Short Communication

Making Transporter Models for Drug-Drug Interaction Prediction Mobile

Sean Ekins, Alex M. Clark and Stephen H. Wright

Collaborations Pharmaceuticals, Inc., 5616 Hilltop Needmore Road, Fuquay-Varina, NC 27526, USA (S.E.); Collaborations in Chemistry, 5616 Hilltop Needmore Road, Fuquay-Varina, NC 27526, USA (S.E.); Collaborative Drug Discovery, 1633 Bayshore Highway, Suite 342, Burlingame, CA 94010, USA (S.E.); Molecular Materials Informatics, Inc., 1900 St. Jacques #302, Montreal H3J 2S1, Quebec, Canada (A.M.C.); Department of Physiology, University of Arizona, Tucson, Arizona, AZ 85724, USA (S.H.W.)
Running Title Page

Running title: Mobile Transporter Models

Corresponding author: Collaborations Pharmaceuticals, Inc., 5616 Hilltop Needmore Road, Fuquay-Varina, NC 27526, USA. E-mail: ekinssean@yahoo.com Phone (215) 687-1320

Number of text pages: 15
Number of tables: 1
Number of figures: 2
Number of references: 27
Number of words in

Abstract: 244
Introduction: 387
Results and Discussion: 528

Abbreviations: ADME/Tox, Absorption, distribution, metabolism, excretion and toxicity; ASBT, Human Apical Sodium-dependent Bile Acid Transporter; ASP, 4-4-dimethlaminostyryl)-N-methyl-pyridinium; MATE1, MATE-2K, Human Multidrug And Toxin Extrusion Proteins; MPP, 1-methyl-4-phenylpyridinium; NTCP, Sodium Taurocholate Co-transporting Polypeptide; OCT2, Organic Cation Transporter; OCTN2, Human Organic Cation/Carnitine Transporter.
Abstract

The past decade has seen increased numbers of studies publishing ligand-based computational models for drug transporters. Although generally using small experimental datasets, 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. Further, 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 have used recently published transporter datasets (MATE 1, MATE 2K, 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 datasets extracted from ChEMBL were used to illustrate how such public data and models can be shared. Predicting drug-drug interactions (DDI) for various transporters using computational models is potentially within reach of anyone with an iPhone / iPad. Such tools could help prioritize which substrates should be used for in vivo DDI testing and enable open sharing of models.
Introduction

Increasingly we are seeing medium or high throughput screens (HTS) used to develop ligand-based models for individual transporters (Diao et al., 2009; Zheng et al., 2009; Diao et al., 2010; Kido et al., 2011; Astorga et al., 2012; Ekins et al., 2012; Greupink et al., 2012; Dong et al., 2013; Sedykh et al., 2013; Wittwer et al., 2013; Xu et al., 2013; Dong et al., 2014). One of the significant limitations of this is that the models developed are rarely accessible outside of the research group developing them likely due to the commercial software required. One way to surmount this is the development of 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). As many computational machine learning methods use molecular “function class fingerprints of maximum diameter 6” (FCFP6) and “extended connectivity (ECFP6) fingerprints,” 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 datasets including those from the ChEMBL database (Clark et al., 2015; Clark et al., 2015). In response to the shift towards mobile computing, we have developed apps for drug discovery, leveraging years of research in cheminformatics (Williams et al., 2011; Ekins et al., 2012; Clark et al., 2013; Ekins et al., 2013; Ekins et al., 2013; Clark et al., 2014). A number of mobile apps have been created for sketching and sharing molecules (e.g. MMDS, MolPrime and ODDT (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 have used transporters modeled previously (Zheng et al., 2009; Diao et al., 2010; Astorga et al., 2012; Ekins et al., 2012; Dong et al., 2013; Dong et al., 2014). In the current study we describe validated models for the Human Multidrug And Toxin Extrusion Proteins (MATE1, MATE-2K) Organic Cation Transporter (OCT2), Human Organic Cation/Carnitine Transporter (OCTN2), Human Apical Sodium-dependent Bile Acid Transporter (ASBT) and Sodium Taurocholate Co-transporting Polypeptide (NTCP).

