J|A|C|S
ARTICLES
Published on Web 10/19/2004

Theoretical Studies on the Deacylation Step of Serine Protease Catalysis in the Gas Phase, in Solution, and in Elastase

Maya Topf*,‡ and W. Graham Richards*,†

Contribution from the Physical and Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, OX1 3QZ, United Kingdom, and Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, Mission Bay Genentech Hall, 600 16th Street, Suite N472D, University of California, San Francisco, California 94143-2240

Received May 21, 2004 E-mail: graham.richards@chem.ox.ac.uk; maya@salilab.org

Abstract: The deacylation step of serine protease catalysis is studied using DFT and ab initio QM/MM calculations combined with MD/umbrella sampling calculations. Free energies of the entire reaction are calculated in the gas phase, in a continuum solvent, and in the enzyme elastase. The calculations show that a concerted mechanism in the gas phase is replaced by a stepwise mechanism when solvent effects or an acetate ion are added to the reference system, with the tetrahedral intermediate being a shallow minimum on the free energy surface. In the enzyme, the tetrahedral intermediate is a relatively stable species (~7 kcal/mol lower in energy than the transition state), mainly due to the electrostatic effects of the oxyanion hole and Asp102. It is formed in the first step of the reaction, as a result of a proton transfer from the nucleophilic water to His57 and of an attack of the remaining hydroxyl on the ester carbonyl. This is the rate-determining step of the reaction, which requires ~22 kcal/mol for activation, approximately 5 kcal/mol less than the reference reaction in water. In the second stage of the reaction, only small energy barriers are detected to facilitate the proton transfer from His57 to Ser195 and the breakdown of the tetrahedral intermediate. Those are attributed mainly to a movement of Ser195 and to a rotation of the His57 side chain. During the rotation, the imidazolium ion is stabilized by a strong H-bond with Asp102, and the $C^{\epsilon 1}$−H···O H-bond with Ser214 is replaced by one with Thr213, suggesting that a "ring-flip mechanism" is not necessary as a driving force for the reaction. The movements of His57 and Ser195 are highly correlated with rearrangements of the binding site, suggesting that product release may be implicated in the deacylation process.

1. Introduction

Understanding the origin of enzyme catalysis remains an active problem for biochemists, despite the significant progress in experimental and computational techniques over the last few decades.[1,2] Most of these metabolic processes involve the breaking and forming of covalent bonds, and therefore, when studying them theoretically, one should use high-level quantum mechanical (QM) calculations to obtain reliable activation energies that can be compared with kinetic measurements.[3] This type of calculation is expensive, though, and can only be performed on a small model of the active site that includes the substrate and the catalytic groups of the enzyme, without considering the distant enzyme environment and the biophase description. However, the enzyme can be treated as a whole if a simpler description, such as a molecular mechanical (MM) force field, is used to describe the surroundings. This is the basis of the combined quantum mechanical/molecular mechanical (QM/MM) method[4] that allows the studying of the dynamical aspects of enzyme reactions. The uncatalyzed reference reaction in aqueous solution can also be studied, using the same model of the active site surrounded by explicit or implicit water, and the catalytic effect can then be estimated by comparing the free energy of activation in the water and that in the protein.[5]

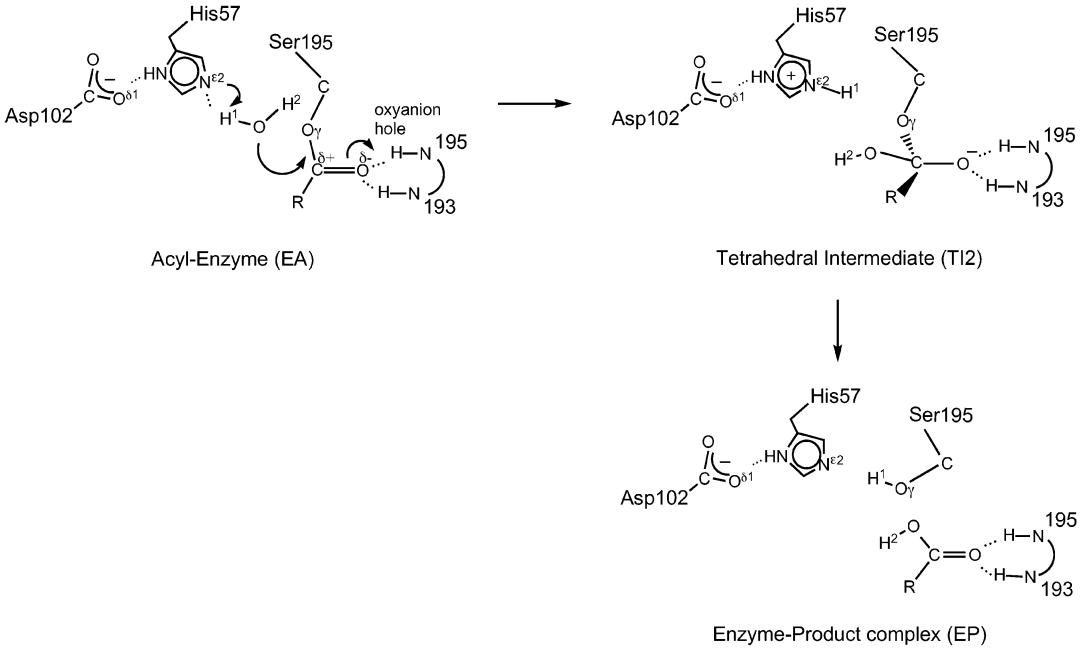
A good case study to examine in order to understand enzyme catalysis is the serine protease (SP) enzyme family, which has been the subject of a large amount of experimental and theoretical work over several decades.[1,3,6] SPs are characterized

‡ University of California, San Francisco.
† University of Oxford.

(1) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, 2nd ed.; Freeman and Company: New York, 1998.
(2) Jencks, W. P. *Catalysis in Chemistry and Enzymology*; Dover Publications: New York, 1987.
(3) Náray-Szabó, G. *J. Mol. Struct. (THEOCHEM)* **2000**, *500*, 157−167.
(4) (a) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249. (b) Gao, J. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. R., Eds.; VCH Publishers: New York, 1996; Vol. 7, pp 119−185.
(5) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, 1st ed.; Wiley: New York, 1991.
(6) (a) Kraut, J. *Annu. Rev. Biochem.* **1977**, *46*, 331−358. (b) Goldblum, A. In *Computational Approaches to Biochemical Reactivity*; Warshel, A., Náray-Szabó, G., Eds.; Kluwer: Amsterdam, 1997; pp 295−340.

10.1021/ja047010a CCC: $27.50

Scheme 1. Schematic Representation of the Deacylation Step in the SP-Catalyzed Reaction, Assuming a Stepwise Mechanism[a]



[a] EA, TI2, and EP denote the acyl−enzyme, the tetrahedral intermediate, and the enzyme−product complex, respectively.

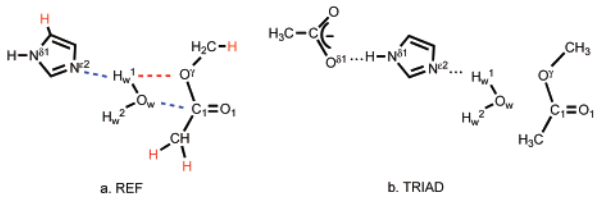
by a catalytic triad of active-site residues consisting of serine, histidine, and aspartate (Ser195, His57, and Asp102, respectively, using chymotrypsin numbering). The most commonly accepted mechanism of SPs is divided into two steps: acylation and deacylation. In the acylation step, formation of an acyl−enzyme (ester) complex occurs sequentially via an initial noncovalent enzyme−substrate complex and a tetrahedral intermediate. In the deacylation step, hydrolysis proceeds via a second tetrahedral intermediate to give a noncovalent enzyme−product complex (Scheme 1). Yet, despite the availability of many experimental data and modeling studies, questions remain about the precise mechanism of serine proteases. For example, the experimental evidence for the formation of the two high-energy tetrahedral intermediates is indirect[7] and is mainly based on analogy of the proposed mechanism with amide hydrolysis in small molecules and on the use of "transition-state analogue" inhibitors.[1] Another example concerns the H-bond network in the active site, especially between His57 (the general base/acid catalyst) and Asp102. Although a double-proton transfer mechanism ("charge−relay system") was accepted for a few years,[8] experimental evidence since then has favored a single-proton transfer.[9] Nevertheless, the nature of the Asp-His H-bond and its role in catalysis is still debatable. Recent NMR studies have suggested that this H-bond is a low-barrier hydrogen bond (LBHB)[10] that can partially explain SP catalysis, although other NMR studies[11] and theoretical calculations[12,13] have argued against this hypothesis.

