# Chapter 19

# A Round Trip from Medicinal Chemistry to Predictive Toxicology

**Giuseppe Felice Mangiatordi, Angelo Carotti, Ettore Novellino, and Orazio Nicolotti**

## Abstract

Predictive toxicology is a new emerging multifaceted research field aimed at protecting human health and environment from risks posed by chemicals. Such issue is of extreme public relevance and requires a multidisciplinary approach where the experience in medicinal chemistry is of utmost importance. Herein, we will survey some basic recommendations to gather good data and then will review three recent case studies to show how strategies of ligand- and structure-based molecular design, widely applied in medicinal chemistry, can be adapted to meet the more restrictive scientific and regulatory goals of predictive toxicology. In particular, we will report:

- Docking-based classification models to predict the estrogenic potentials of chemicals.
- Predicting the bioconcentration factor using biokinetics descriptors.
- Modeling oral sub-chronic toxicity using a customized k-nearest neighbors (k-NN) approach.

**Key words** Docking-based classification models, Estrogenic potentials of chemicals, Bioconcentration factor, Biokinetics descriptors, Oral sub-chronic toxicity

## 1 Introduction

Predicting the effects of xenobiotics, not solely drugs, is far from being a winning bet. Their interplay with living organisms is in fact responsible for biological/toxicological actions which are often not easy to predict. On the other hand, predictions can be made on the basis of (a) in vivo experiments based on direct animal testing, (b) in vitro experiments making use of tissue culture cells, and (c) in silico simulations by employing computer models. It is widely acknowledged that in vivo and in vitro experiments are time demanding and expensive. Great efforts have been thus directed to develop in silico approaches. Such computational strategies allow a

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_19, 

significant save in terms of money, time, and, above all, laboratory animals and provide reliable toxicological evidence in order to minimize or replace in vivo assays according to the "three Rs" principle (replacement, reduction, refinement) [1]. In our opinion, computational methods are thus complementary to experimentation and prospectively capable of replacing empirical testing. The tendency is thus that of moving from experiments to exploratory toxicology which can provide timely go/no-go decisions and represents a viable alternative for the prediction of biological/toxicological effects [2, 3].

In the present survey, we will review some ad hoc examples taken from our recent studies showing how adapting consolidated drug discovery strategies to the scientific and regulatory goals of exploratory toxicology. First of all, we will emphasize the importance of having high-quality data to ensure the derivation of trustable models. In this respect, some practical recommendations will be given. Then, we will discuss how applying molecular docking, perhaps the most popular structure-based method employed by medicinal chemists, to obtain classifiers for discerning estrogenic from non-estrogenic substances. In the second case studies, we will present how QSAR models can be derived and applied to predict the bioconcentration factor, a relevant ecotoxicological endpoint. In this respect, attention will be paid to the appropriate use of biokinetics descriptors and to the definition of the applicability domain to ensure both model transparency and adequacy. Finally, we will describe how customizing a k-NN algorithm to properly model oral sub-chronic toxicity. We will show how the implementation of user-adjustable rules can be very effective to increase the confidence in data prediction, which is the ultimate aim of computational toxicology.

## 2 Looking for High-Quality Data: Some Practical Recommendations

The advent of new regulations concerning the protection of human health and environment has strengthened the role of QSAR. Such methodology has today assumed the *status* of a mature discipline for both scientific and regulatory purposes. The pressing need of regulatory bodies and industries for the derivation of adequate QSAR models has led to issue some best practices, which are, at present, key elements for successful predictive in silico toxicological studies. Some seminal papers [4–6] have clearly demonstrated that the predictive potential of QSAR models is mostly dependent from the quality of chemical descriptors rather than from the sophistication of the employed optimization techniques. A high-quality data is therefore essential for obtaining trustable models. In this respect, several preliminary checks need to be taken into

account for steering away from even small structural mistakes whose occurrence can result in inaccurate molecular descriptors, which in the end are responsible for disappointing predictions. To circumvent this pitfall [7], great attention has been given to the data curation, a preprocessing treatment necessary to discard or amend chemical records, which are difficult to handle with conventional cheminformatics techniques. Normally, data curation is applied to filter out inorganic and organometallic compounds, counterions, salts, and mixtures. In addition, data curation is carried out to standardize the ring aromatization, to uniform specific chemotypes, to assign tautomeric forms, and to remove duplicates.

