The Receptor-Dependent QSAR Paradigm: An Overview of the Current State of the Art

Osvaldo A. Santos-Filho[1,*], Anton J. Hopfinger[2a,2b], Artem Cherkasov[3] and Ricardo Bicca de Alencastro[1]

[1]*Universidade Federal do Rio de Janeiro, Instituto de Química, CT – Bloco A –Lab609, Cidade Universitária, CEP 21949-909, Rio de Janeiro, RJ, Brazil;* [2a]*College of Pharmacy, University of New Mexico, Albuquerque, NM, 87131, USA;* [2b]*The Chem21 Group, Inc., 1780 Wilson Drive, Lake Forest, IL, 60045, USA;* [3]*Prostate Centre at Vancouver General Hospital, Faculty of Medicine, University of British Columbia, 2660 Oak Street, Vancouver, British Columbia, V6H 3Z6, Canada*

Abstract: The original quantitative structure-activity relationship (QSAR) formulation was proposed by Hansch and Fujita in the 1960's and, since then QSAR analysis has evolved as a mature science, due mainly to the advances that occurred in the past two decades in the fields of molecular modeling, data analysis algorithms, chemoinformatics, and the application of graph theory in chemistry. Moreover, it is also worthy of note the exponential progress that have occurred in software and hardware development. In this context, a myriad of QSAR methods exist; from the considered "classical" approaches (known as two-dimensional (2D) QSAR), to three-dimensional (3D) and multidimensional (nD) QSAR ones.

A distinct QSAR approach has been recently proposed, the receptor-dependent-QSAR, where explicit information regarding the receptor structure (usually a protein) is extensively used during modeling process. Indeed, a limited, but growing number of receptor-dependent QSAR methods are reported in the literature. With no intention to be comprehensive, an overview of receptor-dependent QSAR methods will be discussed along with an in-depth examination of their applications in drug design, virtual screen, and ADMET modeling *in silico*.

Key Words: QSAR, 3D-QSAR, receptor-dependent QSAR, structure-based QSAR, drug design.

INTRODUCTION

Computer-assisted drug design (CADD) is the science and art of finding molecules of potential therapeutic value, which relies on computational chemistry methods. The goal is to discover, enhance, or study drugs and related biological molecules with the assistance of modern computational resources. Working as a preliminary step ("pre-synthesis") in rational drug design CADD significantly reduces the cost and time for the development of potential new drug candidates. Moreover, CADD represents a viable alternative in cases where experiments are unfeasible, or too dangerous [1]. In CADD, quantitative structure-activity relationship (QSAR) is a crucial approach for the development of medicines. One would say that nowadays no drug is developed without previous QSAR analyses.

The origin of the QSAR formalism, as we know today, is attributed to the works of Hansch and Fujita [2] and Free and Wilson [3] in 1964. The underlying theory of QSAR is that biological activity is directly related to molecular structure. Therefore, biological activity of congeneric molecular structures are related to specific molecular features (descriptors) by using regression techniques to estimate the relative importance of those features contributing to the biological effect. Resulting QSAR models can then be utilized to help guide chemical synthesis. After the 1960's QSAR methodology has became in a broad subfield of CADD. Thus, several QSAR methodologies have been proposed. Each of them can be characterized by having particular approaches for calculating and selecting the molecular descriptors, and specific statistical algorithms for constructing the resulting models [4-6].

In the broadest sense, QSAR studies can be grouped in two major groups: receptor-independent (RI) and receptor-dependent (RD) QSAR analyses [4]. The first group is regarding the construction of models in the absence of a well defined structure for the molecular target, which can be a nucleic acid, a protein, a receptor, or a cellular membrane model. This group included the "classical" (zero-dimensional), one-dimensional (1D), two-dimensional (2D), three-dimensional (3D), and four-dimensional QSAR approaches [7]. The calculated descriptors are recognizable molecular features, such as atom and molecular counts, molecular weight, sum of atomic properties (0D-QSAR); fragment counts (1D-QSAR); topological descriptors (2D-QSAR); geometrical, atomic coordinates, or energy grid descriptors (3D-QSAR); and the combination of atomic coordinates and sampling of conformations (RI 4D-QSAR) [7].

The second QSAR group, and the focus of this review, is regarding construction of models when the tri-dimensional structure of a binding site is known and can be implemented

*Address correspondence to this author at the Universidade Federal do Rio de Janeiro, Instituto de Química, CT – Bloco A –Lab609, Cidade Universitária, CEP 21949-909, Rio de Janeiro, RJ, Brasil;
E-mails: osvaldo@iq.ufrj.br / o_santosfilho@yahoo.com

1573-4064/09 $55.00+.00

when an initial compound (analog pharmacophore) is bound. RD-QSAR is used to gather binding interaction energies, as descriptors, from the interaction between the analog molecules and the receptor. Since the descriptors are calculated by using 3D structural information of the receptor, this type of QSAR analysis has the potential to be more precise than its receptor-independent counterpart.

Due to the intrinsic dependence of atomic coordinates of both receptor and ligands, RD-QSAR includes multidimensional methods (nD-QSAR), such as 4D-, 5D-, and 6D-QSAR, among others. Without the intention to be comprehensive, in the next sections, some of those methods will be described and discussed in details.

4D-QSAR

As an evolution of his Molecular Shape Analysis [8,9], Hopfinger proposed the 4D-QSAR formalism [10-24], which includes the conformational flexibility and the freedom of alignment by ensemble averaging in the conventional three-dimensional descriptors found in traditional 3D-QSAR methods. Thus, the "fourth dimension" of the method is the dimension of the ensemble sampling.

