UNIVERSITY OF CALIFORNIA, SAN DIEGO

The Role of Drug Transporters in Metabolism

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Bioengineering

By

Henry Chow-Heim Liu

Committee in charge:

Professor Sanjay K. Nigam, Chair
Professor Shu Chien, Co-Chair
Professor Ruben Abagyan
Professor Satish A. Eraly
Professor Bernhard Ø. Palsson

2015

The Dissertation of Henry Chow-Heim Liu is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2015

Tables of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank all my family for providing me unconditional support. Thank you for always standing by my sides and cheering me up when I faced all these challenges. Also, I would like to thank my thesis advisor, Sanjay K. Nigam, for his support in my entire research, and thank you for providing guidance and comments to help me complete my thesis work. In addition, I would like to thank all the current and past members from the Nigam lab. Each of you has provided me different knowledge in academia, ranging from cell cultures, tissue engineering, staining, dissection, and so on. Thank you all for teaching me all the techniques so I can survive on my own. Special thanks to Gleb Martovetsky, Xing Zhang, Anne Goldenberg, Saiyee Cho, Christina Lun, Thomas Gallegos, Kevin Bush, Wei Wu. Besides teaching me lab techniques, you guys also help me proof-read my papers and also provide valuable friendship.

I would like to thank to all my collaborators. Thanks to Ruben Abagyan and Winston Chen for providing me helps in computational chemistry analysis; thanks to Bernard Palsson and Neema Jamshidi for providing assistance in systems biology analysis; thanks to Natasha Balac and Paul Rodriguez for providing me helps in data-mining analysis; and thanks to Satish Eraly and Shu Chien for providing me knowledge in transporter function and history and also the physiological concepts.

The work presented in Chapter 2, in full, is being prepared for submission. Liu, Henry C.; Jamshidi, Neema; Chen, Yuchen; Eraly, Satish A.; Cho, Sai Yee; Bhatnagar, Vibha; Bush, Kevin T.; Abagyan, Ruben; Palsson, Bernard O.; Nigam, Sanjay K. The dissertation author was the primary investigator and author of this paper.

The work presented in Chapter 3, in full, is being prepared for submission. The authors are Liu, Henry C.; Goldenberg, Anne D.; Chen, Yuchen; Lun, Christina; Rodriguez, Paul; Balac, Natasha; Abagyan, Ruben; Nigam, Sanjay K. The dissertation author was the primary investigator and author of this paper.

Vita

2015 Ph.D., Bioengineering, UCSD, La Jolla, CA

2009 B.S., Bioengineering, UCB, Berkeley, CA

Publications

Nigam SK, Bush KT, Martovetsky G, Ahn SY, **Liu HC**, et al. “The organic anion transporter (OAT) family and its role in remote sensing and signaling.” Physiol Rev 95, 83-123.

Wu W, Jamshidi N, Eraly SA, **Liu HC**, et al. “Multispecific drug transporter Slc22a8 (Oat3) regulates multiple metabolic and signaling pathways.” Drug Metab Dispos. 2013 Oct;41(10):1825-34.

Abstract Of The Dissertation

The Role of Drug Transporters in Metabolism

by

Henry Chow-Heim Liu

Doctor of Philosophy in Bioengineering

University of California, San Diego, 2015

Professor Sanjay K. Nigam, Chair
Professor Shu Chien, Co-Chair

In this dissertation, a multi-leveled analysis is performed to understand the physiological functions of several key SLC/ABC transporters. First, a single transporter was studied, and then, the scope of study was extended to a set of 4 transporters; and finally 12 transporters were studied as a group.

First, using the systems biology analysis and computational chemistry analysis, the function of OAT1 is individually studied in the context of cellular metabolism. This

two-layered computational method not only predicts new OAT1-binding metabolites but also provides a clear picture how the function of OAT1 is involved in cellular metabolism. In addition, based on this multi-scaled computational method, the construction of an OAT1-centered metabolic network is achievable. In this network, metabolites known to or predicted to interact with OAT1 are shown along with their immediate first neighbors in metabolic reactions. Identified in this network are metabolites which are direct substrates of OAT1 and metabolites which are indirectly affected by OAT1 (through metabolic cascading effects). In addition, since metabolites within this network can be grouped according to which metabolic pathways they are part of, one can also understand how OAT1 is involved in individual metabolic pathways.

In addition to OAT1, other SLC transporters known to handle organic compounds are OAT3, OCT1, and OCT2. With the tools from computational chemistry and machine learning, a systemic ligand-based approach was used to understand the substrate overlap and substrate preference between transporters, which helps us study how these transporters work in a systemic fashion.

Finally, to understand how the major ABC/SLC transporters are involved in cellular metabolism, metabolic networks are constructed for the 12 ABC/SLC transporters; the resulting networks demonstrate how individual transporters are involved in various aspects of cellular metabolism and how they work, collectively, to contribute to the whole-body hemostasis.

In conclusion, a wide variety of tools are used to study the transporting functions, structural differences, and physiological significance of the major ABC/SLC transporters. Using the combination of these tools provides a novel and multi-scaled way to study the transporters.

Chapter 1 : Introduction

1.1 Overview of major ABC/SLC transporters

ATP-binding cassette (ABC) transporters and solute carrier (SLC) transporters are important drug transporters responsible for the absorption, distribution, metabolism, and elimination (ADME) of pharmaceutical drugs. Even though these transporters have been studied mainly in the context of pharmacology, both clinical and animal data have shown the importance of these transporters in cellular metabolism and the whole-body physiology [1-3].

1.1.1 Polarized expression of transporters in the barrier epithelium

The ABC/SLC transporters are expressed in different epithelia tissues that function as barriers, such as in the kidney, liver, intestine, brain, and placenta. Cells expressing these transporters are polarized, and have a basolateral and an apical side. The basolateral side of the cell sits on the basement membrane, which is a thin, fibrous layer made of extracellular proteins, and faces toward the interstitial fluid; the apical side of the cell faces inward to the lumen. Most ABC/SLC transporters are expressed on only one side of the cell membrane. For example, OAT1 and OAT3 are basolateral transporters which are known to be expressed on the basolateral side of proximal tubule cells in the kidney and on the basolateral side of choroid plexus cells in the brain.

1.1.2 Influx and efflux transporters

The ABC/SLC transporters are also classified as either influx transporters or efflux transporters depending on the direction they transport substrates. The influx

transporters uptake substrates from the extracellular space into the cellular domain, whereas the efflux transporter export substrates from the cell to the extracellular space. OAT1, OAT3, OATP1B1, OATP1B3, OCT1, and OCT2 are influx transporters whereas BCRP, MDR1, MRP2, MRP3, MRP4, MATE1, and MATE2 are efflux transporters.

1.1.3 Pairing of influx and efflux transporters in transport

Influx transporters and efflux transport work together in a serial fashion in order to complete the vectorial transport of substrates from one domain to another. For example, in the kidney, OAT1 and MATE1 pair together in function. OAT1, expressed on the basolateral side of proximal tubule cells, uptakes substrates from the blood into the cell, and once it has entered the cell, substrates are then exported by MATE1 on the other side of the cell into the urine. Thus, the directional transport of compounds through epithelia barriers is possible through the pairing of this two-stepped transport.

1.1.4 The multi-specific nature of the transporters and the substrate preference and overlap

The ABC/SLC transporters are known to interact with a variety of substrates, ranging from pharmaceutical drugs, metabolites, toxins, and waste products. These substrates are of diverse chemical structures. The capability to transport such a wide range of compounds demonstrates the multi-specificity of the transporters. Additionally, each transporter handles a set of substrates. Transporters with similar names, and thus having similar sequence and structures, handle like substrates. For example, OAT1 and OAT3 handle similar sets of organic anions found in urine; OATP1B1 and OATP1B3 handle bile-acid-associated anions; and OCT1 and OCT2 handle organic cations.

1.1.5 The expression of the transporters in different epithelial barriers

Transporters are expressed in different epithelia tissues. OAT1 is mainly expressed in the kidney and in the choroid plexus; OAT3 is found to be more universally expressed and is seen in kidney, brain endothelium, choroid plexus, retina, and testes; OATP1B1, OATP1B3, and OCT1 are predominantly expressed in liver; and BCRP and MDR1 are also found to be expressed in many tissues, including intestine, liver, and kidney [2, 3].

1.1.6 The physiological role of the transporters

A growing pool of data has shown the physiological and endogenous importance of these transporters. In addition to transporting pharmaceutical drugs, these transporters are also found to transport metabolites, antioxidants, microbiome products, nutrients, bile acids, hormones, neuro-transmitters, and signaling molecules. Metabolites important in different metabolic processes are transported in order to participate in the metabolic factory. Without these transporters, metabolic processes could be suspended or, and the cellular metabolism would be disrupted.

1.1.7 Clinical relevance

Realizing the importance of these transporters in physiology, it is no surprise that several clinical issues and syndromes are associated with transporters. For example, genetic polymorphisms of MDR1 are associated with inflammatory bowel disease [4]; genetic mutation of BCRP is associated with hyperuricaemia [5]; genetic mutation of OATP1B1 leads to hyperbilirubinemia [6]; and genetic mutation of MRP2 leads to Dubin-Johnson syndrome [7].

1.2 Experimental approaches to study the transporters

A combination of chemistry, molecular biology, and genetic engineering tools has been used to better understand the structures and functions of these transporters.

1.2.1 In-vivo studies

In-vivo studies are the most straightforward methods to understand the physiological and pharmacological roles of the transporters. In humans, clinical data of single nucleotide polymorphisms (SNPs) of transporter genes are used to see how the malfunction of the transporters due to genetic mutation would affect the metabolic state and drug handling. In animals, knockout studies examine how deletion of the transporters affects animals. These knockout animals are then compared with wild-types ones, and the comparison studies include observing physical states and profiling urine/plasma metabolomics. Metabolomics profiling of *Oat1*-knockout animals revealed a handful of metabolites to be significantly altered [8, 9]. Moreover, it was found that the knockout of *Oat1* would prevent animals from mercury-induced kidney injury [10]. In another study, *Oat3* knockout studies support the importance of the transporter in bioenergetics pathways and signaling pathways [11]. Knockout studies of *Mrp3* identified several phytoestrogen-glucuronides to be altered in plasma/urine of the knockout animals [12].

1.2.2 In-vitro cell assays

Compared to in-vivo studies, in-vitro cell assays provide a more controlled environment to study transporter function. Transfected cell lines and *Xenopus laevis* Oocytes are utilized. Cell lines transfected with the transporter gene of interest are cultured as a monolayer for the transport assay. In the transport assay, usually a fluorescent tracer or radiolabeled compound, which is known to be a substrate of the

transporter, is added along with a compound for testing to determine if the tested compound inhibits the transport of the tracer.

1.3 In silico approaches to study the transporters

In silico methods add another layer on top of the experimental data and help us to understand the functions of the transporters better. Using a combination of systems biology, computational chemistry, and machine learning approaches, physiological and pharmacological roles of the transporters can be unveiled.

1.3.1 Systems biology

Genome-scaled metabolic reconstruction (GEM) is one prominent systems biology tool used to simulate and understand the metabolic state of an organism. Using the mathematical balance of equations, the fluxes of individual metabolic reactions are predicted based on –"omics" data, either transcriptomics, metabolomics, or both. The very first multicellular model was reconstructed for *C. elegans* in 1998 [13], and a few years later, mouse and human models were reconstructed [14, 15]. The GEM analysis was first applied to study OAT1 in the context of cellular metabolism [16], and it was revealed that OAT1 was significantly involved in many metabolic pathways. In addition, the GEM analysis also predicted a handful of metabolites, previously not identified to be OAT1 substrates, and follow-up in-vitro cell assays confirmed these as novel OAT1-binding metabolites, such as spermine and spermidine. GEM analysis for OAT3 was also done, and it was shown that this transporter was linked to multiple metabolic and signaling pathways [11].

1.3.2 Computational chemistry

Several approaches have been undertaken to perform computational chemistry analysis on the organic anion transporter (OAT) family. These include structure model

building and quantitative structure-activity relationship (QSAR) analysis. The structure model building method attempts to build models in a chemical 3D space, and two approaches are available, the protein-based approach and the ligand-based approach. The protein-based approach attempts to recapitulate the 3D structure of the transporters themselves. On the other hand, the ligand-based approach focuses on the ligands that interact with the transporters, and information about common features of the ligands are used to build pharmacophore models. Instead of studying the structure of ligands and transporters as a whole, QSAR analysis focuses on finding individual physiochemical properties of ligands one by one and identifying the correlations between these molecular, atomic properties and the substrate affinity.

In the protein-based structure building approach, because the crystal structure of OATs is lacking, other transporters with known crystal structures that were highly similar to OATs were first used as templates. In one study, fold recognition analysis was first used to find a template transporter that exhibited the most similar secondary structures with hOAT1, and the result showed that the glycerol 3-phosphate transporter (GlpT) from Escherichia coli shared similar 3-dimensional folds with hOAT1, despite having low sequence alignment with hOAT1. The crystal structure of GlpT was then used as the template to construct hOAT1 structure model. With this model, a putative active site, positioned on a central cavity, was found to be the main site for the substrate-transporter interaction [17].

After the hOAT1 protein model was created, it was compared with the models of the rabbit organic cation transporter 2 (rbOct2), built based on the homology model of the glucose transporter 1(glut1) [18] and the rat organic cation transporter 1 (rOct1), built based on LacY [19], and the comparison showed that even with many structural and sequence alignment similarities, small differences remained. A sequence alignment

done on these three protein sequence shows that the amino acid residues surrounding the putative active site of hOAT1 (Met207, Phe438, and Arg466) corresponded to those around the active cavity of OCT1/OCT2 models (Tyr222/Tyr221, Leu447/Tyr446, and Asp475/Asp474). The putative active site of hOAT1 contained more basic amino acids; in contrast, the active cavity of OCT1/OCT2 retains more acidic amino acids. The differences in amino acid composition surrounding the active sites perhaps contributed to substrate specificity among transporters.

The 3-dimensional structure models not only predicted the potential active site of the transporter but also identified amino acid residues surrounding the cavity that were important in substrate recognition. The important role of these amino acids was validated with wet-lab site-directed mutagenesis, and it was found that the mutations of these amino acids would result in decrease or abrogation of chemical interaction with substrates originally known to chemically bind with the transporters. Also, it was shown that a mutation in a single amino acid surrounding the active site had different impacts on different substrates, suggesting that there existed multiple binding domains within a single active cavity. These studies support the multi-specificity nature for OATs and OCTs.

These models for the transporters can also help us understand the transport mechanism. For instance, the hOAT structure model built based on the homology model of GlpT helped to clarify the detailed mechanism of transport [20], and it was found that hOAT1, consisting of 2 hemidomains, had the two hemidomains tilting to each other when the substrates entered the central cleft from the extracellular side of the transporter.

Pharmacophore modeling, a ligand-based structure building approach, was used to build the common physiochemical patterns of substrates in 3D space, and these

common patterns were then used to predict and identify novel substrates. Pharmacophore models for OAT1 and OAT3 were built by using different set of ligands and different software. A pharmacophore model of mOAT3 based on four organic cations was built in the Catalyst software [21]. These four organic cations are drugs that have Ki values for mOat3 at least ten times smaller than those for mOAT1, making them more mOAT3 specific. The designed pharmacophore defines four important features, including two hydrogen bond acceptor features, one hydrophobic feature, and one positive ionizable feature. This pharmacophore emphasizes the importance of having hydrogen bond acceptors within the substrates for mOAT3 binding, which corresponds to the prior finding in 1D-QSAR antiviral drug study [22]. Finally, this pharmacophore was used to screen a commercially available compound database, and seven cationic molecules are found to be good fit with the model. Wet-lab experiments validate four to exhibit great inhibition affinity with mOAT3. Three of the four drugs also chemically interact with mOAT1; however, their mOat1 Ki values are at least fifty times larger than their mOAT3 Ki values. Thus, despite showing many structural similarities, OATs exhibit various binding domains that make them interact with different substrates. Moreover, by selecting the right substrates for pharmacophore definition, one can design models that are more specific to a particular transporter which may helping us to design drugs that only or preferentially interact with one type of transporter.

In addition to building OAT3 pharmacophore models, mOAT1 Pharmacophore models were built by using a set of substrates known to interact with mOAT1 [9, 23]. In two studies, the model generation utilized the results of the in-vivo metabolomic data. In the metabolomic studies, some metabolites found to have significantly altered concentrations in either plasma, urine, or both between the Wild-type and the *Oat1* Knockout mice were identified [8, 9]. These metabolites were used to generate models.

One pharmacophore modeling study used 19 metabolites identified from targeted metabolomics using model generation [23], and the other model was built based on the 15 metabolites identified by an untargeted metabolomic study [9]. In both studies, the modeling was done using MOE software. In these two methods of pharmacophore model studies, the first step was the process of clustering the metabolites into groups, in which members within the groups have structural similarities. The clustering step was necessary because the selected metabolites were diverse in their chemical structures. Then, a pharmacophore model was created for each clustering group. Then, the pharmacophore models were used to screen against a known chemical database, and compounds that fit the models well were selected for further in vitro validations. Twenty compounds from the commercially available database were a good fit with pharmacophore models generated from targeted metabolomics, and these compounds were verified to have a strong chemical interaction with OAT1. The untargeted metabolomics-driven pharmacophores models were used to screen against the NCI database, and many compounds were selected that satisfied the criterion of the designed pharmacophores. Within this top hit list, a compound with the lowest total self-consistent field energy (meaning that this compound has the highest probability to retain the specific conformer from chosen in the chemical simulation), was selected and verified to have good inhibition affinity with mOAT1.

Aside from building the separate models for OAT1 and OAT3, a joint OAT1-OAT3 pharmacophore model was built based on 16 clinical drugs that demonstrated strong in-vitro inhibition with OAT1, OAT3, or both [24]. This pharmacophore model identified the common features shared among the selected OAT1 inhibitors and Oat3 inhibitors , which included a hydrogen bond acceptor region and an aromatic center (the two features were separate by a distance of 5.7 Angstroms).

In addition to focusing on transporter/ligand structures as a whole, the molecular/atomic physiochemical properties of the ligands can be analyzed individually. This was done in QSAR analysis, in which the physiochemical properties of ligands were quantitatively measured, and their correlations with substrate affinities were determined. In one study, over 40 molecularly diverse organic anions that are known to be OAT1 and OAT6 substrates were studied. Their molecular properties, such as LogP (hydrophobicity) molecular weights, and net charges were determined, and the correlations between these properties and the substrate affinity for OAT1 and OAT6 were studied [25]. Both OAT1 ligands and OAT6 ligands shared several similar 1D physiochemical properties. For example, a positive correlation between LogP and substrate affinity for both the transporters was found, explaining the importance of hydrophobicity for both OAT1 and OAT6 recognition. Nonetheless, differences in other properties seemed to differentiate between OAT1 substrates and OAT6 substrates. For example, OAT6 tend to interact with mono-anions (compounds with one anionic functional group) with molecular weights below or around 150.

In addition to using 1D QSAR, 3D-QSAR was applied to study how the physiochemical properties of the anions affect their binding affinities in 3-dimensional space. Comparative molecular field analysis considering the steric and electrostatic potential of the anions confirmed the importance of both hydrophobic and charged interaction in OAT1-substrate recognition; however, hydrophobicity was a more predominate factor than charged states in OAT6 binding. In another study, 1D-QSAR is used to study nine antiviral drugs, known to be substrates for mOAT1, mOAT3, and mOAT6 in order to determine antiviral preference between these three OATs [22]. Some physiochemical properties, such as molecular weight, LogP, number of rotatable bonds, number of hydrogen bond donors, number of hydrogen bond acceptors, polar surface

area (PSA), and the molar refractivity were determined for these antiviral drugs by using Sybyl software and the Chemaxon package. Then, the multivariable linear regression analysis was used to determine how these properties affect the binding affinity of each of the transporter. It was found that these three transporters exhibited different preference in substrate selection. mOAT1 substrates are characterized by higher polar surface area; mOat3 substrates are found to link with more hydrogen bond acceptors and low rotatable bond numbers; and mOat6 prefers less hydrogen bond donors.

1.3.3 Machine learning

Machine learning methods have been widely used in different fields in an attempt to “learn” or identify patterns from large data sets. The methods include support vector, neural network, decision tree, and random forest. These methods have been widely used in various fields. In business world, these tools are applied to study the market trend and customer product relationship; in the setting of industrial engineering, people applied these methods to make decisions to optimize the production process. Recently, these methods have also been applied in the chem- and bioinformatics fields. For example, they are used to study the correlation between DNA sequence and diseases, and artificial neural network was used to identify miRNA biomarkers associated with breast cancer [26].

Machine learning methods can also be used to study the protein-ligand interaction. With physiochemical properties calculated for ligands using computational chemistry software, a large amount of data is generated. It was ypothesized that this large data set retains ecessary information to understand the interaction between ligands and their binding proteins, and it was further believed that by applying machine learning methods, such hidden information would be unveiled. In the transporter field, such methods can be used to generate models that predict if a given compound is a

substrate or non-substrate for a particular transporter. For example, the support vector machine method (SVM) has been used to develop models for two ABC transporters, breast cancer resistance protein (BCRP) and P-glycoprotein (P-gp) [27, 28]; these models were mainly used for the in-silico prediction of new substrates, which had prediction accuracy greater than 70%.

Chapter 2 : Construction and Evaluation of an Organic Anion Transporter-1 (OAT1) Centered Metabolic Network

2.1 Abstract

Genome-scale metabolic reconstruction of transcriptomic data derived from a comparison of wild-type and organic anion transporter-1 (*Oat1*) knockout tissues, together with virtual screening using pharmacophore hypotheses created from known OAT1-interacting drugs, predicted a number of metabolites potentially affected by deletion of the transporter. A set of these metabolites were validated in wet-lab assays of OAT1 function. Re-evaluation of the transport and knockout metabolomics data generated a largely experimentally validated OAT1-centered metabolic network that included several key metabolic pathways. Since many of the network metabolites have been implicated as “uremic toxins in chronic-kidney disease (CKD),” a CKD-associated metabolic network was constructed. This overlapped significantly with the OAT1-centered network, suggesting the importance of OAT1 activity in modulating CKD-induced metabolic changes. This novel combined systems biology/computational chemistry approach should be applicable to other SLC and ABC drug transporters to analyze how they may regulate normal and disease metabolism individually and as a group.

2.2 INTRODUCTION

Despite their well-established roles in the absorption, distribution and elimination of xenobiotics [1], there is growing evidence for the transport of metabolites, nutrients, signaling molecules, antioxidants and other molecules of physiological importance at the organismal and cellular levels by members of the SLC and ABC “drug” transporter families [2, 29, 30].

Among these transporters, organic anion transporter 1 (OAT1/SLC22A6), originally dubbed NKT [31, 32], mediates the rate-limiting steps in the renal elimination of organic anionic drugs and a few cationic drugs [3, 21, 33, 34]. The pharmacological and toxicological relevance of OAT1 has been supported by *in vivo* and *ex vivo* studies in knockouts [10, 22, 35-37], which have also provided new information about the potential role of OAT1 in basal physiology. For example, metabolomic studies comparing the plasma and urine of wild-type (WT) and knockout (KO) mice revealed significant changes in the concentrations of a number of endogenous metabolites [8, 9].

Taken together with the fact that OAT1 and other SLC22 transporters have been implicated in the transport of key biochemical pathway intermediates, as well as in metabolic diseases [3, 16], this suggests an important, if underappreciated, role for SLC “drug” transporters in metabolic processes. Perhaps because of their key role in pharmacokinetics, these transporters are not generally depicted in biochemical pathways involving the metabolites they transport. Such an omission could have clinical significance, with drugs directly or indirectly affecting pathways normally involved in the movement of key metabolites, pathway intermediates, and signaling molecules, thereby fundamentally affecting cell and organ physiology. If so, this might help explain metabolite-drug interactions and perhaps even aspects of certain drug-induced

metabolic syndromes (e.g., those seen with diuretic use or chronic HIV antiviral treatment). It may also be useful for further defining the role of drug transporters in modulating common metabolic diseases, such as diabetes and kidney diseases [2, 38, 39].

Thus, it is important to begin to build a detailed map of metabolism affected by "drug" transporters such as OAT1. In order to address this question, a combined systems biology and computational chemistry approach was employed (Figure 2.1). Transcriptomics data derived from wildtype (WT) and *Oat1*-deficient mouse kidneys was analyzed and used to build a mathematical reconstruction using iMM1415 [40], a mouse genome-scale metabolic reconstruction (GEM) [14], which predicted the involvement of ~150 metabolites represented in several biochemical pathways. Some predictions were then validated with untargeted and targeted metabolomics data obtained from the *Oat1* knockout mouse [8, 9]. These metabolites were quantitatively assessed for their likelihood to bind OAT1 by virtual screening using pharmacophore models based on drugs known to be handled by this transporter. Novel metabolites predicted to interact with OAT1 were assayed using *in vitro* transporter assays, and based on these analyses, an "OAT1-centered metabolic network" was constructed to reflect those metabolites with the ability to interact with OAT1. In this OAT1-centered network, about thirty percent of the metabolites are validated by *in vitro* data and/or *in vivo* data. This network was further evaluated in the context of metabolic changes occurring in chronic kidney disease (CKD), which can cause dramatic alterations in metabolites, including uremic toxins thought to interact with OAT1 [9]. A strong overlap was observed between the CKD-associated metabolic network and the OAT1-centered metabolic network, raising the possibility that targeting the expression and/or function of OAT1 may be useful for modulating some of the metabolic alterations of CKD. Moreover, the approaches

employed can be applied to other SLC and ABC drug transporters to define how they separately and collectively regulate systemic metabolism in health and disease.

2.3 RESULTS

The goal here was, first, to apply a systems level analysis of transcriptomic data to predict metabolites potentially interacting with OAT1 (either directly or indirectly) and, then, to apply computational chemistry constraints to differentiate those metabolites that would interact with OAT1 directly from those interacting indirectly. Finally, this two-tiered systems level analysis, together with metabolomics (from knockout and wild type animals) and transport data from Oat1-transfected cells was used to build an as detailed as possible OAT1-centered metabolic network (Figure 2.1).

2.3.1 Comparison of the WT and KO metabolic reconstructions generated based on iMM1415, a mouse genome-scale metabolic reconstruction model (GEM)

Context specific wild-type (WT) and knockout (KO) models were constructed from gene expression data using the Gene Inactivity Moderated by Metabolism and Expression (GIMME) algorithm [41] based on the mouse GEM, iMM1415 [40]. This systems biology tool simulates the metabolic state of a tissue/organ based on transcriptomic data and the execution of specific biological processes (i.e., objective functions). Such simulation was done by converting the metabolic network into mathematical equations followed by the application of linear programming to calculate a solution of fluxes (i.e., measurement of rate of production or depletion of metabolites) for each of the reactions. The biomass function, which is commonly used for the generation of eukaryotic cell/tissue models [42, 43], was chosen as the objective function in the

generation of context-specific OAT1 WT and KO models. The resulting WT and KO models consisted of 2233 and 2143 reactions, respectively.

