J|A|C|S
ARTICLES
Published on Web 09/04/2003

Theoretical Perspectives on the Reaction Mechanism of Serine Proteases: The Reaction Free Energy Profiles of the Acylation Process

Toyokazu Ishida and Shigeki Kato*

Contribution from the Department of Chemistry, Graduate School of Science, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan

Received November 20, 2002

Abstract: The reaction mechanism of serine proteases (trypsin), which catalyze peptide hydrolysis, is studied theoretically by ab initio QM/MM electronic structure calculations combined with Molecular Dynamics−Free Energy Perturbation calculations. We have calculated the entire reaction free energy profiles of the first reaction step of this enzyme (acylation process). The present calculations show that the rate-determining step of the acylation is the formation of the tetrahedral intermediate, and the breakdown of this intermediate has a small energy barrier. The calculated activation free energy for the acylation is ∼17.8 kcal/mol at QM/MM MP2/(aug)-cc-pVDZ//HF/6−31(+)G*(*)/AMBER level, and this reaction is an exothermic process. MD simulations of the enzyme−substrate (ES) complex and the free enzyme in aqueous phase show that the substrate binding induces slight conformational changes around the active site, which favor the alignment of the reactive fragments (His57, Asp102, and Ser195) together in a reactive orientation. It is also shown that the proton transfer from Ser195 to His57 and the nucleophilic attack of Ser195 to the carbonyl carbon of the scissile bond of the substrate occur in a concerted manner. In this reaction, protein environment plays a crucial role to lowering the activation free energy by stabilizing the tetrahedral intermediate compared to the ES complex. The polarization energy calculations show that the enzyme active site is in a very polar environment because of the polar main chain contributions of protein. Also, the ground-state destabilization effect (steric strain) is not a major catalytic factor. The most important catalytic factor of stabilizing the tetrahedral intermediate is the electrostatic interaction between the active site and particular regions of protein: the main chain NH groups in Gly193 and Ser195 (so-called oxyanion hole region) stabilize negative charge generated on the carbonyl oxygen of the scissile bond, and the main chain carbonyl groups in Ile212 ∼ Ser214 stabilize a positive charge generated on the imidazole ring of His57.

Introduction

One of the most important topics in biochemistry is to understand the nature of enzyme reactions at a molecular level.[1−3] Despite the intensive works over several decades, there is still no quantitative understanding of how an enzyme works. Owing to the rapid progress of computational chemistry, understanding of the structure−function relationship for enzyme reactions is becoming an important issue in theoretical chemistry. In this article, we present theoretical perspectives on the reaction mechanism of serine proteases, one of the most extensively studied family of enzymes,[4] based on theoretical calculations.

The serine proteases are endopeptidases whose catalytic mechanism is based on an active site serine residue. The catalytic ability of serine proteases is considered to be originated from the catalytic triad residues (His57, Asp102, and Ser195) and the oxyanion hole at the active site. At present, the most likely reaction mechanism is the acylation/deacylation mechanism, as summarized in Figure 1.[1,4] Despite extensive studies in the past,[1,2,4] there still remain several questions about this mechanism. One is with regard to the formation of tetrahedral intermediates in both of the reaction steps. Their formations are very likely based on a consideration of simple organic reactions and the effectiveness of tetrahedral intermediate analogues as inhibitors, but there is no direct experimental evidence for their formations.[5,6] Another is about the hydrogen-bond network at the active site, especially the catalytic relay system between His57 and Asp102. Though early works on this system implied the double proton transfer called the *charge relay* mechanism,[7,8] many experimental evidences have favored the

(1) 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.
(2) Jencks, W. P. *Catalysis in Chemistry and Enzymology*; Dover Publications: New York, 1987.
(3) Many standard textbooks of biochemistry illustrate reaction mechanism of serine proteases. For example, see: Voet, D.; Voet, J. G. *Biochemistry*, 2nd ed.; John Wiely & Sons: New York, 1995.
(4) Progress before mid 90's is well described in Wharton, C. W. *Comprehensive Biological Catalysis*; Sinnott, M., Garner, C. D., First, E., Davies, G., Eds.; Academic Press: New York, 1998; Vol. 1, Chapter 9, and most experimental results cited in this article are based on this book.
(5) Fersht, A. R.; Requena, Y. *J. Am. Chem. Soc.* **1971**, *93*, 7079−7087.
(6) Hirohara, H.; Bender, M. L.; Stark, R. S. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 1643−1647.
(7) Blow, D. M.; Birktoft, J. J.; Hartley, B. S. *Nature* **1969**, *221*, 337−340.

10.1021/ja021369m CCC: $25.00

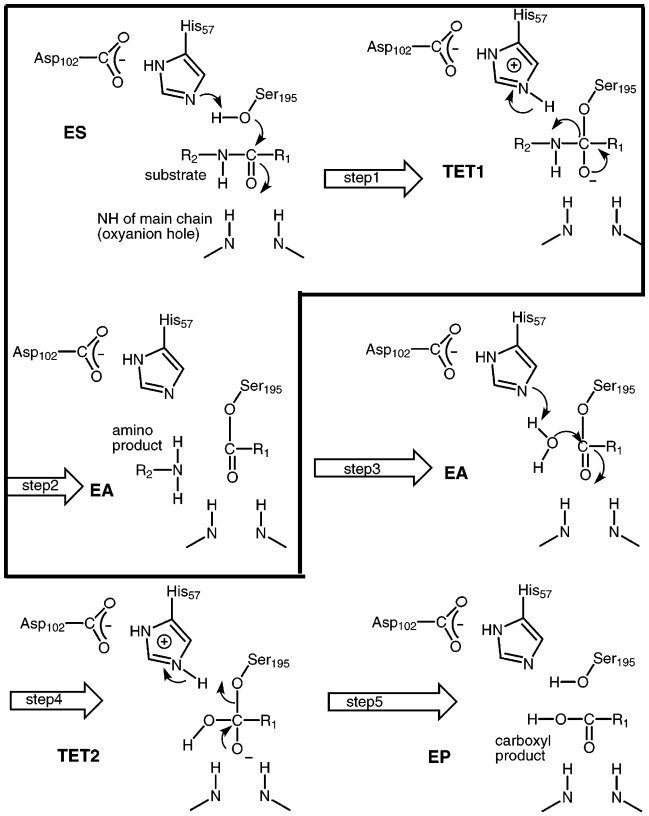

Figure 1. Proposed reaction mechanism of serine proteases. Only the main elements of the catalytic system (catalytic triad of His57, Asp102 and Ser195, oxyanion hole, peptide substrate) are shown in this schematic picture. The following abbreviations are used: ES denotes the enzyme−substrate complex (Michaelis complex); TET1/TET2 denote the first/second tetrahedral intermediate; EA denotes the acyl-enzyme intermediate; and EP denotes the enzyme−product complex. The reaction steps enclosed squarely in this figure (step 1/step 2) are the acylation process which we consider in this article.

single proton transfer.[9−11] Recently, some research groups proposed that a low-barrier hydrogen bond (LBHB) formed between His57 and Asp102 causes significant rate enhancement by stabilizing the transition state.[12,13]

Several groups applied quantum chemical calculations to the catalytic triad system in serine proteases. The earlier studies were based on semiempirical MO methods and considered only the catalytic triad amino residues without the surrounding protein. These calculations supported the double proton transfer.[14] Warshel and co-workers used the EVB−FEP method to calculate the reaction energy difference between in aqueous phase and in enzyme.[15,16] Although their calculations assumed that the proton is transferred from Ser195 to His57 prior to the nucleophilic attack of Ser195 and that the formation of the tetrahedral intermediate is the rate-determining step, they showed that the electrostatic stabilization of the tetrahedral intermediate is the most important catalytic factor. Daggett et al. employed the semiempirical AM1 and PM3 calculations to obtain the entire reaction path.[17] On the basis of a truncated model system, they concluded that the formation of the first tetrahedral intermediate is the rate-determining step in the entire reaction pathway. Recently, Kollman and co-workers applied a new methodology, called the QM-FE approach, to the tetrahedral intermediate formation step in the acylation reaction.[18] Though their calculations showed good agreement with experiments, their QM model neglected the environmental response of protein to determine the reaction path. Many of those theoretical calculations assumed that the transition state (TS) structure is similar to the tetrahedral intermediate. The activation energy barrier was estimated based on the potential energy (not free energy) difference between the enzyme−substrate (ES) complex and the tetrahedral intermediate. Considering those circumstances, it will be necessary to employ an ab initio MO method, treating the whole protein environment in the aqueous phase to obtain deeper insight to the reaction mechanism.

The purpose of the present study is to calculate the reaction free energy profile based on the realistic protein model, and to discuss the enzyme reaction mechanism from the free energy viewpoint. Computer simulations of enzyme reactions[16,19] using combined quantum mechanical/molecular mechanical (QM/MM) methods have attracted much attention in these days.[20,21] The present method is based on ab initio QM/MM calculations combined with Molecular Dynamics−Free Energy Perturbation (MD−FEP) calculations. The enzyme active site is modeled in atomic detail by QM/MM methodology. Particularly, we focus on the molecular level interaction between the active site and the rest of protein environment along the reaction coordinate. The reaction path is determined in a realistic protein environment, and the activation free energy is estimated by MD−FEP calculations along the minimum energy path for the entire solvated protein system. In the current study, we have only dealt with the first part of the entire reaction process (acylation process), which is considered to be rate-determining.

Computational methods as well as the calculation details are summarized in section 2, and section 3 presents the calculated results. In section 4, we discuss a reliability of the present results comparing with the previous experimental and theoretical results. The conclusions of the present work are summarized in section 5.