Since model reliability is strictly dependent on data quality (i.e., garbage in, garbage out), QSAR developers should also pay high attention in appropriately sizing the dataset and in fairly balancing structural classes or categories, which in real-life investigations are often unevenly represented. It would be advisable that the number of compounds in the dataset should not be too small since this could lead to the occurrence of chance correlation and overfitting; both these phenomena can deteriorate the real predictive power of models. Moreover, a small-sized dataset would be unsuitable for validation analyses. On the other hand, there is not an upper limit to define a maximum size. In this respect, a key role is played by the algorithm implemented for deriving QSAR as well as by the available resources (e.g., computer and time). For practical reasons, a too large dataset can be reduced by selecting a given subset of chemically diverse compounds or can be partitioned in clusters from which deriving multiple and independent models. However, some golden rules should be observed to split the initial dataset into a training set for model derivation and into a test and external set for model validation. In case of continuous response variables, at least 40 compounds should be considered: 20 compounds in the training set and 10 compounds in both test and external sets. Moreover, the response variables should cover a range at least five times larger than the experimental error and should be fairly distributed over such entire range. In case of classification or category response variables, at least 20 compounds per class are recommended: the training set should be made of no less than 10 compounds per class while test and external sets no less than 5 each.

Another reason of attrition in QSAR derivation is given by compounds, which are typical chemical singletons, being their structural features far away from those of all the other compounds within a dataset. In other words, they could behave as leverage (or structural) outliers. Other compounds could instead act as activity outliers as they rebut the basic QSAR assumption stating that similar compounds have similar properties. As reported in a number of

seminal works [8, 9], these compounds could originate the so-called cliffs of the descriptor space where a given response property (i.e., biological/toxicological response) changes dramatically for an even subtle structural variation. Actually, both these types of outliers can be real or sometimes due to accidental errors in reporting the chemical structure or in annotating the response variable. Normally, it is wise to remove them prior to model derivation as they will likely cause model instability and deeply affect predictions.

Moreover, high-quality molecular descriptors are essential to derive predictive and interpretable QSAR models [10]. Nowadays, it is quite easy to quickly calculate an overwhelming number of descriptors [11] related to two- or three-dimensional molecular aspects, although their mechanistic interpretation remains somewhat obscure to mid-level QSAR practitioners. Needless to say that medicinal chemists have long debated about chemical desirability, a concept inherent to the chemical meaning of QSAR model [12, 13]. We can guess that descriptors referring to the passage of xenobiotics across cellular membranes, for instance, may be desirable in a toxicological context. In this respect, we do believe that ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties would make the descriptors space more attractive for toxicological purposes and of adequate transparency for molecular and numerical modeling. ADMET properties are in fact important to study the fate and disposition of drugs and to monitor their behavior in the body at therapeutic doses (i.e., pharmacokinetic properties). Importantly, the studies of ADMET properties are not limited to drugs but can be extended to any chemical, including environmental pollutants, potentially affecting human health. In this respect, the term toxicokinetics and, even better, the more inclusive term biokinetics [14] are normally used to describe and, then, to predict unwanted toxic effects of xenobiotics on living system exposed to chemicals at any dosage regimen. The masterpiece by Waterbeemd [15] describes the progress made by medicinal chemistry in the attempt of refining ADMET properties in order to reduce the costly late-stage failures in drug development and thereby accelerating the drug discovery process. Such efforts have resulted in the wide introduction of ADMET-related descriptors implemented in in silico methods to predict the most relevant pharmacokinetic, metabolic, and toxicity endpoints.

## 3 Docking-Based Classification Models to Predict the Estrogenic Potentials of Chemicals

Predicting the endocrine disruptor potential of chemicals and, more specifically, their ability to interfere with the estrogen receptors (ERs) is a theme of utmost relevance [16]. Unlike previous

predictive models [17–19], we have recently described how the current availability of X-ray-solved target structures can be employed [20]. Importantly, accounting for physicochemical information on the biological target allows a larger applicability domain with respect to classical QSAR-like models.

We used a three-dimensional (3D) training dataset (hereafter referred to as EPA-ERDB) consisting of 1677 chemical structures shared by US EPA. For each chemical, the estrogenic/nonestrogenic action was derived from concentration-response data resulting from 18 high-throughput assays probing several sites of the mammalian ER pathway. Challengingly, the 1677 chemicals were unevenly distributed, being only 237 (14.13 %) chemicals designated as ER binders. To possibly cover a broader spectrum of possible biological actions of compounds comprised within the EPA-ERDB training dataset, eight ER crystal structures were retrieved from the Protein Data Bank (PDB) for docking simulations. All four possible ER classes were considered: (1) ERα bound to agonist, (2) ERα bound to antagonist, (3) ERβ bound to agonist, and (4) ERβ bound to antagonist. The 3D conformations of the 1677 chemicals in the training dataset were subjected to docking simulations performed by both GLIDE v.6.5 [21] and GOLD v.5.2 [22], two very popular software largely adopted in drug discovery projects. The ability of the selected docking protocols to discern binders from nonbinders was assessed using typical confusion matrix, which includes information about experimental and predicted matches and mismatches returned for each classification system. Next, docking performance was evaluated using the enrichment factor (EF), which represents the percentage of known binders found at a given percentage of the ranked database. In addition, we reported the EF at the early 1 % of the ranked dataset (i.e., EF1%). Predictive docking-based classification models are expected to return similar values for both EF1% and EFmax (a reference ideal value obtained by dividing the total number of chemicals by the total number of binders). All these data were derived from the obtained receiver operating characteristic (ROC) curves (*see* Fig. 1). The thresholds for defining the classes were set on the basis of the desired sensitivity (SE) values. The value of SE estimates the proportion of true positives that are correctly identified. In order to designate the estrogenic or nonestrogenic potential, two SE values equal to 0.25 and to 0.75 were set as thresholds to define, for each ER crystal, three probability binding classes as follows:

