Chapter 9

3D QSAR MODELS

Ovidiu Ivanciuc
Department of Organic Chemistry, Faculty of Chemical Technology
University "Politehnica" of Bucharest, Oficiul 12 CP 243,
78100 Bucharest, Romania
E-mail: o_ivanciuc@chim.upb.ro

1. INTRODUCTION

Drug discovery was mainly the result of chance discovery and massive screening of large corporate libraries of synthesized or naturally-occurring compounds. Computer-aided drug design is an approach to rational drug design made possible by the recent advances in computational chemistry in various fields of chemistry, such as molecular graphics, molecular mechanics, quantum chemistry, molecular dynamics, library searching, prediction of physical, chemical, and biological properties.

An important step in drug design is to find a lead, a compound that binds to the target receptor. Leads can be generated using techniques of *de novo* drug design or can be discovered by *in vitro* screening of large corporate libraries. We have to mention that the lead identification is only the beginning of a long and expensive process that eventually yields a commercial drug. A lead may have a low affinity for the target receptor, may be too unstable in solution, too toxic, too rapidly eliminated, too quickly metabolized, too difficult or too expensive to synthesize in large quantities. Because the screening procedures generally give leads that are not suitable as commercial drugs, these compounds have to be optimized using various techniques of computer-aided drug design. The availability of three-dimensional (3D) structural information of biological receptors and their complexes with various ligands can be extremely useful in suggesting ways to improve the affinity of the lead to the target. Because in many cases such detailed structural information is still unavailable, the drug design process must rely upon a more indirect approach, the quantitative structure-activity relationships (QSAR) approach.

In the absence of detailed structural data upon biological receptors, a QSAR model establishes a statistical relationship between the biological activity exerted by a series of compounds and a set of parameters determined from the structures of the compounds. The central assumption of a QSAR model is that the numerical value of a specified biological activity measured for a set of molecules depends on the structure of these molecules. In order to correlate with a QSAR model the biological activity and the molecular structure, the structure must be adequately described in a numerical form with a large variety of structural descriptors: empirical (Hammett and Taft substituent constants), physical properties (octanol-water partition coefficient, dipole moment, aqueous solubility), constitutional (counts of various molecular subgraphs), topological indices, geometrical descriptors (molecular surface and volume), quantum (atomic charges, HOMO and LUMO energies), molecular fields (steric, electrostatic, and hydrophobic).

The first studies that use QSAR notions to explain the biological activity of sets of compounds were published by Kopp,[1] Crum-Brown and Frazer,[2] Meyer,[3] and Overton.[4] However, the background for a quantitative understanding of relationships between chemical structure and reactivity was performed by Hammett who introduced the linear free energy relationship (LFER) model in physical organic chemistry. In a study published in 1962,[5] Hansch extended these ideas and developed the most widely and effectively used QSAR model for congeneric series of compounds, expressed by a multilinear regression equation:[5-7]

$$\log(1/C) = a_0 + a_1 \pi + a_2 \pi^2 + a_3 \sigma + a_4 E_S \quad (1)$$

Here, C is the concentration (or dose) of congeneric members that gives a standard response such as EC_{50}, LD_{50}, etc., on a molar basis; π is the hydrophobic substituent parameter defined from 1-octanol/water partition coefficients P as:[8]

$$\pi = \log P_{RX} - \log P_{RH} \quad (2)$$

where subscripts RX and RH denote substituted and unsubstituted compounds, respectively; σ is the Hammett constant for the electronic property of substituents derived from dissociation constants of benzoic acids; E_S is the Taft constant that measures the steric effect of the substituent. Depending on the situation, the hydrophobicity parameter of the whole molecule, log P, can sometimes be used in place of π; also, the term π^2 can be ignored if it lacks statistical significance. Depending on the chemical structure of the molecules, the electronic parameters for aliphatic substituents such as σ_I and σ_R replace the Hammett constant, and the steric effect can be measured by modified E_S substituent constants, the volume of the substituent, or the STERIMOL parameters.

The Hansch model from Eq. (1) was developed from the assumption that the biological activity of a certain molecule depends on the transport from the site of application to its site of action and on the ligand-receptor interaction. The transport depends in a nonlinear (parabolic) manner on the lipophylicity of the molecule, while the binding affinity to the receptor depends on the lipophylicity, electronic properties, steric effect, or other properties of the ligand. Because the effect of structural modification is separated into components, and significant physicochemical factors are indicated

quantitatively, the Hansch model may reveal molecular mechanisms involved in the transport and ligand-receptor interaction of drugs.

The Hansch model was applied with success in hundreds of QSAR studies; however, the empirical parameters employed in this equation are not available for all substituents. Moreover, the use of substituent parameters limits the applicability of this model to series of congeneric compounds. To overcome this problem, computational chemistry techniques were employed to obtain various structural descriptors from the three-dimensional molecular structure. Such geometric, electrostatic, or quantum descriptors, used in a Hansch-like multilinear regression equation, represent a first type of 3D QSAR models.[9,10] In recent years the availability of detailed, three-dimensional structural information of biological receptors and their complexes with various ligands has revolutionized the drug-design process. However, in most instances such information is still unavailable. Therefore, when the structure of the biological target is not known the drug design process must rely upon a more indirect 3D QSAR approach that uses molecular alignment of atoms, pharmacophores, volume, or fields to generate a virtual receptor. Whenever the receptor structure is known, it is possible to investigate the ligand-receptor interactions and to derive 3D QSAR models from the corresponding parameters. The precise definition of 3D QSAR is still lacking, but we can identify two components that are essential for this type of models. The first component in the definition of a 3D QSAR model is the computation of the structural descriptors from the three-dimensional molecular structure; various geometrical, quantum, or molecular field descriptors were proposed in recent years to substitute the Hansch substituent constants. The second component in a 3D QSAR model is an explicit mathematical structure-activity relationship established between a dependent variable (biological activity) and a set of independent variables (3D structural descriptors); the mathematical 3D QSAR equations can be computed with the help of a large number of statistical models, such as multilinear regression, partial least squares (PLS), or neural networks. Some 3D QSAR models contain also a third component, a graphical representation of the three-dimensional information relative to the ligand-receptor interactions encoded into the structure-activity equation. Some representative 3D QSAR descriptors and ligand-receptor models will be presented in this review.

2. MASS DISTRIBUTION DESCRIPTORS

The van der Waals atomic radius is a successful concept for the computation of molecular size and shape descriptors, even if in a quantum chemical description the electron cloud has no well-defined boundary surface. In this theory, each atom of the molecule is represented as a sphere centered at the equilibrium position of the atomic nucleus having a radius equal to the van der Waals radius of the corresponding atom. The exterior surface of all atomic spheres defines the van der Waals surface, which delimits the van der Waals volume of the molecule.

In the QSAR program CODESSA[11] Katritzky defined two geometric descriptors that compute the mass distribution in a molecule. Gravitation indexes for all pairs of atoms G_1 and for all bonded pairs of atoms G_2 are defined as follows:

$$G_1 = \sum_{\text{all pairs}} \frac{m_i m_j}{r_{ij}^2} \tag{3}$$

$$G_2 = \sum_{\text{all bonds}} \frac{m_i m_j}{r_{ij}^2} \tag{4}$$

where m_i and m_j are the atomic weights of atoms i and j, and r_{ij} is the interatomic distance.

3. GEOMETRIC DESCRIPTORS FROM INFORMATION-THEORY OPERATORS

Numerous attempts have been made in theoretical chemistry to develop molecular graph descriptors that express in a numerical form the chemical structure. Such structural descriptors are widely used in modeling physical, chemical, or biological properties, in similarity and diversity assessment, drug design, database mining, and in the screening of virtual combinatorial libraries. Recently, the mathematical equations used to define molecular graph descriptors were extended to molecular matrices derived from the three-dimensional molecular geometry.[12-18] These numerical measures of the chemical structure, called topographic indices, offer a simple and efficient way for the computation of structural descriptors.

The information on distance topological indices U, V, X, and Y[19] were extended to three-dimensional molecular matrices, giving the information-theory operators **U**(**3M**), **V**(**3M**), **X**(**3M**), and **Y**(**3M**).[20] These operators are computed from atomic invariants that measure the information content of the matrix elements associated with the respective atom. Three-dimensional descriptors can be computed for the hydrogen-depleted molecular structure, i.e. the heavy-atom molecular structure, denoted with G, or for the whole molecule, denoted with H. These four operators can be computed from any molecular matrix **3M** derived from three-dimensional molecular geometry, such as the geometric distance matrix or the reciprocal geometric distance matrix.

The three-dimensional distance matrix, **3D** = **3D**(G), of a molecular structure G with N atoms is a real symmetric $N \times N$ matrix with the element $[\mathbf{3D}]_{ij} = [\mathbf{3D}]_{ji}$ representing the shortest Cartesian distance between atoms i and j in G; all geometric distances from the examples presented below are in Å. The reciprocal geometrical distance matrix of a molecular structure G with N atoms, **3RD** = **3RD**(G), is the square $N \times N$ symmetric matrix whose entries $[\mathbf{3RD}]_{ij}$ are equal to the reciprocal of the geometric distance between atoms i and j, i.e. $1/[\mathbf{3D}]_{ij}$, for non-diagonal elements, and is equal to zero for the diagonal elements:

$$[\mathbf{3RD}]_{ij} = \begin{cases} 1/[\mathbf{3D}]_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (5)$$

As mentioned above, these molecular matrices can be computed both for the hydrogen-depleted molecular structure *G*, or for the whole molecule *H*.

The graph vertex operators **VUinf**(**3M**,*G*), **VVinf**(**3M**,*G*), **VXinf**(**3M**,*G*), and **VYinf**(**3M**,*G*) apply the information theory equations to the non-zero elements of the molecular matrix **3M** that correspond to an atom *i*:

$$\mathbf{VUinf}(\mathbf{3M})_i = -\sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{[\mathbf{3M}]_{ij}}{\mathbf{AS}(\mathbf{3M})_i} \log_2 \frac{[\mathbf{3M}]_{ij}}{\mathbf{AS}(\mathbf{3M})_i} \quad (6)$$

$$\mathbf{VVinf}(\mathbf{3M})_i = \mathbf{AS}(\mathbf{3M})_i \log_2 \mathbf{AS}(\mathbf{3M})_i - \mathbf{VUinf}(\mathbf{3M})_i \quad (7)$$

$$\mathbf{VXinf}(\mathbf{3M})_i = \mathbf{AS}(\mathbf{3M})_i \log_2 \mathbf{AS}(\mathbf{3M})_i - \mathbf{VYinf}(\mathbf{3M})_i \quad (8)$$

$$\mathbf{VYinf}(\mathbf{3M})_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} [\mathbf{3M}]_{ij} \log_2 [\mathbf{3M}]_{ij} \quad (9)$$

where $\mathbf{AS}(\mathbf{M})_i$ represents the atom sum of the atom *i*, and the summations in equations (6) and (9) are done for the non-zero elements of the molecular matrix **3M**, $[\mathbf{3M}]_{ij} \neq 0$. The atom sum operator of the atom *i*, $\mathbf{AS}(\mathbf{3M})_i$, of a molecular structure *G* or *H* with *N* atoms, is defined as the sum of the elements in the column *i*, or row *i*, of the molecular matrix **3M**:

$$\mathbf{AS}(\mathbf{3M})_i = \sum_{j=1}^{N} [\mathbf{3M}]_{ij} = \sum_{j=1}^{N} [\mathbf{3M}]_{ji} \quad (10)$$

For a general molecular matrix **3M** derived from three-dimensional molecular geometry, the matrix elements $[\mathbf{3M}]_{ij}$ may have values lower than 1, giving negative terms for certain vertex structural descriptors computed with the graph vertex operators **VUinf**(**3M**), **VVinf**(**3M**), **VXinf**(**3M**), and **VYinf**(**3M**). The Randi⇲-like formula used in the case of the indices *U*, *V*, *X*, and *Y* is therefore replaced by the following equation:

$$f(x,y) = \begin{cases} (xy)^{-1/2} & \text{if } xy > 0 \\ -(|xy|)^{-1/2} & \text{if } xy < 0 \end{cases} \quad (11)$$

The operators **U**(**3M**), **V**(**3M**), **X**(**3M**), and **Y**(**3M**), representing information on matrix elements, are computed with the equations:

$$\mathbf{U}(\mathbf{3M}) = \frac{M}{\mu + 1} \sum_{\text{bonds}} f\left(\mathbf{VUinf}(\mathbf{3M})_i, \mathbf{VUinf}(\mathbf{3M})_j\right) \quad (12)$$

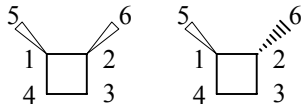
$$\mathbf{V}(\mathbf{3M}) = \frac{M}{\mu + 1} \sum_{\text{bonds}} f\left(\mathbf{VVinf}(\mathbf{3M})_i, \mathbf{VVinf}(\mathbf{3M})_j\right) \quad (13)$$

$$\mathbf{X}(\mathbf{3M}) = \frac{M}{\mu + 1} \sum_{\text{bonds}} f\left(\mathbf{VXinf}(\mathbf{3M})_i, \mathbf{VXinf}(\mathbf{3M})_j\right) \quad (14)$$

$$\mathbf{Y(3M)} = \frac{M}{\mu + 1} \sum_{\text{bonds}} f\left(\mathbf{VYinf(3M)}_i, \mathbf{VYinf(3M)}_j\right) \tag{15}$$

where M is the number of covalent bonds in the molecular structure G or H, μ is the cyclomatic number of G or H, i.e. the number of cycles in the graph, $\mu = M - N + 1$, and the summation goes over all bonds.

