Chapter 4
In Silico ADME Techniques Used in Early-Phase Drug Discovery

Matthew L. Danielson, Bingjie Hu, Jie Shen, and Prashant V. Desai

Abstract The process of drug discovery and development is time consuming and expensive. In silico tools, in combination with in vitro and in vivo models, provide a valuable resource to improve the efficiency of this process. In this chapter, we provide an overview of various in silico tools and models used to identify and resolve absorption, distribution, metabolism, and excretion (ADME) challenges in drug discovery. In general, structure-based in silico techniques such as docking and molecular dynamics simulations have limited applicability in the ADME space due to the promiscuity of many ADME targets and the limited availability of high-resolution 3-D structures. Pharmacophore models, a ligand-based in silico method, can be used to identify key structural features responsible for the interaction with the target of interest. However, due to broad ligand specificity and the probability of multiple binding sites in many ADME targets, pharmacophore models have limited prospective applicability across structurally diverse chemical scaffolds. Conversely, quantitative structure-property relationship (QSPR) models are capable of extracting knowledge from a wide variety of chemical scaffolds and have prospectively shown utility as predictive models for many ADME endpoints measured in the pharmaceutical industry. QSPR models, especially those based on machine learning techniques, are known to have limited interpretability. To address this challenge, the use of QSPR models is typically coupled with information derived from trends between ADME endpoints and physicochemical properties (e.g., lipophilicity, polar surface area, number of hydrogen bond donors, etc.) during drug discovery. Furthermore, knowledge extracted by the matched molecular pair analysis (MMPA) of ADME data provides insight that is used to identify fragment replacements to improve the ADME characteristics of compounds. In conclusion, an effective amalgamation of in silico tools is necessary to influence the design of compounds that will possess favorable ADME properties. Finally, in silico tools should never be used in isolation; they make up one arm of the

M.L. Danielson • B. Hu • J. Shen • P.V. Desai (✉)
Computational ADME, Drug Disposition, Lilly Research Laboratories,
Eli Lilly and Company, Indianapolis, IN 46285, USA
e-mail: desai_prashant_v@lilly.com

integrated and iterative learning cycle that is comprised of in silico, in vitro, and in vivo models that we recommend using to effectively drive a drug discovery project.

Keywords In silico ADME • Quantitative structure-property relationship models • Matched-molecular pair analysis • Predictive models • Physico-chemical properties

The drug discovery and development process is time consuming and expensive, encompassing approximately 15 years and over two billion dollars to bring a drug to market [1]. Stage-appropriate use of models is an integral part of the drug discovery process. Early-phase drug discovery uses various in silico and in vitro models to explore potency, ADME properties, and safety. As drug discovery progresses, preclinical in vivo animal models are used to estimate how a compound will behave in humans, and ultimately model situations are created in a controlled clinical environment (clinical models) before the compound is approved for use in the general population.

In an attempt to reduce the time and cost associated with the drug discovery process, in silico tools are one class of models employed throughout this process. In silico tools have a direct impact on how drug discovery progresses and are especially useful in the early-phase of drug discovery where a clinical candidate is being pursued and optimized. These tools are used to design and prioritize the synthesis of compounds with desirable affinity, specificity, a multitude of ADME properties, and safety with the goal of delivering the best possible compound to test in the clinical setting.

In this chapter, we provide an overview of various in silico models and tools employed to identify and resolve ADME challenges during the process of drug discovery. Generally speaking, in silico ADME tools are classified into two major categories, structure-based and ligand-based. Each class of in silico tools are addressed in subsequent sections.

4.1   Structure-Based In Silico Models

When sufficient structural information exists on the protein of interest, generally in the form of a nuclear magnetic resonance or crystallographic X-ray structure, structure-based drug design techniques are used in early-phase drug discovery. In structure-based drug design, interactions between the protein and the ligand are the focus of the study, and this is commonly referred to as rational drug design. Novel ligands can be designed de novo, meaning the interactions between a hypothetical ligand and the protein are optimized with the goal of creating a compound with high affinity and selectivity. Molecular docking can be used to orient a ligand within the active site of the protein to provide an estimate of the protein-ligand interaction. However, molecular recognition between a protein and a ligand is a complex process

that does not occur in a static structure. Molecular dynamics (MD) and Monte Carlo (MC) simulations are computational techniques used to create trajectories that model the protein-ligand fluctuations and dynamics in atomic detail [2, 3].

4.1.1   Molecular Docking

The goal of molecular docking is to model the potential interaction between a protein and a ligand [4]. Although several docking programs exist [4–11], each docking program can be broken down into two general parts: the search function used to orient and place the ligand inside the binding pocket (binding pose generation) and the scoring function used to quantify the protein-ligand interaction and predict the binding affinity (binding affinity prediction). This chapter provides an overview of the current status of molecular docking but does not go into detail on search algorithms or scoring functions, both areas of active research.

For certain protein targets, the search algorithm may generate bioactive binding poses (root-mean-square deviation <2 Å) during the search process for 90% of compounds, but this percentage can be as low as 40% for other protein systems [12]. This is especially challenging for ADME targets that are known to bind a diverse array of compounds and are promiscuous in nature. For many ADME targets, factors such as the size of the binding pocket (relatively large and hydrophobic), the water network within the active site, and protein flexibility lead to significant challenges while utilizing molecular docking. Figure 4.1 illustrates this point on one class of ADME targets, the cytochrome P450 (CYP) family of enzymes. CYPs are estimated to be involved in the metabolism of approximately 75% of drugs currently on the market with CYP3A4 known to metabolize approximately 50% of such compounds [20]. While several publications exist on CYP3A4 docking [21–26], the abovementioned problems limit its use in early-phase drug discovery programs outside of qualitative idea generation.

In instances where the docking search algorithm identifies a bioactive binding pose, current scoring functions are not accurate enough to reliably predict the binding affinity [27–29]. The correlation between the experimentally measured and predicted binding affinities for a series of compounds binding to the same protein target is usually weak and often influenced by the size of the ligand rather than the underlying physicochemical contributions to the binding affinity [30, 31]. Therefore, bioactive binding poses are not always ranked as the most energetically favorable (or top ranked) during the docking procedure [12]. In addition, the lack of accuracy and separation in binding affinity prediction makes it challenging to predict the binding affinities of compounds within a structure-activity relationship (SAR) series let alone in silico de novo-designed compounds. A recent review by Lill [32] discusses many of the current problems and challenges of molecular docking and goes into greater depth on techniques used to overcome such obstacles.
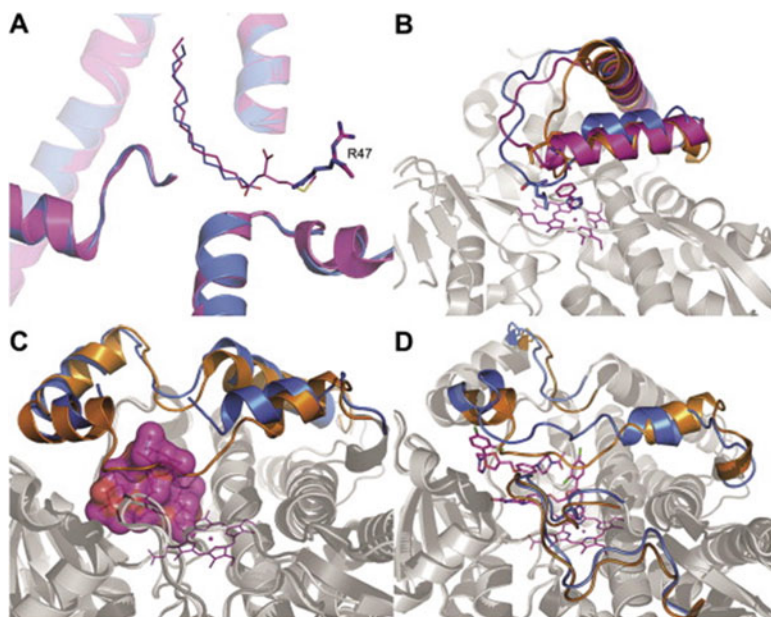
Fig. 4.1 Reproduced from Danielson et al. Potentially increasing the metabolic stability of drug candidates via computational site of metabolism prediction by CYP2C9: The utility of incorporating protein flexibility via an ensemble of structures. Eur J Med Chem 2011 Sep.;46(9):3953–63. Copyright © 2001 published by Elsevier Masson SAS. All rights reserved [13]. Examples of protein flexibility in cytochrome P450 enzymes: (**a**) Changes in Arg47 side-chain rotamer in P450 BM-3 depending on the bound ligand (palmitoleic acid and corresponding protein in *blue*, PDB-code: 1FAG [14]; N-palmitoylmethionine and corresponding protein in magenta: 1ZO9 [15]). (**b**) Alternative loop conformations are observed in CYP119 when different ligands are bound. Compared to the apo structure of CYP119 (F/G loop in orange: 1IO7 [16]), the F/G loop adapts distinct configurations when 4-phenylimidazole (ligand and loop in magenta: 1F4T [17]) or imidazole (blue: 1F4U [17]) is bound. (**c**) In CYP3A4 significant protein flexibility occurs in the F/G portion of the protein (apo: *orange*, 1TQN [18]; erythromycin bound: *blue*, 2J0D [19]) to accommodate erythromycin and part of the F–F′ loop becomes disordered. This motion causes the solvent-accessible volume of the binding site to significantly increase and can dramatically affect ligand binding. (**d**) CYP3A4 exhibits a protein breathing motion increasing the size of the binding pocket to accommodate two ketoconazole (ligands in *magenta*, protein in *blue*: 2V0M [19]) compounds without significant conformational changes of the helices or loop regions composing the binding pocket (apo: *orange*: 1TQN [18])

Post-processing is one such technique designed to overcome the problem of using simplistic scoring functions in docking and can significantly improve the successful prediction of binding affinities [33, 34]. Post-processing techniques incorporate dynamic information of the protein-ligand system after the docking process has been completed. The top-scored binding pose, or several favorably scored poses, is used as input to subsequent MD simulations. In combination with free-energy methods such as free-energy perturbation [35], thermodynamic integration [36], molecular-mechanics Poisson-Boltzmann or generalized Born surface

area [37], or linear interaction energy analysis [38], a more accurate estimation of the free energy of binding is possible [33]. However, this process is relatively time consuming and requires that the bioactive binding pose is within the top-ranked binding poses in order to limit computational time, a criterion that is not always evident when carrying out molecular docking studies on large and rather promiscuous ADME targets.

4.1.2 Molecular Dynamics

Molecular dynamics (MD) is a computational technique used to study the physical movement of atoms. The first MD simulation of a biomolecular system was done in 1977 on bovine pancreatic trypsin inhibitor using a simplistic molecular mechanics potential to describe the properties of the system [39]. Although this simulation was only performed for 9.2 ps, it was a groundbreaking study that showed that integrating Newton's equations of motion over a series of very short-time steps (usually one or two femtoseconds) could transform a once static X-ray structure into a dynamic trajectory from which time-averaged properties could be calculated. Underlying any MD simulation is a physics-based force field that defines all parameters of the system. Several force fields and MD programs exist [40–46], and the parameters are usually defined by high-level quantum chemical calculations or empirically fit to experimental properties. In addition to the force field parameters, a potential function, or mathematical relationship, is needed to describe how the individual atoms of a system interact during the MD simulation. Most force field potentials describe the interactions between atoms in the system in terms of a five-component description of intra- and intermolecular forces. The AMBER force field potential is shown in Eq. (4.1) and consists of bonded (bonds, angles, and dihedral terms) and nonbonded (van der Waals and electrostatic terms) components [42].

$$
\begin{aligned}
V(r^N) = &\sum_{\text{bonds}} K_r \left(r - r_{\text{eq}}\right)^2 && \text{bond term} \\
&+ \sum_{\text{angles}} K_\theta \left(\theta - \theta_{\text{eq}}\right)^2 && \text{angle term} \\
&+ \sum_{\text{dihedrals}} \frac{V_n}{2} (1 + \cos\left[n\phi - \gamma\right])^2 && \text{dihedral term} \\
&+ \sum_{i<j}^{\text{atoms}} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \right) && \text{van der Walls term} \\
&+ \sum_{i<j}^{\text{atoms}} \frac{q_i q_j}{\varepsilon R_{ij}} && \text{electrostatic term}
\end{aligned}
$$

$$(4.1)$$

In this type of potential, intermolecular bonds are treated as a simple Hooke's law springs with a characteristic force constant K_r and equilibrium bond length r_{eq}.

