Conservative Electrostatic Potential Patterns at Enzyme Active Sites: The Anion−Cation−Anion Triad

Timea Gérczei,[†] Bence Asbóth,[‡] and Gábor Náray-Szabó*,[†]

Department of Theoretical Chemistry, Eötvös University Budapest, P.O. Box 32, H-1518 Budapest 112, Hungary, and Institute for Biochemistry and Protein Research, Agricultural Research Center, P.O. Box 170, H-2101 Gödöllö, Hungary

Received July 30, 1998

Electrostatic potentials at enzyme active sites with the (− + −) charge distribution were calculated using the linearized Poisson−Boltzmann equation. It was found for all cases studied (five serine proteases, lipase, acetylcholinesterase, lysozyme, and D-xylose isomerase) that the protein and substrate electrostatic potential patterns on the van der Waals envelope of the latter complement each other. This calls attention to a convergent evolution of the active-site potential in the cases of serine proteases, providing similar patterns for enzymes with very different primary structures. Enzyme activities, as characterized by log k_{cat}/k_M for the same substrate (succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide) of α-chymotrypsin, *β*-trypsin, α-lytic protease, subtilisin *Novo*, and subtilisin *Carlsberg*, respectively, correlate well with the calculated electrostatic interaction energies between the protein environment and the active site. To achieve a better fit between the calculated and experimental quantities, the geometry of the enzyme−substrate complexes had to be optimized by a technique based on molecular dynamics. For the same enzymes, it was found that a quantitative measure of the electrostatic complementarity between the active site and protein environment correlates with the electrostatic interaction energies, as well as the activities. On the basis of this observations we propose the use of electrostatic complementarity between the active site and surrounding protein for the characterization of enzyme catalytic power.

INTRODUCTION

Enzymes often make use of preorganized structural units at their active sites for catalytic rate acceleration. A classic important example, the Asp···His···Ser triad, was identified as the catalytic machinery in serine proteases,[1] lipases,[2,3] and acetylcholinesterase,[4] but many other hydrolytic enzymes also display the same acid−base−Ser/Thr pattern in their active sites.[5] An important part of enzyme evolution may have been to bring the active serine side chain close to histidine, acting as the general base, and aspartate, stabilizing the protonated form of imidazole. Amino acids belonging to the triad lie far in the primary sequence, and protein folding allows them to form an integrated entity with a (− + −) charge distribution during formation of the tetrahedral intermediate. In an early paper, we stressed the significance of another type of enzyme evolution: providing an appropriate template for the active site that stabilizes the transition state via electrostatic (hydrogen-bonding and charge−charge interaction) forces.[6] Conservation of electrostatic patterns in enzymes, visualized by electrostatic potential maps,[7] has been discussed for trypsin isozymes,[8] lysozymes,[9] and superoxide dismutase variants.[10] In a recent preliminary paper,[11] we discussed further examples, like xylose isomerase, an enzyme where a similar (− + −) charge pattern forms in a catalytic reaction step that is, however, not rate limiting.[12]

In the following, we report on electrostatic calculations for a number of enzymes with a (− + −) charge distribution at their active sites during catalysis. Five of them belong to the family of serine proteases and two, acetylcholinesterase and pancreatic lipase A, are esterases, whereas lysozyme and D-xylose isomerase act on carbohydrates: the former utilizes an aspartate and a glutamate side chain for the stabilization of the carbocation of the substrate lying between them, while the latter stabilizes it via an Asp^-···$HisH^+$ dyad with the positive end pointing toward the negatively charged substrate group. As we will point out, there is a close analogy among these enzyme active sites that can be quantified.

Table 1. Protein Data Bank Entries Used for Construction of Geometric Models

enzyme	code	ligand	authors
α-chymotrypsin	6CHA	phenylethaneboronic acid	Tulinsky and Blevins[27]
β-trypsin	1TPP	*p*-amidinophenyl pyruvate	Walter et al.[28]
α-lytic protease	2ALP	no	Fujinaga et al.[29]
subtilisin *Carlsberg*	1CSE	eglin	Bode et al.[30]
subtilisin *NOVO*	1SBN	eglin C	Gruetter et al.[31]
acetylcholinesterase	1ACE	no	Sussman et al.[32]
lipase	1CRL	no	Grochulski and Cygler[33]
lysozyme	1LZT	no	Hodson et al.[34]
xylose isomerase	3XIS	no	Whitlow and Howard[35]

MODELS AND METHODS

Three-dimensional structures of the enzymes treated in this study were taken from the Protein Data Bank (see Table 1).[13] For the five serine proteases depicted in Table 1, we

[†] Eötvös University Budapest.
[‡] Agricultural Research Center.