An example for the computation of the information operators **U**(**3M**), **V**(**3M**), **X**(**3M**), and **Y**(**3M**) is presented for the molecular graph of *cis*-1,2-dimethylcyclobutane **1** and *trans*-1,2-dimethylcyclobutane **2**:



1 **2**

The geometry of the two isomers was optimized with the molecular mechanics method MM+ implemented in HyperChem.[21] The geometrical distance matrix of the carbon skeleton of *cis*-1,2-dimethylcyclobutane **1**, **3D**(**1**,*G*), is:

3D(**1**,*G*)

	1	2	3	4	5	6
1	0.000	1.563	2.205	1.559	1.543	2.648
2	1.563	0.000	1.559	2.205	2.648	1.543
3	2.205	1.559	0.000	1.556	3.370	2.598
4	1.559	2.205	1.556	0.000	2.598	3.370
5	1.543	2.648	3.370	2.598	0.000	2.962
6	2.648	1.543	2.598	3.370	2.962	0.000

The atom sum vector **AS** is computed with formula (10) as the sum of the elements in the column i, or row i, of the matrix **3D**:

AS(**3D**,**1**) = {9.517, 9.517, 11.287, 11.287, 13.122, 13.122}

The information-theory operators **VUinf**(**3D**,*G*), **VVinf**(**3D**,*G*), **VXinf**(**3D**,*G*), and **VYinf**(**3D**,*G*) are applied to the geometric distance matrix **3D** to compute with formulas (6)-(9) new atom invariants:

VUinf(**3D**,**1**) = {2.283, 2.283, 2.257, 2.257, 2.280, 2.280}

VVinf(**3D**,**1**) = {28.651, 28.651, 37.210, 37.210, 46.453, 46.453}

VXinf(**3D**,**1**) = {21.730, 21.730, 25.477, 25.477, 29.919, 29.919}

VYinf(**3D**,**1**) = {9.204, 9.204, 13.991, 13.991, 18.813, 18.813}

The information on geometric distance matrix operators **U**(**3D**), **V**(**3D**), **X**(**3D**), and **Y**(**3D**) are computed with equations (12)-(15) from the above vectors:

U(**3D**,**1**) = 7.916 **V**(**3D**,**1**) = 0.534

X(**3D**,**1**) = 0.746 **Y**(**3D**,**1**) = 1.525

The reciprocal geometric distance matrix **3RD** is computed with equation (5) to give for *cis*-1,2-dimethylcyclobutane **1** the matrix **3RD**(**1**,*G*):

3RD(**1**,*G*)

	1	2	3	4	5	6
1	0.000	0.640	0.454	0.642	0.648	0.378
2	0.640	0.000	0.642	0.454	0.378	0.648
3	0.454	0.642	0.000	0.643	0.297	0.385
4	0.642	0.454	0.643	0.000	0.385	0.297
5	0.648	0.378	0.297	0.385	0.000	0.338
6	0.378	0.648	0.385	0.297	0.338	0.000

The reciprocal geometric distance matrix **3RD** is the basis for the computation of the information-theory atom operators **VUinf**(**3RD**,*G*), **VVinf**(**3RD**,*G*), **VXinf**(**3RD**,*G*), and **VYinf**(**3RD**,*G*) using the corresponding **AS** vector:

AS(**3RD**,**1**) = {2.761, 2.761, 2.420, 2.420, 2.045, 2.045}

VUinf(**3RD**,**1**) = {2.290, 2.290, 2.262, 2.262, 2.262, 2.262}

VVinf(**3RD**,**1**) = {1.755, 1.755, 0.823, 0.823, –0.152, –0.152}

VXinf(**3RD**,**1**) = {6.321, 6.321, 5.473, 5.473, 4.626, 4.626}

VYinf(**3RD**,**1**) = {–2.276, –2.276, –2.388, –2.388, –2.515, –2.515}

The above vectors of atom invariants are utilized to compute the indices **U**(**3RD**), **V**(**3RD**), **X**(**3RD**), and **Y**(**3RD**):

U(**3RD**,**1**) = 7.910 **V**(**3RD**,**1**) = –1.276

X(**3RD**,**1**) = 3.152 **Y**(**3RD**,**1**) = 7.655

The geometrical distance matrix of the carbon skeleton of *trans*-1,2-dimethylcyclobutane **2**, **3D**(**2**,*G*), is:

3D(**2**,*G*)

	1	2	3	4	5	6
1	0.000	1.566	2.211	1.563	1.547	2.613
2	1.566	0.000	1.563	2.211	2.613	1.547
3	2.211	1.563	0.000	1.561	3.352	2.612
4	1.563	2.211	1.561	0.000	2.612	3.352
5	1.547	2.613	3.352	2.612	0.000	3.792
6	2.613	1.547	2.612	3.352	3.792	0.000

The information-theory operators **VUinf**(**3D**,**2**), **VVinf**(**3D**,**2**), **VXinf**(**3D**,**2**), and **VYinf**(**3D**,**2**) are computed from the **AS**(**3D**,**2**) vector:

AS(**3D**,**2**) = {9.500, 9.500, 11.300, 11.300, 13.916, 13.916}

VUinf(**3D**,**2**) = {2.285, 2.285, 2.258, 2.258, 2.264, 2.264}
VVinf(**3D**,**2**) = {28.568, 28.568, 37.272, 37.272, 50.596, 50.596}
VXinf(**3D**,**2**) = {21.708, 21.708, 25.520, 25.520, 31.507, 31.507}
VYinf(**3D**,**2**) = {9.145, 9.145, 14.010, 14.010, 21.353, 21.353}

The information on geometric distance matrix operators **U**(**3D**), **V**(**3D**), **X**(**3D**), and **Y**(**3D**) computed with the above vectors are:

U(**3D**,**2**) = 7.920 **V**(**3D**,**2**) = 0.527
X(**3D**,**2**) = 0.740 **Y**(**3D**,**2**) = 1.502

Using equation (1) one obtains the reciprocal geometric distance matrix **3RD** for *trans*-1,2-dimethylcyclobutane **2**:

3RD(**2**,*G*)

	1	2	3	4	5	6
1	0.000	0.639	0.452	0.640	0.647	0.383
2	0.639	0.000	0.640	0.452	0.383	0.647
3	0.452	0.640	0.000	0.640	0.298	0.383
4	0.640	0.452	0.640	0.000	0.383	0.298
5	0.647	0.383	0.298	0.383	0.000	0.264
6	0.383	0.647	0.383	0.298	0.264	0.000

The above reciprocal geometric distance matrix **3RD** gives the vectors **AS**(**3RD**,**2**), **VUinf**(**3RD**,**2**), **VVinf**(**3RD**,**2**), **VXinf**(**3RD**,**2**), and **VYinf**(**3RD**,**2**):

AS(**3RD**,**2**) = {2.760, 2.760, 2.413, 2.413, 1.974, 1.974}
VUinf(**3RD**,**2**) = {2.291, 2.291, 2.263, 2.263, 2.245, 2.245}
VVinf(**3RD**,**2**) = {1.751, 1.751, 0.805, 0.805, –0.308, –0.308}
VXinf(**3RD**,**2**) = {6.322, 6.322, 5.461, 5.461, 4.432, 4.432}
VYinf(**3RD**,**2**) = {–2.280, –2.280, –2.393, –2.393, –2.495, –2.495}

The above vectors of atom invariants are utilized to compute the indices **U**(**3RD**), **V**(**3RD**), **X**(**3RD**), and **Y**(**3RD**):

U(**3RD**,**2**) = 7.917 **V**(**3RD**,**2**) = 2.320
X(**3RD**,**2**) = 3.178 **Y**(**3RD**,**2**) = 7.653

The extension of the **U**, **V**, **X**, and **Y** operators to molecular matrices derived from three-dimensional molecular geometry offers four new families of structural descriptors for QSPR and QSAR models.

4. TOPOLOGICAL ELECTRONIC INDEX

The topological electronic index was proposed by Kaliszan[22] to characterize quantitatively the differences of solutes polarity in gas-liquid chromatography:

$$T_1^E = \sum_{\text{all pairs}} \frac{\left|Q_i - Q_j\right|}{r_{ij}^2} \tag{16}$$

where Q_i denotes the partial charge on the ith atom, r_{ij} denotes the distance between ith and jth atoms, and the summation is done over all pairs of atoms in the molecule. CODESSA computes a similar index by considering all bonded atoms:[11]

$$T_2^E = \sum_{\text{all bonds}} \frac{\left|Q_i - Q_j\right|}{r_{ij}^2} \tag{17}$$

where Q_i denotes the partial charge on the ith atom, r_{ij} denotes the distance between ith and jth atoms, and the summation is done over all bonded pairs of atoms in the molecule.

5. CHARGED PARTIAL SURFACE AREA DESCRIPTORS

Charged partial surface area (CSPA)[23] descriptors have been defined by Jurs in terms of the solvent-accessible surface area of each atom and the atomic charge computed from the atomic electronegativity or with a quantum chemistry method. The molecule is considered as an assemble of hard spheres defined by the van der Waals radii of the atoms. The solvent-accessible surface area is traced out by the center of a solvent sphere (usually water) that rolls over the van der Waals surface of the molecule. The CSPA descriptors encode features responsible for polar interactions between molecules. The definitions of these descriptors are presented below:

(1) PPSA-1, partial positive surface area:

$$\text{PPSA-1}=\sum_i SA_i^+ \tag{18}$$

where SA_i^+ is the surface area of the positively charged atom i.

(2) PNSA-1, partial negative surface area:

$$\text{PNSA-1}=\sum_i SA_i^- \tag{19}$$

where SA_i^- is the surface area of the negatively charged atom i.

(3) PPSA-2, total charge weighted partial positive surface area:

$$\text{PPSA-2}=\sum_i SA_i^+ \sum_i Q_i^+ \tag{20}$$

where Q_i^+ is the partial positive charge of atom i, the first summation gives the partial positive surface area, and the second summation gives the total positive charges for the molecule.

(4) PNSA-2, total charge weighted partial negative surface area:

$$\text{PNSA-2} = \sum_i SA_i^- \sum_i Q_i^- \tag{21}$$

where Q_i^- is the partial positive charge of atom i, the first summation gives the partial negative surface area, and the second summation gives the total negative charges for the molecule.

(5) PPSA-3, atomic charge weighted partial positive surface area:

$$\text{PPSA-3} = \sum_i SA_i^+ Q_i^+ \tag{22}$$

(6) PNSA-3, atomic charge weighted partial negative surface area:

$$\text{PNSA-3} = \sum_i SA_i^- Q_i^- \tag{23}$$

(7) difference in charged partial surface areas:

$$\text{DPSA-1} = \text{PPSA-1} - \text{PNSA-1} \tag{24}$$

$$\text{DPSA-2} = \text{PPSA-2} - \text{PNSA-2} \tag{25}$$

$$\text{DPSA-3} = \text{PPSA-3} - \text{PNSA-3} \tag{26}$$

(8) fractional charged partial surface areas:

$$\text{FPSA-1} = \frac{\text{PPSA-1}}{MSA} \tag{27}$$

$$\text{FPSA-2} = \frac{\text{PPSA-2}}{MSA} \tag{28}$$

$$\text{FPSA-3} = \frac{\text{PPSA-3}}{MSA} \tag{29}$$

$$\text{FNSA-1} = \frac{\text{PNSA-1}}{MSA} \tag{30}$$

$$\text{FNSA-2} = \frac{\text{PNSA-2}}{MSA} \tag{31}$$

$$\text{FNSA-3} = \frac{\text{PNSA-3}}{MSA} \tag{32}$$

where *MSA* is the molecular surface area computed with the formula:

$$MSA = \sum_{i=1}^{N} SA_i \tag{33}$$

(9) surface weighted charged partial surface areas:

$$\text{WPSA-1} = \frac{\text{PPSA-1} \cdot MSA}{1000} \tag{34}$$

$$\text{WPSA-2} = \frac{\text{PPSA-2} \cdot MSA}{1000} \tag{35}$$

$$\text{WPSA-3} = \frac{\text{PPSA-3} \cdot MSA}{1000} \tag{36}$$

$$\text{WNSA-1} = \frac{\text{PNSA-1} \cdot MSA}{1000} \tag{37}$$

$$\text{WNSA-2} = \frac{\text{PNSA-2} \cdot MSA}{1000} \tag{38}$$

$$\text{WNSA-3} = \frac{\text{PNSA-3} \cdot MSA}{1000} \tag{39}$$

(10) RPCG, relative positive charge:

$$\text{RPCG} = \frac{Q^{+}_{\max}}{\sum_i SA^{+}_i} \tag{40}$$

where $Q^{+}_{\max}$ is the charge of the most positive atom.

(11) RNCG, relative negative charge:

$$\text{RNCG} = \frac{Q^{-}_{\max}}{\sum_i SA^{-}_i} \tag{41}$$

where $Q^{-}_{\max}$ is the charge of the most negative atom.

(12) RPCS, relative positive charged surface area:

$$\text{RPCS} = \text{RPCG} \cdot SA^{+}_{\max} = \frac{Q^{+}_{\max} SA^{+}_{\max}}{\sum_i SA^{+}_i} \tag{42}$$

where $SA^{+}_{\max}$ is the surface area of the most positive atom.

(13) RNCS, relative negative charged surface area:

$$\text{RNCS} = \text{RNCG} \cdot SA^{-}_{\max} = \frac{Q^{-}_{\max} SA^{-}_{\max}}{\sum_i SA^{-}_i} \tag{43}$$

where $SA^{-}_{\max}$ is the surface area of the most negative atom.

All the above CPSA descriptors can be computed from the 3D molecular structure using two available programs, ADAPT[24,25] or CODESSA.[11,26-33] CODESSA offers several related descriptors for hydrogen acceptor or donor atoms in a molecule. We present below a selection of these descriptors.

The HASA-1 descriptor is calculated as the sum of exposed surfaces areas of all possible hydrogen acceptor atoms in the molecule:

$$\text{HASA-1} = \sum SA_a \tag{44}$$

where SA_a is the surface area of the hydrogen acceptor atom. The following atoms are considered as possible hydrogen acceptors: carbonyl oxygen atoms (except in COOR), hydroxy oxygen atoms, amino nitrogen atoms, aromatic nitrogens, and mercapto sulfur atoms.

