

Chapter 13

Virtual Screening Applications in the Search of Novel Antiepileptic Drug Candidates

Alan Talevi and Luis E. Bruno-Blanch

Abstract

Virtual screening (VS) encompasses a wide spectrum of computational approaches oriented to prioritize which compounds from chemical libraries or repositories will be subjected to experimental *in vitro* and *in vivo* testing. VS can be broadly classified into structure- and ligand-based approaches. Which VS methodology is preferable is highly dependent on the targeted system and the availability of experimental data and also on background considerations such as existing technical and economic resources. To the moment, VS has been underexplored for the discovery of new antiepileptic agents. In this chapter we discuss details on the general ligand-based procedures to undertake VS campaigns, with emphasis on some particular considerations for the case of antiepileptic drug development.

Key words Epilepsy, *In silico* screening, Ligand-based approximations, Virtual screening

1 Introduction

The universe of accessible chemical compounds has been growing exponentially during the last 50 years: in 2015, the Chemical Abstract Service, which roughly reflects the chemical entities known by mankind, achieved its 100 millionth entry. Judging from the number of entries in PubChem, around half of those substances are small drug-like molecules. This vast and expanding population of available chemicals makes it probable to find novel therapeutic agents with virtually any desired pharmacological profile. Nevertheless, the systematic, massive screening of such molecular diversity remains challenging. Even automated and miniaturized approaches like high-throughput screening are technologically demanding and prohibitive for small academic centers or pharmaceutical companies.

The term virtual screening (VS) or *in silico* screening refers to the application of a diversity of computational approaches to rank digital chemical repositories or libraries in order to establish which drug candidates are more likely to obtain favorable results when

experimentally tested through *in vitro* and/or *in vivo* models. Since they are meant to reduce the volume of experimental testing and optimize the results, VS techniques possess several advantages in terms of cost-efficiency, bioethics, and environmental impact. We may also mention that, since many computational applications and chemical databases are publicly available online and several of them run smoothly in any current personal computer, the technology gap between developed and emerging countries is considerably low for VS technologies compared with other screening approaches.

VS approaches can be essentially classified within two categories: structure-based (or direct) and ligand-based (or indirect) approximations. A persistent and major obstacle for the implementation of structure-based VS approaches comes from the fact that most validated targets for antiepileptic drugs are ligand- or voltage-operated ion channels whose structure has not been solved experimentally yet. In order to predict protein folding for unsolved structures, a drug designer can turn to homology modeling, which uses a known protein similar to the protein of interest (e.g., a homolog from other species or another member of the same protein family from the same species) as a template to predict the secondary and tertiary structure of the target protein [1]. These models can be in turn be used for design or VS campaigns. For instance, some attempts of homology models of epilepsy-relevant targets such as GABA transporters, GABA transaminase, and SV2A have recently been reported [2–4]. A remarkable exception is carbonic anhydrase, a putative AED target whose human isoforms have already been solved and are being actively used in the field of drug discovery [5–7].

Alternatively, ligand-based approaches can be applied whenever a model of the target structure is not available or to complement structure-based approximations. Concisely, ligand-based approaches can be classified into similarity searches, machine learning approaches, and superposition approximations [1, 8]. These techniques differ in a number of factors, from their requisites to their enrichment metrics or scaffold hopping. Similarity search uses molecular fingerprints derived from 2D or 3D molecular representations, comparing database compounds with one or more reference molecules in a pairwise manner. Notably, only one reference molecule (e.g., the physiologic ligand of a target protein) is required to implement similarity-based VS campaigns. Usually, similarity searches are the only alternative to explore the chemical universe for active compounds in the absence of experimental knowledge of the target protein or related proteins and when the number of known ligands is scarce. Machine learning approaches operate by building models from example inputs to make data-driven predictions on the database compounds. Machine learning approximations require several learning or calibration examples.

Finally, superimposition techniques are conformation-dependent methods that analyze how well a compound superposes onto a reference compound or fit a geometrical, fuzzy model (pharmacophore) in which functional groups are stripped off their exact chemical nature to become generic chemical properties (e.g., hydrophobic point, H-bond donor, etc.); such process is facilitated if the modeler counts on an active rigid analog with limited conformational freedom. As a general rule, the more complex approximations take the lead in terms of scaffold hopping, whereas simpler approaches are computationally less demanding while achieving good active enrichment metrics [9]; the efficacy of a given technique is, however, highly dependent on the chosen molecular target, and frequently different techniques are complementary in nature [10].

Since, as it has been mentioned, most of the human proteins validated as molecular targets for antiepileptic drugs have not been solved experimentally, this chapter will focus on ligand-based VS protocols. Similarity search protocols are basically simple: their sole complexity lies in a number of decisions that the user has to make regarding what similarity metric will be used, what fingerprint system, and the size of the molecular features (molecular substructures) considered. Thus, we will aim our attention at more complex ligand-based approaches that rely on inferring a model from a set of instances or examples, in particular, machine learning approaches.

2 Building a Model

2.1 Dataset Compilation

Naturally, the first step when one tries to infer a model (a generalization on some structure–property relationship with a variable degree of abstraction) from a number of examples is to compile such training or calibration instances. We will refer this as dataset compilation. It is often heard that models are as good as the biological and chemical data from which they are derived [11, 12]. Therefore, the dataset compilation is one of the most important steps when developing a model for VS purposes. When intending to build a model for VS applications, some specific considerations must be taken into account: (a) compile a dataset as diverse as possible; (b) ideally, the dependent variable (the modeled property, i.e., the biological activity) should span at least two or three orders of magnitude [13–16], from the least to the most active compound in the series; (c) the available biological data on the training set compounds should, preferably, be uniformly distributed across the range of activity or at least follow some defined statistical distribution, usually normal distribution – the same principle applies to the model’s independent variables [13]; and (d) the biological activities of all the training instances should be of comparable quality – ideally, they should have been determined in the

same laboratory using the same conditions [14, 15], so that variability in the measured biological activity only reflects treatment variability (this is rarely the case, though).

Let us discuss point by point. Regarding the chemical diversity of the training set, while for computer-aided drug design the use of homologous series of chemicals might be advantageous, in the case of VS, we are expecting to apply the model for the screening of vast collections of chemicals (typically, thousands to millions). As we will comment later, the predictions of the models will only be reliable within the chemical space covered by the training set: thus, if we want to cover a wide chemical space, the training set should ensure the most possible chemical diversity. On the other hand, guaranteeing a wide coverage of the activity space allows the model to capture not only essential molecular features needed to elicit the desired activity, but features that diminish activity as well. Data distribution should be studied in order to avoid poorly populated regions within the studied chemical space as well as highly populated narrow intervals [13]. It is often heard that modelers should avoid data extrapolation; however, interpolation could also be dangerous if too sparse regions exist within the data. Histograms can be of great help to visualize the distribution of both the model dependent and the independent variables; still, analysis of the multivariate space can reveal empty or scarcely populated data regions that separate analysis of the independent variables may not [17]. The second issue related to inadequate data distribution is the existence of *leverage points* (outliers) among the data points. We will resort to the terminology adopted by Cruz-Monteagudo and collaborators to discuss the subject of exceptional data points [18], though their nomenclature is not universal. An outlier is a type of data exception represented by extreme values in the descriptor or property (response) space that cannot be attributed to mislabeling due to annotation or measurement errors (Fig. 1). Outliers exert great influence on the model calibration (hence they are called leverage points) especially when a quantitative model with a continuous response/output (regression model) is used. It is likely that some degree of overfitting will appear when outliers are present among the training examples. Even if no

