

COMPUTATIONAL PREDICTION
OF
MOLECULAR INTERACTIONS

by

Natalia Khuri

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION

of the

Copyright 2014

by

Natalia Khuri

*To my husband Sami Khuri,
for his unwavering support and encouragement over the years.*

Acknowledgements

Over the past four and a half years I have received support and encouragement from a great number of individuals. First and foremost, I would like to express my gratitude to my advisor, Andrej Sali, for the continuous support of my Ph.D studies, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in research and writing of this dissertation and his mentorship has made this a thoughtful and rewarding journey. Andrej has provided invaluable advice and resources for me to succeed and has given me the freedom to pursue various projects.

My time at UCSF would not have been the same without Kathy Giacomini. I thank her for accepting me for a rotation in her lab and supporting my collaborations with several of her students and postdocs. She is an incredible person, a great scientist, and a wonderful mentor and I look forward to my future collaboration with her in the UCSF-Stanford CERSI.

When I was accepted into the graduate program at UCSF, I planned to continue research in genomics and metagenomics. These plans changed radically after I took the first course in “Macromolecular Interactions I” by Robert Stroud. His passion for the subject was so contagious that I decided to take the second course on that topic and eventually joined a structural biology lab. I will never forget writing the mini-oral proposal and working with Dr. Stroud until midnight to come up with the best title for it. I would like to thank him for giving me the opportunity to rotate in his lab, for always welcoming me in his office whenever I show up unannounced, for serving on my qualifying exam and thesis committees, and for his insightful comments and hard questions. I would also like to thank my dissertation committee member

Tanja Kortemme for her support and advice. Her lecture during my recruitment visit and in our bioinformatics class contributed to my realization that proteins are indeed beautiful.

I would also like to thank several UCSF faculty members for being there for me when I needed them. Robert Fletcher for serving on my qualifying exam committee and for listening to the stories of my successes and failures. I am forever grateful for his encouragement to keep exploring new things, learn about unfamiliar topics, and think like a scientist. I would like to acknowledge Patsy Babbitt, who was my first year advisor guiding me through rotations and courses and who also considered me for a position at UCSF. I would also like to thank Katie Pollard for her advice and support before my acceptance into the program and during my years at UCSF. A big thank you goes to Charly Craik for giving me opportunities to collaborate with several members of his lab and encouragement along the way.

I would like to thank my collaborators: Avner Schlessinger, Matthias Wittwer, Arik Zur, Sook Wah Yee, Hao Fan, Massimiliano Bonomi, Eugene Chen, Michael Winters, Anthony O'Donoghue, and Dina Schneidman-Duhovny. A big kudos goes to all the members of the Sali, Stroud, Craik, and Giacomini labs for making me feel at home and helping me succeed. In addition, I really appreciated tremendous support and efficient manner of Hilary Mahon, Sali lab administrative assistant. Without her assistance, my scheduling and reimbursement would have been orders of magnitude slower. She also became a good friend over the years. I would also like to acknowledge support (administrative and moral) from Suzan Bethel, assistant to R. Stroud and Joslyn Polzien, assistant to T. Kortemme.

I would not have been where I am now without the support from Sami Khuri, John Avila, Devaki Bhaya, and Frederick Stern. Finally, I would like to thank all my extended family in Russia, USA, France, and Lebanon.

COMPUTATIONAL PREDICTION OF MOLECULAR INTERACTIONS

Abstract

Molecular interactions are of critical importance to all biological processes such as enzymatic reactions, transport, signaling, protein folding and aggregation, and macromolecular assembly. Proteins evolved to interact with other cellular and extracellular molecules in a precise and time-dependent manner. Structural analyses and modeling of such interactions is a critical step for a mechanistic description of their functions. The research in this dissertation focused on developing and applying integrative computational modeling approaches to understand the structural basis of molecular recognition and interactions, and predict modulators of these interactions. To achieve these goals, several techniques were used, such as computer simulations, physicochemical modeling, as well as graph-theoretic and machine learning techniques. More specifically, computational methods and tools were developed and applied to (1) identify ligands of clinically relevant human membrane transporters expressed in liver, kidney and intestine, (2) study molecular interactions between hemoglobin proteins in red blood cells, and (3) identify human proteins attacked by the HIV-1 protease during infection.

Table of Contents

Acknowledgements	iv
COMPUTATIONAL PREDICTION OF MOLECULAR INTERACTIONS	VII
Abstract	vii
Table of Contents	viii
List of Tables	xv
List of Figures	xvi
Contributions	1
Chapter 1: Introduction	3
IDENTIFICATION OF POTENT, CLINICALLY SIGNIFICANT INHIBITORS OF HUMAN SLC TRANSPORTERS	3
MODELING DYNAMIC PROTEIN-PROTEIN INTERACTIONS IN HIV-1 INFECTION	7
MAPPING DYNAMIC ASSEMBLY / DISASSEMBLY PATHWAYS OF MACROMOLECULAR COMPLEXES	8
REFEENCES	9
Chapter 2: Molecular modeling and ligand docking for Solute Carrier (SLC) transporters	13
ABSTRACT	13
SOLUTE CARRIER TRANSPORTERS	14
SEQUENCES, STRUCTURES, AND MECHANISMS OF SLC TRANSPORTERS	15
COMPARISON AND CLUSTERING OF SLC TRANSPORTERS	18
COMPARATIVE MODELING OF SLC TRANSPORTER STRUCTURES	19
STRUCTURE-BASED DISCOVERY OF SLC TRANSPORTER LIGANDS	21

CASE STUDY 1: THE NOREPINEPHRINE TRANSPORTER (NET, SLC6A2)	22
CASE STUDY 2: THE GABA TRANSPORTER 2 (GAT-2, SLC6A13) ..	24
CASE STUDY 3: THE SLC6 FAMILY	26
CASE STUDY 4: THE LARGE-NEUTRAL AMINO ACID TRANSPORTER (LAT-1, SLC7A5)	27
CASE STUDY 5: THE MULTIDRUG AND TOXIN EXTRUSION TRANSPORTER (MATE, SLC47) FAMILY	30
CONCLUSIONS AND FUTURE OUTLOOK	33
FIGURES	36
REFERENCES	43

Chapter 3: High Selectivity of the γ -Aminobutyric Acid Transporter 2

(GAT-2, SLC6A13) Revealed by Structure-based Approach	53
SUMMARY	53
INTRODUCTION	54
EXPERIMENTAL PROCEDURES	56
Comparative Model Construction	56
Ligand Docking and Virtual Screening	57
Binding Site Assessment	58
Data Sets for Virtual Screening	58
Chemical Novelty Evaluation	59
Subcloning and Transient Transfection of GAT-2 into HEK Cells	59
Uptake Experiments	61
Site-directed Mutagenesis	61
Construction and Visualization of Chemical Similarity Networks	62
RESULTS	62
GAT-2 Models and Their Assessment	62
Mode of GAT-2 Interaction with GABA	63
Model Validation Using Site-directed Mutagenesis	64
Virtual Screening of Small Molecule Libraries against the GAT-2	

Models.....	65
Chemical Similarity of the Predicted Ligands	66
Rationale for Selecting Molecules	66
Experimental Validation of the Top Hits.....	67
Comparison between the GAT-2 and the NET Models.....	69
DISCUSSION	70
Distinct Ligands Are Identified Using Two Different GAT-2 Conformations.....	72
Structure-based Ligand Discovery for Membrane Transporters	73
Physiological and Pharmacological Implications of GAT-2 Inhibitors	74
Toward a Description of Substrate Specificity in the SLC6 Family ...	75
FIGURES	78
TABLES	86
REFERENCES	87
SUPPLEMENTAL INFORMATION	91
SI FIGURE LEGENDS	94
REFERENCES	97

Chapter 4: Discovery of Potent, Selective Multidrug And Toxin Extrusion

Transporter 1 (MATE1, SLC47A1) Inhibitors Through Prescription Drug Profiling and Computational Modeling	99
ABSTRACT.....	99
INTRODUCTION	100
RESULTS	103
High Throughput Screen for MATE1-Inhibitors with ASP+ as Fluorescent Probe.....	103
Physicochemical Properties of MATE1 HTS Hits and Comparison to OCT2 HTS Hits	103
Validation of HTS Screen by Follow-up IC ₅₀ Determination	106
Development and Utilization of Computational Model (RF-Model-I)	

to Ratify and Complement HTS Results.....	107
Use of RF-Model-II to Screen a Large in silico Library	109
Dragon Descriptors Important for RF Models of MATE1 Inhibitors	111
Clinically Relevant MATE1 Inhibitors and Their Implications for	
Metformin DDIs.....	111
DISCUSSION	113
EXPERIMENTAL SECTION	119
Reagents.....	119
Cell Culture.....	120
Fluorescent-Probe Uptake Assay	120
High throughput Screen (HTS) of Transporter Inhibition	121
Experimental IC50 Determination	122
MATE1 IC50 Determinations With 14C-Labelled Metformin in	
HEK293 cells	122
MATE1 IC50 Determinations With 14C-Labelled Metformin in	
polarized MDCK-II cells	123
<i>In silico</i> Modeling	124
In silico Model Refinement	125
DrugBank Ligand Screening.....	126
Principal Component Analysis of Physicochemical Properties	126
Statistical Analysis.....	127
FIGURES	128
REFERENCES	142

Chapter 5: Discovery of Inhibitors of the Human OATP2B1 Transporter by

Virtual Screening Against Multiple Comparative Models and Machine	
Learning	148
ABSTRACT.....	149
INTRODUCTION	149
RESULTS	152
Selection of informative molecular descriptors of known OATP2B1	

inhibitors and noninhibitors	152
Development and validation of binary classification models using machine learning	153
Virtual screening of the library with the RF classifier	154
In vitro validation of inhibitors predicted with RF model	155
Comparative structure modeling of human OATP2B1 transporter ...	156
Docking and clustering of docked compounds	157
Determination of IC ₅₀ values for selected compounds	158
Physicochemical properties of validated hits	159
DISCUSSION	160
MATERIALS AND METHODS	166
Pairwise compound similarity computation and clustering	166
Selection of informative molecular descriptors and machine learning	166
Virtual screening of the DrugBank library using Random forest classifier	167
Comparative modeling of the 3D structure of the human OATP2B1 and fragment-based identification of putative binding sites	168
Virtual screening of the DrugBank library against the comparative models of OATP2B1	168
Reagents	169
Cell culture	169
Inhibition experiments with CHO-OATP2B1	170
Determination of IC ₅₀ values	170
ACKNOWLEDGEMENTS	171
REFERENCES	171
TABLES	177
FIGURES	178
SUPPLEMENTARY INFORMATION	187
SUPPLEMENTARY ON-LINE DATA FILE	194

Chapter 6: ModBase, a database of annotated comparative protein structure

models, and associated resources.....	195
ABSTRACT.....	196
INTRODUCTION.....	196
METHODS.....	198
Model generation by comparative modeling (Modeller and modpipe).....	198
ModBase model sets.....	200
ModWeb: Comparative modeling web-server.....	201
Associated Resources.....	202
AllosMod: a web-server for modeling ligand-induced protein dynamics.....	202
A family of web-servers for computation and application of SAXS profiles.....	203
Pose & Rank: a web-server for scoring protein-ligand complexes ..	206
APPLICATION EXAMPLES.....	207
Coordinating the impact of structural genomics on the human α -helical transmembrane proteome.....	207
Structural determinants of Human Immunodeficiency Virus-1 (HIV-1) protease.....	209
ACCESS AND INTERFACE.....	210
Direct access.....	210
The Protein Model Portal.....	211
Access through external databases.....	212
Future Directions.....	212

Chapter 7: Hemoglobin S Polymerization Studied by Rapid Brownian

Dynamics Simulations.....	217
INTRODUCTION.....	217
METHODS AND RESULTS.....	219
Preparation of the Macromolecular Structures for Simulation.....	219

Tabulation of Protein-Protein Interaction Energies and Forces.....	220
Brownian Dynamics Simulations of Polymerization.....	221
Analysis of time courses	223
Analysis of the HbS Dimers	224
SUMMARY AND FUTURE DIRECTIONS.....	226
FIGURES.....	228
REFERENCES	238

List of Tables

Table 3-1: assessment the GAT-2 models	86
Table 4-1. Overview of IC50 Results for Selected HTS Hits.....	135
Table 4-2. Results of Random Forest-Model-I (RF-Model-I) Validation.	136
Table 4-3. Selected Hits of the DrugBank Library-in silico Screening with RF-Model-II and Their Experimental IC50 Values.....	138
Table 4-4. Overview of the 21 most important descriptors used in the RF-Model-II (mean descriptor values of MATE1 inhibitors and non-inhibitors are presented).....	139
Table 4-5. Summary of IC50 Values for Putative Clinically Relevant Inhibitors of MATE1, MATE2-K or OCT2 in ASP+ Uptake Assay.....	140
Table 4-6. Follow-up IC50 Experiments with Metformin for Putative Clinically Relevant MATE1 Inhibitors in Different Assays.....	141
Table 5-1. Physicochemical properties of OATP2B1 inhibitors identified with different computational methods.	177

List of Figures

Fig. 2-1. Structures of SLC transporters.	36
Fig. 2-2. Sequence similarity network of human SLC families predicted to have the MFS fold.....	38
Fig. 2-3. Binding sites and modes of ligand binding in SLC6 members.	39
Fig. 2-4. Structure-based ligand prediction for GAT-2.	40
Fig. 2-5. Ligand discovery for MATE-1.....	41
Fig. 3-1. GAT-2 – GABA models and their validation by ligand enrichment.	78
Fig. 3-2. Validating predicted binding site residues by mutagenesis.	80
Fig. 3-3. Chemical similarity network of predicted ligands.	81
Fig. 3-4. Predicted binding modes for GAT-2 validated ligands.....	82
Fig. 3-5. Cis-inhibition studies of predicted GAT-2 inhibitors.	84
Fig. 3-6. Comparison of GAT-2 and NET predicted binding sites.....	85
Fig 3-S1. GAT-2-LeuT alignment.	94
Fig 3-S2. The extracellular gate.....	95
Fig 3-S3. E48 in various models of GAT-2.....	96
Fig. 4-1. Methods development and results of high throughput screening.....	128
Fig. 4-2. Comparison of MATE1 and OCT2 inhibitors.	129
Fig. 4-3. Comparison of physicochemical parameters for different groups of inhibitors.....	132
Fig. 4-4. Receiver Operating Characteristic curves for ten external data sets.....	133
Fig. 4-5. Selectivity of putative clinically relevant inhibitors of MATE1, MATE2-K or OCT2.....	134
Fig. 5-1. In silico OATP2B1 inhibitor discovery pipeline.	178

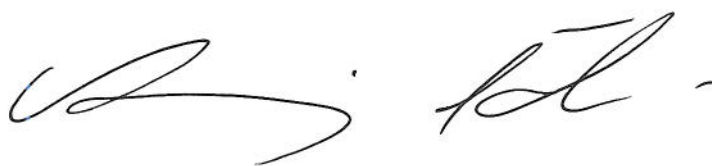
Fig. 5-2. Results of feature selection, ligand-based modeling, and virtual screening of DrugBank library.....	179
Fig. 5-3. 2-concentration inhibition of E3S uptake in human OATP2B1 over expressing cells.....	181
Fig. 5-4. Results of binding site prediction for comparative models of the human OATP2B1 transporter.	183
Fig. 5-5. Results of the structure-based virtual screening of DrugBank library. .	184
Fig. 5-6. Results of IC ₅₀ determination for putative clinically relevant compounds.	185
Fig. 5-S1. Training set diversity and inhibition activity.	187
Fig. 5-S2. Receiver operating characteristic curves for 100 independent tests using kNN and SVM algorithms.	188
Fig. 5-S3. Seven templates utilized for comparative structure modeling and two-way clustering of docking against the seven comparative models.....	189
Fig. 5-S4. Density of DOCK scores in each cluster.	191
Fig. 5-S5. Box plots of physicochemical properties of OATP2B1 inhibitors and noninhibitors identified in this study.....	192
Fig. 5-S6. Structural similarity of predicted OATP2B1 inhibitors.....	193
Fig. 6-1: The computed profiles for filament models of the XLF-XRCC4 complex [353] are fitted to the experimental SAXS profile with FoXS.	213
Fig. 6-2. Cleavage of human proteins by the HIV-1 protease: Crystal structure of the N-terminal domain of human Lupus La protein [354].	214
Fig. 6-3. ModBase Interface Elements: Search Form: search options are	

available through the pull-down menu.	215
Fig. 7-1. Sick cell polymerization process.	228
Fig. 7-2. Pre-computing pairwise forces and torques.	229
Fig. 7-3. Benchmarking time step and neighbor list algorithm.	230
Fig. 7-4. Results of the time-course analyses for 100 independent simulations at two different volume fractions (0.25 and 0.33).	231
Fig. 7-5. Frequency of HbS polymers in 100 independent simulations at two volume fractions (0.25 and 0.33).	232
Fig. 7-6. Analysis of HbS dimers using self-organizing maps (SOMs).	233
Fig. 7-7. Density of all pairwise center-of-mass distances (Å) in 100 independent simulations at volume fraction of 0.25.	234
Fig. 7-8. Contact frequencies of residues in hemoglobin dimers.	235
Fig. 7-9. Second most frequent interface involving the sickle cell mutation.	236
Fig. 7-10. Detailed representation of Asp47 in the structure of deoxy-HbS.	237

Contributions

Several chapters of this dissertation describe studies that have been published in the journals cited herein. The chapters do not necessarily represent the final published form. These studies were carried out with the collaborators listed as co-authors for each chapter. The study described in Chapter 2, “Molecular modeling and ligand docking for solute carrier (SLC) transporters”, was published in 2013 in *Current Topics of Medicinal Chemistry*, 13(7), pp. 843-56. For this study, Natalia developed a method for predicting ligand specificities of 15 human SLC transporters expressed in kidney, analyzed the data, and wrote the review with co-authors. The study described in Chapter 3, “High selectivity of the γ -aminobutyric acid transporter 2 (GAT-2, SLC6A13) revealed by structure-based approach”, was published in November 2012 in *Journal of Biological Chemistry*, 287(45), pp. 37745-56. In this study, Natalia Khuri built a comparative model of the human GAT-2 transporter in the occluded conformation, performed virtual screening of the ZINC library against the comparative model, carried out data analysis and wrote the paper with co-authors. The study described in Chapter 4, “Discovery of potent, selective multidrug and toxin extrusion transporter 1 (MATE1, SLC47A1) inhibitors through prescription drug profiling and computational modeling”, was published in February 2013 in *Journal of Medicinal Chemistry*, 56(3), pp. 781-95. In this study, Natalia designed and carried out the study with co-authors. She developed a computational method for predicting inhibitors of pharmacologically important human transporter, MATE1, carried out analyses of the high-throughput screening data, and wrote the paper with the co-authors. The study described in Chapter 5, “Discovery of Inhibitors of the Human OATP2B1 Transporter by Virtual Screening

Against Multiple Homology Models and Machine Learning”, is currently in preparation and will be submitted to Journal of Medicinal Chemistry. Natalia designed and developed a computational method combining ligand- and structure-based approaches to predict inhibitors of the OATP2B1 transporter. She managed collaborations, designed experiments and is currently writing the manuscript with the co-authors. The study described in Chapter 6, “ModBase, a database of annotated comparative protein structure models and associated resources” was published in January 2014, in Nucleic Acids Research, Vol. 42, pp. D336-46. Natalia designed and carried out the study of structural determinants of the HIV-1 protease, and wrote the review with co-authors. The study described in Chapter 7, “Hemoglobin S Polymerization Studied by Rapid Brownian Dynamics Simulations”, is currently in preparation. Natalia designed and developed a method for simulating polymerization of sickle cell hemoglobin at biologically relevant concentrations and on relevant time scales. She ran simulations, analyzed the data and is currently writing the manuscript. The work of Natalia Khuri described herein was done under the supervision of Andrej Sali and meets the requirements for a standard PhD dissertation.

A handwritten signature in black ink, appearing to read 'Andrej Sali', with a stylized flourish at the end.

Andrej Sali

Chapter 1: Introduction

In cells, proteins interact with other molecules in a precise and time-dependent manner. These interactions determine all cellular functions, such as the catalysis of chemical reactions, motility, and regulation of the transmission of hereditary information. Structural analysis and modeling of these interactions and their interaction interfaces is a critical step toward a mechanistic description of all biological processes.

IDENTIFICATION OF POTENT, CLINICALLY SIGNIFICANT INHIBITORS OF HUMAN SLC TRANSPORTERS

Membrane proteins, such as transporters, are required for the movement of selected molecules into and out of cells. Multiple combinations of ATP-binding cassette (ABC) and Solute Carrier (SLC) transporters are present on cell membranes in all species, facilitating many important physiological processes, such as ion homeostasis, immune functions, cellular metabolism, onset and progression of cancer [1]. These proteins act in concerted fashion to deliver nutrients and to detoxify the cells. Among the molecules transported by these proteins are amino acids, nucleotides, ions, sugars, and vitamins. Transporters are also involved in intestinal absorption, tissue distribution, biliary and urinary excretion of drugs, and regulation of systemic exposure. Overexpression of transporters is implicated in many cancers, where cell's metabolism is altered as a cause or a consequence of an increased need for energy and nutrients [2, 3]. Conversely, when transmembrane transport fails, cells are unable to regulate their internal environment and become vulnerable to endogenous and exogenous toxic ligands and metabolites. Moreover, drug resistance is often directly correlated with up- or down-regulation of transporters, limiting success of therapeutic treatment of cancer, heart diseases, infectious

diseases, and neurological disorders [4]. Finally, mutations in transporter proteins have been causally associated not only with changes in transport rates, but also with progression and prognosis of the disease [5, 6]. It is, therefore, important to characterize membrane transporter proteins and predict their endogenous and exogenous ligands.

The drug discovery process is time-consuming and expensive. Combinatorial chemistry and high-throughput screening (HTS) technologies offered great promise for accelerating the drug discovery pipeline. These advances enabled the synthesis of huge libraries of compounds, which could be screened against a biological target in short period of time. However, many promising HTS hits failed in a subsequent optimization process due to absorption, distribution, metabolism, excretion, and toxicity (ADME/Tox) issues. Therefore, there remains the need for efficient strategies for choosing hits by eliminating compounds with unsuitable properties. As a rapid alternative to HTS, computational methods, such as virtual screening (VS), can be employed to rapidly scan the chemical space of compounds for those with desired affinity against the chosen target(s). Virtual screening is a knowledge-driven approach to search compound databases for novel molecules with desired biological activities. Virtual screening can be used to search for structurally novel ligands or for *ab initio* ligand discovery for unexplored drug targets. The databases available for searching contain up to 10 million molecules, some of which have not yet been synthesized. To deal with the large size of the search space, in a typical VS pipeline, a series of filters is applied sequentially to narrow down and prioritize the hits.

Approaches to VS can be broadly divided into two categories, ligand-based (LBVS) and structure-based (SBVS) [7]. Ligand-based approaches rely on structure-activity relationship (SAR) models of compounds with confirmed biological activities for prediction of candidate molecules with desired biological activity [8]. Methods for LBVS include machine learning,

such as similarity and substructure searching, regression, and classification, as well as two-dimensional pharmacophore and three-dimensional shape matching [9]. In contrast, SBVS requires knowledge of a three-dimensional structure of the target, which is either determined experimentally or modeled computationally. *In silico* libraries are then docked against the binding site in the target structure and poses of candidate compounds are ranked based on their complementarity to the binding site [10].

To answer questions about transporter's specificity and selectivity, high-resolution 3D structures of the protein in multiple conformations are needed. Because X-ray structures of many human transporters have not yet been solved, homology models can be built using available 3D structures of bacterial homologues as templates. Finding suitable templates is difficult, because sequence identity between most human and bacterial homologues is low (~14-30%). A recent study analyzed the human genome to group human SLC transporters into clusters of similar structural folds and identified the best bacterial templates for each group [11].

I used multiple conformations of a bacterial leucine transporter (LeuT) to model the structure of the human γ -amino butyric acid transporter 2 (GAT-2) expressed in a wide range of neuronal and non-neuronal cells, such as dendrites, axon terminals, as well as epithelial cells. A two-pronged approach was taken to validate a homology model of GAT-2 (Chapter 3). First, we identified several residues lining the predicted binding site and performed site-directed mutagenesis of these residues, thus inhibiting GAT-2 transport activity. Second, we identified six structurally and therapeutically novel inhibitors of GAT-2 by docking over 170,000 chemicals against its homology model and selecting several top-ranked compounds for the experimental inhibition assay. In summary, we showed the utility of comparative modeling complemented by computational docking for the functional studies of human SLC transporters.

Unfortunately, many human transporters do not even have suitable high-quality bacterial templates. For instance, our attempt to model the structure of the multidrug and toxin extrusion protein 1 (MATE1) was unsuccessful. Instead, we performed a high throughput screen (HTS) of over 900 drugs and drug-like molecules and identified 84 putative inhibitors (Chapter 4). Using this HTS data, I built predictive classification models linking the physicochemical properties of compounds to their inhibitory activity, i.e. inhibiting or non-inhibiting. The corresponding classifier is based on the Random forest algorithm[12], with feature selection performed with the genetic algorithm. The classifier achieved an area under the Receiver Operating Characteristic (ROC) curve of 0.86, indicating a high predictive value. The model was employed in two ways. First, we identified, by the means of 5-fold cross-validation, compounds with mismatched predicted and observed inhibitory activity and subjected them to follow up screening experiments. This analysis allowed us to rescue several MATE1 inhibitors and identify false positives. Second, we validated the model in a prospective test by predicting five novel MATE1 inhibitors from the DrugBank library[13]. To my knowledge, this is the first application of machine learning in this problem domain. One of the compounds identified by us is currently in clinical trial at UCSF to explore its potential interaction with the widely prescribed anti-diabetic drug metformin.

While both methods for predicting inhibitory ligands of human transporters were successful, their utility is limited by their applicability domain [14-17]. For instance, computational docking tends to find ligands that bind in the same pockets on the protein as substrates and/or competitive inhibitors. HTS followed by machine learning of probabilistic classifiers can only identify compounds with structural properties and mechanisms of action similar to those in the training set. To resolve the shortcomings of the two methods, I have

developed a hybrid approach combining structure- and ligand-based techniques and successfully applied it to predict ligands for a human member of the OATP superfamily of transporters. For identifying binding pockets in homology models of human organic anion-transporting polypeptide 2B1 (OATP2B1), I docked small fragments of organic molecules [18] in the large substrate-translocation pore, and determined “docking profiles” for the known substrates of OATP2B1. These docking profiles successfully predicted structurally and therapeutically novel potent inhibitors in retrospective and prospective experimental studies (Chapter 5).

MODELING DYNAMIC PROTEIN-PROTEIN INTERACTIONS IN HIV-1 INFECTION

Protein-protein interactions occur frequently in cells and are an essential aspect of the cell function. However, some protein-protein interactions can lead to undesirable outcomes. For example, the Gag and Pol polyproteins of the human immunodeficiency virus (HIV) are cleaved by the HIV-1 protease [19]. The aspartic protease is a homodimer that catalyzes these processing events at ten non-homologous sites. These sites are eight amino acid residues in length; the cleavage occurs between the third and fourth residues [20-23]. Many antiretroviral drugs target the HIV-1 protease [24-26]. Interestingly, in addition to processing viral proteins, HIV-1 protease cleaves several human proteins during infection, such as subunit D of the eukaryotic translation initiation factor 3 (eIF3D) [27-30]. It is thus important to develop predictive methods to rapidly identify targets of the HIV-1 protease in the host and their cleavage sites. Using a data-driven approach utilizing primary sequence and structural features of over 120 known cellular substrates of HIV-1 protease [31], we profiled the human proteome and identified several putative cellular substrates (Chapter 6). In retrospective five-fold cross-validation, model's predictions were 94% accurate.