The HDSA-2 descriptor is directly related to the hydrogen bonding between the molecules in condensed media. It is calculated according to the following formula:

$$\text{HDSA-2} = \sum \frac{Q_d SA_d^{1/2}}{MSA} \tag{45}$$

where Q_d is the partial charge on the hydrogen bonding donor atom, SA_d denotes the exposed surface area of this atom, and MSA is the total molecular surface area calculated from the van der Waals radii of the atoms. The summation is performed over all possible hydrogen bonding donor sites in the molecule.

The hydrogen-bonding ability of compounds can also be characterized by HDCA-2, the hydrogen bonding donor charged surface area defined with the equation:

$$\text{HDCA-2} = \sum \frac{Q_d SA_d^{1/2}}{MSA^{1/2}} \tag{46}$$

The summation in the formula of HDCA-2 goes over all possible hydrogen bonding donor and acceptor pairs in the molecule. In the computation of HDCA-2 hydrogen atoms attached to carbons connected directly to carbonyl or cyano group are included as possible hydrogen bonding donor centers.

The fractional hydrogen donor surface area FHDSA-2 is defined as:

$$\text{FHDSA-2} = \frac{\text{HDSA-2}}{MSA^{1/2}} \tag{47}$$

The 3D descriptors from this section can be computed either with electrostatic charges computed from the atomic electronegativity or with partial atomic charges computed with a quantum chemistry method.

6. TLSER -THEORETICAL LINEAR SOLVATION ENERGY RELATIONSHIP

The linear solvation energy relationship (LSER)[34,35] model was developed by Kamlet, Taft, and co-workers to explain solvent effects on various free energy based properties. The LSER approach contains parameters that measure the solute size, polarity/polarizability, hydrogen bond donating and accepting capacities; this model was applied with success to various QSAR studies, such as the prediction of octanol/water partition coefficients, boiling temperatures, and critical constants. Famini used the LFER philosophy with quantum chemically derived descriptors, and proposed the theoretical linear solvation energy relationship (TLSER) model.[36-38] This model uses a new set of theoretical parameters to correlate a property P:

$$P = a_0 + a_1 V_{mc} + a_2 \pi^* + a_3 \varepsilon_\alpha + a_4 \varepsilon_\beta + a_5 q^+ + a_6 q^- \tag{48}$$

where V_{mc} is the molecular van der Waals volume, π^* is the polarizability index, ε_α is the covalent acidity index, ε_β is the covalent basicity index, q^+ is the electrostatic acidity index, and q^- is the electrostatic basicity index. The polarizability index is obtained by dividing the molecular polarizability to V_{mc}, q^+ is the magnitude of the most positively charged hydrogen atom in the molecule, and q^- is the magnitude of the most negative atomic partial charge in the molecule. The covalent hydrogen bonding acidity index ε_α is computed with the formula:

$$\varepsilon_\alpha = 0.30 - \frac{\left|E_{\mathrm{LUMO}} - E_{\mathrm{HOMO},w}\right|}{100} \tag{49}$$

where E_{LUMO} is the LUMO energy of the molecule, and $E_{\mathrm{HOMO},w}$ is the HOMO energy of water. The covalent hydrogen bonding basicity index ε_β is:

$$\varepsilon_\beta = 0.30 - \frac{\left|E_{\mathrm{HOMO}} - E_{\mathrm{LUMO},w}\right|}{100} \tag{50}$$

where E_{HOMO} is the HOMO energy of the molecule, and $E_{\mathrm{LUMO},w}$ is the LUMO energy of water. Several QSPR and QSAR applications of the TLSER model were presented in two reviews.[36,37]

7. QUANTUM SUBSTITUENT PARAMETERS

Using the extended Hückel theory, Esaki defined new electronic substituent parameters computed from the energy of the single occupied molecular orbital (SOMO).[39] The SOMO resistance of the substituent X is:

$$Re_{\mathrm{SOMO}}(\mathrm{X}) = \exp\left[-c^2_{\mathrm{SOMO}}(\mathrm{X})\right] \tag{51}$$

where c_{SOMO} is the SOMO coefficient of the hybrid orbital which participates in bonding to the parent structure, and X is not hydrogen. For hydrogen a modified formula is employed:

$$Re_{\mathrm{SOMO}}(\mathrm{H}) = 1/\left[\exp c^2_{\mathrm{SOMO}}(\mathrm{H}) - 1\right] \tag{52}$$

The frontier electron current intensity is:

$$I_{\mathrm{SOMO}}(\mathrm{X}) = E_{\mathrm{SOMO}}(\mathrm{X}) / Re_{\mathrm{SOMO}}(\mathrm{X}) \tag{53}$$

where $E_{\mathrm{SOMO}}(\mathrm{X})$ is the energy of the SOMO orbital. Frontier substituent constants are obtained from the above parameters adjusted relative to hydrogen as standard:

$$E(\mathrm{X}) = -10\left[E_{\mathrm{SOMO}}(\mathrm{X}) - E_{\mathrm{SOMO}}(\mathrm{H})\right] \tag{54}$$

$$Re(\mathrm{X}) = 10\left[Re_{\mathrm{SOMO}}(\mathrm{X}) - Re_{\mathrm{SOMO}}(\mathrm{H})\right] \tag{55}$$

$$I(\mathrm{X}) = -10\left[I_{\mathrm{SOMO}}(\mathrm{X}) - I_{\mathrm{SOMO}}(\mathrm{H})\right] \tag{56}$$

Esaki demonstrated the advantage of the new quantum parameters by using them, together with other substituent constants, to develop 45 Hansch-type QSAR models.

8. FRONTIER INDICES FOR GROUPS OF ATOMS

Pires assumed that in the ligand-receptor interaction the frontier orbitals of both the ligand and receptor play an important role and proposed a set of three frontier indices for groups of atoms.[40] Consider a group *S* composed of *M* bonded atoms. The group frontier electron density is:

$$G_S^{\mathrm{FED}} = \sum_{i=1}^{M} 2c_{\mathrm{HOMO},i}^2 \tag{57}$$

where $c_{\mathrm{HOMO},i}$ is the coefficient for the highest occupied molecular orbital on the ith atom from the group. The group frontier orbital density is:

$$G_S^{\mathrm{FOD}} = \sum_{i=1}^{M} 2c_{\mathrm{LUMO},i}^2 \tag{58}$$

where $c_{\mathrm{LUMO},i}$ is the coefficient for the lowest unoccupied molecular orbital on the ith atom from the group. A combination of the terms from the above two descriptors gives the group frontier radical density:

$$G_S^{\mathrm{FRD}} = \sum_{i=1}^{M} \left(c_{\mathrm{HOMO},i}^2 + c_{\mathrm{LUMO},i}^2 \right) \tag{59}$$

These three frontier indices computed with the AM1 method were used to model the mutagenic activity of triazenes. Good correlations were obtained when the octanol/water partition coefficient was added to the quantum frontier descriptors.

9. GIPF - GENERAL INTERACTION PROPERTIES FUNCTION

Politzer proposed the general interaction properties function (GIPF) method for the correlation of physical, chemical, and biological properties with molecular surface descriptors computed from the electrostatic potential.[41-46] The electrostatic potential $V(r)$ created in the space around a molecule by its nuclei and electrons is:

$$V(r) = \sum_{i=1}^{N} \frac{Z_i}{\left| R_i - r \right|} - \int \frac{\rho(r')r'}{|r'-r|} \tag{60}$$

where Z_i is the charge on nucleus i, located at R_i, and $\rho(r)$ is the electronic density function. $V(r)$ gives the interaction energy between a positive point charge of unitary magnitude located at r and the unperturbed charge distribution of the molecule. A molecular surface is defined by the 0.001 a.u. contour of the electronic density $\rho(r)$. Politzer found that the hydrogen bond acceptor capability of a molecule is quantitatively related to V_{min}, the most negative potential in space, and $V_{\mathrm{S,min}}$, the most negative surface potential, while the hydrogen bond donor capability is related to the most positive surface potential, $V_{\mathrm{S,max}}$.

The average deviation of the electrostatic potential on the molecular surface is a measure of local polarity:

$$\Pi = \frac{1}{k} \sum_{i=1}^{k} \left| V(r_i) - \overline{V}_{\mathrm{S}} \right| \tag{61}$$

where $V(r_i)$ is the potential at the ith point on the surface and $\overline{V}_{\mathrm{S}}$ is its average value:

$$\overline{V}_{\mathrm{S}} = \frac{1}{k} \sum_{i=1}^{k} V(r_i) \tag{62}$$

It was found that Π is correlated with π*, the LSER polarity/polarizability index. The variance of the positive regions of electrostatic potential on the molecular surface, $V^+(r)$, is:

$$\sigma_+^2 = \frac{1}{m}\sum_{i=1}^{m}\left[V^+(r_i) - \overline{V}_S^+\right]^2 \tag{63}$$

where $\overline{V}_S^+$ is the average of $V(r)$ on the positive regions of the molecular surface:

$$\overline{V}_S^+ = \frac{1}{m}\sum_{i=1}^{m} V^+(r_i) \tag{64}$$

The variance of the negative regions of electrostatic potential on the molecular surface, $V^-(r)$, is:

$$\sigma_-^2 = \frac{1}{n}\sum_{i=1}^{n}\left[V^-(r_i) - \overline{V}_S^-\right]^2 \tag{65}$$

where $\overline{V}_S^-$ is the average of $V(r)$ on the negative regions of the molecular surface:

$$\overline{V}_S^- = \frac{1}{n}\sum_{i=1}^{n} V^-(r_i) \tag{66}$$

The variance of the electrostatic potential over the molecular surface is:

$$\sigma_{tot}^2 = \sigma_+^2 + \sigma_-^2 \tag{67}$$

An electrostatic balance term is obtained from the above σ parameters:

$$\upsilon = \frac{\sigma_+^2\sigma_-^2}{\left[\sigma_{tot}^2\right]^2} \tag{68}$$

The GIPF descriptors, when used in conjunction with a parameter measuring the molecular size (molecular volume or surface area), were effective in correlations with critical constants, boiling temperatures,[45] and octanol/water partition coefficients.[42,43]

10. CoMFA - Comparative Molecular Field Analysis

Although molecular field descriptors where introduced more than 20 years ago, the lack of an efficient statistical algorithm for the analysis of thousands of field variables prevented the widespread use of 3D QSAR models using molecular fields. In 1988 Cramer[47,48] used partial least squares to correlate the molecular field descriptors with biological activities, proposing the now widely used comparative molecular field analysis (CoMFA).[9,10,49] The main steps for the development of a CoMFA model are:

(1) A set of molecules is selected by considering only those compounds that interact in a similar manner with the same receptor.

(2) The geometry of all molecules in the study set is optimized with molecular mechanics or quantum mechanics methods. Any information regarding the active conformations of the ligands in the actual receptor must be used to optimize the

appropriate molecular geometry. For certain molecules several low energy conformations may be considered.

(3) All optimized molecules are aligned (superimposed) using some pharmacophore hypothesis. The CoMFA model depends on the molecules alignment and errors in this step can provide 3D QSAR models that have a low predictive power. Several alignment assumptions have to be investigated in order to identify the best model.

(4) A subset of molecules is selected to generate the CoMFA virtual receptor model.

(5) Atomic partial charges are calculated with an electrostatic (based on electronegativity) or quantum mechanics method for all molecules.

(6) The aligned molecules are placed in a box that is larger in all directions than the volume occupied by the superimposed compounds, and the box is partitioned into a grid with points separated by an adjustable distance, usually between 0.5 and 2 Å.

(7) A neutral atom, such as a carbon atom, is used as a steric probe to compute the steric field on the grid points. In order to compute the value of the steric field on a certain grid point j, the steric probe is placed in that grid point and the van der Waals interaction energy, $E_{vdW,j}$, is computed with the formula:

$$E_{vdW,j} = \sum_{i=1}^{n} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} \right) \tag{69}$$

where n is the number of atoms in the molecule, r_{ij} is the distance between atom i of the molecule and the grid point j where the probe atom is located, A_{ij} and C_{ij} are constants depending on the types of atom i and steric probe.

(8) The electrostatic field values in each grid point are computed by placing a positively or negatively charged atom in the grid point:

$$E_{C,j} = \sum_{j=1}^{n} \frac{q_i q_j}{\varepsilon r_{ij}} \tag{70}$$

where $E_{C,j}$ is the value of the electrostatic field at the grid point j, q_i is the partial charge of atom i in the molecule, q_j is the charge of the probe atom located at the grid point j, and ε is the dielectric constant. The grid points inside the van der Waals molecular surface were excluded from the calculation of the electrostatic field.

(9) The values of the steric and electrostatic fields are collected into a QSAR table in order to be correlated with the biological activity values. The resulting matrix, containing thousands of columns corresponding to individual grid points, cannot be correlated with the multilinear regression method. The PLS method uses this data matrix to generate a 3D QSAR. PLS extracts principal component-like vectors (latent variables) from the matrices of independent and dependent variables. This method takes a matrix containing a large number of potentially useful structural descriptors, that can be highly intercorrelated, and offers a correlation using the latent variables. The optimum number of latent variables is determined by cross-validation.

(10) The PLS model is used to generate contour maps showing favorable and unfavorable regions for electropositive or electronegative regions, as well as favorable and unfavorable steric regions around the molecules.

(11) Predictions for molecules not included in the derivation of the PLS model can be made by computing their steric and electrostatic fields, and by using the grid values into the PLS model.

CoMFA is the most important 3D QSAR model, used in a few hundred drug design studies, mainly to describe the ligand-receptor interactions. However, this method seems to be less appropriate for the correlation of the *in vivo* data. This highly successful model was followed by the introduction of a large number of related 3D QSAR methods.

