Article

Molecular Dynamics Simulations of the Catalytic Pathway of a Cysteine Protease: A Combined QM/MM Study of Human Cathepsin K

Shuhua Ma, Lakshmi S. Devi-Kesavan, and Jiali Gao

More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 3 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

Molecular Dynamics Simulations of the Catalytic Pathway of a Cysteine Protease: A Combined QM/MM Study of Human Cathepsin K

Shuhua Ma, Lakshmi S. Devi-Kesavan, and Jiali Gao*

Contribution from the Department of Chemistry and Supercomputing Institute, Digital Technology Center, University of Minnesota, 207 Pleasant Street SE, Minneapolis, Minnesota 55455

Received June 10, 2007; E-mail: gao@chem.umn.edu

Abstract: Molecular dynamics simulations using a combined QM/MM potential have been performed to study the catalytic mechanism of human cathepsin K, a member of the papain family of cysteine proteases. We have determined the two-dimensional free energy surfaces of both acylation and deacylation steps to characterize the reaction mechanism. These free energy profiles show that the acylation step is rate limiting with a barrier height of 19.8 kcal/mol in human cathepsin K and of 29.3 kcal/mol in aqueous solution. The free energy of activation for the deacylation step is 16.7 kcal/mol in cathepsin K and 17.8 kcal/mol in aqueous solution. The reduction of free energy barrier is achieved by stabilization of the oxyanion in the transition state. Interestingly, although the "oxyanion hole" has been formed in the Michaelis complex, the amide units do not donate hydrogen bonds directly to the carbonyl oxygen of the substrate, but they stabilize the thiolate anion nucleophile. Hydrogen-bonding interactions are induced as the substrate amide group approaches the nucleophile, moving more than 2 Å and placing the oxyanion in contact with Gln19 and the backbone amide of Cys25. The hydrolysis of peptide substrate shares a common mechanism both for the catalyzed reaction in human cathepsin K and for the uncatalyzed reaction in water. Overall, the nucleophilic attack by Cys25 thiolate and the proton-transfer reaction from His162 to the amide nitrogen are highly coupled, whereas a tetrahedral intermediate is formed along the nucleophilic reaction pathway.

1. Introduction

Cysteine proteases are proteolytic enzymes that catalyze the cleavage of peptide bonds by using a cysteine residue in the active site. These enzymes have been identified from diverse organisms, such as bacteria, plants, and animals,[1-6] which have been classified into nearly 50 protein families belonging to seven major clans based on the nature and sequence of the catalytic residues.[7,8] Because of the important role of these enzymes in pathological cellular processes, the reaction mechanism of cysteine proteases has been extensively investigated experimentally and computationally. One member of the papain family of cysteine proteases is human cathepsin K, which is highly expressed in osteoclasts and implicated in bone resorption.[9-16]
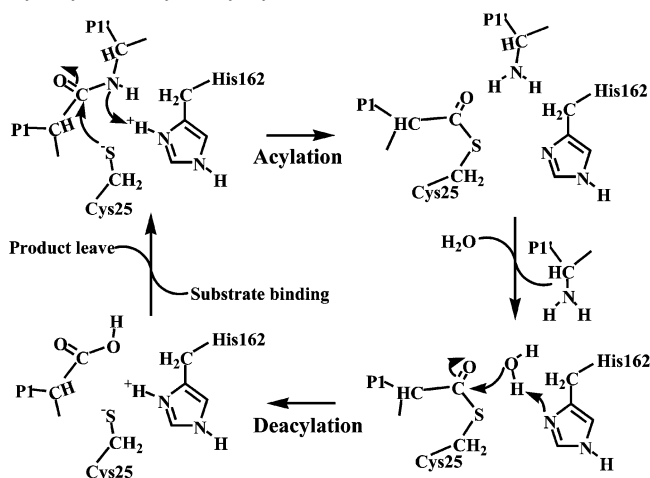
Deficiency of cathepsin K causes the bone-sclerosing disorder pycnodysostosis, which at the molecular level is a result of insufficient degradation of type I collagen during bone remodeling, leading to osteoporosis and increased bone fragility.[13] Cathepsin K is also expressed in synovial fibroblasts, which are involved in pathological erosion of articular cartilage,[10,15] and in various epithelial cells[17] where its biological function is still unknown. Cathepsin K is an attractive therapeutic target for the treatment of osteoporosis and rheumatoid arthritis, and a theoretical study aimed at a deeper understanding of the catalytic mechanism and binding interactions in the active site of cathepsin K is of great interest.

The active site of cathepsin K is characterized by a catalytic triad, consisting of Cys25, His162, and Asn182 residues. In addition, a highly conserved residue, Gln19, is found in the active site that plays a major role in cathepsin K catalysis by forming the "oxyanion hole" to stabilize the transition state. In

(1) Neuberger, A.; Brocklehurst, K. *Hydrolytic Enzymes*; Elsevier; Sole distributors for the U.S. and Canada Elsevier Science: Amsterdam; New York, 1987.
(2) Polgar, L.; Halasz, P. *Biochem. J.* **1982**, *207*, 1.
(3) Drenth, J.; Kalk, K. H.; Swen, H. M. *Biochemistry* **1976**, *15*, 3731.
(4) Baker, E. N. *J. Mol. Biol.* **1977**, *115*, 263.
(5) Baker, E. N. *J. Mol. Biol.* **1980**, *141*, 441.
(6) Storer, A. C.; Menard, R. *Methods Enzymol.* **1994**, *244*, 486.
(7) Rawlings, N. D.; Morton, F. R.; Barrett, A. J. *Nucleic Acids Res.* **2006**, *34*, D270.
(8) Correa, G. C.; Margis-Pinheiro, M.; Margis, R. *Genet. Mol. Biol.* **2001**, *24*, 275.
(9) Tezuka, K.; Tezuka, Y.; Maejima, A.; Sato, T.; Nemoto, K.; Kamioka, H.; Hakeda, Y.; Kumegawa, M. *J. Biol. Chem.* **1994**, *269*, 1106.
(10) Bromme, D.; Okamoto, K. *Biol. Chem. Hoppe-Seyler* **1995**, *376*, 379.
(11) Bromme, D.; Okamoto, K.; Wang, B. B.; Biroc, S. *J. Biol. Chem.* **1996**, *271*, 2126.
(12) Gelb, B. D.; Edelson, J. G.; Desnick, R. J. *Nat. Genet.* **1995**, *10*, 235.

(13) Gelb, B. D.; Shi, G. P.; Chapman, H. A.; Desnick, R. J. *Science* **1996**, *273*, 1236.
(14) Drake, F. H.; Dodds, R. A.; James, I. E.; Connor, J. R.; Debouck, C.; Richardson, S.; Lee-Rykaczewski, E.; Coleman, L.; Rieman, D.; Barthlow, R.; Hastings, G.; Gowen, M. *J. Biol. Chem.* **1996**, *271*, 12511.
(15) Bossard, M. J.; Tomaszek, T. A.; Thompson, S. K.; Amegadzie, B. Y.; Hanning, C. R.; Jones, C.; Kurdyla, J. T.; McNulty, D. E.; Drake, F. H.; Gowen, M.; Levy, M. A. *J. Biol. Chem.* **1996**, *271*, 12517.
(16) Inaoka, T.; Bilbe, G.; Ishibashi, O.; Tezuka, K.; Kumegawa, M.; Kokubo, T. *Biochem. Biophys. Res. Commun.* **1995**, *206*, 89.
(17) Buhling, F.; Gerber, A.; Hackel, C.; Kruger, S.; Kohnlein, T.; Bromme, D.; Reinhold, D.; Ansorge, S.; Welte, T. *Am. J. Respir. Cell Mol. Biol.* **1999**, *20*, 612.

Scheme 1. Illustration of the Reaction Pathway of Peptide Hydrolysis Catalyzed by Cysteine Proteases

Figure 1. Schematic representation of human cathepsin K. The substrate, Ace-Leu-Arg-Phe-Nme, used in the present computation and the catalytic triad in the active site are specifically highlighted.

the catalytic cycle, His162 first acts as a general acid in the peptide bond cleaving (acylation) step, and then as a general base in the subsequent hydrolysis (deacylation) step (Scheme 1).[1,18] Cathepsin K is highly homologous (42% sequence identity) to the plant cysteine protease papain, which is among the most extensively studied proteases. In fact, many experimental findings on papain can be directly compared with computational results on cathepsin K.

Because of the important biological functions of cysteine proteases, the catalytic mechanisms of these enzymes have been extensively investigated. Early studies by Howard and Kollman[19] on nucleophilic additions of formamide by OH^- and by SH^- in the gas phase suggested that a stable tetrahedral intermediate is produced by the hydroxide addition, but there is no stable tetrahedral intermediate by SH^- addition. Solvation effects stabilize both the reactant state and the ion–dipole complex.[19] A similar reaction between *N*-methylacetamide (NMA) and methanethiolate ion in water was carried out employing B3LYP/ aug-cc-pVDZ and combined QM/MM simulations.[20] It was found that the tetrahedral species is not stable without a general acid catalyst. Strajbl et al.[21] performed a comprehensive study of the general base and general acid-catalyzed thiolysis of formamide and the hydrolysis of methyl thioformate using B3LYP/aug-cc-pVDZ//HF-6-31G(d) calculations plus model solvation. The most significant finding is that both the amide thiolysis (acylation) and the hydrolysis of methyl thiolformate (deacylation) have a stepwise character with the formation of stable tetrahedral intermediates. The activation barriers for the acylation and deacylation steps are both 26 kcal/mol. Arad et al.[22] studied the minimum energy path for a peptide hydrolysis reaction in the active site of papain using the semiempirical AM1 method. It was found that it is preferred for a proton transfer from His159 to the substrate prior to or concerted with the nucleophilic attack by Cys25 thiolate ion than a general acid assisted pathway. The papain system has been subsequently examined by a number of groups using combined B3LYP/MM
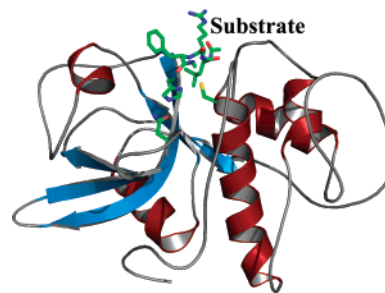
and AM1/MM potentials.[23] These studies suggest that the enzyme–substrate thiolester intermediate is formed by a concerted pathway. Despite these efforts, a detailed mechanism of cysteine protease reactions based on free energy simulations has not emerged, perhaps due to the tremendous computational costs needed to determine two-dimensional free energy surfaces.

In this work, we employ molecular dynamics free energy simulations with a combined quantum mechanical and molecular mechanical (QM/MM) potential to model the hydrolytic mechanism of a peptide substrate catalyzed by human cathepsin K. We carry out two-dimensional free energy simulations on both the acylation and the deacylation steps. Each reaction profile has two reaction coordinates: one specifies the nucleophilic attack and the other describes the proton-transfer reaction. To understand the actions of cathepsin K, the corresponding model reactions in water have been determined using the same method. In the following, we describe the computational details employed in this work. This is followed by results and discussion. We conclude with a summary of major findings of this study.