(a) SE $\leq$0.25, the class with high probability of binding (i.e., binder molecules).

(b) SE >0.75, the class with low probability of binding (i.e., nonbinder molecules).

(c) $0.25 < SE \leq 0.75$, the class with medium probability of binding (i.e., suspicious molecules).

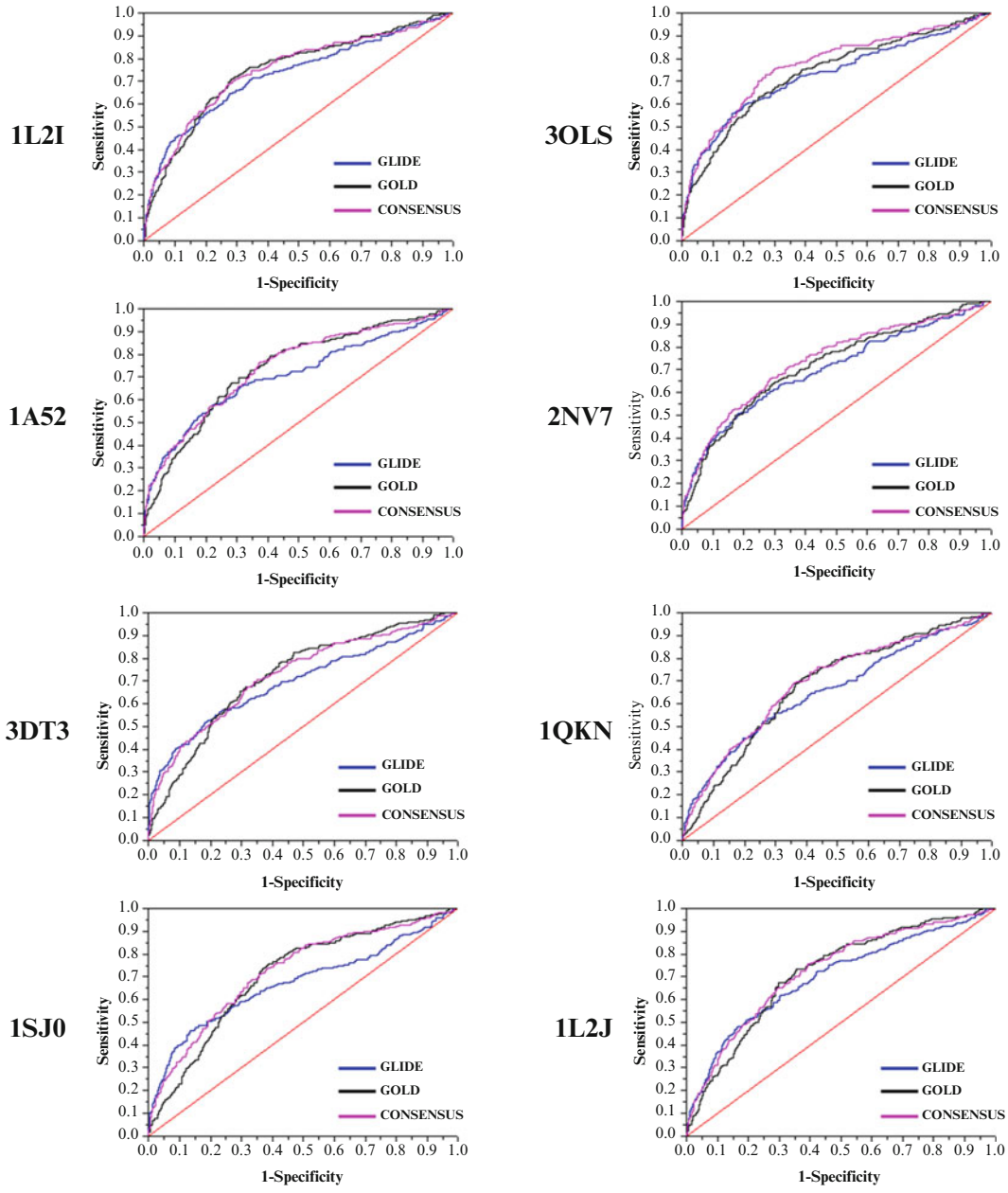


**Fig. 1** ROC curves derived from ERα (PDB entries: 1L2I, 1A52, 3DT3, and 1SJ0) and ERβ structures (PDB entries: 3OLS, 2NV7, 1QKN, and 1L2J) are shown on the *left* and *right hand side*, respectively (taken from 20)

At a given threshold, the goodness of the classification was assessed using two parameters: (a) the positive predictive value (PPV) that is related to the probability that a chemical predicted as a binder (over-threshold) is actually a binder and (b) the negative predictive value (NPV) that is related to the probability that a

chemical predicted as a nonbinder (under-threshold) is actually a nonbinder. However, the pronounced asymmetry of data prompted us to compute the positive (+LR) and the negative likelihood ratio (−LR) for each of the SE-considered thresholds. Briefly, the greater the +LR is at a given threshold, the better the performance of the classification model. It is worthy to say that these likelihood ratios are independent from the data distribution within the training set.

We observed that, unequivocally, GLIDE detects a higher number of binders in the earliest fraction of the rank despite the lower AUC values. For all ER crystal structures, the ability to minimize FPs is higher with GLIDE with respect to GOLD, in agreement with the already discussed EF1% factors. Importantly, an opposite trend can be detected if the second threshold (SE = 0.75) is considered. GOLD returns PPV values higher than GLIDE. In other words, GLIDE ensures better performances in terms of ability to minimize FPs, whereas the interest is mostly oriented to the upper part of the ranking. Our results would suggest that the use of GLIDE or GOLD depends on the pursued goals. As shown, there is not a winning model, but rather a case-by-case evaluation should be made. Docking-based classification models have allowed to employ the wealth of physicochemical information contained in the native protein structures to screen large chemical collections and demonstrated to be helpful for immediately obtaining a preliminary idea of the estrogenic activity by simply comparing the docking score of a target chemical with those reported at the different SE-based thresholds.

