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

The open models and descriptors research was supported by the National Institutes of Health National Center for Advancing Translational Sciences [Grant 9R44TR000942-02]. Mobile app development for transporters was supported by the National Institutes of Health National Institute of Diabetes and Digestive and Kidney Diseases [Grants 5R01DK058251-14 and 1R01DK080801-5]. 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>.

dx.doi.org/10.1124/dmd.115.064956.

[§]This article has supplemental material available at dmd.aspetjournals.org.

ABBREVIATIONS: ASP, 4-(4-dimethylaminostyryl)-N-methyl-pyridinium; CDK, Chemistry Development Kit; DX-619, 7-[(3R)-3-(1-aminocyclopropyl)pyrrolidin-1-yl]-1-[(1R,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

Fig. 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 data sets indicated. DX-619 was predicted with one MATE1 Bayesian model for inhibition of 1-methyl-4-phenylpyridinium transport as a likely inhibitor and scored lower for inhibition of ASP(+) transport.

Oligopeptide transporter small intestine isoform (Homo sapiens)Measurement: **Ki**, Assay Type: **Binding** (9 assays), Target Type: **SINGLE PROTEIN** **Ileal bile acid transporter (Homo sapiens)**Measurement: **IC50**, Assay Type: **Binding** (3 assays), Target Type: **SINGLE PROTEIN**

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.

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

TABLE 1
Results for Bayesian model cross-validation

Results show 5-fold and LOO validation with Bayesian models generated with Discovery Studio and open models implemented in the MMDS mobile app.

Model	Number Active/Total	Reference	Discovery Studio FCFP6 Plus 8 ADME Descriptors	Discovery Studio ECFP6 Only	Open Model ECFP6 Only
MATE1	81/841	Wittwer et al. (2013)	0.65 (LOO 0.72)	0.62 (LOO 0.68)	0.69 (LOO 0.67)
MATE1	24/45	Astorga et al. (2012)	0.75 (LOO 0.79) ^a	0.63 (LOO 0.72)	0.70 (LOO 0.74)
MATE2K	21/45	Astorga et al. (2012)	0.59 (LOO 0.72) ^a	0.60 (LOO 0.81)	0.75 (LOO 0.73)
OCT2	243/904	Kido et al. (2011)	0.80 (LOO 0.81)	0.76 (LOO 0.79)	0.79 (LOO 0.79)
OCTN2	30/49	Diao et al. (2009, 2010)	0.85 (LOO 0.89) ^a	0.71 (LOO 0.77)	0.85 (LOO 0.88)
ASBT	17/38	Zheng et al. (2010)	0.75 (LOO 0.91) ^a	0.73 (LOO 0.87)	0.74 (LOO 0.76)
NTCP	50/107	Dong et al. (2013, 2014)	0.61 (LOO 0.73)	0.56 (LOO 0.68)	0.49 (LOO 0.53)

ADME, absorption, distribution, metabolism, excretion; LOO, leave one out.

^aPreviously published by the authors of this article.

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.)

SEAN EKINS
ALEX M. CLARK
STEPHEN H. WRIGHT

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

- Astorga B, Ekins S, Morales M, and Wright SH (2012) Molecular determinants of ligand selectivity for the human multidrug and toxin extruder proteins MATE1 and MATE2-K. *J Pharmacol Exp Ther* **341**:743–755.
- Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, Krüger FA, Light Y, Mak L, and McGlinchey S, et al. (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res* **42**:D1083–D1090.
- Clark AM, Dole K, Coulon-Spektor A, McNutt A, Grass G, Freundlich JS, Reynolds RC, and Ekins S (2015) Open source bayesian models: I. Application to ADME/Tox and drug discovery datasets. *J Chem Inf Model* **55**:1231–1245.
- 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.

