Theoretical Aspects of Molecular Recognition*

Veronika Harmat[a] and Gábor Náray-Szabó[b,**]

[a]*Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University,*
Pázmány Péter sétány 1/A, 1117 Budapest, Hungary
[b]*Protein Modelling Group HAS-ELTE, Institute of Chemistry, Eötvös Loránd University,*
1117 Budapest, P.O.Box 32, Hungary

Abstract. Molecular recognition is a key process in non-covalent interactions, which determines, among others, host-guest complexation, drug action and protein-protein interaction. A simple and attractive formulation is the lock-and-key analogy defining the host as a lock accommodating the guest as a key. We stress three major aspects of molecular recognition, determining both complementarity between host and guest and similarity within a group of guest molecules. These aspects are: steric, *i.e.* maximization of close contacts, electrostatic, *i.e.* maximization of electrostatic attraction between host and guest, as well as hydrophobic, *i.e.* avoiding hydrophobic hydration, which can be reached by the maximization of apolar contacts between interacting molecules. Some examples are presented from our laboratory: the complexes of acylaminoacyl peptidase with small peptides, the effect of heparin binding on inhibitory potency of C1-inhibitor as well as small-molecule ligand binding to prolyl oligopeptidase and calmodulin.

Keywords: complementarity, steric effects, electrostatics, hydrophobicity

INTRODUCTION

Molecular recognition plays a crucial role in supramolecular chemistry, molecular engineering and structural biology. The first, simple but very effective model of molecular recognition has been proposed by Emil Fischer more than hundred years ago:[1] "... the intimate contact between the molecules ... is possible only with similar geometrical configurations. To use a picture, I would say that the enzyme and the substrate must fit together like a lock and key." In order to make this analogy more concrete, we should define the terms "lock", "key" and "fit". In a broad sense we mean under "lock" the host structure, *i.e.* any crevice on the surface of a macromolecule or a hollow site inside of a molecular aggregate, while the "key" means the guest, a small molecule or a fragment of a larger one, fully or partly embedded by the above empty space.

In the following we suppose that both lock and key are rigid, which makes discussion simpler. However, it should be mentioned that in many cases they are flexible and adopt their final shapes only upon binding, which is best illustrated by the "hand-and-a glove" analogy. If the lock (macromolecular crevice) undergoes conformational changes in order to provide the optimal shape for embedding the relatively rigid key (guest molecule), we speak about induced fit.[2] Alternatively, the guest molecules may adopt different conformations, some of which are appropriate for binding, this is called conformational selection.[3] In most cases, a mixture of both mechanisms can be observed. Flexibility and disorder-order transitions can be important in fine-tuning the strength of host guest interactions via entropic effects (induced local folding of intrinsically unstructured proteins upon binding to their target).[4]

It is not quite simple to define what we mean under the term "fit" or "complementarity". In short, these terms stand for the simplified description of the host-guest interaction in the biophase. Such an interaction is a combination of at least three components: steric, electrostatic and hydrophobic.[5,6] Steric fit means that interacting atoms may not interpenetrate beyond their van der Waals radii and, simultaneously, the host crevice should be filled as perfectly as possible reducing the empty space between host and guest atoms to a minimum. Electrostatic fit refers to a maximum of ionic, hydrogen-bonding and any other type (*e.g.* aromatic ring-cation)[7] of polar interaction, and is well accounted for by the molecular electrostatic potential due to the host and acting on a charge distribution representing the

guest. The term "hydrophobic fit" refers to the association trend between apolar groups in the biophase. Under "biophase" we mean an aqueous medium with dissolved ions and some small molecules surrounding the biopolymer as a solute. This trend may be explained in terms of density differences between water and the host, which makes that the net dispersion attraction between host and guest atoms is larger that between both of them and water. Accordingly, if they associate the net interaction energy becomes larger than in the dissociated state. Another argumentation is macromolecular crowding, an entropy effect reducing water-accessible surfaces of dissolved molecules in order to avoid unfavourable perturbation of water structure around the solute. Crowding shifts equilibria toward a molecular arrangement, where the excluded volume is minimal. Empty space between protein crevice and its ligand increases the excluded volume, which will be reduced if filled by a substituent. For more details see an early review.[8] Quite often, steric, electrostatic and hydrophobic effects on ligand binding are combined, however, it is useful to treat them separately, therefore we selected examples where one of them seems to be dominant.

