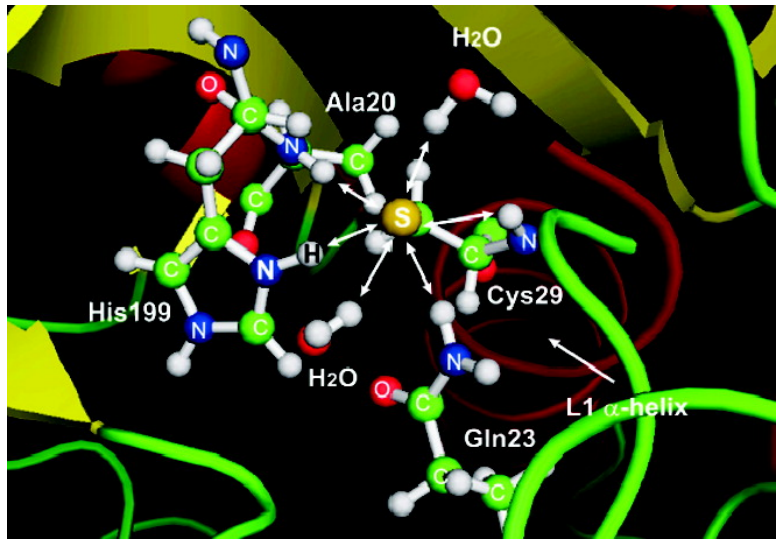
Article

On the Origin of the Stabilization of the Zwitterionic Resting State of Cysteine Proteases: A Theoretical Study

Milena Mladenovic, Reinhold F. Fink, Walter Thiel, Tanja Schirmeister, and Bernd Engels

J. Am. Chem. Soc., **2008**, 130 (27), 8696-8705 • DOI: 10.1021/ja711043x • Publication Date (Web): 17 June 2008



More About This Article

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

- Supporting Information
- Links to the 1 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

J|A|C|S
ARTICLES

Published on Web 06/17/2008

On the Origin of the Stabilization of the Zwitterionic Resting State of Cysteine Proteases: A Theoretical Study

Milena Mladenovic,[†] Reinhold F. Fink,[†] Walter Thiel,[‡] Tanja Schirmeister,[§] and Bernd Engels*,[†]

Institut für Organische Chemie, Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany, Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470 Mülheim an der Ruhr, Germany, and Institut für Pharmazie and Lebensmittelchemie, Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany

Received December 20, 2007; E-mail: bernd@chemie.uni-wuerzburg.de

Abstract: Papain-like cysteine proteases are ubiquitous proteolytic enzymes. The protonated His199/deprotonated Cys29 ion pair (cathepsin B numbering) in the active site is essential for their proper functioning. The presence of this ion pair stands in contrast to the corresponding intrinsic residue pK_a values, indicating a strong influence of the enzyme environment. In the present work we show by molecular dynamics simulations on quantum mechanical/molecular mechanical (QM/MM) potentials that the ion pair is stabilized by a complex hydrogen bond network which comprises several amino acids situated in the active site of the enzyme and 2−4 water molecules. QM/MM reaction path computations for the proton transfer from His199 to the thiolate of the Cys29 moiety indicate that the ion pair is about 32−36 kJ mol^{-1} more stable than the neutral form if the whole hydrogen bonding network is active. Without any hydrogen bonding network the ion pair is predicted to be significantly less stable than the neutral form. QM/MM charge deletion analysis and QM model calculations are used to quantify the stabilizing effect of the active-site residues and the L1 helix in favor of the zwitterionic form. The active-site water molecules contribute about 30 kJ mol^{-1} to the overall stabilization. Disruption of the hydrogen bonding network upon substrate binding is expected to enhance the nucleophilic reactivity of the thiolate.

1. Introduction

Papain-like cysteine proteases are processive and digestive enzymes found in a wide variety of living organisms, from bacteria to humans.[1] For many years these enzymes were believed to be responsible for redundant "house-keeping" activities and as such were considered to have little, if any, value as drug targets. However, recent progress in identifying new members of this class, as well as a better understanding of their physiological roles, has altered this perception. It has become clear that human lysosomal cysteine proteases fulfill specific functions in extracellular matrix turnover, antigen presentation, and processing events. As such they are involved in many diseases and represent promising drug targets for osteoporosis,[2] arthritis,[3] cancer,[4] and Alzheimer's disease.[2] In addition, protozoan analogues of these proteases are now known to be involved in a wide variety of parasitic infections such as malaria[5] and African trypanosomiasis ("sleeping sickness").[2]

Proteolytic activity of cysteine proteases depends on the thiol group of a cysteine residue situated in the active site. In the active enzyme an ion pair is established between this cysteine residue and a nearby histidine. Figure 1 shows the active site of the cysteine protease cathepsin B with the His199H$^+$/Cys29S$^-$ ion pair. The existence of this ion pair is no longer disputed. In the late 1970s and early 1980s, its presence was experimentally proven by potentiometric difference titrations,[6] isotope effect measurements,[7] and proton nuclear magnetic resonance spectroscopy.[8,9] Kinetic studies on different thiol proteases showed that there are two ionizable groups involved in the catalysis, one with a pK_a value of around 4 and another with a pK_a around 8.5. The former is usually attributed to the deprotonation of the Cys29 residue, whereas the latter is believed to reflect the ionization of His199. In contrast, the intrinsic pK_a values of cysteine and histidine residues are about 8−9 and 6−7, respectively. The observed large pK_a shifts prove the strong influence of the enzyme environment and show that the situation in the active site can only be explained if the local enzyme environment is accounted for in detail.

[†] Institut für Organische Chemie, Universität Würzburg.
[‡] Max-Planck-Institut für Kohlenforschung.
[§] Institut für Pharmazie and Lebensmittelchemie, Universität Würzburg.

(1) Otto, H.-H.; Schirmeister, T. *Chem. Rev.* **1997**, *97*, 133.
(2) Lecaille, F.; Kaleta, J.; Brömme, D. *Chem. Rev.* **2002**, *102*, 4459.
(3) Huet, G.; Flip, R. M.; Richet, C.; Thiebet, C.; Demeyer, D.; Balduyck, M.; Duquesnoy, B.; Degand, P. *Clin. Chem.* **1992**, *38*, 1694.
(4) Sloane, B. F.; Moin, K.; Krepela, E.; Rozhin, J. *Cancer Methastasis Rev.* **1990**, *9*, 333.
(5) Rosenthal, P. J.; McKerrow, J. H.; Aikawa, M.; Nagasawa, H.; Leech, J. J.; Malarial, A. *J. Clin. Invest.* **1988**, *82*, 1560.
(6) Lewis, S. D.; Johnson, F. A.; Shafer, J. A. *Biochemistry* **1976**, *82*, 5009.
(7) Creighton, D. J.; Gessouroun, M. S.; Heapes, J. M. *FEBS Lett.* **1980**, *110*, 319.
(8) Lewis, S. D.; Johnson, F. A.; Shafer, J. A. *Biochemistry* **1981**, *20*, 48.
(9) Johnson, F. A.; Lewis, S. D.; Shafer, J. A. *Biochemistry* **1981**, *20*, 44.