The angular term accounts for bond angle bending in the system, and the dihedral term represents the intrinsic torsional energy due to twisting about bonds. The van der Waals term accounts for the attractive London dispersion and repulsive van der Waals nonbonded forces and is calculated by a 12-6 Lennard-Jones potential. Force field assigned atomic partial charges are used to calculate the nonbonded electrostatic interaction between two atoms by solving Coulomb's law. Summing over all pairs, triplets, and quartets of atoms in the system, the force field potential provides an estimate of the energy of the system at a particular configuration. A more detailed description of MD and the algorithms associated with this technique can be found elsewhere in the literature [3, 41–43, 47–49].

Currently, MD simulations are performed on macromolecular systems comprised of thousands of atoms, and several different explicit and implicit water models exist to solvate the system [47–53]. The nanosecond time scale is routinely reached in MD simulations, and in specialized instances protein systems have even been simulated up to the millisecond time scale [54, 55]. With increasing computer power and advances in technologies and methods, millisecond time scale simulations may become routine in the near future. However, this also brings with it additional challenges such as storing, analyzing, and interpreting such a vast array of data. Despite the previously mentioned problems, MD simulations are routinely used to turn a static X-ray crystallographic structure into a dynamic system. Snapshots taken from the MD simulation provide some estimate of protein flexibility and can be used as alternative templates for molecular docking, and this technique has been utilized in several CYP isoforms [13, 56–61]. While MD simulations have become routine in the computational chemistry field, their application in early-phase drug discovery has not. This is especially true for ADME targets due to very limited number of high-resolution X-ray crystallographic structures and their promiscuous nature. Additionally, the time and resource intensive nature of MD simulations and the rather fast-paced movement of chemistry SAR on project teams further limit the application of MD simulations during this phase.

4.2 Ligand-Based In Silico Models and Tools

4.2.1 Quantitative Structure-Property Relationship (QSPR) Models

Quantitative structure-activity relationship (QSAR) models are one of the commonly employed ligand-based techniques to predict the activity of compounds. The field of modern QSAR can be traced back more than 50 years to a model produced by Hansch [62]. QSAR sophistication has grown from its early application on a small congeneric series of compounds using simple linear regression to now being applied to data sets comprised of thousands of diverse compounds utilizing a wide variety of statistical and machine learning algorithms.

When such models are used to predict various properties, including ADME endpoints, they are referred to as quantitative structure-property relationship (QSPR) models. Given the promiscuity and limited structural knowledge of ADME targets, QSPR models are commonly used in the pharmaceutical industry to address ADME-related challenges. The basic premise of QSPR methodology is to develop a relationship between an observed property and structural features of a compound. Considering a set of compounds with observed experimental data (training set), a model is developed that can be used to predict the activity of other compounds (test set) not included in the initial training set. Compounds are represented using a variety of molecular descriptors that describe the chemical structure and properties of the compound. A relationship between the molecular descriptors and the observed response is computed using mathematical techniques such as linear regression, artificial neural network, support vector machine (SVM), and random forest (RF). A general description of such algorithms is summarized in Sect. 4.2.1.4. Figure 4.2 illustrates the general process of building and applying QSPR models to a group of compounds, and each step of the process is further explained below.

Generalized equation representing a typical QSPR model

$$P = f(X)$$

P = Value of property being modeled
- Example: Solubility, microsomal clearance, LC50 for hepatocytes

f = statistical function / method
- Linear- Multiple Linear Regression, Linear Discriminant analysis
- Non-Linear/Machine learning - Recursive partitioning (random forest), Support vector machine (SVM)

X = molecular descriptors
- Examples: MW, no. of hetero atoms, charge, no. of valence electrons on certain atoms, surface area, distance between certain features, hydrophobic surface area etc.
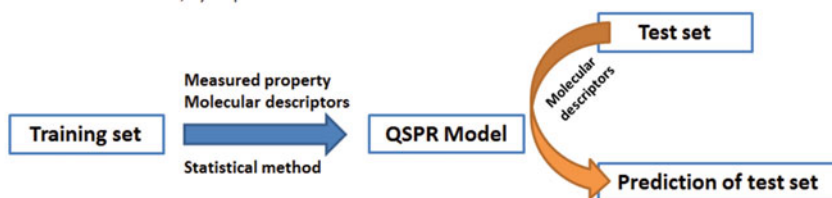
Training set → (Measured property, Molecular descriptors, Statistical method) → QSPR Model → (Molecular descriptors) → Test set / Prediction of test set

Fig. 4.2 Schematic representation of key components when building and applying QSPR models. The *top section* shows the generalized equation representing a typical QSPR model and lists key components required to derive such an equation for a given data set. The *bottom section* depicts a typical workflow used to build and use a QSPR model

4.2.1.1   Data Set Selection and Curation

The first step to create any QSPR model is the selection of the data set that the model will be built upon. A key consideration when choosing any data set to create a model upon is that the data should be accurate, reliable, reproducible, and measured using identical experimental conditions for all compounds. This can be a significant challenge when building QSPR models based on public databases compiled by collating data from multiple labs spanning a variety of experimental protocols. Stouch et al. demonstrated that models based on data sourced from multiple labs showed poor predictive capabilities for compounds tested in a rigorous and consistent manner [63]. For example, in the case of a hERG inhibition model provided by an external vendor, the data were collated from several different laboratories using a variety of assay conditions: different cell types expressing the hERG channel and different activation potentials for the channel, along with combining binding and inhibition data. The predictions from the vendor model had a poor correlation coefficient of 0.01 and a high root-mean-square error (RMSE) of 1.3 log units for the test set evaluated by the authors.

Following the selection of data, the importance of data curation cannot be overemphasized. In order to create the best possible QSPR model, it is critical to minimize the inclusion of potentially erroneous data. The potential sources of erroneous data include false positives/false negatives, under-/overestimated responses, spurious results (e.g., microsomal stability >100%), incorrect structural representation of compounds, data below the analytical detection limits, and impure material. For example, while building a classification model for P-glycoprotein (P-gp) efflux, Desai et al. excluded compounds reported as non-substrates displaying >60% inhibition of a fluorescent P-gp substrate, very slow passive permeability, and very low cell partitioning (all cases suggesting potential false negatives) in addition to compounds with poor mass recovery (potentially spurious data) [64]. When feasible, it is good practice to find and utilize analytical data related to identity and purity of compounds. Such information is commonly available in an industrial setting but not easily found for data compiled from multiple sources and available in public databases like ChEMBL. In a previous study, several public and commercial databases were investigated, and error rates in chemical structure annotation ranged from 0.1% to 3.4% [65].

In order to properly curate the assay data that will be used to build a model, it is critical to understand the experimental protocol and potential caveats associated with that given measurement. One of the common issues leading to potentially erroneous results is poor solubility of the compound in the medium used for the assay (e.g., none or very little of the compound is in solution giving an incorrect assay value). This can potentially be addressed by running a parallel experiment to measure the solubility of the compound in the buffer used for the ADME assay. For example, at Eli Lilly and Company, aqueous kinetic solubility in pH 7.4 phosphate buffer is measured for all compounds tested in high-throughput ADME assays. This information is used to curate the data for various ADME endpoints wherein compounds that are not in solution at the concentration used for the given ADME

assay are not included in the QSPR model. To summarize this section, while it is often an overlooked and underappreciated step, data curation based on detailed understanding of the experimental measurement is a critical step in building high-quality QSPR models.
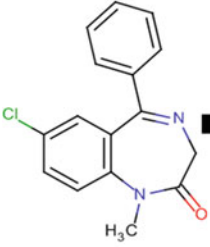
4.2.1.2   Training Set Selection

Following data curation, the next logical step of creating a QSPR model is selecting compounds to construct and train the model. What size or how many compounds needed to be in the training set is a precarious question that is sometimes asked. No easy answer to the question exists, and the size of the training set needed to build a useful model depends on the complexity of the endpoint and the intended use of the model. For example, for models intended to be applied prospectively to compounds spanning a wide range of structural diversity, the training set should reflect similar structural diversity and perhaps as much diversity as possible. Prospective model performance, meaning how well the model predicts compounds not in the training set, also depends on whether the training set encompasses the entire range of the assay response. For models such as microsomal metabolic stability that are based on a continuous response (assay range from 0% to 100%), the ideal situation is to have a training set containing compounds spanning the entire 0–100% range and uniformly distributed if possible. For categorical response such as low or high, an even or close to even distribution of compounds between the categories is desired.

   Models constructed with training sets that span a narrow spectrum of the entire assay response (e.g., a training set containing 95% of compounds that have microsomal metabolic stability of >90% when the assay range spans 0–100%) or with a highly skewed distribution of the categorical response (e.g., 95% of compounds in the training set belong to the "high" class) are likely to result in QSPR models with limited utility when used prospectively.

4.2.1.3   Molecular Descriptors

Following data curation and training set selection, molecular descriptors must be calculated in order to derive the mathematical relationship between chemical structure and assay activity. Molecular descriptors are numerical parameters derived from chemical structures, and a wide variety of descriptors are used to build QSPR models. Physicochemical (e.g., log P, pK_a, MW), topological (e.g., atom connectivity), constitutional (e.g., number of nitrogen), and quantum chemical (dipole moment, atomic charges) are few examples of common types of descriptors. To gain a deeper understanding and comprehension of molecular descriptors, the reader is referred to a publication by Todeschini and Consonni [66].

   In addition to molecular descriptors, molecular fingerprints are often used to represent chemical structures [67, 68]. A molecular fingerprint is comprised of a series of substructures, and the presence/absence of such substructures determines

Fig. 4.3 Snippet of MACCS fingerprint of diazepam

the numerical code for the molecular fingerprint [69–71]. For example, Molecular Access System (MACCS) fingerprint uses a set of structural features to code the compound into a binary representation [72]. Figure 4.3 shows an example snippet of the MACCS fingerprint representation for the drug diazepam. The column titled "key positions" in the figure assigns a number to a particular chemical feature, listed under "fragment description." The "fingerprint code" is a binary value associated to the absence (assigned zero)/presence (assigned one) of the chemical feature. Using the "key positions" and "fingerprint code," one can derive the final fingerprint shown in Fig. 4.3. Only "fingerprint codes" that are present in the compound are kept in order to keep the fingerprint code vector sparse.

Typically, when constructing a QSPR model, a large collection of molecular descriptors and a variety of fingerprints are calculated. The descriptors and fingerprints are subsequently evaluated using statistical approaches to select the optimal combination to relate chemical structure to the activity of the endpoint. When constructing a model for the first time, several versions of the QSPR model may be built using various combinations of descriptors or fingerprints followed by several iterations of prospective model evaluation (Sect. 4.2.1.5) to identify the optimal collection of descriptors or single best fingerprint [73].

4.2.1.4   QSPR Model Training/Building

After data curation, training set preparation, and descriptor/fingerprint selection, the QSPR model is ready to be built. Mathematic algorithms such as linear regression, artificial neural network, SVM, and RF are routinely used to train and build QSPR models [74]. Linear regression (for continuous response) or discriminant (for categorical response) models assume that the measured property value is an additive response to the underlying molecular descriptors. For example, in the QSPR model for solubility shown in Eq. (4.2) [75], it is assumed that solubility is linearly dependent on lipophilicity (log P) and topological polar surface area (TPSA).

$$\log S = -1.0377 \log P - 0.0210 \text{TPSA} + 0.4488 \qquad (4.2)$$

Besides prediction, linear models may provide mechanistic insight and can be interpretable in nature as long as the molecular descriptors are "simple" and intuitive. Thus, in case of the solubility model in Eq. (4.2), the negative coefficient for log P suggests that an increase in the lipophilicity of compounds is expected to decrease solubility.

Given the complexity of most ADME-related responses, linear models appear to only be applicable over a relatively narrow spectrum of compounds that contain conserved structural motifs. In practice, such models are rarely useful prospectively due to their inability to extrapolate and predict compounds outside their immediate domain of applicability. Machine learning methods such as RF [76, 77] and SVM [78, 79] have been applied to QSPR models to combat the abovementioned limitations and are capable of elucidating more complex relationships between structural descriptors and the observed response.

