The (– + –) Charge Distribution: A Common Pattern in the Transition State of Some Enzymes

Gábor Náray-Szabó and Timea Gérczei

Department of Theoretical Chemistry, Eötvös Loránd University Budapest, H-1117 Budapest, Pázmány Péter st. 2, Hungary

X-ray diffraction and molecular modelling studies in the last two decades have drawn attention to the (– + –) charge distribution that plays an important role in some enzymatic processes. In most cases two of the charged moieties are amino-acid side chains of the enzyme, where the positive and negative charges are provided by a protonated His and by the deprotonated carboxylic group of an Asp or Glu side chain, respectively. The third group is the substrate itself, getting negatively charged during the catalytic process. Several enzymes (*e.g.* serine proteases, acetylcholinesterase and lipases) make use of the Ser-His-Asp(Glu) triad as a central machinery in the catalysis offering the serine oxygen, which attacks the substrate to yield the tetrahedral intermediate, as the negative charge in the above pattern. In the ring opening step of xylose isomerase catalysis, this serine oxygen is replaced by the O1 oxygen atom of the glucopyranose substrate. In the case of lysozyme, an Asp and a Glu side chain encounter the positively charged sugar ring of the substrate, thus providing the (– + –) distribution. In the present paper, we discuss the role of the surrounding protein core in the electrostatic stabilization of the above pattern. We call attention to the possibility of convergent evolution which provides not only the conserved charge distribution but also a template by the protein environment stabilizing it electrostatically, *i.e.* through interactions between atomic net charges, hydrogen bonds and α-helix dipole effects.

INTRODUCTION

Since the proposition of the lock-and-key analogy by Emil Fischer,[1] numerous attempts have been made to understand the basis of enzymatic action. The key factor of rate acceleration by enzymes by a factor of 10^{10} or more has been thought to be the fit of transition state to the enzyme active site[2,3] and this view has not changed too much during decades.[4] An important paper, pointing to protein electrostatics as a major component of the fit (interaction) between the enzyme and its substrate, thus determining the catalytic activity, was published two decades ago.[5,6] Later on, the importance of electrostatic effects in enzyme catalysis has been stressed by several authors[7–10] and it has been shown that there is a wide class of enzymes for which the polarity of the transition state is much higher than that of the ground state and, thus, the major source of catalytic rate acceleration is its electrostatic stabilization by the protein environment.[11]

We have earlier called attention to the importance of the (− + −) charge distribution along the active site of serine proteases in electrostatic stabilization of the transition state.[12] In the paper cited, we compared the intraproteic electrostatic potential along the catalytic triad in seven species and, using our Bond Increment method for the calculation of electrostatic potentials,[10] we found a minimum-type pattern for all serine proteases under study. This means that in these enzymes the charge distribution, provided by the serine oxygen, as part of the tetrahedral intermediate, the protonated imidazolium and the deprotonated carboxylate side chains of the catalytic His and Asp residues, respectively, is electrostatically stabilized by the surrounding protein core. As it is known, the trypsin and subtilisin enzyme families followed a convergent evolution that resulted in a similar geometric arrangement of the catalytic triad.[13] Our early study supports the hypothesis that biological evolution of some serine proteases had to result not only in a similar active-site geometry but also in a similar intraproteic electrostatic potential that stabilizes the (− + −) charge distribution. Further evidence, confirming this statement, has been recently published for other enzymes.[14–16]

In this paper, we examine the protein electrostatic potential pattern along the (− + −) charge distribution formed at the active site of six serine proteases and, in addition to our early work,[12] acetylcholin-esterase, lipase A, lysozyme, and D-xylose isomerase. Our goal is to gain information on the validity of the hypothesis quoted above.

METHODS

We took three-dimensional enzyme co-ordinates from the following files of the Protein Data Bank:[17] α-chymotrypsin: 6CHA,[18] β-trypsin: 1TPP,[19] porcine pancreatic elastase: 4EST,[20] *Streptomyces Griseus* protease A: 3SGA,[21]

α-lytic protease: 2ALP,[22] subtilisin NOVO: 1SBN,[23] acetylcholin-esterase: 1ACE,[24] lipase A: 1CRL,[25] lysozyme: 1LZT,[26] and D-xylose isomerase: 3XIS.[27] The protein electrostatic potential was calculated at four points lying along the active site as defined in Table I. These points represent the major geometric features of the active site charge distribution and provide sufficient information on the potential variation, which can also be displayed on the van der Waals surface of the molecular fragments forming the active site. Work in this direction is in progress.