2. Computational Details

A critical prerequisite in computational study of enzymatic reactions is to construct a model for the Michaelis complex. However, this is not an obvious task and often is very difficult because the Michaelis complex structure of the native substrate is typically not available experimentally. It is safe to say that the reliability of the computational results is critically dependent on these structures, yet few publications have addressed this issue.[24] Here, we present a description of the procedure that we typically utilize to approach this question for the study of enzymatic reactions in general, beginning from substrate selection to the final equilibration of the system. We also summarize the potential energy function used in the present study and the technical details in the free energy simulations.

2.1. The Construction of a Model for the Enzyme–Substrate (Michaelis) Complex. The active site of human cathepsin K is located in a "V"-shaped cleft formed by two domains: Cys25 and His162, one from each domain, along with Asn182 form the catalytic triad, and the peptide chain of the substrate lies across the cleft (Figure 1).[25,26] Although there are no specific substrates that have been identified for cathepsin K, the enzyme can effectively cleave the protein strand of type I and II collagens, and kinetic constants have been determined

(18) Brocklehurst, K. In *Cysteine Proteinases Their Inhib., Proc. Int. Symp.*; Turk, V., Ed.; Walter de Gruyter & Co.: Berlin, New York, 1986; p 307.
(19) Howard, A. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 7195.
(20) Byun, K.; Gao, J. *J. Mol. Graphics Modell.* **2000**, *18*, 50.
(21) Strajbl, M.; Florian, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 4471.
(22) Arad, D.; Langridge, R.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 491.

(23) (a) Harrison, M. J.; Burton, N. A.; Hillier, I. H. *J. Am. Chem. Soc.* **1997**, *119*, 12285. (b) Han, W. G.; Tajkhorshid, E.; Suhai, S. *J. Biomol. Struct. Dyn.* **1999**, *16*, 1019.
(24) (a) Garcia-Viloca, M.; Poulsen, T. D.; Truhlar, D. G.; Gao, J. *Protein Sci.* **2004**, *13*, 2341. (b) Ruiz-Pernia, J. J.; Estanislao, S.; Tunon, I. *J. Phys. Chem. B* **2006**, *110*, 20686.
(25) McGrath, M. E.; Klaus, J. L.; Barnes, M. G.; Bromme, D. *Nat. Struct. Biol.* **1997**, *4*, 105.
(26) Thompson, S. K.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14249.

for numerous synthetic peptide substrates.[15,27,28] To model the enzymatic hydrolytic pathway, we need to decide the length and sequence of the peptide substrate, and we build the model system based on experimental data.

We begin by examining the extensively studied, highly homologous enzyme papain, which has a large binding site. In the late 1960s, it was suggested that the active site of papain can accommodate at least seven amino acid residues.[29−31] The residues in the binding pocket of the enzyme are labeled by S2, S1 and S1′, S2′, etc., and those of the substrate by P2, P1 and P1′, P2′, etc., where residues on the C-terminal end from the scissile bond are labeled with primes. The S2, S1, and S1′ binding sites of papain are well defined, and they interact with the substrate both through main chain and side chain contacts. The remaining sites are less organized. Turk et al.[32] compared a number of papain structures, which indicate that there is no common binding pocket beyond the S3 site. Although substrate residues beyond P3 can still interact with the enzyme surface, these interactions are not confined to the same region in different structures. These results are similar to the binding specificity data for cathepsin K.[28] Consequently, we decided to use a minimal, three-residue substrate, that is, the P2−P1−P1′ sequence, in our computational study, which is sufficiently large for substrate recognition and provides a tractable model for the peptide hydrolysis reaction catalyzed by cathepsin K.

The next issue is to select the specific sequence for the substrate. Although the sequence homology between papain and cathepsin K is very high, there are clear differences in substrate recognition. Berger and Schechter[31] reported that the S2 binding site of the papain family has a preference for large hydrophobic residues such as phenylalanine.[33−35] However, the presence of a larger Leu209 residue in cathepsin K creates a pocket that is too shallow for Phe. Instead, an aliphatic Leu is preferred at the P2 position, and this has been confirmed by X-ray crystal structures.[25,26] The P1 residue interacts with the S1 binding site through hydrogen-bonding interactions between the backbone amides of P1 and Asn161. Furthermore, the P1 carbonyl group is loosely bound to the backbone amide of Cys25 and the side chain of Gln19, eventually forming the oxyanion hole at the transition state. Lecaille et al.[36] tested many sequences for cathepsin K and found that the order of preference for P1 residue is Arg > Lys > Nle(norleucine) or Gln > Met > Thr > Leu. X-ray structures show that the side chain of the P1 residue is fully exposed to the solvent.[25,37−39]

The specificity at the P1′-binding position by cathepsin K was examined by Alves et al.,[28] and it was found that larger side chains are preferred with a binding order of Asn > Phe > Leu > Gly at this site.[40,41] In the complex structure of cathepsin K and an inhibitor (called INA),[26] the P1′ amino group of the substrate is oriented toward the carbonyl oxygen of Asn161, and its carbonyl group is hydrogen bonded

to the $N_{\epsilon 1}$ atom of Trp184 side chain. In addition, the structure of another cathepsin K complex (with the inhibitor APC3328)[25] showed that the benzene ring of sulfonyl group at this position can be accommodated in the pocket formed by Ala137, Ser138, Gln143, Asn161, His162, and Trp184.

Combining this analysis and X-ray structures, we decided to use the sequence Ace-Leu-Arg-Phe-NMe, where the N- and C-termini are capped by an acyl (Ace) group and an *N*-methyl amino (NMe) group, respectively. We constructed a model for the enzyme−substrate complex by using the coordinates of the X-ray structure for the enzyme−INA inhibitor complex (PDB: 1AYU, 2.2 Å resolution)[26] as the template along with the structures that contain residue Leu[25,41,42] and Phe analogue[25] at the P2 and P1′ positions. The available cathepsin K−inhibitor complex structures either have no P1 occupancy or use hydrophobic side chains to take this position.

2.2. Potential Energy Function. We use a combined QM/MM potential[43−45] to describe the enzymatic process in this study. In particular, for the acylation step, 71 atoms are treated quantum mechanically, which is depicted in Scheme 2a. The C_α atoms of Leu (substrate), Cys25, and His162 are the boundary atoms, which are treated by the generalized hybrid orbital (GHO) method.[46,47] The rest of the enzyme is modeled by the CHARMM22 force field, and the three-point charge TIP3P model.[48,49] Mechanistically, after the acylation step, the product fragment, NH₂-Phe-NMe, leaves the active site, and water molecules occupy the cavity. In the subsequent deacylation step that restores the enzyme's catalytic cycle, the QM region consists of 47 atoms (Scheme 2b). For comparison with the uncatalyzed processes, we have investigated the corresponding reactions in aqueous solution, for which 30 QM atoms are used for the acylation and 26 QM atoms for the deacylation step, respectively, which are shown in Scheme 2c and d.

We used a mixed empirical function and the semiempirical AM1 Hamiltonian[50,51] in the present calculation. The original AM1 model performs exceptionally well for the model amide hydrolysis reaction in comparison with results from density functional theory (DFT) and the Gaussian-2 theory (G2) (see below). Still, the accuracy can be further improved, particularly on small ions. First, we make small adjustments to the original AM1 parameters to improve the proton affinity for the hydroxide ion. Next, we incorporate an empirical term, V_{SVB}, to correct the reaction energies and barrier heights.[52] Thus, the total potential energy function is expressed as follows:

$$V = \langle \Psi | H_{QM} + H_{QM/MM} | \Psi \rangle + V_{MM} + V_{SVB} \qquad (1)$$

where the first term gives the energy of the QM subsystem and the interaction energy between the QM and MM regions, the second term, V_{MM}, is the potential energy of the MM region, and the third term, V_{SVB}, is an empirical function used to correct the semiempirical QM energy, $\langle \Psi | H_{QM} | \Psi \rangle$, by fitting against experimental results or high level

(27) Lecaille, F.; Weidauer, E.; Juliano, M. A.; Bromme, D.; Lalmanach, G. *Biochem. J.* **2003**, *375*, 307.
(28) Alves, M. F. M.; Puzer, L.; Cotrin, S. S.; Juliano, M. A.; Juliano, L.; Bromme, D.; Carmona, A. K. *Biochem. J.* **2003**, *373*, 981.
(29) Schechter, I.; Berger, A. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157.
(30) Schechter, I.; Berger, A. *Biochem. Biophys. Res. Commun.* **1968**, *32*, 898.
(31) Berger, A.; Schechter, I. *Philos. Trans. R. Soc. London, Ser. B* **1970**, *257*, 249.
(32) Turk, D.; Guncar, G.; Podobnik, M.; Turk, B. *Biol. Chem.* **1998**, *379*, 137.
(33) Kamphuis, I. G.; Kalk, K. H.; Swarte, M. B.; Drenth, J. *J. Mol. Biol.* **1984**, *179*, 233.
(34) Maes, D.; Bouckaert, J.; Poortmans, F.; Wyns, L.; Looze, Y. *Biochemistry* **1996**, *35*, 16292.
(35) Gillmor, S. A.; Craik, C. S.; Fletterick, R. J. *Protein Sci.* **1997**, *6*, 1603.
(36) Lecaille, F.; Choe, Y.; Brandt, W.; Li, Z.; Craik, C. S.; Bromme, D. *Biochemistry* **2002**, *41*, 8447.
(37) Catalano, J. G.; Deaton, D. N.; Furfine, E. S.; Hassell, A. M.; McFadyen, R. B.; Miller, A. B.; Miller, L. R.; Shewchuk, L. M.; Willard, D. H., Jr.; Wright, L. L. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 275.
(38) Adkison, K. K.; Barrett, D. G.; Deaton, D. N.; Gampe, R. T.; Hassell, A. M.; Long, S. T.; McFadyen, R. B.; Miller, A. B.; Miller, L. R.; Payne, J. A.; Shewchuk, L. M.; Wells-Knecht, K. J.; Willard, D. H., Jr.; Wright, L. L. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 978.
(39) Boros, E. E.; Deaton, D. N.; Hassell, A. M.; McFadyen, R. B.; Miller, A. B.; Miller, L. R.; Paulick, M. G.; Shewchuk, L. M.; Thompson, J. B.; Willard, D. H., Jr.; Wright, L. L. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3425.
(40) Schuster, M.; Kasche, V.; Jakubke, H. D. *Biochim. Biophys. Acta* **1992**, *1121*, 207.
(41) Menard, R.; Carmona, E.; Plouffe, C.; Bromme, D.; Konishi, Y.; Lefebvre, J.; Storer, A. C. *FEBS Lett.* **1993**, *328*, 107.
(42) Zhao, B.; Janson, C. A.; Amegadzie, B. Y.; D'Alessio, K.; Griffin, C.; Hanning, C. R.; Jones, C.; Kurdyla, J.; McQueney, M.; Qiu, X.; Smith, W. W.; Abdel-Meguid, S. S. *Nat. Struct. Biol.* **1997**, *4*, 109.
(43) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(44) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
(45) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.
(46) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714. Garcia-Viloca, M.; Gao, J. *Theor. Chem. Acc.* **2004**, *111*, 280.
(47) Amara, P.; Field, M. J.; Alhambra, C.; Gao, J. *Theor. Chem. Acc.* **2000**, *104*, 336.
(48) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(49) MacKerell, A. D., Jr.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586.
(50) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
(51) Dewar, M. J. S.; Yuan, Y.-C. *Inorg. Chem.* **1990**, *29*, 3881.
(52) Devi-Kesavan, L. S.; Garcia-Viloca, M.; Gao, J. *Theor. Chem. Acc.* **2003**, *109*, 133.

