Structural chemometrics applied to small bioactive molecules complexed with their protein receptors

Márcia Miguel Castro Ferreira and Rudolf Kiralj

Laboratório de Quimiometria Teórica e Aplicada (LQTA), Instituto de Química, Universidade Estadual de Campinas, Campinas SP, 13084-971, Brazil

E-mails: marcia@iqm.unicamp.br, rudolf@iqm.unicamp.br, URL: http://lqta.iqm.unicamp.br

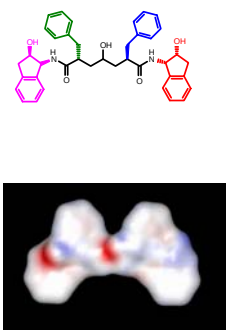
THE OBJECTIVES OF THIS WORK

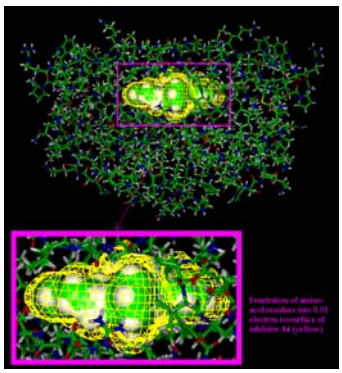
1) To present structural chemometrics applied to three cases of ligand-receptor interactions, based only on structural data from Protein Data Bank → no computer simulations applied;

2) To show that structural properties of small molecules (ligands) are important and sometimes even predominant in determining their biological response or physico-chemical behavior.

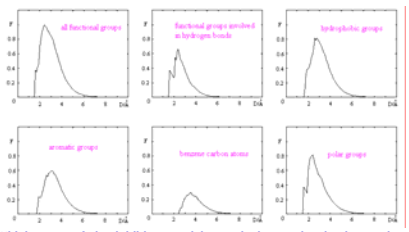
EXAMPLE 1: A PEPTIDIC HIV-1 PROTEASE INHIBITOR L-700,417 IN INTERACTION WITH THE PROTEASE

(R. Kiralj, M. M. C. Ferreira, *J. Mol. Graph. Mod.*, 21, 2003, 499)





The structure of a peptidic HIV-1 protease inhibitor L-700,417 or *N,N*-bis(2(*R*)-hydroxy-1(*S*)-indanyl)-2,6-(*R,R*)-diphenylmethyl-4-hydroxy-1,7-heptandiamide with pseudo C_2 symmetry (left) and structure of HIV-1 protease complexed with this ligand (right). (PDB: 4PHV. R. Bone *et al.*, *J. Am. Chem. Soc.*, 113, 1991, 9382.) The ligand is represented by electrostatic potential in free state (left bottom) and electron density isosurface (calculated at PM3 semi-empirical level) inside the active site of the protease (R. Kiralj, M. M. C. Ferreira, *J. Mol. Graph. Mod.*, 21, 2003, 499.) The inhibitor has four side chains that fit into four pockets of the protease binding cavity.

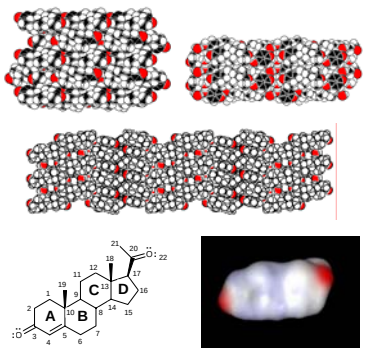


Which parts of the inhibitor participates in intermolecular interactions with the protein? Can these interactions be characterized and quantified using only structural data from the PDB? Function F was defined as a measure of the overlap of the inhibitor's and protease's electron densities: $F = (\rho\tau)^{1/2}$ $(V\text{-}V_0)$, and $\tau = N_{ve}$ $(V\text{-}V_0)$, where: ρ – overall ligand electron density at distance D from the ligand surface (a log-linear function of D); V_0 – ligand molecular volume; V – volume inside the cut-off distance D (a cube root-log function of ρ); N_{ve} – number of valence electrons of the protein and solvent (water) around the ligand, inside the cut-off D. When various groups of the ligand are used in this analysis (all functional groups, hydrogen bonding groups, hydrophobic groups, aromatic groups, benzene carbon atoms and polar groups), differences in the F-D curves are obvious. These differences can be quantified as: F_m – the global maximum of F; D_m – position of F_m in terms of D values; I – area under the curve; M_F – weighted mean of F; D_F – position of M_F; B – D value for the shortest interaction. The presented ligand was among the most active peptidic inhibitors in terms of both experimental and modeled activity (R. Kiralj, M. M. C. Ferreira, *J. Mol. Graph. Mod.*, 21, 2003, 499) .

