*Review*

# Recent Advances in Fragment-Based QSAR and Multi-Dimensional QSAR Methods

**Kyaw Zeyar Myint [1] and Xiang-Qun Xie [1,2,3,]**

[1] Department of Computational Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15260, USA; E-Mail: kym3@pitt.edu

[2] Department of Pharmaceutical Sciences, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA 15260, USA

[3] Pittsburgh Chemical Methodologies & Library Development (PCMLD) and Pittsburgh Drug Discovery Institute, University of Pittsburgh, Pittsburgh, PA 15260, USA

[*] Author to whom correspondence should be addressed; E-Mail: xix15@pitt.edu; Tel.: +1-412-383-5276; Fax: +1-412-383-7436.

**Abstract:** This paper provides an overview of recently developed two dimensional (2D) fragment-based QSAR methods as well as other multi-dimensional approaches. In particular, we present recent fragment-based QSAR methods such as fragment-similarity-based QSAR (FS-QSAR), fragment-based QSAR (FB-QSAR), Hologram QSAR (HQSAR), and top priority fragment QSAR in addition to 3D- and nD-QSAR methods such as comparative molecular field analysis (CoMFA), comparative molecular similarity analysis (CoMSIA), Topomer CoMFA, self-organizing molecular field analysis (SOMFA), comparative molecular moment analysis (COMMA), autocorrelation of molecular surfaces properties (AMSP), weighted holistic invariant molecular (WHIM) descriptor-based QSAR (WHIM), grid-independent descriptors (GRIND)-based QSAR, 4D-QSAR, 5D-QSAR and 6D-QSAR methods.

**Keywords:** QSAR; fragment similarity based; fragment-based; 2D-QSAR; 3D-QSAR; nD-QSAR

## 1. Introduction

Quantitative structure-activity relationship (QSAR) is based on the general principle of medicinal chemistry that the biological activity of a ligand or compound is related to its molecular structure or properties, and structurally similar molecules may have similar biological activities [1]. Such molecular structural information is encoded in molecular descriptors and a QSAR model defines mathematical relationships between descriptors and biological activities of known ligands to predict unknown ligands' activities. QSAR methods have been applied in several scientific studies including chemistry, biology, toxicology and drug discovery to predict and classify biological activities of virtual or newly-synthesized compounds [2–6]. QSAR models can also be used in designing new chemical entities (NCEs) and are now regarded as essential tools in pharmaceutical industries to identify promising hits and generate high quality leads in the early stages of drug discovery [5,7]. In other words, QSAR studies can reduce the costly failures of drug candidates by identifying the most promising hit compounds and reducing the number of costly experiments.

In general, QSAR modeling (Figure 1) involves a systematic process with multiple steps, including dataset preparation, molecular descriptors selection and generation, mathematical or statistical models derivation, model training and validation using a training dataset and model testing on a testing dataset. During the first step, or dataset preparation, it is important to pay attention to the quality of data to develop a reliable QSAR model. Data should come from the same bioassay protocols and it is preferable to collect and use the data generated from a single lab or source in order to avoid data inconsistencies and interlaboratory variability. Moreover, the dataset should have a large enough number of compounds to ensure statistical stability of a QSAR model and the bioactivity should cover a range of values with a good distribution [5]. The second step in QSAR modeling is the selection and generation of molecular descriptors for ligands in the dataset. There are many descriptors available and only some of them are significantly correlated with the activity. Therefore, selection of appropriate descriptors, which best capture the structural variation and information is important to derive a robust QSAR model. Several methods such as evolutionary algorithms (for example, genetic algorithm) and machine learning techniques (for example, forward selection) can be used for descriptor/variable/feature selection. After molecular descriptors are defined and generated for all ligands in the dataset, the next step is to decide a suitable statistical or mathematical model to find the relationship between such descriptors and biological activities. For instance, linear approaches such as multiple linear regression (MLR) or partial least square (PLS) and non-linear methods such as neural networks or support vector machine can be used as correlation or mapping functions. Once a model is chosen, it is then trained on a training dataset which contains a subset of randomly selected compounds from a known dataset, leaving the remaining to be used as testing compounds. During the model training, validation methods such as leave-one-out cross-validation (LOOCV) are often performed to ensure the statistical stability of the QSAR model. The training process is repeated until a satisfactory training performance is achieved. Finally, a testing process is performed in which the trained model is used to predict activity values of those compounds in the testing set.
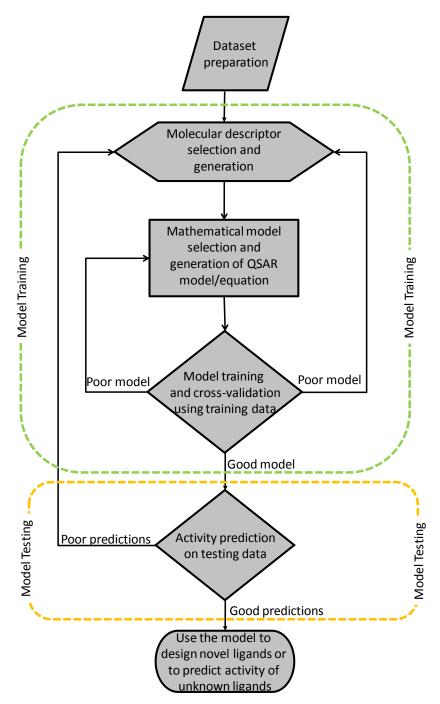
**Figure 1.** A general scheme of a QSAR model development which includes systematic training and testing processes.



A wide range of QSAR methodologies have been invented since the concept was first introduced by Free, Wilson, Hansch, and Fujita [8,9] in 1964. Traditional 2D-QSAR methods such as Free-Wilson and Hansch-Fujita models use 2D molecular substituents or fragments and their physicochemical properties to perform quantitative predictions. Since then, QSAR has experienced a fast development and the first novel 3D-QSAR method called comparative molecular field analysis (CoMFA) was introduced by Cramer *et al.* in 1988. The CoMFA method brought a foundation for the development of other 3D-QSAR methods such as CoMSIA, SOMFA, CoMMA as well as multidimensional (nD)-QSAR methods such as 4D-QSAR, 5D-QSAR, *etc.*, to tackle known 3D-QSAR problems such as

subjective molecular alignment and bioactive conformation problems. In recent years, fragment-based methods have attracted some attention because predicting molecular properties and activities based on molecular fragments is simple, fast and robust. In this review, we present recently available fragment-based QSAR methods and multidimensional (nD)-QSAR methods developed over the past few decades.

## 2. Fragment-Based 2D-QSAR Methods

Over the years, improved methods—that are based on such traditional QSAR methods—have been introduced. 2D methods allow modeling of a wide variety of ligands or compounds including cases where 3D crystal receptor or target structures are not available [7].