Different QM/MM methods have been employed in the last two decades in order to study the catalytic mechanism of SP within the surrounding protein.[5,14,15,58] Yet, most of these studies

(7) (a) Fastrez, J.; Fersht, A. R. *Biochemistry* **1973**, *12*, 1067−1074. (b) Fersht, A. R.; Renard, M. *Biochemistry* **1974**, *13*, 1416−1426.
(8) (a) Blow, D. M.; Birktoft, J. J.; Hartley, B. S. *Nature* **1969**, *221*, 337−340. (b) Hunkapiller, M. W.; Smallcombe, S. H.; Whitaker, D. R.; Richards, J. H. *Biochemistry* **1973**, *12*, 4732−4743. (c) Robillard, G.; Shulman, R. G. *J. Mol. Biol.* **1974**, *86*, 541−558.
(9) (a) Kossiakoff, A. A.; Spencer, S. A. *Biochemistry* **1981**, *20*, 6462−6474. (b) Bachovchin, W. W.; Kaiser, R.; Richards, J. H.; Roberts, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 7323−7326.
(10) (a) Cleland, W. W.; Kreevoy, M. M. *Science* **1994**, *264*, 1887−1890. (b) Frey, P. A.; Whitt, S. A.; Tobin, J. B. *Science* **1994**, *264*, 1927−1930.
(11) (a) Ash, E. L.; Sudmeier, J. L.; De Fabo, E. C.; Bachovchin, W. W. *Science* **1997**, *278*, 1128−1132. (b) Bachovchin, W. W. *Magn. Reson. Chem.* **2001**, *39*, S199−S213.
(12) (a) Warshel, A.; Papazyan, A.; Kollman, P. A. *Science* **1995**, *269*, 102−106. (b) Warshel, A.; Papazyan, A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13665−13670. (c) Westler, W. M.; Weinhold, F.; Markley, J. L. *J. Am. Chem. Soc.* **2002**, *124*, 14373−14381. (d) Schutz, C. N.; Warshel, A. *Proteins* **2004**, *55*, 711−723.
(13) Topf, M.; Varnai, P.; Richards, W. G. *J. Am. Chem. Soc.* **2002**, *124*, 14780−14788.
(14) (a) Warshel, A.; Russell, S. *J. Am. Chem. Soc.* **1986**, *108*, 6569−6579. (b) Warshel, A.; Nray-Szab, G.; Sussman, F.; Hwang, J.-K. *Biochemistry* **1989**, *28*, 3629−3637.
(15) (a) Daggett, V.; Schroeder, S.; Kollman, P. *J. Am. Chem. Soc.* **1991**, *113*, 8926−8935. (b) Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. *Acc. Chem. Res.* **2001**, *34*, 72−79.
(16) Ishida, T.; Kato, S. *J. Am. Chem. Soc.* **2003**, *125*, 12035−12048.
(17) Bender, M. L.; Killheffer, J. V. *CRC Crit. Rev. Biochem.* **1973**, *1*, 149−199.
(18) Topf, M.; Varnai, P.; Schofield, C. J.; Richards, W. G. *Proteins* **2002**, *47*, 357−369.
(19) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.
(20) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.3; Gaussian Inc.: Pittsburgh, PA, 1998.
(21) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(22) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(23) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97−179.
(24) Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **1996**, *17*, 49−56.
(25) (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652. (b) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100. (c) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789. (d) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200−1211.

assumed that the transition-state structure is similar to that of the tetrahedral intermediate and calculated the activation energy based on the potential energy surface (PES), rather than the free energies between the enzyme−substrate complex (or the acyl−enzyme) and the transition state. Recently, Ishida et al. used a QM/MM dynamics/free energy perturbation (FEP) method to calculate the free-energy profile of the acylation step.[16] However, the free energies involved in the deacylation step, which is the rate-determining step for certain ester substrates,[17] have never been studied. We previously employed molecular dynamics (MM/MD)[18] and QM/MM dynamics[13] simulations to study different aspects of the deacylation step, but those studies focused on the dynamics of the acyl−enzyme and the tetrahedral intermediate, separately. The purpose of the present work was to calculate the free-energy profile of the entire deacylation process and to discuss the enzyme mechanism from that point of view.

Scheme 2. Two Reference Models for the Active Site of the Acyl−Enzyme[a]



[a] The REF model (a), which represents the reduced model to study the deacylation step. The TRIAD model (b), which contains an additional acetate ion to represent Asp102. The reaction coordinates are shown with dashes; ξ_1 is in blue, and ξ_2 is in red. The hydrogens that are marked in red are the link atoms in the QM/MM study.

Computational Methods are presented in detail, first. Those are followed by Results and Discussion, where the free energies involved in the reference reaction in the gas phase and in solution are described, first, based on DFT calculations. Then, the free-energy profile of the deacylation step in the protein environment is shown, calculated from QM/MM dynamics/umbrella sampling[19] simulations in the enzyme elastase and refined based on the gas-phase calculations. The results are discussed in detail and compared with previous experimental and theoretical results, mainly of studies on the acylation step. The catalytic effect of the deacylation process is evaluated by comparing the free energy of activation in the enzyme and in water. The conclusions of the work are summarized in the final section.

2. Computational Methods

The potential energy surface (PES) of the reference reaction for the deacylation step in the gas phase was calculated using a model (which is referred to below as REF) that included a water molecule, methyl acetate, and an imidazole base. These represent the nucleophilic water in the SP active site, the ester carbonyl between Ser195 and the peptide substrate, and His57, respectively (Scheme 2a). Additional calculations were performed using a model that also included an acetate ion (which is referred to as TRIAD) to evaluate the effect of the third member of the catalytic triad, Asp102 (Scheme 2b). The free energies in water were calculated by adding a continuum solvent to the gas-phase energies. In the enzyme, ab initio QM/MM dynamics simulations were employed in conjunction with the umbrella sampling method[19] to calculate the free energy as a function of the reaction coordinates, which is the potential of mean force (PMF).[51] All gas-phase and solution calculations were performed using the Gaussian 98 package.[20] All QM/MM calculations were performed with the CHARMM package (version 27),[21] interfaced to GAMESS-US.[22] In all of the calculations, H-bonds were defined as contacts with a maximum hydrogen−acceptor distance of 2.5 Å and a minimum donor−hydrogen−acceptor angle of 120°.[23]

2.1. Gas-Phase Calculations. Molecular geometries of stationary points involved in the deacylation reaction in the gas phase were fully optimized using the REF model (Schemes 1 and 2a). The three minimum points along the reaction pathway [the reactants (EA), the tetrahedral intermediate (TI2), and the products (EP)] were initially optimized at the HF/3-21G and HF/6-31G* levels, with a convergence criterion of root-mean-square (RMS) gradient $\leq 3 \times 10^{-4}$ Hartree/Bohr. To locate the two transition states, an adiabatic mapping procedure

(26) (a) Wiest, O.; Black, K. A.; Houk, K. N. *J. Am. Chem. Soc.* **1994**, *116*, 10336−10337. (b) Zhang, Q.; Bell, R.; Truong, N. *J. Phys. Chem.* **1995**, *99*, 592−599. (c) Goldstein, E.; Beno, B.; Houk, K. N. *J. Am. Chem. Soc.* **1996**, *118*, 6036−6043.
(27) Ochterski, J. W. *Thermochemistry in Gaussian*; Gaussian, Inc: Pittsburgh, PA, 2000.
(28) Frisch, M. J. *Gaussian 98 User's Reference*; 2nd ed.; Gaussian, Inc.: Pittsburgh, PA, 1998.
(29) (a) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211−7218. (b) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1983**, *78*, 4066−4073. (c) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899−926.
(30) (a) Fukui, K. *Acc. Chem. Res.* **1981**, *14*, 363−368. (b) Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154−2161.
(31) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486−1493.
(32) (a) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117−129. (b) Cammi, R.; Cossi, M.; Mennucci, B.; Pomelli, C. S.; Tomasi, J. *Int. J. Quantum Chem.* **1996**, *60*, 1165−1178.
(33) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210−3221.
(34) Strajbl, M.; Florian, J.; Warshel, A. *J. Am. Chem. Soc.* **2000**, *122*, 5354−5366.
(35) Wilmouth, R. C.; Clifton, I. J.; Robinson, C. V.; Roach, P. L.; Aplin, R. T.; Westwood, N. J.; Hajdu, J.; Schofield, C. J. *Nat. Struct. Biol.* **1997**, *4*, 456−462.
(36) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(37) (a) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718−730. (b) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.
(38) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298−305.
(39) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580−10594.
(40) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720−1735.
(41) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
(42) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(43) MacKerell, J. A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, I. W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(44) Freindorf, M.; Gao, J. *J. Comput. Chem.* **1996**, *17*, 386−395.
(45) Gao, J.; Freindorf, M. *J. Phys. Chem. A* **1997**, *101*, 3182−3188.
(46) Byun, K.; Mo, Y.; Gao, J. *J. Am. Chem. Soc.* **2001**, *123*, 3974−3979.
(47) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667−683.
(48) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311−1327.
(49) Eurenius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. *Int. J. Quantum Chem.* **1996**, *60*, 1189−1200.
(50) (a) Brooks, C. L.; Brunger, A.; Karplus, M. *Biopolymers* **1985**, *24*, 843−865. (b) Brunger, A. T.; Huber, R.; Karplus, M. *Biochemistry* **1987**, *26*, 5153−5162.
(51) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Adv. Chem. Phys.* **1988**, *71*, 1−249.
(52) Burgi, H. B.; Dunitz, J. D.; Shefter, E. *J. Am. Chem. Soc.* **1973**, *95*, 5065−5067.
(53) Hunkapiller, M. W.; Forgac, M. D.; Richards, J. H. *Biochemistry* **1976**, *15*, 5581−5588.
(54) (a) Bachovchin, W. W.; Roberts, J. D. *J. Am. Chem. Soc.* **1978**, *100*, 8041−8047. (b) Markley, J. L.; Ibanez, I. B. *Biochemistry* **1978**, *17*, 4627−4640.
(55) (a) Wells, J. A.; Cunningham, B. C.; Graycar, T. P.; Estell, D. A. *Philos. Trans. R. Soc. London, Ser. A* **1986**, *A317*, 415−423. (b) Sprang, S.; Standing, T.; Fletterick, R. J.; Stroud, R. M.; Finer-Moore, J.; Xuong, N. H.; Hamlin, R.; Rutter, W. J.; Craik, C. S. *Science* **1987**, *237*, 905−909. (c) Craik, C. S.; Roczniak, S.; Largman, C.; Rutter, W. J. *Science* **1987**, *237*, 909−913. (d) Carter, P.; Wells, J. A. *Nature* **1988**, *332*, 564−568.
(56) Berezin, I. V.; Kazanskaya, N. F.; Klyosov, A. A. *FEBS Lett.* **1971**, *15*, 121−124.
(57) Schowen, R. L. In *Mechanistic Principles of Enzyme Activity*; Liebman, J. F., Greenberg, A., Eds.; VCH: New York, 1988; pp 119−168.
(58) Stanton, R. V.; Perakyla, M.; Bakowies, D.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 3448−3457.