In general terms, RF models are based on several iterations of the recursive partition approach, and SVM models identify a hyperplane in the high-dimension descriptor space to enable maximum separation of observed responses. Within the pharmaceutical industry, a large amount of ADME data are generated in a consistent manner, and therefore such machine learning methods are preferred to build "global" QSPR models that are designed to be applicable across multiple drug discovery projects that cover a broad spectrum of chemical space [80]. In our experience, such models typically outperform linear QSPR models in extracting structure-property relationship knowledge from large sets of diverse compounds. However, given the complexity of RF and SVM models, they are relatively less interpretable compared to linear models and often offer limited mechanistic insight to go along with predictions. Although generally less interpretable, it should be noted that it is possible to get an estimation of the most influential descriptors for RF models, in turn providing some understanding of key molecular characteristics influencing a given endpoint. For example, in case of an RF model for P-gp efflux, Desai et al. identified that molecular features related to the number of hydrogen bond donors (HBD), TPSA, and hydrogen bond strength were most influential in terms of P-gp efflux of compounds [64].

4.2.1.5   QSPR Model Evaluation

The performance of a QSPR model is evaluated using a variety of parameters depending on the type (continuous vs. categorical) and the intended use of the model. Performance parameters are typically calculated at three stages of the model building process. For example, after building a continuous response model, the first stage is to assess the ability of the model to fit the training set compounds. This metric is commonly referred to as r^2 in the QSAR/QSPR literature. The second stage evaluates the ability of the model to predict the set of compounds left out of

the model building process in an iterative manner (called cross-validation, leave-one-out, or leave-some-out) is referred to as q^2. The third stage is known as external or prospective validation, and the model's ability to predict compounds that were not used during any stages of the model building process is evaluated.

The ability of the model to fit the training set simply serves as a feasibility assessment. It does not provide an assessment of the model's ability to predict compounds outside the training set and therefore isn't particularly useful [81]. Cross-validation is based on prediction of compounds left out of the model but is still an internal validation as it derives the test set from the existing pool of compounds. Depending on the modeling method employed, the cross-validation test set can bias the choice of descriptors and other model-related parameters [82]. Many experts in the QSAR community believe that this type of validation often overestimates a model's ability to predict a true external or prospective test set. Therefore, in order to comprehensively evaluate the utility of a QSPR model, it is critical to assess its predictive ability against an external prospective test set [64, 83–85].

For QSPR models based on continuous data, the square of the correlation coefficient (r^2) between the observed and predicted value (referred to as q^2 when used in the context of cross-validation) is the most common performance parameter reported. RMSE between the observed and predicted values is another key parameter used to assess continuous response model performance. Higher values of r^2 (maximum 1 for a perfect model) and smaller values of RMSE are desirable [86]. In many cases, Spearman's rank correlation coefficient (ρ) is also reported as an indicator of model performance [87]. Depending on the intended use of the QSPR model, one or more of these parameters may be utilized to determine how well a particular model is preforming. For example, if the goal is to identify a model wherein predictions are *correlated* with the observations (not necessarily to predict the absolute value of the property), the r^2 of a prospective test set would serve as a useful parameter. On the other hand, to simply rank order the prospective compounds, a model with high ρ value would be sufficient. If the goal is to accurately predict the absolute value of the property, a model with low RMSE would be necessary. The ideal QSPR model would have favorable performance values for all of the abovementioned metrics.

Classification QSPR models have a different set of performance metrics compared to regression models. Commonly reported performance parameters for classification models are based on the fraction/percent of correct predictions (overall accuracy), the accuracy of each experimental class (sensitivity and specificity), and the accuracy of each predicted classes (PPV and NPV). Table 4.1 provides details to calculate the abovementioned parameters and is referred to as a contingency table or confusion matrix. In addition to these widely used metrics, parameters such as the kappa index are often reported to assess the agreement between prediction and the experimentally determined category. A kappa value of 1 indicates perfect agreement between predictions and experimental values, −1 suggests complete disagreement, and 0 indicates the prediction is no better than random chance. In general, a kappa value >0.4 is considered an indicator of reasonable model performance with useful predictive power [88, 89].

Table 4.1 Contingency table with equations for a classification QSPR model

Predictions	Experiments			
	Positive	Negative		
Positive	TP	FP	Positive predictive value (PPV)	TP/(TP + FP)
Negative	FN	TN	Negative predictive value (NPV)	TN/(TN + FN)
	Sensitivity	Specificity	Overall accuracy = (TP + TN)/N	
	TP/(TP + FN)	TN/(FP + TN)		

TP, TN = true positive, true negative; FP, FN = false positive, false negative; N = total number of compounds

4.2.1.6   Interpretation of Model Prediction

In addition to the abovementioned parameters for model evaluation, several other factors should be considered when assessing the utility and/or applying a QSPR model to a given drug discovery project. In the case of a continuous response model, an applicability domain-related parameter should also be considered in addition to the predicted value if available. Meaning a parameter that indicates if the QSPR model can, or should, predict a compound of interest based on what the model was trained on. If the compound of interest is vastly different than all compounds in the training set, it is expected that such an applicability domain parameter would be unfavorable. Several methods to estimate the applicability domain for a QSPR model have been described in the literature, and they generally provide a qualitative indicator of the confidence for each prediction or a quantitative estimation of the confidence interval around the predicted value [90–93].

In addition to the standard contingency table metrics commonly reported (see Table 4.1), if one is evaluating a classification QSPR model built with a machine learning method (e.g., RF or SVM), the predicted scores of each compound give an estimation of the relative confidence or reliability of prediction [64, 77, 94]. For example, for two compounds predicted to be in the same category, the compound associated with higher score is assumed to be a more reliable prediction compared to the other.

In addition to the abovementioned numerical parameters reported to determine QSPR model applicability/reliability, in order to conduct a thorough assessment of the utility of a model for a given chemical scaffold or drug discovery project, one should always consider:

- The inherent experimental variability in the measurement, especially in case of the high-throughput ADME assays. Model performance has been shown to be directly related to the inherent variability in the measurement of the given assay parameter [95]. For regression QSPR models built on continuous data, one should evaluate the performance of the model based on the proportion of predicted values that falls within the experimental variability of the measured

response and not just rely on an r^2 value. For example, if the inherent variability of an assay is threefold, a model built on these data should be evaluated with this variability in mind. One should check the proportion of the prospective test set that are predicted within threefold of the experimental values. A regression model may not have an r^2 value of 0.9 for this model, but if 90% of the predicted compounds are within threefold of the experimental values, then that model will still be useful.

- Due to the variability in ADME high-throughput assays, we build and advise the use of categorical QSPR models for such data.
- The QSPR model should be evaluated on a prospective test set that spans the entire spectrum of the response, or in the case of a categorical model, the test set should have a balanced distribution of compounds from each category or one that mirrors the training set distribution.
- The assessment of a QSPR model should not be based on a small fraction of compounds, only the most recent compounds, or only the potent compounds from a given chemical scaffold or drug discovery project.
- A QSPR model should not be evaluated based on its performance against a second experimental endpoint not directly predicted by the model. For example, comparing predictions from a QSPR model built on in vitro microsomal metabolic stability data against an in vivo clearance outcome should not be done without establishing if this is permissible. The compound and scaffold of interest may be cleared by mechanisms other than microsomal metabolism, and an in silico microsomal clearance QSPR model should not be expected to accurately predict the in vivo clearance value for such cases.

4.2.2 ADME QSPR Models Used at Eli Lilly and Company

Over the past couple of decades, many publications pertaining to the application of QSPR models for ADME-related physicochemical properties and in vitro/in vivo endpoints have been published. In an attempt at brevity, the reader is referred to review articles that summarize this area of research [96–98]. Table 4.2 provides a brief summary of ADME QSPR models developed and used at Eli Lilly and Company. The data set for each individual model was generated by/for Eli Lilly and Company using consistent experimental conditions for each individual ADME in vitro or in vivo assay. Total data set size ranges from 2,000 to 80,000 depending on the throughput of the particular assay. All ADME QSPR models are built using an SVM algorithm with an optimum molecular fingerprint selected for each assay endpoint.

Table 4.2 Representative list of ADME QSPR models used at Eli Lilly and Company

Endpoint	Data source	Training set size (in thousands)	Model type: Classification (C) Continuous (R)
Kinetic aqueous solubility	Solubility of DMSO stock diluted at various concentrations in phosphate buffer (pH 7.4)	80	C
High-throughput solubility	Solubility of DMSO-dried sample in buffers at pH 2, 6, and 7.4	30–32	C
Passive permeability	Passive permeability across MDCK cells	15	C and R
Hepatic microsomal stability (human/ mouse/rat/dog/ monkey)	Stability in hepatic microsomes	20–80	C and R
Cytochrome P450 competitive inhibition (CYP3A4/CYP2D6/ CYP2C9)	% inhibition of CYPs at 10 μM	65	C
CYP3A4 time-dependent inhibition	Time-dependent inhibition of CYP3A4 at 10 μM	10	C
P-glycoprotein substrate recognition	Efflux by human P-glycoprotein overexpressed in MDCK cells	4	C
In vivo mouse brain unbound concentration	Unbound concentration of compound 5 min post-IV dose	2	R
In vivo mouse brain: plasma partition coefficient of unbound compound (Kpuu)	brain:plasma partition coefficient of unbound drug (Kpuu) in mouse 5 min post-IV dose	2	C and R
Fraction unbound (plasma, brain, microsomes)	Equilibrium dialysis at 1 μM incubation	6–8	R

4.2.3   Prospective Validation of ADME QSPR Models at Eli Lilly and Company

In an industrial drug discovery paradigm where new pharmacological targets are constantly explored, it is important to update global QSPR models to ensure their applicability and prospective prediction performance. Figure 4.4 highlights the outcome of this chronological process at Eli Lilly and Company where prospective performance of ADME QSPR models was maintained for several classification models used over the past several years.

As drug discovery project teams synthesize and test new compounds in various ADME in vitro assays, the global models are updated by curating and adding the new data to their respective training sets. Before updating any particular model, the existing model is prospectively evaluated to measure its predictive performance
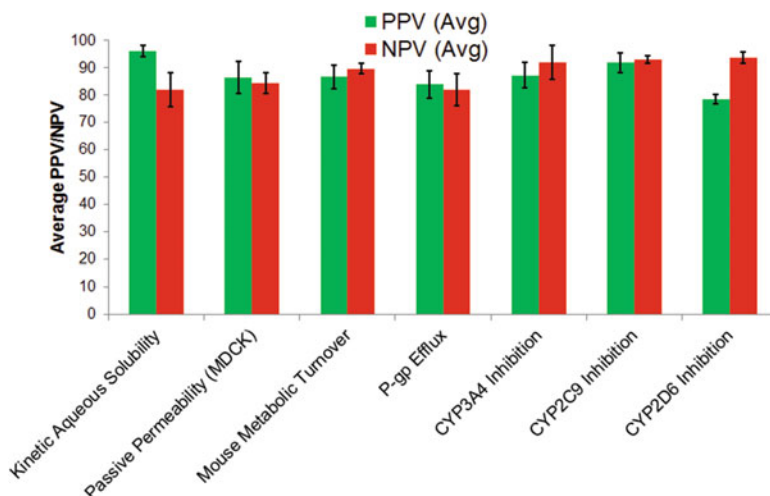
Fig. 4.4 Prospective validation of ADME QSPR classification models used at Eli Lilly and Company. Average PPV and NPV over the last 8–10 versions are shown. Error bars represent the standard deviation. All models were applicable for ~80% of prospective test sets when score cutoffs were used to "accept" a prediction

against data generated *after* the model was built. The result of this assessment for a set of seven Eli Lilly and Company ADME models is shown in Fig. 4.4. The training set for these models range from ~4,000 to 75,000 and increases in number with every model update cycle. Focusing on the mouse metabolic turnover model, the oldest version of the QSPR model in Fig. 4.4 was built using ~40,000 compounds. Before updating the model, it was prospectively evaluated against an additional ~4,000 compounds, and after showing suitable performance, the new data were added to the training set of the existing model to build the next version containing ~44,000 compounds.

All models in Fig. 4.4 are SVM models using fingerprints as descriptors and provide categorical predictions, along with a score representing the reliability of such a prediction. As explained in Sect. 4.2.1.6, predictions associated with higher scores are expected to have greater likelihood of aligning with the measured response. Based on the prospective validation results, suitable score cutoffs (typically 0.7 on a scale of 0–1.0 for both prediction categories) are assigned to "accept" a given prediction, while predictions with scores below the cutoffs for a given category are labeled as "indeterminate." The PPVs/NPVs shown in Fig. 4.4 are calculated for compounds with "acceptable" scores. For all models listed in Fig. 4.4, >80% of the test set compounds had "acceptable" scores, and thus the models were applicable for >80% of the test sets. As shown in Fig. 4.4, the average PPV/NPV for the ADME models ranged from 75% to 85% in prospective testing. Given such consistent prospective performance, the ADME QSPR models are routinely used to design and prioritize compounds for synthesis and testing during early-stage drug discovery. The performance of various versions of the global P-gp

efflux model and its application in identifying and addressing challenges related to central nervous system (CNS) drug discovery projects is described in detail by Desai et al. [64].

