Short Communication

Making Transporter Models for Drug–Drug Interaction Prediction Mobile

Received April 16, 2015; accepted July 21, 2015

ABSTRACT

The past decade has seen increased numbers of studies publishing ligand-based computational models for drug transporters. Although they generally use small experimental data sets, these models can provide insights into structure–activity relationships for the transporter. In addition, such models have helped to identify new compounds as substrates or inhibitors of transporters of interest. We recently proposed that many transporters are promiscuous and may require profiling of new chemical entities against multiple substrates for a specific transporter. Furthermore, it should be noted that virtually all of the published ligand-based transporter models are only accessible to those involved in creating them and, consequently, are rarely shared effectively. One way to surmount this is to make models shareable or more accessible. The development of mobile apps that can access such models is highlighted here. These apps can be used to predict ligand interactions with transporters using Bayesian algorithms. We used recently published transporter data sets (MATE1, MATE2K, OCT2, OCTN2, ASBT, and NTCP) to build preliminary models in a commercial tool and in open software that can deliver the model in a mobile app. In addition, several transporter data sets extracted from the ChEMBL database were used to illustrate how such public data and models can be shared. Predicting drug–drug interactions for various transporters using computational models is potentially within reach of anyone with an iPhone or iPad. Such tools could help prioritize which substrates should be used for in vivo drug–drug interaction testing and enable open sharing of models.

Introduction

We are increasingly seeing medium- or high-throughput screens used to develop ligand-based models for individual transporters (Diao et al., 2009, 2010; Zheng et al., 2009; Kido et al., 2011; Astorga et al., 2012; Ekins et al., 2012b; Greupink et al., 2012; Dong et al., 2013, 2014; Sedykh et al., 2013; Wittwer et al., 2013; Xu et al., 2013). One of the significant limitations of this is that the models developed are rarely accessible outside of the research group developing them, likely because of the commercial software required. One way to surmount this is to develop models using open-source software. We previously showed that such “open models” produce validation statistics that are comparable to commercial tools (Gupta et al., 2010). Because many computational machine learning methods use molecular function class fingerprints of maximum diameter 6 (FCFP6) and extended connectivity fingerprints (ECFP6), we have described their implementation with the Chemistry Development Kit (CDK) (Steinbeck et al., 2003) components (Clark et al., 2014). We also recently described how an open-source Bayesian algorithm can be used with these descriptors to develop and validate thousands of data sets, including those from the ChEMBL database (Clark and Ekins, 2015; Clark et al., 2015). In response to the shift toward mobile computing, we have developed apps for drug discovery, leveraging years of research in cheminformatics (Williams et al., 2011; Ekins et al., 2012a, 2013a,b; Clark et al., 2013, 2014). A number of mobile apps have been created for sketching and sharing molecules, such as the Mobile Molecular DataSheet (MMDS), MolPrime, and Open Drug Discovery Teams apps (Supplemental Table 1). More recently, we created mobile apps that combine Bayesian models and open-source fingerprint descriptors to enable models that can be used within a mobile app without connecting to the Internet (TB Mobile, MMDS, Approved Drugs, and MolPrime) (Supplemental Table 1). A mobile app that enables a scientist to select a molecule and score it with models (e.g., for various transporters of relevance for drug–drug interactions) is now possible. As a proof of concept, we used previously modeled transporters (Zheng et al., 2009; Diao et al., 2010; Astorga et al., 2012; Ekins et al., 2012b; Dong et al., 2013, 2014). In this study, we describe validated models for the human multidrug and toxin extrusion proteins (MATE1, MATE2K), organic cation transporter (OCT2), human organic cation/carnitine transporter (OCTN2), human apical sodium-dependent bile acid transporter (ASBT), and sodium taurocholate cotransporting polypeptide (NTCP).

Materials and Methods

We recently described the development of open-source FCFP6 and ECFP6 descriptors and the Bayesian algorithm that enables us to build models with open-source tools (Clark and Ekins, 2015; Clark et al., 2014, 2015). The CDK codebase

