P-Glycoprotein Transporter Modelling

Subjects: Biochemical Research Methods

Contributor: Liadys Mora Lagares, Marjana Novič

ATP-binding cassette (ABC) transporters play a critical role in both drug bioavailability and toxicity, and with the discovery of the P-glycoprotein (P-gp), this became even more evident, as it plays an important role in preventing intracellular accumulation of toxic compounds. Intensive studies have been conducted to find new therapeutic molecules to reverse the phenomenon of multidrug resistance (MDR), that research has found is often associated with overexpression of P-gp, the most extensively studied drug efflux transporter; in MDR, therapeutic drugs are prevented from reaching their targets due to active efflux from the cell. The development of P-gp inhibitors is recognized as a good way to reverse this type of MDR, which has been the subject of extensive studies over the past few decades. Despite the progress made, no effective P-gp inhibitors to reverse multidrug resistance are yet on the market, mainly because of their toxic effects. Computational studies can accelerate this process, and in silico models such as quantitative structure-activity relationship (QSAR) models that predict the activity of compounds associated with P-gp (or analogous transporters) are of great value in the early stages of drug development, along with molecular modelling methods, which provide a way to explain how these molecules interact with the ABC transporter.

Keywords: P-glycoprotein ; ligand-based ; models ; structure-based ; homology modelling ; molecular dynamics simulations ; machine learning ; computational models

1. Introduction

P-glycoprotein (P-gp) ligand-based drug design relies on knowledge of compounds known to interact with this membrane transporter. The structure–activity relationship (SAR) [1], quantitative structure-activity relationship (QSAR) [2], three-dimensional quantitative structure-activity relationship (3D-QSAR) [3], and pharmacophore models [4] have been used to predict the activity of new compounds toward P-gp. The amount of literature addressing ligand-based approaches of P-gp is enormous and dates back several decades. Since the discovery of verapamil as an agent to reverse multidrug resistance (MDRR) [5], many SAR and QSAR studies have been published. However, in recent years, models have used larger datasets and focus more on using machine-learning approaches.

In general, the ligand-based modelling approach consists of two elements: molecular descriptors and mathematical methods for deriving predictive models, such as linear models, artificial neural networks (ANN), support vector machines (SVMs), and others. The ligand-based methods require, in the first step, the creation of an appropriate model from a training dataset. Then, the model is applied to the screening set of molecules to make predictions for the property under study [6].

Ligand-based methods, including fingerprint-based similarity search, 2D-QSAR, pharmacophores, 3D-QSAR, and others [7][8], are faster and relatively easy to implement compared to structure-based drug design methods; one of their main advantages is their low CPU requirements. Ligand-based approaches allow the use of generalized descriptors, features, and fingerprints, making them effective virtual screening queries; they have evolved along with advances in statistics and other machine-learning algorithms, such as regression, pattern recognition, and neural networks. However, one of their limitations is that they are based on the principles of molecular similarity, which state that structurally similar molecules should elicit similar biological responses; thus, they are limited to the chemical space used in the formulation of the models, i.e., to their domain of applicability. Efforts are still needed to develop methods that can be universally applicable [9].

In the last five years, most studies have used machine-learning approaches to build predictive models, and some also combined ligand-based machine-learning approaches with structure-based techniques such as molecular docking or MD simulations. A good example is the work by Esposito et al. [10], which used a combination of machine learning and MD simulations to predict P-glycoprotein substrates. In their study, the authors used molecular dynamics fingerprints (MDFPs) as orthogonal descriptors for training substrate classification models by machine learning. The performance of the MDFPs proved to be as good as the performance of the commonly used 2D molecular descriptors, which achieved high accuracy

on chemically diverse subsets. However, when challenging the models with external validation sets, only the models trained on MDFPs or property-based descriptors could be applied to regions of chemical space not covered by the training set. The work of Esposito et al. involved one of the largest datasets found in the literature for building P-gp activity models, covering 3930 compounds. On the other hand, Kadioglu et al. [11] used machine-learning approaches such as k-nearest neighbors (kNN), neural network, random forest (RF), and support vector machine (SVM) in combination with molecular docking to develop predictive models for P-gp modulators. The molecular docking step was used as further validation for the best performing model, with twenty of the predicted compounds of each class docked to the human structure of P-gp [12]. The results revealed similar docking poses to those of doxorubicin and elacridar, which are substrate and inhibitor, respectively.