was employed using the Berny algorithm in redundant internal coordinates[24] mode (as implemented in Gaussian 98[20]), starting from the tetrahedral intermediate and driving the system toward the reactants and the products. The reaction coordinate used for locating the first transition state (TS1) was two-dimensional, and it comprised the distance between the imidazolium nitrogen ($N^{\epsilon 2}$) and the hydrogen bound to it (H_w^1) versus the distance between the carbonyl carbon (C_1) and the nucleophilic oxygen of the water molecule (O_w): $\xi_1 = d(N^{\epsilon 2}\cdots H_w^1)$, $d(C_1\cdots O_w)$. The $d(N^{\epsilon 2}\cdots H_w^1)$ ranged between 1.07 and 2.07 Å and $d(C_1\cdots O_w)$ between 1.4 and 2.4 Å, with a step size of 0.1 Å. The reaction coordinate used for locating the second transition state (TS2) was simply the distance between O^{γ} and H_w^1: $\xi_2 = d(O^{\gamma}\cdots H_w^1)$. The latter ranged between 0.95 and 2.25 Å, with a step size of 0.1 Å. Two transition-state optimizations were carried out: one on the geometry of a saddle point that was found on the energy surface of ξ_1, and the other on the maximum point that was found on the energy curve of ξ_2. Because the variation of the energy along the reaction coordinate is very sensitive to the correlation energy, the geometries of the stationary points were reoptimized using the DFT method (at the B3LYP/6-31G*, B3LYP/6-31+G*, and B3LYP/6-31+G**[25] levels), which has been shown to accurately describe the activation energies for different reactions, including proton transfers.[26] Finally, the nature of those points was characterized at each level of theory by performing a vibrational frequency analysis. To account for harmonicity, basis set limitations, and the neglect of electron correlation effects, the frequencies were scaled by 0.8929 (which is the default value at 298.15 K using Gaussian 98[20,28]). These were used to calculate thermodynamic quantities, including zero-point energy (ZPE) corrections, thermal corrections, and Gibbs free energies (G_{gas}).[27] Charges were obtained for the optimized geometries of the stationary points at the B3LYP/6-31+G* level using natural bond orbital (NBO) analysis.[20,29] In addition, to verify that the transition states correspond to the chemical reactions of interest, intrinsic reaction coordinate (IRC) calculations were performed using mass weighted internal coordinates[30] to follow the reaction path.

To calculate the PES of the reference reaction in the TRIAD model (which contained an acetate ion; Scheme 2b), a procedure similar to that employed in the REF model was used but, this time, only at the B3LYP/6-31+G* level.

2.2. Calculations of Solvation Free Energies. The effect of the dielectric environment on the two active site models (REF and TRIAD) was taken into account through the use of a self-consistent reaction field (SCRF) methodology.[31] The solvent was modeled using the polarizable continuum model (PCM),[32] and the solute was placed inside a cavity built with the united atom topological model.[33] The values of the dielectric constants were set to 1.00 for the inside of the solute cavity and 78.4 for the solvent (which is the value at 298.15 K). The free energies in solution (G_{sol}) were calculated by adding the free energy of solvation (ΔG_{solv}) to the gas-phase energies (E_{gas}) of each of the optimized geometries along the reaction coordinate at the B3LYP/6-31+G* level. This assumes that the solute entropy (S) and vibrational enthalpy (H_{vib}) do not modify the shape of the calculated free-energy surfaces in a major way because most of the degrees of freedom of the reacting fragments have similar mobility in the ground and transition states and, therefore, their entropic contributions should cancel.[34]

2.3. Calculations in the Enzyme Environment. QM/MM Dynamics in Conjunction with Umbrella Sampling. The model of the acyl−enzyme was built using the 1.9 Å resolution crystal structure of porcine pancreatic elastase (PPE) complexed with human β-casomorphin-7[35] (BCM7) (PDB entry:[36] 1QIX). The system was set up following the procedure described in our earlier study,[18] and it contained 240 protein residues, 4 substrate residues, and 2800 added TIP3P water molecules[42] (12 053 atoms in total). It was allowed to equilibrate for 50 ps of stochastic boundary molecular dynamics (SBMD) at 300 K[50] (as described in the earlier study, using CHARMM[21]) and was then minimized. The resulting model was used as the foundation for the QM/MM dynamics/umbrella sampling simulations.

In the QM/MM methodology, as described in the literature,[4,37,38] the region described by a QM potential is where the most important electronic changes occur, while the remainder of the enzyme is described by a MM force field. The effective Hamiltonian of the system can be partitioned into quantum and classical components:

$$\hat{H}_{eff} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \quad (1)$$

where $\hat{H}_{QM}$ is the electronic Hamiltonian; $\hat{H}_{MM}$ is the classical Hamiltonian, and $\hat{H}_{QM/MM}$ is the hybrid Hamiltonian. $\hat{H}_{QM/MM}$ includes the interactions between the QM and the MM regions:

$$\hat{H}_{QM/MM} = -\sum_{e,i}\frac{Q_i}{r_{e,i}} + \sum_{m,i}\frac{Z_m Q_i}{r_{m,i}} + \hat{H}_{vdW} \quad (2)$$

where i is summed over all MM partial charges, m over all QM nuclei, and e over all QM electrons. The first term (which is included in the SCF calculation) gives the one-electron interaction between the QM electron density and the MM partial charges; the second is the standard Coulomb interaction between the QM nuclei and the MM charges, and the final term is a van der Waals interaction term between the MM atoms and the QM nuclei. The latter is required as there is no Pauli repulsion between quantum and classical atoms.

In our simulation, the definition of the QM region was based on the REF model described above (Scheme 2a). This included the following 23 atoms: the imidazole ring of His57, fragments of Ser195 and Ile7 (the P_1 residue of the substrate), Wat317 (the crystallographic nucleophilic water, notated here as W), and 4 hydrogen link atoms to satisfy the valence of the QM part.[37] The link atoms, which do not interact with the neutral MM group consisting of the frontier atom and the atoms directly bonded to it,[39] were restrained on the bond between the QM and the MM atoms[40] with a harmonic potential (k = 1000 kcal mol^{-1} Å^{-2}). The choice of this QM region was a compromise between computational efficiency and accuracy. On the basis of many experimental and theoretical works,[1,3,6,9,12,14] including our previous QM/MM study[13] (in which both Asp102 and the oxyanion hole groups were included in the QM region), the effect of the oxyanion hole and Asp102 is mainly electrostatic in nature, and therefore, the inclusion of those groups inside the QM region is not crucial for accurately calculating the reaction barriers.

The QM region was treated at the HF/3-21G level (using GAMESS-US[22]). Calculations at this level of theory have been shown to give satisfactory results for molecular geometries,[41] and in order to validate this, we compared the geometries of the gas-phase stationary points at the HF/3-21G and B3LYP/6-31+G* levels. The rest of the system (12 034 atoms, including 2800 water molecules) was treated with the coupled CHARMM22 potential.[43] The performance of the HF/3-21G//TIP3P potential was found to be in good agreement with the full ab initio results[44] and has been used in condensed-phase simulations.[45,46] The MM electrostatic terms were truncated, with a switch function between 10 and 14 Å, and van der Waals terms with a shift function, with a cutoff distance of 14 Å;[47] the dielectric constant was unity. The bonds within the solvent molecules were fixed using the SHAKE algorithm.[48] Later, a correction based on the gas-phase PES at the B3LYP/6-31+G* level was included in the generation of the PMF reaction.