10.1021/ci980128o CCC: $18.00
Published on Web 02/13/1999

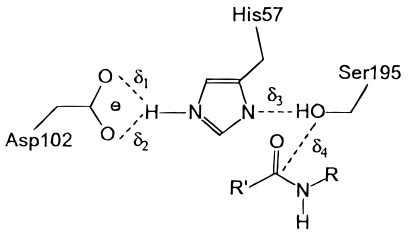


Figure 1. Active-site model for the serine proteases studied in this work. Constrained distances applied in the molecular mechanics and molecular dynamics calculations are denoted by δ_1, δ_2, δ_3, and δ_4, respectively.

Table 2. Constraints Applied in Molecular Mechanics and Molecular Dynamics Simulations (for Definition of Parameters, See Figure 1)

parameter	range, pm	opt value, pm
	Ground State	
δ_1	170−190	176
δ_2	205−225	215
δ_3	185−205	196
δ_4	285−305	291
	Transition State (Modeled by the Anionic Triad)	
δ_1	160−180	166
δ_2	235−255	244
δ_3	95−110	102
δ_4	145−152	148

calculated the electrostatic potential around the anionic triad formed by the aspartate, the adjacent histidine, and the tetrahedral intermediate. The latter one is a good model of the transition state[14] with supposedly minor difference in the position of the transferred proton between the enzymes as well. Though the energies of the two structures may differ and the difference may vary in the serine proteases studied here, the charge distributions are roughly similar and can be best characterized by the (− + −) pattern. A schematic representation of the ground state with distances is shown in Figure 1. We treated the above enzymes also at a more sophisticated level by calculating the electrostatic stabilization energies for the tetrahedral intermediate formed with Suc-Ala-Ala-Pro-Phe-pNA (Suc, succinyl; pNA, *p*-nitroaniline), a substrate for which experimental kinetic data are available for all five enzymes.[15−17]

Table 3: Calculated Total Energies (E), Electrostatic Complementarity Values (P), and Relative Stabilization Energies ($\Delta\Delta G$ in kJ/mol) of the Anionic Triad for Five Serine Proteases[a]

	E_{MM}	E_{MD}	P	$\Delta\Delta G_{MM}$	$\Delta\Delta G_{MD}$	$\Delta\Delta G_{meas}$
α-chymotrypsin	11 713	8 536	−5.1	−159.7	−71.9 (−73.7)[b]	−34.2[15]
β-trypsin	2 601	1 445	−1.2	−42.3	−40.7	−26.7[16]
α-lytic protease	625	531	8.0	−11.6	9.3	0.3[17]
Subtilisin *Carlsberg*	19 654	8 260	−6.9	45.7	−102.6	−36.5[15]
Subtilisin *NOVO*	16 311	14 402	−3.8	−43.7	−62.1	−31.6[15]

[a] Experimental activation energies, $\Delta\Delta G_{meas} = -2.303RT \log(k_{cat}/K_M)$, at 25 °C and pH = 8.00 are also given for comparison. MM and MD refer to active-site modeling by molecular mechanics and molecular dynamics, respectively (see text). [b] For a model containing all side chains of the oxyanion hole.

To get a preliminary model of the enzyme−substrate complex, we performed molecular mechanics optimization with the SYBYL molecular modeling package using the TRIPOS force field.[18] We applied the constrained energy minimization technique,[19] allowing the initial structure of the tetrahedral intermediate (His-57, Asp-102, Ser-195, substrate) and its environment within a sphere of 1-nm radius around the active oxygen atom of Ser-195 to relax. Then, we fitted the structure of the anionic triad to the ground state and restricted optimization to the active site to obtain our final model. The list of constrained geometric data applied in our calculations is given in Table 2. In all cases, we used a constraint force of 0.02 kJ/(mol·pm^2). The convergence criterion for geometry optimization was set to 0.0005 mdyn/pm.