(8) Hunkapiller, M. W.; Smallcombe, S. H.; Whitaker, D. R.; Richards, J. H. *Biochemistry* **1973**, *23*, 4723−4743.
(9) (a) Robillard, G.; Shulman, R. G. *J. Mol. Biol.* **1974**, *86*, 519−540. (b) Robillard, G.; Shulman, R. G. *J. Mol. Biol.* **1974**, *86*, 541−558.
(10) (a) Bachovchin, W. W.; Roberts, J. D. *J. Am. Chem. Soc.* **1978**, *100*, 8041−8047. (b) Bachovchin, W. W. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 7948−7951.
(11) (a) Kossiakoff, A. A.; Spencer, S. A. *Nature* **1980**, *288*, 414−416. (b) Kossiakoff, A. A.; Spencer, S. A. *Biochemistry* **1981**, *20*, 6462−6474.
(12) Cleland, W. W.; Kreevoy, M. M. *Science* **1994**, *264*, 1887−1890.
(13) Frey, P. A.; Whitt, S. A.; Tobin, J. B. *Science* **1994**, *264*, 1927−1930.
(14) (a) Umeyama, H.; Imamura, A.; Nagato, C.; Hanano, M. *J. Theor. Biol.* **1973**, *41*, 485−502. (b) Scheiner, S.; Kleier, D. A.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 2606−2610. (c) Dewar, M. J. S.; Storch, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 2225−2229.
(15) (a) Warshel, A.; Russell, S. *J. Am. Chem. Soc.* **1986**, *108*, 6569−6579. (b) Warshel, A.; Naray-Szabo, G.; Sussman, F.; Hwang, J. K. *Biochemistry* **1989**, *28*, 3629−3637.
(16) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiely & Sons: New York, 1991.
(17) Daggett, V.; Schröder, S.; Kollman, P. *J. Am. Chem. Soc.* **1991**, *113*, 8926−8935.
(18) (a) Stanton, R. V.; Peräkylä, M.; Bakowies, D.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 3448−3457. (b) Peräkylä, M.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 3436−3444. (c) Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. *Acc. Chem. Res.* **2001**, *34*, 72−79.
(19) (a) Villà, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 7887−7907. (b) Kollman, P. A.; Kuhn, B.; Peräkylä, M. *J. Phys. Chem. B* **2002**, *106*, 1537−1542.
(20) (a) Monard, G.; Merz, Jr, K. M. *Acc. Chem. Res.* **1999**, *32*, 904−911. (b) Field, M. J. *J. Comput. Chem.* **2002**, *23*, 48−58.
(21) For example, recent works of ab initio QM/MM method for enzyme system are: (a) Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 2293−2301. (b) Amara, P.; Volbeda, A.; Fontecilla-Camps, J. C.; Field, M. J. *J. Am. Chem. Soc.* **1999**, *121*, 4468−4477. (c) Liu, H.; Zhang, Y.; Yang, W. *J. Am. Chem. Soc.* **2000**, *122*, 6560−6570. (d) Cui, Q.; Karplus, M. *J. Am. Chem. Soc.* **2001**, *123*, 2284−2290.

2. Methods and Computational Details

We employed an ab initio HF method to describe the potential energy profiles of bond making/breaking processes in the enzyme reactions. All calculations reported here were based on our developed programs. An ab initio QM/MM program has been developed based on HONDO package.[22] We chose the standard AMBER (parm.96) potential energy function[23] because of its ab initio MO based parametrization nature for the charge model, which seems to be consistent with the ab initio QM/MM framework.

2.1 QM/MM Potential Energy Function. As the details of QM/MM methodology are described in the literature,[24−26] we give a brief description of the methods, pertinent to the present calculations.

The total Hamiltonian of the whole system can be written as follows

$$\hat{H}_{\text{total}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM/MM}} \tag{1}$$

Here, $\hat{H}_{\text{QM/MM}}$ is given by

$$\hat{H}_{\text{QM/MM}} = \hat{H}^{\text{elec}}_{\text{QM/MM}} + \hat{H}^{\text{vdw}}_{\text{QM/MM}} + \hat{H}^{\text{strain}}_{\text{QM/MM}} \tag{2}$$

where $\hat{H}^{\text{elec}}_{\text{QM/MM}}$ describes the electrostatic interaction between the QM and the MM region, and $\hat{H}^{\text{strain}}_{\text{QM/MM}}$ is the strain term at the QM/MM boundary. The last two terms are calculated by the classical force field. $H^{\text{strain}}_{\text{QM/MM}}$ includes the bonded terms at the boundary (bond, bend, torsion, and improper torsion energy terms). In the present calculations, we employed the standard link atom approach, where hydrogen atoms were added to saturate chemical valences in QM region. To validate this approach, frontier bonds were placed away from the reactive center.

To avoid the calculations of the analytic energy gradients with respect to MM nuclear coordinates, we introduced the approximation

$$E^{\text{elec}}_{\text{QM/MM}} = \left\langle \Psi \middle| -\sum_{MM}\sum_{i} \frac{Z_{\text{MM}}}{|\boldsymbol{r}_{\mathbf{i}} - \boldsymbol{R}_{\mathbf{MM}}|} + \sum_{MM}\sum_{QM} \frac{Z_{\text{MM}}}{|\boldsymbol{r}_{\mathbf{i}} - \boldsymbol{R}_{\mathbf{MM}}|} \middle| \Psi \right\rangle$$

$$\simeq \sum_{QM/MM} \frac{Q^{\text{RESP}}_{\text{QM}} Z_{\text{MM}}}{|\boldsymbol{R}_{\mathbf{QM}} - \boldsymbol{R}_{\mathbf{MM}}|} \tag{3}$$

where Ψ is the wave function of QM region, r and R are electronic and (QM or MM) nuclear coordinates, respectively, Z_{QM} and Z_{MM} are the nuclear charges on QM and MM atomic sites, respectively, and $Q^{\text{RESP}}_{\text{QM}}$ are the effective point charges which are assigned to QM atomic sites. By fixing the effective charges on each QM atomic sites, the energy gradients with respect to MM nuclear coordinates were calculated in a classical manner.

The standard ESP fitting procedures[27] are not appropriate in the present case because the sum of the atomic charges in QM region, which is extracted from the total enzyme system, generally has no integral values. We therefore employed the RESP method[28] which uses generalized Lagrangian restraints technique to fix the net atomic charges of the total system. Because of this method, the sum of the charges on the atomic sites in QM and MM regions is exactly zero (or integers).

To optimize the geometry of enzyme system efficiently, we adapted an iterative procedure. QM region was first optimized with the fixed MM coordinates. In this step, the energy gradients with respect to the QM nuclear coordinates were evaluated analytically, and the quasi-Newton method was employed. Next, by fitting the atomic charges in QM region, the MM region geometry was optimized with the QM geometry fixed. These two processes were repeated iteratively until convergence.

2.2 Free Energy Calculations. After determining the reaction path by QM/MM geometry optimizations, we next evaluated free energy difference along the reaction path. The configurational partition function of the total QM/MM system can be written as follows

$$Z = \int e^{-\beta[\text{E}_{\text{QM}}(\boldsymbol{R}_{\text{QM}}) + \text{E}_{\text{QM/MM}}(\boldsymbol{R}_{\text{QM}},\boldsymbol{R}_{\text{MM}}) + \text{E}_{\text{MM}}(\boldsymbol{R}_{\text{MM}})]} \text{d}\boldsymbol{R}_{\mathbf{QM}} \text{d}\boldsymbol{R}_{\mathbf{MM}} \tag{4}$$

To avoid sampling of the QM region configurational space by ab initio calculations, eq 4 was expanded around the minimum energy configurations determined by the QM/MM optimizations as follows

$$F = -k_{\text{B}}T \ln Z$$

$$\simeq G_{\text{QM}}(\boldsymbol{R}^{\text{min}}_{\text{QM}})$$

$$- k_{\text{B}}T \ln \int e^{-\beta\left[\text{E}_{\text{QM/MM}}\left(\boldsymbol{R}^{\text{min}}_{\text{QM}},\boldsymbol{R}_{\text{MM}}\right)+\text{E}_{\text{MM}}(\boldsymbol{R}_{\text{MM}})\right]} \text{d}\boldsymbol{R}_{\mathbf{MM}} \tag{5}$$

where the first term represents the vibrational free energy of QM region, and the second term is the free energy contribution from MM region. The first term was calculated with the harmonic approximation. The force constant matrix for the QM region was obtained by numerical differentiations of analytically evaluated energy gradients. The second term was evaluated by the FEP technique[29] along the minimum energy path determined by QM/MM optimizations. The free energy difference between i and $i+1$ point on the reaction path was calculated by

$$\Delta F_{\text{i}} = -k_{\text{B}}T \ln\langle e^{-\beta\left[\text{E}_{\text{QM/MM}}\left(\boldsymbol{R}^{\text{min,i+1}}_{\text{QM}},\boldsymbol{R}_{\text{MM}}\right) - \text{E}_{\text{QM/MM}}\left(\boldsymbol{R}^{\text{min,i}}_{\text{QM}},\boldsymbol{R}_{\text{MM}}\right)\right]}\rangle_{\text{i}}$$

$$= k_{\text{B}}T \ln\langle e^{\beta\left[\text{E}_{\text{QM/MM}}\left(\boldsymbol{R}^{\text{min,i+1}}_{\text{QM}},\boldsymbol{R}_{\text{MM}}\right) - \text{E}_{\text{QM/MM}}\left(\boldsymbol{R}^{\text{min,i}}_{\text{QM}},\boldsymbol{R}_{\text{MM}}\right)\right]}\rangle_{\text{i+1}} \tag{6}$$

where $R^{\text{min,i}}_{\text{QM}}$ represents the fixed coordinates of QM region at i-th optimized geometry.

2.3 Calculation Details. 2.3.1 Initial Structure Preparation. The initial coordinates of the trypsin-substrate system were obtained from the X-ray crystal structure of trypsin complex with an inhibitor from Bitter Gourd determined to 1.6 Å resolution (PDB code 1MCT). The inhibitor was removed from the original structure, and the substrate was placed at the X-ray coordinates of the inhibitor. The sequence of the substrate was taken from the inhibitor sequence at the active site (Cys3-Pro4-Arg5-Ile6- Trp7-Met8, this serial number comes from the original inhibitor sequence). Hydrogens were added to the complex system assuming the standard geometries. Considering the pH condition in which trypsin works, all polar residues were assumed to be in the ionized form. Most of the crystal waters observed in the X-ray structure were retained except those located far away from the trypsin but near the inhibitor.

To remove unfavorable steric contact, the first energy minimization was performed with the use of steepest descent algorithm until the RMS energy gradient was below 1.0 kcal/mol·Å^{-1}. At this stage, the total charge of the trypsin-substrate complex system, including one crystallographically defined calcium ion, was +8. To avoid highly positive electrostatic effect on the QM region, we added 8 chloride ions to neutralize the total charge. The counterions were placed at the positions of largest electrostatic potentials. All of these points were located outside

(22) Dupuis, M.; Watts, J. D.; Villar, H. O.; Hurst, G. J. B. HONDO ver. 7.0. *QCPE* **1987**, 544.
(23) (a) 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. (b) 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: Norwell, 1997; Vol. 3.
(24) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.
(25) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718−730.
(26) (a) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298−305. (b) Gao, J. In *Reviews in Computational Chemistry*; VCH: New York, 1996; Vol. 7.
(27) (a) Cox, S. R.; Williams, D. E. *J. Comput. Chem.* **1981**, *2*, 304−323. (b) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894−905. (c) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361−373.
(28) (a) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10 269−10 280. (b) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357−1377.
(29) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395−2417.

positively charged residues (Arg, Lys) at the protein surface. The final energy minimization was performed by the conjugate gradient algorithm until the RMS energy gradients was below 0.005 kcal/mol·Å^{-1}.

The resultant molecular system (trypsin, substrate, counterions, and crystal waters) was used for QM/MM calculations.