Scheme 2. Partition of Quantum and Classical Regions in Combined QM/MM Simulations for (a) the Acylation Step, and (b) the Deacylation Step in Cathepsin K, and for (c) the Acylation Step, and (d) the Deacylation Step in Aqueous Solution

(a)

(b)

(c)

(d)
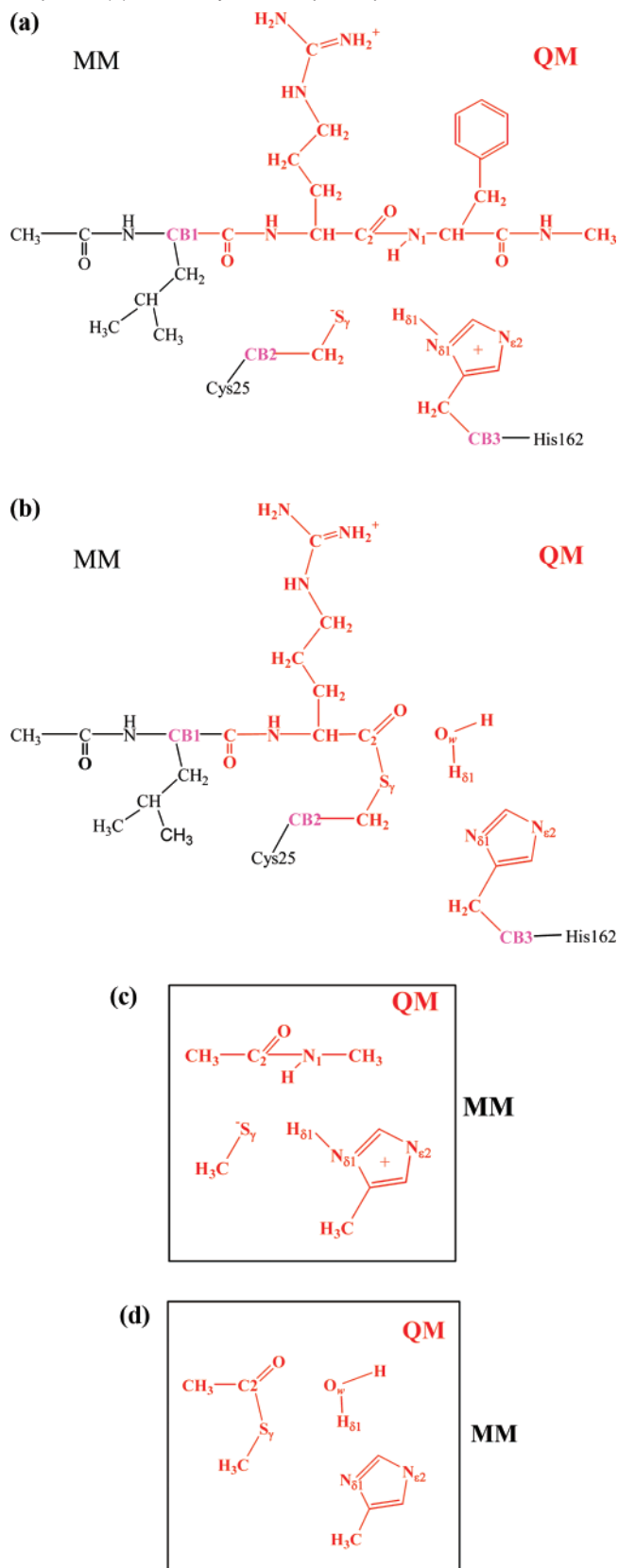
ab initio and DFT data. The last term depends on the reaction coordinate for each reaction. The final parameters and the functional form of the V_{SVB} potential are given in the Supporting Information. Table 1 lists

the computed and experimental energies for the model reactions that have been considered (Scheme 3), and Table 2 provides the heats of formation and proton affinities for the individual compounds. Note that, although a few entries have somewhat greater errors than others in the computed heats of formation, the relative energies for the proton affinities are in reasonable accord with experiments.

2.3. Molecular Dynamics Simulations. The pH−activity profile shows an optimal pH of 5.5 for cathepsin K,[11] based on which protonation states of titratable residues are assigned. All Asp and Glu are unprotonated, and Lys and Arg are protonated. There are two histidine residues in cathepsin K: one is the active site residue His162, which is protonated and forms an ion-pair with the nucleophile Cys25 thiolate anion, and the other is His177, which was set to be neutral with the proton on $N_{\delta 1}$ by considering hydrogen-bond interactions with the neighboring residue. The total charge of the enzyme cathepsin K (+7 au) and substrate (+1 au) is neutralized by 8 chloride ions.

The enzyme−substrate complex is fully solvated by a cubic box ($60 \times 60 \times 60$ Å³) of water molecules. Water molecules that are within 2.5 Å of any heavy atoms of the system are deleted, and crystal waters are kept. The chloride counterions are positioned about 5−6 Å away from positively charged residues on the protein surface, and all of these chloride anions are relatively far away from the active site. The substrate and residues within a 10 Å sphere of the active site Cys25 thiolate ion are first minimized for 30 steps using the adopted basis set Newton−Raphson (ABNR)[53] method to remove close contacts, and this subsystem is relaxed by short molecular dynamics calculations. Next, we perform 30 steps of optimization of the entire system to remove other close contacts, followed by a series of molecular dynamics simulations to gradually heat the system, at 30 ps intervals in the temperature ranges of 0−100 K, 100−200 K, and 200−298 K. During the heating simulations, we apply harmonic restraints to all non-hydrogen atoms of the protein and substrate with a force constant of 20 kcal/mol·Å². After heating, we perform an additional 50 ps of equilibration with periodic boundary conditions (PBC) with the PBOUND option[54] incorporated for QM/MM potentials,[55] and the isothermal−isobaric (NPT) ensemble at 298 K and 1 atm. This yields an average volume of 59.19835³ Å³, which is adopted subsequently in molecular dynamics simulations with the combined QM/MM potential using the NVT ensemble. A cutoff distance of 13 Å is used for nonbonded interactions, and we used a switch function for electrostatic interactions in the range of 12−13 Å, and a shift function for the Lennard-Jones terms. The nonbonded pair list is updated on every 25 steps, and 100 steps for the image coordinates. All bond lengths involving hydrogen atoms are constrained by the SHAKE algorithm,[56] and the dielectric constant is unity. We used the CHARMM22 force field[49] in this stage.

After the enzyme−substrate system was constructed and equilibrated using the MM force field, the system is partitioned into QM and MM regions as described above. We first perform 20 steps of energy minimization of the QM region mainly to ensure the initial self-consistent-field convergence. Next, the system is gradually heated to 298 K again, following the same procedure described for the MM system, now using the QM/MM potential. All simulation conditions are the same as before, except that the images are updated every 25 steps. The velocity Verlet algorithm[57] with a time step of 1 fs and the Nośe−Hoover thermostat algorithm[58,59] are used throughout.

The initial structure for the deacylation step is taken from the "product" of the acylation step, replacing the amino product by water

(53) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(54) Shirley, W. A.; Brooks, C. L., III. *Proteins* **1997**, *28*, 59.
(55) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *Biochemistry* **2003**, *42*, 13558.
(56) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(57) Verlet, L. *Phys. Rev.* **1967**, *159*, 98.
(58) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511.
(59) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.

Table 1. Comparison of the Model Reaction Energies (kcal/mol)[a]

	AM1	AM1-SRP	HF/1*	B3LYP/**	MP2/2*	G2	exp.
(1a)	9.1	9.1	35.0	21.6	6.7	9.6	
(1b)	−141.8	−141.8	−149.2	−140.2	−129.5	−133.8	
(1c)	44.2	44.2	32.8	33.1	46.4	43.9	
(1d)	−406.2	−406.2	−412.7	−398.9	−398.1	−403.3	−403.2
(1e)	−220.2	−220.2	−230.7	−225.6	−222.2	−225.6	−223.4
(2a)	−55.1	−45.5	−22.8		−28.0	−29.6	
(2b)	−6.7	1.0	−28.1		−3.4	−4.8	
(2)	−61.8	−44.6	−50.9		−31.4	−34.4	−47.8
(2c)	−410.8	−394.7	−398.9		−383.8	−389.8	−390.8

[a] 1* = 6-311+G(d,p), 2* = aug-cc-pVDZ, ** = aug-cc-pVDZ//HF. The experimental values for each reaction listed in this table were calculated based on the experimental values for molecules listed in Table 2.

Scheme 3. Model Reactions Considered for Peptide Hydrolysis Catalyzed by Cathepsin K
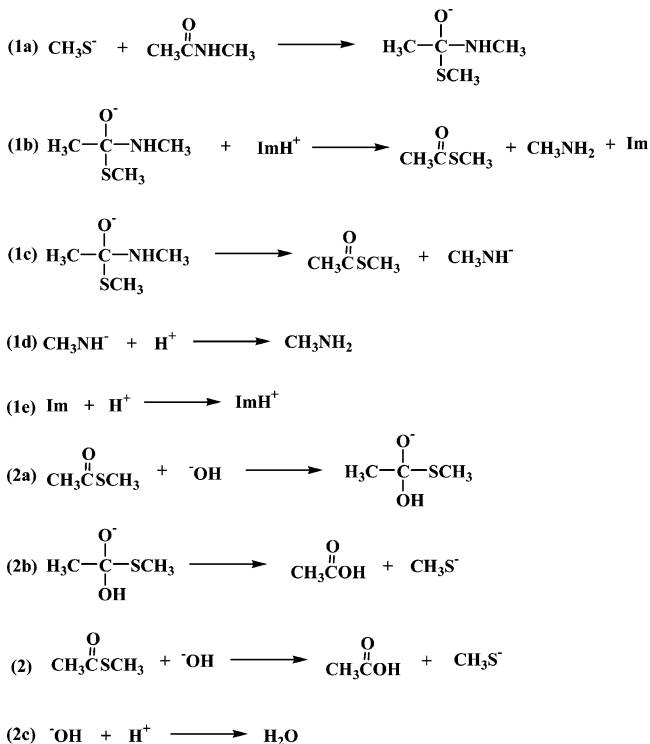
Table 2. Computed and Experimental Heats of Formation (kcal/mol)

	AM1	AM1-SRP	exp.[a]
H_2O	−59.3	−60.4	−57.8
^-OH	−14.1	−31.4	−32.7
CH_3COOH	−103.1	−103.1	−103.3
CH_3COSCH_3	−44.2	−44.2	−37
CH_3S^-	−17.0	−17.0	−14.3
Im	50.8	50.8	33.7
ImH^+	196.3	196.3	177
CH_3NH_2	−7.4	−7.4	−5.5
CH_3NH^-	33.0	33.0	32

[a] Experimental values in the last column were obtained from: (a) *NIST WebBook*; National Institute of Standards and Technology: Gaithersburg, MD, 1991. http://webbook.nist.gov/chemistry. (b) Lias, S.G *Gas-phase and Neutral Thermochemistry*; American Chemical Society: Washington, D.C.; 1988.

molecules. One water molecule is assigned to the QM region. Keeping the same volume of the system as that used in the acylation step, we sequentially minimized all QM atoms for 20 steps to relieve possible close contacts, and this is followed by an additional 25 ps of equilibration. Note that all calculations involving an energy minimization are only used to remove close contacts.