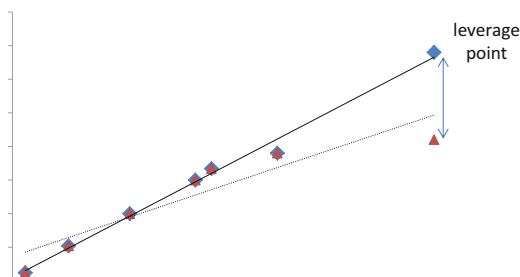


Fig. 1 Example of a leverage point. Note how much the isolated point on the extreme right influences the fit of the regression line

systematic error in the experimental measurement of the modeled property exists, random error will have greater impact on the model in the case of outliers. There are a huge number of methods to detect outlier behavior (and, consequently, remove atypical points from the calibration samples) [19]. The very simple approach by Roy et al. [20] can be used to detect outliers provided that the data is normally distributed, which is relatively frequent in our field (and the stand-alone application is publicly available online!). However, as always when using parametric methods (which assume that the dataset fits a known distribution or probability model), it may be a good idea to run statistical discordance tests to check if the assumed distribution is optimal or close to optimal. Finally, since most of the machine learning applications are based on datasets that have been directly or indirectly compiled from literature, the last requisite on the dataset (homogeneous data quality) is seldom accomplished. Compiling the dataset from literature could also give raise to another type of data exceptions: noise. Noisy data points emerge from large experimental errors or from wrong annotations. To mitigate this issue, conscientiously curate your dataset: read your data sources carefully and dispose of those training examples extracted from inadequate or dubious experimental protocols. At present, there are several databases that compile experimental data for small molecules (e.g., ChEMBL); though such resources are manually curated from primary scientific literature, it will not hurt to review the experimental procedures from which biological data is extracted. When possible, ChEMBL tries to normalize bioactivities into a uniform set of endpoints and units. Remarkably, they currently flag activity values that are outside a range typical for that activity type, potentially missing data and suspected or confirmed author errors. Finally, classification models can be used to alleviate the influence of data heterogeneity, as will be discussed later. See *Note number 1* for an extensive additional discussion on the topic of dataset compilation in relation to antiepileptic drug discovery.

2.2 Partitioning the Dataset into Training and Test Sets

Once the dataset has been compiled, it is typically split into a training (or calibration) set (from which the model will be inferred) and an independent test set which will be used to assess the predictive power of the model. Partitioning the dataset is not a trivial task. The general objective of this procedure is to attain representative subsets of the whole dataset. Often, such subsets are obtained through random sampling (a randomly chosen subset of the dataset is tagged as test set) or activity range algorithms (the dataset is divided in groups according to activity values, and test set compounds are chosen from each group to cover the activity range uniformly). While these approaches are appropriate when training and test sets are comparable in size [21, 22], better results are obtained with more rational partitioning procedures such as sphere exclusion algorithms when test sets are small (but larger than five compounds)

in comparison with the corresponding training sets [21]. Note that typically only 10–20% of the dataset is selected for external validation [12]. It should also be considered that ideally, the number of active and inactive compounds in the training set should be balanced in order to avoid potential bias toward the prediction of the overrepresented category of training examples [12].

2.3 Choosing and Calculating Molecular Descriptors

Briefly, molecular descriptors are numerical variables that reflect chemical information encoded within a symbolic representation of a chemical compound, e.g., a molecular formula, a chemical graph, and a geometric molecular model. There is a wide diversity of molecular descriptors available to characterize relevant aspects of a molecule, from simple functional group counts to time-demanding quantum descriptors. Once the dataset has been compiled and divided into appropriate training and test sets, the nature of the molecular descriptors that will be allowed into the model should be decided. Some studies suggest that the choice of descriptors plays a more important role than the choice of a modeling technique [23, 24]. Two fundamental aspects are considered at this stage of the modeling process: first, the throughput speed associated with different types of descriptors and, second, the interpretability of each type of descriptor. We should keep in mind that this chapter is focused in models which will eventually be used to screen large collections of chemicals. Thus, the selected molecular descriptors should ideally be easy to compute and readily interpretable. Unfortunately, these two aspects are usually inversely related: in general, the more interpretable a model is, the more it is computationally demanding. 3D QSAR models often provide a graphical output which is easy to interpret in familiar chemical terms. Most 3D QSAR methods, however, implement the pharmacophore concept and are conformation and sometimes alignment dependent. What conformation should be used to compute the correspondent 3D descriptors? An ideal solution to account for the conformation dependency would be to determine the bound (also called bioactive) conformation [25, 26]. Defining the bound conformation is often an enormously difficult and time-consuming task. Remember that the focus of this chapter is VS applications; consistently, any conformational analysis should eventually be applied not only to the training and test set compounds but also to the screened database molecules, which could well include millions of compounds. Bound conformations of ligands can be obtained experimentally by NMR or X-ray crystallography. However, as has been pointed out in the introduction of the chapter, most of the targets for anti-epileptic drugs have not been solved yet. Furthermore, crystal structures also have limitations, from data acquisition and data refinement errors to the potential inadequacy of crystal structures to represent the conformational ensemble in solution [26]. Valuable hints on the active conformation might be obtained when

rigid ligands with restricted conformational freedom are available. For instance, some antiepileptic agents like phenytoin and carbamazepine have very few rotatable bonds and have been used to propose pharmacophore models [27]. When no hints on the bioactive conformation can be inferred from experimental data or rigid ligands, the modeler has no alternative, but to sample the potential energy surface of the ligands. A number of methods (all of them computationally demanding) are available for such purpose, including systematic search, stochastic approaches, and molecular dynamics. Note that very frequently very rough approximations are performed in this stage, from using the presumed global energy minimum or a local energy minimum (which is not representative of the bioactive conformation) to energy minimization procedures in vacuum, which ignore solvent effects. If bound conformations can be defined, eventually the conformational energy (energy difference between the bound and unbound conformation, sometimes called strain energy) should be calculated for each of the chemicals in the database subjected to VS, retaining those molecules with calculated strain energy below a user-defined threshold, which is typically below 10 kcal/mol. On the other hand, when alignment-dependent methods are used, defining alignment criteria for structurally diverse compounds can be tough (if not impossible). On the basis of all the previous limitations of 3D QSAR methods, we believe that conformation-independent QSAR methods (2D QSAR) are more easily automated and adapted to the task of VS since they neither require conformational search nor structural analysis [28]. Judging from the success of such approaches, simple representations of the molecular descriptors such as chemical graphs seem to implicitly contain a large amount of biologically relevant molecular information. They are naturally inappropriate, though, to differentiate geometrical isomers. Moreover, they are usually less transparent to interpretation; many 2D models behave like a black box: effective but inscrutable. In any case, 2D approaches could be used as a first screening approach to reduce the number of potential drug candidates, which can be later complemented with more computationally demanding techniques.