MAPPING DYNAMIC ASSEMBLY / DISASSEMBLY PATHWAYS OF MACROMOLECULAR COMPLEXES

Realizing that many details of molecular recognition and interactions will be missed when only static structures are modeled, I have embarked on the development of a computational framework for data-driven simulations of dynamic macromolecular processes, at both the atomic and coarse-grained resolutions (Chapter 7). I developed a method for directly incorporating experimental data and theoretical considerations into simulations, such that the results are consistent with experiments by construction. This goal had been achieved by the integration of existing simulation techniques, such as the Brownian dynamics algorithm [32], with stochastic sampling techniques that recursively estimate states of dynamic systems from given simulation trajectories. I applied this simulation approach to model the process of hemoglobin S polymerization, the primary pathogenic event in the sickle cell disease [33, 34]. More specifically, I incorporated my method into the open-source Integrative Modeling Platform (IMP) software [35], to study the diffusion and hemoglobin interactions at physiologically relevant concentrations and time scales. Dimers of hemoglobin tetramers are the predominant oligomeric species formed in the simulation, although larger aggregates with more than five hemoglobin tetramers also occur. Additionally, the known interaction between $\beta 6\text{Val}$ and $\beta 85\text{Phe}/\beta 88\text{Leu}$ residues [33, 34, 36] is observed frequently (occurring in 43% of dimer interfaces), thus demonstrating that the model can rationalize even some experimental data that were not used in its construction. Finally, the simulation predicted that several unreported protein-protein interaction interfaces play a role in polymerization. Perturbation of these interfaces with small molecules might allow us to modulate the polymerization process.

The rest of this dissertation is organized as follows. In Chapter 2, we present structure- and ligand-based approaches for the discovery of ligands for human solute carrier transporters. Challenges in comparative protein structure modeling, docking of small molecules to proteins, and evaluation of ligand-protein complex models are discussed. In Chapter 3, we report the results of comparative modeling and ligand discovery for the human GAT-2 transporter. In Chapter 4, we describe the ligand-based approach utilizing machine learning and high-throughput experimental screening that identified 84 new inhibitors of the human MATE1 transporter. We present in Chapter 5 how ligand-based and structure-based approaches can be combined to predict inhibitors of other transporters. Machine learning and graph-theoretic prediction of human substrates of the HIV-1 protease is described in Chapter 6. In Chapter 7, we discuss our contribution to accelerating Brownian dynamics simulations by the means of pre-computed forces and torques. Each chapter contains supplemental materials sections with additional information.

REFEENCES

1. Giacomini, K. M.; Huang, S. M. Transporters in drug development and clinical pharmacology. *Clin Pharmacol Ther* 2013, 94, 3-9.
2. Lemstrova, R.; Soucek, P.; Melichar, B.; Mohelnikova-Duchonova, B. Role of solute carrier transporters in pancreatic cancer: a review. *Pharmacogenomics* 2014, 15, 1133-45.
3. Nagamori, S.; Kanai, Y. [Amino acid transporters in cancer]. *Seikagaku* 2014, 86, 338-44.
4. DeGorter, M. K.; Xia, C. Q.; Yang, J. J.; Kim, R. B. Drug transporters in drug efficacy and toxicity. *Annu Rev Pharmacol Toxicol* 2012, 52, 249-73.
5. Klaassen, C. D.; Aleksunes, L. M. Xenobiotic, bile acid, and cholesterol transporters: function and regulation. *Pharmacol Rev* 2010, 62, 1-96.
6. Nakanishi, T.; Tamai, I. Genetic polymorphisms of OATP transporters and their impact on intestinal absorption and hepatic disposition of drugs. *Drug Metab Pharmacokinet* 2012, 27, 106-21.

7. Dror, O.; Shulman-Peleg, A.; Nussinov, R.; Wolfson, H. J. Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design. *Curr Med Chem* 2004, 11, 71-90.
8. Jahn, A.; Hinselmann, G.; Fechner, N.; Zell, A. Optimal assignment methods for ligand-based virtual screening. *J Cheminform* 2009, 1, 14.
9. Villoutreix, B. O.; Renault, N.; Lagorce, D.; Sperandio, O.; Montes, M.; Miteva, M. A. Free resources to assist structure-based virtual ligand screening experiments. *Curr Protein Pept Sci* 2007, 8, 381-411.
10. Fan, H.; Irwin, J. J.; Sali, A. Virtual ligand screening against comparative protein structure models. *Methods Mol Biol* 2012, 819, 105-26.
11. Schlessinger, A.; Yee, S. W.; Sali, A.; Giacomini, K. M. SLC classification: an update. *Clin Pharmacol Ther* 2013, 94, 19-23.
12. Breiman, L. Random Forests. *Machine Learning* 2001, 45, 5-32.
13. Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y.; Eisner, R.; Guo, A. C.; Wishart, D. S. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res* 2011, 39, D1035-41.
14. Dimitrov, S.; Dimitrova, G.; Pavlov, T.; Dimitrova, N.; Patlewicz, G.; Niemela, J.; Mekenyan, O. A stepwise approach for defining the applicability domain of SAR and QSAR models. *J Chem Inf Model* 2005, 45, 839-49.
15. Dragos, H.; Gilles, M.; Alexandre, V. Predicting the predictability: a unified approach to the applicability domain problem of QSAR models. *J Chem Inf Model* 2009, 49, 1762-76.
16. Sahigara, F.; Mansouri, K.; Ballabio, D.; Mauri, A.; Consonni, V.; Todeschini, R. Comparison of different approaches to define the applicability domain of QSAR models. *Molecules* 2012, 17, 4791-810.
17. Sheridan, R. P. Three useful dimensions for domain applicability in QSAR models using random forest. *J Chem Inf Model* 2012, 52, 814-23.
18. Brenke, R.; Kozakov, D.; Chuang, G. Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* 2009, 25, 621-7.
19. Freed, E. O. HIV-1 replication. *Somatic cell and molecular genetics* 2001, 26, 13-33.
20. Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A.; Scolnick, E. M.; Sigal, I. S. Active human immunodeficiency virus protease is required for viral infectivity. *Proceedings of the National Academy of Sciences of the United States of America* 1988, 85, 4686-90.
21. Murthy, K. H.; Winborne, E. L.; Minnich, M. D.; Culp, J. S.; Debouck, C. The crystal structures at 2.2-A resolution of hydroxyethylene-based inhibitors bound to human immunodeficiency virus type 1 protease show that the inhibitors are present in two distinct orientations. *The Journal of biological chemistry* 1992, 267, 22770-8.

22. Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *Journal of molecular biology* 2000, 301, 1207-20.
23. Mahalingam, B.; Louis, J. M.; Hung, J.; Harrison, R. W.; Weber, I. T. Structural implications of drug-resistant mutants of HIV-1 protease: high-resolution crystal structures of the mutant protease/substrate analogue complexes. *Proteins* 2001, 43, 455-64.
24. McDonald, C. K.; Kuritzkes, D. R. Human immunodeficiency virus type 1 protease inhibitors. *Archives of internal medicine* 1997, 157, 951-9.
25. Drag, M.; Salvesen, G. S. Emerging principles in protease-based drug discovery. *Nature reviews. Drug discovery* 2010, 9, 690-701.
26. Flexner, C. HIV-protease inhibitors. *The New England journal of medicine* 1998, 338, 1281-92.
27. Nie, Z.; Phenix, B. N.; Lum, J. J.; Alam, A.; Lynch, D. H.; Beckett, B.; Krammer, P. H.; Sekaly, R. P.; Badley, A. D. HIV-1 protease processes procaspase 8 to cause mitochondrial release of cytochrome c, caspase cleavage and nuclear fragmentation. *Cell death and differentiation* 2002, 9, 1172-84.
28. Algeciras-Schimmich, A.; Belzacq-Casagrande, A. S.; Bren, G. D.; Nie, Z.; Taylor, J. A.; Rizza, S. A.; Brenner, C.; Badley, A. D. Analysis of HIV Protease Killing Through Caspase 8 Reveals a Novel Interaction Between Caspase 8 and Mitochondria. *The open virology journal* 2007, 1, 39-46.
29. Nie, Z.; Bren, G. D.; Rizza, S. A.; Badley, A. D. HIV Protease Cleavage of Procaspase 8 is Necessary for Death of HIV-Infected Cells. *The open virology journal* 2008, 2, 1-7.
30. Castello, A.; Franco, D.; Moral-Lopez, P.; Berlanga, J. J.; Alvarez, E.; Wimmer, E.; Carrasco, L. HIV- 1 protease inhibits Cap- and poly(A)-dependent translation upon eIF4GI and PABP cleavage. *PloS one* 2009, 4, e7997.
31. Impens, F.; Timmerman, E.; Staes, A.; Moens, K.; Arien, K. K.; Verhasselt, B.; Vandekerckhove, J.; Gevaert, K. A catalogue of putative HIV-1 protease host cell substrates. *Biological chemistry* 2012, 393, 915-31.
32. Ermak, D. L.; McCammon, J. A. Brownian dynamics with hydrodynamic interactions. *The Journal of Chemical Physics* 1978, 69, 1352-1360.
33. Perutz, M. F.; Mitchison, J. M. State of haemoglobin in sickle-cell anaemia. *Nature* 1950, 166, 677-9.
34. Perutz, R. R.; Liquori, A. M.; Eirich, F. X-ray and solubility studies of the haemoglobin of sickle-cell anaemia patients. *Nature* 1951, 167, 929-31.
35. Webb, B.; Lasker, K.; Velazquez-Muriel, J.; Schneidman-Duhovny, D.; Pellarin, R.; Bonomi, M.; Greenberg, C.; Raveh, B.; Tjioe, E.; Russel, D.; Sali, A. Modeling of proteins and their assemblies with the Integrative Modeling Platform. *Methods Mol Biol* 2014, 1091, 277-95.

36. Nagel, R. L.; Johnson, J.; Bookchin, R. M.; Garel, M. C.; Rosa, J.; Schiliro, G.; Wajcman, H.; Labie, D.; Moo-Penn, W.; Castro, O. Beta-chain contact sites in the haemoglobin S polymer. *Nature* 1980, 283, 832-4.

Chapter 2: Molecular modeling and ligand docking for Solute Carrier (SLC) transporters

Avner Schlessinger^{1,2,*}, Natalia Khuri^{1,2,4}, Kathy Giacomini^{1,2}, Andrej Sali^{1,2,3,*}

¹Department of Bioengineering and Therapeutic Sciences, and ²California Institute for Quantitative Biosciences, and ³Department of Pharmaceutical Chemistry, ⁴Graduate Group in Biophysics, University of California, San Francisco, 1700 4th Street, San Francisco, CA 94158.

*Corresponding authors: Avner Schlessinger, avnersch@gmail.com, tel: (415) 514 4228; Andrej Sali, email sali@salilab.org, tel: (415) 514 4227.

ABSTRACT

Solute Carrier (SLC) transporters are membrane proteins that transport solutes, such as ions, metabolites, peptides, and drugs, across biological membranes, using diverse energy coupling mechanisms. In human, there are 381 SLC transporters, many of which contribute to the absorption, distribution, metabolism, and excretion of drugs and/or can be targeted directly by therapeutics. Recent atomic structures of SLC transporters determined by X-ray crystallography and NMR spectroscopy have significantly expanded the applicability of structure-based prediction of SLC transporter ligands, by enabling both comparative modeling of additional SLC transporters and virtual screening of small molecules libraries against experimental structures as well as comparative models. In this review, we begin by describing

computational tools, including sequence analysis, comparative modeling, and virtual screening, that are used to predict the structures and functions of membrane proteins such as SLC transporters. We then illustrate the applications of these tools to predicting ligand specificities of select SLC transporters, followed by experimental validation using uptake kinetic measurements and other assays. We conclude by discussing future directions in the discovery of the SLC transporter ligands.

SOLUTE CARRIER TRANSPORTERS

Transporters are membrane proteins that control the uptake and efflux of various solutes, including metabolites, ions, toxins, and drugs. They include solute carrier (SLC) transporters as well as ion dependent pumps and ATP-binding cassette (ABC) transporters. The majority of the SLC transporters are secondary active transporters, such as exchangers, symporters, and antiporters, for which transport is driven by various energy coupling mechanisms [37-43].

Here, we focus on the human SLC transporters, which include 381 members grouped into 52 families based on their sequences, number of transmembrane α -helices (TMHs) (typically 10-14 TMHs), and biological functions [37, 38, 40]. The SLC members play an important role in a variety of cellular functions, often cooperatively with other protein families, including receptors, enzymes, and other transporters, as illustrated by the following examples.

First, members of the SLC families SLC1 (glutamate and neutral amino acid transporter), SLC6 (Na⁺- and Cl⁻-dependent neurotransmitter transporters), SLC17 (vesicular glutamate transporters), SLC18 (vesicular amine transporters), and SLC32 (vesicular inhibitory amino acid transporters) control the concentration of neurotransmitters such as glutamate and serotonin in the synapses, thereby regulating downstream neurosignaling pathways that are activated or inhibited via membrane receptors [44]. Several members of these SLC families are important

drug targets. For instance, the norepinephrine transporter (NET/SLC6A2) transports norepinephrine from synaptic spaces into presynaptic neurons to regulate downstream adrenergic pathways associated with behavioral traits. Thus, several drugs, such as the attention-deficit hyperactivity disorder (ADHD) drug methylphenidate (Ritalin®), inhibit NET activity to increase adrenergic signaling [45, 46].

Second, members of several SLC families, such as SLC21 (organic anion transporters), SLC22 (organic cation/anion/zwitterion transporters), and SLC47 (multidrug and toxin extrusion (MATE) transporters), are highly abundant in the liver, kidney, and blood-brain-barrier (BBB) where they regulate drug absorption, distribution, metabolism, and elimination (ADME) [47-49]. For example, the organic ion transporter 1 (OCT1/SLC22A1) transports anti-cancer and anti-viral drugs into the liver and kidney and may, therefore, mediate drug-drug interactions [50, 51].

Third, mutations in SLC members can lead to differential drug response among individuals (pharmacogenetics) [52, 53]. For instance, the intracellular concentrations of the anti-diabetic drug metformin are affected by genetic variations in OCT1 [54-57].

SEQUENCES, STRUCTURES, AND MECHANISMS OF SLC TRANSPORTERS

By definition, members of each of the 52 SLC transporter families share sequence identity of 20% or more to at least one other family member, and do not exhibit any significant sequence similarity to members of other families [37]. Several additional classifications of the SLC members based on their amino-acid sequences have also been suggested [39, 58]. The Transporter Classification (TC) system classifies transporters from all organisms, based on their sequence similarity and cellular functions (<http://www.tcdb.org/>) [39, 40]. The TC is a hierarchical classification consisting of five components, including “class” (eg, a carrier), “subclass” (eg, secondary-active) and “superfamily” (eg, the Major Facilitative Superfamily

(MFS)). Due to the diversity of the human SLC members, they occur in a number of TC groups [40]. For example, the SLC6 and the SLC8 families are assigned to the neurotransmitter:Sodium Symporter (NSS; 2.A.22) and the Ca²⁺:Cation Antiporter (CaCA; 2.A.19) superfamilies, respectively. Furthermore, a phylogenetic analysis of the SLC transporters using Hidden Markov Models (HMMs) refined the current classification, grouping 15 of the 52 SLC families into four distinct clusters [58].

Currently, the only human SLC transporter with an experimentally determined atomic structure is the Rhesus glycoprotein ammonium transporter (RhCG, SLC42A3) [59]. In addition, there are several high-resolution structures of proteins from prokaryotic and other eukaryotic organisms that are similar to the human SLC members, sharing at least 25% sequence identity with the human homolog [60-65]. These structures revealed that the human SLC transporters are likely to be highly diverse in structure (Fig. 2-1A). For example, the structure of the proton glutamate symporter Glp from *Pyrococcus horikoshii* (an SLC1 homolog) is dissimilar to the structure of the leucine transporter LeuT from *Aquifex aeolicus* (an SLC6 homolog). In addition, the structures of prokaryotic SLC homologs revealed that some SLC families share similar structural features despite weak sequence relationships (less than 15% sequence identity). For example, the prokaryotic homologs of the SLC5 (vSGLT) [60], SLC6 (LeuT) [61], SLC7 (Arginine/agmatine antiporter (AdiC) [62, 63], and ApcT transporter [64]) families, as well as other prokaryotic transporters, including the uracil transporter UraA [65], the sodium-hydantoin transporter Mhp1 [66], the Na⁺/betaine symporter BetP [67], the glutamate/gamma-aminobutyrate antiporter [68], and the L-carnitine/gamma-butyrobetaine antiporter CaiT [69], are classified into the same structural family [70] that is often referred to as the LeuT-like or NSS fold.

The majority of the SLC structures contain evolutionary-related inverted structural repeats that are highly divergent in sequence [39, 42, 71-73]. For example, the LeuT fold repeats typically include five TMHs that are related to each other via a two-fold pseudo-symmetry axis parallel to the membrane plane; the first helix of each repeat is often involved in ligand binding [42, 71]. Although the intricate details of the transport mechanisms differ among the SLC members of the different folds (or even within the same fold), their structures support a common ‘alternating access’ transport mechanism, in which the transporter alternately exposes its primary binding site at either side of the membrane [41, 43, 61, 71, 72, 74-76] (Fig. 2-1B). For example, LeuT structures have been determined in four conformations representing distinct snapshots of a putative transport cycle [76-78]. Recent LeuT structures in complex with various amino acid substrates and inhibitors suggested a competitive inhibition mechanism in which larger and more hydrophobic ligands, such as tryptophan, stabilize an inactive outward-facing conformation [78]. Additional putative second substrate-binding site (‘S2’) has also been shown to be located on the LeuT surface [79-81], while its affinity for substrates remains uncertain [82]. However, a number of inhibitors were shown to stabilize different conformations of LeuT via sites overlapping with the S2 site [82, 83].

Furthermore, various conformations of SLC transporters that are adopted during the transport cycle have been revealed using other experimental techniques, such as single molecule fluorescence resonance energy transfer (smFRET) [79-81, 83] as well as crystallographic characterizations of various LeuT-like structures (eg, Mhp1) [66, 84, 85]. Recently, alternate states of various SLC transporters were modeled by relying on the internal symmetry of their structures in other states [61, 86] and Molecular Dynamics (MD) simulations [87]. The

mechanisms of transport and inhibition, even for such an extensively studied model transporter as LeuT, are far from being fully described.

COMPARISON AND CLUSTERING OF SLC TRANSPORTERS

We recently performed a comprehensive analysis of SLC transporters that included topology prediction, profile-profile sequence alignment, sequence-based clustering to create similarity networks, and clustering analysis [88]. A node of a sequence similarity network corresponds to a protein sequence; edges connect those pairs of sequences that are related above a defined similarity threshold. A variety of different functional annotations, including substrate type, transport mode, organism conservation, and tissue specificity, were mapped on the similarity networks. The goal was to identify relationships among human SLC transporters and extend their classification.

SLC families of homologs with similar structures generally clustered together, despite exhibiting relatively weak sequence similarities that do not traditionally suffice for a reliable transfer of structural and functional annotation (~10% sequence identity) [89]. For example, the SLC5, SLC6, and SLC7, which exhibit significant sequence similarity to prokaryotic proteins with LeuT-like fold, were interconnected. Our analysis indicated additional families, such as the SLC32 (the vesicular inhibitory amino acid transporter family), SLC36 (the proton-coupled amino acid transporter family), and SLC38 (System A and System N sodium-coupled neutral amino acid transporter family) as LeuT-like fold members.

The recent structure of the D-xylose-proton symporter (an SLC2 homolog) [90] and the structure of the lactose permease LacY (an SLC37 homolog) [91], both of which adopt the major facilitator superfamily (MFS) fold, confirmed our prediction that the human SLC2 and SLC37 families are related, increasing our confidence in the analysis (Fig. 2-2). However, some

structural relationships were not detected using the SLC similarity network, indicating that better sequence comparison methods are still needed for membrane proteins. For instance, PepT_{SO}, a homolog of the human peptide transporter family (PEPT/SLC15), also adopts the MFS fold, but is not connected to any other family in our similarity network [92] (Fig. 2-2). The SLC15 family members appear to be more divergent from other previously characterized MFS transporters (eg, they have additional TMH regions and large insertions in the loops between the TMH) [92].

The similarity network also shows that some families with chemically similar substrates cluster together in the sequence space. For instance, the organic ion transporter families SLC22 and SLC21, both of which are important for drug ADME, are grouped together, suggesting that the SLC similarity network can guide functional annotation of uncharacterized SLC members based on their characterized aligned homologs. Particularly, the SLC22 family was also highly inter-connected to the glucose transporter family (GLUTs/SLC2), which includes members that are associated with cancer and diabetes (Fig. 2-2). The predicted functional overlap between the SLC2 and SLC22 families was confirmed experimentally by kinetic measurements of representative members of these families (eg, the fructose transporter SLC2A9 and the urate transporter SLC22A12), which were found to have a common substrate (ie, uric acid) [88]. In summary, the comprehensive sequence-based comparison was shown to capture many functional and structural relationships among SLC transporters. It thus served as a guide for structural modeling of selected SLC members, as described in the next section.

COMPARATIVE MODELING OF SLC TRANSPORTER STRUCTURES

To construct a comparative model of a human SLC transporter (target), most similar known structures (templates) need to be first identified and aligned with the target sequence. Because SLC families (eg, the SLC5, SLC6, and SLC7) can be only distantly related to proteins

of known structure, state-of-the-art fold assignment and alignment techniques need to be applied judiciously. The majority of the SLC families share a similar topology (10-14 TMHs). Thus, a key step toward aligning their sequences involves identifying these TMH segments, which is fortunately relatively accurate [93, 94]; for example, a support vector machine program MEMSAT-SVM achieves ~90% accuracy on a testing set of experimentally determined membrane protein structures [95].

Efforts to model membrane transporters have also relied on profile-profile alignment (eg, HHpred [96]) and fold recognition methods that rely on structural information (eg, Promals3D [97]), as well as combination of these methods and other techniques [98, 99]. For example, the alignment program AlignMe is based on a dynamic programming algorithm that takes as input the hydropathy plot of a protein in addition to its sequence, thereby improving its alignment accuracy [73]. AlignMe was shown to be accurate in identifying proteins of the LeuT fold and also captured sequence similarities among the LeuT inverted structural repeats that were not identified by other means. In addition, sequence similarity networks have been shown to be a useful tool in functional annotation and structural modeling of various protein families, including GPCRs, kinases, and enzymes families [95, 100, 101].

Once an alignment between the target sequence and one or more known structures is obtained, comparative model building can be performed. For example, the program MODELLER computes an atomic model based on satisfaction of spatial restraints that are derived from the target-template alignment, atomic statistical potentials, and molecular mechanics (61). Finally, comparative models can be assessed using a variety of computational criteria.

STRUCTURE-BASED DISCOVERY OF SLC TRANSPORTER LIGANDS

In structure-based ligand discovery, a large virtual library of small organic molecules is computationally screened against an atomic structure of a protein target ('receptor') [102, 103]. For a protein without an experimentally determined structure, virtual screening can be performed against a comparative protein structure model [104]. The small molecules are docked and scored based on their complementarity to the binding site. For example, in the program DOCK, the score is a sum of van der Waals, Poisson–Boltzmann finite-difference electrostatics, and ligand desolvation penalty terms [105, 106]. The top-scoring molecules are then often analyzed manually and selected for experimental testing. Although one limitation of docking is that binding affinities of ligands cannot yet be predicted accurately, the ability of docking programs to prioritize molecules from large databases, such that one needs to consider only the top few hundred molecules instead of the whole library, is its key strength [103, 107].

The accuracy of virtual screening can be estimated by the enrichment for the known ligands relative to decoy compounds, calculated by docking against the tested model ('enrichment calculations') [108-110]. Thus, by selecting models based on their enrichment scores, comparative models can be optimized for their utility in predicting additional ligands [111-114]. Recent studies have successfully combined modeling and docking approaches with experimental validation of predicted ligands for proteins representing a wide range of functions, including GPCRs [112, 113, 115, 116], nuclear hormone receptors [117], and various enzymes [118-120].

Because of limited high-resolution structural information, most previous discovery efforts targeting SLC transporters have used ligand-based methodologies (eg, pharmacophore modeling) [121-124]. However, a recent increase in the number of atomic structures of several prokaryotic

and eukaryotic homologs expanded our ability to apply structure-based ligand discovery to the human SLC members. Nevertheless, constructing comparative models that are sufficiently accurate for virtual screening still remains challenging, due to the following four reasons: First, it can be difficult, and sometimes even impossible, to align the sequences of the human SLC proteins to their prokaryotic homologs of known structure, because they can be highly divergent in sequence (sequence identity lower than 30%) [88]. Second, modeling and docking programs were not optimized for membrane proteins, as opposed to globular proteins [125]. Third, the quality of the available membrane protein structures is, on average, lower than those for globular proteins. Fourth, comparative modeling and ligand docking against highly dynamic transporters, using the X-ray structure of a single and often unbound conformation, is challenging and sometimes inaccurate [109].

Next, we discuss five case studies in which computational modeling, docking, and experimental follow-up were used to describe specificity determinants of select SLC transporters.

CASE STUDY 1: THE NOREPINEPHRINE TRANSPORTER (NET, SLC6A2)

NET transports norepinephrine from the synapse into presynaptic neurons, thereby regulating signaling pathways that are linked to behavioral traits and cardiovascular effects [126, 127]. NET is a known target for a variety of prescription drugs, including antidepressants and psychostimulants that inhibit its activity to increase neurotransmitter's levels in the synapse.

NET is predicted to have one domain containing 12 TMHs and is related to LeuT (sequence identity of ~25%) [77]. The ligands transported by LeuT (ie, substrates) are chemically different from those of NET (amino acids versus positively charged monoamines), suggesting that the LeuT structure is not well suited for virtual screening for NET ligands and

that NET structure needs to be modeled. Thus, we constructed a model for NET based on the LeuT structure in an outward-occluded substrate-bound conformation [77], which likely represents an outward-occluded active conformation of LeuT [78], and computationally validated the model using enrichment calculations.

The modeled binding site includes the following four key features that might determine ligand-binding specificity in the NET's primary binding site (Figs. 3A,B): First, several polar groups are predicted to make polar interactions with the ligands. For example, the sidechain of Asp75 makes ionic interactions with the amine group of norepinephrine. Second, four residues with aromatic groups are capable of forming interactions involving π electrons (Phe72, Tyr152, Phe317, and Phe323). For example, Phe72 is predicted to form π -cation interactions with norepinephrine via its benzyl sidechain as well as a hydrogen bond via the amide oxygen of its main chain. Third, several hydrophobic residues (eg, Ala145 and Val148) are capable of interacting with ligands via van der Waals interactions and the hydrophobic effect. Fourth, the binding site is relatively small, and, hence, the size of the ligands is limited. These four features are highly conserved among other SLC6 members that transport monoamines, including the dopamine transporter (DAT, SLC6A1) and the serotonin transporter (SERT, SLC6A4). For example, the residues corresponding to NET's Asp75 in SERT (Asp98) [128] and DAT (Asp79) [129] were proposed to have similar key roles in ligand recognition [99, 130].

We computationally screened 6,436 drugs from the Kyoto Encyclopedia of Genes (KEGG DRUG) against the NET model. 18 top-scoring molecules were tested experimentally using a cis-inhibition assay in which the molecules are tested for their ability to inhibit transport of a known transporter substrate (radiolabeled or fluorescent) without distinguishing between inhibitors and substrates [110]. Ten drugs were found to be potent NET inhibitors; five of these

were chemically novel ligands of NET (ie, Tanimoto coefficient of <0.5 against all known ligands, using the Daylight fingerprints [131]). These results may rationalize the efficacy of several sympathetic and antidepressant drugs, as well as side effects of diabetes and Alzheimer's drugs. For example, talsaclidine is a muscarinic M1 receptor agonist that was under development for the treatment of Alzheimer's disease, but failed in clinical trials due to various side effects, such as tachycardia, high blood pressure, nausea, diarrhea, excessive sweating, and palpitation [132]. It was proposed that the pharmacological effects of the drug are caused by its effect on the cholinergic system via binding to the M2 and M3 receptors as well as on the adrenergic system via an unknown mechanism [132]. The pharmacodynamic properties of talsaclidine, coupled with our uptake experiments, suggest that inhibition of the NET uptake might occur in clinically relevant concentrations [53, 110]. Thus, the positive and negative pharmacological effects of the drug can be partially rationalized by NET binding. Finally, these observations highlight the utility of virtual screening against a comparative model, even when the target shares less than 30% sequence identity with its template structure and different ligand binding profiles in the primary binding site.

CASE STUDY 2: THE GABA TRANSPORTER 2 (GAT-2, SLC6A13)

GABA is a key inhibitory neurotransmitter in mammals that activates the GABAergic receptors in inhibitory neurons (eg, GABA_A) [133] and in peripheral tissues, such as the liver, kidney and lungs [133-135]. The GABA transporter (GAT) family consists of four transporters that regulate GABAergic signaling pathway [133, 135] targeted by various anticonvulsants and relaxants (eg, tiagabine). GAT-2 (SLC6A13) is primarily expressed in the liver, kidney as well as in other peripheral tissues (eg, lungs [133-135]), and might also regulate ADME of GABAergic drugs. We modeled two different conformations of GAT-2, using the LeuT

structures in the occluded-outward-facing ('occluded') and an outward-facing ('outward') conformations (Fig. 2-3C). The models were iteratively refined with molecular dynamics simulations and side-chain optimization, and were subsequently evaluated using enrichment calculations [113]. We validated the models experimentally, using site-directed mutagenesis, kinetic uptake measurements, and cis-inhibition assays.