## 4 Predicting the Bioconcentration Factor Using Biokinetics Descriptors

The bioconcentration factor (BCF) represents the ratio of the concentration of a substance in an aquatic organism with respect to that in water [23]. It is an endpoint of utmost relevance due to its costs and its (eco)toxicological impact. Its assessment should be done following the experimental test OECD 305, which requires for each substance more than hundreds of fishes, months for test execution, and tens of thousands of Euros [24]. The herein used data [25] comprises 851 chemicals, retrieved from the ANTARES dataset. The obtained dataset was split into three subsets: about 10 % (78 out of 851) of the compounds were randomly selected to form the blind set (BS), required for final validation. The remaining chemicals were split to ensure a uniform distribution of their experimental BCF values, applying the Venetian blinds method [26], to form training set (TS) and validation set (VS) containing 620 and 153 chemicals, respectively. These selection criteria were used to obtain two different and independent sets for model validation and to ensure the most realistic situation for the external compounds, so that statistics could explain the real capability of the model to predict new compounds, as it should be for regulatory purposes.

Many commercial and free software programs are available for the calculation of thousands of two-dimensional (2D) or three-dimensional (3D) descriptors. In the present work, we preferred to calculate a smaller number (i.e., 51) of ADMET (absorption, distribution, metabolism, excretion, and toxicity)-relevant descriptors that are closely related to pharmaceutical properties of organic molecules. To this end, we used QikProp 3.4 [27] included in Schrödinger 2011-1 suite [28]. Note that, as already mentioned, descriptors referring to the permeation of the membrane may be more desirable for a toxicological or pharmacological audience. A number of models were derived using the Monte Carlo approach (simulated annealing), multiple linear regression (MLR), and neural network algorithm (NN). Importantly, the obtained models could be flexibly adapted to play as classifiers using as thresholds those established in Annex XIII of REACH to classify chemicals. All substances that exceed the first threshold of log BCF = 3.3 are classified as bioaccumulative (B), while those having log BCF < 3.3 are classified as nonbioaccumulative (nB) according to the PBT (persistent, bioaccumulative, and toxic) definition; on the other hand, all substances that exceed the second threshold of log BCF = 3.7 are classified as very bioaccumulative (vB).