The 4D-QSAR method was originally developed to investigate structure−activity data sets where the geometry of the receptor was *not* available, i.e. RI 4D-QSAR [10-21]. Recently, the 4D-QSAR paradigm has been extended to explicitly include the geometry of the receptor in building a QSAR model. Thus, the methodology now includes the capacity to do *quantitative* structure-based design [22-24].

The RD 4D-QSAR analysis consists of 12 steps, which are summarized in Table **1**. A detailed description of the method is given in [22]. The main feature of a RD 4D-QSAR analysis is that the resultant pharmacophore sites of the QSAR models generated in the analysis are explicitly dependent upon the combined geometries of the (bound) ligand and the receptor.

Table 1. The Operational Steps in Performing a RD 4D-QSAR Analysis [22]

Step	Description of the Step Operation
1	Receptor pruning and atom charge assignment of the "receptor"
2	Modeling of the data set of compounds
3	Ligand docking
4	Select the trial set of interaction pharmacophore elements, IPEs
5	Constraint of selective receptor and inhibitor atoms
6	Molecular dynamics simulations of each pruned "receptor"−inhibitor complex
7	Alignment in a binding site
8	Grid analysis
9	Trial descriptor pool generation
10	Partial least-squares (PLS) regression analysis
11	Construction of RD 4D-QSAR models/GFA−MLR−GFA
12	"Active" conformation postulation of the ligands

As a case study example, RD 4D-QSAR models were constructed for a set of 39 4-hydroxy-5,6-dihydropyrone analog HIV-1 protease inhibitors [24]. The receptor model used in this QSAR analysis was derived from the HIV-1 protease (PDB ID 1d4s) crystal structure. The bound ligand in the active site of the enzyme, also a 4-hydroxy-5,6-dihydropyrone analogue, was used as the reference ligand for docking the data set compounds. The optimized RD 4D-QSAR models showed to be not only statistically significant (r^2 = 0.86, q^2 = 0.80 for four- and greater-term models) but also possess reasonable predictivity based on test set predictions [24]. The proposed "active" conformations of the docked analogues in the active site of the enzyme are consistent in overall molecular shape with those suggested from crystallographic studies. Moreover, the RD 4D-QSAR models also qualitatively "captured" the existence of specific induced-fit interactions between the enzyme active site and each specific inhibitor. Hydrophobic interactions, steric shape requirements, and hydrogen bonding of the 4-hydroxy-5,6-dihydropyrone analogues with the HIV-1 protease binding site model dominate the RD 4D-QSAR models in a manner, again, consistent with experimental conclusions. From the constructed models, it is possible to infer hypotheses for the development of new lead HIV-1 protease inhibitors [24].

Quasar 4D-, 5D-, AND 6D-QSAR

Biographics Laboratory 3R [25], led by Angelo Vedani, has developed a multidimensional QSAR software system called *quasi-atomistic receptor modeling* (*Quasar* package) [25]. This package includes 4D-, 5D- and 6D-QSAR modeling methods.

Quasi-atomistic receptor model refers to a three-dimensional binding-site surrogate, represented by a surface surrounding a series of ligands (superimposed in their bioactive conformation) at *van der Waals* distance and populated with atomistic properties (hydrogen bond donors and acceptors, hydrogen bond flip-flop particles, salt bridges, neutral and charged hydrophobic particles, void space mapped on it). Vedani's 4D-QSAR approach refers to the possibility to represent each ligand molecule by an ensemble of conformations, orientations or protonation states, similar to Hopfinger's 4D-QSAR approach.

In order to address the induced fit problem, Vedani's group extended his 4D-QSAR approach [26-29] by proposing the 5D-QSAR method [30]. The "fifth dimension" refers to a multiple representation of receptor induced-fit models, constructed from experimental data or random placement of physicochemical properties to aid in the construction of optimal QSAR models. In contrast to other 3D-QSAR techniques, 5D-QSAR approach allows for a receptor surface, individually adapted to each ligand molecule used in the study as well as for H-bond flip-flop particles, mimicking Ser, Thr, Tyr, Cys, His, Asn and Gln residues engaging in differently directed H-bonds with different ligand molecules. A family of receptor surface models is then generated by means of genetic algorithms combined with cross-validation protocols. Whereas, the "sixth dimension" of the *Quasar*

package (6D-QSAR) refers to simultaneous evaluation of different salvation modes [31]. Table **2** shows the *Quasar* operational steps.

Table 2. The Operational Steps in Performing a *Quasar* nD-QSAR Analysis [25-27]

Step	Description of the Step Operation
1	Generation of the receptor surface. A mean van der Waals surface of the receptor (generated about all ligands defining the training set) is adapted to the topology of each ligand molecule. Thus, a local induced fit is mimicked.
2	Generation of an initial family of parent structures. Quasi-atomistic properties are mapped onto equally distributed points of the individual surface.
3	Evolution of a model family. Using a genetic algorithm, the initial family of receptor models is evolved using both cross-over and mutation events.
4	Estimation of ligand relative free energies of solvation during the ligand binding process.
5	Analysis and evaluation of the constructed models.

As a case study example *Quasar* a 6D-QSAR models was constructed for a set of structurally diverse ligands of the estrogen receptor. The statistical results ($q^2 = 0.903$; $p^2 = 0.885$) suggest that the model seems suitable for the identification of features associated with the endocrine-disrupting potential of drugs and chemicals [31].