- Clark AM, Sarker M, and Ekins S (2014) New target prediction and visualization tools incorporating open source molecular fingerprints for TB Mobile 2.0. *J Cheminform* **6**:38.
- Clark AM, Williams AJ, and Ekins S (2013) Cheminformatics workflows using mobile apps. *Chem-Bio Informatics J* **13**:1–18.
- Diao L, Ekins S, and Polli JE (2009) Novel inhibitors of human organic cation/carnitine transporter (hOCTN2) via computational modeling and in vitro testing. *Pharm Res* **26**:1890–1900.
- Diao L, Ekins S, and Polli JE (2010) Quantitative structure activity relationship for inhibition of human organic cation/carnitine transporter. *Mol Pharm* **7**:2120–2131.
- Dong Z, Ekins S, and Polli JE (2013) Structure-activity relationship for FDA approved drugs as inhibitors of the human sodium taurocholate cotransporting polypeptide (NTCP). *Mol Pharm* **10**:1008–1019.
- Dong Z, Ekins S, and Polli JE (2014) Quantitative NTCP pharmacophore and lack of association between DILI and NTCP Inhibition. *Eur J Pharm Sci* **66C**:1–9.
- Ekins S (2014) Progress in computational toxicology. *J Pharmacol Toxicol Methods* **69**:115–140.
- Ekins S, Clark AM, and Sarker M (2013a) TB Mobile: a mobile app for anti-tuberculosis molecules with known targets. *J Cheminform* **5**:13.
- Ekins S, Clark AM, and Williams AJ (2012a) Open Drug Discovery Teams: a chemistry mobile app for collaboration. *Mol Inform* **31**:585–597.
- Ekins S, Clark AM, and Williams AJ (2013b) Incorporating green chemistry concepts into mobile chemistry applications and their potential uses. *ACS Sustain Chem Eng* **1**:8–13.
- Ekins S, Diao L, and Polli JE (2012b) A substrate pharmacophore for the human organic cation/carnitine transporter identifies compounds associated with rhabdomyolysis. *Mol Pharm* **9**:905–913.
- Greupink R, Nabuurs SB, Zarzycka B, Verweij V, Monshouwer M, Huisman MT, and Russel FG (2012) In silico identification of potential cholestasis-inducing agents via modeling of Na (+)-dependent taurocholate cotransporting polypeptide substrate specificity. *Toxicol Sci* **129**:35–48.
- Gupta RR, Gifford EM, Liston T, Waller CL, Bunin B, and Ekins S (2010) Using open source computational tools for predicting human metabolic stability and additional absorption, distribution, metabolism, excretion, and toxicity properties. *Drug Metab Dispos* **38**:2083–2090.
- Imamura Y, Murayama N, Okudaira N, Kurihara A, Inoue K, Yuasa H, Izumi T, Kusuhara H, and Sugiyama Y (2013) Effect of the fluoroquinolone antibacterial agent DX-619 on the apparent formation and renal clearances of 6 β -hydroxycortisol, an endogenous probe for CYP3A4 inhibition, in healthy subjects. *Pharm Res* **30**:447–457.
- Kido Y, Matsson P, and Giacomini KM (2011) Profiling of a prescription drug library for potential renal drug-drug interactions mediated by the organic cation transporter 2. *J Med Chem* **54**:4548–4558.
- Papadatos G and Overington JP (2014) The ChEMBL database: a taster for medicinal chemists. *Future Med Chem* **6**:361–364.
- Sedykh A, Fourches D, Duan J, Hucke O, Garneau M, Zhu H, Bonneau P, and Tropsha A (2013) Human intestinal transporter database: QSAR modeling and virtual profiling of drug uptake, efflux and interactions. *Pharm Res* **30**:996–1007.
- Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, and Willighagen E (2003) The Chemistry Development Kit (CDK): an open-source Java library for chemo- and bio-informatics. *J Chem Inf Comput Sci* **43**:493–500.
- Williams AJ, Ekins S, Clark AM, Jack JJ, and Apodaca RL (2011) Mobile apps for chemistry in the world of drug discovery. *Drug Discov Today* **16**:928–939.
- Wittwer MB, Zur AA, Khuri N, Kido Y, Kosaka A, Zhang X, Morrissey KM, Sali A, Huang Y, and Giacomini KM (2013) Discovery of potent, selective multidrug and toxin extrusion transporter 1 (MATE1, SLC47A1) inhibitors through prescription drug profiling and computational modeling. *J Med Chem* **56**:781–795.
- Xu Y, Liu X, Li S, Zhou N, Gong L, Luo C, Luo X, Zheng M, Jiang H, and Chen K (2013) Combinatorial pharmacophore modeling of organic cation transporter 2 (OCT2) inhibitors: insights into multiple inhibitory mechanisms. *Mol Pharm* **10**:4611–4619.
- Zheng X, Diao L, Ekins S, and Polli JE (2010) Why we should be vigilant: drug cytotoxicity observed with in vitro transporter inhibition studies. *Biochem Pharmacol* **80**:1087–1092.
- Zheng X, Ekins S, Raufman JP, and Polli JE (2009) Computational models for drug inhibition of the human apical sodium-dependent bile acid transporter. *Mol Pharm* **6**:1591–1603.

Address correspondence to: Dr. Sean Ekins, Collaborations Pharmaceuticals, Inc., 5616 Hilltop Needmore Road, Fuquay-Varina, NC 27526. E-mail: ekinssean@yahoo.com