Among others, our attention was mostly engaged by a nine-descriptor model. Apart from robust statistics, particular attention was paid to the definition of the applicability domain (AD). Needless to say that predictions provided by models without a clearly defined AD are meaningless [29–31]. As previously described, its importance has also been remarked in REACH Annex XI, BPR Annex IV, and OECD principles for the derivation of acceptable QSARs. In our studies, we implemented a multi-step filter system to confidently designate chemicals within the AD only those having the matching criteria requested at any step. Such procedure ensures higher confidence and transparency irrespective of the accuracy of predictions [32].

The first independent filter accounted for the dataset structural diversity. Briefly, the occurrence of organic functional group (nested) was assessed using the QSAR Toolbox 3.0 software, released by OECD—2013. The second independent filter accounted for the chemical descriptors range. The minimum and maximum values of the nine descriptors in the model for TS chemicals were used a criterion of interval validity. In this respect, VS or BS chemicals whose descriptors violated even only one range were placed outside AD. The third independent filter was a geometrical trap based on the interpolation region space representing the smallest convex area whose borders describe the perimeter of a polygon containing TS compounds. In particular, the interpolation polygon was drawn using spatial coordinates of the first two principal components of the multivariate descriptor space of the nine-term model. The polygon area was reduced to include the top 98 % TS compounds (considering their closeness to TS centroid)

to avoid the inclusion of underrepresented areas likely increasing the prediction uncertainty. Finally, the leverage method was applied as fourth independent filter. Briefly, the leverage represents the compound distance from the model experimental space (that is the center of TS observations) and, thus, provides a measure of the degree of influence that a particular TS chemical structure has on the model or the degree of extrapolation for the prediction of VS and BS compounds. In this respect, VS and BS compounds having leverages exceeding the widely acknowledged threshold of $h^* = 3p'/n$ (where $p'$ is the number of model variables plus one and $n$ is the number of TS compounds) were placed outside model AD being poor reliable predictable [33].

The simultaneous application of multi-filter system has the effect of leaving outside AD: (a) a number of 20 (13 % of the initial) VS compounds with an indirect gain of $r^2$ from 0.635 to 0.765 and of RMSE from 0.794 to 0.616 and (b) a number of 7 (9 % of the initial) BS compounds with an indirect gain of $r^2$ from 0.623 to 0.659 and of RMSE from 0.841 to 0.817 (*see* Fig. 2).

The harmonic application of consolidated QSAR approaches employing pharmaceutically relevant descriptors and a multi-step filter system to designate chemicals inside/outside AD demonstrated to be very effective for modeling BCF data, an endpoint of utmost importance in both toxicological and regulatory terms.

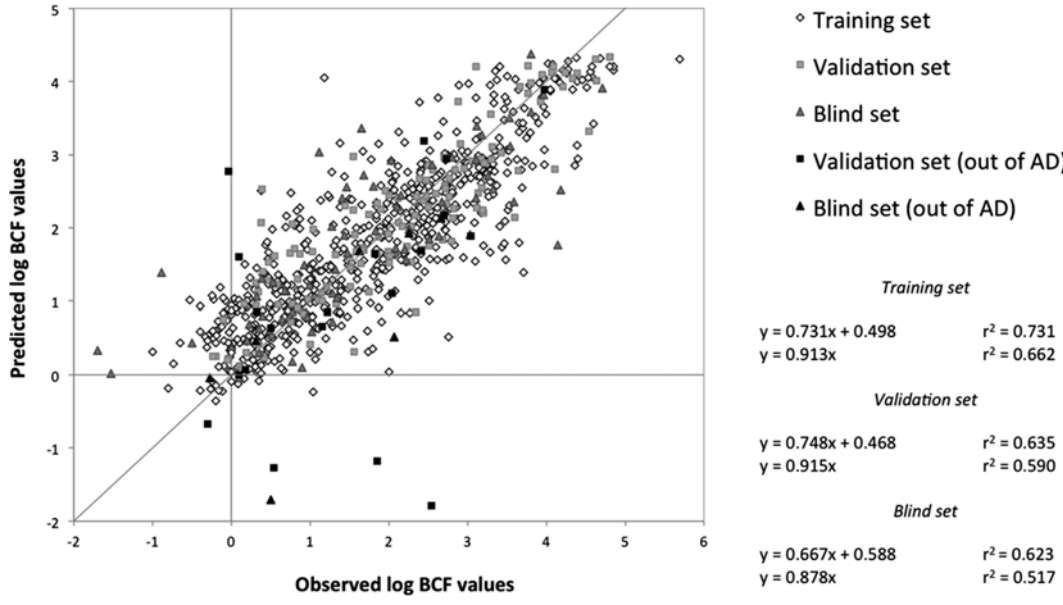


**Fig. 2** Comparison of the experimental and predicted log BCF values obtained through the nine-descriptor BCF model. TS, VS, and BS chemicals are represented by *white diamonds*, *gray squares*, and *upside triangles*, respectively. VS and BS outside AD chemicals are represented by *black squares* and *upside triangles*, respectively. The continuous line represents the case of ideal correlation (taken from 25)