 10.1021/ja711043x CCC: $40.75

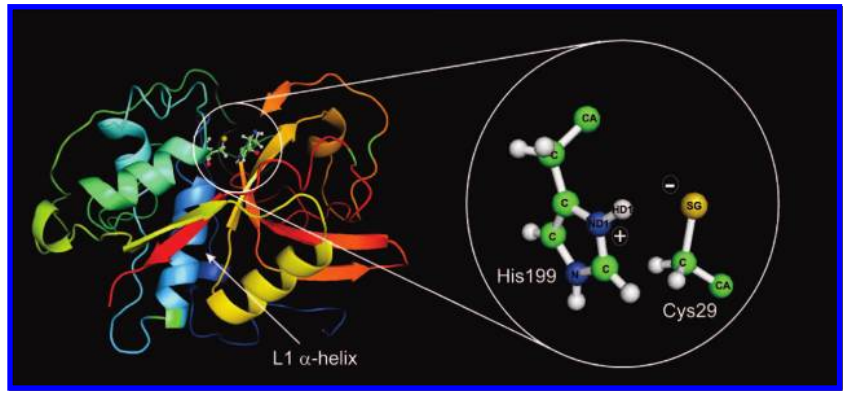


Figure 1. Active site ion pair in cathepsin B (His199H$^+$ and Cys29S$^-$) (right) together with its position within the enzyme (left). Notation for the atoms of the active site is given.

Many residues around the active site are considered to be of importance for the proper functioning of cysteine proteases.[10] Early quantum chemical studies[11,12] on gas-phase models of papain proposed that the macrodipole of the helix near the active site (L1-helix, Figure 1) facilitates the proton transfer from the cysteine to the histidine residue and thus stabilizes the zwitterionic form of the active site amino acids. While initially the entire α-helix was thought to be involved in the stabilization of the ion pair via its macrodipole, later studies indicated that the first turn of the helix accounts already for about 80% of the overall charge stabilization effect, possibly via hydrogen bond interactions.[13,14] Further indication of the importance of the terminal hydrogen bonds has been provided by both mutagenesis studies[15] and Hartree−Fock calculations.[16]

Extensive studies of the energetics of ion pairs in proteins with the use of empirical valence bond (EVB) theory and the Langevin dipole (LD) model[17,18] include work on the His-Cys ion pair in papain. It has been established by this well-tested methodology that the protein environment in papain inverts the stability of the ionic and neutral forms relative to their order in aqueous solution.[17,19,20] Previous theoretical studies have characterized the hydrolysis mechanism in much detail.[21–27] These include QM computations on small model systems[21,22,24] as well as a systematic QM/LD study of the reference solution reaction providing free energy surfaces for the acylation and deacylation steps.[23] The catalytic mechanisms of cysteine proteases have been explored by QM/MM calculations[25–27] which incorporate the influence of the whole enzyme. These QM/MM studies show that the enzyme environment accelerates the hydrolysis by stabilizing transition states and intermediates. However, they neither address the effects which stabilize the ion pair in the enzyme nor describe the switch to the nucleophilic mode since they start from structures with the substrate present. Similar problems arise for previous computations about inhibition processes of cysteine proteases.[28–31] Monte Carlo calculations[16] indicate that solvent water molecules contribute to the stabilization of the ion pair; however, their influence has not been quantified, and their role in the activation is not known.[16,32]

In order to shed more light onto these important issues we characterize the ion pair state and the transfer reaction of the HD1 proton (Figure 1) from the His199 residue to the SG atom of the Cys29 leading to the neutral state, by quantum mechanical (QM), hybrid quantum mechanics/molecular mechanical (QM/MM), and QM/MM molecular dynamics (QM/MM MD) calculations on the cathepsin B enzyme. QM/MM methods[33,34] are suitable because they allow us to describe the cleavage of the ND1−HD1 covalent bond and formation of the HD1−SG bond while accounting for the enzyme environment at the same time. They divide the total system (enzyme and solvent) into the active center and the rest. The active center comprises all parts which are directly involved in the bond breaking/formation processes and thus have to be described by QM approaches, whereas the influence of the protein environment and the solvent can be captured at the MM level. The QM and MM regions interact by electrostatic and van der Waals terms.

The paper is organized as follows: In section 2, we briefly describe the chosen computational methods and model systems. In section 3, we present and interpret the computational results

(10) Storer, A. C.; Ménard, R. *Methods Enzymol.* **1994**, *244*, 486.
(11) van Duijnen, P. T.; Thole, B. T.; Hol, W. G. *Biophys. Chem.* **1979**, *9*, 273.
(12) Broer, R.; van Duijnen, P.Th.; Nieuwpoort, W. C. *Chem. Phys. Lett.* **1976**, *42*, 525.
(13) Aqvist, J.; Luecke, H.; Quiocho, F. A.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A* **1991**, *88*, 2026.
(14) Roos, G.; Loverix, S.; Geerlings, P. *J. Phys. Chem. B* **2006**, *110*, 557.
(15) Sancho, J.; Serrano, L.; Fersht, A. R. *Biochemistry* **1992**, *31*, 22.
(16) Rullmann, J. A.; Bellido, M. N.; van Duijnen, P. T. *J. Mol. Biol.* **1989**, *206*, 101.
(17) Warshel, A. *in Computer Modeling of Chemical Reactions in Enzymews and Solutions* John Wiley & Sons: New York, 1991, pp. 140−144.
(18) Warshel, A.; Russell, S. T. *Q. Rev. Biophys.* **1984**, *17*, 283.
(19) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161.
(20) Hwang, J.-K.; Pan, J.-J. *Chem.Phys. Lett.* **1995**, *243*, 171.
(21) Howard, A. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *110*, 7195.
(22) Byun, K.; Gao, J. *J. Mol. Graph. Modell.* **2000**, *18*, 50.
(23) Strajbl, M.; Florian, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 4471.
(24) Arad, D.; Langridge, R.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 491.
(25) Harrison, M. J.; Burton, N. A.; Hillier, I. H. *J. Am. Chem. Soc.* **1997**, *119*, 12285.
(26) Han, W. G.; Tajkhorshid, E.; Suhai, S. *J. Biomol. Struct. Dyn.* **1999**, *16*, 1019.
(27) Ma, S.; Devi-Kesavan, L. S.; Gao, J. *J. Am. Chem. Soc.* **2007**, *129*, 13633.
(28) Vicik, R.; Helten, H.; Schirmeister, T.; Engels, B. *ChemMedChem* **2006**, *3*, 1021.
(29) Helten, H.; Schirmeister, T.; Engels, B. *J. Phys. Chem. A* **2004**, *108*, 7691.
(30) Helten, H.; Schirmeister, T.; Engels, B. *J. Org. Chem.* **2005**, *70*, 233.
(31) Mladenovic, M.; Schirmeister, T.; Thiel, S.; Thiel, W.; Engels, B. *ChemMedChem* **2007**, *2*, 120.
(32) Dardenne, L. E.; Werneck, A. S.; de Oliveira Neto, M.; Bisch, P. M. *Proteins* **2003**, *52*, 236.
(33) (a) For recent reviews see Senn, H. M.; Thiel, W. *Top. Curr. Chem.* **2007**, *268*, 173. (b) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.
(34) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

which help answer the much disputed questions about the factors stabilizing the active site ion pair. We also address the switch of the enzyme from the stabilized ion-pair state to the highly nucleophilic active mode. Section 4 offers a summary of the major findings.

