Electrostatic complementarity in molecular associations

Gábor Náray-Szabó

CHINOIN Pharmaceutical and Chemical Works, Budapest, Hungary

In this paper, I attempt to summarize the main qualitative features of electrostatic complementarity and similarity, important determinants of molecular recognition. The two aspects, Coulombic and hydrophobic matching, can be formulated in terms of molecular electrostatic potentials and fields. The Coulombic aspect is equivalent to the requirement to produce a potential pattern in the host cavity that is opposite in sign to that emerging from a guest. Hydrophobic complementarity is best described by the similis simili gaudet principle. This means that field patterns near the interacting molecular surfaces must be of similar magnitude. The above rules, which may find useful application in molecular graphics, were studied for different cases of enzyme-ligand interactions in trypsin. A further example, a noncovalent structural model of the catalytic diad in Streptomyces Griseus protease A, supports the observation that the same molecular entities form similar associations even in different environments, as is the case in the complex of small species in a crystal and amino acid residues with structural water molecules in a protein.

Keywords: electrostatic potential and field, recognition complementarity, similarity, trypsin, Streptomyces Griseus *protease A*

INTRODUCTION

Molecular recognition is one of the primary events in host-guest complexation, biopolymer-ligand interactions and other important molecular processes.[1] Before a ligand can bind, there must be a relatively rigid, complementary crevice or cavity inside a molecule or molecular ensemble. Host-guest binding is thus analogous to the fit between a lock and its key, an idea first formulated by Emil Fischer 100 years ago. The main factor of recognition is clearly steric. The cavity has to complement the guest in shape, at least to some extent, in order to be able to accommodate it. Since electrostatics plays a major role in association of molecules,[2,3] it should be also considered in order to describe the recognition process adequately.

Complementarity is rarely perfect; more than one ligand may fit quite well in the same crevice. On this basis, a definition of similarity can be formulated. A group of molecules can be called similar if they are all complementary to the same host site. Through this definition, both complementarity and similarity are related to molecular recognition, so we treat them jointly.

This paper focuses on the electrostatic aspects of the problem. Examples for geometric complementarity and similarity may be found in other works.[4,5] After outlining the background, I will discuss and illustrate cases of complementarity and similarity.

Address reprint requests to Dr. Náray-Szabó, Chinoin Pharmaceutical and Chemical Works, H-1325 Budapest, P.O. Box 110, Hungary.

This paper was presented at the Symposium on Molecular Recognition, held in Sopron, Hungary, 24–27 August 1988.

THEORY

Complementarity is mostly determined by three major factors: steric, Coulombic and hydrophobic. In this paper, I will not consider dynamic and entropy aspects that may also play a role in molecular recognition.[6,7] Most important is the steric fit. Once this is ensured, the interacting species must match electrostatically, too (i.e., the Coulombic interaction between them should be attractive). Hydrophobic complementarity represents minimization of dehydration energies and can be formulated as the matching between regions of host and guest that are of similar polarity (nonpolar-nonpolar, polar-polar or charged-charged). Since polarity of a certain molecular region is well characterized by the electrostatic field here, the hydrophobic aspects of complementarity can be discussed in terms of field matching.

Let us discuss the above three criteria on a somewhat more quantitative footing. The host-guest interaction free energy can be written as follows:

$$\delta G_{int} = \delta H_{vac} - T\delta S_{vac} - \delta H_{solv} - T\delta S_{solv} \quad (1)$$

where δH_{vac} and δH_{solv} are the interaction energy *in vacuo* and the solvation energy change upon association. $T\delta S_{vac}$ and $T\delta S_{solv}$ are the corresponding entropy terms. δH_{vac} can be decomposed as follows:[8]

$$\delta H_{vac} = \delta H_{Coul} + \delta H_{ind} + \delta H_{ct} + \delta H_{ex} + \delta H_{disp} \quad (2)$$

The indices correspond to Coulombic, inductive, charge-transfer, exchange repulsion and dispersion terms, respectively.