ABBREVIATIONS: ASP, 4-[(4-dimethylaminostyryl)-N-methyl-pyridinium; CDK, Chemistry Development Kit; DX-619, 7-[(3R)-3-[[1-amino cyclopropyl]pyrrolidin-1-yl]-1-[(R,2S)-2-fluorocyclopropyl]-8-methoxy-4-oxoquinoline-3-carboxylic acid; MMDS, Mobile Molecular DataSheet; ROC, receiver operator curve.
has been deposited in the latest version of GitHub (http://github.com/cdk/cdk; in the “tools” section, look for class org.openscience.cdk.fingerprint.model.Bayesian). Because of their open nature, future tools can build on them. We previously published several transporter models and described Bayesian models generated using Discovery Studio (Biovia, San Diego, CA) for MATE1, MATE2K, OCTN2, ASBT, and NTCP (Diao et al., 2009, 2010; Zheng et al., 2010; Astorga et al., 2012; Dong et al., 2013, 2014). We have now analyzed several larger published data sets from other groups for MATE1 (Wittwer et al., 2013) and OCT2 (Kido et al., 2011), which we have also used to generate Bayesian models with Discovery Studio to compare the different fingerprints. To illustrate the utility of transporter models built with open ECFP6 descriptors and the Bayesian algorithm, 5-fold cross-validation and leave-one-out validation were used. Receiver operator curve (ROC) values were produced, in which a value of 1 is ideal and a value greater than 0.7 is considered good. Cutoffs for actives and inactives were as previously described (Diao et al., 2009, 2010; Zheng et al., 2010; Kido et al., 2011; Astorga et al., 2012; Dong et al., 2013, 2014; Wittwer et al., 2013). Open transporter models were developed using open-source software (Clark et al., 2014, 2015; Clark and Ekins, 2015) and loaded into the MMDS app (http://molmatinf.com/) (Fig. 1, and several MATE models were then used to score DX-619 (7-{(3R)-3-(1-aminocyclopropyl)pyrrolidin-1-yl]-1-[(1R,2S)-2-fluorocyclopropyl]-8-methoxy-4-oxoquinoline-3-carboxylic acid), a MATE inhibitor (Imamura et al., 2013) (Fig. 1C). Transporter data sets from the ChEMBL database (Bento et al., 2014; Papadatos and Overington, 2014) (Fig. 2) were used to illustrate our work to determine a cutoff for activity and automated model building and validation (Clark et al., 2015). These models are also openly accessible as part of a set of over 2000 models (http://molsync.com/bayesian2) and can be uploaded into MMDS to score molecules of interest.

**Results and Discussion**

We have described the use of seven data sets representing six transporters of interest (MATE1, MATE2K, OCT2, OCTN2, ASBT, and NTCP) due to their potential involvement in drug–drug interactions (Diao et al., 2009, 2010; Zheng et al., 2009; Kido et al., 2011; Astorga et al., 2012; Ekins et al., 2012b; Greupink et al., 2012; Dong et al., 2013, 2014; Sedykh et al., 2013; Wittwer et al., 2013; Xu et al., 2013). We focused on Bayesian models with ECFP6 and FCFP6 descriptors because these were previously used to develop models that were successfully validated using prospective testing. Our goal in this study was to illustrate how such data sets could be used to create computational models that can be made open or more accessible. Bayesian models developed with either Discovery Studio or the open descriptors and open Bayesian algorithm (Clark and Ekins, 2015; Clark et al., 2014, 2015) were evaluated using 5-fold cross-validation and leave-one-out ROC values. Models created with commercial or open Bayesian algorithms and ECFP6 descriptors had comparable 5-fold ROC values that were above 0.70 (Table 1). One would expect some...
variability because of the small data set size, and six of the seven data sets have slightly higher ROC values with open Bayesian algorithms and ECFP6 descriptors. In addition, variability would be expected because the implementations of ECFP6 descriptors are similar but not identical (Clark et al., 2014). These data sets and open-source Bayesian models are also available via http://molsync.com/transporters. The models with the larger training sets, such as for MATE1 and OCT2, had very comparable ROC values across models. The combination of published NTCP data sets (Dong et al., 2013, 2014) for the first time here consistently produced less desirable models.

We previously suggested that hMATE1 may have a complex binding surface, rather than a single binding site (Astorga et al., 2012). Because data sets have been developed for different substrate probes, there is now the potential to build substrate-specific models for a single transporter. This may be useful to predict substrate-specific drug–drug interactions. As an example, we have used the antibiotic molecule DX-619 as a test case, which was predicted in MMDS (Fig. 1, A and B) with one MATE1 Bayesian model for inhibition of 1-methyl-4-phenylpyridinium transport as a likely inhibitor (Bayesian score 0.80) and scored lower for inhibition of ASP transport (Bayesian score 0.48; Fig. 1C). This may suggest that future drug–drug interaction studies could focus on using 1-methyl-4-phenylpyridinium as the substrate probe, rather than ASP.