We used the SYBYL software package[28] for the calculation of the intraproteic electrostatic potential with Pullman net charges[29] on protein backbone and side-chain atoms and a distance-dependent dielectric constant, $\varepsilon = R$ where R is in Å units. Using the BASIC module, we set the charges on the active site atoms (CB and further atoms of amino-acid side chains with substrate attached) to zero and assigned probe charges of +1 to points A, B, C, and D (as defined in Table I). Then, we defined the whole

TABLE I

Definition of sites along the catalytic machinery of the investigated enzymes where the protein electrostatic potential was calculated

| enzyme	A	B	C	D
α-chymotrypsin β-trypsin porcine pancreatic elastase SGPA[*] α-lytic protease	OG of Ser-195	NE2 of His-57	ND1 of His-57	bisector of OD1 and OD2 of Asp-102
subtilisin NOVO	OG of Ser-221	NE2 of His-64	ND1 of His-64	bisector of OD1 and OD2 of Asp-32
acetylcholin-esterase	OG of Ser-200	NE2 of His-440	ND1 of His-440	bisector of OE1 and OE2 of Glu327
lipase A	OG of Ser-209	NE2 of His-449	ND1 of His-449	bisector of OE1 and OE2 of Glu-341
lysozyme	bisector of OD1 and OD2 of Asp52	first trisector of A and D	second trisector of A and D	bisector of OE1 and OE2 of Glu-35
D-xylose isomerase	O1 of cyclic xylose	NE1 of His-54	ND1 of His-54	bisector of OD1 and OD2 of Asp-57

Streptomyces Griseus protease A
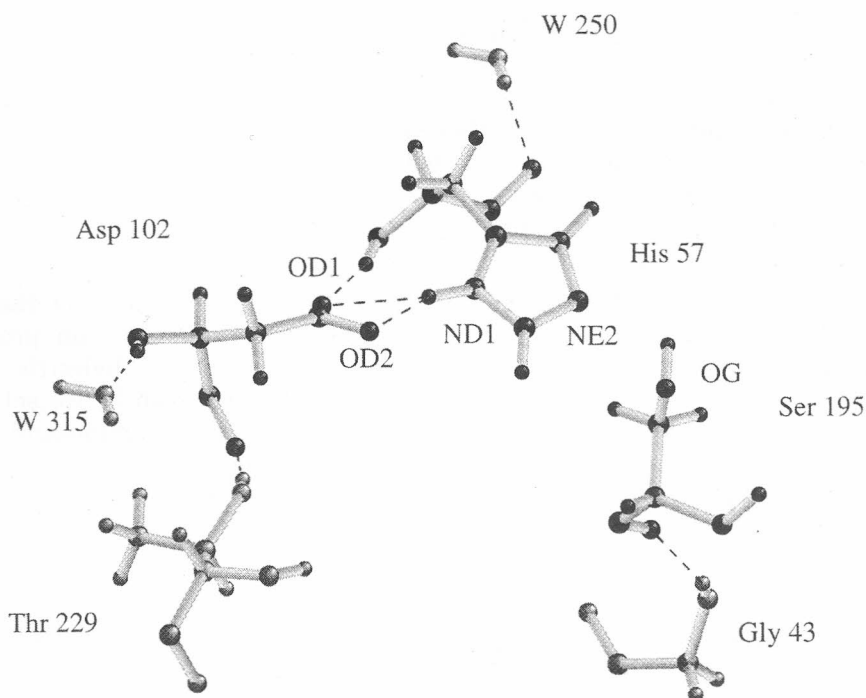
Figure 1. Amino-acid residues hydrogen-bonded (dashed lines) to those of the catalytic triad in β-trypsin. Model drawn using the co-ordinates of the Protein Data Bank File 1TPP[19] by the SYBYL software.[30]

protein as an aggregate, merged the probe charges in it and calculated the electrostatic interaction energy for the probe atom-protein complex yielding directly the protein electrostatic potential at this site. Work is in progress to repeat the above calculations through solution of the Poisson-Boltzmann equation.[30]