The reference reaction in aqueous solution was constructed similarly with an average dimension of 36.86617^3 Å^3 for the water box. The QM atoms shown in Scheme 2c and d are solute for the acylation and deacylation steps, respectively. Because the whole system is neutral, no counterion is needed.

2.4. Free Energy Surfaces with Two Reaction Coordinates. The acylation step of the hydrolysis is initiated by nucleophilic attack at the peptide carbonyl carbon (C_2) by Cys25 thiolate ion (S_γ), which is accompanied by proton transfer from His162 to the P1′ nitrogen (N_1). Mechanistically, these two processes can be either concerted or stepwise, and an understanding of the nature of these processes is a long-standing goal in mechanistic enzymology. The information is also of interest to drug design based on transition state or intermediate structures. This question can be resolved by constructing a Jencks− More O'Farrell two-dimensional (2D) free energy surface. The reaction coordinates are defined as follows (Scheme 2a and c):

$$RC1 = R(C_2 - N_1) - R(C_2 - S_\gamma) \quad (2a)$$

$$RC2 = R(N_{\delta 1} - H_{\delta 1}) - R(H_{\delta 1} - N_1) \quad (2b)$$

where $RC1$ is the reaction coordinate describing the nucleophilic addition−elimination (acylation) process of the amide bond and Cys25, and $RC2$ specifies the $H_{\delta 1}$ proton transfer from the donor atom $N_{\delta 1}$ of His162 to the amide nitrogen N_1.

The same mechanistic question on the deacylation step can be asked, and we construct another two-dimensional free energy surface to characterize these processes. In this case, the reaction coordinates are defined below (Scheme 2b and d):

$$RC3 = R(O_w - H_{\delta 1}) - R(H_{\delta 1} - N_{\delta 1}) \quad (3a)$$

$$RC4 = R(S_\gamma - C_2) - R(C_2 - O_w) \quad (3b)$$

where $RC3$ and $RC4$ correspond to the proton-transfer reaction from water to His162 and the nucleophilic attack of the carbonyl carbon by the water, respectively.

The free energy surfaces are determined by an adaptive umbrella sampling technique[60−63] in molecular dynamics simulations. The computation has been divided into multiple simulations by focusing on the two reaction coordinates within different regions with a harmonic biasing potential. To reduce computational costs and to increase sampling efficiency, we mainly follow the low free energy regions. Consequently, when free energies are 15 kcal/mol greater than that in the low free energy basin, we do not place simulation windows beyond these areas, and we have tried to ascertain that no stable intermediates are formed in the corners of the 2D free energy surface.

(60) Rajamani, R.; Naidoo, K. J.; Gao, J. *J. Comput. Chem.* **2003**, *24*, 1775.
(61) Bartels, C.; Schaefer, M.; Karplus, M. *Theor. Chem. Acc.* **1999**, *101*, 62.
(62) Hooft, R. W. W.; Van Eijck, B. P.; Kroon, J. *J. Chem. Phys.* **1992**, *97*, 6690.
(63) Kottalam, J.; Case, D. A. *J. Am. Chem. Soc.* **1988**, *110*, 7690.
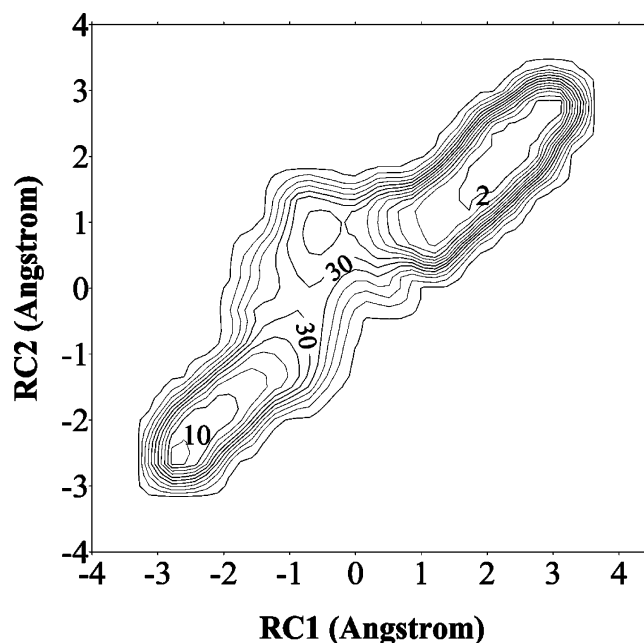
In the acylation step, we used a total of 200 windows for the enzyme reaction and 200 windows for that in aqueous solution. In the deacylation step, the number of simulation windows is 176 for the enzyme reaction and 199 for reference reaction in water. For each window, we performed at least 50 ps of equilibration, followed by an additional 50 ps for averaging and trajectory collection, in which 1000 structures were saved for further analysis. The weighted histogram analysis method (WHAM)[64,65] was used to obtain the probability density and free energy for the unbiased system. Each free energy profile was obtained from about 20 ns of total simulations.

2.5. Enzyme−Substrate Interaction Energy Decomposition. We have used an interaction energy decomposition analysis with the combined QM/MM potential to probe the relative energy contributions from each residue on the stabilization or destabilization of the transition state and product/intermediate state. This approach was first used by Bash et al. in the study of triosephosphate isomerase,[66] and it has been used by different groups on a number of enzyme systems.[52,55,67−70] Specifically, we determine QM/MM interaction energies for each given configuration by sequentially zeroing the atomic charges on each residue. The difference in QM electrostatic energy before and after the charge annihilation is defined as the interaction energy of this residue with the QM subsystem. Next, the interaction energies are respectively averaged over all of the structures saved at the reactant, transition, and product or intermediate states.[55] Reference 55 contains additional computational details.

3. Results and Discussion

3.1. The Reaction Mechanism of Cathepsin K. The overall cysteine protease reaction involves two reaction steps: the formation of a covalent enzyme-thiolester intermediate (acylation) and the hydrolysis of this intermediate (deacylation) to restore the enzyme cycle (Scheme 1). Both in the acylation and in the deacylation steps, the two events, the peptide (or thiolester) bond cleavage and the proton-transfer reaction, can be either concerted or stepwise, and we aim to elucidate the exact sequence of these processes in cathepsin K. An understanding of the reaction mechanism is also of interest for establishing the presence or absence of a distinct tetrahedral intermediate along the reaction pathway. We present the results for the acylation and deacylation steps separately below.

3.1.1. Two-Dimensional Free Energy Surfaces for the Acylation Step. The computed free energy surface for the acylation step in the hydrolysis of the peptide substrate catalyzed by cathepsin K is shown in Figure 2, and the corresponding uncatalyzed reaction in water is given in Figure 3. Comparison of the two reaction profiles shows that the mechanisms for the catalyzed and uncatalyzed reactions are remarkably similar. In the acylation step, the nucleophilic addition of Cys25 thiolate ion to the carbonyl group and the proton transfer from His162 to the amino nitrogen are highly correlated to yield a distinctive tetrahedral intermediate (TI). Strictly speaking, the cathepsin K reaction undergoes a stepwise mechanism; however, the overall nucleophilic addition and proton-transfer processes are
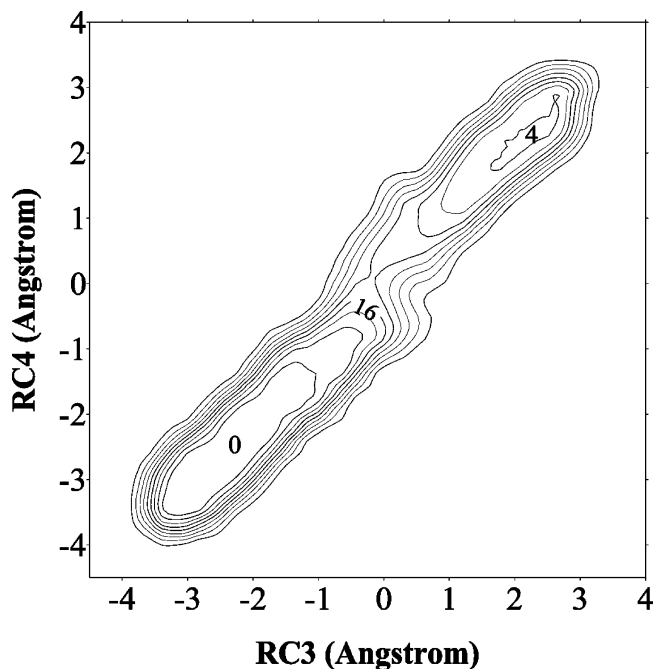
(64) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
(65) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339.
(66) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826.
(67) Davenport, R. C.; Bash, P. A.; Seaton, B. A.; Karplus, M.; Petsko, G. A.; Ringe, D. *Biochemistry* **1991**, *30*, 5821.
(68) Cunningham, M. A.; Ho, L. L.; Nguyen, D. T.; Gillilan, R. E.; Bash, P. A. *Biochemistry* **1997**, *36*, 4800.
(69) Chatfield, D. C.; Eurenius, K. P.; Brooks, B. R. *THEOCHEM* **1998**, *423*, 79.
(70) Dinner, A. R.; Blackburn, G. M.; Karplus, M. *Nature* **2001**, *413*, 752.