Once the molecular descriptors that will be used have been chosen, it is time to curate chemical structures. The required degree of curation depends on the descriptors that are to be used. Do not underestimate the importance of this step: it has been shown that, in average, there are two structural errors per medicinal chemistry publication [12] and a variable rate of errors in compounds indexed in chemical databases which may be as high as 8% [29]. Even slight errors in chemical structures may lead to significant loss of prediction accuracy in the subsequent model [29]. The first step for cleaning chemical records is to remove those data points that are usually not handled by conventional cheminformatics techniques: inorganic and some organometallic compounds,

counterions, salts, and mixtures [12]. Some additional details on this subject are provided in *Note number 2*. Duplicates should also be removed; otherwise the incidence of a single compound on the model would be exaggerated. Note that different compounds might act as duplicates depending on the chosen molecular descriptors: e.g., stereoisomers are distinct but act as duplicates if 2D descriptors are used. Some chemical functions and moieties that can be represented in multiple ways should be standardized, e.g., aromatic rings, nitro groups, etc. [30]. Finally, tautomeric groups should also be curated. In order to decide which tautomer should be kept, the mechanism of action of the compounds could be considered [29] or, alternatively, the dominant tautomer at physiologically relevant conditions should be used. Many of the previous steps can be performed in an automated manner by specialized software applications (e.g., ChemAxon's Standardizer); still, it is advisable to manually verify a randomly picked subset of the compounds to ensure everything has gone well. Also note that many software tools for molecular descriptor calculation impose restrictions to the molecular representations (e.g., explicit or implicit hydrogens, aromatic rings, etc.).

2.4 Building a Model

So far we have explained the cautions needed to compile an adequate training set, curate the correspondent molecular structures, and compute a set of molecular descriptors. At this point, the subset of descriptors that best correlates with the property of interest should be selected (variable or feature selection step), and the dependent variable should be mapped with those preselected descriptors (i.e., the incidence of each descriptor on the dependent variable should be weighted). There is a plethora of techniques to execute these tasks [31–33], and their analysis is out of the scope of this chapter. However, we would like to discuss two particular aspects of model building that are generally pertinent no matter which modeling technique is considered. The first relates to the principle of parsimony and the problem of overfitting. The second relates to the convenience of choosing between regression and classification approaches.

Overfitting means gaining explanatory power on the training examples at the expense of generalizability (predictive power). As in any learning task, memorizing is discouraged since the goal is to extract a generalization from learning examples that can be later applied to other cases/situations. The principle of parsimony affirms that we should use the simplest method that provides the desired performance level. This includes avoiding the use of excessively flexible approaches if they are not required (e.g., avoid using nonlinear methods if linear methods can provide an appropriate solution) and also avoiding the inclusion of more parameters/features than needed [34]. In our personal experience, the models that explain too well the training data fail at predicting external

data, while models with moderately good performance on the training examples tend to behave similarly on external instances. Overfitting can be avoided retrospectively (once a model has already been built, using adequate validation procedures as discussed in the next subsection) or prospectively (during the model-building stage). Prospective avoidance of overfitting usually involves a rule of thumb regarding the ratio of learning examples to predictors (molecular descriptors) included in the model. Though a ratio equal or greater than five is often suggested for multivariate regression approaches [13, 35–37], in our opinion at least ten training compounds per independent variable are safer. Some methods such as partial least squares allow a higher number of descriptors. A final point that should be regarded is the influence of the total number of variables screened for possible correlation with the modeled activity (i.e., the size of the descriptor pool) on the statistical significance of the obtained correlation [38]: the larger the descriptor pool, the greater the probability of arriving to spurious, chance correlations. Software for molecular descriptor calculation can typically compute hundreds to thousands of descriptors. In our experience, we have found that the use of small random subsets of descriptors is a useful strategy to mitigate the chance of spurious correlations (note that this issue is intimately related to the problem of overfitting, since the general idea still is avoiding inclusion of independent variables that reflect meaningless particularities of the training examples; thus, the strategies that have been already discussed to minimize the risk of overfitting are also suitable to reduce the probability of chance correlations). Fisher's randomization test, which is discussed in the next section, is also a valuable tool to assess the probability of random correlations. Some additional hints on the matter of variable selection are presented in *Note number 3*.

With regard to the selection of a classification or regression model, classification models might be more adequate when biological data has been obtained in different labs and it is subjected to large interlaboratory variability, possibly introducing noise to the model [39]; this approach might be valuable to mitigate the influence of leverage data points. Note that, typically, biological data from cellular and animal models is prone to present high variability; as we discuss in the *Notes*, many modeling efforts directed to predict anticonvulsant activity are based on in vivo data obtained from seizure models; even when models of seizure are naturally more controlled than models of epilepsy, which often require sub-chronic administration of pro-convulsant stimuli, high interlab variability is to be expected. Classification models are quantitative models based on relationships between independent variables (in this case molecular descriptors) and a categorical response variable of integer numerical values that represents the class of the corresponding sample. Here, the term “quantitative” is

referred to the numerical value of the independent variables needed to classify the chemicals in the qualitative classes (a categorical response); such variables specify the quantitative meaning of a QSAR-based classification process [36]. As clearly stated by Polanski et al. [40], extensive data independence implies qualitative and not quantitative solutions; classification models might be considered somehow in the middle, because they provide a qualitative response (yes or no, active or inactive, etc.) while retaining quantitative analysis through the numerical independent variables. Those same authors point out that the combination of different data handling schemes seems particularly effective to provide robust solutions. When using classification models, the sources of noise are majorly restricted to those points that lie in the frontier between the categories of objects under consideration (e.g., those whose activity value is near the activity threshold that has been defined to differentiate active from inactive compounds). Also note that scientific literature is generally biased toward the report of highly active compounds, resulting in a general relative abundance of active compounds in comparison to inactive ones. Classification methods can ameliorate this issue since the modeler might include putative inactive compounds within the inactive class: though this is a potential source of error (the inactive nature of such presumed inactive examples has not been verified), it can be assumed that this error will not be significant if the dataset is sufficiently large.

2.5 Model Validation

Model validation implies the *quantitative* assessment of the model robustness and predictive power [41, 42], which serves to detect the occurrence of overfitting and chance correlations. In the context of signal processing applications (within which we can enclose QSAR modeling), robustness refers to *approaches that are not degraded significantly when the assumptions that were invoked in defining the processing algorithm are no longer valid* [40]. Validation techniques can be divided in internal and external validation. In the internal validation approaches, the training set itself is used to assess the model stability and predictive power; in external validation, a holdout sample absolutely independent from the training set is used to test the predictive ability of the model. Though there is a diversity of techniques that can be used for internal validation purposes, the most frequent are cross validation and Y-randomization.