4.2.4 Trends Between Calculated Physicochemical Properties and ADME Parameters

To complement the usefulness of ADME QSPR models, the physicochemical properties of compounds influencing ADME properties is well documented. One of the earliest analysis of ADME properties was performed by Lipinski leading to the "rule of five" suggesting that poor absorption and permeability are more likely if the molecular weight (MW) is >500, the number of NH and OH hydrogen bond donors is >5, the calculated log P (i.e., $c$log P) is >5, and the number of N and O atoms is >10 [99]. The goal of this guideline was not necessarily to rule out certain synthetic ideas but rather steer the synthetic chemistry effort toward chemical space that is more likely to yield compounds with superior ADME properties. Subsequently, several analyses describing the trends between calculated physicochemical properties and in vitro/in vivo ADME parameters have been reported [100–103]. In an exhaustive analysis of a large and structurally diverse set of preclinical compounds profiled at GlaxoSmithKline, Gleeson reported relationships between several ADME assays and calculated physicochemical descriptors [100]. This included in vitro ADME endpoints like solubility, permeability, rat brain tissue and plasma protein binding, P-gp efflux, and inhibition of the CYP isozymes. Several in vivo ADME parameters like oral bioavailability, clearance, volume of distribution, and CNS penetration in the rat were also analyzed. Some of the calculated physicochemical descriptors used in this analyses were $c$log P, $c$log D, the number of hydrogen bond acceptors (HBA) and donors (HBD) (typically counted as number of N + O for HBA and NH + OH for HBD), positive and negative ionization states, molecular flexibility, molar refractivity, MW, TPSA [104], and the number of rotatable bonds. From this descriptor list, ionization state, $c$log P, and MW were identified as the most influential physicochemical properties for ADME properties. The paper suggested that compounds with a MW of <400 and a $c$log P of <4 were preferred with regard to maintaining a favorable ADME profile. In another report by Varma et al. [102], ionization state, lipophilicity, and polar descriptors were found to be the physicochemical determinants of renal clearance in human based on a compiled data set of ~400 marketed drugs. It is important to keep in mind that the conclusions about correlations between physicochemical and ADME properties can be strongly influenced by the size and nature of the database employed. Moreover, many of the physicochemical parameters are not independent of each other. For example, an increase in MW is likely to be associated with increase in the number of heteroatoms like N and O, which in turn are associated with TPSA.
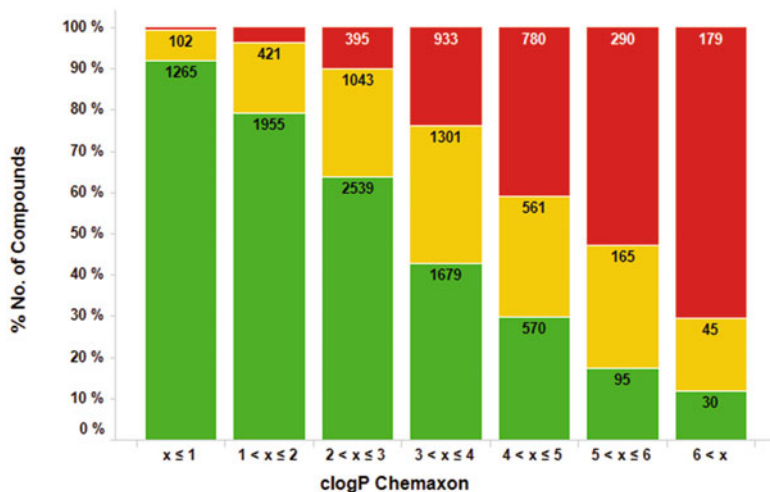
Fig. 4.5 Experimental rat microsomal $Cl_{int,u}$ vs $c\log P$. *Green* = slow, *yellow* = moderate, *red* = rapid $Cl_{int,u}$. Global data analysis suggests compounds with $c\log P$ of <4 are less likely to have rapid $Cl_{int,u}$
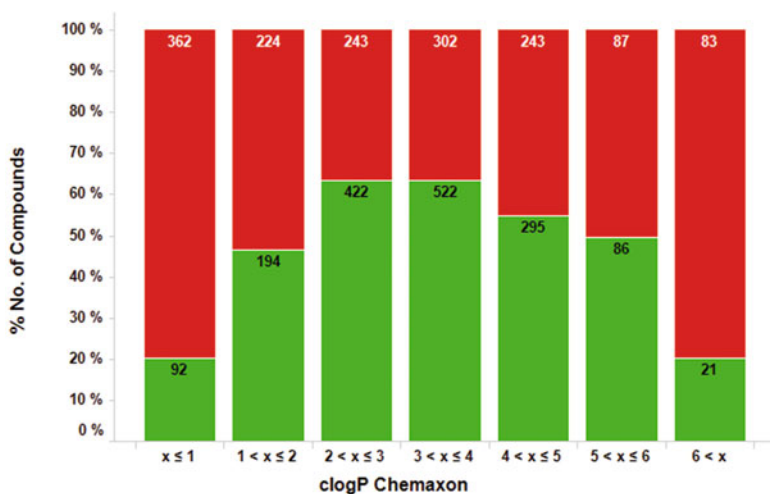

Fig. 4.6 Experimental MDCK permeability vs $c\log P$. *Green* = rapid permeability, *red* = slow. Global data analysis suggests that compounds with $c\log P$ between 2 and 4 are more likely to have rapid permeability

Figures 4.5, 4.6, 4.7, and 4.8 along with summary Table 4.3 detail Eli Lilly and Company's ADME in vitro data in relation to key physicochemical properties over the past 2 years. Figure 4.5 shows the trend that as $c\log P$ increases so does microsomal unbound intrinsic clearance ($Cl_{int,u}$) [105]. This analysis indicates
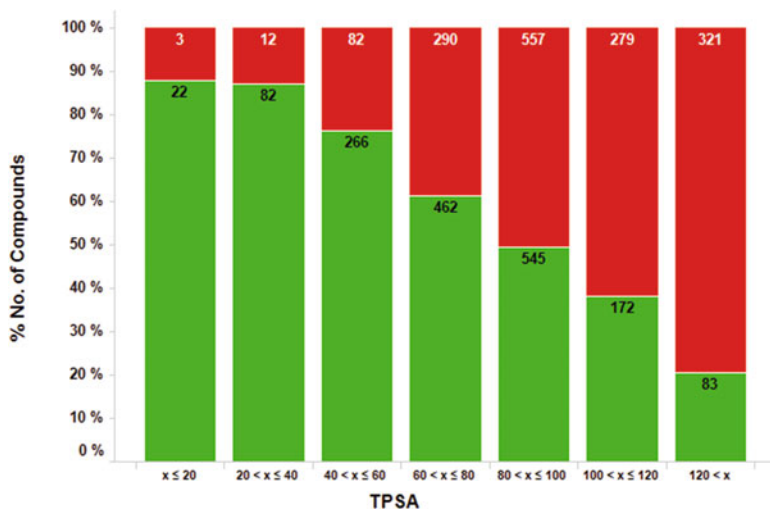
Fig. 4.7 Experimental MDCK permeability vs TPSA. *Green* = rapid permeability, *red* = slow. Global data analysis suggests compounds with TPSA of <100 Å² are more likely to have rapid permeability
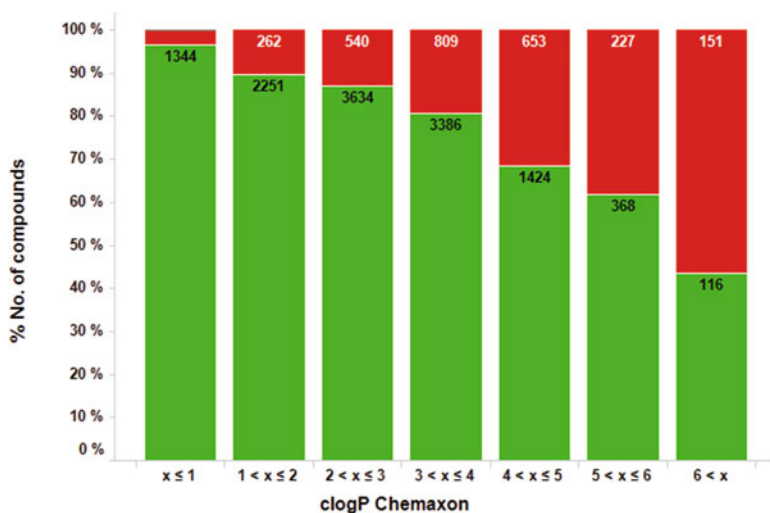

Fig. 4.8 CYP3A4 inhibition vs $c\log P$. *Green* = low inhibition, *red* = high inhibition. Global data analysis suggests compounds with $c\log P$ of <4 are more likely to have low inhibition of CYP3A4

that compounds with $c\log P$ value <4 are more likely to have slow unbound intrinsic clearance (Fig. 4.5) and a low CYP3A4 inhibition potential (Fig. 4.8). Similarly, compounds with $c\log P$ between 2 and 4 (Fig. 4.6) and TPSA <100 Å² (Fig. 4.7) are more likely to have rapid permeability across MDCK cells. Desai

Table 4.3 Trends between calculated physicochemical properties and ADME endpoints from Eli Lilly and Company data

Physicochemical parameter	Desirable range	Trends with ADMET properties
$c\log D$ at pH = 7.4 (Chemaxon)	<3	Higher kinetic aqueous solubility, slow microsomal metabolism and unbound intrinsic clearance ($Cl_{int,u}$), low CYP inhibition potential, high unbound CNS exposure
	1–3	Rapid passive permeability
$c\log P$ (Chemaxon)	<4	Slow microsomal metabolism and unbound intrinsic clearance (Clint,u), low CYP inhibition potential, high unbound CNS exposure
	2–4	Rapid passive permeability
Most basic pK_a	<7.4	Lower risk of P-gp efflux
Molecular weight	<400 Da	Higher unbound CNS exposure
TPSA	60–90 Å^2	Rapid passive permeability, lower risk of P-gp efflux, higher unbound CNS exposure
Number of NH + OH groups	<3	Rapid passive permeability, lower risk of P-gp efflux, higher unbound CNS exposure
Number of N + O atoms	<8	Rapid passive permeability, lower risk of P-gp efflux, higher unbound CNS exposure
Number of negatively charged atoms	0	Higher unbound CNS exposure

et al. have previously published physicochemical trends for efflux by the P-gp transporter and reported having the most basic $pK_a < 8.0$ and TPSA <60 Å^2 as key physicochemical properties of P-gp non-substrates [64].

4.2.5 Pharmacophore Modeling

Another ligand-based modeling technique that is used in drug discovery is pharmacophore modeling. The word pharmacophore has several definitions associated with it despite the concept being around for over 40 years. A medicinal chemist may define a pharmacophore as a structural fragment or functional group related to a chemical compound or series of compounds. Computational chemists often define a pharmacophore as a collection of hydrogen bond acceptors, hydrogen bond donors, aromatic rings, charged atoms, and hydrophobic regions of compounds that provide affinity and specificity to a particular target. The official IUPAC definition states, "A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response" [106].

No matter the definition, the concept of pharmacophore modeling is simple and even intuitive to medicinal chemists working in early drug discovery. The

technique models the interaction between a ligand and a binding site, thereby producing a model of the spatial arrangement of molecular features essential for biological activity. The central premise of a pharmacophore model states that if a compound contains the needed molecular features in a spatial orientation that matches the model, the compound should bind to the target of interest. Pharmacophore models have been created for several ADME targets along with being used to predict activity, selectivity, toxicity, and enrichment in high-throughput screening experiments [20, 74, 107–110].

The scope of this chapter provides an overview of pharmacophore modeling and will only briefly introduce the two general parts of any pharmacophore modeling program. However, extensive literature has been published that describes pharmacophore models in greater detail [111–113]. In general, pharmacophore modeling can be broken down into two general steps: (1) molecular super positioning of ligands and (2) scoring how well a ligand matches the pharmacophore features.