### 2.1. Hologram-QSAR (HQSAR)

One earlier example of a fragment-based method is HQSAR (Hologram QSAR) from Tripos [10,11]. Given a method based on 2D molecular fragments, HQSAR does not require molecular alignment and therefore allows for automated analyses of large data sets without manual intervention. The first step in the HQSAR method is to generate molecular holograms which contain counts of molecular fragments and can be related to 2D fingerprints. As depicted in Figure 2, the input dataset contains 2D structures of compounds and they are split into all possible linear and branched fragments. Then each unique fragment is assigned to a specific large positive integer by using a cyclic redundancy check (CRC) algorithm. All fragments generated are then hashed into array (hologram) bins in the range from 1 to L (total length of hologram). Bin occupancies represent counts of fragments in each bin. In other words, they are structural descriptors, which contain topological and compositional molecular information. During the second step, such fragment counts or hologram bins are correlated to corresponding biological activities (dependent variables) in a form of mathematical equation. Leave-one-out cross-validation (LOOCV) is performed to identify an optimal number of explanatory variables or components which yields an optimal model. Then by using standard partial least square (PLS) analysis, a following mathematical regression equation is derived to correlate hologram bin values or components with corresponding biological activities:
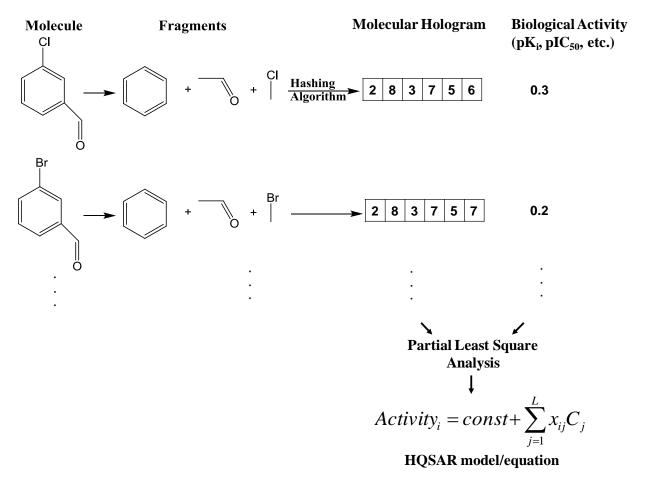
$$BA_i = const + \sum_{j=1}^{L} x_{ij} C_j \tag{1}$$

where $BA_i$ is the biological activity of the ith compound, $x_{ij}$ is the occupancy value of the molecular hologram of the ith compound at position or bin *j*, $C_j$ is the coefficient for the bin *j* derived from the PLS analysis, and *L* is the length of the hologram.

One drawback of HQSAR is a phenomenon called a fragment collision problem which happens during the hashing process of fragments. Although hashing reduces the length of the hologram, it causes bins to have different fragments in the same bin. The hologram length, a user-definable parameter, controls the number of bins in the hologram and alteration of hologram length can causes the pattern of bin occupancies to change. The program provides 12 default lengths which have been

found to give good predictive models on different datasets. Each of these default lengths provides a unique set of fragment collisions [11].

Several HQSAR models for different ligand datasets including cases where the 3D crystal structure of receptor targets or proteins are unavailable have been developed in recent years [12–15]. For example, HQSAR was used to study a set of 9-substituted-9-deazaguanine analogs which inhibit the human purine nucleoside phosphorylase (PNP) enzyme. HQSAR was used to identify structural features with poor and favorable contributions towards molecular interactions in the active site [12]. In addition, HQSAR has been used in virtual screening to identify hits [16–18]. For instance, Salum *et al.* studied a set of 180 indole derivatives having potent anticancer activity. They developed several HQSAR models and compared them to determine optimal cutoff values in virtual screening procedures [7].

**Figure 2.** Hologram-QSAR (HQSAR) model development, which includes molecular hologram generation and partial least square analysis to derive a final predictive HQSAR equation.



$$Activity_i = const + \sum_{j=1}^{L} x_{ij} C_j$$

**HQSAR model/equation**

## 2.2. Fragment-Based QSAR (FB-QSAR)

Recently, Du *et al.* [19] introduced a 2D-QSAR method based on molecular fragments. The method uses a mixed Hansch-Fujita [9] linear free energy equation and Free-Wilson [8] equation. In particular, molecular fragments are first generated from ligands and the total binding free energy $\Delta G_i^o$ between ligand *i* and the receptor is considered as the sum of contributions $\Delta g_{i,\alpha}$ from all fragments:

$$\Delta G_i^o = \sum_{\alpha=1}^{M} b_\alpha \Delta g_{i,\alpha} \tag{2}$$

where $\Delta g_{i,\alpha}$ is the free energy contribution of fragment $F_{i,\alpha}$ and $b_\alpha$ is a weight coefficient for each fragment. The binding free energy of a fragment, $\Delta g_{i,\alpha}$, is described by a set of physical and chemical properties of the fragment:

$$\Delta g_{i,\alpha} = \sum_{l=1}^{L} a_l p_{i,\alpha,l} \tag{3}$$

where $p_{i,\alpha,l}$ is the *l*-th property of fragment $F_{i,\alpha}$ in molecule $m_i$ and $a_l$ is the coefficient of *l*-th property of the fragment.

In their studies, a total of 48 neuraminidase (NA) inhibitor analogs were used to train and test the model. Ten physicochemical properties were calculated for each substituent. Using an iterative double least square (IDLS) procedure, two sets of coefficients, one for fragments ($b_\alpha$ from Equation 2) and another for physicochemical properties ($a_l$ from Equation 3), in the linear equation were solved alternately and iteratively until the model met the convergence criterion. After 176 iterations, the model converged and both sets of coefficients were solved. Such converged coefficients were used for the test calculation and the correlation coefficient (r) was 0.9525 (or $r^2 = 0.91$). They also tested on Free-Wilson and Hansch-Fujita models, which achieved r values of 0.2488 ($r^2 = 0.06$) and 0.9373 ($r^2 = 0.88$), respectively. The quantitative results proved the IDLS procedure enhanced the predictive power, and, given a novel method, more applications are necessary to fully explore its predictive potential.

## 2.3. Fragment-Similarity Based QSAR (FS-QSAR)

More recently, a fragment-similarity based QSAR (FS-QSAR) method [20] was developed to solve the major limitation of the original Free-Wilson method by introducing the fragment-similarity concept in the linear regression equation. Such a similarity concept was applied for the first time to improve the traditional Free-Wilson equation instead of using physicochemical properties which often produce non-unique solutions. In this approach, the fragment similarity calculation was carried out by the similarity. It used the lowest or highest eigen values calculated from BCUT-matrices [21,22], which contained partial charges of individual atoms and their atomic connection information in each individual fragments. The updated equation of the FS-QSAR is as follows:

$$-\log K_i = const + \sum_{j=1}^{N} [\max_{k=1}^{P_j} \{Sim(F_{jk}, F_{jg})\}] \times A_j^{MSF} \tag{4}$$

$N$ = the total number of substituent positions.
$P_j$ = the total number of possible substituents at the jth substituent position.
max = the max function picks the maximum score among similarity scores.
$F_{jk}$ = the kth fragment (a known fragment in the training set) at the jth substituent position.
$F_{jg}$ = a given fragment (the fragment from a testing/unknown compound) at the jth substituent position.
$Sim[F_{jk}, F_{jg}]$ = the fragment similarity function compares $F_{jg}$ to $F_{jk}$ and calculates a similarity score.
$A_j^{MSF}$ = the coefficient of the most similar fragment (MSF) at the jth substituent position.