11. SOMFA - SELF-ORGANIZING MOLECULAR FIELD ANALYSIS

Richards proposed a simplified CoMFA model that generates a 3D QSAR model considering only the first PLS component, the self-organizing molecular field analysis (SOMFA).[50] The most significant steps of the SOMFA algorithm are:

Steps (1)-(6) are identical with those employed for CoMFA.

(7) The mean centered activity MCA_k of a molecule k from the calibration set is computed by subtracting the mean activity of all molecules from each molecule's activity, giving a scale where the most active molecules have positive values and the least active molecules have negative values:

$$\mathrm{MCA}_k = \mathrm{A}_k - \mathrm{MA} \tag{71}$$

where A_k is the activity of molecule k and MA is the mean activity of the molecules from the calibration set.

(8) For each molecule, the shape field is computed by assigning the value 1 to all grid points situated inside the van der Waals molecular surface and 0 for the outside grid points.

(9) The electrostatic potential values at the grid points of each molecule are computed with the formula:

$$\mathrm{ESP}_j = \sum_{i=1}^{n} \frac{q_i}{r_{ij}} \tag{72}$$

where n is the number of atoms in the molecule, q_i is the partial charge of atom i in the molecule, and r_{ij} is the distance between atom i and the grid point j.

(10) A self-organizing master grid (SOMG) is generated for each field considered using the grid values of all molecules in the calibration set. For a field f (shape or electrostatic potential) the value of the self-organizing master grid in the grid point $j = j(x,y,z)$, SOMG(f,j), is computed with the formula:

$$\mathrm{SOMG}(f, j) = \sum_{k=1}^{N} f(j)_k \, \mathrm{MCA}_k \tag{73}$$

where the summation goes for all N molecules in the calibration set, and $f(j)_k$ is the value of the field f in the grid point j for compound k.

(11) For each molecule k in the calibration or prediction set a SOMFA activity derived from the field f computed in each grid point j is determined for both shape and electrostatic fields:

$$\mathrm{SOA}(f)_k = \sum_{j=1}^{n} f(j)_k \mathrm{SOMG}(f, j) \tag{74}$$

where the summation is performed for all grid points n.

(12) The biological activity of the molecule k, BA_k, is computed as a linear combination of the corresponding SOA(f) values for the two molecular fields:

$$\mathrm{BA}_k = c_1 \mathrm{SOA}(\mathrm{shape})_k + (1 - c_1)\mathrm{SOA}(\mathrm{electrostatic})_k \tag{75}$$

where the mixing coefficient c_1 is optimized for each 3D QSAR model.

(13) The shape and electrostatic potential self-organizing master grid (SOMG) can generate visual maps of the important features that determine the biological activity.

We have to mention that starting with the same field values for each grid point, the SOMFA model is identical with the CoMFA model computed with only the first PLS component.

12. RSM - RECEPTOR SURFACE MODEL

Hahn[51,52] proposed a 3D QSAR model that uses a set of active compounds to generate a virtual receptor model; the virtual receptor is a surface formed by points provided with several properties that are used when calculating interaction energy between a molecule and a surface model. The main steps of the receptor surface model (RSM) are:

Steps (1)-(5) are identical with those employed for CoMFA.

(6) A steric surface representing the virtual receptor is generated to enclose the set of aligned molecules. The surface of the virtual receptor is generated by a volumetric field (shape field), characterizing molecular shape, which is produced for each aligned molecule. The shape fields from each individual molecule are combined to produce a final volumetric shape field from which an explicit surface is generated. First, a three-dimensional regularly spaced grid is superposed over the aligned set of molecules; the grid box dimensions are extended several Å in each direction from the coordinates of every molecule. The steric field is computed for each point of the grid, and an isosurface of the field is used to generate the surface of the virtual receptor. Two field functions are used to create the shape of the virtual receptor, namely the van der Waals field function and the Wyvill field function. Each field source corresponds to an atom. The van der Waals field function generated by the atom i from a molecule is:

$$V(r_{ij})_i = r_{ij} - r_{\mathrm{VDW},i} \tag{76}$$

where r_{ij} is the distance from the atom i and the grid point j and $r_{\mathrm{VDW},i}$ is the van der Waals radius of the atom i. This field function, which is computed for every grid point, has the property that inside the van der Waals volume the value is negative, outside the volume the value is positive, and at the van der Waals surface the value $V(r)$ is zero. If a grid point contains a shape field value computed for a different atom, the smaller of the two values is assigned to that grid point. The Wyvill function is a bounded function that decays completely in a finite distance R:

$$V(r_{ij})_i = -\frac{4r_{ij}^6}{9R^6} + \frac{17r_{ij}^4}{9R^4} - \frac{22r_{ij}^2}{9R^2} + 1 \tag{77}$$

where r_{ij} is the distance from the atom i and the grid point j. A field value is the sum of the field values contributed by each atom; if a grid point is outside of R, its shape field value is not computed. The value of R depends on the atom type, and usually its value is twice the van der Waals radius of the atom i. This function has the properties that $V(0) = 1$, $V(R) = 0$, and $V(R/2) = 1/2$. Using the shape field values the marching cubes isosurface algorithm produces a set of triangulated surface points representing the surface of the virtual receptor. The default grid spacing of 0.5 Å yields an average surface density of 6 points/Å^2. This gives an average distance between neighboring points (points in the same triangle) of about 0.47 Å.

(7) After a surface is created, several properties of the virtual receptor associated with each surface point are assigned; these properties include partial charge, electrostatic potential, hydrogen-bonding propensity, and hydrophobicity. These values are used when calculating interaction energy between a molecule and a surface model.

(8) A partial atomic charge is assigned to each point from the virtual receptor site. This atomic charge is determined as complementary to the partial charge of any atom in contact with the surface. If a single molecule is used to generate the virtual receptor, each surface point is assigned a charge which is equal to but opposite in sign to the charge of the closest atom in the molecule. If a set of molecules are employed to generate the virtual receptor, each surface point is assigned a charge which is equal and opposite to the average partial atomic charge of the set. The average for a surface point is found by summing the partial atomic charges of the closest atom in each molecule and dividing by the number of molecules. This method assumes that each molecule contributes equally to the description of the virtual receptor.

(9) The second property assigned to each surface point is the electrostatic potential, considered to be complementary with that of the ligand. If a single molecule is used to generate the virtual receptor, each surface point is assigned an electrostatic potential value which is equivalent to but opposite in sign to the distance-dependent electrostatic potential at that surface point:

$$\mathrm{ESP}_j = \sum_{i=1}^{n} \frac{q_i q_j}{r_{ij}} \tag{78}$$

where the summation goes over all atoms in the molecule. If a set of molecules are employed to generate the virtual receptor, each surface point is assigned an electrostatic potential which is equal and opposite to the average of the values obtained for all molecules.

(10) The third property assigned to each surface point is the hydrogen bond property, corresponding to the tendency of the surface point to be involved in a hydrogen bond. If a hydrogen bond acceptor property would be desirable for a certain surface point a value of 1.0 is assigned to the respective point, while if a hydrogen bond donor hydrogen property would be desirable for a certain surface point a value of –1.0 is assigned to the respective point. Hydrogen bond acceptors are defined as any oxygen or nitrogen atom with a free

lone pair of electrons, and hydrogen bond donors are any hydrogens attached to oxygen or nitrogen. The hydrogen bond values are averaged for all molecules, and the resulting surface point values are in the range [–1.0, 1.0].

(11) The fourth property assigned to each surface point is the hydrophobicity, with a value of one assigned to points in hydrophobic regions and zero to all other points. A hydrophobic point is a point with a low partial charge (absolute values less than 0.15), a low electrostatic potential (absolute values less than 0.01), and a low hydrogen bond donating or accepting propensity (absolute values less than 0.1).

(12) With the virtual receptor model defined in the above steps, the ligand-receptor interaction of a set of molecules is evaluated by computing their interaction energy with the virtual receptor. For a molecule a set of descriptors is obtained by computing the interaction energy between each surface point from the virtual receptor and the atoms in the molecule. These descriptors are stored in a QSAR table together with the biological activity of the investigated molecule.

The total interaction energy contribution of a single atom is computed by summing the individual interaction energies of all valid atom/point pairs. An atom/point pair is valid if the distance between the pair is less than a 6 Å cutoff distance and if the atom is inside the surface with respect to the point. The nonbonded energy between an atom and the surface points is composed of a van der Waals term, an electrostatic term, and a desolvation energy correction term. The van der Waals term is a scaled Lennard-Jones equation:

$$\mathrm{E}_{\mathrm{vdW},ij} = K\left[\left(\frac{\mathrm{RA}}{r_{ij}}\right)^{12} - 2\left(\frac{\mathrm{RA}}{r_{ij}}\right)^{6}\right]D \tag{79}$$

where RA is the hybridization corrected van der Waals radius for the atom, r_{ij} is the distance between the atom i and the surface point j, K is the well depth constant and is set to 0.1 for all van der Waals atom/point interactions, and D is an empirically derived point-density scaling factor, set to 0.01 for the default grid resolution of 0.5 Å and surface point density of 6 points Å^{-2}. RA is computed with the equation:

$$\mathrm{RA} = r_{\mathrm{VDW},i} \cdot C_{\mathrm{h}} \tag{80}$$

where $r_{\mathrm{VDW},i}$ is the van der Waals radius of the atom i and C_{h} is a hybridization correction.

The electrostatic term is a monopole-monopole Coulombic function computed with the equation:

$$\mathrm{E}_{\mathrm{C},ij} = 322.1D\frac{q_i q_j}{r_{ij}} S(r_{ij}) \tag{81}$$

where r_{ij} is the distance between the atom i and the grid point j, q_i is the partial charge of the atom i, q_j is the charge of the surface point j, D is the point-density scaling factor used in the computation of $\mathrm{E}_{\mathrm{vdW}}$, and $S(r_{ij})$ is an atom-based switching function:

$$S(r_{ij}) = \frac{(r_{\mathrm{off}}^2 - r_{ij}^2)^2 (r_{\mathrm{off}}^2 + 2r_{ij}^2 - 3r_{\mathrm{on}}^2)}{(r_{\mathrm{off}}^2 - r_{\mathrm{on}}^2)^3} \tag{82}$$

where r_{on} = 7 Å and r_{off} = 8 Å. If $r_{ij} < r_{\mathrm{on}}$ then $S(r_{ij}) = 1$.

The desolvation energy term is a penalty function for the presence of polar atoms in hydrophobic regions of the receptor surface model. If the fraction of hydrophobic surface points around a polar atom is greater than 90%, then the desolvation energy term is proportional to the exposed surface area of the polar atom; a value of 0.3 kcal/mol Å^2 is used. The desolvation energy is added to the total energy.

(13) The interactions between a ligand and the virtual receptor are used to compute several structural descriptors. The nonbonded energy terms, i.e. the van der Waals energy E_{VDW}, the electrostatic energy E_C, and their sum E_{total}, representing the sum for all surface points of the respective energy terms, are used as descriptors in MLR equations.

Another descriptor is the intramolecular strain energy of the ligand inside the virtual receptor, E_{inside}. To compute this descriptor, the conformation of the ligand is optimized inside the virtual receptor.

These global descriptor that characterize the ligand-receptor interaction were used to model the binding affinity for corticosteroid binding globulin of 31 steroids, and the inhibition of dopamine β-hydroxylase by 52 1-(substituted-benzyl)imidazole-2(3*H*)-thiones.[52]

The Cerius^2 implementation of the receptor surface model[53] offers a larger population of descriptors for the generation of the 3D QSAR model, by collecting into the QSAR table, for each molecule, the nonbonded energy terms that measure the interaction between that molecule and each surface point *j*, i.e. the van der Waals energy $E_{VDW,j}$ and the electrostatic energy $E_{C,j}$, and their sum $E_{total,j}$. From the large set of several hundreds of parameters, the selection of descriptors for the 3D QSAR model is made with the genetic function approximation (GFA) algorithm.[54-57]

The GFA algorithm generates an initial population of QSAR equations by the random selection of molecular descriptors. The length of the equations is determined by the number of molecular descriptors selected and is allowed to increase or decrease with a certain probability. Each equation is fit to the experimental data using linear least-squares regression techniques, and the 3D QSAR models are ranked according to their lack of fit (LOF), which is an adjusted least-squares error (LSE) statistical index:

$$\mathrm{LOF} = \frac{\mathrm{LSE}}{\left(1 - \frac{c + df}{n}\right)^2} \tag{83}$$

where *c* is the number of basis functions (number of terms in the 3D QSAR equation), *d* is the smoothing parameter, *f* is the total number of structural descriptors contained in all basis functions, and *n* is the number of molecules in the calibration set. The addition of a new descriptor to a 3D QSAR equation may reduce the LSE, but it also increases the values of *c* and *f*, and the LOF index may increase. In this way, the LOF index avoids overfitting of the data by limiting the tendency to add new descriptors and favoring simple, more compact models. The user-defined smoothing parameter *d* controls the growing of the 3D QSAR equations.

The GFA algorithm uses a genetic algorithm to perform the genetic crossover of the best equations and the elimination of the poorer equations. Using pairs of 3D QSAR equations with a low LOF, the genetic algorithm cuts and separates pieces of equations

and then recombines the fragments to form new equations. Next, mutation operators are applied randomly in order to increase the diversity of the population. The following probability levels p were used for the equation mutation operators:

(1) add new term: $p = 0.5$; this mutation increases the diversity in the equation population by randomly adding a new term to a child equation.

(2) reduce equation: $p = 0.5$; this mutation operation generates smaller models by eliminating from a child equation the term with the lowest contribution to the model.

(2) extend equation: $p = 0.5$; this mutation operation generates larger models by adding to a child equation a new descriptor. Each not yet used descriptor is tested and the algorithm retains the one that has the highest contribution to the model.

The genetic algorithm uses the LOF index to select equations for crossover and survival. With new generations, the equation population evolves to a set of higher quality 3D QSAR models. The output of a GFA calculation consists of a population of 3D QSAR equations that model the relationship between the structural descriptors and the biological activity.