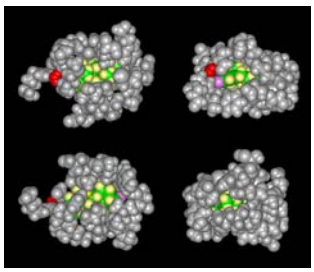
The 3D scores plot from Principal Component Analysis (PCA, left) and the dendogram with samples from Hierarchical Cluster Analysis (HCA, right) applied to variables F_m, D_m, I, M_F, D_F and B. The variables were autoscaled and complete linkage was used in HCA. The first three principal components (PCs) describe 99.2% of the original information. It is visible in both plots that aromatic and hydrophobic groups make a cluster, and polar and H bonding groups another one. Benzene carbon atoms represent an isolated case. Chemical interpretation of these results is based on the fact that significant portion of polar groups are also H bonding groups, and aromatic groups are important hydrophobic groups. The benzene carbon atoms contribute to some intermolecular interactions in which they behave as π donors (C,N,O-H…π interactions). Hydrogen atoms of the same benzene groups may participate in hydrophobic interactions C-H…H-C or C,N,O-H…π where π represents neighboring residues with π systems. It is obvious that the proposed six parameters distinguish well different functional group types as well as types of related intermolecular interactions.

EXAMPLE 2: PROGESTERONE IN INTERACTION WITH PROGESTERONE RECEPTOR

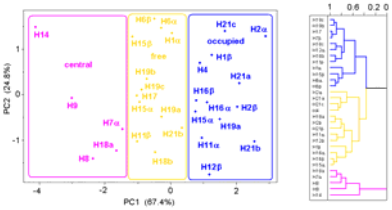
(R. Kiralj, M. M. C. Ferreira, *QSAR Comb. Sci.*, 22, 2003, 430)



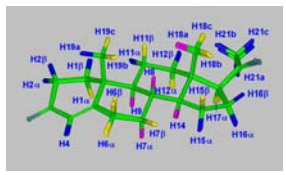
Progesterone structure and numbering (bottom left) and electrostatic potential surface (bottom right) show its amphiphilic character with prononunced hydrophobicity of rings B and C. The progesterone amphiphilicity can be observed in crystals of progesterone (top left, CSD: PROGST10), progesterone-hydroquinone complex hydrate (top right, CSD: PRORES) and progesterone-resorcinol complex (middle, CSD: KEFBEC), with alternating hydrophobic and hydrophilic layers.



Progesterone (C: green; H: yellow; O; pink) at the active site of the progesterone receptor (PR, gray; water: red), with amino-acids that are interacting with the ligand. Four free pockets are visible: at C8 and C19 (top left), at C4 and C6 (top right), at C14-C17 (bottom left), and C18 and C21 (bottom right). It seems that these are the four possibilities for corresponding hydrogen substitution by small groups to increase progesterone-PR binding. Modeling of new progestogens with higher binding potency to PR can be important for treatment of diseases, post-menopausal difficulties and hormonal disorders. (PR-progesterone structure from: PDB: 1A28. S. P. Williams, P. B. Sigler, *Nature*, 393, 1998, 392.)



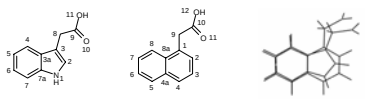
By using molecular graphics (previous picture), it is not possible to determine with certainty which hydrogen atom may be substituted, so the new substituent would fit into an empty pocket nearby. Structural chemometrics may aid in this determination. Four variables were generated using a monomer unit of PR complexed with a progesterone molecule (PDB: 1A28): the number of all/non-hydrogen atoms of PR and water, and the number of valence electrons of all/non-hydrogen atoms of PR and water, all within a cut-off distance 5.5 Å from progesterone molecular surface. Both Principal Component Analysis (left; two principal components: 92.2% of the total variance) and Hierarchical Cluster Analysis (right) reveal that all hydrogen atoms belong to one of the clusters: central (hidden or non-accessible, at junction of the rings), occupied (no free space around) and free (there is a free space around) hydrogens. Free hydrogens are the most probable candidates for substitution. Due to the flexibility of the protein, other substitutions of some occupied hydrogens are also possible.



Superimposed structures of two progesterone molecules from the PR-progesterone complex with *ab initio* structure of free progesterone. High degree of rigidity is visible, even in terms of central (pink), occupied (blue) and free (yellow) hydrogen atoms. Substitution of hydrogen atoms H6α, H6β, H11(α, β), H17α, H18(a-c) showed to vary progestational activity. For example, at position Cα, parabolic activity-substituent size relationships are observed: Cl and Me groups account for the highest activity, and F and Br for decrease in activity (**R. Kiralj, M. M. C. Ferreira, *QSAR Comb. Sci.*, 22, 2003, 430**). Presented structural chemometrics is a simple example showing that this new methodology may be useful in analyzing experimental geometries of small molecules, macromolecules and molecular complexes prior to computational approach.