In cross validation, groups of training examples are iteratively held out from the training set used for model development; the model is thus regenerated without the removed chemicals, and the regenerated model is used to predict the dependent variable for the held-out compounds [43]. The process is typically repeated until every training compound has been removed from the training set at least once (Fig. 2). When only one compound is held out in each

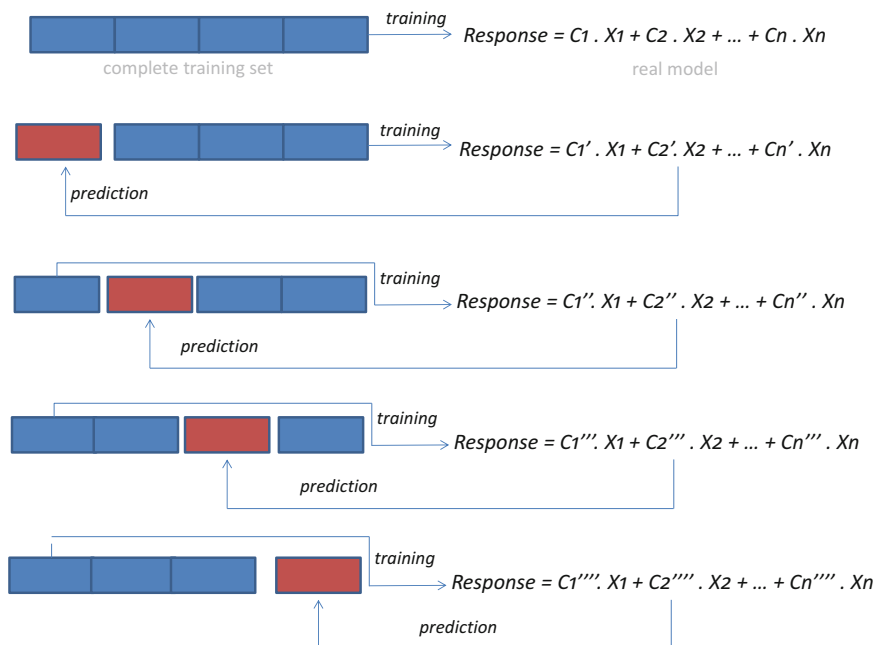


Fig. 2 A fourfold cross validation is schematized as an example. In each cross validation round, 25% of the training compounds (shown in *red*) are randomly removed from the training set and used as internal test sample, while the remaining compounds (in *blue*) are used for training purposes. Note that in the example the same model parameters that are included in the original model are present in the new models, but the regression coefficients change reflecting the variability introduced to the training sample. Here, the process has been repeated until every instance in the original training data has been removed once

cross validation cycle, we will speak of leave-one-out cross validation. If larger subsets of training examples are removed in each round, we will speak of multifold, leave-some-out, leave-group-out, or leave-many-out cross validation. Naturally, the more compounds removed per cycle, the more challenging the cross validation test. Cross validation in general and leave-one-out cross validation in particular tend to be overoptimistic [41, 43, 44]: good cross validation metrics are a necessary but not sufficient condition to prove the predictive power of a model. When leave-many-out cross validation is used, the results for each (held-out) fold or subsample can be averaged or otherwise combined to produce a single estimation. See *Note number 4* for some additional discussion on this subject.

Y-randomization (Fig. 3) involves scrambling the value of the experimental/observed dependent variable across the training instances, thus abolishing the relationship between the response and the molecular structure. Naturally, since the response is now randomly assigned to the training examples, no correlation is expected to be found if the model is regenerated from the scrambled data.

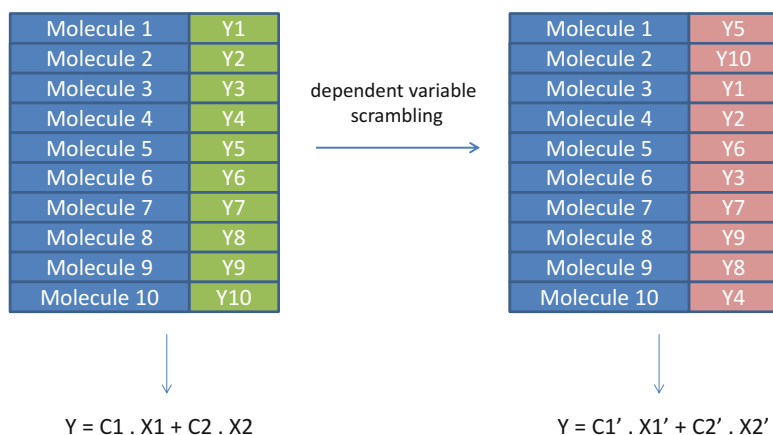


Fig. 3 In randomization test, the Y-response is scrambled among the training samples, and the relationship between the molecular structure and the response is thus abolished, hopefully leading to poorly performing randomized models. Note that process randomization has been illustrated here, as can be deduced from the different parameters included in the randomization models and the real one

External validation, i.e., using an independent test set to establish the model predictive power, has been regarded as the most significant validation step [41, 43], though one condition should be observed for the results to be reliable: the test sample should be representative of the training sample; at least 20 holdout examples are advised when the test set is randomly chosen from the dataset and, if possible, at least 50. Some authors have pointed out, however, that dividing the dataset into training and test sets may result in loss of valuable chemical information which otherwise could have been used for model building [34, 42, 45], suggesting that only internal validation is advised for small (<50 instances) datasets. In that situation, leave-group-out using folds comprising 30% of the training set gives robust results across several small datasets [43].

3 Virtual Screening

3.1 Pilot VS Campaign

Once the model has been built and properly validated and before it is applied in a real VS campaign, it is convenient to assess the model performance in a pilot VS campaign. Note that the active and inactive compounds are generally balanced in the dataset that was used to build the model, but in real VS applications, the inactive compounds greatly outnumber the active ones. This constitutes an intrinsic limitation of the VS approach, determining that even models with very good performance will tend to have low positive predictive value (PPV, i.e., activity probability of a given predicted hit) in VS, since, as follows, such probability is influenced not only by intrinsic features of the model (specificity, S_p ,

and sensitivity; Se) but also by the yield of active Ya of the screened database [46]:

$$\text{PPV} = (\text{Se} \times \text{Ya}) / [\text{Se} \times \text{Ya} + (1 - \text{Sp})(1 - \text{Ya})] \quad (1)$$

While Ya is uncertain in real VS applications, it is unequivocally low. Thus, it is desirable to perform a pilot VS campaign in the first place, seeding a relatively low number (below 5%) of known actives among a large number of observed and presumed inactive compounds. As mentioned previously, finding reported inactive compounds in the literature is hard due to the bias to report positive results, which makes it difficult to accomplish the precedent condition. Therefore, one will often resort to putative inactive compounds as decoys. A good decoy should share certain physicochemical features with the active compounds in order to pose a more rigorous challenge to the model [47]. The enhanced directory of useful decoys (DUD-E) is an online public resource that automatically generates valuable matched decoys for user-supplied ligands (such decoys are matched by a number of physicochemical properties but are topologically dissimilar to the actual active compounds) [48]. Receiver operating characteristic (ROC) curves in which the model Se (true positive rate) is plotted versus 1 minus Sp (true negative rate) for a diversity of score thresholds are a valuable tool to test the performance of a given VS model/method and also for benchmarking purposes [46]. They are also used to optimize the score threshold that will define if a given compound from the screened database will be considered a predicted active or inactive, allowing selecting an adequate Se/Sp balance and thus optimizing the PPV. A scheme describing how ROC curves are built is presented in Fig. 4. With the help of ROC curves, different VS approaches can be statistically compared between themselves and to random behavior; the area under the ROC curves is usually used for those purposes. Though owing to the saturation effect, the total area under the ROC curve is not a suitable metric to assess the VS approaches in relation to the early recognition (i.e., the ability of a VS method to rank actives early in the ordered list) [49], the partial area under the ROC curve is well suited for this purpose. pROC is an excellent free open-source package for R which can be used for partial and total ROC curve comparison [50]. Some additional remarks on this subject are included in *Note number 5*.

