Comparison of Biomolecules on the Basis of Molecular Interaction Potentials

Jordi Rodrigo, Montserrat Barbany, Hugo Gutiérrez-de-Terán, Nuria B. Centeno, Miquel de-Càceres, Cristina Dezi, Fabien Fontaine, Juan J. Lozano[#], Manuel Pastor, Jordi Villà and Ferran Sanz*

Research Group on Biomedical Informatics (GRIB), Institut Municipal d'Investigació Mèdica, Universitat Pompeu Fabra, C/ Dr. Aiguader 80, E-08003 Barcelona

Potenciais Moleculares de Interação (PIMs) são frequentemente utilizados na comparação de séries de compostos que apresentam funções biológicas relacionadas. Esses potenciais representam as energias de interação entre os compostos de interesse e sondas químicas apropriadas. As energias de interação são calculadas computacionalmente em uma grade de pontos onde os compostos estão justapostos. É necessário que se efectuem análises comparativas objectivas e detalhadas das distribuições dos PIMs dentro do contexto de estudos de estrutura-atividade. Por outro lado, estudos baseados em PIMs não precisam estar restritos a séries de ligantes, pois este tipo de estudos também abre novas perspectivas na análise e comparação de macromoléculas biológicas. Essas análises podem ser melhoradas pela aplicação de novos métodos e procedimentos computacionais. Um novo programa de computador chamado MIPSim foi desenvolvido recentemente com o objetivo de analisar e comparar distribuições de MIPs de séries de biomoléculas. Este programa está integrado em outros como GAMESS ou GRID, que podem ser usados para o cálculo de potenciais que serão analisados e comparados. O programa MIPSim incorpora várias definições de coeficientes de similaridade e combina várias medidas de similaridade em uma só. Além disso, através deste programa pode-se explorar, automaticamente, os alinhamentos de similaridade máxima entre pares de moléculas.

Molecular Interaction Potentials (MIP) are frequently used for the comparison of series of compounds displaying related biological behaviors. These potentials are interaction energies between the considered compounds and relevant probes. The interaction energies are computed in the nodes of grids defined around the compounds. There is a need of detailed and objective comparative analyses of MIP distributions in the framework of structure-activity studies. On the other hand, MIP-based studies do not have to be restricted to series of small ligands, since such studies present also interesting possibilities for the analysis and comparison of biological macromolecules. Such analyses can benefit from the application of new methods and computational approaches. The new software MIPSim (Molecular Interaction Potentials Similarity analysis) has recently been introduced with the purpose of analyzing and comparing MIP distributions of series of biomolecules. This program is transparently integrated with other programs, like GAMESS or GRID, which can be used for the computation of the potentials to be analyzed or compared. MIPSim incorporates several definitions of similarity coefficients, and is capable of combining several similarity measures into a single one. On the other hand, MIPSim can perform automatic explorations of the maximum similarity alignments between pairs of molecules.

Keywords: Molecular Interaction Potentials (MIP), Molecular Interaction Fields (MIF), Molecular Electrostatic Potential (MEP), molecular similarity, molecular alignment, MIPSim

Introduction

The biological behavior of chemical entities is determined by chemical features that can be difficult to extract by visual inspection of series of active and inactive compounds. For instance, if we look at the compounds shown in Figure 1, it is difficult to reveal the common features that make all of them to be oxidized in the catalytic site of the cytochrome P-450 1A2 (CYP1A2).

When we try to explain the similarities/differences in the biological behavior of series of compounds, we have to look for how the compounds are "seen" by their biological counterpart (receptor, enzyme, DNA etc.). The

* e-mail: grib@imim.es
[#] Present address: Unidad de Bioinformática. Centro de Biología Molecular "Severo Ochoa". CSIC-UAM. E-28049 Canto Blanco (Madrid). Spain.
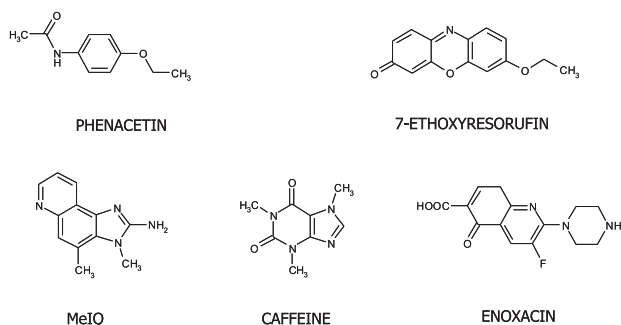
Figure 1. Some substrates of cytochrome P450 1A2.

system that molecules use to "see" each other is their mutual interaction. Two molecules will have the same biological activity if they are "seen" as similar by their common biological counterpart or, expressed in another way, if they have similar interactions with their biological counterpart. Consequently, the study of the molecular interaction capabilities is crucial when looking for molecular features associated with biological activities. The ideal way to study interaction capabilities is the analysis of the corresponding biomolecular complexes. However, many of these complexes are not experimentally available or are difficult to be computationally simulated. Consequently, "indirect" approaches based in the analysis of the structures of series of ligands, without considering the structure of their common biological counterpart, are widely used.

