

# On the interpretation and interpretability of quantitative structure–activity relationship models

Rajarshi Guha

Received: 13 May 2008 / Accepted: 14 August 2008 / Published online: 11 September 2008  
© Springer Science+Business Media B.V. 2008

**Abstract** The goal of a quantitative structure–activity relationship (QSAR) model is to encode the relationship between molecular structure and biological activity or physical property. Based on this encoding, such models can be used for predictive purposes. Assuming the use of relevant and meaningful descriptors, and a statistically significant model, extraction of the encoded structure–activity relationships (SARs) can provide insight into what makes a molecule active or inactive. Such analyses by QSAR models are useful in a number of scenarios, such as suggesting structural modifications to enhance activity, explanation of outliers and exploratory analysis of novel SARs. In this paper we discuss the need for interpretation and an overview of the factors that affect interpretability of QSAR models. We then describe interpretation protocols for different types of models, highlighting the different types of interpretations, ranging from very broad, global, trends to very specific, case-by-case, descriptions of the SAR, using examples from the training set. Finally, we discuss a number of case studies where workers have provide some form of interpretation of a QSAR model.

**keywords** Quantitative structure–activity relationship (QSAR) · Interpretation · Linear regression · Partial least squares (PLS) · Neural network

## Introduction

An unsophisticated forecaster uses statistics as a drunken man uses lamp-posts—for support rather than for illumination.

Andrew Lang (1844–1912) [69]

The use of statistical models to predict biological and physical properties has a long history, starting with linear regression models developed by Hansch [48, 49] in the 1960s. Since then there has been an explosion of predictive models developed using a wide variety of modeling methods. These methods are designed to encode a quantitative structure–activity relationship (QSAR) or a quantitative structure–property relationship (QSPR) by examining the chemical structure of small molecules. Thus, for biological structure–activity relationships (SARs), the receptor is not taken explicitly into account, and, for physical properties, QSPR models do not generally consider the detailed physics of the bulk systems. In other words, QSAR and QSPR models implicitly consider the environment of the molecules being studied, in contrast to methods such as docking and molecular dynamics that are based on a direct physical description of the system. This is not always a bad approximation, since biological activity and physical properties are fundamentally derived from chemical structure. Certainly, some details may be lost by lack of a true physical description, but chemical structure encodes a large amount of information explaining why a certain molecule is active, or is toxic, or is insoluble.

Given the fact that QSAR or QSPR models (hereafter we refer to both types of models as QSAR models) are essentially pattern recognition models, their goal is to identify trends in structural features that correlate with the observed activity. By identifying and encoding these

---

R. Guha (✉)  
School of Informatics, Indiana University, Bloomington,  
IN 47408, USA  
e-mail: rguha@indiana.edu

structure–activity patterns, we can then use the model to predict the property of a new molecule. Ideally, if a model is able to capture most (or all) of the significant correlations between structural features and the observed activities, we should be able to predict reliably the activity of a molecule that the model has not seen before. Conversely, a model that has not been able to encode the underlying structure–activity relationships in a dataset in sufficient detail will not be able to make accurate predictions or new molecules.

Thus, a predictive model provides us with two things: a set of predicted values, and information regarding the SAR(s) that are present in the dataset. In many scenarios, such as virtual screening, investigators employ such models purely for their predictive ability. For such problems, especially when the underlying chemistry is well understood, this use is justified. However, there are many cases in which a predictive model can be used not only to predict the property of a new molecule, but also to *understand* why certain molecules exhibit activity (or inactivity, or toxicity, and so on) and others do not. In other words, not only can we use a model for its predictions, but we can try to extract and understand the SARs that have been encoded within the model. This is especially important for problems where one builds models to explore poorly understood structure–activity relationships. We define the process of extracting the SAR(s) encoded by a predictive model as interpreting the model.

#### Why interpret?

Given that the extraction of a useful interpretation can require some effort, should interpretation be a requirement of the QSAR modeling process? The answer to this depends on the planned use of the model. For example, many QSAR models are built for filtering purposes [12, 18, 25, 55, 87, 109], where the goal is to predict some property rapidly. Such models are generally used as screening tools, allowing one to prioritize molecules from large libraries. One can also consider, in relation to this type of usage, QSAR models that are built for properties that are well understood. Examples include physical properties such as boiling points, vapor pressure and peptide mobilities as well as biological phenomena such as serum albumin binding and intestinal absorption. The mechanistic underpinnings of these properties are well understood<sup>1</sup>, and there is not much utility in developing a QSAR to elucidate the underlying structure–activity relationships. Thus, in such

cases, the focus is on obtaining the most predictive model, and interpretations are secondary.

An alternative use of a QSAR model is explanatory. Such usage can occur when one is considering a biological target for which no receptor structure is known. Even if a structure is available, the actual mechanism of action may be unclear. In such cases, one could develop a QSAR model with the goal of elucidating the underlying structure–activity relationships. Clearly, such situations warrant interpretation of the model, and predictive accuracy may be secondary. This scenario is also important when one is developing global models. In contrast to local models that use structurally homogeneous datasets, usually containing a single, well-defined, SAR, global models may contain more than one SAR. In such a case, an interpretation of the QSAR model can provide insight into the presence and nature of the multiple SARs. Furthermore, since global models can be expected to exhibit poorer predictive accuracy than local models would, an interpretation can enhance their value.

In addition to the broad uses of QSAR models noted above, there are a number of scenarios where an interpretation of a QSAR model can provide insight. An example is the process of inverse QSAR modeling [67]; that is, the use of a QSAR model to suggest structural modifications that will improve activity. A number of approaches to this problem are algorithmic [10] in nature and utilize descriptors [112] designed specifically for this purpose. However, in the absence of any specific descriptors or algorithm, one can perform an interpretation of a QSAR model to extract the details of how specific structural features correlate to observed activity. Given this knowledge, one can then systematically modify structures and obtain predictions, in a manner similar to that described by Lewis [67].

A common feature of QSAR models is the presence of outliers or other anomalous compounds. Many studies identify these using numerical methods (such as leverage and Cook's distance [15]) and, in general, are removed from the dataset. Though valid, it is important from the point of view of an SAR to understand why a compound is regarded as an outlier. What are the structural features that cause it to deviate from the SAR exhibited by the remaining molecules? For the answer to these questions, it can be useful to have a detailed interpretation of a model to understand the main SARs and how an outlying molecule deviates from them.

Finally, one can also view interpretations as a way to confirm the utility of models and descriptors. Thus, for example, when a QSAR model of a well-understood property is being developed for the purposes of filtering, it can be instructive to extract the interpretation to ensure that the known SARs have, indeed, been captured by the model.

<sup>1</sup> It should be noted that even though the problem may be understood from a mechanistic point of view, it is still possible for one to derive poor QSAR models, since the numerical *characterization* of the mechanistic features responsible for the property might be inaccurate or incomplete.

If not, this could be an indication that additional or alternative descriptors are required. Similarly, when new descriptors are being developed, an interpretation of the resultant QSAR model would allow one to confirm that the descriptor is, indeed, capable of characterizing the structural features important for the SAR.

## Outline

In this paper we address the issue of model interpretability in the field of QSAR modeling. In the section “Interpretability” we discuss the factors affecting interpretability, including the relative ease with which certain types of models can be interpreted (section “Models”) and the interpretability of molecular descriptors (section “Descriptors”). As we shall see, it is possible to use a simple modeling technique that is easy to interpret, but the actual interpretation is opaque since the descriptors used to build the model are extremely abstract. In the section “Interpretation methodologies”, we discuss some methods that have been designed specifically to aid the interpretation of QSAR models. In the section “Case studies” we discuss a number of case studies, where investigators have developed QSAR models and then have attempted to extract the SAR encoded by these models.

## Interpretability

Interpretations of QSAR models are contingent upon the fact that one can examine the SAR encoded by a model (i.e., the model should not be a black box). Given a model that is amenable to interpretation, we must then address the fact that the descriptors that constitute the model may or may not have simple physical interpretations. We first address the issue of the interpretability of modeling methods.

## Models

The mathematical and statistical communities are the primary source of modeling methods employed in QSAR studies. There are many ways to categorize these techniques, such as linear and non-linear methods. In the context of interpretation it is convenient to consider two classes of methods [7]: model-based and model-free (also known as algorithmic models).