On the basis of the concept of complementarity it is straightforward to define molecular similarity.[5] We call those molecules similar, that fit into the same biopolymer crevice with about the same pattern of steric, electrostatic and hydrophobic complementarity. A set of superimposed, similar molecules defines the pharmacophore, *i.e.* a molecular framework that carries the essential features responsible for the biological activity of a drug molecule[9] or, in more precise terms, an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target.

In the following sections we present some examples for various types of host-guest complementarity and guest similarity based on X-ray diffraction and molecular modelling studies in our laboratory.

STERIC COMPLEMENTARITY: COMPLEXES OF ACYLAMINOACYL PEPTIDASE WITH SMALL PEPTIDES

Steric fit can be nicely illustrated on the complex of *Aeropyrum pernix* K1 acylaminoacyl peptidase (ApAAP) with product-like inhibitors 2-aminobenzoyl-Gly-Phe and Gly-Phe (see Figure 1).[10] The crystal structures contain the dimer of ApAAP. The monomers A and B have similar conformations and bind the ligands at the S1 and S2 binding sites in similar arrangements. (We use the nomenclature of the substrate residues and substrate binding subsites of proteaseas as proposed by
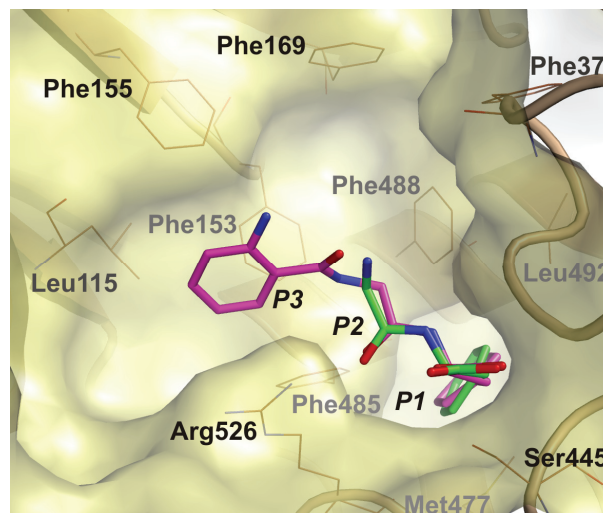

Figure 1. Binding of 2-aminobenzoyl-Gly-Phe and Gly-Phe to the active-site crevice of ApAAp. Complex structures refer to PDB codes 2HU5 and 2HU8, respectively. Note the empty pockets near Phe488 and Leu115 on the molecular surface representation of the enzyme.

Schechter and Berger.)[11] The guest crevice near the catalytic serine (Ser445) (Figure 1) accommodates the P1 phenylalanine and partly the P2 glycine of the ligand. Note that the crevice is precisely filled here, while near Leu115 and Phe448 empty cavities can be observed, which are formed by hydrophobic residues; this makes the steric fit imperfect. Filling these pockets with appropriately bulky substituents may increase the inhibitory potency of the ligand, *i.e.* the binding interaction with the host molecule. One of the earliest application of this "empty-space-filling" principle to increase of the binding power of a new substituted thyroxine analogue to the host, prealbumin, has been published by Blaney and co-workers more than two decades ago.[12] Introduction of two iodine substituents at the appropriate site, to fill the binding crevice perfectly, increase binding power by a factor of two.

