Published in final edited form as:
Curr Comput Aided Drug Des. 2011 March 1; 7(1): 10–22.

Recent Advances in Ligand-Based Drug Design: Relevance and Utility of the Conformationally Sampled Pharmacophore Approach

Chayan Acharya[†], **Andrew Coop**, **James E. Polli**, and **Alexander D. MacKerell Jr.**[*]
Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, 20 Penn Street, Baltimore, MD 21201, USA

Abstract

In the absence of three-dimensional (3D) structures of potential drug targets, ligand-based drug design is one of the popular approaches for drug discovery and lead optimization. 3D structure-activity relationships (3D QSAR) and pharmacophore modeling are the most important and widely used tools in ligand-based drug design that can provide crucial insights into the nature of the interactions between drug target and ligand molecule and provide predictive models suitable for lead compound optimization. This review article will briefly discuss the features and potential application of recent advances in ligand-based drug design, with emphasis on a detailed description of a novel 3D QSAR method based on the conformationally sample pharmacophore (CSP) approach (denoted CSP-SAR). In addition, data from a published study is used to compare the CSP-SAR approach to the Catalyst method, emphasizing the utility of the CSP approach for ligand-based model development.

Keywords

CoMFA; computer-aided drug design; CoMSIA; CSP; drug discovery; lead optimization; pharmacophore

Introduction

Computer-aided drug design (CADD) is a very useful tool in rational drug design to minimize the time for identification, characterization and structure-optimization for novel drug candidates [1-5]. CADD can also be useful for rational design of prodrugs. Prodrugs are typically designed to increase the specificity or bioavailability of the original drug molecules [6-8]. Ligand-based drug design is an indirect approach to facilitate the development of pharmacologically active compounds by studying molecules that interact with the biological target of interest [9]. In contrast, structure-based drug design methods directly use knowledge of the 3D structure of the target molecule to identify or optimize drug candidates [10-12].

Step one of any drug design process is identification of a suitable target molecule associated with a disease. Usually a key protein of a biochemical pathway associated with the disease state serves as a potential drug target [6,13,14]. Depending on the nature of the disease state, molecules, referred to as lead compounds, are identified or designed to inhibit or promote the

[*] Author to whom correspondence should be addressed; Tel.: +1-410-706-7442; Fax: +1-410-706-5017 c.acharya@iecb.u-bordeaux.fr; alex@outerbanks.umaryland.edu.
[†] Present Address: Institut Européen de Chimie et Biologie, Université Bordeaux 1, 2 rue Robert Escarpit, 33607 Pessac Cedex, France

concerned biochemical pathway [15-18]. The next step in the drug discovery process is to optimize the lead molecules to maximize the interaction with the target molecule. CADD can play a crucial role to guide the lead optimization process.

CADD methods can be applied to both ligand-based as well as structure-based drug design. Ligand-based drug design methods are useful in the absence of an experimental 3D structure [19-22]. Due to the lack of an experimental structure, the known ligand molecules that bind to the drug target are studied to understand the structural and physico-chemical properties of the ligands that correlate with the desired pharmacological activity of those ligands [9]. Besides known ligand molecules, ligand-based methods may also include natural products or substrate analogues that interact with the target molecule yielding the desired pharmacological effect [23-27]. Alternatively, in the presence of a 3D structure of the drug target, structure-based methods, such as molecular docking or *in silico* chemical alteration, are usually applied for lead optimization [10,11]. This approach takes advantage of the availability of the target 3D structure to identify the nature of the target-ligand interaction and the structural requirements of the ligand to optimize the interaction.

The present article includes a brief overview of ligand-based modeling approaches followed by recent developments in ligand-based optimization methodologies, with emphasis on the conformationally-sampled pharmacophore (CSP) SAR method (CSP-SAR) developed in our laboratories.

Basics of QSAR

The most popular approaches for ligand-based drug design are the QSAR method and pharmacophore modeling. QSAR is a computational method to quantify the correlation between the chemical structures of a series of compounds and a particular chemical or biological process. The underlying hypothesis behind QSAR method is that similar structural or physiochemical properties yield similar activity [28,29]. Initially a group of chemical entities or lead molecules are identified which show the desired biological activity of interest. A quantitative relationship is established between the physico-chemical features of the active molecules and the biological activity. The developed QSAR model is then used to optimize the active compounds to maximize the relevant biological activity. The predicted compounds are then tested experimentally for the desired activity. The QSAR method thus can be used as a guiding tool for identification of compound modifications with improved activity.

The general methodology of QSAR is built upon a series of consecutive steps (Figure 1):

(1) Identify ligands with experimentally measured values of the desired biological activity. Ideally these ligands are of a congeneric series but should be of adequate chemically diversity to have a large variation in activity.

(2) Identify and determine molecular descriptors associated with various structural and physico-chemical properties of the molecules under study.

(3) Discover correlations between molecular descriptors and the biological activity that can explain the variation in activity in the data set.

(4) Test the statistical stability and predictive power of the QSAR model.

Depending on the goal of the study, the appropriate biological activity is experimentally measured for a series of compounds and this data serves as the dependent variable in QSAR modeling. Once the molecules are selected for the study they are modeled in silico and energy minimized using molecular mechanics or quantum mechanical methods [21,30-33]. Next, relevant molecular descriptors are generated for the set of molecules to describe the chemical features of the molecules that are required for their biological activity. Molecular descriptors can be structural as well as physico-chemical. The goal here is to create a molecular “fingerprint” for each molecule that relates to its activity. Depending on the QSAR method, knowledge-based, molecular mechanical or quantum chemical tools can be used to generate the molecular descriptors. Molecular descriptors are then used to develop a mathematical relation that can explain the variability of the biological activity of the molecules. In the final step, the developed models are subjected to various internal and external validation procedures to test their statistical significance, robustness and predictive power.

Over the years the strategies to execute these steps have evolved to make the QSAR technique an essential part of the drug optimization process. The advancement of QSAR methodology primarily occurred in the range of molecular descriptors applied and how they are related to the activity. The remainder of this review will give an overview of the major QSAR methodologies, emphasizing their key differences, followed by a more detailed description of the CSP-SAR method developed in our laboratories.

Statistical Tools for Model Development and Validation

The success of any QSAR model greatly depends on the choice of molecular descriptors and the ability to generate the appropriate mathematical relationship between the descriptors and the biological activity of interest. Since the early days of QSAR it was clear that the definition of molecular descriptors is the crucial part of the method [28,34]. Recent software developments now allow for generation of large numbers of molecular descriptors that can be used for QSAR methods [35,36]. This also poses a new problem in selection of appropriate descriptors to explain the activity data [34,37]. There are three major statistical methods traditionally applied in linear QSAR methods to select molecular features important for activity:

(i) Multivariable linear regression analysis (MLR)

(ii) Principal component analysis (PCA)

(iii) Partial least square analysis (PLS)

MLR analysis is the simplest method to quantify the molecular descriptor having good correlation with the variation in activity. MLR model development can involve forward or backward stepwise regression according to a statistical test to find the best model (*i.e.* systematically adding or eliminating molecular descriptors, respectively, to determine the ideal model). However for large numbers of descriptors the MLR method can be time consuming and the user needs to be careful to exclude the variable combinations with high internal correlation. Nevertheless, this problem can be solved by using statistical software (such as MATLAB [38], R [39]) where the user can automate the MLR process with adequate conditions. The PCA method was designed to overcome the problems of MLR analysis by extracting information from multiple, possibly redundant variables into a smaller number of uncorrelated variables [40,41]. Thus PCA provides an efficient way for reduction of the number of independent variables used in QSAR models. This method is highly useful for systems with a larger number of molecular descriptors than the number of observations. However, results from PCA are often difficult to analyze with respect to identification of the particular structural or physicochemical characteristics important for activity. PLS is a combination of MLR and PCA techniques where the dependent variable (e.g. the biological activity) is also extracted into new components to optimize the correlation [42]. PLS is advantageous for systems with more than one dependent variable. There are also other variable selection methods available such as genetic algorithms and Bayesian methods used in linear QSAR models; these have been discussed in detail in other review articles [34,43,44].

Biological systems often display non-linear relationship between the molecular descriptors and the activity [45-49]. Neural network is one of the most popular non-linear regression methods used to develop QSAR models for such systems [34,50-56]. Neural network supported QSAR modeling is based on the self-learning property of the neural network. In this method the network learns the association between the molecular descriptors and the related biological activity based on the training set of ligand molecules. A “trained” neural network is also able to perform back-propagation to predict the activity of a new molecule given its structural and physico-chemical properties. Like any regression method neural networks are also susceptible to overfitting and can result in models with poor predictive power. The detection of the optimized architecture of the neural network can also be subjective and time-consuming [34]. A variation of neural network is the Bayesian regularized neural network method which can be used to model nonlinear QSAR data [47,57-60]. This method is advantageous over the regular neural network as it is able to overcome the overfitting problem and it can automatically optimize the neural architecture [34]. Recently a modified Bayesian regularized artificial neural network (BRANN) with a Laplacian prior has also been reported which is able to optimize the number of descriptors used in QSAR models by pruning out ineffective descriptors [61].