EXAMPLE 3: 1-NAPHTHALENIC ACID IN INTERACTION WITH AUXIN BINDING PROTEIN 1

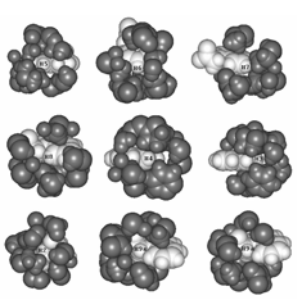
(M. M. C. Ferreira, R. Kiralj, *Croat. Chem. Acta*, submitted)



Scheme with structure and atom numbering of 1*H*-indole-3-acetic acid (IAA, left) and 1-naphthalenic acid (NAA, middle), and superimposed geometries of the two molecules as obtained from molecular mechanics MMFF94 calculations.

Substitution	Higher activity (5,6)	IAA like activity (5,4 – 5,5)	Lower activity (≤ 5,3)	Pocket size
2			Me	small
4	F, Me	Et, Cl	–	medium
5	–	F, Cl, Me, Et, Pr, Bu	Br	large
6	–	F, Me, Et	–	medium
7		F, Cl, N, Me, Et		medium or large

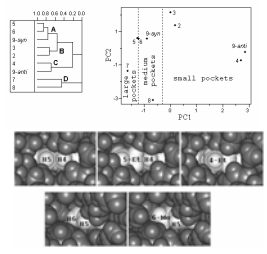
Biological activities of IAA and its derivatives (auxins) and probable size of pockets at the protein receptor cavity into which the substituents can fit. The activity was defined as –log*c*, where *c* is half-optimal concentration of auxin that provokes growth of *Avena Sativa L.* coleoptiles. The activity data were from literature (B. Nigović *et al.*, *Acta Cryst. B*, 56, 2000, 94; S. Antolić *et al.*, *Acta Cryst. B*, 52, 1996, 651).



NAA at the active cavity of auxin binding protein 1 (ABP, PDB: 1LRH. E. J. Woo *et al.*, *Embo J.*, 21, 2002, 2877). Coordination sphere (dark) around hydrogen atoms of NAA (light) is presented. H9a and H9s stand for H9-*anti* and H9-*syn*, respectively, depending if they are placed at the same or opposite side of the aromatic ring as the carboxyl group. Other hydrogen atoms are named according to their C/N atoms.

NAA[a]	IAA adj.[b]	Count1	Count2	Min. d	Count3	Non-H	Count4	Pocket size[d]
9 *anti*	8 *anti*	2	1	2.90	27	0.54	21	very small
9 *syn*	8 *syn*	0	0	3.43	23	0.46	17	medium
2	[2]*	0	0	3.35	26	0.48	20	small
3	[2]*	0	0	3.70	27	0.44	22	small
4	1	3	1	2.91	27	0.46	11	small
8	4	1	1	2.68	17	0.38	15	medium
7	5	2	0	3.31	19	0.35	16	large
6	6	0	0	3.46	24	0.39	18	medium
5	7	0	0	3.54	22	0.42	18	large

[a]Steric properties of the ABP1 active site cavity around NAA hydrogen atoms inside 5.5±0.2 Å cutt-of: Count1, Count2 – No. van der Waals (vdW) contacts with all/non-H atoms; Min. d – min. distance observed for vdW contacts; Count3 – No. non-H atoms; Non-H – number fraction of non-H atoms count; Count4 – No. C atoms. [b]NAA H atoms; *syn* and *anti* atoms are named whether they are oriented up or down with respect to the aromatic plane in orientation as in Scheme. [c]IAA H atoms according to the numbering system from Scheme, equivalent to the analogous NAA H atoms; *syn* and *anti* atoms named with respect to the aromatic ring plane. [d]As observed from molecular graphics. *This IAA H is in the mid-way between two NAA H atoms (see superimposition).



Chemometric analysis of data from previous table (top) and modeling of NAA derivatives at the active site cavity of ABP1 (bottom). Hierarchical cluster analysis (top left) shows four distinct clusters of NAA H atoms placed in very small (B), small (C), medium (A) and large (D) pockets. Principal component analysis (87.6% of variance presented in top right) discriminates H atoms in small, medium and large pockets along PC1. Clusters A-D can be noted if both principal components are considered. These classifications are related well with those based on molecular graphics and auxin activity (see previous tables). Larger pockets allow placement of Me and Et at C4 and C5 of NAA.