2. Computational Methods and Model Systems

To obtain a reliable model of the active cathepsin B protease the following procedure was used. We started from the cathepsin B-E64c enzyme−inhibitor complex (1ITO[35] X-ray structure) and deleted the inhibitor moiety. To incorporate the influence of the surrounding solvent, the system was solvated in a water sphere with a radius of 50 Å. Seven sodium cations were added at the outer surface of the solvent shell to make the system neutral overall. The whole system was then submitted to a series of consecutive constrained optimizations, MD runs (heating from 100 to 300 K and equilibration at 300 K; time step 1 fs; Verlet leapfrog algorithm[36]), and further solvation until saturation with regard to the number of water molecules was reached. In this procedure the enzyme was allowed to "move" only during the optimization and was held fixed during the MD runs. The outer 5 Å layer of water and the His199 and Cys29 residues of the active site were kept fixed all the time. Finally, a 200 ps MD run (NVT; 300 K; Berendsen thermostat;[37] time step 1 fs; Verlet leapfrog algorithm[38]) was performed in which all constraints except those for the residues of the ion pair of the active site (His199 and Cys29) were released. During this run snapshots were taken every 40 ps. The five structures obtained after 40, 80, 120, 160, and 200 ps were used as starting structures in QM/MM MD equilibration and production runs. We forbore from using other cysteine protease X-ray structures without a bound inhibitor (1HUC,[39] 1PPN,[40] 9PAP[41]) since close inspection of the active sites revealed no significant differences in the geometrical arrangement of the ion pair and surrounding residues compared to the 1ITO structure. Moreover, the active site is modified also in these structures, e.g., by oxidation. The preparation of the system was done by means of fully classical MD simulations using the CHARMM force field and program package.[42]

To limit the computational effort for the QM/MM MD calculations the size of the system had to be reduced. This was achieved by restricting the water layer around the protein surface to a thickness of 12 Å. The counterions were kept and placed at the new surface of the water sphere. During the simulations the enzyme and a water layer of 3 Å around the enzyme were allowed to move. The 9 Å thick "frozen" outer water shell was used to create a boundary between the system and the surrounding vacuum, preventing the mobile water molecules from evaporating and thus providing a kind of effective boundary potential. To simplify the simulation, all hydrogen atoms in the system were replaced by deuterium; i.e., their mass was set to $m_d = 2.014$ au. The calculations were done using the ChemShell program[43] and the DL_POLY code.[44] The AM1[45] semiempirical method was employed in the QM part, and the CHARMM force field, in the MM part. The active water molecules were only allowed to undergo rotation and translation; i.e., their vibrational motion was frozen. The QM region contained the residues of the active site, His199 and Cys29. The His199 residue was terminated at the CA position (see Figure 1) by replacing this atom with a H link atom. The Cys29 residue was treated analogously. At the QM/MM boundary the charge-shift scheme was applied, and fixed MM point charges were included into the QM Hamiltonian to provide electrostatic embedding of the QM region. The reversible noniterative leapfrog-type integrator[46] was used for integration of the equations of motion. The simulated temperature was 300 K (NVT conditions), controlled by a Berendsen thermostat[37] (coupling time 0.1 ps) during the 10 ps long equilibration and by a Nosé−Hoover chain thermostat[47] (chain length of 4, characteristic coupling time of the thermostat to the physical system of 0.02 ps) during the 20 ps production phase. The time step was 1 fs.

QM/MM reaction path calculations were performed using density functional theory (DFT) at the RIDFT[48]/B-LYP[49]/TZVP[50] level for the QM part and the CHARMM force field for the MM part. DFT is well-known to describe many properties with an excellent cost-benefit value.[51–53] Nevertheless, this is often based on an error compensation which does not work in all cases,[54–57] and indeed in many cases multireference approaches are necessary to obtain reliable reaction paths.[58–61]

(35) Yamamoto, A.; Tomoo, K.; Matsugi, K.; Hara, T.; Murata, Y. M.; Kitamura, K.; Ishida, T. *Biochim. Biophys. Acta* **2002**, *1597*, 244.
(36) (a) Verlet, L. *Phys. Rev.* **1967**, *159*, 98. (b) Hockney, R. W. *Methods in Computational Physics* **1970**, *9*, 136.
(37) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(38) (a) Schechter, I.; Berger, A. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157. (b) Berger, A.; Schechter, I. *Phil. Trans. R. Soc. London* **1970**, *B257*, 1330.
(39) Musil, D.; Zucic, D.; Turk, D.; Engh, R. A.; Mayr, I.; Huber, R.; Popovic, T.; Turk, V.; Towatari, T.; Katunuma, N.; Bode, W. *EMBO J.* **1991**, *10*, 2321.
(40) Pickersgill, R. W.; Harris, G. W.; Garman, E. *Acta Crystallogr., Sect.B* **1992**, *48*, 59.
(41) Kamphuis, I. G.; Kalk, K. H.; Swarte, M. B.; Drenth, J. *J. Mol. Biol.* **1984**, *179*, 233.
(42) (a) MacKarell, A. D., Jr.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586. (b) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187. (c) MacKerell, A. D., Jr.; Brooks, B. R.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *The Encyclopedia of Computational Chemistry*; Schleyer P. v. R. et al., Eds.; John Wiley & Sons: Chichester, 1998; Vol. 1, p. 271.
(43) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J. *THEOCHEM* **2003**, *632*, 1.
(44) Smith, W.; Forester, T. *J. Mol. Graph.* **1996**, *14*, 136.
(45) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. G.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
(46) Jang, S.; Voth, G. A. *J. Chem. Phys.* **1997**, *107*, 9514.
(47) (a) Nosé, S. J. *J. Chem. Phys.* **1984**, *81*, 511. (b) Nosé, S. J. *Mol. Phys.* **1984**, *52*, 255. (c) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695. (d) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635.
(48) (a) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. *Chem. Phys. Lett.* **1993**, *213*, 514. (b) Eichkorn, K.; Treutler, O.; ¨; Ohm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *242*, 652.
(49) (a) Becke, A. D. *Phys. Rev. B* **1988**, *38*, 3098. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(50) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1997**, *100*, 5829.
(51) Adam, W.; Bottke, N.; Engels, B.; Krebs, O. *J. Am. Chem. Soc.* **2001**, *123*, 5542.
(52) Schlund, S.; Mladenovic, M.; Basilio; Janke, E. M.; Engels, B.; Weisz, K. *J. Am. Chem. Soc.* **2005**, *127*, 16151.
(53) Hupp, T.; Sturm, C.; Basilio Janke, E. M.; Perez Cabre, M.; Weisz, K.; Engels, B. *J. Phys. Chem.* **2005**, *109*, 1703.
(54) Engels, B.; Schoeneboom, J. C.; Munster, A. F.; Groetsch, S.; Christl, M. *J. Am. Chem. Soc.* **2002**, *124*, 287.
(55) Suter, H. U.; Pleβ, V.; Ernzerhof, M.; Engels, B. *Chem. Phys. Lett.* **1994**, *230*, 398.
(56) Suter, H. U.; Engels, B. *J. Chem. Phys.* **1994**, *100*, 2936.
(57) Engels, B. *Theo. Chim. Acta* **1993**, *86*, 429.
(58) Schmittel, M.; Steffen, J. P.; Engels, B.; et al. *Angew. Chem., Int. Ed.* **1998**, *37*, 2371.
(59) Engels, B.; Peyerimhoff, S. D. *J. Phys. Chem.* **1989**, *93*, 4462.
(60) Engels, B.; Hanrath, M.; Lennartz, C. *Computer & Chemistry* **2001**, *25*, 15.
(61) Musch, P. W.; Engels, B. *J. Am. Chem. Soc.* **2001**, *123*, 5557.

For the present problem, however, our procedure was found to be sufficiently accurate for the determination of the reaction energy. The distance between the His199 ND1 atom and the HD1 hydrogen (see Figure 1) was always used as a reaction coordinate. In the QM/MM calculations the whole enzyme and a water sphere of $r = 25$ Å around the active site were included. All QM/MM computations employed the ChemShell software[43] using the TURBOMOLE program suite[62] for QM and the DL_POLY[44] code for MM calculations. Geometry optimizations were carried out with the hybrid delocalized internal coordinate (HDLC) optimizer.[63] In these calculations the outer atoms beyond a sphere of $r = 10$ Å around the inhibitor molecule were kept fixed. The division into QM and MM parts remained the same as in the AM1/CHARMM MD runs.

We emphasize that the present QM/MM reaction profiles are based on the computed energies rather than free energies, for the following two reasons. First, standard free energy calculations (e.g., by thermodynamic integration or umbrella sampling) at the currently applied DFT/CHARMM level are extremely demanding in terms of computational effort, and one would thus have to resort to simpler semiempirical QM/MM methods[27] or to one of the more approximate free energy treatments that have been proposed recently; see, e.g., refs 64−66. Second, we are interested in the origin of the stability of the zwitterionic resting state in the enzyme, and we thus focus on the underlying potential energy surfaces which should be of prime importance in this context at least in a qualitative sense, e.g., with regard to the role of hydrogen bond networks.

