Published in final edited form as:

Clin Pharmacol Ther. 2018 November ; 104(5): 818–835. doi:10.1002/cpt.1174.

Molecular Modeling of Drug–Transporter Interactions—an International Transporter Consortium Perspective†

Avner Schlessinger¹, Matthew A. Welch², Herman van Vlijmen³, Ken Korzekwa⁴, Peter W. Swaan², and Pär Matsson^{5,6}

¹Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY

²Department of Pharmaceutical Sciences, University of Maryland, Baltimore, MD

³Computational Chemistry, Discovery Sciences, Janssen Research & Development, Beerse, Belgium

⁴Department of Pharmaceutical Sciences, Temple University, Philadelphia, PA

⁵Department of Pharmacy, Uppsala University, Sweden

Abstract

Membrane transporters play diverse roles in the pharmacokinetics and pharmacodynamics of small-molecule drugs. Understanding the mechanisms of drug-transporter interactions at the molecular level is therefore essential for the design of drugs with optimal therapeutic effects. This white paper examines recent progress, applications and challenges of molecular modeling of membrane transporters, including modeling techniques that are centered on the structures of transporter ligands, and those focusing on the structures of the transporters. The goals of this manuscript are to illustrate current best practices and future opportunities in using molecular modeling techniques to understand and predict transporter-mediated effects on drug disposition and efficacy.

Keywords

Transporters; Computational biology; Modeling; Drug Transport

Introduction

Membrane transporters from the solute carrier (SLC) and ATP-binding cassette (ABC) superfamilies regulate the cellular uptake, efflux and homeostasis of many essential nutrients and significantly impact the pharmacokinetics of drugs (1-4); further, they may provide targets for novel therapeutics as well as facilitate prodrug approaches (5, 6). Because of their

†A list of ITC members appears on the ITC website, www.itc-transporter.org.

⁶Address correspondence to: Pär Matsson, Department of Pharmacy, Uppsala University, Box 580, SE-75123 Uppsala, Sweden, Phone: +46-(0)18-471 46 30, Fax: +46-(0)18-471 42 23, par.matsson@farmaci.uu.se.

Disclosure

H.v.V. is an employee of Janssen.

often broad substrate selectivity they are also implicated in many undesirable and sometimes life-threatening drug-drug interactions (DDIs) (5, 6).

Despite their clinical significance, most membrane transporters are poorly characterized at the molecular level. For example, atomic-resolution structures have been determined only for a very limited number of mammalian transporters (7, 8). Consequently, elucidation of drug-transporter interactions, DDIs and rational design of therapeutics aimed at this important class of membrane protein targets remains unacceptably serendipitous. Despite the limited number of human transporter structures, computational models—whether based on inhibition or substrate transport data or evolutionary relationships to proteins from other species—can provide important insights into their structural requirements. Thus, the aim of transporter modeling is to gain a deeper understanding of the proteins' structures, their membrane topology, and their different conformational states (e.g. inward/outward facing), which in turn will inform on the mechanism of substrate binding, translocation, and kinetics, as well as the mode of action for inhibitors. Further, transporter structure models may be used to visualize and provide insight into how genetic polymorphism affects transporter function and/or expression. Ultimately, the goal of transporter modeling in the pharmaceutical sciences is to inform drug discovery and development and to provide tools for *a priori* determining potential drug-transporter interactions that may lead to DDIs or pathologies such as drug-induced liver injury mediated by transporters.

The impact of different transporters in current pharmaceutical drug discovery is variable. One of the clearest examples of undesired compound activity is being a substrate of MDR-1 (Multidrug-resistance protein 1; P-glycoprotein; *ABCB1*) for drug candidates targeting the central nervous system (CNS) (9). Unless the compound also has a high passive diffusion rate across the blood-brain barrier, this will usually make the compound unsuitable for CNS applications. Consequently, computational MDR-1 models are commonly used in industry to avoid the synthesis of compounds that are likely to be substrates. Another transporter that is usually a red flag in drug discovery is the bile salt export pump (BSEP; *ABCB11*), inhibitors of which carry a potential for inducing cholestasis (for examples see Supplemental Reading List). Most other transporters are commonly not seen as showstoppers for further development of a compound, although effects on tissue distribution and DDIs can be significant. For instance, the organic anion transporting polypeptides OATP1B1 and OATP1B3 (*SLCO1B1/SLCO1B3*) mediate selective liver targeting of several compounds including marketed drugs (Supplemental Reading List). In silico models for predicting substrates of these OATPs can play an important role in the design of compounds when liver targeting is desired (10). For diseases where drug combinations are the rule rather than the exception, such as cancer and viral infections, predictive models for transporter substrates are especially useful for indicating potential DDI risks.

The 2010 ITC white paper on Transporters in Drug Development contained a brief section and a decision tree on the use of computational models to predict and assess drug-transporter interactions (4). We highlighted the importance of accelerating the pace of atomic level structure determination for transporter proteins and the need to combine computational technologies to increase fidelity of transporter models; further, we called for expanding transporter substrate and inhibitor datasets that would allow models to predict transporter

affinity rather than inhibition. Over the past 10 years, an abundance of new membrane protein structures have been determined, including homologues of key transporters involved in drug disposition and dynamics. At the same time, newer modeling and simulation algorithms have been developed to aid in constructing improved structural models ('structure-based approaches') as well as predictive models based on substrate and inhibitor data ('ligand-based approaches'). For ligand-based models, the development of new structure-independent algorithms such as Bayesian inference has proven useful. Likewise, recent development in computing power and molecular dynamics (MD) simulation methodologies have greatly increased the timescale and confidence of structure-based simulations. However, shortcomings still exist in the way we obtain input data for transporter modeling through experimental assays and the derivation of kinetic parameters from these, and experimental data is still scarce for many transporters of relevance for drug development. The present review aims to provide an update and an expansive view of contemporary techniques in transporter modeling and data acquisition through in vitro assays. In the following sections, we discuss the technology and applications to transporter modeling, current capabilities, pitfalls and future directions.

Experimental data and implications for modeling

The type and quality of available experimental data dictates what computational modeling techniques can and should be used, and what types of inferences can be drawn from the models. For example, very different ligand-based models will be obtained if the underlying data are of substrate *transport* (i.e., the ligands are themselves transported) or transporter *inhibition* (the ligand is decreasing the transport of another molecule). Further, model fidelity will differ depending on whether the underlying experimental data were determined at a single concentration or in a more detailed concentration-dependence assay. Similarly, the quality of inferences made from a structure-based model is closely linked to the resolution of the experimentally determined protein structure, and the quality of sequence alignments between template structures and modeled target transporters.

Inhibition or substrate transport data

A major consideration when generating datasets for use in modeling (either as the basis for ligand-based models or as validation of structure-based virtual screening) is whether to use a *substrate* or an *inhibition* assay. Unfortunately, inhibition assays are not generally predictive of substrate transport, and vice versa. This is partly due to different molecular interactions between the ligands and the transporter: a transported substrate can be a weak inhibitor depending on the strength of its molecular interactions with the transporter binding site; conversely, inhibitors are defined by their binding to the transporter in a way that affects the transport of other molecules, which does not necessarily translate to itself being transported (11).

Depending on the transporter, different correlation is seen between substrate and inhibitor activity. For example, for many transporters of exogenous molecules (e.g. MDR-1/P-gp, breast cancer resistance protein—BCRP, and OATP1B1/1B3/2B1) inhibitors often tend to be substrates as well (4). Other transporters vary greatly in their inhibitor–substrate

relationships. Most di- and tripeptides are inhibitors of the oligopeptide transporter PEPT1 (*SLC15A1*) but not all are transported substrates (12). Inhibitors of the organic cation transporter 2 (OCT2; *SLC22A2*) could, depending on their molecular properties be divided into those that are both substrates and inhibitors and those that only inhibit the transporter (13), and most of the reported inhibitors of the plasma membrane monoamine transporter (PMAT; *SLC29A4*) (14) are non-substrates.

The ultimate use of the transporter model is, therefore, an important consideration when the type of ligand data is selected. If inhibition is the primary concern, for example due to the potential of drug interactions or toxicity, an inhibitor model may be most useful. If the model is to be used to predict transporter activity (e.g. to optimize or limit tissue uptake) a substrate activity assay is almost always required.

In addition to considerations of the differences between substrate and inhibitor binding when generating datasets for modeling, allosteric binding sites and/or transport modes constitute an additional layer of complexity for certain transporters, and will often impact assay selection. For example, in the case of MDR-1, no single quantitative structure-activity relationship (QSAR) or pharmacophore model can describe the spatial arrangement of structural features responsible for substrate and inhibitor affinity (9, 15, 16). This is reflected in the existence of multiple models, which account for multiple binding sites and allosteric mechanisms of interaction. Recent crystal structures demonstrate that multiple allosteric binding sites are a common feature in ABC transporters (17, 18). Similar binding site diversity has been shown in SLC transporters: for example, Wright and colleagues investigated the potential influence of the substrate on the inhibition profiles of OCT2 and multidrug and toxin extrusion protein 1 (MATE1; *SLC47A1*) (19, 20) and found that 1) the choice of substrate significantly influences both quantitative and qualitative inhibitory interactions with cationic drugs; and 2) ligand interactions with OCT2 and MATE1 are not restricted to competition for a common ligand binding site, consistent with a binding surface characterized by multiple, possibly overlapping interaction sites. Consequently, they concluded that development of predictive models of DDIs with OCT2 and MATE1 must take into account the substrate dependence of ligand interaction with these proteins (19, 20); this conclusion most likely applies to many other SLC and ABC transporters as well.