Figure 2. Computed free energy surface for the acylation reaction step catalyzed by cathepsin K. The reaction coordinates, *RC*1 (nucleophilic addition) and *RC*2 (proton transfer), are defined by eq 2 and are given in angstroms. Free energies are given in kcal/mol, and the contour levels are set at 4 kcal/mol. The contour plot was made to ensure that the tetrahedral intermediate is shown in 4 kcal/mol energy increment. Thus, the minimum energy of the figure is about −1 kcal/mol.
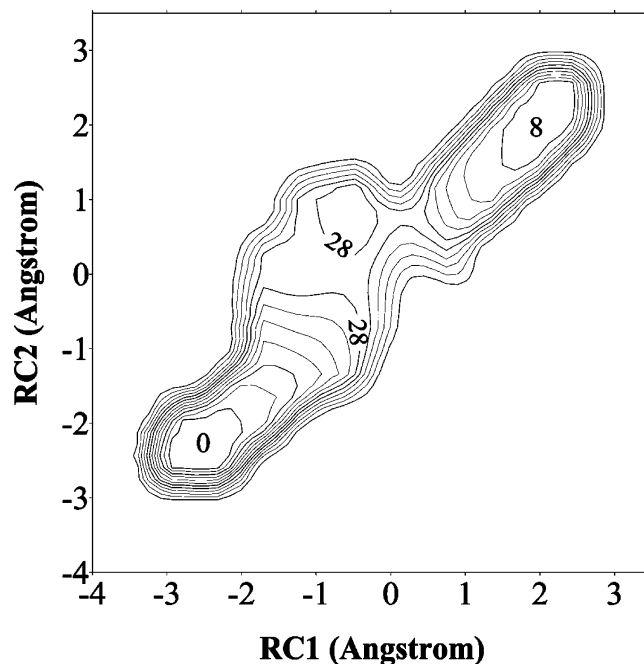
Figure 3. Computed free energy surface for the nucleophilic addition of methanethiolate ion and *N*-methyl acetamide assisted by an imidazolium ion in water. The reaction is a model system for the same process catalyzed by cathepsin K shown in Figure 2. The reaction coordinates, *RC*1 (nucleophilic addition) and *RC*2 (proton transfer), are defined by eq 2 and are given in angstroms. Free energies are given in kcal/mol, and the contour levels are set at 4 kcal/mol.

fully coupled. Elimination of the amino group from the TI has only a free energy barrier of about 2 kcal/mol to yield a covalent enzyme-thioester species. For the corresponding reaction in water, the barrier is slightly higher by 4 kcal/mol. In the active site, the TI structure is located at (−0.56, 0.78) in Figure 2,
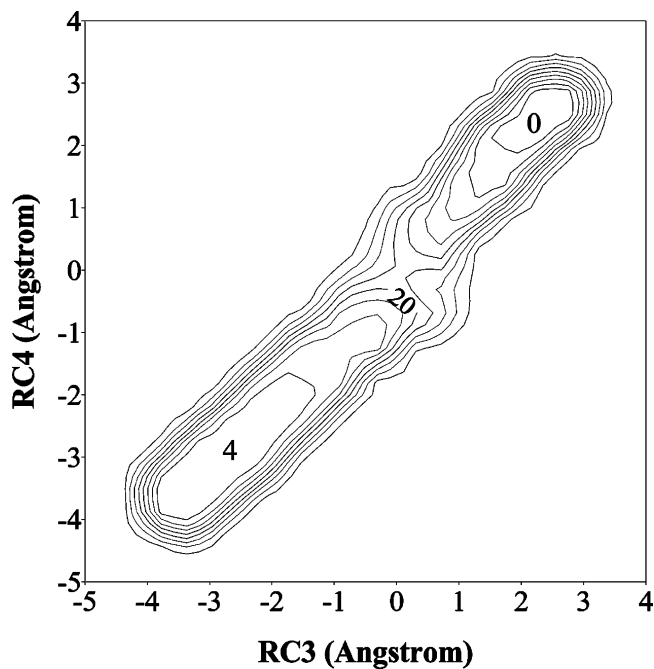
Figure 4. Computed free energy surface for the deacylation reaction step catalyzed by cathepsin K. The reaction coordinates, $RC3$ (proton transfer) and $RC4$ (nucleophilic addition), are defined by eq 3 and are given in angstroms. Free energies are given in kcal/mol, and the contour levels are set at 4 kcal/mol.

Figure 5. Computed free energy surface for the deacylation reaction step for the model reaction in aqueous solution. The reaction coordinates, $RC3$ (proton transfer) and $RC4$ (nucleophilic addition), are defined by eq 3 and are given in angstroms. Free energies are given in kcal/mol, and the contour levels are set at 4 kcal/mol.

where the first value specifies the reaction coordinate for the nucleophilic addition–elimination process and the second value denotes the proton-transfer reaction coordinate. Proton transfer is essentially completed when the TI is formed, which is formally a zwitterionic species that has the leaving group already protonated before departure. It may be anticipated that an inhibitor possessing a zwitterionic feature would be desirable to mimic the binding interactions in the active site.

In an earlier study, Kollman and co-workers[19] suggested that the proton transfer and nucleophilic attack could be concerted in aqueous solution. They have speculated a second possibility, involving a proton transfer prior to the thiolate nucleophilic attack.[22] The present results in Figures 2 and 3 do not support a separate proton transfer and nucleophilic step, but clearly proton transfer proceeds prior to nucleophilic addition. This is in agreement with the work by Strajbl et al.,[21] who also considered the pathway leading to a tetrahedral anion by keeping the proton on an imidazole.[21] Such an intermediate was not located in the present study.

3.1.2. Two-Dimensional Free Energy Surface for the Deacylation Step. The computed two-dimensional potentials of mean force for the deacylation step in cathepsin K and the corresponding reaction in water are depicted in Figures 4 and 5. Similar to the acylation step, the reaction mechanisms are remarkably similar between the enzymatic and the uncatalyzed processes. Unlike the acylation step, there is no tetrahedral intermediate in the deacylation reaction. Figures 4 and 5 show that the reaction is a synchronous concerted process with respect to the nucleophilic attack at the carbonyl group by water and the proton transfer from this water to His162.

The conclusion of a concerted mechanism in the deacylation step is in good accord with previous studies by Theodorou et al.[71] and by Strajbl et al.,[21] who examined model deacylation reactions of a methanethiolester by water, assisted by imidazole

or water as a general base, in aqueous solution. It was found that the mechanism for the reaction by a neutral nucleophile assisted by an imidazole base has concerted character with a barrier of 20 kcal/mol. In our simulation, the tetrahedral structure at the carbonyl center was found to be the transition state, and the mechanistic difference with the acylation step can be attributed to the fact that the thiolate ion is a much better leaving group than an amino group, which must be stabilized by a proton before its departure.

3.2. Quasiclassical Free Energy of Activation. The free energy surfaces presented above were obtained from classical molecular (CM) dynamics simulations, which we denote as $W_{CM}(RC1,RC2)$ and $W_{CM}(RC3,RC4)$. In these calculations, the nuclear quantum effects, in particular bound vibrational free energies, are not included. To compare the computational results with experiment, it is necessary to include nuclear QM contributions to the potential of mean force. Previously, we have utilized the ensemble-averaged variational transition state theory (EA-VTST) to incorporate both zero-point effects and tunneling in applications to several enzymatic reactions.[72–76] It was found that the dominant quantum effects are zero-point energy, which can reduce the classical free energy barriers by 1–3 kcal/mol.[72,75] We have used the EA-VTST method to make QM corrections to the classical free energies shown in Figures 2–5. Below, we summarize the computational procedure for including

(71) Theodorou, L. G.; Lymperopoulos, K.; Bieth, J. G.; Papamichael, E. M. *Biochemistry* **2001**, *40*, 3996.
(72) Garcia-Viloca, M.; Alhambra, C.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2001**, *114*, 9953.
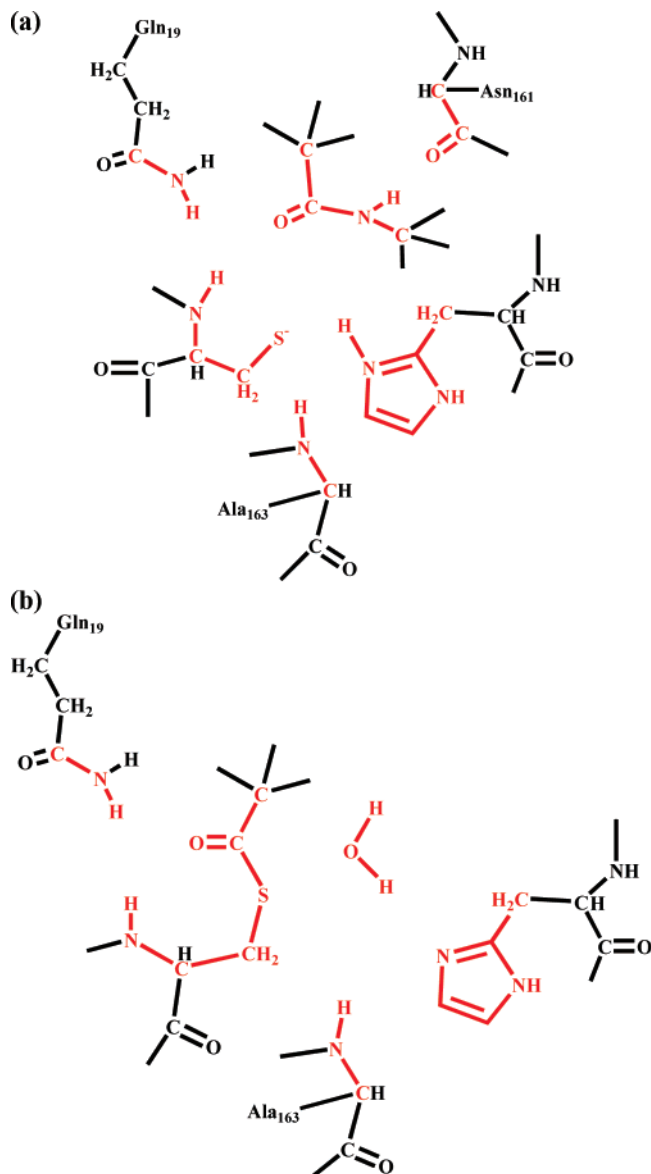(73) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186.
(74) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 3188.
(75) Pu, J.; Gao, J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 3140.
(76) Alhambra, C.; Corchado, J.; Sanchez, M. L.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2001**, *105*, 11326.

Scheme 4. Primary Zone Atoms (in Red) Used for the Zero-Point Energy Calculation (a) in the Acylation Step, and (b) in the Deacylation Step

(a)

(b)

Table 3. Computed Average Quantum Vibrational Free Energies (kcal/mol) to the Transition State and the Product State Relative to the Reactant State

	acylation		deacylation	
reaction step	cathepsin K	water	cathepsin K	water
reactant	0.0	0.0	0.0	0.0
transition state	−1.1	−1.4	−1.1	−1.1
product	0.9	0.5	−0.7	−0.4

reaction coordinate in Figures 2−5.[72] This is justified because the reaction coordinate motion involving proton transfer has a greater frequency than the heavy atom coordinate along the minimum free energy path. In all, a total of 398 configurations for the reactant state, 366 configurations for the transition states, and 1793 configurations for the product state were included in the acylation step. For the deacylation step, the numbers of configurations are, respectively, 1695, 295, and 1029. Finally, we determined the average frequencies as a function of the minimum free energy path, z, from which the quantum mechanical vibrational corrections are added to the classical free energies. The resulting free energy profile is called the quasiclassical (QC) potential of mean force because tunneling has been excluded. We expect that the quantitative contribution from tunneling on the activation barrier is small.[75]

The computed average quantum vibrational corrections for the cathepsin K catalyzed peptide hydrolysis reaction are listed in Table 3. It can be seen that in both acylation and deacylation steps of the enzyme reaction, QM vibrational energy reduces the free energy of activation by about 1 kcal/mol. On the other hand, it increases the free energy of the product (tetrahedral intermediate) for the acylation step by 0.9 kcal/mol, and decreases that for the deacylation product by 0.7 kcal/mol. For the reactions in aqueous solution, the quantum mechanical vibrational energy reduces the free energy barrier by about 1.4 kcal/mol for acylation step and 1.1 kcal/mol for deacylation step, whereas it increases the reaction energy by 0.5 kcal/mol for acylation step and decreases it by 0.4 kcal/mol for the deacylation reaction. Overall, quantum mechanical effects are similar for the reactions in water and in the enzyme, and there is little net quantum mechanical contribution to enzyme catalysis.