FREE ENERGY FORCE FIELD (FEFF) 3D-QSAR ANALYSIS

Tokarski and Hopfinger have proposed the free energy force field (FEFF) 3D-QSAR method in which all of the enthalpy and entropy contributions to the ligand-receptor interaction in a solvent medium are taken into account. Moreover, the enthalpy and entropy contributions are treated as independent variables in developing QSAR models for ligand-receptor binding processes.

Energy terms of a force field is used for estimating the thermodynamics of the ligand-receptor interaction, resulting in a "QSAR force field" for the particular ligand-class/ receptor system [32]. FEFF 3D-QSAR method is a true RD QSAR approach using the solved 3D structure of the receptor in the calculation of ligand-receptor interaction values.

The FEFF 3D-QSAR formalism [32] can be summarized as follows. The ligand–receptor interaction can be expressed as

$$L.M + R.M \overset{K}{\rightleftharpoons} (LR).M \quad (1)$$

(Unbound state) (Bound state)

where L is the ligand, R is the receptor, M is the solvent medium, and *K* is the equilibrium, or binding, constant. The binding free energy, ΔG, of a ligand, L, to a receptor, R, in a solvent medium, M, is equal to the difference in free energy of the bound state and unbound states and can be expressed as

$$\Delta G = G_{LR} - (G_L + G_R) = - RT \ln K \quad (2)$$

where G is the binding free energy, G_{LR} is the free energy of the bound, or complex, state, G_L is the free energy of the unbound ligand, G_R is the free energy of the unbound receptor, R is the gas constant and T is the temperature of the system. The free energy of an enzyme-ligand complex can be approximately broken down into a set of component interactions as follows,

$$G_{LR} = [G_{LR}(LL) + G_{LR}(RR) + G_{LR}(MM) + G_{LR}(LR) + G_{LR}(LM) + G_{LR}(RM)] \quad (3)$$

where G_{LR} (XY) refers to the interaction between X and Y where they each can be L, M or R.

The interaction terms can be divided into their respective enthalpy, H_{LR}, and entropy, S_{LR}, contributions.

$$G_{LR} = H_{LR} - TS_{LR}. \quad (4)$$

At low solute concentration the enthalpy terms, $H_{LR}(XY)$, can be represented by their respective internal energies, $E_{LR}(XY)$,

$$H_{LR} = E_{LR} = [E_{LR}(LL) + E_{LR}(RR) + E_{LR}(MM) + E_{LR}(LR) + E_{LR}(LM) + E_{LR}(RM)] \quad (5)$$

and the entropy term, $S_{LR}(XY)$, contributions can be expressed in the same manner,

$$S_{LR} = [S_{LR}(LL) + S_{LR}(RR) + S_{LR}(MM) + S_{LR}(LR) + S_{LR}(LM) + S_{LR}(RM)] \quad (6)$$

The unbound ligand, G_L, and receptor, G_R, free energies have the following components

$$G_L = [G_L(LL) + G_L(LM) + G_L(MM)], \quad (7)$$

$$G_R = [G_R(RR) + G_R(RM) + G_R(MM)]. \quad (8)$$

The enthalpy contributions of L and R at low concentration, $H_L(XY)$ and $H_R(XY)$ can also be represented by their internal energies, $E_L(XY)$ and $E_R(XY)$ as in Equations 5 and 6. The complete set of contributions to the internal energy and entropy, and their representations, are given in Table **3**.

The terms in Table **3** can be selected as the independent variables used in the FEFF 3D-QSAR analysis. However, the free energy of binding, ΔG, can also be represented by the individual free energy force field terms for L, R, and LR.

$$\Delta G = \alpha 1 \Delta E_{stretch} + \alpha 2 \Delta E + \alpha 3 \Delta E_{torsion} + \alpha 4 \Delta E_{vdW} + \alpha 5 \Delta E_{electrostatic} + \alpha 6 \Delta E_{hydrogen\ bonding} + \alpha 7 \Delta E_{solvation} - \alpha 8 T \Delta S, \quad (9)$$

where $\Delta E_{stretch}$ is the unbound to bound change in internal energy for bond stretching, ΔE_{bend} is the change in bond angle bending energy, $\Delta E_{torsion}$ is the change in torsional energy, ΔE_{vdW} is the change in van der Waals interaction energy, $\Delta E_{electrostatic}$ is the change in electrostatics interaction energy, $\Delta E_{hydrogen\ bonding}$ is the change in hydrogen bonding energy, $\Delta E_{solvation}$ is the change in solvation energy, and ΔS is the total change in the entropy of the L, R,M system which can be partitioned into component contributions.

The hydration shell model proposed by Hopfinger [33, 34] was included in the potential energy function to calculate

Table 3. Breakdown of the FEFF Interaction Terms, XY, for Ligand (L) – Receptor (R) in a Solvent Medium (M)

Binding Component	Type of Interaction Energy, XY	Change in Internal Energy	Change in Entropy
Ligand L	Intramolecular ligand conformation energy LL	$\Delta E_L(LL) = E_{LR}(LL) - E_L(LL)$	$\Delta S_L(LL) = S_{LR}(LL) - S_L(LL)$
	Ligand salvation energy LM	$\Delta E_L(LM) = E_{LR}(LM) - E_L(LM)$	$\Delta S_L(LM) = S_{LR}(LM) - S_L(LM)$
Solvent M	Solvent reorganizational energy MM	$\Delta E_M(MM) = E_{LR}(MM) - [E_L(MM) + E_R(MM)]$	$\Delta S_M(MM) = S_{LR}(MM) - [S_L(MM) + S_R(MM)]$
Receptor R	Intramolecular receptor conformational energy RR	$\Delta E_R(RR) = E_{LR}(RR) - E_R(RR)$	$\Delta S_R(RR) = S_{LR}(RR) - S_R(RR)$
	Receptor solvation energy RM	$\Delta E_R(RM) = E_{LR}(RM) - E_R(RM)$	$\Delta S_R(RM) = S_{LR}(RM) - S_R(RM)$
Ligand-receptor RL	Intermolecular ligand-receptor LR	$\Delta E_{LR}(LR) = E_{LR}(LR)$	$\Delta S_{LR}(LR) = S_{LR}(LR)$

the solvation energies. Only the L and R components to the free energy of aqueous salvation can be extracted from this model. Thus, only these energy terms can be used as trial descriptors in building a FEFF 3D-QSAR model.