The influence of particular moieties in the MM region (e.g., amino acid residues or water molecules close to the active site) on the energy difference between the neutral and the zwitterionic state was assessed by means of a charge deletion analysis, through additional single-point QM/MM calculations in which the atomic charges of the respective MM moieties were set to zero.[67–69]

Additional information about the influence of single residues was obtained from QM calculations on active-site model systems. The standard gas-phase model contained the amino acids His199, Cys29, Gln23, and Asn219 and two additional water molecules, and smaller models were generated by removing selected residues. In these QM calculations, His199 was represented by an imidazolium ion, Gln23 and Asn219 were represented by acetamide, and Cys29 was represented by its side chain and the peptide bond to Ser28 (i.e., amide group of the Cys29 backbone bound to acetyl). Single-point QM computations were performed at optimized QM/MM geometries. For the sake of comparison, some of these QM calculations employed the COSMO[70] continuum solvent model (water, $\varepsilon = 78$).

(62) Ahlrichs R., Bär M., Baron H.-P., Bauernschmitt R., Böcker S., Ehrig M., Eichkorn K., Elliott S., Furche F., Haase F., et al. TURBOMOLE, (since 1988) Version 5.6; University of Karlsruhe, Germany.
(63) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177.
(64) Rosta, E.; Klähn, M.; Warshel, A. *J. Phys. Chem. B* **2006**, *110*, 2934.
(65) Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. *J. Chem. Theory Comput.* **2006**, *2*, 452.
(66) Zhang, J.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483.
(67) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826.
(68) Mulholland, A. J.; Richards, W. G. *Proteins* **1997**, *27*, 9.
(69) Wong, K. F.; Watney, J. B.; Hammes-Schiffer, S. *J. Phys. Chem. B* **2004**, *108*, 12231.

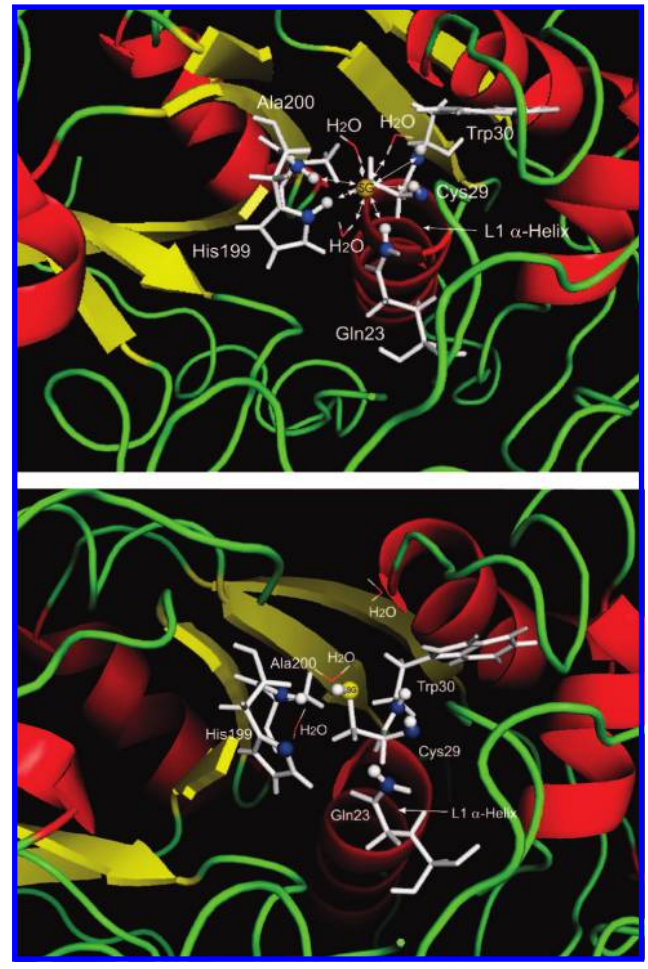


Figure 2. Structural arrangement of the active site of cathepsin B together with the nearby solvent water molecules and protein surroundings taken from a typical frame of the 80, 120, and 200 ps QM/MM MD simulations (top, zwitterionic state) and 40 and 160 ps (bottom, neutral state). The hydrogen bond network between the sulfur of the Cys29 and the surrounding water and enzyme atoms is shown. The amide groups of the enzyme involved in H-bonds, as well as S^- of Cys29, are given in ball-and-stick representation.

3. Results

The zwitterionic and the neutral form of the active site are connected via a proton transfer from the His199 residue to the thiolate group of the Cys29 residue. To study the stability of the zwitterionic state we performed MD simulations in which the potential energy is computed by a QM/MM approach including the ion pair within the QM part. Figure 2 gives typical geometrical arrangements for the active site in the zwitterionic (top) and in the neutral status (bottom). Insight into the structural changes occurring along the QM/MM MD trajectories can be gained from Figures 3 and 4. They give the time dependence of selected atomic distances which characterize the composition of the hydrogen-bond networks within the active site. For all systems the first 200 frames (10 ps) represent equilibration within the Berendsen thermostat[37] while the remaining 400 frames (20 ps) come from the production phase performed by using the Nose−Hover thermostat[47] (see above). Frames were extracted from the QM/MM MD calculations every 0.05 ps. In the QM/MM MD simulations, the proton transfer between the His199 and Cys29 residues is allowed and is indeed observed in two of the five trajectories (40 ps and 160 ps, Figure 4). The transfer is indicated by the drop in the distance abbreviated as