We summarize in Figure 6 the relative quasiclassical free energies for the critical points along the minimum free energy path of the peptide hydrolysis reaction catalyzed by cathepsin K, which are compared with those for the reactions in water. In Figure 6, we have included the results both for the acylation and for the deacylation steps, setting the free energy of the reactant state in the deacylation reaction to be identical to that of the product state in the acylation step. Here, we have neglected quantum effects from the reaction coordinate at the transition state, which are less than 0.3 kcal/mol.

Figure 6 shows that the acylation is the rate-limiting step for the overall peptide bond cleavage reaction, both for the reaction by cathpsin K and for the general acid assisted process in water. However, there is a small difference in the actual rate-limiting step. In cathepsin K, the formation of the tetrahedral intermediate has the highest barrier at 19.8 kcal/mol, whereas, in water, the rate-limiting step is the C−N bond cleavage from the TI with a barrier of 29.3 kcal/mol. The net reduction in the free energy barrier by cathepsin K is about 10 kcal/mol (or 14 kcal/mol relative to the reaction by a neutral thiol species by taking into

quantum mechanical vibrational correction ΔW_{vib} in the computed free energy profile.

In the EA-VTST approach,[76,77] we first partition the system into a primary and a secondary zone. Atoms in the primary zone have kinetic energy and are used to compute the instantaneous normal modes for structures saved during the simulations. The secondary zone serves as a bath to the primary zone in each configuration. The atoms in the primary zone for both the acylation and the deacylation reaction steps in cathepsin K are shown in Scheme 4a and b. For the reference reaction in water, all QM atoms are treated as the primary zone atoms. Next, we computed the normal modes for the primary zone atoms in each of these structures,[78] first projecting out the proton-transfer

(77) Garcia-Viloca, M.; Alhambra, C.; Corchado, J. C.; Sanchez, M. L.; Villa, J.; Gao, J.; Truhlar, D. G. *CHARMMRATE-version 2.0*; University of Minnesota, Minneapolis, 2002.

(78) Corchado, J. C.; Chuang, Y.-Y.; Fast, P. L.; Villa, J.; Hu, W.-P.; Liu, Y.-P.; Lynch, G. C.; Nguyen, K. A.; Jackels, C. F.; Melissas, V. S.; Lynch, B. J.; IRossi, I.; Coitino, E. L.; Fernandez-Ramos, A.; Pu, J.; Albu, T. V.; Steckler, R.; Garrett, B. C.; Isaacson, A. D.; Truhlar, D. G. *POLYRATE-version 9.0*; University of Minnesota, Minneapolis, 2002.
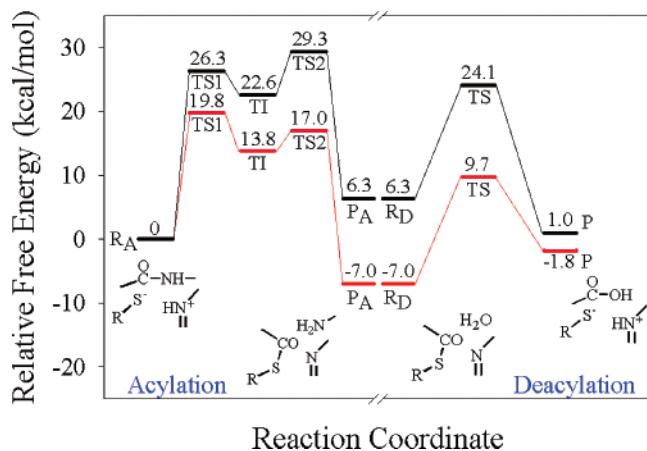
Figure 6. Relative quasiclassical free energies for the critical points along the minimum free energy path of the peptide hydrolysis reaction catalyzed by cathepsin K and the corresponding reference reactions in aqueous solution. The zero-point energy in TI and TS2 states was not calculated; they were lowered by the same magnitude as in TS1. Red, in enzyme; black, in aqueous solution.

account the pK_a of methanethiol in water). Experimentally, the rate constants for the acylation and deacylation steps in enzyme have been measured separately, and these two rate constants typically differ slightly for both amides and esters substrate.[71,79] The most relevant data are from Alves et al.,[28] who obtained a k_{cat} of 4.2 s^{-1} for the substrate Abz-K-(LRF)-SKQ-EDDnp in cathepsin K. This translates to a phenomenological free energy of activation of 16.6 kcal/mol. The computational results are in reasonable accord with experiment.

The tetrahedral intermediate is a stable species both in the enzymatic and in the aqueous reactions, with barriers of 2−7 kcal/mol to fall back into the reactant or forward into the product minima (Figure 6). The free energy of reaction in the acylation step is exoergonic by −7.0 kcal/mol in cathepsin K, but it is endoergonic (6.3 kcal/mol) in water. The deacylation of the thiolester species is fully concerted with proton transfer to the His162 general base. The free energy barriers are 16.7 kcal/mol in the enzyme and 17.8 kcal/mol in the corresponding reaction in water. In this reaction step, the enzyme environment increases the free energy of reaction by 5.2 kcal/mol, making the overall peptide bond cleavage about −2 kcal/mol exoergonic, while it is just 1 kcal/mol in water.

The enzymatic reaction by cathepsin K has not been studied computationally, but the closely related enzyme papain has been extensively investigated. In that case, both concerted and stepwise mechanisms for the nucleophilic addition and the proton-transfer process have been reported, implicating the complexity in characterizing precisely the reaction mechanism of cysteine proteases. Harrison et al.[23] obtained a barrier of 20 kcal/mol using combined AM1/MM energy optimizations. For model reactions, Warshel and co-workers[21] presented a most comprehensive analysis of a model amide cleavage reaction by a thiol group in water. These authors considered both the general base and the general acid assisted reactions as well as the direct nucleophilic attack by a thiolate ion. For the reaction corresponding to that depicted in Figure 3, Strajbl et al. obtained a free energy barrier of 24 kcal/mol,[21] which is about 5 kcal/mol lower than the present result from free energy simulations.
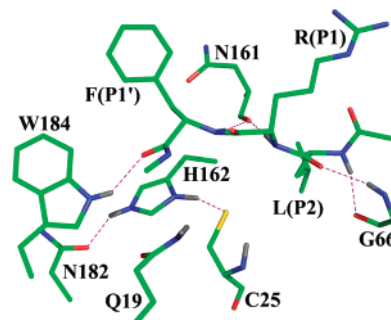
Figure 7. Substrate binding pocket of human cathepsin K at Michaelis state.

Nevertheless, the reasonable agreement between the present results and that of Strajbl et al.[21] further demonstrates that the combined QM/MM model is adequate for treating the cathepsin K reactions.

There are no direct experimental data for comparison with the general acid/base assisted processes (acylation and deacylation) in water. Radzicka et al.[80] reported an activation barrier of 23.8 kcal/mol for the uncatalyzed peptide hydrolysis in neutral aqueous solution. Strajbl et al.[21] analyzed the experimental data by Keillor et al.[81] and suggested that the free energy of activation for the acylation step is about 25 kcal/mol for the nucleophilic attack of 2-mercaptomethyl-1-methylimidazole on formide. Based on the experimental results from Hershfield[82] and Guthrie,[83] the free energy of activation for deacylation step in aqueous solution was estimated to be 25.4 kcal/mol with H_2O as a nucleophile.[21] These results are consistent with our simulation for the aqueous reactions.

3.3. Effects of Enzyme−Substrate Interactions on the Acylation Reaction Step. In the next two sections, we analyze the roles of individual amino acids on the two reaction steps in cathepsin K catalysis by considering the change in hydrogen-bonding interactions and differential stabilization between the transition state and the reactant state (Figures 8 and 9). We have grouped these interactions into sub-categories for convenience of discussion: (a) interactions involved in substrate binding, (b) hydrogen bonds that stabilize the Cys25 nucleophile and the catalytic triad, and (c) distant interactions that have a large impact on transition state stabilization.

To make the analysis tractable, we focus on the three pivotal regions along the reaction pathway for the acylation and deacylation steps, that is, the reactant state (RS), the transition state (TS), and the product state (PS). Based on the free energies in Figures 2 and 4, we have made the following definitions for structural analyses: in the acylation step, the RS (Michaelis complex) structures are defined by the reaction coordinates in the range of $RC1 = -2.70 \pm 0.2$ Å and $RC2 = -2.50 \pm 0.2$ Å, the TS structures by $RC1 = -0.72 \pm 0.1$ Å and $RC2 = -0.14 \pm 0.1$ Å, and the PS structures by $RC1 = 1.50-3.1$ Å and $RC2 = 1.0-2.8$ Å (Figure 2). By these criteria, there are 398 RS structures, 366 TS structures, and 1793 PS structures. Similarly, in the deacylation step, the RS structures are defined by $RC1 = -2.70$ to -1.50 Å and $RC2 = -3.00$ to -1.80 Å; the TS region by $RC1 = -0.05 \pm 0.1$ Å and $RC2 = 0.08 \pm$

(79) Whitaker, J. R.; Bender, M. L. *J. Am. Chem. Soc.* **1965**, *87*, 2728.