ELECTROSTATIC COMPLEMENTARITY: HEPARIN POTENTIATION OF INHIBITORY POTENCY

C1 inhibitor, a member of the serpin family, is a major regulator of inflammatory processes in blood. Genetic deficiency of the C1 inhibitor results in hereditary angioedema, a dominantly inheritable, potentially lethal disease. Recently we reported the first crystal structure of the serpin domain of human C1 inhibitor, representing a latent form, which explains functional consequences of several naturally occurring mutations.[13] On the basis of the surface charge pattern, heparin affinity measurements, and computer docking of a heparin di-
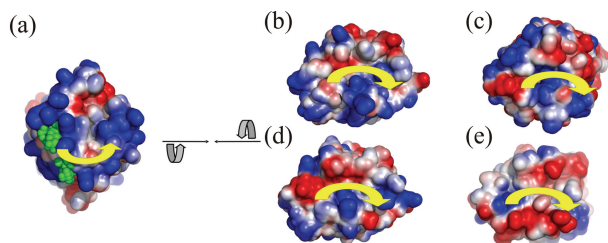
Figure 2. Surface electrostatic potential pattern on the contact surfaces of the C1 inhibitor (a) and target proteases. (b): factor XIa; (c): C1s; (d): plasma kallikrein; (e): factor XIIa (homology model). Blue, positive (deep blue: +50 kT/e, light blue: <+50 kT/e); red, negative (deep red: −50 kT/e, light red: > −50 kT/e). The docked heparin disaccharides are shown in green space filing representation. PDB codes of the structures are 2OAY, 1XX9, 1ELV, 2ANW, respectively. Yellow arrows indicate the presumed position of the reactive centre loop of C1 inhibitor in the Michaelis complex. Grey arrows indicate that one of the interacting molecules must be rotated by 180° in order to get in the right position for overlap with the other at the contact surface. Note that the upper negative region of (a) overlaps with the lower regions (positive for (b) and (c), negative for (e) of the target proteases.

saccharide to C1 inhibitor, a heparin binding site is proposed in the contact area of the C1 inhibitor/protease encounter complex (see Figure 2a). Charge pattern of contact enzyme surfaces and C1 inhibitor suggests neutralization by polyanions, *e.g.* heparin. A major positively charged patch is located in the presumed enzyme-binding region of C1 inhibitor. Binding of polyanions in this area results in the neutralization of positive charges or even providing excess negative ones. In this way, polyanions get "sandwiched" between the inhibitor and the enzyme.

    We calculated the electrostatic potential pattern on the contact surface of the interacting proteins and compared them in order to test the above mechanism (*cf.* Figure 2). The active site cleft and specificity loops of the serine protease domains are visualised. Those proteins (*cf.* the caption to Figure 2) were included in the study, for which independent experimental data were available in the literature. Polyanions increase inhibitory capacity against factor XIa 60–115-fold (Figure 2b), against C1s (Figure 2c) 15–60-fold, and against plasma kallikrein twofold (Figure 2d). Inhibitory capacity for factor XIIa (Figure 2e) is reduced by a factor of 2 to 4. These results correlate well with the calculated electrostatic potential pattern on the contact site. Factor XIa is characterised by the most extended positive electrostatic potential pattern, thus polyanions have the most significant effect here. C1s and plasma kallikrein have less positive potential in the patches.
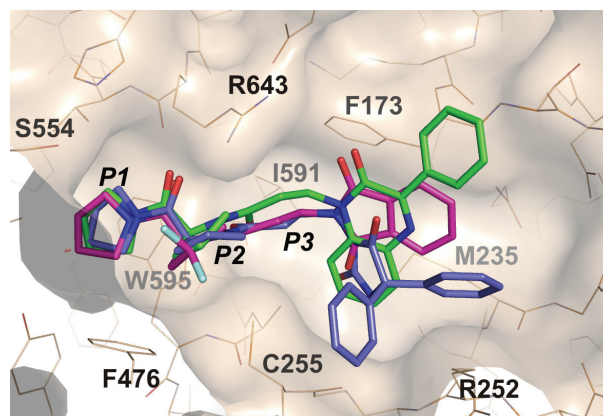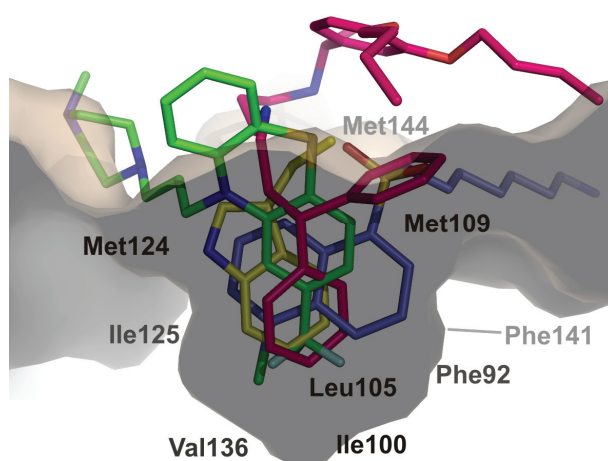