2. Improving Feature Selection

In addition to the combination of ligand-based and structure-based approaches, some efforts seem to have been performed recently-first, to optimize the methodology for selecting features to build the models, and second, to find a mechanistic interpretation that could lead to future optimization of the structures. In the work of García et al. [13], the use of a boosting feature selection method is proposed to improve the performances of P-gp classifiers and to avoid the not uncommon problem that the performance of a model built with the selected subset of features is worse than the performance obtained with all features. The authors used decision trees and support vector machines (SVM) to build P-gp inhibitor and substrate models and proved that the boosting feature selection method performed better compared to standard feature selection algorithms. The dataset used in this research consisted of 1935 P-gp inhibitors and 484 substrates. To obtain informative structural rules for the analyzed endpoint, a recent study by Wang et al. [14] developed an online decision tree-based prediction server for P-gp substrates and inhibitors called *PgpRules*, which provides two separate prediction services for P-gp substrates and inhibitors. Models were built with the classification and regression trees (CART) algorithm employing fingerprints and traditional molecular descriptors; the dataset used consisted of 925 P-gp substrates and 2056 P-gp inhibitors. The performances of the generated models were good, providing classification accuracy of over 0.70 for both endpoints, but the novelty of this server is that it annotates the rules with the key structural features for the endpoint, providing a guide for structural changes in the optimization process of drug development. Similarly, a SMILE-based classification model by Prachayasittikul et al. [15] developed using the CORrelation And Logic software (CORAL), enables the discovery of important chemical features that may contribute to the inhibitory activity of the compounds. The model, built using a dataset of 2254 compounds along with SMILES attributes instead of traditional molecular descriptors, exhibited acceptable predictive performance with accuracy, sensitivity, and specificity values greater than 70% and an MCC value greater than 0.6 for training, calibration, and validation sets.

Many of the models found in the literature mainly use fingerprints or molecular descriptors as features to build the models. Recently however, Hinge et al. [16] decided to develop a binary classification model using descriptors of a new type, namely, solvation free-energy descriptors, to prove that the molecular solvation free energy theory can be used to successfully identify P-gp inhibitors. The authors used various machine-learning approaches to build the models, such as gradient boosting machines (GBM), generalized linear models (GLM), support vector machines (SVM), and weighted κ-nearest neighbors (weighted-kNN). Among these, the SVM classification model showed the best performance, employing a combination of ten three-dimensional reference interaction site model with the Kovalenko-Hirata closure approximation (3D-RISM-KH) solvation free-energy descriptors, along with other thirteen 2D descriptors. The dataset used in this research consists of 1274 compounds derived from the work of Broccatelli et al. [17], which has been used several times to construct new models. The study demonstrates that the combination of 3D-RISMKH-based descriptors with 2D descriptors increases the accuracy of the model in predicting P-gp inhibitors compared to previous classifiers, all the way to 95.6–96.9%.

3. Reducing Heterogeneity in the Data

Other studies have focused on collecting more homogeneous datasets for the prediction of the P-gp efflux ratio, because in many of the previous studies, the collected data came from different assay conditions (e.g., different cell lines), resulting in a high degree of data heterogeneity. Following this line, Ohashi et al. [18] developed regression models to predict the value of P-glycoprotein mediated efflux and classification models to predict P-gp-mediated transport potential (low, medium, or high potential), based on a dataset of 2397 compounds collected under the same experimental conditions; data were obtained using an in vitro assay developed in-house and presented in the same study. The authors built five classification models using various machine-learning methods, such as random forest, support vector machine, artificial neural network, k-nearest neighbors, and Adaboost; and the random forest method had the best performance in both regression and classification models. This model provides information on whether the compound is a strong or weak P-gp

substrate, which is an advantage over the usual binary models. Chen et al. [19] also developed a predictive model for the P-gp substrate efflux ratio but using a small dataset of 63 compounds. The authors used the hierarchical support vector regression (HSVR) method, and although the dataset came from multiple literature sources, it was carefully curated to select experimental data under the same assay conditions. The predictions by HSVR showed high accuracy and agreed well with the observed experimental values. Another recent contribution to the prediction of P-gp efflux potential was the work of Watanabe et al. [20], who developed an in silico prediction model for P-gp efflux potential in brain capillary endothelial cells (BCEC). The authors built three predictive models for the in vitro P-gp net efflux ratio using their own dataset and publicly available data. The model was constructed using the gradient boosting (GB) machine-learning method, and the proposed model was validated with new experimental brain-penetration data of 28 P-gp substrates. It showed good predictive accuracy compared with previous similar models.