3.2 Virtual Screening and Applicability Domain Estimation

After making sure that the model is suitable for real VS applications and selecting an appropriate cutoff score value, one can proceed to the real VS campaign. There exist numerous publicly available databases containing drug-like molecules, such as ZINC [51], DrugBank [52], Sweetlead [53], or the Universal Natural Product Database [54]. The election of the screened database depends on

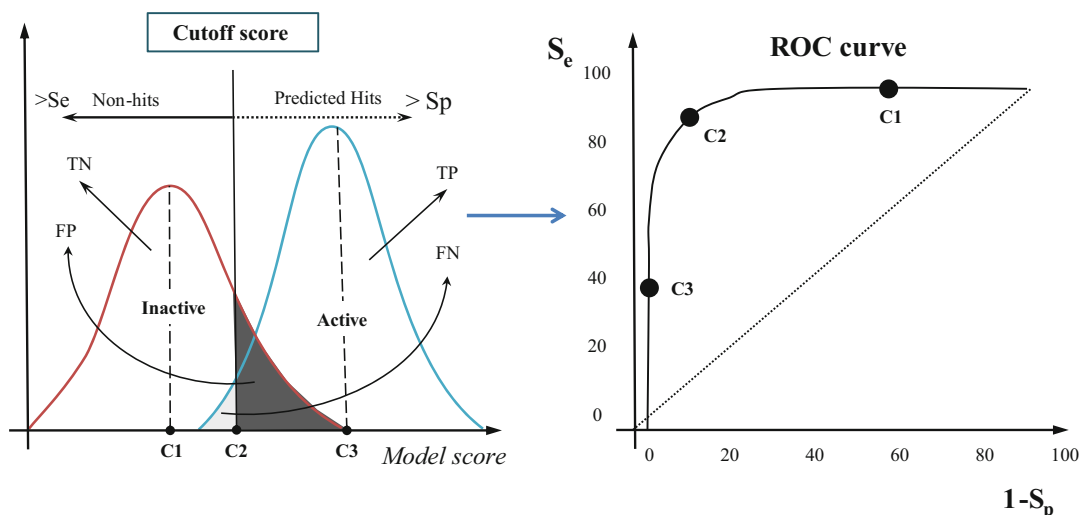


Fig. 4 Schematic representation of ROC curve construction

which compounds the researcher is focused on. For instance, DrugBank and Sweetlead are the collections of choice when the VS is oriented to drug repurposing, since they compile approved and investigational drugs from the FDA and other regulatory agencies. The Universal Natural Product Database, as its name implies, compiles (more than 200,000) compounds from natural sources. It must be checked if every screened compound belongs to the applicability domain of the model (in order to determine whether a given prediction is or is not reliable); for this, there are a number of methods available, including distance-based, parametric, and nonparametric approaches, among others [17]. Most of these approximations can be easily implemented with most statistical software packages; if the molecular descriptors included in the training set follow a normal distribution, the parametric approach discussed by Roy et al. might be applied [20], and it is readily available online. Note, however, that it has been observed that applicability domain assessment often limits the chemical space coverage of the resulting (reliable) predictions [23]; the same authors have observed that consensus scoring (combining the scores or ranks from different VS approximations) might reduce the necessity of applicability domain estimation while retaining a wider coverage. Remarkably, consensus scoring could also mitigate the influence of noisy data achieving robust solutions [40].

3.3 ADME Filters

The pharmaceutically relevant processes of absorption, distribution, metabolism, and excretion (ADME) determine the pharmacokinetics of a given drug and therefore the extent of drug action. Including ADME filters as secondary criteria to select drug candidates is particularly important when implementing VS applications

to discover novel AEDs, whose pharmacology is often critically influenced by biodistribution and metabolism issues. Besides the frequently used Lipinski rules (or other similar rules, e.g., Veber's rules) to filter out compounds with likely unfavorable oral absorption, having in mind the role of the blood–brain barrier regulating the influx of chemicals into the brain, specific rules or algorithms to predict central nervous system bioavailability should be considered. For example, the “rule of 2” states that a log octanol–water partition ($\log P$) coefficient of 2 is optimal to assure brain bioavailability for those compounds that enter the brain through passive diffusion [55]. Other more complex (yet simple) filters such as the central nervous system desirability score proposed by Wager et al. could also be included [56]. If the VS campaign is focused on novel treatments for refractory epilepsy, it could be a good idea to include *in silico* filters to predict affinity for ABC transporters (some models on these antitargets are discussed in other chapters of this volume). Whenever using a model built by other developers, ensure that such model's predictions include applicability domain assessment. At last, many known AEDs are involved in CYP induction or suffer extensive CYP metabolism; accordingly, filters to predict affinity for the main CYP isoforms (e.g., CYP3A4, CYP2D6, CYP2C9) or nuclear receptors might also result useful.

4 Notes

1. *Dataset compilation.* Traditionally, the requirements on the training set examples were even more stringent than the ones listed in Sect. 2.1. For instance, it is usually affirmed that, especially in the case of 3D QSAR, all the training examples should share the same mechanism of action (and the same binding mode) and all the inactive compounds in the training set should be truly (and not putative) inactive [14–16, 57]. It is argued that all 3D QSAR methods were conceived to describe only one interaction step in the lifetime of ligands [14], a statement which is partially supported by the fact that many 3D QSAR methods are highly alignment dependent (the results depend on the position and orientation of the molecular representation in space). Furthermore, it is also said that only *in vitro* biological data should be considered, since *in vivo* data reflects a number of parallel processes (e.g., transport, metabolism, binding to multiple targets), while by definition it is not possible to reach equilibrium in an *in vivo* system [14, 15]. It is true that *in vitro* data is cleaner than *in vivo* data, in the sense that interpretation of the test results is more straightforward and less affected by confounding factors; all the other systems undergo significant time-dependent changes. Personally, however, we believe that such excessive reductionist approach could be