The umbrella sampling calculations were performed using reaction coordinates similar to those used in the gas phase, with the assumption that the reaction in the enzyme follows the same mechanism, including a nucleophilic attack of a water on the ester carbonyl and a proton transfer to Ser195 assisted by the imidazole base (Schemes 1 and 2a). Thus, consistent with a two-step reaction, ξ_2 was defined as $d(O^{\gamma}\cdots H_w^1)$ and ξ_1 was defined as $d(N^{\epsilon 2}\cdots H_w^1) + d(C_1\cdots O_w)$ (i.e., in one dimension, to save computational time and to avoid complication

in the generation of the PMF). Adequate sampling of the entire reaction coordinate was ensured by dividing it into a series of overlapping "windows", where each successive window was centered at points 0.10 Å apart on the reaction coordinate. A harmonic restraining function, $V_i(\xi)$, was employed to keep ξ_1 and ξ_2 within 0.15 Å of the center of each window, using the RESD mode[49] in CHARMM (version 27),[21] with a force constant of 500 kcal mol^{-1} Å^{-2}. The simulation was divided into two parts. The first part included 18 windows along ξ_1, starting from EA ($4.2 > \xi_1 > 2.5$ Å), and the second part included 15 windows along ξ_2, starting from TI2 ($2.3 > \xi_2 > 0.9$ Å). The conformational space within each window was sampled using SBMD at 300 K, with an integration time step of 1 fs and frictional coefficients of 62 and 200 ps^{-1} for water oxygens and heavy atoms, respectively, in the protein buffer region.[50] Different equilibration and sampling times for the generation of a stable free-energy curve were tested to ensure enough sampling of conformations with low probability. It was found that for a force constant of 500 kcal mol^{-1} Å^{-2}, 1 ps of equilibration and 4 ps of productive sampling would not change the distribution of the energy within each window. The values of ξ_1 and ξ_2 and the protein coordinates were recorded for analysis every 1 fs.

Generation of the PMF. The free energy as a function of the reaction coordinate was calculated for each umbrella sampling window according to the following formula:[51]

$$W_i(\xi) = -k_B T \ln \rho_i^*(\xi) - V_i(\xi) + C_i \qquad (3)$$

where $V_i(\xi)$ is the harmonic restraining function; C_i is a constant $\{-k_B T \ln\langle e^{(-v_i(\xi)/k_B T)}\rangle\}$, and $\rho_i^*(\xi)$ is the biased probability distribution. The $\rho_i^*(\xi)$ value was calculated from the values of the ξ_1 and ξ_2, distributed into 0.001 Å wide "bins". For each reaction coordinate, an approximation of the PMF, $W_{total}(\xi)$, was obtained by fitting together each $[W_i(\xi) + C_i]$ of the overlapping windows. A splicing algorithm was used to shift the free-energy curve on the energy axis to join its neighboring curve until they all joined smoothly. Each shift was calculated as the average energy difference between every two curves in the part of the reaction where they overlapped, weighted by the number of samples in both windows. This procedure automatically removed any regions of poor sampling in the overlapping regions. Following that, an additional correction from a higher level of theory was added to the PMF. The correction was developed in an approach similar to the one presented in a recent QM/MM study.[46] The PMF, using a QM/MM potential, can be separated into a solvent-independent (gas phase) term and a solute−solvent interaction component

$$W_{total}(\text{prot}) = G_{QM}(\text{gas}) + \Delta G_{Xs} + W_{MM} \qquad (4)$$

where $G_{QM}(\text{gas})$ is the gas-phase free energy of the solute (the reaction region); W_{MM} is the PMF of the solvent (the rest of the protein), and ΔG_{Xs} is the QM/MM interaction free energy. This separation of energy terms allows the use of a high-level (HL) method to determine the intrinsic energy of the solute molecule in the gas phase, while the time-consuming statistical simulations can be carried out with the low-level (LL) quantum chemical model for the evaluation of the solvation energy. Assuming $\Delta G_{Xs}^{LL} \sim \Delta G_{Xs}^{HL}$, that is, the QM/MM interaction free energy is similar at high- and low-levels of theory, it follows from eq 4 that

$$W_{total}^{HL}(\text{prot}) = W_{total}^{LL}(\text{prot}) + G_{QM}^{HL}(\text{gas}) - G_{QM}^{LL}(\text{gas}) \qquad (5)$$

where $G_{QM}^{HL}(\text{gas})$ and $G_{QM}^{LL}(\text{gas})$ are the gas-phase free energies at a high (B3LYP/6-31+G*) and a low (HF/3-21G) QM level; $W_{total}^{LL}(\text{prot})$ is the PMF calculated from the umbrella sampling simulation at the HF/3-21G//CHARMM level, and $W_{total}^{HL}(\text{prot})$ is the PMF, including the correction from the gas-phase free energies.

In the calculation of $G_{QM}^{HL}(\text{gas})$ and $G_{QM}^{LL}(\text{gas})$, the difference between the zero-point energy (ZPE) corrections and entropic contributions at the ground state and at the transition state was found to be similar at both levels of theory; that is, those corrections at the B3LYP/6-31+G* level and the HF/3-21G level cancel each other out, with an error on the order of ∼1−1.5 kcal/mol. Thus, the ZPE corrections and entropic contributions were neglected from eq 5, and the final PMF was calculated as follows:

$$W_{total}^{HL}(\text{prot}) = W_{total}^{LL}(\text{prot}) + E_{QM}^{HL}(\text{gas}) - E_{QM}^{LL}(\text{gas}) \qquad (6)$$

where $E_{QM}^{LL}(\text{gas})$ and $E_{QM}^{HL}(\text{gas})$ are points on the PES of the reference reaction in the gas phase, obtained from the IRC calculation at the HF/3-21G and B3LYP/6-31+G* levels, respectively, that correspond to the values of the reaction coordinates, ξ_1 and ξ_2, in the QM/MM simulation.

Table 1. Relative Free Energies (G_{gas} in kcal/mol) of the Stationary Points along the Reference Reaction for the Deacylation Step in the Gas Phase, Calculated Using the REF and the TRIAD Models (Scheme 2)

	EA	TS1	TI2	TS2 (TS′)[a]	EP
REF					
HF/3-21G	0.0	32.5	31.7	41.5	−3.7
HF/6-31G*	0.0	57.6	57.5	62.3	0.1
B3LYP/6-31G*	0.0			47.9[a]	−3.3
B3LYP/6-31+G*	0.0			49.4[a]	−0.7
B3LYP/6-31+G**	0.0			50.5[a]	−0.5
TRIAD					
B3LYP/6-31+G*	0.0	25.9	26.4	27.9	−6.0

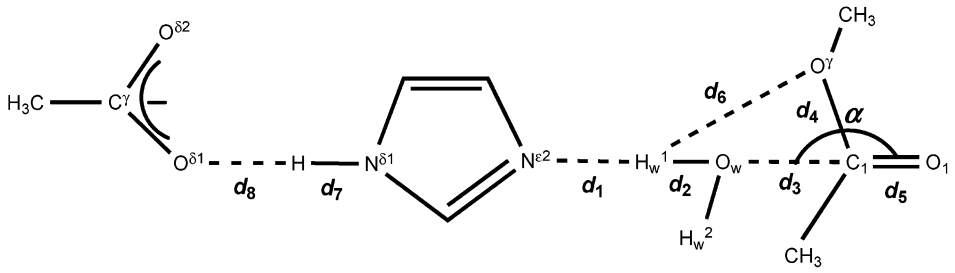
[a] Only one transition state (TS′) is found on the PES at a location similar to TS2 at the HF/6-31G* level.

3. Results and Discussion

3.1. Reference Reaction for the Deacylation Step in the Gas Phase. The free-energy profiles of the reference reaction for the deacylation step in the gas phase are summarized in Table 1. The results indicate that when using the HF method (with both 3-21G and 6-31G* basis sets) in the REF model (Scheme 2a), the reaction follows a stepwise mechanism, where the first transition state (TS1) is energetically and structurally very close to the tetrahedral intermediate (TI2): TS1 is at 57.6 kcal/mol at the HF/6-31G* level, and TI2 is only 0.1 kcal/mol lower (which is within the range of error). The value of ξ_1 [$d(N^{\epsilon 2}\cdots H_w{}^1)$, $d(C_1\cdots O_w)$] is [1.10, 1.58 Å] in TS1 and [1.07, 1.54 Å] in TI2. The value of ξ_2 [$d(O^{\gamma}\cdots H_w{}^1)$] in the second transition state (TS2) is 1.80 Å at the same level of theory. However, when correlation effects are partially accounted for by using the DFT method, the activation free energy decreases by 9.7 kcal/mol (57.6 kcal/mol at the HF/6-31G* level vs 47.9 kcal/mol at the B3LYP/6-31G* level), and an intermediate cannot be stabilized on the PES. Only a single transition state (TS′) is found, with values of ξ_1 and ξ_2 similar to those found in TS1 and TS2, respectively, using the HF/6-31G* level (ξ_1 = [1.07, 1.51 Å], ξ_2 = 1.80 Å). This shallowness of the surface can be explained by the lack of electrostatic stabilization of the negatively charged tetrahedral structure and the positively charged imidazolium ring, both of which are stabilized by only one H-bond (between O_w and $N^{\epsilon 2}-H_w{}^1$).

While the inclusion of the correlation energy causes a large decrease in the activation energy, adding a set of s and p diffuse functions on heavy atoms (B3LYP/6-31+G* level) only causes a small increase in the barrier (1.5 kcal/mol). The effect of adding a set of p polarization functions on the hydrogens (B3LYP/6-31+G** level) is even smaller (1.1 kcal/mol).