Atomic charges for the tetrahedral intermediate were obtained by semiempirical AM1 molecular orbital calculations with the MOPAC package[20] on a model containing the substrate and the His-57, Asp-102, and Ser-195 side chains cut out of the protein and saturating their dangling bonds by hydrogen atoms. For protein atoms without the anionic triad

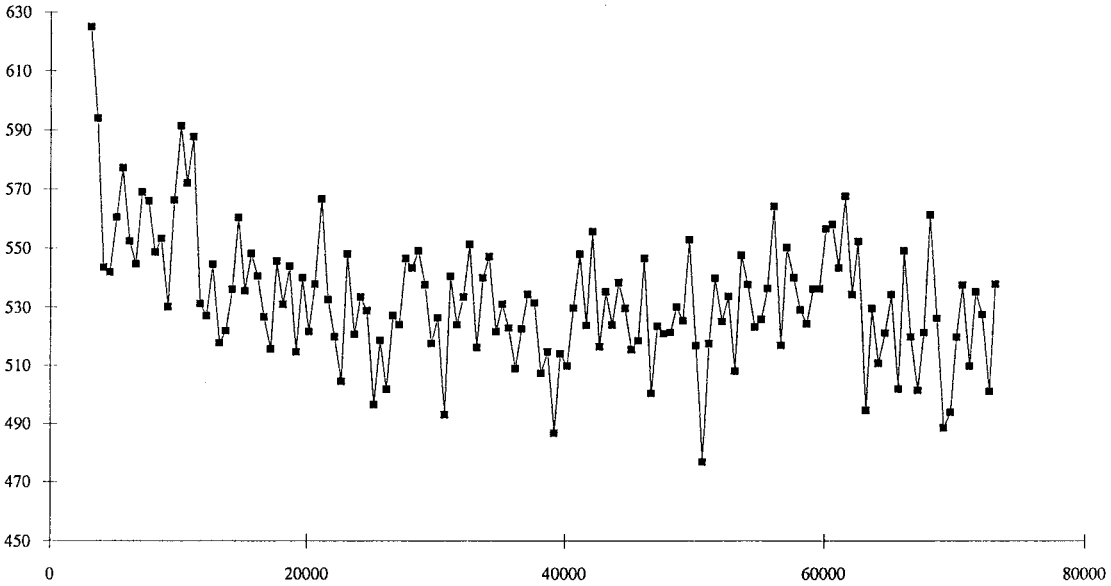


Figure 2. Trajectory for the molecular dynamics simulation of ligand binding by α-lytic protease. Units on the *y* and *x* axes are in kcal/mol and fs, respectively.

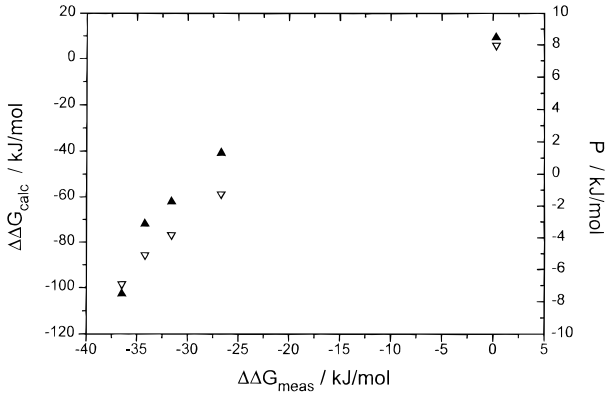
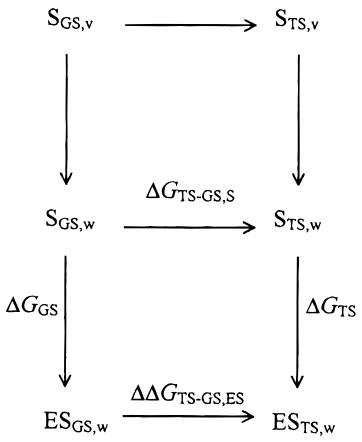


Figure 3. Plot of the calculated electrostatic binding energies ($\Delta\Delta G_{calc}$, full triangles, $r = 0.943$) and electrostatic complementarity values (P, empty triangles, $r = 0.993$) versus the experimental activation energies ($\Delta\Delta G_{meas}$) of the Suc-Ala-Ala-Pro-Phe-pNa substrate to serine proteases studied in this work. All entries are in kJ/mol.