2.3.2 Ab Initio QM/MM Potential Energy Calculations. Considering the proposed reaction mechanism, as well as the computational limitations inherent in ab initio calculations, the present QM region was limited to the side chains of catalytic triad (His57, Asp102, and Ser195) and the scissile peptide portion of the substrate (peptide bond between Arg5 and Ile6). The boundaries between QM and MM regions ($C_\alpha-C_\beta$ bond of His57 and Asp102, $C_\alpha-NH$, $C_\alpha-CO$ bonds of Ser195, $C_\alpha-C_\beta$, $C_\alpha-NH$ bonds of substrate Arg5, and $C_\alpha-C_\beta$, $C_\alpha-CO$ bonds of substrate Ile6) were saturated by link hydrogen atoms. The van der Waals parameters for the QM region were the same as the standard AMBER force field except two QM atoms. For the atoms to be changed, its atomic types in the product state, we assigned constant van der Waals parameters throughout the calculations (OS and H type vdW parameters for OG and HG of Ser195, respectively). With this partition, the QM region consists of 40 atoms including 8 link atoms and the whole QM/MM system has 3749 atoms (3244 atoms of trypsin, 112 of substrate, 384 of crystal waters, and one calcium ion/8 chloride ions).

In all ab initio QM/MM geometry optimizations, we used the HF method with 6-31G* basis set.[30] Note that it provides a consistent treatment of the electrostatic interaction with MM region atomic charges which are determined by RESP method at HF/6-31G* level. Diffuse functions were added to the carboxylic group of Asp102, OG of Ser195, the scissile peptide portion of the substrate, and p-functions were added to HG of Ser195 and HD of His57 (hereafter, we abbreviate this basis set as 6−31(+)G*(*)). No cutoff was introduced between QM and MM interactions. The convergence criteria used in ab initio QM/MM optimizations were the default of HONDO package (the maximum gradient of 5×10^{-4} au) in QM region and the RMS gradient less than 1×10^{-2} kcal/mol·Å^{-1} in MM region.

The reaction path of the acylation was calculated assuming the following two steps: the first step is the formation of the tetrahedral intermediate (TET) state (step 1 in Figure 1), and the second is the formation of the acylenzyme (EA) product (step 2 in Figure 1). We chose the reaction coordinate for the former reaction step 1 to be a linear combination of two independent bond distances: one is the bond length of OG-HG in Ser195, which describes the proton transfer from Ser195 to His57, and the other is the bond distance between OG of Ser195 and the carbonyl carbon of the scissile peptide, which describes the nucleophilic attack coordinate of Ser195 to the substrate. For the latter (step 2), we adopted a linear combination of two bond distances: one is the scissile bond length of the substrate, and the other is the bond length of NE−HE in His57 describing the proton transfer from His57 to the main chain NH group of the substrate. With these definitions, we determined the minimum energy reaction path for each reaction step in two-dimensional coordinate space. The restrained geometry optimizations were performed in the internal coordinate space with fixing the selected reaction coordinates. At each point along the reaction path, the vibrational frequencies were calculated by numerical differentiation method at the same level (QM/MM HF/6−31(+)G*(*)) as the reaction path optimizations. Vibrational frequencies were used to estimate the free energy contributions of QM region (ZPE, entropy contribution). We further calculated the electron correlation energy at each geometry along the reaction path by single point MP2 calculations with cc-pVDZ basis set.[31] In these calculations, diffuse functions were added to the same atoms, as described above. Hereafter, we denote this basis set as (aug)-cc-pVDZ.

2.3.3 Free Energy Calculations. At first, the reactant structure obtained from QM/MM optimizations was solvated in a sphere of TIP3P[32] water molecules with 30 Å radius centered on the OG atomic site of Ser195. Any water molecules that came within 3.0 Å of protein atoms were removed. This resulted in the enzyme−substrate complex being solvated by 128 crystal waters and 3502 additional water molecules. To prepare a reasonable initial solvated structure, NVT Monte Carlo simulations were carried out at 310 K in standard Metropolis manner.[33] All of the water molecules were moved randomly with the acceptance probability of ∼40% and 50 M configurations were generated to achieve thermal equilibrium. In these MC and next MD calculations, no cutoff of nonbonded interaction was introduced.

Next, this solvated system was equilibrated by relaxing protein structure by MD simulations for 500 ps. Nosé−Hoover-chain (NHC) algorithm[34] was employed to generate NVT ensemble, and the system temperature was maintained at 310 K by attaching five chains of thermostat with thermostat mass corresponding to $\tau = 0.35$ ps. To save the computational cost, we used the reversible reference system propagation algorithm (r-RESPA)[35] extended to non-Hamiltonian NHC system.[36] The long-range nonbonded forces (electrostatic and van der Waals) were integrated in a long time scale (1.5 fs), whereas short-range bonded forces (bond, bend, torsion, and improper term) were integrated in a short time scale (0.3 fs). After the equilibration, we calculated the free energy differences along the reaction path with double-wide sampling technique.[37] At each point along the reaction path from the reactant to the product, the simulations consisted of 10 ps of equilibration run and 60 ps of averaging run. Total 16 increments (windows) were used in MD−FEP calculations for each reaction path (first/second). The free energy differences were calculated in two directions defined by eq 6: $i \rightarrow i + 1$ *forward* perturbation step (from the reactant to the product) and $i \rightarrow i - 1$ *backward* step. To fix the optimized geometry in QM region during MD−FEP calculations, we used the SHAKE/RATTLE constrained MD technique.[38,39] At each time step, the internal coordinates of QM region were exactly retained at the optimized geometries. The relative free energies were calculated as the average of the *forward* and *backward* simulation results.

3. Results

3.1 Optimized Structures of ES Complex and Free Enzyme. The optimized structure around the active site of ES complex is shown in Figure 2, and selected geometric parameters are summarized in Table 1. An apparent hydrogen bond is formed between NE of His57 and OG of Ser195. Two hydrogen bonds are also observed between His57 and Asp102: one is between ND of His57 and OD1 of Asp102, and another is between NH of main chain His57 and OD2 of Asp102. The calculated hydrogen bond length between OD1 of Asp102 and ND of His57 is 2.723 Å, which is slightly longer than the LBHB hypothesis predicted value (2.5 ∼ 2.6 Å).[13] These hydrogen bonds formed among the catalytic triad in ES complex are regarded as a moderate hydrogen bond. The interatomic distance between OG of Ser195 and the carbonyl carbon of the scissile bond is 2.672 Å. This carbonyl carbon is little distorted from

(30) Hehre, W. J.; Radom, L.; van Schleyer, P. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiely & Sons: New York, 1986.
(31) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796−6806.
(32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(33) (a) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987. (b) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; Academic Press: London, 1996.
(34) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635−2643.
(35) Tuckerman, M.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990−2001.
(36) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. *Mol. Phys.* **1996**, *87*, 1117−1157.
(37) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050−3054.
(38) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(39) Anderson, H. C. *J. Comput. Phys.* **1983**, *52*, 24.

the trigonal form: the dihedral angle ω in the scissile peptide is 151.2°, and the improper torsion angle at carbonyl carbon is 168.3°.

To analyze conformational changes induced by the substrate binding, we optimized the free enzyme structure. The selected geometric parameters around the active site are also included in Table 1.

3.2 MD Simulations of the ES Complex and Free Enzyme. To further analyze conformational changes induced by the substrate binding, we performed NVT-MD simulations for 1 ns in aqueous phase. The equilibrium values of bond lengths and angles (r_{eq}, θ_{eq}) in QM region were fixed to QM/MM optimized values. The geometric structures were saved every 100 steps (every 0.15 ps) for analyses.

In the case of ES complex, two stable hydrogen bonds are observed between the side chains of His57 and Asp102 during almost all simulation time. The dihedral angle χ_2 of His57 has relatively constant value ($\sim -90° \pm 10°$) and the rotational orientation of the imidazole ring of His57 is restricted by the stable hydrogen bonds between His57 and Asp102. The weak/moderate hydrogen bond between His57 and Ser195 is observed in ~70% of the total simulation time. The average interatomic distance between OG of Ser195 and the carbonyl carbon of the scissile bond of the substrate is $\sim 3 \pm 0.2$ Å, and the scissile peptide bond remains in planar form during almost all the simulation time.

In the case of free enzyme, the relative conformations between His57 and Asp102 (hydrogen bonds and dihedral angle χ_2 of His57, etc) are very similar to those in ES complex. However, in most of the simulation time, the hydrogen bond between His57 and Ser195 is broken and occasionally weak/moderate hydrogen bonding is formed during a short time (2~3 ps). The present trajectory analyses show that OG of Ser195 forms hydrogen bond with the main chain NH group of Ser195 or Gly193 during most of the simulation time.

Most of chloride counterions moved around the positively charged residues at the protein surface.

3.3 First Step Reaction Path (from ES to TET). The minimum energy reaction path connecting ES to TET is depicted in the 2D-reaction coordinate diagram as shown in Figure 3. The potential energy profile and its components along the minimum energy path are given in Figure 4 (above). The total potential energy has 14.5 kcal/mol energy barrier at the reaction coordinate of 1.1~1.2 Å. This region corresponds to the point where the proton is donated to His57. The free energy profiles along the minimum energy path are shown in Figure 4 (below). The profile of the free energy is similar to that of the potential energy: the energy barrier is observed at the same region on the reaction coordinate. The calculated barrier height is ~17.8 kcal/mol. Though the calculated potential energy profiles by HF and MP2 methods show a similar trend in each case, MP2 correction lowers the barrier height. The free energy calculation results of this process are summarized in Table 2.

The selected geometric parameters of transition state (TS) and TET are given in Table 1. The numerical frequencies calculations show that TS and TET have one and zero imaginary frequency, implying that TET has a stable structure. From Table 1, the hydrogen bond distances between His57 and Asp102 become short with the progress of the reaction, and a short (2.605 Å) and relatively strong hydrogen bond is formed in TET.

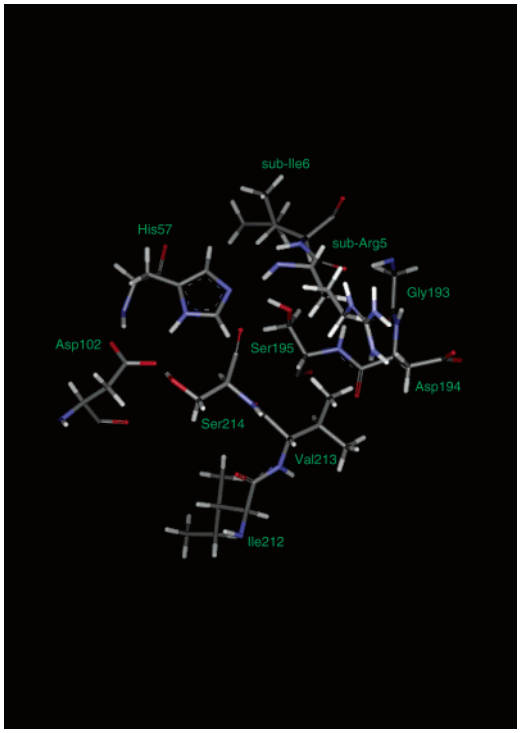

Figure 2. Optimized structure of ES complex (ab initio QM/MM HF/6−31(+)G*(*)/AMBER level). Only selected amino residues around the active site (catalytic triad of His57, Asp102, and Ser195, the scissile portion of the substrate (Arg5, Ile6), catalytically important regions (Gly193~Ser195, the so-called oxyanion hole and Ile212~Ser214) are shown in this figure.