Materials and Methods

Recently we described the development of open source FCFP6 and ECFP6 descriptors and the Bayesian algorithm which enables us to build models with open source tools (Clark et al., 2014; Clark et al., 2015; Clark et al., 2015). The CDK codebase 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). Due to their open nature, future tools can build on them. Previously we published several transporter models and described Bayesian models generated using Discovery Studio (Biovia, San Diego, CA) for (MATE 1, MATE 2K, OCTN2, ASBT and NTCP) (Diao et al., 2009; Diao et al., 2010; Zheng et al., 2010; Astorga et al., 2012; Dong et al., 2013; Dong et al., 2014). We have now analyzed several larger published datasets 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” (LOO) cross validation was used. Receiver Operator Curve (ROC) values were produced, where 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; Diao et al., 2010; Zheng et al., 2010; Kido et al., 2011; Astorga et al., 2012; Dong et al., 2013; Wittwer et al., 2013; Dong et al., 2014). Open transporter models were developed using open source software (Clark et al., 2014; Clark et al., 2015; Clark et al., 2015) and loaded into the Mobile Molecular Data Sheet (MMDS (http://molmatinf.com/)) (Fig 1A-C) and then several MATE models were used to score DX-619, a MATE inhibitor (Imamura et al., 2013) (Fig 1C). Transporter datasets from ChEMBL (Bento et al., 2014; Papadatos et al., 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 in order to score molecules of interest.

Results and Discussion

We have described the use of seven datasets representing six transporters of interest (MATE 1, MATE 2K, OCT2, OCTN2, ASBT and NTCP) due to their potential involvement in drug-drug interactions (Diao et al., 2009; Zheng et al., 2009; Diao et al., 2010; Kido et al., 2011; Astorga et al., 2012; Ekins et al., 2012; Greupink et al., 2012; Dong et al., 2013; Sedykh et al., 2013; Wittwer et al., 2013; Xu et al., 2013; Dong et al., 2014). We focused on Bayesian models with ECFP6 and FCFP6 descriptors as these were previously used to develop models that were successfully validated using prospective
testing. Our goal in this study was to illustrate how such datasets 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 et al., 2014; Clark et al., 2015; Clark et al., 2015) were evaluated using 5-fold and LOO ROC values. Models created with commercial or open Bayesian algorithms and ECFP6 descriptors had comparable 5-fold ROC values which were above 0.70 (Table 1). One would expect some variability due to small dataset size and 6 out of 7 have slightly higher ROC values with open Bayesian algorithms and ECFP6 descriptors. Also variability would be expected as the implementations of ECFP6 descriptors are similar but not identical (Clark et al., 2014). These datasets 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 datasets (Dong et al., 2013; Dong et al., 2014) for the first time here produced less desirable models consistently.

We previously suggested hMATE1 may have a complex binding surface, rather than a single binding site (Astorga et al., 2012). Because datasets 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 1A, B) with one MATE1 Bayesian model for inhibition of 1-methyl-4-phenylpyridinium (MPP) transport as a likely inhibitor (Bayesian score 0.80) and scored lower for inhibition of 4-4-dimethlaminostyryl)-N-methyl-
pyridinium (ASP) transport (Bayesian score 0.48, Fig 1C). This may suggest that future drug-drug interaction studies could focus on using MPP 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 datasets curated from the literature for important drug targets, including transporters. We recently described, using a novel automated approach, how a cut-off for activity can be selected for such datasets, and models built and validated (Fig 2A,B) (Clark et al., 2015). These models could be uploaded alongside the transporters from the users own, or other published, datasets in order 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. As 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

SE kindly acknowledges Professor James Polli, and his students for collaborations that led to several published datasets used in this study. Biovia are kindly acknowledged for
providing Discovery Studio, and Colleagues at Collaborative Drug Discovery, Inc. are thanked for assistance with development of the open models and descriptors work.

**Authorship Contributions.**

*Participated in research design:* Ekins, S. and Clark, A.M.

*Conducted experiments:* Ekins, S., Clark, A.M., and Wright S.H.

*Contributed new reagents or analytic tools:* Ekins, S. and Clark, A.M.

*Performed data analysis:* Ekins, S. and Clark, A.M.

*Wrote or contributed to the writing of the manuscript:* Ekins, S., Clark, A.M., and Wright, S.H.
References


Footnotes

a. The open models and descriptors work was supported by NIH National Center for Advancing Translational Sciences [9R44TR000942-02]. Mobile app development for transporters was supported by NIH National Institute of Diabetes and Digestive and Kidney Diseases [5R01DK058251-14 and 1R01DK080801-5].

b. All Discovery Studio models are available from the authors upon request. Open Models are accessible at http://molsync.com/transporters. ChEMBL models are accessible at http://molsync.com/bayesian2.

c. Reprint requests: Sean Ekins, Ph.D, D.Sc. Collaborations Pharmaceuticals, Inc., 5616 Hilltop Needmore Road, Fuquay Varina, NC 27526, USA ekinssean@yahoo.com
Figure Legends

Figure 1. Making transporter models mobile. A. OCT2 Model details in the MMDS mobile app B. Selecting MATE models C. Predictions for DX-619 in MMDS. Larger scores indicate the higher probability of the compound being an inhibitor of the transporters using the datasets indicated. DX-619 was predicted with one MATE1 Bayesian model for inhibition of MPP transport as a likely inhibitor and scored lower for inhibition of ASP(+) transport.