Scheme 1



Kollmann, all type charges were used. To draw conclusions on the mere effect of the surrounding protein core, neither structural water molecules nor the groups forming the oxyanion hole were implicated in the calculations. However, to have an estimate on the effect of the oxyanion hole, we also made calculations for a model of α-chymotrypsin including oxyanion side chains, too. As is seen from Table 3, the relative electrostatic stabilization energy remains practically unchanged for this latter model. To obtain a better geometric model for the ground state, we carried out molecular dynamics (MD) simulations to find the appropriate binding position of the substrate by allowing a more complete relaxation. By using the structure minimized by molecular mechanics and applying the same constraints, we carried out a 100-ps MD simulation at 300 K with a distance-dependent dielectric constant, $\epsilon = 4$. A typical trajectory is depicted in Figure 2. In all cases, the MD-optimized structures had lower energies than those optimized by simple molecular mechanics (cf. Table 3). For subtilisin *Carlsberg*, the protein−active site interaction energy, as obtained by the simple molecular mechanics geometry optimization, is repulsive. This might be due to the fact that for this enzyme we used a model for the active site that was derived from the enzyme in a complex with eglin, a low molecular weight protein. In the case of the enzyme−eglin complex, the active-site structure considerably differs from that of the complexes with the small molecular weight substrate studied by us.

Electrostatic binding energy differences were calculated by the discretized Poisson−Boltzmann method[21,22] using the DelPhi package. The thermodynamic cycle, used in the calculations, is depicted in the Scheme 1.[23] Here $S_{GS,v}$ and $S_{TS,v}$ stand for the active site in its ground and transition states in vacuo (the latter modeled by the anionic triad), while $S_{GS,w}$ and $S_{TS,w}$ refer to the same states in water, respectively. $ES_{GS,w}$ and $ES_{TS,w}$ denote the ground and transition states of the enzyme−substrate complex in water. If the substrate molecule is identical for each calculation, we may neglect the upper cycle and have

$$\Delta\Delta G_{TS-GS,ES} = \Delta G_{TS} - \Delta G_{GS} + \Delta G_{TS-GS,S} \qquad (1)$$

where we consider $\Delta G_{TS-GS,S}$ to be constant for all enzymes since the substrates are the same.

In DelPhi calculations, we applied the same charges as in the simulations outlined above. We considered the protonation states corresponding to pH = 8.00 where the kinetic constants were measured, i.e., Asp(−1), Lys(+1), Arg(+1), Glu(−1), and His($+^1/_2$). An ionic strength of 0.145 M was used in all calculations. The protein and solvent dielectric constants were set to $\epsilon = 4$ and 80, respectively. We applied a focusing-type boundary condition consisting of four steps, and in the last step, we focused on the active site with a grid size of 20 pm.

For the calculation of electrostatic complementarity in point *i*, we used the formula proposed by Nakamura et al.[24]

$$P_i = \mathrm{sign}(V_i^P V_i^A)(V_i^P V_i^A)^{1/2} \qquad (2)$$

where V_i^P and V_i^A denote the molecular electrostatic potential in point *i* emerging from the protein environment (active site excluded) and the active site, respectively. Electrostatic complementarity between the active site and its environment is defined as an average for the set of N points, $\{i\}$, distributed evenly at a constant density as generated by the GRASP software[25] on the van der Waals surface of the former including only regions around potentially hydrogen-bonding atoms (N, O, and bound hydrogen for the present substrate)

$$P = \sum P_i/N \qquad (3)$$

A more negative value of P refers to better complementarity and corresponds to a larger value of the electrostatic interaction energy between associating partners.[26] In their original proposition, Nakamura et al.[24] averaged P_i on the whole van der Waals surface of the ligand. However, they argue that some unrealistic results may occur if the potential surface has regions with large positive or negative values, while other regions have values near zero. Since the electrostatic potential near nonpolar bonds involving carbon and hydrogen atoms is small, for a large substrate like Suc-Ala-Ala-Pro-Phe-pNa, P defined by the original formula suffers from the above problem and does not reflect the variation of enzyme environments.