Flux variability analysis (FVA) [44], which measures all the boundaries of solution spaces, was used to compare metabolic differences and to find the maximum and minimum flux values for each reaction in the network, enabling the calculation of flux span ratios of the KO over WT for each reaction. Reactions with flux span ratios equal to 1 indicate no change in reaction activities due to the deletion of *Oat1*; flux span ratios less than 1 imply decreased reaction activities while ratios greater than 1 indicate increased reaction activities (Table S1). The reactions that require transport of metabolites would be expected to change in the OAT1 KO, and 1026 reactions with altered activities were identified, including 321 exchange/transport reactions, which were responsible for the handling of 177 metabolites. After excluding water and other uninformative molecules (see Methods), 146 metabolites remained (Table S2), which were predicted to be linked to OAT1-mediated transport.

2.3.2 Validating the models by metabolomics data from the Oat1-KO

Multiple approaches were used to validate the metabolic reconstruction. Metabolomics data from the *Oat1*-KO were interrogated to determine whether experimental observations corresponded with computational predictions. Previous metabolomics profiles identified 38 metabolites with significantly altered plasma and/or urine concentrations between the WT and *Oat1*-KO (Table S3) [8, 9]. These metabolites were assumed to be interacting with OAT1, either directly (as ligands) or indirectly via secondarily altered metabolic pathways. Among the 38 metabolites, 19 were present in the metabolic reconstruction model —of the remaining 19, some were derived from the

gut microbiome [9], reactions for which are generally not included in the mammalian GEMs. To determine how well the models performed, the 19 metabolites that were, in fact, present in iMM1415 were examined; 10/19 were correctly predicted to change in the knockout model (Figure 2.2B). Thus, about half of the metabolites predicted to change in concentration by the metabolic reconstruction were actually found to change *in vivo*. Thus, it appears that the systems biology model built from the transcriptomics data makes reasonably good predictions of the *in vivo* metabolic differences between WT and KO.

2.3.3 Using pharmacophores as chemical constraint-based filters

Before performing wet lab validation on the 146 predicted metabolites, we aimed to define which of these were likely to directly interact with OAT1. We reasoned that the chemical features of known OAT1-transported drugs could be used to identify likely OAT1-transported metabolites, which could then be prioritized for later wet lab validation to assess direct interaction with OAT1. Therefore, OAT1 pharmacophore models based on known drug ligands were built.

In order to construct pharmacophore models for OAT1, 61 drugs having a published K_m or K_i less than 100 μM for OAT1 were selected as “actives” for model building and model validation; two third of the group is used to build the models, and one third (20 drugs) of actives is split to validating set. (Table S5). Since the drugs possess diverse chemical structures (consistent with the multi-specific nature of OAT1), they were clustered into groups using their atomic property fields (APF) (e.g., hydrogen bond donors, hydrogen bond acceptors, SP2 hybridization, lipophilicity, sizes of large atoms, positive and negative charges, etc) [45]. Thus, the training actives were grouped into 7

distinct clusters (Figure 2.3A) and pharmacophore models were built for each based upon the alignment of its members. Figure 2.3B demonstrates how members of cluster 1 were first aligned, and “pharmacophore model 1” was built to represent the 3D atomic properties shared among the members of that cluster. Then, the pharmacophore models were validated based on the validating set (known positives) and drugs from Drugbank database (serving as true negatives), and a ROC curve was generated (Figure 2.3D), and the calculated area under curve (AUC) was 80.58, demonstrating that our pharmacophore system was well validated.

The 7 pharmacophores (Figure 2.3C) were then used as 3D chemical space constraints on the predicted metabolites. 74 out of the 146 metabolites satisfied the constraints and were thus predicted to have direct interaction with OAT1. Of these metabolites, 18 have direct interaction with OAT1 based on previous experimental observations. Compared to the original list of 146 metabolites (prior to pharmacophore filtering) the percentage of metabolites known to have direct interactions was enriched about two-fold (from 14.4% to 24.3% after filtering).

2.3.4 Wet-lab validation and identification of novel OAT1 ligands

Based on their ability to fit the pharmacophore models, a number of metabolites computationally predicted to “directly interact” with OAT1 were then validated in wet-lab transport assays. Of the 8 metabolites that were not tested previously for their interaction with OAT1 were selected for further validation, four interact with Oat1 in transfected cells (Figure 2.4A). These metabolites were dihydrofolic acid, palmitoleic acid, 16-hydroxy-hexadecanoic acid, and prostaglandin E1 with calculated K_i values of 93 μM, 200 μM, 13 μM, and 12 μM, respectively (Figure 2.4B); the values are well within the documented range for many substrates of OAT1.

2.3.5 Demonstrating the enrichment power of our two-tiered in-silico approach and our overall approach

To determine how good our in-silico approach was in enriching metabolites known to have direct interaction with OAT1, a hypergeometric test is applied. Before the two-tiered in-silico prediction, we have a total of 1503 metabolites to start with (iMM1415 database), and among them, 59 metabolites are positives, or ones known to have direct interaction with OAT1 (the 4 newly identified metabolites from our wet-lab experimentation are also included). After applying the two-tiered in-silico approach, it gives us a total of 74 metabolites (top-ranked GEM-predicted metabolites), and among them 22 are positives (again, including the 4 newly identified ones). Then applying the hypergeometric test (Population size: 1503; number of successes in population: 59; sample size: 74; number of successes in sample: 22), the p-value was calculated to be 1.53e-15. Such small p-value shows that our two-tiered in-silico approach is significant in enrichment of “positive” metabolites.

In addition, this statistical test is also used to determine if our overall prediction and identification of novel OAT1 metabolites, the combined in-silico and in-vitro approach, is significant in enrichment. After the two-tiered in-silico prediction, 8 metabolites were tested for wet-lab validation, and 4 of them were found to “positives”. Applying the hypergeometric test (Population size: 1503; number of successes in population: 59; sample size: 8; number of successes in sample: 4), the p-value is found to be 1.34e-4.

2.3.6 Construction of an OAT1-centered metabolic network

In order to link OAT1 to multiple metabolic pathways, a network was built based on the results of the aforementioned systems biology/pharmacophore approach and the

wet-lab data (validated here or published previously), using Metscape, a Cytoscape plugin used to construct and visualize metabolic networks based on the KEGG database [46]. The resulting broader OAT1-centered metabolic network (see materials and methods) consisted of a total of 253 metabolites, including 176 experimentally validated and/or computationally predicted and 77 “plus-one” (directly connected) metabolites (Figure 2.5A). Of these 176 metabolites, 74 had wet-lab support for interactions with OAT1, either by *in vivo* metabolomics from the knockout (Table S3) or *in vitro* assays that were performed in this study or published (Table S4). These metabolites were thus placed in the group termed “wet-lab” support (Table S7) and had the highest level of confidence for being part of an OAT1-centered metabolic network.

Three other groups of metabolites were included—in order of level of confidence (Table S7). The metabolites with the next level of confidence were those first predicted by GEM and which also passed pharmacophore filters; these were termed “metabolites with high confidence of interacting with OAT1” (Table S7). Metabolites only predicted by GEM (but having structures such that they did not pass the drug-based pharmacophore filters with high confidence) were classified as “metabolites likely to be affected indirectly”. Finally, the remaining “plus-one” metabolites were termed “OAT1-first-neighbor-compounds” (Figure 2.5B). Together, these metabolite nodes were connected by 289 edges. The network revealed that, in the revised OAT1-centered metabolite pathways, the majority of metabolites were interconnected to constitute a main component, and there were also a number of small self-connected components (Figure 5A); some of network parameters were measured and are shown in Figure 2.5C. The 253 metabolites within the network participated in more than 20 different designated metabolic pathways, suggesting the broad importance of OAT1 in metabolism. These pathways included carbohydrate metabolism (e.g., TCA cycle, galactose metabolism,

etc.), lipid metabolism (glycosphingolipid metabolism, bile acid biosynthesis, etc.), amino acid metabolism (alanine, aspartate and glutamate, etc metabolism), nucleotide metabolism (purine and pyrimidine metabolism), and cofactor and vitamin metabolism (vitamin A, B2, B3, B5, B6, and B9 metabolism) (Figure 2.5A).

The most highly represented metabolic pathways having at least 7 metabolites are shown and ranked according to the validation percentage in Figure 2.5D (validation percentage is equal to the number of "wet-lab supported" metabolites of the pathway divided by the total number of metabolites in the pathway). Among these well-represented pathways, the two with the highest validation percentage were tyrosine metabolism and TCA cycle, (58.3% and 57.1%, respectively). The TCA cycle is noteworthy, since it includes metabolites known to be classical substrates of OAT1, such as α-ketoglutarate, citrate, fumarate, and succinate [3, 33]. The next top-ranked pathway was tryptophan metabolism, which had considerable wet-lab support for 10 of 19 metabolites (52.6%), including anthranilate, xanthurenic acid, kynurenine, and indole-acetic acid, which are also putative uremic toxins associated with CKD [9, 47, 48].

2.3.7 Network analysis of the OAT1-centered metabolic network

NetworkAnalyzer, another Cytoscape plugin, was used to measure the global and local network parameters of the network. The whole network consisted of 32 connected "components" (253 nodes), including one main component (170 nodes), in which the majority of metabolites were interconnected, and 31 other small self-connected components. To determine if our networks were either random or scale-free, degree distribution, *P(k)*, was measured, which calculates the probability of a selected node *n* having exactly *k* degrees (degree of a node is equal to the number of links it has to other nodes) [49]. Networks with nodes that follow a Poisson distribution, in which most nodes have almost the same number of degrees, are said to be random networks;

in contrast, networks that follow a power law, where $P(k) \sim k^{-\gamma}$, are called scale-free networks; most biological networks are thought to have a power law distribution with γ ranging between 2 and 3 [50]. The OAT1-centered metabolic network was found to follow the power law distribution (Figure 2.6A) and had γ value of 2.151; thus, our network, like most biological networks, was not random, but was scale-free. In a scale-free network, most nodes have only a limited number of links, and a few nodes have a very large number of links (these highly connected nodes are also called "hubs") [50]. To identify the hubs, we measured the "degree" of individual nodes to investigate their local impact in the main component of the network since the main component was considerably more complex than the rest of the network. Metabolites with high degree, were found, including pyruvate, L-glutamine, L-glycine, L-glutamate, and L-serine (Figures 2.6B, 2.6C). Pyruvate is not only known to interact with OAT1 *in vitro* [16], but it also serves as the precursor for the TCA cycle, which is the classical energy pathway linked with OAT1 function. In addition, the amino acids with high degrees are also important precursors for several metabolic pathways, and their importance in metabolism, as well as in the OAT1-centered network demonstrates, again, the significance of OAT1 in cellular metabolism.

To identify "hub" metabolic pathways, the degree values for metabolites from metabolic pathway having at least 7 metabolites were averaged (Figure 2.6D). The main "hub" metabolic pathways (with average degree greater than 3.5) included vitamin B9 (folate) metabolism; urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine; glycine, serine, alanine and threonine metabolism; and TCA cycle (Figure 2.6D).

2.3.8 Metabolic pathways altered in chronic kidney disease (CKD)

CKD causes “uremic syndrome,” which affects many metabolic pathways [51, 52] and is thought to be partly due to the accumulation of so-called uremic toxins. OAT1 has been implicated in the metabolic abnormalities associated with diabetic nephropathy [38], and *Oat1* levels change in animal models of kidney disease [53-56]. Some of the small molecule uremic toxins of greatest clinical concern (e.g. indoxyl sulfate, p-cresol sulfate, kyurenine) are known OAT1 substrates; they also accumulate in *Oat1* knockout mice [9]. The recent application of targeted and untargeted metabolomics to patients with CKD and to animal models for this disease has made it possible to produce a list of metabolic alterations associated with CKD [57]. This list includes metabolites involved in key metabolic pathways, uremic toxins, and molecules considered to be biomarkers of CKD. We used this list to generate a putative CKD-associated metabolic network. After trimming, this CKD network contained a total of 322 metabolites (Figure 2.7).

2.3.9 Overlap of the OAT1-centered metabolic network with the CKD-associated network

As mentioned above, OAT1 has been implicated in CKD. To evaluate the potential role of OAT1 in the metabolic alterations of CKD, the OAT1-centered network was compared with the CKD network (Figure 2.8). It was found that the two networks overlapped greatly. While no causality is implied, a role for OAT1 is thus supported in the modulation of many metabolites and metabolic pathways affected in CKD. There were 113 overlapping metabolites between the models (44% of metabolites in the OAT1-centered metabolic network). Applying the hypergeometric test, such overlap was found to be significant (the p-value was calculated to be 2.58e-95). Among the overlapped metabolites, 35 metabolites are known uremic toxins. Among them, some are known to

directly interact with OAT1 [9, 58]; these included xanthurenate, xanthine, L-citrulline, and hippurate. Apart from the "wet-lab supported" group of metabolites, other overlapping metabolites (e.g. sphinganine, hexadeocanoic, and dopamine) from the OAT1-centered metabolic network were predicted by the GEM reconstruction.

2.4 DISCUSSION

We have described, validated, and shown disease relevance of an OAT1-centered metabolic network. This drug transporter is also the focus of regulatory agencies concerned about side effects of drugs due to interaction at the level of the transporter [2]. Indeed, it is becoming increasingly clear that OAT1 and other drug transporters—which are differentially and highly expressed in various epithelial tissues lining body fluid compartments, as well as endothelial and other cells and are also highly conserved through evolution—play key roles in regulating levels of endogenous metabolites [3, 9, 11, 16].

Only a limited number of endogenous substrates of these "drug" transporters have been identified, which has made it more difficult to identify the metabolic networks in which these transporters might be involved. Furthermore, metabolic reconstruction has been hampered in most cases by relevant 'omics' data and computational methods, but OAT1 is a notable exception. OAT1 is largely responsible for the uptake of anionic substrates from the blood and their elimination into the urine, from the viewpoint of organ physiology, any change in the function of OAT1 (e.g., *Oat1* knockout) would concomitantly alter the concentration of metabolites transported by OAT1 not only in the bloodstream, but in the proximal tubule cells of the kidney as well, which could result in alterations in gene expression. Thus, with the availability of transciptomic and metabolomics (untargeted and targeted) data from the *Oat1* knockout mice, we have

now been able to link one of the most important multispecific drug transporters, OAT1, to a number of metabolic pathways; we are thus able to propose an OAT1-centered metabolic network, much of which is validated by *in vitro* and/or *in vivo* data. Furthermore, we have been able to separate direct interactions with the transporter from potential indirect cascade effects. We have also shown that metabolites in this network substantially overlap with those known to accumulate in CKD patients (including many putative uremic toxins) as well as in animal models of this complex metabolic disorder.

To summarize our novel strategy, we combined multiple systems biology and computational chemistry methods with wet lab data from knockout mice and *in vitro* assays (Figure 2.1). We first used transcriptomic changes from the *Oat1* knockout kidney to develop a quantitative reconstruction with the well-validated systems biology tool iMM1415 (a mouse genome-scale metabolic reconstruction (GEM)) to predict metabolic changes due to the loss of *Oat1* (Figure 2.9). The predictions were in reasonable agreement with the *in vivo* metabolomics results. Chemical constraints were applied at the molecular and atomic levels to identify those metabolites predicted by GEM that were likely to be directly transported by OAT1; these constraints were built into pharmacophore models based on a set of 61 drugs known to directly interact with OAT1 at an affinity of < 100 μM (Figure 2.10). This approach narrowed the list to 74 metabolites likely to interact with OAT1, and a set of 8 of these were tested using *in vitro* assays; half (4) were identified as directly interacting with OAT1, including prostaglandin E1, dihydrofolic acid, palmitoleic acid, and 16-hydroxy-hexadecanoic acid (Figure 2.4B). These metabolites are important in whole-body physiology and cellular metabolism. For example, prostaglandin E1, an endogenous vasodilator, serves to increase peripheral blood flow [59], while dihydrofolic acid is required to synthesize both purines and pyrimidines. Palmitoleic acid, a long-chained fatty acid serving as a potential lipokine, is

important in the regulation of lipid metabolism [60]. Based on these results, OAT1 is implicated in the modulation of a number of metabolic pathways, including the urea cycle, TCA cycle, purine and pyrimidine metabolism, tyrosine metabolism, fatty acid, and prostaglandin metabolism.

We argue the types of multifaceted analyses described here for OAT1, enabling the construction of a “drug transporter”-centered metabolic network, can be applied to other SLC and ABC drug transporters to generate a more comprehensive picture of the role these transporters play in metabolism. Eventually, this approach could potentially connect cellular metabolism in different organs via molecules transported by multispecific drug transporters (as well as other transport systems such as those involving other types of transporters or channels).

Given the broad substrate specificity of many SLC and ABC transporters, as well as their abundance in tissues like the kidney and liver, it is likely that they play an important role in modulating metabolite levels in complex metabolic diseases such as chronic liver and kidney disease, metabolic syndrome and diabetes [2, 38, 39, 61]. For example, some metabolites known to be transported by OAT1 are potential classical uremic toxins (e.g., indoxyl sulfate, kynurenate, spermine, and uric acid) which accumulate in CKD [47, 48]. Our analyses, for instance, indicate the importance of the OAT1-centered network in CKD; half of the metabolites in the OAT1-centered network overlapped with the CKD network. This provides a rationale for potential therapies aimed at ameliorating the uremic syndrome of CKD by enhancing *Oat1* expression and/or function.

In addition, the complexity of the OAT1-centered network points to the possibility of unexpected metabolic changes that could be induced by chronic treatment with drugs that compete with key metabolites for elimination; these changes could go well beyond

the relatively straightforward concept of transporter-level competition for the ligand binding site. We were hence able to separate direct versus indirect interactions of metabolites with OAT1; thus, a drug that tightly binds OAT1 may not only alter metabolites that directly compete for transport but also others in the OAT1-centered network that are not directly transported by OAT1. Since thiazide diuretics and HIV antivirals are transported by OAT1 [22, 36, 37], the OAT1-centered network may help in understanding the drug-induced metabolic syndromes associated with chronic treatment with these drugs [62-65]. Importantly, once drug-transporter metabolic networks are created for other SLC and ABC multispecific transporters, the systems biology approach employed here may be useful for explicitly predicting the metabolic alterations expected for new drugs in healthy or diseased populations with globally-altered metabolism (e.g. CKD, liver disease, metabolic syndrome, diabetes).

2.5 EXPERIMENTAL PROCEDURES

The overall approach taken in the study is depicted in the schematic flowcharts (Figure 1), which consist of 5 stages: 1) systems biology analysis (Figure 2.9), 2) computational chemistry analysis (Figure 2.10), 3) wet-lab validation, 4) Construction of an OAT1-centered metabolic network (Figure 2.11), and 5) evaluation of the OAT1-centered metabolic network in the context of CKD (Figure 2.12).

2.5.1 Materials

Water-soluble probenecid was purchased from Molecular Probes. The fluorescent tracers, 6-carboxyfluorescein (6CF), and tested metabolites (prostaglandin E1, pristanic acid, elaidic acid, trans-vaccenic acid, dihydrofolic acid, palmitoleic acid, beta-nicotinamide mononucleotide, and 16-hydroxy-hexadecanoic acid) were purchased from Sigma.

2.5.2 Analysis of transcriptomic data with mouse genome-scale metabolic network (GEM)

The transcriptomic and metabolomics data from previous studies [8-10, 35, 37, 66] were used for the metabolic reconstructions. The mouse GEM, iMM1415 [40], which contains the biochemical transformations for numerous tissues and cells in mice, was utilized. When establishing the lower and upper bound for exchange reactions, uptake constraints for WT and KO models were set to be the same (10μmol/h) for most metabolites, and to make our model more renal-specific, a few metabolites that were listed by the previously published renal object function, which was used to analyze blood pressure regulation, [67] were set to be either secreted or absorbed accordingly. To create context-specific models, gene expression data was analyzed by Microarray Suite Version 5.0 to assign present/absent calls and this data was incorporated into the model using the gene inactivity moderated by metabolism and expression (GIMME) algorithm [41]. A reduced model was created by removing reactions classified as absent based on the analysis of the gene expression data. Then, reactions needed to achieve a given objective function were reinserted to provide a functional, context-specific model. In our analysis, biomass maintenance function, which was defined to represent the metabolic composition of mouse tissues, was used as the objective function to generate the context-specific models. The biomass-pseudo-reaction was used to generate the GIMME models because it ensured certain functionality of the models. The biomass was not used as an objective in the analysis however, since results were derived from FVA with a sub-optimal minimal lower bound that was placed on the biomass function.

2.5.3 Comparing the metabolic differences between KO and WT with Flux Variability Analysis (FVA) and validation of the model

FVA was used to compare functional differences between the WT and KO mice based on maximal achievable reaction flux ranges [44]. FVA was performed for the WT and KO reconstructions. Flux spans were calculated as the differences between the maximum and minimum of reaction fluxes and the pairwise ratio between the KO and WT flux spans were calculated, resulting in a vector of flux span ratios for the set of reactions shared between the two models. This analysis was used to determine which metabolites have altered transport activities due to the loss of *Oat1*. Lists of significantly altered metabolites from the analysis of targeted and untargeted metabolomics data from the *Oat1* knockout mouse [8, 9] were used to determine how well the GEM prediction models performed. A successful model was expected to generate a list of "GEM-predicted OAT1 metabolites" that would be further evaluated by wet lab experimental studies (below).

2.5.4 Integrating the in vivo expression data in Recon2 and comparing the WT and KO with FVA

In addition to the iMM1415 metabolic reconstruction, Recon2, a later version of Recon1 that contained more reactions and metabolites [68], was also used for model generation. This enabled inclusion of additional metabolites and reactions in the reconstruction. Recon2, however, represents human metabolism. In order to use it for modeling mouse, the human Entrez gene IDs in Recon2 were converted to the mouse homologs based on the online database BioDBnet (http://biodbnet.abcc.ncifcrf.gov/). Then, the WT and KO model was generated and a comparison of the models was done following the approaches described above. The FVA differences between the WT and

KO were smaller when using Recon2 versus iMM1415 as we observed only 521 reactions to be altered (The FVA results are shown in Table S1).

2.5.5 Pharmacophore model building and validation based on drugs known to interact with OAT1

Computational chemistry analysis was performed with ICM software developed by Molsoft L.L.C (San Diego, CA). The software was used to perform clustering, alignment, and pharmacophore building based on the atomic property field (APF) [45] of OAT1-interacting drugs or tracers (pharmaceuticals) with known K_m (substrate affinity) or K_i (inhibitory affinity) of less than 100 μM of were selected. A total of 61 pharmaceuticals were selected; among them, two third of them (41 drugs) is split to training set for the model generation, and one third (20 drugs) of actives is split to validating set. After the model generation, the validating set (serving as true positives) and Drugbank database (serving as true negatives) are screened against the pharmacophores. And with the screening results, a ROC curve is generated.

2.5.6 Screening the list of metabolites predicted by metabolic reconstructions with pharmacophore models

The GEM-predicted OAT1 metabolites were compared to each of the 7 pharmacophore models and ranked by how well they fit with the 3D molecular space defined by the drugs known to interact with OAT1. The 30 metabolites that best fit each of the 7 pharmacophores were selected as having potential to directly interact with OAT1. Many metabolites fit more than one pharmacophore model. After eliminating overlaps, 74 metabolites remained, and these were termed “Top-ranked GEM-predicted OAT1 metabolites” (passing OAT1 pharmacophore filter).

2.5.7 Uptake Inhibition Assay

Metabolites predicted to have direct interactions with OAT1 (ie. which had been predicted to be altered in the GEM analysis and which passed the pharmacophore filters) were tested for further wet-lab validation based on availability and solubility; they were prostaglandin E1, pristanic acid, elaidic acid, trans-vaccenic acid, dihydrofolic acid, palmitoleic acid, beta-nicotinamide mononucleotide, and 16-hydroxy-hexadecanoic acid. Briefly, *Oat1*-transfected CHO cells cultured on 96-well plates were incubated in the presence of 10 μM 6-carboxy-fluorecein (6CF) [22, 25, 36] with or without individual metabolites (with controls treated with the OAT1 inhibitor probenecid). The IC_{50} curves for novel ligands were plotted in Prism Software (GraphPad Inc., San Diego, CA), and the IC_{50} values were converted to K_i (inhibition affinity) using the Cheng-Prusoff equation:

$$Ki = \frac{IC_{50}}{1+\frac{[S]}{Km}} \qquad \text{(Equation 1)}$$

2.5.8 Construction of an OAT1-centered Metabolic Network

KEGG IDs for OAT1-connected metabolites (metabolites known or predicted to interact with OAT1) were input into Metscape, a Cytoscape plugin [46]. Metabolites present in Metscape were used as “input metabolites” to build a metabolic network. The construction of the network also introduced many “plus-one” or “first-neighbor” metabolites and the network was then trimmed to eliminate uninformative nodes. The following criteria were used to eliminate nodes: small molecules (such as water, carbon dioxide, etc), energy-related molecules (NADH, ATP, etc), and large-sized peptides known not to interact with OAT1 or related transporters (somatostatin, kinetensin, etc). In addition, unnecessary “dead-ended” and “inter-connecting” plus-ones were removed to create a more concise network (In other words, “dead-ended” plus-one nodes, which connected only to one node, were removed, as were “inter-connecting” plus-ones that

did not affect connections between wet-lab validated or predicted nodes). The final network thus consisted of metabolites that fell into four categories, which in the order of level of confidence, the categories were: 1), “wet-lab supported”, 2), “predicted to interact with high confidence”, 3), “predicted to be affected indirectly”, and 4), “plus-one after trimming”.

2.5.9 Network analysis

To measure the global and node network parameters of the OAT1-centered metabolic network, another Cytoscape plugin, NetworkAnalyzer [69], was used to study the topology of the metabolic network. The following global parameters were measured for the network: clustering coefficient, connected components, network diameter, network radius, network centralization, shortest paths, characteristic path length, average number of neighbors, and network density [49, 70]. In addition to examining the topology of the network as a whole, individual nodes were also analyzed by measuring their cluster coefficient, degree, neighborhood connectivity, stress, and topological coefficient [49, 70].

2.5.10 Effect of Chronic Kidney Disease (CKD) on the OAT1-centered Network

A CKD-associated metabolic network was constructed in Metscape based on 114 metabolites found to be associated with CKD in humans and/or animal models (reviewed in [57]). As above, a metabolic network was then created which included the metabolites and their immediate neighbors, and the network was trimmed as described above to eliminate water, CO_2, ATP, NADH, peptides, etc. After the disease-associated metabolic network was created, it was compared with the OAT1-centered metabolic network.