Another modification of the RSM is the comparative receptor surface analysis (CoRSA)[58] that uses the nonbonded energy terms that measure the interaction between a molecule and each surface point j, i.e. the van der Waals energy $E_{VDW,j}$ and the electrostatic energy $E_{C,j}$, and their sum $E_{total,j}$, in a partial least squares data analysis. This 3D QSAR algorithm was used to model the activity of a set of compounds that act as calcium channel agonists for the guinea pig left atrium assay.

13. MOLREF - MOLECULAR REFERENCE

Alsberg proposed a 3D QSAR model, the molecular reference (MOLREF),[59] that compares the structure of test molecules with a reference molecule. The reference may be a representative molecule for the whole set of compounds, or a pseudomolecule that contains all relevant structural characteristics. MOLREF uses a set of descriptors that measure the similarity between each molecule and the reference; subsequently, the relationship between the **X** and **Y** matrices is computed with the PLS method.

The MOLREF algorithm contains the following steps:

(1) The geometry of all molecules in the study set is optimized with molecular mechanics or quantum mechanics methods.

(2) A compound that contains the relevant structural characteristics for the whole set of compounds is selected as the reference molecule. If such a molecule does not exist a pseudomolecule is generated.

(3) All molecules are superimposed over the reference using some pharmacophore hypothesis.

(4) Each atom in the reference and model molecules is characterized by the x, y, and z coordinates, van der Waals radius, and atomic charge. Other atomic descriptors can be added if these five parameters are not sufficient to generate a good QSAR model.

(5) A model molecule *i* is selected. For each atom *j* from the reference find the nearest neighbor atom *k* from the model molecule. The atomic descriptors of atom *k* are added to the row *i* of matrix **X**. If the reference contains *n* atoms and each atom is characterized by *m* atomic descriptors, then each row from the matrix **X** has $n \times m$ columns.

(6) Repeat step (5) for all model molecules.

(7) A 3-D QSAR model is obtained by applying the PLS algorithm to the matrices **X** and **Y**.

MOLREF was used to model the toxicity of 42 benzene derivatives (mainly phenols) against *Daphnia magna*; benzene was employed as the reference, but similar results were obtained when phenol was selected as reference.

14. MOLECULAR SIMILARITY INDICES

The 2D and 3D similarity indices of chemical structures are widely used in searching molecular databases, in pharmacophore identification, and classification of molecules in structure-activity studies.[60-64] Recent applications of these similarity indices as structural descriptors in 3D QSAR offers an efficient way of condensing a large number of field, grid or surface descriptors into a single number.

The molecular similarity may be quantified theoretically by comparing the electron densities, ρ_A and ρ_B, of two molecules *A* and *B*, and calculating an index of similarity, R_{AB}, as first introduced by Carbó:[65-68]

$$R_{AB} = \frac{\int \rho_A \rho_B \mathrm{d}v}{\left(\int \rho_A^2 \mathrm{d}v\right)^{1/2} \left(\int \rho_B^2 \mathrm{d}v\right)^{1/2}} \tag{84}$$

where the integrations are considered over all space. Several QSPR and QSAR application of this similarity index were presented in two reviews.[67,68]

Willett revealed that the Carbó index formula computed for a property distributed on a grid surrounding a molecule is essentially equivalent to the Cosine coefficient.[62,64] The Cosine similarity coefficient of two molecules *A* and *B*, $C(P)_{AB}$, computed for a property $P = P(j)$ defined in a grid point $j = j(x,y,z)$ is:

$$C(P)_{A,B} = \frac{\sum_{j=1}^{n} P(j)_A P(j)_B}{\left[\sum_{j=1}^{n} P(j)_A^2 \sum_{i=1}^{n} P(j)_B^2\right]^{1/2}} \tag{85}$$

where *n* is the number of grid points.

Hodgkin and Richards[69] pointed that in the Carbó index the denominator is a normalizing constant and R_{AB} varies in the range 0 to 1. Such an index of similarity is required to have a value of 1 when the electron density distributions in the two molecules are identical. However, substitution of $\rho_A = a\rho_B$ into the above equation, where *a* is a

constant, gives an index of unity. Thus the Carbó index represents the similarity of the shapes of the density distributions but not of the magnitudes as well.

Although originally proposed as a method of comparing molecules in terms of electron density, Hodgkin and Richards[69] proposed to use the formula of the Carbó index with other quantum properties, such as molecular electrostatic potential (MEP) or molecular electrostatic field (MEF). The use of electrostatic potentials and electrostatic fields is particularly attractive since they are better discriminators than charge and problems can be avoided if only values external to a van der Waals volume of the molecule are considered. The electrostatic potentials and electrostatic fields can be calculated over a grid of points surrounding a molecule.

In an attempt to increase the magnitude sensitivity of similarity calculations, Hodgkin and Richards proposed the Hodgkin index:[69]

$$H_{AB} = \frac{2\int \rho_A \rho_B \mathrm{d}v}{\int \rho_A^2 \mathrm{d}v + \int \rho_B^2 \mathrm{d}v} \tag{86}$$

Analogously with the Carbó index, the Hodgkin index can be used with grid-based properties such as MEP or MEF, when its formula is identical with that of the Dice coefficient, as found by Willett.[62,64] The Dice similarity coefficient of two molecules A and B, $D(P)_{AB}$, computed for a property $P = P(j)$ defined in a grid point $j = j(x,y,z)$ is:

$$D(P)_{A,B} = \frac{2\sum_{j=1}^{n} P(j)_A P(j)_B}{\sum_{j=1}^{n} P(j)_A^2 + \sum_{i=1}^{n} P(j)_B^2} \tag{87}$$

Richards[70] proposed a linear index for the computation of grid-based molecular similarity descriptors:

$$L(P)_{AB} = \frac{1}{n}\sum_{j=1}^{n}\left(1 - \frac{\left|P(j)_A - P(j)_B\right|}{\max\left(\left|P(j)_A\right|,\left|P(j)_B\right|\right)}\right) \tag{88}$$

where $\max(|P(j)_A|,|P(j)_B|) = P_{max}$ equals the larger value of the property P (P_A or P_B) at the grid point j where the similarity is being calculated. A related formula was defined by Good[71] who introduced an exponential similarity index:

$$E(P)_{AB} = \frac{1}{n}\sum_{j=1}^{n}\exp\left(-\frac{\left|P(j)_A - P(j)_B\right|}{\max\left(\left|P(j)_A\right|,\left|P(j)_B\right|\right)}\right) \tag{89}$$

So and Karplus[72,73] proposed to encode the 3D chemical structure in molecular similarity matrices computed with the Carbó index. The electrostatic and steric molecular fields were used to generate the similarity matrices, and a genetic neural network was employed to select the relevant descriptors.

15. MIMIC

The MIMIC approach is a molecular-field similarity algorithm for aligning molecules by matching their steric and electrostatic fields.[74] After the alignment of the molecules, two methods can be used to obtain information on feature-based and field-based pharmacophoric patterns. Feature-based pharmacophoric patterns correspond to traditional pharmacophore models and are computed from the contribution of individual atoms to the total similarity. Field-based pharmacophoric patterns are generated by computing the percent contribution to the similarity of individual points in a regular lattice surrounding the molecules; this algorithm identifies steric and electrostatic field-based patterns of similarity onto a 3D grid, van der Waals surface, or solvent-accessible surface. The identification of such patterns is useful in detecting important structural features and pharmacophoric-field patterns that contribute to the bioactivity of the molecules.

MIMIC uses a molecular steric volume (MSV) field that describes the size and shape of a molecule, and a molecular electrostatic potential (MEP) field representing the electrostatic potential of a molecule. A brief description of the most important characteristics of this molecular-field similarity algorithm is provided below. At a point $r = r(x,y,z)$ the MSV field of molecule A is:

$$F_A^{\mathrm{MSV}}(r) = \sum_{i \in A} f_i^{\mathrm{MSV}}(r) \tag{90}$$

where f_i^{MSV} represents the steric volume of atom i and is represented by a Gaussian function located at $R_i = R_i(x_i,y_i,z_i)$, the center of atom i:

$$f_i^{\mathrm{MSV}}(r) = \alpha_i \exp(-\beta_i |r - R_i|^2) \tag{91}$$

where the parameters α_i and β_i are optimized for each atom type. The MEP field of molecule A at a point r is:

$$F_A^{\mathrm{MEP}}(r) = \sum_{i \in A} \frac{q_i}{|r - R_i|} \tag{92}$$

where q_i is the partial charge on atom i located at R_i. A three-Gaussian expansion of the $1/r$ term is used to avoid discontinuity of the function at the nuclei.

For any pair of molecules A and B a similarity measure with respect to the field MF, Z_{AB}^{MF}, is computed with the formula:

$$Z_{AB}^{\mathrm{MF}} = \int F_A^{\mathrm{MF}}(r) F_B^{\mathrm{MF}}(r) \mathrm{d}r \tag{93}$$

A molecular similarity index (MSI) is computed for each pair of molecules A and B using a Cosine coefficient formula:

$$S_{AB}^{\mathrm{MF}} = \frac{Z_{AB}^{\mathrm{MF}}}{\left[Z_{AA}^{\mathrm{MF}} Z_{BB}^{\mathrm{MF}}\right]^{1/2}} \tag{94}$$

The MSI index takes values in the range [0, 1] for the steric field MSV and [–1, 1] for the electrostatic field MEP. The pairwise similarity index for molecules A and B is obtained as a weighted sum of the corresponding MSI values for each field:

$$S_{AB} = \lambda S_{AB}^{\mathrm{MSV}} + (1-\lambda) S_{AB}^{\mathrm{MEP}} \tag{95}$$

where $\lambda = 0.66$, that corresponds to a 2:1 weighting of the steric field compared to the electrostatic field. The determination of the best molecular-field alignment of two molecules *A* and *B* requires an exploration of the similarity space of the two molecules. The alignment with the highest similarity for molecules *A* and *B* represents a pairwise alignment solution. However, it was found that the optimum alignment for each pair of molecules will not necessarily produce the optimal multi-molecule alignment, leading to inconsistencies when three or more molecules are considered simultaneously. In an experiment involving three non-nucleoside HIV-1 reverse transcriptase inhibitors it was found that the optimum simultaneous three-molecule alignment reproduces the experimental inhibitor alignment. The similarity determined as a simultaneous matching of several molecules is highly important for the development of consistent relative binding orientations of sets of bioactive molecules. The molecular-field similarity of the molecule set $\mathbf{M} = \{A, B, C, \ldots\}$ is defined as the average of the pairwise similarities:

$$S_{\mathbf{M}} = \left[\frac{m^2 - m}{2} \right]^{-1} \sum_{I<J} S_{IJ} \tag{96}$$

where *m* is the number of molecules in set **M**, and S_{IJ} is the pairwise molecular similarity index for molecules *I* and *J*. The solution for the alignment of the *m* molecules from the set **M** is obtained by maximizing the value of the index $S_{\mathbf{M}}$. This equation represents an efficient way of computing the multi-molecule similarity and to compare the pairwise molecule alignments with the multi-molecule optimized results. A pairwise consistency index (PCI) is proposed to measure the quality of combining the pairwise alignments:

$$\mathrm{PCI} = \frac{S_{\mathbf{M}}^{\mathrm{o}}}{S_{\mathbf{M}}} \tag{97}$$

where $S_{\mathbf{M}}^{\mathrm{o}}$ is the multi-molecule similarity computed with Eq. (96) by superposing the *m* molecules in set **M** according to their optimized pairwise alignments, and $S_{\mathbf{M}}$ is the similarity computed from full multi-molecule optimization according to Eq. (96).

After the similarity-based alignment of a set of molecules is finished, two procedures are applied to extract information regarding the principal structural features of the molecules under study. The first procedure identifies pharmacophoric patterns or pharmacophores from atomic similarities computed with the formula:

$$S_{A_i B}^{\mathrm{MF}} = \frac{\int F_{A_i}^{\mathrm{MF}}(r) F_B^{\mathrm{MF}}(r) \mathrm{d}r}{\left[Z_{AA}^{\mathrm{MF}} Z_{BB}^{\mathrm{MF}} \right]^{1/2}} \tag{98}$$

where A_i represents the contribution of the *i*th atom in molecule *A*. The total similarity index is defined as the sum of atomic similarities:

$$S_{AB}^{\mathrm{MF}} = \sum_i S_{A_i B}^{\mathrm{MF}} \tag{99}$$

Regions of high atomic similarity in pairwise or multi-molecule alignments represent pharmacophoric patterns that can be more general than those from the classical

pharmacophores. Once identified, pharmacophoric patterns can be used to evaluate molecules representing potential novel leads.

The second procedure can be used for the graphical display of molecular similarity fields by computing the percent contribution to the similarity of individual points in a regular lattice surrounding the molecules. The similarity fields can be computed onto a 3D set of grid points into which the aligned molecules are embedded, or onto various molecular surfaces, such as van der Waals or solvent-accessible surfaces. For each grid or molecular surface point r_i a field similarity value is computed with the equation:

$$S_{AB}^{\mathrm{MF}}(r_i) = \frac{F_A^{\mathrm{MF}}(r_i) F_B^{\mathrm{MF}}(r_i)}{\left[\sum_j F_A^{\mathrm{MF}}(r_j) F_A^{\mathrm{MF}}(r_j)\right]^{1/2} \left[\sum_j F_B^{\mathrm{MF}}(r_j) F_B^{\mathrm{MF}}(r_j)\right]^{1/2}} \tag{100}$$

where $F_A^{\mathrm{MF}}(r_i)$ and $F_B^{\mathrm{MF}}(r_i)$ are the molecular field values at the point r_i for molecules *A* and *B*, and the summations are carried out over all the grid or surface points. When the summation is carried out over the entire set of points, the above equation gives:

$$S_{AB}^{\mathrm{MF}} = \sum_i S_{AB}^{\mathrm{MF}}(r_i) \tag{101}$$

The percent of the total field-based similarity at each field point is given by the formula:

$$C_{AB}^{\mathrm{MF}}(r_i) = 100 \frac{S_{AB}^{\mathrm{MF}}(r_i)}{S_{AB}^{\mathrm{MF}}} \tag{102}$$

The visual recognition of steric and electrostatic similarity patterns can be accomplished with the aid of the graphical representation of $C_{AB}^{\mathrm{MF}}(r_i)$ values at each of the points on a surface surrounding the superposed molecules. Equally well, the structural features that determine the biological activity can be deduced from iso-similarity surfaces computed with $C_{AB}^{\mathrm{MF}}(r_i)$ values.