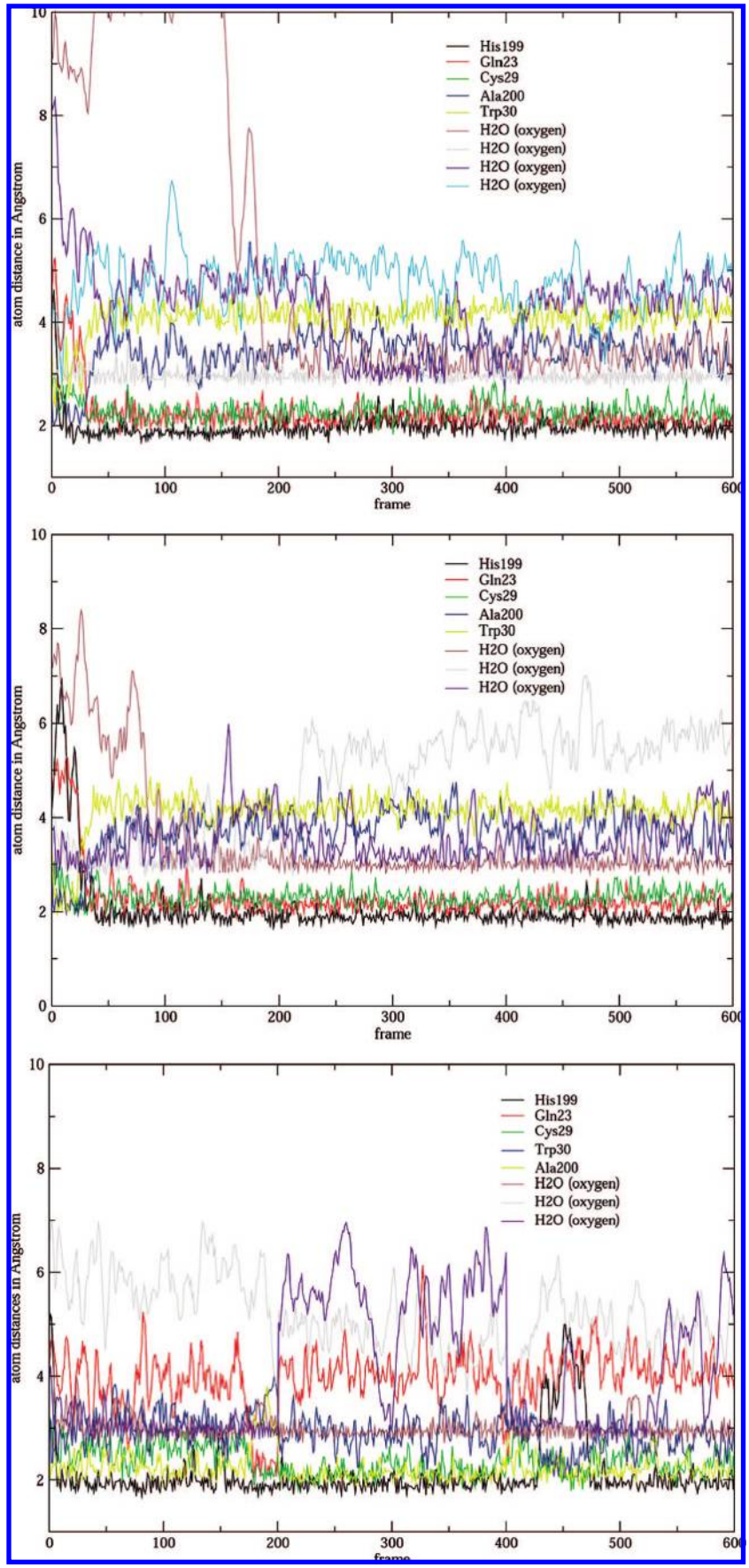


Figure 3. Time dependence of the hydrogen bond network between the sulfur of Cys29 and the surrounding water and enzyme atoms (for an explanation of the legend see text) for the 80 ps (top), 120 ps (middle), and 200 ps (bottom) QM/MM MD trajectories. Distances to water molecules are presented as distances between the sulfur of Cys29 and the oxygen atom of water. Other distances are between sulfur and hydrogen. No protonation of S^- of Cys29 occurs.

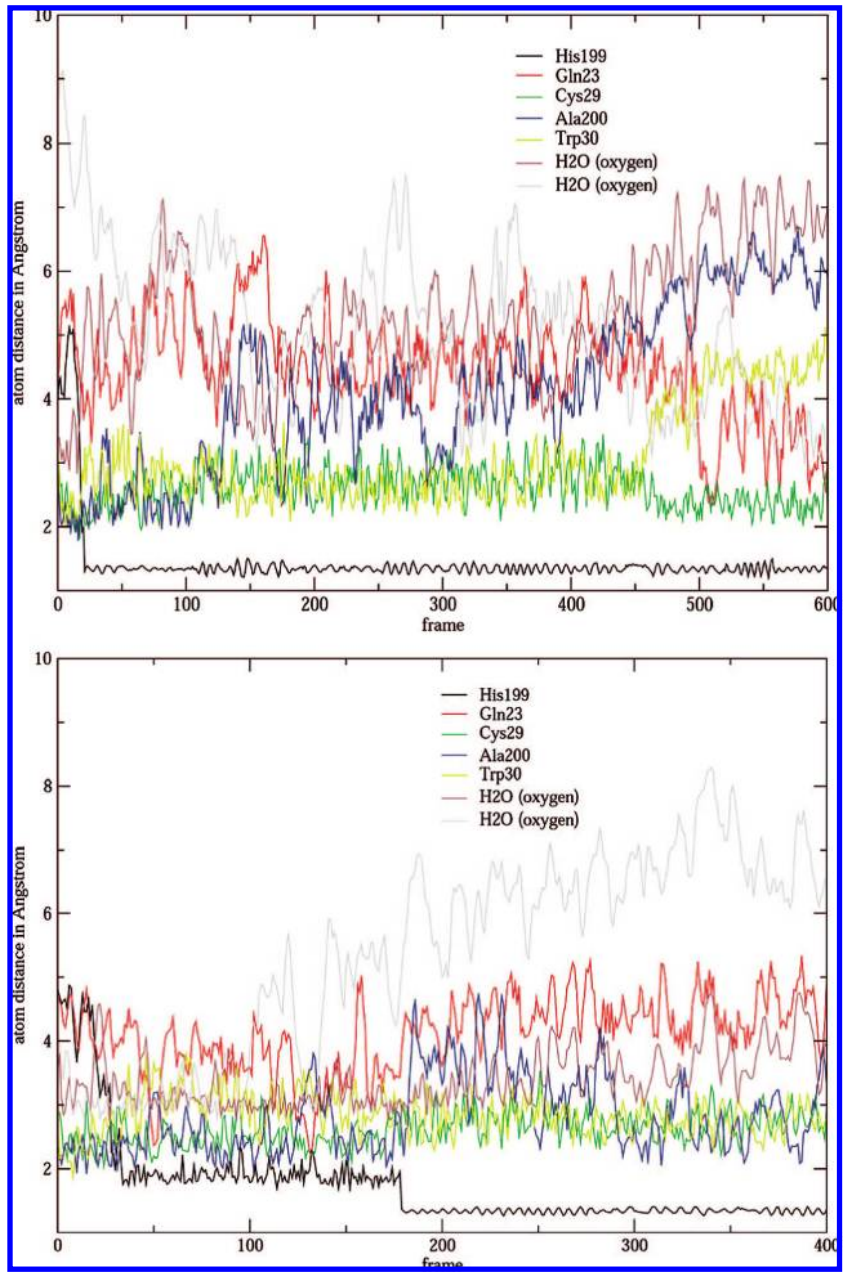


Figure 4. Time dependence of the hydrogen bond network between the sulfur of Cys29 and the surrounding water and enzyme atoms (Figure 2) for the 40 ps (top) and 160 ps (bottom) QM/MM MD trajectories. Distances to water molecules are presented as distances between the sulfur of Cys29 and the oxygen atom of water. Other distances are between sulfur and hydrogen. Protonation of S^- of Cys29 occurs.

His199 (Figure 4, black line) to about 1 Å. This His199 line represents the distance between the sulfur center and the proton located at the His199 residue at the beginning of the MD run. In the remaining three trajectories, however, the zwitterionic form remains stable (Figure 3). For these trajectories the value for the His199 distance oscillates around 2 Å.

First we concentrate on the systems in which no proton transfer appears (Figure 3: 80, 120, and 200 ps). The time evolution of the corresponding H-bond networks is characterized by the atomic distance variations given in Figure 3. It clearly shows that in all three systems the negatively charged sulfur of the Cys29 residue is stabilized by four to five strong and stable hydrogen bonds. Their bond lengths are around 2 Å. The negatively charged sulfur center always forms strong hydrogen bonds to the His199 residue and to the amide group of the Cys29 backbone. Additionally, it forms bonds either to the Gln23 residue (80 and 120 ps, Figure 3 top and middle) or to the Ala200 residue backbones (200 ps, Figure 3 bottom). Approximately two additional strong H-bridges are formed to solvent water molecules. Please note that H-bonds to water molecules are described by the distance between the S^- and the oxygen atom of water, since the water molecules are mobile and can, by rotation, easily switch the hydrogen atom involved in the H-bond. The actual H-bond distance is typically about 1 Å shorter than the associated SO distance plotted in the figures.

Beside these strong interactions, there are also somewhat weaker hydrogen bonds (with distances of 3−4 Å) providing additional stabilization for the deprotonated Cys29 residue. These are usually formed with Trp30 and additional water molecules. It is important to note that the potential hydrogen bond donors (His199, Cys29 backbone, Gln23, Ala200, Trp30, see Figure 4 top) are present in every direction around the Cys29

residue. Therefore, the thiolate residue can be stabilized regardless of the direction in which it moves. If it moves downward in Figure 2 (top) it will be stabilized mostly by Gln23, His199, and Cys29 (80 and 120 ps). If it moves upward the Ala200 and Trp30 residues take over (200 ps).