2.5.11 Statistics

To determine if the overall in-silico approach is significant in enrichment of metabolites known to have direct interaction with OAT1, a hypergeometric test is performed to calculate the p-value for the enrichment process. In addition, the hypergeometric test is also used to determine the significance of overlap between two metabolic networks, i.e. the OAT1 network and the CKD network.

2.6 Acknowledgements

This work was supported by National Institutes of Health Grants R01-GM098449 and U54-HD07160 to S. K. Nigam. The work presented in Chapter 2, in full, is being prepared for submission. Liu, Henry C.; Jamshidi, Neema; Chen, Yuchen; Eraly, Satish A.; Cho, Sai Yee; Bhatnagar, Vibha; Bush, Kevin T.; Abagyan, Ruben; Palsson, Bernard O.; Nigam, Sanjay K. The dissertation author was the primary investigator and author of this paper.

2.7 Figures

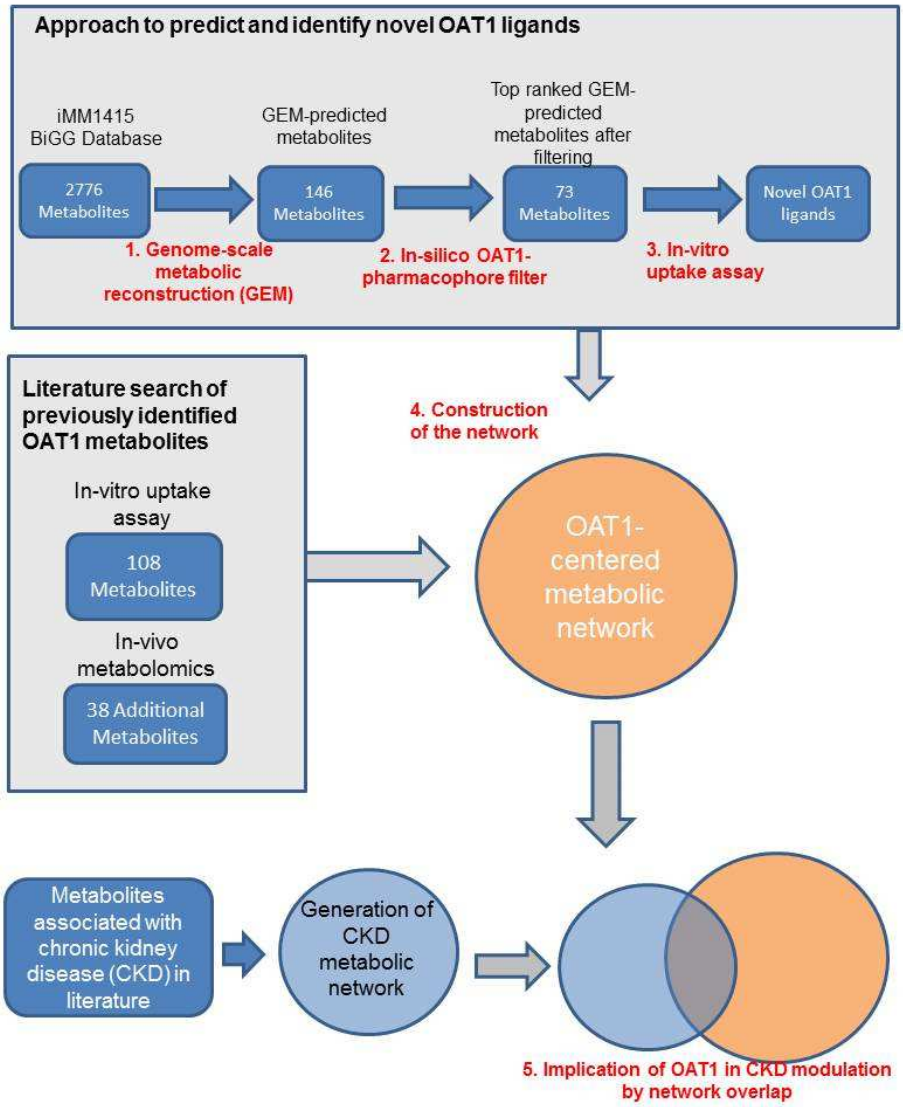


Figure 2.1: Flowchart of methods. The methods consist of 5 stages. 1) systems biology analysis and mathematical simulation of metabolic reconstruction (Figure S1), 2) computational chemistry analysis and generation of pharmacophore hypotheses and the application of pharmacophore filtering (Figure S2), 3) wet-lab validation and identification of new OAT1 metabolites, 4) construction of a substantially validated OAT1-centered metabolic network (Figure S3), and 5) evaluation of the role of the OAT1-centered metabolic network in disordered metabolism of chronic kidney disease (uremia) (Figure S6).

A

Prediction	*# Reactions & metabolites predicted to change*
Total reactions	1026
Exchange/transport reactions	321
Metabolites predicted to be altered in transport	177
GEM-predicted metabolites (excluding water, ATP, etc.)	14

B

Metabolites identified by previous targeted and untargeted metabolomics and are also present in iMM1415 BiGG Database	*Altered in urine/plasma?*	*Predicted to change*
3-Hydroxybutyrate	yes	yes
4-Hydroxyphenyllactate	yes	yes
3-Hydropxypropionate	yes	yes
4-Hydroxyphenylacetate	yes	yes
Orotate	yes	yes
4-Hydroxyphenylpyruvate	yes	yes
alpha-Ketoglutarate	yes	yes
Urate	yes	yes
Uracil	yes	yes
N-Acetylaspartate	yes	yes
Benzoate	yes	--
3-Hydroxyisobutyrate	yes	--
2-Oxo-3-methylvalerate	yes	--
2-Oxoisocaproate	yes	--
Pantothenic acid	yes	--
4-Pyridoxic acid	yes	--
Kynurenine	yes	--
Methionine	yes	--
Thymidine	yes	--

Figure 2.2: The results of GEM analysis based on transcriptomics of the knockout animals are consistent with in vivo metabolomics. A.) The number of metabolic reactions and metabolites predicted to be altered due to *Oat1* knockout based on GEM analysis. B.) Comparison of metabolomics data from *Oat1* knockout mouse (plasma, urine) with computational predictions of GEM transcriptomic-based metabolic network. See Methods for details. Column 1: List of metabolites identified by targeted and untargeted metabolomics analysis and represented in iMM1415 (a GEM model). Column 2: Metabolites that are altered either in plasma or urine or both by metabolomics. Column 3: Predictions of changes in fluxspans made by GEM based on the transcriptomic data from wildtype and *Oat1* knockout mice (-- is either no change predicted or no prediction possible).

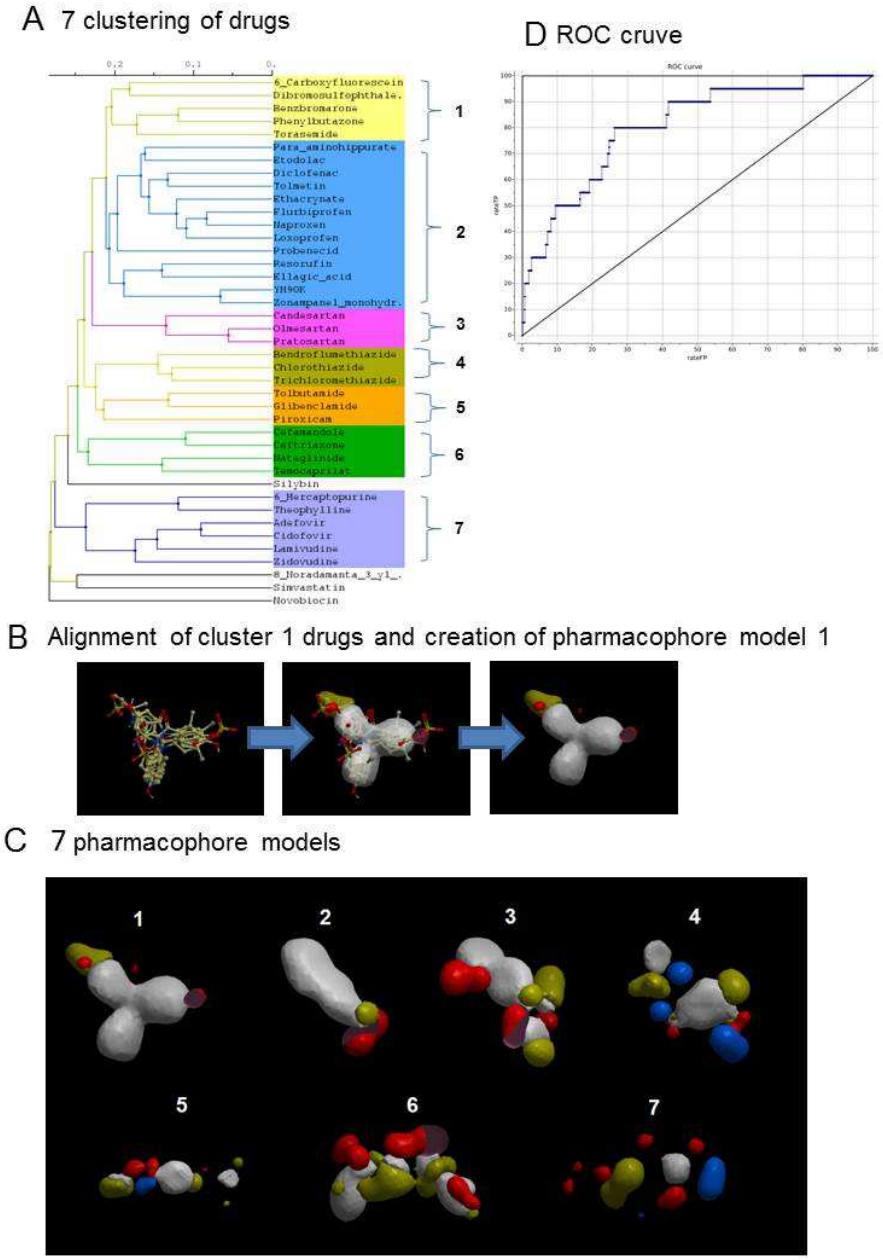
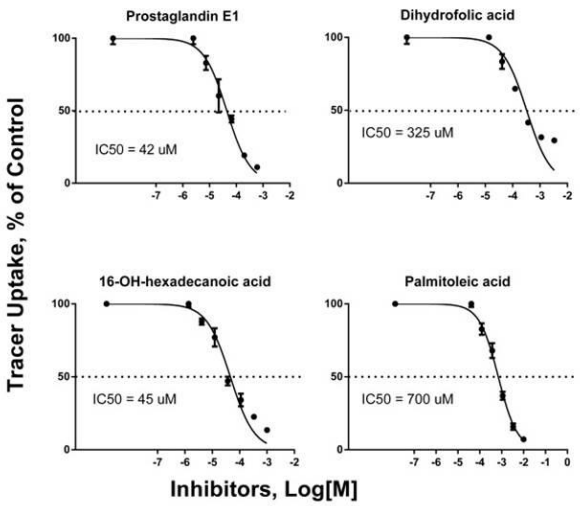


Figure 2.3: The clustering of drugs and the alignment and pharmacophore generation for individual clusters. A.) Clustering of OAT1 drugs based on atomic property field (APF). B.) The alignment of drugs from cluster 1 and the creation of pharmacophore model 1. C.) The seven pharmacophore models generated; blue – hydrogen bond donor; red – hydrogen bond acceptor; white – aromaticity; yellow – hydrophobacity; light red – negative charges; Light blue –positive charges. D.) The ROC curve generated for the validation of the system of the pharmacophore models, and area under the curve (AUC) was calculated to be 80.58, demonstrating such system was well validated.

A

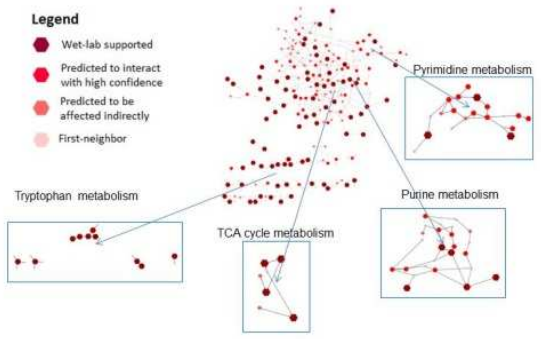


B

Novel OAT1 ligands	*Measured K_i (μM)*	*Metabolic pathways*
Prostaglandin E1	12	Arachidonic acid metabolism
Dihydrofolic acid	93	Folate biosynthesis
Palmitoleic acid	200	Fatty acid metabolism
16-Hydroxy-hexadecanoic acid	13	Fatty acid metabolism

Figure 2.4: The wet lab validation of predicted metabolites and the identification of novel OAT1 metabolites. A.) The IC_{50} curves for the metabolites characterized in the *in vitro* uptake inhibition assay using CHO stably expressing *Oat1*. The assay was done by testing for inhibition of uptake of 10 μM 6CF, a fluorescent tracer that is also an OAT1 substrate. B.) The table summarizes the calculated Ki and metabolic pathways the characterized metabolites were involved in.

A



B

Metabolites at different confidence levels	Number of metabolites
Wet-lab supported	74
Predicted to interact with high confidence	41
Predicted to be affected indirectly	61
First-neighbor	77

C

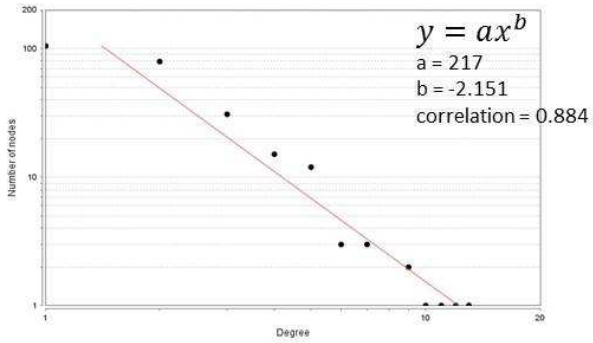
Network parameters	Oat1-centered metabolic network
Number of nodes	253
Clustering coefficient	0.105
Connected components	32
Network diameter	15
Network radius	1
Network centralization	0.043
Shortest paths	28900 (44%)
Characteristic path length	6.104
Avg. number of neighbors	2.285
Network density	0.009

D

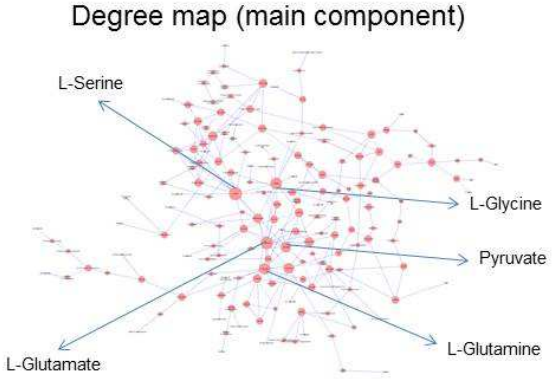
The top well-represented metabolic pathways (ranked by validation percentage)	Total metabolites	Wet-lab supported	Validation percentage
Tyrosine metabolism	12	7	58.3%
TCA cycle	7	4	57.1%
Tryptophan metabolism	19	10	52.6%
Valine, leucine and isoleucine degradation	10	5	50.0%
Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine	29	13	44.8%
Vitamin B9 (folate) metabolism	8	3	37.5%
Glycine, serine, alanine and threonine metabolism	15	5	33.3%
Bile acid biosynthesis	9	3	33.3%
Vitamin B3 (nicotinate and nicotinamide) metabolism	10	3	30.0%
Purine metabolism	25	5	20.0%
Pyrimidine metabolism	20	3	15.0%

Figure 2.5: The OAT1-centered metabolic network contains several biochemical pathways essential for cellular metabolism. A.) The network consisted of metabolites known to or predicted to, interact with OAT1; these metabolites were classified according to the level of confidence with which they interacted with OAT1. In the order of level of confidence, the categories were: 1), “wet-lab supported”, 2), “predicted to interact with high confidence”, 3), “predicted to be affected indirectly”, and 4), “plus-one. The “wet-lab support” metabolites were supported by either *in vitro* or *in vivo* data. The metabolites “predicted to interact with high confidence” were ones predicted by GEM and passed the pharmacophore filter criteria. The metabolites “predicted to be affected indirectly” were ones that were predicted by GEM only but did not pass the pharmacophore filter criteria. The “plus-one” metabolites were predicted by Metscape as the first neighbors of the aforementioned three groups and did not fit any of the other categories described above. The level of confidence of metabolites interacting with the transporter is reflected by the size and color of nodes; the larger and darker the metabolites are, the higher the confidence of interacting with OAT1. In the network, the well-represented metabolic pathways are shown. B.) The number of metabolites falls into the four confidence categories. C.) The global network parameters of the OAT1-centered metabolic network. D.) The total number of metabolites, the number of wet-lab supported metabolites, and the validation percentage for well-represented pathways are listed in Figure 5D. Most represented metabolic pathways are ranked according to the validation percentage. Tyrosine, TCA cycle, and tryptophan metabolism are the most well-validated pathways represented in the network, and they contain several metabolites known as classical OAT1 substrates.

A



B



C

Highlighted metabolite	*Degree*
Pyruvate	9
L-Glutamine	10
L-Serine	13
L-Glutamate	12
Glycine	11
Average of above highlighted metabolites	11
Average of all metabolites	**2.8**

D

The top well-represented metabolic pathways	*Average Degree*
Tyrosine metabolism	2.1
TCA cycle	3.7
Tryptophan metabolism	1.5
Valine, leucine and isoleucine degradation	1.4
Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine	4.3
Vitamin B9 (folate) metabolism	6.5
Glycine, serine, alanine and threonine metabolism	4.2
Bile acid biosynthesis	2.7
Vitamin B3 (nicotinate and nicotinamide) metabolism	3.0
Purine metabolism	3.1
Pyrimidine metabolism	2.6

Figure 2.6: The degree distribution and the node degree map of OAT1-centered metabolic network. A.) The degree distribution for the network was fitted with the power law equation, $y = ax^b$, using the least-squared method. The measured correlation coefficient *a* is equal to 217, and the measured exponent *b* is equal to -2.151, demonstrating that the network had degree distribution that followed a power law and was thus found to be scale-free B.) The map shows the value of degree for individual nodes of the main component of the network (and the higher the value of the parameter, the bigger the node). C.) The values of degree for the highlighted metabolites are listed along with the average values of the highlighted metabolites and of all metabolites in the network.

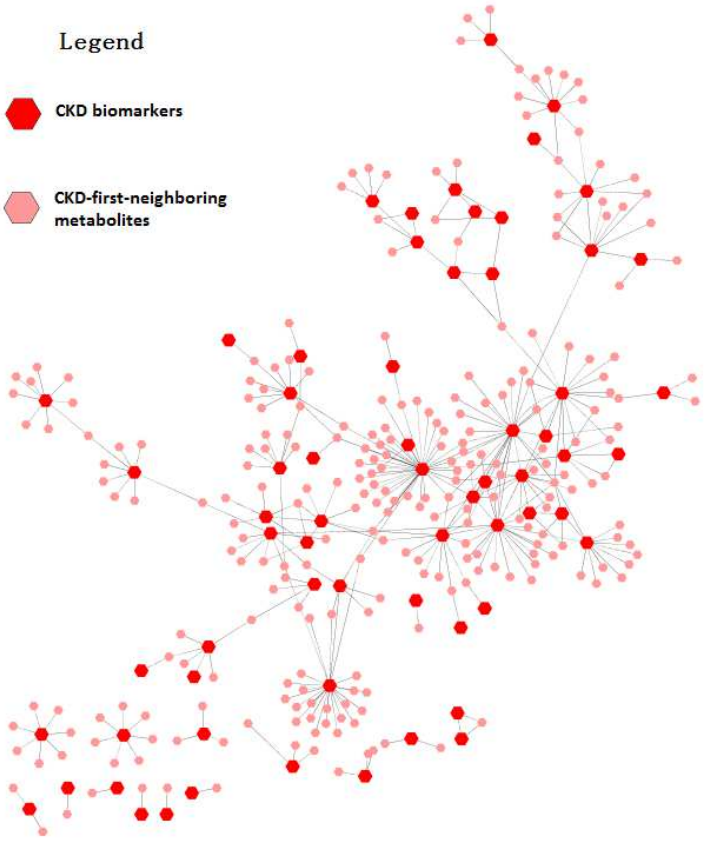
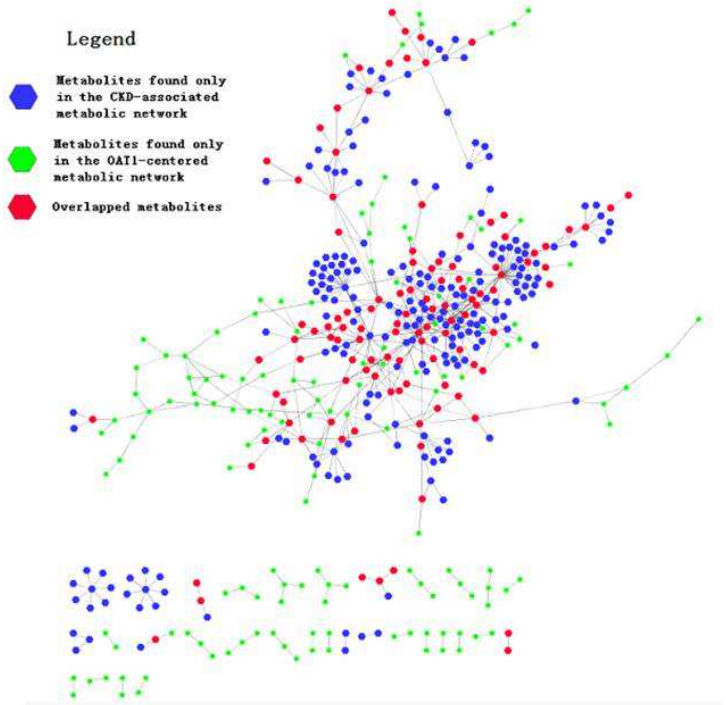


Figure 2.7: The CKD-associated metabolic network. The network is based on a list of CKD biomarkers found to be altered in metabolomics analyses of CKD patients and of animal models of kidney disorder (Zhao 2013). Red nodes are CKD biomarkers, and pink nodes are their first neighbors.

A



B

Metabolites in the merged network	*Number of metabolites*	*Examples*
Overlapped metabolites	113	L-Kynurenine, Hypoxanthine, alpha-Ketoglutarate, Xanthine.
OAT1 only metabolites	140	4-Pyridoxate, Homovanillate.
CKD only metabolites	209	Ethanolamine, Linoleate.

Figure 2.8: The overlap of the OAT1-centered metabolic network and the chronic kidney disease (CKD)-associated metabolic network. The CKD-associated metabolic network is built based on the metabolites known to be biomarkers for chronic kidney disease [57] (See Figure S5). A. The overlapped, combined model is shown. Overlapping metabolites (113) are marked red, and they represented 44.9% of metabolites in OAT1-centered metabolic network; metabolites found only in the CKD metabolic network are marked blue; and metabolites found only in the OAT1-centered metabolic network are marked green. Among the overlapping ones were metabolites that are both listed as potential uremic toxins and known to directly interact with OAT1 [9, 58]; these included metabolites such as xanthurenate, xanthine, L-citrulline, and hippurate. B. The number of metabolites for the three metabolite groups is shown along with some examples.

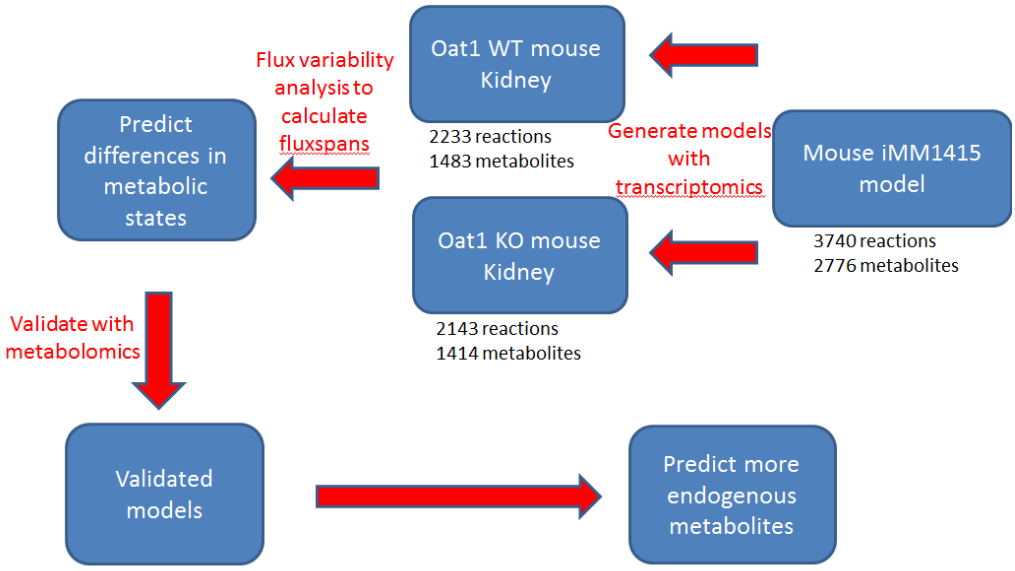


Figure 2.9: The schematic flowchart for the GEM analysis. The transcriptomics data of the animals were incorporated into the universal mouse GEM model, iMM1415, to generate two mouse kidney models, Oat1 WT and Oat1 KO. Then, flux variability analysis was used to determine the functional differences between the two models; these predictive differences were validated by in-vivo metabolomics data (which identified endogenous metabolites that were affected by the knockout of Oat1) of the animals and were used to predict more metabolites that may interact with OAT1.