The molecular super positioning (also known as alignment) of ligands is time consuming and represents a significant challenge to creating any pharmacophore model. This step inherently involves the alignment of flexible compounds that have multiple possible conformations. Precomputing ligand conformers is common in many of the pharmacophore program available today [111–113]. When conformers are pre-generated, pattern-matching techniques are then used to create the ligand alignment. Many pharmacophore programs use a rigid-body alignment technique that is some type of a maximum common substructure search [114] implemented with the Bron-Kerbosh clique detection algorithm [115] that accounts for the spatial arrangement of pharmacophore features. Scoring functions differ between software, but they generally account for things such as number of matching pharmacophore points along with the spatial orientation and the internal energy of the matching ligand conformer along with some sort of volume or binding site matching term. Throughout the pharmacophore building process, several parameters must be set and optimized, thereby complicating the process of creating an optimal pharmacophore model or one that the entire community uses or accepts for that matter. The reference ligand, or set of ligands, used to create the pharmacophore alignment is often subjective and requires the skill and knowledge of a computational expert.

However, it can be especially challenging to create useful pharmacophore models for targets that are known to be flexible and promiscuous in binding many compounds. Most ADME targets fall into this class, but there is no lack of pharmacophore models published for such targets [107, 109, 116–118]. For example, pharmacophore models have been published for several CYP enzymes, including CYP3A4, that are known to be extremely flexible and recognize diverse compounds. Figure 4.9 displays a pharmacophore model for the organic anion-transporting polypeptide 1B1 (OATP1B1), a liver-specific uptake transporter that lacks high-resolution structural information.

While many pharmacophore publications exist, in many instances pharmacophore models are created using a small subset of compounds known to bind to such targets (10–15 compounds maximum). Such models may perform well on very similar compounds (meaning if the alignment was done with a statin
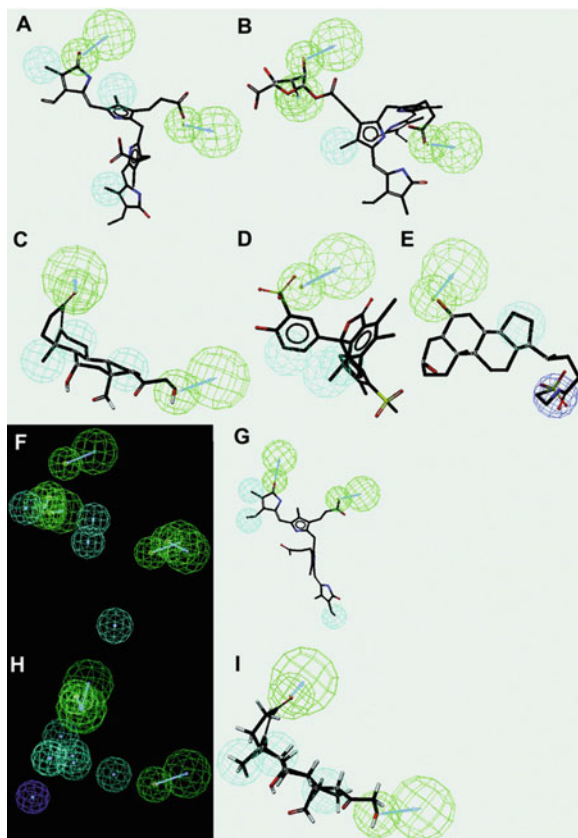
Fig. 4.9 Reproduced from Ekins et al. Comparative pharmacophore modeling of organic anion-transporting polypeptides: a meta-analysis of rat Oatp1a1 and human OATP1B1. J Pharmacol Exp Therap 2005, 314(2):533–541 [116]. Pharmacophores generated from substrate data for human OATP1B1 expressed in oocytes (showing bilirubin mapped to features) (**a**), human embryonic kidney cells (showing bilirubin monoglucuronide mapped to features) (**b**), rat Oatp1a1 expressed in oocytes (showing aldosterone mapped to features) (**c**), CHO cells (showing BSP mapped to features) (**d**), HeLa cells (showing taurohyodeoxycholate mapped to features) (**e**), merged OATP1B1 model using pharmacophores described in **a** and **b** (**f**), meta-analysis model using all cell type compound data for human OATP1B1 (showing bilirubin mapped to features) (**g**), and merged Oatp1a1 model using pharmacophores described in **c**, **d**, and **e** (**h**), showing aldosterone mapped to features (**i**). Pharmacophore features include hydrophobes (*cyan*), negative ionizable (*blue*), and hydrogen bond acceptors (*green*)

compound, the pharmacophore model more than likely will predict other statin-like compounds as likely to interact with the target), but they are not particularly useful in a drug discovery setting where diverse chemistry is being explored on many projects.

The other extreme also is problematic for ADME targets, meaning creating a pharmacophore model based on hundreds of compounds. This is due to the fact that

generating a "unique" pharmacophore pattern for ligand binding is extremely challenging given the diversity of compounds. More often than not, the number of unique matching pharmacophores for several hundred diverse structures will be very few and limited. For example, a pharmacophore model constructed on 500 OATP1B1 inhibitors may only have three pharmacophore points that match the majority of the 500 compounds. When this occurs, the pharmacophore model is not useful as it is incapable of differentiating between active and inactive compounds in the data set. In order for any pharmacophore model to be useful, it has to be shown to not only differentiate active vs inactive compounds but additionally it must have predictive power that informs the design of de novo compounds. This validation criterion is not examined in many published ADME pharmacophore models, and it is essential to evaluate before making the claim that a useful model has been created.

4.2.6 Site of Metabolism Prediction

Understanding and modulating drug metabolism is one of the fundamental concepts of ADME. Several computational techniques exist to predict the site of metabolism (SOM) on compounds. It should be noted that publications and research on SOM prediction exist for metabolizing enzymes other than CYPs [119–122]. However, due to their significance in metabolizing compounds, SOM predictions by CYP enzymes dominate the published literature and will be the focus of this section.

Prior studies predicting SOM of compounds interacting with CYPs have utilized a variety of computational methods such as quantum chemical calculations, pharmacophore models, QSAR, molecular docking, MD simulations, and basic empirical/chemical rules [13, 121, 123–138]. Recent reviews published on CYP SOM prediction provide a good summary of prior studies and techniques used [139, 140]. Although previous studies have been performed to predict SOM, there is no consensus about which method performs "best." In general, the top performing methods claim to accurately predict the experimental SOM 80% of the time or greater.

Recent thinking suggests that the SOM of a compound is influenced by two factors: (1) the intrinsic reactivity of each site in the compound to oxidation and (2) the accessibility of individual atoms to the CYP heme group, the site where oxidation occurs in the enzyme. The intrinsic reactivity is normally estimated using Hartree-Fock, semiempirical methods such as the Austin Model 1, or density functional theory quantum mechanical calculations of the chemical reaction. Accessibility to the CYP heme group is routinely estimated with solvent-accessible surface area calculations, molecular docking, and other structural features.

Several commercial SOM prediction programs exist that allow users to profile compounds to overcome metabolic liabilities. While this may be possible, caution should be used when proposing such a strategy using SOM tools in isolation. In a publication by Vaz et al. [141], they address problems associated with the metabolic

"blocking" strategy. Metabolic "blocking" occurs when a halogen atom, typically a fluorine atom, is attached to the atom/region of the compound susceptible to metabolism in order to reduce the metabolic turnover. Despite literature examples where this strategy was shown to be successful, the general strategy of "blocking" typically shifts the SOM to another atom or region of the compound due to the promiscuous nature of CYPs. In many instances, halogenating a site, typically an aromatic ring, makes the compound more lipophilic. This ultimately can lead to no change, or even increase, in affinity for CYPs and thus expose other sites on the compound to oxidation. In addition, the more lipophilic compound could potentially fit the CYP pocket better and hence become potential CYP inhibitors. By possibly fixing one ADME problem (metabolism) by introducing additional lipophilicity through "blocking," another problem may also arise in the form of solubility limitations.

When trying to mediate metabolic ADME problems, we suggest that multiple in silico tools and methods are used to provide a balanced ADME profile of a compound. In addition to SOM prediction software, in silico models of unbound intrinsic clearance, metabolic stability, log P, and solubility should be monitored with any proposed structural change to mediate a metabolic liability. Besides altering the reactivity of a particular site, we suggest evaluating options to reduce the affinity of a compound for CYPs as well. A reduction in log P by modifying hydrophobic groups into polar moieties and/or removing hydrophobic fragments from the compound is more likely to provide the reduction in metabolic turnover needed for a particular project.

4.2.7 SPR/STR Knowledge Extraction Using Matched Molecular Pair Analysis

Knowledge-driven modification of compounds is desirable to achieve the optimal potency and ADME properties. For each drug discovery project, a useful QSAR/QSPR model is able to accurately predict the activity of a compound. However, the model provides limited information pertaining to what modifications should be made to the compound in the next cycle of drug design. The matched molecular pair analysis (MMPA) technique is a promising approach to address this issue. MMPA was first coined by Kenny and Sadowski [142] to describe any systematic method of identifying structural matched molecular pairs (MMPs) from a set of compounds and associated property change. In this context, MMPs are generally defined as pairs of compounds that differ only by a single, localized structural transformation, and Fig. 4.10 shows an example [144].

The basic premise of MMPA is essentially an extraction of information within a chemical series featuring a common core. The property of interest can be plotted against the substituents at a given position of the core in order to identify the effects of the structural transformation on the property [145]. Various automated methods,
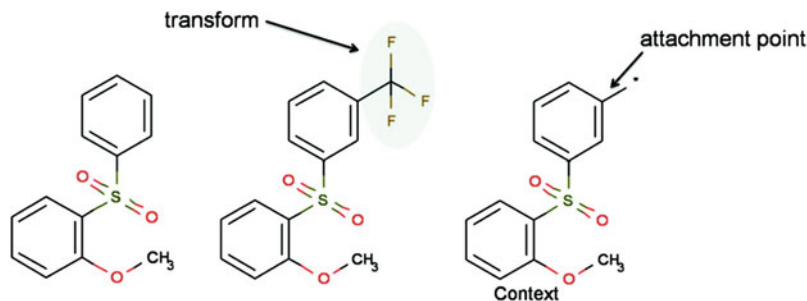
Fig. 4.10 Permission to use from Papadatos et al. [143]. Example of a matched molecular pair. The transformation is H to CF_3 (a single-point change) and is highlighted in *blue*. The *asterisk* in the context denotes the attachment point

including supervised and unsupervised methods, have been developed to identify MMPs and quantify the associated biological changes on large data sets. Supervised methods require predefined molecular transformations to identify the MMPs in the data set [144, 146]. However, any possible MMPs that are outside the predefined structural transformation dictionary cannot be identified. Unsupervised methods have the potential to identify all MMPs within a compound data set without a predefined molecular transformation dictionary [147–151]. It decomposes the compounds into fragments first and then indexes the fragments for rapid sorting and identifies the core scaffolds and R-group substituents. For a more detailed summarization of current MMPA methods, the reader is referred to a review by Griffen et al. [145].

After the MMPA algorithm identifies all possible MMPs, the results are tabulated to show differences between MMPs for a measured endpoint. The effect of a specific chemical substitution is typically summarized by the mean response change, the sample standard deviation of the response change, and the standard error of the mean for each endpoint. The total number of pairs identified for each substituent is also reported to assess the significance of the effects. Leach et al. recommended at least 20 MMPs should be identified for a useful molecular transformation [144]. More recently, Kramer et al. have recommended the use of paired *t*-test to calculate the number of pairs necessary to achieve statistical significance with a given average activity difference. They also demonstrated the importance of building pairs from identical assays measured in the same laboratory [152].

To provide quick and easy understandable guidance, the effects of a molecular transformation on different endpoints can be summarized by a simple symbolic colored arrow or circle that informs the medicinal chemists what compounds to be synthesized [153]. In addition, the structural transformations information can be summarized as rules in a knowledge database. By querying a compound of interest against the knowledge database with MMP rules in place, virtual compounds can be proposed to determine if the property of interest is likely to improve with the associated structural modification.

MMPA methods have been used to assess the mean effect of different substituents on various ADME parameters such as solubility [143, 144, 154], permeability [147, 149], clearance [149], and CYP inhibition [147]. Not surprisingly, common structural modifications, such as replacing hydrogen with a methyl group or changing a methyl to an ethyl substituent, were the most frequently observed MMPs [149].

In general, the structural changes that displayed favorable changes for an endpoint could also be explained by the associated change in physicochemical properties. For example, Gleeson et al. reported that replacing an aliphatic hydrogen atom with a hydroxyl, ethyl, or benzyl group leads to a decrease in CYP3A4 pIC50 > 0.2 log unit in 55%, 15%, and 10% of MMPs. This finding correlates well with the change in $c$log D (pH 7.4) of the substituents [147], meaning that as the compound becomes less lipophilic, it is less likely to be an inhibitor of CYP3A4. This observation is aligned with our internal analysis of trends between lipophilicity and CYP3A4 inhibition (Fig. 4.8).