dangerous when dealing with complex disorders such as epilepsy. There are many good reasons to take dogmatic principles with caution. First, very frequently biological data emerging from phenotypic models (e.g., *in vivo* or cellular models) are used to obtain QSAR models, and in spite of this, the models achieve considerable explicatory and predictive ability (see, e.g., [58–62]). A common mechanism could be presumed when compounds of the same chemical series are being considered, but the modeler cannot be truly sure regarding the specific action mechanism explaining the phenotypic observation or the number or identity of the pharmacologically active chemical species. The complex nature of the biological response makes it impossible to describe, *a priori*, a well-defined action mechanism or to discriminate the influence of other processes on the modeled activity (transport processes, bioactivation, etc.). And yet, the so obtained models work. Sometimes one will not have a satisfactory explanation to an observation, but facts should not be ignored in favor of a preestablished principle. Second, it is now understood that multifactor, complex disorders (e.g., mood disorders, neurodegenerative disorders, or epilepsy) are usually better addressed with drugs with complex pharmacology (i.e., multi-target drugs) [63–65]; in fact, once and again it is mentioned that the “one target, one drug” paradigm has been disappointing in terms of innovative treatments [66–68]. Most currently approved antiepileptic drugs are in fact multi-target agents [64]. Therefore, it is possible that we should resort to tailored multi-target drug discovery and/or return to phenotypic-based drug discovery to identify new pharmacological solutions to epilepsy. Under this perspective, going against the doctrine and using biological data obtained from animal models might be in fact a better approach toward VS for novel antiepileptic agents. Note that many successful QSAR and VS applications focused on antiepileptic drugs have indeed used *in vivo* biological data for modeling purposes [27, 28, 69–72], including reports by leading experts in the QSAR field [28]. Third, not all available molecular descriptors are conformation and/or alignment dependent, and some of them are capable of describing more general properties than those relevant for a single binding event. For instance, molecular complexity and molecular size have been directly and inversely correlated with drug promiscuity [73]; these molecular features can be captured by descriptors such as information indices or molecular weight, respectively. Fourth, QSAR theory has greatly evolved in the last years; multitasking QSAR models are suitable to predict multiple features and complex behaviors, exploiting latent commonalities across tasks [74, 75]. Finally, as it has already been discussed, classificatory models might be able to mitigate potential noise linked to experimental error or

simultaneous incidence of multiple parallel processes on the bioactivity data. All in all, there is no reason to exclude a model where the mechanism is not known or if there are multiple mechanisms [35]. A final remark must be done in relation to activity cliffs: while continuous structure–activity relationships where gradual change in the molecular structure results in moderate changes in the biological response are benign to modeling efforts, sometimes small modifications in the molecules introduce huge changes in biological activity [17]. Whereas highly informative (they provide valuable information on molecular features essential to the activity) activity cliffs can be really problematic for modeling purposes; some algorithms have been developed to identify these “instances that should be misclassified” [76].

2. *Curation of chemical structures.* While the properties of salts can be very different from those of the corresponding neutral molecules, before excluding salts from the QSAR analysis, some questions should be (if possible) answered. The first question would be if the descriptors that will be used are sensitive to charge. If they are insensitive to charge, just neutralize the cations or anions that are left when removing counterions. If at least some of the descriptors that you will use are sensitive to charge, it might be a good idea to go through the original publications from which biological data of your dataset have been extracted and see if the pH of the tested solutions and the test system are known or can be figured out from the media composition. In any case, the (experimental or at least theoretical) pKa/s of the dataset compounds will be needed to assign their protonation states. The pH of the drug solution is more likely to affect in vivo than in vitro data. In vitro assays usually require buffered systems, thus limiting the influence of the drug solution pH; in general, the protonation state in in vitro models would be thus determined by the test media pH. In contrast, in vivo absorption and thus pharmacokinetics can be greatly influenced by the pH of the drug solution. If the administered dose is known and assumed to be bioavailable, one might choose to assign the protonation state depending on the pH of the target organ; e.g., in the case of antiepileptic drugs, the pH of the brain is around 7 [77]. Always remember to analyze your set of chemicals at those physiologically relevant conditions for the administration route and therapeutic goal; also keep in mind that some pharmaceutically/pharmacologically relevant conditions may vary due to a physiopathological process. If data to afford the previous analysis is not available, neutralizing ions is acceptable [30].
3. *Excluding redundant variables.* When building the model, it is often advised that simultaneous inclusion of highly correlated (redundant) independent variables should be avoided.

Correlated independent variables lead to multicollinearity, which can increase the standard errors associated with the regression coefficients and cause problems in interpreting the results of a regression equation. Orthogonalization procedures can be applied to obtain orthogonal descriptors [78]. Redundant variables can be avoided prospectively by setting a high tolerance value (or by defining a threshold value for its inverse, the variance inflation factor, VIF). Descriptor calculation software usually includes some option to exclude descriptors correlated above a programmer- or user-defined threshold. However, it must be taken into account that even highly correlated descriptor pairs might be included in the model without losing statistical significance ([79] and references therein): it has been suggested that the VIF should be compared to the model inflation factor (MIF) and only if $VIF > MIF$, one of the seemingly redundant descriptors should be excluded.

4. *Validation.* In our opinion, averaging the predictive performance on the folds used for the multifold cross validation is more useful, since it allows computing a confidence interval (or an estimate of the standard deviation) to evaluate, for instance, if the statistical behavior of the original model (the one built using the entire training set) significantly differs from the behavior of the cross validation models on the held-out samples. It is advisable to perform stratified multifold cross validation, in which the folds are selected so that the mean response value is approximately equal in all the folds. For instance, if you are validating a classification model, the folds might comprise an equal number of examples from each class. In the case of Y-randomization, the confidence interval of the metric used to characterize the performance of the randomized model should not contain the value of the performance metric for the original, nonrandomized model (although other ways of analyzing the randomization results have been proposed [13]).

Finally, note that internal validation procedures can be performed at varying confidence levels. One can study the whole process from the variable selection step, or, alternatively, one can validate the influence of the experimental design on the weighting scheme used to measure the contribution of an independent variable to the response. The former is, of course, a more strict challenge to the model robustness, since one is leaving aside any influence of the original training data on the variable selection. The importance of applying cross-validation at the variable selection step has been signaled in the specialized literature [42, 43]. It has also been highlighted that the randomization technique can be of two types: *process randomization*, in which the values of the dependent variable are scrambled and variable selection is done freshly using the whole descriptor pool, and *model randomization*, in which case the response is scrambled but the new QSAR models so

obtained include the same set of independent variables as present in the nonrandom model [80] which is of course a less strict validation approach. In conclusion, use validation techniques to validate the whole modeling procedure, including variable selection, for more reliable results.

5. *ROC curve analysis*. The adequate balance between the true positive rate and the true negative rate is context dependent and not a statistical matter. Sp and Se evolve in an opposite way and therefore they cannot be optimized simultaneously. If the budget for experimental testing is limited, Sp can be prioritized to reduce the false positive rate (predicted hits that will yield negative results) at the expense of sacrificing some true positives; if, in contrast, chemical novelty of the hits is the priority and funding abounds, Sp may be relaxed in favor of Se.

5 Final Remarks

We have presented an overview of the most relevant considerations that must be made in order to develop QSAR models and apply them in VS campaigns. Regarding particular considerations for the case of antiepileptic drug discovery, the complex nature of the disease and the multi-target nature of most of the existing antiepileptic drugs suggest that modeling *in vivo* data (i.e., phenotypic-based drug discovery) might lead to more efficacious drug candidates. As discussed in separate chapters, the classical “the more potent the better” paradigm might not apply to the particular task of antiepileptic drug discovery. Finally, considering the critical role of pharmacokinetics in antiepileptic drug pharmacology, it is advised to include *in silico* ADME filters during the screening process to predict the brain bioavailability and potential biodistribution and metabolism issues of the selected drug candidates, along with potential drug interactions.