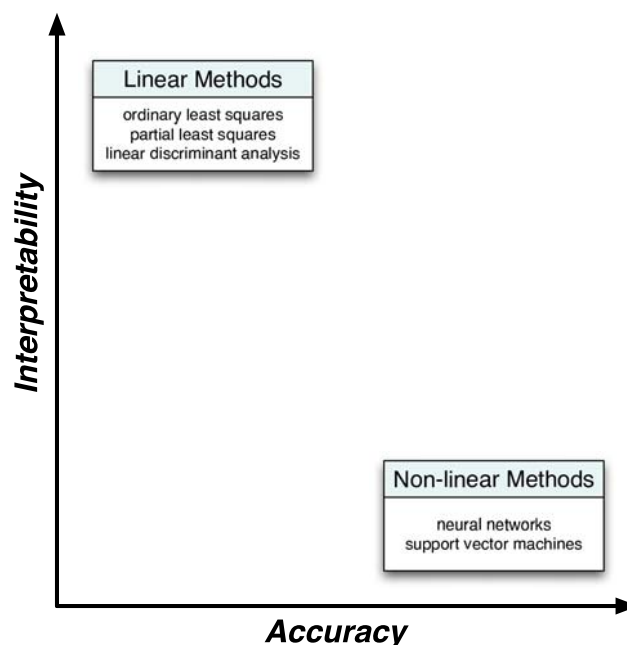
Model-based methods aim to describe the data in terms of a statistical distribution. In other words, they attempt to model the underlying process that gives rise to the observations. Examples of such methods include ordinary linear regression and partial least squares. Obviously, certain assumptions are made (such as normality in the case of

linear regression). Within these assumptions, the resultant model can be said to have *explanatory* power in addition to predictive ability.

Model-free methods, on the other hand, make no attempt to explain the observed data. Rather, they focus on predictive ability. Examples of such methods include *k*-nearest neighbors (*k*-NN) and random forests. In a number of cases these methods do not (and cannot) provide any explanation of why a certain observation is predicted the way it is. On the other hand, certain methods such as decision trees and random forests are able to provide some insight into the underlying SARs by virtue of design.

When we consider the types of models that can be interpreted, we observe that, in many cases, the interpretability of a model is a trade-off with predictive accuracy, shown schematically in Fig. 1. For example, linear regression models can be interpreted in a detailed fashion, but, generally, have lower accuracy, especially for biological activities. On the other hand, one can achieve high accuracy using a neural network model, but extracting the encoded SAR can be very difficult. In some cases, a model is interpretable by virtue of the underlying design (such as decision trees and Bayesian networks) and does not require extra effort to extract the SAR, whereas other methods (linear regression, random forests and neural networks) require an *interpretation protocol*.

Another aspect of the interpretability of a model is the nature of the problem. Specifically, is one performing a classification or a regression? In the former case, interpretations are generally broader by virtue of the fact that a



**Fig. 1** A schematic diagram highlighting the trade-off between accuracy and interpretability

categorization is being performed, rather than a ranking. In the latter case, interpretations can be more detailed and explain why one molecule is predicted to be more (or less) active than another. In this paper, we focus on regression models.

Given a model that is interpretable, we can then consider the level of detail that is possible or desired. For example, one can examine the regression coefficients in a linear regression model to understand which descriptor plays an important role in the predictive ability of the model. Similarly, one can use randomization methods to examine the role of descriptors in the context of predictive ability of a neural network or random forest model. We describe such interpretations as *broad*. In other words, we may understand which descriptor is important for the predictive ability of the model, but we do not gain insight into how model descriptors interact with each other or specific examples of the encoded SAR from the training set. On the other hand, certain models can be analyzed to provide a *detailed* interpretation. In such interpretations we not only understand which descriptors are important, but we also gain insight into how the effects of one descriptor may balance that of another, as well as specific examples from the training set that highlight the SAR encoded by individual descriptors. It should be noted that, for certain model types, one may be restricted to a broad interpretation. On the other hand, if a model type does allow a detailed interpretation, it is probably worth the effort to extract it.

One final aspect of models that can affect interpretability is their scope. It is common knowledge that models developed on homogeneous datasets will exhibit good predictive ability and statistical significance, but they will have limited applicability [41]. From an interpretation point of view, such models are nice because there will most probably be a single distinct SAR, and, thus, an interpretation can be straightforward. On the other hand, when one builds models on very heterogeneous datasets, it is possible that there may be multiple SARs present. For such global models, an interpretation is certainly more challenging. However, at the same time, an interpretation is much more important, since it can shed light on the presence of multiple SARs (though this may also be perceived using numerical methods such as clustering) in addition to providing details of the encoded SARs. A number of examples are available in the literature, where workers have extracted detailed interpretations from global models [38]. It should be noted that an interpretation on its own does not guarantee that one will extract all the SARs that may be present in the dataset. Clearly, if a set of features responsible for a specific SAR is not captured by the molecular descriptors, that SAR will be ‘invisible’ to the model.

## Descriptors

A wide variety of molecular descriptors [102] is available for use in QSAR model development. Many programs are able to generate hundreds, and even thousands, of descriptors. Though the goal of a descriptor is to characterize a molecular feature numerically, there are many different ways to do so, some of which are based on physical aspects of the molecule, whereas others are more abstract. Broadly, we can consider three classes of descriptors: topological, geometric and electronic, which differ in their interpretability.

Topological descriptors consider the connectivity of the molecule, though certain topological descriptors such as  $\chi$  [60] and E-state [61] descriptors will take into account the nature of specific chemical groups. Certain topological descriptors are interpretable, especially, when they are fragment based (such as the various  $\chi$  descriptors). In such cases, one can usually draw a connection to molecular size and branching. On the other hand, topological descriptors that are obtained from mathematical graph theory tend to be more abstract. Examples include the eigenvalues of the adjacency matrix and Cluj numbers [26]. Though such graph invariants may lead to useful predictive models, it is usually difficult to connect the values of such descriptors to some physical feature of the molecules. As a result, topological descriptors have a reputation for being uninterpretable. Indeed, Randic et al. [84] noted that graph theoretical descriptors “*may be of lesser interest for structure-property and structure-activity relationships*”. A number of workers have attempted to provide interpretations for certain classes of descriptors, such as row sums of the topological distance matrix [84], Wiener, Hosoya Z and connectivity ( $^1\chi$ ) indices [31, 83]. Garcia-Domenech et al. [35] provide a comprehensive review of various topological descriptors and discuss the issue of their interpretability.

Geometric and electronic descriptors, which we refer to as “physical” descriptors, on the other hand characterize molecules in a physically interpretable manner. Examples include moments of inertia, and highest occupied molecular orbital (HOMO) and lowest-energy unoccupied molecular orbital (LUMO) energies. A number of physical descriptors are designed to characterize the distribution of molecular properties over a molecule, either in terms of the molecular surface (such as charged partial surface area [94] and hydrophobicity descriptors [95]) or in terms of distances (such as radial distribution function [52, 53] and autocorrelation descriptors [73]) or in terms of molecular interaction fields [101, 116]. It should be noted that, in certain cases, a physical descriptor may be not be interpretable by virtue of the underlying mathematical form. For example, the Burden-CAS-University of Texas

(BCUT) descriptors [78] are eigenvalues of the atomic property-weighted Burden matrix [11]. Linking the eigenvalues back to the molecular structure can be a difficult task, though Masek et al. [70] have recently shown that one could derive connection tables from the values of a set of BCUT descriptors. Though this can be considered one form of interpretability, it is indirect and more challenging than descriptors with a clear physical connection to molecular structure.

Fingerprints can also be considered a class of descriptors, combining constitutional and topological features of a molecule. Though fingerprints were originally developed for searching chemical databases, they have proven to be useful as descriptors in developing a variety of predictive modeling studies [12, 33, 45, 51, 92, 117]. The goal of any fingerprint is to characterize the substructures present in a molecule. The primary difference between fingerprints is what substructures are characterized and how they are represented within a bit string. A well-known example of pure substructural fingerprints (also termed substructural keys) are the 166-bit Molecular Access System keys [28]. In these types of fingerprints each bit position corresponds to a specific substructural feature. From the point of view of interpretation these descriptors can be very useful, since there is no abstraction. When combined with certain methods such as decision trees or Naïve Bayes, one can easily understand which structural aspects of a molecule affect its predicted property, as encoded by the model. Another type of fingerprint that addresses substructures, albeit in a slightly abstracted form, are pharmacophore fingerprints such as Chemically Advanced Template Search [86]. In these types of fingerprints, each bit position corresponds to the occurrence of two pharmacophore groups separated by a given topological or geometric distance (though occurrence counts may also be employed, rendering a real-valued fingerprint). Another well-known class of substructural fingerprints is circular fingerprints. The features characterized by these types of fingerprints represent *neighborhoods*, centered on individual atoms and extending out to  $k$  bonds (where  $k$  usually varies from 2 to 6). A variety of circular fingerprints has been described [5, 9, 82]. One of the most popular examples of this class of fingerprints is the extended connectivity fingerprint (ECFP) and functional class fingerprint (FCFP) implemented in Pipeline Pilot (Scitegic Inc.), in which atoms in the neighborhood are described in terms of Daylight atomic invariants [113] or in terms of functional class (hydrogen bond donor, aromatic, etc.). It is important to note that most circular fingerprints are hashed fingerprints, where a given substructure is hashed into the bit string. When dealing with hashed fingerprints in general, one cannot usually link a bit position to a specific substructure (indeed, a bit position may link to one or more different substructures).