The similarity function used in Equation (4) is defined as:

$$Sim(F_{jk}, F_{jg}) = e^{-\left|EV(F_{jk}) - EV(F_{jg})\right|} \tag{5}$$

where $EV(F_{jk})$ = lowest or highest eigen value of BCUT matrix of a fragment ($F_{jk}$).

The algorithm was developed and then tested on different datasets including 83 COX2 analogs and 85 triaryl bis-sulfone analogs. For statistical modeling, the model was repeatedly tested on five different testing sets which were generated by random selection of compounds. The average squared correlation coefficient, $r^2$, over five testing sets was 0.62 for COX2 analogs and 0.68 for bis-sulfone analogs. For comparison, the original Free-Wilson method was also tested, achieving the average $r^2$ values of 0.46 for COX2 dataset and 0.42 for bis-sulfone dataset. Moreover, for better comparison the BCUT-similarity function was replaced by Tanimoto coefficient (Tc) method, the traditional 2D molecular similarity function, and the average $r^2$ was 0.62 for both COX2 and bis-sulfone analogs. The FS-QSAR method was proved to have an effective predictive power compared to the traditional 2D-QSAR method since it solved the major limitation of the original Free-Wilson method by introducing the similarity concept into the regression equation. However, the predictive accuracy of FS-QSAR may not be as high as other higher dimension QSAR methods, but the method provides an objective, unique and reproducible 2D-QSAR model.

### 2.4. Top Priority Fragment QSAR

Casalegno *et al.* [23] introduced a fragment-based QSAR approach to predict pesticide aquatic toxicity to the rainbow trout. The method prioritizes fragments' contributions to toxicity with the assumption that one fragment among others present in a compound is mainly responsible for the toxicity. They used 282 carefully selected pesticides which were partitioned into 240 training and 42 testing molecules. In the first stage, all 282 molecules were broken into small substructures or atomic centered units (ACUs). Then, a numerical criterion based on the training set toxicity data was applied to assign one fragment or top-priority fragment (TPF), made up of one or more ACUs, to each training molecule. Once the TPFs were extracted, a 'priority matrix' was used to extract all priority relationships. A priority matrix contains information among training TPFs and can be used to find out which TPF has a priority to be assigned to a testing molecule. In the last stage, testing molecules were submitted to check for the presence of TPFs and information from the priority matrix was used to identify the ones(s) with highest priority, and final prediction was made based on average fragment toxicity. The final $r^2$ for the training set was 0.85 and 0.75 for the test set proving the model's effectiveness.

### 2.5. Other Fragment-Related QSAR Studies

In recent years, some new fragment-based QSAR methods have been discovered as well as applications to biological interests. Zhokhova *et al.* [24] introduced a method which uses fragmental descriptors with labeled atoms and applied it to their QSAR/QSPR (quantitative structure-property relationship) studies. In their approach, the fast stepwise multiple linear regression (FSMLR) and three-layer artificial neural network (ANN) methods implemented in the NASAWIN program [25]

were used to generate fragmental descriptors with labeled atoms and to construct QSAR/QSPR models. Andrade *et al.* [26] used HQSAR and other 2D-QSAR programs to study a series of hydrazides as antituberculosis agents. They used DRAGON 5.4 [27], BuildQSAR [28], PIROUETTE [29] programs for generation and selection of 2D molecular descriptors. Tsygankova *et al.* [30] also did the QSAR studies of barbituric acid derivatives using 2D fragments as descriptors with different regression approaches such as step-by-step regression to construct correlation equations.

## 3. 3D-QSAR

The 3D-QSAR methods have been developed to improve the prediction accuracies of 2D methods. 3D methods are computationally more complex and demanding than 2D approaches. In general, there are two families of 3D-QSAR methods: alignment-dependent methods and alignment-independent methods. Both families need experimentally or computationally derived bioactive conformations of ligands as templates for studies. Such 3D conformers are one of the most important factors to produce reliable 3D-QSAR models and are also the major drawbacks of 3D methods. Examples of both families are discussed below.

*3.1. Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA)*

One of well-known methods is a three dimensional QSAR method called CoMFA developed by Cramer *et al.* [31]. It is a method to describe 3D structure-activity relationship quantitatively by considering 3D structures, and steric and electrostatic fields of ligands which are superimposed to generate such molecular fields. In other words, CoMFA is an alignment-dependent method in which molecular field interaction energy terms are correlated with biological activities/responses using multivariate statistical analyses. Figure 3 illustrates a general CoMFA modeling process where active molecules are first placed in a 3D grid. Using a probe atom, steric and electrostatic energies are measured at each grid point for each molecule. Partial least square (PLS) analysis is then performed to correlate such field energy terms to activity values and make predictions. Such features and calculations make CoMFA an improved and different method from other traditional QSAR approaches.

Another 3D QSAR method named CoMSIA by Klebe *et al.* is similar to CoMFA in terms of using a probe atom along grid points. However, additional molecular fields have been implemented in the CoMSIA approach. In particular, electrostatic, steric, hydrophobic, hydrogen bond acceptor (HBA), and hydrogen bond donor (HBD) properties are generated using a Gaussian distance function [32]. Using such a Guassian-type potential function instead of Lennard-Jones and Coulombic functions provides accurate information at grid points for calculating molecular fields [33].

However, the major drawback of both methods is that all molecules have to be aligned and such alignment can affect the final CoMFA/CoMSIA model and predictions. A good alignment is necessary and quality of such alignment can be subjective, time-consuming [34] and CoMFA/CoMSIA models are sometimes non-reproducible [33]. Nevertheless, several CoMFA/CoMSIA models have been developed for many drug design and molecular modeling studies [6,35–39]
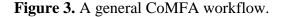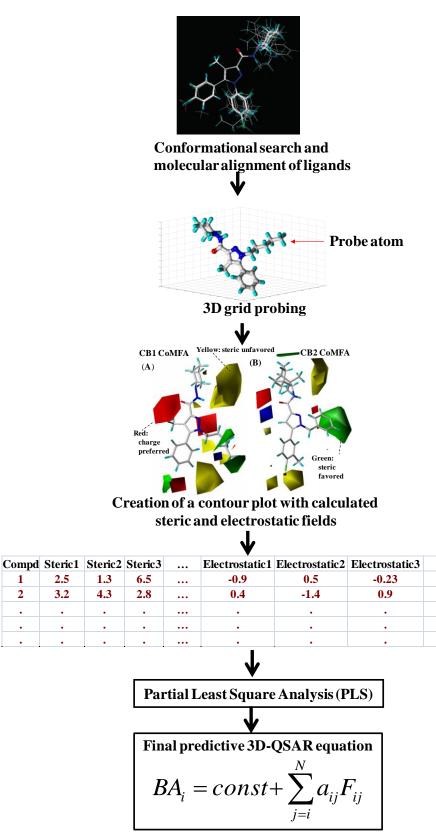
**Figure 3.** A general CoMFA workflow.