4. Three-Class Classification Models

Recently, Mora Lagares et al. [21] developed a classification model that provides a qualitative prediction of P-glycoprotein inhibition/substrate activity, as it is a three-class classification model that, unlike the currently available classifiers, is able to distinguish whether the molecule under study is a substrate, inhibitor, or non-active compound. The model was developed using a counter propagation artificial neural network (CP ANN) based on a set of 2D molecular descriptors and an extensive dataset of 2512 compounds. The P-gp activity model provided good classification performance and was implemented in the online platform VEGAHUB [22], which is freely available to the public at https://www.vegahub.eu/portfolio-item/vega-qsar/ (accessed on 1 November 2022).

5. Models Including Other Transporters

Other recent studies not only aimed to predict P-gp activity but also involved the prediction of other transporters; e.g., Estrada-Tejedor et al. [23] used modified self-organizing maps (SOM) to predict drug resistance related to P-gp activity and other transporters, such as MPR1 (ABCC1) and BCRP (ABCG2). The dataset used consisted of 1204 compounds. The authors used a strategy that combines a new clustering algorithm with SOM, called consensus self-organizing maps (CSOM), to build a multi-labelled unsupervised classification model, and then compared its performance with a k-NN classifier. The performance of the model was similar to those of conventional supervised machine-learning algorithms. However, the improvement in the accuracy of substrate classification, the main advantage of CSOM, relies on its ability to identify those substrates that are more likely to be misclassified and discard those examples, thereby reducing uncertainty. On the other hand, Namasivayam et al. [24] developed a computer-aided pattern analysis (C@PA) method to discover new inhibitors for several ABC transporters, such as P-gp (ABCB1), MRP1 (ABCC1), and BCRP (ABCG2). Based on experimental data collected from 93 reports between 2004 and 2021 evaluating small inhibitors for all three transporters, the authors sought to identify the critical fingerprints for triple inhibition of ABCB1, ABCC1, and ABCG2; and the structural features that must be present for promiscuity toward the three transporters. The dataset of 1049 compounds was divided into eight classes based on their activity profiles toward the transporters. C@PA included identification of basic scaffolds, substructure search and statistical distribution, and extraction of novel scaffolds to screen a large virtual compound library. As a result of screening a public library of drug-like compounds, 45,000 hits were found for novel, broad-spectrum ABC transporter inhibitors, of which 23 were selected for biological evaluation and 5 were found to be novel lead molecules as triple ABCB1, ABCC1, and ABCG2 inhibitors.

References

1. Seelig, A.; Landwojtowicz, E. Structure–activity relationship of P-glycoprotein substrates and modifiers. Eur. J. Pharm. Sci. 2000, 12, 31–40.
2. Dearden, J.; Al-Noobi, A.; Scott, A.; Thomson, S. QSAR studies on P-glycoprotein-regulated multidrug resistance and o n its reversal by phenothiazines. SAR QSAR Environ. Res. 2003, 14, 447–454.
3. Kim, K.H. 3D-QSAR analysis of 2, 4, 5-and 2, 3, 4, 5-substituted imidazoles as potent and nontoxic modulators of P-gly coprotein mediated MDR. Bioorg. Med. Chem. 2001, 9, 1517–1523.
4. Pajeva, I.K.; Wiese, M. Pharmacophore model of drugs involved in P-glycoprotein multidrug resistance: Explanation of structural variety (hypothesis). J. Med. Chem. 2002, 45, 5671–5686.
5. Tsuruo, T.; Iida, H.; Tsukagoshi, S.; Sakurai, Y. Overcoming of vincristine resistance in P388 leukemia in vivo and in vitr o through enhanced cytotoxicity of vincristine and vinblastine by verapamil. Cancer Res. 1981, 41, 1967–1972.