The other hydrogen bonds are formed in the oxyanion hole region. In the ES complex, only one hydrogen bond is formed between the carbonyl oxygen of the scissile bond and the main chain NH group of Gly193. As the reaction proceeds, another hydrogen bond between the main chain NH group of Ser195 is formed and negative charge centered on the carbonyl oxygen is strongly stabilized by two hydrogen bonds (RESP charge of the carbonyl oxygen is −0.6511 in ES complex and −0.8678 in TET). The structural changes are also observed in the scissile bond. With the change of the scissile bond length (from 1.392 to 1.516 Å), the carbonyl bond length becomes slightly longer (from 1.203 to 1.279 Å). The carbonyl carbon of the scissile peptide in TET has a tetrahedral geometry. These structural changes are shown in Figure 5.

3.4 Deformation of Tetrahedral Intermediate (from TET to TET′). To further proceed the acylation, it is necessary for TET to change its conformation so as to be favorable to abstract the proton from the protonated His57. The resultant TET structure is abbreviated as TET′. The main conformational changes of this process are the rotation of main chain NH group around the scissile bond axis. The selected geometric parameters of TET′ are summarized in Table 1. The NH group of main chain in the scissile bond rotates by 128° (the dihedral angle ω in the scissile peptide is 162.7° in TET and −110.6° in TET′) and the atomic distance between NE of His57 and NH of the scissile bond becomes shorter (3.894 Å in TET and 3.453 Å in TET′). In accordance with these changes, the hydrogen bonds between His57 and Asp102 are slightly weakened. The optimized structure is given in Figure 5.

The free energy difference between TET and TET′ was calculated by the FEP calculations.[16,29] The free energy changes

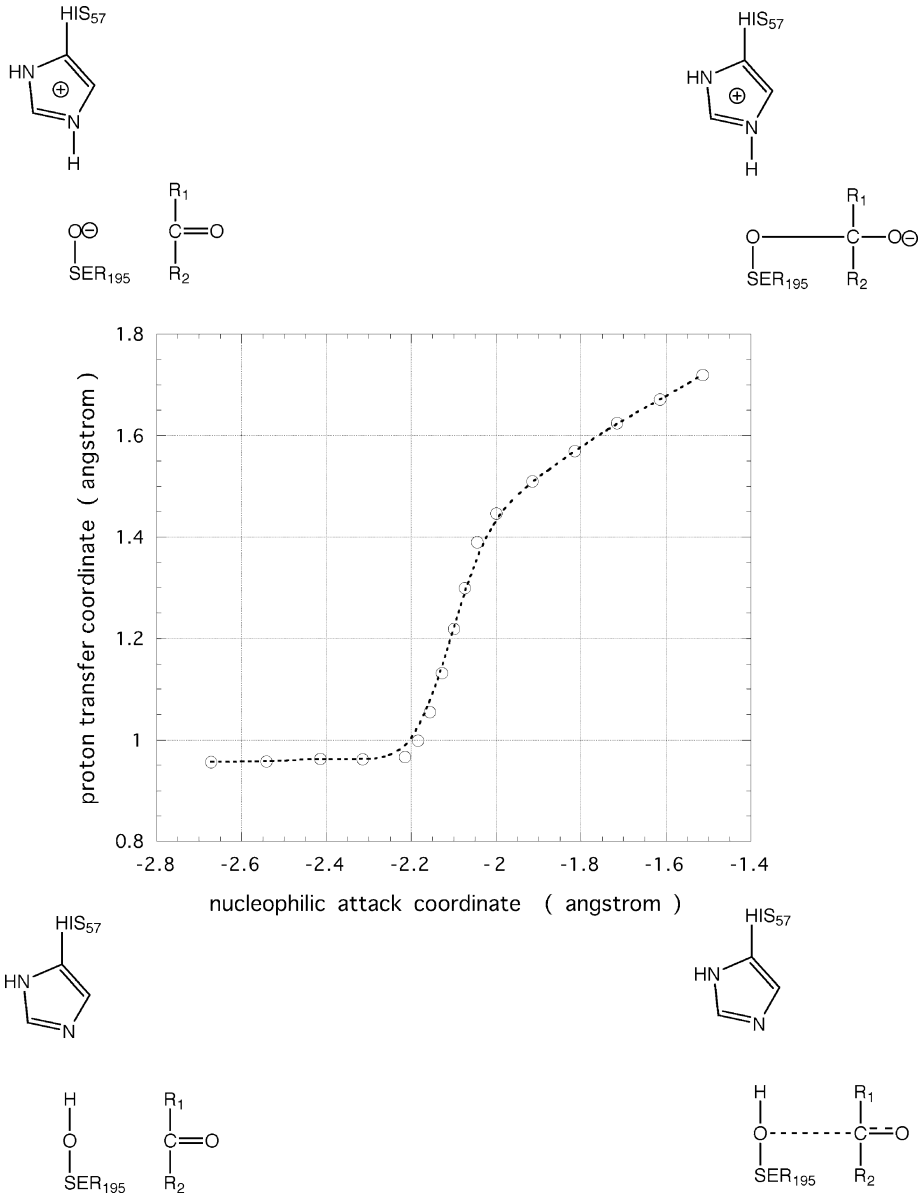

Figure 3. Minimum energy reaction path of the first step reaction (from ES to TET) depicted in 2D-reaction coordinate diagram. The *x*-axis defines the nucleophilic attack coordinate (interatomic distance between OG of Ser195 and the carbonyl carbon of the scissile bond) and the *y*-axis defines the proton-transfer coordinate (bond length of OG-HG in Ser195). The four structures depicted on each square corner represent the asymptotic structures along each reaction direction.

of the potential surface from E_{TET} to $E_{TET'}$ were calculated by gradually changing the potential surface using the relationship

$$E_i(\lambda_i) = \lambda_i E_{TET} + (1 - \lambda_i)E_{TET'} \quad 0 \le \lambda_i \le 1 \qquad (7)$$

We calculated only the free energy difference between two states by gradually changing λ_i with the step size of 0.0625 (16 windows).

The calculated free energies are summarized in Table 2. The main component of these changes is the electrostatic interaction between the rotating NH group and protein environment near the active site. No energy barrier is observed in this process. The total free energy change is exothermic by ~13.3 kcal/mol.

3.5 Second Step Reaction Path (from TET′ to EA). The minimum energy reaction path connecting TET′ to product EA is shown in Figure 6. Because the scissile bond is considered to be almost broken when its distance becomes 2.5 Å, we referred to this point as EA. The potential energies along the minimum energy path are given in Figure 7 (above). The total potential energy has a barrier with 8.1 kcal/mol at the reaction coordinate of ~0.5 Å. At this region, the proton is *in flight* between two nitrogens. The free energies along the minimum energy path are shown in Figure 7 (below). The location of energy barrier is shifted to the reactant side and the barrier height is lowered to ~4 kcal/mol.

The selected geometries of the transition state, denoted as TS′, and EA are included in Table 1. Numerical frequencies calculations showed that TS′ and EA have one and zero imaginary frequency, respectively. From Table 1, hydrogen bonds between His57 and Asp102 become longer with the progress of reaction, and in EA, a moderate strength hydrogen bond observed in ES complex disappears. The other two hydrogen bonds formed in the oxyanion hole still remain stable during the reaction. The acyl-portion of the resultant EA is stabilized by these two moderate strength hydrogen bonds.

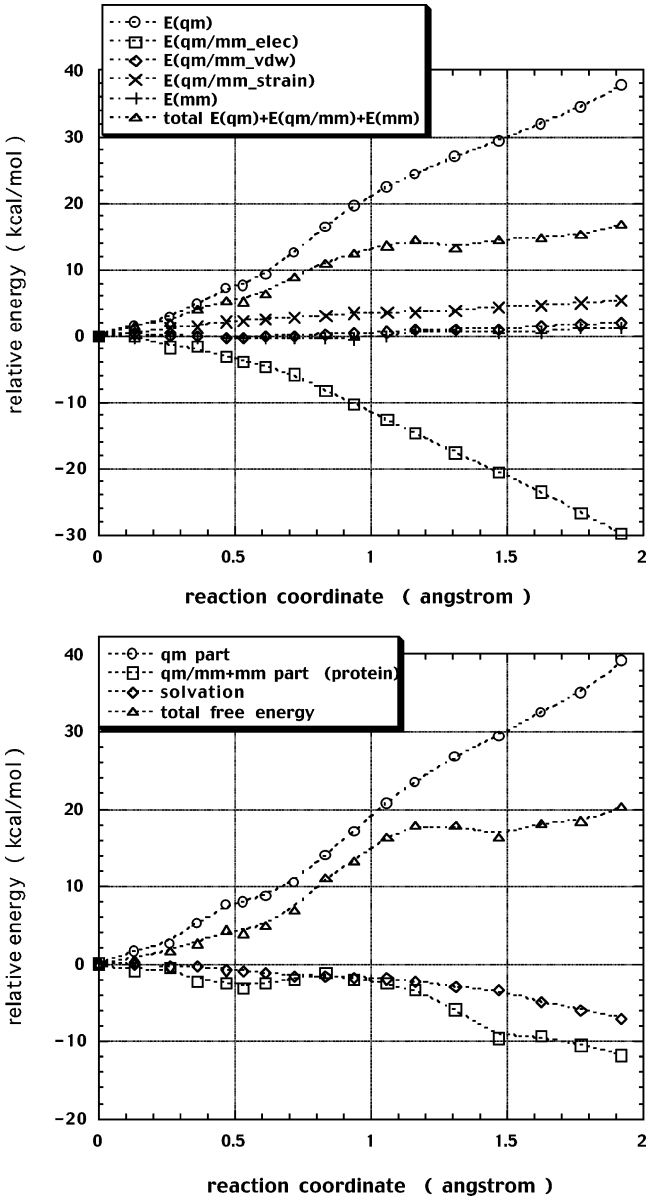

Figure 4. Energy profiles along the minimum energy path in the first step reaction (from ES to TET). (above) the potential energy profiles. QM energy term was calculated at MP2/(aug)-cc-pVDZ//HF/6−31(+)G*(*) level and each energy component was based on the definition of eq 1. (below) the free energy profiles. QM free energy term was calculated at MP2/(aug)-cc-pVDZ//HF/6−31(+)G*(*) level plus ZPE, entropy correction at HF/6−31(+)G*(*) level. These energy components were based on the definition of eq 5. The protein energy term consists of QM/MM interaction energy.

Structural changes are also observed in the carbonyl group in the acyl-portion. As the reaction proceeds, the carbonyl bond length becomes shorter and its length (1.202 Å) in EA becomes quite similar to that observed in ES complex (1.203 Å). The optimized structures of TET′ and EA are given in Figure 5.