**Conformational search and
molecular alignment of ligands**

**3D grid probing**

← **Probe atom**

**Creation of a contour plot with calculated
steric and electrostatic fields**

| Compd | Steric1 | Steric2 | Steric3 | ... | Electrostatic1 | Electrostatic2 | Electrostatic3 | ... | Activity |
|-------|---------|---------|---------|-----|----------------|----------------|----------------|-----|----------|
| 1 | 2.5 | 1.3 | 6.5 | ... | -0.9 | 0.5 | -0.23 | ... | 1.43 |
| 2 | 3.2 | 4.3 | 2.8 | ... | 0.4 | -1.4 | 0.9 | ... | 0.98 |
| . | . | . | . | ... | . | . | . | ... | . |
| . | . | . | . | ... | . | . | . | ... | . |
| . | . | . | . | ... | . | . | . | ... | . |

**Partial Least Square Analysis (PLS)**

**Final predictive 3D-QSAR equation**

$$BA_i = const + \sum_{j=i}^{N} a_{ij} F_{ij}$$

$BA_i$ = biological activity of the ith compound
$a_{ij}$ = the coefficient of the calculated j[th] field energy
     (steric/electrostatic) of compound *i*
$F_{ij}$ = the calculated j[th] field energy of compound *i*

## *3.2. Topomer CoMFA*

Recently, Cramer *et al.* introduced a new QSAR method named the Topomer CoMFA [40] which is a rapid fragment-based 3D-QSAR method to predict significant R-groups, which can optimize the biological activities as well as optimized structural changes for lead scaffold hopping. It uses the compound library collection as a source of molecular fragments to identify such substituents or R-groups. The Topomer CoMFA method, unlike CoMFA, does not require the subjective alignment of 3D ligand conformers and uses automated alignment rules. A topomer describes both a conformation and orientation of a molecular fragment and it is generated based on 2D structure without any relation to a receptor site or other ligands [34,40]. After such topomers are generated, CoMFA analysis is then carried out where electrostatic and steric fields are calculated using a probe atom around the 3D grid. Subsequently, partial least square (PLS) with leave-one-out cross-validation is performed to generate a predictive model. 15 3D-QSAR analyses retrieved from the literature yielded an average $q^2$ of 0.520 compared to literature average $q^2$ of 0.636 [40]. Topomer CoMFA has the potential to optimize biological activities of ligands via fragments and has been used in lead-optimization and R-groups virtual screening studies [34,40]

## *3.3. Self-Organizing Molecular Field Analysis (SOMFA)*

Robinson *et al.* [41] introduced another alignment-dependent 3D-QSAR method called SOMFA, which is based on both molecular shape and electrostatic potentials. Briefly, 3D grids are created as in other 3D-QSAR methods and for each grid point, molecular shape and electrostatic potential values are calculated. Shape values are binary meaning 1 for being inside the van der Waals envelope and 0 outside. The key step is that the electrostatic potential value at each grid point is multiplied by the mean centered activity for that molecule as a weighing factor which causes the most active and least active molecules to have higher values than other common and less interesting molecules which are closer to the mean activity. The SOMFA grid value at a given *x,y,z* is defined as:

$$SOMFA_{x,y,z} = \sum_{i}^{Training\_Set} Property_i(x,y,z) Mean\_Centered\_Activity \qquad (6)$$

Using such a property master grid, an estimate of the activity of the $i^{th}$ molecule as defined by a certain property can be derived as:

$$SOMFA_{property,i} = \sum_x \sum_y \sum_z Property_i(x,y,z) SOMFA_{x,y,z} \qquad (7)$$

In the final stage, correlations between calculated SOMFA property values ($SOMFA_{property\,i}$) and biological activities are derived via multiple linear regression and a final predictive model is produced. Robinson *et al.* tested the model using two datasets: 31 steroid compounds and 35 sulfonamides. The corresponding correlation coefficient values ($r^2$) of 0.5776 (r = 0.76) and 0.5329 (r = 0.73) were achieved, respectively. Compared to other methods such as CoMFA [31], MS-WHIM [42] and few others on steroid dataset, SOMFA had the lowest standard deviation of errors of prediction (SDEP), which is the root-mean-square error of the predictions. In short, SOMFA is similar to CoMFA in terms of using grids and necessity of molecular alignment but is not as statistically rigorous as CoMFA [1],

as the SOMFA model is conceptually simple without heavy statistical elements such as partial least square (PLS).

### 3.4. Alignment-Free 3D-QSAR Methods

In the last few decades, other 3D-QSAR methods which do not rely on alignments were introduced. Some examples include autocorrelation of molecular surfaces properties (AMSP) [43], comparative molecular moment analysis (CoMMA) [44], WHIM (Weighted Holistic Invariant Molecular) method [45,46], Molecular surface (MS)-WHIM [42], and GRIND [47].

3.4.1. Autocorrelation of Molecular Surfaces Properties (AMSP)

Wagener *et al.* introduced the AMSP method to map the physical properties of ligands to a van der Waals surface and individual atoms, respectively. It uses a 3D descriptor based on spatial autocorrelation of molecular properties at distinct points on the molecular surface. The points are randomly distributed to have a continuous surface and the autocorrelation coefficient is obtained by summing the products of property values at various pairs of points at particular distances. For a series of distance intervals ($d_{lower}$, $d_{upper}$), a vector of autocorrelation coefficients is obtained as follows:

$$A(d_{lower}, d_{upper}) = \frac{1}{L} \sum_{ij} p_i p_j \qquad (d_{lower} < d_{ij} < d_{upper}) \tag{8}$$

where $p_i$ is the molecular property value at point *i*, $p_j$ is the molecular property value at point *j* and L is the total number of distances in the interval [43].

Therefore, the vector contains a compressed expression of the distribution of a property on the molecular surface. After autocorrelation vectors were obtained, a multilayer neural network was then trained using such vectors to derive a predictive model of biological activity of 31 steroid compounds. The correlation coefficient value, r, of 0.82 ($r^2 = 0.6724$) was achieved with a cross-validated $r^2$ of 0.63. In summary, the advantages of such autocorrelation vectors are the facts that they are shown to be invariant to translation and rotation since only spatial distances are used and have condensed description of molecular surface. However, original information cannot be reconstructed from such condensed vectors and the pharmacophore nature of a ligand may not be clear or interpretable [43].