6. Moro, S.; Bacilieri, M.; Deflorian, F. Combining ligand-based and structure-based drug design in the virtual screening ar ena. Expert Opin. Drug Discov. 2007, 2, 37–49.
7. Vedani, A.; Briem, H.; Dobler, M.; Dollinger, H.; McMasters, D.R. Multiple-conformation and protonation-state represent ation in 4D-QSAR: The neurokinin-1 receptor system. J. Med. Chem. 2000, 43, 4416–4427.
8. Vedani, A.; Dobler, M. 5D-QSAR: The key for simulating induced fit? J. Med. Chem. 2002, 45, 2139–2149.
9. Prathipati, P.; Dixit, A.; Saxena, A.K. Computer-aided drug design: Integration of structure-based and ligand-based appr oaches in drug design. Curr. Comput.-Aided Drug Des. 2007, 3, 133–148.
10. Esposito, C.; Wang, S.; Lange, U.E.; Oellien, F.; Riniker, S. Combining machine learning and molecular dynamics to pr edict P-glycoprotein substrates. J. Chem. Inf. Model. 2020, 60, 4730–4749.
11. Kadioglu, O.; Efferth, T. A Machine Learning-Based Prediction Platform for P-Glycoprotein Modulators and Its Validation by Molecular Docking. Cells 2019, 8, 1286.
12. Alam, A.; Kowal, J.; Broude, E.; Roninson, I.; Locher, K.P. Structural insight into substrate and inhibitor discrimination b y human P-glycoprotein. Science 2019, 363, 753–756.
13. Garcia, G.C.; Garcia-Pedrajas, N. Boosted feature selectors: A case study on prediction P-gp inhibitors and substrates. J. Comput.-Aided Mol. Des. 2018, 32, 1273–1294.
14. Wang, P.H.; Tu, Y.S.; Tseng, Y.F.J. PgpRules: A decision tree based prediction server for P-glycoprotein substrates and inhibitors. Bioinformatics 2019, 35, 4193–4195.
15. Prachayasittikul, V.; Worachartcheewan, A.; Toropova, A.P.; Toropov, A.A.; Schaduangrat, N.; Prachayasittikul, V.; Nant asenamat, C. Large-scale classification of P-glycoprotein inhibitors using SMILES-based descriptors. Sar Qsar Enviro n. Res. 2017, 28, 1–16.
16. Hinge, V.K.; Roy, D.; Kovalenko, A. Prediction of P-glycoprotein inhibitors with machine learning classification models a nd 3D-RISM-KH theory based solvation energy descriptors. J. Comput.-Aided Mol. Des. 2019, 33, 965–971.
17. Broccatelli, F.; Carosati, E.; Neri, A.; Frosini, M.; Goracci, L.; Oprea, T.I.; Cruciani, G. A Novel Approach for Predicting P -Glycoprotein (ABCB1) Inhibition Using Molecular Interaction Fields. J. Med. Chem. 2011, 54, 1740–1751.
18. Ohashi, R.; Watanabe, R.; Esaki, T.; Taniguchi, T.; Torimoto-Katori, N.; Watanabe, T.; Ogasawara, Y.; Takahashi, T.; Tsu kimoto, M.; Mizuguchi, K. Development of Simplified in Vitro P-Glycoprotein Substrate Assay and in Silico Prediction M odels To Evaluate Transport Potential of P-Glycoprotein. Mol. Pharm. 2019, 16, 1851–1863.
19. Chen, C.; Lee, M.H.; Weng, C.F.; Leong, M.K. Theoretical Prediction of the Complex P-Glycoprotein Substrate Efflux B ased on the Novel Hierarchical Support Vector Regression Scheme. Molecules 2018, 23, 1820.
20. Watanabe, R.; Esaki, T.; Ohashi, R.; Kuroda, M.; Kawashima, H.; Komura, H.; Natsume-Kitatani, Y.; Mizuguchi, K. Dev elopment of an In Silico Prediction Model for P-glycoprotein Efflux Potential in Brain Capillary Endothelial Cells toward t he Prediction of Brain Penetration. J. Med. Chem. 2021, 64, 2725–2738.
21. Mora Lagares, L.; Minovski, N.; Novič, M. Multiclass Classifier for P-Glycoprotein Substrates, Inhibitors, and Non-Active Compounds. Molecules 2019, 24, 2006.
22. Benfenati, E.; Manganaro, A.; Gini, G.C. VEGA-QSAR: AI Inside a Platform for Predictive Toxicology. In Proceedings of the AI* IA, Turin, Italy, 5 December 2013; pp. 21–28.
23. Estrada-Tejedor, R.; Ecker, G.F. Predicting drug resistance related to ABC transporters using unsupervised Consensus Self-Organizing Maps. Sci. Rep. 2018, 8, 6803.
24. Namasivayam, V.; Silbermann, K.; Wiese, M.; Pahnke, J.; Stefan, S.M. : Computer-Aided Pattern Analysis to Predict M ultitarget ABC Transporter Inhibitors. J. Med. Chem. 2021, 64, 3350–3366.