Once an initial QSAR model has been developed is must be validated. There are mainly two types of validation required to establish a QSAR model: (i) internal validation, and (ii) external validation. The most popular internal validation method is leave-one-out cross validation [62]. In this method one of the observations is kept as validation data while the rest of the data comprises the training set to estimate the coefficients of the QSAR model. The activity of the test compound is then predicted using the model based on the training set compounds. This procedure is repeated for all other compounds until each of them served once as a test compound. The predictive power of the model is then assessed by calculating the cross-validated r^2 or Q^2 using the following equation:

$$Q^2 = 1 - \frac{\Sigma\left(y_{pred} - y_{obs}\right)^2}{\Sigma(y_{obs} - y_{mean})^2} \qquad (1)$$

The drawback of this method is that the time to perform the calculation increases with the square of the size of the training set. Another variation of this method is *k*-fold cross validation [63]. Instead of leaving one compound out, this method forms the training set by leaving a subset of *k* number of molecules out of it at a time. This method is able to reduce the time of calculation by diminishing the size of the effect training set. However, the *k*-fold method is subjective to the choice of the value of *k*. In addition, both of the cross validation methods fail to use all available data at the same time for model validation. External validation method involves predicting the activity of the test set molecules which were not used for model development [64]. This method is very rigorous as it can use all available data from the training set for model validation.

Classical or 2D QSAR

In classical QSAR method, also known as the Hansch-Fujita approach, various electronic, hydrophobic and steric features are correlated with the biological activity for a congeneric series of compounds [65,66]. A typical example of an equation relating activity to physical properties is shown in equation 2,

$$\log\left(\frac{1}{C}\right) = k_1\pi - k_2\pi^2 + k_3\sigma + k_4E_s + k_5 \qquad (2)$$

where C is the concentration of the compound required to produce the biological activity, π is hydrophobic substituent constant (i.e. partition coefficient), σ is Hammett electronic substituent constant and E_s is the steric substituent constant. Hansch analysis is also known as 2D QSAR as it usually involves 2D molecular descriptors. In 1964 Free-Wilson independently developed a mathematical model relating the presence of various chemical substituents to biological activity [67]. Each type of chemical group was assigned an activity contribution, depending on the type and location of a substituent (such as *meta* or *para*) and the related impact on biological activity. Biological activity of a substituted compound was then estimated as a summation of the activity of the parent molecule, μ, and the contribution from the individual substituents (Eqn. 3)

$$\log\left(\frac{1}{C}\right)=\sum_{ij} a_{ij}+\mu \tag{3}$$

where a_{ij} represents the activity contribution from the substitution i at location j. Both the Hansch and Free-Wilson approaches served as predictive tools in classical QSAR studies for many years [68]. Later a combined Hansch/Free-Wilson method was developed where equations 2 and 3 were linearly combined to describe the biological activity [68,69]. The strength of the classical QSAR is that by using very simplistic mathematical relations involving various physico-chemical properties and chemical substituents it is able to explain and predict biological activity of a series of similar molecules.

The molecular descriptors used for correlation with the activity were mostly representative of fragments of the parent molecule, including substituents on the parent. The advantage of using fragment-based descriptors is their ready availability for a wide-range of substituents, computational ease and the ability to keep the mathematical implementation fairly understandable [70]. Accordingly, classical QSAR is quite effective for congeneric series of molecules. At the same time, a major drawback of these approaches is that their application is indeed limited to a congeneric series. In these methods the quality of the activity prediction rapidly decreases as the physical properties of new functional group varies further from that included in the training set. Also the use of fragment-based descriptors is usually inadequate to capture the 3D conformational features of the molecule crucial for its activity [20,21].

3D QSAR

As the name suggests, 3D QSAR method includes descriptors that describe 3D features of a molecule to develop a QSAR model. Various geometric, physical characteristics and quantum chemical descriptors may be used to describe the 3D features of the ligands in the 3D QSAR method. Such molecular descriptors are then combined to create a pharmacophore that can explain the biological activity of the ligands. A pharmacophore is defined as the 3D spatial orientation of various features, such as hydrogen bond donors or acceptors, which are essential for the desired biological activity [71-73]. The developed pharmacophore model is tested for the stability and statistical significance to obtain the final 3D QSAR model.

There are several review articles available that elaborately discuss various techniques of 3D QSAR modeling [2,3,9,19,[28],[34],70,71,74-76]. To avoid redundancy, the following section will briefly describe the major 3D QSAR techniques currently in use for drug design. The concluding section will provide a detailed description of the CSP-SAR method developed in our laboratory along with applications of the method.

CoMFA

Comparative Molecular Field Analysis (CoMFA) [77] is one of the most widely used 3D QSAR methods. CoMFA was the first QSAR method to relate 3D shape-dependent steric and electrostatic properties of a molecule to its biological activity. In this method the molecules are aligned based on their 3D structures on a 3D grid and the values of steric and electrostatic potential energies are calculated at each grid point. Usually CoMFA assumes that the minimum-energy conformer is the bioactive conformer. For systems with known crystal structures, crystal coordinates may be used to define bioactive conformers. Field values corresponding to the potential energy terms are calculated at each grid point for every molecule and correlated with the biological activity. PCA or PLS methods are usually used for model development in CoMFA. The CoMFA model is then tested for statistical significance and robustness. The success and predictive ability of CoMFA models are highly sensitive on the alignment of the bioactive conformers [28,78-80]. As the bioactive conformation is not necessarily the lowest-energy conformation in the absence of the receptor [81-83], the assumption made by CoMFA in the selection of bioactive conformers and the corresponding alignment method may produce erroneous models. By neglecting the dynamical nature of the ligands CoMFA limits its applicability. Another limitation of CoMFA is the form of its energy function, as it does not explicitly account for hydrophobicity or hydrogen bond interactions [28,80,84]. CoMFA uses Lennard-Jones and Coulombic potential function to calculate steric and electrostatic interaction, respectively, which can cause unrealistically high values for these energy terms due to the hyperbolic natures of the energy functions. An arbitrary cutoff value for these potential functions is assigned in CoMFA to avoid such behavior [85,86].

CoMSIA

Comparative Molecular Similarity Indices (CoMSIA) [84] is a 3D QSAR technique similar to CoMFA. However, unlike CoMFA, the molecular field expression of CoMSIA includes hydrophobic, hydrogen-bond donor and acceptor terms in addition to steric and coulombic contributions. CoMSIA also calculates the similarity indices instead of interaction energies by comparing each ligand molecule with a common probe with a radius of 1Å, and charge, hydrophobicity and hydrogen bond properties equal to 1 [87]. CoMSIA uses bell-shaped Gaussian function to describe steric, electrostatic and hydrophobic components of the energy function. Unlike CoMFA, this allows CoMSIA to avoid the use of an arbitrary cutoff value for the energy calculations. Similarity indices corresponding to CoMSIA molecular fields define the ligand-protein binding interaction [88].

CATALYST

Efforts have been made to include the conformational flexibility in 3D QSAR methodology. CATALYST [89] is one of the most popular 3D QSAR software packages that use conformational variation during model development. CATALYST uses the poling algorithm [90] to sample conformational space for the ligand molecules. Typically 250 conformers are generated in this process with a default cutoff value of 20 kcal/mol above the energy of global minimum conformation. Spatial orientations of the functional groups are used to develop the pharmacophore hypothesis and the estimated and observed activity values are compared to evaluate the QSAR models. The most common properties or functional groups used to define the pharmacophoric features are:

(i) Hydrogen-bond acceptor

(ii) Hydrogen-bond donor

(iii) Positively charges group (basic)

(iv) Negatively charged group (acidic)

(v) Aromatic ring

(vi) Aliphatic hydrophobic moieties

(vii) Aromatic hydrophobic moieties

The pharmacophore generation process is divided into constructive and subtractive phases. During the constructive phase compounds having activity greater than a cutoff value are used to build a pharmacophore hypothesis. In the subtractive phase any pharmacophore that fits more than half of the inactive compounds is rejected. A cost value is assigned to each selected pharmacophore based on its prediction error, feature weight and complexity. CATALYST is equipped to overcome most of the drawbacks of the previous 3D QSAR methods. However, there are a few limitations in CATALYST. The conformation generator of CATALYST creates a maximum of 250 conformers which may not include all possible accessible conformations for flexible ligand molecules. Hence, CATALYST may fail to include the bioactive conformer of the active compounds, which in turn may lead to incorrect pharmacophore models. CATALYST also does not generate models that include both the physico-chemical properties and pharmacophoric features.

CSP-SAR

Principle

CSP-SAR is a novel method for developing 3D QSAR models based on the Conformationally Sampled Pharmacophore (CSP) method developed in our laboratory [21,91,92]. This method is designed for ligands with conformational flexibility and avoids problems associated with ligand alignment. As discussed before, the active or bound conformation of a ligand molecule need not to be the lowest-energy conformation in the absence of the target molecule; even the active conformer may not belong to the ensemble of low energy conformers [93]. In order to maximize the potential inclusion of the bioactive conformers of the ligand molecules in the model, a rigorous sampling of the conformational space for each ligand is essential. Unlike other pharmacophore development methods, CSP considers all accessible conformations of each ligand molecule for pharmacophore development. Thus, CSP maximizes the probability of including the bioactive conformer in the model.