(80) Radzicka, A.; Wolfenden, R. *J. Am. Chem. Soc.* **1996**, *118*, 6105.
(81) Keillor, J. W.; Neverov, A. A.; Brown, R. S. *J. Am. Chem. Soc.* **1994**, *116*, 4669.
(82) Hershfield, R.; Schmir, G. L. *J. Am. Chem. Soc.* **1973**, *95*, 3994.
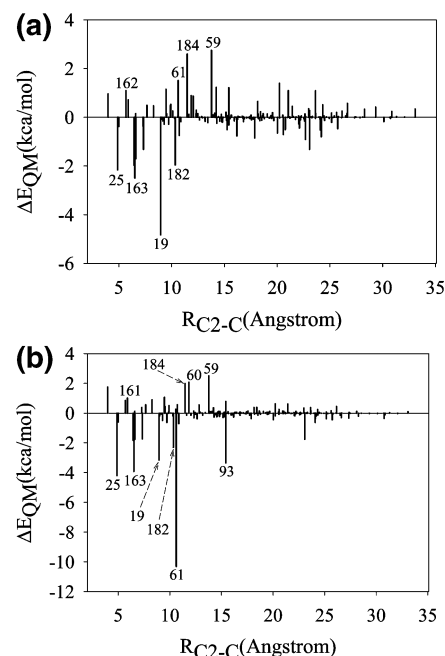(83) Guthrie, J. P. *J. Am. Chem. Soc.* **1978**, *100*, 5892.
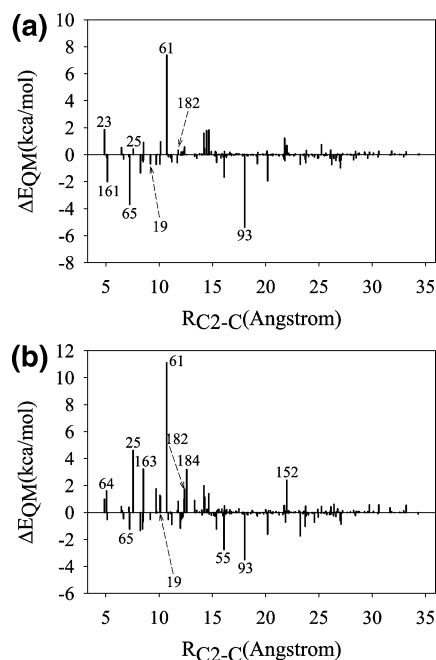
Figure 8. (a) The residual contribution of human cathepsin K to transition state stabilization or destabilization for acylation step. The labeled ones include residues around active site and those which result in ΔE_{QM} more than 2.0 kcal/mol. R_{C2-C} is the distance between atom C_2 on substrate and carbonyl carbon (C) of each residue. (b) The residual contribution of human cathepsin K to the reaction energy for acylation step. The labeled ones include residues around active site and those which result in ΔE_{QM} more than 2.0 kcal/mol. R_{C2-C} is the distance between atom C_2 on substrate and carbonyl carbon (C) of each residue.

Figure 9. (a) The residual contribution of human cathepsin K to the free energy of activation for deacylation step. The labeled ones include residues around active site and those which result in ΔE_{QM} more than 2.0 kcal/mol. R_{C2-C} is the distance between atom C_2 on substrate and carbonyl carbon (C) of each residue. (b) The residual contribution of human cathepsin K to the reaction energy for deacylation step. The labeled ones include residues around active site and those which result in ΔE_{QM} more than 2.0 kcal/mol. R_{C2-C} is the distance between atom C_2 on substrate and carbonyl carbon (C) of each residue.

0.1 Å; and the PS structures by $RC1 = 1.40-2.60$ Å and $RC2 = 1.60-2.80$ Å (Figure 4). Here, we collected 1695 structures for the RS, 295 for the TS, and 1029 structures for the PS in the deacylation step. All quantities reported below have been averaged over these structures.

3.3.1. Substrate Binding. Key residues involved in substrate binding include Gln19, Gly66, Asn161, and Trp184. The average hydrogen-bonding distances are listed in Tables 4 and 5, while Figure 7 illustrates the binding pocket of the cathepsin K active site, which accommodates the tri-peptide substrate, Ace-Leu(P2)-Arg(P1)-Phe(P1′)-NMe. In examining the change of hydrogen-bonding interactions along the reaction pathway, it is useful to keep in mind that the fragment of P2−P1 residues is covalently linked to Cys25 after the acylation step, and the P1′ fragment will dissociate from the active site before the next, deacylation step occurs. Thus, the P2−P1 region is relatively immobile, whereas the leaving group shows greater flexibility.

Figure 7 and Table 4 show that the backbone of the P2 residue (Leu) and the S2 residue (Gly66) are fully matched by hydrogen-bonding interactions at distances of 2.9 and 3.2 Å. These interactions exhibit little variations along the entire reaction path in going from the Michaelis complex (RS) to the thiolester covalent intermediate (PS) in the acylation step. The backbone carbonyl group of Asn161 accepts two hydrogen bonds from the substrate, one from the amino group of Arg(P1) at 2.8 Å and another from that of Phe(P1′) at 3.2 Å in the Michaelis complex. The O(Asn161)−N(Phe) interaction in the thiolester product state is elongated to 5.5 Å as a result of the amide bond cleavage. Energy decomposition analyses show that Asn161 stabilizes both transition states and the TI by about 2 kcal/mol

in the acylation reaction (Figure 8). Mutations to Gly and Ala of the corresponding Asn residue in papain showed reduced activities by 33- and 240-folds.[84]

The carbonyl group of the P1′ residue is hydrogen bonded with the side chain of Trp184 at a distance of 3.3 Å, which is decreased to 3.0 Å at the transition states before losing its interaction in the PS. Trp184 does not have major effects on transition state stabilization; its interaction is weakened by 4 kcal/mol in the product state (Figure 8).

3.3.2. Stabilization of the Nucleophile Cys25 Thiolate Anion. Cysteine proteases follow a common mechanism that utilizes a thiol group in the nucleophilic addition−elimination reaction. The catalytic cysteine is invariably coupled to a histidine residue, which in turn is hydrogen bonded to a strictly conserved asparagine residue in the papain family. In cathepsin K, this catalytic triad consists of Cys25, His162, and Asn182. Experimental evidence and computational results have shown that the high nucleophilicity of the thiol group results from the fact that Cys25 is present as a thiolate anion at the optimal pH range for the enzyme activity, and it is stabilized by a thiolate−imidazolium ion pair with His162. We have calculated the potential of mean force for the proton-transfer reaction between Cys25 and His162 (to be published). It shows that in apo cathepsin K, the zwitterionic state of Cys25 and His162 is slightly more stable than the neutral state of these two residues.

In the Michaelis complex, the thiolate anion is 2.8 Å from the N_δ atom of the imidazolium side chain of His162. As the

---

(84) Menard, R.; Khouri, H. E.; Plouffe, C.; Dupras, R.; Ripoll, D.; Vernet, T.; Tessier, D. C.; Lalberte, F.; Thomas, D. Y.; Storer, A. C. *Biochemistry* **1990**, *29*, 6706.

Table 4. Selected Average Interaction Distances r_{DA} (Å) between Donor (D) and Acceptor (A) Atoms of the Ace-Leu-Arg-Phe-NMe Substrate and Residues in the Active Site of Cathepsin K at the Key Points (Reactant, Transition States, Tetrahedral Intermediate, and Product States) along the Acylation Reaction Pathway

| interaction pairs						
D	A	reactant	TS1	TI	TS2	product
Substrate−Enzyme Interactions						
N, Leu(P2)	O, Gly66	2.9	2.9	2.9	3.0	3.0
N, Gly65	O, Leu(P2)	3.4	4.5	4.5	4.6	4.6
N, Gly66	O, Leu(P2)	3.2	3.1	3.2	3.1	3.1
N, Arg(P1)	O, Asn161	2.8	3.0	3.0	3.0	3.1
NE2, Gln19	O, Arg(P1)	5.2	2.9	2.9	2.8	3.0
N, Cys25	O, Arg(P1)	6.2	3.9	3.8	3.6	3.5
N, Phe(P1′)	O, Asn161	3.2	3.4	3.5	3.4	5.5
NE2, Gln19	O, Phe(P1′)	4.0	3.8	3.7	3.8	4.3
NE1, Trp184	O, Phe(P1′)	3.3	3.0	3.1	3.0	4.2
N, terminal	O, Gln21	5.6	6.5	6.5	6.3	6.7
N, Gly64	O, Arg(P1)	4. 7	8.0	7.8	7.9	7.7
N, Gly65	O, Arg(P1)	5.4	6.8	6.8	6.9	6.7
NE, Arg(P1)	O, Gly64	4.0	3.6	3.4	3.8	3.9
NH1, Arg(P1)	O, Gly64	4.5	3.5	4.4	3.4	5.2
NH2, Arg(P1)	O, Gly64	5.5	4.3	4.2	4.6	4.8
Catalytic Triad						
NE2, Gln19	SG, Cys25	3.9	4.7	4.8	4.8	4.8
N, Cys25	SG, Cys25	3.2	3.3	3.3	3.3	3.4
N, Trp26	SG, Cys25	4.0	3.6	3.6	3.6	3.6
N, Ala163	SG, Cys25	3.7	3.5	3.6	3.7	3.7
ND1, His162	SG, Cys25	2.9	3.3	3.6	3.6	3.7
NE2, NE2	OD1, Asn182	2.9	2.9	2.9	2.9	2.9
Reaction Coordinates						
C2, QM	N1, QM	1.4	1.5	1.6	1.9	3.9
C2, QM	SG, QM	4.1	2.2	2.1	1.9	1.7
$RC1 = R(C2 - N1)$ $- R(C2 - SG)$		−2.7	−0.7	−0.5	0.0	2.2
ND1, QM	HN1, QM	1.1	1.2	1.8	1.8	2.9
HD1, QM	N1, QM	3.6	1.4	1.1	1.0	1.0
$RC2 = R(ND1 - HD1)$ $- R(HD1 - N1)$		−2.5	−0.2	0.7	0.8	1.9

Table 5. Selected Average Interaction Distances r_{DA} (Å) between Donor (D) and Acceptor (A) Atoms of the Substrate and Residues in the Active Site of Cathepsin K at the Key Points (Reactant, Transition State, and Product) along the Deacylation Reaction Pathway

interaction pairs				
D	A	reactant	TS	product
Substrate−Enzyme Interactions				
N, Leu(P2)	O, Gly66	2.9	2.9	2.9
N, Gly65	O, Leu(P2)	4.5	4.5	4.4
N, Gly66	O, Leu(P2)	3.1	3.1	3.3
N, Arg(P1)	O, Asn161	3.1	3.1	3.0
NE2, Gln19	O, Arg(P1)	3.1	2.8	4.1
N, Cys25	O, Arg(P1)	3.6	3.7	5.7
OH2, QM	O, Asn161	6.8	3.7	3.5
N, Gly64	O, Arg(P1)	7.5	7.9	7.5
N, Gly65	O, Arg(P1)	6.5	6.8	6.6
NE, Arg(P1)	O, Gly64	4.3	3.9	3.7
NH1, Arg(P1)	O, Gly64	5.7	4.5	4.2
NH2, Arg(P1)	O, Gly64	5.4	4.9	4.7
Catalytic Triad				
NE2, Gln19	SG, Cys25	5.0	4.8	4.1
N, Cys25	SG, Cys25	3.4	3.3	3.2
N, Trp26	SG, Cys25	3.6	3.5	3.8
N, Ala163	SG, Cys25	3.7	3.6	3.7
ND1, His162	SG, Cys25	3.8	3.5	2.9
NE2, NE2	OD1, Asn182	3.0	3.0	2.9
Reaction Coordinates				
OH2, QM	HD1, QM	1.0	1.2	3.1
HD1, QM	ND1, QM	3.4	1.3	1.2
$RC3 = R(OH2 - HD1)$ $- R(HD1 - ND1)$		−2.4	−0.1	1.9
SG, QM	C2, QM	1.7	1.8	3.5
OH2, QM	C2, QM	4.4	1.7	1.4
$RC4 = R(SG - C2)$ $- R(OH2 - C2)$		−2.7	0.1	2.1

nucleophilic addition takes place, the histidine side chain is tilted by 20° to interact with the amino nitrogen of the P1′ residue, which is accompanied by proton transfer from His162 to the P1′-amino group. The present results suggest that His162 has two functional roles in the acylation reaction step: the first is to provide stabilization of the thiolate anion in the catalytic triad to enhance nucleophilicity of Cys25, and the second role is to act as a general acid to stabilize the tetrahedral intermediate and to assist the subsequent elimination of the amino group.