References

1. Weaver DF (2008) Principles and practice of computer-aided drug design as applied to the discovery of antiepileptic agents. In: Computational neuroscience in epilepsy, 1st edn. Academic, London
2. Baglo Y, Gabrielsen M, Sylte I et al (2013) Homology modeling of human γ -butyric acid transporters and the binding of pro-drugs 5-aminolevulinic acid and methyl aminolevulinic acid used in photodynamic therapy. *PLoS One* 8:e65200
3. Khan HN, Rashid H, Kulsoom S (2013) Homology modeling of γ -aminobutyrate-aminotransferase, a pyridoxal phosphate-dependent enzyme of Homo sapiens: molecular modeling approach to rational drug design against epilepsy. *Af J Biotechnol* 10:5916–5926
4. Lee J, Daniels V, Sands ZA et al (2015) Exploring the interaction of SV2A with racetams using homology modelling, molecular dynamics and site-directed mutagenesis. *PLoS One* 10:e0116589
5. Temperini C, Innocenti A, Mastrolorenzo A et al (2007) Carbonic anhydrase inhibitors. Interaction of the antiepileptic drug sulthiame with twelve mammalian isoforms: kinetic and X-ray crystallographic studies. *Bioorg Med Chem Lett* 17:4866–4872

- Carta F, Supuran CT, Scozzafava A (2014) Sulfonamides and their isosters as carbonic anhydrase inhibitors. *Future Med Chem* 6:1149–1165
- De Luca L, Ferro S, Damiano FM et al (2014) Structure-based screening for the discovery of new carbonic anhydrase VII inhibitors. *Eur J Med Chem* 71:105–111
- Lavechia A (2015) Machine-learning approaches in drug discovery: methods and applications. *Drug Discov Today* 20:318–331
- Zhang Q, Muegge I (2006) Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting, and consensus scoring. *J Med Chem* 9:1536–1548
- Krüger DM, Evers A (2010) Comparison of structure- and ligand-based virtual screening protocols considering hit list complementarity and enrichment factors. *ChemMedChem* 5:148–158
- Natasenamath C, Isarankura-Na-Ayudhya C, Prachayasittikul V (2010) Advances in computational methods to predict the biological activity of compounds. *Expert Opin Drug Discov* 5:633–654
- Tropsha A (2010) Best practices for QSAR model development, validation and exploitation. *Mol Inf* 29:476–488
- Kiralj R, Ferreira MMC (2009) Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. *J Braz Chem Soc* 20:770–787
- Langer T, Bryant SD (2008) 3D quantitative structure-property relationships. In: *The practice of medicinal chemistry*, 3rd edn. Academic, London
- Sipl W (2010) 3D QSAR: applications, recent advances, and limitations. In: *Recent advances in QSAR studies. Method and applications*, 1st edn. Springer, Dordrecht
- Roy K, Kar S, Das RN (2015) Background of QSAR and historical developments. In: *Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment*, 1st edn. Academic, London
- Jaworska J, Nokolova-Jeliazkova N, Aldenberg T (2005) QSAR applicability domain estimation by projection of the training set descriptor space: a review. *Altern Lab Anim* 33:445–459
- Cruz-Montegudo M, Medina-Franco JL, Pérez-Castillo Y et al (2014) Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde? *Drug Discov Today* 19:1069–1080
- Zhang J (2013) Advancements of outlier detection: a survey. *ISCT Trans Scal Inf Syst* 13:e2
- Roy K, Kar S, Ambure P (2015) On a simple approach for determining applicability domain of QSAR models. *Chemometr Intell Lab Syst* 145:22–29
- Golbraikh A, Shen M, Xiao Z et al (2003) Rational selection of training and test sets for the development of validated QSAR models. *J Comput Aided Mol Des* 17:241–253
- Talevi A, Bellera CL, Castro EA et al (2010) Optimal partition of datasets of QSPR studies: a sampling problem. *MATCH Commun Math Comput Chem* 63:585–599
- Zhu H, Tropsha A, Foruches D et al (2008) Combinatorial QSAR modeling of chemical toxicants tested against *Tetrahymena pyriformis*. *J Chem Inf Model* 48:766–784
- Tetko IV, Sushko I, Pandey AK et al (2008) Critical assessment of QSAR models of environmental toxicity against *Tetrahymena pyriformis*: focusing on applicability domain and overfitting by variable selection. *J Chem Inf Model* 48:1733–1746
- Verma J, Khedkar VM, Coutinho EC (2010) 3D-QSAR in drug design. *Curr Top Med Chem* 10:95–115
- Avgy-David HH, Senderowitz H (2015) Toward focusing conformational ensembles on bioactive conformations: a molecular mechanics/quantum mechanics study. *J Chem Inf Model* 55:2154–2167
- Tasso SM, Moon SC, Bruno-Blanch LE et al (2004) Characterization of the anticonvulsant profile of valpromide derivatives. *Bioorg Med Chem* 12:3857–3869
- Shen M, LeTiran A, Xiao Y et al (2002) Quantitative structure–activity relationship analysis of functionalized amino acid anticonvulsant agents using k nearest neighbor and simulated annealing PLS methods. *J Med Chem* 45:2811–2823
- Young D, Martin T, Venkatapathy R et al (2008) Are the chemical structures in your QSAR correct? *QSAR Comb Sci* 27:1337–1345
- Fourches D, Muratov E, Tropsha A (2010) Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J Chem Inf Model* 50:1189–1204
- Yee LC, Wei YC (2012) Current modeling methods used in QSAR/QSPR. In: *Statistical modeling of molecular descriptors in QSAR/QSPR*, vol 2, 1st edn. Wiley-Blackwell, Weinheim
- González MP, Terán C, Saíz-Urra L et al (2008) Variable selection methods in QSAR: an overview. *Curr Top Med Chem* 8:1606–1627
- Roy K, Kar S, Das RN (2015) Statistical methods in QSAR/QSPR. In: *A primer on QSAR/QSPR modeling*, 1st edn. Springer, Dordrecht