FEFF 3D-QSAR was successfully applied to a set of peptidometic rennin inhibitors [32], to a set of glucose analogue inhibitors of glucogen phosphorylase [35], and to a set of *Plasmodium falciparum* dihydrofolate reductase inhibitors [36].

MEMBRANE INTERACTION (MI)-QSAR

A methodology called *membrane* interaction (MI)-QSAR analysis, which is a combination of structure-based design with classic intramolecular QSAR analysis, was proposed to model chemically and structurally diverse compounds interacting with cellular membranes [37].

In MI-QSAR analysis the assumption is made that the phospholipid regions of a cellular membrane constitute the "receptor" required in structure-based design that permits incorporation of structural and chemical diversity into a training set. A set of *membrane−solute intermolecular properties* are determined and added to a set of comprehensive intramolecular solute QSAR descriptors to enlarge the trial QSAR descriptor pool and, ostensibly, to provide the information needed to incorporate chemical and structural diversity into the QSAR analysis.

The MI-QSAR descriptor terms have proven to be significant in generating models for predicting eye irritation to chemicals [37, 38], Caco-2 cell permeation coefficients of molecules [39, 40], blood brain barrier partitioning of molecules [41], characterization of skin penetration of organic chemicals [42], and other ADMET properties [43].

An important strength of the MI-QSAR approach is to be able to construct simple and statistically significant relationships and a corresponding general mechanistic equation. That is, MI-QSAR analysis is able to generate meaningful ADME property models employing a limited number of descriptors that can be directly interpreted in terms of physically reasonable mechanisms of action. There is no need to resort to generating very large numbers of [only] *intramolecular solute* descriptors and then producing a model that meets the statistical constraints of acceptance by performing some type of data reduction.

COMBINE

Ortiz and colleagues developed a RD 3D-QSAR method called comparative binding energy (COMBINE) analysis [44], which relies on the use of a series of structures of ligand–receptor complexes (modeled or experimentally determined) to quantify interaction energies by molecular mechanics computations.

The method can be summarized as follow [44]:

1. Selection of experimental binding data values (IC_{50}, K_d, K_i, etc.) for the training set of ligand-receptor complexes.
2. Selection or modeling of the structures receptor-ligand complexes.
3. Energy minimization of all models.
4. Calculation of ligand-receptor interaction energy.
5. Partitioning of each ligand-receptor interaction energy into van der Waals and electrostatic contributions per receptor residue.
6. Correlation of the binding data values with the different components of the interaction energy by using partial least squares (PLS) regression.
7. Internal and external validation of the QSAR model.

COMBINE models not only score the complexes, but also highlight interactions important for the differences in binding amongst the training set, and thus serve as an aid to molecular design. Moreover, in COMBINE analysis the training set can consist of complexes with one receptor and different ligands, one ligand and different receptors, or several different ligands and several different receptors.

COMBINE models have been successfully constructed in a number of cases [45-50]. It has also been demonstrated that regression-based QSAR models derived with COMBINE analysis are suitable for fast virtual screening of compound databases [51].

According to reported results, COMBINE can yield predictive models and mechanistic insights can be obtained. However, the predictive ability of the method can be expected to be significantly enhanced by improvements in the description of the electrostatic term, the inclusion of suitable

descriptors for solvation and entropic effects, and the optimization of particular aspects of the methodology, such as the choice of ligand fragment definitions and the details of the variable selection protocol.

AFMoC

Klebe and colleagues developed adaptation of fields for molecular comparison (AFMoC) [52], which can be considered as a reverse comparative molecular field analysis (CoMFA), since molecular interaction fields (MIFs) of the receptor are computed on a grid over the binding site using knowledge-based pair-potentials and multiplied by Gaussian functions at the positions of atoms in docked ligands. PLS analysis is applied in a similar way to CoMFA to weight the contributions of different atom-types on the grid and derive an expression for binding affinity. Thus, one could say that AFMoC is a receptor-dependent "version"of CoMFA. AFMoC allows one to gradually move from general knowledge-based potentials to receptor-specific adapted potentials.

Recently, an extension of AFMoC called consensus adaptation of fields for molecular comparison ($AFMoC^{con}$) [53], which takes into account multiple ligand conformations in an ensemble of receptor configurations. $AFMoC^{con}$ approach yields reasonable accurate binding affinity predictions and seems to alleviate the need to choose an "appropriate" receptor structure among several alternatives.

The AFMoC approach generates receptor-specific adapted potential fields inside binding pockets that can be subsequently used for binding affinity predictions [52, 54, 55] and as the objective function in docking [56]. The methodology consists of four basic steps [52]:

1. Potential field calculation and ligand alignment;
2. Interaction field calculations;
3. Correlating interaction field values with binding affinities and prediction of binding affinities;
4. Binding affinity prediction for novel ligands.