Once all accessibly conformations of the ligands have been generated, typically via molecular dynamics simulations (see below), it is necessary to extract descriptors of the conformational properties of the ligands for use in model development. This requires the selection of pharmacophore features for the set of ligand molecules, and represents a critical step in the CSP approach. Typically, these features include hydrogen bond donor and acceptors, hydrophobic groups or any other structural feature that may be important for the biological activity. Available SAR data on the system of interest may guide but not limit this selection procedure. For example, the CSP approach was successfully applied to opioids using previously defined functional groups, such as the basic nitrogen known to be essential for opioid activity, as well as identifying novel functional groups during model development [21,91,92]. Studies involving relatively large ligands that have not previously been subjected to significant SAR studies pose more difficulty in the functional group selection process. In the absence of any existing model functional group selection involves the user considering all functional groups that might have any effect on the biological activity. After the identification of all possible chemical features that may serve as pharmacophoric points, all possible distances, angles and dihedral angles between the feature points need to be considered. Once these descriptors have all been identified they are regressed against each other to eliminate redundant descriptors from further analysis (e.g. if descriptors have a correlation coefficient (r^2) greater than, for example, 0.8 one of those may be removed from further consideration). The remaining descriptors are then systematically regressed against the biological data, with those having correlation coefficient (r^2) less than a cutoff value (typically 0.01) with respect to the biological data discarded from further analysis. Initially an extensive calculation of the structural descriptors is highly recommended. For relatively large ligand molecules the number of possible structural descriptors can be quite high, being on the order of 100,000 or more. However, automation of this procedure readily allows the selection of descriptors for additional analysis to be performed.

The nature the descriptors, which include selected pharmacophore features in combination with all accessible conformations of each ligand, is the key feature of the CSP approach. This combination requires that the descriptors be treated as probability distributions that include, for example, all possible distances between two pharmacophore features or all possible angles between three pharmacophore features, and so on. To better elucidate this concept, we will expand on published results of a CSP study of bile acid conjugates and their transporter (Apical Sodium-dependent Bile acid Transporter or ASBT) [20]. Presented in Figure 2 are three conjugates of the bile acid **9**, **2** and **21**, on which three pharmacophore points are shown (note that in the original study, a total of 30 pharmacophore points were initially considered on a total of 13 compounds). For the present example, three conjugates shown in Figure 2 will be considered. Each of these conjugates was subjected to MD simulations to obtain all possible conformations from which probability distributions of descriptors based on the pharmacophore features in Figure 3 were determined. One-dimensional descriptors associated with the NG-OA distance and OA-NG-CG angles are shown in Figure 3 for compounds **9** (red), **2** (blue) and **21** (turquoise) [20]. As is evident, each conjugate samples a range of conformations as represented by the probability distributions. It is these distributions that represent the individual descriptors and the degree of overlap between the descriptors (see following paragraph) may be used as independent variables for model development. In addition, the descriptors may be developed in two or more dimensions. An example of 2D probability distributions for the two structural descriptors shown in Figure 3 is shown in Figure 4. From the distributions it is evident that **9** and **2** share high degree of structural similarity with respect to the given descriptors, while **21** did not sample conformational space similar to either **9** or **2**. Accordingly, based on this qualitative analysis, **9** and **2** would be predicted to have similar activity versus **21**. Notably, this analysis did not require any alignment of the ligands, simply a comparison of the probability distributions of the selected pharmacophore features. The lack of a requirement for structural alignment represents another strength of the CSP approach.

While use of the CSP approach in a qualitative manner is of utility, as described below, quantitative analysis is required to predict inhibition constants, potencies and so on. This requires that the degree of overlap of the probability distributions of the individual ligands be determined, yielding overlap coefficients that may be used directly in regression analysis. 1D overlap coefficient of a single structure descriptor between two ligands can be calculated using the following relation for discrete probability density functions [20,94],

$$OC=\sum_{i=1}^{N}\min\left(P^{A_i},P^{B_i}\right) \tag{4}$$

where P^{Ai} andP^{Bi} are the probability in bin i for compounds A and B and N is the total number of bins. Similarly 2D overlap coefficients between two different structural descriptors can be calculated based on Equation 5, [92]

$$OC=\frac{\sum_{ij}P_{ij^k}\cdot .P_{ij^l}}{\sqrt{\sum_{ij}\left(P_{ij^k}\right)^2\cdot\sum_{ij}\left(P_{ij^l}\right)^2}} \tag{5}$$

where *P* represents the normalized probability at pixel *ij* from the 2D distributions for compounds *k* (i.e., the reference compound) and *l*. Usually the most potent compound is chosen to be the reference compound. Accordingly, overlap coefficients are quantitative measures of similarity of the ligands with respect to the reference compound in the context of their sampling conformation space (see below). In terms of the 2D distributions shown above in Figure 4, 2D overlap coefficients for **2** and **21** were calculated with respect to **9**; **2** yielded an overlap coefficient of 0.688 whilst this value for **21** is 0. Experimental data used for this study showed that **9** (0.953μM) and **2** (2.26μM) both were potent inhibitors of hASBT while **21** displayed moderate potency (31.8μM). In that study, which included 13 ligands in combination with multiple regression analysis, the CSP-SAR method was able to obtain correlations with the experimental data both quantitatively as well as qualitatively, leading to a physical understanding of the relationship of the compounds to their biological activity.

A final advantage of the CSP approach is that the structural descriptors, the overlap coefficients, may be readily combined with descriptors based on physical properties. For example, physical properties such as polar surface area, dipole moment, free energy of solvation among others may be calculated for each ligand and included in the regression analysis. This may involve calculating the physical property for each conformation of a ligand and using the average values for regression analysis. The capability to readily include physical properties is clearly an additional strength of the CSP approach.

Computational Method

The primary requirement of the CSP method is adequate conformational sampling for the ligand molecules. In order to achieve a complete sampling of conformational space rigorous molecular dynamics (MD) simulations [95] are an essential part of CSP. However, other sampling methods such as systematic grid search [96-98], fragment-based search [36,99], random search or Monte Carlo (MC) simulations [100-103], distance geometry [103], genetic algorithm [104-106], simulated annealing [107,108], taboo search [109] etc. can also be applied for conformational sampling purpose as long as exhaustive sampling of conformational space can be assured. A detail description of these searching algorithms can be found in several review articles and book chapters [95,110-112].

Empirical force fields [82] are an integral part of *in silico* modeling. Any molecular force field such as CHARMM [113], AMBER [114], MMFF [115-119] or OPLS [120] which is suitable for small molecules can be used for CSP-SAR modeling. However, it is important that the force field used accurately model the structural properties of the molecules of interest. Test of this accuracy may be performed by quantifying the ability of the force field to reproduce minimum energy geometries with those obtained from quantum mechanical (QM) calculations or high resolution crystals structures such as those obtained from the Cambridge Structural Database [121]. In addition, the use of QM methods allows the ability of a force field to reproduce the change in energy as a function of ligand conformation to be validated and optimized, as required. Proper treatment of the conformational energies is particular important for the CSP approach as it is based on conformational distributions. Methods for force field validation and optimized have been described elsewhere [36,122].

CSP-SAR models developed by our laboratory used MD simulations as a tool for conformational search. MD generates consecutive conformations of a molecular system using Newton's second law of motion in which the force acting on a system along with velocities of the atoms in the system are used to predict new conformations by integrating over time. The time evolution of the position and velocity of the molecular system is estimated from the analytical solution of the differential equation of motion. For flexible ligands replica-exchange MD simulations [123-126] are preferably employed for sampling the conformational space. Replica-exchange MD simulation methods reduce the probability of a molecular system getting trapped in local minimum energy region during a simulation facilitating complete sampling of the accessible conformational space. In this method a number of replicas of the same system are simultaneously simulated at different temperatures and with coordinates or other properties swapped between the replicas performed at regular interval. The probability of the exchange of two replicas is subjected to the Metropolis Criterion [127] thereby assuring that the system maintains a proper Boltzmann distribution. MD simulations of each replica are typically performed using 20ns Langevin dynamics [128] with an integration time step of 0.002ps in the presence of an implicit solvent model [129-132], such as Generalized Born Continuum Solvent Model (GBMV) [133,134]. Usually 20ns simulations yield conformational convergence for flexible ligands with moderate size (~ 650 Daltons); testing that additional simulation time does not lead to additional sampling is often adequate to verify that the full range of accessible conformations of the molecule has been sampled. Coordinate frames are saved from the MD trajectories and used to determine the conformational distribution of the structural descriptors from which the overlap coefficients are calculated. During *in silico* modeling the protonation state of any ionizable chemical group should be properly assigned based on the pH of the experimental condition used to measure the biological activity.