The occluded and outward-facing GAT-2 models are similar to each other, differing only by minor backbone and side-chain rearrangements that make the putative binding site of the outward conformation slightly more accessible to a ligand. However, almost all predicted polar interactions between GABA and GAT-2 are conserved in both the occluded and the outward-facing models (eg, the carboxy group of GABA interacts with the sodium ion Na⁺, Gly51, and Gly53) (Fig. 2-3C). Furthermore, the amino group of GABA forms a key hydrogen bond with the main chain oxygen of Glu48 as well as a polar interaction with the negatively charged side-chain of the same residue. During model refinement, the conformation of the Glu48 side-chain correlated with the enrichment scores for the models, suggesting its importance for ligand recognition (Fig 3C). Site-directed mutagenesis indeed confirmed this residue as important for the function, demonstrating the utility of combining model refinement and ligand docking to predict a key functional residue. Interestingly, while Glu48 and almost all other putative binding site residues are highly conserved among three of the four GAT's (GAT-2, GAT-3, and BGT1) [136], suggesting a common ligand recognition mechanism for these transporters. The GAT-1 binding site, however, is predicted to be unique and likely interacts with GABA via a different mechanism [74, 137].

A virtual screen of 594,166 drugs, metabolites, and fragment-like molecules in the ZINC database [138] against the two GAT-2 conformations retrieved two groups of chemically distinct

ligands (Fig. 2-4). Particularly, molecules predicted to interact with GAT-2 using the outward-facing model (eg, baclofen) were larger and more hydrophobic than those predicted using the occluded model (eg, homotaurine), and covered a different area of the chemical space (Fig. 2-4). We tested 31 small organic molecules experimentally, using the cis-inhibition assay. Twelve ligands were validated, six of which were chemically novel (eg, homotaurine). The validated ligands suggest that GAT-2 is a high-selectivity/low-affinity transporter that is resistant to inhibition by typical GABAergic inhibitors. For example, even ligands that are within one heavy atom of GABA, such as GABOB, were found to inhibit GAT-2 significantly more weakly than GABA (K_m of $26.2 \mu\text{M}$ for GABA versus IC_{50} of $402 \mu\text{M}$ for GABOB) [114]. This difference suggests that GAT-2 has a different binding profile from those of GAT-3 and BGT1 [133-135], despite their conserved primary binding sites. Because GAT-2 is highly abundant in the BBB, it is possible that it evolved to be highly selective to prevent toxins and other molecules from entering the CNS. Furthermore, the pharmacological effects of some of the discovered ligands can still be partially rationalized by GAT-2 binding. For example, the heme-precursor 5-aminolevulinic acid (5-ALA) is involved in neurological side-effect (eg, porphyria) [139, 140] and is also used for the detection of CNS tumors [141-143]. Thus, it is plausible that GAT-2 transports 5-ALA across the BBB into the CNS.

CASE STUDY 3: THE SLC6 FAMILY

The SLC6 transporter family of Na^+ - and Cl^- - dependent transporters can be classified into four groups based on their amino acid sequences and functions – the monoamine transporters, GABA transporters, amino acid transporters, and “orphan” transporters that may also transport amino acids [46, 126]. Interestingly, the SLC6 members are closely related in sequence despite their variability in function. Comparisons between structural models of

representative members of the groups within the SLC6 family in complex with their ligands can help describe key specificity determinants within this family (Fig. 2-3). For example, GAT-2 and NET share sequence identity of ~45%, but bind chemically different ligands in their putative primary binding sites (GABA-like molecules and aromatic monoamines for GAT-2 and NET, respectively). The following three differences in the predicted binding sites of NET and GAT-2 likely rationalize their differences in ligand-binding selectivity (Fig. 2-3B,C): First, the number of aromatic residues (two in GAT-2 and four in NET). Second, the number and location of the charged groups (Glu48 and Na1 in GAT-2, and Asp75 in NET). Third, the size and shape of the binding site (ie, Ala145, Val148, Gly422, and Gly423 in NET are replaced by the larger Val122, Leu125, Val393, and Cys394 in GAT-2). Moreover, these proposed ligand-binding determinants are consistent with previous studies characterizing other monoamine SLC6 transporters [74, 99] SERT [128] and DAT [129] as well as other GATs [74, 99, 136], such as GAT-3 (mouse GAT-4) [136]. In the future, characterization of SLC6 members in additional conformations and at higher resolution will enable us to further address functional variability within the SLC6 family. For example, why do GAT-2 and GAT-3 exhibit substantial difference in binding affinity to ligands, despite almost identical binding sites?

CASE STUDY 4: THE LARGE-NEUTRAL AMINO ACID TRANSPORTER (LAT-1, SLC7A5)

LAT-1 is a Na⁺-independent exchanger of large-neutral amino acids (eg, tyrosine), thyroid hormones (eg, triiodothyronine (T₃)), and prescription drugs (eg, the anti-convulsant gabapentin) [144]. LAT-1 is found in the brain, testis, and placenta, and is highly abundant in the BBB where it is responsible for the transport of key metabolites and drugs into the CNS [145-147]. LAT-1 is also highly upregulated in a variety of cancerous tumors, such as non-small cell

lung cancer and glioblastoma multiforme (GBM) [148, 149], for which it is thought to provide essential amino acids that are used as nutrients and signaling molecules for proliferation [150, 151]. LAT-1 is inactive in isolation and becomes active upon binding to the single helix glycoprotein SLC3A2 [152]. The putative transmembrane domain (12 predicted TMHs) of LAT-1 is similar in sequence to prokaryotic transporters of the acid/polyamine/organo-cation transporter (APC) family that adopt a LeuT-like fold [70, 153]. Therefore, LAT-1 was modeled based on the structures of the outward-occluded arginine-bound arginine/arginine transporter AdiC from *E. coli* [85] and the inward-apo conformation of the amino acid, polyamine, and organo-cation transporter ApcT from *Methanococcus jannaschii* [64, 153].

The model of LAT-1 complex with its natural ligands provides structural basis for the similarities and differences in ligand binding profiles among the amino acid transporters, such as LAT-1 and AdiC. In particular, backbone atoms in residues interacting with the carboxy- and amino-group of the amino acid ligands are highly conserved between the LAT-1 model and the AdiC template structure (eg, Thr62, Ile63, and Ile64 in LAT-1 correspond to Ala22, Ile23, and Met24 in AdiC). Conversely, two key differences in their binding sites can explain why their ligands are amino acid ligands with different physicochemical properties (large and hydrophobic amino acids for LAT-1 versus small and polar amino acids for AdiC). First, larger residues in the AdiC binding site are replaced with smaller residues in LAT-1, making LAT-1 binding site much larger (eg, Val148, Gly255, and Ser342 in LAT-1 correspond to Met104, Ile205, and Trp293 in AdiC). Second, polar residues in AdiC correspond to residues with aromatic and hydrophobic side-chains in LAT-1 (eg, Thr361 in AdiC corresponds to Trp405 in LAT-1) that are capable of forming van der Waals interactions and a hydrophobic effect with ligands.

We identified four small-molecule ligands for LAT-1 (ie, 3,5-diiodo-L-tyrosine, 3-iodo-L-tyrosine, fenclonine, and acivicin) by virtual screening of 19,166 endogenous metabolites and prescription drugs from the KEGG database, followed by kinetic uptake experiments of the 12 top-scoring hits. Two of the four ligands (ie, the anti-cancer agent acivicin and the tryptophan hydroxylase inhibitor fenclonine) were also confirmed as substrates by a trans-stimulation assay; in the trans-stimulation assay, cells overexpressing the transporter exchange the intracellular known radiolabeled substrate with the tested extracellular molecule, but only if the tested molecule is a substrate [85, 153]. These results rationalize some of the positive and negative pharmacological effects of these ligands. For example, acivicin failed in clinical trials for cancer therapy due to CNS-related toxic side effects that resulted from active transport [154]. Our results suggest that such toxicities were facilitated by LAT-1 active transport of acivicin through the BBB.

A potential drug that is a ligand of a cancer-upregulated exchange or import transporter can act via one or both of the following mechanisms. First, a substrate “hijacking” the transporter to deliver a drug against a target in the cell (eg, a metabolic enzyme). Second, a transporter inhibitor that selectively blocks transport activity denying the cancer cell key nutrients (eg, the glucose transporter (GLUT1) and LAT-1). Both of the newly discovered LAT-1 ligands, 3-iodo-L-tyrosine and acivicin, are capable of inhibiting cancer cell proliferation in a cell proliferation assay of a GBM cell line, which naturally expresses LAT-1 at high levels. This finding suggests that the two newly discovered LAT-1 ligands inhibited the GBM cell line proliferation via both mechanisms, including nutrient deprivation by 3-iodo-L-tyrosine (mechanism 1) and cytotoxicity by acivicin (mechanism 2). Future structure-based ligand discovery studies will be applied simultaneously to LAT-1 and other proteins that might function together in cancer (eg GLUT1)

or BBB permeability (P-gp). For instance, a comparative model of another key BBB transporter, the ATP-dependent efflux transporter P-glycoprotein (P-gp), was constructed based on the structure of its mouse homolog and then used, via molecular docking, to predict drugs that may get exported from the BBB [155].

CASE STUDY 5: THE MULTIDRUG AND TOXIN EXTRUSION TRANSPORTER

(MATE, SLC47) FAMILY

The multidrug and toxin extrusion transporter (MATE) family of proteins consists of two members, MATE1 and MATE2 [48, 156]. MATE proteins contribute to the excretion of diverse organic cations through the membranes of various cell types by exchanging their substrates with an oppositely directed proton gradient [156, 157]. In the human kidney, MATE1 and MATE2-K (a splice variant of MATE2) are localized to the apical membrane of the proximal and distal convoluted tubule [156]. The MATEs have ligands similar to those of key SLC22 transporters (OCT1-3), and together these transporters contribute to the tissue distribution and excretion of many drugs (eg, metformin) (Fig. 2-5A). Several genetic variants of MATE1 have been associated with differential response to metformin among diabetic patients [158, 159].

Different methodologies, such as classical experimental procedures to pharmacophore and QSAR modeling, have been previously employed to find common and selective substrates and inhibitors of MATE and OCT proteins [121, 160-162]. We recently identified 84 MATE1 ligands using iterative ligand-based computational modeling and experimental testing by the cis-inhibition experiments. Particularly, we first performed an experimental high-throughput screen (HTS) for 910 compounds from 124 therapeutic classes, including antidepressants, antibacterials, antivirals, and antihypertensives [163]. This experimental data was then used to develop a computational QSAR model of ligand binding, using the Random Forest algorithm. The QSAR

model performed significantly better in identifying true ligands than random selection (as judged by the area under the receiver operating curve (AUC) of 0.7, p-value < 0.0001). Next, selected positive and negative predictions were validated by low-throughput experiments focused on a single tested compound at a time. For example, eight inhibitors (eg, a histamine H2 receptor antagonist ranitidine) did not show inhibition in the original high-throughput screen, but were predicted as inhibitors by the QSAR model, and were later confirmed by low-throughput experiments. Thus, the QSAR model was used to save false negatives from the high-throughput screen. Likewise, several false positives in the high-throughput screen, such as the painkiller phenacetin and the vitamin niacin, were also corrected using the computational model and additional focused testing. We then refined the QSAR model based on the additional compounds identified by the low-throughput experiments, significantly improving its performance (ie, AUC=0.78). Next, the refined QSAR model was used to screen 6,122 compounds from the DrugBank library [164]. Finally, five of the eight chemically novel predicted ligands that were selected for experimental testing were confirmed experimentally as MATE1 inhibitor; for example, maraviroc, an antiretroviral drug in the CCR5 receptor antagonist class that is used in the treatment of HIV inhibited MATE1 (IC50 of 17.3 μ M).

With the aid of the QSAR model, we also identified several features that are important for MATE1 inhibition. MATE1 compounds are large and bulky as evidenced by a higher molecular weight, a higher number of bonds, and a higher number of rings among MATE1 inhibitors versus non-inhibitors. We found that lipophilicity is the key determinant of the MATE1 ligands. Interestingly, while charge is clearly important for the MATE1 ligands, about half of the inhibitors are acids or uncharged at the physiological pH.

The MATEs are predicted to have 13 TMHs and are related to the transporter NorM from *Vibrio cholerae*, whose structure was determined at atomic resolution and exhibits a 12-TMH topology [165]. A recent comparative model of the rabbit MATE1 based on the NorM structure, followed by experimental validation, confirmed that a ‘core’ domain containing the N-terminal 12 TMHs is responsible for transport and suggested that TMH 13 likely influences transporter turnover [166, 167]. We constructed a comparative model of MATE-1 that consists of a large pore with negative electrostatic potential and hydrophobic patches. This preliminary comparative model rationalizes the results obtained with our ligand-based QSAR models. For example, the negative electrostatic potential on the surface of MATE1 pore complements the cationic nature of the MATE1 ligands (Fig. 2-5B).

MATE-2K was recently modeled based on the NorM structure, followed by rationalizing the effect of four non-synonymous protein coding single nucleotide polymorphisms (nsSNPs), including Lys64Asn, Pro162Leu, Gly211Val, and Gly393Arg [168, 169]. Particularly, Lys64 is predicted to be in the extracellular loop near the membrane-spanning residues; thus, mutation of the positively charged lysine to an uncharged asparagine residue might have an effect on the interaction of MATE-2K with the membrane phospholipid head groups or on posttranslational modifications; Pro162 and Gly211 are located in TMH 4 and 6, respectively. Therefore, mutations of these residues to leucine and valine, respectively, might reduce the main chain rigidity (Pro) or flexibility (Gly) as well as disrupt helix–helix packing. Finally, Gly393 is predicted to be located in the channel close to the putative proton-binding site. Mutation of Gly393 to a positively charged residue might, therefore, perturb the interaction of the transporter with the cationic ligands and protons.

CONCLUSIONS AND FUTURE OUTLOOK

In this review, we first outlined the cellular and pharmacological functions of the human SLC transporters, as well as their sequences, structures and mechanisms of transport (Fig. 2-1). We then described various sequence- and structure-based tools that are used, in combination with experiment, to characterize interactions between the SLC transporters and their small-molecule ligands (Fig. 2-2). We illustrated the utility of computational strategies, including structure-based (NET, GAT-2, and LAT-1) and ligand-based (MATE-2) modeling, to predict ligands for biomedically important SLC transporters (Figs. 2-3, 2-4, 2-5). Finally, we demonstrated how a comparative model can be used to rationalize the effect of nsSNP on pharmacogenetics of the antidiabetic drug metformin, a substrate for the kidney and liver drug transporter MATE-2. We now present our outlook on future directions in the field of structure-based ligand discovery for human SLC transporters.

Virtual screening against comparative models of different transporter conformations was found useful for identifying chemically novel transporter ligands (Fig. 2-3) [114]. To identify additional unexplored classes of molecules targeting SLC transporters, virtual screening should be performed against structures representing additional states in the conformational cycle of the transporter, characterized experimentally or computationally. Applications of structural symmetry [42] and MD simulations [87] have been helpful to model experimentally undefined conformations. However, because SLC transporters are exceptionally diverse in their structures and mechanisms (including transport of ligands on widely different time scales), new approaches to characterizing their dynamics with sufficient accuracy for structure-based ligand discovery are still needed.

The applicability of structure-based ligand discovery tools to human SLC transporters has been steadily increasing, for two reasons. First, recent experimental determinations of high-resolution transporter structures have been instrumental in extending the structural coverage of human SLC transporters via comparative modeling. The number of membrane transporter structures representing uncharacterized conformations and unexplored families is expected to grow considerably in the next few years, both as a result of traditional structural biology and structural genomics efforts [170, 171]. Second, the accuracy and automation of computational tools for structure-based ligand discovery is improving in general [109, 110, 113, 114, 172, 173].

It is now appreciated that SLC transporters function in a coordinated manner with each other as well as other proteins, including receptors and enzymes, in various major biological and pharmacological processes. For example, several SLC members directly mediate drug-drug interactions in the liver, kidney, and BBB, which often affect clinical outcome [53, 174]. This led the FDA to strongly recommend that drug candidates should be tested for drug interactions with seven drug transporters, five of which are SLC members (eg, OAT1) [53]; this number is expected to grow [174]. Furthermore, SLC transporters such as LAT-1 and GLUT1 can function cooperatively in reprogrammed cancer metabolic pathways, by providing nutrients and signaling molecules to the transforming cancer cells; they may thus be reasonable targets for cancer therapy [150, 151].

The increasing applicability of structure-based ligand discovery enables us to characterize interactions between small molecules and multiple protein targets that make up a 'system'. Such systems can be a biological pathway associated with cancer (eg, the mTOR pathway) or an organ important for drug clearance (eg, kidney). Therefore, future discovery efforts should aim at describing interactions of small molecules with multiple proteins as well as appropriately

integrating these interactions with additional ‘omics’ data, such as protein-protein interactions and gene expression data, to better describe these systems and to ultimately relate them to clinical observations.

FIGURES

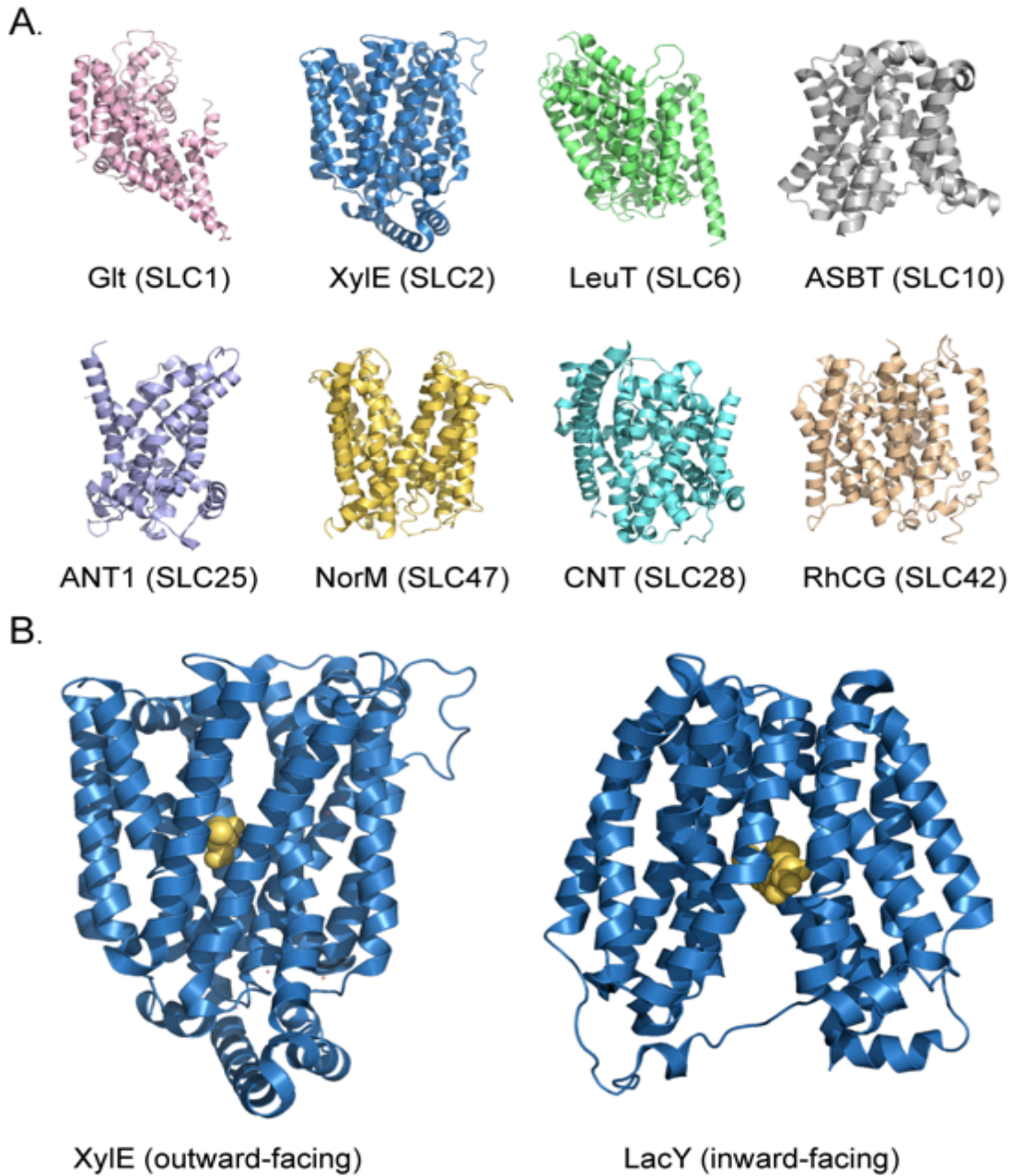


Fig. 2-1. Structures of SLC transporters.

(A) Structures of SLC transporters representing select known structural classes and related human families. The structures are of the aspartate transporter Glt [175], the xylose transporter XylE [90], the leucine transporter LeuT [176], the sodium / bile acid symporter ASBT [177], the

ADP/ATP translocase 1 ANT1 [178], the drug and toxin transporter NorM [165], the concentrative nucleoside transporter CNT [179], and the human rhesus glycoprotein RhCG ammonium transporter [180]. (B) The alternating access mechanism for structures with the MFS fold. The structures are of XylE in a ligand-bound outward-facing conformation (left) and the lactose permease LacY in a ligand-bound inward-facing conformation [91] (right). All structures are visualized with PyMol [181].

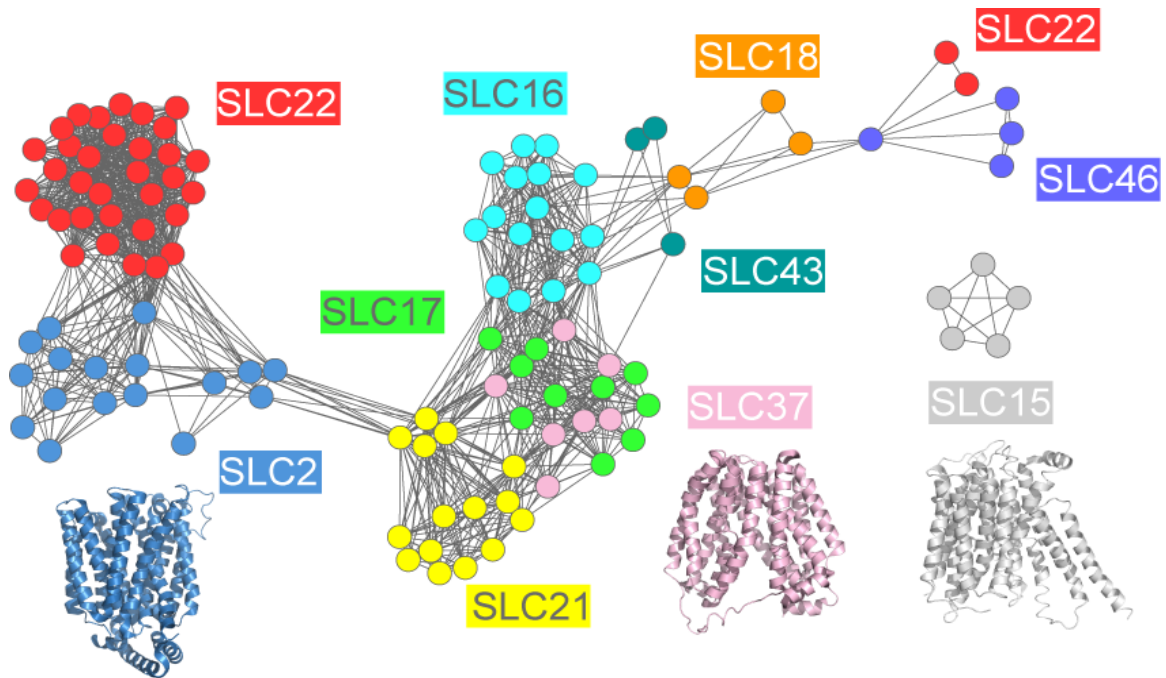


Fig. 2-2. Sequence similarity network of human SLC families predicted to have the MFS fold.

(A) The relationships between SLC sequences are visualized using Cytoscape 2.6.1.94 [88, 182].

(A) The nodes represent SLC sequences, including splice variants, that are similar to each other or to sequences of proteins with known MFS structures; the colors of the nodes indicate the SLC family [37]. The edges between the nodes correspond to a pairwise alignment with sequence identity of at least 10% and an E-value of less than 1 [88, 183]. The structures of Xyle (blue) [90], LacY (pink) [91], and PEPTSO (grey) [92] are visualized using PyMol [181] and their colors are based on the color of the most similar human family in the similarity network.

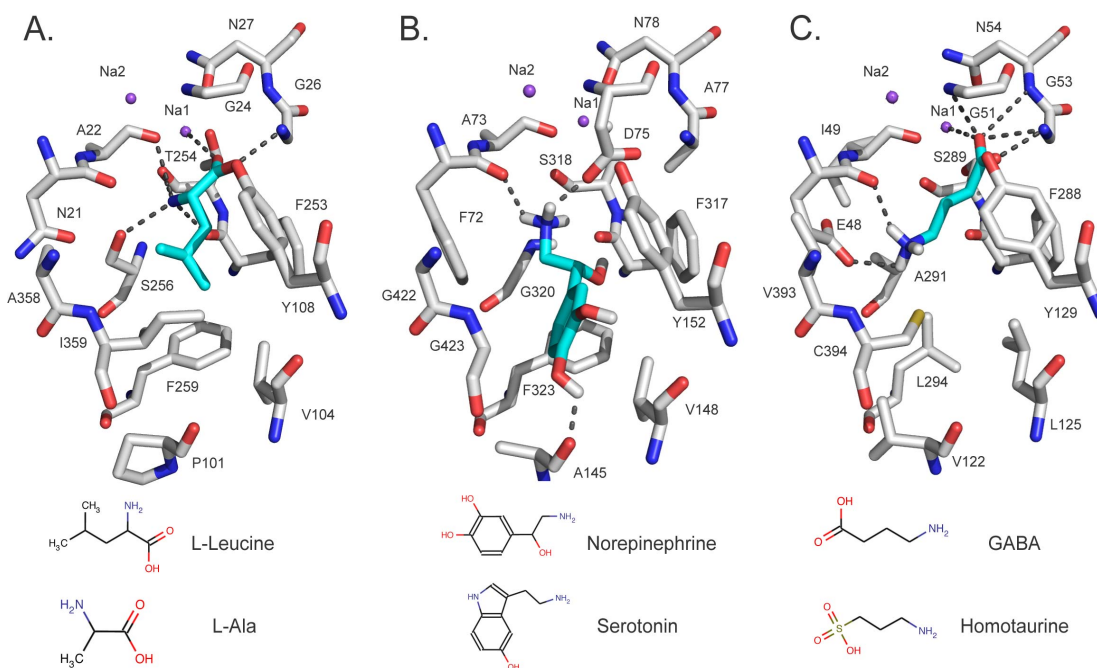


Fig. 2-3. Binding sites and modes of ligand binding in SLC6 members.

The X-ray structure of LeuT as well as the comparative models of (B) NET and (C) GAT-2, in the ligand-bound occluded conformation, are visualized with PyMol [181]. Atoms are displayed as sticks, with oxygen, nitrogen, and hydrogen atoms in red, blue, and white, respectively. The sodium ions Na1 and Na2 are visualized with purple spheres. The ligands L-Leucine, norepinephrine, and GABA are illustrated in cyan sticks and their hydrogen bonds with the binding-site residues of LeuT (Ala22, Gly26, Thr254, Ser256, and Na1), GAT-2 (Glu48, Gly51, Gly53, Asn54, and Na1), and NET (Ala145, Phe72, and Asp75) are displayed as dotted gray lines. Sketches of two representative ligands of each transporter are shown at the bottom.