AFMoC overcomes the prerequisite to involve complete ligand training sets. In addition, since interaction fields based on atom types are used, interpreting the PLS results in terms of ligand structure optimization to achieve better binding affinities is straightforward. Moreover, the information contained in the different atom-type based interaction fields is mutually orthogonal. This is an important advantage over the information comprised by generic fields such as "van der Waals" or "electrostatic" interaction used in CoMFA. Finally, since structural information of experimentally determined complexes is onverted into statistical pairpotentials, the latter (and the derived interaction fields) are expected to contain besides enthalpic also entropic effects, resulting from (de-)solvation [52].

CoLiBRI

Tropsha's group developed a new computational drug discovery method termed complimentary ligands based on receptor information (CoLiBRI) that combines structure-based and ligand-based paradigms [57]. With this approach, both receptor active sites and their corresponding ligands are characterized in the same multidimensional, chemical descriptor space. The idea is that mapping both binding pockets and the ligands onto the same chemical space would preserve the complementarity relationships between binding sites and the ligands. Thus, similar binding sites would correspond to similar ligands and, consequently, the relative location of a novel binding site in the chemistry space with respect to other binding sites could be used to predict the location of the ligand(s) complementary to that site in the ligand chemistry space. This virtual ligand(s) could then be used as a query in chemical similarity searches to identify putative ligands of the same receptor in available chemical databases.

Molecular descriptors based on transferable atom equivalents (TAE)/RECON developed by Breneman and colleagues are [58-60] used. Such an approach is based on Bader's quantum theory atoms in molecules [61].

CoLiBRI consists of the following steps:

1. Randomly selection of a subset of nVar descriptors, which is regarding the hypothetical topological requirements for a ligand to bind with a receptor.
2. Exclusion of the receptor.
3. Calculate the predicted ligand point in the ligand space for the excluded receptor based on the relative based on the relative orientation of ligands known to bind with the k most similar receptors of the excluded receptor.
4. Predict the ligand(s) of the excluded receptor as ligand(s) closest to the predicted binding point in the ligand space.
5. Repeat steps 2-4 in order to find the best topological requirements that minimize the PMR value of the CoLiBRI model. Such an optimization process is driven by a generalized simulated annealing using PMR as the objective function.
6. Calculate the predictive ability (PMR) of the "model".
7. Select the best CoLiBRI model for nVar and k.

As an illustrative example, CoLiBRI was applied to a set of 800 X-ray characterized receptor-ligand complexes [57] in the PDBBinding database [62, 63]. It was shown knowledge of the active site structure affords identification of its complimentary ligand among the top 1% of a large chemical database in over 90% of all test active sites when a binding site of the same protein family was present in the training set. In the case where test receptors are highly dissimilar and not present among the receptor families in the training set, the prediction accuracy is decreased; however, CoLiBRI was still able to quickly eliminate 75% of the chemical database as improbable ligands. CoLiBRI affords rapid prefiltering of a large chemical database to eliminate compounds that have little chance of binding to a receptor active site [57].

CoRIA

Coutinhos's group proposed the comparative residue interaction analysis (CoRIA) methodology [64], which is a QSAR approach for taking into account receptor-ligand interactions as well as the thermodynamics of the corresponding binding process.

The technique was tested to a set of 36 cyclooxygenase-2 (COX-2) inhibitors belonging to different structural classes. The constructed QSAR models were internally cross-validated and boot strapping, and externally validated using a test set of 13 molecules. Moreover, the models are shown to be robust with good r^2 and q^2 values [64].

The methodology can be summarized as follows [64]:

1. Docking of the ligands to the active site of the receptor.
2. Computation of the energy terms in the CoRIA approach is similar to those terms of the FEFF 3D-QSAR approach (solvation energy, strain energy, non-bonded interaction energies) and a couple of intramolecular miscellaneous terms for the ligands, such as the free energy for desolvation of the ligand in water (Fh2o) and in octanol (Foct), Jurs descriptors [65], molecular refractivity, molecular volume, lipophilicity (log P), and molecular surface area.
3. Optimization of the QSAR models with genetic function approximation algorithm (GFA) [66].
4. Analysis and validation of the QSAR models.

COMBINING DOCKING, MOLECULAR FINGERPRINTS-BASED CLUSTER ANALYSIS AND 'INDUCTIVE' DESCRIPTORS FOR DERIVING STRUCTURE-BASED QSAR

Recently, Santos-Filho proposed a computer-aided drug design approach [67] which includes docking, molecular fingerprints based cluster analysis, and Cherkasov's approach for estimating 'inductive' descriptors [68].

Structurally diverse ligand data sets from the literature and from known databases were used to define its applicability domain and to evaluate its advantages and limitations. When needed, molecular fingerprints-based cluster analysis was used for resampling the data sets. Extensive docking calculations were carried out in order to "capture" optimum intermolecular interactions of the ligands with their respective receptor binding sites. PLS and genetic algorithm were used for constructing the QSAR models.

The results presented in that study illustrate the complex scenario common to any drug design project, where distinct tools are complementarily used in order to explore multiple aspects of the chemical space. The constructed models showed to be interpretable, with high statistical and predictive significance, and could be used for guiding ligand modification for the development of potential new inhibitors for several targets.

It was shown how docking and QSAR analysis can be used together for the construction of drug design hypothesis, when working with structurally diverse data sets. Moreover, the applicability of molecular fingerprint based cluster analysis for screening and resampling that kind of data was also described. It was shown that the method is reliable and accurate enough for calculating QSAR models for structurally diverse ligand data sets.