3.6 QM/MM Interaction Energy Decomposition Analysis. As seen in Figures 4 and 7, the electrostatic interaction between QM and MM region plays a crucial role to stabilize TET/TET′ throughout the acylation. To analyze the origin of this interaction from a structural viewpoint, we decomposed the QM/MM nonbonded interaction energy into each residue contribution. The structures were based on the QM/MM optimized ones, and the electrostatic interaction was calculated using the assigned RESP charges on each QM atomic site. The contribution from each residue was calculated by

$$\Delta E_{\mathrm{QM/MM}}^{\mathrm{i-residue}} = (E_{\mathrm{QM/MM}}^{\mathrm{i-residue}})_{\mathrm{product}} - (E_{\mathrm{QM/MM}}^{\mathrm{i-residue}})_{\mathrm{reactant}} \quad (8)$$

Figure 8 (above/below) shows the decomposition results in the first/second step reaction, respectively. It is apparent from Figure 8 that only some selected residues have favorable interaction and that most of residues have the opposite trend in the two cases. The interactions favorable to stabilize TET in the first step reaction are unfavorable to stabilize TET′ in the second step.

The selected energy components are summarized in Table 3. Gly193 and Ser214 have unfavorable van der Waals interaction in the first step reaction. These residues form stable hydrogen bonds with the active site in compensation for van der Waals stabilization. Though the total energy component of Ser214 has unfavorable, the electrostatic interaction between the carbonyl part of Ser214, and the imidazole ring of His57 gives favorable contribution.

3.7 Polarization Effect in QM Region Active Site. To validate the hydrophobicity at the active site, we calculated the polarization effect in the active site QM region with ab initio QM/MM framework.[40] In the present model, the polarization effect is determined by the change of QM region wave function under the influence of protein partial charges. The polarization energy can be decomposed into two components

$$E_{\mathrm{pol}} = E_{\mathrm{destab}} + E_{\mathrm{stab}} \quad (9)$$

where E_{destab} is the destabilization energy by the polarization of QM region wave function, and E_{stab} is the electrostatic stabilization energy gained as a result of QM region polarization. These two terms are calculated as follows

$$E_{\mathrm{destab}} = \langle \Psi^{\mathrm{protein}} | \hat{H}_{\mathrm{QM}} | \Psi^{\mathrm{protein}} \rangle - \langle \Psi^{\mathrm{gas}} | \hat{H}_{\mathrm{QM}} | \Psi^{\mathrm{gas}} \rangle \quad (10)$$

$$E_{\mathrm{stab}} = \langle \Psi^{\mathrm{protein}} | \hat{H}_{\mathrm{QM/MM}}^{\mathrm{elec}} | \Psi^{\mathrm{protein}} \rangle - \langle \Psi^{\mathrm{gas}} | \hat{H}_{\mathrm{QM/MM}}^{\mathrm{elec}} | \Psi^{\mathrm{gas}} \rangle \quad (11)$$

where Ψ^{protein} means the wave function determined by QM/MM calculations in the enzyme system, and Ψ^{gas} is that in gas phase. The second term of eq 11 represents the QM/MM electrostatic interaction energy with the permanent (nonperturbed gas phase) charge distribution in QM region. On the basis of this definition, we calculated each energy component along the minimum energy path in the first/second-step reactions and the results are summarized in Table 4. The energetic relations between E_{destab} and E_{stab} show a good agreement with linear response approximation.

3.8 Reference Reaction in Aqueous Phase. In the case of enzyme reactions, it is meaningful to compare the reaction profiles of enzyme reactions with corresponding reference reactions in solution phase. We focus here on the role of protein environment in the enzyme reaction. We therefore calculated the free energy profile for the aqueous phase reaction with the same configurations of reactive fragments along the reaction path obtained for the enzyme reaction. The components of the model reference reaction are the same as the QM region in the trypsin-substrate complex with eliminating Asp102: imidazole

(40) (a) Gao, J.; Xia, X. *Science*, **1992**, *258*, 631−635. (b) Gao, J.; Freindorf, M. *J. Phys. Chem. A* **1997**, *101*, 3182−3188.

Table 1. Selected Geometrical Parameters of ab Initio QM/MM Full-Geometry Optimized Structures[a] along the Minimum Energy Reaction Path (in Ångstrom)

	trypsin	ES	TS (step1)	TET	TET′	TS′ (step2)	EA
OG(Ser195)−HG(Ser195)[b]	0.964	0.957	1.447	1.720	(1.827)	(2.196)	(2.518)
OG(Ser195)−NE(His57)	2.785	2.766	2.529	2.722	2.770	2.883	3.088
NE(His57)−HG(Ser195)[b]	1.871	1.834	1.090	1.020	(1.013)	(1.063)	(2.302)
OD1(Asp102)−ND(His57)	2.777	2.723	2.640	2.605	2.650	2.782	2.834
OD1(Asp102)−HD(His57)	1.760	1.718	1.617	1.568	1.609	1.765	1.820
HD(His57)−-ND(His57)	1.018	1.021	1.037	1.050	1.043	1.026	1.014
OD2(Asp102)−ND(His57)	3.806	3.503	3.506	3.460	3.638	3.940	3.838
OD2(Asp102)−HD(His57)	2.991	2.676	2.674	2.632	2.836	3.200	3.070
OD2(Asp102)−N(His57)	2.687	2.731	2.710	2.713	2.708	2.693	2.729
OG(Ser195)−C(scissile bond)		2.672	2.000	1.513	1.482	1.425	1.315
N−C(scissile peptide bond)		1.392	1.435	1.516	1.512	1.727	2.450
C═O(in scissile bond)		1.203	1.218	1.279	1.286	1.264	1.202
O(scissile peptide)−N(Ser195)		3.841	3.348	3.094	3.119	3.070	3.028
O(scissile peptide)−HN(Ser195)		2.864	2.379	2.135	2.177	2.129	2.100
O(scissile peptide)−N(Gly193)		2.772	2.737	2.705	2.670	2.674	2.706
O(scissile peptide)−HN(Gly193)		1.768	1.723	1.684	1.652	1.660	1.709
OG(Ser214)−OD1(Asp102)	2.520	2.530	2.533	2.538	2.526	2.507	2.514
HG(Ser214)−OD1(Asp102)	1.517	1.527	1.533	1.540	1.529	1.503	1.513
O(carbonyl,Ser214)−CE(His57)	2.954	2.972	2.813	2.785	2.796	2.819	2.961
O(carbonyl,Val213)−CE(His57)	4.026	4.292	3.567	3.520	3.672	4.021	4.114
NE(His57)−NH(scissile bond)[b]		(3.710)	(3.948)	(3.894)	3.453	3.027	3.315
NE(His57)−H(from Ser195)[b]		(1.834)	(1.090)	(1.020)	1.013	1.063	2.302
H(from Ser195)−NH(scissile bond)[b]		(2.840)	(3.136)	(3.088)	2.587	1.983	1.013

[a] All geometry optimizations were performed at QM/MM HF/6−31(+)G*(*) level. [b] Note that HG of Ser195 is donated to His57 in the first step reaction and then donated to NH group in the scissile bond of the substrate.

Table 2. Free Energy Calculation Results in Each Reaction Step (in kcal/mol)

	ES → TET	TET → TET′	TET′ → EA
QM part contribution[a]	39.18	−4.58	−22.39
(E(QM) part)	(52.43)	(−4.03)	(−33.81)
(E_{MP2})	(−14.70)	(−0.04)	(12.42)
(ZPE)	(−0.13)	(−0.40)	(0.18)
(TΔS)	(−1.58)	(0.12)	(1.17)
protein contribution[b]	−11.79 ± 0.92	−15.59 ± 0.23	11.14 ± 1.24
solvent contribution[b]	−7.02 ± 0.01	6.85 ± 0.02	0.99 ± 0.71
total free energy[b]	20.37 ± 0.93	−13.32 ± 0.25	−10.26 ± 1.95

[a] QM part contribution includes the zero point energy and entropy contributions in 310 K. MP2 electron correlation energies were calculated at MP2/(aug)-cc-pVDZ//HF/6−31(+)G*(*) level. ZPE and entropy contributions were calculated by numerical differentiation method at HF/6−31(+)G*(*) level. [b] The error estimates indicate the difference between forward and backward MD−FEP simulation runs.

(as base), ethanol (as nucleophile) and *N*-methyl-acetamide (as peptide or amide). In the solution phase reaction, it seems to be favorable to take a stepwise reaction path, rather than a concerted one, because nature gives rare opportunity to bring three reactive fragments in a solvent cage at the same time. However, our main concern is the environmental effect of protein, and this simple comparison of two free energy profiles (in enzyme/in aqueous phase) reveals the role of environment effect clearly.

The calculation procedures are the same as the enzyme reaction. Three fragments were placed in a sphere of 785 TIP3P waters with 15.0 Å radius centered on the hydroxyl oxygen of ethanol. The free energy calculations along the reaction path (minimum energy path in enzyme reaction) were performed under the same condition as the enzyme reaction.

The calculated free energy profiles in the first-step reaction are shown in Figure 9. It is apparent that no energy barrier appears along the reaction coordinate. The rate-determining step is the formation of the tetrahedral intermediate complex instead of the base-abstraction of the proton. Note that in the present calculations, the free energy factors which align the reactive fragments in a favorable orientation are completely neglected.[18,19] We only focus the activation free energy difference between the reaction in enzyme and in solution (this corresponds to the energy difference between $\Delta g^{\neq}_{cat}$ and hypothetical $\Delta g^{\neq}_{cage}$).

4. Discussion

4.1 Overall Free Energy Profile of the Acylation Process. In this article, we calculated the free energy profiles of the acylation reaction of serine proteases by ab initio QM/MM calculations combined with MD-FEP calculations. On the basis of the calculated results, the overall reaction free energy profile of the acylation process is schematically shown in Figure 10. Below, we discuss the validity of this free energy profile comparing with the experimental results.

In ES complex, it has been a matter of discussion whether there is a hydrogen bond between His57 and Ser195.[41,42] The present simulations show that the substrate binding induces slight conformational changes around the active site, which favor the reactive residues to bring together in a reactive orientation. Also, these favorable conformations occur frequently in ES complex. These results show that the enzyme provides a reactive environment which is preorganized to the reactive residues.

Although the formation of the tetrahedral intermediate is generally assumed, there is no direct experimental evidence. Considering the calculated results, TET′ structure seems to correspond to the putative *tetrahedral intermediate*, and TET structure seems to be an artificial structure, originated from the limited availability of the present QM/MM optimization, which is difficult to sample enough configurational space. When a tetrahedral-like intermediate is formed, the carbonyl carbon of the scissile bond loses its sp2-like character, which is the origin of the stable peptide bond at ES state. Because of the change

(41) Brayer, G. D.; Delbaere, L. T. J.; James, M. N. G. *J. Mol. Biol.* **1979**, *131*, 743−775.
(42) Tsukada, H.; Blow, D. M. *J. Mol. Biol.* **1985**, *184*, 703−711.