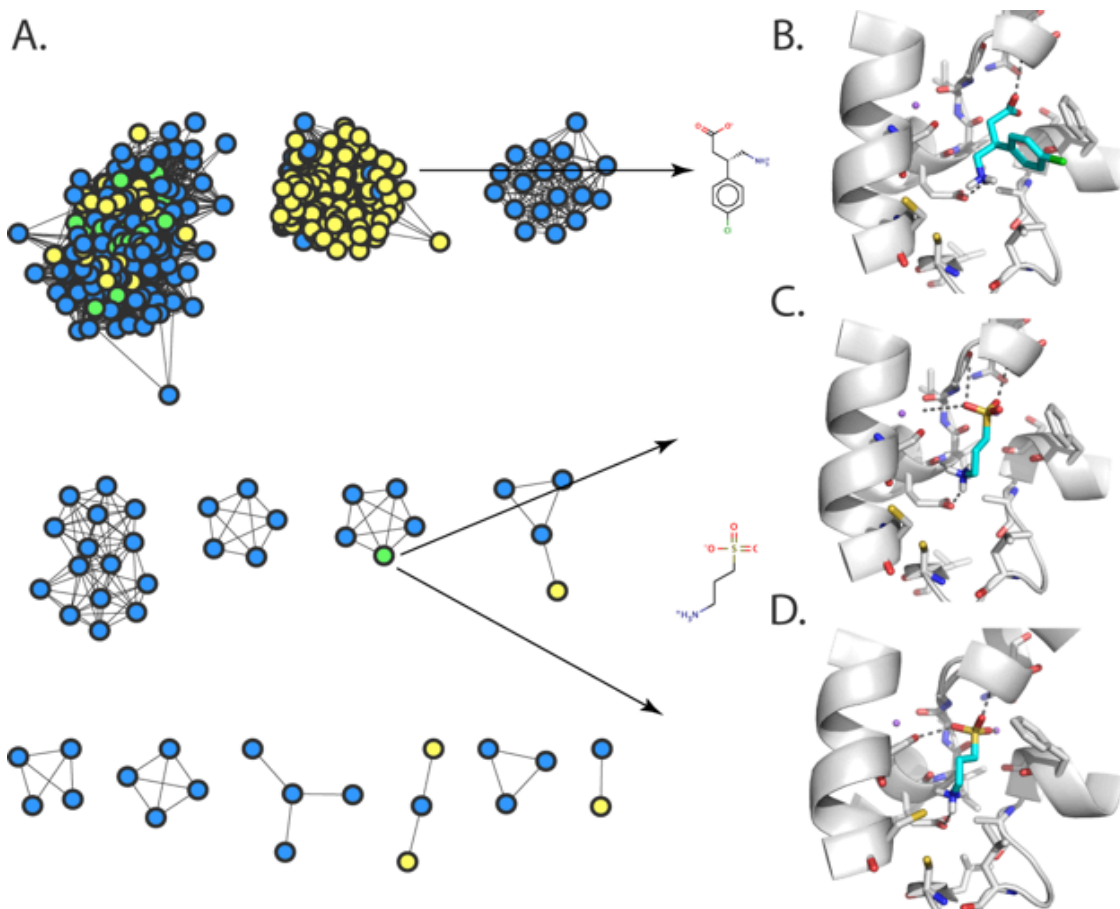


Fig. 2-4. Structure-based ligand prediction for GAT-2.

(A) A network view of predicted GAT-2 ligand drugs and their similarities. The nodes represent top small molecule hits predicted to bind GAT-2, using the occluded model (blue), the outward-facing model (yellow), or both models (green). Predicted structures of GAT-2 in the outward-facing (B and C) and occluded (D) conformations, in complex with the representative experimentally confirmed hits baclofen (B) and homotaurine (C and D). Small molecules ligands are colored in cyan, with oxygen, nitrogen, sulfur, and hydrogen atoms in red, blue, yellow, and white, respectively. The sodium ions Na1 and Na2 are visualized as purple spheres. The GAT-2 TMH regions are illustrated in white ribbons. Important residues for binding are depicted as sticks; predicted hydrogen bonds between ligands and GAT-2 are displayed as dotted gray lines.

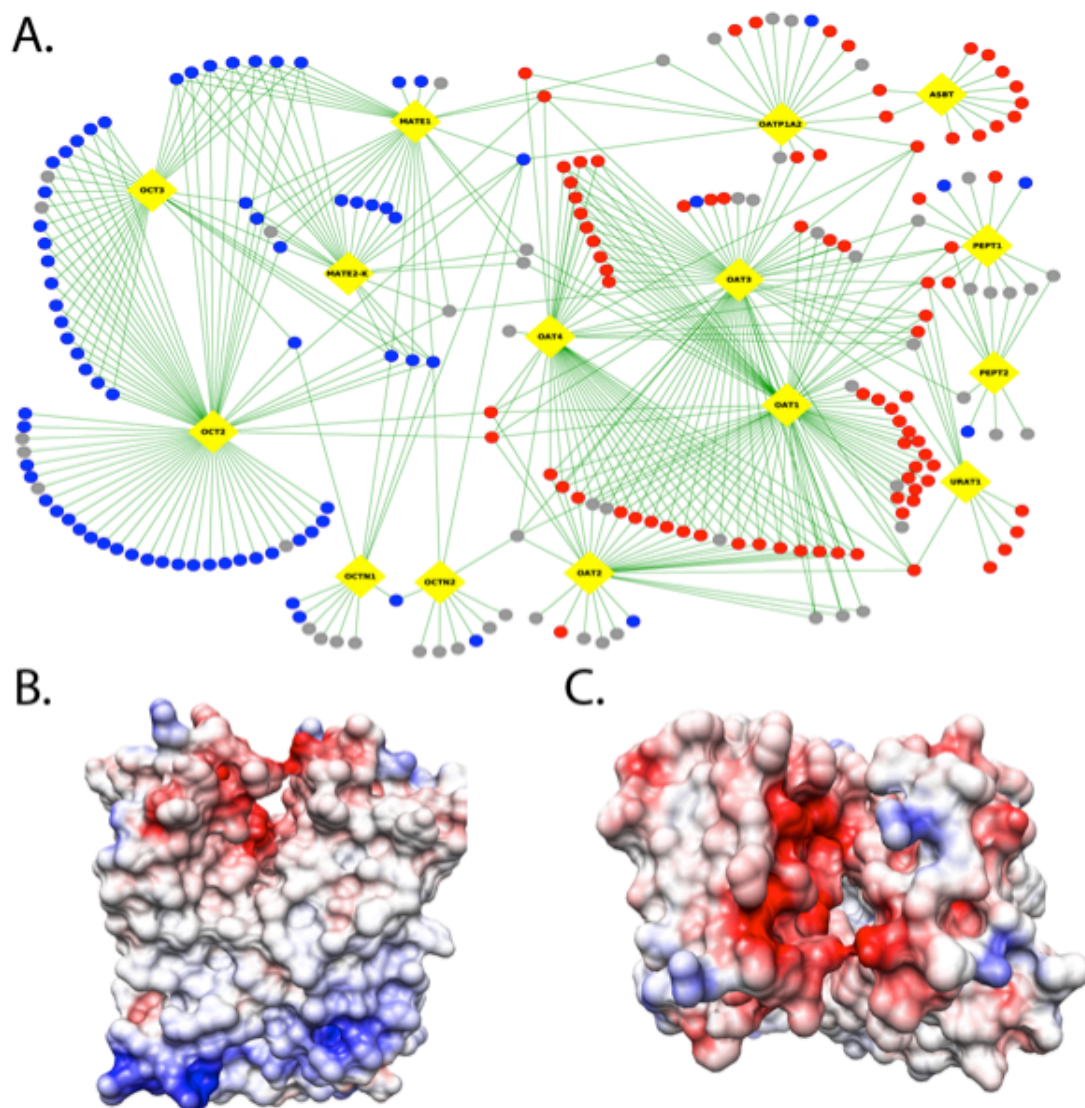


Fig. 2-5. Ligand discovery for MATE-1.

(A) An interaction network for prescription drugs and SLC transporters in kidney. SLC transporters are depicted as yellow diamonds. Positively charged drugs are shown as solid blue circles, negatively charged drugs as solid red circles, and uncharged drugs as gray circles. Edges between transporters and drugs correspond to known transporter-drug interactions. (B) A preliminary human MATE-1 comparative model based on the structure of its prokaryotic homolog NorM. The model is visualized using the Coulombic Surface Coloring in UCSF

Chimera [184]. Negative and positive electrostatic potentials are illustrated in red and blue, respectively.

REFERENCES

1. Hediger MA, et al. (2004) The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins Introduction. *Pflugers Arch* 447(5):465-468.
2. Povey S, et al. (2001) The HUGO Gene Nomenclature Committee (HGNC). *Hum Genet* 109(6):678-680.
3. Saier MH, Jr. (2000) A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol Mol Biol Rev* 64(2):354-411.
4. Saier MH, Jr., Yen MR, Noto K, Tamang DG, & Elkan C (2009) The Transporter Classification Database: recent advances. *Nucleic Acids Res* 37(Database issue):D274-278.
5. Forrest LR & Rudnick G (2009) The rocking bundle: a mechanism for ion-coupled solute flux by symmetrical transporters. *Physiology (Bethesda)* 24:377-386.
6. Forrest LR, Kramer R, & Ziegler C (2011) The structural basis of secondary active transport mechanisms. (Translated from eng) *Biochimica et biophysica acta* 1807(2):167-188 (in eng).
7. Jardetzky O (1966) Simple allosteric model for membrane pumps. *Nature* 211(5052):969-970.
8. Gether U, Andersen PH, Larsson OM, & Schousboe A (2006) Neurotransmitter transporters: molecular function of important drug targets. *Trends Pharmacol Sci* 27(7):375-383.
9. Blakely RD & DeFelice LJ (2007) All aglow about presynaptic receptor regulation of neurotransmitter transporters. *Mol Pharmacol* 71(5):1206-1208.
10. Chen NH, Reith ME, & Quick MW (2004) Synaptic uptake and beyond: the sodium- and chloride-dependent neurotransmitter transporter family SLC6. *Pflugers Arch* 447(5):519-531.
11. You G, Morris, M. E., Wang, B. (2007) *Drug Transporters: Molecular Characterization and Role in Drug Disposition* (Wiley, Hoboken) p 897.
12. Otsuka M, et al. (2005) A human transporter protein that mediates the final excretion step for toxic organic cations. (Translated from eng) *Proc Natl Acad Sci U S A* 102(50):17923-17928 (in eng).
13. Meyer zu Schwabedissen HE, Verstuyft C, Kroemer HK, Becquemont L, & Kim RB (2010) Human multidrug and toxin extrusion 1 (MATE1/SLC47A1) transporter: functional characterization, interaction with OCT2 (SLC22A2), and single nucleotide polymorphisms. (Translated from eng) *Am J Physiol Renal Physiol* 298(4):F997-F1005 (in eng).
14. Leabman MK, et al. (2003) Natural variation in human membrane transporter genes reveals evolutionary and functional constraints. *Proc Natl Acad Sci U S A* 100(10):5896-5901.

15. Koepsell H, Schmitt BM, & Gorboulev V (2003) Organic cation transporters. *Rev Physiol Biochem Pharmacol* 150:36-90.
16. Cropp CD, Yee SW, & Giacomini KM (2008) Genetic variation in drug transporters in ethnic populations. (Translated from eng) *Clin Pharmacol Ther* 84(3):412-416 (in eng).
17. Giacomini KM, et al. (2010) Membrane transporters in drug development. *Nature reviews. Drug discovery* 9(3):215-236.
18. Shu Y, et al. (2008) Effect of genetic variation in the organic cation transporter 1, OCT1, on metformin pharmacokinetics. *Clin Pharmacol Ther* 83(2):273-280.
19. Shu Y, et al. (2007) Effect of genetic variation in the organic cation transporter 1 (OCT1) on metformin action. *J Clin Invest* 117(5):1422-1431.
20. Shu Y, et al. (2003) Evolutionary conservation predicts function of variants of the human organic cation transporter, OCT1. *Proc Natl Acad Sci U S A* 100(10):5902-5907.
21. Dresser MJ, Gray AT, & Giacomini KM (2000) Kinetic and selectivity differences between rodent, rabbit, and human organic cation transporters (OCT1). *J Pharmacol Exp Ther* 292(3):1146-1152.
22. Schlessinger A, et al. (2011) Structure-based discovery of prescription drugs that interact with the norepinephrine transporter, NET. (Translated from eng) *Proc Natl Acad Sci U S A* 108(38):15810-15815 (in eng).
23. Choi JH, et al. (2011) A Common 5'-UTR Variant in MATE2-K Is Associated With Poor Response to Metformin. (Translated from Eng) *Clin Pharmacol Ther* (in Eng).
24. Fredriksson R, Nordstrom KJ, Stephansson O, Hagglund MG, & Schiöth HB (2008) The solute carrier (SLC) complement of the human genome: phylogenetic classification reveals four major families. *FEBS Lett* 582(27):3811-3816.
25. Gruswitz F, Chaudhary, S., Ho, J. D., Schlessinger, A., Pezeshki, B., Ho, C. M., Sali, A., Westhoff, C. M., Stroud, R.M. (2009) Renal Ammonia Transport and the Structure of the Human Rhesus Glycoprotein RhCG at 2.1 Å. submitted.
26. Karpowich NK & Wang DN (2008) Structural biology. Symmetric transporters for asymmetric transport. *Science* 321(5890):781-782.
27. Forrest LR, et al. (2008) Mechanism for alternating access in neurotransmitter transporters. *Proc Natl Acad Sci U S A* 105(30):10338-10343.
28. Gao X, et al. (2009) Structure and mechanism of an amino acid antiporter. *Science* 324(5934):1565-1568.
29. Fang Y, et al. (2009) Structure of a prokaryotic virtual proton pump at 3.2 Å resolution. *Nature* 460(7258):1040-1043.
30. Shaffer PL, Goehring A, Shankaranarayanan A, & Gouaux E (2009) Structure and mechanism of a Na⁺-independent amino acid transporter. *Science* 325(5943):1010-1014.
31. Lu F, et al. (2011) Structure and mechanism of the uracil transporter UraA. (Translated from eng) *Nature* 472(7342):243-246 (in eng).

32. Weyand S, et al. (2008) Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* 322(5902):709-713.
33. Ressler S, Terwisscha van Scheltinga AC, Vonnrhein C, Ott V, & Ziegler C (2009) Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP. *Nature* 458(7234):47-52.
34. Ma D, et al. (2012) Structure and mechanism of a glutamate-GABA antiporter. (Translated from eng) *Nature* 483(7391):632-636 (in eng).
35. Schulze S, Koster S, Geldmacher U, Terwisscha van Scheltinga AC, & Kuhlbrandt W (2010) Structural basis of Na(+)-independent and cooperative substrate/product antiport in CaiT. (Translated from eng) *Nature* 467(7312):233-236 (in eng).
36. Lomize MA, Lomize AL, Pogozheva ID, & Mosberg HI (2006) OPM: orientations of proteins in membranes database. *Bioinformatics* 22(5):623-625.
37. Krishnamurthy H, Piscitelli CL, & Gouaux E (2009) Unlocking the molecular secrets of sodium-coupled transporters. *Nature* 459(7245):347-355.
38. Abramson J & Wright EM (2009) Structure and function of Na(+)-symporters with inverted repeats. *Curr Opin Struct Biol* 19(4):425-432.
39. Khafizov K, Staritzbichler R, Stamm M, & Forrest LR (2010) A study of the evolution of inverted-topology repeats from LeuT-fold transporters using AlignMe. (Translated from eng) *Biochemistry* 49(50):10702-10713 (in eng).
40. Kanner BI & Zomot E (2008) Sodium-coupled neurotransmitter transporters. *Chem Rev* 108(5):1654-1668.
41. Guan L & Kaback HR (2006) Lessons from lactose permease. *Annu Rev Biophys Biomol Struct* 35:67-91.
42. Krishnamurthy H & Gouaux E (2012) X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. (Translated from eng) *Nature* 481(7382):469-474 (in eng).
43. Nyola A, et al. (2010) Substrate and drug binding sites in LeuT. *Curr Opin Struct Biol* 20(4):415-422.
44. Singh SK, Piscitelli CL, Yamashita A, & Gouaux E (2008) A competitive inhibitor traps LeuT in an open-to-out conformation. *Science* 322(5908):1655-1661.
45. Zhao Y, et al. (2011) Substrate-modulated gating dynamics in a Na(+)-coupled neurotransmitter transporter homologue. *Nature*.
46. Zhao Y, et al. (2010) Single-molecule dynamics of gating in a neurotransmitter transporter homologue. *Nature* 465(7295):188-193.
47. Claxton DP, et al. (2010) Ion/substrate-dependent conformational dynamics of a bacterial homolog of neurotransmitter:sodium symporters. *Nature structural & molecular biology* 17(7):822-829.

48. Shi L, Quick M, Zhao Y, Weinstein H, & Javitch JA (2008) The mechanism of a neurotransmitter:sodium symporter--inward release of Na⁺ and substrate is triggered by substrate in a second binding site. *Mol Cell* 30(6):667-677.
49. Shimamura T, et al. (2010) Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. (Translated from eng) *Science* 328(5977):470-473 (in eng).
50. Gao X, et al. (2010) Mechanism of substrate recognition and transport by an amino acid antiporter. (Translated from eng) *Nature* 463(7282):828-832 (in eng).
51. Radestock S & Forrest LR (2011) The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. (Translated from eng) *Journal of molecular biology* 407(5):698-715 (in eng).
52. Faraldo-Gomez JD & Forrest LR (2011) Modeling and simulation of ion-coupled and ATP-driven membrane proteins. (Translated from eng) *Curr Opin Struct Biol* 21(2):173-179 (in eng).
53. Nyola A, et al. (2010) Substrate and drug binding sites in LeuT. *Curr Opin Struct Biol* 20(4):415-422.
54. Punta M, et al. (2007) Membrane protein prediction methods. (Translated from eng) *Methods* 41(4):460-474 (in eng).
55. Nugent T & Jones DT (2012) Membrane protein structural bioinformatics. (Translated from eng) *J Struct Biol* 179(3):327-337 (in eng).
56. Nugent T & Jones DT (2009) Transmembrane protein topology prediction using support vector machines. (Translated from eng) *BMC Bioinformatics* 10:159 (in eng).
57. Soding J, Biegert A, & Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. (Translated from eng) *Nucleic acids research* 33(Web Server issue):W244-248 (in eng).
58. Pei J, Kim BH, & Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. (Translated from eng) *Nucleic acids research* 36(7):2295-2300 (in eng).
59. Albers T, Marsiglia W, Thomas T, Gameiro A, & Grewer C (2012) Defining substrate and blocker activity of alanine-serine-cysteine transporter 2 (ASCT2) Ligands with Novel Serine Analogs. (Translated from eng) *Molecular pharmacology* 81(3):356-365 (in eng).
60. Beuming T, Shi L, Javitch JA, & Weinstein H (2006) A comprehensive structure-based alignment of prokaryotic and eukaryotic neurotransmitter/Na⁺ symporters (NSS) aids in the use of the LeuT structure to probe NSS structure and function. *Mol Pharmacol* 70(5):1630-1642.
61. Atkinson HJ, Morris JH, Ferrin TE, & Babbitt PC (2009) Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS ONE* 4(2):e4345.

62. Brown SD & Babbitt PC (2012) Inference of functional properties from large-scale analysis of enzyme superfamilies. (Translated from eng) *The Journal of biological chemistry* 287(1):35-42 (in eng).
63. Schlessinger A, et al. (2010) Comparison of human solute carriers. *Protein science : a publication of the Protein Society* 19(3):412-428.
64. Rost B, Liu J, Nair R, Wrzeszczynski KO, & Ofran Y (2003) Automatic prediction of protein function. (Translated from eng) *Cell Mol Life Sci* 60(12):2637-2650 (in eng).
65. Sun L, et al. (2012) Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. (Translated from eng) *Nature* 490(7420):361-366 (in eng).
66. Abramson J, et al. (2003) Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* 301(5633):610-615.
67. Newstead S, et al. (2011) Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. (Translated from eng) *Embo J* 30(2):417-426 (in eng).
68. Shoichet BK, McGovern SL, Wei B, & Irwin JJ (2002) Lead discovery using molecular docking. *Curr Opin Chem Biol* 6(4):439-446.
69. Shoichet BK (2004) Virtual screening of chemical libraries. *Nature* 432(7019):862-865.
70. Shoichet BK, Stroud RM, Santi DV, Kuntz ID, & Perry KM (1993) Structure-based discovery of inhibitors of thymidylate synthase. *Science* 259(5100):1445-1450.
71. Mysinger MM & Shoichet BK (2010) Rapid context-dependent ligand desolvation in molecular docking. *Journal of chemical information and modeling* 50(9):1561-1573.
72. Kuntz ID (1992) Structure-based strategies for drug design and discovery. *Science* 257(5073):1078-1082.
73. Jacobson M & Sali A (2004) Comparative Protein Structure Modeling and its Applications to Drug Discovery. *Annual Reports in Medicinal Chemistry*, (Academic Press), Vol Volume 39, pp 259-276.
74. Huang N, Shoichet BK, & Irwin JJ (2006) Benchmarking sets for molecular docking. *J Med Chem* 49(23):6789-6801.
75. Fan H, et al. (2009) Molecular docking screens using comparative models of proteins. *J Chem Inf Model* 49(11):2512-2527.
76. Evers A, Gohlke H, & Klebe G (2003) Ligand-supported homology modelling of protein binding-sites using knowledge-based potentials. *Journal of molecular biology* 334(2):327-345.
77. Cavasotto CN, et al. (2008) Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. *Journal of medicinal chemistry* 51(3):581-588.
78. Carlsson J, et al. (2011) Ligand discovery from a dopamine D3 receptor homology model and crystal structure. (Translated from eng) *Nature chemical biology* 7(11):769-778 (in eng).

79. Schlessinger A, et al. (2012) High Selectivity of the gamma-Aminobutyric Acid Transporter 2 (GAT-2, SLC6A13) Revealed by Structure-based Approach. (Translated from eng) *The Journal of biological chemistry* 287(45):37745-37756 (in eng).
80. Shoichet BK & Kobilka BK (2012) Structure-based drug screening for G-protein-coupled receptors. (Translated from eng) *Trends in pharmacological sciences* 33(5):268-272 (in eng).
81. Kufareva I, Rueda M, Katritch V, Stevens RC, & Abagyan R (2011) Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. (Translated from eng) *Structure* 19(8):1108-1126 (in eng).
82. Schapira M, Abagyan R, & Totrov M (2003) Nuclear hormone receptor targeted virtual screening. (Translated from eng) *Journal of medicinal chemistry* 46(14):3045-3059 (in eng).
83. Diller DJ & Li R (2003) Kinases, homology models, and high throughput docking. (Translated from eng) *Journal of medicinal chemistry* 46(22):4638-4647 (in eng).
84. Song L, et al. (2007) Prediction and assignment of function for a divergent N-succinyl amino acid racemase. (Translated from eng) *Nature chemical biology* 3(8):486-491 (in eng).
85. Kalyanaraman C, et al. (2008) Discovery of a dipeptide epimerase enzymatic function guided by homology modeling and virtual screening. (Translated from eng) *Structure* 16(11):1668-1677 (in eng).
86. Astorga B, Ekins S, Morales M, & Wright SH (2012) Molecular determinants of ligand selectivity for the human multidrug and toxin extruder proteins MATE1 and MATE2-K. (Translated from eng) *The Journal of pharmacology and experimental therapeutics* 341(3):743-755 (in eng).
87. Andersen J, Kristensen AS, Bang-Andersen B, & Stromgaard K (2009) Recent advances in the understanding of the interaction of antidepressant drugs with serotonin and norepinephrine transporters. *Chem Commun (Camb)* (25):3677-3692.
88. Ekins S, Ecker GF, Chiba P, & Swaan PW (2007) Future directions for drug transporter modelling. (Translated from eng) *Xenobiotica* 37(10-11):1152-1170 (in eng).
89. Chang C & Swaan PW (2006) Computational approaches to modeling drug transporters. (Translated from eng) *Eur J Pharm Sci* 27(5):411-424 (in eng).
90. Eyre TA, Partridge L, & Thornton JM (2004) Computational analysis of alpha-helical membrane protein structure: implications for the prediction of 3D structural models. (Translated from eng) *Protein Eng Des Sel* 17(8):613-624 (in eng).
91. Hahn MK & Blakely RD (2007) The functional impact of SLC6 transporter genetic variation. *Annu Rev Pharmacol Toxicol* 47:401-441.
92. Pacholczyk T, Blakely RD, & Amara SG (1991) Expression cloning of a cocaine- and antidepressant-sensitive human noradrenaline transporter. *Nature* 350(6316):350-354.
93. Celik L, et al. (2008) Binding of serotonin to the human serotonin transporter. Molecular modeling and experimental validation. *J Am Chem Soc* 130(12):3853-3865.

94. Beuming T, et al. (2008) The binding sites for cocaine and dopamine in the dopamine transporter overlap. *Nat Neurosci* 11(7):780-789.
95. Severinsen K, et al. (2012) Binding of the Amphetamine-like 1-Phenyl-piperazine to Monoamine Transporters. (Translated from eng) *ACS Chem Neurosci* 3(9):693-705 (in eng).
96. Anonymous (Daylight Chemical Information Systems, Inc.Laguna Niguel, CA 92677).
97. Adamus WS, Leonard JP, & Troger W (1995) Phase I clinical trials with WAL 2014, a new muscarinic agonist for the treatment of Alzheimer's disease. *Life sciences* 56(11-12):883-890.
98. Madsen KK, White HS, & Schousboe A (2010) Neuronal and non-neuronal GABA transporters as targets for antiepileptic drugs. (Translated from eng) *Pharmacology & therapeutics* 125(3):394-401 (in eng).
99. Christiansen B, Meinild AK, Jensen AA, & Brauner-Osborne H (2007) Cloning and characterization of a functional human gamma-aminobutyric acid (GABA) transporter, human GAT-2. (Translated from eng) *The Journal of biological chemistry* 282(27):19331-19341 (in eng).
100. Nakashita M, Sasaki K, Sakai N, & Saito N (1997) Effects of tricyclic and tetracyclic antidepressants on the three subtypes of GABA transporter. (Translated from eng) *Neurosci Res* 29(1):87-91 (in eng).
101. Melamed N & Kanner BI (2004) Transmembrane domains I and II of the gamma-aminobutyric acid transporter GAT-4 contain molecular determinants of substrate specificity. (Translated from eng) *Molecular pharmacology* 65(6):1452-1461 (in eng).
102. Kanner BI (2003) Transmembrane domain I of the gamma-aminobutyric acid transporter GAT-1 plays a crucial role in the transition between cation leak and transport modes. (Translated from eng) *The Journal of biological chemistry* 278(6):3705-3712 (in eng).
103. Irwin JJ & Shoichet BK (2005) ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45(1):177-182.
104. Albers JW & Fink JK (2004) Porphyrinic neuropathy. (Translated from eng) *Muscle Nerve* 30(4):410-422 (in eng).
105. Lindberg RL, et al. (1999) Motor neuropathy in porphobilinogen deaminase-deficient mice imitates the peripheral neuropathy of human acute porphyria. (Translated from eng) *J Clin Invest* 103(8):1127-1134 (in eng).
106. Musiol R, Serda M, & Polanski J (2011) Prodrugs in photodynamic anticancer therapy. (Translated from eng) *Curr Pharm Des* 17(32):3548-3559 (in eng).
107. Stummer W, et al. (2006) Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. (Translated from eng) *Lancet Oncol* 7(5):392-401 (in eng).
108. Fukuda H, Casas A, & Batlle A (2005) Aminolevulinic acid: from its unique biological function to its star role in photodynamic therapy. (Translated from eng) *Int J Biochem Cell Biol* 37(2):272-276 (in eng).

