acid residues. The internal degrees of freedom in the QM region were fixed exactly in the QM/MM optimized geometries using the SHAKE/RATTLE constrained MD techniques.[76,77] MD simulations were performed in classical MM level to sufficiently sample conformational space, and electrostatic interactions between QM and MM regions were approximated with the ESP charge model. The ESP charges of QM atoms were optimized to reproduce the optimized QM/MM reaction path.[78,79] We calculated the free energy differences (eq 3) using the double-wide sampling technique.[80] The simulations consisted of a 20−40 ps equilibration run and a 50−200 ps averaging run at each point along the path. Longer MD-FEP runs were performed in regions showing large free energy changes and neighboring the TS region rather than in nonreactive regions. A total of 25 windows, which correspond to intervals between calculated points along the reaction path, were used in the simulations. The relative free energies were calculated as an average of the *forward* and the *backward* simulation results.

For the selected points corresponding to ES, TS, and product along the reaction coordinate, we performed additional 1000 ps MD simulations to evaluate the cross-correlation movements and to directly compute the QM/MM interaction energies. The cross-correlation coefficient for the displacement of any two atoms *i* and *j* is given by[81]

(70) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(71) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643.
(72) Cheng, A.; Merz, K. M., Jr. *J. Phys. Chem.* **1996**, *100*, 1927–1937.
(73) Tuckerman, M.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990–2001.
(74) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. *Mol. Phys.* **1996**, *87*, 1117–1157.
(75) (a) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395–2417. (b) Reynolds, C. A.; King, P. M.; Richards, W. G. *Mol. Phys.* **1992**, *76*, 251–275.
(76) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
(77) Anderson, H. C. *J. Comput. Phys.* **1983**, *52*, 24–34.
(78) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361–373.
(79) (a) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280. (b) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357–1377.
(80) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050–3054.
(81) Swaminathan, S. L.; Harte, W. E., Jr.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 2717–2721.

$$C_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\sqrt{\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle}} \quad (4)$$

where Δr_i is the displacement from the mean position of the *i*th atom. The elements C_{ij}, which were collected in a symmetric matrix form, were computed as an average over successive backbone (N−C_α−C═O) atoms to give one entry per pair of amino acid residues. These matrix elements were calculated from a stable 1 ns MD trajectory, and the MD trajectory was divided into four 250 ps blocks. A covariance matrix comprised of C_{ij} elements was derived for each block. The four covariance matrixes were averaged to generate a mean covariance matrix, which was plotted as a two-dimensional (2D) cross-correlation diagram for each mutant reaction. At the same time, the resulting MD trajectories were used to directly perform QM/MM sampling calculations for each reaction system. After obtaining nanosecond order MD trajectories, we performed *ab initio* QM/MM computations for the entire solvated enzyme complexes and evaluated QM and QM/MM interaction energies along the MD trajectories. We considered geometries that were sampled more than 5000 times at each point along the reaction coordinate.

2.3. All-Electron QM Calculations by the Fragment Molecular Orbital (FMO) Method. All-electron QM calculations were performed for the entire protein complex to estimate the strength of electrostatic interaction using the fragment molecular orbital (FMO) framework.[82,83] For this purpose, we used the large-QM-region models refined along the selected reaction coordinate in each mutant variants. In all FMO calculations, we employed the two-body expansion (FMO2) due to the complex enzyme system size, and the total energy of the system was approximated by

$$E_{\text{total}} = \sum_I E'_I + \sum_{I>J} \Delta E_{IJ} \quad (5)$$

where ΔE_{IJ} is the *pair interaction energy* that includes higher-body effects through the external field of the system. As demonstrated in a previous paper,[55] the electrostatic interaction is the leading energy term in the pair interaction energy in the wild-type BsCM reaction. To clarify the difference in catalytic activity between the wild-type reaction and the mutant variants, we focused on the electrostatic factor during the catalytic process. As noted above, the electrostatic interaction component was evaluated in the following manner using the *ab initio* MO formulation:

$$\Delta E_{IJ}^{(A)} = \sum_{\mu\nu\in I} D_{\mu\nu}^{I,(A)} \sum_{\alpha\in J} \left\langle \mu \left| \frac{-Z_\alpha}{|\mathbf{r} - \mathbf{R}_\alpha|} \right| \nu \right\rangle + \sum_{\rho\sigma\in J} D_{\rho\sigma}^{J,(A)} \sum_{\beta\in I} \left\langle \rho \left| \frac{-Z_\beta}{|\mathbf{r} - \mathbf{R}_\beta|} \right| \sigma \right\rangle + \sum_{\mu\nu\in I} \sum_{\rho\sigma\in J} D_{\mu\nu}^{I,(A)} D_{\rho\sigma}^{J,(A)} (\mu\nu|\rho\sigma) + \Delta E_{IJ}^{nuc} \quad (6)$$

where (*A*) designates the *isolated* and *complex* states of the system. The former corresponds to the electrostatic interaction of the unperturbed or *isolated* density distribution, whereas the latter corresponds to the density distribution mutually polarized inside the protein complex. *I* and *J* describe the substrate and the amino acid residues, respectively.

All FMO calculations were performed using the GAMESS implemented version[84,85] with local modifications by the author. The technical details of FMO2 computations are as follows. Both the atomic and molecular orbital accuracies were increased to 10^{-12} using ICUT═12, ITOL═24, and CUTOFF═10^{-12}, and the SCF convergence was tightened to 10^{-7}. The same values were used during the monomer SCF cycle where monomer densities converge. Since using diffuse functions in the fragment-based methods often leads to problems, we used the 6-31G* basis set in all FMO2 calculations. The option to remove *s* contaminants from *d* functions was used. We used the following approximations to treat the surrounding electrostatic potential (ESP). Atomic point charges determined through Mulliken population analysis were used to compute ESP when the interfragment distances were larger than the RESPPC threshold (RESPPC═2.0). In addition, the separated dimer energies were calculated by the quantum mechanical electrostatic interaction using eq 6 when the interfragment distances were larger than the RESDIM criteria (RESDIM═2.5). We did not use the atomic population approximation for the ESP in this article (RESPAP═0). All approximations were applied when the distance between the two fragments was larger than a threshold. We used the unitless distance between two fragments that corresponds to the distance measured relative to the sum of the van der Waals radii of the closest contact atoms. For fragmentation, we followed a one amino residue per fragment partition scheme due to the direct application of the interaction energy decomposition analysis based on eq 6. The protein backbone was divided into fragments at the C_α position, keeping peptide bonds intact. The hybrid sp^3 orbitals of the carbon atom were used to appropriately divide the molecular orbital space at bond fraction points. The substrate was treated as a single fragment in all FMO2 calculations. General input data for FMO calculations were prepared using the FMOutil modeling utility.[86]

3. Results

3.1. Free Energy Profile of Mutant Variants Obtained by QM/MM Calculations Combined with MD-FEP. Reaction free energy changes along the QM/MM optimized reaction coordinate are shown in Figure 2. The reaction coordinate in this case is defined as a linear combination of two scissile bond lengths. As clearly shown in recent large scale and high level computations,[51,55] the electron correlation effect is an important factor to quantitatively predict the activation free energies. MP2 and DFT/B3LYP calculations using basis sets larger than 6-31G* essentially reproduce a similar energy profile, although both methods underestimate the activation energy. Because the reproducibility of the relative free energy differences is an essential prerequisite for the main purpose of this computational study, we performed the geometry optimizations at the MP2/6-31(+)G** level followed by single point MP4/6-31(+)G** energy corrections. Figure 2a compares the free energy profile for the enzymatic catalytic process with the reference aqueous phase reaction. The computational activation free energy for the enzyme reaction and its difference with respect to the reference reaction were ∼16.4 and ∼5 kcal/mol, respectively. Compared to the experimental activation free energy ($\Delta G^{\ddagger}_{\text{act}}$, ∼15 kcal/mol) and its difference with respect to the reference case ($\Delta\Delta G^{\ddagger}_{\text{act}}$, ∼10 kcal/mol),[18] the present computation gives

(82) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, U. *Chem. Phys. Lett.* **1999**, *313*, 701–706.
(83) *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; Fedorov, D. G., Kitaura, K., Eds. CRC Press: Boca Raton, 2009.
(84) (a) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832–6840. (b) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *121*, 2483–2490.
(85) Schmidt, M. W.; Baldridge, K. K.; Boats, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
(86) FMOutil, available from http://staff.aist.go.jp/d.g.fedorov/