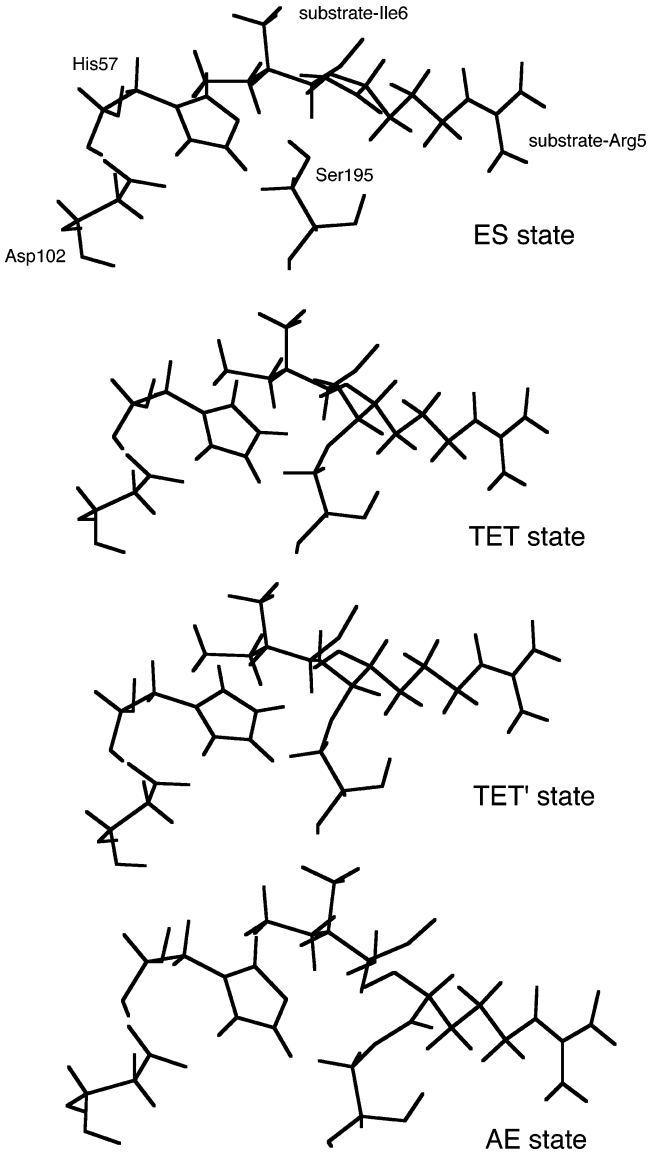

Figure 5. Optimized structures of the active site amino residues along the reaction path. Only a selected residues (which were considered in QM calculation) are shown. These selected four states correspond to reactant, product state in first/second step reaction.

of this electronic character, it becomes easy to rotate the NH group around the scissile bond axis, and the resultant orientation is favorable to accept the proton from His57. Also, the calculated free energy profile of this change is completely exothermic. So it is natural to assume the formation of the tetrahedral intermediate once over the barrier. Therefore, the actual acylation reaction is considered to proceed in the following way: once the interatomic distance between the hydroxyl group of Ser195 and the carbonyl carbon of the substrate becomes short, the proton transfer from Ser195 and the nucleophilic attack of Ser195 proceed synchronously, and when a tetrahedral-like intermediate is formed, the rotation of the scissile peptide bond proceeds in accordance with the reaction. Finally, the tetrahedral intermediate whose orientation is favorable to the next reaction step is formed.

The formation and breakdown of the tetrahedral intermediate has some stereochemical difficulties. When a tetrahedral-like intermediate is formed, the lone pair on nitrogen in the scissile bond is located outside His57, which is stereochemically unfavorable to accept the proton from protonated His57. This stereochemical orientation is more favorable to reprotonate Ser195 and regenerate a substrate than to proceed the acylation process.[43] The His57-ring-flip mechanism was proposed recently to solve this problem.[44] This mechanism states that after the initial formation of the tetrahedral intermediate, protonated His57 rotates by $\sim$180° around the $C_\beta-C_\gamma$ bond, and the resultant hydrogen bond between HD of His57, and the scissile amide is responsible for the following reaction. In the present results, the hydrogen bonds between His57 and Asp102 stabilize the relative orientation of imidazole ring of His57 at ES complex. Also at the tetrahedral intermediate, strong hydrogen bonds are formed among these residues. Because of these nature, it seems less likely to flip His57 along the reaction pathway, which accompanies the disruption and reformation of the hydrogen bond network at the active site in a short time. The present free energy profile could naturally explain these catalytic process, without invoking other proposals such as His57-ring-flip mechanism.

From Figure 10, the rate-determining step is the formation of the tetrahedral intermediate with the activation free energy of $\sim$17.8 kcal/mol. The energy barrier is observed at the proton abstraction step. Though His57 is weak base and Ser195 is weak acid, the nucleophilic attack of Ser195 proceeds without much barrier. This efficiency is due to the electrostatic effect of protein, accompanying with concerted heavy atoms motion in the active site. Although the experimental activation energy depends on substrates and conditions (temperature, pH, etc.), the estimated values are 15$\sim$20 kcal/mol for various amide substrates, based on the rate constants using the transition state theory.[1] The present result is in good agreement with the experimental results. The reaction scheme for the simple enzyme reaction can be written as

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{P}$$

In the present case, based on the assumption of the stationary concentration of the tetrahedral intermediate, the overall acylation rate constant k_2 can be expressed in the following way

$$\mathrm{ES} \underset{k_{-\mathrm{a}}}{\overset{k_\mathrm{a}}{\rightleftharpoons}} \mathrm{TET} \xrightarrow{k_\mathrm{b}} \mathrm{EA} + \mathrm{P}$$

$$k_2 = \frac{k_\mathrm{a} k_\mathrm{b}}{k_{-\mathrm{a}} + k_\mathrm{b}}$$

From Figure 10, k_b is approximately $\sim 10^5$ times larger than $k_{-\mathrm{a}}$, and the rate constant k_2 is approximately equal to k_a. This means that the reaction rate of the formation of the acylenzyme is dominated by the formation of the tetrahedral intermediate. Also, based on the barrier height of the breakdown process ($\sim$ 3.3 kcal/mol), the estimated lifetime of the tetrahedral intermediate is $\sim 10^{-10}$ sec. Therefore, once the tetrahedral intermediate is formed, it is kinetically preferable to breakdown into the acylenzyme instead of going back to the reactant ES complex, without much accumulation of the tetrahedral intermediate. In experiments, this intermediate is supposed to be

(43) Dutler, H.; Bizzozero, S. A. *Acc. Chem. Res.* **1989**, *22*, 322−327.
(44) Ash, E. L.; Sudmeier, J. L.; Day, R. M.; Vincent, M.; Torchilin, E. V.; Haddad, K. C.; Bradshaw, E. M.; Sanford, D. G.; Bachovchin, W. W. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10 371−10 376.

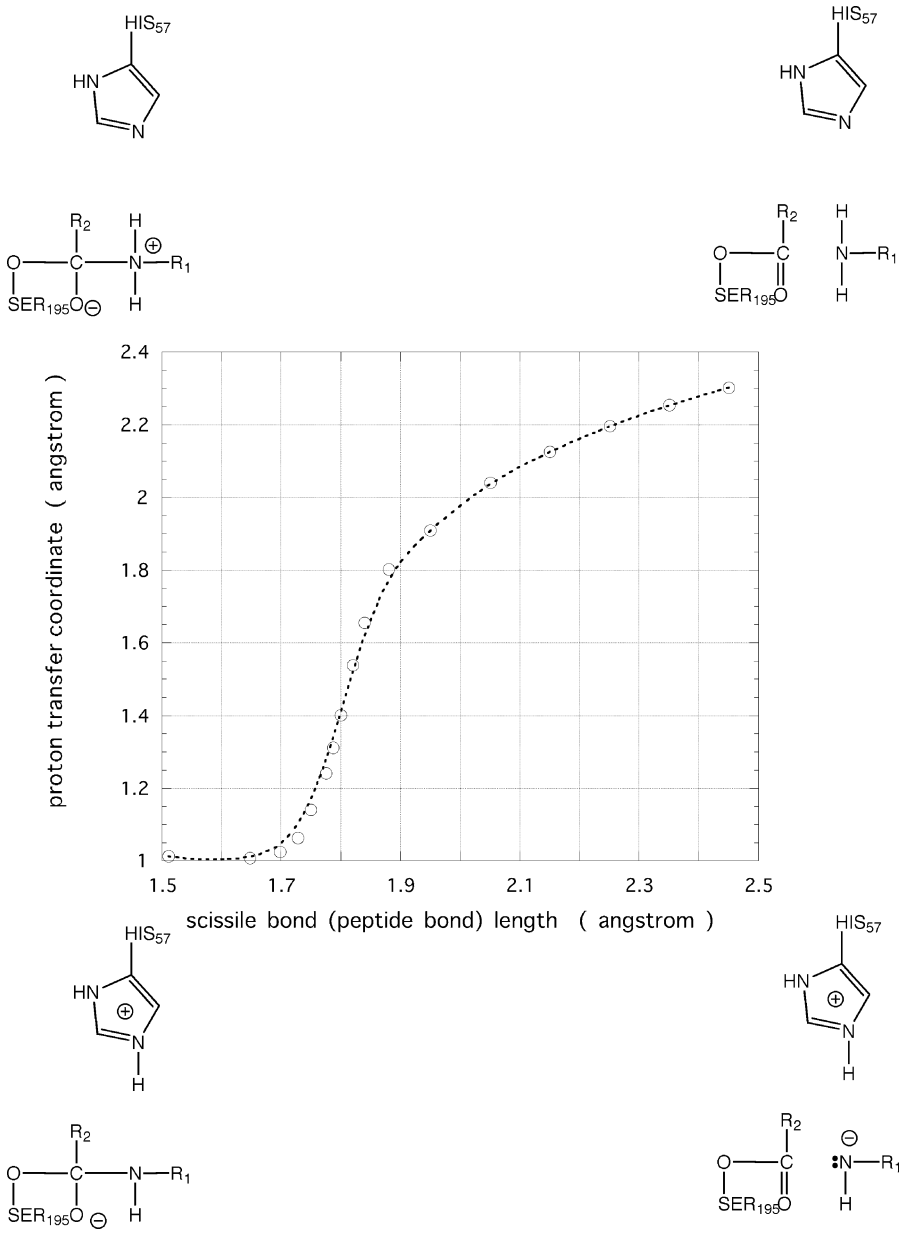

Figure 6. Minimum energy reaction path of the second step reaction (from TET′ to EA) depicted in 2D-reaction coordinate diagram. The *x*-axis defines the scissile bond coordinate (peptide bond length in the substrate) and the *y*-axis defines the proton-transfer coordinate (bond length of NE−HE in His57). The four structures depicted on each square corner represent the asymptotic structures along each reaction direction.

formed but not to be accumulated in the acylation process.[5,6] Also the acylenzyme is relatively stable and can be isolated by using a substrate analogue for which further reaction cannot occur under low-temperature condition. The calculated free energy profile could explain these experimental results qualitatively.

4.2 Factors that Stabilize the Tetrahedral Intermediate. Many previous calculations by other groups were based on an assumption that the transition state structures of the formation/breakdown of the tetrahedral intermediate were very similar to that of the tetrahedral intermediate.[15,17,18] In the present calculations, the transition state for the formation of the tetrahedral intermediate has a structure where the proton is donated to His57, and the bond distance between OG of Ser195 and the carbonyl carbon of the scissile peptide is ∼2.0 Å. This structure is similar to a structure of the tetrahedral intermediate rather than that of ES complex. The transition state for the breakdown of the tetrahedral intermediate has a structure where the proton is still located on His57, and the scissile bond just starts to be longer. This structure is similar to a structure of the tetrahedral intermediate than that of acylenzyme.