As a result, though they can be very useful for developing predictive models, they do not allow one easily to understand the substructures that are important for predictivity.

Finally, one can also consider ‘property descriptors’, which represent a specific molecular property. Examples would include  $\log P$  and molar refractivity. In many cases, these descriptors are calculated rather than being experimentally observed. However, they represent the property directly, without any intervening abstraction and, thus, do not require interpretation as such.

Clearly, the choice of descriptors strongly influences the extent to which one can interpret a model and extract SAR trends. Descriptor selection (also known as feature selection) is an important step in the model development process, and many methods have been employed [1, 71, 88, 97, 108]. One aspect of the feature selection process is that it tends to be automated and will identify descriptors based on their ability to create a predictive model and not based on their interpretability. One could certainly manually choose a descriptor set, utilizing prior knowledge of the system that is being modeled. If sufficient information is available, this may be the best approach to identifying an interpretable descriptor subset. However, in many cases, one does not necessarily know all the details of the system being studied, and the goal is to use a model to characterize the SARs in the dataset. In such cases, automated feature selection is preferred, and it is imperative to have a descriptor pool that contains descriptors that will allow interpretation. Given the trade off between accuracy and interpretability, it has been suggested [29] that one build two models using the same set of descriptors. Specifically, the method uses a genetic algorithm to select descriptors that are simultaneously optimal for a linear regression and a neural network regression model. This implies that neither model is the optimal model. However, it does ensure that the same SARs are encoded by both models. The result of this method is that the linear model can be used for ease of interpretability, whereas the non-linear model can be used for its higher accuracy.

### Interpretation methodologies

As noted above, certain methods such as decision trees can be interpreted by design. For other models, one is required to utilize an interpretation protocol to extract the encoded SAR.

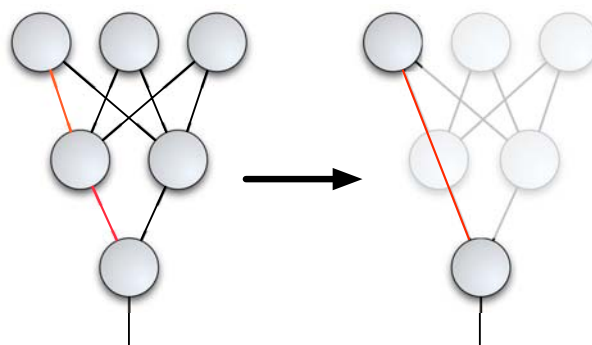
We first consider linear regression models. In the absence of any extra information one can gain a broad view of the SARs encoded in the model by simply considering the magnitude and signs of the model’s coefficients. This is an example of a broad interpretation, and though easy to perform, it is rather superficial. A detailed interpretation

can be obtained by use of a technique based on partial least squares (PLS) [93]. Briefly, the technique develops a PLS model using the descriptors from the original linear regression model. The PLS model provides one with a set of latent variables (also called components), which are linear combinations of the input descriptors. We can then arrange the components in order of their ability to explain the variance in the  $Y$  variable. Within a component, we can then identify the most important descriptors by their magnitude and the nature of their effect on the predicted property by their signs. By examining successive components, we can understand how the model correlates different descriptors to the property. Furthermore, the procedure is able to highlight molecules that confirm the trend in a given component and molecules that cannot be explained by a given component. We see that, by considering the next component (which usually focuses on a different descriptor), one can address the SAR exhibited by molecules that could not be explained by the preceding component. Thus, in contrast to simply looking at the regression coefficients in the original model, the PLS technique is able to dissect the regression model and explain in detail how the descriptors have encoded the SAR in the dataset. Furthermore, it allows us to highlight specific molecules in the dataset that exhibit (or do not exhibit) specific aspects of the encoded SARs. Of course, one can also build the PLS model directly, without building a prior linear regression model. A variant of the PLS approach is the use of hierarchical PLS [30], which allows one to use a large collection of descriptors without prior feature selection. The method then considers ‘blocks’ of descriptors, which are analyzed in a hierarchical manner and have been used to develop interpretable models for a variety of problems (carcinogenicity [81], human immunodeficiency virus (HIV) protease inhibitors [62] and human ether-a-go-go-related gene (hERG) inhibitors [38]). Katritzky et al. [57] discussed the use of principal components analysis (PCA) as a way to interpret QSAR models. However, it appears that their approach does not directly interpret the QSAR model itself, but rather, it is used to summarize a data matrix obtained from predictions from the QSAR model. As a result, it is more a broad and indirect interpretation method.

Random forests are a popular QSAR modeling method for a variety of reasons, including resistance to over-fitting and implicit feature selection. Interpretation of random forest models is generally restricted to a broad interpretation based on variable importance plots [8]. The model is initially built using the supplied descriptor pool, and the mean square error (MSE) is evaluated. The model is then rebuilt, with the first descriptor scrambled. The MSE is recorded, and the procedure is repeated, each time scrambling one of the descriptors. Once the procedure has been

completed, we plot the descriptors versus the decrease in MSE, compared with that for the original model. Thus, if a descriptor is important for the model’s predictive ability, then the model built using the scrambled version of the descriptor will exhibit a much higher MSE than will a model built using a scrambled version of a descriptor that is not important for the model’s predictive ability. An alternative interpretation approach for random forests is to explore the forest itself [104]. This is more complex than the descriptor importance method and can be subjective. However, with the use of suitable graphical tools [105], this method can be used to provide more detail than descriptor importance.

We next consider neural networks. A neural network encodes the SARs present in the dataset within the weights and biases that define the connections. A number of efforts have been made to interpret these weights and biases [14, 37, 75, 90, 98, 99], though it was only recently that neural network QSAR models were considered. One approach to interpreting neural networks is based on linearizing the network [46] and is shown schematically in Fig. 2. This protocol was designed to be analogous to the PLS method for linear regression models. Thus, the hidden neurons in a neural network model were considered to be the equivalent to latent variables in a PLS model. The method then uses the weights and biases to order the hidden neurons in terms of their increasing degree of contribution to the output layer. The method has two main drawbacks. First, in considering the hidden neurons as a form of latent variables, it ‘linearizes’ the network, and, thus, valuable details of the encoded SAR may be lost. Second, the method is applicable to only fully connected, three-layer feed-forward networks. Neural network models can also be broadly interpreted using a descriptor scrambling procedure in a manner analogous to random forests and has been described [44]. It is clear that the descriptor



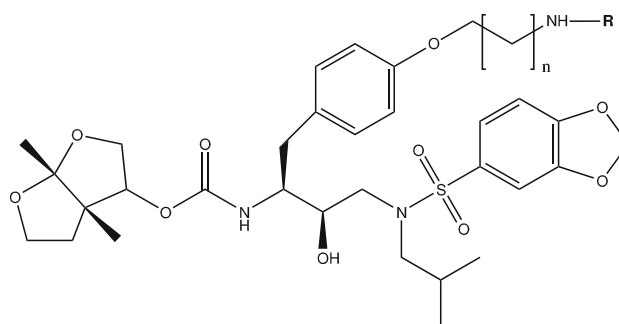
**Fig. 2** A representation of an interpretation scheme for feed-forward neural networks, using a linearization scheme [46]. The shaded nodes in the left hand network indicate that the interpretation procedure only implicitly uses the hidden node information

randomization approach is quite similar to the sensitivity analysis described above. While it does provide one with an overview of which descriptors are important (in effect, a form of feature selection), it does not really provide a detailed view of the SARs encoded by the model.

Finally, we consider support vector machines (SVMs), which are known to exhibit high predictive accuracy but are black boxes, due to their use of the kernel trick [91]. There has been recent work that has addressed the interpretability of these models. For example, Cho et al. [17] described an approach using a specialized kernel function and a nomogram [4], though this was more a visualization than an interpretation. Navia-Vázquez and Parrado-Hernández [74] described an approach to interpreting SVM classification models based on segmenting the input space using the prototypes extracted from the trained model. In the area of QSAR modeling, Usdun et al. [106] described an approach to visualizing and interpreting support vector regression (SVR) models. Their approach considered the ‘kernel matrix’ (the result obtained from mapping the input space to the higher dimensional space via the kernel transform) and evaluated the correlation between each input descriptor and each row of the kernel matrix. Visualization of the resultant ‘correlation’ matrix allows one to extract the importance of each input descriptor to the kernel matrix. This approach is similar in idea to the descriptor scrambling methods for random forests and neural networks. In contrast to those methods, the correlation matrix method for SVMs does not directly identify descriptors that are important to overall predictive ability. Usdun et al. also described a procedure which generated a ‘loading plot’, based on the \* values obtained from the quadratic programming step of the SVR algorithm [107]. Taken together, the interpretation procedure allows one to obtain a relatively detailed view of the SAR encoded in SVR models.