Generating inhibition data

Inhibition assays are usually simpler than assays of substrate transport, since a common endpoint is used for all ligands. This allows (semi-) high-throughput assays to be developed for a single probe substrate (the transport of which is quantitated by fluorescence, radioactivity, LC-MS, etc.). The effect that co-incubation with other compounds (the potential inhibitors) has on the transport of the probe substrate is then assessed. This setup allows relatively large numbers of potential ligands to be tested (often in the range of hundreds to thousands) (13, 21, 22). The assay formats typically used are similar to other high-throughput experiments used in screening for drug–target binding, and similar considerations are necessary regarding assay stability and signal-to-noise ratios. Common criteria for *in vitro* screening assays are discussed in, e.g., (23). For example, it is recommended that computational filters are applied on the compound set used in the

screening campaign, to identify potential pan-assay interference (PAINS) compounds—i.e., compounds that may give rise to false signals in the assay through mechanisms that are not related to a true ligand–transporter interaction (24, 25). Interfering compounds should either be removed from the compound libraries prior to screening, or flagged for thorough verification of the inhibition mechanism; examples include rhodanines and enones which frequently give false-positive results in experimental testing through various mechanisms (e.g., molecular aggregation). Specialized filtering programs can be used to avoid these issues (See also Supplemental Reading List).

Generating substrate data

Substrate assays, in which transporter activity is measured for individual substrates, are much more labor intensive since both transport and downstream analytical assays must often be optimized for each compound. Traditionally, assays have often been based on radiometric, UV, or fluorescence detection, but the generalizability and sensitivity of LC-MS-based analytics has made this the predominant technique in transporter substrate measurements.

The cellular transport of a drug molecule is typically the result of multiple parallel processes: the drug may be a substrate of one or more transporters but it will also, with varying rates, diffuse across the phospholipid bilayers of cell membranes (26). Therefore, such passive lipoidal permeability of the substrates constitutes another significant experimental constraint when generating transporter modeling datasets, with different assay types being more or less suitable depending on the passive diffusion rates of the substrates. Examples of when different assay types have been used to generate data for subsequent modeling are given below; please see Table 1 and supplemental reading list for additional information.

Many transporters have evolved for the transport of poorly permeable substrates across cellular membranes. Such transporters can usually be characterized by an uptake assay, in which the rate of substrate entry is measured in cells or membrane vesicles that express the transporter (either naturally or through transient or stable transfection) (4). Rates in the transporter-expressing system are typically compared to those in a control system that lacks expression of the studied transporter. Examples where natural transporter expression has been used to generate transporter data for computational modeling include the ABC efflux transporter MDR-1, the intestinal bile acid transporter ASBT (apical sodium–bile acid transporter; *SLC10A2*) and the oligopeptide transporter PEPT1, all of which are endogenously expressed in the human colon carcinoma cell line Caco-2 (Examples are listed in the Supplemental Reading List). The human embryonic kidney cell line HEK293 is a commonly used host for transfection of transporters, and datasets generated in HEK293 transfectants have been used in computational modeling of several human transporters, including OATP1B1/1B3/2B1, OCT1 (*SLC22A1*), MATE1, and PMAT. Transient expression in oocytes has been used for many transporters including nucleoside transporters CNT1, CNT2 (*SLC28A1/2*), and ENT1 (*SLC29A1*) and rat Oatp1a5 (*Slco1a5*) (Supplemental Reading List). In some cases monolayer-forming cells are preferred, and transporter modeling data sets have, for example been generated in ASBT-expressing

Madine-Darby Canine Kidney (MDCK) cell monolayers. Transfected cell systems can be advantageous when generating data for modeling, since the effects observed can be attributed to the specific recombinant transporter. Depending on the background expression of transporters in the host cell, however, confounding transport mechanisms may be present. Advances in genome editing, e.g. using CRISPR-Cas9 to selectively knock out genes (27, 28), and the establishing of haploid cell libraries (29), are providing additional tools for rapid and selective probing of specific transporters that will open up new opportunities for molecular modeling in the near future.

Intact cells are challenging to use for efflux transporters of *poorly permeable substrates*, since the substrate must enter the cell before it can access the transporter. The poor rate of passive entry into the cell may thus obscure the interaction with the efflux transporter. In such cases, inverted membrane vesicles obtained from transporter-expressing cells are a commonly used alternative to cell-based assays. These vesicles are mixtures of inside-out and right-side-out vesicles, but since only the inverted vesicles react with externally applied ATP, energy-dependent uptake rates only reflect the inside-out activity. Datasets from vesicles have been used to model, for example, BCRP, multidrug resistance-associated proteins 3 (MRP3; *ABCC3*) and 4 (MRP4; *ABCC4*) (HEK cell vesicles), rat Mrp2 (canalicular membrane vesicles), and the dopamine transporter DAT (*SLC6A3*; caudate putamen membrane vesicles) (Supplemental Reading List).

Some efflux transporters, including MDR-1 and BCRP, also affect the distribution of *membrane-permeable* substrates. These transporters have important roles in keeping exogenous substrates from entering peripheral tissues such as the central nervous system, preventing intestinal absorption, or excreting substrates into the bile or urine. For permeable efflux transporter substrates, the use of cellular or vesicular uptake assays pose a significant challenge as background permeability across the lipid membrane can diminish differences relative to control cells that lack the transporter, thus masking the effect of the transporter. Instead, it is typically preferable to measure permeability across confluent monolayers of cells (e.g. Caco-2, MDCK, or LLC-PK1 cells). Most commonly, flux ratios are established that compare permeabilities in the apical-to-basolateral and basolateral-to-apical directions; non-equal permeability is a sign of carrier mediated transport. Datasets from such assays have been used to build models for, e.g., MDR-1 (Supplemental Reading List).

Resolution and merging of data

The resolution of the various assays used to obtain drug–transporter interaction data is also an important consideration for downstream modeling efforts. Many of the transporter models reported in the scientific literature (for examples, see Table 1) have been built with data generated at a single inhibitor concentration using high-throughput inhibition assays. Models based on such screening-type data may, naturally, be less exact than models based on potency values obtained from concentration-dependent inhibition experiments; inhibition potencies are typically reported as the assay concentration leading to half-maximal inhibition (IC₅₀) or as drug–transporter dissociation constants (K_i). While screening assays do not provide the fidelity of a full concentration-dependence experiment, the higher

throughput allows for generation of substantially larger datasets and improved model coverage.

Analogous to inhibition experiments, substrate transport data are reported at different resolution, ranging from single-concentration experiments to full concentration profiles from which the Michaelis-Menten kinetic parameters $V_{\max}$ (maximum transport rate) and K_m (substrate concentration at half-maximal rate) can be obtained.

The dearth of larger datasets for many transporters—in particular for substrate transport—makes merging of data from multiple sources the only viable option to access large enough datasets for ligand-based modeling. In such cases, care should be taken to ensure that the data have been generated in similar assays. A number of factors can lead to inter-assay and inter-laboratory variability, including differences in the expression of transporters in different cell types and in different clones of the same cell line (47), different probe substrates (potentially interacting with different binding sites as described above), and differences in other assay conditions. Typically, due to the inherent variability between datasets, binning of data into classes such as low/medium/high affinity or potency (48) is necessary when data from different laboratories or assays are combined for a modeling project.

Importantly, although the kinetic parameters derived from a transport experiment are often reported as the system-independent parameters K_i and K_m (reflecting the affinity of the transporter for the inhibitor or substrate, respectively), the results are usually more correctly described as the *apparent* constants $K_{i,app}$ and $K_{m,app}$ unless the system is very well characterized. For example, these parameters will only describe affinity if they are based on the drug concentration *at the transporter binding site*. Different levels of nonspecific binding or membrane partitioning can be expected between different assays, leading to variability in the substrate and/or inhibitor concentration that reaches the binding site. Any decrease in the concentration available for interaction with the transporter will increase the apparent K_m or K_i values (49). For transporters that efflux substrates directly from the membrane (e.g., MDR-1 and BCRP) the apparent K_m and K_i values depend on the level of transporters in the membrane. This is because the transporter decreases the concentration of drug in the membrane, and higher expression levels will require higher drug levels to saturate the membrane (50). This will be true whenever the transporter alters the concentration in the driving compartment (51).

Other reasons for variability between datasets may be due to the non-Michaelis-Menten characteristics of some transporters (52). Similar to most drug metabolizing enzymes, transporters that show broad substrate selectivity (e.g., MDR-1, BCRP, OATP1B1/1B3/2B1) can show non-hyperbolic saturation kinetics and mixed inhibition kinetics. This is likely due to the simultaneous binding of multiple substrates and inhibitors. The most significant implication is that inhibition constants can vary depending on the probe substrate that is used, as exemplified above for MDR-1, OCT2 and MATE1 (15, 16, 19, 20). Also, interactions between two compounds can vary between complete inhibition, partial inhibition, and activation, and the molecular mechanisms driving such variable interaction modes are not well understood.

Taken together, many considerations go into the choice, setup and analysis of data from experimental transporter assays. Importantly, any assumptions and biases associated with the experiments will translate to ligand-based models trained on the generated data. Consequently, a thorough understanding of the strengths and weaknesses of the experimental assays is necessary when designing computational models that build on experimental data.

Ligand-Based Molecular Modeling

Ligand-based molecular modeling describes ligand–transporter interactions through correlating the molecular properties and structure of a set of ligands to an observed transporter activity—typically the translocation of substrates or the inhibition of transport activity as described above. Once a model is developed based on prior empirical data, it can be used to predict the activity of other ligands based on their molecular properties and structure. In addition, information about transporter protein structure can also be inferred by analyzing the most predictive molecular features included in the model.

A central step in all ligand-based modeling is the description of the structure and/or molecular properties of the ligands (Figure 1). To ensure that the molecular descriptors accurately describe the molecule in the environment of interest, general considerations such as ionization state, stereochemistry, and absence/presence of counterions need to be accounted for; this is typically done using specialized pre-filtering software. For three-dimensional (3D) ligand modeling techniques such as 3D-quantitative structure-activity relationship (3D-QSAR) and pharmacophore modeling, additional steps of ligand conformation generation and spatial ligand alignment methods are required (53). The bioactive conformations are typically unknown for substrates, but inherently conformationally constrained molecules can aid in spatial alignment.

Numerous different techniques and algorithms are available and have to different extent been applied in ligand-based transporter modeling, exemplified in Table 1 and in the following sections. Side-by-side comparisons of different modeling approaches are rare in the transporter literature, and the optimal choice of algorithm and way of representing structures typically depends on the dataset. It is thus recommended to evaluate multiple options when starting a transporter modeling project. Importantly, unbiased evaluation of the models' predictive power is essential; guidelines for model evaluation are provided in the sections below.