The advantages of the proposed 3D-QSAR approach over other ones are as follows: (a) No hypothetical receptor structure is used, as in the case of 6D-QSAR. Instead, actual 3D-structures are used. (b) No ambiguous alignments are needed, since extensive docking simulations are carried out, so that the orientations of the ligands are the optima docked poses. (c) 'Induced' descriptors "capture" intermolecular interactions between ligands and receptors. (d) As compared to other methods, the computational overhead is less.

Not including additional conformational of the ligands seems to be the main limitation of the approach. Probably, the proposition of "active" conformations for the ligands, based on conformational ensemble, would improve still more the statistical significance of the models. However, conformational ensemble calculation would certainly increase the computational cost.

CONCLUSION

This review is not intended to be comprehensive. However, we have tried to outline the most important RD multi-dimensional QSAR methods available at the time of writing the manuscript. Methods that explicitly make use of both receptors and ligands have been applied and validated on a wide-range of problems of relevance to drug design. With advances in high-throughput virtual screening methods, mainly by the development of more accurate scoring functions for docking algorithms, as well as the inclusion of electronic features derived from quantum chemistry, we expect that structure-based QSAR methods will become increasingly valuable owing to their accuracy, reliability and general applicability.

ACKNOWLEDGEMENTS

OAS-F and RBA are grateful for financial support from the Brazilian Agencies FAPERJ, CAPES, and CNPq. AJH gratefully acknowledges support from the National Institutes of Health, USA, through the NIH Roadmap for Medical Research, Grant 1 R21 GM075775, and from both Avon and the Procter and Gamble Company. AC acknowledges the financial support for the PREPARE Project (http://www.prepare.med.ubc.ca/).

REFERENCES

[1] Tollenaere, J. P. In *Guidebook on Molecular Modeling in Drug Design*; Cohen, N. C., Ed.; Academic Press: San Diego, **1996**, pp. 337-356.

[2] Hansch, C.; Fujita, T. ρ-σ-π analysis. A method for correlation of biological activity and chemical structure. *J. Am. Chem. Soc.*, **1964**, *86*, 1616-1626.

[3] Free, S. M., Jr.; Wilson, J. W. A mathematical contribution to structure-activity studies. *J. Med. Chem.*, **1964**, *7*, 395-399.

[4] Esposito, E. X.; Hopfinger, A. J.; Madura, J. D. In *Methods in Molecular Biology: Chemoinformatics: Concepts, Methods, and Tools for Drug Discovery*. Bajorath, J., Ed.; Humana Press: Totowa, **2004**, Vol. *275*, pp. 131-213.

[5] Tropsha, A. In *Comprehension Medicinal Chemistry II*; Mason, J., Ed.; Elsevier: New York, **1997**; Vol. *4*, pp. 149-165.

[6] Doweyko, A. In *Comprehension Medicinal Chemistry II*; Mason, J., Ed.; Elsevier: New York, **1997**; Vol. *4*, pp. 575-594.

[7] Terfloth, L. In *Chemoinformatics: A Textbook*, Gasteiger, J.; Engel, T., Eds.; Wiley-VCH: Weinheim, **2003**, pp. 401-437.

[8] Hopfinger, A. J. A QSAR investigation of dihydrofolate reductase inhibition by Baker triazines based upon molecular shape analysis. *J. Am. Chem. Soc.*, **1980**, *102*, 7196-7206.

[9] Hopfinger, A. J. Inhibition of dihydrofolate reductase: structure-activity correlations of 2,4-diamino-5-benzylpyrimidines based upon molecular shape analysis. *J. Med. Chem.*, **1981**, *24*, 818-822.

[10] Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.*, **1997**, *119*, 10509-10524.

[11] Albuquerque, M. G.; 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.

[12] Prabha, V. P.; Hopfinger, A. J. Prediction of ligand−receptor binding free energy by 4D-QSAR analysis: application to a set of glucose analogue inhibitors of glycogen fhosphorylase, *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 1141-115.

[13] Santos-Filho, O. A; Hopfinger, A. J. A search for sources of drug resistance by the 4D-QSAR analysis of a set of antimalarial dihydrofolate reductase inhibitors. *J. Comput.-Aided Mol. Des.*, **2001**, *15*, 1-12.

[14] Duca, J. S.; Hopfinger, A. J. Estimation of molecular similarity based on 4D-QSAR analysis: formalism and validation. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 1367-1387.

[15] Ravi, M.; Hopfinger, A. J.; Hormann, R. E.; Dinan, L. 4D-QSAR analysis of a set of ecdysteroids and a comparison to CoMFA modeling. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 1587-1604.

[16] Krasowski, M. D.; Hong, X., Hopfinger, A. J.; Harrison, N. L. 4D-QSAR analysis of a set of propofol analogues: mapping binding sites for an anesthetic phenol on the $GABA_A$ receptor. *J. Med. Chem.*, **2002**, *45*, 3210-3221.

[17] Hong, X.; Hopfinger, A. J. 3D-pharmacophores of flavonoid binding at the benzodiazepine $GABA_A$ receptor site using 4D-QSAR analysis. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 324-336.

[18] Senese, C. L.; Hopfinger, A. J. Receptor-independent 4D-QSAR analysis of a set of norstatine derived inhibitors of HIV-1 protease. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 1297-1307.

[19] Liu, J.; Pan, D.; Tseng, Y.; Hopfinger, A. J. 4D-QSAR analysis of a series of antifungal P450 inhibitors and 3D-pharmacophore comparisons as a function of alignment. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 2170-2179.

[20] Senese, C. L.; Hopfinger, A. J. A simple clustering technique to improve QSAR model selection and predictivity: application to a receptor independent 4D-QSAR analysis of cyclic urea derived inhibitors of HIV-1 protease. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 2180-2193.