109. Kanai Y, et al. (1998) Expression cloning and characterization of a transporter for large neutral amino acids activated by the heavy chain of 4F2 antigen (CD98). (Translated from eng) *The Journal of biological chemistry* 273(37):23629-23632 (in eng).
110. Roberts LM, et al. (2008) Subcellular localization of transporters along the rat blood-brain barrier and blood-cerebral-spinal fluid barrier by in vivo biotinylation. (Translated from eng) *Neuroscience* 155(2):423-438 (in eng).
111. Alexander GM, Schwartzman RJ, Grothusen JR, & Gordon SW (1994) Effect of plasma levels of large neutral amino acids and degree of parkinsonism on the blood-to-brain transport of levodopa in naive and MPTP parkinsonian monkeys. (Translated from eng) *Neurology* 44(8):1491-1499 (in eng).
112. Wang Y & Welty DF (1996) The simultaneous estimation of the influx and efflux blood-brain barrier permeabilities of gabapentin using a microdialysis-pharmacokinetic approach. (Translated from eng) *Pharm Res* 13(3):398-403 (in eng).
113. Kaira K, et al. (2008) Prognostic significance of L-type amino acid transporter 1 expression in resectable stage I-III nonsmall cell lung cancer. (Translated from eng) *Br J Cancer* 98(4):742-748 (in eng).
114. Kobayashi K, et al. (2008) Enhanced tumor growth elicited by L-type amino acid transporter 1 in human malignant glioma cells. (Translated from eng) *Neurosurgery* 62(2):493-503; discussion 503-494 (in eng).
115. Kroemer G & Pouyssegur J (2008) Tumor cell metabolism: cancer's Achilles' heel. (Translated from eng) *Cancer Cell* 13(6):472-482 (in eng).
116. Nicklin P, et al. (2009) Bidirectional transport of amino acids regulates mTOR and autophagy. (Translated from eng) *Cell* 136(3):521-534 (in eng).
117. Verrey F, et al. (2004) CATs and HATs: the SLC7 family of amino acid transporters. (Translated from eng) *Pflugers Arch* 447(5):532-542 (in eng).
118. Geier EG, Schlessinger, A., Fan, H., Gable, J. E., Irwin, J. J., Sali, A., Giacomini, K. M. (Submitted) Structure-based ligand discovery for the Large-neutral Amino Acid Transporter 1, LAT-1.
119. Hidalgo M, et al. (1998) A Phase I and pharmacological study of the glutamine antagonist acivicin with the amino acid solution aminosyn in patients with advanced solid malignancies. (Translated from eng) *Clin Cancer Res* 4(11):2763-2770 (in eng).
120. Dolgih E, Bryant C, Renslo AR, & Jacobson MP (2011) Predicting binding to p-glycoprotein by flexible receptor docking. (Translated from eng) *PLoS Comput Biol* 7(6):e1002083 (in eng).
121. Masuda S, et al. (2006) Identification and functional characterization of a new human kidney-specific H⁺/organic cation antiporter, kidney-specific multidrug and toxin extrusion 2. (Translated from eng) *J Am Soc Nephrol* 17(8):2127-2135 (in eng).
122. Otsuka M, et al. (2005) Identification of essential amino acid residues of the NorM Na⁺/multidrug antiporter in *Vibrio parahaemolyticus*. (Translated from eng) *J Bacteriol* 187(5):1552-1558 (in eng).

123. Becker ML, et al. (2009) Genetic variation in the organic cation transporter 1 is associated with metformin response in patients with diabetes mellitus. (Translated from eng) *Pharmacogenomics J* 9(4):242-247 (in eng).
124. Jablonski KA, et al. (2010) Common variants in 40 genes assessed for diabetes incidence and response to metformin and lifestyle intervention in the diabetes prevention program. (Translated from eng) *Diabetes* 59(10):2672-2681 (in eng).
125. Nies AT, Koepsell H, Damme K, & Schwab M (2011) Organic cation transporters (OCTs, MATEs), in vitro and in vivo evidence for the importance in drug therapy. (Translated from eng) *Handb Exp Pharmacol* (201):105-167 (in eng).
126. Damme K, Nies AT, Schaeffeler E, & Schwab M (2011) Mammalian MATE (SLC47A) transport proteins: impact on efflux of endogenous substrates and xenobiotics. (Translated from eng) *Drug Metab Rev* 43(4):499-523 (in eng).
127. Kido Y, Matsson P, & Giacomini KM (2011) Profiling of a prescription drug library for potential renal drug-drug interactions mediated by the organic cation transporter 2. (Translated from eng) *Journal of medicinal chemistry* 54(13):4548-4558 (in eng).
128. Wittwer MB, Zur, A. A., Khuri, N., Kido, Y., Kosaka A., Zhang, X., Morrissey K. M., Sali A., Huang Y., Giacomini K. M. (In press) Discovery of Potent, Selective Multidrug And Toxin Extrusion Transporter 1 (MATE1, SLC47A1) Inhibitors Through Prescription Drug Profiling and Computational Modeling. *Journal of Medicinal Chemistry*.
129. Wishart DS, et al. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 34(Database issue):D668-672.
130. He X, et al. (2010) Structure of a cation-bound multidrug and toxic compound extrusion transporter. *Nature* 467(7318):991-994.
131. Zhang X & Wright SH (2009) MATE1 has an external COOH terminus, consistent with a 13-helix topology. (Translated from eng) *Am J Physiol Renal Physiol* 297(2):F263-271 (in eng).
132. Zhang X, et al. (2012) Twelve transmembrane helices form the functional core of mammalian MATE1 (multidrug and toxin extruder 1) protein. (Translated from eng) *The Journal of biological chemistry* 287(33):27971-27982 (in eng).
133. Kajiwara M, et al. (2009) Identification of multidrug and toxin extrusion (MATE1 and MATE2-K) variants with complete loss of transport activity. (Translated from eng) *J Hum Genet* 54(1):40-46 (in eng).
134. Stroud RM, et al. (2009) 2007 Annual progress report synopsis of the Center for Structures of Membrane Proteins. *J Struct Funct Genomics*.
135. Love J, et al. (2010) The New York Consortium on Membrane Protein Structure (NYCOMPS): a high-throughput platform for structural genomics of integral membrane proteins. *Journal of structural and functional genomics* 11(3):191-199.
136. Mysinger MM, et al. (2012) Structure-based ligand discovery for the protein-protein interface of chemokine receptor CXCR4. (Translated from eng) *Proc Natl Acad Sci U S A* 109(14):5517-5522 (in eng).

137. Fan H, et al. (2011) Statistical potential for modeling and ranking of protein-ligand interactions. (Translated from eng) *Journal of chemical information and modeling* 51(12):3078-3092 (in eng).
138. Zamek-Gliszczynski MJ, Hoffmaster KA, Tweedie DJ, Giacomini KM, & Hillgren KM (2012) Highlights from the international transporter consortium second workshop. (Translated from eng) *Clin Pharmacol Ther* 92(5):553-556 (in eng).
139. Boudker O, Ryan RM, Yernool D, Shimamoto K, & Gouaux E (2007) Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. *Nature* 445(7126):387-393.
140. Yamashita A, Singh SK, Kawate T, Jin Y, & Gouaux E (2005) Crystal structure of a bacterial homologue of Na⁺/Cl⁻-dependent neurotransmitter transporters. *Nature* 437(7056):215-223.
141. Hu NJ, Iwata S, Cameron AD, & Drew D (2011) Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT. (Translated from eng) *Nature* 478(7369):408-411 (in eng).
142. Pebay-Peyroula E, et al. (2003) Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. *Nature* 426(6962):39-44.
143. Johnson ZL, Cheong CG, & Lee SY (2012) Crystal structure of a concentrative nucleoside transporter from *Vibrio cholerae* at 2.4 Å. (Translated from eng) *Nature* 483(7390):489-493 (in eng).
144. Gruswitz F, et al. (2010) Function of human Rh based on structure of RhCG at 2.1 Å. (Translated from eng) *Proc Natl Acad Sci U S A* 107(21):9638-9643 (in eng).
145. DeLano WL (2002) *The PyMOL Molecular Graphics System* San Carlos, CA, USA).
146. Shannon P, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498-2504.
147. Madhusudhan MS, Webb BM, Marti-Renom MA, Eswar N, & Sali A (2009) Alignment of multiple protein structures based on sequence and structure features. *Protein Eng Des Sel* 22(9):569-574.
148. Pettersen EF, et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605-1612.

Chapter 3: High Selectivity of the γ -Aminobutyric Acid Transporter 2 (GAT-2, SLC6A13) Revealed by Structure-based Approach

Avner Schlessinger[§], Matthias B. Wittwer[‡], Amber Dahlin[‡], Natalia Khuri^{‡§},

Massimiliano Bonomi^{‡§}, Hao Fan^{‡§}, Kathleen M. Giacomini[‡], and Andrej Sali^{‡§}

From the [‡]Department of Bioengineering and Therapeutic Sciences and the [§]California Institute for Quantitative Biosciences, University of California, San Francisco, California 94158

SUMMARY

The solute carrier 6 (SLC6) is a family of ion-dependent transporters that mediate uptake into the cell of osmolytes such as neurotransmitters and amino acids. Four SLC6 members transport GABA, a key neurotransmitter that triggers inhibitory signaling pathways via various receptors (e.g., GABAA). The GABA transporters (GATs) regulate the concentration of GABA available for signaling and are thus targeted by a variety of anticonvulsant and relaxant drugs. Here, we characterize GAT-2, a transporter that plays a role in peripheral GABAergic mechanisms, by constructing comparative structural models based on crystallographic structures of the leucine transporter LeuT. Models of GAT-2 in two different conformations were constructed and experimentally validated, using site-directed mutagenesis. Computational screening of 594,166 compounds including drugs, metabolites, and fragment-like molecules from the ZINC database revealed distinct ligands for the two GAT-2 models. 31 small molecules, including high scoring compounds and molecules chemically related to known and predicted GAT-2 ligands, were experimentally tested in inhibition assays. Twelve ligands were found, six of which were chemically novel (e.g., homotaurine). Our results suggest that GAT-2 is a high selectivity/low affinity transporter that is resistant to inhibition by typical GABAergic

inhibitors. Finally, we compared the binding site of GAT-2 with those of other SLC6 members, including the norepinephrine transporter and other GATs, to identify ligand specificity determinants for this family. Our combined approach may be useful for characterizing interactions between small molecules and other membrane proteins, as well as for describing substrate specificities in other protein families.

INTRODUCTION

The solute carrier 6 family (5) consists of 20 Na⁺- and Cl⁻-dependent membrane transporters that regulate a variety of biological activities such as neurotransmission and metabolism (1). This transporter family can be classified into four groups based on their amino acid sequences: monoamine trans- porters, GABA transporters, amino acid transporters, and “orphan” transporters that may also transport amino acids (1, 2). Mutations in SLC6 members are associated with a wide spectrum of disorders such as obsessive-compulsive disorder (serotonin transporter, SLC6A4), obesity (SLC6A14), and orthostatic hypotension (NET, SLC6A2) (2). SLC6 members are therefore targets for many prescription drugs, including antidepressants (e.g., venlafaxine (Effexor®)) and stimulants (e.g., methylphenidate (Ritalin®)) that often act on more than one transporter (3). GABA is a key inhibitory neurotransmitter in the mammalian brain and acts by binding to the GABAergic receptors in inhibitory neurons (e.g., GABAA) in the CNS (4). GABAergic mechanisms have also been found in the stomach, pancreas, intestine, testis, ovary, uterus, liver, urinary bladder, and kidney (5). Malfunctions of the GABAergic system have been associated with mucous overproduction in asthma (6), as well as with protective and regenerative effects on islet β cells in diabetes (7). The GABA transporter (GAT) family consists of four transporters that regulate the concentration of GABA available for signaling via the GABAergic receptors (4, 8). Therefore, the GATs are emerging

drug targets for a variety of disorders, primarily those that have been associated with neurosignaling (4). For example, GAT-1 (SLC6A1) and GAT-3 (SLC6A11) are key targets for anticonvulsants and relaxants (e.g., tiagabine (Gabitril®)). These drugs increase the concentration of GABA in the synaptic cleft by inhibiting GAT-1- or GAT-3-mediated GABA reuptake. The GABA transporter 2 (GAT-2, SLC6A13) is primarily expressed in the liver, kidney, and other peripheral tissues such as the testis, retina, and lungs (4, 8, 9). Thus, GAT-2 might be physiologically important for regulating key peripheral GABAergic mechanisms such as those associated with asthma and diabetes (6, 7). Furthermore, GAT-2 might also play a pharmacological role in disposition and metabolism of GABAergic drugs in the liver and kidney or be a drug target itself (e.g., in asthma therapy).

GAT-2 contains one large domain with 12 predicted membrane-spanning helices (10). No structures of human SLC6 members, including GAT-2, have been determined at atomic resolution; however, x-ray structures of a bacterial homolog, the leucine transporter LeuT, have been determined in four different conformations that were proposed to represent different snapshots of the transport cycle (11–13). Additionally, LeuT complex structures with various substrates and inhibitors suggested a competitive inhibition mechanism in which larger ligands (e.g., tryptophan) stabilize an inhibited outward facing conformation (12). Interestingly, it was also shown that an additional substrate-binding site (S2) is located on the surface of LeuT (14 – 16) and that various inhibitors stabilize different conformations in LeuT via sites overlapping with the S2 site (11, 17). These observations are in agreement with the notion that LeuT and its human homologs transport ligands across the cell membrane via the “alternating access” transport mechanism (13, 19 –25).

Here, we characterize the function of GAT-2 using an integrated computational and experimental approach. We constructed structural models for GAT-2 in two different conformations (i.e., occluded and outward facing states) in complex with GABA and experimentally validated the models using site-directed mutagenesis. We then performed virtual ligand screening against the modeled binding site to predict small molecules, including metabolites, prescription drugs, and fragment-like compounds, that interact with GAT-2. The predicted hits and additional molecules were then validated, using inhibition of radiolabeled substrate uptake experiments. We also compared the predicted GAT-2-binding site with those of other SLC6 members, including the norepinephrine transporter (NET), and the other human GATs. Finally, we discuss the application of these results to describe the specificity determinants in the SLC6 family, as well as the utility of our approach to identify residues important for function and chemically novel ligands for GAT-2 and other transporters.

EXPERIMENTAL PROCEDURES

Comparative Model Construction

GAT-2 was modeled based on x-ray structures of LeuT from *Aquifex aeolicus* in the occluded/outward facing (“occluded”) and the outward facing conformations (Protein Data Bank codes 2A65 (26) and 3F3A (12), respectively), using MODELLER-10v8 (see Fig. 2-1 and supplemental materials). For each conformation, 100 models were generated based on the GAT-2-LeuT alignment (Supplemental Fig. 3-S1), using the standard “automodel” class of MODELLER-10v8 (27). The models were assessed using Z-DOPE, a normalized atomic distance-dependent statistical potential based on known protein structures (see Table 1) (28). Moreover, the occluded GAT-2 conformation was modeled with nonprotein atoms, including the

leucine molecules, ions (chloride and sodium), and other heteroatoms (i.e., the detergents), based on their corresponding coordinates in the template structure; the outward facing GAT-2 conformation was modeled similarly with nonprotein atoms, including the tryptophan molecules, sodium ions, and detergents that were used for crystallization. However, because the S1-binding site in the outward facing structure of LeuT is partially occupied by the detergent molecules B-octylglucoside and tetradecane, we removed the atoms of these molecules from the template structure. Finally, the models were refined by repacking the side chains on a fixed backbone with SCWRL4 (29), as well as by being subjected to 10,000 steps of conjugate gradient minimization under the Amber99SB-ILDN force field (Refs. 30 and 31 and supplemental materials).

Ligand Docking and Virtual Screening

Virtual screening against the GAT-2 models was performed using a semiautomatic docking procedure (32–36). All of the docking calculations were performed with DOCK 3.5.54 (37, 38). The docking poses of database molecules were ranked by DOCK score consisting of van der Waals, Poisson-Boltzmann electrostatic, and ligand desolvation penalty terms. Importantly, binding affinity cannot yet be predicted accurately by docking (35, 39). Thus, poses of the 200–500 highest ranked compounds from each one of the computational screens were analyzed manually (36,40).

The receptor structure was prepared by removing all non- protein atoms, except for the sodium ions. Binding site residues were identified as residues with at least one atom within 10 Å of any heavy atoms of the ligand leucine from the initial model, using the program FILT (from the DOCK3.5 distribution). The solvent-accessible molecular surface of the protein-binding site was then calculated with the program DMS (41) using a probe radius of 1.4 Å. Receptor-derived spheres were calculated using the program SPHGEN (42) (part of the UCSF DOCK suite),

whereas the ligand-derived spheres were generated from the positions of the heavy atoms of the crystallographic ligand, if available. In total, 45 matching spheres were used to orient ligands in the binding site.

Binding Site Assessment

The final models that were used for virtual screening were selected based on their ability to discriminate known ligands from decoys using docking. In particular, we calculated the enrichment for the known ligands among the top scoring decoy compounds, generated by the Directory of Useful Decoys protocol (32, 43). 11 GAT-2 ligands were collected from the literature (4, 9), as well as the UniProt (44) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (45) data-bases. For each known ligand and 36 Directory of Useful Decoys-generated decoys, the best docking pose was computed. The corresponding docking scores were used to calculate EF1 (supplemental materials, Equation 1) and logAUC (supplemental materials, Equation 2), where final models were selected based on their logAUC score. For example, a random selection of known ligands from a database consisting of known ligands and decoys yields a logAUC of 14.5. Finally, the generated models were evaluated using enrichment calculations that included the ligands discovered in this study. The most enriching models were virtually identical to the models used for our original virtual screening.

Data Sets for Virtual Screening

We used the following three compound data sets for virtual screening. First, the KEGG DRUG database is a comprehensive information resource for approved drugs in Japan, the United States, and Europe. It includes all the marketed drugs in Japan, including prescription drugs and over the counter drugs. We used a filtered version that included 6,436 molecules

suitable for docking (36). Second, the KEGG Ligand Compound database includes metabolites, biopolymers, and other chemical substances that are related to biological systems. We also used a filtered version of KEGG Ligand Compound that included 12,730 molecules. Third, the ZINC fragment-like set includes 575,000 purchasable organic molecules with fragment-like physicochemical properties, including a molecular mass of 250 Dalton or lower, five or fewer rotatable bonds, and an xlogP value of 3.5 or less, where xlogP is the octanol/water partition coefficient (logP) calculated by an atom additive method (43, 46).

Chemical Novelty Evaluation

For assessing chemical similarity between two compounds, we computed the Daylight 1024 hashed fingerprints (Daylight Chemical Information Systems, Inc., Laguna Niguel, CA). Tanimoto coefficients (Tc) were calculated between each docking discovered ligand and 14 annotated GAT-2 ligands in the ChEMBL database (48), using the program Pipeline Pilot. Tc values of < 0.5 suggest that the molecule is a chemically novel GAT-2 ligand.

Subcloning and Transient Transfection of GAT-2 into HEK Cells

The full-length human GAT-2 cDNA clone was purchased from ATCC and subcloned into the pcDNA5/FRT expression vector (Invitrogen) according to the manufacturer's protocol. For directional subcloning, the full-length GAT-2 cDNA was excised from the host vector using the restriction enzymes XhoI and HindIII (New England Biolabs, Ipswich, MA). The expression vector was similarly cleaved with the same restriction enzymes, and the two DNA fragments were ligated using T4DNA ligase (Invitrogen) according to the manufacturer's protocol. The ligated DNA was transformed into DH5a cells (Invitrogen) according to the manufacturer's protocol and plated on agar plates (TEKnova, Hollister, CA) overnight at 37 °C. The colonies

were selected the following day and amplified in 5 ml of ampicillin-containing LB broth (TEKnova, Hilden, Germany) overnight. The cultures were then pelleted by centrifugation, and DNA was isolated and purified from the pellets using a Qiagen DNA extraction kit. The DNA was then sequenced (Quintara Biosciences, Albany, CA) to determine the validity of the pcDNA-5-GAT-2 clones. Clones with 100% identity to the GenBank™ reference human GAT-2 sequence were selected.

For transfection, HEK-Flpin cells were seeded in DMEM (Cell Culture Facility, University of California, San Francisco) supplemented with 10% FBS in poly-D lysine-coated 24-well plates at a density of 600,000 cells/ml. Approximately 24 h later, the cells were transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. In brief, for each plate 48 µl of Lipofectamine 2000 were added to 1152 µl of OptiMEM-medium (Cell Culture Facility, University of California, San Francisco), and the solution was incubated at room temperature for 5 min. In the meantime, OptiMEM solutions for each plate containing 19.2 µg of either pcDNA5/FRT (empty vector) or vector containing the DNA of interest (e.g., GAT-2- or site-directed mutant sequences) in a total volume of 1200 µl were prepared. After the 5-min incubation period, the two solutions were combined and incubated for 20 min at room temperature. During this time, the medium in the 24-well plates was changed to 0.5 ml/well prewarmed OptiMEM. Finally, 100 µl of the Lipofectamine-DNA solution were added to each well, and the plates were put back into the incubator. 12 h later the medium was changed to DMEM (Cell Culture Facility, University of California, San Francisco) substituted with 10% FBS, and uptake experiments were performed ~48 h after transfection.

Uptake Experiments

The cells were washed with 0.5 ml of Hanks' buffered saline solution (HBSS; Cell Culture Facility, University of California, San Francisco) per well and then incubated for 10 min in 0.5 ml of HBSS/well at 37 °C. Then the buffer was removed and replaced with 0.5 ml of prewarmed HBSS containing 20 nM ³H-labeled (PerkinElmer Life Sciences) and 1 μM unlabeled GABA (Sigma-Aldrich) with or without the compound of interest at 50, 500, or 5000 μM. After incubating the plate at 37 °C for 2 min, the uptake was stopped by removing the medium and washing twice with 1 ml of ice-cold HBSS/well. The cells were lysed in 700 μl of lysis buffer (0.1 N NaOH and 0.1% SDS in bidistilled water) by shaking for 2.5 h. 600 μl of the lysate were added to 3 ml of EcoLite scintillation fluid (MP Bio, Solon, OH) and counted on a LS6500 Scintillation Counter (Beckman Coulter, Pasadena, CA). The counts were corrected for protein concentration, and the uptake was expressed as a percentage of control (labeled and unlabeled GABA). Protein concentrations were determined with a BCA assay kit (Thermo Scientific, Rockford, IL).

Site-directed Mutagenesis

Site-directed mutagenesis was performed using the QuikChange kit (Agilent Technologies, Santa Clara, CA) according to the manufacturer's protocol. In brief, 5 μl of 10× reaction buffer were mixed with 1 μl of DNA solution at 5 ng/μl, 1.25 μl of primer solution (containing 100 ng/μl of both forward and reverse primers each), 1 μl of dNTP mix, and 41.75 μl of double distilled H₂O. Then 1 μl of PfuTurbo® DNA polymerase was added, and the PCR was started as described in the manufacturer's protocol with 16 cycles and 15 min at 68 °C for the elongation step. PCR products were digested with DpnI at 37 °C for 1 h followed by transfection

into Epicurian Coli XL1-Blue supercompetent bacteria. The bacteria were plated, and the clones were selected and screened for the correct sequence.

Primers for site-directed mutagenesis were designed with the PrimerX web-based program using the following parameters: melting temperature between 78 and 85 °C, GC content between 40 and 60%, length between 25 and 45 bp, and both 5'- and 3'-flanking regions 11–25 bp in length with the primers terminating in G or C and the mutation site at the center of the primer.

Construction and Visualization of Chemical Similarity Networks

The graphs representing the chemical similarity networks (see Fig. 2-3) were constructed and visualized using Cytoscape 2.8.1 (49), relying on the ChemViz (v.1.1) and clusterMaker (v.1.9) plugins. In particular, we calculated the Tc values between all pairs of the 376 small molecules, using the default parameters in ChemViz, which calculates molecular descriptors for the compounds using the Chemistry Development Kit open source library (50). The layout of the final network was obtained using the edge-weighted spring-embedded algorithm in Cytoscape, using the calculated Tc values as weights (see Fig. 2-3A). The edges indicate similarities between molecules with Tc of at least 0.30. We also clustered the molecules using the Markov clustering algorithm (51) of the clusterMaker plugin, using the default parameters (see Fig. 2-3B).

RESULTS

GAT-2 Models and Their Assessment

GAT-2 was modeled based on the structures of the leucine transporter LeuT from *A. aeolicus* in the occluded/outward facing (“occluded”) (26) and the outward facing conformations

(12) (Fig. 2-1, A and B). The comparative models contain the whole transmembrane domain of the protein, including the 12 transmembrane helices and the S1-binding site residues. The refined models were assessed based on their ability to discriminate between known ligands and likely nonbinders (“decoys”), using “enrichment curves” derived from ligand docking calculations (Table 1; Fig. 2-1, C and D; and supplemental materials) (36). The final refined GAT-2 models in the occluded and outward facing conformations obtained logAUC scores of 53.4 and 37.4, respectively (Fig. 2-1 and Table 2-1). This result suggests that the occluded model is more accurate than the outward facing model and that both models are suitable for selecting ligands for experimental testing (32, 36). The logAUC values for the refined models are substantially better than those calculated for the template structures (25.9 and 19.7 for the occluded and outward facing models, respectively) and the initial models (29.5 and 29.1), as well as for random selection of ligands (14.5) (Table 1). We also assessed the models using the enrichment factor (EF1), which is the fraction of the annotated ligands among the 1% top scoring docking hits compared with their fraction in the entire docking database (supplemental materials) (32, 33, 43). The EF1 values for the occluded and outward facing models (17.0 and 7.9, respectively) also indicate that the models can potentially discriminate between known ligands and nonbinders.

Mode of GAT-2 Interaction with GABA

The occluded and outward facing GAT-2 models are structurally similar and include minor backbone and side chain rearrangements, similarly to the corresponding LeuT template structures (12). For example, the residues forming the extracellular gate (i.e., Tyr-129 and Phe-288) are 2 Å further away from each other in the outward facing conformation than in the occluded conformation, making additional volume accessible to a ligand (supplemental Fig. 3-

S2). The corresponding space of the outward facing LeuT structure (i.e., between Tyr-108 and Phe-253) is also partially occupied by a detergent that was used for crystallization (12). In addition, the majority of the key polar interactions between GABA and GAT-2 are conserved in both the occluded and the outward facing models; for example, the carboxyl group of GABA forms polar interactions with the sodium ion Na1, Gly-51, and Gly-53 in the models of both conformations, as well as with Asn-54 only in the occluded model and with Ile-49 only in the outward facing conformation. Furthermore, the amine group of GABA forms a hydrogen bond with the main chain oxygen of Glu-48, as well as polar interaction with the negatively charged side chain of the same residue. During refinement, the conformation of the Glu-48 side chain correlated with the enrichment scores for the models (i.e., models with buried Glu-48 obtained the worst enrichment scores and vice versa (supplemental Fig. 3-S3)), suggesting the importance of this residue in ligand recognition. Thus, our combined model refinement/ligand docking approach was used to predict Glu-48 as a key residue for transport.

Model Validation Using Site-directed Mutagenesis

To validate the GAT-2 models, we mutated residues predicted to be involved in GABA binding or to be in close proximity to the binding site (Fig. 3-1). In particular, we looked at their effect on the uptake of the radiolabeled substrate 3H-GABA into transiently transfected HEK293 cells (“Experimental Procedures”) (Fig. 3-2). Among the most significant mutations were the G51L and G51A that completely abolished transport (Fig. 3-2). This result confirms the importance of Gly-51 in mediating the network of polar interactions between the residues in the GAT-2-binding site, the sodium ion Na1, and the GAT-2 ligands. This network is likely to be conserved among SLC6 members that transport amino acids, including the GATs (e.g., Gly-63 in GAT-1 (52)), but not in monoamine transporters, which have aspartate in the corresponding

position (e.g., Asp-75 in NET (36)). In addition, mutations of Glu-48 hampered transport but did not prevent substrate uptake completely. A reduction of 3H-GABA uptake by ~50% was observed for E48A and by ~90% for E48L and E48Y. The position of Glu-48 in GAT-1 corresponds to a tyrosine in GAT-1 (Tyr-60); thus, the latter mutation (E48Y) suggests that GAT-2 and GAT-1 achieve specificity for GABA via different amino acid residues (52). Interestingly, although Val-132 does not directly interact with the ligand, the V132I mutation abolished transport completely, suggesting that it might have an indirect effect. Importantly, Val-132 corresponds to Ile-111 in LeuT in the S2 substrate-binding site (11), which has not yet been confirmed to exist on the surface of the human SLC6 members. Immunohistochemistry with a GAT antibody confirmed that all of the mutant proteins localized to the membrane similarly to the wild-type protein (data not shown). Finally, the location of the binding site is in agreement with those of the other GATs (20, 53). For example, the mouse GAT-3 (i.e., GAT-4) binding site is almost identical to that of our GAT-2 model (54).

Virtual Screening of Small Molecule Libraries against the GAT-2 Models

We computationally screened filtered libraries of 6,436, 12,730, and 575,000 small molecules from the KEGG DRUG, KEGG Ligand Compound (45), and ZINC fragment-like (43) databases, respectively, against the refined models of the occluded and the outward facing conformations (Experimental Procedures). The KEGG DRUG set includes over the counter and prescription drugs that are marketed in Europe, Japan, and the United States; the KEGG Ligand Compound library consists of a variety of small molecules, biopolymers, and other chemical substances that are found in biological systems; the ZINC fragment-like data set includes purchasable small molecule ligands with fragment-like chemical properties (43, 46). Several known GAT-2 ligands were ranked highly in our screen, which increases our confidence in the

models. For example, GABA was ranked 3 and 36 in the KEGG DRUG screens against the occluded and outward facing models, respectively.