1D and 2D probability distributions of various pharmacophoric feature points are obtained by analyzing the trajectories from the MD simulations. Overlap coefficients of the conformational distributions are combined with the physico-chemical properties of the ligands to obtain a set of molecular descriptors. The molecular descriptors are subjected to single-variable as well as multivariable linear regression (MLR) analysis against the biological activity of interest. All possible combinations of the molecular descriptors are subjected to MLR analysis to identify the combination of descriptors (candidate models) that can explain the variability of the biological activity of the ligands. To avoid overfitting, any combination of independent variables having correlation between each other greater than 0.8 are not included for multivariable regression. Akaike information criteria [135,136] is applied to rank the candidate models for systems with more than one statistically significant quantitative models. Simple SHELL scripts may be used to automate the process of capturing snapshots from the MD trajectories and calculating the overlap coefficients of the structural features. MLR analysis for all possible combinations of molecular descriptors and calculation of AIC values of the selected candidate models can also be automated using statistical software like R in conjunction with a SHELL script. The combination of CSP approach with 3D QSAR method, named CSP-SAR, thus potentially can capture information on the bioactive conformation in model development which facilitates an understanding of the biological interactions dictating activity without any available ligand-target 3D structure.

Applications

The CSP method was developed and first successfully applied on δ-opioid ligands [21,91, 92]. CSP was used to study both peptidic as well as nonpeptidic opioids and the derived pharmacophore model distinguished δ-opioid agonists from the antagonists. Using qualitative CSP models for δ opioid ligands Bernard and coworkers [21] discovered that DPI2505, a compound previously suggested to be a δ antagonist, may act as an agonist. The qualitative model was also used to design novel δ opioid ligands. Subsequent application of quantitative CSP for δ-opioid ligands [92] yielded efficacy and affinity models that were able to distinguish between ligands that differed by a single substitution on an aromatic ring. These efforts also discovered a novel hydrophobic moiety imporant for δ efficacy and affinity that had not been identified in previous studies. This represented a significant advance in our understanding of opioid SAR, as previous thinking assumed that the hydrophobic moiety was limited to aromatic groups, whereas as the CSP approach showed that aliphatic moieties could also serve as the hydrophobic groups in certain ligands. Notably, the models of opioid developed by the CSP methods encompassed low molecular weight, nonpeptidic opioids as well as peptidic ligands.

Previous opioid models were not able to bridge this gap. The ability of the CSP method to overcome this is based on the inclusion of all conformations in model development, the lack of the need to align molecules, a particular problem when both non-peptidic and peptidic ligands are being studied and the inclusion of a large number of possible pharmacophore features in model development. Indeed, that later consideration led to the identification of the novel hydrophobic moieties in the selected opioids.

Recently, the inhibition requirement of hASBT using amino-piperidine conjugates of bile acids was studied using the CSP-SAR method [20]. CSP-SAR models developed for hASBT inhibition successfully identified structural and physico-chemical descriptors that explained the variance of the biological activity. Despite the fact that the inhibitors used in this study had a narrow range of activity, the conformational sampling feature of the CSP-SAR method was able to facilitate identification of the information from the molecular descriptors necessary to explain the activity. The quantitative CSP-SAR models developed in this study was able to distinguish between very-potent inhibitors (<16μM) from moderately-potent (>16 μM) inhibitors with some exceptions. However, further qualitative analysis was able to overcome the limitation of the quantitative models. Qualitative CSP-SAR demonstrated that very subtle chemical modifications in some inhibitors led to the formation of salt-bridge interaction resulting in conformational restriction associated to poorer binding affinity. This study established the strength of CSP-SAR method to capture the effect of such small chemical modification on biological activity and it emphasizes the utility of both quantitative and qualitative CSP approaches.

The CSP method has also been applied and discussed by other researchers in the context of 3D QSAR [137-145]. CSP models developed by Bernard and coworkers, demonstrated the importance of including extensive conformational sampling in model development. This motivated other researchers to include conformational sampling during the development of 3D QSAR models for other flexible systems. Gilbert and coworkers applied the concept of CSP method in their work by considering a set of representative conformations of the flexible ligands to develop selective inhibitors of DAT/SERT using CoMFA and CoMSIA methods [137]. Mallik and coworkers [140] used the CSP method for developing 3D pharmacophore for the 13-residue cyclic peptide, compstatin, an anti-complement peptide and other related peptidic analogues. Using the CSP methodology the researchers were able to distinguish between active and inactive analogues. The researchers also extended the original CSP work by Bernard and coworkers, by including dihedral angles as a pharmacophoric descriptor to capture 3D structural features of the peptidic ligands. The inclusion of multiple conformers instead of using only the lowest-energy conformer yielded a stable and predictive model. Kalaszi and coworkers [142] developed a novel 3D QSAR method based on thermodynamic properties to predict bioactive conformation of flexible ligands using conformational analysis of the ligand molecules. In two recent studies Lexa and coworkers [143] and Kirschner and coworkers [144] used replica exchange molecular dynamics to explore the conformational space accessible by peptidic ligands with breast cancer inhibiting properties. They used ligand-based 3D QSAR method to identify the bioactive conformations of the active ligands. Conformational analysis of the larger active peptides allowed them to explain the activity of the existing ligands as well as discover novel smaller peptidic ligands with full biological activity. The successful works of these researchers confirm the validity and importance of CSP approach in ligand-based 3D QSAR modeling for flexible molecules. In addition, CSP has also been mentioned in several review articles [13,141,145] as a novel method to utilize the dynamical behavior of flexible biomolecules to explain ligand-protein binding.

Validation of the CSP method

To assess the performance of CSP method as compared to more traditional 3D QSAR approaches, additional calculations were performed as part of the present study. These involved a comparative study of the inhibition pharmacophore model for hASBT based on the thirteen ligands in **G1** and **G3** groups as described by Gozalez and coworkers [20] with a model developed presently using the Catalyst approach. Similar to the observation of Gonzalez and coworkers inclusion of the compounds in **G2** group did not yield statistically significant model (r^2 = 0.55).

Catalyst model development was performed using Discovery Studio 2.1 Catalyst™ (Accelrys, San Diego, CA). The best conformation generation method as implemented in Catalyst™ was used to generate up to 250 conformers of each ligand based on a 20 kcal/mol energy cutoff. Ten hypotheses were generated using the conformers of the ligands and their K_i values using five molecular features, such as hydrogen bond donor, hydrogen bond acceptor, hydrophobic, positively ionizable group and negatively ionizable group. Out of the ten hypotheses, the hypothesis yielding the lowest total cost was selected for further analysis. The best inhibition model generated by Catalyst consisted of five features including one hydrogen-bond acceptor, one hydrogen-bond donor, two hydrophobic moieties and one positively ionizable group. The most potent inhibition in the set, compound **9**, mapped all the five features of the pharmacophore; 3-OH represented the hydrogen-bond acceptor, 7-OH represented the hydrogen-bond donor, C-19 and D-ring represented the two hydrophobes and the basic piperidine nitrogen depicted the positively ionizable group feature. Top three CSP-SAR inhibition models also consisted of structural descriptors representing similar features, e.g. 7-OH, basic piperidine nitrogen (positively ionizable group) and hydrophobic moieties close to C-19 and D-ring such as centroid of B and C rings and C-20. In addition, CSP-SAR models included structural descriptors involving the relative orientation of α-substituent with respect to the steroidal nucleus. However, CSP-SAR models did not any descriptor that explicitly considered 3-OH. CSP-SAR models also included physico-chemical descriptors such as GB energy (electrostatic component of solvation free energy) and logP (octanol/water partition coefficient). This is a clear advantage of CSP-SAR method over Catalyst as there is no simple tool in Catalyst that can combine structural features with physico-chemical descriptors. The Catalyst model yielded r^2 of 0.849 while the r^2 of the best model reported by Gonzalez and coworkers was 0.813. However, the CSP-SAR model yielded better RMSD value than the Catalyst model (Table 1). Table 1 represents the observed and estimated K_i values of the ligands based on CSP and Catalyst methods. Moreover, the best CSP-SAR model included only two descriptors to explain the activity while five descriptors were used by Catalyst model for the same set of compounds. From the comparison of the inhibition models developed by the two methods it is evident that CSP and Catalyst yielded very similar fitting quality. Nevertheless, the Catalyst method did not provide any tool to explanation of the variation in activity due to subtle chemical modifications; while CSP-SAR qualitative model was able to explain such variation in activity via salt-bridge interaction.

Limitations

One limitation of the CSP-SAR method is that the selection of pharmacophoric features of the ligands is user dependent. The selection of functional groups is often facilitated by previous studies though all the chemical groups present on the ligands must be considered. This limitation may be overcome by considering probability distributions between all possible distances, angles and dihedral angles involving all chemical groups that may impact on biological activity. A second limitation is the computational requirement. As extensive sampling of conformational space is required, extended MD simulations must be performed on each ligand. While this step is computationally demanding, the accessibility of commodity computing minimizes this limitation. In addition, once the conformational sampling of a ligand is completed and the generated conformations stored, further analysis may be performed to identify additional structural or physio-chemical properties that correlate with biological activity without redoing the MD simulation or other sampling procedure.