Molecular Interaction Potentials (MIP), also called Molecular Interaction Fields (MIF), are molecular properties frequently used for the description and comparison of series of compounds in the framework of SAR (Structure-Activity Relationships) and 3D-QSAR (Quantitative Structure-Activity Relationships on the basis of 3D molecular features) studies. MIP are computationally generated interaction energies between the considered compounds and probes that represent molecular fragments playing particular roles in the inter-molecular recognition. A popular program for MIP computation is GRID,[1] which uses probes such as methyl group, carbonyl oxygen, hydroxyl, amino nitrogen, etc. The simplest probe for MIP computations is the proton (constituted by an atomic unit of mass plus an atomic unit of positive charge). The MIP computed with the proton probe is called Molecular Electrostatic Potential (MEP). The physicochemical formalism used by the GRID program in order to compute interaction energies belongs to the molecular mechanics category and consequently it is relatively inexpensive from the computational point of view. If we wish a more rigorous computation of the MIP, we can use quantum mechanical methods that imply a higher computational cost. The

compound-probe interaction energies are usually computed in the nodes of grids defined around the compounds and they are frequently represented by means of isopotential surfaces. As a first application example,[2] Figure 2 shows the common features of the MEP distributions of the five CYP1A2 substrates shown in Figure 1. All of them have a highly attractive (negative) MEP zone at a distance of 2.2-3.1 Å from the group to be oxidized, and another highly negative MEP zone at 6.4-7.5 Å from the first one, being both zones located in the plane defined by the heterocyclic system, one at each side of it. This kind of MIP-based pharmacophores allow for the building up of molecular superpositions ("alignments") different from the classical structural ones. Figure 3 shows the MEP-based alignment of the aforementioned CYP1A2 substrates. In this example the MEP distributions were obtained by means of quantum mechanical computations using the 3-21G basis set. It has to be pointed that some time after the proposal of this
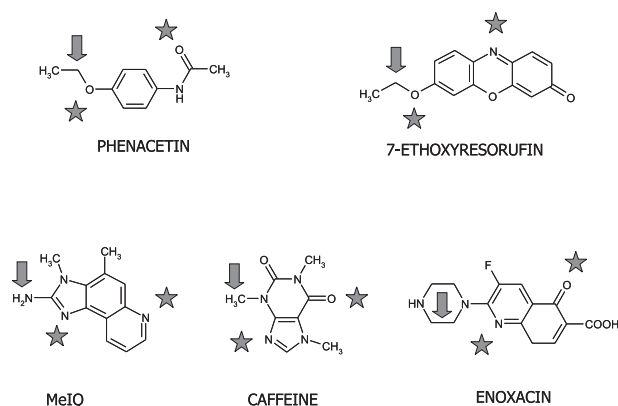

Figure 2. Common MEP features of the compounds shown in Figure 1. Arrows indicate the oxidation position. Stars indicate the MEP minima that define a common interaction template.
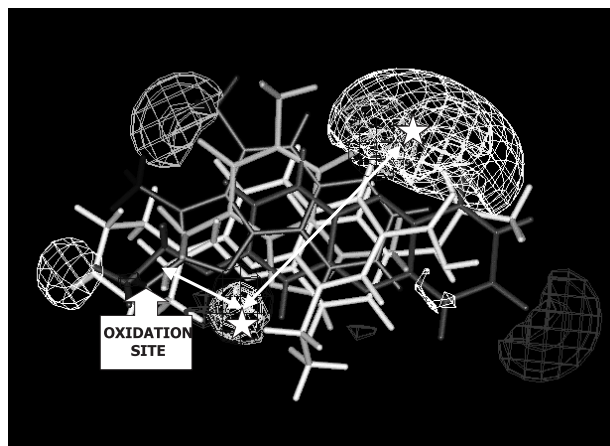

Figure 3. Superposition of the compounds shown in Figure 1 on the basis of their common MEP features described in Figure 2.