Figure 3. Binding of derivatives of Z-Pro-prolinal to the active site of prolyl oligopeptidase. Carbon atom colors: green: 1-{3-Oxo-3-[(2*S*)-2-(pyrrolidinocarbonyl)pyrrolidin-1-yl]propyl}-3-phenylquinoxalin-2(1*H*)-one; purple: 3-{4-Oxo-4-[(2*S*)-2-(pyrrolidinocarbonyl)pyrrolidin-1-yl]butyl}-5,5-diphenylimidazolidine-2,4-dione; magenta: 2-{3-[(2*S*)-4,4-Difluoro-2-(pyrrolidinocarbonyl)pyrrolidin-1-yl]-3-oxopropyl}-isoindole-1,3(2*H*)-dione. The molecular surface of the enzyme is presented.

STERIC AND HYDROPHOBIC SIMILARITY: PROLYL OLIGOPEPTIDASE AND CALMODULIN INHIBITORS

As we mentioned in the introduction, molecular similarity may be defined in terms of the binding mode of various molecules or their fragments to the same host. A conservative binding mode may be observed for three derivatives of Z-Pro-prolynal in the active-site crevice of prolyl oligopeptidase (*cf.* Figure 3).[14] Comparison of the structures of the complexes reveals that the binding molecules closely fit in the S1-S2-S3 region, filling the crevice formed near Cys255, Phe476, Ser554, Ile591, Arg643 and Phe173. These inhibitors can be superimposed and the common region defines the pharmacophore, which is quite precisely defined in this case. It is formed equivocally by the P1-P2-P3 molecular entities and it can be anticipated that further molecules, which fit to this pattern, will also bind relatively strongly to the enzyme.

    The opposite is observed in case of the hydrophobic fit of various, quite dissimilar molecules, to the crevice of calmodulin, formed near the side chains Phe92, Leu105, Met109, Met124, Ile125 and Met144 (see Figure 4).[15] Calmodulin regulates several intracellular processes by binding to various proteins and modifying their action. Its interdomain flexibility and plasticity of its main binding sites (hydrophobic pockets) enable binding of target peptides of various sequences, which is related to its functional promiscuity. Indeed the plasticity of the hydrophobic pocket ensures binding

of them possessing an inhibitory role on calmodulin mediated processes. These fragments cannot be superimposed; therefore no pharmacophore can be defined, which would guide the design of new analogues.

CONCLUSION

We defined the terms "fit" or "complementarity", often used in the concept of lock-and-key analogy of molecular host-guest interactions. In more precise terms "fit" means maximum attraction between host and guest in the biophase. Based on this definition we make distinction among three types of complementarity, steric, electrostatic and hydrophobic. Furthermore, we call similar those guest molecules, which fit into the same host. Application of these simple definitions is illustrated on the analysis of protein-ligand interactions based on structures determined by X-ray diffraction in our laboratory.

EXPERIMENTAL

Figures were produced using the program PyMOL.[16] Surface electrostatic potential was calculated using APBS[17] using default parameters (nonlinear Poisson-Boltzmann equation, ionic strength of 0.15 mol dm^{-3}, pH = 7.0).