Figure 2. Selected examples of extracted datasets from ChEMBL 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 grey) and actives (above threshold: dark grey). The ROC integral for subset models at various thresholds is plotted, as is the overall desirability composite score. To the right is shown the ROC curve for ECFP6 and FCFP6 models, built using the whole dataset at the determined threshold. A representative diverse selection of "active" and "inactive" molecules is shown underneath.
Table 1. Results for Bayesian model cross validation. 5-fold and Leave one out (LOO) validation with Bayesian models generated with Discovery Studio and Open Models implemented in the mobile app MMDS. * = previously published by us

<table>
<thead>
<tr>
<th>Model</th>
<th>References</th>
<th>Discovery Studio FCFP6 + 8 ADME descriptors</th>
<th>Discovery Studio ECFP6 only</th>
<th>Open Model ECFP6 only</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATE1 (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>MATE1 (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>MATE-2K (21/45)</td>
<td>(Astorga et al., 2012)</td>
<td>0.59 (LOO 0.72)*</td>
<td>0.60 (LOO 0.81)</td>
<td>0.75 (LOO 0.73)</td>
</tr>
<tr>
<td>OCT2 (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 (30/49)</td>
<td>(Diao et al., 2009; Diao et al., 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 (17/38)</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 (50/107)</td>
<td>(Dong et al., 2013; Dong et al., 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>
Figure 1

A. Bayesian Model

B. ROC Curve

C. DX-619
Figure 2

**Oligopeptide transporter small intestine isoform** *(Homo sapiens)*

Measurement: Ki; Assay Type: Binding (9 assays); Target Type: SINGLE PROTEIN

202 compounds
- Test model
- Desirability

ROC integrals
- Folding
  - ECFP6: 0.913
  - FCP6: 0.504

<table>
<thead>
<tr>
<th>Compound</th>
<th>ECFP6</th>
<th>FCP6</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>0.879</td>
<td>0.824</td>
</tr>
<tr>
<td>256</td>
<td>0.896</td>
<td>0.873</td>
</tr>
<tr>
<td>512</td>
<td>0.902</td>
<td>0.886</td>
</tr>
<tr>
<td>1034</td>
<td>0.917</td>
<td>0.890</td>
</tr>
<tr>
<td>2068</td>
<td>0.915</td>
<td>0.902</td>
</tr>
<tr>
<td>4096</td>
<td>0.918</td>
<td>0.907</td>
</tr>
<tr>
<td>8192</td>
<td>0.916</td>
<td>0.906</td>
</tr>
<tr>
<td>16384</td>
<td>0.912</td>
<td>0.905</td>
</tr>
<tr>
<td>32768</td>
<td>0.912</td>
<td>0.904</td>
</tr>
<tr>
<td>65536</td>
<td>0.913</td>
<td>0.905</td>
</tr>
<tr>
<td>All</td>
<td>0.913</td>
<td>0.904</td>
</tr>
</tbody>
</table>

**Bile acid transporter** *(Homo sapiens)*

Measurement: IC50; Assay Type: Binding (1 assay); Target Type: SINGLE PROTEIN

109 compounds
- Test model
- Desirability

ROC integrals
- Folding
  - ECFP6: 0.847
  - FCP6: 0.837

<table>
<thead>
<tr>
<th>Compound</th>
<th>ECFP6</th>
<th>FCP6</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>0.789</td>
<td>0.809</td>
</tr>
<tr>
<td>128</td>
<td>0.801</td>
<td>0.841</td>
</tr>
<tr>
<td>256</td>
<td>0.835</td>
<td>0.856</td>
</tr>
<tr>
<td>512</td>
<td>0.836</td>
<td>0.829</td>
</tr>
<tr>
<td>1024</td>
<td>0.848</td>
<td>0.833</td>
</tr>
<tr>
<td>2048</td>
<td>0.847</td>
<td>0.837</td>
</tr>
<tr>
<td>4096</td>
<td>0.843</td>
<td>0.840</td>
</tr>
<tr>
<td>8192</td>
<td>0.847</td>
<td>0.856</td>
</tr>
<tr>
<td>16384</td>
<td>0.845</td>
<td>0.838</td>
</tr>
<tr>
<td>32768</td>
<td>0.846</td>
<td>0.838</td>
</tr>
<tr>
<td>65536</td>
<td>0.844</td>
<td>0.839</td>
</tr>
<tr>
<td>All</td>
<td>0.847</td>
<td>0.837</td>
</tr>
</tbody>
</table>