34. Hawkins DM (2004) The problem of overfitting. *J Chem Inf Comput Sci* 44:1–12
35. Cronin MTD (2010) Quantitative structure-activity relationships (QSARs) – applications and methodology. In: *Recent advances in QSAR studies. Method and applications*, 1st edn. Springer, Dordrecht
36. Gramatica P (2010) Chemometric methods and theoretical molecular descriptors in predictive QSAR modeling of the environmental behavior of organic pollutants. In: *Recent advances in QSAR studies. Method and applications*, 1st edn. Springer, Dordrecht
37. Kalisz R, Bączek TB (2010) QSAR in chromatography: quantitative structure-retention relationships (QSRRs). In: *Recent advances in QSAR studies. Method and applications*, 1st edn. Springer, Dordrecht
38. Topliss JG, Costello RJ (1972) Chance correlations in structure-activity studies using multiple regression analysis. *J Med Chem* 15:1066–1068
39. Talevi A, Bellera C, Di Ianni M et al (2012) An Integrated drug development approach applying topological descriptors. *Curr Comput Aided Drug Des* 8:172–181
40. Polanski J, Bak A, Gieleciak R et al (2006) Modeling robust QSAR. *J Chem Inf Model* 46:2310–2318
41. Tropsha A, Gramatica P, Gombar VK (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
42. Roy K, Mitra I (2011) On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Comb Chem High Throughput Screen* 14:450–474
43. Gramatica P (2007) Principles of QSAR models validation: internal and external. *QSAR Comb Sci* 26:694–701
44. Golbraikh A, Tropsha A (2002) Beware of q²! *J Mol Graph Model* 20:269–276
45. Hawkins DM, Basak SC, Mills D (2003) Assessing model fit by cross-validation. *J Chem Inf Comput Sci* 43:579–586
46. Triballeau N, Acher F, Brabet I et al (2005) Virtual screening workflow development guided by the “Receiver Operating Characteristic” curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J Chem Inf Model* 48:2534–2547
47. Nicholls A (2008) What do we know and when do we know it. *J Comput Aided Mol Des* 22:239–255
48. Mysinger MM, Carchia M, Irwin JJ et al (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* 55: 6582–6594
49. Truchon JF, Bayly CI (2007) Evaluating virtual screening methods: good and bad metrics for the “early recognition” problem. *J Chem Inf Model* 47:488–508
50. Robin X, Turck N, Hainard A et al (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77
51. Irwin JJ, Sterling T, Mysinger MM et al (2012) ZINC: a free tool to discover chemistry and biology. *J Chem Inf Model* 52:1757–1768
52. Wishart DS, Knox C, Guo AC et al (2008) DrugBank: a knowledge base for drugs, drug actions and drug targets. *Nucleic Acids Res* 36:D901–D906
53. Novick PA, Ortiz OF, Poelman J et al (2013) SWEETLEAD: an *in silico* database of approved drugs, regulated chemicals, and herbal isolates for computer-aided drug discovery. *PLoS One* 8:e79568
54. Gu J, Gui Y, Chen L et al (2013) Use of natural products as chemical library for drug discovery and network pharmacology. *PLoS One* 8:e62839
55. Misra A, Ganesh S, Shahiwala A et al (2003) Drug delivery to the central nervous system: a review. *J Pharm Pharm Sci* 6:252–273
56. Wager TT, Hou X, Verhoest PR et al (2010) Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. *ACS Chem Neurosci* 1:435–449
57. Alagarsamy V (2013) Structure-activity relationship and quantitative structure-activity relationship. In: *Textbook in medicinal chemistry*, vol 1, 1st edn. Elsevier, New Delhi
58. Nikolic K, Agbaba D (2009) QSAR study and design of novel organoselenium and α -tocopherol derivatives with enhanced chemotherapeutic activity. *Lett Drug Des Discov* 6:228–235
59. Bello-Ramírez AM, Buendía-Orozco J, Nava-Ocampo AA (2003) A QSAR analysis to explain the analgesic properties of aconitum alkaloids. *Fundam Clin Pharmacol* 17:575–580
60. Patel SR, Gangwal R, Sangamwar AT et al (2014) Synthesis, biological evaluation and 3D-QSAR study of hydrazide, semicarbazide and thiosemicarbazide derivatives of 4-(adamantan-1-yl)quinoline as anti-tuberculosis agents. *Eur J Med Chem* 85:255–267

61. Richard AM, Benigni R (2002) AI and SAR approaches for predicting chemical carcinogenicity: survey and status report. *SAR QSAR Environ Res* 13:1–19
62. Myshkin E, Brennan R, Khasanova T et al (2012) Prediction of organ toxicity endpoints by QSAR modeling based on precise chemical-histopathology annotations. *Chem Biol Drug Des* 80:406–416
63. Roth BL, Sheffler DJ, Kroeze WK (2004) Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat Rev Drug Discov* 3:353–359
64. Bianchi MT, Pathmanathan J, Cash SS (2009) From ion channels to complex networks: magic bullet versus magic shotgun approaches to anticonvulsant pharmacotherapy. *Med Hypotheses* 72:297–305
65. Talevi A (2015) Multi-target pharmacology: possibilities and limitations of the “skeleton key approach” from a medicinal chemist perspective. *Front Pharmacol* 6:205
66. Sams-Dodd F (2005) Target-based drug discovery: is something wrong? *Drug Discov Today* 10:139–147
67. Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 4:682–690
68. Swinney DC, Anthony J (2011) How were new medicines discovered? *Nat Rev Drug Discov* 10:507–519
69. Talevi A, Bellera CL, Castro EA et al (2007) A successful virtual screening application: prediction of anticonvulsant activity in MES test of widely used pharmaceutical and food preservatives methylparaben and propylparaben. *J Comput Aided Mol Des* 21:527–538
70. Talevi A, Enrique AV, Bruno-Blanch LE (2012) Anticonvulsant activity of artificial sweeteners: a structural link between sweet-taste receptor T1R3 and brain glutamate receptors. *Bioorg Med Chem Lett* 22:4072–4074
71. Sutherland JJ, Weaver DF (2003) Development of quantitative structure–activity relationships and classification models for anticonvulsant activity of hydantoin analogues. *J Chem Inf Model* 43:1028–1036
72. Hashemianzadeh M, Safarpour MA, Ghlamjani-Moghddam K et al (2008) DFT-based QSAR study of valproic acid and its derivatives. *QSAR Comb Sci* 27:469–474
73. Sturm N, Desaphy J, Quinn RJ et al (2012) Structural insights into the molecular basis of the ligand promiscuity. *J Chem Inf Model* 52:2410–2421
74. Speck-Planche A, Cordeiro MN (2015) Multitasking models for quantitative structure-biological effect relationships: current status and future perspectives to speed up drug discovery. *Expert Opin Drug Discov* 10:245–256
75. Liu Q, Zhou H, Liu L et al (2011) Multi-target QSAR modelling in the analysis and design of HIV-HCV co-inhibitors: an in-silico study. *BMC Bioinformatics* 12:294
76. Smith MR, Martínez T (2011) Improving classification accuracy by identifying and removing instances that should be misclassified. The 2011 International Joint Conference on Neural Networks, IEEE, pp 2690–2697
77. Vorstrup S, Jensen KE, Thomsen C et al (1989) Neuronal pH regulation: constant normal intracellular pH is maintained in brain during low extracellular pH induced by acetazolamide-³¹P NMR study. *J Cereb Blood Flow Metab* 9:417–421
78. Duchowicz PR, Fernández FM, Castro EA (2010) Orthogonalization methods in QSPR. In: *QSPR-QSAR studies on desired properties for drug design*, 1st edn. Research Signpost, Trivandrum
79. Peterangelo SC, Seybold PG (2004) Synergistic interactions among QSAR descriptors. *Int J Quantum Chem* 96:1–9
80. Roy K, Kar S (2015) How to judge predictive quality of classification and regression based QSAR models? In: *Frontiers in computational chemistry*, vol 2, 1st edn. Elsevier, Amsterdam