3.4.2. Comparative Molecular Moment Analysis (CoMMA)

Silverman *et al.* [44] introduced the CoMMA method, which calculates the zeroth-, first-, and second-order spatial moments of the charge (such as quadrupolar moments) and the mass distribution (such as moments of inertia). Such molecular moment descriptors may be classified in three different categories: descriptors relating solely to molecular shape, descriptors relating only to molecular charge and descriptors relating to both shape and charge. The authors calculated 13 such descriptors and used them in partial least square analysis to generate predictive QSAR models for 31 steroid compounds. A range of statistical performance was obtained depending on different partial charge calculation methods used to derive electrostatic moments. Cross-validated $r^2$ values ranging from 0.412 to 0.828 were obtained using electrostatic moment descriptors calculated from Gasteiger charges or Guassian molecular orbital *ab initio* methods. The results showed that using quantum chemistry calculation-

based moments produced better predictive models than using only Gasteiger charge-based moments. Despite CoMMA's comparable statistical performances to CoMFA's, there are some limitations which may account for the limited number of published CoMMA applications. One reason is that the value of these descriptors, which measures the displacement between the center of mass and center of dipole with respect to the principal inertial axes, equals infinity for symmetric molecules whose dipole moment is zero [5].

3.4.3. Weighted Holistic Invariant Molecular (WHIM) Descriptor-Based QSAR

WHIM descriptors contain 3D molecular information such as molecular size, shape, symmetry and distribution of molecular surface point coordinates [45,46]. Molecular surface (MS)-WHIM is a WHIM-based 3D descriptor derived directly from molecular surface properties [42]. For WHIM descriptors, two types of matrices are defined: a molecular matrix containing cartesian coordinates of the *n* atoms and diagonal matrices containing the weights which are physicochemical properties associated with the *n* atoms of the molecule [42]. Each element of the diagonal matrix is defined as:

$$s_{jk} = \frac{\sum_{i=1}^{n} w_i \left(q_{ij} - \overline{q_j}\right)\left(q_{ik} - \overline{q_k}\right)}{\sum_{i-1}^{n} w_i} \tag{9}$$

where *n* is the number of atoms, $w_i$ is the weight of ith atom, $q_{ij}$ is the $j^{th}$ coordinate of the $i^{th}$ atom and $\overline{q_j}$ is the average of the $j^{th}$ coordinates [45].

In this expression, atoms can be weighted by mass, van der Waals volume, atomic electronegativity, electrotopological index of Kier and Hall, atomic polarizability and molecular electrostatic potential [33]. Elements in each diagonal matrix are subjected to principal component analysis (PCA) to obtain the scoring matrix, which is used to calculate PCA eigen values and eigen value proportion. Such values and proportions are then correlated with the molecular size and shape, respectively. One major advantage of the WHIM approach is that it provides a 3D QSAR descriptor which is invariant to translation and rotation of 3D molecular structures. In MS-WHIM, properties associated with the molecular surface points are used as different weighting schemes to compute statistical parameters. In particular, the unitary value and molecular electrostatic potential (MEP) are computed at each point of the Connolly molecular surface [48], and they are considered as weights. The unitary value contains information about the molecular surface shape and MEP provides the electrostatic information about the electron density distribution [42]. Although the WHIM approach is not sensitive to molecular orientation, MS-WHIM descriptor values are affected by the facts that the Connolly surface points are dependent on the 3D orientation of the molecule and indices for different weighting schemes are sensitive to surface point density [42]. The authors tested both WHIM and MS-WHIM on 31 steroid compounds and achieved the SDEP (standard deviation error of prediction) values of 1.750 and 0.742, respectively while CoMFA's SDEP was 0.837. The results suggested that MS-WHIM prediction performance was comparable to CoMFA's. SDEP was defined as follows:

$$SDEP = \sqrt{\frac{\sum \left(y_{pred} - y_{obs}\right)^2}{n}} \qquad (10)$$

WHIM/MS-WHIM descriptors are invariant to 3D molecular orientation but both methods, like other 3D-QSAR methods, rely on ligand conformation, which may be subjective if ligand-receptor co-crystal structures are not known for the target of interest.

3.4.4. Grid-Independent Descriptors (GRIND)-Based QSAR

In an attempt to provide alignment-free descriptors which are easy to understand and interpret, Pastor *et al.* introduced grid-independent descriptors [47]. The method utilizes specific probes such as the O probe (carbonyl oxygen) and N1 probe (amide nitrogen) to calculate molecular interaction fields (MIFs) at grid points. At each node of the grid, the energy between the probe and target ligand (E) is calculated as:

$$E = \sum E_{es} + \sum E_{hb} + \sum E_{lj} \qquad (11)$$

where $E_{es}$ is the electrostatic energy, $E_{hb}$ is the hydrogen-bonding energy, and $E_{lj}$ is the Lennard-Jones potential energy [49].

In this method, electrostatic interactions, hydrophobic interactions, hydrogen bond acceptor and hydrogen bond donor fields are considered to get a set of positions which defines a 'virtual receptor site' (VRS). VRS regions are then encoded into GRIND via an auto- and cross-correlation transform so that those regions are no longer dependent upon their positions in the 3D space. In other words, autocorrelation descriptors of the fields are calculated and only the highest products of molecular interaction energies are stored while others are discarded. This difference is responsible for the 'reversibility' of GRIND and the descriptors can be back-projected in 3D space using another related program called ALMOND [50]. The statistical performance of GRIND is comparable to other methods, but the advantage is that it is alignment-free and easy to interpret. However, bioactive conformations of ligands are valuable information to derive the virtual receptor site (VRS) and limitations on such information may affect final predictive models like other 3D methods.

*3.5. Multi-Dimensional (nD) QSAR Methods*

Multi-dimensional (nD) QSAR methods are essentially extensions of 3D-QSAR methods. These methods incorporate additional physical characteristics or properties (or a new dimension) to tackle the drawbacks of 3D-QSAR methods. One example is 4D-QSAR by Hopfinger *et al.* [51] which samples molecular conformations and alignments during the generation of a QSAR model. While incorporating some CoMFA features, it introduces the fourth dimension, which is the conformational Boltzmann sampling, and enables the method to be used as a receptor-independent (RI) method as well as receptor-dependent (RD) method in which the geometry of the receptor is known. It should be noted that their 4D-QSAR method does not solve the alignment problem but it allows a rapid evaluation of individual trial alignments [51]. Such 4D-QSAR implementation can be found in XMAP program [51,52]. Recently, it has been shown that 5D- and 6D-QSAR can be used for multiple representations of the receptor as well as its solvation states [53–55]. In the reported 5D-QSAR method, Vedani *et al.*

introduced a multiple representation of induced-fit hypotheses, *i.e.*, the adaptation of the receptor binding pocket to the individual ligand topology, as the fifth dimension. In other words, they generated a family of quasi-atomistic receptor surrogates [56] which are optimized by using a genetic algorithm. The binding energy was calculated as:

$$E_{binding} \approx E_{ligand\text{-}receptor} - E_{solvation,ligand} - T\Delta S - E_{internal\,strain} - E_{induced\,fit} \qquad (12)$$

where $E_{ligand\text{-}receptor}$ is the force field energy of the ligand-receptor interaction, $E_{solvation,ligand}$ is the ligand desolvation energy, $T\Delta S$ is the change in the ligand entrophy upon receptor binding, $E_{internal\,strain}$ is the change in ligand internal energy upon receptor binding, and $E_{induced\,fit}$ is the energy uptake required for adapting the receptor surrogate [54].