Descriptor-based statistical models/QSAR

Descriptor-based statistical modeling and QSAR are techniques that use physicochemical descriptors and two-dimensional (2D) connectivity/structure fingerprints to describe and predict the effect of a ligand on the activity of a specific transport protein (Figure 2). The models can vary from simple linear equations to complex multivariate and non-linear statistical (machine learning) models. The physicochemical descriptors that have frequently been associated with ligand-transporter interactions are those that govern non-covalent interactions (lipophilicity, hydrogen bonding, charge, aromaticity) and spatial properties (molecular size, flexibility, polar surface area, atomic connectivity) (21, 54). Beyond the basic physicochemical descriptors, the utility of molecular fingerprints that describe the

structure and connectivity of molecules should be emphasized for their ability to increase the accuracy of models and identify 2D structural ligand motifs that affect transport activity (55).

In a typical use-case of descriptor-based statistical modeling, molecules are empirically tested in transport assays for their ability to modulate the transport of a model substrate or be transported by a specific transporter of interest. Molecular descriptors are then generated for the library of tested molecules and a model correlating the biological measurement to molecular descriptors is generated using statistical modeling algorithms. Algorithms typically used in model generation are multiple linear regression (MLR) (30, 31), partial least-squares (PLS) (21, 34), k-nearest neighbor (kNN) (35, 36), random forest (RF) (56, 57), support vector machines (SVM) (37, 38) and Bayesian classification (40, 41). Different statistical algorithms have advantages and limitations that can make affect their appropriateness for different datasets. For example, MLR and PLS result in models that are linear combinations of the molecular descriptors included. Linearity can facilitate interpretation of the models, for example answering the question of which molecular properties are the most important ones for the observed transporter effect. However, non-linear techniques such as RF and SVM often result in a better model performance.

3D-QSAR

3D-QSAR extends the descriptor-based QSAR statistical models to also consider the spatial 3D molecular structures of ligands and their relationship with transport activity. This technique correlates biological activity to the non-covalent interaction field that surrounds a molecule defined by van der Waals and electrostatic forces. Briefly, 3D-QSAR involves spatially aligning the bioactive or lowest energy molecular conformations of the molecular data set. The electrostatic and steric interactions are then calculated with a probe placed at evenly spaced intervals, typically 1–2 Å apart, in a 3D lattice grid. That spatial interaction information is then correlated to biological activity, typically using a multiple linear regression method such as PLS. The two approaches most commonly used are CoMFA and CoMSIA and readers can refer to this recent review for guidelines and techniques (53). Recent examples of published 3D-QSAR analyses for transporters include MDR-1 and MRP1 (42), BCRP (43), and organic anion transporters (OATs) (44).

Pharmacophore modeling

Ligand-based pharmacophore modeling is another 3D molecular modeling approach and aims to identify the essential pharmacophoric molecular features (hydrogen bond acceptors, hydrogen bond donors, hydrophobic features, and negatively and positively ionizable atoms) of ligands that are necessary for molecular recognition by a target protein. The technique assumes that a set of molecules that share similar biological activity will also share pharmacophoric molecular features in the same superimposable spatial arrangement and these features correspond to the shared protein binding site.

A known bioactive ligand conformation or conformationally constrained ligand serves as a pharmacophore template for ligand alignment. The method then maximizes the number of shared pharmacophore features across the set of active ligands. The ligands are scored by

their ability to fit the pharmacophore; the fit is defined by the number of pharmacophore feature matches and how closely the features overlap. For more in-depth reviews of pharmacophore modeling and examples of models developed for drug transporters, we refer the reader to this selection: (58, 59).

Proteochemometric models

Whereas the previous methods described the relationship between ligands and a single transport protein, proteochemometrics extends these methods to describe the interplay between ligands and multiple proteins. The protein targets are described by physicochemical properties derived from their respective amino acid sequence (60). While typically applied to libraries of mutants, this technique has been used for closely related transporters such as the OATP1B family (10).

Impact of substrate/inhibitor selectivity

The vast majority of the known human transporters are thought to have relatively defined substrate spectra (e.g., transporters of ions, amino acids, sugars, and signaling molecules) (61). In contrast, the subset of ‘drug transporters’ often accept much wider ranges of substrates. Examples include transporters implicated in tumor multidrug resistance (e.g., MDR-1, BCRP and members of the MRP family) that also have physiological roles in protecting tissues from toxins and excreting metabolites, and members of the solute carrier families SLC10, SLC15, SLC22, SLCO and SLC47 (4, 62).

The multidrug transporters, particularly the ABC efflux transporters have large enough binding cavities to accept several ligands simultaneously, and different ligands may interact with different parts of the substrate binding site (18). This has fundamental implications for modeling drug-transporter interactions, since several different submodels may be needed to account for series of molecules that interact differently with the transporter. If the training molecules bind to different sites—despite having the same observed activity—the modeling algorithm will need to account for all different interaction possibilities, typically resulting in models describing the common denominator of these. This applies most directly to pharmacophore-based modeling techniques, which are based on the assumption that structural features that occur in the same spatial arrangement in multiple ligands do so because they interact with the same residues in the binding site. However, modeling approaches that are agnostic to the interaction mechanism—such as descriptor and fingerprint-based statistical modeling—will also have problems with such mechanistically diverse datasets, unless the overall physicochemical properties driving binding are the same for the different sub-sites.

Qualitative and quantitative models

Depending on the intended use of the model and the input data available, models can be either qualitative (substrate–nonsubstrate; inhibitor–noninhibitor) or quantitative (e.g., rate of transport, K_m , IC_{50} , K_i). The choice is typically guided by the nature of the measured data. For example, if the model is based on probe transport inhibition data collected at a single screening concentration (e.g., 13, 21, 37, 40), the training data may not be of sufficient resolution for a quantitative model. In contrast, if the model is based on potency

data (e.g., IC_{50} or K_i), a quantitative model is the natural choice (e.g., 30, 42-44, 46). Notably, two distinct approaches can be used for qualitative modeling from (semi-)quantitative data: either the class labels (e.g., inhibitor–noninhibitor) are assigned prior to the modeling; or, the model is trained directly on the quantitative data, and class labels are assigned based on the predictions post-hoc. The use of discrete or continuous response variables implies different statistical modeling modalities; thus, while the final prediction is qualitative in both approaches, prediction accuracy may differ and the best choice will depend on the application.

Model coverage – local versus global models

Ligand based modeling approaches use information from molecules with known activity (the training set) to make inferences about unknown molecules. The molecules included in the training set are thus an essential part of the model: accurate predictions can only be expected for molecules with similar physical properties or structures as the molecules used for training (i.e., the same ‘chemical space’).

Typically, there is a tradeoff between local prediction accuracy and the coverage of a model: a model that is trained on a specific series of chemically related molecules (i.e., a ‘local’ model) will likely perform better in predicting new molecules in the same series, but have limited predictive power for other chemical scaffolds. In contrast, a model trained on structurally diverse molecules (for example selected from all marketed drugs; i.e., a ‘global’ model) will have greater coverage (a wider ‘applicability domain’), but typically at the expense of decreased local accuracy. Typically, local models are used to guide compound optimization, for example to increase/decrease transport or transporter inhibition in a chemical series. If, instead, the aim is to identify transporter substrates or inhibitors with novel chemical scaffolds, models with wider coverage are preferable (e.g., 10, 22, 33, 36-39).

An intermediate approach, aimed at increasing local coverage, is to select multiple chemically similar molecules at points spread throughout the larger model space, thereby challenging the machine learning algorithm to fine-tune predictions and not equate a certain region of chemical space with a certain activity. Non-linear machine learning techniques such as random forests and neural networks can be advantageous in this respect, while strictly linear techniques may be less suited to capture local variations.

Application domain

Most modeling techniques do not directly address whether or not the predicted molecules are well-described by the model. It is therefore important to assess the model’s applicability domain, to inform the user of how reliable a certain prediction is. Common approaches are to calculate the chemical similarity of the predicted molecules to the training set (36), for example as distances in a simplified, low-dimensional property space obtained by principal component analysis of the model descriptors, or by assessing if the same substructural fingerprints are represented in the training and prediction sets (typically quantified with the Tanimoto coefficient, T_c) (39). Alternatively, distance-to-model metrics calculate the

amount of unexplained variance in the predicted dataset; if this is similar to the unexplained variance in the training data then the prediction set is within the application domain (63).

Evaluating ligand-based model performance

Central in all modeling is the evaluation of model performance. A number of strategies are commonly used to this end. The gold standard is an external test set, i.e., a set of molecules for which the modeled activity is known, but which were not part of the model development (Figure 3a; Table 1). The final model is used to predict the modeled property for the test set molecules, and model performance is assessed by comparing the predictions to the measured activities (Figure 3b). Depending on the type of data modeled different performance metrics are used. For continuous data, the coefficient of determination (R^2) or the root-mean squared error of prediction (RMSE) are common metrics; for categorical data, a number of metrics are used to assess the number of molecules assigned to the correct class, including accuracy (the fraction of correctly classified objects), sensitivity (the fraction of the truly *active* objects that were predicted as active), the specificity (the fraction of the *inactive* objects that were predicted as inactive), and the precision (the fraction of the *predicted actives* that are, in fact, active). Additional metrics such as the Correct Classification Rate (CCR) and Matthew's correlation coefficient (MCC) describe the balance between sensitivity and specificity/precision in a single metric.

As an alternative (or a complement) to a training–test set division, cross-validation is often used to assess model performance (Figure 3c). In cross-validation, the dataset is divided into subsets. In turn, one of these sets is excluded and a model is trained on the remaining data, which is then used to predict the withheld subset; the procedure is repeated until all subsets have been excluded once. Cross-validation results thus represent the prediction performance of the model (13). Leave-one-out (LOO) cross-validation is the extreme variant in which a single molecule is withheld in each round, and models are trained on all remaining compounds; given that a large dataset will often contain some structurally very similar molecules, LOO cross-validation will typically give overly optimistic estimates of model performance (since the predicted molecule will be matched by a similar one in the training data). It is not possible to define a partitioning that is the ‘most relevant’ for all datasets; however, five to ten cross validation subsets are commonly used.