When calculating the electrostatic potential, we studied the contribution of various protein regions to the overall pattern in detail. We defined the close environment of the active site triad as those amino acids that are hydrogen-bonded to any atom of any of its residues (cf. Figure 1). Furthermore, we handled separately α-helices, which are known to provide strong electrostatic fields at their termini,[31] and the ensemble of surface ionizable side chains also having a considerable influence on the protein electrostatic potential.[30] We call these hydrogen-bond, helix and surface regions, respectively. The total intraproteic potential was obtained as a sum of the above terms added to the contribution of other enzyme atoms not included in the above classification.

RESULTS AND DISCUSSION

We display intraproteic electrostatic potential patterns for β-trypsin, subtilisin NOVO, acetylcholin-esterase, lipase A, lysozyme and D-xylose isomerase in Figures 2–7, respectively. These enzymes represent the different classes treated in this work. It is seen that hydrogen-bond, helix and surface regions provide patterns that are sometimes essentially different from the resulting one. In all the cases, except for β-trypsin and lysozyme, the contribution due to the hydrogen-bond is gradually decreasing. This is also true of α-chymotrypsin and α-lytic protease, not shown in the figures. The contribution due to the α-helix region is decreasing in acetylcholin-esterase, lysozyme and D-xylose isomerase, furthermore in *Streptomyces Griseus* hydrolase A and α-lytic protease, not presented in the figures. The contribution, due to the surface ionizable side chains, shows a variable trend. It may gradually decrease (in acetylcholin-esterase, lipase A and D-xylose isomerase) or show a minimum-type pattern as in β-trypsin, porcine pancreatic elastase and α-lytic protease. However, the sum of potentials showing variable patterns along the catalytic machinery (including the contribution of other protein atoms) provides in all, but one, cases a minimum-type curve (*cf.* Figure 8). It should be mentioned that in this work we do not consider electronic polarisation effects neither the effect of the water solvent surrounding the protein. Thus, the supposed additivity of the electrostatic potential contributions of hydrogen-bond, helix, surface and other regions of the protein is only an approximation.
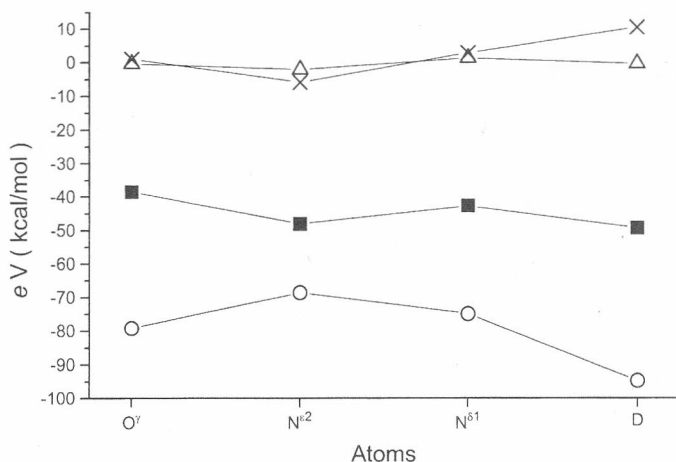

Figure 2. Protein electrostatic potential along the catalytic triad of β-trypsin.--o-- hydrogen-bond region, --Δ-- α-helix, --×-- surface ionizable side-chain contribution, --■-- total intraproteic potential.
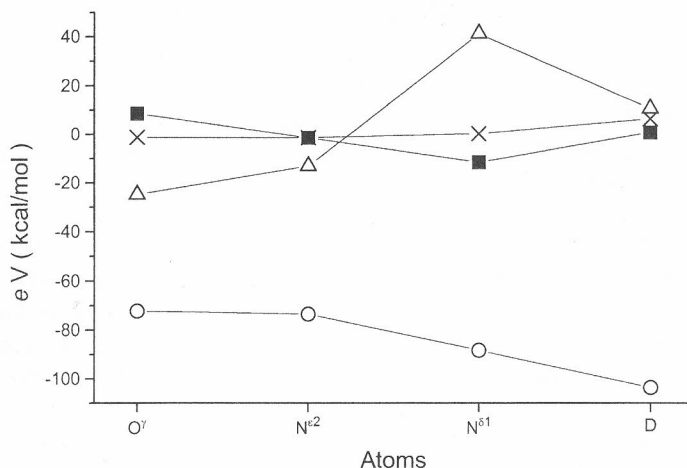
Figure 3. Protein electrostatic potential along the catalytic triad of subtilisin NOVO. Notations as in Figure 2.