In the trajectories of the 40 ps and 160 ps systems (Figure 4) there is proton transfer from His199 to the sulfur of the Cys29 residue. In both systems proton transfer occurs during the equilibration period (Berendsen thermostat[37]) where larger potential energy fluctuations are possible. By looking at the time evolution of the hydrogen bond distances in these two systems (Figure 4) one notices that the pattern of 4−5 strong, stable H-bonds is missing. This is also seen in Figure 2 (bottom). In the 40 ps system, protonation of the S^- takes place at the very beginning. This QM/MM MD run starts with four H-bonds (Ala200, Trp30, Cys29, and H_2O), but the distance variations are very large. Once the water molecule starts to drift away (the brown line in Figure 4) no nearby hydrogen bond donor can take over, and as a consequence the proton transfer takes place and the H-bond network falls apart. The influence of the water molecule leaving the H-bond network is even better reflected in the 160 ps system. The collapse of the hydrogen bond network in this case is obviously due to the water molecule which starts to drift away after 5 ps (gray line in Figure 4).

These QM/MM-MD simulations underline the importance of the hydrogen bonds between the Cys29 sulfur atom and the surrounding enzyme residues and water molecules but cannot deliver a quantitative picture. To quantify the various effects and to isolate and analyze them, QM/MM and pure QM calculations were performed.

The QM/MM computations employed the same partitioning of the system as the QM/MM MD simulations; however, the RIDFT[48]/B-LYP[49]/TZVP[50] level of theory was used for the description of the QM part. The distance between the His199 ND1 atom and the HD1 hydrogen ($R_{ND1-HD1}$, for atom notation see Figure 1) was used as a reaction coordinate, and the minimum energy path for the proton transfer was obtained by fixing this distance to a given value while all other geometrical parameters were relaxed. The DFT/CHARMM calculations started from the final structure obtained in the 80 ps AM1/CHARMM MD run; i.e., we concentrated on the situation in which the thiolate forms stronger H-bonds to the His199 and Gln23 residues and to the amide group of the Cys29 backbone.

In the QM/MM MD simulations the composition and the nuclear arrangement of the first solvent layer changed considerably when the proton transfer took place. In the zwitterionic structure at least two water molecules form strong hydrogen bonds to the negatively charged sulfur center of the Cys29 residue (Figure 2 top). As discussed the proton transfer is induced by one water molecule moving away. For the neutral state only one water molecule forms a bond to the thiolate, albeit weaker than that in the zwitterionic system (Figure 2 top). This behavior indicates that the $R_{ND1-HD1}$ distance is not the only important coordinate in the course of the proton transfer reaction. Preparatory computations indeed showed that a one-dimensional description of the proton transfer starting from the zwitterionic state is problematic since the arrangement of the two water molecules changed abruptly leading to noncontinuous curves. However, using the neutral states as starting points, continuous curves were obtained if the positions of the water molecules were carefully optimized in each step. To assess the importance of the water molecules, the reaction path was first computed with two water molecules in the direct vicinity of the sulfur of the Cys29 residue (system 1, Figure 5, top), and then with only one water molecule (system 2, Figure 5, bottom), by manually moving the second one away from the sulfur center to the bulk water. Note that both water molecules adopt quite different positions and that the surrounding water shells differ to some extent in systems 1 and 2. The numerical results are summarized in Table 2 where the computed energies are given with respect to the zwitterionic state. Figure 5 shows the computed reaction profiles.

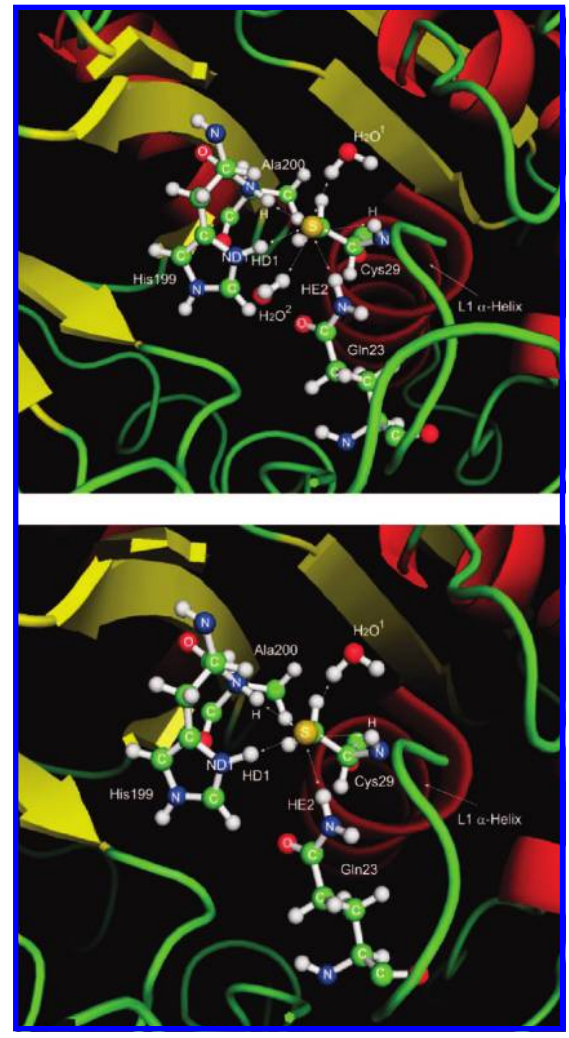


Figure 5. Structural arrangement of the active site of cathepsin B with two (top) and one (bottom) water molecules in the direct vicinity of the sulfur of the Cys29 residue. The protein surroundings are also indicated.

Table 1. Selected Distances (in Å) of the Structural Arrangements Depicted in Figure 2 (Top, Ionic State in QM/MM MD; Bottom, Neutral State in QM/MM MD)[a]

state	$R_{ND1-HD1}$	R_{S-HD1}	$R_{S-H(Gnl23)}$	$R_{S-H(Cys29)}$	$R_{S-H2O}{}^1$	$R_{S-H2O}{}^2$	$R_{S-H(Ala200)}$
ionic	1.05	2.02	2.19	2.31	2.98	3.38	3.57
neutral	4.26	1.44	4.48	2.60	3.89	5.27	4.13

[a] The values are averaged over the QM/MM MD runs. The definition of the distances can be taken from Figures 2−4 and the text. SO distances are given for hydrogen bonds involving water.

For system 1 (with two water molecules) the zwitterionic state is 31 kJ mol^{-1} lower in energy than the neutral state. Upon lengthening R_{ND1-H1} from 1.08 Å to 1.45 Å in 0.1 Å steps, the energy increases by about 7 kJ mol^{-1} per step. In this range the increase in R_{ND1-H1} leads to a decrease in R_{S-HD1} by about