From Figures 4 and 7, the protein stabilizes the tetrahedral intermediate more than two transition states. This effect is mainly due to the electrostatic interaction between the particular sequences of amino residues, which are located around the active site. Analyses shown in Figure 8 reveal that catalytically responsible residues are categorized into some groups. Among them, the most important regions are summarized into two groups: one is the so-called oxyanion hole region that consists of mainly the backbone of Gln192∼Ser195, and the other is the region that mainly consists of the main chain carbonyl groups of Ile212∼Ser214, located near the catalytic triad. The former

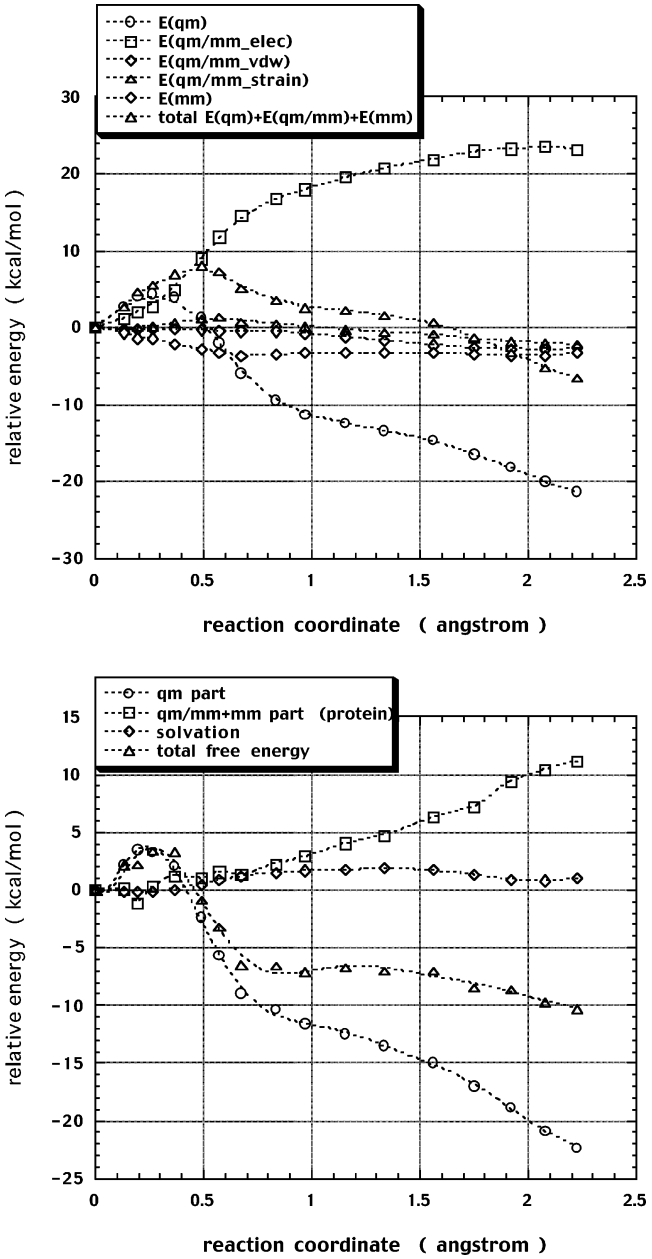

Figure 7. Energy profiles along the minimum energy path in the second step reaction (from TET′ to EA). (above) the potential energy profiles. QM energy term was calculated at MP2/(aug)-cc-pVDZ//HF/6−31(+)G*(*) level and these energy components were based on the definition of eq 1. (below) the free energy profiles. QM free energy term was calculated at MP2/(aug)-cc-pVDZ//HF/6−31(+)G*(*) level plus ZPE, entropy correction at HF/6−31(+)G*(*) level. These energy components were based on the definition of eq 5. The protein energy term consists of QM/MM interaction energy.

region stabilizes negative charge centered at the carbonyl oxygen of the scissile bond by forming two hydrogen bonds with main chain NH groups. The latter stabilizes positive charge created on the imidazole ring of His57 by forming *a kind of hydrogen bond* with main chain carbonyl groups. These hydrogen bonds are formed and broken as the reaction proceeds without large conformational changes of protein. These results imply that the sequences of amino residues around the active site are designed to stabilize the tetrahedral intermediate more than the reactant/product. Although the importance of the oxyanion hole is generally recognized, the role of Ile212∼Ser214 remains elusive at present. On the basis of an extensive comparative study of

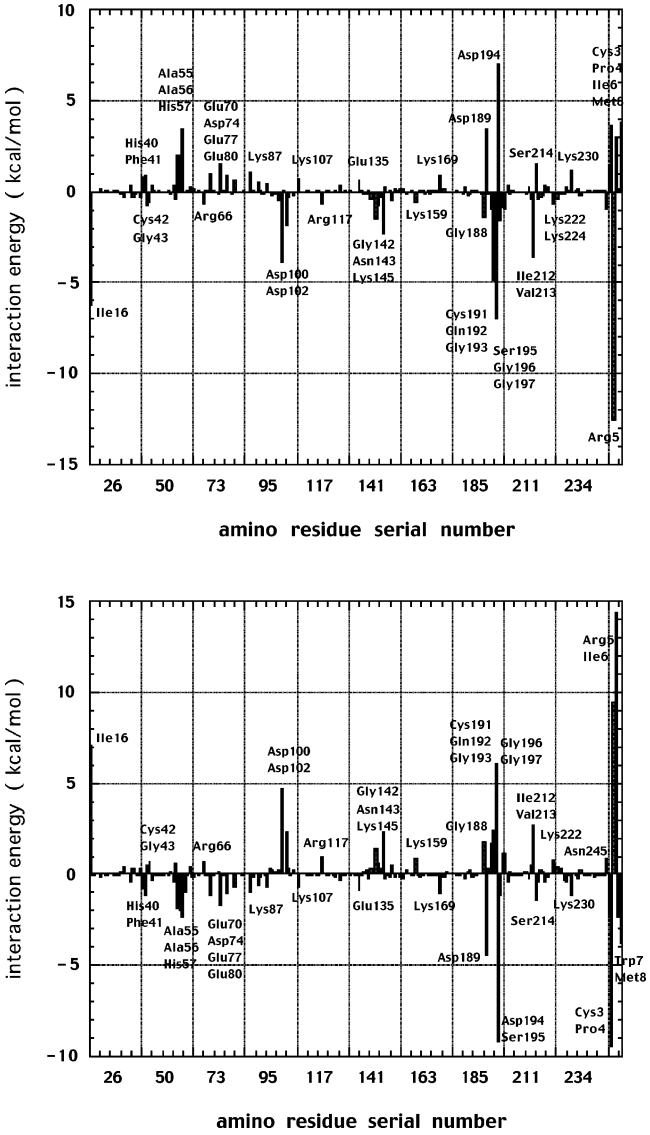

Figure 8. Interaction energy decomposition between QM (active site) and MM (the rest of protein environment) region: (above) decomposition in the first step (ES → TET). (below) decomposition in the second step (TET′ → EA). All contributions of crystal waters and counterions were omitted. Cys3∼Met8 are amino residues of the substrate. The energy component of Asp102, His57, and Ser195 in this Figure were originated from the rest of backbone contributions. The energy component of Arg5 was mainly due to guanidino side chain.

Table 3. QM/MM Interaction Energy Decomposition Results of Selected Amino Residues (in kcal/mol)[a]

	ES → TET		TET′ → EA	
	vdw	charge	vdw	charge
Cys191	0.12	−1.09	−0.27	1.77
Gln192	−1.17	−3.69	−0.15	3.08
Gly193	1.12	−8.08	−0.93	6.38
Ser195	−0.09	−1.44	0.07	−1.26
Gly196	−0.24	−0.62	0.04	2.12
Ile212	−0.04	−0.33	0.02	0.46
Val213	−0.26	−3.34	0.15	2.35
Ser214	1.02	0.51	−0.88	−0.32

[a] Only two important catalytic regions (oxyanion hole and Ile212∼Ser214 region) are shown. See details in text.

serine hydrolases, Derewenda et al. proposed the importance of the hydrogen bond between imidazole ring of His57 and the carbonyl group of Ser214 for facilitating general acid/base

Table 4. QM Region Polarization Energy Components along the Minimum Energy Reaction Path (in kcal/mol)[a]

	ES	TS (step1)	TET	TET′	TS′ (step2)	EA
E_{destab}	13.66	14.40	14.93	13.67	13.74	12.77
E_{stab}	−27.29	−28.77	−29.85	−27.29	−27.43	−25.51
E_{pol}	−13.63	−14.37	−14.93	−13.62	−13.69	−12.74
$\langle\Psi^{gas}\vert\hat{H}^{\mathrm{elec}}_{\mathrm{QM/MM}}\vert\Psi^{gas}\rangle$	−79.67	−92.72	−106.97	−88.28	−85.40	−67.08
$\langle\Psi^{protein}\vert\hat{H}^{\mathrm{elec}}_{\mathrm{QM/MM}}\vert\Psi^{protein}\rangle$	−106.97	−121.49	−136.93	−115.58	−112.84	−92.59

[a] Each energy component is based on the definition of eq 9, 10, 11, and calculated at the QM/MM HF/6−31(+)G*(*)level.

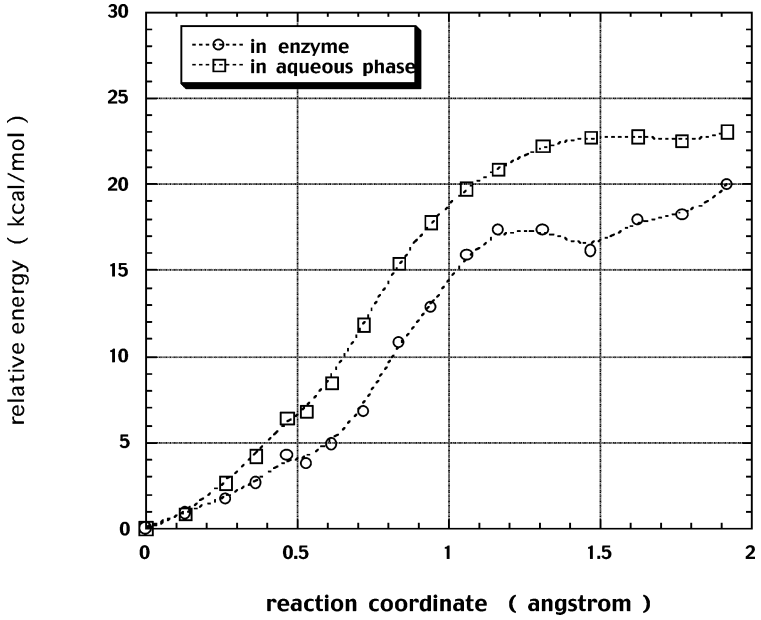

Figure 9. Comparison of the free energy profiles along the minimum energy path in the first step reaction (in enzyme/in aqueous phase). In the hypothetical reference reaction in aqueous phase, QM energy term was the potential energy calculated at MP2/(aug)-cc-pVDZ level (ZPE, entropy correction were not included).

catalysis.[45] The present calculations also support the importance of this type of hydrogen bond for efficient catalysis.