The only exception from the overall trend is β-trypsin (*cf.* Figure 2) for which the intraproteic electrostatic potential decreases as approaching the bisector of atoms OD1 and OD2 of Asp-102 from the direction of ND1 of His-57. At first glance, this is at variance with our previous results[12] where we found a minimum-type potential along the catalytic triad, but with a strong
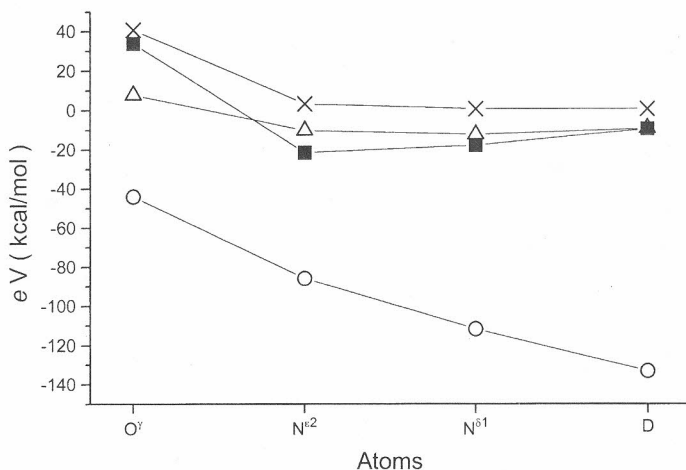

Figure 4. Protein electrostatic potential along the catalytic triad of acetylcholin-esterase. Notations as in Figure 2.
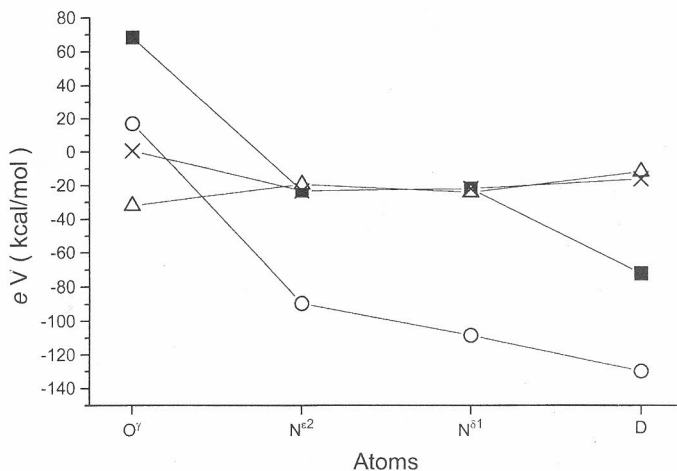
Figure 5. Protein electrostatic potential along the catalytic triad of lipase A. Notations as in Figure 2, crosses and full squares inter changed.

decrease at the bisector of the OD1 and OD2 atoms of the Asp-102 side chain. If we drop point B from the potential curve in Ref. 12, we get a similar, yet minimum-type pattern, as in Figure 2. The slight difference may be due to the different method and model, applied in this work. Further studies using more sophisticated computational methods are needed to clarify this point.
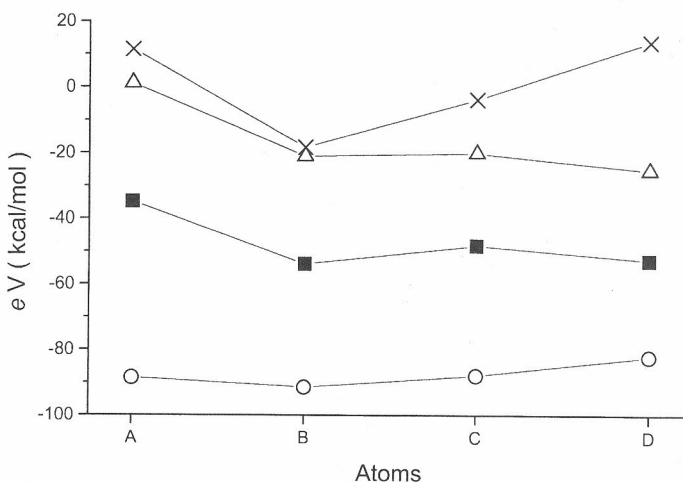