Table 2. Selected Geometrical Parameters of the Stationary Points along the Reference Reaction for the Deacylation Step in the Gas Phase



	α (deg)[a]	d_1 (Å)	d_2	d_3	d_4	d_5	d_6	d_7	d_8
				REF Model[b]					
HF/3-21G									
EA	129	1.85	0.98	3.89	1.37	1.20	3.97	0.99	
TS2	113	1.05	1.80	1.50	1.58	1.26	1.65	1.00	
EP	122	1.73	2.87	1.34	3.38	1.20	0.99	0.99	
B3LYP/6-31+G*									
EA	125	1.93	0.99	4.06	1.37	1.21	4.15	1.01	
TS′	114	1.07	1.83	1.51	1.68	1.25	1.80	1.01	
EP	121	1.82	2.95	1.35	3.56	1.21	1.00	1.01	
				TRIAD Model[c]					
B3LYP/6-31+G*									
EA	132	1.83	1.00	4.52	1.35	1.22	4.49	1.06	1.66
TS1	107	1.10	1.51	1.86	1.43	1.25	2.83	1.60	1.06
TI2	110	1.05	1.68	1.59	1.49	1.27	2.59	1.60	1.05
TS2	112	1.03	1.98	1.52	1.55	1.27	2.05	1.61	1.04
EP	121	1.68	3.29	1.34	3.46	1.22	1.02	1.08	1.58

[a] α is the $O_w\cdots C_1{=}O_1$ angle. [b] Without the acetate ion. [c] Including the acetate ion.

Moving from the REF model to the TRIAD model (Scheme 2b), that is, adding an acetate ion to the system, has a 2-fold effect on the reaction mechanism: stabilizing a tetrahedral intermediate and lowering the transition state by 23.5 kcal/mol (from 49.4 kcal/mol in TS′ of REF to 25.9 kcal/mol in TS1 of TRIAD at the B3LYP/6-31+G* level). Note that after the inclusion of ZPE and thermal corrections, the relative free energy of TI2 becomes slightly higher than that of TS1 (26.4 and 25.9 kcal/mol, respectively) due to the shallowness of the surface (but within the range of error).

Examination of the geometries and charges along the reaction coordinate in the REF model (Table 2, Figure 1) shows that during the attack of the nucleophilic water on ester carbonyl, the $O_w\cdots H_w{}^1$ distance (d_2) elongates from 0.99 to 1.83 Å and a change in the charge distribution in the system occurs. The negative charge on O_w decreases from $-0.99e$ in EA to $-0.85e$ in TS′, while the negative charges on O_1 and O^γ, which are covalently bound to C_1, increase from -0.62 and $-0.56e$ in EA to -0.75 and $-0.77e$ in TS′, respectively. As a result, the $C_1{=}O_1$ bond (d_5) is elongated from 1.21 to 1.25 Å. The "angle of attack", $O_w\cdots C_1{=}O_1$ (α), is 114°, which is slightly larger than the ideal angle (107°) for a nucleophilic addition to a carbonyl group (Burgi angle[52]). This can be explained by the lack of an intermediate in the reaction; to enable the break of the $C_1{-}O^\gamma$ bond, a proton has to transfer directly from the nucleophilic water to O^γ (rather than via His57), and this can only be facilitated by increasing the value of α. The $O^\gamma\cdots H_w{}^1$ distance (d_6) is reduced from 4.15 to 1.80 Å, reaching a value similar to d_2, so that at the transition state (TS′), when the proton is almost transferred to $N^{\epsilon 2}$ ($d_1 = 1.07$ Å), it is located at the midpoint between its donor (O_w) and its acceptor (O^γ).

The geometries of the TRIAD model are slightly different from those of the REF model (Table 2). The charge on O_w is still around $-1.0e$ at the first transition state (TS1); α is 107°, and $C_1\cdots O_w$ is far from a single C−O bond ($d_3 = 1.86$ Å), indicating that the proton is fully transferred to $N^{\epsilon 2}$ and a tetrahedral intermediate is stabilized on the PES. This process is accompanied by a proton transfer from $N^{\delta 1}$ to the acetate ion (d_7 changes from 1.06 to 1.60 Å and d_8 from 1.66 to 1.06 Å) and is referred to as the "double-proton transfer" mechanism,[53] which, as discussed above, is no longer regarded as the correct description of the SP mechanism.[1,3,9] However, in the gas phase where ion pairs are not stable, this mechanism is likely to occur, and indeed, by using the same geometries but with the proton located on $N^{\delta 1}$, we calculated the AcIm pair to be ∼4.5 kcal/mol more stable than the Ac^-Im^+ pair at TS1. The tetrahedral structure, on the other hand, is still poorly stabilized in the TRIAD model (only by the $N^{\epsilon 2}{-}H_w{}^1\cdots O_w$ H-bond), as in the REF model.

3.2. Reference Reaction for the Deacylation Step in Aqueous Solution. The free energies involved in the reference reaction to the deacylation step in aqueous solution, calculated at the B3LYP/6-31+G* level, based on the geometries of the stationary points optimized in the gas phase, are presented in Table 3. The results demonstrate how the inclusion of solvent effects in the REF model causes a decrease in the activation barrier (49.4 kcal/mol in the gas phase and 27.7 kcal/mol in solution), due to a better stabilization of the positively charged imidazolium and the negatively charged tetrahedral structure. A free-energy barrier of 27.7 kcal/mol for the deacylation process is slightly above the barrier calculated for the acylation process in a recent theoretical study,[34] suggesting that the overall

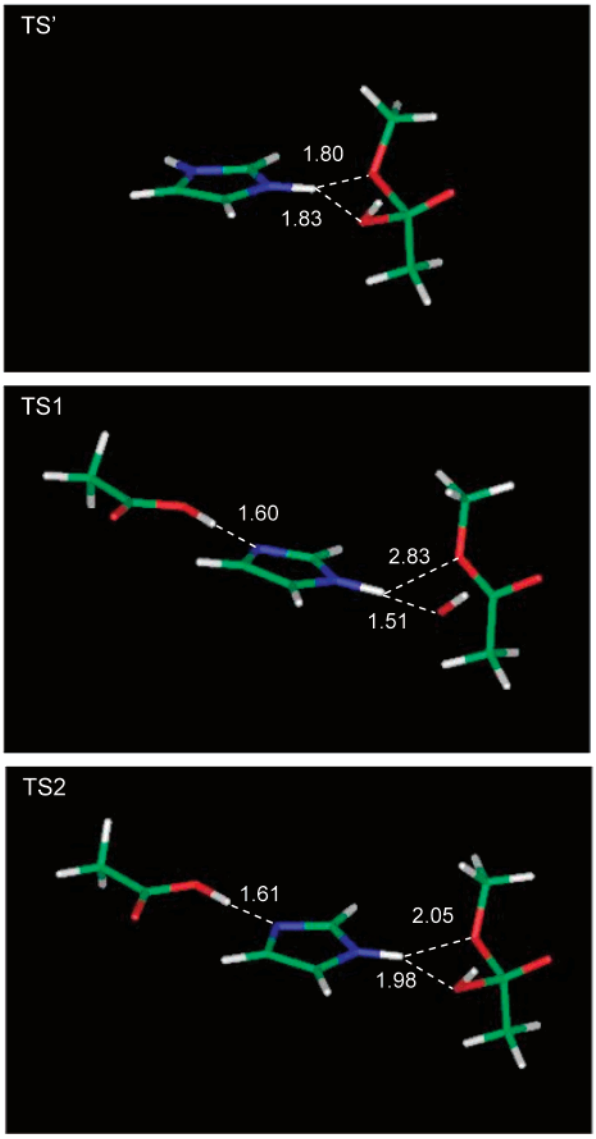


Figure 1. Transition states of the nucleophilic attack on the ester carbonyl optimized at the B3LYP/6-31+G* level. (a) The REF (TS′) model. (b) The TRIAD (TS1 and TS2) model. Distances are indicated in angstroms (Å). The figure was created with the iMol molecular graphics program (http://www.pirx.com/iMol/).

Table 3. Relative Free Energies in Aqueous Solution (G_{sol} in kcal/mol) of the Reference Reaction for the Deacylation Step, Calculated Using the REF, TRIAD, and TRIAD1 Models at the B3LYP/6-31+G* Level

model	EA	TS1	TI2	TS2 (TS′)	EP
REF	0.0	26.9[a]	25.5[a]	27.7	−2.9
TRIAD	0.0	19.6	14.9	15.8	−6.5
TRIAD1[b]	0.0	24.0	21.7	23.6	−6.5

[a] These energies were calculated from the geometries of the corresponding stationary points in the TRIAD model, with the acetate ion removed. [b] Same model as TRIAD but with the $H^{\delta 1}$ proton placed on $N^{\delta 1}$.

rate-determining step of the SP reference reaction in aqueous solution is the deacylation step. However, these numbers are not entirely accurate because the free energies in solution were calculated from the gas-phase geometries, and therefore, the reaction pathway is concerted. To test the possibility of a stepwise reaction, the relative free energies of TS1 and TI2 in solution were calculated using the gas-phase geometries of the

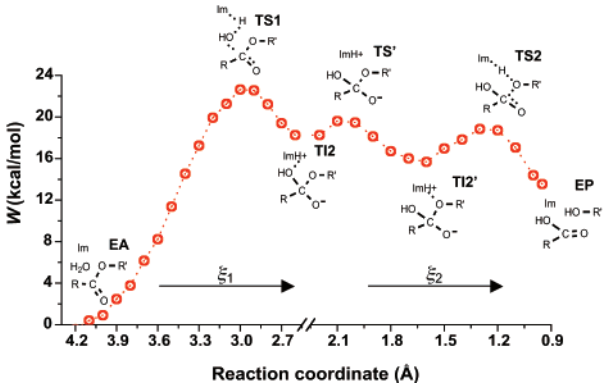


Figure 2. PMF of the deacylation step in the PPE−BCM7 complex, calculated from the QM/MM dynamics/umbrella sampling simulation at the HF/3-21G//CHARMM level, including a correction based on the PES of the reference reaction in the gas phase at the B3LYP/6-31+G* level.

corresponding stationary points in the TRIAD model, with and without the acetate ion. Indeed, a tetrahedral intermediate was viable on the free-energy surface; TI2 is 1.4 kcal/mol lower in energy than TS1 when the acetate ion is removed (25.5 vs 26.9 kcal/mol) and 4.7 kcal/mol lower when the acetate ion is included (14.9 vs 19.6 kcal/mol).