Comparing with the free energy profile of the hypothetical reference reaction in aqueous phase, the stabilization effect for the tetrahedral intermediate is more effective. The reason of this effectiveness is mainly due to the role of Asp102, which stabilizes the imidazole ring of His57 directly by its carboxylic group. Because of these orientations of selected residues, the enzyme could stabilize the tetrahedral intermediate without much energy penalty, or *reorganization energy* to deform its structure along the reaction coordinate. From Figure 5, the conformations around the active site are relatively similar throughout the acylation. Although some movement of both enzyme and substrate are necessary along the catalytic pathway, the principle of least motion for the catalytic residues seems to be a key element of enzyme catalysis.

From the 3D-structural viewpoint, trypsin consists of two domains that are mainly composed of β-barrel structures.[46] Two of the catalytic triad residues (His57 and Asp102) are located at one domain, and the other residue (Ser195) and catalytically important regions (oxyanion hole and Ile212∼Ser214) are in the other. Also the substrate binding occurs at the contact region in two domains. Because the oxyanion hole and the binding specificity residue (Asp189) are located in the same domain

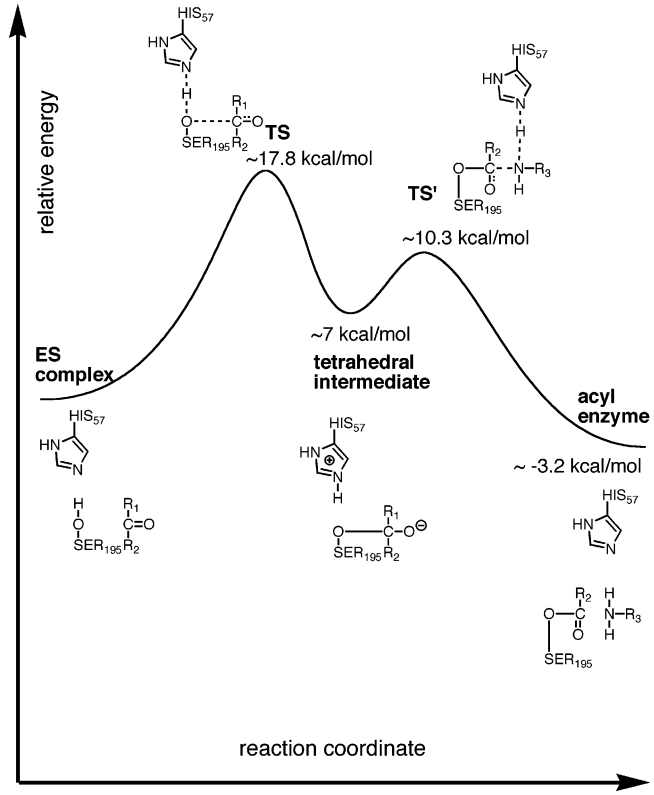

Figure 10. Schematic free energy diagram along the reaction coordinate in the acylation process.

(45) Derewenda, Z. S.; Derewenda, U.; Kobos, P. M. *J. Mol. Biol.* **1994**, *241*, 83−93.
(46) Branden, C.; Tooze, J. *Introduction to Protein Structure*, 2nd ed.; Garland Publishing: New York, 1999.

where Ser195 exists, the important role of the oxyanion hole is apparent: the main role is to bind a substrate tightly and to stabilize the tetrahedral intermediate of a substrate. For the region of Ile212∼Ser214, two important catalytic factors are considered. First is the electrostatic stabilization effect for the tetrahedral intermediate. Second is the binding effect between two domains. As seen in Table 1, strong hydrogen bond between Asp102 and Ser214 is observed along the entire reaction pathway. Also from MD simulations of ES complex, strong hydrogen bond between Asp102 and Ser214 is observed during almost all simulation time (average bond length is ∼2.6 Å). The calculated reaction paths show that the acylation progresses in a concerted manner. To achieve a concerted path, it is important to keep a relative orientation of the catalytically important residues. Because of this binding effect of Ser214, which maintains relative orientation of the catalytic triad, the acylation seems to proceed in a concerted manner, and the resultant tetrahedral intermediate is stabilized effectively.

4.3 Other Catalytic Factors. Many working hypotheses[1−3,16] have been proposed for the enzyme-catalyzed reactions, and among them, major hypotheses for serine proteases are summarized as follows: (1) electrostatic complementarity to the transition state,[47,48] (2) steric strain which destabilizes the ground state,[47] and (3) desolvation effect in nonpolar active site.[48,49] The present results strongly support the first proposal. Below, we discuss other proposals.

As for the ground-state destabilization hypothesis, structural works with transition-state analogues suggest that a distortion of the scissile peptide from planar is a major driving force for catalysis.[50] Also, X-ray structures of several trypsin-BPTI complexes have revealed that there exists extensive contact between protease and inhibitor, and the carbonyl group in the scissile bond is slightly distorted from trigonal.[51] These experimental facts give a part of proof for the formation of the tetrahedral intermediate. However, NMR experiments show that the carbonyl carbon of the scissile bond remains relatively planar in the ES complex and that there is little or no distortion of the scissile peptide.[52,53] In the present MD trajectory calculations of ES complex, the scissile peptide of substrate remains relatively planar during most of the simulation time. From the QM/MM optimized structure of ES complex, the carbonyl carbon of the scissile peptide is slightly distorted from trigonal (improper torsion angle is 168.3°) and the dihedral angle ω in this peptide is 151.2°. To compare the conformational changes with the free substrate, we optimized the same substrate without trypsin by the same procedures for ES complex. From these results, the dihedral angle ω in the free substrate is 178.9°, and the carbonyl carbon remains in just the trigonal form. The energy penalty (elimination of the resonance stabilization) to distort the scissile peptide of the substrate is only 1.3 kcal/mol (QM region energy in the substrate at QM/MM MP2/(aug)-cc-pVDZ//HF/6−31(+)G*(*) level). Though the ground-state destabilization effect is not a negligible factor for the total QM region energy, this factor is not a major catalytic element compared with the electrostatic stabilization effect.

The desolvation hypothesis states that an enzyme catalyzes its reaction by removing solvents and creating a gas-phase-like environment at the active site for the reacting substrates which are usually solvated strongly in aqueous phase. To validate the desolvation hypothesis in the enzyme active site, we calculated the polarization energy contribution in ab initio QM/MM framework. From Table 4, the polarization energy has a similar values (−13 ∼ − 15 kcal/mol) along the minimum energy path though the QM/MM electrostatic energy varies largely in accordance with the acylation. Overall, the polarization energy contributes about 13∼16% to the total QM/MM electrostatic interaction energy during the acylation. These results are similar to the case of small bio-related molecules in aqueous phase.[40] These calculated results show that the enzyme active site is in very polar environment mainly because of the surrounding polar backbone contributions. In contrast to His57 and Ser195 which are located in solvent accessible region, Asp102 is buried in an interior of protein, whose environment is supposed to be nonpolar hydrophobic. However, in the optimized structure of ES complex, Asp102 has three hydrogen bond sites (main chain NH group of His57, ND in imidazole ring of His57, and OG of Ser214), and this environment seems to be polar. From the present results, the assumption that an interior of protein is nonpolar hydrophobic seems to be inappropriate. Although the present results show that the enzyme active site is polar, the proposal that the exclusion of solvents in the active site accelerates chemical reaction is another problem. However, the present calculations show that the desolvation is not a major catalytic factor in serine proteases.

In conclusion, the main catalytic effect in the acylation process of serine proteases is the electrostatic complementarity to the tetrahedral intermediate. The calculated free energy profile shown in Figure 10 is consistent with the assumption that the tetrahedral intermediate is formed but not accumulated in the reaction pathway.

5. Conclusion

In this article, we have studied the reaction mechanism catalyzed by serine proteases (trypsin) by using ab initio QM/MM calculations combined with MD-FEP calculations. The overall reaction profile of the acylation process is found to be exothermic. The calculations show that the tetrahedral intermediate is a relatively stable species (not TS structure) though it is not detected by any direct experimental techniques. There are two transition state structures along the acylation process. These two transition states (formation and breakdown of the tetrahedral intermediate) are similar to the tetrahedral intermediate in its structure. The catalytic role of protein environment is to align the reactive amino residues in a favorable orientation for chemical reaction and to stabilize the tetrahedral intermediate more effectively than the reactant/product. The main component of this catalytic advantage is the electrostatic effect. The oxyanion hole region which mainly consists of the main chain NH groups of Gly193 and Ser195 stabilizes negative charge generated on the carbonyl oxygen in the tetrahedral intermediate. The carbonyl groups in the main chain of Ile212∼Ser214

(47) Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 5250−5254.
(48) Warshel, A.; Florián, J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5950−5955.
(49) Warshel, A.; Åqvist, J.; Creighton, S. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5820−5824.
(50) Delbaere, L. T. J.; Brayer, G. D. *J. Mol. Biol.* **1985**, *183*, 89−103.
(51) (a) Rühlmann, A.; Kukla, D.; Schwager, P.; Bartels, K.; Huber, R. *J. Mol. Biol.* **1973**, *77*, 417−436. (b) Huber, R.; Bode, W. *Acc. Chem. Res.* **1978**, *11*, 114−121.
(52) Baillargeon, M. W.; Laskowski, M., Jr.; Neves, D. E.; Porubcan, M. A.; Santini, R. E.; Markley, J. L. *Biochemistry* **1980**, *19*, 5703−5710.
(53) Richarz, R.; Tschesche, H.; Wüthrich, K. *Biochemistry* **1980**, *19*, 5711−5715.

stabilize positive charge generated on the imidazole ring of His57. These stabilization effects are mainly due to the formation of several hydrogen bonds. The enzyme pays less energy costs, or *reorganization energy*, to stabilize the tetrahedral intermediate by deforming its structure along the reaction coordinate because the catalytic responsible residues are oriented properly around the active site. Also, the present results show that the ground-state destabilization effect (steric strain) has no significant contribution to catalytic activity and that the enzyme active site is in polar environment.

The present calculations show that the rate-determining step is the formation of the tetrahedral intermediate and the breakdown process of this intermediate has small energy barrier. The calculated activation free energy is ~17.8 kcal/mol at QM/MM MP2/(aug)-cc-pVDZ//HF/6−31(+)G*(*)/AMBER level, which is in good agreement with the experimentally estimated values. Because the electron correlation effect plays an important role to determine the activation barrier heights, more elaborate methods such as MP3 and MP4 might be required to obtain more accurate free energy profile.

Acknowledgment. The author (T.Ishida) is grateful to Dr. Akihiro Morita for his valuable comments and continuous encouragement. The numerical calculations were carried out on the Computer Center of the Institute for Molecular Science (IMS) and the Data Processing Center of Kyoto University. This work was supported by the Grant-in-Aid for Scientific Research from the Ministry of Education in Japan.

JA021369M