## 5 Modeling Oral Sub-chronic Toxicity Using a Customized k-Nearest Neighbors (k-NN) Approach

Repeated dose toxicity (RDT) is an important endpoint to toxicologically profile a given chemical after repeated administration. RDT studies are focused on the no observed (adverse) effect level (NO(A)EL) and on the lowest observed (adverse) effect level (LO(A)EL). The former is the higher experimental dose at which there is no appreciable response [34]; the latter indicates the lowest dosage at which adverse effects occur in comparison with a control group (e.g., onset of an adverse effect) [35]. The NO(A)EL and LO(A)EL are assessed by means of in vivo studies that can be based on various protocols accounting for different exposure period, animal model (rodent or non-rodent species) and exposure route (oral, inhalation or dermal) [36]. As a result, regulators explicitly require data relative to repeated dose toxicity.

We recently conducted a toxicological study [37] focused on RDT data for sub-chronic oral exposure (i.e., 90 days) in rats. Training set data was retrieved from different sources (i.e., Munro database, Hazard Evaluation Support System, EPA's Integrated Risk Information System). In particular, 254 chemicals were selected being the ones having unequivocal values of chronic toxicity studies (from 84 to 98 days) of oral exposure (gavage, diet, or drinking water). An external dataset comprising 179 chemicals was also used to challenge the predictive power of our models. External dataset data were taken and properly selected from the RepDose database.

A customized k-nearest neighbors (k-NN) approach for predicting sub-chronic oral toxicity in rats was used (*see* Fig. 3). The basic idea was that of predicting a given response on the basis of those observed in the most structurally similar chemicals. The straight application of the k-NN was however very disappointing. To overcome this limitation, the algorithm was ad hoc adapted by implementing several rules to better control the reliability of predicted chemicals. The gain in prediction and confidence was obtained for a given percentage of the dataset; the reasonable price to pay was that a number of compounds (those unmatching the new rules implemented in the k-NN) were left unpredicted as a precautionary measure. However, the use of restrictive conditions in modeling such a complex endpoint meets both the scientific and regulatory purposes established by international bodies for the protection of human health. In fact, providing few but highly reliable predictions represents a valuable prioritization strategy to generate trustable toxicological information on the substances and, at the same time, to support the use of alternative methods and thus to reduce the number of animals needed for in vivo testing.