### Case studies

We now consider a number of published QSAR studies built for the following systems: anti-HIV targets (HIV protease and reverse transcriptase), anti-malarials (artemisinin analogs), dihydrofolate reductase (DHFR) inhibitors, and some models developed for absorption, distribution, metabolism, excretion and toxicity (ADMET) properties. We primarily focus on studies that have developed linear models, since they are easily interpretable and one would expect some form of interpretation to be provided. We do not consider studies that used 3D methods, such as comparative molecular field analysis (CoMFA) and variants, since they, by definition, allow one to investigate molecular interactions directly.



**Fig. 3** The amprenavir scaffold that was used by Miller et al. [72] to derive HIV protease inhibitors

### Anti-HIV targets

A number of studies have focused on molecules that target HIV protease. Ravichandran et al. [85] developed a series of models to characterize the ability of a set of arylsulfonamides originally synthesized by Miller et al. [72] to inhibit HIV-1 protease. They described three linear regression models, developed using physical descriptors (heat of formation, log *P*, molar refractivity and solvent-accessible surface area). The original experimental work investigated the role of substitutions at the *para* position of the P1 phenyl group on the amprenavir scaffold (Fig. 3). As shown in the synthetic study [72], the length of the tether, as well as the nature of the *para* substituent, played a role in determining the inhibitory activity. Ravichandran et al. provided a broad interpretation of one of their models by simply considering the regression coefficients. Thus, they noted that the positive value of the coefficient for log *P* indicated that substitutions that increase hydrophobicity will lead to better inhibitors. Similarly, they stated that an increase in heat of formation would enhance inhibitory activity, and an increase in volume would be detrimental. Those observations are trivial conclusions that one can draw from the regression equation and do not really provide any insight into the actual mode of action of these compounds. Furthermore, they do not indicate what structural features of the training set molecules play a dominant role in the activities of the compounds. As noted by Miller et al. [72], an increase in tether length from  $n = 1$  to  $n = 3$  resulted in increased activities for primary amine derivatives but no such trend for carbamate derivatives. This could be correlated to the log *P* and could have been investigated in detail using the PLS interpretation technique. Leonard and Roy [66] described a set of models built to model the anti-HIV activity of a set of thiocarbamates. Though they used linear regression and PLS to develop predictive models, the interpretations provided were derived directly from the signs and magnitudes of the linear regression coefficients. Furthermore, though they

attempt to correlate SAR trends from the regression coefficients to specific molecular features, the lack of detail inherent in such an approach does not really allow strong conclusions. Furthermore, they noted the presence of three outliers, and, in the absence of a detailed interpretation, were unable to explain their behavior as outliers. A number of other models are also described in that study and are interpreted in a similar, general, manner. However, the interpretations reported are valid and correspond to experimental observations, but they lack a more detailed analysis linking specific descriptors to structural features in the training set and how these affect the observed activity.

De Lucca et al. [22] reported an experimental study of the structure–activity relationships of a series of tetrahydropyrimidin-2-one analogs. The authors noted the importance of the lipophilic P2 groups as well as hydrogen bonding groups at the *meta* position in the *N*-benzyl substituents. In general, hydrogen bonding played an important role in the activity of a number of analogs. There have been a number of computational studies with this dataset, and we focus on those that employed 2D-QSAR methods (as opposed to 3D approaches). Katritzky et al. [58] presented the results of a QSAR study in which they developed a series of QSAR models ranging in size from one to four descriptors. The models employed physically interpretable descriptors. However, the authors restricted themselves to interpretations derived simply from the model's regression coefficients. As a result, the general trends exhibited by the model matched the SARs described by De Lucca et al. However, the details of the SAR encoded by their models were not fully extracted, and even though the authors do provide some explanation of outliers in their models, it would have been useful to explore the components of the SAR explicitly, using examples from the training set. Similarly, Garg and Bhatarai [36] developed a series of one-descriptor models using log *P* and molar refractivity descriptors. Since the models used one descriptor, interpretation was relatively simple and matched the experimental observations noted by De Lucca.

We next consider models built to characterize inhibitors of HIV integrase. Sahu et al. [89] described models built to predict the inhibitory activity of caffeoyl naphthalene sulfonamide derivatives. Though this class of compounds exhibits a number of structural features common to known integrase inhibitors, it does not exhibit significant inhibitory activity. The goal of the models was to identify features that would allow the design of more potent inhibitors. The authors employed linear regression and used physically interpretable descriptors. Unfortunately, the authors provided very little analysis of the SARs encoded by the model beyond the trivial conclusions drawn by inspecting the regression coefficients. If the goal of the model was to suggest structural modifications to increase

potency, it is not clear how the analysis presented would help. On the other hand, Yuan and Parrill [114] described the development of a linear regression model using a combination of geometric and topological descriptors, to predict the concentrations that would reduce the effect by 50% (IC<sub>50</sub>s) of a diverse set of compounds. Though their interpretation simply examined regression coefficients, they were able to gain some insight by combining an analysis of the model with results from clustering. However, their conclusions would have been strengthened by highlighting the SARs that they extracted using examples from their training sets.

In addition to the linear regression QSARs described above, a large number of studies have been published that used non-linear methods, such as neural network methods and support vector machines [3, 32, 80, 100, 111]. In general, these studies focused on the accuracy provided by the non-linear method. When interpretations were provided, they tended to focus broadly on descriptor importance.

#### Anti-malarials

We next consider a number of QSAR models developed to study anti-malarial compounds. Cruz-Monteagudo et al. [21] developed models to predict the anti-malarial activity of a set of 42 bisbenzamidines. They described a three-descriptor model, using the GETAWAY descriptors [20]. They describe a “desirability analysis” whose aim is to determine the range of values of the descriptors that lead to the best activity. However, they do not correlate these values to the structural features that would give rise to them. Furthermore, in their interpretation of the model, the conclusions drawn are rather trivial and broad, though they match experimental observations. They do provide a detailed description of the descriptors themselves. In addition, they were able to link indirectly the descriptor information to the fact that loss of planarity correlated with a decrease in anti-malarial activity, and they highlighted this with some specific examples. One drawback of the study is that they removed outliers, without being able to explain why they should be removed. This is striking, since one of the outliers that was removed was the most potent compound in the collection studied. A detailed analysis of the model and comparison to other molecules using the PLS technique could be used to shed light on why the outlier was regarded as such.

Zahouily et al. [115] described a linear regression and neural network QSAR model built for a set of 63 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl derivatives. The experimental data that this study was based on indicated that the hydrophobic nature and steric characteristics of the substituents played an important role



in determining anti-malarial activity [68]. As a result, the authors employed a set of physical descriptors, including size, shape, log *P* and counts of hydrogen bonding donors and acceptors. However, their interpretation of the linear regression model was restricted to a broad interpretation, based on examination of the regression coefficients. Once again, the broad conclusions drawn from the model matched the experimental observations [68], but they did not provide insight into how the model had captured the SAR, in terms of specific examples, or any analysis of the interplay between molecular features that would explain aspects of the SAR. It is interesting to note that the authors provided a broad interpretation of the neural network model, using a modification of the descriptor randomization method described by Chastrette et al. [14].

A number of studies have been performed on collections of artemisinin [63] analogs. For example, Katritzky et al. [59] built a number of linear regression models using both physical and information theoretic and topological descriptors. Though the authors do describe connections between the more abstract descriptors and physical properties of the molecules, the actual interpretation is rather broad. In contrast to the other examples, where workers have simply considered the regression coefficients, this work employs the *t* statistic for each of the coefficients, to decide the relative importance of the descriptors. This could be considered more rigorous than considering just the regression coefficients, assuming that the correlations between the descriptors are very small (which was not discussed in the paper). In general, their conclusions match previous observations but do not provide much detail explaining the activity of the specific molecules studied.