Cross-validation results are often used to guide further model optimization, for example through variable selection: several models are trained using different sets of descriptors (e.g., by iteratively adding or removing variables), and the model with the best cross-validated predictions is retained. While this can be a good way to guide model refinement, it leads to ‘information leakage’. Since the cross-validation predictions are used to optimize the model, these predictions no longer reflect the true external prediction performance. It is therefore important to keep a portion, typically one-third, of the complete dataset outside all model optimization, for use in estimating the performance of the final, optimized model. Alternatively, a ‘double loop’ cross validation procedure can be used, where models are optimized using cross validation as above, and the optimized models are evaluated using an additional, outer cross validation loop so that each cross-validated model is evaluated using molecules that were not used in its training (13).

Once an optimized model is obtained, it is often used to prospectively screen for new ligands by applying it to large compound libraries ('virtual screening'). The performance of a model in virtual screening is typically assessed by different enrichment scores, describing improvements in hit rates compared to random selection from the database (Figure 3d). The top hits in the virtual screen are subjected to experimental testing; if a number of randomly selected molecules from the dataset are also tested, the frequency of true hits in each set can be compared to calculate the enrichment relative to a random selection. Prior to such experiments, enrichment scores can be calculated by seeding the database with known actives, and observing where they appear in the rank-ordered list of hits. Importantly, such seed molecules must not be part of the training data; otherwise, information leakage will lead to artificially high enrichment scores.

Protein Structure-Based Molecular Modeling

Understanding transporter structure, dynamics, and mode of interaction with substrates and inhibitors can help address a variety of fundamental questions in transporter pharmacology. Particularly, protein structure can be used (i) to discover small molecule ligands such as endogenous compounds that deorphanize the transporter function, or approved drugs that guide drug-drug interaction studies and drug repurposing; (ii) to rationally design new drugs or useful chemical tools against therapeutic targets; and (iii) to explain or predict the functional consequence of non-synonymous variants on transporter function, thereby contributing to the understanding of disease mechanisms or differential drug response among patients (Figure 4).

Despite the biological importance of membrane transporters, due to multiple technical challenges, structures of transporters are significantly under-represented in the protein databank (64). For example, atomic structures have been determined experimentally for fewer than 5% of all human SLCs (7, 65). The vast majority of resolved structures are from prokaryotic proteins, which may not always provide suitable templates for eukaryotic transporters due to the absence of post-translational modifications, which may affect protein structure, folding and dynamics. Nonetheless, recent determination of transporter structures from human and other organisms, combined with modern computational power and methods, have led to an increased applicability of computational modeling to transporters (65-67).

Structural modeling

The two major classes of structural modeling include *ab initio* methods that use sequence information alone, without directly using known structures (68). Although *ab initio* modeling can help characterize the structural class of the transporter and even identify functional residues, these techniques are unlikely to generate models sufficiently accurate for key biomedical applications such as rational drug design. The second type of structural modeling is homology modeling, which relies on detectable sequence similarity with at least one experimentally determined structure serving as a modeling template. While homology models can generally address similar basic questions in transporter pharmacology to those studied with experimentally determined structures, homology modeling also have some

limitations. Particularly, the variable regions between the template structure and target protein can be responsible for regulating specific biological functions. For example, the L type amino acid transporter (LAT1; *SLC7A5*), which is typically modeled based on its prokaryotic homolog, the arginine/agmatine transporter AdiC (69), interacts with the single transmembrane protein SLC3A1 to transport substrates *in vivo* – an interaction that does not appear to be conserved with AdiC (70). However, homology modeling is presently the most appropriate approach to model transporter structure for structure-based ligand discovery and is thus the focus of this review.

Key steps in homology modeling comprise template selection, target-template alignment, model building and assessment, and model refinement (71) (Figure 5). Some online services provide automatically generated homology models for the human proteome (72, 73); however, when modeling a human transporter, each step needs to be applied judiciously (65). In brief, the utility of the homology model correlates with the quality of the alignment, which can be measured by the sequence identity shared between the target and template. For example, models that are based on sequence identity of 30% or less can be used to visualize the location of mutated residues and rationalize their function, but they are less likely to be relevant for the characterization of catalytic mechanisms and rational drug design (68). Notably, because of the limited number of template structures, transporter models are often generated based on relatively low levels of sequence similarity to the template structure (sequence identity of 30% or even less), limiting the potential applications of these models.

Modeling challenging targets can be done by using multiple sources of information derived from different experimental methodologies (74). For example, simple hybrid or integrative modeling approaches involve fitting homology models into a density map determined by cryogenic electron microscopy (cryoEM), which can also capture complexes involving membrane transporters (e.g., LAT2/4F2hc) (75). In recent years, multiple computational methods have been introduced that are specifically designed for membrane proteins, improving the performance of each modeling step. For instance, the alignment program AlignMe takes into account the target's and template's hydropathy profiles to generate alignments that are less likely to introduce gaps within the transmembrane helices (76). RosettaMembrane has been optimized to model and assess membrane protein structures relying on distant sequence similarity, capturing unique conformations of transmembrane helices and the loops connecting them (77).

Protein dynamics

Membrane transporters are dynamic proteins for which different conformational states are required for their transport activity. Transporters constitute distinct structural classes, which use different transport mechanisms and distinct energy coupling mechanisms, and they also transport substrates at a range of timescales (7, 61) (Figure 4). Thus, modeling transporter dynamics is a highly active research area. Distinct conformations of the transporters of interest can be modeled with homology modeling, relying on template structures that capture different snapshots of the transport cycle (78). Interestingly, many transporters have an internal pseudo-symmetry where the structure consists of two structurally similar halves. This fact can be utilized to efficiently derive alternative conformations (79), and has been

used to model transporters belonging to a variety of families and folds (80) (Figure 4). For example, the *Vibrio Cholerae* Na⁺/succinate transporter VcINDY structure was determined in the inward conformation (81), and has been modeled in the outward conformation using this technique (Figure 4) (82), providing a template for modeling the human homologs of the SLC13 family in this conformation (83).

Perhaps the most common approach to visualize transporter dynamics at an atomic level is molecular dynamics (MD) simulations. MD simulations compute the position of the atoms in a biomolecular system as a function of time, using Newtonian laws of motion with atomic interactions described by a specific force-field (84). The large size of many transporters (e.g., 60–80 kDa for prototypical drug-transporting SLC transporters, and 140–175 kDa for ABC efflux transporters), combined with the relatively long timescales of the transport process (e.g., μ s–ms–s) and the demanding computational requirements of MD simulations, makes it difficult to describe complete transport cycles at a reasonable computing time. Recently, superior computer architectures and improved force fields have enabled researchers to capture a full cycle for a small bacterial transporter related to the human sugar transporter SWEET (85). Alternatively, state-of-the-art efficient simulation approaches such as metadynamics, umbrella sampling, and steered MD—each providing different ways of capturing local energy barriers—have enabled simulations of conformational changes in larger transporters, as well as those whose timescales are still unreachable to unguided MD simulations (66, 67, 84). Notably, applying MD simulations to characterize human transporters remains challenging due to the limited number of known atomic human transporters structures and the low resolution of homology models.

Ligand discovery

In molecular docking or virtual screening, organic molecules are sampled in multiple configurations, docked to the protein structure, which can either be experimentally determined or modeled, and scored based on their complementarity to the binding site (86). Docking algorithms use a range of approaches from physics, statistics, and machine learning, and have become important tools in drug discovery over the past decades (87).

Several considerations should be taken when using docking in transporter studies. First, one weakness of docking is that due to various approximations made (e.g., with estimating desolvation energy), it cannot accurately predict ligand binding affinity; however, docking's ability to efficiently identify putative binders from large compound libraries—which can contain millions of molecules—is a major strength, since this allows the user to prioritize, and manually analyze only the top-scoring molecules (86).

Second, because of the limited structural coverage of membrane transporters, models are often not of sufficient resolution for accurate ligand docking. One way to provide a useful control that mitigates this shortcoming is to estimate the potential relevance of the model for virtual screening with ligand enrichment calculations (see also ligand-based modeling above) (Figure 3d). Specifically, known ligands and decoy compounds are docked against the model's binding site and the enrichment of the known ligands in the whole dataset is calculated (88, 89). Moreover, by generating multiple models and selecting the final model based on its enrichment score, models are often optimized for ligand discovery (65, 90–92).

Third, due to inaccuracies in the various computational methods, experimental testing of a few candidate compounds is needed to ensure that the model is predictive. Top-ranked compounds are visually analyzed to eliminate the molecules with energetically unfavorable or strained conformations commonly found in large computational screens (86).

Fourth, the modeler should consider removing potentially problematic compounds from the virtual library (24, 25). Specifically, before docking, the small-molecule library can be analyzed with PAINS filters as discussed under ‘Experimental data and implications for modeling’ above (See also Supplemental Reading List). This will help to identify compounds frequently giving false-positive results in large screens due to reactivity under assay conditions that is unrelated to the intended binding to the transporter: such confounding effects include covalent binding, redox effects, autofluorescence, or aggregation. While the docking procedure itself is not affected by such assay artifacts, the downstream verification of predicted transporter inhibitors can be problematic if these contain assay-interfering structural features. Notably however, there are two issues with blindly removing predicted PAINS: (i) the computational filters are not always accurate and should be followed up with experimental testing once active compounds have been confirmed experimentally; and (ii) the activity of many PAINS substructures can depend on the structural context of the molecule or the molecular environment of the assay, and even compounds that are flagged under certain conditions may still be worthwhile pursuing (86, 93). For example, commonly prescribed cancer drugs, such as fulvestrant, lapatinib, and sorafenib can form colloidal aggregates in some cellular assays (94). Therefore, multiple orthogonal assays should be used to confidently assign biologically relevant transporter inhibitors and substrates in discovery campaigns.