[21] Pasqualoto, K. F. M.; Ferreira, E. I.; Santos-Filho, O. A.; Hopfinger, A. J. Rational design of new antituberculosis agents: receptor-independent four-dimensional quantitative structure−activity relationship analysis of a set of isoniazid derivatives. *J. Med. Chem.*, **2004**, *47*, 3755-3764.

[22] Pan, D.; Tseng, Y.; Hopfinger, A. J. Quantitative structure-based design: formalism and application of receptor-dependent RD-4D-QSAR analysis to a set of glucose analogue inhibitors of glycogen phosphorylase. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 1591-1607.

[23] Pan, D.; Liu, J.; Senese, C.; Hopfinger, A. J.; Tseng, Y. Characterization of a ligand-receptor binding event using receptor-dependent four-dimensional quantitative structure-activity relationship analysis. *J. Med. Chem.*, **2004**, *47*, 3075-3088.

[24] Santos-Filho, O. A.; Hopfinger, A. J. Structure-based QSAR analysis of a set of 4-hydroxy-5,6-dihydropyrones as inhibitors of HIV-1 protease: an application of the receptor-dependent (RD) 4D-QSAR formalism. *J. Chem. Inf. Model.*, **2006**, *46*, 345-354.

[25] Biographics Laboratory 3R. Friedensgasse 35, CH-4056 Basel, Switzerland.

[26] Vedani, A.; Briem, H.; Dobler, M.; Dollinger, K.; McMasters, D. R. Multiple conformation and protonation-state representation in 4D-QSAR: the neurokinin-1 receptor system. *J. Med. Chem.*, **2000**, *43*, 4416-4427.

[27] Vedani, A.; Dobler, M. Multidimensional QSAR in drug research: predicting binding affinities, toxicity, and pharmacokinetic parameters. In *Progress in Drug Research*; Jucker, E., Ed.; Birkhäuser: Basel/Boston/Berlin, **2000**; pp. 105-135.

[28] Vedani, A.; McMasters, D. R.; Dobler, M. Multi-conformational ligand representation in 4D-QSAR: Reducing the bias associated with ligand alignment. *Quant. Struct.-Act. Relat.*, **2000**, *19*, 149-161.

[29] Streich, D.; Neuburger-Zehnder, M.; Vedani, A. Induced fit–the key for understanding LSD activity. A 4D-QSAR study on the 5-HT2A receptor system. *Quant. Struct.-Act. Relat.*, **2000**, *19*, 565-573.

[30] Vedani, A.; Dobler, M. 5D-QSAR: the key for simulating induced fit? *J. Med. Chem.*, **2002**, *45*, 2139-2149.

[31] Vedani, A.; Dobler, M.; Lill, M. A. Combining protein modeling and 4D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *J. Med. Chem.*, **2005**, *48*, 3700-3703.

[32] Tokarski, J. S; Hopfinger, A. J. Prediction of ligand-receptor binding thermodynamics by free energy force field (FEFF) 3D-QSAR analysis: application to a set of peptidometic renin inhibitors. *J. Chem. Inf. Model.*, **1997**, *37*, 792-811.

[33] Hopfinger, A. J. In *Conformational Properties of Macromolecules*; Academic Press: New York, **1973**; p. 71.

[34] Forsythe, K. H.; Hopfinger, A. J. The influence of solvent on the secondary structures of poly(L-alanine) and poly (L-proline). *Macromolecules*, **1973**, *6*, 423-437.

[35] Venkataragan, P.; Hopfinger, A. J. Prediction of ligand-receptor binding thermodynamics by free energy force field three-dimensional quantitative structure-activity relationship analysis: application to a set of glucose analogue inhibitors of glycogen phosphorylase. *J. Med. Chem*, **1999**, *42*, 2169-2179.

[36] Santos-Filho, O. A.; Hopfinger, A. J. Free energy force field (FEFF) 3D-QSAR analysis of a set of *Plasmodium falciparum* dihydrofolate reductase inhibitors. *J. Comput. Aided Mol. Des.*, **2001**, *15*, 787-810.

[37] Kulkarni, A. S.; Hopfinger, A. J.; Osborne, R.; Bruner, L. H.; Thompson, E. D. Prediction of eye irritation from organic chemicals using membrane-interaction QSAR analysis. *Toxicol. Sci*., **2001**, *59*, 335-345.

[38] Kulkarni, A. S.; Hopfinger, A. J. Membrane-interaction QSAR analysis: application to the estimation of eye irritation by organic compounds. *Pharm. Res.*, **1999**, *16*, 1244-1252.

[39] Kulkarni, A. S.; Han, Y.; Hopfinger, A. J. Predicting Caco-2 cell permeation coefficients of organic molecules using membrane-interaction QSAR analysis. *J. Chem. Inf. Comput. Sci*., **2002**, *42*, 331-342.

[40] Santos-Filho, O. A.; Hopfinger, A. J. Combined 4D-fingerprint and cluster based membrane-interaction QSAR analysis for constructing consensus Caco-2 cell permeation virtual screens. *J. Pharm. Sci.*, **2008**, *97*, 566-583.

[41] Iyer, M.; Mishra, R.; Han, Y.; Hopfinger, A. J. Predicting blood-brain barrier partitioning of organic molecules using membrane-interaction QSAR analysis. *Pharm. Res*., **2002**, *19*, 1611-1621.

[42] Santos-Filho, O. A.; Hopfinger, A. J.; Zheng, T. Characterization of skin penetration processes of organic molecules using molecular similarity and QSAR analysis. *Mol. Pharm.*, **2004**, *1*, 466-476.