However, a number of studies that developed 2D-QSAR models have provided reasonably detailed interpretations. For example, Pinheiro et al. [79] described the development of PLS models to predict the anti-malarial activity of a set of artemisinin derivatives. Their study used physically interpretable descriptors, and, due to their use of PLS, they were able to highlight the SARs, using examples from the training set. They were also able to justify the conclusions by use of a docking procedure and prior experimental observation. Guha and Jurs [43] developed a linear regression model and a neural network model to predict the activities of a set of artemisinin analogs. They provided a detailed interpretation of the linear model, using the PLS technique [93], and were able to highlight specific characteristics of the SAR with reference to examples in the training set. The interpretation was limited to some degree, due to the fact that the linear model was comprised primarily of topological descriptors. Girones et al. [39] described a series of linear regression models, in which the features were derived from a molecular quantum similarity measure (MQSM) [13] matrix, by the use of principal

components analysis. The authors then used the principal components plot to identify specific molecules that highlighted different aspects of the SAR. An interesting feature of that study was that rather than directly analyzing the model, the principal component plots were used as a clustering method. One reason for this was that the principal components of the MQSM merged a number of molecular features. As a result, even though the principal components of the MQSM matrix contained all the relevant information, it is difficult to relate that back to the actual structural features.

In contrast to the studies described above, a number of studies [2, 6, 76] did not provide any interpretation of the models. These cases presented models (primarily linear regression) developed using abstract descriptors, whose physical interpretability was difficult. Furthermore, most of these types of studies used relatively small datasets. Given the small number of molecules, an interpretation of the encoded SAR would have made the models significantly more useful.

#### DHFR inhibitors

Dihydrofolate reductase (DHFR) is an enzyme catalyzing the conversion of dihydrofolate to tetrahydrofolate, which is involved in the synthesis of purines, thymidylate, and so on [96]. As a result, it is an important therapeutic target for a variety of diseases, such as bacterial infections and malaria. Crystal structures of the enzyme bound to different ligands have been obtained [34, 64, 96], and the molecular interactions required to inhibit the enzyme successfully, have been described [96]. Thus, QSAR models of inhibitors should be able to identify ligand features that will correspond to the previously described interactions.

A number of studies [23, 24, 54] of DHFR inhibitors have employed Hansch substituent constants [50] as descriptors. The majority of these studies developed linear regression models and did not utilize any specific interpretation protocol. However, due to the physical nature (characterizing features such as hydrophobicity and molar refractivity) of the Hansch constituents, the interpretation of the models was relatively simple (especially since they were usually one- or two-descriptor models). Furthermore, in a number of cases, the SARs were further examined by analysis of crystal structures [24], confirming the conclusions drawn from the linear models. On the other hand, a study by Chin and So [16] developed linear regression and neural network models to study the inhibition of DHFR by a set of 68 2,4-diamino-5-(substituted-benzyl) pyrimidine derivatives. However, the authors simply described the statistics of the models, and the extent of the interpretation was to note that certain substituents were selected in the

model based on their importance to the activity of the molecules. Of course, the use of the Hansch constants does limit the detail of any interpretation of such models, as they do not necessarily characterize sub-structural features. The work by Otzen et al. [77] is an interesting study, since it combined experiment and QSAR modeling. More specifically, they developed QSAR models for a set of sulfone derivatives to investigate their ability to inhibit dihydropterotic acid synthase. The used descriptors derived from experiment [fraction of unionized sulfone and proton nuclear magnetic resonance (NMR) shifts]. The simplicity of the model and the use of physical descriptors certainly make the model interpretable. Though the authors did not use any explicit interpretation protocol, they were able to correlate the effects of the descriptors in the model to experimental observations.

### ADMET models

A number of QSAR models have been developed to predict various ADMET properties. For example, much work has been carried out on the predictive modeling of cytochrome P450 metabolites. Though some methods have employed structures of the active site or mechanistic models to predict such metabolites, a number of studies have been performed that are exclusively ligand based. For example, Sheridan et al. [92] developed a set of random forest model to predict the regioselectivity of oxidation by different members of the cytochrome P450 (CYP) family. The model was developed using a variety of substructure and physical property descriptors. Sheridan et al. employed the descriptor importance measure [8] to highlight important sub-structural and physical features that correlated to oxidation probabilities. In this case they were able to correlate such sub-structural features to previous studies which had identified common mechanisms (such as *O*-dealkylation of aromatic methoxy by 2D6) or structural features relevant to CYP oxidation (such as carbons in piperidines not being oxidation sites for CYP 3A4) [92]. Though broad in nature, the choice of descriptors allowed them to provide a thorough interpretation of the SARs encoded by their models. Gleeson et al. [40] described a set of PLS and recursive partitioning models to predict the inhibition of CYP enzymes by a set of drug and drug-like molecules. They specifically chose to employ physically interpretable descriptors and utilized the PLS coefficients to characterize the encoded SARs. Given the use of PLS, the interpretation in this study is quite detailed. However, the authors did not attempt to highlight specific examples from their training set that would have given concrete examples of the SARs they described. Similarly, Verma et al. [110] described a QSAR study of the absorption [by parallel artificial

membrane assay (PAMPA) and Caco-2 cells] of a set of drugs. The authors developed linear regression models using log *P* and various indicator variables, but they used only the regression coefficients to draw broad conclusions regarding the structural features that contributed to or hindered absorption, and they did not perform detailed analyses of the training set itself.

Gunturi et al. [47] performed a study on the prediction of the serum albumin binding affinity of a set of drug and drug-like compounds [19]. Though the focus of the paper was a feature selection mechanism, the authors developed linear regression models. The authors did provide a comprehensive description of the descriptor meanings and identified significant descriptors based on the frequency of occurrence in their feature selection routine. Unfortunately, the authors did not really perform any analyses of the models themselves and rather simply drew broad conclusions from the nature of the selected descriptors.

### Discussion

Recent reports assessing the utility and validity of QSAR modeling [27, 56] appear to indicate that the use of QSAR models purely for their predictive abilities has led to a somewhat inflated view of their utility. Part of the disenchantment with QSAR modeling can be ascribed to poor practices. However, it is also true that QSAR methods, being indirect descriptions of a physical or biological system, may not be as informative as structure-based or fully physical methods. Yet, given the wide variety of molecular descriptors and modeling methods, there is no doubt that a QSAR model can capture many details of a structure–activity relationship, with much less computational effort. The path to this goal is not necessarily easy and clear-cut. Though the individual steps of a QSAR modeling protocol have been discussed by many workers, the plethora of papers on each of these steps indicates that no definite answer has been reached. Indeed, it is highly probable that there is no ‘best’ solution to set composition, feature selection and modeling technique. Given these observations, it is all the more important to extract as much as we can from a QSAR model, though it is also true that all QSAR models are not designed to provide insight into the underlying SARs (such as filtering [12, 18, 25, 55, 87, 109] models). In these cases, the underlying structure–activity relationship is usually well known. As a result, interpretability is a secondary issue, and the focus is on speed and accuracy.

One aspect that we have not considered in detail here is the difference between QSAR models for biological activities and physical properties. The structural features that are responsible for a variety of physical phenomena are

relatively well understood. In such cases, it is usually a matter of identifying descriptors that will be able to capture these features and a modeling method that will exhibit high accuracy when predicting the physical property. In a sense, such models correspond to the idea of filtering models described above. However, there are cases of physical properties which may be a combination of several effects, in which a QSPR model can be used to elucidate the details of the individual effects.

When models for biological activities are being developed, interpretations become more significant. One reason for this is that many QSAR models are focused on a specific system and, thus, are not generally applicable (in contrast to something like boiling point). Furthermore, there are usually multiple phenomena underlying a given biological activity. As a result, a QSAR model has the potential to extract these phenomena and rationalize the biological activity for a set of compounds. The key term here is ‘rationalize’. That is, we not only want to predict the biological activity, but we should also be able to understand the cause of the activity. In other words, what is the SAR that causes some molecules to be active and some to be inactive?

In this context, it is understandable why many QSAR and QSPR models may not be interpreted. Many of the models built for biological activities are usually based on small training sets. It is unfair to expect that models built on such datasets, which may, in some cases, also be heterogeneous, will exhibit a high degree of predictive ability in general (even though they may exhibit very good training set and validation statistics [41, 103]). Given these characteristics, one would not use such models as filtering tools. What is more important is that when building models on small datasets, one should make the effort to provide a comprehensive interpretation of the SARs identified by the model. This can be justified in a number of ways. First, this would force modelers to employ interpretable descriptors or else provide meaningful discussion of how an abstract descriptor correlates to chemical structure. Second, by extracting the encoded SAR, one could gain some measure of confidence in the model itself (beyond training set or prediction set statistics). In other words, if the SAR that one extracts from a model does not make physical sense, it might be indicative of a problem with the data, descriptors or modeling method. Third, when we are building models for systems where prior knowledge of the SAR is available, an interpretation allows us to confirm that our models make sense in light of previous work. The strategy of ensuring that mathematical models make physical sense is not new. However, the ease of descriptor calculation and graphical user interface based modeling packages have made it very easy for modelers to generate a large number of models quickly, without having to think about the specifics of each

step. We reiterate that this should not be construed as a general statement—there are a number of examples where workers have thoroughly analyzed descriptors and models [3, 42, 43, 65, 92, 95]. It is just not as widespread as one would like.