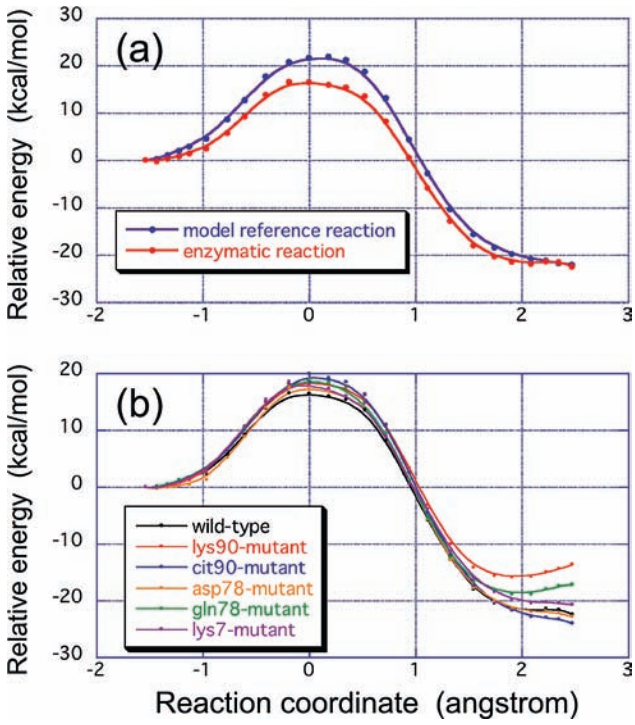

Figure 2. (a) Free energy profiles along the reaction coordinate evaluated by *ab initio* QM/MM calculations combined with MD-FEP simulations at 303 K (QM/MM MP4/6-31(+)G**//MP2/6-31(+)G**/AMBER (parm.96) level). The overall free energy profile of the enzymatic reaction is shown in red, and the corresponding reference solution-phase reaction is shown in blue. The *x*-axis indicates the reaction coordinate defined by the linear combination of two scissile bond (C−O/C−C) distances (in Å), and the *y*-axis represents the relative energy of the reaction (kcal/mol). (b) Comparison of the free energy profiles between the wild-type enzyme reaction (black) and the remaining five mutant variants: Lys90 (red), Cit90 (blue), Asp78 (yellow), Gln78 (green), and Lys7 mutants (purple). In all *ab initio* QM/MM MD-FEP calculations at 303 K, we define the same type of reaction coordinate and the same calculation levels (QM/MM MP4/6-31(+)G**//MP2/6-31(+)G**/AMBER (parm.96) level).

both qualitative and quantitative estimations. Note that the solution-phase reaction was simply modeled by replacing the protein environment with an aqueous solution. This means that in the current model, we do not evaluate the effect of substrate preorganization through the so-called near-attack conformation energy,[45] which have been estimated in the order of 4−5 kcal/mol in earlier computations.[44,48] With this additional energy contribution, the estimated $\Delta\Delta G^{\ddagger}_{act}$ becomes 9−10 kcal/mol, which agrees well with experimental results. Based on these wild-type modeling results, we subsequently performed free energy simulations for five mutant reactions. As noted in the methodological section, each mutant system was assumed to catalyze the same chemical reaction as observed in the wild-type enzyme.

Figure 2b shows free energy profiles for the wild-type and the five mutant reactions. All mutant reactions display activation free energies larger than that of the wild-type reaction. Several publications have reported that Arg90 in the domain 2 is the most influential residue for catalysis.[41,42,53,55] This point was confirmed by the present simulations. The largest free energy barrier was observed for the Cit90 mutant, which neutralizes the positive charge of the guanidinium side chain among the six reaction systems. As a reference for experimental findings, we summarized the known experimental kinetic measurements for these reaction systems in Table 1. Both the first-order (k_{cat}) and the second-order (k_{cat}/K_M) rate constants have already been determined for the Cit90 mutant reaction.[28] Although the catalytic rate of an enzymatic reaction is well characterized by both kinetic constants, a simple decomposition of complex catalytic elements inside both rate constants is not always clear. However, in the following discussion, we consider the enzymatic catalysis as described on the basis of the k_{cat} values and we compare them with computational data. Comparison between the wild-type and the Cit90 mutant reactivities showed that the reaction rate drastically decreased by $\sim 10^5$ for Cit90, which corresponds to a ~6 kcal/mol increase in the activation energy barrier.[28] The present simulation properly reproduces this experimental observation. Another mutant system for which kinetic parameters have been measured is the Asp78 mutant enzyme.[22] The Asp78 mutant enzyme exhibited a catalytic activity similar to the wild-type enzyme (Table 1), but its catalytic rate decreased by approximately $10-10^2$. Our simulations showed that Asp78 displayed the second most efficient reaction system with respect to the activation free energy, which was only ~1 kcal/mol higher than that of the wild-type reaction, consistent with the available experimental data (Table 1).

Unfortunately, there are presently no experimental kinetic parameters for k_{cat} for the three remaining mutant systems. However, estimates from another important kinetic parameter (k_{cat}/K_M) suggested that the activation free energies potentially increased by a few kcal/mol for all mutant enzymes.[22] The present QM/MM calculations combined with MD-FEP simulations reasonably reproduce all of these energy relationships. Mutation effects were analyzed using these computed structures from both the geometrical and the electronic structure viewpoints in the following sections.

3.2. Geometry Changes at Local Active Sites. The local structural parameters that characterize substrate binding and recognition were examined. The major hydrogen-bond distances between the substrate and the surrounding active site residues were selected for this purpose, as summarized in Figure 1d and Table 2. As demonstrated previously,[55] there are 10 characteristic hydrogen-bonded pairs at the center of the enzyme active site. Figure 3 shows the three-dimensional active site structures in each reaction system obtained from the QM/MM MD-FEP simulations.

Because the wild-type enzyme reaction is used as a reference for the following discussion, first we briefly summarize structural changes of hydrogen-bonded network in the wild-type reaction. As described in the computational procedure, the QM substrate geometries were fixed in the QM/MM optimized ones and forced to transform adiabatically from reactant geometry to product one through the constrained MD technique. Therefore, the internal coordinates of the substrate were the same as the QM/MM optimized coordinates. Along the reaction coordinate, two carboxylate groups were tightly stabilized by strong hydrogen bonds with Arg7, Tyr108, and Arg116, which are located in domain 2 of the enzyme, and Arg63 in domain 1. Although Cys75 was directed toward the substrate hydroxyl group, only a weak interaction was achieved during the reaction. A small reorganization of the hydrogen bond network was observed in the Glu78-Arg90-substrate alignment. Mainly because of attractive interaction between Arg90 and Glu78 side chains, Glu78 could not strongly interact with the reactant hydroxyl group. As a result, a strong hydrogen bond was formed between Glu78 and the hydroxyl group after product formation only.

Next, the effect of point mutations at residue 90 was examined. Several experimental observations have reported that a point mutation of the original Arg90 drastically decreased the

Table 1. Kinetic Parameters from Available Experiments and Activation Free Energies Using *ab Initio* QM/MM Calculations Combined with Constrained-MD-FEP Simulations (in kcal/mol)

reaction	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} s^{-1})	$\Delta G^{\ddagger}_{exp}$ (kcal/mol)[d]	$\Delta G^{\ddagger}_{calc}$ (kcal/mol)[e]
wild-type[a]	46 ± 3	67 ± 5	$(6.9 \pm 0.7) \times 10^5$	15.4	16.5 ± 0.14
Lys90 mutant[b]			31 ± 1		18.8 ± 0.84
Cit90 mutant[c]	0.0026	270	9.6	∼21	19.9 ± 0.34
Asp78 mutant[b]	35.7 ± 2.4	1297 ± 139	$2.75 \times 10^4 \pm 3.5 \times 10^3$	∼16	17.5 ± 0.63
Gln78 mutant[b]			75 ± 2		18.9 ± 0.44
Lys7 mutant[b]			717 ± 11		18.5 ± 0.72

[a] Wild-type data are taken from refs 18 and 26. [b] These mutant data are taken from ref 22. [c] Cit90 mutant data are taken from ref 28. [d] Experimental activation free energies are estimated on the basis of the transition state theory using k_{cat} values. [e] In all QM/MM reaction path calculations, the QM region was limited to the substrate only. The QM/MM geometry optimizations were performed at MP2/6-31(+)G*(*) level followed by single point MP4 energy corrections with the same basis sets. The error estimates indicate the difference between the forward and backward MD-FEP simulation runs.