Leach et al. also found that the addition of heavy halogens on aromatic rings was detrimental to solubility and a numerical estimate for such effects was also calculated. For instance, adding bromine to an aromatic ring led to over an order of magnitude reduction of aqueous solubility [144]. Therefore, if a drug discovery team is trying to increase the solubility of their scaffold, they should avoid adding heavier halogens, such as bromine, to their compounds.

While molecular substitutions that track closely with the molecular properties can be useful in guiding the design of new compounds, they may not be overly insightful to a well-versed medical chemist. It is more interesting to identify the substituents that display changes not associated with their physicochemical property changes. For example, despite the considerable increase in lipophilicity caused by phenyl substitutions of an aliphatic hydrogen (Δclog D at pH 7.4 of +1.8 log units), the average change in pIC_{50} of CYP1A2 inhibition for 147 pairs of compounds was quite insignificant (ΔpIC_{50} of 0.11) [147].

Another type of MMP is called "switch" transformations, which acts to turn on or turn off the activity. Regardless of the starting value of the endpoint, such MMP transformation results in approximately the same ending value. For example, it was reported that the replacement of a hydrogen by a 4-piperidine group resulted in a microsomal clearance value of ~20 μL/min/mg for all the studied compounds regardless of the starting microsomal clearance values [149].

One should be aware that MMPA results depend on both the transformation and the chemical context. This is manifested by the observation that although many of the molecular transformations are statistically significant with large mean activity changes, most of them also have high variability [149]. Therefore, making conclusions based on the average activity change across the entire MMPA data may be misleading for the chemical series of interest [143, 147]. For example, global context independent MMPA indicated that substituting a pyrimidine for a hydrogen atom increased CYP2C9 inhibition [147]. However, when the same substitution occurred for an aliphatic hydrogen (context dependent), a decrease in CYP2C9 inhibition was observed [147].
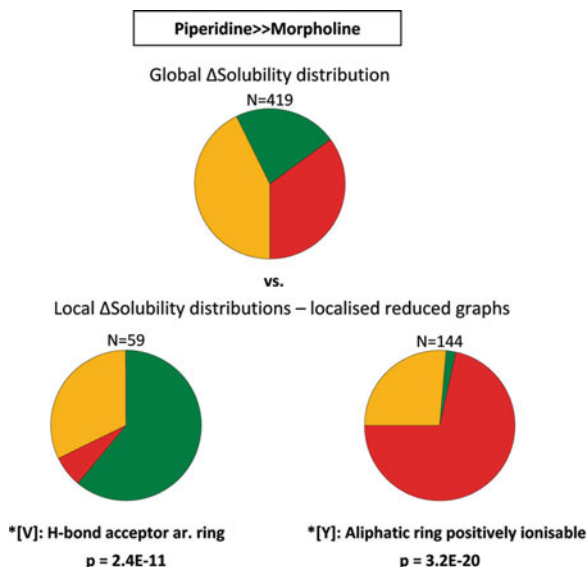
Fig. 4.11 Permission to use from Papadatos et al. [143]. Global and local MMPA distributions for the piperidine to morpholine transformation for a solubility data set. The *colors* reflect the effect of each transformation with *red*, *amber*, and *green* denoting unfavorable (decrease), zero, and favorable (increase) changes in solubility. Different outcomes are observed depending on the context of the compound; if the attachment point is a polar aromatic ring [V], then there is an increase in solubility, while if the attachment point is a positively ionizable aliphatic ring [Y], then solubility decreases

Another example also showed the importance of the chemical context for the MMP transformation. It was observed that transforming a piperidine ring into a morpholine ring has conflicting effects on solubility depending on whether the transformation was added to a polar aromatic ring or a positively ionizable aliphatic ring (Fig. 4.11) [143]. Several recent publications have proposed adding two dimensional contextual information about the compound or three dimensional (3-D) information pertaining to binding environment into the MMPA analysis to address the issue of context dependency in MMPA [155, 156].

4.3 Integrated and Iterative Use of Models in Early Drug Discovery

As mentioned in the introduction to this chapter, the application of in silico, in vitro, and in vivo models is inherent to the drug discovery process. It should be noted that the use of such models in isolation is unlikely to be fruitful and may even be misleading. Therefore, models should be applied in an integrated and iterative
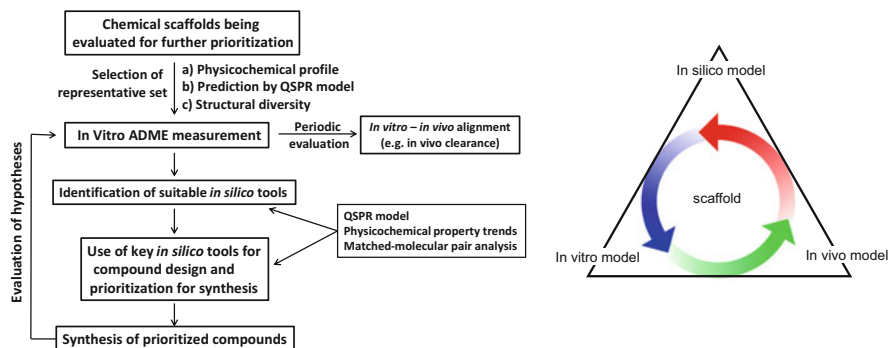
Fig. 4.12 Integrated and iterative use of models in early-phase drug discovery. The left schematic shows the recommended process to identify and integrate in silico, in vitro, and in vivo models. The schematic on the right illustrates the importance of the iterative learning cycle

fashion to build structure-activity and structure-property knowledge toward identifying the best clinical candidate possible for any given drug discovery project.

Once a scaffold has been identified that interacts with the desired pharmacological target, to assess the applicability of in silico ADME models for that particular scaffold, one needs to select a set of compounds that will be tested in vitro. As depicted in Fig. 4.12, this representative set should span the range of predicted in silico values, include various physicochemical characteristics, and include as much structural diversity as possible in order to systematically evaluate in silico model(s). While it would be preferred to select "active" compounds against the biological target for this assessment, this is not a requirement. It is more important to focus on including diversity as mentioned above. The in silico-in vitro analyses will help assess whether the in silico model(s) are applicable for a particular scaffold or along with predicted physicochemical properties can be used to guide and prioritize the synthesis of compounds. In an analogous manner, it is equally important to explore the relationship between in vitro ADME models and the in vivo profile of compounds in order to select an appropriate suite of in vitro tools to prioritize the selection of compounds for in vivo assessment. This iterative learning cycle (shown in Fig. 4.12) provides an efficient strategy to identify and resolve various challenges related to optimizing compound potency and ADME properties rather than using a filtration approach where only the active compounds progress for in vitro and in vivo ADME measurements.

To detail how this integrated and iterative process unfolds in the pharmaceutical industry, consider this example. The typical goal of most small compound drug discovery project is to identify compounds that can attain, and maintain, sufficient in vivo unbound concentration to engage the pharmacological target following oral dosing. To that end, it is important to balance compound potency with key ADME parameters like solubility, permeability, and clearance from the body. For this example, let us assume that the discovery project team has access to global QSPR models for solubility, permeability, and microsomal stability.

   The first step to establish the in silico-in vitro connectivity is to select a set of compounds from the scaffold and subsequently compare the outcome from corresponding in vitro measurements. This set of compounds should represent a range of predicted property (solubility, permeability, and microsomal stability), calculated phys-chem properties (e.g., $c\log P$, TPSA), and be structurally diverse. This step will determine if the global ADME QSPR models are applicable for the scaffold in question and if they provide reasonable predictive performance to enable the prioritization and design of compounds predicted to have a balanced ADME profile in terms of the three ADME endpoints mentioned above.

   Before implementing this strategy, it is important to test a small set of compounds spanning a range of measured solubility, permeability, and microsomal stability in the in vivo models to determine whether the oral exposure of these compounds is aligned with their in vitro profile. For example, if the in vivo clearance is rapid for compounds with low microsomal turnover in vitro, it would suggest that the primary clearance mechanism for such compounds is likely to involve non-oxidative pathways and/or excretion via renal or biliary route. Typically, elimination routes outside the oxidation pathway would not be identified using a microsomal stability assessment (in silico or in vitro). In such cases, one might consider testing the compounds in an in vitro hepatocyte clearance model (that will account for various non-CYP metabolic enzymes) to see if better alignment is observed with in vivo clearance. Once a suitable suite of in silico and in vitro tools have been identified that align with key in vivo characteristics, an efficient and robust strategy to integrate these models in an iterative manner can be implemented.

4.4   Summary

In this chapter, a variety of structure- and ligand-based in silico methods used to identify and resolve challenges related to the optimization of key ADME properties have been described. Given the promiscuity of many ADME targets and the limited availability of high-resolution 3-D structures, structure-based in silico techniques like docking and MD simulation have significant challenges and therefore have limited applicability for this purpose. Ligand-based in silico methods such as pharmacophore models can be useful to identify key structural features responsible for the interaction with the target of interest. However, due to broad ligand specificity and likelihood of multiple binding sites (e.g., P-glycoprotein) for many ADME targets, pharmacophore models also have limited prospective applicability across structurally diverse chemical scaffolds.

   QSPR models, especially machine learning models, can extract knowledge from a wide variety of chemical scaffolds and a large number of compounds enabling their utility as predictive models for many ADME endpoints. Not surprisingly, QSPR models are one of the most commonly employed in silico tools for ADME optimization during the drug discovery process, especially in an industrial setting

where a large number of structurally diverse compounds are routinely measured in a variety of ADME assays. At the same time, QSPR models have limited interpretability and thus typically don't provide direct clues to design new compounds to address ADME challenges.

To address that limitation of QSPR models, trends with calculated physicochemical properties like molecular weight, $c\log P$, TPSA, and others are effectively utilized during the design process to optimize the ADME characteristics of a given chemical scaffold. Similarly, knowledge extracted by the MMPA of existing ADME data also provides clues that identify fragment replacements toward improving the ADME characteristics.

To summarize, an effective amalgamation of in silico tools is valuable in guiding the design of compounds with favorable ADME properties on a drug discovery project. These models must be verified to show they provide valid predictions or the integrated in silico-in vitro-in vivo cycle breaks down. Finally, in silico tools should never be used in isolation. They make up one arm of the integrated and iterative learning cycle that we recommend using in order to effectively drive a drug discovery project.

References

1. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. J Health Econ. 2016;47:20–33.
2. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. BMC Biol. 2011;9:71.
3. Paquet E, Viktor HL. Molecular dynamics, Monte Carlo simulations, and langevin dynamics: a computational review. Biomed Res Int. 2015;2015:183918.
4. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struct. 2003;32:335–73.
5. Friesner RA, Banks JL, Murphy RB, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem. 2004;47 (7):1739–49.
6. Hu B, Lill MA. PharmDock: a pharmacophore-based docking program. J Cheminform. 2014;6(1):1–14.
7. Kuntz ID, Blaney JM, Oatley SJ, et al. A geometric approach to macromolecule-ligand interactions. J Mol Biol. 1982;161(2):269–88.
8. Rarey M, Kramer B, Lengauer T. Time-efficient docking of flexible ligands into active sites of proteins. Proc Int Conf Intell Syst Mol Biol. 1995;3:300–8.
9. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31 (2):455–61.
10. Verdonk ML, Cole JC, Hartshorn MJ, et al. Improved protein-ligand docking using GOLD. Proteins. 2003;52(4):609–23.
11. Zavodszky MI, Sanschagrin PC, Korde RS, et al. Distilling the essential features of a protein surface for improving protein-ligand docking, scoring, and virtual screening. J Comput Aided Mol Des. 2002;16(12):883–902.
12. Warren GL, Andrews CW, Capelli AM, et al. A critical assessment of docking programs and scoring functions. J Med Chem. 2006;49(20):5912–31.

13. Danielson ML, Desai PV, Mohutsky MA, et al. Potentially increasing the metabolic stability of drug candidates via computational site of metabolism prediction by CYP2C9: the utility of incorporating protein flexibility via an ensemble of structures. Eur J Med Chem. 2011;46 (9):3953–63.

14. Li H, Poulos TL. The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. Nat Struct Biol. 1997;4(2):140–6.

15. Hegde A, Haines DC, Bondlela M, et al. Interactions of substrates at the surface of P450s can greatly enhance substrate potency. Biochemistry. 2007;46(49):14010–7.