When an acetate ion is added to the aqueous solution, and if a concerted double-proton transfer occurs ($Ac^-Im + HOH \rightarrow AcImH + OH^-$), the activation energy decreases by ∼7.3 kcal/mol (from 26.9 kcal/mol in REF to 19.6 kcal/mol in TRIAD), which is ∼16 kcal/mol smaller than the corresponding effect in the gas phase. If only a single-proton transfer occurs ($Ac^-Im + HOH \rightarrow Ac^-ImH^+ + OH^-$), this effect is estimated to be only ∼3 kcal/mol based on calculations using a third model (referred to as TRIAD1), in which $H^{\delta 1}$ was manually placed on $N^{\delta 1}$ (from 26.9 kcal/mol in REF to 24.0 kcal/mol in TRIAD1). However, these results might be inaccurate, considering that the transition states were not optimized in the polar dielectric continuum. Even if a double-proton transfer does occur in water, it may still not occur in the protein, where Asp102 is surrounded by polar residues and has a uniquely low pK_a value (between 0 and 3), making the ionized form energetically more preferable.[9,13,14,54] In any case, the effect of adding an acetate ion to the REF model in water is between ∼3 and 7 kcal/mol, which is close to the effect of adding Asp102 in the enzyme (∼6 kcal/mol), based on site-directed mutagenesis studies in subtilisin and in trypsin.[55]

3.3. Influence of the Enzyme Environment on the Deacylation Step. The PMF of the deacylation step, calculated from the QM/MM dynamics/umbrella sampling simulation of the PPE−BCM7 complex, is presented in Figure 2. Although the simulation was performed at the HF/3-21G//CHARMM level, an energy correction term from the B3LYP/6-31+G* level was added in the calculation of the PMF to improve the accuracy of the results. This correction was based on the assumption that the gas-phase geometries are similar between the HF/3-21G and B3LYP/6-31+G* levels, an assumption that has been shown to be satisfactory in the chemical reaction studied here (Table 2). In addition, the RMS deviation (RMSD) of all non-hydrogen atomic positions in the protein from their original positions in EA ($\xi_1 = 4.2$) was calculated from the average structure of each umbrella sampling window. The RMSDs, along the reaction coordinates, were continuous (reaching 2.9 Å at EP), indicating that the choice of the coordinates was appropriate. However,

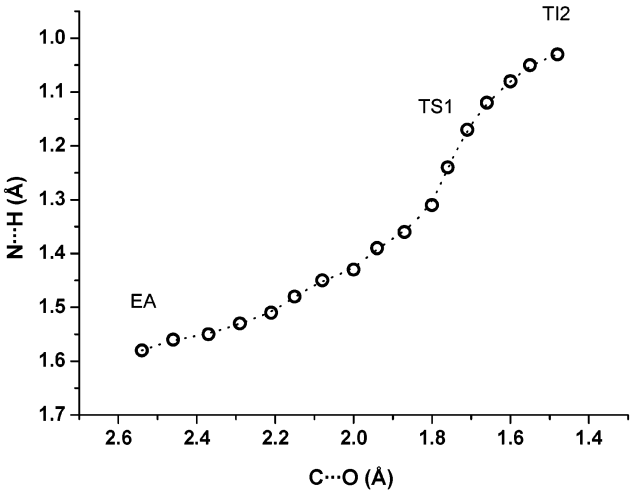


Figure 3. Change in $O_w\cdots C_1$ vs $N^{\epsilon 2}\cdots H_w{}^1$ during the attack of the nucleophilic water on the ester carbonyl in the active site of PPE−BCM7.

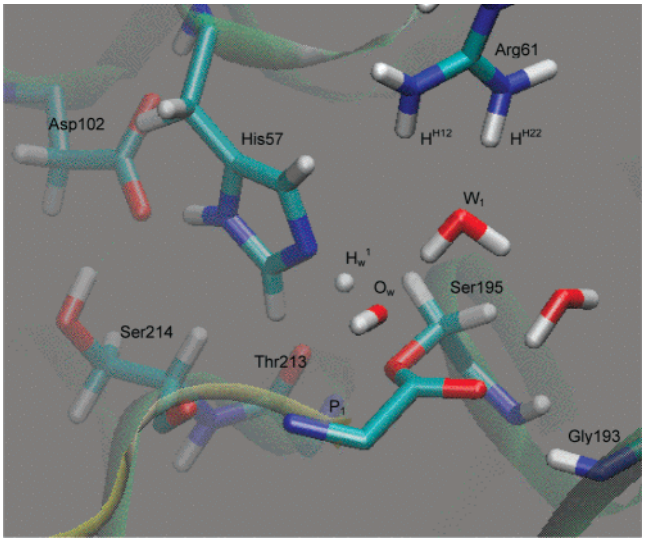


Figure 4. Snapshot of the QM/MM dynamics/umbrella sampling simulation, showing the elastase active site during the transition state (TS1) of the attack of the nucleophilic water (W) on the ester carbonyl (the protein main chain in green, the peptide substrate in yellow). The proton ($H_w{}^1$) is located approximately in the midpoint between His−$N^{\epsilon 2}$ and O_w. The latter is H-bonded to a water molecule (W_1), which is also H-bonded to Arg61. In the oxyanion hole (lower right side), a bulk water interacts with the carbonyl oxygen, in addition to the backbone NH groups of Gly193 and Ser195. His57−$C^{\epsilon 1}$ is located between the backbone carbonyls of Thr213 and Ser214. The figure was created with the VMD molecular graphics program.[63]

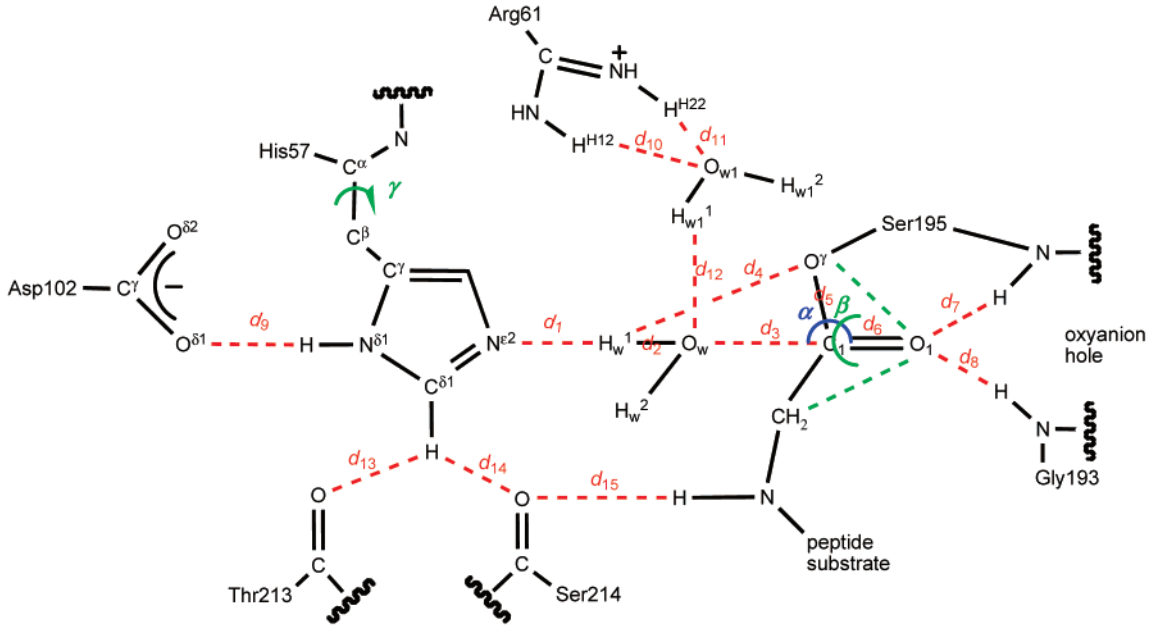
whether the reaction in the enzyme favors a stepwise mechanism cannot be answered directly from the PMF since the reaction coordinates were defined in such a way that the tetrahedral intermediate (TI2) must be formed. Yet, considering that the DFT results show how the reaction becomes stepwise when electrostatic effects (such as an acetate ion or the continuum solvent model) are added to the system, it is very likely that the reaction in the enzymatic environment (which is highly polar) is stepwise.