MEP-based pharmacophore, a consistent alignment was obtained by means of docking simulations in the catalytic site of a 3D model of the CYP1A2 enzyme, using a series of 12 amines that included MeIQ.[3]

Regression models that link biological activities with MIP-based descriptions of series of compounds have become popular 3D-QSAR analyses, which can be carried out using software such as CoMFA,[4] GRID/GOLPE,[5] VolSurf[6] or ALMOND.[7] However, the present paper is focused on the direct analysis of the similarity between MIP distributions in the framework of discussions about the biological behavior of biomolecules.

Methods

A methodological challenge when comparing or looking for common features of MIP distributions is the ability of performing such tasks in an objective and automatic fashion. The authors of the present article have considered several approaches for such a purpose. One of them is based in the automatic search of the most favorable points for interaction (MIP minima), as well as their geometrical relationships.[8] This approach, later incorporated in the MEPSim software as its MEPMIN module,[9] was initially restricted to the analysis of MEP distributions. The more recent MIPSim software[10] extends the same kind of analysis to any MIP distribution. The aforementioned software VolSurf[6] and ALMOND,[7] which perform automatic generation and analysis of alignment-independent MIP-based molecular descriptors, generate some of the descriptors after automatically locating the most favorable zones for probe interaction in a way analogous to that of MEPMIN.

Another useful task that calls for being automated is the exploration for maximum MIP similarity alignments. Such alignments become useful when looking for the relative docking positions of series of compounds in their common receptor (see, for instance, reference 3). This analysis was initially implemented in the MEPCOMP module[11] of MEPSIM,[9] and has been improved and extended in the MIPSim package.[10]

MIPSim[10] (Molecular Interaction Potentials Similarity analysis) has the purpose of analyzing and comparing MIP distributions of series of biomolecules. This program is transparently integrated with other software, like GRID[1] or the quantum mechanical package GAMESS,[12] which can be used for the computation of the molecular interaction potentials to be analyzed or compared. In addition to these integration capabilities, MIPSim incorporate several innovative algorithms that deserve to be remarked: i) the computation of the molecular similarity incorporates an algorithm[13,14] that does not require the coincidence of the grid size and the node positions of the molecules to be compared. A similarity coefficient having such characteristics is defined as:

$$A_{ab} = \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} p_i p_j e^{-\alpha r_{ij}^2}$$

$$Sim_{ab} = \frac{A_{ab}}{\sqrt{A_{aa}^{self} \cdot A_{bb}^{self}}}$$

being Sim_{ab} the molecular similarity coefficient, p_i and p_j the MIP values in nodes i and j of molecules a and b respectively, r_{ij} the distance between such nodes, and n_a and n_b the total number of grid nodes around each molecule. Figure 4 makes easier the understanding of this definition; ii) several similarity coefficients between a pair of molecules (for instance, those obtained using MIPs computed with different probes) can be combined in a single one by means of the expression:

$$Sim_{comb} = \frac{\sum_{i=1}^{k} w_i \, Sim_i}{\sum_{i=1}^{k} w_i}$$

being $\{Sim_i\}$ the k coefficients to be combined, and $\{w_i\}$ arbitrary weight coefficients that can be adjusted to obtain the best agreement with experimental evidences; iii) the exploration of maximum similarity alignments is carried out in an exhaustive manner by means of a random generation of initial relative positions of the compared molecules, followed by gradient-based maximizations. The
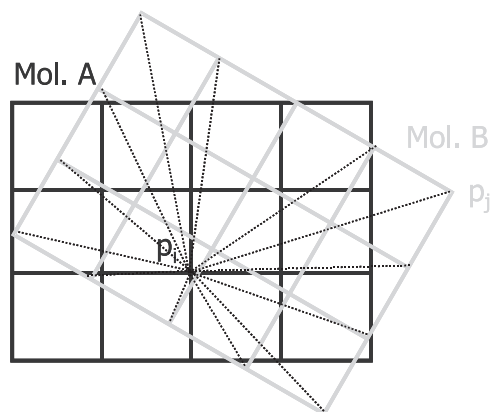

Figure 4. Scheme to illustrate the MIP similarity computation as implemented in MIPSim. The two rectangular grids represent MIP points computed around two compounds. All the possible pairs of points between both grids are considered for the similarity computation but having different weights, which are inversely proportional to the distance between the two points.

convergence of the maximization trajectories with other previously followed is permanently checked in order to abort the redundant ones. When such convergence happens after a predetermined number of iterations of random initial position plus gradient-based maximization, the molecular alignment space is considered sufficiently explored.