[43] Liu, J.; Li, Y.; Pan, D.; Hopfinger, A. J. Predicting permeability coefficient in ADMET evaluation by using different membranes-interaction QSAR. *Int. J. Pharm.*, **2005**, *304*, 115-123.

[44] Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of drug binding affinities by comparative binding energy analysis. *J. Med. Chem*., **1995**, *38*, 2681-2691.

[45] Kmunicek, J.; Luengo, S.; Gago, F.; Ortiz, A. R.; Wade, R. C.; Damborsky, J. Comparative binding energy analysis of the substrate specificity of haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10. *Biochemistry,* **2001**, *40*, 8905-8917.

[46] Wang, T.; Wade, R. C. Comparative binding energy (COMBINE) analysis of OppA-peptide complexes to relate structure to binding thermodynamics. *J. Med. Chem.,* **2002**, *45*, 4828-4837.

[47] Wang, T.; Wade, R. C. Comparative binding energy (COMBINE) analysis of influenza neuraminidase-inhibitor complexes. *J. Med. Chem.,* **2001**, *44*, 961-971.

[48] Perez, C.; Pastor, M.; Ortiz, A. R.; Gago, F. Comparative binding energy analysis of HIV-1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design. *J. Med. Chem.,* **1998**, *41*, 836-852.

[49] Murcia, M.; Morreale, A.; Ortiz, A. R. COMBINE analysis considering multiple receptors: a step towards structure-activity models for protein families. *J. Med. Chem*., **2006**, *49*, 6241-6253.

[50] Lushington, G. H.; Guo, J. X.; Wang, J. L. Whither combine? New opportunities for receptor-based QSAR. *Curr. Med. Chem*., **2007**, *14*, 1863-1877.

[51] Murcia, M.; Ortiz, A. R. Virtual screening with flexible docking and COMBINE based models. Application to a series of factor Xa inhibitors. *J. Med. Chem*., **2004**, *47*, 805-820.

[52] Gohlke, H.; Klebe, G. DrugScore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledgebased pair-potentials to a particular protein. *J. Med. Chem.,* **2002**, *45*, 4153-4170.

[53] Breu, B.; Silber, K.; Gohlke, H. Consensus adaptation of fields for molecular comparison (AFMoC) models incorporate ligand and receptor conformational variability into Tailor-made scoring functions. *J. Chem. Inf. Model.*, **2007**, *47*, 2383-2400.

[54] Matter, H.; Will, D. W.; Nazare, M.; Schreuder, H.; Laux, V.; Wehner, V. Structural requirements for factor Xa inhibition by 3-oxybenzamides with neutral P1 substituents: combining X-ray crystallography, 3DQSAR, and tailored scoring functions. *J. Med. Chem.,* **2005**, *48*, 3290-312.

[55] Silber, K.; Heidler, P.; Kurz, T.; Klebe, G. AFMoC enhances predictivity of 3D QSAR: a case study with DOXP-reductoisomerase. *J. Med. Chem.,* **2005**, *48*, 3547-3563.

[56] Radestock, S.; Bohm, M.; Gohlke, H. Improving binding mode predictions by docking into protein-specifically adapted potential fields. *J. Med. Chem.,* **2005**, *48*, 5466-5479.

[57] Oloff, S.; Zhang, S.; Sukimar, N.; Breneman, C.; Tropsha, A. Chemometric analysis of ligand receptor complementarity. Identifying complementary ligands based on receptor information (CoLiBRI). *J. Chem. Inf. Model.*, **2006,** *46,* 844-851.

[58] Breneman, C. M.; Thompson, T. R.; Rhem, M.; Dung, M. Electron density modeling of large systems using the transferable atom equivalent method. *Comput. Chem.,* **1995**, *19*, 161-169.

[59] Mazza, C. B.; Sukumar, N.; Breneman, C. M.; Cramer, S. M. Prediction of protein retention in ion-exchange systems using molecular descriptors obtained from crystal structure. *Anal. Chem.,* **2001**, *73*, 5457-5461.

Received: 11 February, 2009 **Revised: 07 April, 2009** **Accepted: 07 April, 2009**

[60] Song, M.; Breneman, C. M.; Bi, J.; Sukumar, N.; Bennett, K. P.; Cramer, S.; Tugcu, N. Prediction of protein retention times in anionexchange chromatography systems using support vector regression. *J. Chem. Inf. Comput. Sci.,* **2002**, *42*, 1347-1357.

[61] http://www.chemistry.mcmaster.ca/faculty/bader/aim/

[62] Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. The PDBbind database: methodologies and updates. *J. Med. Chem.*, **2005**, *48*, 4111-4119.

[63] Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.*, **2004**, *47*, 2977-2980.

[64] Datar, P. A.; Khedkar, S. A.; Malde, A. K.; Coutinho, E. C. Comparative residue interaction analysis (CoRIA): a 3D-QSAR approach to explore the binding contributions of active site residues with ligands. *J. Comput. Aided Mol. Des.*, **2006**, *20*, 343-360.

[65] Staton, 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.

[66] Rogers, D.; Hopfinger, A. J. Applications of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.,* **1994**, *34*, 854-866.

[67] Santos-Filho, O. A.; Cherkasov, A. Using molecular docking, 3D-QSAR, and cluster analysis for screening structurally diverse data sets of pharmacological interest. *J. Chem. Inf. Model.*, **2008**, *48*, 2054-2065.

[68] Cherkasov, A. 'Inductive' descriptors. 10 successful years in QSAR. *Curr. Comput. Aided Drug Des.*, **2005**, *1*, 21-42.