Ligand optimization

One step in generating useful tool compounds or lead molecules for drug development involves the optimization of initial hits through the development of structure–activity relationship (SAR) models. SAR models can be developed by iterative computational modeling and experimental testing of analogs, combined with medicinal chemistry reasoning. Structure-guided ligand optimization commonly involves the application of multiple MD simulations (e.g., 50–100ns) to assess the compound-transporter interactions of top hits visually, as well as to estimate the binding free energy of the complex with MM/PBSA (95) or more advanced techniques such as free energy perturbation (FEP) (96) or thermodynamic integration (97). Notably, ligand optimization may require precise description of the binding site atoms, including sidechains, ions, and water molecules; thus, it is often challenging—or even impossible—to obtain accurate enough results from homology models or even low-resolution experimentally determined structures. Notably, the ligand affinity (predicted or observed) does not always translate to being a substrate that gets transported by the transporter, and obtaining the affinities of the ligand in multiple transporter conformations may enable computational prediction of substrates with structure-based approaches (78).

Mutation effects

Non-synonymous variants can have significant effects on key protein features, such as hydrogen bond networks, conformational dynamics, stability, and interactions with other molecules (98). In some cases, the mutation effect on function can be rationalized by visualization of the mutated residue on the protein structure. Specifically, if a mutation involves dramatic change in the biophysical properties of the substrate binding site, it can be straightforward to determine the potential functional consequences (99). For example, some epilepsy mutations in NaCT (SLC13A5) can occur in or near the ion or substrate binding site, directly affecting transport function (Figure 4) (99). For other cases, more sophisticated methods are employed. For instance, some machine learning algorithms are trained on biophysical features (predicted or observed), such as secondary structure elements, protein flexibility, and residue conservation, directly estimating the functional consequence of the mutation (100-102). Physics-based techniques estimate the folding free energy (ΔG), providing a quantitative prediction of the mutation's effect on protein stability and dynamics based on MD simulations of the wild type and mutant proteins (103).

Future of Transporter Modeling

The accuracy of membrane transporter models depends on the abundance and quality of the experimental data used for modeling. Recent technical breakthroughs in a variety of assay technologies have enabled the generation of vast amounts of data, thereby facilitating significant increase in transporter modeling studies.

For example, advances in mass spectrometry have enabled generation of larger datasets for transporter *substrates* rather than merely *inhibitors*, while an increased understanding of transporter kinetics and the way confounding factors (e.g. interference of passive permeability, allosteric binding sites) are handled has led to more accurate data analysis resulting in increased consistency of data treatment across laboratories (52). In turn, these developments have increased the fidelity of ligand-based models due to their intrinsic dependency on robust data sets. As analyzing transporter data and generating datasets continues to increase in resolution, we can expect to see a simultaneous enhancement in the quality and performance of ligand-based models. In addition to greater confidence in data treatment, improved modeling algorithms have resulted in ligand-based models with greater accuracy and precision. Ultimately, global models that take into consideration multiple local datasets will expand our understanding of allosteric transport and inhibition sites, as well as determine ligand-specific transporter interactions.

Improved computer hardware, computational methods, and an increased availability of diverse types of omics data are continually improving both ligand- and structure-based modeling. The near future can be expected to see an increased use of integrative modeling techniques that combine multiple types of data and modeling algorithms (e.g. combining homology-based and ligand-based models) to allow greater understanding of transporter structure–function relationships and kinetic mechanisms. New developments of these technologies are expected as the number and coverage of datasets and eukaryotic transporter structures become increasingly available.

In the area of structure-based modeling, advances in cryogenic electron microscopy (cryoEM) have led to much improved resolution for macromolecular structures solved using this approach, as well as to the expansion of its application to challenging membrane protein targets (104). For example, recently determined structures of membrane-bound receptors, ion channels, and membrane transporters, have provided important biological discoveries and an experimental basis for biomedical modeling studies. While solving structures with cryoEM technologies is still challenging for many important transporter targets, these technologies are advancing rapidly and are expected to facilitate significant increase in the coverage of the transporter structural space in the near future. This will open up the scene for modeling of the many currently uncharacterized transporters, and will thus be of profound importance for our understanding of cellular drug and solute transport.

Conclusion

Molecular modeling of drug–transporter interactions is increasingly used to generate new hypotheses, rationalize experimental observations, and to prospectively screen for potential transporter-mediated drug–drug interactions and for new ligands of therapeutically interesting transporters. Here we summarize the methods most commonly used in ligand-based and structure-based modeling of drug–transporter interactions, and, where possible, provide guidance on best practices. We stress the importance of thoroughly evaluating model performance, and of understanding not only the strengths and limitations of the computational and statistical methods applied but also those of the experimental assays used to generate data that the models rely on. Finally, we discuss technological advances that will significantly boost the amount and quality of data available for modeling and the algorithms used to analyze them. In combination with such advances, the emerging knowledge on the function of transporters, the increased interest in selective tissue targeting, and the higher potential for DDIs in drug combination regimens, give predictive modeling of transporters an ever more important role in pharmaceutical drug discovery and development, as well as in the understanding of fundamental transporter biology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This white paper is based on the session held at the 3rd International Transporter Consortium (ITC) Workshop – Transporters in Drug Development ASCPT Pre-conference, 13-14th March 2017.

Funding

This work was supported in part by grants from the National Institutes of Health (R01 GM108911 to A.S. and R01 DK56631 to P.S.) and the Åke Wiberg Foundation (P.M.).

References

- (1). Adibi SA The oligopeptide transporter (Pept-1) in human intestine: biology and function. *Gastroenterology* 113, 332–40 (1997). [PubMed: 9207295]

- (2). Bradshaw DM & Arceci RJ Clinical relevance of transmembrane drug efflux as a mechanism of multidrug resistance. *J Clin Oncol* 16, 3674–90 (1998). [PubMed: 9817290]
- (3). Reith ME, Xu C & Chen NH Pharmacology and regulation of the neuronal dopamine transporter. *Eur J Pharmacol* 324, 1–10 (1997). [PubMed: 9137907]
- (4). International Transporter, C. et al. Membrane transporters in drug development. *Nat Rev Drug Discov* 9, 215–36 (2010). [PubMed: 20190787]
- (5). Han H et al. 5'-Amino acid esters of antiviral nucleosides, acyclovir, and AZT are absorbed by the intestinal PEPT1 peptide transporter. *Pharm Res* 15, 1154–9 (1998). [PubMed: 9706043]
- (6). Kramer W et al. Liver-specific drug targeting by coupling to bile acids. *J Biol Chem* 267, 18598–604 (1992). [PubMed: 1526993]
- (7). Colas C, Ung PM & Schlessinger A SLC Transporters: Structure, Function, and Drug Discovery. *Med Chem Comm* 7, 1069–81 (2016).
- (8). Locher KP Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat Struct Mol Biol* 23, 487–93 (2016). [PubMed: 27273632]
- (9). Raub TJ P-glycoprotein recognition of substrates and circumvention through rational drug design. *Mol Pharm* 3, 3–25 (2006). [PubMed: 16686365]
- (10). De Bruyn T et al. Structure-based identification of OATP1B1/3 inhibitors. *Mol Pharmacol* 83, 1257–67 (2013). [PubMed: 23571415]
- (11). Severance AC, Sandoval PJ & Wright SH Correlation between Apparent Substrate Affinity and OCT2 Transport Turnover. *J Pharmacol Exp Ther* 362, 405–12 (2017). [PubMed: 28615288]
- (12). Vig BS et al. Human PEPT1 pharmacophore distinguishes between dipeptide transport and binding. *J Med Chem* 49, 3636–44 (2006). [PubMed: 16759105]
- (13). Kido Y, Matsson P & Giacomini KM Profiling of a prescription drug library for potential renal drug-drug interactions mediated by the organic cation transporter 2. *J Med Chem* 54, 4548–58 (2011). [PubMed: 21599003]
- (14). Ho HTB, Pan Y, Cui Z, Duan H, Swaan PW & Wang J Molecular analysis and structure-activity relationship modeling of the substrate/inhibitor interaction site of plasma membrane monoamine transporter. *J Pharmacol Exp Ther* 339, 376–85 (2011). [PubMed: 21816955]
- (15). Ekins S & Swaan PW Development of computational models for enzymes, transporters, channels and receptors relevant to ADME/TOX. *Rev Comp Chem* 20, 333–415 (2004).
- (16). Stouch TR & Gudmundsson O Progress in understanding the structure-activity relationships of P-glycoprotein. *Adv Drug Del Rev* 54, 315–28 (2002).
- (17). Ho H et al. Structural basis for dual-mode inhibition of the ABC transporter MsbA. *Nature*, (2018).
- (18). Aller SG et al. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* 323, 1718–22 (2009). [PubMed: 19325113]
- (19). Belzer M, Morales M, Jagadish B, Mash EA & Wright SH Substrate-dependent ligand inhibition of the human organic cation transporter OCT2. *J Pharmacol Exp Ther* 346, 300–10 (2013). [PubMed: 23709117]
- (20). Martinez-Guerrero LJ & Wright SH Substrate-dependent inhibition of human MATE1 by cationic ionic liquids. *J Pharmacol Exp Ther* 346, 495–503 (2013). [PubMed: 23785176]
- (21). Matsson P, Pedersen JM, Norinder U, Bergstrom CA & Artursson P Identification of novel specific and general inhibitors of the three major human ATP-binding cassette transporters P-gp, BCRP and MRP2 among registered drugs. *Pharm Res* 26, 1816–31 (2009). [PubMed: 19421845]
- (22). Wittwer MB et al. Discovery of potent, selective multidrug and toxin extrusion transporter 1 (MATE1, SLC47A1) inhibitors through prescription drug profiling and computational modeling. *J Med Chem* 56, 781–95 (2013). [PubMed: 23241029]
- (23). Powell DJ, Hertzberg RP & Macarromicronn R Design and Implementation of High-Throughput Screening Assays. *Methods Mol Biol* 1439, 1–32 (2016). [PubMed: 27316985]
- (24). Irwin JJ et al. An Aggregation Advisor for Ligand Discovery. *J Med Chem* 58, 7076–87 (2015). [PubMed: 26295373]