Results and Discussion

An example of MIPSim application is the comparison of xanthine and adenine, two structurally related and biologically important heterocyclic systems. Using the carbonylic oxygen and the amide nitrogen as GRID probes for the computation of MIP distributions, and combining the corresponding similarities between the two molecules with the above described Sim_{comb} coefficient using the same weight coefficients for both probes, MIPSim locates six high similarity alignments ($Sim_{comb} > 0.9$) that have a clear chemical and biological coherence (coincidence of the hydrogen bond acceptor and donor vectors, as well as the heterocyclic systems). Two of these alignments are shown in Figure 5.
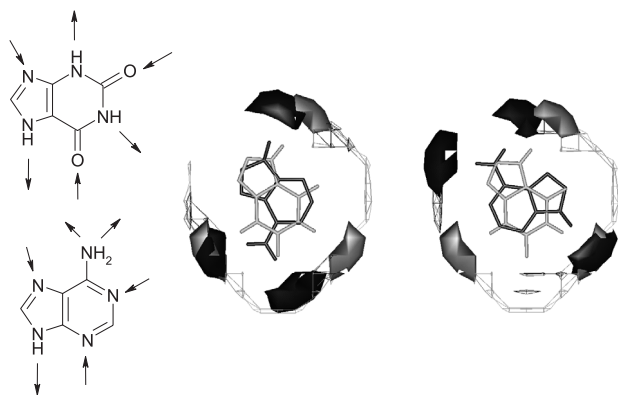

Figure 5. Two of the xanthine and adenine alignments automatically provided by MIPSim. Arrows in the formulae indicate the HBA and HBD vectors. The alignment in the middle has $Sim_{comb} = 0.98$ and the one on the left 0.94.

The usefulness of the MIP analyses and comparisons is not restricted to the case of small biological ligands like those of the aforementioned examples. Interesting examples of their application to biological macromolecules has been recently published, including methodologies for the comparison of related proteins.[15] A simple example of this kind of applications could be the comparison of highly related 5-HT$_{2A}$, 5-HT$_{2B}$ and 5-HT$_{2C}$ receptors. In a recent work of the same research group,[16] homology models of such receptors, docking of some ligands, and GRID/GOLPE models for a series of structurally related antagonists were obtained and

discussed, resulting especially noteworthy the GRID/GOLPE models obtained for the 5-HT$_{2A}$ binding of the antagonists. On the other hand, if we observe the docking models of the natural agonist serotonin into the binding site of the three receptors (shown in Figure 6), it is difficult to reveal the differences that allow for the existence of selective agonists for such receptors. Methods based on statistical regression[4-7] are not relevant for this kind of analysis. However, if we remove the serotonin from the binding sites and we perform a MIP scanning using the N3+, OH and hydrophobic DRY probes, we observe the existence of a couple of wide areas of favorable interactions
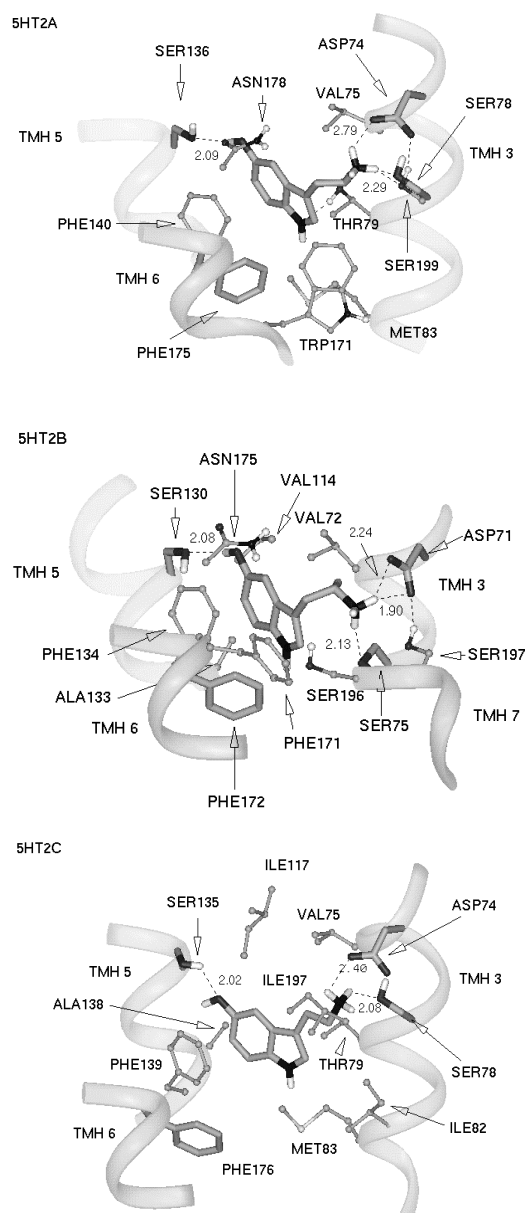