The 5D-QSAR method was tested on a set of 65 *NK-1* receptor antagonists and a set of 131 *Ah* receptor ligands, achieving predictive $r^2$ values of 0.837 and 0.832, while 4D-QSAR model resulted in 0.834 and 0.795, respectively [54]. They concluded that the binding affinities of new molecules were predicted more accurately with 5D-QSAR than with other lower dimension models. In the reported 6D-QSAR model, the simultaneous consideration of different solvation models was introduced by mapping parts of the surface area with different solvent properties [55]. 3D, 4D, 5D and 6D models were explored as comparison studies and the results showed the 6D-QSAR model produced the best predictive $r^2$ of 0.885 [55]. Both 5D- and 6D-QSAR methods are implemented in the *Quasar* and *VirtualToxLab* software [56,57].

## 4. Comparison of 2D or Fragment-Based QSAR *versus* 3D or nD-QSAR Methods

In general, the predictive quality of 3D-QSAR methods depends on several factors such as the quality of molecular alignments/superimpositions, and information on ligand bioactive conformations. Especially molecular superimpositions are subjective and ligand bioactive conformations always remain unclear when there is no structural information on the corresponding receptor-ligand complexes. Conventional CoMFA results may often be non-reproducible because the model depends on the orientation of alignment of molecules, which can be varied and subjective. Although various improved methods and other procedures, which were discussed earlier in the paper, have been introduced to overcome major limitations of 3D-QSAR methods, *i.e.*, subjective molecular alignment and bioactive conformation problems, many of them still require manual interventions and superimpositions [58,59]. From this prospect, 2D fragment-based QSAR methods have certain advantages over multi-dimensional QSAR methods since fragment-based or 2D-QSAR methods are simple and robust and do not require subjective (or time consuming) molecular alignment or putative binding conformation or determination of 3D structures. However, the disadvantage is that some of 2D-QSAR methods such as Hansch-Fujita method may provide non-unique solutions and the overall predictive quality may not be as good as some multi-dimensional methods which are computationally more complex and demanding. A summary of QSAR methods discussed in the paper is listed in Table 1. It should be noted that the performance of each QSAR model depends on the choice of dataset and different datasets can result in different predictive $q^2$ or $r^2$ or SDEP values.

**Table 1.** Summary of different QSAR methods and source information.

| Method | nD | Dataset | Statistical model | Performance | Reference/Website |
|---|---|---|---|---|---|
| HQSAR | 2D | 21 Steroids | PLS | $q^2 = 0.71$; $r^2 = 0.85$ [11] | [11] http://www.tripos.com |
| FB-QSAR | 2D | 48 NA analogs | IDLS | $r = 0.95$ ($r^2 = 0.91$) [19] | [19] |
| FS-QSAR | 2D | 85 bis-sulfone analogs; 83 COX2 analogs | MLR | $r^2 = 0.68$; $r^2 = 0.62$ [20] | [20] |
| TPF-QSAR | 2D | 282 pesticides | PM-based prediction | $r^2 = 0.75$ [23] | [23] |
| CoMFA | 3D | 21 Steroids 54 HIV-1PR inhibitors | PLS | $q^2 = 0.75$; $r^2 = 0.96$ [11] $q^2 = 0.68$; $r^2 = 0.69$ [60] | [31] http://www.tripos.com [60] |
| CoMSIA | 3D | Thermolysin inhibitors 54 HIV-1PR inhibitors | PLS | $q^2 = [0.59, 0.64]$ [32] $q^2 = 0.65$; $r^2 = 0.73$ [60] | [61,62] http://www.tripos.com [60] |
| Topomer CoMFA | 3D | 15 datasets from literature | PLS | average $q^2 = 0.636$ [40] | [40] http://www.tripos.com |
| SOMFA | 3D | 31 steroids; 35 sulfonamides | MLR | $r^2 = 0.58$; $r^2 = 0.53$ [41] | [41] |
| AMSP | 3D | 31 steroids | MNN | $q^2 = 0.63$; $r^2 = 0.67$ [43] | [43] |
| CoMMA | 3D | 31 steroids | PLS | $q^2 = [0.41, 0.82]$ [44] | [44] |
| WHIM | 3D | 31 steroids | PCA | SDEP = 1.750 [42] | [45] http://www.vcclab.org/lab/indexhlp/whimdes.html |
| MS-WHIM | 3D | 31 steroids | PCA | SDEP = 0.742 [42] | [42] |
| GRIND | 3D | 31 steroids 175 hERG inhibitors | PLS; PCA PLS; SVM | $q^2 = 0.64$; SDEP = 0.26 [47] $q^2 = 0.41$; $r^2 = 0.57$; SDEP = 0.72 [63] | [47] http://www.moldiscovery.com/soft_grid.php [63] |

**Table 1.** *Cont.*

| Method | nD | Dataset | Statistical model | Performance | Reference/Website |
|--------|----|---------|-------------------|-------------|-------------------|
| 4D-QSAR | 4D | 20 DHFR inhibitors;<br>42 PGF$_2$a analogs;<br>40 2-substituted dipyridodiazepione inhibitors<br>33 p38-MAPK inhibitors | PLS<br>GL-PLS | $r^2 = [0.90, 0.95]$;<br>$r^2 = [0.73, 0.86]$;<br>$r^2 = [0.67, 0.76]$ [51]<br>$q^2 = [0.67, 0.85]$ [64] | [51] http://www.seascapelearning.com/4DsgiSW/<br>[64] |
| 5D-QSAR | 5D | 65 NK-1 antagonists;<br>131 Ah ligands | MLR | $r^2 = 0.84$;<br>$r^2 = 0.83$ [54] | [54]<br>http://www.biograf.ch |
| 6D-QSAR | 6D | 106 estrogen receptor ligands | MLR | $q^2 = 0.90$;<br>$r^2 = 0.89$ [55] | [55]<br>http://www.biograf.ch |
| HQSAR = Hologram QSAR<br>FB-QSAR = Fragment-based QSAR<br>FS-QSAR = fragment-similarity-based QSAR<br>TPF-QSAR = Top priority fragment QSAR<br>CoMFA = Comparative molecular field analysis<br>CoMSIA = Comparative molecular similarity indices analysis<br>SOMFA = Self-organizing molecular field analysis<br>AMSP = Autocorrelation of molecular surface properties<br>CoMMA = Comparative molecular moment analysis<br>WHIM = Weighted holistic invariant molcular QSAR<br>MS-WHIM = Molecular surface WHIM<br>GRIND = Grid independent descriptor | | | PLS = Partial least square<br>IDLS = Iterative double least square<br>PM = Priority matrix<br>MNN = Multilayer neural networks<br>MLR = Multiple linear regression<br>PCA = Principal component analysis | | $q^2$ = cross-validated $r^2$<br>SDEP = standard deviation of errors of prediction |

## 5. Conclusion

We have provided an overview of different QSAR methods and recent development in fragment-based approaches using selected studies as an illustration. Since each QSAR method has its own advantages and disadvantages, researchers should choose appropriate methods for modeling their systems. However, given a wide range of choices, it is a challenging task to pick appropriate models for one's studies. This paper outlines many basic principles of new fragment-based QSAR methods as well as other 3D- and nD- QSAR models and illustrates some examples which may be helpful references to many researchers.