The rate-determining step is the formation of the tetrahedral intermediate, with an activation free energy ($\Delta G^{\ddagger}$) of 22.5 kcal/mol. To compare this result to the experimental results, the free energy of activation was calculated from the experimental rate constants (k_{cat}) using the Eyring equation (which relates k_{cat} to $\Delta G^{\ddagger}$). Because the acyl−enzyme does not accumulate in the hydrolysis of amides, the rate constants used are of the deacylation step in chymotrypsin with ester substrates, rather than with amide substrates.[1,56] For k_{cat} ranges between 0.01 and 192 s^{-1}, $\Delta G^{\ddagger}$ ranges between 14.3 and 20.2 kcal/mol. Thus, the calculated free energy of activation is slightly above the upper bound of the experimental results (and is most likely not overestimated by more than 1−2 kcal/mol, within the error range). Considering that the calculated activation free energy for the reference reaction in solution is 26.9 kcal/mol (at the B3LYP/6-31+G* level), the catalytic effect of the deacylation process in the elastase is about 4−5 kcal/mol. The latter result is very similar to the calculated catalytic effect of the acylation step, based on the QM/MM study in trypsin mentioned above[16] (5 kcal/mol), suggesting that the catalytic effect of deacylation is similar to that of acylation.

Selected geometrical parameters of the stationary points along the reaction coordinate, averaged over the last 2.5 ps of the relevant umbrella sampling windows (every 5 fs), are presented in Table 4. In addition, the values of His57−$N^{\epsilon 2}\cdots H_w{}^1$ (d_1) versus $O_w\cdots C_1$ (d_3), during the attack of the nucleophilic water on the ester carbonyl, are presented in Figure 3. These values reveal that the favored path for the formation of the tetrahedral intermediate in the enzyme environment is a concerted path, with a "late" proton ($H_w{}^1$) transfer from the nucleophilic water to His57−$N^{\epsilon 2}$. The movement of the water toward C_1 is correlated with the elongation of the $O_w-H_w{}^1$ (d_2) bond toward $N^{\epsilon 2}$. Also correlated is the elongation of C=O from a double bond to a single bond, due to the negative charge transfer to the tetrahedral structure (d_6 = 1.22, 1.26, and 1.31 Å in EA, TS1, and TI2, respectively). In the transition state (TS1), the $O_w\cdots C_1{=}O_1$ angle is very close to the Burgi angle ($\alpha = 106°$), and $H_w{}^1$ is located approximately at the midpoint between $N^{\epsilon 2}$ and O_w (d_1 = 1.24 Å and d_2 = 1.22 Å) (Figure 4). Thus, supported by early kinetic isotope effect experiments,[57] it is very likely that a protonic bridging occurs in the transition state.

The tetrahedral intermediate is calculated to be 15.7−18.3 kcal/mol less stable than the acyl−enzyme (Figure 2), a result similar to the one reported for the acylation step in a QM/FE study on trypsin.[58] However, in the latter study, it was assumed that the tetrahedral intermediate is energetically very close to the transition state, while our results suggest that this might not be the case in the deacylation step, where it is calculated to be 4.2−6.8 kcal/mol more stable than the transition state. It appears in two conformations that are separated by a small barrier of 1.3 kcal/mol. In the first conformation (TI2), $N^{\epsilon 2}-H_w{}^1$ is H-bonded to O_w (d_2 = 1.63 Å). In the second conformation (TI2′, which is 2.6 kcal/mol more stable), it is H-bonded to O^{γ}, instead (d_2 = 2.43 Å and d_4 = 1.64 Å). The switching between those conformations is mainly enabled by a small rotation of the His57 side chain about its $C^{\alpha}-C^{\beta}$ bond (γ = 86 and 78° in TI2 and TI2′, respectively). This movement is also correlated with a change in the C_1-O_w and C_1-O^{γ} bond lengths (d_3 = 1.54 Å and d_5 = 1.45 Å in TI2, and d_3 = 1.49 Å and d_5 = 1.50 Å in TI2′), indicating that the C_1-O^{γ} bond has to elongate prior to breaking. On this basis, and supported by our previous QM/MM study,[13] it is likely that the mobility of the His57 side chain helps the reaction to proceed to the enzyme−product complex (EP). Note that these results are based

Table 4. Selected Geometrical Parameters of Each Step of the Reaction, Averaged over the Last 2.5 ps (every 5 fs) of the Relevant Umbrella Sampling Windows



	α[a]	β[b]	γ[c]	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}
EA	93 ±6	174 ±5	91 ±7	1.71 ±0.10	1.01 ±0.05	2.70 ±0.14	2.67 ±0.21	1.34 ±0.02	1.22 ±0.02	2.34 ±0.21	1.94 ±0.16	1.57 ±0.07	2.36 ±0.45	1.92 ±0.21	2.28 ±0.58	3.40 ±0.27	2.40 ±0.22	2.05 ±0.16
TS1	106 ±3	139 ±5	90 ±5	1.24 ±0.06	1.22 ±0.07	1.76 ±0.06	2.42 ±0.12	1.39 ±0.03	1.26 ±0.02	2.72 ±0.17	2.06 ±0.13	1.54 ±0.07	2.18 ±0.23	1.93 ±0.19	1.85 ±0.16	3.19 ±0.25	2.38 ±0.16	2.30 ±0.22
TI2	107 ±3	131 ±5	86 ±8	1.04 ±0.03	1.63 ±0.12	1.54 ±0.05	2.20 ±0.03	1.45 ±0.03	1.31 ±0.02	2.63 ±0.18	1.85 ±0.16	1.55 ±0.08	2.83 ±0.41	1.90 ±0.15	2.75 ±0.52	3.17 ±0.27	2.41 ±0.19	2.21 ±0.15
TS'	109 ±3	128 ±4	84 ±6	1.03 ±0.04	1.82 ±0.16	1.53 ±0.05	2.00 ±0.05	1.46 ±0.04	1.31 ±0.03	2.69 ±0.18	1.88 ±0.15	1.54 ±0.08	3.61 ±0.41	2.27 ±0.62	3.02 ±0.30	2.76 ±0.18	2.39 ±0.25	2.32 ±0.20
TI2'	110 ±2	123 ±4	78 ±6	1.07 ±0.10	2.43 ±0.20	1.49 ±0.04	1.64 ±0.11	1.50 ±0.04	1.31 ±0.03	2.96 ±0.20	2.02 ±0.21	1.57 ±0.10	2.42 ±0.34	2.01 ±0.23	3.22 ±0.35	2.30 ±0.23	2.52 ±0.27	2.60 ±0.25
TS2	116 ±6	137 ±15	74 ±7	1.28 ±0.13	2.68 ±0.29	1.30 ±0.02	1.19 ±0.08	2.43 ±0.88	1.26 ±0.04	3.46 ±0.72	1.89 ±0.14	1.57 ±0.09	2.52 ±0.46	2.10 ±0.41	3.38 ±0.58	2.41 ±0.30	3.02 ±0.56	2.48 ±0.22
EP	120 ±3	149 ±12	82 ±6	1.80 ±0.11	3.30 ±0.23	1.37 ±0.03	0.99 ±0.03	2.57 ±0.13	1.20 ±0.02	2.49 ±0.27	1.82 ±0.10	1.56 ±0.07	1.56 ±0.07	1.56 ±0.07	3.73 ±0.60	2.75 ±0.35	2.70 ±0.38	2.63 ±0.23

[a] α is the $O_w\cdots C_1{=}O_1$ angle. [b] β is the $C^\alpha\cdots C_1{=}O_1\cdots O^\gamma$ dihedral angle. [c] γ is the $N\cdots C^\alpha{-}C^\beta\cdots C^\gamma$ dihedral angle of His57.

not only on the simulation along the ξ_2 reaction coordinate, where His57 is "forced" to move toward Ser195−O^γ, but also on the simulation along the ξ_1 reaction coordinate, where this movement is independent of the reaction coordinate.

The free-energy barrier of the second proton transfer from His57 to Ser195, which leads to the breakdown of the tetrahedral intermediate, is only 3.2 kcal/mol (Figure 2). To achieve this, the side chain of His57 rotates further ($\gamma = 74°$ in TS2), and Ser195 moves significantly (Figure 5a) to abstract the $N^{\epsilon 2}$−H_w^1 proton ($d_4 = 1.19$ Å in TS2). The enzyme−product complex (EP) is 12.9 kcal/mol less stable than EA. This could be explained by two main changes in the S_1 binding site: (a) the loss of the H-bond between the backbone carbonyl of Ser214 and the NH group of the P_1 residue ($d_{15} = 2.05$ and 2.63 Å, in EA and EP, respectively), and (b) the formation of a carboxylate group (which is less stabilized than the EA ester carbonyl in this pocket). While EA is based on a crystal structure and represents a more "realistic" picture of the acyl−enzyme intermediate, EP represents a structure that has yet to proceed to the "free" enzyme. Therefore, the final products (the free enzyme and the solvated peptide) are expected to have lower energy.