Conclusions

Ligand-based drug design is inherently a complicated problem as this approach is restricted to considering only one side of the actual biochemical process. It has been shown in many cases that receptor molecules and/or ligands undergo significant conformational changes to facilitate their interaction [146-150]. While traditional pharmacophore approaches often did not take into account ligand conformational flexibility by only using minimum energy conformations of the ligands, more recent methods include a large number of conformations during model development. Though such methods offer significant improvements, they are still limited by including a finite range of conformations as well as requiring alignment of the ligands under study. The CSP method largely overcomes these limitations by including all accessible conformations of the ligands and using the overlap of probability distributions of pharmacophore features during model development. In addition, the CSP-SAR method may readily be combined with physicochemical properties. The utility of this approach has been demonstrated in a number of studies in our laboratories as well as by other workers.

Clearly, ligand-based drug design is an effective method to understand the features of ligands important for their biological activity in the absence of the receptor structure. Investigation of the structural and physico-chemical features of the ligands of a drug target can indicate the nature of interactions that are essential for the desired pharmacological response. The method can also predict novel molecular structures with features facilitating the interaction with the target molecule. As stated above, there are several different methodologies to perform ligand-based modeling. However, proper understanding of the underlying principle of the chosen method is highly recommended for successful application of these methods to complex biological systems.

Acknowledgments

This work was supported in part by National Institutes of Health grant DK67530 (to JEP and ADM), NIDA Grant DA13583 (to AC and ADM), NIDA Grant DA19634 (to AC and ADM), Accelrys Inc., and the Computer-Aided Drug Design Center, School of Pharmacy, University of Maryland, Baltimore for computational resources.

References

1. Chang C, Ekins S, Bahadduri P, Swaan PW. Pharmacophore-based discovery of ligands for drug transporters. Adv. Drug Deliv. Rev 2006;58(12-13):1431–50. [PubMed: 17097188]
2. Ekins S, Mirny L, Schuetz EG. A Ligand-Based Approach to Understanding Selectivity of Nuclear Hormone Receptors PXR, CAR, FXR, LXRα, and LXRβ. Pharm. Res 2002;19(12):1788–1800. [PubMed: 12523656]
3. Ekins S, Waller CL, Swaan PW, Cruciani G, Wrighton SA, Wikel JH. Progress in predicting human ADME parameters in silico. J. Pharmacol. Toxicol. Methods 2000;44(1):251–272. [PubMed: 11274894]
4. van de Waterbeemd H, Gifford E. ADMET in silico modelling: towards prediction paradise? Nat. Rev. Drug Discov 2003;2(3):192–204. [PubMed: 12612645]
5. Ekins, S. Computer applications in pharmaceutical research and development. John Wiley & Sons; Hoboken, NJ: 2006.
6. Balakrishnan A, Polli JE. Apical sodium dependent bile acid transporter (ASBT, SLC10A2): a potential prodrug target. Mol. Pharm 2006;3(3):223–30. [PubMed: 16749855]
7. Takakura Y, Hashida M. Macromolecular drug carrier systems in cancer chemotherapy: macromolecular prodrugs. Crit. Rev. Oncol. Hematol 1995;18(3):207–231. [PubMed: 7695833]

8. Tolle-Sander S, Lentz KA, Maeda DY, Coop A, Polli JE. Increased acyclovir oral bioavailability via a bile acid conjugate. Mol. Pharm 2004;1(1):40–8. [PubMed: 15832499]
9. Kurogi Y, Guner OF. Phamacophore Modeling and Three-dimensional Database Searching for Drug Design using Catalyst. Curr. Med. Chem 2001;8:1035–1055. [PubMed: 11472240]
10. Marrone TJ, Briggs JM, McCammon JA. Structure-based drug design: computational advances. Annu. Rev. Pharmacol. Toxicol 1997;37:71–90. [PubMed: 9131247]
11. Gane PJ, Dean PM. Recent advances in structure-based rational drug design. Curr. Opin. Struct. Biol 2000;10(4):401–4. [PubMed: 10981625]
12. Jhoti, H.; Leach, AR. Structure-based Drug Discovery. Springer; 2007.
13. Zhong S, Macias AT, MacKerell AD Jr. Computational identification of inhibitors of protein-protein interactions. Curr. Top. Med. Chem 2007;7(1):63–82. [PubMed: 17266596]
14. Chen F, Hancock CN, Macias AT, Joh J, Still K, Zhong S, MacKerell AD Jr. Shapiro P. Characterization of ATP-independent ERK inhibitors identified through in silico analysis of the active ERK2 structure. Bioorg. Med. Chem. Lett 2006;16(24):6281–7. [PubMed: 17000106]
15. Hafner C, Schmitz G, Meyer S, Bataille F, Hau P, Langmann T, Dietmaier W, Landthaler M, Vogt T. Differential Gene Expression of Eph Receptors and Ephrins in Benign Human Tissues and Cancers. Clin. Chem 2004;50(3):490–499. [PubMed: 14726470]
16. Dobrzanski P, Hunter K, Jones-Bolin S, Chang H, Robinson C, Pritchard S, Zhao H, Ruggeri B. Antiangiogenic and antitumor efficacy of EphA2 receptor antagonist. Cancer Res 2004;64(3):910–9. [PubMed: 14871820]
17. Cheng N, Brantley D, Fang WB, Liu H, Fanslow W, Cerretti DP, Bussell KN, Reith A, Jackson D, Chen J. Inhibition of VEGF-dependent multistage carcinogenesis by soluble EphA receptors. Neoplasia 2003;5(5):445–56. [PubMed: 14670182]
18. Torres GE, Gainetdinov RR, Caron MG. Plasma membrane monoamine transporters: structure, regulation and function. Nat. Rev. Neurosci 2003;4(1):13–25. [PubMed: 12511858]
19. Mason JS, Good AC, Martin EJ. 3-D Pharmacophores in Drug Discovery. Curr. Pharm. Des 2001;7:567–597. [PubMed: 11375769]
20. González PM, Acharya C, MacKerell AD Jr. Polli JE. Inhibition Requirements of the human Apical Sodium-dependent Bile acid Transporter (hASBT) using Aminopiperidine Conjugates of glutamyl-Bile Acids. Pharm. Res 2009;26(7):1665–1678. [PubMed: 19384469]
21. Bernard D, Coop A, MacKerell AD Jr. Conformationally sampled pharmacophore for peptidic delta opioid ligands. J. Med. Chem 2005;48(24):7773–80. [PubMed: 16302816]
22. Loew GH, Villar HO, Alkorta I. Strategies for indirect computer-aided drug design. Pharm. Res 1993;10(4):475–86. [PubMed: 8483829]
23. Koehn FE, Carter GT. The evolving role of natural products in drug discovery. Nat. Rev. Drug Discov 2005;4(3):206–20. [PubMed: 15729362]
24. Lee KH. Anticancer drug design based on plant-derived natural products. J. Biomed. Sci 1999;6(4):236–50. [PubMed: 10420081]
25. Lee KH, Huang BR, Tzeng CC. Synthesis and anticancer evaluation of certain alpha-methylene-gamma-(4-substituted phenyl)-gamma-butyrolactone bearing thymine, uracil, and 5-bromouracil. Bioorg. Med. Chem. Lett 1999;9(2):241–4. [PubMed: 10021937]
26. Kuntz ID. Structure-based strategies for drug design and discovery. Science 1992;257(5073):1078–82. [PubMed: 1509259]
27. Guner OF. Pharmacophore perception, development, and use in drug design. 1999
28. Akamatsu M. Current State and Perspectives of 3D-QSAR. Curr. Top. Med. Chem 2002;2:1381–1394. [PubMed: 12470286]
29. Verma RP, Hansch C. Camptothecins: A SAR/QSAR Study. Chem. Rev 2009;109:213–235. [PubMed: 19099450]
30. Duchowicz PR, Castro EA, Fernandez FM, Gonzalez MP. A new search algorithm for QSPR/QSAR theories: Normal boiling points of some organic molecules. Chem Phys Lett 2005;412:376–380.
31. Wade RC, Henrich S, Wang T. Using 3D protein structures to derive 3D-QSARs. Drug Discovery Today: Technologies 2004;1(3):241–246.

32. Halloway MK. A priori prediction of ligand affinity by energy minimization. Perspectives in Drug Discovery and Design 1998;9(11):63–84.

33. Bohl CE, Chang C, Mohler ML, Chen J, Miller DD, Swaan PW, Dalton JT. A ligand-based approach to identify quantitative structure-activity relationships for the androgen receptor. J. Med. Chem 2004;47(15):3765–76. [PubMed: 15239655]

34. Winkler DA. Rapley R, Harbron S. Overview of Quantitative Structure-Activity Relationships (QSAR). Molecular Analysis and Genome Discovery. 2004

35. Discovery Studio. Accelrys Inc.; (http://www.accelrys.com/dstudio)

36. Molecular Operating Environment, Chemical Computing Group. Montreal, Quebec, Canada: http://www.chemcomp.com

37. Topliss JG, Edwards RP. Chance factors in studies of quantitative structure-activity relationships. J. Med. Chem 1979;22(10):1238–44. [PubMed: 513071]

38. MATLAB. The MathWorks, Inc.; (http://www.mathworks.com/matlabcentral)

39. Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics 1996;5(3):299–314.