Prolyl oligopeptidase from porcine muscle was purchased from László Polgár in the Institute of Enzymolo-

Figure 4. Binding of different structures to the hydrophobic pocket of the C-terminal lobe of calmodulin. Carbon atoms of the bound antagonist molecules: green: trifluoperazine; purple: *N*-(6-aminohexyl)-5-chloro-1-naphthalenesulfonamide; magenta: *N*-(3,3-diphenylpropyl)-*N*'-[1R-(3,4-bisbutoxyphenyl)-ethyl]propylenediamine. The molecular surface of the hydrophobic pocket in the calmodulin/trifluoperazine complex is shown. As a reference the anchoring Trp800 side chain of target protein calmodulin dependent kinase II (yellow) is shown. PDB codes of the structures are 1MUX, 1QIV, 1LIN, 1CDL, respectively.

and adapting to different molecular fragments, which can bind in different relative orientations as seen in the complexes of calmodulin with small molecules − most

Table 1. Crystallographic data collection and refinement statistics of prolyl oligopeptidase-inhibitor complexes. **1**: 2-{3-[(2*S*)-4,4-Difluoro-2-(pyrrolidinocarbonyl)pyrrolidin-1-yl]-3-oxopropyl}isoindole-1,3(2*H*)-dione, **2**: 1-{3-Oxo-3-[(2*S*)-2-(pyrrolidinocarbonyl)pyrrolidin-1-yl]propyl}-3-phenylquinoxalin-2(1*H*)-one), **3**: 3-{4-Oxo-4-[(2*S*)-2-(pyrrolidinocarbonyl)pyrrolidin-1-yl]butyl}-5,5-diphenylimidazolidine-2,4-dione

	1	**2**	**3**
Data collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions / Å	a = 72.44, b = 101.44, c = 112.25	a = 72.44, b = 101.57, c = 112.17	a = 72.55, b = 101.46, c = 112.33
Resolution range / Å	74.53-2.89	34.38-2.73	34.50-2.47
Completeness / %	96.5	93.5	94.1
No. of reflections,	17989	21070	28405
redundancy	2.7	4.2	3.4
Refinement			
R factor/R_{free} (5% of data)	0.1779/0.2570	0.1768/0.2533	0.1744/0.2438
No. of atoms (protein/ inhibitor/water)	5574/29/52	5656/33/99	5675/36/132
Average B factors / Å^2	31.154	35.505	26.463
RMS deviation from ideal geometry			
Bond lengths / Å	0.003	0.002	0.006
Bond angles / °	0.809	0.718	1.182
Dihedral angles / °	28.14	27.99	27.95

gy, Hungarian Academy of Sciences. Crystals were grown by co-crystallization using the conditions published by Fülöp *et al.*[18] optimized for the present complexes. Protein and inhibitor concentrations of 10 mg/mL and 1 mmol dm^{-3} respectively, were applied. Data were collected at room temperature, using a Rigaku R-AXIS IIC diffractometer. Rigid body fitting was carried out with the program AMoRe[19] of the CCP4 suite,[20] using the protein part of an isostrucutral prolyl oligopeptidase complex structure (PDB code 1QFS). Model building and refinement were carried out using programs O[21] and X-PLOR, version 3.851,[22] respectively. Table 1 shows data collection and refinement statistics.