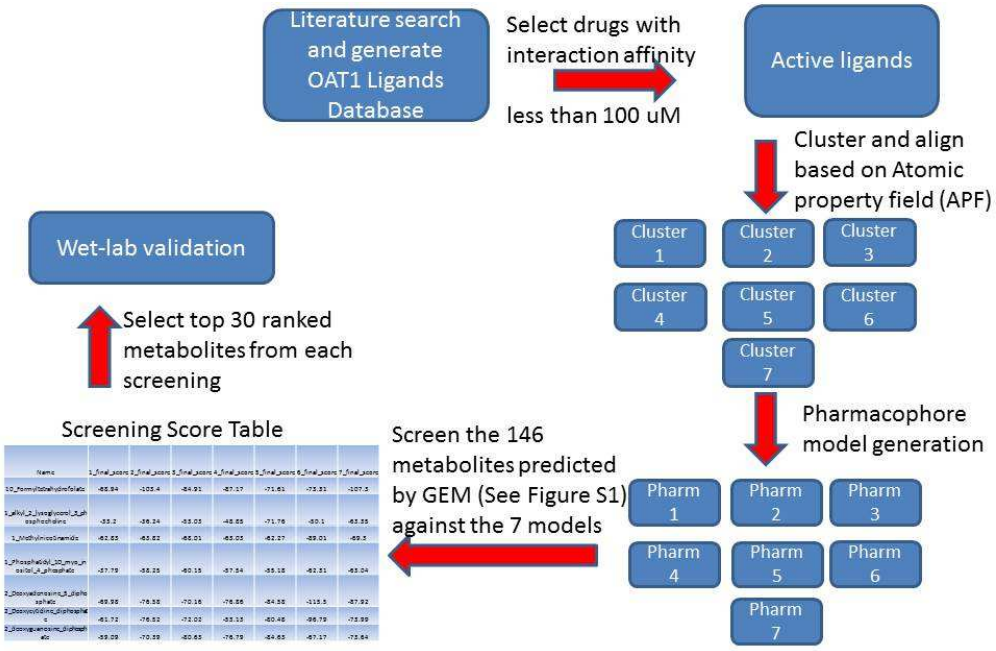


Figure 2.10: The schematic flowchart for the ligand-based pharmacophore building and screening approach. Briefly, a literature search was done to identify high-affinity OAT1 pharmaceuticals, and those having Km or Ki less than 100 μM with OAT1 interaction were selected as active ligands for pharmacophore generation. These active ligands were clustered into 7 groups based on atomic property field (APF), and a single pharmacophore model was generated based on each of the clusters, giving us a total of 7 pharmacophore models. The metabolites predicted to be affected by the knockout of Oat1 in the genome-scaled metabolic reconstruction (GEM) analysis were screened against the 7 models, and the top 30 metabolites from each screening were compiled as "Top-ranked GEM predicted" (passing the OAT1 pharmacophore filtering) metabolites and some of them were selected for further wet-lab validation.

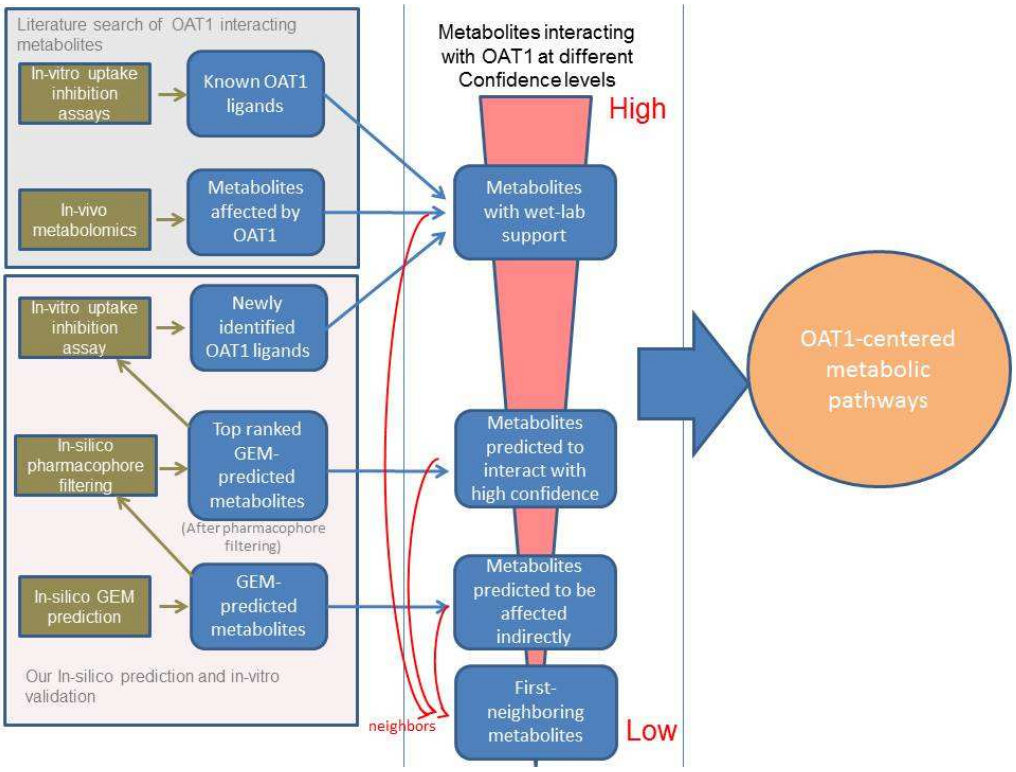


Figure 2.11: The schematic flowchart for generation of OAT1-centered metabolic network using both wet-lab data and in-silico predictions. The OAT1-centered metabolic network was generated based on metabolites known to, or predicted to, interact with OAT1 originating from 5 sources: previous in-vitro uptake assays, previous in-vivo metabolomics, the in-vitro uptake assays done here, the GEM prediction followed by the pharmacophore filtering, and the GEM prediction alone. The metabolites are categorized into 3 groups according to the sources they originated from. The metabolites from the first 3 sources were merged together and were labeled as “Wet-lab supported” metabolites as their interaction with OAT1 was supported by wet-lab data. The metabolites predicted by the GEM analysis and then top-ranked by the pharmacophore filtering, also known as “Top-ranked GEM-predicted metabolites”, fell into the category of “Metabolites predicted to interact with high confidence”. Finally, metabolites that were only predicted by the GEM analysis were categorized into the group of “Metabolites predicted to be affected indirectly”. Besides these three groups, an additional group of metabolite was introduced in the network generation; this group of metabolites included the first neighbors of the aforementioned metabolites, and since there is currently no evidence supporting their interactions with OAT1, they were deemed to have the lowest confidence to interact with OAT1.

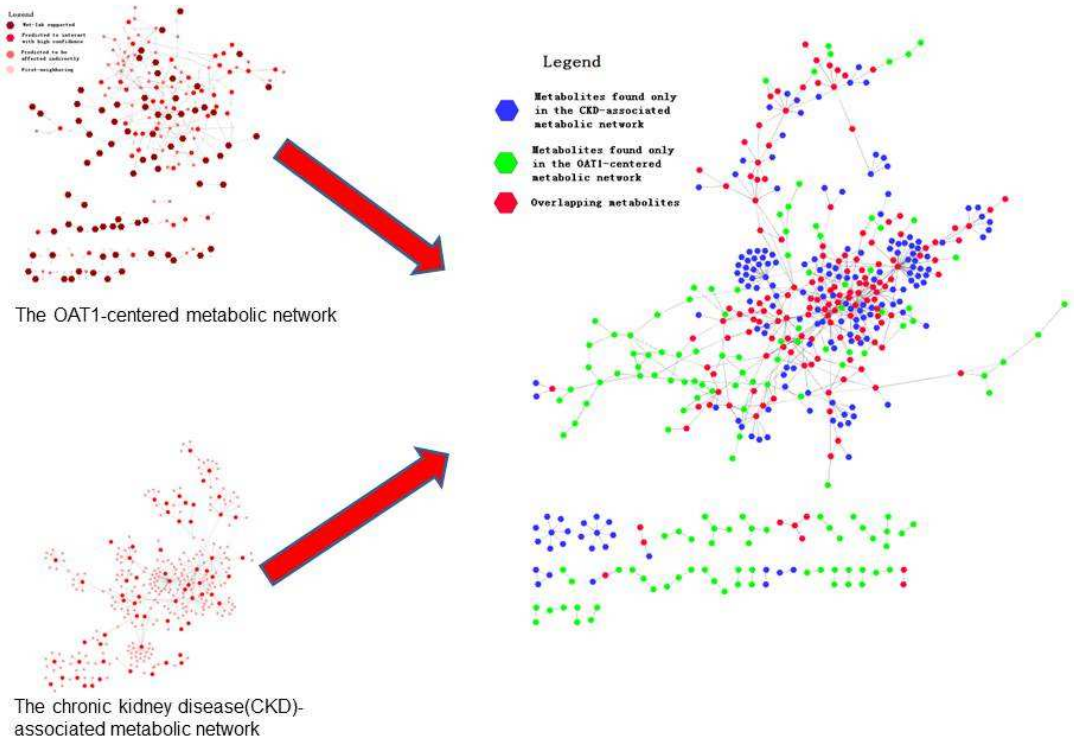


Figure 2.12: The overlap of the OAT1-centered metabolic network and the CKD-associated metabolic network. The two networks are merged together to generate a greater network, and the overlap of the two is depicted to potentially implicate OAT1 activity in modulating the metabolic changes of CKD.

Chapter 3 : The Ligand-based Computational Chemistry Analysis of OAT1, OAT3, OCT1, and OCT2

3.1 Abstract

Organic anion transporters (OATs) and organic cation transporters (OCTs) are membrane transporters of the solute carriers 22 (SLC22) family that are widely expressed in the epithelium of different tissues. They are important in the clearance of drugs, toxins, and metabolites. Despite having high overall sequence similarity, they interact with distinct groups of substrates. While, as suggested by their names, it is generally accepted that OATs interact with organic anions, and OCTs mainly interact with organic cations, there is growing evidence that the basis for substrate selection may be more complex [21]. To better analyze differences in substrate preference between OAT1, OAT3, OCT1, and OCT2, we performed machine learning analysis and statistical analysis on physiochemical descriptors of ~250 drugs known to interact with one or more of these four SLC22 transporters and applied machine learning methods to understand the basis of drug interaction with a particular transporter. While the overall results were consistent with accepted scientific understanding (i.e. the predominant differences between the functions of OATs and OCTs were due to the charge of the substrates), apart from charge, another property distinguishing OATs and OCTs was the degree of complexity of substrates; OATs interact with planar structures whereas OCTs interact

with substrates having more three-dimensional structures (SP3 character). Although decision tree analysis and other machine learning tools did not uncover clear molecular properties that distinguished OCT1 and OCT2 substrates, there were differences between OAT3 and OAT1 substrates. In particular, OAT3 interacted with drugs with more acyclic tetravalent bonds and more zwitterionic/cationic drugs than OAT1. To further study the cationic substrate preference of OAT3, pharmacophore models were built, and it was found that OAT3 contained one unique pharmacophore, which possessed some cationic properties similar to pharmacophore models of OCT substrates. Further wet-lab uptake assays based on the virtual screening of this pharmacophore model not only validated the model system but also identified novel cationic drugs that selectively interact with OAT3 but not OAT1. These analyses indicated that the boundary between OATs and OCTs is more complex than charge alone. These analyses more clearly distinguish the basis of drug binding to individual OATs and OCTs. By applying this type of approach to studying drug transporters binding it may be possible to predict the molecular properties of a drug necessary for interacting with one or more transporters. This may be useful for targeting specific tissues or elimination routes.

3.2 Introduction

Organic anion transporter 1 (OAT1 or SLC22A6), organic anion transporter 3 (OAT3 or SLC22A8), organic cation transporter 1 (OCT1 or SLC22A1), and organic cation transporter 2 (OCT2 or SLC22A2) are four important multi-specific transporters responsible for the excretion of a wide variety of drugs, toxins, and metabolites in the kidneys and other tissues [2, 3, 71]. They are perhaps the best studied drug transporters of the SLC22 family of solute carriers. Although sharing overall sequence similarity and tertiary structures [19], the four transporters have distinct substrate preferences. As their names suggest, OATs, belonging to the "organic anion" transporters family, mainly interact with anions whereas OCTs, belonging to the "organic cation" transporter family, mainly interact with cations.

Nevertheless, the grouping of OATs and OCTs into two different transporter sub-families, organic anions and organic cations, respectively, can be misleading when it comes to individual drugs, toxins and metabolites. For example, recent studies demonstrate that OATs have the capacity to interact with cationic drugs [21], and both OATs and OCTs appear to interact in vitro and in vivo with zwitterionic or mildly "cationic" metabolites such as creatinine [72, 73]. However, these studies are limited to a few substrates. This raises questions about the general conception of OATs as organic anion transporters and OCTs as organic cation transporters. Given that there are a large number of well-established substrates published in scientific literature, a systematic computational and statistical analysis now seems possible. Here, we attempted to address this issue by performing machine learning analyses based on the physiochemical descriptors of the drugs known to interact with one or more of these four transporters.

Since the crystal structures of the four transporters are unknown at this time, only ligand-based computational chemistry approaches were utilized in our studies. Among the ligand-based approaches, one commonly used method is the development of quantitative structure-activity relationship models (SAR and QSAR), which is an attempt to identify the correlation between the activity, or binding affinity between ligands and transporters, and the values of physiochemical descriptors of ligands. Previously developed QSAR models for OAT1, OAT3, and OAT6, which were built based on inputs of approximately 10 descriptors, identified several physiochemical properties of ligands important for binding to transporters [22]. In addition to QSAR models, another approach that has gained popularity is the application of machine learning tools. Among these tools, the support vector machine method (SVM) has been used to develop models for two ABC transporters, breast cancer resistance protein (BCRP) and P-glycoprotein (P-gp) [27, 28]; these models were mainly used for the in-silico prediction of new substrates. Besides SVM, other powerful machine learning tools, such decision trees and random forests, have been used widely for different applications.

To understand the functional differences between OAT1, OAT3, OCT1, and OCT2, several machine learning methods were applied to understand what physiochemical properties within the set of ~250 drugs affect the interaction between individual transporters and drugs as well as the relative importance of these properties in contributing to selectivity for interaction with a particular transporter. The results based on machine learning analysis were found to be consistent with the results based on conventional statistical analysis and provide quantitative support for subtleties in the molecular basis of substrate interactions among and between individual OATs and OCTs.

The results were further supported by pharmacophore models of OAT and OCT substrates. Pharmacophore modeling studies aim to find the common features shared among the ligands in three-dimensional space. Previous studies using this method have built several pharmacophore models of renal transporters: OAT1 models based on 16 drugs that highly inhibited the function of OAT1 [24] and the targeted and untargeted metabolites found to be altered in OAT1 Knockout plasma [9, 23]; and OAT3 models based on four cationic drugs [21]. All the previous models were built for the purpose of predicting novel ligands based on a limited set of known substrates. In order to gain a comprehensive understanding of binding interactions and substrate selections, it was necessary to construct pharmacophore models based on as many pharmaceutical drugs as possible. This would provide a more comprehensive chemical space to build complete pharmacophore models for the multi-specific transporters because each individual compound contributes some information to the broader representation of the whole chemical space of binding.

Our results demonstrate that, consistent with previous knowledge, the main functional difference between OATs and OCTs are charge and electro-charge related properties. OATs interact with negative and more anionic substrates; on the other hand, OCTs interact with positive and more cationic substrates. In addition to charge-related factors, a structure property, SP3, is newly identified to be an important determinant in distinguishing between OAT and OCT substrates; OATs interact with planar structures, whereas OCTs interact with more 3-dimenional structures, indicating that in addition to charge, the topology of ligands is another important factor. In addition, subtle but important differences exist between OATs, and that apart from having anionic substrate preferences, OATs have a propensity to bind some cations that structurally overlap with OCT substrates.

3.3 Materials and Methods

The overall computational workflow is shown in Figure 3.1.

3.3.1 Materials

Water-soluble probenecid was purchased from Molecular Probes. The fluorescent tracers, 5-carboxyfluorescein (5CF) and 6-carboxyfluorescein (6CF), and cationic drugs (loperamide hydrochloride, nebivolol hydrochloride, darifenacin hydrobromide, paliperidone, cisapride monohydrate, and halofantrine hydrochloride) were purchased from Sigma.

3.3.2 Selecting drugs that interact with OAT1, OAT3, OCT1, and OCT2 with high affinity

Through a comprehensive literature and internet search, pharmaceutical drugs and tracers found to interact in-vitro with any of the four transporters were selected for this study. The information compiled from this literature search, as well as, the information about the individual drugs used in this study was organized into a database (Supplemental Table). In all, ~250 drugs were analyzed. Since Km data is only available for a subset of drugs, the binding affinity (Ki) and the substrate affinity (Km) were used as a measurement to reflect how well a given drug interacts with a given transporter; this is referred to as “interaction affinity”.

Reported values of Ki or Km in the literature vary for several reasons: different labs, different assays, different assay conditions, different species, etc. In most cases, we found that values varied only slightly across species (rat, mouse, human, etc.), and thus we treated them similarly. Outliers were discarded, and most commonly reported values were used. Because of these variations, we grouped the affinity values into categories described in the following paragraph.

3.3.3 Determining the charge states of drugs that interact with individual transporters

Interaction affinity was categorized as high affinity, mid affinity, low affinity, or extremely low affinity: “high affinity” is less than 100 uM, “mid affinity” is between 100 uM and 1000 uM, “low affinity” is between 1000 uM and 2000 uM, and “extremely low affinity” is between 2000 uM and 12000 uM. Then, charge states of drugs in the physiological pH, 7.4, were determined in the computational environment of ICM software. We labeled the drugs as “cationic” when their net charge was greater than zero; “anionic” when their net charge was less than zero; “neutral” when net charge was equal to zero and they contained no charged atoms; or “zwitterionic neutral” when net charge was zero but they contained an equal number of positively charged and negatively charged atoms. The results were then plotted.

3.3.4 Determine the charge species percentage for drugs that interact with individual transporters at high affinity

The nomenclature of charge states alone does not describe the overall percentage of charge species of a particular drug as a drug can have more than one charge species co-existing at a given pH point. To address this problem, the percentage of charge species was measured based on the pH/concentration curves provided by the website of chemicalize.org (Chemaxon), and the species percentages were calculated at three different pH values, 7.2, 7.4, and 7.6. Then, the drugs that had high-affinity with a given transporter were grouped together, and their charge species percentages were graphed to construct charge-species bar diagrams. In the charge-species bar diagram, individual rows represented the percentage of various charge species for each drug, and combined, they delineated the propensity of individual transporters to bind to various

charge species. The charge-species bar diagrams were done for each transporter at three pH levels, 7.2, 7.4, and 7.6. Finally, total positive species percentage, total negative species percentage, total neutral species percentage, and total zwitterionic neutral species percentage were calculated for individual charge-species bar diagrams.

3.3.5 Venn diagram of the drugs that interact with the four transporters at high affinity

Since pairwise comparison study in machine learning analysis was limited to non-overlapping drugs, we used Venn diagrams as a visualization tool to identify the overlapping drugs between each pair of transporters. Four Venn diagrams were created comparing the high affinity drugs for each transporter and drugs were classified as overlapping vs non overlapping for each pairwise comparison. Overlapping drugs were excluded from the machine learning data set. Instances in our machine learning models consisted of only non-overlapping drugs in all cases.

3.3.6 Collecting the Data

Attributes in our machine learning models were physiochemical properties calculated in ICM and tabulated in KNIME. ICM is a commercially available software supported by Molsoft (San Diego). Using ICM, we calculated about 50 physiochemical attributes of the drugs such as molecular quantum numbers, atom counts, bond counts, polarity counts, and topology counts. KNIME is an open source platform for machine learning [74]. It includes extensions capable of collecting data from three notable open source cheminformatics toolkits, RDKIT, Indigo, and CDK. Through the KNIME platform using RDKit, Indigo and CDK, we added attributes to represent about 100 chemical features such as molecular weight, molecular volume, Log P, Log S, polar surface area,

etc. In addition to these physiochemical attributes, a class variable was added to represent which transporter a given drug would interact with.

3.3.7 Preprocessing the Data

After collecting the data, Weka, KNIME, and Excel tools were used to preprocess the data. Weka is an open source collection of machine learning algorithms developed by the University of Waikato and is bundled with tools for preprocessing data to make it more easily understood by the machine learning algorithms. The raw data extracted from KNIME and ICM contained some attributes that were overlapping, empty, or constant and those were eliminated. The second step was to use Weka's attribute selection feature, Chi Square Evaluator, which has the capability of ranking the attributes according to their contribution to predicting the class variable. The Chi Square procedure is applied individually to each variable by first binarizing real-valued variables and then testing the expected minus observed counts with respect to the class, where the expected counts are taken under the assumption of independence. Larger counts give a higher chi-square statistic and suggest non-independence [75].

3.3.8 Machine learning analyses

After compiling our data and preprocessing our data, the next step was to employ machine learning algorithms to develop models. Drugs that had a "high affinity" for the transporters OAT1, OAT3, OCT1, and OCT2 were treated as "instances", and the physiochemical properties of the drugs were used as "attributes". Six pairwise comparison studies were conducted: OAT1 versus OCT1, OAT1 versus OCT2, OAT3 versus OCT1, OAT3 vs OCT2, OAT1 versus OAT3,and OCT1 versus OCT2, and in each comparison study, "overlapping" drugs, or ones that had high-affinity with the two transporters being studied, were eliminated from the analysis.

Several machine learning models that are available in WEKA were tried: decision trees, decision rules, support vector machine, Bayesian models, and neural networks. We were able to get classification models that were well-validated using several different techniques but preference was for models which could explain transporter binding data. The neural network model is a “black box” model, so it was not as useful for this purpose as decision rules or decision trees but still was an accurate classifier. Comparable classification success rates (Table 2) with several different algorithms demonstrate that there is a boundary between transporter selectivity. Also, of note, within a given model depending on the features of the model selected, different decision trees were generated likely due in part to the overlap in molecular characteristics captured by various attributes.. Multiple iterations of the algorithms and parameters were explored to arrive at models with the best validation scores.

In a decision tree, each node is a variable, and each branch represents a data split that depends on the value of the variable. An instance of the data determines a path down the tree, which ultimately leads to a leaf node that represents a class prediction. The decision tree is induced by ranking how well each variable can split the data at a decision node (starting with the root), splitting the data, and repeating the process for each branch. As the data gets split more and more, eventually each node will mostly reflect one class or the other, and the branching will stop. Typically, trees are induced and then lower levels are pruned back to improve performance in a cross-validation procedure. Compared to other techniques, a decision tree is more interpretable because the decisions are easily described as rules, even though the combination of data splits can be highly non-linear.

A random forest is an ensemble of decision trees, in which each tree is trained with different bootstrap samples (1000 in our case). The ensemble is averaged together

to produce an aggregate classification. The trees are made slightly de-correlated by limiting the choice of variables during tree induction so that different combinations of variables can fill out the tree branches. An additional benefit of the bootstrap is that one can estimate the detrimental effect of variable permutations on predictions for each “left out of bag” sample. That effect is averaged and normalized over all trees, leading to a measure of variable importance. Because a decision tree is nonlinear in the way it partitions the input, the variable importance is potentially a measure of both interaction and main effects [76].

3.3.9 Performing Student’s T tests for pairwise comparison of transporters

In addition to the machine learning approach, we used statistical tests to study the differences between substrate transporter interactions. In each of the pairwise comparison studies, t-tests were performed on the physiological properties to determine if the mean values for each are statistically different between the two groups of drugs. Then, the physiological properties were ranked according to their p-values.

3.3.10 Machine learning analysis on mid affinity drugs

To gain a deeper understand of our population of drugs and transporters and as a method of comparison with high affinity drugs, we also collected physiochemical attributes for drugs that were in the “mid affinity” range, and machine learning analysis was applied using a similar approach (a more stringent criteria was applied to define “mid affinity” drugs for the machine learning analyses, which would be drugs with affinity between 100 uM and 600 uM, than the charge state analysis which defined the mid affinity range as 100 uM to 1000 uM). This population of drugs was smaller than the high affinity population. The study with the smallest instance number was the comparison

study of OCT1 and OCT2, with the number of instances being, 6 and 5, respectively, after the overlapping drugs were removed. In some pairwise comparisons, the number of drugs interacting with one transporter was much larger than the other, so the model was adjusted to compensate. For example, when comparing OAT1 and OCT1, there were 23 drugs interacted with OAT1, and 10 drugs interacted with OCT1. This discrepancy was managed by adding the same OCT1 interacting drugs twice to the training set.

3.3.11 Building pharmacophore models based on high affinity binding drugs

The pharmacophore models were built in ICM. The ICM software was used to perform clustering, alignment, and pharmacophore building based on the atomic property field, APF. APF considers the 3D representation of atomic properties such as hydrogen bond donors, hydrogen bond acceptors, SP2 hybridization, lipophilicity, size of large atoms, and positive and negative charges [77]. High affinity drugs were chosen as “actives”. Since our actives were diverse in their 3D molecular structures, hierarchical clustering of actives based on APF was first done to separate them into groups. Then, actives among the same clustering group were aligned, and a pharmacophore model was generated for the aligned drugs. Each clustering group needed to be comprised of a minimum of 3 drugs with dissimilarity (1-similarity) less than or equal to 0.25. The value of similarity measures how similar two compounds are in APF and ranges from 0 to 1. Thus, drugs that were too dissimilar would not be considered for model generation.

3.3.12 Constructing the APF properties quantitative measurement table for the pharmacophore models

APF properties were determined for each pharmacophore model, and the vectors of each APF property across all the models were added to calculate the total for that property.

3.3.13 Performing the in-silico screening and the uptake inhibition assay

The OAT3 pharmacophore 9, which was found to possess uniquely cationic propensity among the OAT3 pharmacophores, were used to screen against *Drugbank* Database to identify cationic drugs that interact with OAT3. Among the 100 top hits, 6 cationic drugs were selected based on their availability, solubility, and charges for further wet-lab validation; they were loperamide hydrochloride, nebivolol hydrochloride, darifenacin hydrobromide, paliperidone, cisapride monohydrate, and halofantrine hydrochloride. For the uptake inhibition assay, *Oat3*-transfected Chinese hamster ovary (CHO) cells were cultured on 96-well plates, and the selected top hits and probenecid (as a control) were individually added into the cells in a dosage-dependent manner along with 20 uM 5CF, the fluorescent tracer that is also OAT3 substrate. To determine if these selected drugs also interact with OAT1, the same assay was done for Oat1-transfected CHO cells, and 10 uM 6CF was used as the fluorescent tracer instead.

3.4 Results

Through an extensive literature search, the number of pharmaceutical drugs and tracers identified to interact with OAT1, OAT3, OCT1, and OCT2 were 103, 105, 96, and 81, respectively at all affinity levels. Unless otherwise specified, machine learning

analysis,, statistical analysis, and pharmacophore modeling were done using drugs that interact with the transporters in the high affinity range (less than 100 uM).

3.4.1 The capability of OAT3 to interact with drugs that are cationic or zwitterionic neutral

As expected, OATs, mainly interacted with more anionic drugs, and OCTs primarily interacted with more cationic drugs (Figure 3.2). In addition to interacting with drugs of the same charge, as per their names (OATs are considered as “anionic” entities and interact with compounds of the same charge, or anions; in comparison, OCTs are considered as “cationic” entities and would interact with cations), the four transporters also had some ability to interact with “oppositely charged” drugs to some degree. Of all four transporters, OAT3 was shown to have the greatest capability to interact with cations and zwtterionic neutrals in the high and mid affinity range; in contrast, the other three transporters only interacted with oppositely charged or zwtterionic neutral drugs at the low or very low affinity range.