δH_{ex} and δH_{disp} are responsible for the steric fit. The exchange repulsion energy increases drastically if two nonbonded atoms get closer to each other than the sum of their van der Waals radii. On the other hand, δH_{disp}

represents attractive and nondirectional dispersion forces depending on inverse higher powers of the interatomic distance. Its value becomes optimal if the crevice is filled by the ligand atoms as perfectly as possible. In an aqueous medium, this is explained in terms of density differences between water and the host, the latter being more dense; therefore, interacting atoms may get closer to each other than in the hydrated case.[9] As a superposition of the above two effects, the steric interaction energy will be optimal if the steric complementarity is perfect.

The requirement of Coulombic fit is accounted for by δH_{Coul} in equation (2). Once the crevice is more or less perfectly filled by the ligand, the Coulombic interaction between the host and the guest has to be an optimum (i.e., positively and negatively charged or inversely polarized groups should get as close to each other as possible). The interaction energy can be approximated by the following empirical formula:[10]

$$\delta H_{Coul} = \text{const.} \Sigma V_h(i) \times V_g(i) \quad (3)$$

where $V_h(i)$ and $V_g(i)$ are the magnitudes of electrostatic potential produced by the host and the guest in appropriate reference points, i. Strictly physically, equation (3) has no meaning, since the potential product may not have an energy dimension, but heuristically it may be supposed that $V_h(i)$ [or $V_g(i)$] is roughly proportional to an appropriately defined charge at point i. This is why it works in a considerable number of cases.[11] After further simplification, we can formulate electrostatic complementarity as the requirement to produce a potential pattern by the guest, whose sign is opposite that belonging to the host in as many reference points as possible.[12]

Hydrophobic complementarity is related mainly to the solvation free energy term in equation (1), $\delta H_{solv} - T\delta S_{solv}$. Since most interesting cases of host-guest associations take place in an aqueous medium, here we will consider only water as a solvent. It is known that the hydration energy of a solute at a given site is approximately proportional to the value of the molecular electrostatic field here.[13] Therefore, low-field regions of the interacting molecules, where the hydration energy is small, tend to associate in order to minimize unfavorable entropy effects by extruding water from the contact surface. On the other hand, high-field regions also look for each other if they are associated with oppositely charged groups in the complex.

We do not consider $T\delta S_{vac}$ in equation (1), which is connected (e.g., with freezing of rotational degrees of freedom upon association). However, rigid, entropically unfavored structures ensure better complementarity, since the free energy change due to the above freezing effect can be neglected in the host-guest complexation. Two further terms in equation (2), δH_{ind} and δH_{ct}, are also neglected; they vary approximately parallel to the Coulombic term or are of minor importance in most interesting cases.

Let us characterize potential and field patterns of the interacting molecules by using reference points defined as follows.[14,15] They are located near potential hydrogen-bonding sites and hydrophobic CH-bonds inside the van der Waals envelope, in hypothetical lone-pair centers 100 or 150 pm apart from N, O, F or S, Cl atoms, respectively, and in the direction of CH, NH, OH or SH bonds at a distance of 100 pm from the hydrogen atom. We call the host and guest complementary to each other if the above-defined potential and field patterns match. Matching means that in complementary reference points of the host and the guest, the potentials are of opposite sign and the fields are of the same order of magnitude. Molecular electrostatic potentials were calculated using our bond increment method.[16-19]

COMPLEMENTARITY

Coulombic: Modified inhibitors of trypsin[20]

Analysis of the electrostatic pattern in the P3P2P1 (Pro-13, Cys-14, Lys-15) contact region in the complex between basic pancreatic trypsin inhibitor (BPTI) and trypsin has shown that there is no Coulombic complementarity in the peptide carbonyl regions. Both enzyme and inhibitor potentials are negative in the corresponding reference points. A fragment of the enzyme-inhibitor contact surface is depicted in Figure 1, where carbonyl reference points bear the number 4, 5, 10 and 11. The absence of complementarity is due to two factors. First, the positive charge on Lys-15 in BPTI is insufficient to overbalance the negative potential in the lone-pair region around backbone carbonyl atoms. Second, the arrangement of proton donor and acceptor groups in the interacting molecules is not suitable for an optimal hydrogen bonding. As seen in Figure 1, the geometry of the NH(Gln-192)...OC(Cys-14) and NH(Gly-216 backbone)...OC(Pro-13) hydrogen bonds is considerably distorted.