Figure 6. Protein electrostatic potential along the catalytic triad of lysozyme. Notations as in Figure 5.
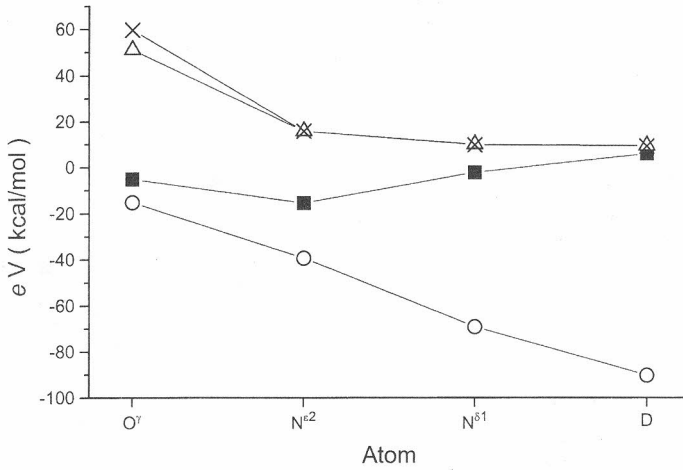
Figure 7. Protein electrostatic potential along the catalytic triad of D-xylose isomerase. Notations as in Figure 2.

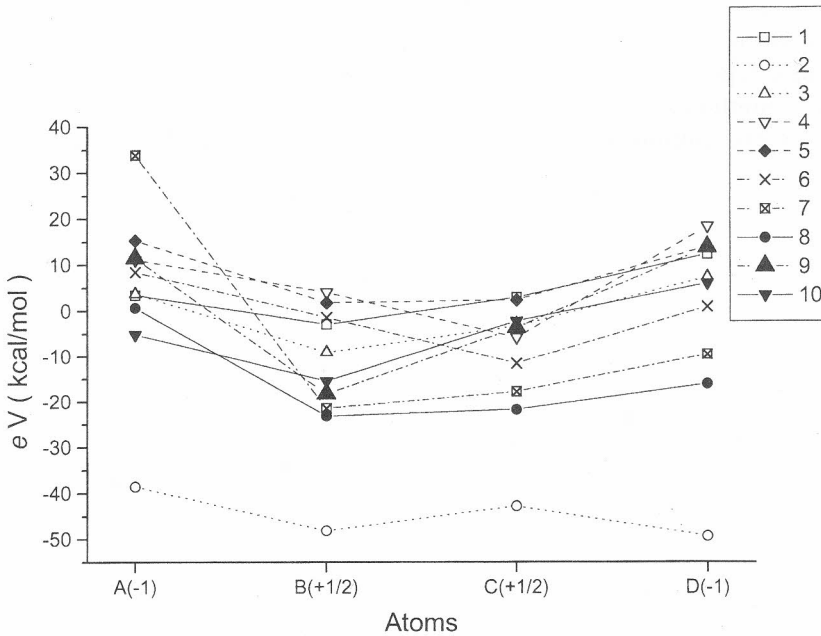Figure 8. Comparison of intraproteic electrostatic potential patterns for the enzymes studied in this work. **1**: α-chymotrypsin, **2**: β-trypsin, **3**: porcine pancreatic elastase, **4**: *Streptomyces Griseus* hydrolase A, **5**: α-lytic protease, **6**: subtilisin NOVO, **7**: acetylcholin-esterase, **8**: lipase A, **9**: lysozyme, **10**: D-xylose isomerase.