3.4.2 The capability of OAT3 to interact with drugs of positively charged species and zwitteronic-neutral species

Charge-species composition diagrams (Figure 3.3) demonstrate the unique interacting power of OAT3. Based on the charge-species bar diagrams for individual transporters at pH 7.4, it was noted that the charge-species OATs mainly interacted with were negatively charged species; and the OCTs mainly interacted with positively charged species, demonstrating that OATs are “anion” entities, and OCTs are “cation” entities. Then, the next most prevalent charge species that both OATs and OCTs interacted with were neutral species. All four transporters interacted with zwitterionic-neutral species and oppositely charged species but to the least extent compared to other

charge species. OAT3 exhibited more power in interacting with zwitterionic-neutral and opposite-charged species than the other transporters. This could be seen from the table that summarized the total percentage of various charge species for each transporter (Figure 3.4). At the physiological pH, it was shown that OAT1 could not interact with any positively charged species; in contrast, OAT3 was able to interact with positively charged species (which constituted 3.55% of species OAT3 interacting with). Both OCT1 and OCT2 interacted with negatively charged species, and the total negatively charged species percentages were 3.80% and 3.17%, respectively. Finally, the four transporters interacted with zwitterionic-neutral species in varying degrees; the total zwitterionic-neutral species percentages for OAT1, OAT3, OCT1, and OCT2 were the following: 2.75%, 5.44%, 1.78%, and 2.15%. This shows that OAT3 had the most ability to interact with zwitterionic-neutral species. To determine how well individual transporters interacted with oppositely charged and zwitterionic-neutral species together, we explored the total of oppositely charged species percentage and zwitterionic-neutral species percentage for each transporter. Among the four transporters, OAT3 had a much higher total percentage of oppositely charged species added to zwitterionic-neutral species than the rest of the transporters (the value for OAT3 was 8.98%, whereas the values for OAT1, OCT1, and OCT2 were 2.75%, 5.58%, and 5.33%, respectively).

3.4.3 Changing pH can potentially change the ability of transporters to interact with drugs of oppositely charged species and zwitteronic neutral species

Furthermore, we explored how varying the pH of the solution in-silico might change the composition of charge species that each of the transporters interacted with. At different pH levels, the compositions of charge species for drugs would vary. In a

more acidic environment, drugs would be protonated and contain more positively charged species, and in a more basic environment, drugs would be deprotonated and contain more negatively charged species. This would affect the charge species composition of drugs that individual transporters interacted with, as seen in Table 1. The sum of the oppositely charged species percentage and zwitterionic-neutral species percentage for both OAT1 and OAT3 increased as pH decreased, and the values of such sum for OCT1 and OCT2 increased when pH shifted toward the basic direction. In addition, it was found that the drugs that interacted with OAT3 were the most versatile compared to drugs interacting with the other 3 transporters as the charge states of OAT3 drugs would change most dramatically throughout the pH range of 7.2 to 7.6; when pH was either lowered or increased, the sum of the total oppositely charged and zwitterionic-neutral species for OAT3 changed from 8.98% to 9.93% when pH was lowered and decreased from 8.98% to 8.31% when pH was increased. In contrast, those values for OAT1, OCT1, and OCT2 did not change as drastically.

3.4.4 Substrate overlap between OATs and between OCTs

The Venn diagrams were created using all drugs that interacted with the four transporters with an affinity lower than 100 uM (Figure 3.5). The Venn diagrams visually display the functional similarities between transporter-substrate interaction. OAT1 and OAT3 shared many substrates, 43 drugs; similarly, OCT1 and OCT2 shared 41 substrates. When we compared across members from the OATs and members from the OCTs, there was less overlap, showing the functional differences between OATs and OCTs. Overlapping drugs, the ones interacting with two transporters, were excluded from our data mining analysis. The reason for this exclusion was to attempt to identify subtle differences between transporters, and excluding the overlapping drugs could help us study the extremity of transporting functions between transporters.

3.4.5 Machine Learning Analysis

Results of applying classification algorithms using decision trees, neural networks, support vector machine, d decision rules, and naïve Bayes for the comparison of OAT1 drugs and OCT1 drugs are presented in Table 3.1.

Decision trees developed using the J48 algorithm and random forest will be presented in detail for all the pairwise comparisons since they provide a logical way to demonstrate how physiochemical properties of the substrates affect the binding interaction between substrates and transporters.

3.4.6 The use of ROC area as a measurement to determine more differences exist between the OATS than between the OCTS

Table 3.2 shows the summary of weighted average ROC areas for the six decision tree models based on high-affinity drugs when performing ten-fold cross validation. Most decision tree models were well validated, and only two trees had ROC areas less than 0.80, which were the trees for OAT1/OAT3 and OCT1/OCT2. This was likely due to the fact that substrates for the two OATs and two OCTs were highly similar among each other, and any model built to identify and predict the differences between them becomes difficult. The ROC areas for OAT1/OAT3 and for OCT1/OCT2 were 0.795 and 0.545, respectively, and this revealed that functional differences between OAT1 and OAT3 were more easily discriminated than between OCT1 and OCT2. This is an important point for the analyses that follow.

3.4.7 The pair-wise comparison between an OAT and an OCT showed the differences between OATs and OCTs were mostly due to charge

When an OAT was compared with an OCT in decision tree analysis, it was found that the first two physiochemical attributes that separated an OAT from an OCT were the number of negative charges and the number of positive charges (Figure 3.6). This is consistent with previous experimental data across mammalian species for many OAT1 and OCT1 substrate that include not only drugs, but also metabolites and toxins. Drugs that had the number of negative charge greater than zero were classified to interact with OATs; in contrast, drugs that had the number of positive charge greater than zero interacted with OCTs. After charges, the next determinant attribute seen in most trees was SP3 character, in which drugs with greater value would be classified as OCT drugs, and vice versa.

3.4.8 The OAT/OCT decision trees excluding charge showed the OATs and OCTs were different in physiochemical properties other than charge

Importantly, to understand properties that were important in separating OAT interacting drugs from OCT interacting drugs in additional to charge, we constructed decision trees that excluded the properties positive charge and negative charge (Figure 3.7). The resulting trees split on a variety of properties among which are the number of acyclic double bonds(“adb”), number of acyclic oxygen(“ao”), and number of acyclic nitrogen”(an”), and “SP3 character” were dominant (a description of these attributes will follow). In the OAT1/OCT1 tree, the first attribute that split was “adb”; drugs that had

zero “adb” would be classified as OCT1 drugs. Then, the next attribute was “an”, and if drugs had “an” greater than 3, they would be classified as OCT1 drugs. When we examined the three other OAT vs OCT trees, they followed a similar trend as the OAT1/OCT1 tree; OCT substrates had higher number of “an” than OAT substrates, and OAT substrates had higher numbers of “adb” and “ao” than OCT substrates. Interestingly, statistics from the accuracy of these models were not as strong as ones including charge but were still significant; results of those correctly classified were between 76% to 81% and results of ROC area were between 0.75 to 0.83. In addition, the attributes “ao”, “adb” and “SP3 character” are identified and were confirmed as important attributes in the t-test statistical analysis.

3.4.9 The pair-wise comparison between OAT1 and OAT3 showed that differences remained between the two OATs

When OAT1 and OAT3 were compared (Figure 3.6), the first attribute separating OAT1 substrates and OAT3 substrates was the number of acyclic tetravalent nodes (“aqv”). Drugs that have the number of acyclic tetravalent nodes greater than 7 would be classified as interacting with OAT3. The next attribute separating the OATs was the number of phosphorous atoms (“p”). Drugs that had at least one or more phosphorus would be classified to interact with OAT1. A third attribute that emerged from the tree was the number of positive charges; drugs with a positive charge were associated with an OAT3 classification. The abovementioned properties will be explained in more detail. In contrast to the comparison of the two OATs, the model generated for comparing the two OCTs had poor validation performance, and, hence, the results for that model would not be discussed further.

3.4.10 The results of random forest were in agreement with the results of decision trees

Figure 3.8 below shows the variable importance plots derived from the Random Forest model for the pairwise OAT and OCT comparisons. The charge state information was also found to dominate in the ranking, and this supports the results that the higher nodes in the decision tree are robustly important for classification across the bootstrap samples in the Random Forest. After the charge attributes, the variables found to be important were “adb”, “ao”, and “SP3 characters”, which are consistent with the decision trees. After 5 or 6 variables, the importance levels drop off and the marginal differences are minimal.

For the pairwise comparison between OATs the results also confirm and justify the decision tree interpretation. However, for Oct1 vs Oct 2, the results are not aligned, which is not surprising given that the classification performance is barely above chance.

3.4.11 The statistical analysis confirmed the results of the machine learning analyses and provided further support for the differences between various transporters

When performing t-test analyses on individual attributes for each pairwise transporter comparison, we identified a number of attributes to be statistically different between substrates interacting with each pair of transporters. The attributes that had the least p-values for each comparison are summarized in Figure 3.9 and are consistent with the machine-learning analyses. The two properties that had the lowest p-values were the number of positive charge and the number of negative charge, corresponding to the results from the machine learning analyses. After positive and negative charge, the next attributes that came out from the ranking were numbers of acyclic double bond

(“adb”), acyclic oxygen (“ao”), hydrogen bond acceptor site (“hbam”), and SP3 character. For the pair-wise comparison of the two OATs, the twp properties seen in the OAT1/OAT3 decision tree had p values less than 0.01, and among them, number of acyclic tetravalent nodes (“aqv”) and number of positive charges were found to have the lowest p-values in the ranking. Again, our results from both decision trees and random forest are consistent with the statistical analysis.

3.4.12 Explanation of properties found to be relevant in results

Based on the results of machine learning and statistical tests, we found that OAT (either OAT1 or OAT3) had higher numbers of negative charge, acyclic double bonds, acyclic oxygen, and hydrogen bond acceptor sites than an OCT (either OCT1 or OCT2). All these properties are associated with the anionic propensity. For example, most acyclic double bonds within the structures were in the forms of carbonyl (O=C), thial (S=C), sulfoxide (S=O), and the electro-negative oxygen and sulfur within these double bonds are prominent hydrogen bond accepting sites. The number of acyclic oxygen is another property that expresses the anionic propensity as the acyclic oxygen also serves as a potential hydrogen bond accepting site.

In addition to having differences in properties associated with charges and ionization, OCTs and OATs are different in geometry-related properties, particularly in the SP3 character value. SP3 character is defined as the number of SP3 hybridized carbon divided by total number of atoms and is one measure of the degree of three-dimensionality of a compound. If a drug has a higher SP3 character value, it is more three-dimensional; likewise, a lower SP3 character value is taken to imply the drug is more planar [78, 79]. In machine learning models and statistical analyses, drugs with a stronger affinity for OCT had a greater SP3 character value than OATs, supporting the view that the “OCT interacting drugs” are more three-dimensional than “OAT drugs”. As

measured by SP3 character, compared with most other drugs in the data set, amantadine, nandrolone, and atropine are three OCT drugs that have highly three-dimensional structures, each with SP3 character value of 0.357, 0.326, and 0.227, respectively. On the other hand, OAT drugs have much lower values of SP3 character, with none of the OAT drugs having SP3 character values greater than 0.300.

Some differences are also observed among the two OATs; OAT3 tended to interact with drugs with more acyclic tetravalent nodes and more positive charges whereas OAT1 tended to interact with drugs with more phosphorus atoms. An acyclic tetravalent node usually is composed of a carbon forming tetravalent bonds with four elements. In the decision tree model, 11 drugs were classified as OAT3 drugs from this node; among them were verapamil, pravastatin, enalapril, and methotrexate, and with the higher number of acyclic tetravalent nodes, these drugs having longer and more hydrophobic chains. The next attribute separating OAT1 and OAT3 substrates was the number of phosphorous atoms ("p"). Drugs that had at least one or more phosphorus atoms were classified as interacting with OAT1; the three drugs in this category were cidofovir, tenofovir, and adefovir. When looking at the chemical structures of these drugs, it was found the phosphorus atoms were in part of phosphate groups. Since the phosphate group contains several oxygen atoms binding with phosphorus, and some of them were deprotonated at the normal pH range the phosphate group is highly anionic. Thus, the number of phosphorus atoms was directly correlated with the anionic propensity. In summary, even though both OAT1 and OAT3 were known to have functional overlap, there were some differences between them identified in our analyses. OAT3 preferred to interact with drugs with long, hydrophobic chains and more positive charge, and OAT1 was functionally more anionic than OAT3.

3.4.13 The machine learning based on mid affinity drugs confirmed the results of the analyses based on high affinity drugs

In addition to understanding the molecular interaction between transporters and drugs that bind with high affinity, we also tried to study how transporters interact with drugs in the mid affinity range (100uM to 600uM). The decision trees based on mid affinity drugs (Figure 3.10) showed that major factors in classifying a drug as an OAT or an OCT substrate were due to charge as in the high affinity group, but the separation was less impressive, suggesting that the functional differences between OATs and OCTs in the mid-affinity binding interaction did not depend on the charge of substrate as much as they did in the high-affinity binding interaction. The decision tree that compares OAT1 and OAT3 in mid-affinity range only had one node, which split on positive charge. Drugs with a positive charge were classified as having an OAT3 interaction.

3.4.14 The pharmacophore models showed the structural similarities corresponding to the overlap in functions for OATs and for OCTs

Since it was found that OAT3 substrates also possessed some cationic characteristics based on the machine-learning analyses, pharmacophore models for OAT3, OCT1, and OCT2 were built to compare the functional similarities/differences between the OAT and the OCTs in the three-dimensional space. Because the drugs interacting with each transporter were diverse in their 3D structures, the first step was clustering the drugs into groups based on atomic property field (APF). Then, the drugs among the same clustering groups were aligned, and the pharmacophore models were

created (Figure 3.11, 3.12, 3.13). In the pharmacophore models, different colors captures various APF properties: blue - hydrogen bond donor; red – hydrogen bond acceptor; white – aromaticity; yellow – hydrophobicity; light red – negative charges; light blue – positive charges. The models showed that OAT3 and OCTs interacted with drugs that had hydrophobic and aromatic centers. However, a slight difference in compound backbone appeared as the hydrophobic chains for OCT1 and OCT2 models would sometimes enclose cationic spheres (seen in OCT1 pharmacophore model 3, 4, 5, and 6), which is not observed in most OAT3 models. Overall, OAT3 models were more anionic, and OCT models were more cationic. This could also be seen from Figure 3.14, which showed the quantitative measurements of the seven properties for individual models; as measured by the mean, the table showed that OATs had higher “hydrogen bond acceptors” and higher “negative charges”; in contrast, OCTs had higher “hydrogen bond donors” and higher “positive charges”.

3.4.15 The pharmacophore models revealed structural similarities between OAT3 substrates and OCT substrates

Even though the majority of pharmacophore models for OAT3 had similar features, there was one exception, which was pharmacophore model 9 for OAT3 (Figure 3.11). Unlike other OAT models, this model contained a hydrophobic chain that tended to enclose a sphere enriched with hydrogen bond donors and positive charges, which was a pattern shared among many OCT1 and OCT2 models. Thus, this model (OAT3 pharmacophore model 9) was found to be very OCT-like, and the quantitative APF measurement of this model was found to have greater values of “positive charges” and “electropositive charges”.

Interestingly, the list of drugs used to construct model 9 for OAT3 was found to be highly similar to the list of drugs that was independently separated based on the first attribute or node in the OAT1/OAT3 decision tree (Figure 3.6). Out of the 9 drugs used to construct the pharmacophore model, 6 of them contained more than 7 acyclic tetravalent nodes and were classified as “OAT3” drugs in the decision tree. This demonstrates that the results from the decision trees and the pharmacophore models identified the same differences found between the functions of OAT1 and OAT3, and the differences were due to the capability of OAT3 to interact with OCT-like substrates.

3.4.16 The in-silico screening based on the OAT3 cationic pharmacohore identified new cationic drugs that preferentially interact with OAT3 but not OAT1

Using the pharmacophore model 9 of OAT3, a virtual screening of Drugbank Database was done to identify new OAT3 cationic ligands. 6 top hits were selected for further wet-lab validation, 4 of them were found to interact with OAT3 with strong inhibition (Figure 3.15), which is another way to validate our system of pharmacophore models. In addition, when these 6 cationic drugs were tested in the OAT1 uptake assay, it was found that only two of them inhibited the OAT1 function, and in a much lower affinity. The preference of these compounds to interact with OAT3 but not OAT1 demonstrates the unique capability of OAT3 to interact with cationic drugs. The measured IC50 values of tested compounds against OAT1 and OAT3 are summarized in Table 3.3.

3.5 Discussions

Recent knockout and in vitro data on a limited set of substrates suggest that substrate specificity of the OATs and OCTs of most clinical and pharmaceutical interest

is more complex than whether or not the ligand is an anion or a cation. However, the "rules" may not a straightforward. To systemically address to this question, an extensive literature search was first done to build the most complete transporter-ligand database (any compound found to interact with the transporters of study was curated), and within the data, all drugs known to interact with OAT1, OAT3, OCT1, and OCT2 were selected and used to study the functional differences and similarities between the transporters by applying machine learning tools. Among the machine learning tools, decision trees and random forests were helpful to understand this question of substrate specificity.

The results of the decision tree analyses were in agreement with the results of the random forest, and were further verified by the conventional statistical tests. The results indicated that, while the main difference between the substrate preferences of OATs and OCTs was to the charge of ligands, the structure of ligands could also affect the interaction between the transporters and the ligands, and it was found that OCTs would interact with more three-dimensional structures, whereas OATs would interact with planar compounds. This suggests that the binding pockets of OCTs are more three-dimensional than those of OATs and may require further investigation.

In addition to finding differences between the two sub-families, some differences among the sub-families are also identified. Based on the machine-learning models and pharmacophore models, OAT1 and OAT3 are found to be different as the latter possesses some capability to interact with cations, making it functionally similar with OCT1 and OCT2. In the high affinity range, OAT3 could recognize ligands in more diverse structures than OAT1, suggesting that OAT3 has more different binding pockets than OAT1. Moreover, some of these binding pockets could be similar to some of the binding pockets of OCTs, providing OAT3 with some cationic characteristics. With its greater multi-specific property, OAT3, unlike OAT1, can thus interact with cationic drugs

and zwitterionic neutral-charged drugs that exhibit cationic characteristics. This indicates that with that, the boundary that separates OATs and OCTs is not as clear as the current literature suggests.

Understanding the substrate preference of transporters to interact with a particular substrate group is valuable as it not only helps distinguish the functional differences between transporters but also suggests, pending accurate 3D structures, complementary properties of the transporters. Each transporter is responsible for a set of substrates, but when they join together, overall homeostasis is reached. In addition, finding the differences and similarities between the transporters with respect to substrate preference can help us to predict and identify new compounds that interact with the transporters, since the set of rules defined by decision trees can be further used for in-silico screening of new substrates/inhibitors. These rules can also be used to design new, potent, selective substrates that can target a particular transporter. These could be drugs that are aimed at a particular tissue or body fluid, or alternatively, selective inhibitor of transport.

3.6 Conclusion

In the high affinity range, OAT3 could recognize ligands in more diverse structures than OAT1, suggesting that OAT3 has more different binding pockets than OAT1. Moreover, some of these binding pockets could be similar to some of the binding pockets of OCTs, providing OAT3 with some cationic characteristics. With its greater multi-specific property, OAT3, unlike OAT1, can thus interact with cationic drugs and neutral-charged drugs that exhibit cationic characteristics. The multi-role that OAT3 plays, as both anion transporters and cation transporters, suggests that the barrier that separates OATs and OCTs according to their functions is not so definitive, and should

these renal transporters be named as either anion transporters or cation transporters raise more concerns. In addition, the decision trees built in this analysis could be further used to predict novel substrates *in silico* and to design drugs that can specifically interact with one transporter.

3.7 Acknowledgment

The work presented in Chapter 3, in full, is being prepared for submission. The authors are Liu, Henry C.; Goldenberg, Anne D.; Chen, Yuchen; Lun, Christina; Rodriguez, Paul; Balac, Natasha; Abagyan, Ruben; Nigam, Sanjay K. The dissertation author was the primary investigator and author of this paper.

3.7 Tables

Table 3.1: The validation performance of various machine learning tools done for the comparison of OAT1/OCT1 high-affinity drugs. Models based on different tools generate well-validated results, and since decision trees provide us the most logical way to explain how physiochemical properties affect substrate preferences between transporters, we focus them further.

Machine Learning Models	Correctly Classified	ROC Area
Decision Tree	86.5672 %	0.905
Support Vector Machine	93.2836 %	0.934
Neural Network – Multilayer Perceptron	88.0597 %	0.945
Naïve Bayes	88.806 %	0.933
Decision Rule	86.5672 %	0.918

Table 3.2: The validation performance of various decision tree analyses. The table summarizes the results using ten-fold cross validation of machine learning decision tree models for: a.) high affinity drugs (with affinity less than 100 uM), b.) high affinity drugs without using charge as an attribute, and c.) mid affinity drugs (with affinity between 100 to 600 uM). Note the poor results in the OCT1/OCT2 analysis due to a small data set of 6 and 5 instances.

Transporters Compared	a. High affinity drugs (with affinity less than 100 uM)		b. High affinity drugs (with affinity less than 100 uM) excluding charge as an attribute		c. Mid affinity drugs (with affinity between 100 and 600 uM)	
	% correctly classified	ROC area	% correctly classified	ROC area	% correctly classified	ROC area
OAT1/OCT1	86.5672 %	0.905	80.6 %	0.823	82.5 %	0.874
OAT1 /OCT2	83.3333 %	0.932	78.9 %	0.835	82.2 %	0.868
OAT3/OCT1	86.3309 %	0.88	77.7 %	0.764	80.0 %	0.88
OAT3/OCT2	93.2773 %	0.932	72.3 %	0.774	70.8 %	0.779
OAT1/OAT3	69.7674	0.795	NA	NA	86.4 %	0.722
OCT1/OCT2	66.6667	0.639	NA	NA	45.5 %	0.450

Table 3.3: The IC50 values of the cationic drugs tested for the function of OAT3 and OAT1. The data for probenecid uptake inhibition is shown as a control, and it is found that most of the tested cationic drugs selectively inhibit OAT3 function.

Drug Name	**OAT3 IC_{50} (uM)**	**OAT1 IC_{50} (uM)**
Loperamide	95	no significant inhibition
Paliperidone	260	1082
Nebivolol	169	no significant inhibition
Derifenacin	198	807
Cisapride	No inhibition	No inhibition
Halofantrine	No inhibition	No inhibition
Probenecid	33	2.4

3.8 Figures

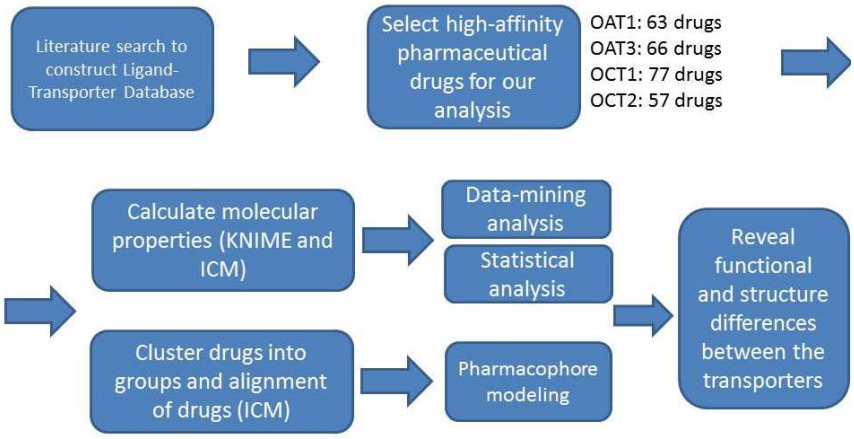


Figure 3.1: The overall strategy applied in this study. A comprehensive literature search was done to identify all the pharmaceutical drugs and tracers known to interact with any of the four transporters. Among these pharmaceutical entities, ones that have high-affinity interaction were selected for further data-mining analysis and pharmacophore modeling.

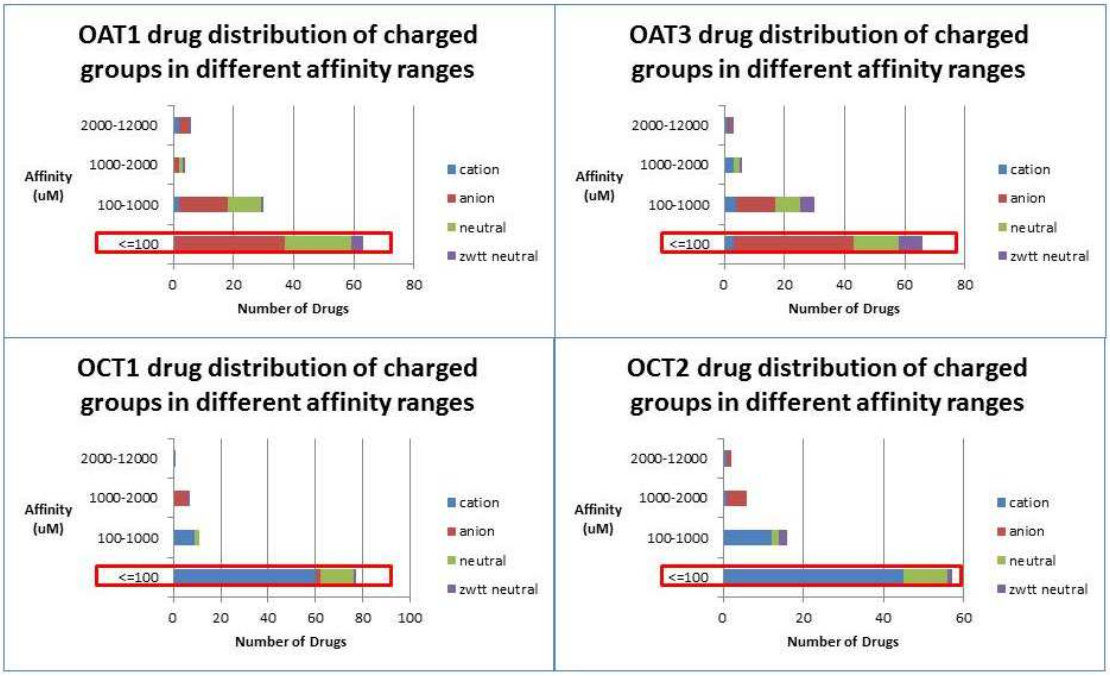


Figure 3.2: The distribution of charge states for pharmaceuticals that interacted with each of the transporters at various binding affinity ranges. The charge states of the pharmaceuticals were defined by considering the number of positive charges and negative charges calculated in ICM at the environment of pH = 7.4.