The MIMIC model was applied to a study involving three non-nucleoside HIV-1 reverse transcriptase inhibitors.

16. SUPERMOLECULE SIMILARITY DESCRIPTORS

De Benedetti introduced new size, shape and electrostatic similarity descriptors computed from a reference supermolecule.[75-78] The supermolecule, that is an ensemble of the most active and diverse ligands, contains information related to the ligand-receptor complementarity. The calculation of the supermolecule similarity descriptors involves the following steps:

(1) The QSAR is developed for a set of molecules that interact with the same receptor, having the same mechanism of action.

(2) The geometry of all molecules is optimized with molecular mechanics or quantum mechanics methods.

(3) Atomic partial charges are calculated for all molecules with a quantum mechanics method.

(4) The supermolecule is generated by the superposition of the most active and structurally most different compounds.

(5) A molecule is superimposed on the supermolecule using some pharmacophore hypothesis.

(6) The aligned molecule and supermolecule are placed in a box enclosing both of them and extending 3 Å from the larger coordinates value on each Cartesian axis, and the box is partitioned into a grid with points separated by 0.5 Å; the distance between two grid points defines the grid spacing, GS.

(7) The van der Waals volumes of the molecule (V_{mol}) and supermolecule (V_{sup}) are computed by counting the number of grid points included within the volume of each of them:

$$V_{mol} = Ngm \times GS^3 \quad (103)$$

$$V_{sup} = Ngs \times GS^3 \quad (104)$$

where Ngm is the number of grid points included within the volume of the molecule, and Ngs is the number of grid points included within the volume of the supermolecule.

(8) The intersection volume (V_{in}) is computed by counting the number of grid points falling inside the volumes of both the molecule and supermolecule (Ngc, number of common grid points):

$$V_{in} = Ngc \times GS^3 \quad (105)$$

(9) The outer van der Waals volume (V_{out}) of the molecule with respect to the volume of the supermolecule is:

$$V_{out} = Ngo \times GS^3 \quad (106)$$

where Ngo is the number of grid points that belong only to the molecule.

(10) Using the above volumes, two normalized indices are computed for the molecule:

$$V_{Dnorm} = (V_{in} - V_{out})/V_{sup} \quad (107)$$

$$V_{in}/V_{mol} \quad (108)$$

(11) For each molecule, the indices V_{in}, V_{Dnorm}, and V_{in}/V_{mol} are used as size and shape descriptors in QSAR models.

(12) Steps (5)-(11) are repeated for all molecules from the set.

(13) The molecular electrostatic potential (MEP) is computed at the intersection of a rectilinear grid which has the same extension and spacing as used for the volume calculations. Each grid point $g = g(x,y,z)$ is uniquely identified by is Cartesian coordinates x, y, and z. A probe atom is placed at each grid point and the MEP values are calculated using atom centered point charges:

$$MEP_{g,i} = \sum_{j=1}^{na} \frac{q_g q_j}{d_{g,j}} \quad (109)$$

where $MEP_{g,i}$ is the MEP value in the grid point g for molecule i, q_g is the charge of the probe atom, which is set to 1.0, q_j is the charge of atom j from molecule i, $d_{g,j}$ is the distance between the jth atom and the probe, and na is the number of atoms in the

molecule *i*. In order to avoid singularities the evaluation of MEP is restricted to grid points outside the van der Waals volume of the molecule.

(14) At each grid point *g* the supermolecule MEP ($MEP_{sup,g}$) is calculated as the average MEP values of the molecules that form the supermolecule:

$$MEP_{sup,g} = \frac{1}{nz}\sum_{i=1}^{ns} MEP_{g,i} \tag{110}$$

where *ns* is the number of molecules that form the supermolecule, and *nz* is the number of molecules with non-zero $MEP_{g,i}$ values. Whenever the grid point *g* falls inside the van der Waals volume of one or more molecules from the supermolecule, the summation is averaged only over the remaining molecules with the grid point outside their volumes.

(15) The MEP similarity index ($MEP_{sim,i}$) of molecule *i* with respect to the supermolecule is computed with the Hodgkin similarity index:

$$MEP_{sim,i} = \frac{2\sum_{g=1}^{np} MEP_{sup,g} MEP_{g,i}}{\sum_{g=1}^{np} MEP_{sup,g}^2 + \sum_{g=1}^{np} MEP_{g,i}^2} \tag{111}$$

(16) In analogy with the MEP_{sim} index, a surface MEP index (MEP_{sur}) is computed over a set of points distributed on the van der Waals surface of the supermolecule. A gridded sphere is generated around each atom of each molecule from the supermolecule. Each sphere has a radius equal to the van der Waals radius of the corresponding atom. All points falling at the intersection region of any two spheres is discarded, and the remaining grid points are used to compute the MEP_{sur} index for each molecule.

The descriptors V_{in}, V_{Dnorm}, V_{in}/V_{mol}, MEP_{sim}, and MEP_{sur} were used to model pNMS, the cologarithm of the binding affinity constants measured by displacement of the muscarinic antagonist [^{3}H]N-methylscopolamine which labels both high affinity and low affinity states of the receptor, and pOXO-M, the cologarithm of the binding affinity constants measured by displacement of the muscarinic agonist [^{3}H]oxotremorine-M which labels the high affinity agonist state of the receptor.[78]

17. MLP - MOLECULAR LIPOPHILIC POTENTIAL

A common assumption in drug design is that the ligand-receptor interactions involve the same intermolecular forces as those acting in the partitioning of a solute between water and an immiscible organic phase, usually *n*-octanol. For these reasons, the logarithm of the *n*-octanol/water partition coefficient, log *P*, is a widely used QSAR descriptor for the characterization of the molecular hydrophobicity. However, the characterization of the molecular lipophilicity with a single number is not sufficient in 3D QSAR, when the ligand-receptor interactions are investigated. By analogy with the molecular electrostatic potential, a grid-based definition of several types of molecular lipophilic potentials (MLP) were proposed in the literature.[79,80] At a given point in space,

the MLP value is the result of the intermolecular interactions encoded by the lipophilicity of all molecular fragments. Therefore, the MLP computation of MLP is based on two components, namely a fragmental system of lipophilicity, and a distance function describing the variation of lipophilicity in space. A general equation of the MLP is:

$$\mathrm{MLP}_j = \sum_{i=1}^{N} f_i F(d_{ij}) \tag{112}$$

where j is a grid point, i is the label of the molecular fragment, N is the total number of fragments in the molecule, f_i is the lipophilic constant of fragment i, F is a distance function, and d_{ij} is the distance between fragment i and grid point j. Unlike the MEP which has a physical significance and which is dependent on the electronic density distribution, an MLP is purely a mathematical tool with no physical existence.

The first MLP field was proposed by Audry, who used a hyperbolic distance function:[81]

$$\mathrm{MLP}_j = \sum_{i=1}^{N} \frac{f_i}{1 + d_{ij}} \tag{113}$$

This MLP definition is based on the assumption that log P can be computed from the hydrophobic constants of certain molecular fragments:

$$\log P = \sum_{i=1}^{N} f_i \tag{114}$$

The values for fragment hydrophobicities can be computed with the Hansch[8,82,83] or Rekker[84-88] systems. However, the calculation of the MLP on the van der Waals surface requires hydrophobic constants defined for atoms not for functional groups. Atomic parameters present several advantages, i.e. for different conformations of the same molecule, the MLP has different values for the relevant space regions. Several atomic hydrophobicity parameters systems were proposed, but the most used are those defined by Broto[89] and Ghose.[90-92] According to these systems, log P for a molecule with N atoms can be expressed as:

$$\log P = \sum_{i=1}^{N} a_i \tag{115}$$

where a_i is the lipophilic contribution of a certain atom type. With a hyperbolic distance function and the atomic system of hydrophobic constants, the MLP in the grid point j is:

$$\mathrm{MLP}_j = \sum_{i=1}^{N} \frac{a_i}{1 + d_{ij}} \tag{116}$$

Using different combinations of distance functions and hydrophobic constants, several other MLP systems were proposed by Furet,[93] Fauchère,[94] Testa,[95] and Rozas.[96,97] For example, the MLP system defined by Testa[79,95] uses the atomic increments of Broto and a modification of the exponential function proposed by Fauchère:

$$\mathrm{MLP}_j = \sum_{i=1}^{N} a_i e^{-d_{ij}/2} \tag{117}$$

In this formula, the distance function is limited by the use of a cut-off of 4 Å to avoid the influence of too-distant atoms.

We have to mention that the intermolecular forces encoded into the MLP field can add important information to the molecular fields usually used in CoMFA or other related 3D QSAR models. As presented in a previous section, the CoMFA analysis reduces the description of ligand-receptor complexes to the intermolecular interactions described by the steric and electrostatic fields. These fields are purely enthalpic, while MLP describes the entropy component of the binding free energy associated with solvation/desolvation of the ligand and receptor.

18. HINT - HYDROPATHIC INTERACTIONS

Kellogg introduced a new model for biological interactions, HINT (hydropathic interactions),[98-104] that is based on atom-atom interactions computed with the formula:

$$b_{ij} = a_i S_i a_j S_j f(r_{ij}) \tag{118}$$

where a_i is the hydrophobic atomic constant for atom i, S_i is the solvent-accessible surface area for atom i, a_j is the hydrophobic atomic constant for atom j, S_j is the solvent-accessible surface area for atom j, and $f(r_{ij})$ is a function of the distance between atoms i and j. The distance function usually used is an exponential:

$$f(r_{ij}) = \exp(-r_{ij}) \tag{119}$$

The sum of all ligand-receptor atom-atom interactions defines the total HINT binding score, a descriptor used in 3D QSAR models:

$$B = \sum_{i=1}^{\text{rec}} \sum_{j=1}^{\text{lig}} a_i S_i a_j S_j T_{ij} f(r_{ij}) \tag{120}$$

where atom i belongs to the receptor, atom j belongs to the ligand, T_{ij} is a sign-flip function that discriminates between acid-base interactions, which are favorable, and acid-acid and base-base interactions, which are unfavorable, and r_{ij} is the distance between atoms i and j. The cutoff distance for interactions is 10 Å.

The model proposed by Kellogg can be converted to HINT 3D property maps computed on a grid of points that surround the ligand:[98,104]

$$A_j = \sum_{i=1}^{n} a_i S_i f(r_{ij}) \tag{121}$$

where A_j is the value of the hydropathic field in the grid point j, and the summation goes over all n atoms in the ligand. Together with the MLP fields, HINT 3D maps can be added to the two standard fields from CoMFA. Several experiments show that HINT field represent a significant component of the final CoMFA model whenever hydrophobic interactions are significant in the binding mode, or where membrane transport is encoded in the biological measure.

19. MOLECULAR SURFACE LIPOPHILICITY

Brickmann[105] proposed an MLP approach particularly designed for mapping local hydrophobicity properties on a solvent-accessible surface of a molecule. He used the atomic hydrophobicity parameters proposed by Ghose[90,91] in the following formula:

$$MLP_j = \frac{\sum_{i=1}^{n} a_i F(d_{ij})}{\sum_{i=1}^{n} F(d_{ij})} \tag{122}$$

where d_{ij} is the distance between atom i and the grid point j, a_i is the lipophilic contribution of a certain atom type, and the summation goes over all n atoms in the molecule. The distance function $F(d_{ij})$ is a Fermi-type function centered on each atom:

$$F(d_{ij}) = \left(1 + e^{a(d_{ij} - d_{\text{cutoff}})}\right)^{-1} \tag{123}$$

This atom-centered function has two adjustable parameters, a and d_{cutoff}, which, for simplicity, are taken to be the same for all atoms. The precise values of these parameters are somewhat arbitrary, but a value of 1.5 was suggested for a, and 4.0 Å for d_{cutoff}. A modified Fermi function was proposed:

$$F(d_{ij}) = \frac{1 + \exp(-a \cdot d_{\text{cutoff}})}{1 + \exp[a(d_{ij} - d_{\text{cutoff}})]} \tag{124}$$

This approach was further elaborated to give a molecular free energy surface density (MOLFESD) approach for a quantitative treatment of hydrophobicity.[106,107]

20. HMLP - HEURISTIC MOLECULAR LIPOPHILICITY POTENTIAL

Du, Arteca, and Mezey proposed the heuristic molecular lipophilicity potential (HMLP),[108] defined on the molecular surfaces from the electrostatic potential and requiring no empirical indices of atomic lipophilicity. This non-empirical, unified lipophilicity and hydrophilicity potential model is computed with the formula:

$$L_i(r_k) = V(r_k) \sum_{\substack{j=1 \\ j \neq i}}^{n} M_j(r_k, R_j, b_j) \tag{125}$$

where $L_i(r_k)$ is the lipophilic potential of atom i computed on the point r_k situated on its exposed surface, $V(r_k)$ is the MEP at point r_k, and the summation is over all atoms in the molecule except atom i. In the screening function $M_j(r_k,R_j,b_j)$ R_j is the nuclear position of atom j, and b_j is the atomic surface-MEP descriptor of atom j:

$$b_j = \sum_{q \in S_j} V(r_q) \Delta s_q \tag{126}$$

where Δs_q is the area element on atom j, and the summation goes over all exposed surface elements of atom j. The atom-based screening function $M_j(r_k,R_j,b_j)$ can be computed with one of the three formulas:

$$M_j(r_k,R_j,b_j) = \frac{r_0^\gamma}{b_0}\frac{b_j}{\left\|R_j - r_k\right\|^\gamma} = \zeta\frac{b_j}{\left\|R_j - r_k\right\|^\gamma} \tag{127}$$

$$M_j(r_k,R_j,b_j) = \frac{b_j}{b_0}\exp\left(-\frac{\left\|R_j - r_k\right\|}{d_0}\right) \tag{128}$$

$$\begin{aligned} M_j(r_k,R_j,b_j) &= \frac{b_j}{\left|b_j\right|}\exp\left(-\frac{b_0}{\lambda_0}\frac{\left\|R_j - r_k\right\|}{\left|b_j\right|}\right) \\ &= \mathrm{sign}(b_j)\exp\left(-\xi\frac{\left\|R_j - r_k\right\|}{\left|b_j\right|}\right) \end{aligned} \tag{129}$$

In the above three functions (r_0, b_0, γ), (b_0, d_0) and (b_0, λ_0) are parameters. The parameters ζ, γ, ξ, and d_0 have to be optimized in order to obtain good lipophilic and hydrophilic descriptors for QSAR.