In Table II, we compared the approximate electrostatic energy of the catalytic site inside the enzyme. We assigned charges of $-e$, $+1/2e$, $+1/2e$, and $-e$ to points A, B, C, and D, respectively, and calculated the corresponding electrostatic energy as $E(- + -) = e(-V_A + 0.5V_B + 0.5V_C - V_D)$, where e is the unit charge and V_X denotes the intraproteic potential at point X. The results are summarized in Table II. This quantity is compared with the electrostatic energy of the (0 0 −) charge distribution, $E(0\ 0\ -) = -eV_D$. The difference, $\Delta = E(0\ 0\ -) - E(- + -)$, is positive, indicating stabilization of the (− + −) charge pattern for all enzymes studied. Thus, our rough preliminary calculations indicate that the intraproteic electrostatic field stabilizes the

TABLE II

Gross stabilization of the (− + −) vs. the (− 0 0) charge distribution by the protein environment (electrostatic energies and in kcal/mol)

enzyme	(− + −)	(0 0 −)	Δ
α-chymotrypsin	−16	−12	4
β-trypsin	43	50	7
porcine pancreatic elastase	−17	−7	25
Streptomyces Griseus protease A	−30	−18	12
α-lytic protease	−27	−14	13
subtilisin *NOVO*	−16	−1	15
acetylcholin-esterase	−44	10	34
lipase A	−7	16	23
lysozyme	−36	−14	22
D-xylose isomerase	−10	−6	4

catalytic triad in its transition state, *i.e.* protein electrostatic is a major factor in enzymatic rate acceleration (*cf.* Ref. 11 for further examples). Another lesson we can learn from these studies is that protein evolution followed a way that should ensure not only the appropriate steric arrangement of the catalytic triad (Ser...His...Asp or Glu in serine proteases, acetylcholinesterase and lipases, Asp...substrate...Glu in lysozyme and Asp...His...substrate in D-xylose isomerase) but also an intraproteic electrostatic potential ensuring its stabilization in the transition state. Arrangement of amino-acid residues directly hydrogen-bonded to the catalytic machinery, α-helices and surface ionisable side chains are essential determinants of this minimum-type pattern. Though their contributions are of various character, sometimes they provide minimum-type potential curves, and sometimes they do not, the overall effect is always stabilizing.

CONCLUSIONS

In this paper, we call attention to the effect of protein electrostatic potential on the energy of the transition state of the enzymatic reaction. In all the cases studied here, the effect is stabilizing, *i.e.* there is a net electrostatic energy gain for the highly polar transition state as compared to the less polar ground state. We have, thus, provided a further evidence that protein electrostatic potentials stabilize transition states and thereby accelerate the enzymatic process.

REFERENCES

1. E. Fischer, *Ber. Deutsch. Chem. Ges.* **27** (1894) 2984.
2. J. B. S. Haldane in: *Enzymes,* Longmans, London, 1930, p. 182.
3. L. Pauling, *Chem. Engn. News* **24** (1946) 1375.
4. R. L. Schowen in: *Transition States of Biochemical Processes,* R. D. Gandour and R. L. Schowen (Eds.), Plenum, New York, 1978, Ch. 2.
5. A. Warshel and M. Levitt, *J. Mol. Biol.* **103** (1976) 227.
6. A. Warshel, *Acc. Chem. Res.* **14** (1981) 284.
7. D. M. Hayes and P. A. Kollman, *J. Am. Chem. Soc.* **98** (1976) 7811.
8. M. F. Perutz, *Science* **201** (1978) 1187.
9. W. G. J. Hol, P. T. van Duijnen, and H. J. C. Berendsen, *Nature* **273** (1978) 443.
10. G. Náray-Szabó, *Int. J. Quant. Chem.* **16** (1979) 265.
11. G. Náray-Szabó, M. Fuxreiter, and A. Warshel in: *Theoretical Aspects of Biochemical Reactivity,* G. Náray-Szabó and A. Warshel (Eds.), Kluwer, Dordrecht, 1996, Ch. 7.
12. J. Ángyán and G. Náray-Szabó, *J. Theor. Biol.* **103** (1983) 349.
13. J. Kraut, *Annu. Rev. Biochem.* **46** (1977) 331.
14. S. Dao-Pin, D. I. Liao, and S. J. Remington, *Proc. Natl. Acad. Sci. U.S.A.* **86** (1989) 5361.
15. A. Desideri, M. Falconi, F. Polticelli, M. Bolognesi, K. Djinovic, and G. Rotilio, *J. Mol. Biol.* **223** (1992) 337.
16. D. R. Knighton, C. C. Kan, E. Howland, C. A. Janson, Z. Hostomska, K. M. Welsh, and D. M. Matthews, *Nature Struct. Biol.* **1** (1994) 186.
17. E. E. Abola, F. C. Bernstein, S. H. Bryant, T. F. Koetzle, and J. Weng in: *Crystallographic Databases – Information Content, Software Systems, Scientific Applications,* F. H. Allen, G. Bergerhoff, and R. Sievers (Eds.), Data Commission of the International Union of Crystallography, Bonn, 1987, p. 107.
18. A. Tulinsky and R. A.Blevins, Protein Data Bank entry 6CHA, Summer Version, 1992.
19. J. Walter, W. Bode, and R. Huber, Protein Data Bank entry 1TPP, Summer Version, 1992.
20. L. H. Takahashi, R. Radhakrishnan, R. E. Rosenfield, jr., and E. F. Meyer, jr., Protein Data Bank entry 4EST, Summer Version, 1992.
21. A. R. Sielecki and M. N. G. James, Protein Data Bank entry 3SGA, Summer Version, 1992.
22. M. Fujinaga, L. T. J. Delbaere, G. D. Brayer, and M. N. G. James, Protein Data Bank entry 2ALP, Summer Version, 1992.