RESULTS AND DISCUSSION

We compare enzyme active sites with the (− + −) charge distribution at three levels of sophistication. For five serine proteases, we have kinetic data for identical substrates; thus,

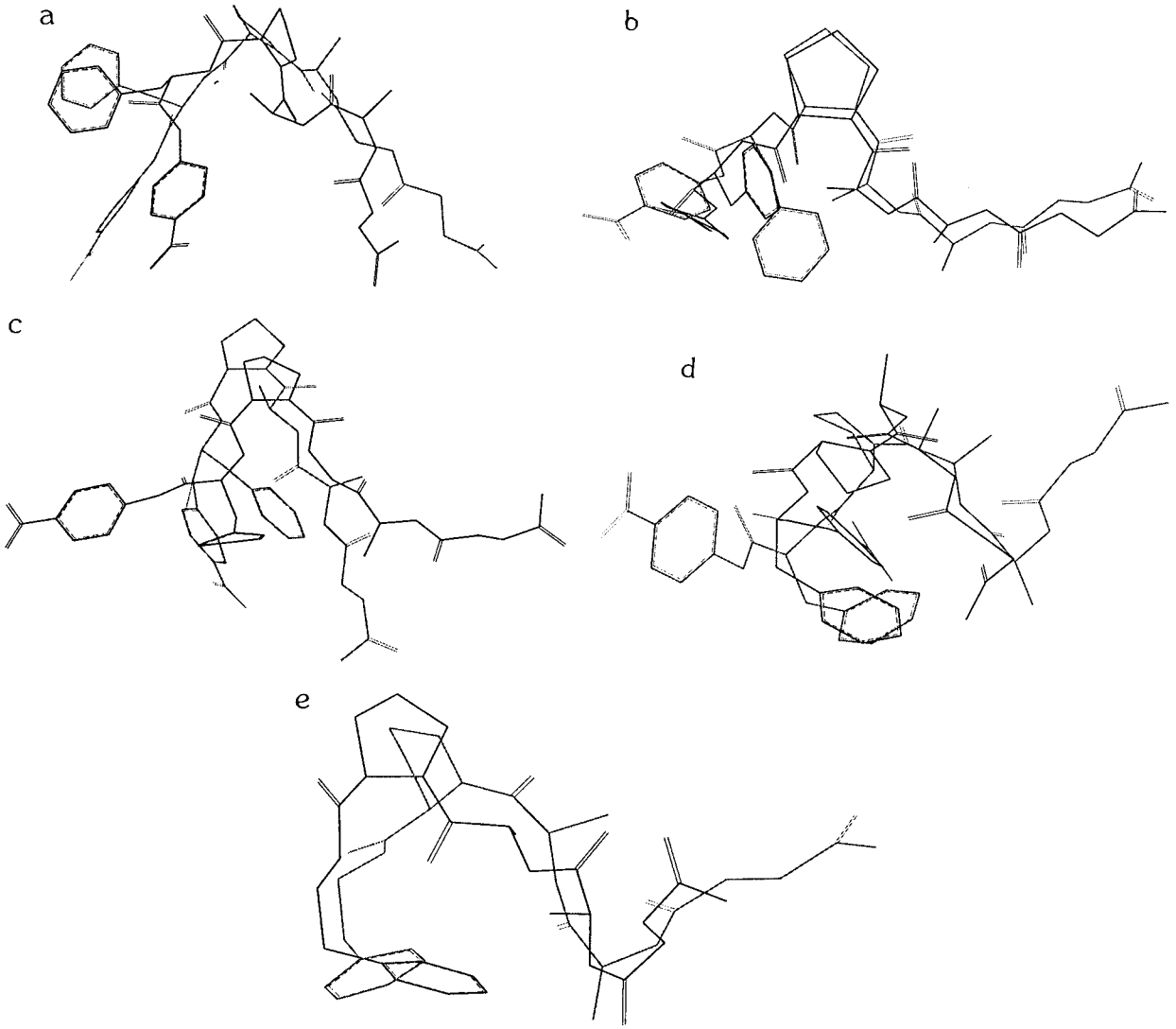


Figure 4. Superimposed substrate conformations obtained by molecular mechanics (light lines) and molecular dynamics (heavy lines) energy minimization. (a) α-Chymotrypsin, (b) *β*-trypsin, (c) α-lytic protease, (d) subtilisin *Carlsberg*, (e) subtilisin *NOVO*. Note the extended structure for α-lytic protease.

a quantitative comparison is possible via calculating the electrostatic stabilization energy of the transition-state structures modeled by the anionic triad. At the second, semiquantitative, level of approximation, we compare complementarity measures, defined in eq 3, for the same group of enzymes; at last, we present graphs of the electrostatic potential maps for all nine enzymes studied in this work.