Table 2. Hydrogen-Bonding Parameters Sampled from the Constrained QM/MM MD-FEP Simulations for Each Reaction System (in Å)

H-bond type	state	wild-type	Lys90 mutant	Cit90 mutant	Asp78 mutant	Gln78 mutant	Lys7 mutant
1[a] (Tyr108-COO)	ES	2.575 ± 0.086	2.626 ± 0.107	2.772 ± 0.165	2.611 ± 0.099	2.617 ± 0.106	2.590 ± 0.090
	TS	2.593 ± 0.096	2.589 ± 0.093	2.679 ± 0.130	2.593 ± 0.087	2.608 ± 0.097	2.550 ± 0.084
	Prod	2.591 ± 0.117	2.597 ± 0.093	2.595 ± 0.094	2.676 ± 0.120	2.598 ± 0.092	2.578 ± 0.087
2 (residue 7-COO)	ES	2.774 ± 0.105	2.774 ± 0.095	3.243 ± 0.174	2.749 ± 0.084	2.741 ± 0.087	2.715 ± 0.075
	TS	2.716 ± 0.072	2.732 ± 0.076	3.037 ± 0.144	2.710 ± 0.075	2.720 ± 0.071	2.678 ± 0.067
	Prod	2.819 ± 0.115	2.801 ± 0.115	2.745 ± 0.089	2.756 ± 0.095	2.768 ± 0.097	2.757 ± 0.081
3 (residue 7-COO)	ES	3.094 ± 0.187	2.901 ± 0.132	2.743 ± 0.088	3.091 ± 0.164	3.233 ± 0.195	2.980 ± 0.150
	TS	3.009 ± 0.155	2.870 ± 0.138	2.729 ± 0.079	3.196 ± 0.265	2.936 ± 0.125	3.396 ± 0.252
	Prod	4.004 ± 0.143	3.926 ± 0.162	3.820 ± 0.166	3.379 ± 0.218	3.757 ± 0.180	4.564 ± 0.130
4 (residue 90-COO)	ES	2.735 ± 0.075	2.810 ± 0.103	3.053 ± 0.260	2.753 ± 0.083	2.812 ± 0.103	2.808 ± 0.091
	TS	2.744 ± 0.084	2.781 ± 0.096	2.939 ± 0.186	2.743 ± 0.077	2.779 ± 0.089	2.759 ± 0.088
	Prod	2.837 ± 0.115	3.767 ± 0.305	2.869 ± 0.124	2.796 ± 0.096	2.751 ± 0.076	2.720 ± 0.076
5 (residue 90-ether O)	ES	3.044 ± 0.186	3.020 ± 0.146	3.174 ± 0.199	3.140 ± 0.202	2.969 ± 0.174	2.947 ± 0.161
	TS	3.149 ± 0.260	3.000 ± 0.170	3.106 ± 0.198	3.270 ± 0.324	2.746 ± 0.083	3.038 ± 0.247
	Prod	3.863 ± 0.262	4.410 ± 0.299	4.257 ± 0.225	3.403 ± 0.283	3.038 ± 0.213	4.454 ± 0.251
6 (residue 78-residue 90)	ES	3.116 ± 0.460	2.778 ± 0.087	5.092 ± 0.529	2.753 ± 0.097	3.940 ± 0.345	4.103 ± 0.390
	TS	3.007 ± 0.356	2.807 ± 0.092	4.899 ± 0.524	2.757 ± 0.106	4.037 ± 0.268	3.841 ± 0.453
	Prod	2.827 ± 0.135	2.798 ± 0.100	4.974 ± 0.341	2.960 ± 0.346	3.828 ± 0.377	2.856 ± 0.148
7 (residue 78-OH)	ES	2.927 ± 0.259	2.636 ± 0.089	2.774 ± 0.150	3.533 ± 0.696	2.740 ± 0.158	2.658 ± 0.111
	TS	2.760 ± 0.144	2.652 ± 0.103	2.711 ± 0.135	5.711 ± 0.617	2.655 ± 0.103	2.669 ± 0.125
	Prod	2.639 ± 0.104	2.597 ± 0.088	2.602 ± 0.095	5.667 ± 0.292	2.644 ± 0.106	5.146 ± 0.336
8 (Cys75-OH)	ES	3.852 ± 0.381	3.666 ± 0.351	3.431 ± 0.295	3.829 ± 0.391	3.740 ± 0.601	3.898 ± 0.393
	TS	3.683 ± 0.327	3.635 ± 0.340	3.438 ± 0.301	4.007 ± 0.426	3.360 ± 0.212	3.880 ± 0.366
	Prod	3.752 ± 0.273	3.769 ± 0.343	3.706 ± 0.450	3.516 ± 0.254	3.927 ± 0.386	3.269 ± 0.158
9 (Arg63-COO)	ES	2.811 ± 0.125	3.212 ± 0.710	3.921 ± 0.162	2.766 ± 0.101	2.957 ± 0.451	4.462 ± 0.595
	TS	2.939 ± 0.176	2.878 ± 0.188	3.998 ± 0.168	2.778 ± 0.102	6.102 ± 0.984	2.776 ± 0.107
	Prod	2.847 ± 0.252	2.776 ± 0.103	6.754 ± 0.476	2.823 ± 0.120	2.836 ± 0.165	5.406 ± 0.753
10 (Arg116-COO)	ES	2.804 ± 0.132	2.797 ± 0.106	5.072 ± 0.324	2.794 ± 0.108	2.763 ± 0.093	2.765 ± 0.092
	TS	2.789 ± 0.106	2.903 ± 0.299	4.724 ± 0.314	2.783 ± 0.098	2.791 ± 0.104	2.780 ± 0.101
	Prod	2.769 ± 0.098	2.772 ± 0.096	2.776 ± 0.104	2.763 ± 0.096	2.763 ± 0.095	2.729 ± 0.084

[a] The Arabic numerals indicate the number of hydrogen bonds shown in Figure 1.

catalytic ability,[22,23,26,28] and the present free energy simulations support these experimental findings. The estimated roles of Arg7, Tyr108, Arg116, and Arg63 were nearly the same for the Lys90 mutant as observed in the wild-type reaction. A major difference was observed in the Lys90-Glu78-substrate alignment compared to that in the wild-type reaction. The loss of the sterically hindered guanidinium side chain of the original Arg90 canceled the electrostatic interaction with the substrate ether oxygen, which is a potential hydrogen bond donor for the negative induced charge appearing in the putative TS. In addition, Lys90 strongly interacted with Glu78 throughout the reaction process. Glu78 also strongly interacted with the substrate hydroxyl group. As a result, both ES and TS were equally stabilized by the multiple hydrogen-bonded environment at the active site. These are major structural factors of the anticatalytic effect. The positive charge concentrated on the original Arg90 side chain was not conserved at all in the case of the Cit90 mutant, although the side chain structure of the heavy atoms were relatively similar to the original enzyme. As a result, interactions within the Cit90-Glu78-substrate network were not efficient. Similar to the Lys90 mutant reaction, the carboxylate side chain of Glu78 strongly interacted with the substrate hydroxyl group along the reaction pathway and ES and TS were stabilized to the same extent. Another characteristic observation of the Cit90 mutant active site was the loss of remote site interactions. Strong hydrogen bonds between the substrate carboxyl group and the Arg116 and Arg63 side chains were not observed during the TS formation. This implies a small functional movement of two helices located on the solvent exposed surface.

Previous experimental observations have suggested that the anionic residue in position 78 was crucial for the stabilization of the reaction intermediate.[22–24] In contrast, present simulation results do not always support these observations. For the Asp78 mutant, mutation-induced structural change seems to be a minor factor because only one methylene group is removed from the original Glu78 side chain. This small structural perturbation was found to have negligible effects on the active-site hydrogen-bonded network. The side chains of Arg7, Tyr108, Arg116, and Arg63 have played a role in stabilizing the orientation of the

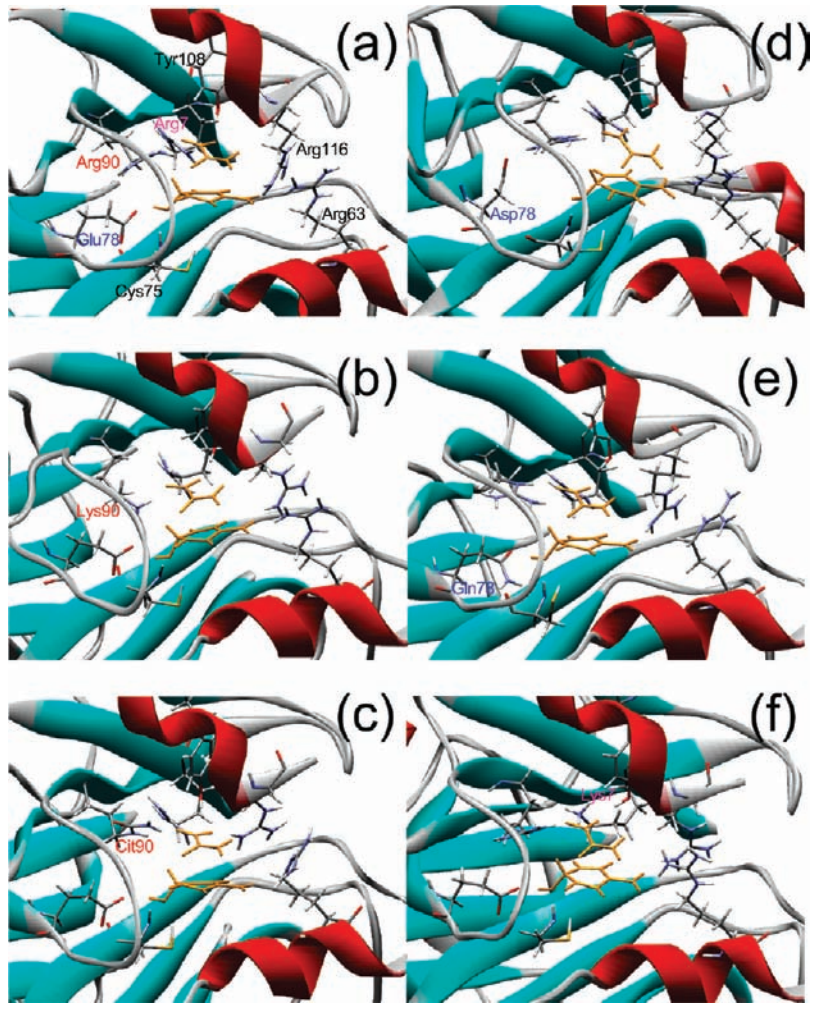

Figure 3. Typical active site structures in the TS region for each reaction system, sampled from *ab initio* QM/MM MD-FEP simulations along the selected reaction coordinate: (a) wild-type, (b) Lys90, (c) Cit90, (d) Asp78, (e) Gln78, and (f) Lys7 mutant enzymes. In all figures, only important residues that form hydrogen-bond pairs with the substrate (see Figure 1d) are drawn as stick models. The yellow sticks designate the TS geometries of the substrate.