Figure 6. Docking models of serotonin into the binding site of the 5-HT$_{2A}$, 5-HT$_{2B}$ and 5-HT$_{2C}$ receptors.

with the OH probe in the case of the 5-HT$_{2B}$ receptor, and a couple of areas of favorable hydrophobic interactions in the case of the 5-HT$_{2C}$ receptor, being these features exclusive of each of these receptors. Consequently, there are clear differences in the interaction characteristics of the three binding sites despite their common natural ligand. These differences can be exploited in the design of selective ligands for such receptors.

Conclusions

In conclusion, we postulate the execution of detailed and objective comparative analyses of MIP distributions in the framework of structure-activity studies. These studies can benefit from the application of new methods and software that are currently available. On the other hand, the MIP-based studies do not have to be restricted to series of small ligands as done in the usual 3D-QSAR studies, since they present also interesting possibilities for the analysis and comparison of related biological macromolecules.

Note: The MIPSim program is available at http://www1.imim.es/modeling/mipsim/index.html

Acknowledgements

References

1. Goodford, P; *J. Med. Chem.* **1985**, *28*, 849.

2. López-de-Briñas, E.; Lozano, J. J.; Centeno, N. B.; Segura, J.; González, M.; de la Torre, R.; Sanz, F. In *Molecular Modelling and Prediction of Bioactivity*; Gundertofte, K.; Jørgensen, F.S., eds.; Kluwer Academic / Plenum Publishers: New York, 2000, p. 141.

3. Lozano, J. J.; Pastor, M.; Cruciani, G.; Gaedt, K.; Centeno, N. B.; Gago, F.; Sanz, F.; *J. Comput. Aided Mol. Des.* **2000**, *14*, 341.

4. Cramer, R. D.; Patterson, D.; Bunce, J. D.; *J. Am. Chem. Soc.* **1988**, *110*, 5959.

5. Baroni, M.; Constantino, G.; Cruciani, G.; Valigi, R.; Clementi, S.; *Quant. Struct-Act. Rel.* **1993**, *12*, 9.

6. Cruciani, G.; Pastor, M.; Guba W.; *Eur. J. Pharm. Sci.* **2000**, *11*, S29.

7. Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S.; *J. Med. Chem.* **2000**, *43*, 3233.

8. Sanz, F.; Manaut, F.; José, J.; Segura, J.; Carbó, M.; De la Torre, R.; *J. Mol. Struc. - Theochem* **1988**, *170*, 171.

9. Sanz, F.; Manaut, F.; Rodríguez, J.; Lozoya, E.; López-de-Briñas, E.; *J. Comput. Aided Mol. Des.* **1993**, *7*, 337.

10. De Càceres, M.; Villà, J.; Lozano, J. J.; Sanz, F.; *Bioinformatics* **2000**, *16*, 568.

11. Manaut, F.; Sanz, F.; José, J.; Milesi, M.; *J. Comput. Aided Mol. Des.* **1991**, *5*, 371.

12. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A.; *J. Comput. Chem.* **1993**, *14*, 1347.

13. Kaersley, S. K.; Smith, G. M.; *Tetrahedron Comput. Methodol.* **1990**, *3*, 615.

14. Klebe, G.; Mietzner, T.; Weber, F.; *J. Comput. Aided Mol. Des.* **1994**, *8*, 751.

15. Kastenholz, M. A.; Pastor, M.; Cruciani, G.; Haaksma, E. E.; Fox, T.; *J. Med. Chem.* **2000**, *43*, 3033.

16. Brea, J.; Rodrigo, J.; Carrieri, A.; Sanz, F.; Cadavid, M. I.; Enguix, M. J.; Villazón, M.; Mengod, G.; Caro, Y.; Masaguer, C. F.; Raviña, E.; Centeno, N. B.; Carotti, A.; Loza, M. I.; *J. Med. Chem.* **2002**, *45*, 54.