Chemical Similarity of the Predicted Ligands

Relationships among predicted small molecules are illustrated using chemical similarity networks. We calculated the chemical similarity among the 200 top scoring hits of the KEGG DRUG screens against the outward facing and the occluded models. Except for four drugs, all 376 molecules were related to each other, including 24 molecules that were common to both lists (“Experimental Procedures”). This result indicates that the hits occupy a continuous area in the chemical space, even though there was no significant filtering bias in the virtual screens (Fig. 3-3A). These 376 molecules were grouped into 13 distinct clusters based on the similarity among their chemical structures (Fig. 3-3B). For example, cluster 9 includes small drugs containing an alkyne group (e.g., the hypnotic/sedative drug methylpentynol). Interestingly, molecules covering particular areas of the similarity network are predicted using screens against particular models. For example, molecules predicted only in the outward facing model screen are localized in the center of the similarity network (Fig. 3-3A) corresponding to cluster 2 (Fig. 3-3B). The majority of the molecules in this cluster are too large to fit into the binding site of the occluded model (e.g., pemirolast contains three aromatic rings) (Fig. 3-3B). Some hits are predicted to bind GAT-2 based on both models (e.g., homotaurine) (Fig. 3-4).

Rationale for Selecting Molecules

The top 200–500 highest ranked hits in each of the six computational screens (i.e., three data sets against two models) were examined manually. In particular, we analyzed similarities of the predicted poses of these ligands to those in the predicted complexes of GAT-2 with known

ligands, as well as frequent scaffolds and common pharmacological function (35, 36, 39, 40). Some of our selected compounds were similar to known GAT-2 ligands in their structure, function, or predicted mode of interaction with GAT-2. For example, the anticonvulsant γ -amino- β -hydroxybutyric acid (GABOB) is a derivative of GABA that is predicted to interact with the key residues Glu-48 and Gly-51, as well as with the sodium ion Na⁺, similarly to GABA (Figs. 1A and 4B).

Other molecules, however, were different from known GAT-2 ligands in structure, function, or predicted mode of interaction. For instance, homotaurine contains a sulfonic acid group, making it dissimilar to known GAT-2 ligands (Tc of 0.30) (supplemental Table S1). Nevertheless, our confidence in the homotaurine prediction was increased for two reasons. First, homotaurine was predicted as GAT-2 ligand by virtual screens against GAT-2 models in both the occluded and outward facing conformations. Second, the poses of homotaurine in the predicted complexes were similar to the poses of GABA in the corresponding GAT-2-GABA complexes (Fig. 3-4, A and C). In total, 31 of the predicted compounds were tested using a cis-inhibition assay that does not distinguish between inhibitors and substrates.

Experimental Validation of the Top Hits

To experimentally validate GAT-2 inhibition by the predicted compounds, we developed a cell-based cis-inhibition assay. This assay is based on the uptake of ³H-GABA in HEK293 cells transiently transfected with GAT-2 and on the capacity of inhibitors and substrates to reduce intracellular accumulation of the probe substrate. To choose an appropriate GABA concentration and a suitable uptake time, we performed time course and uptake kinetics experiments (supplemental Fig. 3-S4). The functionality of the assay was demonstrated by comparing kinetic parameters with those previously reported for GAT-2 in published data (4, 8,

9) and by competitive inhibition of 3H-GABA uptake with unlabeled GABA in addition to the known GAT-2 inhibitors β -alanine and desipramine.

Twelve of the predicted molecules inhibited 3H-GABA uptake by 20% or more at concentrations ranging from 500 μ m to 5 mm (Fig. 3-5 and supplemental Figs. S5 and S6). These values are in agreement with previous uptake inhibition data for GABAergic molecules (4, 9). Six of the identified ligands (homotaurine, baclofen, gabapentin, lorazepam, pyridoxal phosphate, and ZAPA) are chemically novel GAT-2 ligands (Fig. 3-5 and supplemental Table S1). Homotaurine, gabapentin, and ZAPA were predicted using the models in both conformations, whereas the larger molecules baclofen, lorazepam, and pyridoxal phosphate were predicted using only the outward facing model, demonstrating that virtual screening against different conformations can cover different parts of the chemical space (Fig. 3-4D). However, the affinity of the three ligands binding to the outward conformation is not as strong as that of the occluded model ligands. For example, the relaxant drug baclofen, a GABAB receptor agonist (55) that was recently approved for treating alcoholism (56), contains a chlorophenyl group, making it a chemically novel GAT-2 ligand (Tc of 0.36) (supplemental Table S1). In addition, the anticonvulsants GABOB and vigabatrin were also identified as GAT-2 ligands. However, although these molecules are highly similar in structure to GABA (i.e., Tc of 0.93 and 0.75, respectively), they are much weaker inhibitors of GAT-2 than GABA itself. For example, GABOB and vigabatrin inhibited 3H-GABA uptake by 59% at 500 μ m and by 53% at 5 mm, respectively.

Similarly, other GABA analogs including pregabalin and amicar, as well as various amino acids, did not exert any effect on 3H-GABA uptake (Fig. 3-5 and supplemental Table S1). Moreover, classical inhibitors of transporters exhibit no (i.e., rifampicine and cyclosporine) or

only weak (erythromycin) inhibitory effects on GAT-2 transport further demonstrating the high tolerance of GAT-2 against inhibition. We also tested compounds that are structurally related to known and predicted GAT-2 ligands (supplemental Table S1). However, of these predicted molecules only the chemically novel hit pyridoxal phosphate inhibited 3H-GABA uptake (Fig. 3-5). In summary, our results suggest that GAT-2 is more selective for inhibitors than the other GATs.

Finally, based on the observation that GAT-2 is highly selective for small GABA-like compounds, we computationally screened the ZINC fragment-like small molecule set (575,000 molecules) against the GAT-2 model. We experimentally validated three hits. 3-Aminobutanoic acid reduced 3H-GABA uptake by at least 50% at 500 μ m, and 5-aminovaleric acid exhibited a 30% inhibition at 500 μ m (Fig. 3-5). Furthermore, 5-aminolevulinic acid (5-ALA) inhibits 3H-GABA uptake by 44% at 5 mm (Fig. 3-5) and completely abolished transport at 20 mm (supplemental Fig. 3-S6). The relatively weak inhibition by these previously unknown GABA-like ligands again indicates that GAT-2 is a highly selective transporter.

Comparison between the GAT-2 and the NET Models

We compared the binding site of GAT-2 to that of NET (36), which shares about 45% sequence identity with GAT-2 (1, 2), to rationalize their variation in substrate specificity (i.e., GAT-2 and NET transport small GABA-like zwitterions and monoamines, respectively) (Fig. 3-6). The NET model, which was also constructed based the LeuT structure, is highly similar to models of the other SLC6 monoamine transporters serotonin transporter (57) and dopamine transporter (58). The predicted GAT-2 and NET structures are similar (root mean square deviation of 0.1 Å), sharing several features, including the location of the sodium ions and the

S1-binding site, as well as the arrangement of several key binding site residues (i.e., Tyr-129, Phe-288, and Ser-289 in GAT-2 correspond to Tyr-152, Phe-317, and Ser-318 in NET).

However, the structural models of NET and GAT-2 include the following main differences that may rationalize their substrate specificities: First, Phe-72 and Phe-323 in NET are substituted for Glu-48 and Leu-294 in GAT-2. As a result, (i) the GAT-2-binding site consists of fewer aromatic residues than that of NET, in which the additional phenylalanine residues can make π - π interactions with ligands such as norepinephrine (36) and (ii) the corresponding region of the GAT-2-binding site is more acidic than that of NET; thus, it is capable of forming key polar interactions with the amine groups of the GAT-2 ligands, such as GABA (Fig. 3-6A, gray dotted lines). Second, the negatively charged Asp-75 in NET is replaced by Gly-51 in GAT-2, which is predicted to adopt a similar conformation to that of Gly-24 in the LeuT x-ray structure. Consequently, (i) the volume occupied by the aspartate side chain in NET is accessible for ligands in GAT-2, and (ii) the negative charge of the aspartate side chain is removed, and the positive charge of the sodium ion Na1 becomes accessible for ligands, changing the corresponding accessible surface of the binding site from acidic (NET) to basic (GAT-2). Third, Ala-145, Val-148, Gly-422, and Gly-423 in NET are replaced by the larger Val-122, Leu-125, Val-393, and Cys-394 in GAT-2, making the volume previously occupied by the aromatic rings of NET ligands inaccessible in GAT-2 (Fig. 3-6).

DISCUSSION

The function of transporters is determined by their structure, dynamics, and localization (12, 14, 17, 20–25). For example, the shape and physicochemical properties of the transporter binding site (i.e., specificity determinants) govern the molecules that bind to the transporter (i.e., binding specificity), which may help determine the molecules that get transported (i.e., substrate

specificity). The mechanism of transport describes relationships between specificity determinants to binding specificity and substrate specificity. A key step toward describing the mechanisms of transport by solute carriers includes the characterization of transporter structures in different conformational states in complex with their ligands, through computation and/or experiment.

Four key findings are presented in this study. First and importantly, distinct ligands were identified using GAT-2 models in two different conformations: the occluded and the outward facing (Fig. 3-3). This finding highlights the importance of characterizing the dynamics of the transport process for GAT-2 and other transporters to improve our ability to find unknown ligands with novel scaffolds. Second, by applying our combined modeling/docking approach to GAT-2, a membrane protein that shares only about 23% sequence identity with its template structure (Fig. 3-1 and supplemental Fig. 3-S1), we correctly predicted key residues for function (Figs. 3-1 and 3-2 and supplemental Fig. 3-S3) and selectivity of chemically novel GAT-2 ligands (Fig. 3-5 and supplemental Table 3-S1). This suggests that our combined experimental and computational structure-based approach can be useful for identifying functionally important residues in other proteins, as well as unknown interactions between these proteins and chemically novel small molecule ligands. Third, several drugs (e.g., baclofen) and metabolites (e.g., GABOB) that target proteins other than transporters (e.g., the GABA receptor GABAB) are also ligands of GAT-2 (Figs. 4 and 5, Table 1, and supplemental Table S1). GAT-2 inhibition by these molecules might contribute to their pharmacological (i.e., efficacy and/or side effects) and physiological functions, which is an example of polypharmacology, a phenomenon in which a drug binds multiple targets (59, 60). Fourth, a comparison of the GAT-2 and NET binding sites confirms previously proposed specificity determinants of the human SLC6 family (20, 21) and

identifies unknown factors important for substrate specificity for this key transporter family (Fig. 3-6). We now discuss each of the four points in turn.

Distinct Ligands Are Identified Using Two Different GAT-2 Conformations

An important step toward a description of the transport mechanism for the SLC6 family includes the computational or experimental characterization of their structures in various conformations. It has been suggested that a LeuT substrate needs to both bind to the S1-binding site of the outward facing conformation and fit within the binding cavity of the occluded transporter state (12). Tryptophan, which is much larger than leucine, does not fit into the cavity in the occluded state and consequently traps LeuT in the outward facing conformation, thereby acting as a competitive inhibitor (12). Whether the model approximates an active or inhibited conformation of GAT-2 is expected to determine the type of ligands that are predicted by virtual screening against the model.

Our virtual screening, chemical similarity network of the hits (Fig. 3-3), structural comparison of the predicted complexes (Fig. 3-4), and uptake kinetic experiments (Fig. 3-5) indicate that small molecule ligands predicted using comparative models of the two different conformations are indeed chemically different. Molecules predicted to interact with GAT-2 using the outward facing model (e.g., baclofen) are larger and more hydrophobic than those predicted to bind to the occluded model (e.g., homotaurine); however, our experiments show that the larger molecules are not stronger inhibitors (Fig. 3-5). This observation suggests that the differences between substrates and competitive inhibitors of GAT-2 includes more features than just size and hydrophobicity and that characterization of additional conformations of SLC6 members in complex with their ligands is needed. Importantly, although the outward facing model is less accurate than the occluded model (Table 1), it is useful for identifying chemically novel ligands

(supplemental Table S1). Recent studies using experimental techniques such as single molecule FRET (15, 16) and electron paramagnetic resonance (17) in combination with MD simulations (14) revealed that different inhibitors stabilize additional LeuT conformations. Furthermore, additional crystallographic structures of LeuT (13) and other proteins with the LeuT-like fold, such as the sodium-hydantoin transporter Mhp1 (61, 62), the amino acid antiporter AdiC (63), and the sodium/galactose transporter vSGLT (64–66), have revealed additional conformations that GAT-2 might adopt during transport or inhibition. Taken together with our results, future studies should screen against GAT-2 models in additional conformations to identify new classes of GAT-2 ligands including substrates and inhibitors.

Structure-based Ligand Discovery for Membrane Transporters

In virtual screening, large libraries of organic molecules are docked computationally against experimentally determined atomic structures of target proteins. For proteins with unknown structure, comparative or homology modeling can be applied when the target sequences are detectably related to an experimentally determined protein structure. Recent advances and automation in molecular docking and comparative modeling enabled the application of structure-based approaches to ligand discovery. For example, such protocols were applied to identify prescription drugs that interact with the NET (36), as well as potent and novel inhibitors for the Dopamine receptor D3 (40). In this study, we apply comparative modeling and virtual screening to characterize a relatively unstudied transporter, GAT-2, which is distantly related to its template structure LeuT. By constructing a large number of models and selecting the final models based on their enrichment scores, we predicted key residues for transport by GAT-2 (Fig. 3-1 and supplemental Fig. 3-S3). These residues were validated experimentally via site-directed mutagenesis and kinetic measurements (Fig. 3-2). We also identified unknown

ligands, including endogenous metabolites and prescription drugs that interact with GAT-2, to further characterize its physiological and pharmacological roles (Figs. 3-4 and 3-5 and supplemental Table 3-S1). Recent structures of homologs of varied human SLCs (66) increase our ability to discover ligands for bio medically important transporters as well as other proteins.

Physiological and Pharmacological Implications of GAT-2 Inhibitors

Of the four GABA transporters, GAT-2 is the least studied. Its localization in the liver, kidney, pancreas, retina, and lung suggests that it plays an important role in GABAergic signaling in peripheral tissues. Identifying structure-function relationships of the transporter will enhance our understanding of the functional role of the transporter. Although six of the GAT-2 ligands identified in our screen contain novel scaffolds (e.g., baclofen), most of the identified ligands were chemically similar to GABA (supplemental Table S1). Strikingly, even ligands that are within one heavy atom of GABA, such as GABOB (supplemental Table S1) and 2,4-diaminobutyric acid (4, 9), have significantly lower affinity to GAT-2 than GABA itself. Therefore, we hypothesize that a potential inhibitor needs to fulfill specific structural requirements (e.g., size and configuration of charges) to bind GAT-2, unlike inhibitors of other SLC6 monoamine transporters (2, 67–69) and other GATs (4, 8, 9).

Because of the blood-brain barrier (BBB), peripherally expressed transporters such as GAT-2 are exposed to higher systemic concentrations of xenobiotics compared with transporters within the CNS. Although speculative, it is possible that GAT-2 may therefore have evolved to be more resistant to chemical inhibition. In particular, resistance to xenobiotic inhibitors, which would otherwise result in a reduction in its function, might be important. Notably, general transporter inhibitors such as cyclosporine were not able to inhibit GAT-2 even at high concentrations, and because of specific structural requirements, it seems unlikely that GAT-2 is

inhibited by many commonly used drugs at their pharmacological concentrations. Nevertheless, several pharmacological agents may still be substrates of GAT-2. In particular, the reduction of GAT-2-mediated 3H-GABA uptake by several of the above mentioned compounds (e.g., vigabatrin) might indicate that they are substrates of this transporter, which would have important toxicological and pharmacological consequences. For example, the heme precursor 5-ALA is involved in the development of the neurological symptoms of porphyria (70, 71). Medical uses of 5-ALA include the photodynamic detection of various tumors (especially in the CNS) and its use as a photosensitizer for photodynamic therapy of many diseases (11, 72, 73). Nevertheless, to date it is not fully understood how 5-ALA crosses the BBB. However, it has been shown that GABA and 5-ALA share a common facilitator in *Saccharomyces cerevisiae* (47). In addition, based on mRNA expression data from our laboratory (data not shown), GAT-2 is enriched at the BBB and might thus be a candidate transporter for the translocation of 5-ALA through the BBB.

Toward a Description of Substrate Specificity in the SLC6 Family

Identification of structural relationships in transporter-ligand complexes among members of the SLC6 family and their correlation with experimental ligand binding results facilitates the description of specificity determinants within this transporter family. Although GAT-2 and NET are highly related in sequence (sequence identity of ~45%), their substrates are chemically different (small linear amino acids and monoamines for GAT-2 and NET, respectively). The differences in the physicochemical properties of the small molecule substrates of these two SLC6 members are reflected in the following key differences in their corresponding binding sites (Fig. 3-6): (i) the number of aromatic residues (two in GAT-2 and four in NET), (ii) the number and location of the charged groups (Glu-48 and Na¹ in GAT-2 and Asp-75 in NET), and (iii) the size

and shape of the binding site (i.e., Ala-145, Val-148, Gly-422, and Gly-423 in NET are replaced by the larger Val-122, Leu-125, Val-393, and Cys-394 in GAT-2). Thus, our comparison between the models of GAT-2 and NET, representatives of two groups within the SLC6 family (i.e., the GABA transporters and the monoamine transporters, respectively), in complex with their ligands highlights key features used by SLC6 members to achieve substrate specificity (Fig. 3-6).

In addition, our results further support the finding that even though four SLC6 members transport GABA in humans, they achieve their specificities using different mechanisms. For example, the side chain of Glu-48 in GAT-2 is predicted to make key polar interactions with GABA, whereas this position in GAT-1 is occupied by a tyrosine (Tyr-60). Interestingly, the E48Y mutation in GAT-2 (Fig. 3-2) and the Y60E mutation in GAT-1 (52) significantly affected the functions of these transporters. Similarly, the mutation E61Y in the mouse homolog of GAT-3 (i.e., GAT-4) resulted in a negative effect on its transport activity (54). Furthermore, despite exhibiting high sequence similarity, the transporters GAT-2, GAT-3, and the betaine/GABA transporter (BGT-1) have considerable differences in affinity to GABA-like molecules (4, 9). Particularly, the residues in close proximity to the S1-binding site in GAT-2 and GAT-3 are highly similar (the only difference is Val-132 in GAT-2, which corresponds to Ile-150 in GAT-3) (supplemental Table S2). Surprisingly, the V132I GAT-2 mutant, which mimics GAT-3, almost completely lost its transport capability (Fig. 3-2). Although the corresponding residue in LeuT (Ile-111) is found in the S2 binding site (11), it has not been shown that human SLC6 members contain an additional high affinity substrate-binding site (i.e., S2); however, it is plausible that Val-132 is a part of a lower affinity binding site that is coupled allosterically to S1.

In summary, through modeling GAT-2 in two different conformations, our study revealed distinct ligands of GAT-2 and suggested that GAT-2, a peripheral transporter, selectively restricts binding of inhibitors to a greater degree than other GABA transporters. Finally, our results have broad implications for the characterization of SLC6 structures. In particular, the interactions of multiple conformations of transporters with ligands are needed to describe at higher resolution the specificity and mechanisms of transport. Furthermore, our approach is generally useful for describing substrate specificities in protein families other than the SLC6 family, including other transporters, receptors, and enzymes.

FIGURES

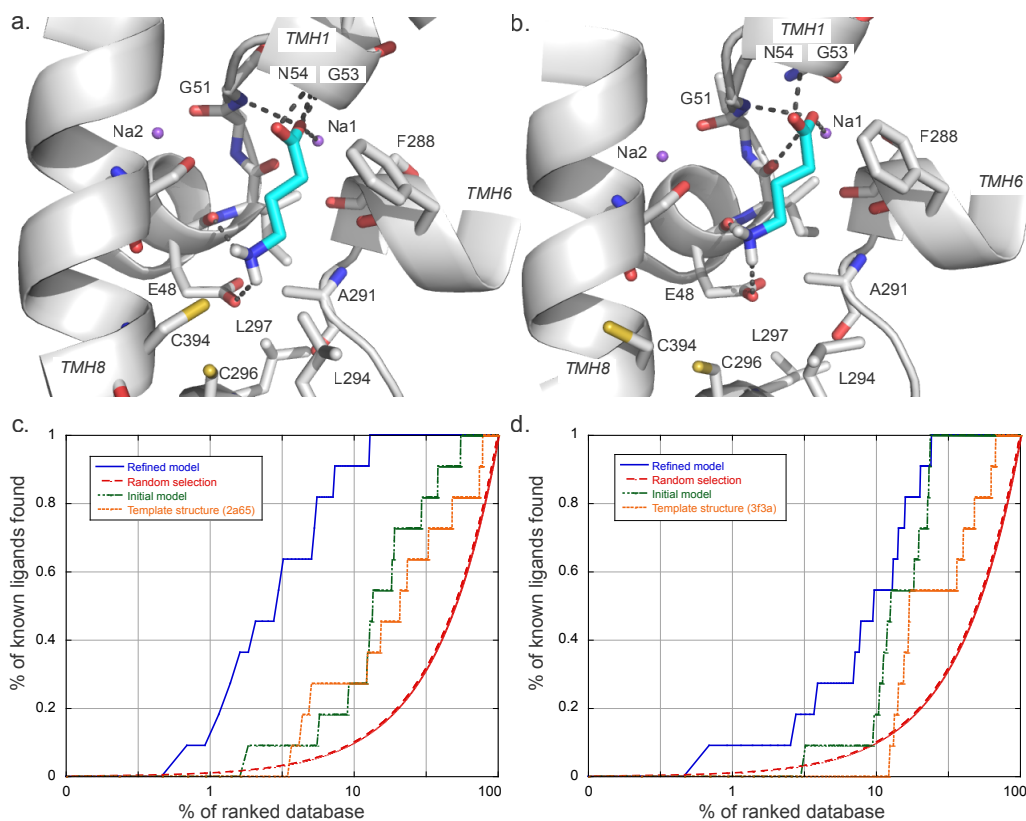


Fig. 3-1. GAT-2 – GABA models and their validation by ligand enrichment.

Predicted structures of the GAT-2 – GABA complex in the occluded (a) and the outward-facing (b) conformations. GABA is colored in cyan, with oxygen, nitrogen, and hydrogen atoms in red, blue, and white, respectively. The sodium ions Na1 and Na2 are visualized as purple spheres. The transmembrane helices of GAT-2 are depicted as white ribbons. Key residues are displayed as sticks; the hydrogen bonds between GABA and GAT-2 are shown as dotted gray lines; they involve the residues E48, G51, and G53, as well as the sodium ion Na1 for both conformations models; GABA forms polar interactions with N54 only in the occluded conformation model. Enrichment plots for different structures of the occluded (c) and the outward-facing (d) models:

the refined GAT-2 models (blue), random selection (red), the initial GAT-2 models (green), and the LeuT template structures (orange).

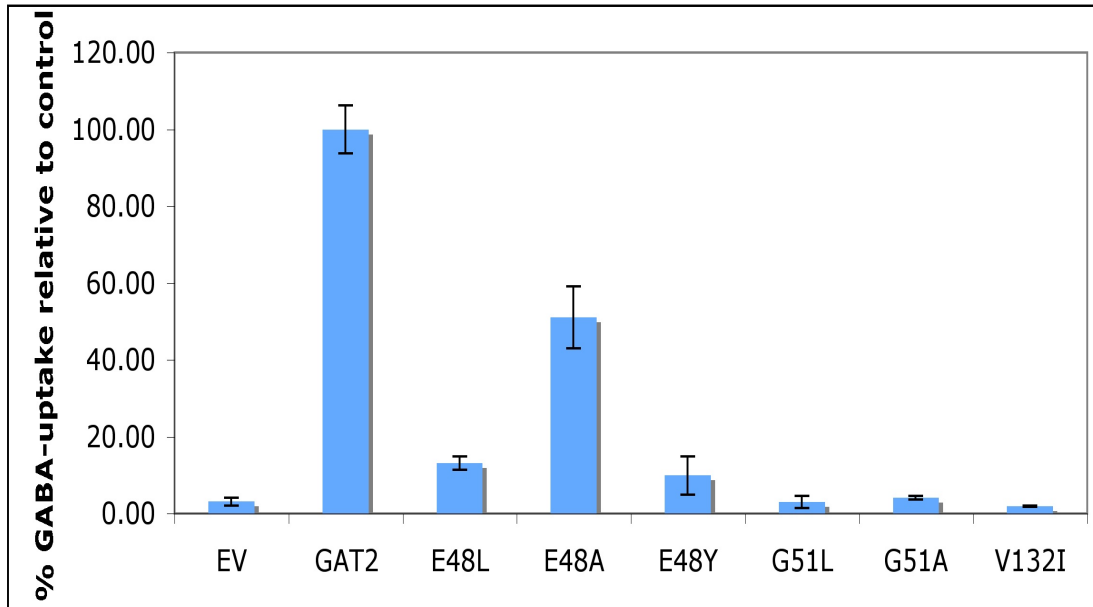


Fig. 3-2. Validating predicted binding site residues by mutagenesis.

Influence of site-directed mutations on 3H-GABA transport as compared to the wild type sequence (GAT2) and empty vector (EV). The results were obtained in HEK293-cells transiently transfected with the cDNAs for the reference and mutated proteins.

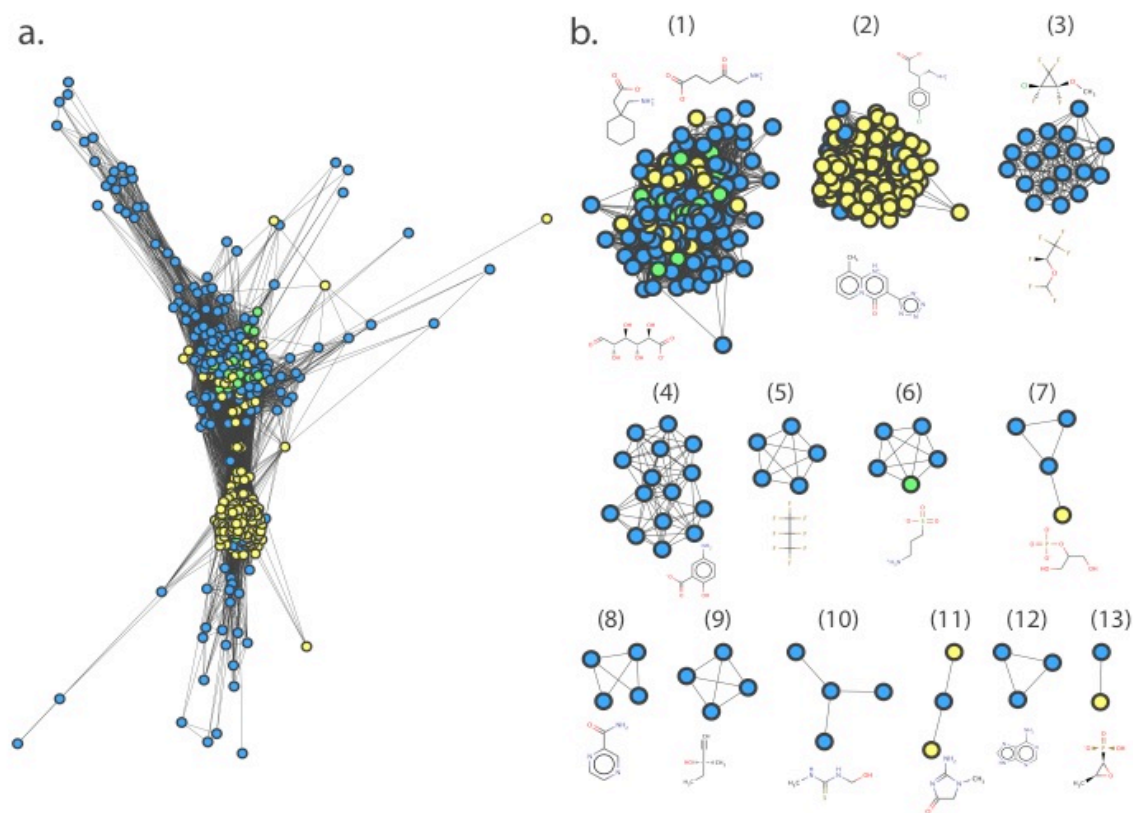


Fig. 3-3. Chemical similarity network of predicted ligands.