reactive substrate through strong hydrogen bonds, as observed in the wild-type enzyme. However, the side chain of the Asp78 residue never acts as a hydrogen bond partner with the substrate but always forms strong hydrogen bonds with the guanidinium side chain of Arg90. As a result, the alignment of the hydrogen-bond network in Arg90-Asp78-substrate never formed. This situation was also observed for the Gln78 mutant. Again, an effective hydrogen-bond alignment in the Arg90-Gln78-substrate sequence was not observed along the reaction pathway. In addition, both Arg90 and Gln78 interacted with the substrate independently and simultaneously. The loss of anionic charge on residue 78 apparently affected remote site interactions between the second carboxyl group of the substrate and the Arg63 side chain during TS formation. In these mutants, ES and TS were equally stabilized mainly as a result of strong electrostatic effect induced by Arg90.

Compared to residue mutations at positions 78 and 90, experimental work has not clearly identified the catalytic effect of the Arg7 side chain to date, although kinetic measurements showed an apparent reduction of catalytic activity.[22] In fact, interaction residue analyses reported in previous computational studies suggested an important catalytic contribution from position 7.[41,42,53,55] The free energy simulations clearly demonstrated that there were two major structural differences even in the case of a perturbation inside the small active site. Although the catalytic alignment of the original Arg90-Glu78-substrate sequence was conserved in this case, the changes in hydrogen-bond structures were quite different from the pattern observed for the wild-type enzyme. The disappearance of the hydrogen bonds between the side chain of Arg90 and Glu78 in the ES complex caused the two residues to independently interact with the substrate. Moreover, the Glu78 carboxyl group strongly stabilized the substrate hydroxyl group during TS formation, resulting in a loss of catalytic activity upon relative stabilization of ES. In addition, the strong hydrogen bonds were found to disappear at the remote interaction site, suggesting a loss of effective functional interaction between the two helices located on different protein domain surfaces.

In summary, these free energy simulation results clarify two important factors involved in the catalytic process. The first is the effective control of fine hydrogen-bond distances in the original Arg90-Glu78-substrate alignment, and the second relates to interactions between two helices located in the lid position above the binding pocket or at different protein domain surfaces, which are apart from the active site.

3.3. Electronic Structures of Local Active Site. Further, interaction energies between the reactive substrate and the surrounding protein environment were investigated. We clearly demonstrated that the electrostatic energy was the leading interaction energy term in molecular interactions in the wild-

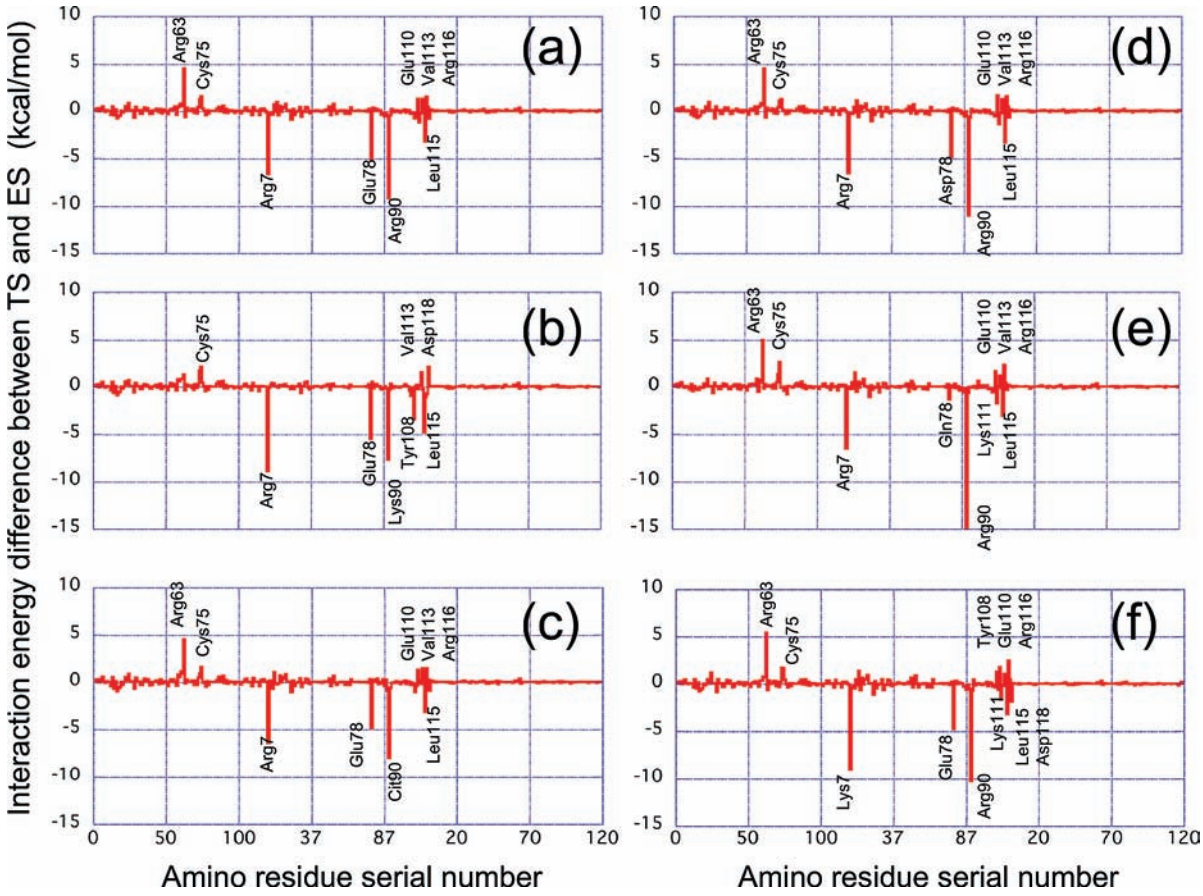

Figure 4. Interaction energy differences between TS and ES complex for each reaction system determined by FMO2-MP2/6-31G* computations: (a) wild-type enzyme of original Arg90 residue, (b) Lys90, (c) Cit90, (d) Asp78, (e) Gln78, and (f) Lys7 mutant enzymes. In all figures, the *x*-axis corresponds to the amino acid residue serial number, and the *y*-axis designates the interaction energy differences between the two selected states determined by QM/MM structural modeling (in kcal/mol).

type reaction process.[55] To identify differences in electrostatic environment among the five mutant variants, we followed the same systematic theoretical analysis as above. First, molecular geometries were reoptimized along the reaction coordinate using the large-QM region models. Then, after the geometrical parameters and the energetics at the QM/MM level were validated, all-electron QM calculations were performed for entire protein complexes at the FMO2-MP2/6-31G* level. The main focus was to identify the electrostatic effect induced by mutations in the catalytic process; therefore, interaction energy differences between the ES and TS are discussed exclusively. The results of each reaction system are shown in Figure 4.

Because the energetic components in the wild-type reaction have already been reported in detail elsewhere,[55] only the differences between the wild-type and five mutant reactions are discussed here. As noted above, position 90 mutations resulted in an apparent catalytic reduction. Consistent with this fact, the transition state stabilization (TSS) energies of the Lys90 and Cit90 mutants decreased compared to the wild-type enzyme. Regarding position 78 mutations, the Asp78 mutant appeared to have a rather similar profile to the wild-type reaction. However, the profile of the Gln78 mutant was significantly different from other residue 78 mutants. The loss of a negative charge on the Glu78 residue resulted in an interaction energy increase between Arg90 and the substrate. The unchanged residues did not display any apparent electrostatic effect in the Lys7 mutant despite a greater interaction between Lys7 and the substrate.