The imperfection in the Coulombic fit in the lone-pair region is partly overbalanced by the binding of a structural water molecule, W403, to the contact surface. The overall electrostatic fit in the new reference points *a*, *b*, and *c*, located around W403, is considerably better (c.f. Figure 1). This suggests that the >C=O...HOH structural unit might be replaced by a $>CHCH_2$ CH_2 OH group. Clearly, this is not possible for BPTI, but, for example, the corresponding $P_1P_2P_3$ tripeptide aldehyde, which is known to be a good inhibitor,[21] could be modified. Adding two methyl units to the carbonyl would have a further advantage. W403 could be expelled from the contact region to bulk water, resulting in a considerable entropy gain. A similar incorporation of a structural water into the inhibitor is thought to be responsible for the especially strong binding of pepstatin to pepsin and penicillopepsin.[22]

The above case may serve as an example of a novel type of nonclassical bioisosterism. Table 1 summarizes further possible isosteric replacements. The gain in the free energy is estimated after Andrews *et al.*,[23] who determined average binding energies of common functional groups to estimate the goodness of fit of a drug to the receptor. The large increase in δG for the elongation of the P_1 side chain by two methylene units is

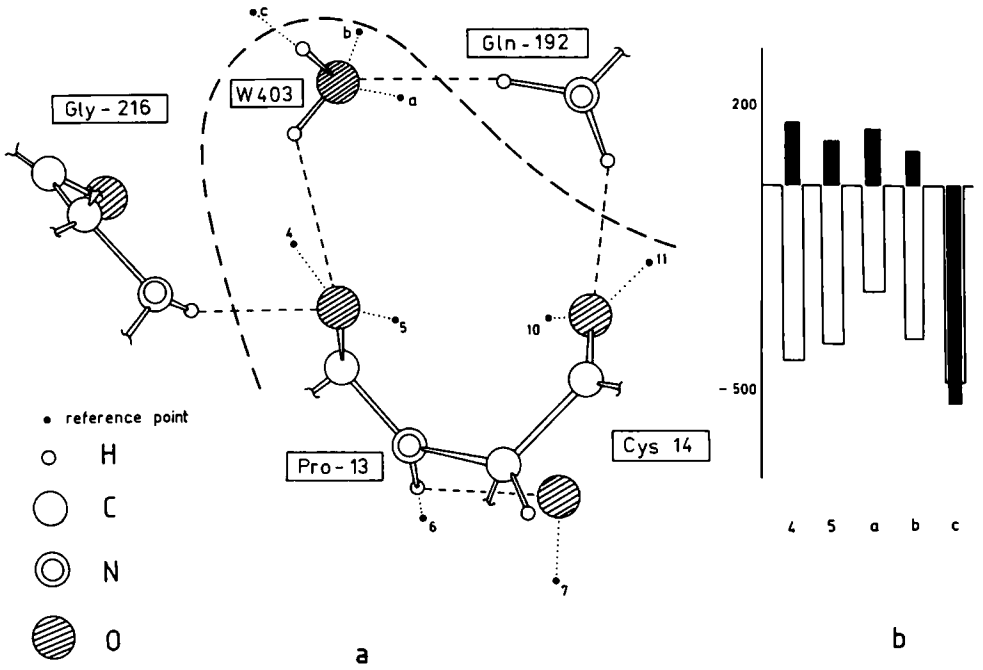

Figure 1. (a) Geometric arrangement of the Pro-13 region of BPTI in the trypsin-inhibitor complex. Hydrogen bonds and atoms of the hypothetical isostere are indicted by dashed lines. (b) Electrostatic complementarity around the carbonyl group (reference points 4 and 5) and the isostere (reference points a, b and c). Empty bars refer to the enzyme, while full bars represent the potential (with opposite sign) of the ligand

due to the probable displacement of W414 and W416 from the contact region to bulk water. Note that δG is an average estimate. Since the Coulombic fit is better for the isostere than for the parent compound, the true values should be larger for the backbone carbonyls of the P_2 and P_1 sites; they may become positive.