3.3.3. Oxyanion Hole and Transition State Stabilization. A characteristic feature of the reaction mechanism in serine proteases is the stabilization of the transition state and tetrahedral intermediate by an oxyanion hole through hydrogen-bonding interactions. The question on the existence of a similar oxyanion hole in cysteine proteases has been debated, and experimental evidence has been inconclusive. The side chain of Gln19 is hydrogen bonded to the thiolate anion of Cys25 in the Michaelis complex at an average distance of 3.9 Å between $N_{\epsilon 2}$ and S_{γ} atoms. This interaction is broken as the thiolate group approaches the P1 carbonyl group, and Gln19 becomes hydrogen bonded to the oxygen anion at an average distance of 2.8−2.9 Å, significantly shorter than the average distance (5.2 Å) in the Michaelis complex. A second major structural change of the active site, accompanying the nucleophilic addition, is the average distance between the backbone of Cys25 and the P1

carbonyl oxygen, moving from an average distance of 6.2 Å in the Michaelis complex to 3.5−3.9 Å as the thiolester enzyme−substrate covalent complex is formed. Thus, Cys25 and Gln19 donate two hydrogen bonds to the oxygen anion of the nucleophilic addition reaction, which forms an oxyanion hole that stabilizes the transition states and TI. The energy decomposition method cannot provide a clear picture for assessing the energy contributions because these amide groups interact both with the thiolate anion in the reactant state and with the oxyanion in the transition state, resulting in a relatively small energy change of only −0.7 kcal/mol from the Michaelis complex to the transition state.

These structural changes are depicted in Figure 10, which superimposes the Michaelis complex and the tetrahedral intermediate by overlapping the backbone atoms (residues 20−180). The enzyme itself only undergoes minor adjustments, but the backbone atoms of the substrate move as large as 2.0 Å (C_2) and 2.6 Å (N_1) toward the nucleophile thiolate. Figure 10 shows that the oxyanion hole is already defined in the Michaelis complex state. However, there are no apparent interactions with the carbonyl group in the Michaelis complex because the Gln19 is bonded to the Cys25 anion and the substrate amide is 2 Å away. Thus, binding interactions of the oxyanion with Gln19 and the backbone of Cys25 in cathepsin K are created by structural changes accompanying the nucleophilic addition. This structural change gives rise to a transition state-enforced stabilization.[87]

Asn182 maintains hydrogen-bonding interactions with His162 at an average distance of about 2.9 Å throughout the acylation
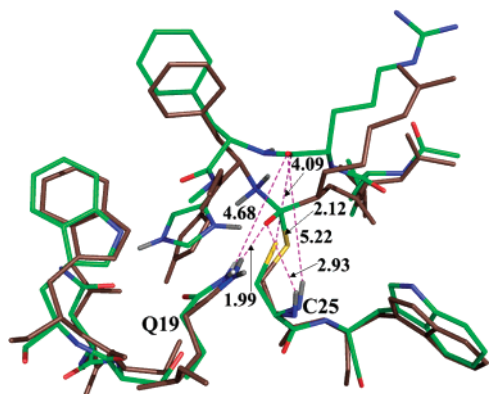
Figure 10. Superimposition of the active site of cathepsin K and substrate at Michaelis and tetrahedral intermediate states. The ochre is tetrahedral intermediate state structure.

step. Surprisingly, there is only a small effect of -0.5 kcal/mol from Asn182 in transition state stabilization based on energy decomposition analysis. The effect is somewhat greater in the deacylation step (by -2.0 kcal/mol). Thus, computational results do not point to major energetic contributions to catalysis from Asn182, which is somewhat surprising in view of the fact that it is strictly conserved across species. However, this is in accord with mutation results on the papain: mutations of the equivalent Asn residue to Gln and to Ala mutations only decrease the second-order rate constant, k_{cat}/K_M, by a factor of 3.4 and 150,[85] respectively.

3.3.4. Nonspecific Interactions. A number of residues are found to have rather large energy contributions to transition state stabilization or destabilization (Figure 8). Gly65 interacts with the transition state more favorably than it does with the Michaelis complex by 3.7 kcal/mol (Figure 8), although this difference mainly originates from the movement of the Arg side chain of the substrate (a shortening of the average distance by about 0.5 Å at a value of about 7.5 Å). To the opposite effect is Asp61, which does not directly form hydrogen bond with Arg(P1); the average distance is about $7-8$ Å, but it contributes 7.5 kcal/mol based on Figure 8a destabilizing interactions in going from the Michaelis complex to the transition state. Note that the energy decomposition analysis presented here does not include the effect of solvent/enzyme electrostatic screening and the energy change resulting from interactions with other residues in the MM region. Thus, it is not an indication that mutation of Asp61 to a neutral residue could lead to transition state stabilization. In fact, a number of negatively charged residues (Glu35, Asp55, Glu93, and Glu94) show stabilizing interactions at the transition state, mainly due to shortening of average distances with the P1-Arg side chain, although they are far away from the active site.

3.4. Effects of Enzyme−Substrate Interactions on the Deacylation Reaction Step. Many features of the hydrogen-bonding interactions in the active site in the deacylation step are similar to those found in the acylation step. Thus, we briefly mention the most significant variations. In the deacylation step, the change in electrostatic character in the active site is reversed in comparison with the acylation step. Now, the neutral enzyme−substrate covalent intermediate plus a water molecule and a neutral His162 are separated into an ion pair between the

Cys25 thiolate leaving group and a protonated His162 (from the nucleophile water). Consequently, some of the effects observed for the acylation step are reversed here, particularly those distant anionic residues, such as Asp61 (Figures 8 and 9), which shows a large stabilizing contribution to the product formation as opposed to a destabilizing effect in the acylation step. The average distances between the side chain of the substrate Arg(P1) residue and backbone carbonyl groups of Gly64 and Gly65 decrease by $0.7-1.6$ Å along the deacylation reaction pathway, leading to stabilizing contributions.

The oxyanion hole in the deacylation step is again well-preserved with large energy contributions from Gln19 and Cys25 to stabilize the transition state. Residues Trp26 and Ala163, which are found to stabilize the Cys25 thiolate anion in the Michaelis complex of the acylation step, contribute to the stabilization of the transition state and the product state in the deacylation step. Trp184, which forms a hydrogen bond with the P1′ residue in the acylation step, participate in a hydrogen bond with the nucleophilic water molecule in the deacylation step. As the nucelophilic attack at the carbonyl center of the thiolester occurs, the distance between Trp184 increases from 3.2 Å in the reactant state to 6.0 Å in the transition state. Thus, Trp184 is found to destabilize the deacylation transition state by 2.6 kcal/mol (Figure 9a).

Mutation of Gln19 to Ala caused a 60-fold decrease in k_{cat}/K_M for hydrolysis of CBZ-Phe-Arg-MCA by papain, and 2.4 kcal/mol increase of the free energy of activation, the mutant Gln19Ser leads to a decrease of k_{cat}/K_M by 600-fold, and the increase of 3.8 kcal/mol of the free energy of activation was found.[86] Our result is consistent with the mutated result in that Gln19 does not reduce much of the barrier height for acylation step, but reduces 5 kcal/mol of barrier height for deacylation step (Figure 9). From Figures 2 and 4, the barrier height for acylation step of wild-type cathepsin K is 2 kcal/mol higher than that of deacylation step, which means that mutation of Gln19 will lead to deacylation step as rate-limiting, whose free energy of activation is about 3 kcal/mol higher than that of wild type. It would be interesting to test this proposal by experiment.

4. Conclusions

We have carried out molecular dynamics simulations on both acylation and deacylation steps along the reaction pathway of the hydrolysis of a peptide substrate catalyzed by human cathepsin K, as well as the corresponding model reactions in aqueous solution. The two-dimensional potentials of mean forces show that in both enzyme human cathepsin K and aqueous solution, the hydrolysis reaction shares the same mechanism; that is, the acylation step is rate-limiting with a barrier height 19.8 kcal/mol in human cathepsin K and 29.3 kcal/mol in aqueous solution. The acylation step is a combination of the concerted and stepwise process; that is, the nucleophilic attack from Cys25 thiolate to carbonyl carbon C_2 is concerted with the proton-transfer reaction from side chain of His162 to N_1 on substrate, whereas the amide bond cleavage itself is stepwise with the formation of a tetrahedral intermediate. The free energy of activation for the proton-transfer reaction is 3 kcal/mol higher

(85) Vernet, T.; Tessier, D. C.; Chatellier, J.; Plouffe, C.; Lee, T. S.; Thomas, D. Y.; Storer, A. C.; Menard, R. *J. Biol. Chem.* **1995**, *270*, 16645.

(86) Menard, R.; Carriere, J.; Laflamme, P.; Plouffe, C.; Khouri, H. E.; Vernet, T.; Tessier, D. C.; Thomas, D. Y.; Storer, A. C. *Biochemistry* **1991**, *30*, 8924.
(87) Alhambra, C.; Wu, L.; Zhang, Z.-Y.; Gao, J. *J. Am. Chem. Soc.* **1998**, *120*, 3858.

than the C_2-N_1 bond breaking reaction. In the deacylation step, the proton-transfer reaction from H_2O to the imidazole ring of His162 is fully concerted with the nucleophilic attack from O of H_2O to the carbonyl carbon C_2 at a free energy of activation 16.7 kcal/mol in enzyme and 17.8 kcal/mol in aqueous solution. Our calculated results demonstrate that quantum mechanical correction to the vibrational free energy along the reaction path decreases the barrier height by 1 kcal/mol for both acylation and deacylation steps in enzyme. Comparing the free energy profiles obtained for the catalyzed reaction by human cathepsin K and the uncatalyzed one in aqueous solution, we conclude that the enzyme catalyzes the hydrolysis of peptide substrate reaction by reducing the free energy of activation for acylation step by 10 kcal/mol. This is achieved by stabilization of the oxyanion from two amide hydrogen bonds. Structural analyses suggest that, although the basic structural elements of the oxyanion hole are already present in the Michaelis complex, hydrogen-bonding interactions with the carbonyl oxygen (i.e., the oxyanion) from Gln19 and the backbone of Cys25 are induced by structural changes in the formation of the sulfur—carbon bond. Analyses of energy components and hydrogen-bond distances between the QM region and MM enzymatic atoms show that both residues in the active site and those far way affect transition state stabilization.

Supporting Information Available: Summary of the development, results, and validation of the combined QM/MM potential energy function for cysteine protease reactions, and the complete citations for refs 26 and 49. This material is available free of charge via the Internet at http://pubs.acs.org.

JA074222+