When we do consider models that have attempted to provide some form of interpretation, we can see a significant variation in the quality of the interpretations. In cases where the modeling technique does not allow significantly detailed interpretation, broad conclusions can be justified. However, many papers describe linear regression and PLS models. These are simple to interpret and can provide much insight into the encoded SARs if thoroughly interpreted. Yet, the bulk of papers describing such models simply consider the regression coefficients and draw very broad conclusions. In very few cases are the conclusions (i.e., SARs) explained with respect to specific examples in the training set. It may appear that such focused interpretations, addressing individual molecules, contradicts the goal of QSAR in generalizing from structures to property. We do not believe that this is the case. As described below, such interpretations can act as a check on the generalizations made by the model. Indeed, the presence of molecules in the training set that do not follow the encoded SAR (i.e., an outlier) can provide useful information on the limitations of the generalization made by the model. In addition, investigating individual molecules and how they represent the SARs encoded by the model allows us to provide specific examples of the encoded SARs, as opposed to simply saying that a molecule is active or inactive. Given that this activity is restricted to the training set, it does not detract from the use of the model in predicting activities of new molecules. Rather, by having specific example of the SARs encoded by the model, one can use it to help understand why a new molecule is regarded as active.

Finally, we must consider the possibility of over-interpreting a model. To a large extent, this depends on how one connects the descriptors to the chemical structures. Though not such a problem with physical descriptors, it is possible to read more into some topological descriptors than is really contained within them. From the point of view of the model, methods that allow only broad interpretations can be at risk. Though one can identify the most important descriptors from a random forest model, such an interpretation does not give us enough information to describe a specific SAR. On the other hand, for linear models, it is possible (though not commonly performed) to extract SARs and identify molecules in the training set that exhibit these SARs. In such cases, one is prevented from over-interpreting by virtue of the molecules in the training set. Of course, this does not prevent one from generalizing further—but that is probably not the fault of the interpretation.

## Conclusions

This paper has attempted to highlight the need for interpretations of QSAR models (focusing on the 2D variety) and has presented various aspects of the practice of model interpretation, along with cases studies. Given the fact that QSAR models indirectly characterize a physical or biological system, some form of interpretation allows us to go from pure numerical descriptions of the problem to a more physical description of the structure–activity relationships. In many scenarios, not interpreting a model could be considered a waste of the modeling effort. Of course, many factors affect the extent to which we can gain an understanding of the chemistry and biology underlying the SAR. Depending on the descriptors and modeling method, we can obtain interpretations ranging from very broad and general to a very detailed, case-by-case interpretation. Current software technologies have made the process of building QSAR models very easy, but these tools focus on the numeric aspects of the problem. Though accuracy and statistical significance are important, practitioners of QSAR modeling should not be blindsided by high  $r^2$  values and good  $F$  statistics. In the end, the goal of a QSAR model is to capture SARs present in data, and it should be the goal of the modeler to extract and understand the SARs encoded by the model. As the opening quote suggests, practitioners of QSAR modeling should aim for illumination and not just support.

**Acknowledgments** I would like to thank Prof. Gerald Maggiora and Dr. David Stanton for useful comments on the issues underlying interpretability.

## References

- Agrafiotis DK, Cedeño W (2002) Feature selection for structure-activity correlation using binary particle swarms. *J Med Chem* 45:1098–1107
- Agrawal V, Sharma R, Khadikar P (2002) QSAR studies on antimalarial substituted phenyl analogues and their  $n^{\omega}$  oxides. *Bioorg Med Chem* 10(5):1361–1366
- Arakawa M, Hasegawa K, Funatsu K (2006) QSAR study of anti-HIV HEPT analogues based on multiobjective genetic programming and counter-propagation neural network. *Chemom Intel Lab Syst* 83:91–98
- Banks J (1985) Nomograms. In: *Encyclopedia of statistical sciences*, vol 6. Wiley, New York
- Bender A, Mussa H, Glen R, Reiling S (2004) Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): evaluation of performance. *J Chem Inf Comput Sci* 44(5):1708–1718
- Besalu E, Gallegos A, Carbo-Dorca R (2001) Topological quantum similarity indices and their use in QSAR: application to several families of antimalarial compounds. *Commun Math Comp Chem* 44:41–64
- Breiman L (2001) Statistical modeling: two cultures. *Stat Sci* 16:199–231
- Breiman L, Friedman J, Olshen R, Stone C (1984) *Classification and regression trees*. Chapman & Hall/CRC, Boca Raton, FL
- Bremser W (1978) HOSE—a novel substructure code. *Anal Chim Acta* 103:355–365
- Brown N, McKay B, Gasteiger J (2006) A novel workflow for the inverse QSPR problem using multiobjective optimization. *J Comput Aided Mol Des* 20:333–341
- Burden F (1989) Molecular identification number for substructure searches. *J Chem Inf Comput Sci* 29:225–227
- Byvatov E, Baringhaus KH, Schneider G, Matter H (2007) A virtual screening filter for identification of cytochrome P450 2C9 (CYP2C9) inhibitors. *QSAR Comb Sci* 26:618–628
- Carbo-Dorca R, Leyda L, Arnau M (1980) How similar is a molecule to another? An electron density measure of similarity between two molecular structures. *Int J Quantum Chem* 17(6):1185–1189
- Chastrette M, Zakarya D, Peyraud J (1994) Structure-musk odor relationships for tetralins and indans using neural networks (on the contribution of descriptors to the classification). *Eur J Med Chem* 29:343–348
- Chatterjee S, Hadi A (1986) Influential observations, high leverage points, and outliers in linear regression. *Stat Sci* 1(3):379–416
- Chin T, So S (2004) Development of neural network QSPR models for Hansch substituent constants. 2. Applications in QSAR studies of HIV-1 reverse transcriptase and dihydrofolate reductase inhibitors. *J Chem Inf Comput Sci* 44:154–160
- Cho BH, Yu H, Lee J, Chee YJ, Kim IY, Kim SI (2008) Non-linear support vector machine visualization for risk factor analysis using nomograms and localized radial basis function kernels. *IEEE Trans Inf Technol Biomed* 12:247–256
- Chohan K, Paine S, Mistry J, Barton P, Davis A (2005) A rapid computational filter for cytochrome P450 1A2 inhibition potential of compound libraries. *J Med Chem* 48:5154–5161
- Colmenarejo G, Pedraglio A, Lavandera J (2001) Cheminformatic models to predict binding affinities to human serum albumin. *J Med Chem* 44(25):4370–4378
- Consonni V, Todeschini R, Pavan M, Gramatica P (2002) Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. Part 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. *J Chem Inf Comput Sci* 42(3):693–705
- Cruz-Monteagudo M, Borges F, Perez Gonzalez M, Dias Soeiro Cordeiro MN (2007) Computational modeling tools for the design of potent antimalarial bisbenzamidines: overcoming the antimalarial potential of pentamidine. *Bioorg Med Chem* 15:5322–5339
- De Lucca G, Liang J, De Lucca I (1999) Stereospecific synthesis, structure-activity relationship, and oral bioavailability of tetrahydropyrimidin-2-one HIV protease inhibitors. *J Med Chem* 42(1):135–152
- Dias Selassie C, Li RL, Poe M, Hansch C (1991) Optimization of hydrophobic and hydrophilic substituent interactions of 2,4-diamino-5-(substituted-benzyl)pyrimidines with dihydrofolate reductase. *J Med Chem* 34(1):46–54
- Dietrich SW, Blaney JM, Reynolds MA, Jow PYC, Hansch C (1980) Quantitative structure-selectivity relationships. Comparison of the inhibition of *Escherichia coli* and bovine liver dihydrofolate reductase by 5-(substituted benzyl)-2,4-diaminopyrimidines. *J Med Chem* 23(11):1205–1212
- Diller DJ, Hobbs DW (2007) Understanding hERG inhibition with QSAR models based on a one dimensional molecular representation. *J Comput Aided Mol Des* 21:379–393
- Diudea M (1997) Cluj matrix invariants. *J Chem Inf Comput Sci* 37:300–305