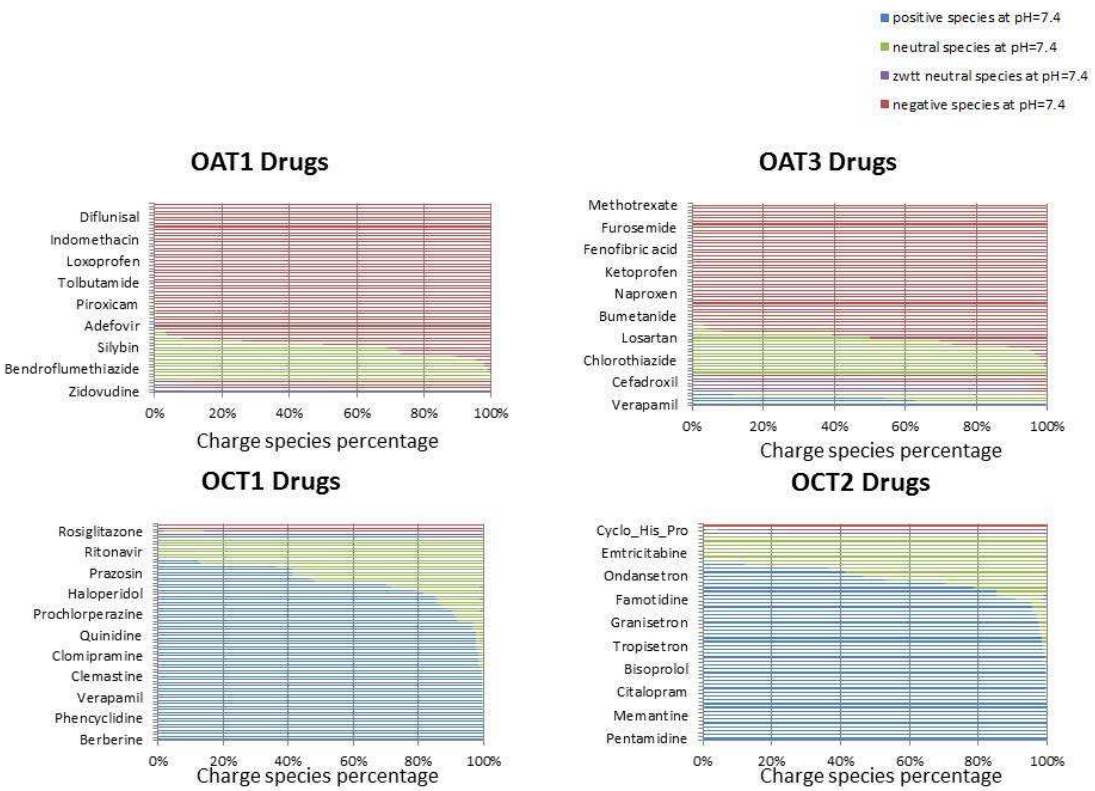


Figure 3.3: The charge-species composition diagrams for the transporters. The charge species composition for individual pharmaceuticals was measured based on the pH/concentration curves found in Chemicalize.org (an online compound database supported by Chemaxon), which were then grouped according to which transporters the pharmaceuticals interacted with. The diagrams could indicate the capability of the transporters to interact with various charge species.

OAT1 Drugs	pH = 7.2	pH = 7.4	pH = 7.6
Total positive %	0.00%	0.00%	0.00%
Total negative %	78.05%	79.42%	80.56%
Total neutral %	18.88%	17.83%	16.97%
Total zwtt neutral %	3.07%	2.75%	2.46%
Total positive charge + zwtt neutral %	3.07%	2.75%	2.46%

OAT3 Drugs	pH = 7.2	pH = 7.4	pH = 7.6
Total positive %	4.04%	3.55%	3.17%
Total negative %	69.25%	69.53%	70.24%
Total neutral %	20.82%	21.48%	21.44%
Total zwtt neutral %	5.89%	5.44%	5.15%
Total positive charge + zwtt neutral %	9.93%	8.98%	8.31%

OCT1 Drugs	pH = 7.2	pH = 7.4	pH = 7.6
Total positive %	76.54%	74.24%	71.73%
Total negative %	3.65%	3.80%	3.94%
Total neutral %	18.06%	20.18%	22.46%
Total zwtt neutral %	1.75%	1.78%	1.87%
Total negative charge + zwtt neutral %	5.40%	5.58%	5.81%

OCT2 Drugs	pH = 7.2	pH = 7.4	pH = 7.6
Total positive %	75.98%	74.28%	72.36%
Total negative %	3.01%	3.17%	3.32%
Total neutral %	18.75%	20.39%	22.25%
Total zwtt neutral %	2.26%	2.15%	2.07%
Total negative charge + zwtt neutral %	5.27%	5.33%	5.38%

Figure 3.4: The summary table of the total percentage of various charge species for each transporter based on the results of charge-species composition diagrams. The capability of OAT3 to interact with positively charged and zwitterionic species is observed.

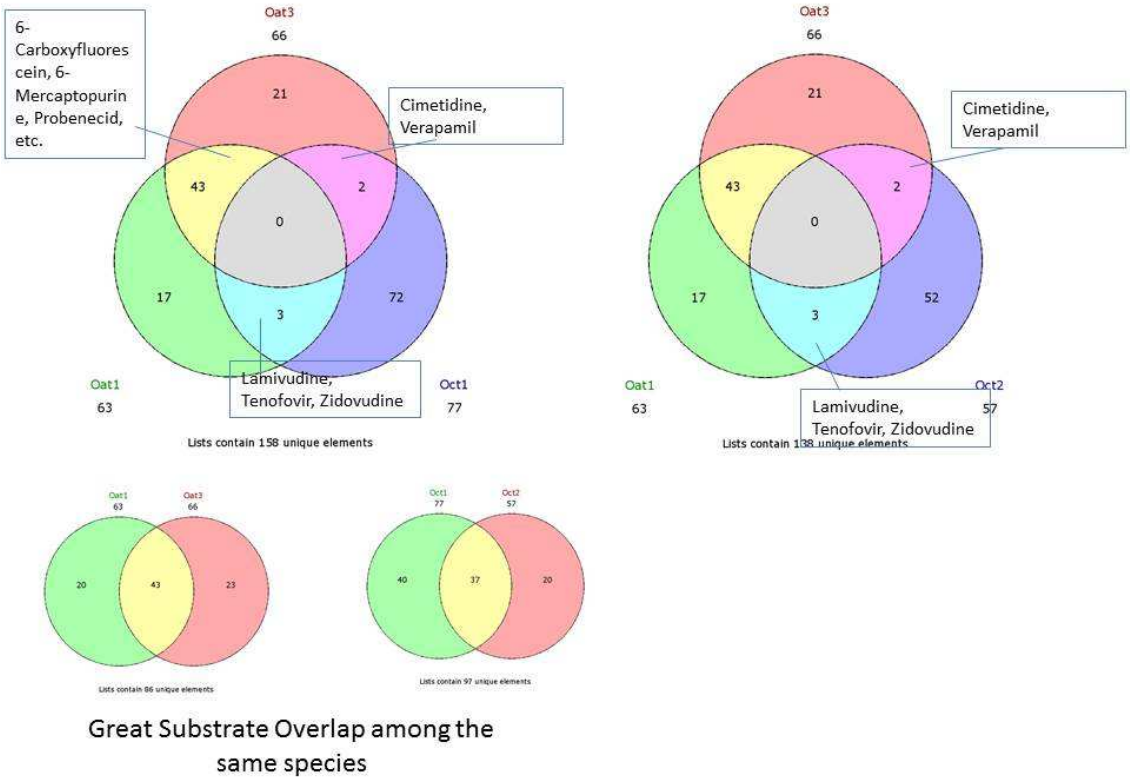


Figure 3.5: Substrate overlap among transporters. The Venn diagrams demonstrate the substrate specificity and substrate multi-specificity between the transporters. In addition, the drugs found to be overlapping were excluded for the following machine learning analysis. Note: While cimetidine and verapamil can bind OAT1 as well (Ahn et al., 2009), the affinity is roughly 10-fold less.

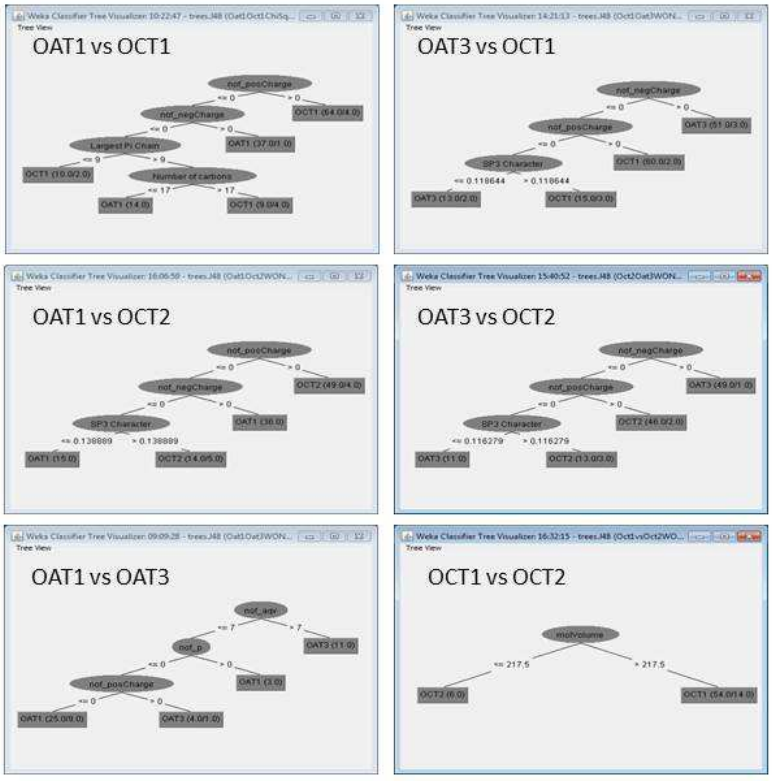


Figure 3.6: The decision trees based on drugs that interact with the transporters at high-affinity range. The decision trees show that the main difference between OATs and OCTs are due to charge and charged-associated properties. Besides charge, the geometry property, named SP3 character, was found to be another important factor in separating OAT and OCT drugs. In addition, some differences were found between two OATs, specifically in number of aqv, p, and posCharge (these attributes are further explained in the text).

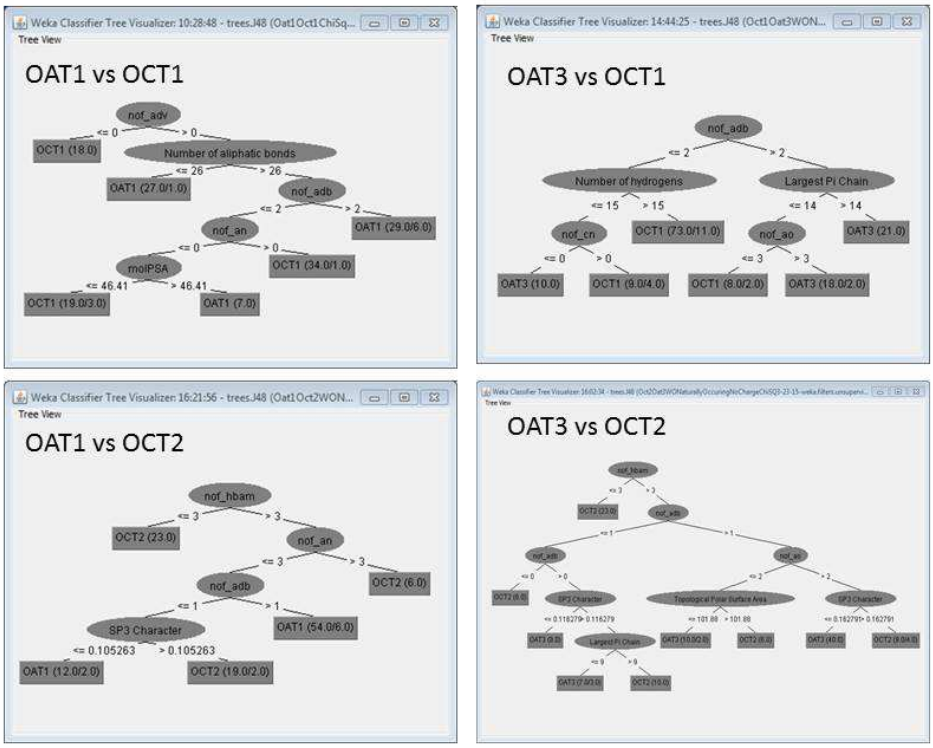


Figure 3.7: The decision trees excluding charge properties. The attributes of positive charge and negative charge were excluded for building the model so as to identify other important properties in separating OAT and OCT drugs. Again, it was found that some charge-associated attributes and SP3 character were the determinants.

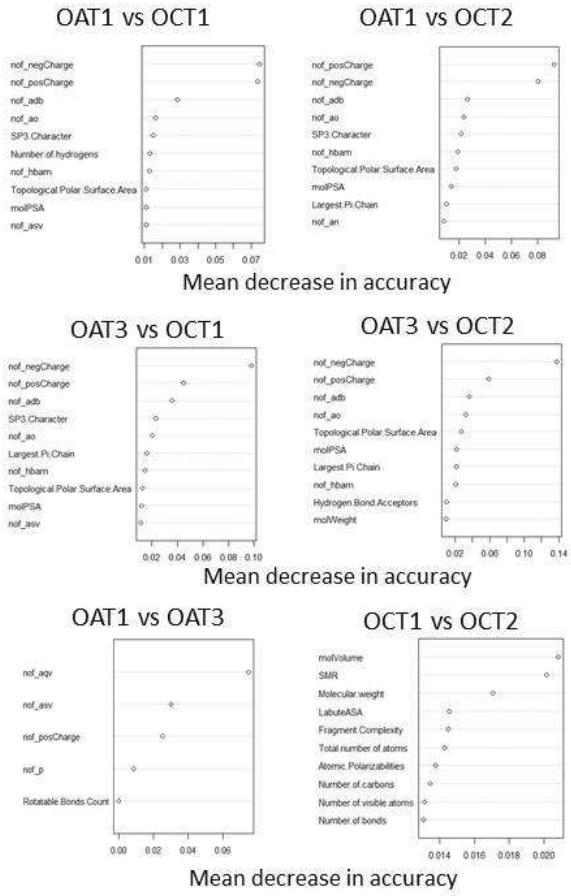


Figure 3.8: The mean decrease in accuracy based on the random forest analyses. These results are highly comparable to the results from decision trees.

T-tests analysis: to determine if individual attributes are statistically different in pairwise comparison of all 6 transporters

Pair-wise comparison	Number of attributes with P<0.005	Number of attributes with P<0.01	Number of attributes with P<0.05	Top 8 attributes ranked according to p-values							
OAT1 vs OCT1	27	29	37	nof_posCharge	nof_negCharge	nof_adb	nof_ao	nof_hbam	SP3 Character	Number of hydrogens	nof_asv
				5.95E-23	4.92E-13	2.4E-12	3.17E-09	2.28E-08	1.34E-07	7.93E-07	1.11E-06
OAT3 vs OCT1	17	18	23	nof_negCharge	nof_posCharge	nof_adb	nof_ao	nof_hbam	molPSA	SP3 Character	Topological Polar Surface Area
				1.10099E-16	2.44E-15	4.98E-13	3.07E-10	7.08E-09	1.38E-06	2.7E-06	3.63E-06
OAT1 vs OCT2	16	21	28	nof_posCharge	nof_negCharge	nof_adb	nof_ao	nof_hbam	SP3 Character	molPSA	nof_adv
				1.45E-19	1.31E-13	1.52E-11	1.15E-10	1.08E-09	4.59E-07	8.03E-07	1.64E-06
OAT3 vs OCT2	17	20	27	nof_negCharge	nof_posCharge	nof_ao	nof_adb	nof_hbam	Hydrogen Bond Acceptors	molPSA	nof_adv
				3.52E-17	8.12E-14	1.71E-11	3.24E-11	1.27E-09	1.03E-06	1.32983E-06	1.66E-06
OAT1 vs OAT3	1	2	18	nof_aqv	nof_posCharge	nof_asv	nof_asb	nof_rbc	molVolume	Fragment Complexity	Number of hydrogens
				0.00257	0.00586	0.01339	0.01364	0.01432	0.01828	0.01924	0.01964
OCT1 vs OCT2	6	12	30	nof_s	Molecular weight	LabuteASA	molWeight	Number of heavy atoms	SMR	nof_hac	Vertex adjacency information magnitude
				0.00171	0.00382	0.00451	0.00465	0.00495	0.00497	0.00514	0.00527

Figure 3.9: The results of statistical tests. The student's T tests calculated the p-values for each attribute for each pairwise transporter comparison and identified properties of significance. The results are found to be consistent with the results from the machine learning analyses.

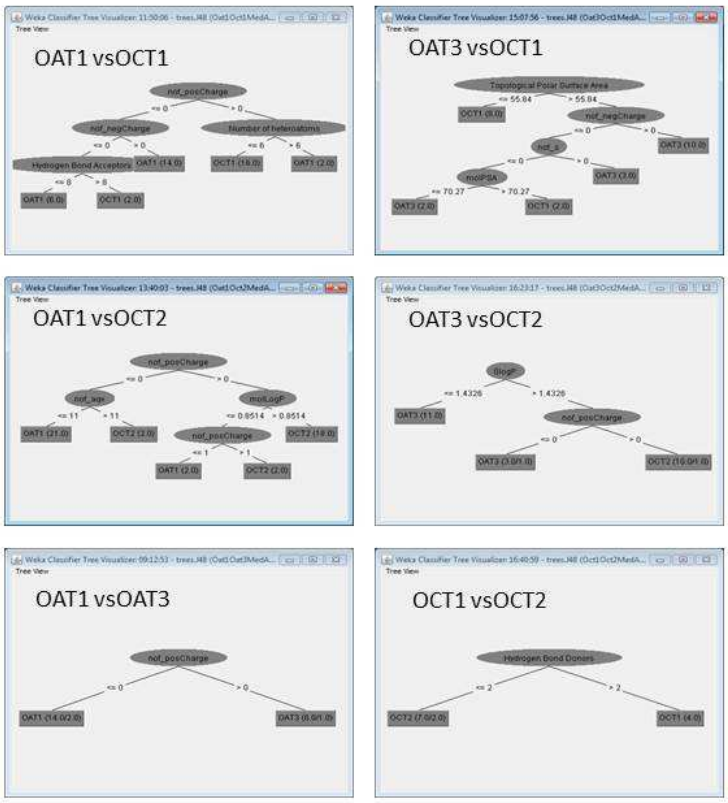


Figure 3.10: The decision trees based on drugs that interact with the transporters at mid-affinity range (between 100 and 600 uM). The trees show that, in the mid-affinity range, the main differences between an OAT and an OCT drugs were still due to charges, but to a lesser degree than drugs in the high-affinity range.

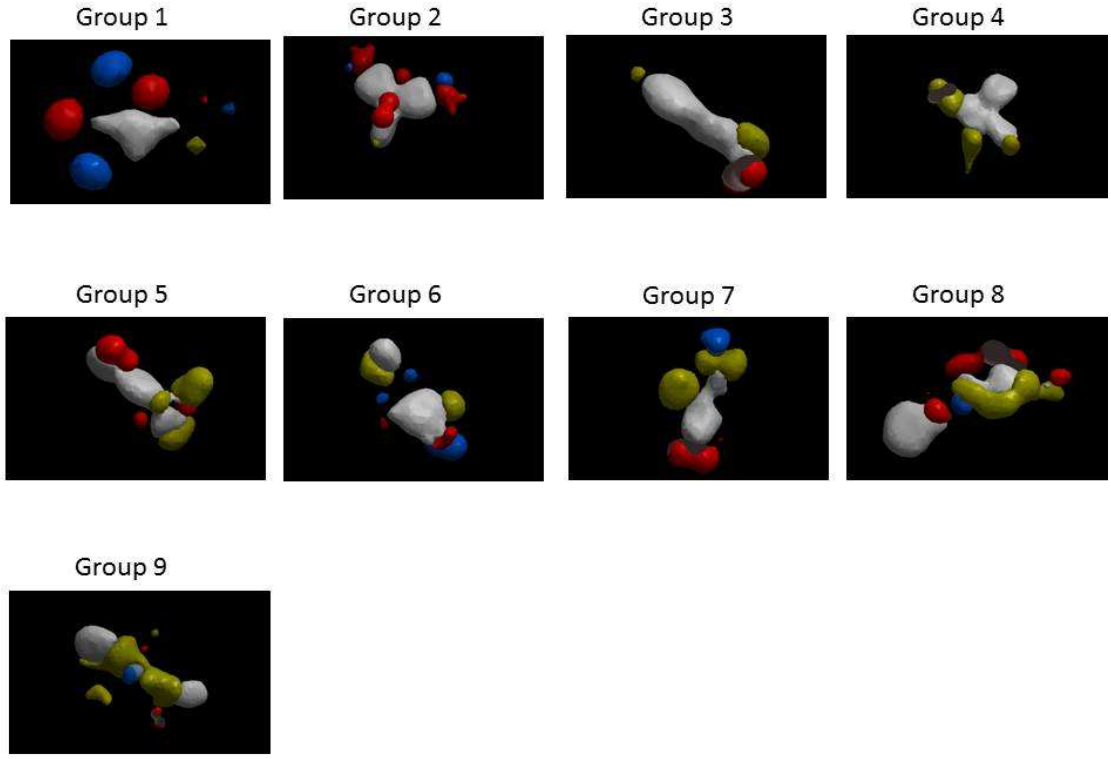


Figure 3.11: The pharmacophore models for OAT3. The OAT3 models and OCT1 models are found to be distinctive as the OAT3 models contained more characteristics of negative charges, electronegativity, and hydrogen bond acceptors, and vice versa for OCT1 models. However, one of the OAT3 models was an exception as it contained several characteristics only found in OCT models.

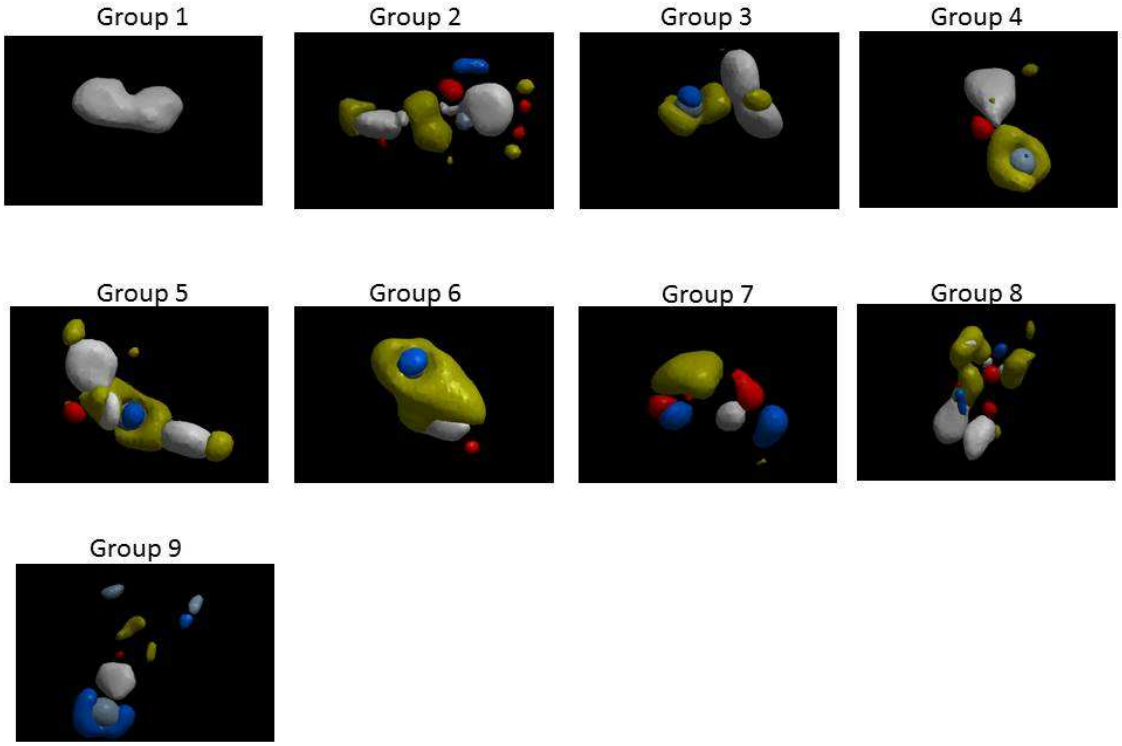


Figure 3.12: The pharmacophore models for OCT1.

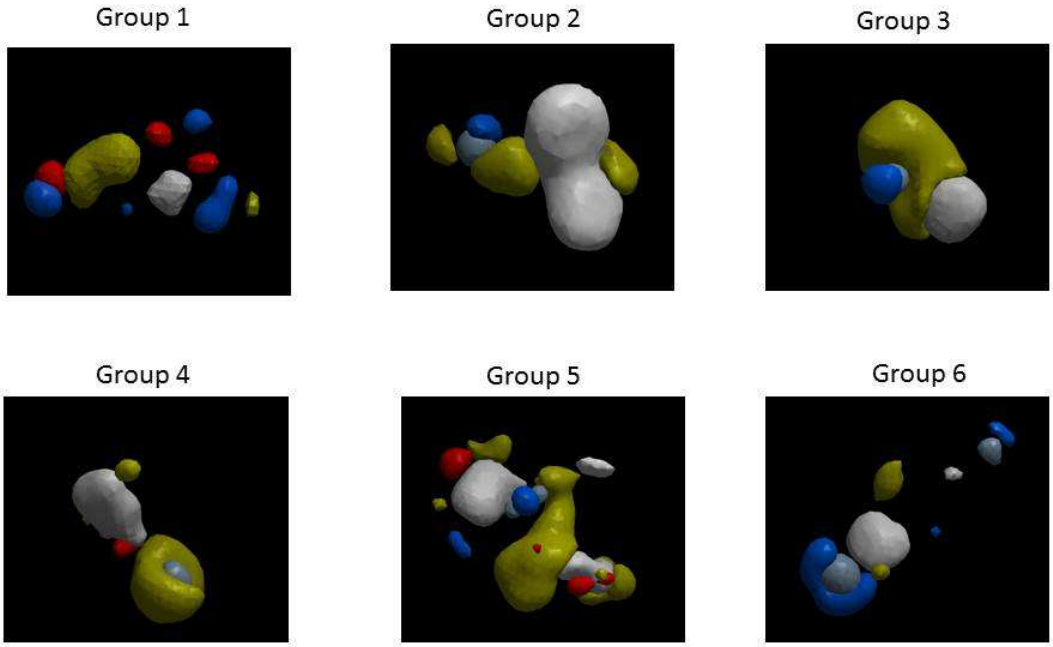


Figure 3.13: The pharmacophore models for OCT2.