Hydrophobic: Specificity of trypsin and subtilisin mutants

Investigating the specificity of the Asp/Ser-189 mutant of trypsin toward various substrates, it was found that the value of log k_{cat}/K_m, the measure of the activation free energy of the enzymatic reaction, is larger for charged-charged and polar-polar pairs of enzyme and substrate side chains than for charged-polar ones.[24] This may be unusual, since it is known that in the gas phase the charged-polar (monopole-dipole) interaction is stronger than the polar-polar (dipole-dipole) one.[8] A

Table 1. Possible isosteric replacements in tripeptide aldehyde inhibitors of trypsin. δG is the lower bound for the increase in free energy of binding (kJ/mol)

Residue	Group	Isostere	δG
Pro-13 (P_3)	>CO	$>CH(CH_2)_2$ OH	+6
Cys-14 (P_2)	>CO	>CHOH	−7
Lys-15 (P_1)	>CO	>CHOH	−7
	$-NH_3^+$	$-(CH_2)_2NH_3^+$	+13

further manifestation of the effect has been the considerable increase of log k_{cat}/K_m for the Ser-189...Lys(P1) pair with increasing pH. Experimental data by Wells and coworkers for single and double mutants in the specificity pocket of subtilisin[25,26] further support our hypothesis, called the *similis simili gaudet* (similiar likes similar) principle—that side chains of similar nature (charged, polar or nonpolar) tend to associate better in the enzyme-substrate complex than different ones.[24,27]

Though recently it became possible to predict enzyme specificities quantitatively,[6,7] it is not without interest to use the concept of hydrophobic complementarity to rationalize the above principle. We use the average electrostatic fields near protein side chains calculated from the point charges by Weiner *et al.*[28] on van der Waals surface points defined by the Connolly algorithm[29] (c.f. Table 2). Note that side chains can be grouped according to their F_{av} values. The average field is the largest for charged, of medium magnitude for polar and small for nonpolar side chains. Maybe Tyr is the only case for which a small F_{av} is obtained, though it is considered to be rather polar than nonpolar. Electrostatic fields of the enzyme pocket and the substrate have to complement each other; those side chains associate stronger that have similar F_{av} values.

Color Plates 1 and 2 illustrate the above rule. In native trypsin the charged Lys P_1 side chain of the substrate has a charged Asp-189 counterpart in the specificity pocket, both producing strong fields (displayed by deep blue on Color Plate 1). In the Asp/Ser-189 mutant, the

complementarity is absent, and the field around Ser is smaller (displayed by light blue near the deep blue surface of Lys in Color Plate 2). Owing to the lack of complementarity, log k_{cat}/K_m is smaller in the mutant by 6.6 kcal/mol than in the wild-type enzyme.[24] If the Lys side chain of the substrate at the P1 site is deprotonated, log k_{cat}/K_m increases in the mutant by 2.7 kcal/mol.,[24] since both interacting side chains are of medium polarity (c.f. Table 2).

SIMILARITY

Coulombic: Classification of benzamidine inhibitors of trypsin

It has been shown that differences in inhibitory potencies of *para* and *meta* substituted benzamidines to trypsin can be explained in terms of electrostatics.[30] The electrostatic interaction energies calculated from the protein electrostatic potential and the charge distribution of the inhibitor are in a fair linear correlation with the experimental binding free energies. Furthermore, it has been shown that the electrostatic patterns of the different derivatives let us distinguish between two classes of substituents.[31] As displayed in Color Plates 3 and 4, for two representative substituents the benzamidine moieties show perfect similarity, but the substituent regions are different: positive for the hydroxl and negative for the nitro group.

Since the potential in the specificity pocket of trypsin where benzamidines are accommodated is everywhere negative, it is expected that the hydroxy-substituted derivative will bind better (i.e., inhibit trypsin more effectively than the nitro-substituted one. According to the sign of the potential around substituents, they can be grouped in two classes—S_I and S_{II}, each possessing similar inhibitory power:

S_I: 4-NH_2, 4-Me, 4-OH, 4-OMe, 4-OEt, 3-NH_2, 3-OH, 3-OMe, 3-OEt

S_{II}: 4-NO_2, 4-COOMe, 4-COOEt, 4-COMe, 4-CONHMe, 3-NO_2, 3-COOMe, 3-CONHMe

with

$$pK_i(S_I) > pK_i(S_{II}) \quad (4)$$

Comparing equation (4) to experimental data, it has been found that there are only two outliers, the 3-COMe and 4-OEt derivatives.[31] The same distinction between the S_I and S_{II} groups can be made on the basis of Hammett substituent constants. However, the latter give no information on the steric aspects of electrostatic similarity, since a substituent is represented by just one number instead of a whole potential pattern.