- (25). Baell JB & Holloway GA New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* 53, 2719–40 (2010). [PubMed: 20131845]
- (26). Sugano K et al. Coexistence of passive and carrier-mediated processes in drug transport. *Nat Rev Drug Discov* 9, 597–614 (2010). [PubMed: 20671764]
- (27). Karlgren M et al. A CRISPR-Cas9 Generated MDCK Cell Line Expressing Human MDR1 Without Endogenous Canine MDR1 (cABCB1): An Improved Tool for Drug Efflux Studies. *J Pharm Sci* 106, 2909–13 (2017). [PubMed: 28450237]
- (28). Simoff I et al. Complete Knockout of Endogenous Mdr1 (Abcb1) in MDCK Cells by CRISPR-Cas9. *J Pharm Sci* 105, 1017–21 (2016). [PubMed: 26869442]
- (29). Winter GE et al. The solute carrier SLC35F2 enables YM155-mediated DNA damage toxicity. *Nat Chem Biol* 10, 768–73 (2014). [PubMed: 25064833]
- (30). Ecker G et al. Structure-activity relationship studies on benzofuran analogs of propafenone-type modulators of tumor cell multidrug resistance. *J Med Chem* 39, 4767–74 (1996). [PubMed: 8941391]
- (31). Hirano H et al. High-speed screening and QSAR analysis of human ATP-binding cassette transporter ABCB11 (bile salt export pump) to predict drug-induced intrahepatic cholestasis. *Mol Pharm* 3, 252–65 (2006). [PubMed: 16749857]
- (32). Gantner ME, Di Ianni ME, Ruiz ME, Talevi A & Bruno-Blanch LE Development of conformation independent computational models for the early recognition of breast cancer resistance protein substrates. *Biomed Res Int* 2013, 863592 (2013). [PubMed: 23984415]
- (33). Broccatelli F et al. A novel approach for predicting P-glycoprotein (ABCB1) inhibition using molecular interaction fields. *J Med Chem* 54, 1740–51 (2011). [PubMed: 21341745]
- (34). Pedersen JM et al. Early identification of clinically relevant drug interactions with the human bile salt export pump (BSEP/ABCB11). *Toxicol Sci* 136, 328–43 (2013). [PubMed: 24014644]
- (35). Broccatelli F QSAR models for P-glycoprotein transport based on a highly consistent data set. *J Chem Inf Model* 52, 2462–70 (2012). [PubMed: 22946765]
- (36). Sedykh A et al. Human intestinal transporter database: QSAR modeling and virtual profiling of drug uptake, efflux and interactions. *Pharm Res* 30, 996–1007 (2013). [PubMed: 23269503]
- (37). Tan W et al. Combined QSAR and molecule docking studies on predicting P-glycoprotein inhibitors. *J Comput Aided Mol Des* 27, 1067–73 (2013). [PubMed: 24322389]
- (38). Warner DJ et al. Mitigating the inhibition of human bile salt export pump by drugs: opportunities provided by physicochemical property modulation, in silico modeling, and structural modification. *Drug Metab Dispos* 40, 2332–41 (2012). [PubMed: 22961681]
- (39). Chen EC et al. Discovery of Competitive and Noncompetitive Ligands of the Organic Cation Transporter 1 (OCT1; SLC22A1). *J Med Chem* 60, 2685–96 (2017). [PubMed: 28230985]
- (40). Welch MA, Kock K, Urban TJ, Brouwer KL & Swaan PW Toward predicting drug-induced liver injury: parallel computational approaches to identify multidrug resistance protein 4 and bile salt export pump inhibitors. *Drug Metab Dispos* 43, 725–34 (2015). [PubMed: 25735837]
- (41). Pan Y, Chothe PP & Swaan PW Identification of novel breast cancer resistance protein (BCRP) inhibitors by virtual screening. *Mol Pharm* 10, 1236–48 (2013). [PubMed: 23418667]
- (42). Pajeva IK, Globisch C & Wiese M Combined pharmacophore modeling, docking, and 3D QSAR studies of ABCB1 and ABCC1 transporter inhibitors. *ChemMedChem* 4, 1883–96 (2009). [PubMed: 19768722]
- (43). Pick A et al. Structure-activity relationships of flavonoids as inhibitors of breast cancer resistance protein (BCRP). *Bioorg Med Chem* 19, 2090–102 (2011). [PubMed: 21354800]
- (44). Kaler G et al. Structural variation governs substrate specificity for organic anion transporter (OAT) homologs. Potential remote sensing by OAT family members. *J Biol Chem* 282, 23841–53 (2007). [PubMed: 17553798]
- (45). Palmeira A, Rodrigues F, Sousa E, Pinto M, Vasconcelos MH & Fernandes MX New uses for old drugs: pharmacophore-based screening for the discovery of P-glycoprotein inhibitors. *Chem Biol Drug Des* 78, 57–72 (2011). [PubMed: 21235729]

- (46). Dong Z, Ekins S & Polli JE Structure-activity relationship for FDA approved drugs as inhibitors of the human sodium taurocholate cotransporting polypeptide (NTCP). *Mol Pharm* 10, 1008–19 (2013). [PubMed: 23339484]
- (47). Hayeshi R et al. Comparison of drug transporter gene expression and functionality in Caco-2 cells from 10 different laboratories. *Eur J Pharm Sci* 35, 383–96 (2008). [PubMed: 18782614]
- (48). Liu HC et al. Molecular Properties of Drugs Interacting with SLC22 Transporters OAT1, OAT3, OCT1, and OCT2: A Machine-Learning Approach. *J Pharmacol Exp Ther* 359, 215–29 (2016). [PubMed: 27488918]
- (49). Mateus A, Treyer A, Wegler C, Karlgren M, Matsson P & Artursson P Intracellular drug bioavailability: a new predictor of system dependent drug disposition. *Sci Rep* 7, 43047 (2017). [PubMed: 28225057]
- (50). Tachibana T et al. Model analysis of the concentration-dependent permeability of P-gp substrates. *Pharm Res* 27, 442–6 (2010). [PubMed: 20135207]
- (51). Korzekwa K & Nagar S Compartmental Models for Apical Efflux by P-glycoprotein: Part 2-A Theoretical Study on Transporter Kinetic Parameters. *Pharm Res* 31, 335–46 (2014). [PubMed: 23959852]
- (52). Zamek-Gliszczynski MJ et al. ITC recommendations for transporter kinetic parameter estimation and translational modeling of transport-mediated PK and DDIs in humans. *Clin Pharmacol Ther* 94, 64–79 (2013). [PubMed: 23588311]
- (53). Verma J, Khedkar VM & Coutinho EC 3D-QSAR in drug design--a review. *Curr Top Med Chem* 10, 95–115 (2010). [PubMed: 19929826]
- (54). Pedersen JM, Matsson P, Bergstrom CA, Norinder U, Hoogstraate J & Artursson P Prediction and identification of drug interactions with the human ATP-binding cassette transporter multidrug-resistance associated protein 2 (MRP2; ABCC2). *J Med Chem* 51, 3275–87 (2008). [PubMed: 18457386]
- (55). Chen L, Li Y, Zhao Q, Peng H & Hou T ADME evaluation in drug discovery. 10. Predictions of P-glycoprotein inhibitors using recursive partitioning and naive Bayesian classification techniques. *Mol Pharm* 8, 889–900 (2011). [PubMed: 21413792]
- (56). Klepsch F, Vasanathan P & Ecker GF Ligand and structure-based classification models for prediction of P-glycoprotein inhibitors. *J Chem Inf Model* 54, 218–29 (2014). [PubMed: 24050383]
- (57). Pinto M, Trauner M & Ecker GF An In Silico Classification Model for Putative ABCC2 Substrates. *Mol Inform* 31, 547–53 (2012). [PubMed: 23198001]
- (58). Leach AR, Gillet VJ, Lewis RA & Taylor R Three-dimensional pharmacophore methods in drug discovery. *J Med Chem* 53, 539–58 (2010). [PubMed: 19831387]
- (59). Chang C, Ekins S, Bahadduri P & Swaan PW Pharmacophore-based discovery of ligands for drug transporters. *Adv Drug Deliv Rev* 58, 1431–50 (2006). [PubMed: 17097188]
- (60). Lapinsh M, Prusis P, Gutcaits A, Lundstedt T & Wikberg JE Development of proteo-chemometrics: a novel technology for the analysis of drug-receptor interactions. *Biochim Biophys Acta* 1525, 180–90 (2001). [PubMed: 11342268]
- (61). Schlessinger A et al. Comparison of human solute carriers. *Protein Sci* 19, 412–28 (2010). [PubMed: 20052679]
- (62). Hillgren KM et al. Emerging transporters of clinical importance: an update from the International Transporter Consortium. *Clin Pharmacol Ther* 94, 52–63 (2013). [PubMed: 23588305]
- (63). Sushko I et al. Applicability domains for classification problems: Benchmarking of distance to models for Ames mutagenicity set. *J Chem Inf Model* 50, 2094–111 (2010). [PubMed: 21033656]
- (64). Cesar-Razquin A et al. A Call for Systematic Research on Solute Carriers. *Cell* 162, 478–87 (2015). [PubMed: 26232220]
- (65). Schlessinger A, Khuri N, Giacomini KM & Sali A Molecular Modeling and Ligand Docking for Solute Carrier (SLC) Transporters. *Curr Top Med Chem* 13, 843–56 (2013). [PubMed: 23578028]
- (66). Faraldo-Gomez JD & Forrest LR Modeling and simulation of ion-coupled and ATP-driven membrane proteins. *Curr Opin Struct Biol* 21, 173–9 (2011). [PubMed: 21333528]