The calculated total energies of the optimized enzyme−substrate complexes obtained both by molecular mechanics and by molecular dynamics calculations and the electrostatic binding energies of the anionic triads modeling the transition-state complexes as well as the experimental free energies of activation are given in Table 3, while Figure 3 displays their correlations. Note that neither E_{MM} nor E_{MD} is directly used in calculating $\Delta\Delta G_{MM}$ and $\Delta\Delta G_{MD}$, respectively. These entries represent merely the effect of the molecular dynamics simulation on the optimization of the substrate geometry. In all cases, we obtained lower energies with MD simulations, indicating that, most probably, they yield more reliable substrate-binding geometries than a simple optimization by molecular mechanics.

Though the absolute magnitudes of the calculated stabilization energies are overestimated by a factor of about 2,

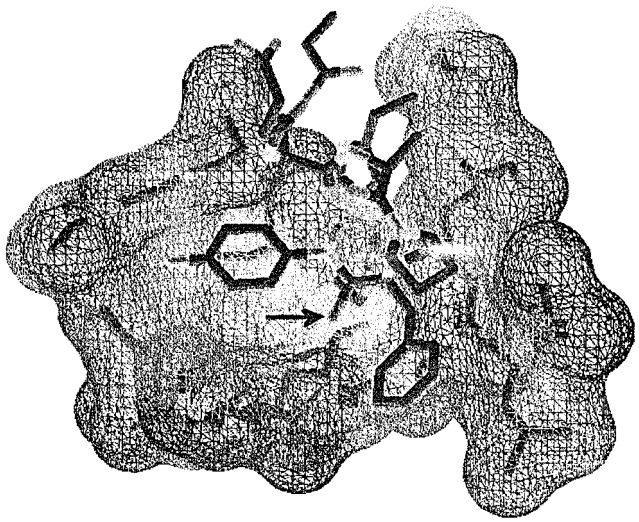


Figure 5. Electrostatic complementarity between α-lytic protease and the Suc-Ala-Ala-Pro-Phe-pNA substrate. Protein regions characterized by positive and negative potentials are denoted by dark and light shading, respectively. Note the negatively charged oxygen atom (marked with an arrow) of the tetrahedral intermediate clashing with a negative potential region of the enzyme.