Table 2. Average electrostatic fields on the van der Waals surface of amino acid side chains (in V/nm)

Side chain	F_{av}	Side chain	F_{av}
Charged			
Asp^-	26.7	Lys (neutral)	7.3
Glu^-	22.8		
Lys^+	18.1	*Nonpolar*	
Arg^+	15.8	Met	5.6
Polar		Tyr	5.0
Asn	12.7	Trp	4.4
Ser	11.8	Ala	1.3
Arg (neutral)	11.8	Val	1.3
Gln	10.9	Leu	1.3
Cys	10.2	Ile	1.2
Thr	9.7	Phe	0.6
His	9.0	Gly	0.0

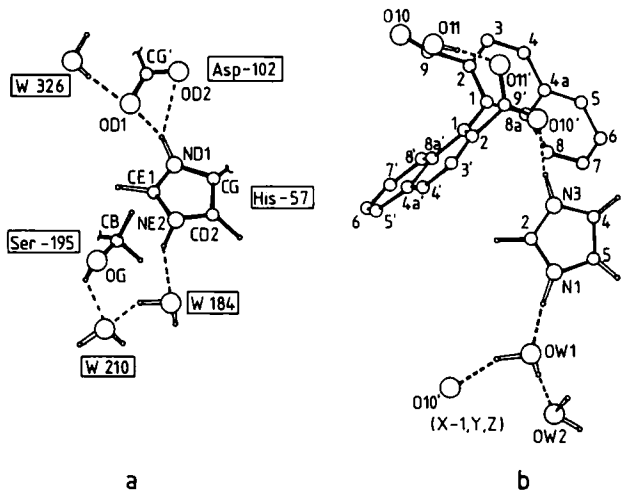

Figure 2. (a) Geometric arrangement of the carboxylate-imidazolium ion pair in Streptomyces Griseus protease A. (b) The 1,1'-binaphtyl-2,2'-dicarboxylic acid: imidazole: water adduct. Hydrogen bonds are indicated by dashed lines

Noncovalent structural model for the Asp-His diad in the active site of serine proteases

At last, we discuss a simple model where, besides geometric similarity to the Asp-102 . . .His-57 couple in *Streptomyces Griseus* protease A (SGPA), the electrostatic potential and field patterns produced by the environment are also remarkably similar.[32,33] As shown in Figure 2, the geometric arrangement of the carboxylate-imidazolium ion pair is almost equivalent both in the 1:1:1 adduct of 1,1'-binaphtyl-2,2'-dicarboxylic acid, imidazole and water and in the enzyme. The electrostatic patterns produced by the otherwise very different environment, a symmetrically packed, tightly bound crystal of low molecular weight compounds and a loose asymmetric protein molecule, are also nearly identical. We calculated electrostatic field components of the protein and crystal environment, directed from X to H in the XH bond in imidazole.[33] The obtained values (in V/nm, for the crystal in parentheses) are CD2-H: +4.2(+6.3), NE2-H: −9.0(−28.4), CE1-H: −2.7(+1.8), and ND1-H: −32.5(−29.2). Since the fields at the CD2 proton are similar in both systems, it is expected that the anisotropy effect and the resulting chemical shift in the ^1H nuclear magnetic resonance