## Acknowledgement

## References

1. Esposito, E.X.; Hopfinger, A.J.; Madura, J.D. Methods for applying the quantitative structure-activity relationship paradigm. *Methods Mol. Biol.* **2004**, *275*, 131–213.
2. Bradbury, S.P. Quantitative structure-activity relationships and ecological risk assessment: an overview of predictive aquatic toxicology research. *Toxicol. Lett.* **1995**, *79*, 229–237.
3. Hansch, C.; Leo, A. *Exploring QSAR—Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC., USA, 1995.
4. Hansch, C.; Telzer, B.R.; Zhang, L. Comparative QSAR in toxicology: Examples from teratology and cancer chemotherapy of aniline mustards. *Crit. Rev. Toxicol.* **1995**, *25*, 67–89.
5. Perkins, R.; Fang, H.; Tong, W.; Welsh, W. Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology. *Environ. Toxicol. Chem.* **2003**, *22*, 1666–1679.
6. Chen, J.-Z.; Han, X.-W.; Liu, Q.; Makriyannis, A.; Wang, J.; Xie, X.-Q. 3D-QSAR studies of arylpyrazole antagonists of cannabinoid receptor subtypes CB1 and CB2. A combined NMR and CoMFA approach. *J. Med. Chem.* **2006**, *49*, 625–636.
7. Salum, L.; Andricopulo, A. Fragment-based QSAR: Perspectives in drug design. *Mol. Divers.* **2009**, *13*, 277–285.
8. Free, S.J.; Wilson, J. A mathematical contribution to structure-activity studies. *J. Med. Chem.* **1964**, *7*, 395–399.
9. Hansch, C.J.; Fujita, T. $\rho-\sigma-\pi$ Analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616–1626.
10. SYBYL8.0. In *Discovery Software for Computational Chemistry and Molecular Modeling*; Tripos: St. Louis, MO, USA, 2008.
11. Lowis, D. *HQSAR: A New, Highly Predictive QSAR Technique*, *Tripos Technique Notes*; Tripos: St. Louis, MO, USA, 1997.
12. Castilho, M.S.; Postigo, M.P.; de Paula, C.B.V.; Montanari, C.A.; Oliva, G.; Andricopulo, A.D. Two- and three-dimensional quantitative structure-activity relationships for a series of purine nucleoside phosphorylase inhibitors. *Bioorg. Med. Chem.* **2006**, *14*, 516–527.

13. Salum, L.B.; Polikarpov, I.; Andricopulo, A.D. Structural and chemical basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies. *J. Mol. Graph. Model.* **2007**, *26*, 434–442.

14. Honorio, K.M.; Garratt, R.C.; Andricopulo, A.D. Hologram quantitative structure-activity relationships for a series of farnesoid X receptor activators. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3119–3125.

15. Castilho, M.S.; Guido, R.V.C.; Andricopulo, A.D. 2D Quantitative structure-activity relationship studies on a series of cholesteryl ester transfer protein inhibitors. *Bioorg. Med. Chem.* **2007**, *15*, 6242–6252.

16. Lo Piparo, E.; Koehler, K.; Chana, A.; Benfenati, E. Virtual Screening for aryl hydrocarbon receptor binding prediction. *J. Med. Chem.* **2006**, *49*, 5702–5709.

17. Tropsha, A.; Golbraikh, A. Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Curr. Pharm. Design* **2007**, *13*, 3494–3504.

18. Prakash, O.; Ghosh, I. Developing an antituberculosis compounds database and data mining in the search of a motif responsible for the activity of a diverse class of antituberculosis agents. *J. Chem. Inf. Model.* **2005**, *46*, 17–23.

19. Du, Q.-S.; Huang, R.-B.; Wei, Y.T.; Pang, Z.-W.; Du, L.-Q.; Chou, K.-C. Fragment-based quantitative structure-activity relationship (FB-QSAR) for fragment-based drug design. *J. Comput. Chem.* **2009**, *30*, 295–304.

20. Myint, K.Z.; Ma, C.; Wang, L.; Xie, X.Q. The Fragment-similarity-based QSAR (FS-QSAR): A Novel 2D-QSAR method to predict biological activities of triaryl Bis-sulfone and COX2 analogs. **2010**, *submitted*

21. Burden, F. Molecular identification number for substructure searches. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 225–227.

22. Xie, X.Q.; Chen, J. Data-mining a small molecule drug screening representative subset from NIH PubChem database. *J. Chem. Inf. Model.* **2008**, *48*, 465–475.

23. Casalegno M.; Sello G. Benfenati E: Top-Priority Fragment QSAR Approach in Predicting Pesticide Aquatic Toxicity. *Chem. Res. Toxicol.* **2006**, *19*, 1533–1539.

24. Zhokhova, N.; Baskin, I.; Palyulin, V.; Zefirov, A.; Zefirov, N. Fragmental descriptors with labeled atoms and their application in QSAR/QSPR studies. *Doklady Chem.* **2007**, *417*, 282–284.

25. Ford, M.G. *Euroqsar 2002 Designing Drugs and Crop Protectants: Processes, Problems, and Solutions*; Blackwell: Melbourne, Australia, 2003.

26. Andrade, C.; Salum, L.B.; Castilho, M.; Pasqualoto, K.; Ferreira, E.; Andricopulo, A. Fragment-based and classical quantitative structure–activity relationships for a series of hydrazides as antituberculosis agents. *Mol. Divers.* **2008**, *12*, 47–59.

27. *Dragon*, v5.4; Talete_Srl: Milan, Italy, 2008.

28. Oliveira, D.B.D.; Gaudio, A.C. BuildQSAR: A new computer program for QSAR analysis. *Quant. Struct.-Act. Relation.* **2000**, *19*, 599–601.

29. *Pirouette Multivariate Data Analysis for IBM PC Systems*; Infometrix: Seattle, WA, USA, 2001.

30. Tsygankova, I.; Zhenodarova, S. Quantitative structure-activity relationship for barbituric acid derivatives: Potential of the fragment approach. *Russ. J. Gene. Chem.* **2007**, *77*, 940–928.