The atomic lipophilic of atom i is defined with the formula:

$$l_i = \sum_{k\in S_i} L_i(r_k)\Delta s_k \tag{130}$$

where the summation is over the exposed surface area S_i of atom i, and Δs_k is the area element associated with the point r_k situated on the exposed surface of atom i. If $l_i > 0$ then atom i is lipophilic, while if $l_i < 0$ then atom i is hydrophilic. The molecular lipophilic index L_M and the molecular hydrophilic index H_M are the sum of the corresponding values for all lipophilic and hydrophilic atoms, respectively:

$$\mathrm{L_M} = \sum_{l_i>0} l_i \tag{131}$$

$$\mathrm{H_M} = \sum_{l_i<0} l_i \tag{132}$$

These two HMLP descriptors were used to model log P values of alkanes, alcohols, amines, and carboxylic acids.

21. CONCLUSIONS

In this review we have presented two important directions of development of 3D QSAR models. The first direction is represented by the new structural descriptors computed from the three-dimensional molecular structure that can be used in a Hansch-like multilinear regression equation or other statistical model. The second direction

contains the methods that compute grid descriptors to generate a virtual receptor, and usually use molecular alignment of atoms, pharmacophores, volume, or fields. All these techniques are of essential value whenever the three-dimensional structures of biological receptors and their complexes with the ligands are not known. Important steps in recent developments in 3D QSAR are molecular fields, PLS, molecular similarity indices, genetic and evolutionary algorithms for feature selection.

REFERENCES

(1) Kopp, H. Ueber den zusammenhang zwischen der chemischen constitution und einigen physikalischen eigen-schoften bei flüssigen verbindungen. *Ann. Chem. Pharm.* **1844**, *50*, 71-144.

(2) Crum-Brown, A.; Frazer, T. On the Connection between Chemical Constitution and Physiological Action. Part 1. On the Physiological Action of the Ammonium Bases, Derived from Strychia, Brucia, Thebaia, Codeia, Morphia and Nicotia. *Trans. Royal Soc. Endinburgh* **1868-1869**, *25*, 257-274.

(3) Meyer, H. Zur Theorie der Alkoholnarkose, welche Eigenschaft de Anästhetica bedingt ihre narkotische Wirkung. *Arch. Exp. Pathol. Pharmakol*. **1899**, *42*, 109-118.

(4) Overton, C.E. Studien über die Narkose. Zugleich ein Beitrag zur Allgemeine Pharmakologie; Gustav Fisher Verlag: Jena, Germany, 1901.

(5) Hansch, C.; Maloney, P.P.; Fujita, T.; Muir, R.M. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* **1962**, *194*, 178-180.

(6) Hansch, C.; Fujita, T. ρ-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc*. **1964**, *86*, 1616-1626.

(7) Hansch, C. A Quantitative Approach to Biochemical Structure-Activity Relationships. *Acc. Chem. Res*. **1969**, *2*, 232-239.

(8) Fujita, T.; Iwasa, J.; Hansch, C. A New Substituent Constant, π, Derived from Partition Coefficients. *J. Am. Chem. Soc*. **1964**, *86*, 5175-5180.

(9) Kubinyi, H. QSAR and 3D QSAR in Drug Design. Part 1: Methodology. *Drug Discovery Today* **1997**, *2*, 457-467.

(10) Kubinyi, H. QSAR and 3D QSAR in Drug Design. Part 2: Applications and Problems. *Drug Discovery Today* **1997**, *2*, 538-546.

(11) CODESSA 2.13, Semichem, 7204 Mullen, Shawnee, KS 66216, U.S.A., E-mail andy@semichem.com, www http://www.semichem.com.

(12) Randi⇲, M. On Characterization of Three-Dimensional Structures. *Int. J. Quantum Chem.: Quantum Biol. Symp*. **1988**, *15*, 201-208.

(13) Bogdanov, B.; Nikoli⇲, S.; Trinajsti⇲, N. On the Three-Dimensional Wiener Number. *J. Math. Chem*. **1989**, *3*, 299-309.

(14) Randi⇲, M.; Jerman-Blazi⇲, B.; Trinajsti⇲, N. Development of 3-Dimensional Molecular Descriptors. *Comput. Chem*. **1990**, *14*, 237-246.

(15) Nikoli, S.; Trinajsti, N.; Mihali, Z.; Carter, S. On the Geometric-Distance Matrix and the Corresponding Structural Invariants of Molecular Systems. *Chem. Phys. Lett*. **1991**, *179*, 21-28.

(16) Diudea, M.V.; Horvath, D.; Graovac, A. Molecular Topology. 15. 3D Distance Matrices and Related Topological Indices. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 129-135.

(17) Randi, M.; Razinger, M. Molecular Topographic Indices. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 140-147.

(18) Randi, M. Molecular Profiles. Novel Geometry-Dependent Molecular Descriptors. *New J. Chem*. **1995**, *19*, 781-791.

(19) Balaban, A.T.; Balaban, T.-S. New Vertex Invariants and Topological Indices of Chemical Graphs Based on Information on Distances. *J. Math. Chem*. **1991**, *8*, 383-397.

(20) Ivanciuc, O.; Balaban, A.T. Design of Topological Indices. Part 21. Molecular Graph Operators for the Computation of Geometric Structural Descriptors. *Rev. Roum. Chim*. **2000**, *45*, 000-000.

(21) HyperChem 5.1, Hypercube, Inc., Florida Science and Technology Park, 1115 N.W. 4th Street Gainesville, Florida 32601, U.S.A., Voice: 352-371-7744, Fax: 352-371-3662, Email info@hyper.com, www http://www.hyper.com.

(22) O mia owski, K.; Halkiewicz, J.; Kaliszan, R. Quantum Chemical Parameters in Correlation Analysis of Gas-Liquid Chromatographic Retention Indices of Amines. II. Topological Electronic Index. *J. Chromatogr*. **1986**, *361*, 63-69.

(23) Stanton, D.T.; Jurs, P.C. Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Relationship Studies. *Anal. Chem*. **1990**, *62*, 2323-2329.

(24) Stuper, A.J.; Jurs, P.C. ADAPT: A Computer System for Automated Data Analysis Using Pattern Recognition Techniques. *J. Chem. Inf. Comput. Sci*. 1976, 16, 99-105.

(25) ADAPT, P.C. Jurs, 152 Davey Lab, Chemistry Department, Penn State University, University Park, PA 16802 U.S.A., Tel: 814-865-3739, E-mail pcj@psu.edu, www http://zeus.chem.psu.edu/ADAPT.html.

(26) Murugan, R.; Grendze, M.P.; Toomey, J.E., Jr.; Katritzky, A.R.; Karelson, M.; Lobanov, V.S.; Rachwal, P. Predicting Physical Properties from Molecular Structure *CHEMTECH* **1994**, *24*, 17-23.

(27) Katritzky, A.R.; Lobanov, V.S.; Karelson, M. QSPR: The Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure. *Chem. Soc. Rev.* **1995**, 279-287.

(28) Karelson, M.; Lobanov, V.S.; Katritzky, A.R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027-1043.

(29) Katritzky, A.R.; Lobanov, V.S.; Karelson, M.; Murugan, R.; Grendze, M.P.; Toomey Jr., J.E. Comprehensive Descriptors for Structural and Statistical Analysis. 1. Correlations Between Structure and Physical Properties of Substituted Pyridines. *Rev. Roum. Chim*. **1996**, *41*, 851-867.

(30) Katritzky, A.R.; Mu, L.; Karelson, M. A QSPR Study of the Solubility of Gases and Vapors in Water. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1162-1168.

(31) Katritzky, A.R.; Wang, Y.; Sild, S.; Tamm, T.; Karelson, M. QSPR Studies on Vapor Pressure, Aqueous Solubility, and the Prediction of Water-Air Partition Coefficients. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 720-725.

(32) Ivanciuc, O.; Ivanciuc, T.; Balaban, A.T. Quantitative Structure-Property Relationship Study of Normal Boiling Points for Halogen-/ Oxygen-/ Sulfur-Containing Organic Compounds Using the CODESSA Program. *Tetrahedron* **1998**, *54*, 9129-9142.

(33) Ivanciuc, O.; Ivanciuc, T.; Filip, P.A.; Cabrol-Bass, D. Estimation of the Liquid Viscosity of Organic Compounds with a Quantitative Structure-Property Model. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 515-524.

(34) Kamlet, M.J.; Doherty, R.M.; Abboud, J.-L.M.; Abraham, M.H.; Taft, R.W. Solubility: A New Look. *CHEMTECH* **1986**, *16*, 566-576.

(35) Kamlet, M.J.; Doherty, R.M.; Abraham, M.H.; Marcus, Y.; Taft, R.W. Linear Solvation Energy Relationships. 46. An Improved Equation for Correlation and Prediction of Octanol/Water Partition Coefficients of Organic Nonelectrolytes (Including Strong Hydrogen Bond Donor Solutes). *J. Phys. Chem*. **1988**, *92*, 5244-5255.

(36) Famini, G.R.; Wilson, L.Y. Using Theoretical Descriptors in Linear Solvation Energy Relationships. In: *Quantitative Treatments of Solute/Solvent Interactions*, *Theoretical and Computational Chemistry*, Vol. 1; Politzer, P., Murray, J.S., Eds.; Elsevier: Amsterdam, 1994, pp. 213-241.

(37) Cramer, C.J.; Famini, G.R.; Lowrey, A.H. Use of Calculated Quantum Chemical Properties as Surrogates for Solvatochromic Parameters in Structure-Activity Relationships. *Acc. Chem. Res*. **1993**, *26*, 599-605.

(38) Donovan, W.H.; Famini, G.R. Using Theoretical Descriptions in Structure Activity Relationships: Retention Indices of Sulfur Vesicants and Related Compounds. *J. Chem. Soc. Perkin Trans. 2* **1996**, 83-89.

(39) Esaki, T. Quantitative Drug Design Studies. II. Development and Application of New Electronic Substituent Parameters. *J. Pharm. Dyn*. **1980**, *3*, 562-576.

(40) Pires, J.M.; Floriano, W.B.; Gaudio, A.C. Extension of the Frontier Reactivity Indices to Groups of Atoms and Application to Quantitative Structure-Activity Relationship Studies. *J. Mol. Struct. (Theochem)* **1997**, *389*, 159-167.

(41) Murray, J.S.; Brinck, T.; Grice, M.E.; Politzer, P. Correlations between Molecular Electrostatic Potentials and Some Experimentally-Based Indices of Reactivity. *J. Mol. Struct. (Theochem)* **1992**, *256*, 29-45.

(42) Murray, J.S.; Brinck, T.; Politzer, P. Partition Coefficients of Nitroaromatics Expressed in Terms of Their Molecular Surface Areas and Electrostatic Potentials. *J. Phys. Chem*. **1993**, *97*, 13807-13809.

(43) Brinck, T.; Murray, J.S.; Politzer, P. Octanol/Water Partition Coefficients Expressed in Terms of Solute Molecular Surface Areas and Electrostatic Potentials. *J. Org. Chem*. **1993**, *58*, 7070-7073.

(44) Murray, J.S.; Lane, P.; Brinck, T.; Politzer, P. Relationships between Computed Molecular Properties and Solute-Solvent Interactions in Supercritical Solutions. *J. Phys. Chem*. **1993**, *97*, 5144-5148.

(45) Murray, J.S.; Lane, P.; Brinck, T.; Paulsen, K.; Grice, M.E.; Politzer, P. Relationships of Critical Constants and Boiling Points to Computed Molecular Surface Properties. *J. Phys. Chem*. **1993**, *97*, 9369-9373.

(46) Murray, J.S.; Brinck, T.; Lane, P.; Paulsen, K.; Politzer, P. Statistically-Based Interaction Indices Derived from Molecular Surface Electrostatic Potentials: A General Interaction Properties Function (GIPF). *J. Mol. Struct. (Theochem)* **1994**, *307*, 55-64

(47) Cramer, R.D.; Patterson, D.E.; Bunce, J.D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc*. **1988**, *110*, 5959-5967.

(48) Clark, M.; Cramer III, R.D.; Jones, D.M.; Patterson, D.E.; Simeroth, P.E.; Comparative Molecular Field Analysis (CoMFA). 2. Toward Its Use with 3D-Structural Databases. *Tetrahedron Comput. Methodol*. **1990**, *3*, 47-59.

(49) Kubinyi, H. Comparative Molecular Field Analysis (CoMFA). In: *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F., Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998.

(50) Robinson, D.D.; Winn, P.J.; Lyne, P.D.; Richards, W.G. Self-Organizing Molecular Field Analysis: A Tool for Structure-Activity Studies. *J. Med. Chem*. **1999**, *42*, 573-583.