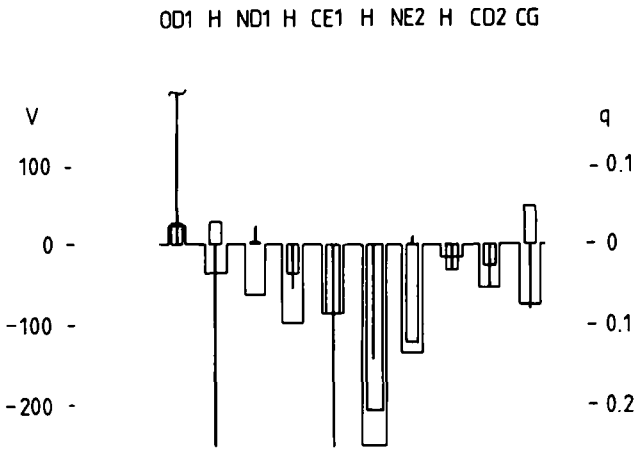

Figure 3. Electrostatic complementarity between atomic charges (q_i) of the carboxyl-imidazolium couple (solid bars, in electrons) and the electrostatic potential (in kJ/mol) of Streptomyces Griseus protease A protein environment (narrow empty bars) and the 1,1'-binaphtyl-2,2'-dicarboxylic acid crystal environment (broad empty bars)

spectrum, which is known to be linearly proportional to the external field along the C-H bond, is also the same. Thus, the crystal may serve as a simple model for the extremely complicated NMR spectra detected for serine proteases, especially SGPA, in solution.

ACKNOWLEDGEMENT

am indebted to Mr. I. Kolossváry of Budapest for making his POTROT program available.

REFERENCES

1 Lehn, J.M. Supramolecular chemistry: Receptors, catalysts and carriers. *Science* 1985, **227**, 849–856
2 Politzer, P., and Truhlar, D.G. (Eds.) *Chemical Applications of Atomic and Molecular Electrostatic Potentials*. Plenum, New York, 1981
3 Warshel, A., and Russell, S. T. Calculations of electrostatic interactions in biological systems and in solution *Quart. Rev. Biophys.* 1984, **17**, 283–422
4 Goodford, P. J. Drug design by the method of receptor fit. *J. Med. Chem.* 1984, **27**, 557–564
5 Cohen, N. C. β-Lactam antibiotics: Geometrical requirements for antibacterial activities. *J. Med. Chem.* 1983, **26**, 259–264
6 Warshel, A., and Sussman, F. Toward computer-aided site-directed mutagenesis of enzymes. *Proc. Natl. Acad. Sci. USA* 1986, **83**, 3806–3810
7 McCammon, J.A. Computer-aided molecular design. *Science* 1987, **238**, 486–491
8 Hobza, P., and Zahradnik, R. *Weak Intermolecular Interactions in Chemistry and Biology*. Elsevier, Amsterdam, 1980
9 Kollman, P. Drug-receptor binding forces, in Horn, A.S., and De Ranter, C.J. (Eds.) , *X-Ray Crystallography and Drug Action*. Clarendon, Oxford, 1984, 63
0 Kollman, P. A general analysis of nonconvalent intermolecular interactions. *J. Am. Chem. Soc.* 1977, **99**, 4875–4894
11 Nagy, P. Orientation and structure-building role of the water molecules bound at the contact surface of the dihydrofolate reductase-methotrexate complex. *J. Comp. Aided Mol. Des.* 1988, **2**, 65–76
12 Nakamura, H., Komatsu, K., Nakagawa, S., and Umeyama, H. Visualization of electrostatic recognition by enzymes for their ligands and cofactors. *J. Mol. Graphics* 1985, **3**, 2–11
13 Peinel, G., Frischleder, H., and Birnstock, F. The electrostatic molecular potential—a too for the prediction of electrostatic molecular interaction properties. *Theoret. Chim. Acta* 1980, **57**, 245–251
14 Náray-Szabó, G., and Nagy, P. Electrostatic recognition in proteins, in Rein, R. (Ed.), *Molecular Basis of Cancer, Part B: Macromolecular Recognition, Chemotherapy and Immunology*, Alan Liss, New York, 1985, 105–113
15 Nagy, P., and Náray-Szabó, G. Electrostatic lock-and-key model for the analysis of inhibitor recognition by dihydrofolate reductase. *Canad. J. Chem.* 1985, **63**, 1694–1698
16 Náray-Szabó, G. Electrostatic isopotential maps for large biomolecules. *Int. J. Quant. Chem.* 1979, **16**, 265–272
17 Náray-Szabó, G., Grofcsik, A., Kósa, K., Kubinyi, M., and Martin, A. Simple calculation of electrostatic isopotential maps from bond fragments. *J. Comput. Chem.* 1981, **2**, 58–62
18 Nagy, P., Ángyán, J.G., Náray-Szabó, G., and Peinel, G. Molecular electrostatic fields from bond fragments. *Int. J. Quant. Chem.* 1987, **31**, 927–939
19 Ángyán, J.G., and Náray-Szabó, G. Comparison of protein electrostatic potential along the catalytic triad of serine proteases. *J. Theor. Biol.* 1983, **103**, 349–356
20 Náray-Szabó, G., and Nagy, P. Electrostatic lock-and-key model for the study of biological isoterism: Role of structural water in the binding of basic pancreatic trypsin inhibitor to β-trypsin. *Enzyme* 1986, **36**, 44–53
21 Bajusz, S., Barabás, E., Tolnay, P., Széll, E., and Bagdy, D. Inhibition of thrombin and trypsin by tripeptide aldehydes. *Int. J. Peptide Protein Res.* 1978, **12**, 217–221
22 Rich, D.H. Pepstatin-derived inhibitors of aspartate proteinases. A close look at an apparent transition-state analogue inhibitor. *J. Med. Chem.* 1985, **28**, 263–273
23 Andrews, P.R., Craik, D.J., and Martin, J.L. Functional group contributions to drug-receptor interactions. *J. Med. Chem.* 1984, **27**, 1648–1657
24 Gráf, L., Jancsó, A., Szilágyi, L., Hegyi, G., Pintér, K., Náray-Szabó, G., Hepp, J., Medzihradszky, K., and Rutter, W. J. Electrostatic complementarity within the substrate-binding pocket of trypsin. *Proc. Natl. Acad. Sci. USA* 1988, **85**, 4961–4965
25 Estell, D.A., Graycar, T.P., Miller, J.V., Powers, D.B., Burnier, J.P., Ng, P.G., and Wells, J.A. Probing steric and hydrophobic effects on enzyme-sub-