31. Cramer, R.; Patterson, D.; Bunce, J. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.

32. Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.

33. Dudek, A.Z.; Arodz, O.; Galvez, J. Computational methods in developing quantitative structure-activity relationships (QSAR): A review. *Comb. Chem. High T. Scr.* **2006**, *9*, 213–228.

34. Cramer, R.D.; Cruz, P.; Stahl, G.; Curtiss, W.C.; Campbell, B.; Masek, B.B.; Soltanshahi, F. Virtual screening for r-groups, including predicted pIC50 contributions, within large structural databases, using topomer CoMFA. *J. Chem. Inf. Model.* **2008**, *48*, 2180–2195.

35. Avram, S.; Milac, A.L.; Flonta, M.L. Computer-aided drug design for typical and atypical antipsychotic drugs: A review of application of QSAR and combinatorial chemistry methods - tools for new antipsychotics design. *Curr. Comput.-Aided Drug Design* **2005**, *1*, 347–364.

36. Patcharawee, N.; Nahoum, G.A.; Blair, F.J.; Simon, P.M.; Jiraporn, U. 3D-QSAR studies on chromone derivatives as HIV-1 protease inhibitors: Application of molecular field analysis. *Arch. Pharm.* **2008**, *341*, 357–364.

37. Labrie, P.; Maddaford, S.P.; Fortin, S.; Rakhit. S.; Kotra, L.P.; Gaudreault, R.C. A comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) of anthranilamide derivatives that are multidrug resistance modulators. *J. Med. Chem.* **2006**, *49*, 7646–7660.

38. Jeong, J.A.; Cho, H.; Jung, S.Y.; Kang, H.B.; Park, J.Y.; Kim, J.; Choo, D.J.; Lee, J.Y. 3D QSAR studies on 3,4-dihydroquinazolines as T-type calcium channel blocker by comparative molecular similarity indices analysis (CoMSIA). *Bioorg. Med. Chem. Lett.* **2010**, *20*, 38–41.

39. Dayan, F.E.; Singh, N.; McCurdy, C.R.; Godfrey, C.A.; Larsen, L.; Weavers, R.T.; van Klink, J.W. Perry NB: β-triketone inhibitors of plant p-hydroxyphenylpyruvate dioxygenase: Modeling and comparative molecular field analysis of their interactions. *J. Agric. Food Chem.* **2009**, *57*, 5194–5200.

40. Cramer, R.D. Topomer CoMFA: A design methodology for rapid lead optimization. *J. Med. Chem.* **2003**, *46*, 374–388.

41. 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.

42. Bravi, G.; Gancia, E.; Mascagni, P.; Pegna, M.; Todeschini, R.; Zaliani, A. MS-WHIM, new 3D theoretical descriptors derived from molecular surface properties: A comparative 3D QSAR study in a series of steroids. *J. Comput.-Aided Mol. Design* **1997**, *11*, 79–92.

43. Wagener, M.; Sadowski, J.; Gasteiger, J. Autocorrelation of molecular surface properties for modeling corticosteroid binding globulin and cytosolic ah receptor activity by neural networks. *J. Am. Chem. Soc.* **1995**, *117*, 7769–7775.

44. Silverman, B.D.; Platt, D.E. Comparative molecular moment analysis (CoMMA): 3D-QSAR without molecular superposition. *J. Med. Chem.* **1996**, *39*, 2129–2140.

45. Todeschini, R.; Gramatica, P. New 3D molecular descriptors: the WHIM theory and QSAR applications. In *3D QSAR in Drug Design*; Kubinyi, H., Folkers, G., Martin, Y.C., Eds.; Kluwer Academic Publishers/Escom: Dordrecht, The Netherlands, 1998; Volume 2, pp. 355–380.

46. Todeschini, R.; Lasagni, M. Marengo E: New molecular descriptors for 2D and 3D structures. *J. Chemometrics* **1994**, *8*, 263–272.

47. Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRid-INdependent descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *43*, 3233–3243.

48. Connolly, M. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548–558.

49. Pastor, M.; Cruciani, G.; Watson, K.A. A strategy for the incorporation of water molecules present in a ligand binding site into a three-dimensional quantitative structure-activity relationship analysis. *J. Med. Chem.* **1997**, *40*, 4089–4102.

50. ALMOND: Molecular Discovery Ltd.: Perugia, Italy, Available at: http://www.moldiscovery.com/soft_almond.php (accessed on 25 September 2010).

51. 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.

52. Scheiber, J.; Enzensperger, C.; Lehmann, J.; Stiefl, N.; Baumann, K. Alignment-free 4D-QSAR: Applying the XMAP technique in prospective analyses. In *QSAR & Molecular Modeling in Rational Design of Bioactive Molecules*; Aki-Sener, E., Yalcin, I., Eds.; CADDD Society: Ankara, Turkey, 2006.

53. Fischer, P.M. Computational chemistry approaches to drug discovery in signal transduction. *Biotechnol. J.* **2008**, *3*, 452–470.

54. Vedani, A.; Dobler, M. 5D-QSAR: The key for simulating induced fit? *J. Med. Chem.* **2002**, *45*, 2139–2149.

55. Vedani, A.; Dobler, M.; Lill, M.A. Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *J. Med. Chem.* **2005**, *48*, 3700–3703.

56. Vedani, A.; Dobler, M.; Zbinden, P. Quasi-atomistic receptor surface models: A bridge between 3-D QSAR and receptor modeling. *J. Am. Chem. Soc.* **1998**, *120*, 4471–4477.

57. Biograf: VirtualToxLab; Basel, Switzerland, 2009. Available at: http://www.biograf.ch/downloads/VirtualToxLab.pdf (accessed on 25 September 2010).

58. Hillebrecht, A.; Klebe, G. Use of 3D QSAR models for database screening: A feasibility study. *J. Chem. Inf. Model.* **2008**, *48*, 384–396.

59. Matter, H.; Potter, T. Comparing 3D pharmacophore triplets and 2D fingerprints for selecting diverse compound subsets. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1211–1225.

60. Khedkar, V.; Ambre, P.; Verma, J.; Shaikh, M.; Pissurlenkar, R.; Coutinho, E. Molecular docking and 3D-QSAR studies of HIV-1 protease inhibitors. *J. Mol. Model.* **2010**, *16*, 1251–1268.

61. Li, Q.; Jørgensen, F.S.; Oprea, T.; Brunak, S.; Taboureau, O. hERG classification model based on a combination of support vector machine method and GRIND descriptors. *Mol. Pharm.* **2008**, *5*, 117–127.

62. Romeiro, N.; Albuquerque, M.; Alencastro, R.; Ravi, M.; Hopfinger, A. Construction of 4D-QSAR models for use in the design of novel p38-MAPK inhibitors. *J. Comput.-Aided Mol. Design* **2005**, *19*, 385–400.