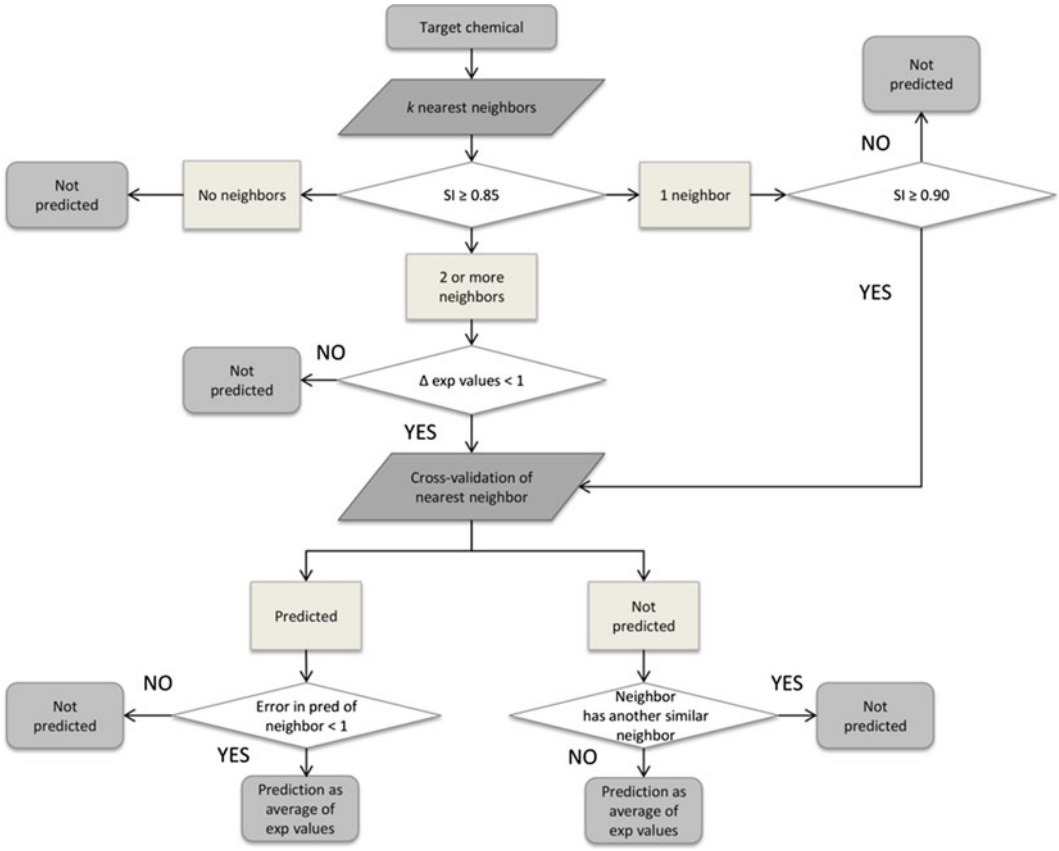


**Fig. 3** Flowchart for the selection of the output predictions. SI or similarity index between the target chemical and its nearest neighbors; Δ exp values are the difference between experimental values of nearest neighbors; error in pred is the error in prediction returned in cross validation of a neighbor in the TS (taken from 37)

## 6 Conclusions and Perspectives

Exploratory toxicology employs in silico methods for their importance in scientific and regulatory context. Indeed, the need of protecting human health and environment has prompted public authorities, such as the US Environmental Protection Agency (US EPA) and the European Chemicals Agency (ECHA) to play a frontline role in the promotion of programs of predictive toxicology to assess the risk posed by chemicals. For instance, the European Commission (EC) has issued, in Annex XI of REACH and Annex IV of BPR, four conditions for using in silico in place of in vivo testing: (1) results have to be derived from a QSAR model whose scientific validity has been well established, (2) the substances are expected to fall within the applicability domain of the QSAR model, (3) results need to be adequate for the purpose of classification and labeling and/or risk assessment, and (4) adequate and reliable documentation of the applied method has to be

provided. Importantly, these recommendations for the implementation of the so-called non-testing methods are perfectly known to medicinal chemists, whose community is continuously discussing roles and goals. It is well known that medicinal chemists have in recent years already openly deplored the frequent temptation of discussing highly speculative computational predictions that are often the result of over-interpreted but not properly validated models. In this respect, a blacklist of simply decorative and colorful QSAR models has been matter of a strong skepticism, as recently pointed out by Cramer [38]. In this continuing debate, we do believe that modern medicinal chemists should be strongly committed to face the new challenge of exploratory toxicology, which implies more restrictive scientific and regulatory purposes (i.e., chemical prioritization, selecting compounds for further experimental testing, reducing the number of false negatives, harmful compounds predicted as safe). By discussing three case studies, we reported how successfully adapting consolidated structure- and ligand-based strategies, largely applied in drug discovery programs, to the goal of exploratory toxicology. Needless to repeat that a critical case-by-case assessment is necessary to prove the result reliability and to make trustable the adopted approach. Indeed, an informed interpretation of the results can make the difference. However, we are just at the beginning of a new fascinating journey requiring new scientific efforts and challenges.

## References


1. Russell WMS, Burch RL (1959) The principles of humane experimental technique. Johns Hopkins Bloom Sch Public Health, Baltimore
2. Hornberg JJ, Laursen M, Brenden N et al (2014) Exploratory toxicology as an integrated part of drug discovery. Part I: why and how. Drug Discov Today 19:1131–1136
3. Nicolotti O, Benfenati E, Carotti A et al (2014) REACH and in silico methods: an attractive opportunity for medicinal chemists. Drug Discov Today 19:1757–1768
4. Young D, Martin T, Venkatapathy R et al (2008) Are the chemical structures in your QSAR correct? QSAR Comb Sci 27:1337–1345
5. Zhu H, Tropsha A, Fourches D et al (2008) Combinatorial QSAR modeling of chemical toxicants tested against Tetrahymena pyriformis. J Chem Inf Model 48:766–784
6. 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
7. 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
8. Maggiora GM (2006) On outliers and activity cliffs—why QSAR often disappoints. J Chem Inf Model 46:1535
9. Cruz-Monteagudo 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
10. Nicolotti O, Carotti A (2006) QSAR and QSPR studies of a highly structured physicochemical domain. J Chem Inf Model 46: 264–276
11. Todeschini R, Consonni V (eds) (2000) Handbook of molecular descriptors. Wiley, Weinheim
12. Nicolotti O, Gillet VJ, Fleming PJ et al (2002) Multiobjective optimization in quantitative structure-activity relationships: deriving accurate and interpretable QSARs. J Med Chem 45:5069–5080