OAT3	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	mean
hydrogen bond donors	143.2	68.1	7.9	88.2	35.9	220.1	197.1	123.9	66.1	105.6
hydrogen bond acceptors	154.1	484.4	175.6	189.3	387.1	169.2	244.1	427.1	252.7	275.9
Sp2 hybridization	839.7	1560.9	1009.0	1441.7	1795.9	894.6	881.1	1445.6	850.4	1191.0
lipophilic	101.5	628.7	493.4	925.4	923.1	581.0	588.4	664.1	1164.0	674.4
size(large)	213.7	773.6	636.5	1058.2	1148.9	656.9	752.1	934.1	1388.8	840.3
charge: pos	0.0	0.0	5.6	0.0	0.0	0.0	57.7	0.0	73.3	15.2
charge: neg	0.0	-230.9	-194.0	-138.6	-197.9	-38.5	-173.2	-181.4	-176.0	-147.8
electro: pos	219.7	587.8	529.8	964.1	987.3	588.5	773.3	686.3	1331.4	740.9
electro: neg	-117.1	-289.5	-102.4	-140.4	-151.6	-325.5	-176.7	-246.4	-99.7	-183.3

OCT1	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	mean
hydrogen bond donors	48.2	90.1	89.4	74.4	80.2	77.4	216.4	193.6	554.4	158.2
hydrogen bond acceptors	31.5	195.5	32.2	96.0	150.2	68.2	229.5	217.2	58.0	119.8
Sp2 hybridization	1056.6	1020.2	845.2	670.3	951.3	211.2	556.2	1639.8	1109.8	895.6
lipophilic	570.6	716.8	850.4	748.8	1241.6	896.3	339.7	1511.1	377.7	805.9
size(large)	704.8	937.2	951.4	882.2	1355.4	963.0	495.4	1759.9	472.5	946.9
charge: pos	102.6	46.2	110.8	115.5	115.5	69.3	17.7	46.2	346.4	107.8
charge: neg	0.0	0.0	0.0	0.0	0.0	0.0	-56.2	0.0	0.0	-6.2
electro: pos	823.4	829.4	997.7	979.2	1323.5	1106.7	503.9	1827.4	887.1	1030.9
electro: neg	-23.7	-75.0	-20.1	-55.5	-66.2	-21.4	-61.3	-80.1	-48.0	-50.1

OCT2	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Mean
hydrogen bond donors	211.2	68.8	101.4	82.7	134.0	502.0	183.3
hydrogen bond acceptors	222.3	29.6	47.9	86.2	214.6	39.8	106.7
Sp2 hybridization	622.4	918.7	288.0	667.7	1024.3	1078.9	766.7
lipophilic	345.9	778.4	825.3	738.5	1186.3	362.9	706.2
size(large)	501.3	899.0	881.8	868.0	1402.0	453.1	834.2
charge: pos	23.1	98.1	79.4	128.3	77.0	269.4	112.6
charge: neg	-69.3	0.0	0.0	0.0	0.0	0.0	-11.5
electro: pos	498.7	946.2	992.6	977.6	1347.4	826.5	931.5
electro: neg	-75.3	-7.1	-23.2	-49.6	-64.5	-24.7	-40.7

Figure 3.14: The quantitative APF property measurement table. The table shows the uniqueness of OAT3 pharmacophore model 9 as it was found to have higher value of positive charge and electro-positive charge than the rest of OAT3 pharmacophore models. In addition, the APF property values of this model are found to be comparable with several OCT1 models, such as OCT1 model 5 and 6.

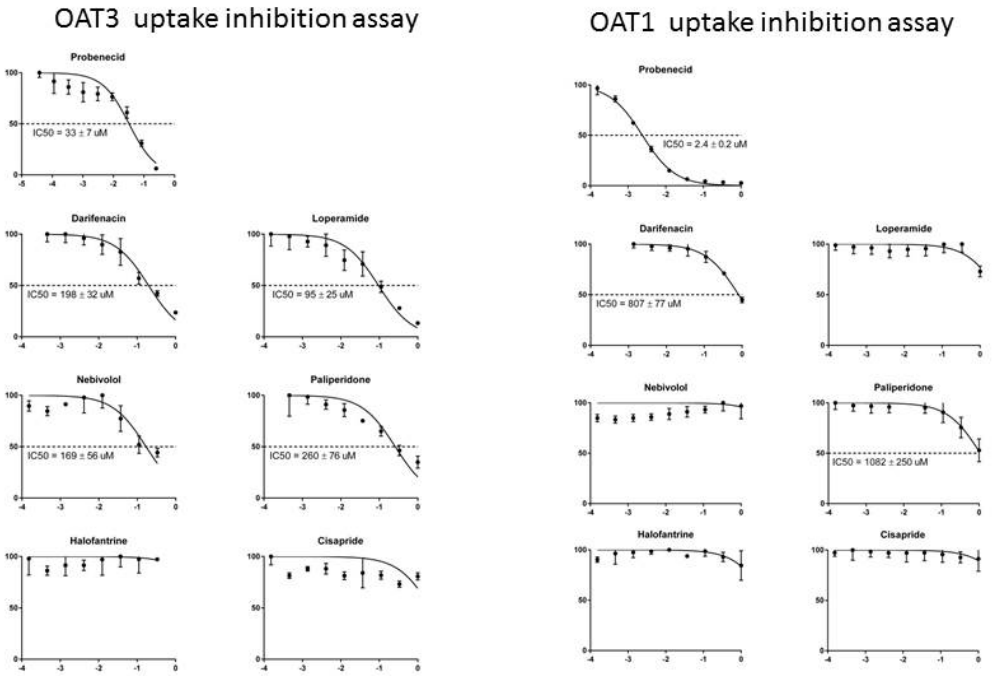


Figure 3.15: Uptake inhibition assay based on the virtual screening of OAT3 cationic pharmacophore against the Drugbank database. For OAT1 inhibition assay, 10 uM 6CF was used as fluorescent tracer, and for OAT3 assay, 20 uM 5CF was used.

Chapter 4 : Building Metabolic Networks for Multiple ABC/SLC Transporters

4.1 Abstract

ABC and SLC transporters are important in both pharmacology and cellular metabolism. To fully understand the endogenous roles these transporters play, individual metabolic networks were built for the 12 ABC/SLC transporters. The networks demonstrate the importance of the transporters, specifically in the context of cellular metabolism. In addition, the merger of these networks demonstrates how these individual transporters, when working together, contribute to the maintenance of homeostasis for the whole body and connection to the pathophysiology of several infamous metabolic disorders/abnormalities, including uremic toxin accumulation, chronic kidney disease, and type-1 diabetes.

4.2 Introduction

ATP binding cassette (ABC) transporters and Solute carrier (SLC) transporters are widely expressed in different epithelia organs and are responsible for the physiological homeostasis. Being either influx or efflux transporters, they are responsible for either uptake or secretion of compounds in or out to the cell. These transporters have received great attention from pharmacology and clinics as they play a huge role in absorption, distribution, metabolism and elimination (ADME) of drugs [80].

Among them, OAT1, OAT3, OATP1B1, OATP1B3, MATE1, MATE2K, MRP2, MRP3, MRP4, BCRP, OCT1, and OCT2 are the main ones responsible of handling of a variety of pharmaceutical drugs, ranging from nonsteroidal anti-inflammatory, antiviral reverse-transcriptase inhibitors, antibiotics, and angiotensin II receptor antagonist. These structurally and functionally related transporters are different in their tissue localizations, expression levels, and substrate specificity [2, 3]. Depending on their locations, they are involved in different processes of drug ADMEs. Expressed in the intestine, BCRP, MDR1, and MRP2 are responsible for the absorption of drugs into bloodstream after drug dosage; OAT3, BCRP, MDR1, MRP2, and MRP4, expressed in blood-brain barrier are involved in the distribution of drugs into the tissues, brains in this case; and the hepatic transporters, OATP1B1 and OATP1B3, and the renal transporters, OAT1, OAT3, and MATE2K, participate in the metabolism and excretion of drugs through the hepatic or renal pathway.

Recognized as drug transporters, these proteins have been mainly studied in the context of pharmacokinetics, and their roles in endogenous metabolism are not the main focus previously. Nonetheless, besides interacting with clinical drugs, they also transport numerous key metabolites, signaling molecules, hormones, and dietary nutrients, into or

out to cells. Their importance in transporting these metabolites is manifested in animal models, in which the knock-outs of some of these transporters, such as OAT1, OAT3, MRP3, alter the concentration of several metabolites in the bloodstream and/or urine [8, 9, 12, 73]. In addition to alteration of metabolite profiles in plasma/urine, the knock-out of OAT3 leads to the increase in blood pressure [73]. All these evidences demonstrate the endogenous roles of the transporters in homeostasis of normal physiology and raise the possibility that the malfunction of them could result in the imbalance of homeostasis and lead to disease state. For instance, the complete knock-out of OATP1B1 and OATP1B3 is found to cause Rotor syndrome, an autosomal recessive bilirubin disorder [81]; the hereditary MRP2 deficiency is associated with Dubin-Johnson syndrome, another autosomal recessive disorder related to hyperbilirubinemia [82]; and patients with diabetic nephropathy are associated with lower gene expression of OAT1 and OAT3 [38].

Even though the transporters have been found to transport individual metabolites in in-vitro assays (molecular levels) and are associated pathophysiology (tissue or organ levels); however, there is no clear connection linking the transporter functionality at molecular levels and the onset of pathophysiology at tissue/organ levels. How the malfunction of transporters leads to the progression of metabolic diseases remains unclear. In order to clarify this relationship, there is a need to study the interaction between the transporters and the metabolites in a systemic way.

To address to this need, in this study, metabolic networks, based on metabolites identified to be associated with, were built for each of these transporters, providing a systemic way to understand how each of transporters is involved with cellular metabolism. In addition, to understand how transporters work together, individual metabolic networks are merged together to a final, larger network, which consists of a total of 451 metabolites. To draw a clear relationship between the function of

transporters and the onset of pathophysiology, individual networks and the merged all-transporter network are overlapped with an uremic-toxin-associated metabolic network, built based on uremic toxins found to be accumulated in renal failures and identified to be cytotoxic to various tissues/organs. The huge overlap between the diseased network and the merged transporter-associated network shows that these transporters, altogether, are responsible for the excretion of these toxins and suggest the function of them is hugely tight to the chronic metabolic diseases. In addition, these transporter-associated networks are also compared two other metabolic-diseased networks, the chronic kidney disease (CKD) network and the type-1-diabetes (T1D) network, and it was found that the individual and the merged transporter-associated networks overlapped well with the two disease models. The percentage of overlapped metabolites of the CKD and T1D models with the merged all-transporter network constituted 64.3% and 73.1%, respectively, of the disease models.

4.3 Materials and Method

4.3.1 Data curation and construction of metabolic networks for individual transporters

A comprehensive literature search was done to identify metabolites found to interact with OAT1, OAT3, OATP1B1, OATP1B3, MATE1, MATE2K, BCRP, MRP2, MRP3, MRP4, OCT1, and OCT2 (Table 4.1). Information regarding to such interaction comes from two types of wet-lab experiments, in-vitro transport assays and in-vivo metabolomics. Based on the list of metabolites, metabolic network for individual transporters were constructed using Metscape, a Cytoscape plugin [46]. Briefly, these metabolites were input into Metscape to generate a network, and in the process of network building, their immediate precursor and products were introduced into the

network as “plus-ones” based on the knowledge of KEGG database. Then, uninformative plus-ones, including small molecules and ions (such as water, carbon dioxide, etc) are removed to generate a final, trimmed network.

After the 12 transporter-associated metabolic networks were built, it was found that some of the networks overlapped greatly. Networks that overlapped greatly were merged together.

Then, all the transporter-associated metabolic networks are merged together to provide us a final, merged network.

4.3.2 Construction of metabolic networks associated with metabolic diseases

An uremic-toxin associated metabolic network was constructed based on the list of metabolites identified as uremic toxin by Eutox [47, 48, 83]. In parallel, a Type-1-diabetes (T1D) associated metabolic network was constructed based on the list of metabolites found to be significantly different (with p-value less than 0.05) between healthy people and T1D patients who had not received any insulin treatment within 8 hours [84], and a chronic kidney disease (CKD) associated metabolic was constructed based on a list of metabolite biomarkers found to be different between CKD patients/animals and the healthy [57]. As above, the networks were further trimmed by removing uninformative plus-ones. Then, this disease network was compared with individual and the merged transporter-associated networks.

4.4 Results

A total of 12 metabolic networks were generated for each of the transporters. Among the networks, OAT1 and OAT3 networks were the largest two. Also, after comparing individual networks, it was observed that, as expected, OATP1B1 and

OATP1B3 networks had great overlap, and OATP1B3 network was found to be a subset of OATP1B1 network, so they were merged into an OATP1Bs network; MATE2K network was found to be a subset of MATE1 network, so they were merged into the MATEs network; MRP2, MRP3, and MRP4 networks share high degree of overlaps, and MRP2 and MRP3 were found to be a subset of MRP4; so they were merged into MRPs network. The top 4 networks in sizes were shown (Figure 4.1).

To study how these transporters function as a whole in cellular metabolism, the 7 transporter-associated networks were merged together to generate a final, merged network. The merged network, all-transporter network, consisted of 451 metabolites connected by 486 edges (Figure 4.2). The network sizes of the individual networks and the final merged network are summarized in Table 4.2.

To understand the connection of the transporters in pathophysiology, several metabolic networks associated with metabolic disorders were built (Figure 4.3) and were compared with individual transporter-associated networks and the all-transporter network. The degree of overlap between the disease network and the individual networks was reflected by the number of metabolites found to be overlapped (Table 4.3), and it was found that when comparing the all-transporter network with the disease models, there were a high degree of overlap. When comparing the all-transporter network with the uremic toxin model, the two networks shared a total of 109 metabolites in common, which represented 67.7 percent of the uremic-toxin metabolic network, showing that the 12 transporters, working together, were responsible for more than two thirds of uremic toxin handling and excretion. The percentage of the overlapped metabolites out of the other two diseased-associated networks, the CKD network and the T1D network, were 64.3% and 73.1%, respectively.

4.5 Discussion

Recognized as drug transporters, ABC and SLC transporters have been widely studied in the context of pharmacology; nonetheless, these transporters also play an important role in cellular metabolism, supported by both in-vitro and in-vivo data [2, 33, 34]. To better understand how ABC and SLC transporters are involved in cellular metabolism and physiology, metabolic networks were constructed for 12 classical transporters. The networks consist of transporter-interacting metabolites, which are the ones known to be direct substrates of the transporters or ones found to be altered when the transporters are malfunctioned, and their immediate precursors and products. Each network delineates the connection between the function of the transporters and the cellular metabolism.

The sizes of these networks vary widely, ranging from ones consisted of more than 150 nodes (OAT1 and OAT3) to ones made by less than 50 nodes (MATEs network). This wide range in network sizes could be explained by the different level of substrate multi-specificity seen in these transporters. For instance, OAT1 and OAT3, with their multi-specificity, would transport a wide array of metabolites. However, this reasoning seemed insufficient to explain the extreme smallness of network sizes for MATEs (MATE1 and MATE2), which were considered to be important in the eliminating process of metabolites and toxins in kidneys. MATE1 and MATE2, being expressed on the apical side of proximal tubule cells of kidneys and having the capability to interact with anions, are considered to pair with OAT1 and OAT3 in function, in which OAT1 and OAT3 first uptake substrates from the bloodstream into the cytoplasm of the cells, and then these substrates are exported by MATE1 and MATE2 into the lumen from the other side of the cells. This function pairing is supported by our ligand-transporter database,

which shows that many drugs found to bind to OAT1 and OAT3 are also found to interact with MATE1 and MATE2. This shows that the OATs and the MATEs have great overlap in substrate pools. Since the OATs are known to interact with many metabolites, we would assume the same for MATEs. However, browsing through the ligand-transporter database, it was found that for MATEs, there were more published data for drugs than for metabolites. The disproportionality between the number of drugs and metabolites studied for MATEs indicates the imbalanced focus when studying the transporters and also hinders the opportunity to understand the full endogenous role of MATE1 and MATE2.

To understand how these transporters function as a group, the individual networks were merged into a final, merged network. This network includes many energy-driving and amino acid metabolic pathways (such as glycolysis, propanoate, and TCA cycle), amino acid metabolism (such as tyrosine, tryptophan, and histidine metabolism), vitamins and cofactors metabolism (such as retinol, thiamin, and riboflavin metabolism), and lipid metabolism (androgen and estrogen biosynthesis, arachidonic acid, and C21-steroid hormone). Lipid metabolism is of interest since many lipid metabolisms are involved in the production of key signaling metabolites, including estradiol, prostaglandin, leukotriene, cGMP, and cAMP. These metabolites are important not only at cellular levels but also at tissue/organ levels as they serve as messengers for cross-talk between tissues/organs. After produced intra-cellularly within one tissue, these metabolites are then transported out by the ABC/SLC efflux transporters into the bloodstream, and after traveling through some distance, they are uptake by other influx transporters into other tissues. The ability to transport these signaling metabolites from one place to another demonstrates that these transporters not only carry transporting function but are also involved in part of whole-body homeostasis.

When comparing the merged all-transporter network with the disease-associated networks, it was found that it overlapped with each of the disease networks hugely. The overlapped metabolites constituted more than 60% of each of the disease networks, and the exact number is the following for the UT model, the CKD model, and the T1D model 67.7%, 64.3% and 73.1%. This shows that the 12 transporters are implicated with all these metabolic diseases. For the example of uremic toxin accumulation, the collective function of these 12 transporters is mainly responsible for the handling and excretion of more than two-third of these toxins. Once one or more transporters are malfunctioned, this could interrupt the overall excretion system for the uremic toxins, and since such toxins could no longer exit the body, they would be accumulated in the bloodstream. Such accumulation would be detrimental to the whole body since each of the toxins could be toxic to different tissues/organs.

Building the metabolic networks for these ABC and SLC transporters are not only valuable in the context of cellular metabolism but are also significant in the clinical pharmacology. Since these transporters also transport a variety of pharmaceutical drugs, drug-metabolite interaction could also occur as drugs and metabolites are competing for the entry by the same transporters. This interaction, especially in the long term, would affect and become harmful to the overall physiology. Thus, having such networks provide clinical values as when physicians provide medications to their patients, they could determine which medicine to be used in order to have the least metabolic impact in patients, especially those who already live with some chronic metabolic diseases.

4.6 Tables

Table 4.1: The main drug transporters.

Transporter	Abbreviation	Gene encoded	Direction	Tissue mainly expressed
Organic anion transporter 1	OAT1	SLC22A6	influx	Kidney
Organic anion transporter 3	OAT3	SLC22A8	influx	Kidney, blood-brain-barrier
Organic anion transporter polypeptide 1B1	OATP1B1	SLCO1B1	influx	Liver
Organic anion transporter polypeptide 1B3	OATP1B3	SLCO1B3	influx	Liver
Multidrug and toxin extrusion protein 1	MATE1	SLC47A1	influx; efflux	Kidney, Liver
Multidrug and toxin extrusion protein 2	MATE2K	SLC47A2	efflux	Kidney, Liver
Breast cancer resistance protein	BCRP	ABCG2	efflux	Universally expressed
Multidrug resistance-associated protein 2	MRP2	ABCC2	efflux	Liver, kidney, gut
Multidrug resistance-associated protein 3	MRP3	ABCC3	efflux	Liver, kidney, gut
Multidrug resistance-associated protein 4	MRP4	ABCC4	efflux	kidney
Organic cation transporter 1	OCT1	SLC22A1	influx	Liver
Organic cation transporter 2	OCT2	SLC22A2	influx	Kidney

Table 4.2: The network size of the individual transporter-associated networks, the merged all-transporter-associated networks, and the disease-associated networks.

Transporter-associated networks	Metabolite	edge
OAT1 network	289	306
OAT3 network	167	147
OATP1Bs network	96	87
MATEs network	44	35
MRPs network	33	28
BCRP network	77	63
OCTs network	83	76
AllTransporters network	451	486
Disease associated network	Metabolite	edge
UT network	161	155
CKD network	322	367
T1D network	134	134

Table 4.3: The overlap between the disease-associated networks and the transporter-associated networks individually and as a whole.

Metabolic networks	Metabolite #	# overlapped with UT model	overlapped % out of UT model	# overlapped with CKD model	overlapped % out of CKD model	# overlapped with T1D model	overlapped % out of T1D model
OAT1 network	289	94	58.4%	145	45.0%	91	67.9%
OAT3 network	167	33	20.5%	67	20.8%	48	35.8%
OATP1Bs network	105	12	7.5%	47	14.6%	17	12.7%
MATEs network	44	13	8.1%	16	5.0%	5	3.7%
MRPs network	81	23	14.3%	30	9.3%	24	17.9%
BCRP network	77	9	5.6%	17	5.3%	10	7.5%
OCTs network	91	16	9.9%	29	9.0%	14	10.4%
AllTransporters network	451	109	67.7%	207	64.3%	98	73.1%
Uremic Toxin network	161						
CKD network	322						
T1D network	134						

4.7 Figures

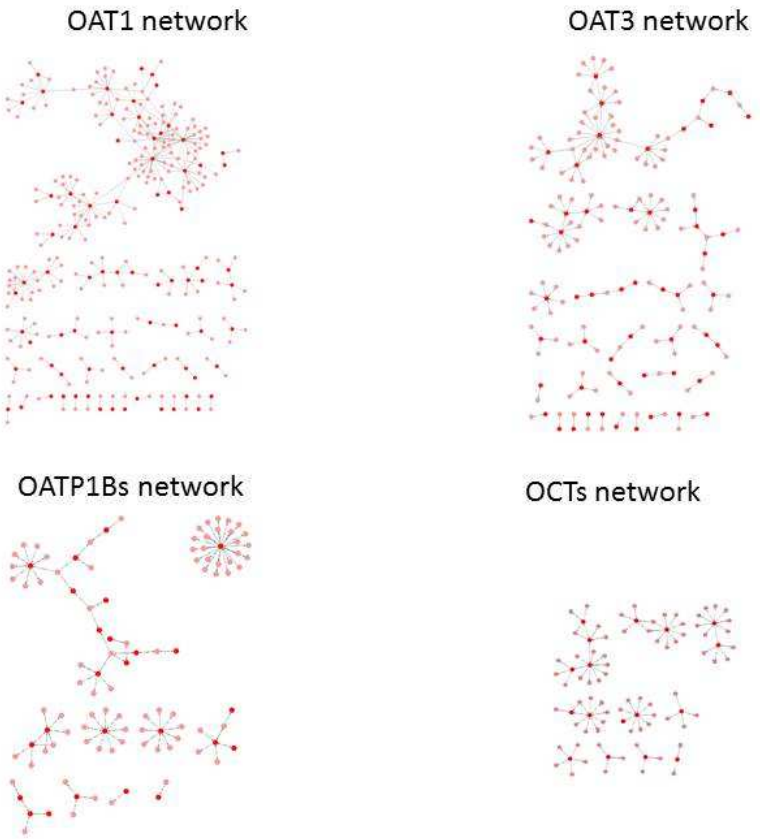


Figure 4.1: The metabolic networks of OAT1, OAT3, OATP1Bs, and OCTs. Metabolites found to be interacting with the transporters are colored red, and the plus-ones are colored pink.

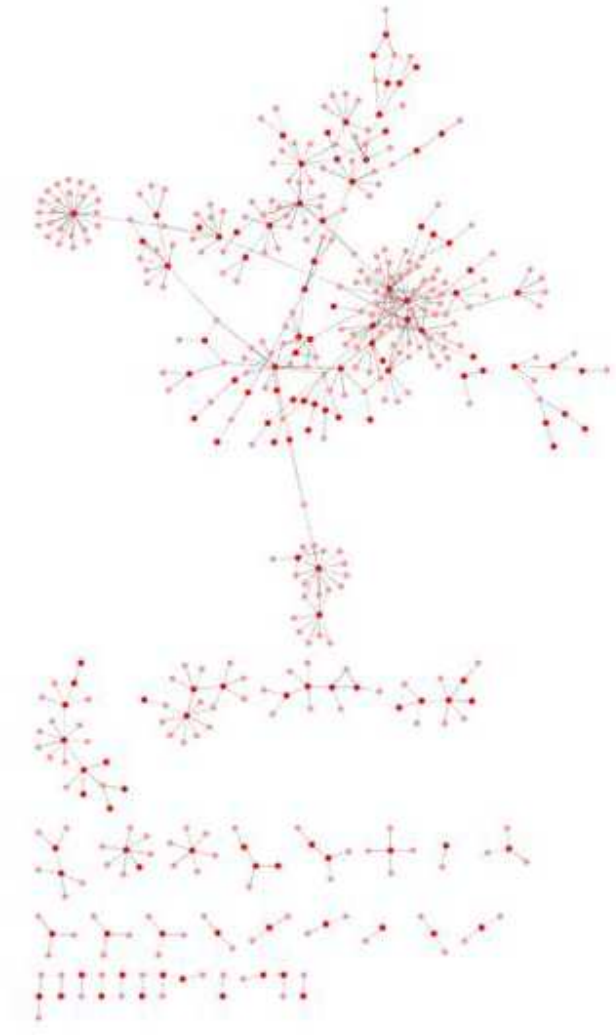


Figure 4.2: The all-transporter-associated metabolic network. The individual transporter-associated metabolic networks are merged together into a final, merged network.

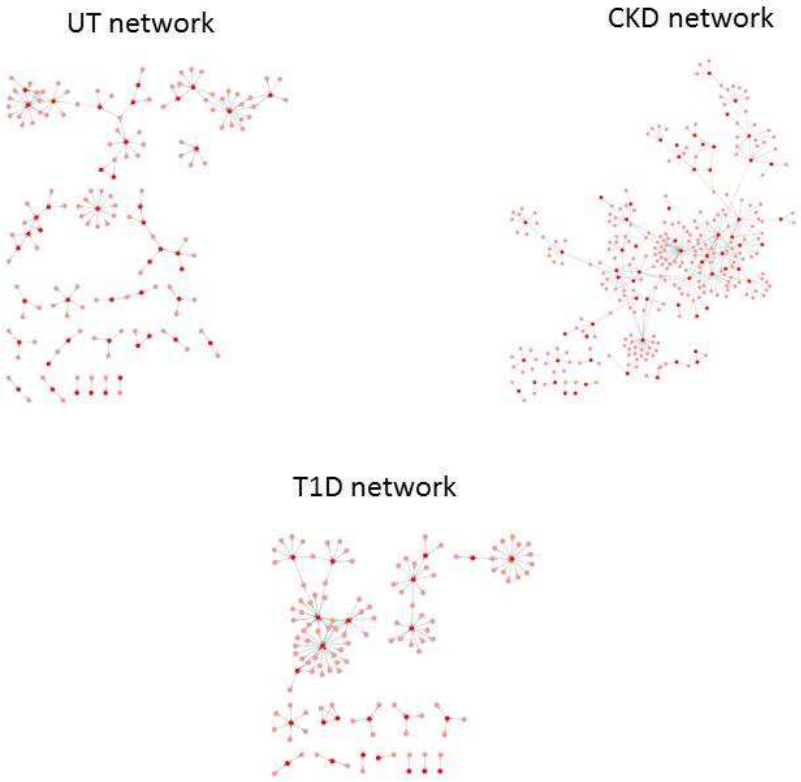


Figure 4.3: The metabolic networks associated with several metabolic diseases. The disease networks include uremic toxin (UT), chronic kidney disease (CKD), and type-1 diabetes (T1D).