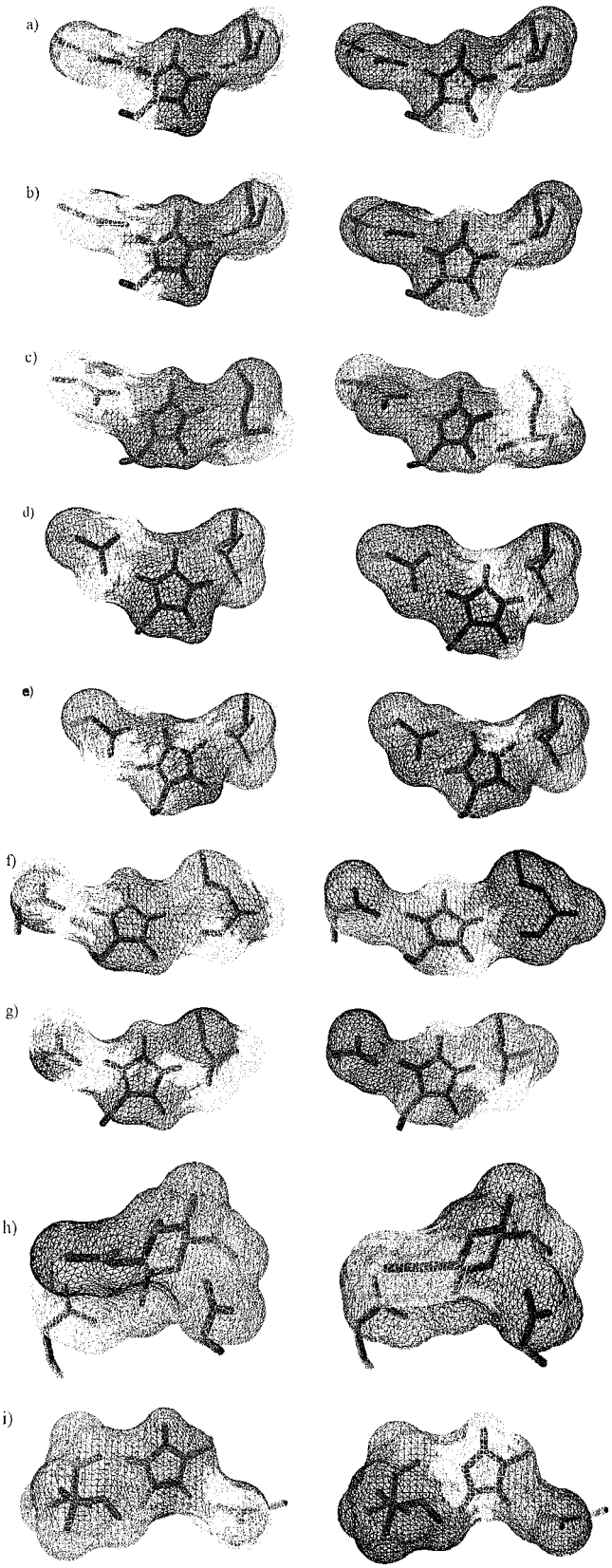


Figure 6. Electrostatic complementarity between enzyme active sites (left) and protein environment (right). Protein regions characterized by positive and negative potentials are illustrated by dark and light shading, respectively. (a) β-Trypsin, (b) α-chymotrypsin, (c) α-lytic protease, (d) subtilisin *Carlsberg*, (e) subtilisin *NOVO*, (f) acetylcholinesterase, (g) lipase, (h) lysozyme, (i) xylose isomerase.

the trends are well-reproduced. Deviations from linearity may be due to the difference between the charge distributions and geometries of the anionic triad and the transition state, deviations of the calculated substrate-binding geometry from the experimental one, and other factors that cannot be considered in the present study (e.g., entropy). Our results indicate that the variation of the protein environment around the bound substrate essentially influences the catalytic rate via its electrostatic effect. It is also seen in Table 3 that slight differences in the charge distribution of the anionic triad do not affect the electrostatic binding energies to a large extent. The catalytic activity of α-lytic protease on the studied substrate, Suc-Ala-Ala-Pro-Phe-pNa, is very low, and this is correctly reflected by the calculated active site−environment interaction energy, which is positive. This can be understood by inspection of the model of the enzyme−substrate complex that basically differs from other enzymes studied in this work (cf. Figure 4). While the structures of the enzyme−substrate complexes are similarly bent for α-chymotrypsin, β-trypsin, and subtilisin *Carlsberg* and *Novo*, in the case of α-lytic protease, the leaving group has a different, extended arrangement.

Table 3 also includes values of the electrostatic complementarity between the active site and its environment, as defined in eqs 2 and 3. As shown in Figure 3, a linear relation holds between P and the experimental stabilization energies, which indicates that the better the complementarity, the stronger the interaction between the active site and environment. It should be noted that the positive activation energy for α-lytic protease is also reflected by the calculated electrostatic interaction energy (cf. Table 3). A probable reason for that is the extended conformation of the substrate at the active site (cf. Figure 4c), resulting in a poor electrostatic complementarity with the enzyme (cf. Figure 5).

In Figure 6, we visualize the electrostatic complementarity on the van der Waals surfaces of various enzymes studied in this work. Matching of electrostatic potential patterns emerging from active sites and their environment is good, though not perfect. Complementarity of the electrostatic patterns is quite similar for all, otherwise very different, enzymes. Thus, we may say that the appropriate electrostatic complementarity between enzyme active sites and their environment is a prerequisite for the catalytic activity.

ACKNOWLEDGMENT

This work was supported by a grant from OTKA (National Fund for Scientific Research, Hungary), No. 22191.