strate interactions by protein engineering. *Science* 1986, **233**, 659–663

26 Wells, J.A., Powers, D.B., Bott, R.R., Graycar, T.P., and Estell, D.A. Designing substrate specificity by protein engineering of electrostatic interactions. *Proc. Natl. Acad. Sci. USA* 1987, **84**, 1219–1223

27 Náray-Szabó, G. Electrostatic catalysis in enzymes. *J. Mol. Catal.* 1988, **47**, 281–287

28 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S., Jr., and Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 1984, **106**, 765–784

29 Connolly, M.L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* 1983, **221**, 709–713

30 Náray-Szabó G. Quantum chemical calculation of the enzyme-ligand interaction energy for trypsin inhibition by benzamidines. *J. Am. Chem. Soc.* 1984, **106**, 4584–4589

31 Náray-Szabó, G. Similarity analysis of bioactive molecules: Electrostatic lock-and-key model and topological molecular transforms for trypsin inhibitors. *J. Mol. Struct. THEOCHEM* 1986, **134**, 401–409

32 Czugler, M., Ángyán, J.G., Náray-Szabó, G., and Weber, E. Noncovalent structural models for the Asp-His dyad in the active site of serine proteases and for solid-state switching of protonation states: Crystal structure of the associates of 1,1′-binaphtyl-2,2′-dicarboxylic acid with imidazole in dihydrated and in anhydrous forms. *J. Am. Chem. Soc.* 1986, **108**, 1275–1281

33 Nagy, P., Ángyán, J. G., and Náray-Szabó, G. Electrostatic complementarity in molecular aggregates. Part 6. The carboxylate-imidazolium diad in *Streptomyces Griseus* protease A and in the 1,1′-binaphtyl-2,2′-dicarboxylic acid dihydrate-imidazole adduct. *J. Mol. Struct. THEOCHEM* 1987, **149**, 169–176

34 Kolossváry, I. *Potrot: A microcomputer program for the visualization of molecular electrostatic potentials.* Department of Analytical Chemistry, Technical University Budapest, Budapest, 1987