Chapter 5 : Conclusion

In this dissertation, a multi-leveled analysis is performed to understand the physiological functions of several key ABC/SLC transporters. These transporters, known as drug transporters, have been widely studied for the pharmaceutical purpose; however, recent studies have identified their importance in endogenous metabolism. To understand the physiological and metabolic roles of these transporters, a multi-leveled approach is applied. First, the study focused on a single transporter. Then, the scope of study extended to a set of 4 transporters, and finally to a set of 12 transporters. The results of this study show that these transporters not only handle a wide array of drugs, but also have important roles as regulators in cellular metabolism and contributors to the whole-body homeostasis.

With the aids from the systems biology analysis and the computational chemistry analysis, the function of OAT1 is individually studied in the context of cellular metabolism. This two-layered computational method not only helped us to predict new OAT1-binding metabolites but also provided us a clear picture of the function of OAT1 in the cellular metabolism. With this novel approach, we have tested the eight top hits, for which four were characterized to be new OAT1-binding metabolites. These four newly identified metabolites are involved in different metabolic pathways and are associated to pathophysiology of several metabolic diseases. With such high success rate, the process of identifying new substrates of transporters becomes easier and more efficient. Instead of blindly testing for compounds from the universal compound database, the computational method filters out compounds unlikely to bind with OAT1 and enriches compounds with higher chance to bind to OAT1. Based on the prediction, researchers

can then focus on these top hits for wet-lab characterization and thus prevent wasting time on testing something that would never work.

In addition, based on this multi-scaled computational method, a construction of an OAT1-centered metabolic network is achievable. Within this network, metabolites known to or predicted to interact with OAT1 are shown and are connected by plus-ones. Such network provides an informative way to understand the link between the function of OAT1 and cellular metabolism. As a major transporter of organic anions in kidney cells, OAT1 is responsible for the influx of many key metabolites in several metabolic pathways. Once the entry of these metabolites is disrupted, cellular metabolism could be affected, causing the whole-body physiology to be imbalanced. The OAT1-centered network provides us a way to understand how the disruption of OAT1 function could possibly lead to such metabolic alteration. In this network, metabolites that are directly transported by OAT1 are shown along with their immediate first neighbors in metabolic reactions. One can understand no only metabolites that are direct substrates of OAT1, but also metabolites that are indirectly affected by OAT1 (through metabolic cascading effects). In addition, since metabolites within this network can be grouped according to which metabolic pathways they are involved in, one can also understand the extent of OAT1 involvement in each metabolic pathway.

Based on this network, the function of OAT1 was found to be hugely associated with cellular metabolism, ranging from tyrosine metabolism, tryptophan metabolism, the TCA cycle, purine and pyrimidine metabolism, to urea cycle. Among these highly represented pathways, the TCA cycle is worthy of further discussion as it is an important energy-driving source for metabolism; several metabolites of the TCA cycle are known to be classical OAT1 substrates, including succinate, alpha-ketoglutarate, and pyruvate, which shows the function of the transporter is tightly linked to energy-driving process in

metabolism. Another pathway significantly represented in the OAT1-centered network is tryptophan metabolism; 19 metabolites present in this network are part of tryptophan metabolism, such as indoxyl sulfate, kynuerenate, and xanthurenate. Many of these metabolites are derived from toxins or by-products secreted by gut enterobiomes [9]. The capability of OAT1 to interact with these symbiotic metabolites demonstrates the role of OAT1 in remote sensing, which is a concept of communication between two different species based on the exchanged metabolites. In this case, OAT1 does not simply exchange metabolites but also relay important information between the host and the gut enterobiomes; the exchanged metabolites act as encrypted information sent from the host that can be perceived by the gut enterobiomes, and vice versa. A third pathway in which OAT1 is heavily involved is urea cycle, which is a metabolic process that converts ammonia to urea, a less toxic one. The urea cycle occurs only in liver and kidney. Also, OAT1 is shown to transport several key factors in this metabolic pathway. Thus, the occurrence of this metabolic process hinges on the transporting function of OAT1. In summary, we not only developed a novel, multi-scaled method to study OAT1, we also have identified the biological importance of OAT1 in the context of cellular metabolism. Moreover, this novel, multi-scaled method could be and should be applied to studying all other ABC/SLC transporters, which could help us understand the metabolic roles and physiological significance of all these transporters.

In addition to OAT1, other SLC transporters known to handle organic compounds are OAT3, OCT1, and OCT2. As described by their names, OAT1 and OAT3 are known to interact with organic anions, and in contrast, OCT1 and OCT2 are known to interact with organic cations. Therefore, within the two OATs and the two OCTs share more overlapping substrates. Studying the substrate overlaps and substrate preference among the transporters is an important step to help us understand how these

transporters work in a systemic fashion. With the tools from computational chemistry and machine learning, a systemic ligand-based approach was used to understand the substrate overlap and substrate preference between transporters. All drugs known to interact with any of the four transporters were analyzed using the machine learning tools. With this systemic approach, a comprehensive understanding of the substrate preference of the transporters is possible as each compound should contribute to providing some information regarding to the binding of the substrate and the transporter.

Several machine learning methods were applied to explore such binding phenomenon. Among them, the decision tree analysis was used to differentiate drugs binding from the two transporters, which characterize the substrate binding preference between the two transporters. Decision trees have been widely used in operations research for decision making, and here decision trees are used in an attempt to answer the biological question.

Using the machine learning approach, differences in functions of these transporters are unveiled. According to what their names suggest, OATs and OCTs are fundamentally different in the substrates they interact with. OAT-binding drugs are anionic whereas OCT-binding drugs are cationic. Besides the difference in charge state, another property of ligands is found to be an important factor in substrate recognition by transporters. This important property is SP3 character of ligands, which refers to the degree of three-dimensionality of a compound. It is found that OAT-binding drugs are more planar, and OCT-binding drugs are more three-dimensional/stereoscopic. This interesting finding suggests that the binding pockets of OATs and OCTs are structurally different, in which the binding pockets of OATs are perhaps flat whereas those of OCTs are perhaps more three-dimensional. This is worthy of further investigation as it can clarify our understanding in how the transporters differentiate substrates from non-

substrates. Besides identifying the major differences between OATs and OCTs, some differences are also observed among OAT1 and OAT3. OAT3-binding drugs are found to be more cationic and electro-positive than OAT1-binding drugs based on the machine learning analyses. The pharmacophore models built based on OAT1 drugs and based on OAT3 drugs also support this notion.

The pharmacophore models built for the OAT and OCT transporters show a striking result: One of the OAT3 pharmacophore models is structurally similar to multiple OCTs pharmacophore models. The “cationic” binding pocket of OAT3 suggests that OAT3 retains several OCT-like characteristics, such as the capability to interact with cations. Thus, unlike OAT1, the capability of OAT3 to interact with both cations and anions makes the transporter functionally unique and physiologically important. Perhaps the diverse substrate pool of OAT3 is associated with the universal expression of the transporter. OAT3 is found to be expressed in various tissues, including kidney, brain endothelium choroid plexus, retina, and testes [3]. Due to this, OAT3 is involved in dealing with different types of biological fluids, yet it has to able to “recognize” its substrates. In the brain, OAT3 encounters more cationic substrates since quite many neurotransmitters and psychoactive drugs are cations. The causality relationship for the development of this unique characteristic of OAT3 remains unclear: either OAT3 adapted this OCT-like characteristic because it is widely expressed across different tissues, or the expression in various tissues of OAT3 was due to its unique characteristic that is in demand to deal with “extreme” environment. This interesting question is worth of further investigation as it can tell us how the “transporter systems” and “organism” evolve together in a symbiotic relationship.

Understanding the substrate preference of transporters can help us design potent inhibitors that will selectively bind to a transporter of interest and thus selectively

inhibit the function of that particular transporter. This strategy becomes more and more popular in current pharmacology since ABC/SLC transporters play a huge role in absorption, distribution, metabolism, and excretion (ADMEs) of many pharmaceutical drugs. By inhibiting the transport functions of these transporters, the ADMEs of the drugs can be manipulated in order to maximize the effect of drugs locally and temporally.

Substrate preference of transporters lead to the notion that transporters will have slightly or drastically different substrate pool they interact with. Besides interacting with drugs and other exogenous compounds, these transporters are also responsible for the transport of endogenous metabolites, as shown in the OAT1-centered metabolic network which is greatly composed of endogenous metabolites. Thus, in an attempt to understand how the ABC/SLC transporters are involved in cellular metabolism, metabolic networks are constructed for the 12 major ABC/SLC transporters. Such networks are built based on the metabolites known to interact with the transporters of study; the resulting networks demonstrate the extents of these transporters are involved in different aspect of cellular metabolism.

Examining these metabolic networks confirms what we know about substrates overlapping and substrate specificity between transporters. OAT1, OAT3, OATP1B1, and OATP1B3 are known to transport similar substrate pools, mostly organic anions, and their metabolic networks are found to have a high degree of overlap. Amongst these four, OAT1 and OAT3, and OAT1PB1 and OATP1B3, have more overlap. Again, what kind of substrates they interact with can also be explained by the organ of expression. OAT1 and OAT3, mainly expressed in the kidney, interact with urine-related metabolites; and OATP1B1, and OATP1B3, mainly expressed in the liver, interact with bile-acid-related metabolites. OCT1- and OCT2-associated networks are found to be consisting of more cationic metabolites as expected.

The individual 12 metabolic networks are merged together to generate a final, merged network. This merged network consists of more than 400 metabolites, which are involved in more than 30 different metabolic pathways. This demonstrates how these ABC/SLC transporters, each with its own substrate pool, work collectively to contribute maintaining the homeostasis of the body. Even though each transporter can only transport a particular set of metabolites, when joined together, they are responsible of transport of all metabolites needed to maintain the ongoing of cellular metabolism. When the transport function of a single transporter is disrupted, either due to genetic disorder or chemical knockout, its function can be compensated by other transporters that handle a similar pool of substrate in the short term. However, some metabolic changes are expected to occur in the long run. Unfortunately, any impaired transporter function is harmful to the whole-body physiology and probably are the causes for serious metabolic disorders, such as diabetes and chronic kidney diseases.

In conclusion, the author has attempted to use a wide variety of tools to study the transporting functions, structural difference, and physiological significance of the major ABC/SLC transporters. Some of these tools have not been used in the biomedical researches, and some of the tools are combined together to create a novel and multi-scaled approach to study transporters. The novel methods prove to be advantageous as it helps us to answer biological questions that could not be answered using the previous methods.

References

1. Huang, S.M., L. Zhang, and K.M. Giacomini, *The International Transporter Consortium: a collaborative group of scientists from academia, industry, and the FDA.* Clin Pharmacol Ther, 2010. 87(1): p. 32-6.

2. Nigam, S.K., *What do drug transporters really do?* Nat Rev Drug Discov, 2015. 14(1): p. 29-44.

3. Nigam, S.K., et al., *The Organic Anion Transporter (OAT) Family: A Systems Biology Perspective.* Physiol Rev, 2015. 95(1): p. 83-123.

4. Krupoves, A., et al., *Associations between ABCB1/MDR1 gene polymorphisms and Crohn's disease: a gene-wide study in a pediatric population.* Inflamm Bowel Dis, 2009. 15(6): p. 900-8.

5. Kolz, M., et al., *Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations.* PLoS Genet, 2009. 5(6): p. e1000504.

6. Huang, M.J., et al., *Risk factors for severe hyperbilirubinemia in neonates.* Pediatr Res, 2004. 56(5): p. 682-9.

7. Kanda, D., et al., *Novel large-scale deletion (whole exon 7) in the ABCC2 gene in a patient with the Dubin-Johnson syndrome.* Drug Metab Pharmacokinet, 2009. 24(5): p. 464-8.

8. Eraly, S.A., et al., *Decreased renal organic anion secretion and plasma accumulation of endogenous organic anions in OAT1 knock-out mice.* J Biol Chem, 2006. 281(8): p. 5072-83.

9. Wikoff, W.R., et al., *Untargeted metabolomics identifies enterobiome metabolites and putative uremic toxins as substrates of organic anion transporter 1 (Oat1).* J Proteome Res, 2011. 10(6): p. 2842-51.

10. Torres, A.M., et al., *Deletion of multispecific organic anion transporter Oat1/Slc22a6 protects against mercury-induced kidney injury.* J Biol Chem, 2011. 286(30): p. 26391-5.

11. Wu, W., et al., *Multispecific drug transporter Slc22a8 (Oat3) regulates multiple metabolic and signaling pathways.* Drug Metab Dispos, 2013. 41(10): p. 1825-34.

12. van de Wetering, K., et al., *Targeted metabolomics identifies glucuronides of dietary phytoestrogens as a major class of MRP3 substrates in vivo.* Gastroenterology, 2009. 137(5): p. 1725-35.

13. Consortium, C.e.S., *Genome sequence of the nematode C. elegans: a platform for investigating biology.* Science, 1998. 282(5396): p. 2012-8.

14. Duarte, N.C., et al., *Global reconstruction of the human metabolic network based on genomic and bibliomic data.* Proc Natl Acad Sci U S A, 2007. 104(6): p. 1777-82.

15. Sheikh, K., J. Forster, and L.K. Nielsen, *Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of Mus musculus.* Biotechnol Prog, 2005. 21(1): p. 112-21.

16. Ahn, S.Y., et al., *Linkage of organic anion transporter-1 to metabolic pathways through integrated "omics"-driven network and functional analysis.* J Biol Chem, 2011. 286(36): p. 31522-31.

17. Perry, J.L., et al., *A three-dimensional model of human organic anion transporter 1: aromatic amino acids required for substrate transport.* J Biol Chem, 2006. 281(49): p. 38071-9.

18. Zhang, X., et al., *A conserved glutamate residue in transmembrane helix 10 influences substrate specificity of rabbit OCT2 (SLC22A2).* J Biol Chem, 2005. 280(41): p. 34813-22.

19. Popp, C., et al., *Amino acids critical for substrate affinity of rat organic cation transporter 1 line the substrate binding region in a model derived from the tertiary structure of lactose permease.* Mol Pharmacol, 2005. 67(5): p. 1600-11.

20. Tsigelny, I.F., et al., *Conformational changes of the multispecific transporter organic anion transporter 1 (OAT1/SLC22A6) suggests a molecular mechanism for initial stages of drug and metabolite transport.* Cell Biochem Biophys, 2011. 61(2): p. 251-9.

21. Ahn, S.Y., et al., *Interaction of organic cations with organic anion transporters.* J Biol Chem, 2009. 284(45): p. 31422-30.

22. Truong, D.M., et al., *Multi-level analysis of organic anion transporters 1, 3, and 6 reveals major differences in structural determinants of antiviral discrimination.* J Biol Chem, 2008. 283(13): p. 8654-63.

23. Kouznetsova, V.L., et al., *Elucidation of common pharmacophores from analysis of targeted metabolites transported by the multispecific drug transporter-Organic anion transporter1 (Oat1).* Bioorg Med Chem, 2011. 19(11): p. 3320-40.

24. Duan, P., et al., *Potent inhibitors of human organic anion transporters 1 and 3 from clinical drug libraries: discovery and molecular characterization.* Mol Pharm, 2012. 9(11): p. 3340-6.

25. 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, 2007. 282(33): p. 23841-53.

26. McDermott, A.M., et al., *Identification and validation of oncologic miRNA biomarkers for luminal A-like breast cancer.* PLoS One, 2014. 9(1): p. e87032.

27. Hazai, E., et al., *Predicting substrates of the human breast cancer resistance protein using a support vector machine method.* BMC Bioinformatics, 2013. 14: p. 130.

28. Wang, Z., et al., *P-glycoprotein substrate models using support vector machines based on a comprehensive data set.* J Chem Inf Model, 2011. 51(6): p. 1447-56.

29. Ahn, S.Y. and S.K. Nigam, *Toward a systems level understanding of organic anion and other multispecific drug transporters: a remote sensing and signaling hypothesis.* Mol Pharmacol, 2009. 76(3): p. 481-90.

30. Wu, W., A.V. Dnyanmote, and S.K. Nigam, *Remote communication through solute carriers and ATP binding cassette drug transporter pathways: an update on the remote sensing and signaling hypothesis.* Mol Pharmacol, 2011. 79(5): p. 795-805.

31. Lopez-Nieto, C.E., et al., *Molecular cloning and characterization of NKT, a gene product related to the organic cation transporter family that is almost exclusively expressed in the kidney.* J Biol Chem, 1997. 272(10): p. 6471-8.

32. Lopez-Nieto, C.E. and S.K. Nigam, *Selective amplification of protein-coding regions of large sets of genes using statistically designed primer sets.* Nat Biotechnol, 1996. 14(7): p. 857-61.

33. VanWert, A.L., M.R. Gionfriddo, and D.H. Sweet, *Organic anion transporters: discovery, pharmacology, regulation and roles in pathophysiology.* Biopharm Drug Dispos, 2010. 31(1): p. 1-71.

34. Burckhardt, G. and B.C. Burckhardt, *In vitro and in vivo evidence of the importance of organic anion transporters (OATs) in drug therapy.* Handb Exp Pharmacol, 2011(201): p. 29-104.

35. Vallon, V., et al., *Overlapping in vitro and in vivo specificities of the organic anion transporters OAT1 and OAT3 for loop and thiazide diuretics.* Am J Physiol Renal Physiol, 2008. 294(4): p. F867-73.

36. Nagle, M.A., et al., *Analysis of three-dimensional systems for developing and mature kidneys clarifies the role of OAT1 and OAT3 in antiviral handling.* J Biol Chem, 2011. 286(1): p. 243-51.

37. Nagle, M.A., et al., *Organic anion transport pathways in antiviral handling in choroid plexus in Oat1 (Slc22a6) and Oat3 (Slc22a8) deficient tissue.* Neurosci Lett, 2013. 534: p. 133-8.

38. Sharma, K., et al., *Metabolomics reveals signature of mitochondrial dysfunction in diabetic kidney disease.* J Am Soc Nephrol, 2013. 24(11): p. 1901-12.

39. Prentice, K.J., et al., *The furan fatty acid metabolite CMPF is elevated in diabetes and induces beta cell dysfunction.* Cell Metab, 2014. 19(4): p. 653-66.

40. Sigurdsson, M.I., et al., *A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1.* BMC Syst Biol, 2010. 4: p. 140.

41. Becker, S.A. and B.O. Palsson, *Context-specific metabolic networks are consistent with experiments.* PLoS Comput Biol, 2008. 4(5): p. e1000082.

42. Forster, J., et al., *Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network.* Genome Res, 2003. 13(2): p. 244-53.

43. Bordbar, A., et al., *A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology.* BMC Syst Biol, 2011. 5: p. 180.

44. Mahadevan, R. and C.H. Schilling, *The effects of alternate optimal solutions in constraint-based genome-scale metabolic models.* Metab Eng, 2003. 5(4): p. 264-76.

45. Totrov, M. and R. Abagyan, *Flexible ligand docking to multiple receptor conformations: a practical alternative.* Curr Opin Struct Biol, 2008. 18(2): p. 178-84.

46. Karnovsky, A., et al., *Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data.* Bioinformatics, 2012. 28(3): p. 373-80.

47. Vanholder, R., et al., *Review on uremic toxins: classification, concentration, and interindividual variability.* Kidney Int, 2003. 63(5): p. 1934-43.

48. Duranton, F., et al., *Normal and pathologic concentrations of uremic toxins.* J Am Soc Nephrol, 2012. 23(7): p. 1258-70.

49. Lacroix, V., et al., *An introduction to metabolic networks and their structural analysis.* IEEE/ACM Trans Comput Biol Bioinform, 2008. 5(4): p. 594-617.

50. Barabasi, A.L. and Z.N. Oltvai, *Network biology: understanding the cell's functional organization.* Nat Rev Genet, 2004. 5(2): p. 101-13.

51. Ching-Ha Kwan, B. and S. Beddhu, *Metabolic syndrome and chronic kidney disease.* Minerva Urol Nefrol, 2006. 58(1): p. 1-12.

52. Guarnieri, G., et al., *Metabolic syndrome and chronic kidney disease.* J Ren Nutr, 2010. 20(5 Suppl): p. S19-23.

53. Monica Torres, A., et al., *Altered renal elimination of organic anions in rats with chronic renal failure.* Biochim Biophys Acta, 2005. 1740(1): p. 29-37.

54. Liu, T., et al., *Changes in expression of renal Oat1, Oat3 and Mrp2 in cisplatin-induced acute renal failure after treatment of JBP485 in rats.* Toxicol Appl Pharmacol, 2012. 264(3): p. 423-30.

55. Komazawa, H., et al., *Renal uptake of substrates for organic anion transporters Oat1 and Oat3 and organic cation transporters Oct1 and Oct2 is altered in rats with adenine-induced chronic renal failure.* J Pharm Sci, 2013. 102(3): p. 1086-94.

56. Saito, H., *Pathophysiological regulation of renal SLC22A organic ion transporters in acute kidney injury: pharmacological and toxicological implications.* Pharmacol Ther, 2010. 125(1): p. 79-91.

57. Zhao, Y.Y., *Metabolomics in chronic kidney disease.* Clin Chim Acta, 2013. 422: p. 59-69.

58. Deguchi, T., et al., *Characterization of uremic toxin transport by organic anion transporters in the kidney.* Kidney Int, 2004. 65(1): p. 162-74.

59. Weiss, T., et al., *Endothelial function in patients with peripheral vascular disease: influence of prostaglandin E1.* Prostaglandins Leukot Essent Fatty Acids, 2002. 67(5): p. 277-81.

60. Hodson, L. and F. Karpe, *Is there something special about palmitoleate?* Curr Opin Clin Nutr Metab Care, 2013. 16(2): p. 225-31.

61. Schuetz, J.D., P.W. Swaan, and D.J. Tweedie, *The role of transporters in toxicity and disease.* Drug Metab Dispos, 2014. 42(4): p. 541-5.

62. Sinxadi, P.Z., et al., *Mitochondrial genomics and antiretroviral therapy-associated metabolic complications in HIV-infected Black South Africans: a pilot study.* AIDS Res Hum Retroviruses, 2013. 29(7): p. 1031-9.

63. Monier, P.L. and R. Wilcox, *Metabolic complications associated with the use of highly active antiretroviral therapy in HIV-1-infected adults.* Am J Med Sci, 2004. 328(1): p. 48-56.

64. Anderson, P.L., T.N. Kakuda, and K.A. Lichtenstein, *The cellular pharmacology of nucleoside- and nucleotide-analogue reverse-transcriptase inhibitors and its relationship to clinical toxicities.* Clin Infect Dis, 2004. 38(5): p. 743-53.

65. Lewis, W., *Mitochondrial DNA replication, nucleoside reverse-transcriptase inhibitors, and AIDS cardiomyopathy.* Prog Cardiovasc Dis, 2003. 45(4): p. 305-18.

66. Sweeney, D.E., et al., *Functional maturation of drug transporters in the developing, neonatal, and postnatal kidney.* Mol Pharmacol, 2011. 80(1): p. 147-54.

67. Chang, R.L., et al., *Drug off-target effects predicted using structural analysis in the context of a metabolic network model.* PLoS Comput Biol, 2010. 6(9): p. e1000938.

68. Thiele, I., et al., *A community-driven global reconstruction of human metabolism.* Nat Biotechnol, 2013. 31(5): p. 419-25.

69. Doncheva, N.T., et al., *Topological analysis and interactive visualization of biological networks and protein structures.* Nat Protoc, 2012. 7(4): p. 670-85.

70. Newman, M.E.J., *Networks : an introduction*. 2010, Oxford ; New York: Oxford University Press. xi, 772 p.

71. Koepsell, H., *The SLC22 family with transporters of organic cations, anions and zwitterions.* Mol Aspects Med, 2013. 34(2-3): p. 413-35.

72. Imamura, Y., et al., *Prediction of fluoroquinolone-induced elevation in serum creatinine levels: a case of drug-endogenous substance interaction involving the inhibition of renal secretion.* Clin Pharmacol Ther, 2011. 89(1): p. 81-8.

73. Vallon, V., et al., *A role for the organic anion transporter OAT3 in renal creatinine secretion in mice.* Am J Physiol Renal Physiol, 2012. 302(10): p. F1293-9.

74. Berthold, M., et al., *KNIME - The Konstanz information miner Version 2.0 and Beyond*. 2009, Konstanz: Bibliothek der Universität Konstanz. Online-Ressource.

75. Agresti, A. and B.A. Coull, *Order-restricted tests for stratified comparisons of binomial proportions.* Biometrics, 1996. 52(3): p. 1103-11.

76. Svetnik, V., et al., *Random forest: a classification and regression tool for compound classification and QSAR modeling.* J Chem Inf Comput Sci, 2003. 43(6): p. 1947-58.

77. Totrov, M., *Atomic property fields: generalized 3D pharmacophoric potential for automated ligand superposition, pharmacophore elucidation and 3D QSAR.* Chem Biol Drug Des, 2008. 71(1): p. 15-27.

78. Lovering, F., J. Bikker, and C. Humblet, *Escape from flatland: increasing saturation as an approach to improving clinical success.* J Med Chem, 2009. 52(21): p. 6752-6.

79. Over, B., et al., *Natural-product-derived fragments for fragment-based ligand discovery.* Nat Chem, 2013. 5(1): p. 21-8.

80. Giacomini, K.M., et al., *International Transporter Consortium commentary on clinically important transporter polymorphisms.* Clin Pharmacol Ther, 2013. 94(1): p. 23-6.

81. van de Steeg, E., et al., *Complete OATP1B1 and OATP1B3 deficiency causes human Rotor syndrome by interrupting conjugated bilirubin reuptake into the liver.* J Clin Invest, 2012. 122(2): p. 519-28.

82. Uchiumi, T., et al., *Mutation and functional analysis of ABCC2/multidrug resistance protein 2 in a Japanese patient with Dubin-Johnson syndrome.* Hepatol Res, 2013. 43(5): p. 569-75.

83. Meert, N., et al., *Inconsistency of reported uremic toxin concentrations.* Artif Organs, 2007. 31(8): p. 600-11.

84. Dutta, T., et al., *Concordance of changes in metabolic pathways based on plasma metabolomics and skeletal muscle transcriptomics in type 1 diabetes.* Diabetes, 2012. 61(5): p. 1004-16.