23. M. G. Gruetter, D. W. Heinz, and J. P. Priestle, Protein Data Bank entry 1SBN, January, 1994.

24. J. L. Sussman, M. Harel, and I. Silman, Protein Data Bank entry 1ACE, Summer Version, 1992.

25. P. Grochulski and M. Cygler, Protein Data Bank entry 1CRL, January, 1994.

26. J. M. Hodsdon, G. M. Brown, L. C. Sieker, and L. H. Jensen, Protein Data Bank entry 1LZT, Summer Version, 1992.

27. M. Whitlow and A. Howard, Protein Data Bank entry 3XIS, Summer Version, 1992.

28. SYBYL Version 6.0a, TRIPOS Associates, St. Louis, MO, February 1993.

29. H. Berthod and A. Pullman, *J. Chem. Phys.* **62** (1965) 942.

30. K. A. Sharp and B. Honig, *Annu. Rev. Biophys. Chem.* **19** (1990) 301.

31. W. G. J. Hol, P. T. van Duijnen, and H. J. C. Berendsen, *Nature* **273** (1978) 443.

SAŽETAK

(− + −) raspodjela naboja: Uobičajeni obrazac u enzimatskim prelaznim stanjima

Gábor Náray-Szabó i Timea Gérczei

U protekla dva desetljeća proučavanja, difrakcijom X-zraka i molekularnim modeliranjem, svratila se pozornost na (− + −) raspodjelu naboja koja igra važnu ulogu u nekim energijskim procesima. U većini slučajeva, dva od nabijenih dijelova su aminokiselinski lanci sa strane enzima, gdje pozitivne i negativne naboje osiguravaju protonirane odnosno deprotonirane karboksilne skupine iz Asp ili Glu postranog lanca. Treća grupa je sam substrat koji postaje negativno nabijen tijekom katalitičkog procesa. Neki spojevi (*e.g.* serin proteaze, acetilholinesteraze i lipaze) koriste Ser-His-Asp(Glu) trijade kao središnje ustrojstvo u katalizi.

U slučaju lizozima dva Asp postrana lanca uključuju pozitivno nabijeni šećerni prsten iz substrata, osiguravajući tako (− + −) raspodjelu. U ovom radu diskutiramo ulogu okružujuće proteinske sredine u elektrostatskoj stabilizaciji ove sheme. Pažnja će biti skrenuta na mogućnost konvergentne evolucije koja osigurava ne samo sačuvanu raspodjelu naboja nego i proteinsko okružje stabilizirajući ga elektrostatski, odnosno kroz interakcije između čistih atomskih naboja vodikovih veza i α-heliks dipolnih efekata.