The analysis of interaction energy components may be useful for identifying the particular roles of amino acid residues. However, the results largely depend on molecular structure along the reaction coordinate. As previously reported,[55] the QM/MM optimized geometries involved in wild-type reaction were representative structures sampled through MD simulations along the selected reaction coordinate. However, the five mutant enzymes exhibited relatively large structural fluctuations around the enzyme active site during TS formation (Figure 3 and Table 2). Also, the QM/MM optimized static structures do not always represent averaged structures. Therefore, the fluctuations of the electrostatic interaction energies were estimated through direct QM/MM sampling calculations. Because the constrained MD-FEP simulations were performed along the reaction coordinate, the electrostatic interaction energy was simultaneously re-evaluated at the *ab initio* QM/MM level using thousands of MD trajectories. The electrostatic interaction energy was calculated as[56]

$$\langle \Psi^{\mathrm{QM/MM}} | \hat{H}^{\mathrm{elec}}_{\mathrm{QM/MM}} | \Psi^{\mathrm{QM/MM}} \rangle \qquad (7)$$

where $|\Psi^{\mathrm{QM/MM}}\rangle$ denotes the QM/MM wave function, and $\hat{H}^{\mathrm{elec}}_{\mathrm{QM/MM}}$ describes the electrostatic interactions between QM and MM regions. The electrostatic energy distributions of ES, TS, and product states are summarized in Figure 5. For the wild-type reaction, the interaction energy distributions showed a quasi-Gaussian shape for each state, implying that the substrate structure fluctuated around the geometries optimized at each

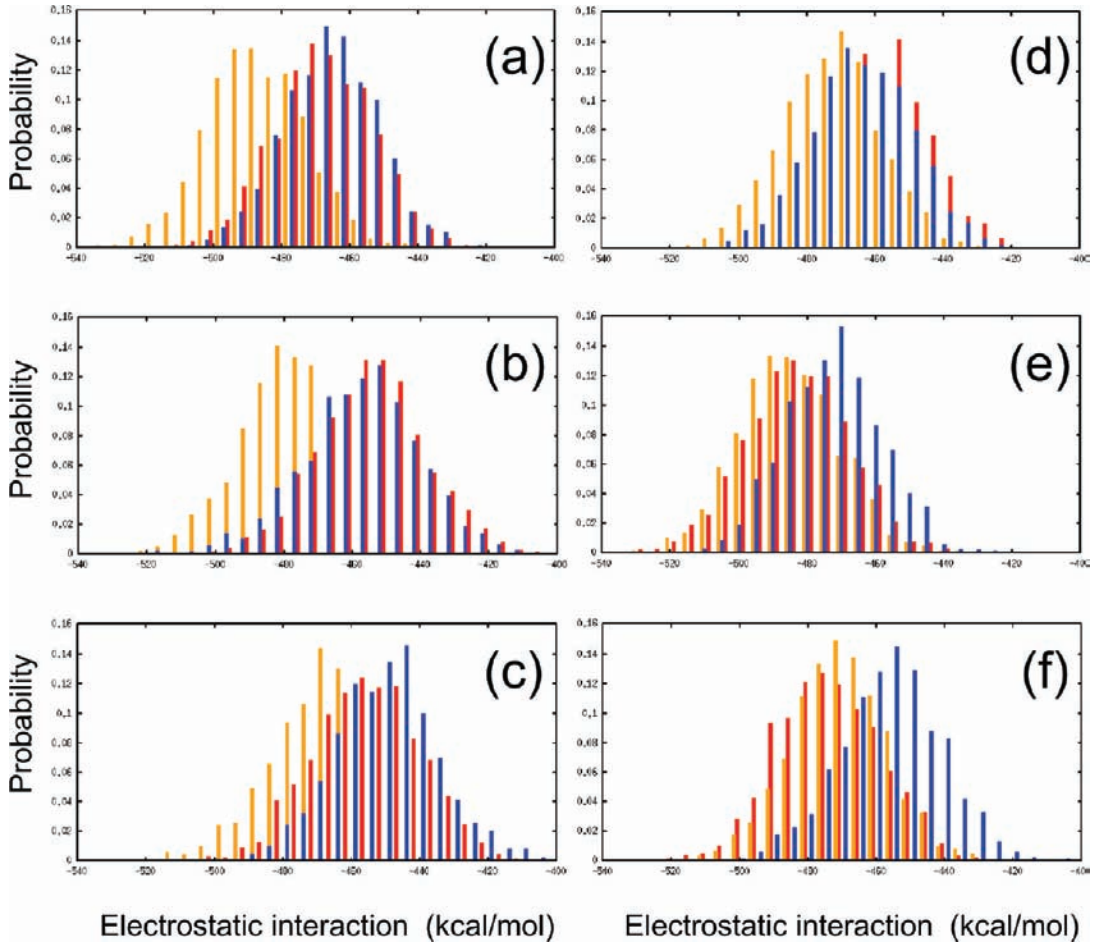

Figure 5. QM/MM electrostatic interaction energy distribution along the reaction coordinate using direct *ab initio* QM/MM sampling calculations (QM/MM RHF/6-31(+)G**/AMBER (parm.96) level): (a) wild-type enzyme of original Arg90 residue, (b) Lys90, (c) Cit90, (d) Asp78, (e) Gln78, and (f) Lys7 mutant enzymes. In all figures, the *x*-axis corresponds to the electrostatic interaction energy, and the *y*-axis represents the probability distribution. Histograms shown in red, yellow, and blue correspond to the ES complex, TS, and the product state, respectively.

step along the reaction coordinate. A large and negative electrostatic interaction was observed in the TS region only. Because ES and product states showed relatively similar distribution profiles, the net TSS factor originated from the differences in energy distribution between TS and ES. The residue 90 mutations were found to largely affect the strength of electrostatic energy inside the active site. Approximately the same amount of electrostatic interaction was stored in Lys90 and Cit90 mutant reactions, but the degree of TSS decreased significantly, especially in the Cit90 mutant reaction. Interestingly, the residue 78 mutations showed different profiles. Nearly the same electrostatic energies were observed in each state along the reaction pathway, suggesting that only a small amount of TSS energy was supplied for residue 78 mutations, in particular for Gln78 mutant reaction. A similar behavior was observed for the Lys7 mutant. Although the Lys7 side chain was strongly bonded to the buried part of the substrate carboxyl group, but not the reactant scissile portion, the Lys7 mutation largely affected the electrostatic environment of the active site. The same electrostatic distribution was observed in ES and TS states.

3.4. Conformational Changes in Global Protein Motion. We next studied the global protein motions along the selected reaction coordinate. Identifying the collective protein motions, which may contribute to enzyme function, is a highly debated issue, and several proposals aiming at extracting protein functional motions have recently been proposed.[87−90] Because our primary concern was to easily capture point mutation effects on protein dynamics along the selected reaction coordinate, we analyzed the cross-correlation coefficient for the displacement of any two C_α atoms in each amino acid unit, as defined in eq 4. In general, a global protein motion can be analyzed through correlated motions because proximal residues composing a well-defined secondary structure or a domain region usually exhibit a correlated movement that reflects communication among separate domains.[81] The elements of C_{ij} can be collected and presented in a matrix form. Typical data for the domain 1 region of the wild-type BsCM complex are shown in Figure 6c. Several apparent correlated motions were found inside each protein domain, suggesting that the protein tertiary structure was composed of several secondary structural motifs (Figure 6). Three helices composed of residues 17−35, 58−65, and 110−115 and four sheets composed of residues 3−11, 45−49, 73−77, and 88−96 were observed in each protein domain. On the basis of the 2D diagram of a monomer unit, we analyzed the overall correlation map of each enzyme reaction.

(87) Hammes-Schiffer, S.; Benkovic, S. J. *Annu. Rev. Biochem.* **2006**, *75*, 519–541.
(88) Olsson, M. H. M.; Parson, W. W.; Warshel, A. *Chem. Rev.* **2006**, *106*, 1737–1756.
(89) Nagel, Z. D.; Klinman, J. P. *Nat. Chem. Biol.* **2009**, *5*, 543–550.
(90) Schwartz, S. D.; Schramm, V. L. *Nat. Chem. Biol.* **2009**, *5*, 551–558.