3.4 Hydrogen-Bonding Network. The results of the present work support our previous studies concerning the H-bonding network surrounding the catalytic triad, and they also reveal new and interesting details. One interesting detail is that, in addition to the nucleophilic water (W), a second water molecule, W_1, is present in the active site throughout the entire deacylation process, H-bonded via its oxygen to H^{H22} and H^{H12} of Arg61 (Table 4, Figure 4). Via one of its hydrogens, W_1 is H-bonded

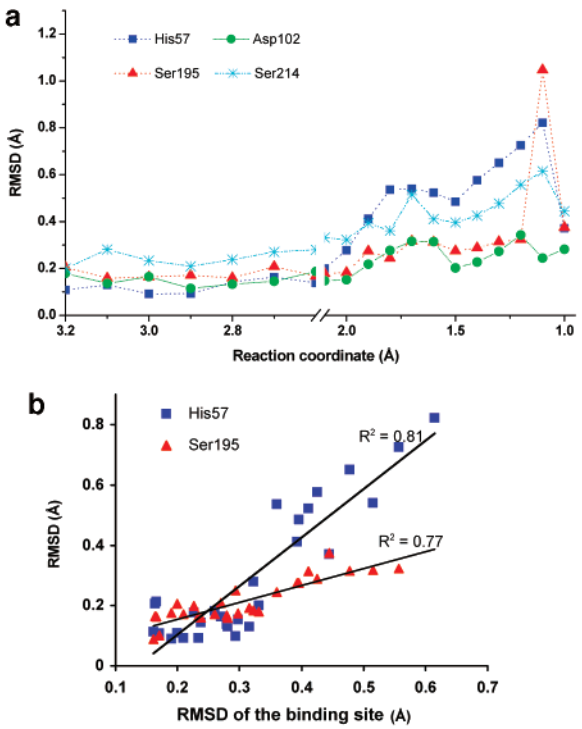


Figure 5. (a) RMSD of all non-hydrogen atomic positions of the catalytic residues from their original positions in EA at different points along the reaction coordinate (from TS1 to EP). The calculation of each RMSD was based on the alignment between the average structure of each umbrella sampling window and the first average umbrella sampling window (ξ_1 = 4.2 Å), using all residues within 8 Å from the carbonyl carbon (C_1). (b) Correlation between the RMSDs of the binding site and the RMSDs of His57 and Ser195 during the entire reaction pathway (EA to EP). The binding site included Ser214, Phe215, Val216, and the P1−P4 residues. The RMSD of Ser195 in TS2 (ξ_2 = 1.1 Å) was not included.

to O_w of the nucleophilic water, stabilizing it during the nucleophilic attack (d_{12} = 2.28 and 1.85 Å in EA and TS1, respectively). This interaction is lost after the formation of TI2, when O_w moves away and forms a bond with C_1 (d_{12} = 2.72 Å). Via its second hydrogen, W_1 is H-bonded to an exchanging bulk water. We have previously proposed that Arg61, which is conserved in some SPs (including plasminogen, chymotrypsinogen, and human neutrophil elastase) and is replaced by potential H-bond donors (Lys, Ser, Thr, Asn, Tyr, and Pro) in others, may play a role in stabilizing or orienting water molecules in the active site of the acyl−enzyme.[18] On the basis of the current results, we also suggest that during deacylation, Arg61 stabilizes a specific water molecule that participates in the H-bonding network and that has an indirect catalytic role.

The oxyanion hole H-bonding network is also interesting. Because it is exposed to the bulk, a water molecule is always present close enough to act as an alternative H-bond donor (Figure 4), in addition to the backbone NH groups of Gly193 and Ser195. Perhaps this is to compensate for the loss of the H-bond between O_1 and Ser195−NH during the reaction. In contrast to the short and stable H-bond with Gly193 (d_8 < ~2.2 Å), the H-bond with Ser195 gets weaker following the nucleophilic attack (d_7 > ~2.6 Å) and is completely lost when Ser195 (including its backbone) moves toward His57 to abstract the $N^{\epsilon 2}-H_w{}^1$ proton (d_7 = 2.96 and 3.46 Å in TI2′ and TS2, respectively). This may also explain why a third H-bond is formed between Thr220 and O_1 in the oxyanion hole of subtilisin.[59,60] The latter has been shown to have a much smaller effect on catalysis than the H-bond formed with the amide NH_2 of Asn155 (the equivalent to NH of Gly193 in elastase). Replacing Asn155 with alanine reduces the catalytic efficiency by 300−400-fold,[60,61] while replacing Thr220 with alanine reduces catalytic efficiency only by ~8-fold. Thus, we suggest that the main H-bond responsible for stabilizing the tetrahedral intermediate in the oxyanion hole during the deacylation step is the one formed with Gly193−NH, while additional stabilization is gained by H-bonding to Ser195 and/or to random bulk water molecules.

The last issues discussed here concern the stabilization of His57 and the coupling between the local chemical changes and the more global changes in the active and binding sites. In TI2, the imidazolium ion of His57 is stabilized by three H-bonds: with Asp102−$O^{\delta 1}$ (d_9 = 1.55 Å), with the backbone carbonyl of Ser214 (a $C^{\epsilon 1}-H\cdots O$ H-bond; d_{14} = 2.41 Å), and with O_w (d_2 = 1.63 Å). The breaking of the latter two H-bonds during the rotation of His57 about its $C^{\alpha}-C^{\beta}$ bond requires ~1−2 kcal/mol (Figure 2). These two are replaced by H-bonds with the backbone carbonyl of Thr213 (a $C^{\epsilon 1}-H\cdots O$ H-bond; d_{13} = 2.41 Å) and with Ser195−O^{γ}, respectively, allowing the reaction to proceed in the direction of the products. On this basis, the completion of the reaction most likely does not require the "ring-flip mechanism", as proposed earlier by Ash et al., based on NMR data.[62]

The described H-bonding network surrounding the imidazolium ion, which also includes a H-bond between Ser214−O^{γ} and Asp102−$O^{\delta 1}$ (Figure 4), leads to coupling between the movements of His57, Asp102, and Ser214 (Figure 5a). Furthermore, because Ser214 is also part of the binding site of the enzyme, stabilizing the P_1 residue via its backbone carbonyl, the movements of His57 in the second part of the reaction also affect structural rearrangements of the binding site−peptide complex. Evidence for this includes (a) the lengthening of the Ser214···P_1 H-bond (d_{15}) and (b) the high correlation between the RMSDs of the binding site and of His57 from their original position in EA throughout the reaction (R^2 = 0.81; Figure 5b). The RMSDs of the binding site are also correlated with the RMSDs of Ser195 (R^2 = 0.77), which is covalently bound to P_1. On the basis of these results, we suggest that peptide hydrolysis may be synchronized with product release.

4. Conclusion

In this paper, we have studied the deacylation step of the SP reaction mechanism by using ab initio QM/MM calculations combined with MD/umbrella sampling calculations. The results suggest a concerted mechanism for the gas-phase reaction; however, the structure of the transition state is tetrahedral-like. When solvent effects or an acetate ion are added to the reference system, a reduction of ~23 kcal/mol occurs in the activation barrier, and the mechanism becomes stepwise. This reduction mainly reflects the effect of Asp102, the oxyanion hole, and the water molecules in the enzyme active site, suggesting that

(59) Rao, S. N.; Singh, U. C.; Bash, P. A.; Kollman, P. A. *Nature* **1987**, *328*, 551−554.
(60) Braxton, S.; Wells, J. A. *J. Biol. Chem.* **1991**, *266*, 11797−11800.
(61) O'Connell, T. P.; Day, R. M.; Torchilin, E. V.; Bachovchin, W. W.; Malthouse, J. G. *Biochem. J.* **1997**, *326*, 861−866.
(62) 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*, 10371−10376.
(63) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.

a tetrahedral intermediate is viable in the enzyme environment (although there is no direct experimental evidence for this). In contrast to the aqueous solution, the tetrahedral intermediate in the enzyme is shown to be a relatively stable species, ~7 kcal/mol lower than the transition state. The stabilization of the nucleophilic water is achieved by another water that is bridged between the nucleophilic oxygen and the side chain of Arg61, suggesting an indirect catalytic role for this residue. Progression of the reaction toward the products is mainly enabled by a rotation of the His57 side chain about its $C^{\alpha}-C^{\beta}$ bond, which requires the breaking of the imidazolium H-bonds with O_w and with the Ser214, both of which are replaced by H-bonds with Ser195$-O^{\gamma}$ and Thr213, respectively. Movement of Ser195 assists the completion of the reaction, causing changes in the H-bonds of the oxyanion hole. This reorganization of the active site residues and H-bonds, as well as the changes in the free energy, demonstrates how the system is using minimal "reorganization energy" along the reaction coordinate, giving a catalytic effect of 5 kcal/mol. Interestingly, the movements of His57 during the reaction causes rearrangements of the binding site (via the interaction with Ser214), suggesting that product release may also be implicated in the deacylation process.

Acknowledgment. We are grateful to Ben Webb, Péter Várnai, Frank Alber, and Christopher J. Schofield for helpful discussions. M.T. is partially funded by the Overseas Research Students (ORS) Awards Scheme. The work was supported, in part, by a grant from the National Foundation for Cancer Research. We thank the Oxford Supercomputing Centre for the generous allocation of computer time.

JA047010A