The relationships between the top-ranked small molecule drugs from KEGG DRUG are visualized using Cytoscape 2.8.1. The nodes represent the small molecules predicted to bind GAT-2, using the occluded model (blue), the outward-facing model (yellow) or both models (green). Each edge represents pairwise chemical similarity with Tanimoto coefficient (Tc) of at least 0.3. (a) A similarity network using the Edge-Weighted Spring-Embedded layout algorithm in Cytoscape, which preserves all the relationships among the small molecules [182]. (b) A network with the 13 clusters of the small molecule drugs. The molecules were clustered using the Markov Clustering algorithm [185] in Cytoscape. Representative small molecules structures of the clusters are visualized using MarvinView 5.4.1.1 (<http://www.chemaxon.com/>).

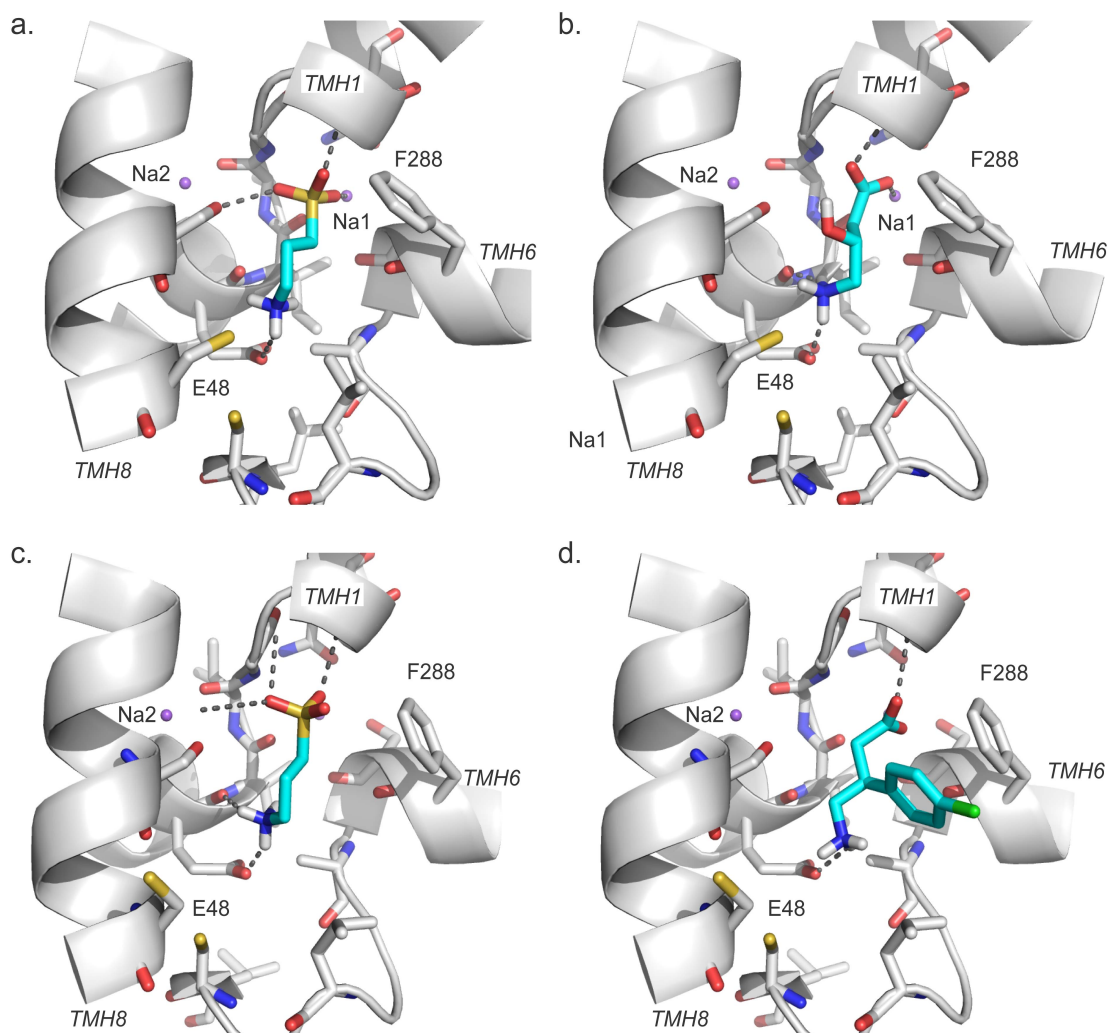


Fig. 3-4. Predicted binding modes for GAT-2 validated ligands.

Predicted binding modes of newly identified GAT-2 ligands with GAT-2 models in the occluded (a-b) and the outward-facing (c-d) conformations. Ligands are colored in cyan, with sulfur, chloride, oxygen, nitrogen, and hydrogen atoms in yellow, green, red, blue, and white, respectively. The sodium ions Na1 and Na2 are visualized as purple spheres. The transmembrane helices of GAT-2 are depicted as white ribbons. Key residues are displayed as sticks; the hydrogen bonds between ligands and key GAT-2 residues (eg, E48) are shown as dotted gray

lines. The representative previously unknown ligands are homotaurine (a, c), GABOB (b), and Baclofen (d).

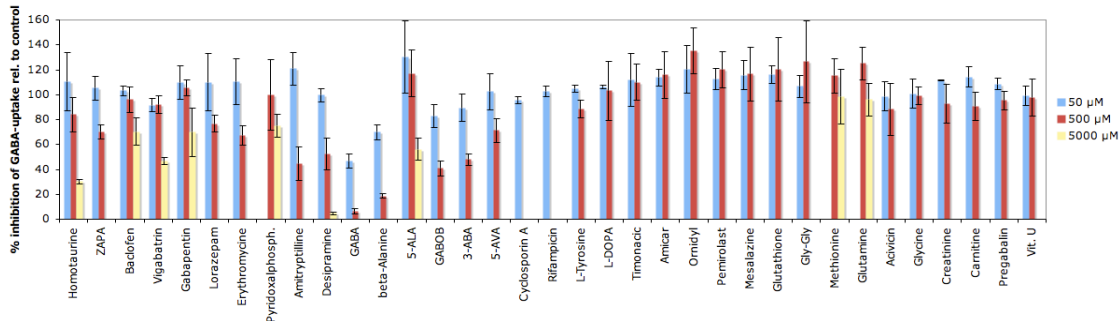


Fig. 3-5. Cis-inhibition studies of predicted GAT-2 inhibitors.

Uptake of 3H-GABA in transiently GAT2-transfected HEK293-cells in presence of various small molecule compounds. The tested concentrations were 50 and 500 μM . A concentration of 5000 μM was only used where solubility and toxicity allowed. The highest possible concentration for pyridoxal phosphate was 1000 μM , which yielded a 15%-reduction in 3H-GABA uptake (data not shown).

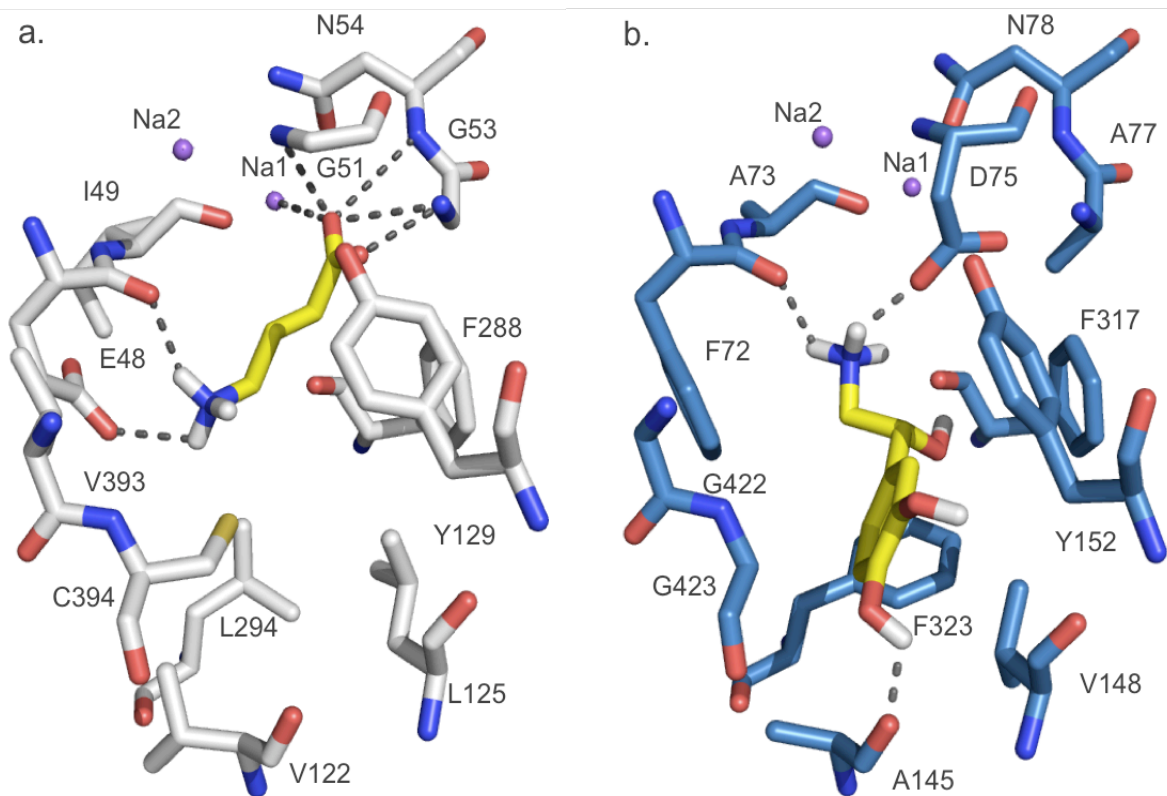


Fig. 3-6. Comparison of GAT-2 and NET predicted binding sites.

The final model of GAT-2 (white) (a) and the model of NET [110] (blue) (b) in the occluded conformation are shown with their corresponding substrates (ie, GABA and norepinephrine, respectively)(yellow). Atoms are illustrated by sticks, with oxygen, nitrogen, and hydrogen atoms in red, blue, and white, respectively. The sodium ions Na1 and Na2 are visualized with purple spheres. GABA and norepinephrine are depicted in orange sticks and their hydrogen bonds with GAT-2 (involving E48, G51, G53, N54, and Na1) and NET (involving A145, F72, and D75) are shown as dotted gray lines.

TABLES

Occluded conformation			
Modela	Z-DOPEb	logAUCc	EF1d
Template Structure	-2.5	25.9	0
Initial model	-0.93	29.5	0
Refined model	-0.36	53.4	17.0
Outward-facing conformation			
Template Structure	-2.39	19.7	0
Initial model	-1.13	29.1	0
Refined model	-1	37.4	7.9

Table 3-1: assessment the GAT-2 models

a. Model marks the model used for the assessment; Template structure corresponds to the structures of the leucine transporter LeuT in the occluded (PDB id 2A65) and the outward-facing (PDB id 3F3A) conformations; Initial model represents the initial models by MODELLER that achieved the best Z-DOPE [186] score, without refinement; Refined model marks our final refined models. b. Z-DOPE provides the score of the models using Z-DOPE, a normalized atomic distance-dependent statistical potential based on known protein structures [186]. Per-residue Z-DOPE score of the initial score was also compared to that of the template structure. c. logAUC marks the area under the logarithmic scale of the enrichment curve. d. EF1 represents the enrichment factor at 1% of the ranked database.

REFERENCES

1. Chen, N. H., Reith, M. E., and Quick, M. W. (2004) *Pflugers Arch* 447, 519-531
2. Hahn, M. K., and Blakely, R. D. (2007) *Annu Rev Pharmacol Toxicol* 47, 401-441
3. Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2006) *Nucleic Acids Res* 34, D668-672
4. Madsen, K. K., White, H. S., and Schousboe, A. (2010) *Pharmacology & therapeutics* 125, 394-401
5. Erdo, S. L., and Wolff, J. R. (1990) *J Neurochem* 54, 363-372
6. Xiang, Y. Y., Wang, S., Liu, M., Hirota, J. A., Li, J., Ju, W., Fan, Y., Kelly, M. M., Ye, B., Orser, B., O'Byrne, P. M., Inman, M. D., Yang, X., and Lu, W. Y. (2007) *Nat Med* 13, 862-867
7. Soltani, N., Qiu, H., Aleksic, M., Glinka, Y., Zhao, F., Liu, R., Li, Y., Zhang, N., Chakrabarti, R., Ng, T., Jin, T., Zhang, H., Lu, W. Y., Feng, Z. P., Prud'homme, G. J., and Wang, Q. (2011) *Proc Natl Acad Sci U S A* 108, 11692-11697
8. Nakashita, M., Sasaki, K., Sakai, N., and Saito, N. (1997) *Neurosci Res* 29, 87-91
9. Christiansen, B., Meinild, A. K., Jensen, A. A., and Brauner-Osborne, H. (2007) *The Journal of biological chemistry* 282, 19331-19341
10. Pacholczyk, T., Blakely, R. D., and Amara, S. G. (1991) *Nature* 350, 350-354
11. Nyola, A., Karpowich, N. K., Zhen, J., Marden, J., Reith, M. E., and Wang, D. N. (2010) *Curr Opin Struct Biol* 20, 415-422
12. Singh, S. K., Piscitelli, C. L., Yamashita, A., and Gouaux, E. (2008) *Science* 322, 1655-1661
13. Krishnamurthy, H., and Gouaux, E. (2012) *Nature* 481, 469-474
14. Shi, L., Quick, M., Zhao, Y., Weinstein, H., and Javitch, J. A. (2008) *Mol Cell* 30, 667-677
15. Zhao, Y., Terry, D., Shi, L., Weinstein, H., Blanchard, S. C., and Javitch, J. A. (2010) *Nature* 465, 188-193
16. Zhao, Y., Terry, D. S., Shi, L., Quick, M., Weinstein, H., Blanchard, S. C., and Javitch, J. A. (2011) *Nature*
17. Claxton, D. P., Quick, M., Shi, L., de Carvalho, F. D., Weinstein, H., Javitch, J. A., and McHaourab, H. S. (2010) *Nature structural & molecular biology* 17, 822-829
18. Nyola, A., Karpowich, N. K., Zhen, J., Marden, J., Reith, M. E., and Wang, D. N. (2010) *Curr Opin Struct Biol* 20, 415-422
19. Forrest, L. R., and Rudnick, G. (2009) *Physiology (Bethesda)* 24, 377-386
20. Kanner, B. I., and Zomot, E. (2008) *Chem Rev* 108, 1654-1668
21. Krishnamurthy, H., Piscitelli, C. L., and Gouaux, E. (2009) *Nature* 459, 347-355

22. Forrest, L. R., Zhang, Y. W., Jacobs, M. T., Gesmonde, J., Xie, L., Honig, B. H., and Rudnick, G. (2008) *Proc Natl Acad Sci U S A* 105, 10338-10343
23. Jardetzky, O. (1966) *Nature* 211, 969-970
24. Guan, L., and Kaback, H. R. (2006) *Annu Rev Biophys Biomol Struct* 35, 67-91
25. Abramson, J., and Wright, E. M. (2009) *Curr Opin Struct Biol* 19, 425-432
26. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y., and Gouaux, E. (2005) *Nature* 437, 215-223
27. Sali, A., and Blundell, T. L. (1993) *J Mol Biol* 234, 779-815
28. Shen, M. Y., and Sali, A. (2006) *Protein Sci* 15, 2507-2524
29. Krivov, G. G., Shapovalov, M. V., and Dunbrack, R. L., Jr. (2009) *Proteins* 77, 778-795
30. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) *Proteins* 78, 1950-1958
31. Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y., and Wriggers, W. (2010) *Science* 330, 341-346
32. Huang, N., Shoichet, B. K., and Irwin, J. J. (2006) *J Med Chem* 49, 6789-6801
33. Fan, H., Irwin, J. J., Webb, B. M., Klebe, G., Shoichet, B. K., and Sali, A. (2009) *J Chem Inf Model* 49, 2512-2527
34. Irwin, J. J., Shoichet, B. K., Mysinger, M. M., Huang, N., Colizzi, F., Wassam, P., and Cao, Y. (2009) *Journal of medicinal chemistry* 52, 5712-5720
35. Shoichet, B. K. (2004) *Nature* 432, 862-865
36. Schlessinger, A., Geier, E., Fan, H., Irwin, J. J., Shoichet, B. K., Giacomini, K. M., and Sali, A. (2011) *Proc Natl Acad Sci U S A* 108, 15810-15815
37. Lorber, D. M., and Shoichet, B. K. (2005) *Curr Top Med Chem* 5, 739-749
38. Mysinger, M. M., and Shoichet, B. K. (2010) *J Chem Inf Model* 50, 1561-1573
39. Carlsson, J., Coleman, R. G., Setola, V., Irwin, J. J., Fan, H., Schlessinger, A., Sali, A., Roth, B. L., and Shoichet, B. K. (2011) *Nature chemical biology* 7, 769-778
40. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988) *Journal of Molecular Graphics* 6, 13-27
41. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., and Ferrin, T. E. (1982) *J Mol Biol* 161, 269-288
42. Irwin, J. J., and Shoichet, B. K. (2005) *J Chem Inf Model* 45, 177-182
43. (2008) *Nucleic Acids Res* 36, D190-195
44. Okuda, S., Yamada, T., Hamajima, M., Itoh, M., Katayama, T., Bork, P., Goto, S., and Kanehisa, M. (2008) *Nucleic Acids Res* 36, W423-426

45. Carr, R. A., Congreve, M., Murray, C. W., and Rees, D. C. (2005) *Drug Discov Today* 10, 987-992
46. Daylight Chemical Information Systems, Inc., Laguna Niguel, CA 92677
47. Overington, J. (2009) *J Comput Aided Mol Des* 23, 195-198
48. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) *Genome Res* 13, 2498-2504
49. Steinbeck, C., Hoppe, C., Kuhn, S., Floris, M., Guha, R., and Willighagen, E. L. (2006) *Curr Pharm Des* 12, 2111-2120
50. Enright, A. J., Van Dongen, S., and Ouzounis, C. A. (2002) *Nucleic acids research* 30, 1575-1584
51. Kanner, B. I. (2003) *The Journal of biological chemistry* 278, 3705-3712
52. Kuntz, I. D. (1992) *Science* 257, 1078-1082
53. Carter, L. P., Koek, W., and France, C. P. (2009) *Pharmacology & therapeutics* 121, 100-114
54. Garbutt, J. C., Kampov-Polevoy, A. B., Gallop, R., Kalka-Juhl, L., and Flannery, B. A. (2010) *Alcohol Clin Exp Res* 34, 1849-1857
55. Abramson, J., and Wright, E. M. (2009) *Curr Opin Struct Biol*
56. Keiser, M. J., Setola, V., Irwin, J. J., Laggner, C., Abbas, A. I., Hufeisen, S. J., Jensen, N. H., Kuijter, M. B., Matos, R. C., Tran, T. B., Whaley, R., Glennon, R. A., Hert, J., Thomas, K. L., Edwards, D. D., Shoichet, B. K., and Roth, B. L. (2009) *Nature* 462, 175-181
57. Roth, B. L., Sheffler, D. J., and Kroeze, W. K. (2004) *Nat Rev Drug Discov* 3, 353-359
58. Weyand, S., Shimamura, T., Yajima, S., Suzuki, S., Mirza, O., Krusong, K., Carpenter, E. P., Rutherford, N. G., Hadden, J. M., O'Reilly, J., Ma, P., Saidijam, M., Patching, S. G., Hope, R. J., Norbertczak, H. T., Roach, P. C., Iwata, S., Henderson, P. J., and Cameron, A. D. (2008) *Science* 322, 709-713
59. Shimamura, T., Weyand, S., Beckstein, O., Rutherford, N. G., Hadden, J. M., Sharples, D., Sansom, M. S., Iwata, S., Henderson, P. J., and Cameron, A. D. (2010) *Science* 328, 470-473
60. Gao, X., Zhou, L., Jiao, X., Lu, F., Yan, C., Zeng, X., Wang, J., and Shi, Y. (2010) *Nature* 463, 828-832
61. Faham, S., Watanabe, A., Besserer, G. M., Cascio, D., Specht, A., Hirayama, B. A., Wright, E. M., and Abramson, J. (2008) *Science* 321, 810-814
62. Watanabe, A., Choe, S., Chaptal, V., Rosenberg, J. M., Wright, E. M., Grabe, M., and Abramson, J. (2010) *Nature* 468, 988-991
63. Forrest, L. R., Kramer, R., and Ziegler, C. (2011) *Biochimica et biophysica acta* 1807, 167-188
64. Bednarczyk, D. (2010) *Anal Biochem* 405, 50-58

65. Andersen, J., Kristensen, A. S., Bang-Andersen, B., and Stromgaard, K. (2009) *Chem Commun (Camb)*, 3677-3692
66. Kristensen, A. S., Andersen, J., Jorgensen, T. N., Sorensen, L., Eriksen, J., Loland, C. J., Stromgaard, K., and Gether, U. (2011) *Pharmacol Rev* 63, 585-640
67. Albers, J. W., and Fink, J. K. (2004) *Muscle Nerve* 30, 410-422
68. Lindberg, R. L., Martini, R., Baumgartner, M., Erne, B., Borg, J., Zielasek, J., Ricker, K., Steck, A., Toyka, K. V., and Meyer, U. A. (1999) *J Clin Invest* 103, 1127-1134
69. Bermudez Moretti, M., Correa Garcia, S. R., Chianelli, M. S., Ramos, E. H., Mattoon, J. R., and Batlle, A. (1995) *Int J Biochem Cell Biol* 27, 169-173
70. Musiol, R., Serda, M., and Polanski, J. (2011) *Curr Pharm Des* 17, 3548-3559
71. Stummer, W., Pichlmeier, U., Meinel, T., Wiestler, O. D., Zanella, F., and Reulen, H. J. (2006) *Lancet Oncol* 7, 392-401
72. Fukuda, H., Casas, A., and Batlle, A. (2005) *Int J Biochem Cell Biol* 37, 272-276
73. Melamed, N., and Kanner, B. I. (2004) *Molecular pharmacology* 65, 1452-1461

SUPPLEMENTAL INFORMATION

Template selection. The structure of LeuT was determined by X-ray crystallography in different conformations, with a variety of ligands [76, 77]. We selected the template structures (PDB identifier 2A65 [176] and 3F3A [78] for the occluded and outward-facing conformations, respectively) based on the following criteria: (i) highest resolution (1.65 Å and 2 Å), (ii) substrate (leucine)- and inhibitor (tryptophan)-bound, and (iii) an occluded and an outward-facing conformations.

GAT-2-LeuT alignment. An initial GAT-2-LeuT alignment was extracted from a comprehensive comparison of the SLC6 family including eukaryotic and prokaryotic members [99]. The alignment was corrected with the updated sequence of the active isoform of GAT-2 (April 2011) and was subsequently refined; four long segments distant from the primary binding site (extracellular loop 2, the loop between transmembrane helices 11 and 12, and the N- and C-termini) were excluded from modeling (Fig. 3-S1). The sequence identity between the modeled fraction of GAT-2 and LeuT is 23%.

Sidechain refinement. For each one of the initial models the sidechains of the binding site residues were repacked on a fixed backbone using SCWRL4 [187]. The coordinates of the sodium ions from the initial models were used as steric constraints for the sidechains. We ran SCWRL4 on different combinations of residues, including E48 alone, four binding site residues (ie, E48, F288, L294, Q391), and residues located within 4 Å from the coordinates of the ligand leucine in the initial model. For the outward facing model, we also ran SCWRL4 on an additional residue separately (ie, D447), three binding site residues (E48, Q391, and D447), or the whole protein.

MD simulations. All MD simulations were performed with GROMACS4 MD code [188]. Each model, including pre- or post-sidechain optimization models, was refined using the following MD protocol. The model was subjected to 10000 steps of conjugate gradient minimization under the Amber99SB-ILDN force field [189, 190]. To account for the membrane hydrophobic environment, an implicit model for the solvent based on a generalized Born formalism was used. A dielectric constant equal to 2 was used to model the membrane interior. Whenever ions were present in the model, the system was simulated “in vacuum” and coulomb interactions were screened by using a dielectric constant equal to 2. All bond lengths were constrained to their equilibrium values using the LINCS algorithm [191]. A time step of 2 fs was adopted. Temperature was enforced using a stochastic thermostat [192] with a coupling time of 0.1 ps. A cutoff of 1.0 nm was used for the Lennard-Jones and the electrostatic interactions.

Enrichment. The enrichment factor was defined as:

$$EF_n = \frac{(\text{ligand}_{\text{selected}} / N_n)}{(\text{ligand}_{\text{total}} / N_{\text{total}})} \quad (1)$$

where $\text{ligand}_{\text{total}}$ is the number of known ligands in a database containing N_{total} compounds, and $\text{ligand}_{\text{selected}}$ is the number of ligands found in a given subset of N_n compounds [108, 109, 138]. Additionally, we used the area under the enrichment curve as a measure of virtual screening accuracy [108, 109, 138]. The enrichment curve can be obtained by plotting the percentage of actual ligands correctly predicted (y-axis) within the top ranked subset of all database compounds (x-axis on logarithmic scale). We calculated the area under the curve (logAUC) of the enrichment plot for $\Delta x=0.1$:

$$\log\text{AUC} = \frac{1}{\log_{10} 100 / 0.1} \sum_{0.1}^{100} \frac{\text{ligand}_{\text{selected}}(x)}{\text{ligand}_{\text{total}}} \Delta x \quad \text{where } x = \log_{10} \frac{N_n}{N_{\text{total}}} \quad (2)$$

SI FIGURE LEGENDS

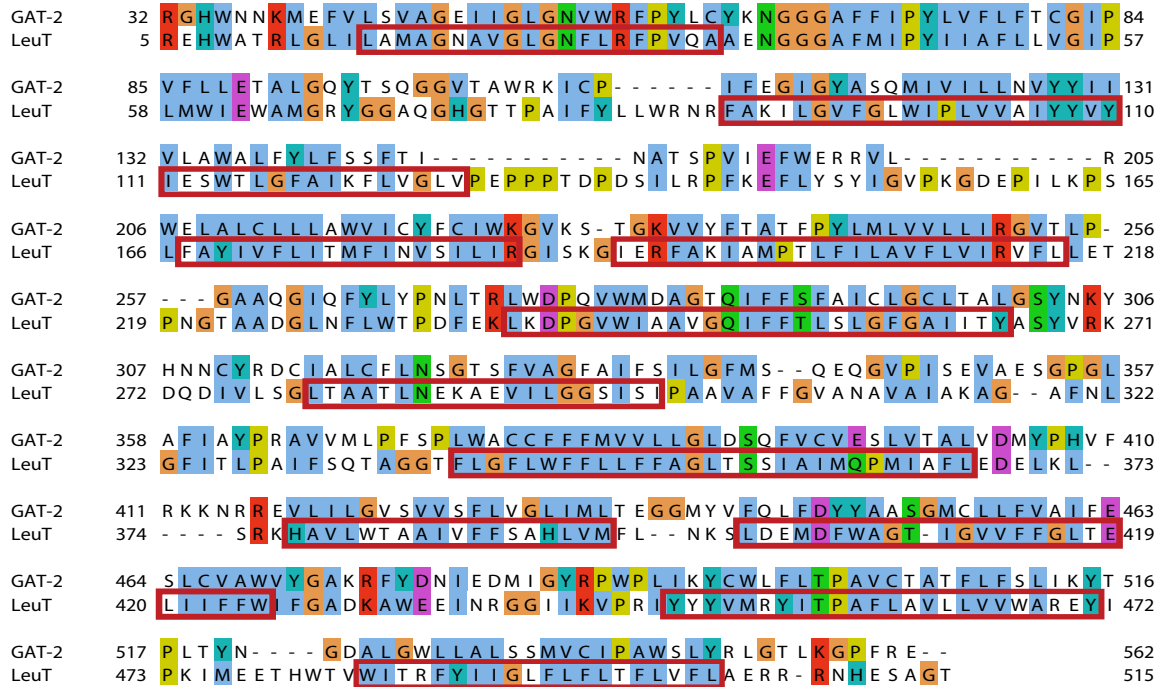


Fig 3-S1. GAT-2-LeuT alignment.

Putative transmembrane regions (in red rectangles) were obtained from the comprehensive SLC6 study [99], and were originally based on the PDB_TM database [193]. The alignment was corrected and refined with an updated GAT-2 sequence. The LeuT sequence was derived from occluded X-ray structure (ie, PDB id 2A65).