40. Wold S. Principal component analysis. Chemom. Intell. Lab. Syst 1987;2:37–52.

41. Kubinyi H. QSAR and 3D QSAR in drug design Part 1: methodology. Drug Discov. Today 1997;2 (11):457–467.

42. Geladi P, Kowalski BR. Partial least-squares regression: a tutorial. Anal. Chim. Acta 1986;185:1–17.

43. Bajorath J. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. J. Chem. Inf. Comput. Sci 2001;41(2):233–45. [PubMed: 11277704]

44. Zheng W, Tropsha A. Novel variable selection quantitative structure--property relationship approach based on the k-nearest-neighbor principle. J. Chem. Inf. Comput. Sci 2000;40(1):185–94. [PubMed: 10661566]

45. Bultinck P, Winter HD, Langenaeker W, Tollenaere JP. Computational medicinal chemistry for drug discovery. 2004

46. Fernandez M, Caballero J, Tundidor-Camba A. Linear and nonlinear QSAR study of N-hydroxy-2-[(phenylsulfonyl)amino]acetamide derivatives as matrix metalloproteinase inhibitors. Bioorg. Med. Chem 2006;14(12):4137–50. [PubMed: 16504515]

47. Caballero J, Fernandez M. Linear and nonlinear modeling of antifungal activity of some heterocyclic ring derivatives using multiple linear regression and Bayesian-regularized neural networks. J Mol Model 2006;12(2):168–81. [PubMed: 16205958]

48. Karelson M, Sild S, Maran U. Non-Linear QSAR Treatment of Genotoxicity. Mol. Simul 2000;24 (4):229–242.

49. Devillers J. Linear versus nonlinear QSAR modeling of the toxicity of phenol derivatives to Tetrahymena pyriformis. SAR QSAR Environ. Res 2004;15(4):237–49. [PubMed: 15370415]

50. Salt DW, Yildiz N, Livingstone DJ, Tinsley CJ. The use of artificial neural networks in QSAR. Pestic. Sci 1992;36(2):161–170.

51. Jalali-Heravi M, Parastar F. Use of Artificial Neural Networks in a QSAR Study of Anti-HIV Activity for a Large Group of HEPT Derivatives. J. Chem. Inf. Comput. Sci 2000;40(1):147–154. [PubMed: 10661561]

52. Agrafiotis DK, Cedeno W, Lobanov VS. On the use of neural network ensembles in QSAR and QSPR. J. Chem. Inf. Comput. Sci 2002;42(4):903–911. [PubMed: 12132892]

53. Manallack DT, Ellis DD, Livingstone DJ. Analysis of linear and nonlinear QSAR data using neural networks. J. Med. Chem 1994;37(22):3758–67. [PubMed: 7966135]

54. Manallack DT, Tehan BG, Gancia E, Hudson BD, Ford MG, Livingstone DJ, Whitley DC, Pitt WR. A consensus neural network-based technique for discriminating soluble and poorly soluble compounds. J. Chem. Inf. Comput. Sci 2003;43(2):674–9. [PubMed: 12653537]

55. Douali L, Villemin D, Cherqaoui D. Comparative QSAR based on neural networks for the anti-HIV activity of HEPT derivatives. Curr. Pharm. Des 2003;9(22):1817–26. [PubMed: 12871199]

56. Douali L, Villemin D, Cherqaoui D. Neural networks: Accurate nonlinear QSAR model for HEPT derivatives. J. Chem. Inf. Comput. Sci 2003;43(4):1200–7. [PubMed: 12870912]

57. Burden FR, Winkler DA. Robust QSAR models using Bayesian regularized neural networks. J. Med. Chem 1999;42(16):3183–7. [PubMed: 10447964]

58. Burden FR, Ford MG, Whitley DC, Winkler DA. Use of Automatic Relevance Determination in QSAR Studies Using Bayesian Neural Networks. J. Chem. Inf. Comput. Sci 2000;40:1423–1430. [PubMed: 11128101]

59. Winkler DA, Burden FR. Modelling blood-brain barrier partitioning using Bayesian neural nets. J. Mol. Graph. Model 2004;22(6):499–505. [PubMed: 15182809]

60. Buntine WL, Weigend AS. Bayesian back-propagation. Complex Sys 1991;5:603–643.

61. Burden FR, Winkler DA. An Optimal Self-Pruning Neural Network and Nonlinear Descriptor Selection in QSAR. QSAR Comb. Sci 2009;28(10):1092–1097.

62. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence 1995;2(12): 1137–1143.

63. Weiss, SM.; Kulikowski, CA. Computer systems that learn. Morgan Kaufmann; San Mateo, CA: 1991.

64. Cronin MTD, Livingstone D. Predicting Chemical Toxicity and Fate. 2004

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

66. Topliss JG. Some Observations on Classical QSAR. Perspectives in Drug Discovery and Design 1993;1:253–268.

67. Free SM Jr. Wilson JW. A Mathematical Contribution to Structure-Activity Studies. J. Med. Chem 1964;7(4):395–399. [PubMed: 14221113]

68. Tmej C, Chiba P, Huber M, Richter E, Hitzler M, Schaper KJ, Ecker G. A combined Hansch/Free-Wilson approach as predictive tool in QSAR studies on propafenone-type modulators of multidrug resistance. Arch. Pharm. (Weinheim) 1998;331(7-8):233–40. [PubMed: 9747179]

69. Japertas P, Didziapetris R, Petrauskas A. Fragmental Methods in the Design of New Compounds. Applications of The Advanced Algorithm Builder. Quant. Struct-Act. Relat 2002;1:23–37.

70. Winkler DA. The role of quantitative structure-activity relationships (QSAR) in biomolecular discovery. Brief. Bioinform 2002;3(1):73–86. [PubMed: 12002226]

71. Chang C, Swaan PW. Computational approaches to modeling drug transporters. Eur. J. Pharm. Sci 2006;27(5):411–24. [PubMed: 16274971]

72. IUPAC Compendium of Chemical Terminology. 1997

73. Wermuth, CG.; Langer, T. Pharmacophore identification. In 3D-QSAR in Drug Design. Theory, Methods, and Applications. ESCOM Science Publishers; 1993.

74. Hopfinger AJ, Wang S, Tokarski JS, Jin B, Albuquerque M, Madhav PJ, Duraiswami C. Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism. J. Am. Chem. Soc 1997;119:10509–10524.

75. de Groot MJ, Ekins S. Pharmacophore modeling of cytochromes P450. Adv. Drug Deliv. Rev 2002;54 (3):367–83. [PubMed: 11922953]

76. Van Drie JH. Pharmacophore discovery: Lessons learned. Curr. Pharm. Des 2003;9:1649–1664. [PubMed: 12871063]

77. Cramer RD, Patterson DE, Bunce JD. Comparative Molecular-Field Analysis (Comfa) .1. Effect of Shape on Binding of Steroids to Carrier Proteins. J. Am. Chem. Soc 1988;110(18):5959–5967.

78. Gohda K, Mori I, Ohta D, Kikuchi T. A CoMFA analysis with conformational propensity: an attempt to analyze the SAR of a set of molecules with different conformational flexibility using a 3D-QSAR method. J. Comput. Aided Mol. Des 2000;14(3):265–75. [PubMed: 10756481]

79. Yasuo K, Yamaotsu N, Gouda H, Tsujishita H, Hirono S. Structure-based CoMFA as a predictive model - CYP2C9 inhibitors as a test case. J. Chem. Inf. Model 2009;49(4):853–64. [PubMed: 19391630]

80. Ghose AK, Viswanadhan VN, Wendoloski JJ. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. J. Combin. Chem 1999;1:55–68.

81. Bostrom J, Norrby PO, Liljefors T. Conformational energy penalties of protein-bound ligands. J. Comput. Aided Mol. Des 1998;12(4):383–96. [PubMed: 9777496]

82. MacKerell AD Jr. Empirical force fields for biological macromolecules: overview and issues. J. Comput. Chem 2004;25(13):1584–604. [PubMed: 15264253]

83. Hasegawa K, Arakawab M, Funatsu K. Rational choice of bioactive conformations through use of conformation analysis and 3-way partial least squares modeling. Chemom. Intell. Lab. Syst 2000;50 (2):253–261.