REFERENCES AND NOTES

(1) Kraut, J. *Annu. Rev. Biochem.* **1977**, *46*, 331.
(2) Brady, L.; Brzozowski, M.; Derewenda, S.; Dodson, E.; Dodson, G.; Tolley, S.; Turkenburg, J.; Christiansen, L.; Huge-Jensen, B.; Norskov, L.; Thim, L.; Menge, U. *Nature* **1990**, *343*, 767.
(3) Winkler, F. K.; D'Arcy, A.; Hunziker, W. *Nature* **1990**, *343*, 771.
(4) Sussman, J. L.; Harel, M.; Frolow, F.; Oefner, C.; Goldman, A.; Toker, L.; Silman, I. *Science* **1991**, *253*, 872.
(5) Dodson, G.; Wlodawer, A. *Trends Biochem. Sci.* **1998**, *23*, 347.
(6) Ángyán, J.; Náray-Szabó, G. *J. Theor. Biol.* **1983**, *103*, 349.
(7) Náray-Szabó, G.; Ferenczy, G. G. *Chem. Rev.* **1995**, *95*, 829.
(8) Soman, K.; Yang, A. S.; Honig, B.; Fletterick, R. *Biochemistry* **1989**, *28*, 9918.
(9) Dao-Pin, S.; Liao, D. I.; Remington, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5361.
(10) Desideri, A.; Falconi, M.; Polticelli, F.; Bolognesi, M.; Djinovic, K.; Rotilio, G. *J. Mol. Biol.* **1992**, *223*, 337.

(11) Náray-Szabó, G.; Gérczei, T. *Croat. Chem. Acta* **1996**, *69*, 955.
(12) Fábián, P.; Asbóth, B.; Náray-Szabó, G. *THEOCHEM* **1994**, *307*, 171.
(13) Abola, E. E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. In *Crystallographic Databases—Information Content, Software Systems, Scientific Applications;* Allen, F. H., Bergerhoff, G., Sievers, R., Eds.; 1987; p 107.
(14) Asbóth, B.; Polgár, L. *Biochemistry* **1983**, *22*, 117.
(15) Stambolieva, N. A.; Ivanov, I. P.; Yomtova, V. M. *Arch. Biochem. Biophys.* **1992**, *294*, 703.
(16) Bone, R.; Fujishinge, A.; Kettner, C. A.; Agard, D. A. *Biochemistry* **1991**, *30*, 10388.
(17) Venekei, I.; Gráf, L.; Rutter, W. J. *FEBS Lett.* **1996**, *379*, 139.
(18) SYBYL Molecular Modeling Software, Version 6a; TRIPOS Associates: St. Louis, MO, 1993.
(19) Daggett, V.; Schröder, S.; Kollman, P. A. *J. Am. Chem. Soc.* **1991**, *113*, 8926.
(20) Stewart, J. J. P. *QCPE Bull.* **1989**, *9*, 10. QCPA Program 455, MOPAC Version 6.0.
(21) Gilson, M. K.; Sharp. K. A.; Honig, B. *J. Comput. Chem.* **1987**, *9*, 327.
(22) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins* **1986**, *1*, 47.
(23) Bashford, D. M.; Karplus, M. *Biochemistry* **1990**, *29*, 10219.
(24) Nakamura, H.; Komatsu, K.; Nakagawa, S.; Umeyama, H. *J. Mol. Graph.* **1985**, *3*, 2.
(25) Nicholls, A.; Honig, B. *GRASP: Graphical Representation and Analysis of Surface Properties*; Columbia University: New York, 1992.
(26) Douglas, J. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1980**, *102*, 4295.
(27) Tulinsky, A.; Blevins, R. A. *Protein Data Bank*, Entry 6CHA, 1992.
(28) Walter, J.; Bode, W.; Huber, R. *Protein Data Bank*, Entry 1TPP, 1992.
(29) Fujinaga, M.; Delbaere, L. T. J.; Brayer, G. D.; James, M. N. G. *Protein Data Bank*, Entry 2ALP, 1992.
(30) Bode, W.; Papaniokos, E.; Musil, D. *Eur. J. Biochem.* **1987**, *166*, 673.
(31) Gruetter, M. G.; Heinz, D. W.; Priestle, J. P. *Protein Data Bank*, Entry 1SBN, 1992.
(32) Sussman, J. L.; Harel, M.; Silman, I. *Protein Data Bank*, Entry 1ACE, 1992.
(33) Grochulski, P.; Cygler, M. *Protein Data Bank*, Entry 1CRL, 1992.
(34) Hodson, J. M.; Brown, G. M.; Sieker, L. C.; Jensen, L. H. *Protein Data Bank*, Entry 1LZT, 1992.
(35) Whitlow, M.; Howard, A. *Protein Data Bank*, Entry 3XIS, 1992.

CI980128O