- (67). Li J, Wen PC, Moradi M & Tajkhorshid E Computational characterization of structural dynamics underlying function in active membrane transporters. *Curr Opin Struct Biol* 31, 96–105 (2015). [PubMed: 25913536]
- (68). Baker D & Sali A Protein structure prediction and structural genomics. *Science* 294, 93–6 (2001). [PubMed: 11588250]
- (69). Singh N & Ecker GF Insights into the Structure, Function, and Ligand Discovery of the Large Neutral Amino Acid Transporter 1, LAT1. *Int J Mol Sci* 19, (2018).
- (70). Fotiadis D, Kanai Y & Palacin M The SLC3 and SLC7 families of amino acid transporters. *Mol Aspects Med* 34, 139–58 (2013). [PubMed: 23506863]
- (71). Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F & Sali A Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29, 291–325 (2000). [PubMed: 10940251]
- (72). Kiefer F, Arnold K, Kunzli M, Bordoli L & Schwede T The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res* 37, D387–92 (2009). [PubMed: 18931379]
- (73). Pieper U et al. ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res* 39, D465–74 (2011). [PubMed: 21097780]
- (74). Lasker K et al. Integrative structure modeling of macromolecular assemblies from proteomics data. *Mol Cell Proteom* 9, 1689–702 (2010).
- (75). Rosell A et al. Structural bases for the interaction and stabilization of the human amino acid transporter LAT2 with its ancillary protein 4F2hc. *Proc Natl Acad Sci USA* 111, 2966–71 (2014). [PubMed: 24516142]
- (76). Stamm M, Staritzbichler R, Khafizov K & Forrest LR AlignMe--a membrane protein sequence alignment web server. *Nucleic Acids Res* 42, W246–51 (2014). [PubMed: 24753425]
- (77). Chen KY, Sun J, Salvo JS, Baker D & Barth P High-resolution modeling of transmembrane helical protein structures from distant homologues. *PLoS Comput Biol* 10, e1003636 (2014). [PubMed: 24854015]
- (78). Colas C et al. Chemical Modulation of the Human Oligopeptide Transporter 1, hPepT1. *Mol Pharm* 14, 4685–93 (2017). [PubMed: 29111754]
- (79). Forrest LR et al. Mechanism for alternating access in neurotransmitter transporters. *Proc Natl Acad Sci USA* 105, 10338–43 (2008). [PubMed: 18647834]
- (80). Forrest LR Structural Symmetry in Membrane Proteins. *Ann Rev Biophys* 44, 311–37 (2015). [PubMed: 26098517]
- (81). Mancusso R, Gregorio GG, Liu Q & Wang DN Structure and mechanism of a bacterial sodium-dependent dicarboxylate transporter. *Nature* 491, 622–6 (2012). [PubMed: 23086149]
- (82). Mulligan C et al. The bacterial dicarboxylate transporter VcINDY uses a two-domain elevator-type mechanism. *Nat Struct Mol Biol* 23, 256–63 (2016). [PubMed: 26828963]
- (83). Colas C, Schlessinger A & Pajor AM Mapping Functionally Important Residues in the Na(+)/Dicarboxylate Cotransporter, NaDC1. *Biochemistry* 56, 4432–41 (2017). [PubMed: 28731330]
- (84). Lindahl E & Sansom MS Membrane proteins: molecular dynamics simulations. *Curr Opin Struct Biol* 18, 425–31 (2008). [PubMed: 18406600]
- (85). Latorraca NR, Fastman NM, Venkatakrishnan AJ, Frommer WB, Dror RO & Feng L Mechanism of Substrate Translocation in an Alternating Access Transporter. *Cell* 169, 96–107 e12 (2017). [PubMed: 28340354]
- (86). Irwin JJ & Shoichet BK Docking Screens for Novel Ligands Conferring New Biology. *J Med Chem* 59, 4103–20 (2016). [PubMed: 26913380]
- (87). Kitchen DB, Decornez H, Furr JR & Bajorath J Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 3, 935–49 (2004). [PubMed: 15520816]
- (88). Fan H, Irwin JJ, Webb BM, Klebe G, Shoichet BK & Sali A Molecular docking screens using comparative models of proteins. *J Chem Inf Model* 49, 2512–27 (2009). [PubMed: 19845314]
- (89). Schlessinger A et al. Structure-based discovery of prescription drugs that interact with the norepinephrine transporter, NET. *Proc Natl Acad Sci USA* 108, 15810–5 (2011). [PubMed: 21885739]

- (90). Evers A, Gohlke H & Klebe G Ligand-supported homology modelling of protein binding-sites using knowledge-based potentials. *J Mol Biol* 334, 327–45 (2003). [PubMed: 14607122]
- (91). Cavasotto CN et al. Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. *J Med Chem* 51, 581–8 (2008). [PubMed: 18198821]
- (92). Carlsson J et al. Ligand discovery from a dopamine D3 receptor homology model and crystal structure. *Nat Chem Biol* 7, 769–78 (2011). [PubMed: 21926995]
- (93). Jasial S, Hu Y & Bajorath J How Frequently Are Pan-Assay Interference Compounds Active? Large-Scale Analysis of Screening Data Reveals Diverse Activity Profiles, Low Global Hit Frequency, and Many Consistently Inactive Compounds. *J Med Chem* 60, 3879–86 (2017). [PubMed: 28421750]
- (94). Owen SC, Doak AK, Wassam P, Shoichet MS & Shoichet BK Colloidal aggregation affects the efficacy of anticancer drugs in cell culture. *ACS Chem Biol* 7, 1429–35 (2012). [PubMed: 22625864]
- (95). Case DA et al. The Amber biomolecular simulation programs. *J Comp Chem* 26, 1668–88 (2005). [PubMed: 16200636]
- (96). Cappel D et al. Relative Binding Free Energy Calculations Applied to Protein Homology Models. *J Chem Inf Model* 56, 2388–400 (2016). [PubMed: 28024402]
- (97). Samsudin F, Parker JL, Sansom MSP, Newstead S & Fowler PW Accurate Prediction of Ligand Affinities for a Proton-Dependent Oligopeptide Transporter. *Cell Chem Biol* 23, 299–309 (2016). [PubMed: 27028887]
- (98). Wang Z & Moulton J SNPs, protein structure, and disease. *Hum Mut* 17, 263–70 (2001). [PubMed: 11295823]
- (99). Klotz J, Porter BE, Colas C, Schlessinger A & Pajor AM Mutations in the Na(+)/citrate cotransporter NaCT (SLC13A5) in pediatric patients with epilepsy and developmental delay. *Mol Med* 22, (2016).
- (100). Bromberg Y & Rost B SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res* 35, 3823–35 (2007). [PubMed: 17526529]
- (101). Yachdav G et al. PredictProtein--an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res* 42, W337–43 (2014). [PubMed: 24799431]
- (102). Kumar P, Henikoff S & Ng PC Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* 4, 1073–81 (2009). [PubMed: 19561590]
- (103). Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F & Serrano L The FoldX web server: an online force field. *Nucleic Acids Res* 33, W382–8 (2005). [PubMed: 15980494]
- (104). Fernandez-Leiro R & Scheres SH Unravelling biological macromolecules with cryo-electron microscopy. *Nature* 537, 339–46 (2016). [PubMed: 27629640]
- (105). Coleman JA, Green EM & Gouaux E X-ray structures and mechanism of the human serotonin transporter. *Nature* 532, 334–9 (2016). [PubMed: 27049939]
- (106). Deng D et al. Crystal structure of the human glucose transporter GLUT1. *Nature* 510, 121–5 (2014). [PubMed: 24847886]
- (107). Canul-Tec JC et al. Structure and allosteric inhibition of excitatory amino acid transporter 1. *Nature* 544, 446–51 (2017). [PubMed: 28424515]
- (108). Taylor NMI, Manolaridis I, Jackson SM, Kowal J, Stahlberg H & Locher KP Structure of the human multidrug transporter ABCG2. *Nature* 546, 504–9 (2017). [PubMed: 28554189]
- (109). Colas C, Pajor AM & Schlessinger A Structure-Based Identification of Inhibitors for the SLC13 Family of Na(+)/Dicarboxylate Cotransporters. *Biochemistry* 54, 4900–8 (2015). [PubMed: 26176240]

a Molecular descriptors

Size, hydrophobicity, polarity, hydrogen bonding, connectivity/topology, electronics etc

b Molecular fingerprints **c** 3D-QSAR / Molecular interaction fields **d** Pharmacophores **Figure 1. Ligand based modeling modalities.**

In descriptor-based modeling (QSAR), numerical representations of ligand structure and/or molecular properties are used as input in a statistical model that relates these to the experimentally determined transporter parameter. The descriptors can, for example, represent molecular properties such as molecular weight, lipophilicity, polarity or charge, or describe the connectivity of atoms in the ligand molecules (**a**). In fingerprint descriptors (**b**) the molecular structure is described as a binary bit string encoding the presence or absence of atoms with a certain local neighborhood (left) or of certain substructural fragments (right). In 3D-QSAR (**c**), ligands are aligned and their interactions with the transporter are described by an interaction field, calculated by placing interaction probes with certain chemical properties in a grid around the ligands. Pharmacophore models (**d**) are defined by aligning pharmacophoric functionalities in a set of ligands. A common spatial arrangement

of such motives in several known binders indicates their importance for ligand–transporter binding.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2. Ligand based modeling procedure.

Ligand-based methods predict drug-transporter interactions based on the molecular properties of the ligands. (a) The selection of example molecules and the type of transporter data to model (e.g., inhibition or substrate transport; continuous— $IC_{50}/K_i/K_m$ —or discrete parameters—substrate/nonsubstrate, inhibitor/noninhibitor) are key steps in the development of ligand-based models, since they will define the application domain of the model. (b) If the mode of interaction is known for the training set (e.g., interactions with a specific binding site), separate models can be developed for each group of ligands. Common ligand-based methods include descriptor-based statistical modeling (c), and pharmacophore and 3D-QSAR modeling (d). In the former, multivariate statistical methods are used to relate the measured transporter interaction with a numerical representation of ligand structure/molecular properties. In the latter techniques, 3D structures of the ligands are aligned, and the spatial arrangement of potentially binding molecular features is assessed. Models are

iteratively optimized, typically by assessing cross-validated prediction errors, and the performance of the optimized models is evaluated using a withheld subset of the data (i.e., the test set, typically $\sim 1/3$) (**e**). Acceptable models can be combined to generate synergistic consensus models (**f**). The final models provide insight into the molecular interactions between ligands and the transporter binding site and can, after assessing the applicability domain of the model, be used for prospective screening for new transporter ligands (**g**).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3. Evaluation of model performance.