84. 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(24):4130–46. [PubMed: 7990113]

85. Folkers G, Merz A, Rognan D. CoMFA: Scope and Limitations in 3D QSAR in Drug Design. 1993

86. Flower DR. Predicting Chemical Toxicity and Fate. 2002

87. Flower DR. Drug design: cutting edge approaches. 2002

88. Klebe G, Abraham U. Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. J. Comput. Aided Mol. Des 1999;13(1):1–10. [PubMed: 10087495]

89. Catalyst. Accelrys Inc.; SanDiego, CA: 2002.

90. Smellie A, Teig SL, Towbin P. Poling: Promoting Conformational Variation. J. Comput. Chem 1995;16:171–187.

91. Bernard D, Coop A, MacKerell AD Jr. 2D conformationally sampled pharmacophore: a ligand-based pharmacophore to differentiate delta opioid agonists from antagonists. J. Am. Chem. Soc 2003;125 (10):3101–7. [PubMed: 12617677]

92. Bernard D, Coop A, MacKerell AD Jr. Quantitative conformationally sampled pharmacophore for delta opioid ligands: reevaluation of hydrophobic moieties essential for biological activity. J. Med. Chem 2007;50(8):1799–809. [PubMed: 17367120]

93. Nicklaus MC, Wang S, Driscoll JS, Milne GWA. Conformational changes of small molecules binding to proteins. Bioorg. Med. Chem 1995;3:411–428. [PubMed: 8581425]

94. Reiser B, Faraggi D. Confidence Intervals for the Overlapping Coefficient: the Normal Equal Variance Case. The Statistician 1999;48(3):413–418.

95. Leach, AR. Molecular modelling: principles and applications. Addison-Wesley Longman Ltd; 2001.

96. Ghose AK, Jaeger EP, Kowalczyk PJ, Peterson ML, Treasurywala AM. Conformational searching methods for small molecules. I. Study of the sybyl search method. J. Comp. Chem 1993;14(9):1050–1065.

97. Lipton M, Still WC. The multiple minimum problem in molecular modeling. Tree searching internal coordinate conformational space. J. Comp. Chem 1988;9(4):343–355.

98. Wiberg KB, Boyd RH. Application of strain energy minimization to the dynamics of conformational changes. J. Am. Chem. Soc 1972;94(24):8426–8430.

99. Chen IJ, Foloppe N. Conformational Sampling of Druglike Molecules with MOE and Catalyst: Implications for Pharmacophore Modeling and Virtual Screening. J. Am. Chem. Soc 2008;48(9): 1773–1791.

100. Saunders M. Stochastic exploration of molecular mechanics energy surfaces. Hunting for the global minimum. J. Am. Chem. Soc 1987;109(10):3150–3152.

101. Ferguson DM, Raber DJ. A new approach to probing conformational space with molecular mechanics: random incremental pulse search. J. Am. Chem. Soc 1989;111(12):4371–4378.

102. Metropolis N, Ula S. The Monte Carlo Method. J. Am. Stat. Assoc 1949;44(247):335–341. [PubMed: 18139350]

103. Blaney, JM.; Dixon, JS. Distance geometry in molecular modeling, in. Reviews in Computational Chemistry. VCH Publishers; New York: 1994.

104. Nair N, Goodman JM. Genetic Algorithms in Conformational Analysis. J. Chem. Inf. Comput. Sci 1998;38(2):317–320.

105. Clark DE, Jones G, Willett P. Pharmacophoric pattern matching in files of three-dimensional chemical structures: Comparison of conformational-searching algorithms for flexible searching. J. Chem. Inf. Comput. Sci 1994;34(1):197–206.

106. Parrill AL. Evolutionary and genetic methods in drug design *Drug Discov*. Today 1996;1(12):514–521.

107. Corcho FJ, Filizola M, Perez JJ. Evaluation of the iterative simulated annealing technique in conformational search of peptides. Chem Phys Lett 2000;319(1-2):65–70.

108. Guarnieri F, Weinstein H. Conformational Memories and the Exploration of Biologically Relevant Peptide Conformations: An Illustration for the Gonadotropin-Releasing Hormone. J. Am. Chem. Soc 1996;118:5580–5589.

109. Cvijovicacute D, Klinowski J. Taboo Search: An Approach to the Multiple Minima Problem. Science 1995;267(5198):664–666. [PubMed: 17745843]

110. Foloppe N, Chen IJ. Conformational Sampling and Energetics of Drug-like Molecules. Curr. Med. Chem. [Epub ahead of print]. 2009

111. Perola E, Charifson PS. Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. J. Med. Chem 2004;47(10):2499–510. [PubMed: 15115393]

112. Becker OM, MacKerell AD Jr. Roux B, Watanabe M. Computational biochemistry and biophysics. 2001

113. Brooks BR, Brooks CL III, MacKerell AD Jr. Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RV, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: The biomolecular simulation program. J. Comp. Chem 2009;30(10):1545–614. [PubMed: 19444816]

114. Case, DA.; Darden, TA. AMBER 9. University of California; San Francisco: 2006.

115. Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J. Comp. Chem 1996;17(5-6):490–519.

116. Halgren TA. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. J. Comp. Chem 1996;17(5-6):520–552.

117. Halgren TA. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. J. Comp. Chem 1996;17(5-6):553–586.

118. Halgren TA. Merck molecular force field. IV. conformational energies and geometries for MMFF94. J. Comp. Chem 1996;17(5-6):587–615.

119. Halgren TA. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. J. Comp. Chem 1996;17(5-6):616–641.

120. Jorgensen WL, Tirado-Rives J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. J. Am. Chem. Soc 1988;110(6):1657–1666.

121. The Cambridge Structural Database. (http://www.ccdc.cam.ac.uk/products/csd/)

122. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell AD Jr. CHARMM General Force Field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comp Chem 2010;31(4)

123. Sugita Y, Okamoto Y. Replica-exchange Molecular Dynamics Method for Protein Folding. Chem Phys Lett 1999;314:141–151.

124. Feig M, Karanicolas J, Brooks CL 3rd. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J. Mol. Graph. Model 2004;22(5):377–95. [PubMed: 15099834]

125. Mitsutake A, Sugita Y, Okamoto Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. Biopolymers 2001;60(2):96–123. [PubMed: 11455545]

126. Rhee YM, Pande VS. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. Biophys. J 2003;84(2 Pt 1):775–86. [PubMed: 12547762]

127. Metropolis N, Rosembluth A, Rosembluth M, Teller A. Equation of state calculations by fast computing machins. J. Chem. Phys 1953;21:1087–1092.

128. Allen, MP.; Tildesley, DJ. Computer simulation of liquids. Clarendon Press; Oxford: 1987. p. 385

129. Bashford D, Case DA. Generalized born models of macromolecular solvation effects. Annu. Rev. Phys. Chem 2000;51:129–52. [PubMed: 11031278]

130. Tsui V, Case DA. Theory and applications of the generalized Born solvation model in macromolecular simulations. Biopolymers 2000;56(4):275–91. [PubMed: 11754341]

131. Im W, Feig M, Brooks CL. An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins. Biophys. J 2003;85(5):2900–2918. [PubMed: 14581194]

132. Im WP, Lee MS, Brooks CL. Generalized born model with a simple smoothing function. J. Comput. Chem 2003;24(14):1691–1702. [PubMed: 12964188]

133. Lee MS, Feig M, Salsbury FR, Brooks CL. New analytic approximation to the standard molecular volume definition and its application to generalized born calculations. J. Comput. Chem 2003;24 (11):1348–1356. [PubMed: 12827676]

134. Lee MS, Salsbury FR, Brooks CL. Novel generalized Born methods. J. Chem. Phys 2002;116(24): 10606–10614.

135. Akaike H. Likeliood of a Model and Information Criteria. Journal of Econometrics 1981;16:3–14.

136. McQuarrie, ADR.; Tsai, CL. Regression and Time Series Model Selection. World Scientific; Singapore: 1998.

137. Gilbert KM, Boos TL, Dersch CM, Greiner E, Jacobson AE, Lewis D, Matecka D, Prisinzano TE, Zhang Y, Rothman RB, Rice KC, Venanzi CA. DAT/SERT selectivity of flexible GBR 12909 analogs modeled using 3D-QSAR methods. Bioorg. Med. Chem 2007;15(2):1146–59. [PubMed: 17127069]

138. Bandyopadhyay D, Agrafiotis DK. A self-organizing algorithm for molecular alignment and pharmacophore development. J. Comput. Chem 2008;29(6):965–82. [PubMed: 17999384]

139. Codd EE, Carson JR, Colburn RW, Dax SL, Desai-Krieger D, Martinez RP, McKown LA, Neilson LA, Pitis PM, Stahle PL, Stone DJ, Streeter AJ, Wu WN, Zhang SP. The novel, orally active, delta opioid RWJ-394674 is biotransformed to the potent mu opioid RWJ-413216. J. Pharmacol. Exp. Ther 2006;318(3):1273–9. [PubMed: 16766719]

140. Mallik B, Morikis D. Development of a quasi-dynamic pharmacophore model for anti-complement peptide analogues. J. Am. Chem. Soc 2005;127(31):10967–76. [PubMed: 16076203]

141. Latour RA. Molecular simulation of protein-surface interactions: Benefits, problems, solutions, and future directions (Review). Biointerphases 2008;3(3):FC2–FC12. [PubMed: 19809597]

142. Kalaszi A, Imre G, Jakli I, Farkas O. Identification of the bioactive conformation for mucin epitope peptides. Journal of Molecular Structure: THEOCHEM 2007;823(1-3):16–27.