(51) Hahn, M. Receptor Surface Models. 1. Definition and Construction. *J. Med. Chem*. **1995**, *38*, 2080-2090.

(52) Hahn, M.; Rogers, D. Receptor Surface Models. 2. Application to Quantitative Structure-Activity Relationships Studies. *J. Med. Chem*. **1995**, *38*, 2091-2102.

(53) Cerius^2 3.0 QSAR+, Molecular Simulations Inc., 9685 Scranton Road, San Diego, CA 92121-3752, Fax: 619/458-0136, 1997.

(54) Rogers, D.; Hopfinger, A.J. Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854-866.

(55) Rhyu, K.-B.; Patel, H.C.; Hopfinger, A.J. A 3D-QSAR Study of Anticoccidal Triazines Using Molecular Shape Analysis. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 771-778.

(56) Klein, C.D.P.; Hopfinger, A.J. Pharmacological Activity and Membrane Interactions of Antiarrhytmics: 4D-QSAR/QSPR Analysis. *Pharm. Res*. **1998**, *15*, 303-311.

(57) Albuquerque, M.G.; Hopfinger, A.J.; Barreiro, E.J.; de Alencastro, R.B. Four-Dimensional Quantitative Structure-Activity Relationship Analysis of a Series of Interphenylene 7-Oxabicycloheptane Oxazole Thromboxane A_2 Receptor Antagonists. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 925-938.

(58) Ivanciuc, O.; Ivanciuc, T.; Filip, P.A.; Cabrol-Bass, D. Comparative Receptor Surface Analysis (CoRSA) of Calcium Channel Agonist Activity of 1,4-Dihydro-2,6-dimethyl-3-nitro-4-pyridyl-5-pyridinecarboxylates. *Rev. Roum. Chim*. **2000**, *45*, 000-000.

(59) Alsberg, B. Molecular Reference (MOLREF): A New Tool in Quantitative Structure-Activity Relationships (QSAR). *Chemom. Intell. Lab. Syst*. 1990, 8, 173-181.
(60) Johnson, M.A. A Review and Examination of the Mathematical Spaces Underlying Molecular Similarity Analysis. *J. Math. Chem*. **1989**, *3*, 117-145.
(61) Johnson, M.A.; Maggiora, G.M., Eds. *Concepts and Applications of Molecular Similarity*; John Wiley & Sons: New York, 1990.
(62) Turner, D.B.; Willett, P.; Ferguson, A.M.; Heritage, T.W. Similarity Searching in Files of Three-Dimensional Structures: Evaluation of Similarity Coefficients and Standardisation Methods for Field-Based Similarity Searching. *SAR QSAR Environ. Res*. **1995**, *3*, 101-130.
(63) Skvortsova, M.I.; Baskin, I.I.; Stankevich, I.V.; Palyulin, V.A.; Zefirov, N.S. Molecular Similarity. 1. Analytical Description of the Set of Graph Similarity Measures. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 785-790.
(64) Willett, P.; Barnard, J.M.; Downs, G.M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 983-996.
(65) Carbó, R.; Leyda, L.; Arnau, M. How Similar Is a Molecule to Another? An Electron Density Measure of the Similarity between Two Molecular Structures. *Int. J. Quantum Chem*. **1980**, *17*, 1185-1189.
(66) Carbó, R.; Domingo, L. LCAO-MO Similarity Measures and Taxonomy. *Int. J. Quantum Chem*. **1987**, *32*, 517-545.
(67) Carbó, R.; Calabuig, B. Molecular Similarity and Quantum Chemistry. In: *Concepts and Applications of Molecular Similarity*; Johnson, M.A., Maggiora, G.M., Eds.; John Wiley & Sons: New York, 1990, pp. 147-171.
(68) Carbó-Dorca, R.; Besalú, E.; Amat, L.; Fradera, X. Quantum Molecular Similarity Measures: Concepts, Definitions, and Applications to Quantitative Structure-Property Relationships. In: *Advances in Molecular Similarity*; Carbó-Dorca, R., Mezey, P.G., Eds.; JAI Press: Greenwich, CT, 1996, Vol 1, pp. 1-42.
(69) Hodgkin, E.E.; Richards, W.G. Molecular Similarity Based on Electrostatic Potential and Electric Field. *Int. J. Quantum Chem.: Quantum Biol. Symp*. **1987**, *14*, 105-110.
(70) Reynolds, C.A.; Burt, C.; Richards, W.G. A Linear Molecular Similarity Index. *Quant. Struct.-Act. Relat*. **1992**, 11, 34-35.
(71) Good, A.C. The Calculation of Molecular Similarity: Alternative Formulas, Data Manipulation and Graphical Display. *J. Mol. Graphics* **1992**, *10*, 144-151.
(72) So, S.-S.; Karplus, M. Three-Dimensional Quantitative Structure-Activity Relationships from Molecular Similarity Matrices and Genetic Neural Networks. 1. Method and Validations. *J. Med. Chem*. **1997**, *40*, 4347-4359.
(73) So, S.-S.; Karplus, M. Three-Dimensional Quantitative Structure-Activity Relationships from Molecular Similarity Matrices and Genetic Neural Networks. 2. Applications. *J. Med. Chem*. **1997**, *40*, 4360-4371.
(74) Mestres, J.; Rohrer, D.C.; Maggiora, G.M. A Molecular-Field-Based Similarity Study of Non-Nucleoside HIV-1 Reverse Transcriptase Inhibitors. *J. Comput.-Aided Mol. Design* **1999**, *13*, 79-93.

(75) De Benedetti, P.G.; Cocchi, M.; Menziani, M.C.; Fanelli, F. Theoretical Quantitative Size and Shape Activity and Selectivity Analyses of 5-HT_{1A} Serotonin and α_1-Adrenergic Receptor Ligands. *J. Mol. Struct. (Theochem)* **1994**, *305*, 101-110.

(76) Cocchi, M.; Menziani, M.C.; Fanelli, F.; De Benedetti, P.G. Theoretical Quantitative Structure-Activity Relationship Analysis of Congeneric and Non-Congeneric α_1-Adrenoceptor Antagonists: A Chemometric Study. *J. Mol. Struct. (Theochem)* **1995**, *331*, 79-93.

(77) De Benedetti, P.G.; Menziani, M.C.; Cocchi, M.; Fanelli, F. Prototropic Molecular Forms and Theoretical Descriptors in QSAR Analysis. *J. Mol. Struct. (Theochem)* **1995**, *333*, 1-17.

(78) Cocchi, M.; De Benedetti, P.G. Use of the Supermolecule Approach to Derive Molecular Similarity Descriptors for QSAR Analysis. *J. Mol. Model*. **1998**, *4*, 113-131.

(79) Carrupt, P.-A.; Gaillard, P.; Billois, F.; Weber, P.; Testa, B.; Meyer, C.; Pérez, S. The Molecular Lipophilicity Potential (MLP): A New Tool for log *P* Calculations and Docking, and in Comparative Molecular Field Analysis (CoMFA). In: *Lipophilicity in Drug Action and Toxicology*; Pliška, V., Testa, B., van de Waterbeemd, H., Eds.; VCH: Weinheim, 1996, pp. 195-217.

(80) Folkers, G.; Merz, A. Hydrophobic Fields in Quantitative Structure-Activity Relationships. In: *Lipophilicity in Drug Action and Toxicology*; Pliška, V., Testa, B., van de Waterbeemd, H., Eds.; VCH: Weinheim, 1996, pp. 219-232.

(81) Audry, E.; Dubost, J.P.; Colleter, J.C.; Dallet, P. A New Approach to Structure-Activity Relations: the "Molecular Lipophilicity Potential". *Eur. J. Med. Chem*. **1986**, *21*, 71-72.

(82) Hansch, C.; Leo, A.J. *Substituent Constants for Correlation Analysis in Chemistry*; Wiley: New York, 1979.

(83) Leo, A.J. Calculating log P_{oct} from Structure. *Chem. Rev*. **1993**, *93*, 1281-1306.

(84) Nys, G.G.; Rekker, R.F. Statistical Analysis of a Series of Partition Coefficients with Special Reference to the Predictability of Folding of Drug Molecules. Introduction of Hydrophobic Fragmental Constants (f-Values). *Chim. Ther*. **1973**, *8*, 521-535.

(85) Nys, G.G.; Rekker, R.F. The Concept of Hydrophobic Fragmental Constants (f-Values). II. Extension of Its Applicability to the Calculation of Lipophilicities of Aromatic and Hetero-aromatic Structures. *Chim. Ther*. **1974**, *9*, 361-375.

(86) Rekker, R.F.; de Kort, H.M. The Hydrophobic Fragmental Constant; An Extension to a 1000 Data Point Set. *Eur. J. Med. Chem*. **1979**, *14*, 479-488.

(87) Rekker, R.F.; Mannhold, R. *Calculation of Drug Lipophilicity*; VCH: New York, 1992.

(88) Mannhold, R.; Rekker, R.F.; Dross, K.; Bijloo, G.; de Vries, G. The Lipophilic Behaviour of Organic Compounds: 1. An Updating of the Hydrophobic Fragmental Constant Approach. *Quant. Struct.-Act. Relat*. **1998**, *17*, 517-536.

(89) Broto, P.; Moreau, G.; Vandycke, C. Molecular Structures: Perception, Autocorrelation Descriptor and SAR Studies. System of Atomic Contributions for the

Calculation of the *n*-Octanol/Water Partition Coefficients. *Eur. J. Med. Chem*. **1984**, *19*, 71-78.
(90) Ghose, A.R.; Crippen, G.M. Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure-Activity Relationships I. Partition Coefficients as a Measure of Hydrophobicity. *J. Comput. Chem*. **1986**, *7*, 565-577.
(91) Viswanadhan, V.N.; Ghose, A.K.; Revankar, G.R.; Robins, R.K. Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally Occurring Nucleoside Antibiotics. *J. Chem. Inf. Comput. Sci*. **1989**, *29*, 163-172.
(92) Ghose, A.K.; Viswanadhan, V.N.; Wendoloski, J.J. Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *J. Phys. Chem. A* **1998**, *102*, 3762-3772.
(93) Furet, P.; Sele, A.; Cohen, N.C. 3D Molecular Lipophilicity Potential Profiles: A New Tool in Molecular Modeling. *J. Mol. Graphics* **1988**, *6*, 182-189.
(94) Fauchère, J.-L.; Quarendon, P.; Kaetterer, L. Estimating and Representing Hydrophobicity Potential. *J. Mol. Graphics* **1988**, *6*, 203-206.
(95) Gaillard, P.; Carrupt, P.-A.; Testa, B.; Boudon, A. Molecular Lipophilicity Potential, a Tool in 3D QSAR: Method and Applications. *J. Comput.-Aided Mol. Design* **1994**, *8*, 83-96.
(96) Rozas, I.; Du, Q.; Arteca, G.A. Interrelation between Electrostatic and Lipophilicity Potential on Molecular Surfaces. *J. Mol. Graphics* **1995**, *13*, 98-108.
(97) Rozas, I.; Martin, M. Molecular Lipophilic Potential on van der Waals Surfaces as a Tool in the Study of 4-Alkylpyrazoles. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 872-878.
(98) Kellogg, G.E.; Semus, S.F.; Abraham, D.J. HINT: A New Method of Empirical Hydrophobic Field Calculation for CoMFA. *J. Comput.-Aided Mol. Design* **1991**, *5*, 545-552.
(99) Kellogg, G.E.; Abraham, D.J. KEY, LOCK, and LOCKSMITH: Complementary Hydropathic Map Predictions of Drug Structure from a Known Receptor-Receptor Structure from Known Drugs. *J. Mol. Graphics* **1992**, *10*, 212-217.
(100) Meng, E.C.; Kuntz, I.D.; Abraham, D.J.; Kellogg, G.E. Evaluating Docked Complexes with the HINT Exponential Function and Empirical Atomic Hydrophobicities. *J. Comput.-Aided Mol. Design* **1994**, *8*, 299-306.
(101) Wei, D.T.; Meadows, J.C.; Kellogg, G.E. Effects of Entropy on QSAR Equations for HIV-1 Protease: 1. Using Hydropathic Binding Descriptors. 2. Unrestrained Complex Structure Optimizations. *Med. Chem. Res*. **1997**, *7*, 259-270.
(102) Kellogg, G.E. Finding Optimum Field Models for 3D QSAR. *Med. Chem. Res*. **1997**, *7*, 417-427.
(103) Kellogg, G.E.; Abraham, D.J. Development of Empirical Molecular Interaction Models that Incorporate Hydrophobicity and Hydropathy. The HINT Paradigm. *Analusis*, **1999**, *27*, 19-23.

(104) Waller, C.L.; Kellogg, G.E. Adding Chemical Information to CoMFA Models with Alternative 3D QSAR Fields. *Network Science* **1996**,
http://www.netsci.org/Science/Compchem/feature10.html.
(105) Heiden, W.; Moeckel, G.; Brickmann, J. A New Approach to Analysis and Display of Local Lipophilicity/Hydrophilicity Mapped on Molecular Surfaces. *J. Comput.-Aided Mol. Design* **1993**, *7*, 503-514.
(106) Pixner, P.; Heiden, W.; Merx, H.; Moeckel, G.; Möller, A.; Brickmann, J. Empirical Method for the Quantification and Localization of Molecular Hydrophobibity. *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 1309-1319.
(107) Brickmann, J.; Jäger, R. The Molecular Free Energy Surface Density (MOLFESD) Approach for a Quantitative Treatment of Hydrophobicity. *Analusis* **1999**, *27*, 15-18.
(108) Du, Q.; Arteca, G.A.; Mezey, P.G. Heuristic Lipophilicity Potential for Computer-Aided Rational Drug Design. *J. Comput.-Aided Mol. Design* **1997**, *11*, 503-515.