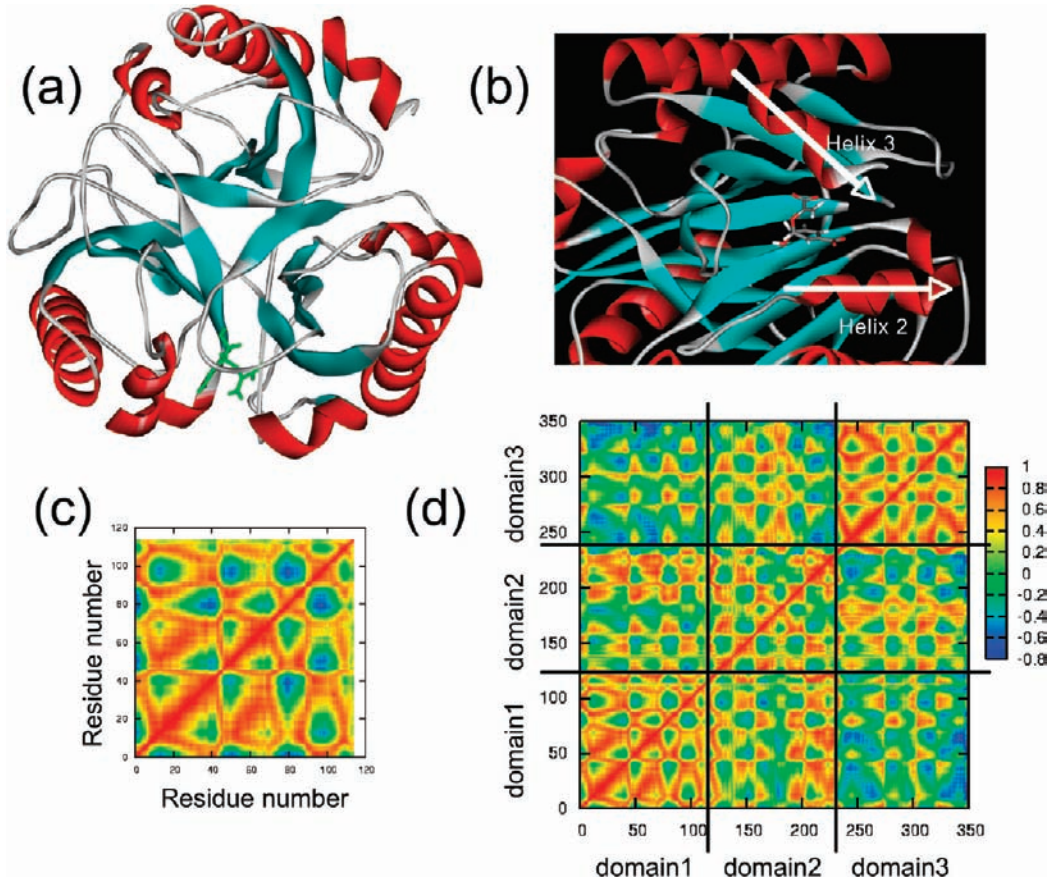

Figure 6. (a) Overall protein structure of the BsCM complex. Substrate (chorismate, represented by green sticks) is located on the boundaries between two domain units. (b) Close-up of the wild-type active site structure. The substrate is surrounded by two helical motifs on the surface of solvent-exposed region. (c) Cross-correlation diagram of one protein unit extracted from the wild-type BsCM complex. (d) Cross-correlation diagram of the ES complex state of the BsCM reaction profile. In both diagrams, *x*- and *y*-axes correspond to the amino acid residue number, and in this case the original residue number assigned in each domain was replaced by serial number throughout all trimer units.

Figure 7a−c shows the dynamic cross-correlation maps of the three selected states for the wild-type reaction along the reaction coordinate. The intrinsic protein structure is composed of three identical domains, leading to several separate regions in each map. A characteristic finding for the ES complex was the presence of an apparent correlated movement through separate protein domains that corresponded to helix 2 (58−65) in domain 1 and helix 3 (110−115 of original sequence, 223−228 of serial number used in Figure 7) in domain 2. Upon TS formation, overall correlation motions inside the enzyme were significantly enhanced compared to the ES complex. The correlation between domains 1 and 2 was strongly enhanced. After formation of the product prephanate, correlations between domains almost disappeared and only local correlation motions inside each three domain unit were mainly observed.

The effect of residue 90 mutations on catalysis was examined next. Figure 7d−f shows the dynamic cross-correlation maps of the Lys90 mutant, and Figure 7g−i shows those of the Cit90 mutant. Introducing a point mutation at residue 90 position caused the degree of correlation motion inside each protein domain to increase slightly, although the pattern of the correlation movement remained unchanged. One notable difference between wild-type and Lys90/Cit90 mutant enzymes was a decrease in interactions between domains 1 and 2. Along the reaction coordinate, the correlations inside each unit were largely enhanced, especially upon TS formation. These trends were the same as for wild-type reactions, although the pattern of the correlation movements between the separate domains was very different in each of the three reaction types. When the reaction is completed and product is formed, the correlation inside the enzyme decreased significantly and major correlated movements were only observed inside each of the three protein domains, similar to the wild-type reaction.

Figure 8a−c shows the dynamic cross-correlation maps of the Asp78 mutant, and Figure 8d−f shows those of Gln78. Although the chemical structure of the Asp78 mutant was significantly similar to that of the wild-type enzyme compared to the other mutant variants, the calculated cross-correlation movements exhibited different patterns along the reaction pathway. Again, one major difference was the loss of the correlation movement between the two helices located on different domains in the ES complex. The results were slightly different for the Gln78 mutant enzyme. More correlated movements were observed in the ES complex than in the TS. Figure 8g−i shows the dynamic cross-correlation maps of the Lys7 mutant reaction. Although the Lys7 residue did not form an apparent hydrogen bond with the substrate scissile portion nor change the hydrogen-bond pattern around the buried carboxyl group, the calculated 2D cross-correlation map of the Lys7 mutant was rather different from that of the wild-type reaction system. As the reaction proceeded, correlation movements inside the enzyme increased, although the pattern of correlations between separate domains differed from that of the original wild-type profile.

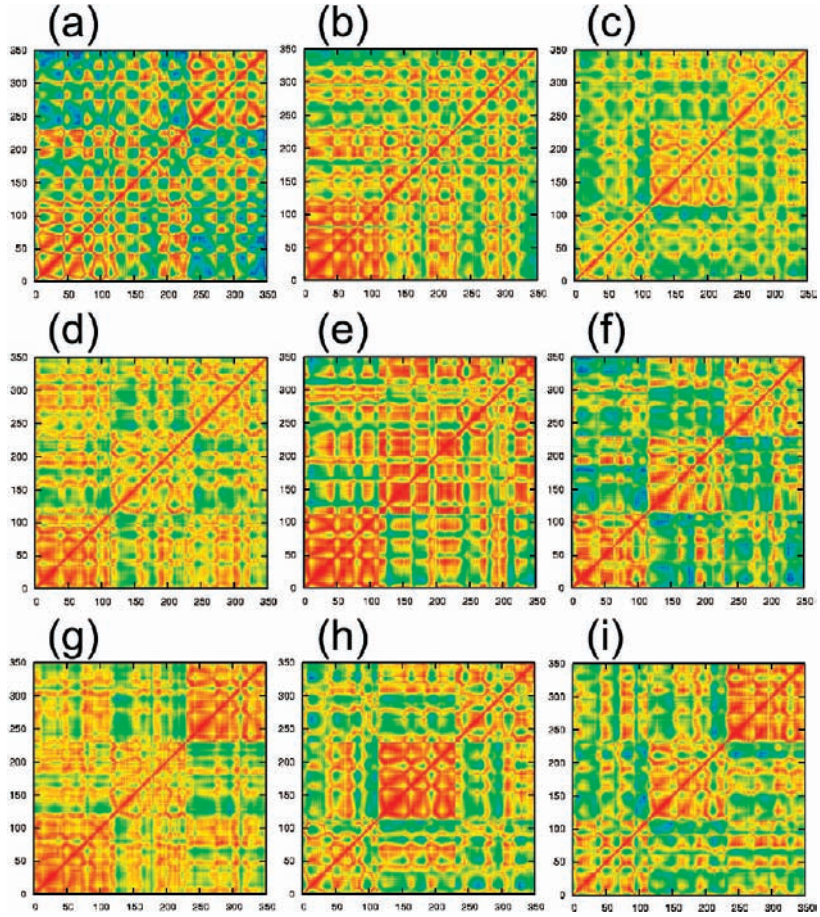

Figure 7. (a−c) Cross-correlation diagrams of the wild-type enzyme reaction along the reaction coordinate: ES complex (a), TS (b), and product states (c). (d−f) Cross-correlation diagrams of the Lys90 mutant enzyme reaction along the reaction coordinate: ES complex (d), TS (e), and product states (f). (g−i) Cross-correlation diagrams of the Cit90 mutant enzyme reaction along the reaction coordinate: ES complex (g), TS (h), and product states (i).

3.5. Electronic Structures of Global Protein Structures. Finally, the effect of electronic polarization on entire protein structures was investigated. Local electrostatic effects between substrate and surrounding amino acid residues were discussed through interaction energy analysis and summarized in the one-dimensional energy diagram shown in Figure 4 (see above). As clearly demonstrated previously,[55] QM/MM and all-electron QM treatments essentially produced the same energy diagrams. Because most QM/MM approaches assume fixed partial charges assigned on each atomic site, interaction energy components between any amino acid residue pair were considered almost unchanged along the reaction coordinate if the protein structure did not undergo a large conformational change.[39–43,51] However, because both the substrate and proteins are polarized in response to electronic reorganization in the TS during chemical reaction, certain degrees of electronic perturbation may affect the electronic structure of the overall protein architecture. In principle, this effect can be investigated by analyzing the interaction energy components in all amino acid component pairs.