13. Bickerton GR, Paolini GV, Besnard J et al (2012) Quantifying the chemical beauty of drugs. Nat Chem 4:90–98
14. Bouvier d'Yvoire M, Prieto P, Blaauboer BJ et al (2007) Physiologically-based Kinetic Modelling (PBK Modelling): meeting the 3Rs agenda. The report and recommendations of ECVAM Workshop 63. Altern Lab Anim ATLA 35:661–671
15. Van de Waterbeemd H, Gifford E (2003) ADMET in silico modelling: towards prediction paradise? Nat Rev Drug Discov 2:192–204
16. Diamanti-Kandarakis E, Bourguignon J-P, Giudice LC et al (2009) Endocrine-disrupting chemicals: an endocrine society scientific statement. Endocr Rev 30:293–342
17. Shi LM, Fang H, Tong W et al (2001) QSAR models using a large diverse set of estrogens. J Chem Inf Comput Sci 41:186–195
18. Devillers J, Marchand-Geneste N, Carpy A et al (2006) SAR and QSAR modeling of endocrine disruptors. SAR QSAR Environ Res 17:393–412
19. Jacobs MN (2004) In silico tools to aid risk assessment of endocrine disrupting chemicals. Toxicology 205:43–53
20. Trisciuzzi D, Alberga D, Mansouri K et al (2015) Docking-based classification models for exploratory toxicology studies on high-quality estrogenic experimental data. Future Med Chem 7:1921–1936. doi:10.4155/FMC.15.103
21. Small-molecule drug discovery suite 2014-4: glide, version 6.5 (2014) Schrödinger, LLC, New York
22. Jones G, Willett P, Glen RC et al (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267:727–748
23. Arnot JA, Gobas FA (2006) A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms. Environ Rev 14:257–297
24. OECD (2012) Test no. 305: bioaccumulation in fish: aqueous and dietary exposure, OECD guidelines for the testing of chemicals, section 3. OECD Publishing, 12.10.12, 72. (http://dx.doi.org/10.1787/9789264185296-en). / http://dx.doi.org/10.1787/9789264185296-enS. Consulted July 2015
25. Gissi A, Gadaleta D, Floris M et al (2014) An alternative QSAR-based approach for predicting the bioconcentration factor for regulatory purposes. ALTEX 31:23–36
26. Consonni V, Ballabio D, Todeschini R (2009) Comments on the Definition of the Q2 Parameter for QSAR Validation. J Chem Inf Model 49:1669–1678
27. Schrödinger Release 2011-1: QikProp, version 3.4 (2011) Schrödinger, LLC, New York
28. Schrödinger Release 2011-1 (2011) Schrödinger, LLC, New York
29. Jaworska J, Nikolova-Jeliazkova N, Aldenberg T (2005) QSAR applicability domain estimation by projection of the training set descriptor space: a review. Altern Lab Anim ATLA 33:445–459
30. Aptula AO, Roberts DW (2006) Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity. Chem Res Toxicol 19:1097–1105
31. Roberts DW, Aptula AO, Patlewicz G (2006) Mechanistic applicability domains for nonanimal based prediction of toxicological endpoints. QSAR analysis of the Schiff base applicability domain for skin sensitization. Chem Res Toxicol 19:1228–1233
32. Schultz TW, Hewitt M, Netzeva TI et al (2007) Assessing applicability domains of toxicological QSARs: definition, confidence in predicted values, and the role of mechanisms of action. QSAR Comb Sci 26:238–254
33. Gramatica P (2010) Chemometric methods and theoretical molecular descriptors in predictive QSAR modeling of the environmental behavior of organic pollutants. In: Puzyn T, Leszczynski J, Cronin MT (eds) Recent advances in QSAR studies. Springer, Netherlands, pp 327–366
34. Sand S, Victorin K, Filipsson AF (2008) The current state of knowledge on the use of the benchmark dose concept in risk assessment. J Appl Toxicol 28:405–421
35. Sakuratani Y, Zhang HQ, Nishikawa S et al (2013) Hazard Evaluation Support System (HESS) for predicting repeated dose toxicity using toxicological categories. SAR QSAR Environ Res 24:351–363
36. SCCS—Scientific Committee on Consumer Safety (2012). The SCCS's notes of guidance for the testing of cosmetics substances and their safety evaluation 8th revision. (http://ec.europa.eu/health/scientific_committees/consumer_safety/docs/ sccs_s_006.pdf). Consulted April 2014
37. Gadaleta D, Pizzo F, Lombardo A et al (2014) A k-NN algorithm for predicting the oral subchronic toxicity in the rat. ALTEX 31:423–432
38. Cramer RD (2012) The inevitable QSAR renaissance. J Comput Aided Mol Des 26:35–38