16. Park SY, Yamane K, Adachi S, et al. Thermophilic cytochrome P450 (CYP119) from Sulfolobus solfataricus: high resolution structure and functional properties. J Inorg Biochem. 2002;91(4):491–501.

17. Yano JK, Koo LS, Schuller DJ, et al. Crystal structure of a thermophilic cytochrome P450 from the archaeon Sulfolobus solfataricus. J Biol Chem. 2000;275(40):31086–92.

18. Yano JK, Wester MR, Schoch GA, et al. The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. J Biol Chem. 2004;279 (37):38091–4.

19. Ekroos M, Sjogren T. Structural basis for ligand promiscuity in cytochrome P450 3A4. Proc Natl Acad Sci U S A. 2006;103(37):13682–7.

20. Williams JA, Hyland R, Jones BC, et al. Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. Drug Metab Dispos. 2004;32(11):1201–8.

21. Lill MA, Dobler M, Vedani A. Prediction of small-molecule binding to cytochrome P450 3A4: flexible docking combined with multidimensional QSAR. ChemMedChem. 2006;1 (1):73–81.

22. Liu T, Qian G, Wang W, et al. Molecular docking to understand the metabolic behavior of GNF-351 by CYP3A4 and its potential drug-drug interaction with ketoconazole. Eur J Drug Metab Pharmacokinet. 2015;40(2):235–8.

23. Nookala AR, Li J, Ande A, et al. Effect of methamphetamine on spectral binding, ligand docking and metabolism of anti-HIV drugs with CYP3A4. PLoS One. 2016;11(1):e0146529.

24. Subhani S, Jamil K. Molecular docking of chemotherapeutic agents to CYP3A4 in non-small cell lung cancer. Biomed Pharmacother. 2015;73:65–74.

25. Sun H, Sharma R, Bauman J, et al. Differences in CYP3A4 catalyzed bioactivation of 5-aminooxindole and 5-aminobenzsultam scaffolds in proline-rich tyrosine kinase 2 (PYK2) inhibitors: retrospective analysis by CYP3A4 molecular docking, quantum chemical calculations and glutathione adduct detection using linear ion trap/orbitrap mass spectrometry. Bioorg Med Chem Lett. 2009;19(12):3177–82.

26. Tie Y, McPhail B, Hong H, et al. Modeling chemical interaction profiles: II. Molecular docking, spectral data-activity relationship, and structure-activity relationship models for potent and weak inhibitors of cytochrome P450 CYP3A4 isozyme. Molecules. 2012;17 (3):3407–60.

27. Huang SY, Grinter SZ, Zou X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. Phys Chem Chem Phys. 2010;12 (40):12899–908.

28. Jain AN. Scoring functions for protein-ligand docking. Curr Protein Pept Sci. 2006;7 (5):407–20.

29. Seifert MH. Optimizing the signal-to-noise ratio of scoring functions for protein–ligand docking. J Chem Inf Model. 2008;48(3):602–12.

30. Ferrara P, Gohlke H, Price DJ, et al. Assessing scoring functions for protein-ligand interactions. J Med Chem. 2004;47(12):3032–47.

31. Kuntz ID, Chen K, Sharp KA, et al. The maximal affinity of ligands. Proc Natl Acad Sci U S A. 1999;96(18):9997–10002.

32. Lill MA. Efficient incorporation of protein flexibility and dynamics into molecular docking simulations. Biochemistry. 2011;50(28):6157–69.

33. Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. Med Res Rev. 2006;26(5):531–68.
34. Naim M, Bhat S, Rankin KN, et al. Solvated interaction energy (SIE) for scoring protein-ligand binding affinities. 1. Exploring the parameter space. J Chem Inf Model. 2007;47(1):122–33.
35. Reddy MR, Reddy CR, Rathore RS, et al. Free energy calculations to estimate ligand-binding affinities in structure-based drug design. Curr Pharm Des. 2014;20(20):3323–37.
36. Garbett NC, Chaires JB. Thermodynamic studies for drug design and screening. Expert Opin Drug Discov. 2012;7(4):299–314.
37. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov. 2015;10(5):449–61.
38. Aqvist J, Medina C, Samuelsson JE. A new method for predicting binding affinity in computer-aided drug design. Protein Eng. 1994;7(3):385–91.
39. McCammon AJ, Gelin BR, Karplus M. Dynamics of folded proteins. Nature. 1977;267(5612):585–90.
40. Banks JL, Beard HS, Cao Y, et al. Integrated Modeling Program, Applied Chemical Theory (IMPACT). J Comput Chem. 2005;26(16):1752–80.
41. Brooks BR, Brooks 3rd CL, Mackerell Jr AD, et al. CHARMM: the biomolecular simulation program. J Comput Chem. 2009;30(10):1545–614.
42. Case DA, Cheatham 3rd TE, Darden T, et al. The amber biomolecular simulation programs. J Comput Chem. 2005;26(16):1668–88.
43. Christen M, Hunenberger PH, Bakowies D, et al. The GROMOS software for biomolecular simulation: GROMOS05. J Comput Chem. 2005;26(16):1719–51.
44. Jorgensen WL, Tirado-Rives J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. J Comput Chem. 2005;26(16):1689–700.
45. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. J Comput Chem. 2005;26(16):1781–802.
46. Van Der Spoel D, Lindahl E, Hess B, et al. GROMACS: fast, flexible, and free. J Comput Chem. 2005;26(16):1701–18.
47. Banks J, Brower RC, Ma J. Effective water model for Monte Carlo simulations of proteins. Biopolymers. 1995;35(3):331–41.
48. Berendsen HJ, Van Gunsteren WF, Zwinderman HR, et al. Simulations of proteins in water. Ann N Y Acad Sci. 1986;482:269–86.
49. Nutt DR, Smith JC. Molecular dynamics simulations of proteins: can the explicit water model Be varied? J Chem Theory Comput. 2007;3(4):1550–60.
50. Brannigan G, Lin LC, Brown FL. Implicit solvent simulation models for biomembranes. Eur Biophys J. 2006;35(2):104–24.
51. Rick SW. A reoptimization of the five-site water potential (TIP5P) for use with Ewald sums. J Chem Phys. 2004;120(13):6085–93.
52. Vorobjev YN. Advances in implicit models of water solvent to compute conformational free energy and molecular dynamics of proteins at constant pH. Adv Protein Chem Struct Biol. 2011;85:281–322.
53. Yang Y, Lightstone FC, Wong SE. Approaches to efficiently estimate solvation and explicit water energetics in ligand binding: the use of WaterMap. Expert Opin Drug Discov. 2013;8(3):277–87.
54. Dror RO, Jensen MO, Borhani DW, et al. Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. J Gen Physiol. 2010;135(6):555–62.
55. Pierce LC, Salomon-Ferrer R, Augusto FOC, et al. Routine access to millisecond time scale events with accelerated molecular dynamics. J Chem Theory Comput. 2012;8(9):2997–3002.
56. Brandman R, Lampe JN, Brandman Y, et al. Active-site residues move independently from the rest of the protein in a 200 ns molecular dynamics simulation of cytochrome P450 CYP119. Arch Biochem Biophys. 2011;509(2):127–32.

57. Bren U, Oostenbrink C. Cytochrome P450 3A4 inhibition by ketoconazole: tackling the problem of ligand cooperativity using molecular dynamics simulations and free-energy calculations. J Chem Inf Model. 2012;52(6):1573–82.

58. de Graaf C, Oostenbrink C, Keizers PH, et al. Free energies of binding of R- and S-propranolol to wild-type and F483A mutant cytochrome P450 2D6 from molecular dynamics simulations. Eur Biophys J. 2007;36(6):589–99.

59. Hritz J, de Ruiter A, Oostenbrink C. Impact of plasticity and flexibility on docking results for cytochrome P450 2D6: a combined approach of molecular dynamics and ligand docking. J Med Chem. 2008;51(23):7469–77.

60. Jerabek P, Florian J, Stiborova M, et al. Flexible docking-based molecular dynamics/steered molecular dynamics calculations of protein-protein contacts in a complex of cytochrome P450 1A2 with cytochrome b5. Biochemistry. 2014;53(42):6695–705.

61. Panneerselvam S, Yesudhas D, Durai P, et al. A combined molecular docking/dynamics approach to probe the binding mode of cancer drugs with cytochrome P450 3A4. Molecules. 2015;20(8):14915–35.

62. Hansch C, Maloney PP, Fujita T, et al. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. Nature. 1962;194 (4824):178–80.

63. Stouch TR, Kenyon JR, Johnson SR, et al. In silico ADME/Tox: why models fail. J Comput Aided Mol Des. 2003;17(2–4):83–92.

64. Desai PV, Sawada GA, Watson IA, et al. Integration of in silico and in vitro tools for scaffold optimization during drug discovery: predicting P-glycoprotein efflux. Mol Pharm. 2013;10 (4):1249–61.

65. Young D, Martin T, Venkatapathy R, et al. Are the chemical structures in your QSAR correct? QSAR Comb Sci. 2008;27(11–12):1337–45.

66. Todeschini R, Consonni V. Handbook of molecular descriptors. Federal Republic of Germany. WILEY-VCH Verlag, Weinheim; 2000.

67. Liu R, Zhou D. Using molecular fingerprint as descriptors in the QSPR study of lipophilicity. J Chem Inf Model. 2008;48(3):542–9.

68. Myint KZ, Wang L, Tong Q, et al. Molecular fingerprint-based artificial neural networks QSAR for ligand biological activity predictions. Mol Pharm. 2012;9(10):2912–23.

69. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model. 2010;50 (5):742–54.

70. Shen J, Cheng F, Xu Y, et al. Estimation of ADME properties with substructure pattern recognition. J Chem Inf Model. 2010;50(6):1034–41.

71. Wale N, Watson IA, Karypis G. Comparison of descriptor spaces for chemical compound retrieval and classification. Knowl Inf Syst. 2008;14(3):347–75.

72. McGregor MJ, Pallai PV. Clustering of large databases of compounds: using the MDL "keys" as structural descriptors. J Chem Inf Comput Sci. 1997;37(3):443–8.

73. Shahlaei M. Descriptor selection methods in quantitative structure-activity relationship studies: a review study. Chem Rev. 2013;113(10):8093–103.

74. van de Waterbeemd H, Gifford E. ADMET in silico modelling: towards prediction paradise? Nat Rev Drug Discov. 2003;2(3):192–204.

75. Ali J, Camilleri P, Brown MB, et al. Revisiting the general solubility equation: in silico prediction of aqueous solubility incorporating the effect of topographical polar surface area. J Chem Inf Model. 2012;52(2):420–8.

76. Breiman L. Bagging predictors. Mach Learn. 1996;24:123–40.

77. Svetnik V, Liaw A, Tong C, et al. Random forest: a classification and regression tool for compound classification and QSAR modeling. J Chem Inf Comput Sci. 2003;43(6):1947–58.

78. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–97.

79. Erickson JA, Mader MM, Watson IA, et al. Structure-guided expansion of kinase fragment libraries driven by support vector machine models. Biochim Biophys Acta. 2010;1804 (3):642–52.