A similar computational model building and sharing approach could be taken for drug-metabolizing enzymes, nuclear receptors, and ion channels of interest for drug–drug interaction and toxicity prediction (Ekins 2014). The ChEMBL database has thousands of data sets curated from the literature for important drug targets, including transporters. We recently described, using a novel automated approach, how a cutoff for activity can be selected for such data sets and how models can be

Fig. 2. Selected examples of extracted data sets from the ChEMBL database and the analysis leading to the detection of a suitable activity threshold (data and models available at http://molsync.com/bayesian2). Each example shows a plot of population versus activity, for which the solid curve shows the integral, which is colored to show inactives (below threshold: light gray) and actives (above threshold: dark gray). The ROC integral for subset models at various thresholds is plotted, as is the overall desirability composite score. The ROC curve for ECFP6 and FCFP6 models is shown to the right, which is built using the whole data set at the determined threshold. A representative diverse selection of “active” and “inactive” molecules is shown underneath.
built and validated (Fig. 2) (Clark et al., 2015). These models could be uploaded alongside the transporters from the user’s own, or other published, data sets to profile compounds of interest. The analysis and prediction of a molecule’s propensity for drug–drug interactions mediated via transporters or other proteins could be useful to suggest future experiments. Because the models can be stored in the mobile app and would not have to communicate over the Internet, such calculations would be secure and may be useful for proprietary studies. This suggests an approach to making transporter models accessible to a larger audience of scientists, with the potential to transform how models are built and shared with implications for drug discovery.

Acknowledgments

The authors thank Dr. James Polli and his students for collaborations that led to several published data sets used in this study, Biovia for providing Discovery Studio, and Colleagues at Collaborative Drug Discovery, Inc. for assistance with development of the open models and descriptors work.

Collaborations Pharmaceuticals, Inc., and Collaborations in Chemistry, Fuquay-Varina, North Carolina (S.E.); Collaborative Drug Discovery, Burlingame, California (S.E.); Molecular Materials Informatics, Inc., Montreal, Quebec, Canada (A.M.C.); and Department of Physiology, University of Arizona, Tucson, Arizona (S.H.W.).

Authorship Contributions

Participated in research design: Ekins, Clark. Conducted experiments: Ekins, Clark, Wright. Contributed new reagents or analytic tools: Ekins, Clark. Performed data analysis: Ekins, Clark. Wrote or contributed to the writing of the manuscript: Ekins, Clark, Wright.

References

Clark AM and Ekins S (2015) Open source Bayesian models II. Mining A "big dataset" to create and validate models with ChEMBL. J Chem Inf Model, in press.

TABLE 1
Results for Bayesian model cross-validation

<table>
<thead>
<tr>
<th>Model</th>
<th>Number Active/Total</th>
<th>Reference</th>
<th>Discovery Studio ECFP6 Plus 8 ADME Descriptors</th>
<th>Discovery Studio ECFP6 Only</th>
<th>Open Model ECFP6 Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATE1</td>
<td>81/841</td>
<td>Wittwer et al. (2013)</td>
<td>0.65 (LOO 0.72)</td>
<td>0.62 (LOO 0.68)</td>
<td>0.69 (LOO 0.67)</td>
</tr>
<tr>
<td>MATE2</td>
<td>24/45</td>
<td>Astorga et al. (2012)</td>
<td>0.75 (LOO 0.79)*</td>
<td>0.63 (LOO 0.72)</td>
<td>0.70 (LOO 0.74)</td>
</tr>
<tr>
<td>MATE2K</td>
<td>21/45</td>
<td>Astorga et al. (2012)</td>
<td>0.50 (LOO 0.72)*</td>
<td>0.60 (LOO 0.81)</td>
<td>0.75 (LOO 0.73)</td>
</tr>
<tr>
<td>OCT2</td>
<td>243/904</td>
<td>Kido et al. (2011)</td>
<td>0.80 (LOO 0.81)</td>
<td>0.76 (LOO 0.79)</td>
<td>0.79 (LOO 0.79)</td>
</tr>
<tr>
<td>OCTN2</td>
<td>30/48</td>
<td>Diao et al. (2009, 2010)</td>
<td>0.85 (LOO 0.89)*</td>
<td>0.71 (LOO 0.77)</td>
<td>0.85 (LOO 0.88)</td>
</tr>
<tr>
<td>ASBT</td>
<td>17/58</td>
<td>Zheng et al. (2010)</td>
<td>0.75 (LOO 0.91)*</td>
<td>0.73 (LOO 0.87)</td>
<td>0.74 (LOO 0.76)</td>
</tr>
<tr>
<td>NTCP</td>
<td>50/107</td>
<td>Dong et al. (2013, 2014)</td>
<td>0.61 (LOO 0.73)</td>
<td>0.56 (LOO 0.68)</td>
<td>0.49 (LOO 0.53)</td>
</tr>
</tbody>
</table>

ADME, absorption, distribution, metabolism, excretion; LOO, leave one out.
*Previously published by the authors of this article.

Mobile Transporter Models 1645