We have previously reported in detail the interaction energy components during the catalytic process through all-electron QM calculations and demonstrated that the electrostatic interaction energy was the major component for the wild-type reaction.[55] The same QM-based energy analysis was followed for all of the interaction residue pairs inside the enzymes, and the effect of point mutations with respect to the response of electronic structure against the electrostatic perturbation on the entire protein was discussed. To identify the degree of electrostatic interaction energies, we calculated electrostatic energy components for all amino acid residue pairs and evaluated the differences between ES and TS along the reaction coordinate. Because of the high computational cost of all-electron QM calculations within the FMO framework, the discussion is limited to geometries optimized along the selected reaction coordinate. The results for each mutant and the reference wild-type reactions are summarized in the 2D interaction energy diagram (Figure 9).

Figure 9a shows the differences in electrostatic interaction energies for the wild-type enzyme reaction. In this figure, the data on the selected plane ($x = 350$ and $y = 350$) have the same meaning as the energy diagram shown in Figure 4. As explained above, a negative peak contributes to the TSS, and a positive one has an anticatalytic effect. Figure 9 shows that most interaction energy components originate from the protein−substrate interaction energy, whereas interactions within any enzyme amino acid residues have no apparent catalytic or anticatalytic contributions. These results clearly demonstrate that electronic polarization has little effect on the net TSS energy, thus suggesting that the electronic structure of each amino acid residue inside the protein is rather stable against external electrostatic perturbations. The electronic response of the protein structure against the point mutations was discussed next. For Lys90, Asp78, Lys7 mutants, which conserved the residual charge on each side chain group, the overall character of the 2D interaction energy diagram was very similar to that of the

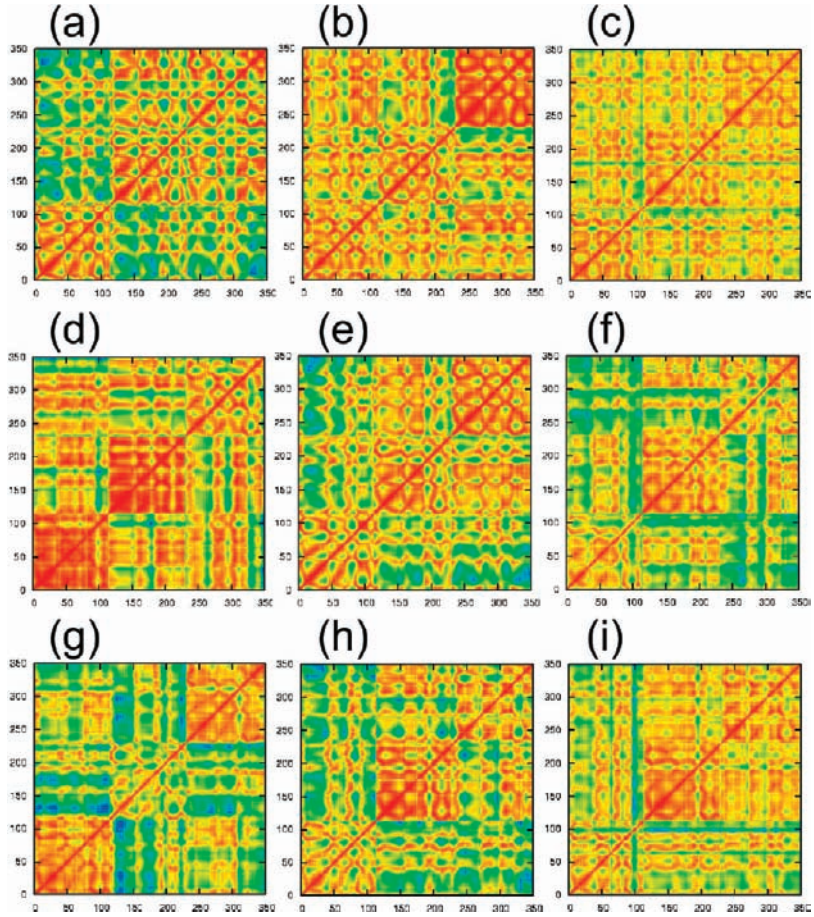

Figure 8. (a−c) Cross-correlation diagrams of the Asp78 mutant enzyme reaction along the reaction coordinate: ES complex (a), TS (b), and product states (c). (d−f) Cross-correlation diagrams of the Gln78 mutant enzyme reaction along the reaction coordinate: ES complex (d), TS (e), and product states (f). (g−i) Cross-correlation diagrams of the Lys7 mutant enzyme reaction along the reaction coordinate: ES complex (g), TS (h), and product states (i).

wild-type reaction, although the degree of interaction between domains 1 and 2 was slightly affected. However, for Cit90 and Gln78 mutants that alter the side chain charge state, the resulting 2D interaction energy diagrams showed a qualitatively different character. Several residues on domain 2 were significantly affected, especially in the Cit90 mutant reaction. These results imply that point mutations of critical residue that destroy the charge balance may affect the protein electronic structures in some cases.

4. Discussion

4.1. Comparison between Computational and Experimental Data. On the basis of the computational modeling, the simulation results were compared with experimental observations. Experimental data have suggested that active site residues may stabilize the developing charge on the enol ether oxygen and the cyclohexadiene ring in the polar transition state. To validate this hypothesis, Cload et al. have performed systematic mutational experiments and measured kinetic parameters for 16 mutants of the BsCM reaction.[22] In their discussion, they assumed that the mutations did not cause any significant structural change, mainly through circular dichroism spectroscopy measurements.

Structural studies suggested that hydrogen bonds between the buried carboxyl group and Arg7 play an important role in positioning the enolpyruvyl side chain during the formation of the chairlike TS structure.[13,22] Consistent with this idea, the substitution of Arg7 with Lys7 was shown to reduce the k_{cat}/K_M parameter by $\sim 10^3$. This experimental observation may be because the ε-amino group of the Lys7 mutant reduces hydrogen bonding to the buried substrate carboxyl group.[22] However, the present simulation predicted a different catalytic mechanism. Substituting the original Arg7 with Lys7 did not change interactions among the carboxyl groups but affected the hydrogen-bond alignment of the Arg90-Glu78-substrate sequence. The resulting structural change stabilized both ES and TS to the same extent.

Regarding mutations at residue 78 position, one mechanistic assumption suggested that Glu78 inside the active site may stabilize the positive charge that develops in the vicinity of the hydroxyl group in a dipolar transition state. Experiments have shown that Gln78 and Ala78 mutants exhibited significantly reduced catalytic activities (>10^4-fold) compared to that of the wild-type reaction, whereas the Asp78 mutant displayed only a 2-fold decrease in its k_{cat} value.[22,24] On the basis of these measurements, Cload et al. suggested that a carboxyl group was required in the vicinity of the substrate hydroxyl group, most likely to orient the substrate and the guanidinium side chain of Arg90.[22] Another systematic mutation study at residue 78 position has been reported by Kast et al.[24] Through targeted randomizing mutagenesis and selections, they probed the role of a negative charge at position 78 and confirmed that the isosteric but neutral Gln78 variant was essentially inactive. They concluded that an electrostatic gradient in the active site was a

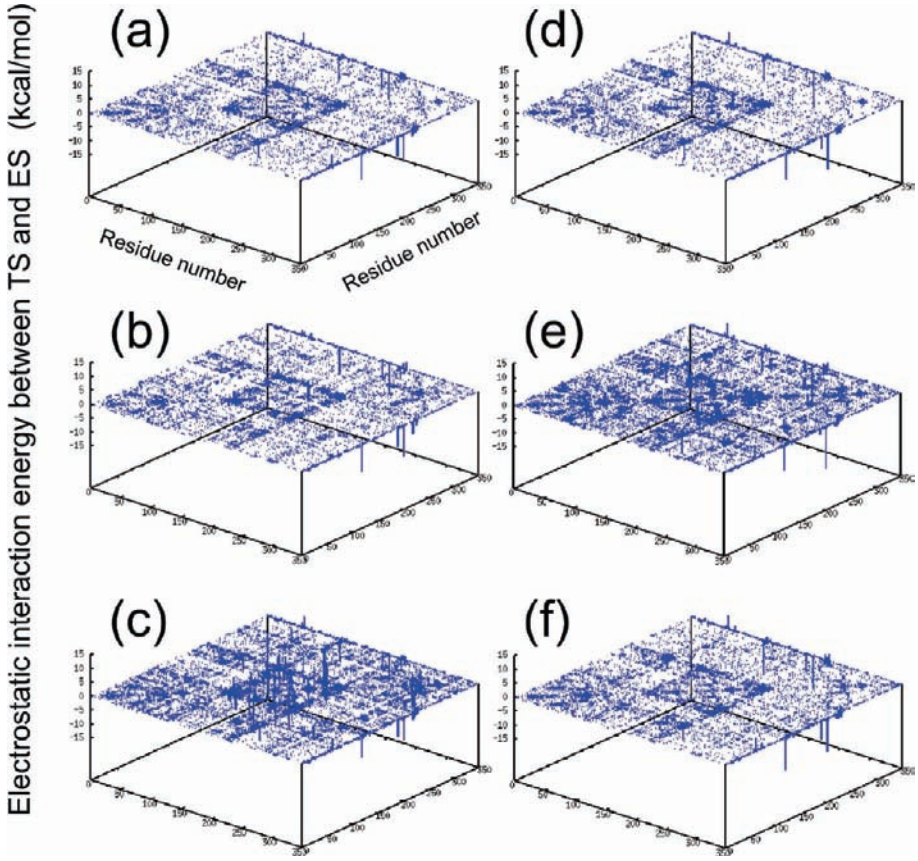

Figure 9. Differences in the electrostatic interaction energies for all of the amino acid pairs between TS and ES complex for each reaction system determined by FMO2-RHF/6-31G* computations: (a) wild-type enzyme of original Arg90 residue, (b) Lys90, (c) Cit90, (d) Asp78, (e) Gln78, and (f) Lys7 mutant enzymes. In all figures, the *x*- and *y*-axes correspond to the amino acid residue serial number (for convenience, we use the serial number throughout trimer units of BsCM, instead of original residue number assigned in each domain unit), and the *z*-axis designates the electrostatic interaction energy differences between the two selected states optimized by QM/MM structural modeling (in kcal/mol).