Table 2. Selected Distances (in Å) and Relative Energies (in kJ mol^{-1}) for Points along the Reaction Paths for Proton Transfer between the His199 and Cys29 Residues Obtained in QM/MM Calculations[a]

entry	ΔE	$R_{ND1-HD1}$	R_{S-HD1}	$R_{S-H(Gln23)}$	$R_{S-H(Cys29)}$	$R_{S-H(Ala200)}$	$R_{S-H2O}{}^1$	$R_{S-H2O}{}^2$
System 1: QM/MM 2H_2O								
1	**0**	**1.08**	**2.06**	**2.22**	**2.37**	**3.31**	**2.18**	**2.80**
2	4	1.17	1.91	2.23	2.38	3.30	2.18	2.87
3	15	1.26	1.76	2.24	2.43	3.29	2.21	3.01
4	22	1.35	1.65	2.25	2.47	3.29	2.23	3.07
5	29	1.45	1.55	2.28	2.48	3.3	2.27	3.2
6	30	1.55	1.49	2.28	2.53	3.36	2.28	3.22
7	31	1.65	1.45	2.28	2.53	3.47	2.30	3.25
8	**31**	**1.68**	**1.45**	**2.29**	**2.53**	**3.42**	**2.29**	**3.26**
System 2: QM/MM 1 H_2O								
9	**0**	**1.10**	**1.99**	**2.25**	**2.33**	**3.1**	**2.29**	
10	3	1.17	1.87	2.26	2.37	3.13	2.35	
11	9	1.26	1.75	2.25	2.4	3.15	2.41	
12	13	1.35	1.64	2.26	2.43	3.17	2.43	
13	15	1.45	1.56	2.27	2.45	3.20	2.50	
14	16	1.55	1.50	2.27	2.49	3.26	2.60	
15	18	1.65	1.45	2.25	2.55	3.34	3.09	
16	**19**	**1.73**	**1.43**	**2.26**	**2.54**	**3.37**	**3.07**	

[a] For notation of the various distances, see Figures 1 and 5. Bold numbers correspond to fully optimized structures.

the same amount showing that the proton moves on a straight line between both centers. For R_{ND1-H1} being larger than 1.45 Å, the energy does not change much, and the variation in R_{S-HD1} is much smaller indicating that the new covalent SH bond has been formed so that a further increase of R_{ND1-H1} simply increases the distance between the newly formed thiol group and the deprotonated His199 residue. The proton transfer is accompanied by a weakening of the hydrogen bonding network. This is seen in the elongation of the H-bonds $R_{S-H(Gln23)}$, $R_{S-H(Cys29)}$, $R_{S-H(Ala200)}$ and of the distance to the more strongly bonded water molecule. All increase by 0.1 Å or more. The distance to the less strongly bonded water molecule increases by about 0.4 Å implying that it does not contribute much to the stabilization of the neutral system. Please note that the distances given in Table 1 (AM1 approach) and those given in Table 2 (BLYP approach) do not differ significantly. This indicates that already the semiempirical AM1 approach used in the QM/MM MD simulation provides a reliable estimate of the strengths of the hydrogen bonds.

To verify the influence of both water molecules on the stabilization we also performed QM/MM computations in which the second water molecule is shifted toward the bulk water so that its influence on the active site becomes considerably smaller (system 2). As expected, the energy difference between the zwitterionic and the neutral state is lowered to about 19 kJ mol^{-1}. In these computations the hydrogen bonds between the sulfur of the Cys29 residue and the enzyme proton donors are similar to those computed for two water molecules (system 1). Judging from the distances, the H-bond strength of the remaining water is between the H-bond strengths of the two water molecules in system 1: the R_{S-H2O} distances for the zwitterionic (neutral) state are 2.29 (3.07) Å in system 2 compared with 2.18 (2.29) Å and 2.80 (3.26) Å in system 1, respectively.

The stabilization of the zwitterionic form afforded by the water molecules can be estimated from the computed QM/MM reaction profiles (see Table 2 and Figure 6) and the QM/MM MD simulations. A comparison of systems 1 and 2 shows that the first and second additional water molecules stabilize the zwitterionic state relative to the neutral one by about 19 and 12 kJ mol^{-1}, respectively. Since the QM/MM MD simulations of

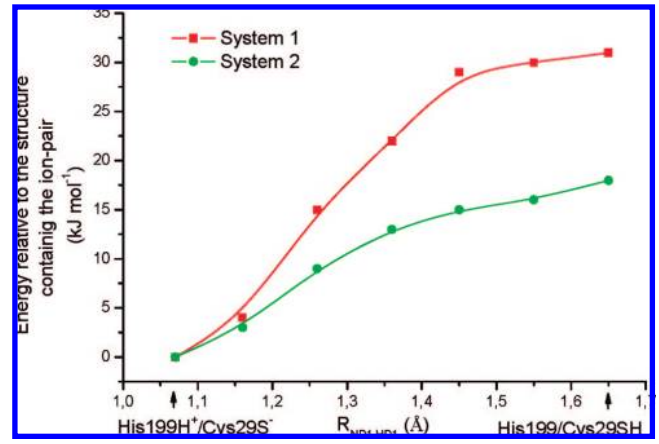


Figure 6. Reaction profiles for proton transfer from His199 to the Cys29 residue in two systems (see text) calculated using the QM/MM method. The energy is given relative to the structure with the active site in ion pair form (His199H$^+$/Cys29S$^-$).

Table 3. Electrostatic Contribution ΔE (kJ mol^{-1}) of Specific Moieties to the Stabilization of the Zwitterionic State Relative to the Neutral State[a]

entry	moiety switched off	ΔE
1	none (system 1)	0
2	NH group of Cys29	20
3	H_2O^1	18
4	H_2O^2	12
5	Gln23	13
6	Asn219	23
7	complete L1 helix except Cys29	15
8	complete MM region	115
9	sum of entries 2−6	86
10	sum of entries 2−7	101
11	CO group of Asn219	9
12	remainder of Asn219	11

[a] Positive values indicate that the zwitterionic state becomes less stable when the charges of the given moiety are set to zero in the charge deletion analysis (see text).

the zwitterionic state typically have even more water molecules in the vicinity of the sulfur (2−4 on average), we estimate that this state is actually even slightly more stable than indicated by system 1, by around 32−36 kJ mol^{-1} relative to the neutral state.

The influence of particular moieties of the enzyme on the energy difference between the neutral and the zwitterionic state can be assessed by a charge deletion analysis (see section 2). The zwitterionic state is generally destabilized relative to the neutral state when the electrostatic QM/MM interactions with a moiety close to the active site are turned off by setting its MM charges to zero. The corresponding energy changes (see Table 3) were computed at the QM/MM optimized geometries (see Figure 5, top panel). Consistent with the preceding results from QM/MM reaction paths, the two water molecules are found to stabilize the zwitterionic state electrostatically by 18 and 12 kJ mol^{-1}, respectively. Other neighboring moieties provide a similar electrostatic stabilization in the range 15−23 kJ mol^{-1}: in the case of the NH group of Cys29 and Gln23, this is due to direct (hydrogen-bonding) interactions within the active site, while the more distant Asn219 residue acts indirectly through polarization of His199 (which accounts for about half of the Asn219 effect, with the other half-coming from other interactions, see entries 11 and 12). The total electrostatic stabilization from the moieties discussed so far (entries 2−6 in Table 3)

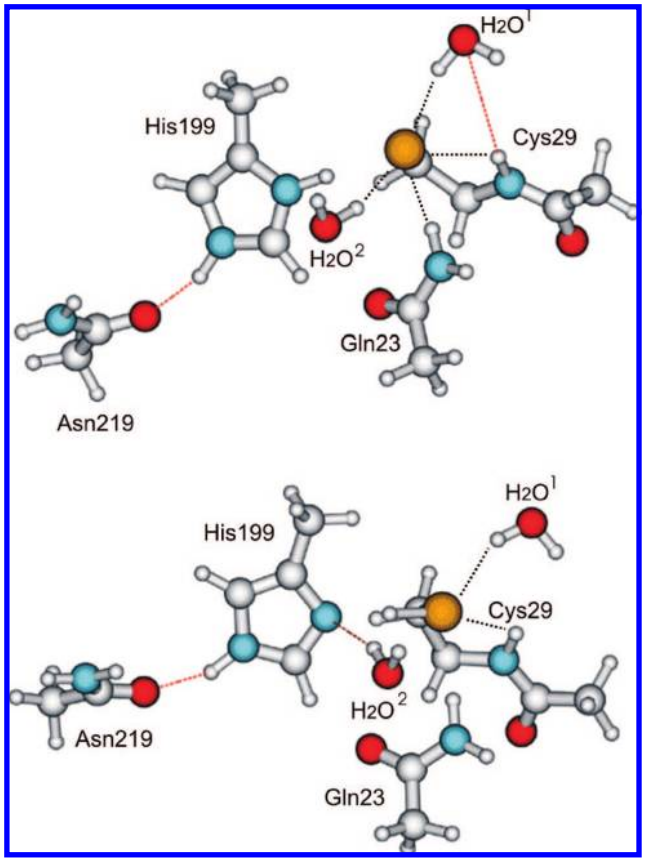


Figure 7. Model system for QM calculations. Top: Zwitterionic state, comprising His199 (represented by methylimidazolium ion), Cys29 (with amide group of backbone), Gln23 (represented by acetamide), Asn219 (represented by acetamide), and two water molecules. Bottom: Correspondingly for the neutral state.

amounts to 86 kJ mol^{-1} and thus accounts for the major part of the overall electrostatic stabilization by the complete MM region (115 kJ mol^{-1}), with about half the remainder being provided by the rest of the L1 helix (15 kJ mol^{-1}).