27. Doweiko A (2008) QSAR: dead or alive? *J Comput Aided Mol Des* 22:81–89
28. Durant J, Leland B, Henry D, Nourse J (2002) Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci* 42(6):1273–1280
29. Dutta D, Guha R, Chen T, Wild D (2007) Ensemble feature selection: consistent descriptor subsets for multiple QSAR models. *J Chem Inf Model* 47(3):989–997
30. Eriksson L, Johansson E, Lindgren F, Sjöström M, Wold S (2002) Megavariate analysis of hierarchical QSAR data. *J Comput Aided Mol Des* 16:711–726
31. Estrada E (2002) Physicochemical interpretation of molecular connectivity indices. *J Phys Chem A* 106:9085–9091
32. Fernandez M, Caballero J (2006) Modeling of activity of cyclic urea HIV-1 protease inhibitors using regularized-artificial neural networks. *Bioorg Med Chem* 14:280–294
33. Franke L, Schwarz O, Müller-Kuhrt L, Hoernig C, Fischer L, George S, Tanrikulu Y, Schneider P, Werz O, Steinhilber D, Schneider G (2007) Identification of natural-product-derived inhibitors of 5-lipoxygenase activity by ligand-based virtual screening. *J Med Chem* 50(11):2640–2646
34. Gangjee A, Yu J, McGuire J, Cody V, Galitsky N, Kisliuk R, Queener S (2000) Design, synthesis, and X-ray crystal structure of a potent dual inhibitor of thymidylate synthase and dihydrofolate reductase as an antitumor agent. *J Med Chem* 43:3837–3851
35. Garcia-Domenech R, Galvez J, de Julian-Ortiz J, Pogliani L (2008) Some new trends in chemical graph theory. *Chem Rev* 108(3):1127–1169
36. Garg R, Bhatarai B (2004) A mechanistic study of 3-aminoindazole cyclic urea HIV-1 protease inhibitors using comparative QSAR. *Bioorg Med Chem* 12(22):5819–5831
37. Garson D (1991) Interpreting neural network connection strengths. *AI Expert* 6(7):47–51
38. Gavaghan CL, Arnby CH, Blomberg N, Strandlund G, Boyer S (2007) Development, interpretation and temporal evaluation of a global QSAR of hERG electrophysiology screening data. *J Comput Aided Mol Des* 21:189–206
39. Girones X, Gallegos A, Carbo-Dorca R (2001) Antimalarial activity of synthetic 1,2,4-trioxanes and cyclic peroxy ketals, a quantum similarity study. *J Comput Aided Mol Des* 15:1053–1063
40. Gleeson MP, Davis AM, Chohan KK, Paine SW, Boyer S, Gavaghan CL, Arnby CH, Kankkonen C, Albertson N (2007) Generation of in-silico cytochrome P450 1A2, 2C9, 2C19, 2D6, and 3A4 inhibition QSAR models. *J Comput Aided Mol Des* 21:559–573
41. Gramatica P (2007) Principles of QSAR models validation: internal and external. *QSAR Comb Sci* 26(5):694–701
42. Guha R, Jurs P (2004) Development of linear, ensemble, and nonlinear models for the prediction and interpretation of the biological activity of a set of PDGFR inhibitors. *J Chem Inf Comput Sci* 44(6):2179–2189
43. Guha R, Jurs P (2004) The development of QSAR models to predict and interpret the biological activity of artemisinin analogues. *J Chem Inf Comput Sci* 44:1440–1449
44. Guha R, Jurs P (2005) Interpreting computational neural network QSAR models: a measure of descriptor importance. *J Chem Inf Model* 45:800–806
45. Guha R, Schürer S (2008) Utilizing high throughput screening data for predictive toxicology models: protocols and application to MLSCN assays. *J Comput Aided Mol Des* 22(6–7):367–384
46. Guha R, Stanton D, Jurs P (2005) Interpreting computational neural network QSAR models: a detailed interpretation of the weights and biases. *J Chem Inf Model* 45:1109–1121
47. Gunturi S, Narayanan R, Khandelwal A (2006) In silico ADME modelling 2: computational models to predict human serum albumin binding affinity using ant colony systems. *Bioorg Med Chem* 14:4118–4129
48. Hansch C (1969) A quantitative approach to biochemical structure-activity relationships. *Acc Chem Res* 2:232–239
49. Hansch C, Fujita T (1964)  $\epsilon$ - $\sigma$ - $\pi$  analysis. A method for the correlation of biological activity and chemical structure. *J Am Chem Soc* 86:1616–1626
50. Hansch C, Leo A, Hoekman D (1995) Exploring QSAR: hydrophobic, electronic, and steric constants. American Chemical Society, Washington, DC
51. Hassan M, Brown RD, Varma-O'Brien S, Rogers D (2006) Cheminformatics analysis and learning in a data pipelining environment. *Mol Divers* 10(3):283–299
52. Hemmer M, Gasteiger J (2000) Prediction of three-dimensional molecular structures using information from infrared spectra. *Anal Chim Acta* 420:145–154
53. Hemmer M, Steinhauer V, Gasteiger J (1999) Deriving the 3D structure of organic molecules from their infrared spectra. *Vib Spectrosc* 19:151–164
54. Hirst J (1996) Nonlinear quantitative structure-activity relationship for the inhibition of dihydrofolate reductase by pyrimidines. *J Med Chem* 39(18):3526–3532
55. Ivanciuc O, Braun W (2007) Robust quantitative modeling of peptide binding affinities for MHC molecules using physical-chemical descriptors. *Protein Pept Lett* 14:903–916
56. Johnson S (2008) The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). *J Chem Inf Model* 48(1):25–26
57. Katritzky A, Petrukhin R, Tatham D, Basak S, Benfenati E, Karelson M, Maran U (2001) Interpretation of quantitative structure-property and -activity relationships. *J Chem Inf Comput Sci* 41:679–685
58. Katritzky A, Oliferenko A, Lomaka A, Karelson M (2002) Six-membered cyclic ureas as HIV-1 protease inhibitors: a QSAR study based on CODESSA PRO approach. *Bioorg Med Chem Lett* 12:3453–3457
59. Katritzky A, Kulshyn O, Stoyanova-Slavova I, Dobehev D, Kuanar M, Fara D, Karelson M (2006) Antimalarial activity: a QSAR modeling using CODESSA PRO software. *Bioorg Med Chem* 14:2333–2357
60. Kier L, Hall L (1986) Molecular connectivity in structure-activity analysis. Wiley, New York
61. Kier L, Hall L (1999) Molecular structure description: the electrotopological state. Academic Press, Burlington, MA
62. Kiralj R, Ferreira M (2003) A priori molecular descriptors in QSAR: a case of HIV-1 protease inhibitors. I. The chemometric approach. *J Mol Graph Model* 21:435–448
63. Klayman D (1985) Qinghaosu (artemisinin): an antimalarial drug from China. *Science* 228:1049
64. Klon A, Heroux A, Ross L, Pathak V, Johnson C, Piper J, Borhani D (2002) Atomic structures of human dihydrofolate reductase complexed with NADPH and two lipophilic antifolates at 1.09 Å and 1.05 Å resolution. *J Mol Biol* 320:677–693
65. Kramer C, Beck B, Kriegl JM, Clark T (2008) A composite model for hERG blockade. *ChemMedChem* 3:254–265
66. Leonard JT, Roy K (2007) Comparative classical QSAR modeling of anti-HIV thiocarbamates. *QSAR Comb Sci* 26:980–990
67. Lewis RA (2005) A general method for exploiting QSAR models in lead optimization. *J Med Chem* 48:1638–1648
68. Lin TS, Zhu LY, Xu SP, Divo AA, Sartorelli AC (1991) Synthesis and antimalarial activity of 2-aziridinyl- and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives. *J Med Chem* 34(5):1634–1639