80. Maltarollo VG, Gertrudes JC, Oliveira PR, et al. Applying machine learning techniques for ADME-Tox prediction: a review. Expert Opin Drug Metab Toxicol. 2015;11(2):259–71.
81. Cherkasov A, Muratov EN, Fourches D, et al. QSAR modeling: where have you been? Where are you going to? J Med Chem. 2014;57(12):4977–5010.
82. Gramatica P. Principles of QSAR models validation: internal and external. QSAR Comb Sci. 2007;26(5):694–701.
83. Golbraikh A, Tropsha A. Beware of q2! J Mol Graph Model. 2002;20(4):269–76.
84. Kubinyi H, Hamprecht FA, Mietzner T. Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. J Med Chem. 1998;41 (14):2553–64.
85. Sheridan RP. Time-split cross-validation as a method for estimating the goodness of prospective prediction. J Chem Inf Model. 2013;53(4):783–90.
86. Alexander DLJ, Tropsha A, Winkler DA. Beware of R(2): simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. J Chem Inf Model. 2015;55 (7):1316–22.
87. Spearman C. The proof and measurement of association between two things. Int J Epidemiol. 2010;39(5):1137–50.
88. Hu Y, Unwalla R, Denny AR, et al. Development of QSAR models for microsomal stability: identification of good and bad structural features for rat, human and mouse microsomal stability. J Comput Aided Mol Des. 2010;24(1):23–35.
89. Lee PH, Cucurull-Sanchez L, Lu J, et al. Development of in silico models for human liver microsomal stability. J Comput Aided Mol Des. 2007;21(12):665–73.
90. Jaworska J, Nikolova-Jeliazkova N, Aldenberg T. QSAR applicability domain estimation by projection of the training set descriptor space: a review. Altern Lab Anim. 2005;33 (5):445–59.
91. Keefer CE, Kauffman GW, Gupta R. Interpretable, probability-based confidence metric for continuous quantitative structure-activity relationship models. J Chem Inf Model. 2013;53 (2):368–83.
92. Sheridan RP. The relative importance of domain applicability metrics for estimating prediction errors in QSAR varies with training set diversity. J Chem Inf Model. 2015;55 (6):1098–107.
93. Toplak M, Močnik R, Polajnar M, et al. Assessment of machine learning reliability methods for quantifying the applicability domain of QSAR regression models. J Chem Inf Model. 2014;54(2):431–41.
94. Susnow RG, Dixon SL. Use of robust classification techniques for the prediction of human cytochrome P450 2D6 inhibition. J Chem Inf Comput Sci. 2003;43(4):1308–15.
95. Wenlock MC, Carlsson LA. How experimental errors influence drug metabolism and pharmacokinetic QSAR/QSPR models. J Chem Inf Model. 2015;55(1):125–34.
96. Cheng F, Li W, Liu G, et al. In silico ADMET prediction: recent advances, current challenges and future trends. Curr Top Med Chem. 2013;13(11):1273–89.
97. Stepensky D. Prediction of drug disposition on the basis of its chemical structure. Clin Pharmacokinet. 2013;52(6):415–31.
98. Wang J, Urban L. In vitro–in silico tools to predict pharmacokinetics of poorly soluble drug compounds. In: Wagner C, Dressman JB, editors. Predictive ADMET: integrative approaches in drug discovery and development. New York: Wiley; 2014. p. 233–61.
99. Lipinski CA, Lombardo F, Dominy BW, et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 1997;23(1–3):3–25.
100. Gleeson PM. Generation of a set of simple, interpretable ADMET rules of thumb. J Med Chem. 2008;51(4):817–34.
101. Meanwell NA. Improving drug candidates by design: a focus on physicochemical properties as a means of improving compound disposition and safety. Chem Res Toxicol. 2011;24 (9):1420–56.

102. Varma MVS, Feng B, Obach SR, et al. Physicochemical determinants of human renal clearance. J Med Chem. 2009;52(15):4844–52.
103. Wager TT, Hou X, Verhoest PR, et al. Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. ACS Chem Neurosci. 2010;1(6):435–49.
104. Ertl P, Rohde B, Selzer P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. J Med Chem. 2000;43(20):3714–7.
105. Austin RP, Barton P, Cockroft SL, et al. The influence of nonspecific microsomal binding on apparent intrinsic clearance, and its prediction from physicochemical properties. Drug Metab Dispos. 2002;30(12):1497–503.
106. Wermuth CG, Ganellin CR, Lindberg P, et al. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). Pure Appl Chem. 1998;70:1129–43.
107. de Groot MJ, Ekins S. Pharmacophore modeling of cytochromes P450. Adv Drug Deliv Rev. 2002;54(3):367–83.
108. Demel MA, Schwaha R, Kramer O, et al. In silico prediction of substrate properties for ABC-multidrug transporters. Expert Opin Drug Metab Toxicol. 2008;4(9):1167–80.
109. Guner OF, Bowen JP. Pharmacophore modeling for ADME. Curr Top Med Chem. 2013;13 (11):1327–42.
110. Yamashita F, Hashida M. In silico approaches for predicting ADME properties of drugs. Drug Metab Pharmacokinet. 2004;19(5):327–38.
111. Dixon SL, Smondyrev AM, Knoll EH, et al. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. J Comput Aided Mol Des. 2006;20(10–11):647–71.
112. Molecular Operating Environment (MOE), 2013.08 (2016). 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7: Chemical Computing Group Inc.
113. Wolber G, Langer T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J Chem Inf Model. 2005;45(1):160–9.
114. Raymond JW, Willett P. Maximum common subgraph isomorphism algorithms for the matching of chemical structures. J Comput Aided Mol Des. 2002;16(7):521–33.
115. Koch I. Enumerating all connected maximal common subgraphs in two graphs. Theor Comput Sci. 2001;250(1–2):1–30.
116. Chang C, Pang KS, Swaan PW, et al. Comparative pharmacophore modeling of organic anion transporting polypeptides: a meta-analysis of rat Oatp1a1 and human OATP1B1. J Pharmacol Exp Ther. 2005;314(2):533–41.
117. Li WX, Li L, Eksterowicz J, et al. Significance analysis and multiple pharmacophore models for differentiating P-glycoprotein substrates. J Chem Inf Model. 2007;47(6):2429–38.
118. Ritschel T, Hermans SM, Schreurs M, et al. In silico identification and in vitro validation of potential cholestatic compounds through 3D ligand-based pharmacophore modeling of BSEP inhibitors. Chem Res Toxicol. 2014;27(5):873–81.
119. Hughes TB, Miller GP, Swamidass SJ. Site of reactivity models predict molecular reactivity of diverse chemicals with glutathione. Chem Res Toxicol. 2015;28(4):797–809.
120. Kirchmair J, Williamson MJ, Afzal AM, et al. FAst MEtabolizer (FAME): a rapid and accurate predictor of sites of metabolism in multiple species by endogenous enzymes. J Chem Inf Model. 2013;53(11):2896–907.
121. Peng J, Lu J, Shen Q, et al. In silico site of metabolism prediction for human UGT-catalyzed reactions. Bioinformatics. 2014;30(3):398–405.
122. Smith PA, Sorich MJ, Low LS, et al. Towards integrated ADME prediction: past, present and future directions for modelling metabolism by UDP-glucuronosyltransferases. J Mol Graph Model. 2004;22(6):507–17.
123. Kingsley LJ, Wilson GL, Essex ME, et al. Combining structure- and ligand-based approaches to improve site of metabolism prediction in CYP2C9 substrates. Pharm Res. 2015;32 (3):986–1001.

124. Li J, Cai J, Su H, et al. Effects of protein flexibility and active site water molecules on the prediction of sites of metabolism for cytochrome P450 2C19 substrates. Mol BioSyst. 2016;12(3):868–78.
125. Liu R, Liu J, Tawa G, et al. 2D SMARTCyp reactivity-based site of metabolism prediction for major drug-metabolizing cytochrome P450 enzymes. J Chem Inf Model. 2012;52 (6):1698–712.
126. Liu X, Shen Q, Li J, et al. In silico prediction of cytochrome P450-mediated site of metabolism (SOM). Protein Pept Lett. 2013;20(3):279–89.
127. Matlock MK, Hughes TB, Swamidass SJ. XenoSite server: a web-available site of metabolism prediction tool. Bioinformatics. 2015;31(7):1136–7.
128. Moors SL, Vos AM, Cummings MD, et al. Structure-based site of metabolism prediction for cytochrome P450 2D6. J Med Chem. 2011;54(17):6098–105.
129. Rudik AV, Dmitriev AV, Lagunin AA, et al. Metabolism site prediction based on xenobiotic structural formulas and PASS prediction algorithm. J Chem Inf Model. 2014;54(2):498–507.
130. Rydberg P, Olsen L. Ligand-based site of metabolism prediction for cytochrome P450 2D6. ACS Med Chem Lett. 2012;3(1):69–73.
131. Sheng Y, Chen Y, Wang L, et al. Effects of protein flexibility on the site of metabolism prediction for CYP2A6 substrates. J Mol Graph Model. 2014;54:90–9.
132. Tarcsay A, Keseru GM. In silico site of metabolism prediction of cytochrome P450-mediated biotransformations. Expert Opin Drug Metab Toxicol. 2011;7(3):299–312.
133. Tarcsay A, Kiss R, Keseru GM. Site of metabolism prediction on cytochrome P450 2C9: a knowledge-based docking approach. J Comput Aided Mol Des. 2010;24(5):399–408.
134. Tyzack JD, Mussa HY, Williamson MJ, et al. Cytochrome P450 site of metabolism prediction from 2D topological fingerprints using GPU accelerated probabilistic classifiers. J Cheminform. 2014;6:29.
135. Vasanthanathan P, Hritz J, Taboureau O, et al. Virtual screening and prediction of site of metabolism for cytochrome P450 1A2 ligands. J Chem Inf Model. 2009;49(1):43–52.
136. Yamazoe Y, Ito K, Yoshinari K. Construction of a CYP2E1-template system for prediction of the metabolism on both site and preference order. Drug Metab Rev. 2011;43(4):409–39.
137. Zamora I, Afzelius L, Cruciani G. Predicting drug metabolism: a site of metabolism prediction tool applied to the cytochrome P450 2C9. J Med Chem. 2003;46(12):2313–24.
138. Zheng M, Luo X, Shen Q, et al. Site of metabolism prediction for six biotransformations mediated by cytochromes P450. Bioinformatics. 2009;25(10):1251–8.
139. Zaretzki J, Bergeron C, Rydberg P, et al. RS-predictor: a new tool for predicting sites of cytochrome P450-mediated metabolism applied to CYP 3A4. J Chem Inf Model. 2011;51 (7):1667–89.
140. Zaretzki J, Rydberg P, Bergeron C, et al. RS-predictor models augmented with SMARTCyp reactivities: robust metabolic regioselectivity predictions for nine CYP isozymes. J Chem Inf Model. 2012;52(6):1637–59.
141. Vaz RJ, Zamora I, Li Y, et al. The challenges of in silico contributions to drug metabolism in lead optimization. Expert Opin Drug Metab Toxicol. 2010;6(7):851–61.
142. Kenny PW, Sadowski J. Structure modification in chemical databases. In: Oprea TI, editor. Chemoinformatics in drug discovery. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2005. p. 271–85.
143. Papadatos G, Alkarouri M, Gillet VJ, et al. Lead optimization using matched molecular pairs: inclusion of contextual information for enhanced prediction of HERG inhibition, solubility, and lipophilicity. J Chem Inf Model. 2010;50(10):1872–86.
144. Leach AG, Jones HD, Cosgrove DA, et al. Matched molecular pairs as a guide in the optimization of pharmaceutical properties; a study of aqueous solubility, plasma protein binding and oral exposure. J Med Chem. 2006;49(23):6672–82.
145. Griffen E, Leach AG, Robb GR, et al. Matched molecular pairs as a medicinal chemistry tool: miniperspective. J Med Chem. 2011;54(22):7739–50.

146. Haubertin DY, Bruneau P. A database of historically-observed chemical replacements. J Chem Inf Model. 2007;47(4):1294–302.
147. Gleeson P, Bravi G, Modi S, et al. ADMET rules of thumb II: a comparison of the effects of common substituents on a range of ADMET parameters. Biorg Med Chem. 2009;17 (16):5906–19.
148. Hussain J, Rea C. Computationally efficient algorithm to identify matched molecular pairs (MMPs) in large data sets. J Chem Inf Model. 2010;50(3):339–48.
149. Keefer CE, Chang G, Kauffman GW. Extraction of tacit knowledge from large ADME data sets via pairwise analysis. Biorg Med Chem. 2011;19(12):3739–49.
150. Sheridan RP, Hunt P, Culberson JC. Molecular transformations as a way of finding and exploiting consistent local QSAR. J Chem Inf Model. 2006;46(1):180–92.
151. Warner DJ, Griffen EJ, St-Gallay SA. WizePairZ: a novel algorithm to identify, encode, and exploit matched molecular pairs with unspecified cores in medicinal chemistry. J Chem Inf Model. 2010;50(8):1350–7.
152. Kramer C, Fuchs JE, Whitebread S, et al. Matched molecular pair analysis: significance and the impact of experimental uncertainty. J Med Chem. 2014;57(9):3786–802.
153. Ritchie TJ, Ertl P, Lewis R. The graphical representation of ADME-related molecule properties for medicinal chemists. Drug Discov Today. 2011;16(1):65–72.
154. Zhang L, Zhu H, Mathiowetz A, et al. Deep understanding of structure–solubility relationship for a diverse set of organic compounds using matched molecular pairs. Biorg Med Chem. 2011;19(19):5763–70.
155. Posy SL, Claus BL, Pokross ME, et al. 3D matched pairs: integrating ligand-and structure-based knowledge for ligand design and receptor annotation. J Chem Inf Model. 2013;53 (7):1576–88.
156. Weber J, Achenbach J, Moser D, et al. VAMMPIRE: a matched molecular pairs database for structure-based drug design and optimization. J Med Chem. 2013;56(12):5203–7.