The charge deletion analysis (CDEL) thus confirms that the two water molecules contribute substantially to the stabilization of the zwitterionic state, but it also reveals important contributions from other residues. In addition, the analysis suggests that the smallest sensible gas-phase model system should include the QM region from the QM/MM calculations (His199 and Cys29 side chain, see section 2) plus the neighboring moieties (entries 2−6 in Table 3). This is borne out by gas-phase QM calculations on this model system (Figure 7) and the bare QM region (Figure 1) which yield energies of +4 and −81 kJ mol^{-1}, respectively, for the neutral relative to the zwitterionic state, compared with +31 kJ mol^{-1} from the full QM/MM calculation. Including the neighboring moieties at the QM level in gas-phase model calculations thus stabilizes the zwitterionic state by 85 kJ mol^{-1}, in almost perfect agreement with the value from charge deletion analysis (86 kJ mol^{-1}, see above), but still falls short of the total stabilization achieved in the enzyme. When the QM computations are done in combination with the COSMO continuum solvation approach[70] using a dielectric constant of 78 (water), we obtain energies of +22 and −8 kJ mol^{-1} for the neutral vs the zwitterionic state of the model system and the bare QM region, respectively.

Finally, we have carried out a series of single-point QM test calculations on truncated gas-phase systems where some of the neighboring moieties (entries 2−6 in Table 3) were removed from the model system (Figure 7). The effects on the relative stability of the neutral and zwitterionic state are generally consistent with those found by charge deletion analysis (Table 3), but normally somewhat smaller: for example, the QM relative energies change by 14, 11, 15, and 10 kJ mol^{-1} upon removal of the NH group of Cys29, H_2O^1, Gln23, and Asn219, respectively.

Combining the available results leads to the following picture. In contrast to expectations based on the "normal" pK_a values of HisH$^+$ (pK_a = 6−7) and of CysSH (pK_a = 8−9), all simulations indicate that the zwitterionic form of the active site (Cys29−S$\cdots$H$^+$−His199) is more stable than the neutral form (Cys29−SH$\cdots$His199). The reverse energy ordering results from a strong hydrogen bonding network involving active-site residues (NH group of Cys29, Gln23, Ala200, and Trp30) and 2−4 water molecules, two of which form strong hydrogen bonds to the thiolate. The NH backbone group of Cys29 and the residue of Gln23 each contribute 15−20 kJ mol^{-1} to the stabilization of the zwitterion (CDEL, QM), while the two strongly bonded water molecules contribute about 30 kJ mol^{-1} (QM/MM, CDEL), one being somewhat more important than the other (18 vs 12 kJ mol^{-1}). Hence, there is a cage-like network around the negatively charged sulfur atom which will stabilize it regardless of the direction in which it moves (QM/MM MD). Apart from this hydrogen bonding network, the interaction of Asn219 with the His199 imidazole ring provides an indirect stabilization by about 10 kJ mol^{-1} (CDEL, QM), and the L1 α-helix without Cys29 contributes about 15 kJ mol^{-1} (CDEL). The latter value may be taken as an estimate for the effect of the helix macrodipole, which is clearly non-negligible but of similar magnitude as others (e.g., less important than the combined effect of the two water molecules).

Why does the strongly stabilized thiolate center become reactive once a substrate enters the active site? We have not studied this process computationally, but the present and previous theoretical work[22,25–27] suggests the following scenario. When the substrate enters the active site cleft, it will dislodge the water molecules that are situated inside and it will also disrupt the hydrogen bonds of the S$^-$ thiolate to the His199 and Gln23 side chains and to the NH group of its own backbone since it interacts noncovalently with these amino acids. The loss of the stabilizing hydrogen bond network is expected to enhance the nucleophilic reactivity of the thiolate toward the nearby carboxyl carbon of the substrate. [22,25–27]

4. Summary and Conclusion

In the present work QM, QM/MM, and QM/MM MD approaches were used to study the effects stabilizing the active site ion pair of cysteine proteases and, in particular, to investigate the proton transfer from the ND1 center of the His199 residue (Figure 1) to the sulfur (SG center) of the Cys29 residue. QM/MM MD computations show that the ion pair is stabilized by a complex hydrogen bonding network involving the S$^-$ thiolate, the His199 and Gln23 residues, the NH backbone groups of Cys29, Ala200, and Trp30, and 2−4 water molecules. These hydrogen bond donors form a cage stabilizing the thiolate regardless of the direction in which it moves. In each case the interactions with His199, Cys29, and at least two water molecules are important. If the thiolate moves upward (see Figure 2) it is additionally stabilized by Ala200 and Trp30. If it moves downward these latter interactions get weaker and the hydrogen bond to Gln23 takes over. The QM/MM MD

simulations also show that the drifting away of a single water molecule already destabilizes the ion pair so much that the proton transfer may take place, leading to the neutral state.

A quantitative assessment of the various effects is obtained by QM/MM and QM computations. They indicate that the enzyme environment stabilizes the ion pair by more than 100 kJ mol^{-1} compared with the gas phase and by around 40 kJ mol^{-1} compared with the ion pair in aqueous solution (QM region in COSMO environment). Most of the stabilization in the enzyme comes from the active-site hydrogen bonding network involving Cys29, Gln23, H_2O^1, and H_2O^2, with some additional contribution from Asn219 and from the macrodipole of the L1 α-helix. According to the QM/MM results, the zwitterionic form is about 30 kJ mol^{-1} more stable than the neutral form in the enzyme.

The present data, together with other computations,[27,31,71] suggest a plausible explanation for the efficiency of the subsequent nucleophilic reaction. Once the substrate enters the active site cleft, it will disrupt the stabilizing hydrogen bonding network around the thiolate such that the negatively charged sulfur center becomes effectively more nucleophilic. This will facilitate the acylation step of the hydrolysis which is also known to be accelerated by electrostatic transition state stabilization.[27]

Acknowledgment. Financial support by the DFG (Deutsche Forschungsgemeinschaft) (B.E., T.S.: SFB630: TPA4, TPC3; B.E., R.F., and T.S.: SPP 1178) and the VolkswagenStiftung (B.E., R.F., and W.T.) is gratefully acknowledged. Dedicated to Professor Dr. Peter Luger on the occasion of his 65th birthday.

Supporting Information Available: Cartesian coordinates of optimized QM regions and full refs 42 and 62. This material is available free of charge via the Internet at http://pubs.acs.org.

JA711043X

(70) (a) Klamt, A.; Schürmann, G. J. *J. Chem. Soc., Perkin. Trans. 2* **1993**, *5*, 799. (b) Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187.

(71) Mladenovic, M.; Junold, K.; Fink, R. F.; Thiel, W.; Schirmeister, T.; Engels, B. *J. Phys. Chem. B* **2008**, *112*, 5458.