69. Mackay A (1977) *Scientific quotations: harvest of a quiet eye*. Crane, Russak & Co, New York
70. Masek B, Shen L, Smith K, Pearlman R (2008) Sharing chemical information without sharing chemical structure. *J Chem Inf Model* 48(2):256–261
71. Miller A (2002) *Subset selection in regression*, 2nd edn. Chapman & Hall/CRC, Boca Raton, FL
72. Miller JF, Brieger M, Furfine ES, Hazen RJ, Kaldor I, Reynolds D, Sherrill RG, Spaltenstein A (2005) Novel P1 chain-extended HIV protease inhibitors possessing potent anti-HIV activity and remarkable inverse antiviral resistance profiles. *Bioorg Med Chem Lett* 15(15):3496–3500
73. Moreau G, Broto P (1980) Autocorrelation of molecular structures: application to SAR studies. *Nouv J Chim* 4:757–764
74. Navia-Vázquez A, Parrado-Hernández E (2006) Support vector machine interpretation. *Neurocomputing* 69:1754–1759
75. Ney H (1995) On the probabilistic interpretation of neural network classifiers and discriminative training criteria. *IEEE Trans Pattern Anal Mach Intel* 17:107–119
76. Nguyen-Cong V, Van Dang G, Rode B (1996) Using multivariate adaptive regression splines to QSAR studies of dihydroartemisinin derivatives. *Eur J Med Chem* 31:797–803
77. Otzen T, Wempe E, Kunz B, Bartels R, Lehwarck-Yvetot G, Hansel W, Schaper K, Seydel J (2004) Folate-synthesizing enzyme system as target for development of inhibitors and inhibitor combinations against candida albicans-synthesis and biological activity of new 2,4-diaminopyrimidines and 4'-substituted 4-aminodiphenyl sulfones. *J Med Chem* 47:240–253
78. Pearlman RS, Smith KM (1999) Metric validation and the receptor-relevant subspace concept. *J Chem Inf Comput Sci* 39:28–35
79. Pinheiro J, Kiralj R, Ferreira M, Romero O (2003) Artemisinin derivatives with antimalarial activity against Plasmodium Falciparum designed with the aid of quantum chemical and partial least squares methods. *QSAR Comb Sci* 22:830–842
80. Polanski J, Zouhiri F, Jeanson L, Desmaele D, D'Angelo J, Mouscadet J, Gieleciak R, Gasteiger J, Le Bret M (2002) Use of the Kohonen neural network for rapid screening of ex vivo anti-HIV activity of styrylquinolines. *J Med Chem* 45:4647–4654
81. Purdy R (1996) A mechanism-mediated model for carcinogenicity: model content and prediction of the outcome of rodent carcinogenicity bioassays currently being conducted on 25 organic chemicals. *Environ Health Perspect* 104:1085–1094
82. Randić M (1978) Fragment search in acyclic structures. *J Chem Inf Comput Sci* 18(2):101–107
83. Randić M, Zupan J (2001) On interpretation of well-known topological indices. *J Chem Inf Comput Sci* 41: 550–560
84. Randić M, Balaban A, Basak S (2001) On structural interpretation of several distance related topological indices. *J Chem Inf Comput Sci* 41:593–601
85. Ravichandran V, Jain PK, Mourya VK, Agrawal RK (2007) QSAR study on some arylsulfonamides as anti-HIV agents. *Med Chem Res* 16:342–351
86. Renner S, Fechner U, Schneider G (2006) Pharmacophores and pharmacophore searches, methods and principles in medicinal chemistry, vol 32, chap Alignment-free pharmacophore patterns—a correlation vector approach. Wiley-VCH, Weinheim, Germany, pp 49–79
87. Roche O, Schneider P, Zuegge J, Guba W, Kansy M, Alanine A, Bleicher K, Danel F, Gutknecht EM, Rogers-Evans M, Neidhart W, Stalder H, Dillon M, Sjogren E, Fotouhi N, Gillespie P, Goodnow R, Harris W, Jones P, Taniguchi M, Tsujii S, von der Saal W, Zimmermann G, Schneider G (2002) Development of a virtual screening method for identification of “frequent hitters” in compound libraries. *J Med Chem* 45(1):137–142
88. Rubner J, Schulten K, Tavan P (1990) A self organizing network for complete feature selection. In: *International conference on parallel processing in neural systems and computers*. Elsevier, Dusseldorf
89. Sahu KK, Ravichandran V, Mourya VK, Agrawal K (2007) QSAR analysis of caffeoyl naphthalene sulfonamide derivatives as HIV-1 integrase inhibitors. *Med Chem Res* 15:418–430
90. Sato M, Tsukimoto H (2001) Rule extraction from neural networks via decision tree induction. In: *Neural networks, proceedings international joint conference*, vol 3. IEEE Computer Society, Los Alamitos, CA, USA, pp 1870–1875
91. Schölkopf B, Smola A (2002) *Learning with kernels*. MIT Press, Cambridge, MA
92. Sheridan RP, Korzekwa KR, Torres RA, Walker MJ (2007) Empirical regioselectivity models for human cytochromes P450 3A4, 2D6, and 2C9. *J Med Chem* 50:3173–3184
93. Stanton D (2003) On the physical interpretation of QSAR models. *J Chem Inf Comput Sci* 43(5):1423–1433
94. Stanton D, Jurs P (1990) Development and use of charged partial surface area structural descriptors in computer assisted quantitative structure property relationship studies. *Anal Chem* 62:2323–2329
95. Stanton D, Mattioni BE, Knittel J, Jurs P (2004) Development and use of hydrophobic surface area (HSA) descriptors for computer assisted quantitative structure-activity and structure-property relationship studies. *J Chem Inf Comput Sci* 44:1010–1023
96. Summerfield R, Daigle D, Mayer S, Mallik D, Hughes D, Jackson S, Sulek M, Organ M, Brown E, Junop M (2006) A 2.13 Å structure of E. coli dihydrofolate reductase bound to a novel competitive inhibitor reveals a new binding surface involving the M20 loop region. *J Med Chem* 49(24):6977–6986
97. Sutter J, Dixon S, Jurs P (1995) Automated descriptor selection for quantitative structure-activity relationships using generalized simulated annealing. *J Chem Inf Comput Sci* 35:77–84
98. Taha I, Ghosh J (1999) Symbolic interpretation of artificial neural networks. *IEEE Trans Knowl Data Eng* 11:448–463
99. Takahashi T (1991) An information theoretical interpretation of neuronal activities. In: *Neural networks, proceedings International Joint Conference*, vol 2. IEEE Computer Society, Los Alamitos, CA, USA, pp 645–648
100. Tang LJ, Zhou YP, Jiang JH, Zou HY, Wu HL, Shen GL, Yu RQ (2007) Radial basis function network-based transform for a nonlinear support vector machine as optimized by a particle swarm optimization algorithm with application to QSAR studies. *J Chem Inf Model* 47:1438–1445
101. Tian F, Zhou P, Lv F, Song R, Li Z (2007) Three-dimensional holograph vector of atomic interaction field (3D-HoVAIF): a novel rotation-translation invariant 3D structure descriptor and its applications to peptides. *J Pept Sci* 13:549–566
102. Todeschini R, Consonni V (2002) *Handbook of molecular descriptors*. Wiley-VCH, Berlin
103. Tropsha A, Gramatica P, Gombar V (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb Sci* 22:69–77
104. Urbanek S (2006) *Graphics of large datasets*, chap Trees, pp. 177–2002. *Statistics and Computing*. Springer, Berlin Heidelberg
105. Urbanek S, Unwin A (2002) Making trees interactive with KLIMT—a COSADA software project. *Stat Comp Graph Newsl* 13(1):13–16
106. Usdun B, Melssen WJ, Buydens LMC (2007) Visualisation and interpretation of support vector regression models. *Anal Chim Acta* 595:299–309
107. Vapnik V (1998) *Statistical learning theory*. Wiley, New York

108. Venkatraman V, Dalby AR, Yang ZR (2004) Evaluation of mutual information and genetic programming for feature selection in QSAR. *J Chem Inf Comput Sci* 44:1686–1692
109. Ventura C, Martins F (2008) Application of quantitative structure-activity relationships to the modeling of antitubercular compounds. 1. The hydrazide family. *J Med Chem* 51(3):612–624
110. Verma RP, Hansch C, Selassie CD (2007) Comparative QSAR studies on PAMPA/modified PAMPA for high throughput profiling of drug absorption potential with respect to Caco-2 cells and human intestinal absorption. *J Comput Aided Mol Des* 21:3–22
111. Vilar S, Santana L, Uriarte E (2006) Probabilistic neural network model for the in silico evaluation of anti-HIV activity and mechanism of action. *J Med Chem* 49:1118–1124
112. Visco D, Pophale R, Rintoul M, Faulon J (2002) Developing a methodology for an inverse quantitative structure-activity relationship using the signature molecular descriptor. *J Mol Graph Model* 20:429–438
113. Weininger D, Weininger A, Weininger J (1989) SMILES. 2. Algorithm for generation of unique SMILES notation. *J Chem Inf Comput Sci* 29:97–101
114. Yuan H, Parrill A (2002) QSAR studies of HIV-1 integrase inhibition. *Bioorg Med Chem* 10(12):4169–4183
115. Zahouily M, Lazar M, Elmakssoudi A, Rakik J, Elaychi S, Rayadh A (2006) QSAR for anti-malarial activity of 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives. *J Mol Model* 12:398–405
116. Zhang H, Li H, Ma Q (2007) QSAR study of a large set of 3-pyridyl ethers as ligands of the  $\alpha_4\beta_2$  nicotinic acetylcholine receptor. *J Mol Graph Model* 26:226–235
117. Zhou D, Alelyunas Y, Liu R (2008) Scores of extended connectivity fingerprint as descriptors in QSPR study of melting point and aqueous solubility. *J Chem Inf Model* 48(5):981–987