143. Lexa KW, Alser KA, Salisburg AM, Ellens DJ, Hernandez L, Bono SJ, Michael HC, Derby JR, Skiba JG, Feldgus S, Kirschner KN, Shields GC. The search for low energy conformational families of small peptides: Searching for active conformations of small peptides in the absence of a known receptor. Int. J. Quantum Chem 2007;107(15):3001–3012.

144. Kirschner KN, Lexa KW, Salisburg AM, Alser KA, Joseph L, Andersen TT, Bennett JA, Jacobson HI, Shields GC. Computational design and experimental discovery of an antiestrogenic peptide derived from alpha-fetoprotein. J. Am. Chem. Soc 2007;129(19):6263–8. [PubMed: 17441722]

145. Bodnar RJ. Endogenous opiates and behavior: 2006. Peptides 2007;28(12):2435–513. [PubMed: 17949854]

146. Remy I, Wilson IA, Michnick SW. Erythropoietin receptor activation by a ligand-induced conformation change. Science 1999;283(5404):990–3. [PubMed: 9974393]

147. Kuiper GG, Carlsson B, Grandien K, Enmark E, Haggblad J, Nilsson S, Gustafsson JA. Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. Endocrinology 1997;138(3):863–70. [PubMed: 9048584]

148. Kunishima N, Shimada Y, Tsuji Y, Sato T, Yamamoto M, Kumasaka T, Nakanishi S, Jingami H, Morikawa K. Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. Nature 2000;407(6807):971–7. [PubMed: 11069170]

149. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. Science 2000;289(5480):739–45. [PubMed: 10926528]

150. Humphries MJ. Integrin activation: the link between ligand binding and signal transduction. Curr. Opin. Cell Biol 1996;8(5):632–40. [PubMed: 8939662]

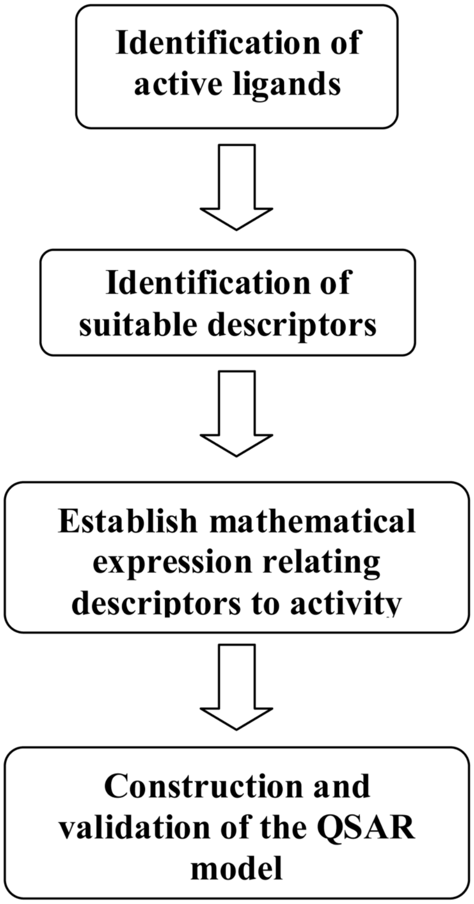
Identification of active ligands
Identification of suitable descriptors
Establish mathematical expression relating descriptors to activity
Construction and validation of the QSAR model

Figure 1.
Typical workflow of QSAR methods

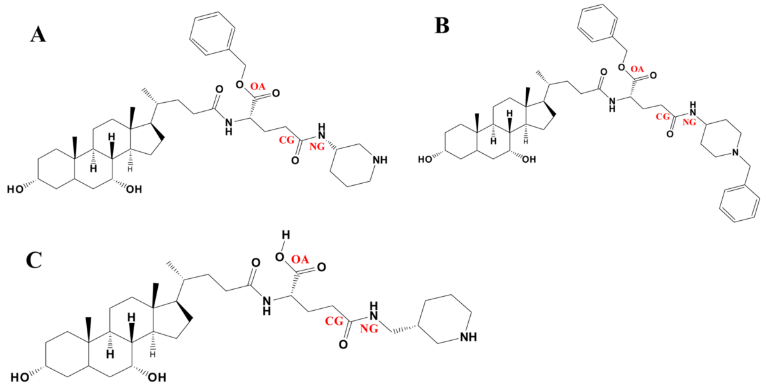

Figure 2.

Structures of three bile acid conjugates (A) **9**, (B) **2** and (C) **21** used by Gonzalez and coworkers [20]. OA, CG and NG represent three pharmacophore feature points used in the study.

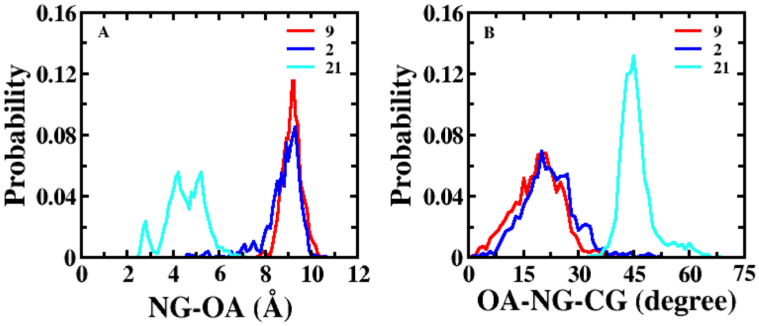

Figure 3.

1D probability distributions of distance between pharmacophoric points NG (basic nitrogen) and OA (α-acid) and angle between pharmacophoric points OA, NG and CG (amide carbon) for hASBT inhibitors; compound **2** (blue), **9** (red) and **21** (turquoise) [20].

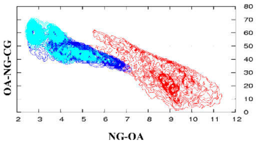

Figure 4.
2D probability distributions of distance between pharmacophoric points NG (basic nitrogen) and OA (α-acid) and angle between pharmacophoric points OA, NG and CG (amide carbon) for hASBT inhibitors; compound **2** (blue), **9** (red) and **21** (turquoise) [20].

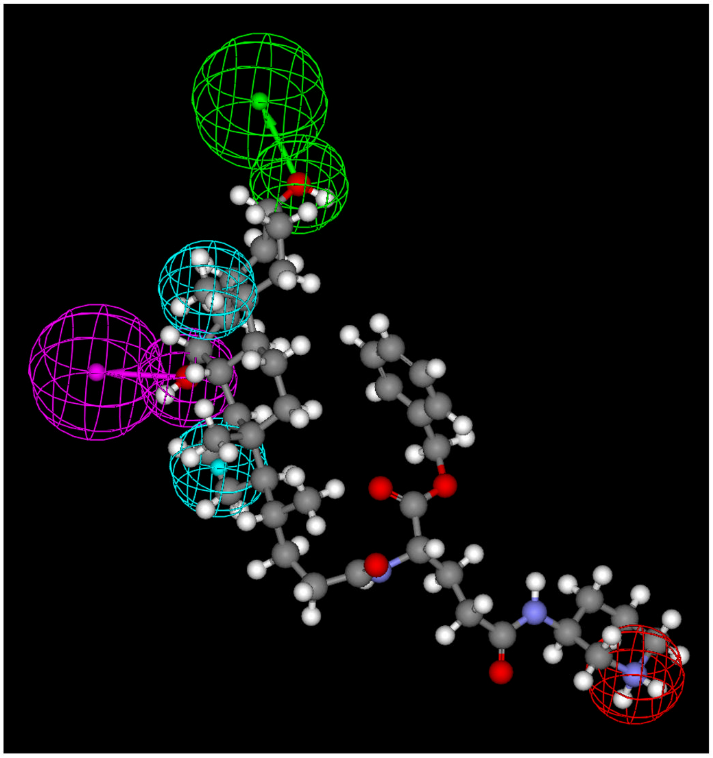

Figure 5.
Catalyst pharmacophore model 1 using **G1** and **G3** compounds illustrating one hydrogen-bond acceptor (green), one hydrogen-bond donor (purple), two hydrophobic moieties (cyan) and one positively ionizable group (red) mapped on **9**.

Table 1

Binding affinities of aminopiperidine conjugates used by Gonzalez and coworkers to hASBT.

Compound	Observed K_i (μM)	CSP-SAR estimated K_i (μM)	Difference[a]	Catalyst estimated K_i (μM)	Difference[a]
9	0.95	−1.5	−2.5	0.47	−0.48
8	1.4	4.3	2.9	1.3	−0.1
1	1.5	6.4	4.9	3.2	1.7
2	2.3	3.1	0.8	2.2	−0.1
4	3.8	4.8	1	3.7	−0.1
6	4.3	8.5	4.2	5.8	1.5
13	9.7	5	−4.7	11	1.3
10	9.9	10	0.1	19	9.1
12	10	11	1	3.7	−6.3
11	16	6.3	−9.7	17	1
23	18	16	−2	12	−6
22	19	21	2	22	3
21	32	31	−1	37	5
		Average	−0.23		0.73
		RMSD	3.76		3.93

[a] The error column shows $K_{i,estimated} - K_{i,observed}$ values.