The external performance of ligand- and structure-based models should be evaluated prior to their use in prospective virtual screening. In ligand-based modeling, this is typically done using a training-test set division (**a**), in which part of the dataset is withheld from the model training, and only used to assess the final model's predictivity. Typically, the model is trained on ~2/3 of the data (black/orange), and the model is used to predict the modeled parameter for the remaining 1/3 (dark/light blue). (**b**) Model performance is assessed by calculating e.g., the root-mean squared error of prediction (for models of continuous data; left panel) or the prediction accuracy (for categorical data; right panel). Cross-validation (**c**) is a complementary procedure to training-test set division. It is performed by dividing the full dataset into subsets, and iteratively training models based on part of the data while withholding one subset at a time for model evaluation. The process is repeated until each subset has been withheld (and predicted) once. Model performance statistics are then calculated by comparing the measured data with the predictions. (**d**) Ligand discovery

models are typically evaluated by docking a library of known ligands mixed with molecular decoys, i.e., molecules with similar physicochemical properties as the ligands, but structurally different from these. The enrichment of the known ligands among the top hits in the docking is then visualized in enrichment plots and the area under the curve is calculated (AUC). In an optimal docking experiments, the known ligands will be identified at the top of the list of ranked hits, resulting in high hit enrichment (orange dotted line) compared to a random selection of molecules from the entire screening database (black dashed line).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4. Transporter structure and function.

(a) Four recently determined structures of human transporters, with the gene name in parenthesis. The structures include the solute carriers serotonin transporter (SERT, *SLC6A4*) (105), the glucose transporter 1 (GLUT1, *SLC2A1*) (106), and the excitatory amino acid transporter 1 (EAAT1, *SLC1A3*) (107); and the ABC transporter the breast cancer resistance protein (BCRP, *ABCG2*) (108). (b) X-ray structure and computational model of the Na⁺/succinate transporter VcINDY (SLC13 homolog) in the inward-(81) and outward-(82)-facing conformations, respectively. (c) Homology models of SLC13 members based on VcINDY can explain the functional consequence of the epilepsy causing mutation in Thr227 in NaCT (*SLC13A5*) (99), as well as (d) guide the discovery of novel inhibitors for NaDC3 (*SLC13A3*) using virtual screening and functional testing (109). The purple sphere represents a sodium ion in the binding site. The green sticks correspond to small molecule ligands including citrate (c) and a new NaDC3 inhibitor (d).

Figure 5. Workflow for structural modeling and analysis.

The structures of transporters are usually modeled with homology modeling, in which the 3D structure a target protein is modeled based on an experimentally determined template structure with detectable sequence similarity to the target protein. Models that share sequence identity of 30% or more with their template structures are expected to be significantly more accurate than models with lower sequence identity, and can potentially be used in structure-based ligand screening. The relevance of the model to structure-based virtual screening can be estimated by the model's ability to separate known ligands

(substrates or inhibitors) from likely non-ligands using docking and enrichment calculations. Notably, hypotheses generated based on the structural model (e.g., functionally important residue or a putative inhibitor) should be tested experimentally using a relevant experimental method which could be biophysical, biochemical, or cell-based assays.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Examples of ligand-based transporter models.

Examples are provided representing different modeling approaches and different types of input data (substrate transport and inhibition; continuous and categorical data).

Algorithm	Modeling Objective	Input Data	Data Source	Datasets	Training Set	Test Set	Reference
Multiple linear regression (MLR) and linear discriminant analysis (LDA) models							
MLR	MDR-1 Inhibition	Continuous: ED50	Resistant T-lymphoblast cell line CCRP-CEM vcr1000	Training: 20 series	$R^2=0.968$	No external validation	(30)
MLR	BSEP Inhibition	Continuous: % Inhibition	Inverted membrane vesicles from BSEP-expressing Sf9 insect cells	Training: 38	$R^2=0.952$	No external validation	(31)
MLR-LDA	BCRP Substrates	Binary: S/NS	Various literature sources	Training: 164 Test: 98	Acc: 0.82 Sen: 0.80 Spe: 0.85 ROC-AUC: 0.804	Acc: 0.75 Sen: 0.70 Spe: 0.76 ROC-AUC: 0.804	(32)
Partial least squares (PLS) models							
PLS-DA and LDA	MDR-1 Inhibition	Binary: I/NI	Various literature sources	Training: 772 Test: 85 External: 418	Acc: 0.88 Sen: 0.84 Spe: 0.91	Acc: 0.85 Sen: 0.82 Spe: 0.94	(33)
OPLS-DA	BSEP Inhibition	Binary: SI/NI	Inverted membrane vesicles from BSEP-expressing Sf9 insect cells	Training: 163 Test: 86	Acc: 0.91 Sen: 0.80 Spe: 0.94 MCC: 0.751	Acc: 0.89 Sen: 0.76 Spe: 0.94 MCC: 0.728	(34)
PLS-DA	BCRP Inhibition	Binary: I/NI	BCRP-transfected Saos-2 cells	Training: 80 Test: 42	Sen: 0.82 Spe: 0.80	Sen: 0.87 Spe: 0.79	(21)
PLS-DA	OCT2 Inhibition	Binary: SI/NI	OCT2-transfected HEK293 cells	Training: 600 Test: 300	Acc: 0.77 Sen: 0.76 Spe: 0.77	Acc: 0.74 Sen: 0.72 Spe: 0.74	(13)
k-Nearest Neighbor (kNN) classification models							
GA-kNN	MDR-1 Substrates	Binary: S/NS	Efflux ratios from MDCK-MDR1	Training: 150 Test: 37	Acc: 0.83 ROC-AUC: 0.90 MCC: 0.65	Acc: 0.81 Sen: 0.86 Spe: 0.81 MCC: 0.63	(35)
kNN	Multiple Transporters Inhibition and Substrates	Binary: I/NI and S/NS	Various literature sources	Training: 3,768	Substrate CCR: 73–97%	Substrate ROC-AUC: 0.70 ±0.02 Inhibitor ROC-AUC: 0.65 ±0.02	(36)

Algorithm	Modeling Objective	Input Data	Data Source	Datasets	Training Set	Test Set	Reference
Support vector machine (SVM) classification models							
SVM	MDR-1 Inhibition	Binary: I/NI	Various literature sources	Training: 857 Test: 418	Acc: 0.84 Sen: 0.87 Spe: 0.81 MCC: 0.68	Acc: 0.87 Sen: 0.94 Spe: 0.74 MCC: 0.70	(37)
SVM	BSEP Inhibition	Binary: I/NI	Inverted membrane vesicles from BSEP expressing Sf21 insect cells	Training: 437 Test: 187	Sen: 0.87 Acc: 0.87 Spe: 0.87 MCC: 0.73	Sen: 0.90 Acc: 0.87 Spe: 0.84 MCC: 0.74	(38)
Random forest (RF) classification models							
RF	MATE1 Inhibition	Binary: I/NI	MATE1-expressing HEK293 cells	Training: 450 Test: 450	Not specified	ROC-AUC: 0.78	(22)
RF	OCT1 Inhibition	Binary: I/NI	OCT1-expressing HEK293 cells	Training: 183 Test: 1780	Not specified	Sen: 0.82 Spec: 0.65 ROC-AUC: 0.84	(39)
Bayesian classification models							
Bayesian model	MRP4 Inhibition	Binary: I/NI	Inverted membrane vesicles from MRP4 expressing HEK293	Training: 57 Test: 29	Sen: 0.97 Acc: 0.97 Spe: 0.96 MCC: 0.93	Sen: 0.59 Acc: 0.69 Spe: 0.83 MCC: 0.42	(40)
Bayesian model	BCRP Inhibition	Binary: I/NI	MCF-7/Adrnp cells	Training: 124 Test: 79	Sen: 0.95 Acc: 0.90 Spe: 0.70 MCC: 0.689	Sen: 0.95 Acc: 0.90 Spe: 0.71 MCC: 0.69	(41)
3D-QSAR models							
CoMFA and CoMSIA	MDR-1 and MRP1 Inhibition	Continuous: IC ₅₀	COR.L23/R cells	Training: 111	CoMFA q ² = 0.712 CoMSIA q ² = 0.732	No external validation	(42)
CoMFA and CoMSIA	BCRP Inhibition	Continuous: IC ₅₀	MDCK BCRP-expressing and MCF-7 mitoxantrone-resistant	Training: 31	CoMFA q ² = 0.619 CoMSIA q ² = 0.624	No external validation	(43)
CoMFA	Oat1 and Oat6 Substrates	Continuous: pK _i	Oat1 and Oat6 uptake in <i>Xenopus laevis</i> oocytes	Training: 28 Test: 7	CoMFA q ² = 0.643	Predictive residual sum of squares: 7.766	(44)

Algorithm	Modeling Objective	Input Data	Data Source	Datasets	Training Set	Test Set	Reference
Pharmacophore models							
Common feature pharmacophore	MDR-1 Inhibition	Binary: I/Ni	K562Dox cells and MDR-1 membrane ATPase	Training: 26	No training set statistics	12 inhibitors from 21 predicted inhibitors	(45)
Common feature pharmacophore	NTCP Inhibition	Continuous: pK _i	NTCP-expressing HEK293 cells	Training: 23	R ² =0.763	No external validation	(46)
Proteochemometric models							
Random forest	OATP1B1/1B3 Inhibition	Binary: I/Ni	OATP1B1- and 1B3-expressing CHO cells	Training: 2000 Test: 54	Sen: 0.75 Spe 0.85 ROC-AUC: 0.94 MCC: 0.43	No external validation	(10)

I: Inhibitor, NI: Non-inhibitor, SI: Strong inhibitor, S: Substrate, NS: Non-substrate, TP: true positive, FP: false positive, TN: true negative, FN: false negative. Sensitivity (Sen) = TP / (TP + FN). Specificity (Spe) = TN / (TN + FP). Accuracy (Acc) = (TP + TN) / (TP + FP + FN + TN). Matthew's Correlation Coefficient (MCC) = (TP × TN − FN × FP) / sqrt ((TP + FN)(TP + FP)(TN + FN)(TN + FP)). ROC-AUC: Receiver Operating Characteristic Area under the Curve. Correct Classification Rate: (CCR = 0.5 × sensitivity + 0.5 × specificity)