REFERENCES

1. E. Fischer, *Ber. Dt. Chem. Ges.* **27** (1894) 2985−2993.
2. Y. Liu, L. Li, and H. Y. Zhang, *Induced Fit* in: J. L. Atwood and J. W. Steed (Eds.), *Encyclopedia of Supramolecular Chemistry*, Taylor and Francis, 2004, DOI: 10.1081/E-ESMC-120019161.
3. S. E. Dobbins, V. I. Lesk, and M. J. E. Sternberg, *Proc. Natl. Acad. Sci. U.S.A.* **105** (2008) 10390−10395.
4. P. Tompa, *Trends Biochem. Sci.* **27** (2002) 527−533.
5. G. Náray-Szabó, *J. Mol. Recogn.* **6** (1993) 205−210.
6. Y. Tsuchiya, K. Kinoshita, and H. Nakamura, *Prot. Eng. Des. Sel.* **19** (2006) 421−429.
7. M. Gaberšček and J. Mavri, *Chem. Phys. Lett.* **308** (1999) 421−427.
8. G. Náray-Szabó and G. G. Ferenczy, *Chem. Rev.* **95** (1995) 829−847.
9. P. Ehrlich, *Ber. Dtsch. Chem. Ges.* **42** (1909) 17−47.
10. A. L. Kiss, B. Hornung, K. Rádi, Zs. Gengeliczki, B. Sztáray, T. Juhász, Z. Szeltner, V. Harmat, and L. Polgár, *J. Mol. Biol.* **368** (2007) 509−520.
11. I. Schechter and A. Berger, *Biochem. Biophys. Res. Commun.* **27** (1967) 157−162.
12. J. M. Blaney, E. C. Jorgensen, M. L. Connolly, T. E. Ferrin, R. Langridge, S. J. Oatley, J. M. Burridge, and C. C. F. Blake, *J. Med. Chem.* **25** (1982) 785−790.
13. L. Beinrohr, V. Harmat, J. Dobó, Zs. Lőrincz, P. Gál, and P. Závodszky, *J. Biol. Chem.* **282** (2007) 21100−21109.
14. K. Kánai, P. Arányi, Z. Böcskei, G. G. Ferenczy, V. Harmat, G. Náray-Szabó, K. Simon, and I. Hermecz, unpublished data.
15. V. Harmat, Z. Böcskei, G. Náray-Szabó, I. Bata, A. S. Csutor, I. Hermecz, P. Arányi, B. Szabó, K. Liliom, B. G. Vértessy, and J. Ovádi, *J. Mol. Biol.* **297** (2000) 747−755.
16. W. L. DeLano, The PyMOL Molecular Graphics System (2002) DeLano Scientific, Palo Alto, CA, USA. http://www.pymol.org.
17. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, *Proc. Natl. Acad. Sci. USA* **98** (2001) 10037−10041.
18. V. Fülöp, Zs. Böcskei, and L. Polgár, *Cell* **94** (1998) 161−170.
19. J. Navaza, *Acta Cryst. A* **50** (1994) 157−163.
20. Collaborative Computing Project 4, *Acta Cryst. D* **50** (1994) 760−763.
21. T. A. Jones, J. Y. Zou, S. W. Cowan, and M. Kjeldgaard, *Acta Cryst. A* **47** (1994) 110−119.
22. A. T. Brunger, XPLOR: *A System for X-ray crystallography and NMR*, Version 3, Yale University, New Haven, CT, 1992.

SAŽETAK

Teorijski aspekti molekulskog prepoznavanja

Veronika Harmat[a] i Gábor Náray-Szabó[b]

[a]*Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University,*
Pázmány Péter sétány 1/A, 1117 Budapest, Hungary
[b]*Protein Modelling Group HAS-ELTE, Institute of Chemistry, Eötvös Loránd University,*
1117 Budapest, P.O. Box 32, Hungary

Molekulsko prepoznavanje je ključni proces u nekovalentnim interakcijama, koji određuje, među ostalim, domaćin-gost (engl. *host-guest*) kompleksiranje, djelovanje lijekova i protein-protein interakcije. Jednostavna i atraktivna formulacija je princip ključ-brava koja definira molekulu domaćina kao bravu u koju može ući odgovarajuća molekula gosta kao ključ. Autori naglašavaju tri glavna aspekta molekulskog prepoznavanja, određujući zajedno komplementarnost između domaćina i gosta, kao i sličnost unutar grupe gost-molekula. To su sljedeći aspekti: sterički, tj. maksimiranje bliskih kontakata; elektrostatski, tj. maksimiranje elektrostatskog privlačenja između domaćina i gosta; i hidrofobni, tj. izbjegavanje hidrofobne hidracije, koja se može postići maksimiranjem apolarnih kontakata između molekula u interakciji. Pokazano je nekoliko primjera iz laboratorija: kompleksi acilaminoacil peptidaze s malim peptidima, efekt vezanja heparina na inhibicijski potencijal C1-inhibitora, kao i ligandno vezanje malih molekula na prolil oligopeptidazu i kalmodulin.