major catalytic factor in the BsCM reaction.[24] The present simulations clearly demonstrated that the major role of Glu78 is to appropriately align the catalytic components in the hydrogen-bonded network of Arg90-Glu78-substrate hydroxyl group, while the carboxylate negative charge may contribute to the TSS to some extent. These simulation results support the idea proposed by Cload et al.

Both structural studies and kinetic measurements predicted that one of the crucial catalytic residues was Arg90 located adjacent to the substrate ether oxygen.[13,22,26] Cload et al. reported that substituting Arg90 with Lys reduced k_{cat}/K_M by $\sim 10^4$ relative to the wild-type reaction, thereby decreasing the catalytic activity. In addition, the Ala90 mutant had no apparent catalytic activity.[22] They suggested that hydrogen bonds formed with Arg90 may stabilize the negative charge, which develops at the TS ether oxygen, and the required chairlike TS formation. Kast et al. also reported the detailed kinetics and structural characterization of Arg90 mutants based on combinatorial mutagenesis and in vivo selection.[26] They also observed that removing the Arg side chain reduced the catalytic activity by more than 5 orders of magnitude and concluded that a developing negative charge in the highly polarized TS was electrostatically stabilized by a sterically placed cation. This conclusion was largely supported by recent deterministic experiments that introduced a Cit residue in the original Arg position. Kienhöfer et al. have performed kinetic measurements of k_{cat} and K_M parameters, which showed $\sim 10^4$ orders decreases in k_{cat} concomitant with modest increases in K_M. They concluded that the Arg90 guanidinium group formed a complementary electrostatic interaction with the developing negative charge at the TS ether oxygen.[28] Our simulation supports the idea, which suggests that the Arg90 residue mainly stabilizes TS electrostatically. However, simplifying the role of Arg90 is not an easy task. Replacing the original Arg by Lys also destroyed the effective catalytic alignment in the active site. Consequently, both ES and TS were equally stabilized by the surrounding polar residues. In addition, introducing Cit changed the charge state of the positive side chain, affecting remote site interactions between domains 1 and 2 and resulting in an increase of activation free energy.

4.2. Effects of Point Mutation on Enzymatic Activity. In this study, we have investigated the effect of point mutation on enzymatic reactivity. By selecting chorismate mutase, one of the simplest reaction systems, we analyzed the catalytic and anticatalytic factors introduced by mutations from the viewpoint of electronic structure and protein dynamics. Our systematic modeling procedure clearly showed that the electrostatic perturbation effect on the electronic structure of the entire enzyme complex was limited to the local active site region, whereas the effect on the protein dynamics, correlated movement of each protein amino acid residue, sometimes changed site interactions that were remote from the mutation site. Experimental results of mutations are usually explained by decomposing side chain effects into two simple but different concepts, which consist of steric (or shape) and charge elements.[22–24,26,28] However, as clearly demonstrated in this study, both elements are closely

related to each other, making it difficult to dissociate the two catalytic factors.

The major catalytic apparatus depends on two factors in the BsCM reaction. First, multiple hydrogen bonds are formed between two carboxyl groups and several key amino acids including (1) Arg7, Tyr108, and Arg116 on the C-term helix of domain 2, and (2) Arg63 on the second helix on domain 1. These hydrogen bonds were crucial to properly maintain substrate orientation at the active site throughout the reaction. Second, a detailed balancing of the hydrogen-bond distances achieves a proper Arg90-Glu78-substrate alignment. To maximize the electrostatic stabilization energy stored in the molecular interactions between enzyme structure and substrate, only the TS should be stabilized energetically compared to the ES complex. By controlling these simple geometrical parameters, the wild-type enzyme achieves the above requirements efficiently and effectively. When this fine balance is broken by the introduction of point mutations, both residues 78 and 90 tend to form hydrogen bonds without cooperative communications. As a result, the active site environment tends to bind both the ES and TS to the same extent, leading to an increase in activation free energy. Even when the ideal catalytic alignment is broken, the enzyme always tends to reorganize its structure to maximize the reactive substrate stabilization. This conformational effect sometimes appears to occur through a change in remote site interactions and in the scissile portion of the chemical reaction. For BsCM, the active site pockets are located at protein domain boundaries, where two helix motifs are aligned on the protein surface. The well-balanced interactions of these two helices allow substrate recognition and stabilization.

These computational results clearly imply that the wild-type active sites are optimally designed to stabilize the TS geometry along the reaction coordinate only. Electrostatic interactions between substrate and protein are significantly enhanced in the TS region compared to mutant reactions, and conformation changes around the active site are usually limited to a very small region, which implies that only a small amount of reorganization energy penalty is exacted during the reaction. In many mutant reactions, a large-scale conformational change usually adds to the reorganization energy cost. In addition, the electrostatic stabilization energy is stored less efficiently through the hydrogen-bond network within the mutant active site. These two factors are the major origin for the activation free energy increase in mutant variants. Because these two elements are closely related to each other, explaining the role of amino acids through simple concepts of charge or shape is usually difficult or sometimes leads to a misinterpretation of enzyme functions.

Finally, we would like to briefly consider the prospects for this computational approach to be applied to other important classes of enzymes. As suggested by earlier pioneering works and this study,[11,37,38,44,52,56,57] the electrostatic factor may work as a leading term in many cases of enzymatic catalysis in which protein environments act as effective electrostatic media on the whole. However, the precise control of the *transition state stabilization* seems to be achieved by small numbers of active site residues, and this local catalytic event may be effectively controlled by the collective motion of amino acid residues through protein dynamics. This implies that the catalytic activity of enzymes is mainly determined by the fine structure of the active site, and some external effects due to point mutations are properly modulated by the remaining protein domains that are apart from the active site.

5. Conclusion

In this article, we investigated the effects of point mutation in the chorismate mutase catalyzed interconversion of chorismate into prephenate. We analyzed the atomistic details of mutation effects from the viewpoints of electronic structure and protein dynamics using a systematic computational modeling strategy based on *ab initio* QM/MM calculations combined with MD-FEP simulations and all-electron QM calculations for the entire protein complex. First, we determined reaction free energy profiles for the wild-type enzyme and five mutant variants using these *ab initio* QM/MM combined with MD-FEP simulations. After validating the computational energy profiles by qualitatively and quantitatively comparing computations and experiments, we analyzed active site electronic structures and protein motions along the selected reaction coordinate for the six reaction systems. The effect of electrostatic perturbation into the electronic structure of the whole enzyme was found to be limited to the local active site region. This suggests that proteins are electrostatic media for optimum catalytic environments, which are generally robust against external electrostatic perturbations. In contrast to the electronic character of protein, enzyme structures are slightly flexible and change their conformations to effectively stabilize the transition state structure in response to steric perturbation induced by mutations. This suggests that local structural modifications may affect the global protein dynamics through correlated motions of particular amino acid residues even remotely separated from the mutation site. Compared to the mutant variants, the wild-type enzyme is most effectively designed to stabilize the transition state of the chemical reaction along the reaction coordinate. Electrostatic interactions between substrate and protein are significantly enhanced in the TS region compared with other mutant reactions. Moreover, changes of protein motion are limited to a very small active site region, implying that only a small amount of reorganization energy penalty is exacted not to expend additional free energy cost.

Acknowledgment. The author thanks Dr. Hiroshi Ishikita for fruitful discussions. The author is also grateful to Prof. Kazuo Kitaura for his continuous encouragement. This work was supported by a Grant-in-Aid for Young Scientists from MEXT, KAKENHI (no. 19750021), and the Next Generation Super Computing Project, Nanoscience Program, MEXT, Japan. Part of the numerical calculations were performed at the Computer Center of the Institute for Molecular Science (IMS).

JA100744H