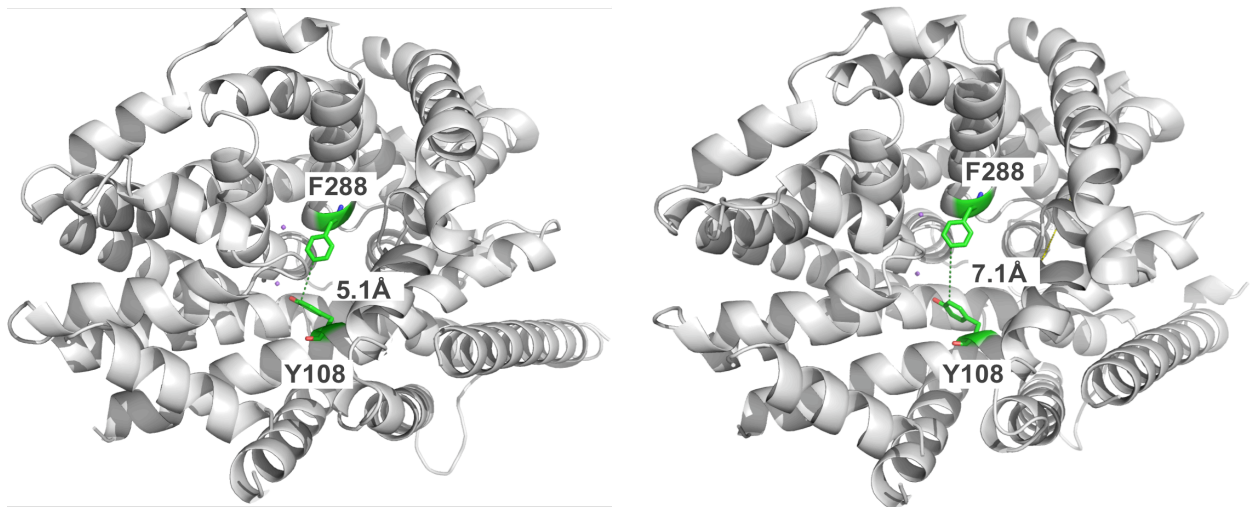


Fig 3-S2. The extracellular gate.

Top view of the extracellular gate formed by F288 and Y108 (green sticks) in the occluded (left panel) and outward-facing (right panel) conformations. All other GAT-2 residues are illustrated by white ribbons.

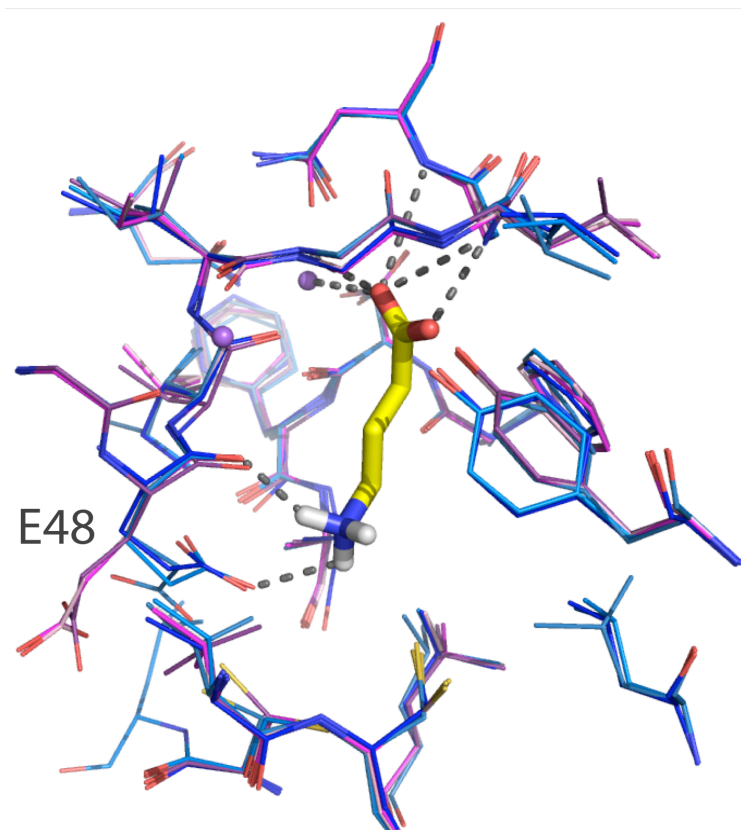


Fig 3-S3. E48 in various models of GAT-2.

The three occluded models with the best (blue) and worst (pink) enrichment scores are shown. Atoms are illustrated by lines, with oxygen, nitrogen, and hydrogen atoms in red, blue, and white, respectively. The sodium ions Na1 and Na2 are visualized with purple spheres. GABA is depicted in yellow sticks and its hydrogen bonds with GAT-2 in the final refined (involving E48, G51, G53, N54, and Na1) are shown as dotted gray lines.

REFERENCES

1. Nyola, A., Karpowich, N. K., Zhen, J., Marden, J., Reith, M. E., and Wang, D. N. (2010) *Curr Opin Struct Biol* 20, 415-422
2. Krishnamurthy, H., and Gouaux, E. (2012) *Nature* 481, 469-474
3. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y., and Gouaux, E. (2005) *Nature* 437, 215-223
4. Singh, S. K., Piscitelli, C. L., Yamashita, A., and Gouaux, E. (2008) *Science* 322, 1655-1661
5. Beuming, T., Shi, L., Javitch, J. A., and Weinstein, H. (2006) *Mol Pharmacol* 70, 1630-1642
6. Krivov, G. G., Shapovalov, M. V., and Dunbrack, R. L., Jr. (2009) *Proteins* 77, 778-795
7. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) *Journal of chemical theory and computation* 4, 435-447
8. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) *Proteins* 78, 1950-1958
9. Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y., and Wriggers, W. (2010) *Science* 330, 341-346
10. Hess, B. (2007) *Journal of chemical theory and computation* 4, 116-122
11. Bussi, G., Donadio, D., and Parrinello, M. (2007) *J Chem Phys* 126, 014101

12. Huang, N., Shoichet, B. K., and Irwin, J. J. (2006) *J Med Chem* 49, 6789-6801
13. Irwin, J. J., and Shoichet, B. K. (2005) *J Chem Inf Model* 45, 177-182
14. Fan, H., Irwin, J. J., Webb, B. M., Klebe, G., Shoichet, B. K., and Sali, A. (2009) *J Chem Inf Model* 49, 2512-2527
15. Tusnady, G. E., Kalmar, L., and Simon, I. (2008) *Nucleic acids research* 36, D234-239

Chapter 4: Discovery of Potent, Selective Multidrug And Toxin Extrusion Transporter 1 (MATE1, SLC47A1) Inhibitors Through Prescription Drug Profiling and Computational Modeling

Matthias B. Wittwer,^{‡,a} Arik A. Zur,^{‡,a} Natalia Khuri,^{‡,b} Yasuto Kido,^c Alan Kosaka,^d
Xuexiang Zhang,^d Kari M. Morrissey,^a Andrej Sali,^b Yong Huang,^{*,d} Kathleen M. Giacomini^{*,a}
[‡] these authors contributed equally to this study

a University of California, San Francisco, Department of Bioengineering and Therapeutic Sciences, RH 581, 1550 4th Street, San Francisco, CA 94158, United States of America

b University of California, San Francisco, Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences (QB3), BH 503B, 1700 4th Street, San Francisco, CA 94158, United States of America

c Drug-Drug Interaction Group, Drug Metabolism and Pharmacokinetics, Shionogi & Co., Ltd., 3-1-1, Futaba-cho, Toyonaka-shi, Osaka 561-0825, Japan

d Optivia Biotechnology Inc., 115 Constitution Drive, Suite 7, Menlo Park, CA 94025

ABSTRACT

The human multidrug and toxin extrusion (MATE) transporter 1 contributes to the tissue distribution and excretion of many drugs. Inhibition of MATE1 may result in potential drug-drug interactions (DDIs) and alterations in drug exposure and accumulation in various tissues. The primary goals of this project were to identify MATE1 inhibitors with clinical importance or

in vitro utility and to elucidate the physicochemical properties that differ between MATE1 and OCT2 inhibitors. Using a fluorescence assay of ASP+ uptake in cells stably expressing MATE1, over 900 prescription drugs were screened and 84 potential MATE1 inhibitors were found. We identified several MATE1 selective inhibitors including four FDA-approved medications that may be clinically relevant MATE1 inhibitors and could cause a clinical DDI. In parallel, a QSAR model identified distinct molecular properties of MATE1 versus OCT2 inhibitors and was used to screen the DrugBank in silico library for new hits in a larger chemical space.

INTRODUCTION

Transporters on the plasma membrane of cells are major determinants of drug absorption, elimination, and accumulation in various tissues. In the past decade, many studies have been devoted to understanding and characterizing transporters involved in pharmacokinetic processes, primarily in the intestine, kidney, and liver (e.g., P-gp, OATs, OATPs).(1) Information from these studies has enhanced our understanding of drug absorption and elimination phenomena.(2) In addition, transporter-related drug accumulation and drug toxicity studies have formed the basis for understanding and predicting clinically important drug-drug interactions (DDIs). With respect to the kidney, several solute carrier transporters (SLCs) for both anions (e.g., OAT1 and OAT3) and cations (e.g., OCT2, MATE1, and MATE2-K) have been shown to be of clinical importance for drug elimination.(1,3–5) This manuscript focuses on the human multidrug and toxin extrusion transporters MATE1 and MATE2-K as well as the organic cation transporter OCT2 which are important in the disposition of positively charged drugs.

The MATE transporter family consists of two members, namely MATE1 (SLC47A1) and MATE2 (SLC47A2), and has only relatively recently been described.(6,7) Both transporters contribute to the urinary excretion of structurally diverse organic cations at the apical membrane

of renal cells. MATE1 is widely expressed in body tissues, including the kidney, liver, skeletal muscle, adrenal gland, and testis. In the kidney, MATE1 is localized to the apical membrane of the proximal and distal convoluted tubule.⁶ Three major splice variants of MATE2 have been identified, MATE2, MATE2-B, and MATE2-K. Of these, only MATE2 and MATE2-K demonstrate functional activity in cell-based experiments.^(7,8) MATE2-K consists of only 566 amino acid residues (as opposed to 602 amino acid residues in MATE2) and is almost exclusively expressed in the apical membrane of human proximal tubule. Both MATE1 and MATE2-K act as electroneutral exchangers of their substrates with an oppositely directed proton gradient as a driving force.

Of the organic cation transporters OCT1 through 3, OCT2 is most strongly expressed in the kidney while OCT1 is predominantly expressed in the liver and OCT3 is expressed in many different tissues.⁽³⁾ In epithelia, OCTs are generally localized to the basolateral membrane and mediate the uptake of organic cations into the cells by facilitated diffusion driven by an inside-negative membrane potential. In the kidney, OCTs and MATEs have been shown to act concertedly in the excretion of organic cations. Their substrate and inhibitor specificity are supposed to overlap heavily.^(9,10)

Recent studies to identify and compare the affinity and potency of MATE and OCT substrates and inhibitors have been published (for a review see Nies et al.).⁽³⁾ Several articles concerning ligand selectivity for MATE1 compared to MATE2-K have been published using different methodologies, ranging from classical experimental procedures¹¹ to pharmacophore-based ligand identification.⁽¹²⁾ These studies, which provide interesting models for predicting inhibitors of MATEs, were based on small libraries with a maximum of about 60 compounds and were not focused on prescription drugs. Previous studies to identify inhibitors of OCT1 and

OCT2 are more extensive with screens that included larger numbers of inhibitors and many prescription drugs.(13–15) However, little work has been performed on substrate and inhibitor overlap between OCTs and MATEs. Even though few examples of selective inhibitors have been described¹⁶ more probes for both in vitro and in vivo applications are required.

Therefore, the primary goal of the current study was to use high throughput screening (HTS) to identify inhibitors of MATE1 that can be used as in vitro and in vivo probes. The screen was complemented by a quantitative structure-activity relationship (QSAR) model based on the random forest (RF) methodology⁽¹⁷⁾ for the prediction of MATE1 inhibitors. This approach led to the identification of novel potent and selective inhibitors of MATE1. A special emphasis was placed on drugs that may potentially cause clinical drug-drug interactions. The International Transporter Consortium (ITC) has issued guidelines for selected transporters (OCT2, P-gp, BCRP, OAT1, OAT3, OATP1B1, OATP1B3) that define when a clinical DDI study should be conducted.⁽²⁾ According to these guidelines, if the ratio $C_{max,unbound} / IC_{50}$ is greater than or equal to 0.1 then a clinical DDI study should be performed. Although, to date, no guidelines for MATEs are available, the ITC is considering making similar recommendations for these transporters. Therefore, in this manuscript we used the threshold of ≥ 0.1 to identify drugs that may cause clinically significant DDIs.

A secondary goal was to compare properties of inhibitors of MATE1 with those of OCT2 which was screened in a previously published study from our laboratory.⁽¹⁵⁾ This study provides novel insights into the inhibitor specificity profiles of organic cation transporters, including their charge selectivity and required physicochemical properties.

RESULTS

High Throughput Screen for MATE1-Inhibitors with ASP+ as Fluorescent Probe

A high throughput screening (HTS) to identify inhibitors of MATE1 was performed using the fluorescent probe ASP+. The uptake assay of ASP+ in cells over-expressing MATE1 was characterized and optimal experimental conditions were derived (i.e., duration of the uptake experiment and ASP+ concentration; Methods) (Fig. 4-1A and 4-1B). In particular, 1.5 minutes was selected to perform the screen because it was in the linear range of transport (Fig. 4-1A). An ASP+ concentration of 2 μM was selected for the screening studies because it was below the K_m of 34 μM (Fig. 4-1B). The functionality of the assay for inhibition studies was confirmed by determining the IC_{50} value of cimetidine, a well-established MATE1 inhibitor. The IC_{50} was $1.2 \pm 0.25 \mu\text{M}$ which is in agreement with published data^{6,18,19} (Fig. 4-1C). For the screening, the probe uptake was determined in the presence of compounds of the ICONIX library which contains 910 FDA approved drugs from 124 therapeutic classes.⁽¹⁵⁾ We identified 84 compounds that exhibited 50% or more inhibition of ASP+ uptake at 20 μM . These compounds were considered as strong inhibitors (hits) of MATE1 (Fig. 4-1D) and selected compounds were subjected to follow-up IC_{50} determinations.

Physicochemical Properties of MATE1 HTS Hits and Comparison to OCT2 HTS Hits

The MATE1 inhibitors identified in the present study were compared to the OCT2 inhibitors identified previously using the same fluorescent probe and the same library.⁽¹⁵⁾ Interestingly, OCT2 exhibited a broader inhibition pattern (Fig. 2A) and considerably fewer compounds were found to inhibit MATE1 as compared to OCT2 (Fig. 4-2B). Fig. 2A shows the amount of compounds per percent inhibition and is thus reflective of the distribution of inhibitors

and non-inhibitors over the dataset. For example, it becomes apparent that not only there are more OCT2-inhibitors but also that they have more variable percent-inhibition compared to MATE1 inhibitors. 45 selective MATE1 inhibitors that did not affect OCT2 transport were identified based on the screening results. In contrast, 205 selective OCT2 inhibitors and 39 dual MATE1/OCT2 inhibitors were found (Fig. 4-2B).

The existence of distinct groups of transporter inhibitors suggests that elucidating differences in structural and physicochemical properties may help in understanding the properties that lead to transporter-selectivity. We therefore used several computational methods to characterize and differentiate MATE1 or OCT2 selective inhibitors as well as inhibitors of both transporters (dual inhibitors) and non-inhibitors.

First, to examine differences between these groups, we performed a principal component analysis (PCA) and compared OCT2 and MATE1 selective inhibitors to dual inhibitors and to the whole ICONIX-library (Fig. 4-2C). More specifically, we analyzed various physicochemical properties, such as molecular weight, molecular volume, number of heavy atoms, number of rotatable bonds, number of hydrogen bond donors and acceptors, SLogP, topological polar surface area and charge at pH 7.4 (Supplementary Information, Table 2). The first two principal components (PC1 and PC2 in Fig. 4-2C) explain ~70% of variability in physicochemical properties between transporter selective inhibitors and dual inhibitors or non-inhibitors. The first principal component is governed by polarity, size and hydrogen bonding, the second principal component is mainly represented by lipophilicity. The bottom right quadrant in Fig. 4-2C, which contained large, hydrophobic, positively charged compounds, contained a large number of dual and MATE1 selective inhibitors. Smaller, positively charged compounds cluster in the bottom left quadrant. Compounds in the upper left quadrant are negatively charged. The PCA analysis

proved helpful in identifying interesting compounds for further testing and, in particular, compounds that showed to be selective MATE1 or OCT2 inhibitors. Many promising inhibitors with strong inhibition values in the screen, such as indinavir, dihydroergotamine, and vecuronium bromide, clustered in the lower right quadrant. Leucomycin was selected for further testing based on proximity to these compounds. Follow-up testing revealed that this drug was not only a MATE1 inhibitor but was also selective over MATE2-K and OCT2 (Table 4-1). The cluster in the lower left quadrant contained many compounds with medium to high IC₅₀ values, such as ethinyl estradiol, but also some highly active inhibitors like rimantadine. Finally, the upper left quadrant contained many true negative and false positive drugs with respect to MATE1 inhibition, including niacin and ciclopirox. The latter was selected for a confirmatory experiment based on its association with this cluster.

As a second analysis, we examined the physicochemical and structural properties of the inhibitors of MATE1, OCT2 or dual inhibitors, compared with the non-inhibitors (Fig. 4-3A-E). No statistically significant differences were observed for the numbers of H-bond donors, H-bond acceptors or rotatable bonds (Supplementary Information, Fig. 1). With respect to molecular weight, the number of heavy atoms and the molecular volume, MATE1 selective inhibitors (but not OCT2 selective inhibitors or inhibitors of both transporters) showed statistically significant ($p < 0.05$) higher values compared to non-inhibitors. Highly statistically significant differences ($p < 0.001$) were observed for SLogP, topological polar surface area (TPSA), and charge. In particular, all three groups exhibited significantly higher values of SLogP, a measure of lipophilicity, in comparison to non-inhibitors. Interestingly, TPSA values were lower for OCT2 selective inhibitors compared to both MATE1 selective inhibitors and non-inhibitors. Finally, at

pH 7.4 positively charged compounds appeared more frequently in the groups of OCT2-selective and dual inhibitors compared to the non-inhibitor group.

For the third analysis, we binned the compounds of each group (i.e., MATE1 selective inhibitors, OCT2 selective inhibitors, dual inhibitors, and non-inhibitors) into bases, acids, zwitterions, neutral, and unknown (Fig. 4-3F-J). As expected for cation transporters, such as MATE1 and OCT2, bases were overrepresented in the inhibitor groups compared to the whole ICONIX library. The fraction of inhibitors that were bases was highly enriched for OCT2 selective ($p < 1 \times 10^{-14}$) and dual inhibitors ($p < 1 \times 10^{-7}$) in comparison to the ICONIX library. Bases were also over-represented among the MATE1 selective inhibitors, but at lower significance levels ($p < 0.05$). Interestingly, zwitterions (e.g., famotidine, telmisartan) were over represented in the OCT2 selective inhibitor group ($p < 0.01$), but not in the other groups. As expected, acids were overrepresented in the non-inhibitor groups to a highly statistically significant extent ($p < 1 \times 10^{-16}$), and the same held true for neutral compounds though the significance level was much lower ($p < 0.05$).

Validation of HTS Screen by Follow-up IC₅₀ Determination

To test the quality of the screening assay as well as to enhance and validate the *in silico* model development, we determined the IC₅₀ values of various promising drugs against MATE1, MATE2-K, and OCT2 (Table 4-1). These drugs were selected based on their predicted MATE1 IC₅₀ values (prIC₅₀), their pharmacological interest (e.g., novelty or how commonly they are used) and/or their potential to inhibit other cation transporters (based on published data). Interestingly, for the library used in this study, we identified more compounds that exhibited MATE1 selectivity over MATE2-K than over OCT2 (e.g., gabaxate, granisetron, rimantadine, vecuronium bromide). In Table 4-1, only topotecan seems to have a slight preference for both

MATEs compared to OCT2. Furthermore, ethinyl estradiol proved to be a better OCT2 than MATE inhibitor and chlorhexidine as well as bithionol were pan-inhibitors, i.e., inhibiting MATE1, MATE2-K, and OCT2. Moreover, the IC₅₀ values for OCT1 and OCT3 of chlorhexidine were 0.21 and 0.41 and for bithionol 3.9 and 5.5 μ M, respectively (data not shown). Nevertheless, selective inhibitors for MATE1 were identified and include prazosine, buspirone, leucomycine, and the drugs inhibiting at clinically relevant concentrations discussed below. Many of the compounds in Table 4-1 have not, to the best of our knowledge, been described as inhibitors of organic cation transporters (e.g., domperidone, camostat, dihydroergotamine, epinastine, gabaxate) or have only been shown to interact with the OCTs but not the MATEs (e.g., prazosin, granisetron). It is noteworthy that some compounds in Table 4-1 were close to clinical significance. For example, tubocurarine has a C_{max,unbound} /IC₅₀ of 0.07, which is near the threshold of 0.1.

Development and Utilization of Computational Model (RF-Model-I) to Ratify and Complement HTS Results

To aid in the targeted validation of the screening hits, we developed an RF classification model of MATE1 inhibitors and non-inhibitors as identified in the HTS. We estimated the performance of the RF algorithm by means of internal cross-validation, using out-of-bag samples, and external validation using ten test sets (Methods). The Receiver Operating Characteristic (ROC) curves for external validation tests are shown in Fig. 4A. The average area under the ROC curve (AUC) for the ten tests was 0.70 \pm 0.05 (permutation test p-value < 0.0001), indicating good quality models. For comparison, we built MATE1 classification models using a k-nearest neighbor (kNN) algorithm and a partial least-squares discriminant analysis (PLS-DA) algorithm. The kNN algorithm is a simple but effective method for similarity clustering and

classification. The PLS-DA technique was evaluated because it was successfully used in QSAR modeling of other transporters.^{15,20} The average AUCs for the kNN (AUC=0.62 ±0.05) and PLS-DA (AUC=0.66 ±0.04) models were significantly lower than for the RF models (Supplementary Fig. 4-2 and Supplementary Table 4-3).

Next, we used the entire HTS data set to train an RF model (RF-Model-I), which we employed in several ways. First, we tested the model's ability to rescue false negative compounds with respect to MATE1 inhibition in the screen. As Table 4-2A illustrates, the model performed well at this task over a broad range of screening-based inhibition values (5-49% inhibition). Remarkable examples were ranitidine and irinotecan, which in the follow-up experiments inhibited MATE1 in the low micromolar range but did not inhibit ASP+ uptake by more than 50% in the screen. The only incorrectly predicted false negative compound was methyl ergonovine, which was surprising because it belongs to the ergot alkaloid-family, which includes several inhibitors of MATE1 (e.g., dihydroergotamine). The rescued negative hits were tested for inhibition of other transporters and several were OCT2 selective. Both, carvedilol and noscapine had IC₅₀ values against OCT2 of 7.5 and 2.6 μM, respectively, and did not inhibit MATE2-K up to 500 μM (data not shown). Although we did not find any additional selective and/or clinically relevant inhibitors, this approach proved useful in rescuing otherwise overlooked inhibitors.

Second, we tested the model's ability to predict false positive screening hits. To do so, we studied compounds that had been assigned a very low score by RF model-I despite >50% inhibition in the initial screen. As Table 4-2B shows, the model performed well at this task. It is noteworthy that compounds with IC₅₀ values above 100 μM (e.g., phenacetin) might still be substrates, because the screening methodology used in this study is primarily powered to identify

inhibitors. For example, when we assessed the inhibition of ASP+ uptake by metformin, we found an IC50 value of 250 μ M (data not shown).

Third, we assessed the model's performance in identifying true negative compounds. Table 4-2C illustrates that the model is also suited well for this task, over a wide array of compounds including cations (e.g., lidocaine) and zwitterions (e.g., furosemide).

Finally, we also identified mismatched compounds based on the RF-Model-I predictions, i.e., compounds that were either predicted to be inhibitors by the model but could not be confirmed experimentally (e.g., irbesartan) or inhibitors confirmed by experiments that were not detected by the model (e.g., propranolol). Table 4-2D gives an overview of this compound category.

We used the feedback from the experiments to refine the RF models and to generate an improved version (i.e., RF-Model-II; see next paragraph) for further use. Upon refinement with experimental results, the performance of the RF models significantly improved, as assessed in ten external tests (Fig. 4-4A-B). The average AUC of the refined RF models increased to 0.78 \pm 0.02 (Student's t-test p-value = 0.0003812). Again, we compared the performance of the RF algorithm to those for the kNN and PLS methods. Although the AUCs improved for all three methods, our analysis shows that the RF models performed better in external validation tests (Supplementary Table 3 and Supplementary Fig. 2).

Use of RF-Model-II to Screen a Large *in silico* Library

After improving the initial model based on feedback from experimental testing, we used the new QSAR model (RF-Model-II) for screening the DrugBank *in silico* library(21–23) to explore a larger chemical space than represented by the ICONIX library. Because the applicability domain of the RF model is limited to molecules with physicochemical properties

similar to those for the training data set,(24–27) we projected the DrugBank compounds onto the principal components' space of the ICONIX library. We then removed DrugBank molecules with physicochemical properties outside of the 95% confidence interval of the first two principal components computed for the ICONIX library (Supplementary Fig. 3).

Next, we classified the remaining 6,122 DrugBank compounds using our RF-Model-II. Five *in silico* hits, selected based on their diversity from ICONIX-compounds, their novelty, and on their therapeutic interest, were tested successfully as inhibitors (Table 4-3). Four of the five selected compounds were MATE1 selective inhibitors, consistent with our model. Only pimozide was a false positive. Risperidone and maraviroc were potent and selective MATE1 inhibitors. Even though not fulfilling our criteria for clinical significance, maraviroc gets close ($C_{max,unbound} / IC_{50} = 0.02$ with 300 mg b.i.d.). In addition, concentrations in the liver (where MATE1 is also expressed) might be even higher due to the elevated concentrations in the portal vein. Maraviroc seems especially intriguing because it is a novel drug from a new therapeutic class (CCR5 receptor-antagonist/entry inhibitor) used to treat HIV-infections.(23,24) To confirm that none of the five *in silico* hits could be found using 2D similarity and substructure searches, we used DrugBank's web interface to retrieve all compounds similar to MATE1 inhibitors. Each MATE1 inhibitor was used as a query for the search. We used 0.6 as a threshold for the Tanimoto similarity coefficient (T_c), which is the default search option in DrugBank. Only pimozide was found using 2D similarity and substructure searching. The pairwise similarity coefficients of *in silico* hits ranged from below 0.3 (maraviroc and dapiprazole) to 0.724 (pimozide and droperidol) (Table 4-3). In conclusion, the *in silico* screening approach using RF-Model-II against the DrugBank-library proved to be helpful and yielded several interesting hits that were confirmed experimentally.

Dragon Descriptors Important for RF Models of MATE1 Inhibitors

The final set of twenty-one descriptors of RF-Model-II was selected using a backward feature elimination algorithm (Table 4-4). The importance of each descriptor was evaluated by the decrease in the mean accuracy of the model in the descriptors' permutation tests (Supplementary Fig. 4-4, Methods).(30) Two topological descriptors, the Balaban distance connectivity index and the Balaban V index, as well as molecular weight were the most important for the performance of the model (Supplementary Fig. 4-5). Topological and constitutional descriptors dominated the final set of descriptors. Interestingly, none of the 3D descriptors was present among the optimal selection and only one substructure fragment was selected. Ten topological and seven constitutional descriptors differed significantly between MATE1 inhibitors and non-inhibitors, as assessed using the nonparametric Mann-Whitney U test (Table 4-4). Unfortunately, it is not possible to interpret topological indices in terms of physicochemical properties. The analysis of constitutional descriptors showed that MATE1 inhibitors tend to be larger (higher molecular weight, higher number of bonds, rings and longer circuits) and less electronegative than non-inhibitors. The importance of lipophilicity for MATE1 inhibitors is again evident from higher MLOGP coefficients. Additionally, the number of nine-membered nitrogen-containing rings (nR09 descriptor) is greater among MATE1 inhibitors. The nR09 descriptor also had higher values among MATE1 selective inhibitors versus OCT2 selective (p-value = 0.0003082) and dual (p-value = 0.004386) inhibitors.

Clinically Relevant MATE1 Inhibitors and Their Implications for Metformin DDIs

A major goal of this study was to identify selective, potentially clinically relevant MATE1 inhibitors. Specificity was assessed by determining IC₅₀ values for MATE1 as well as MATE2-K, OCT1, OCT2, and OCT3.

We identified four drugs that, at clinically relevant plasma concentrations, selectively inhibited MATE1 (i.e., the IC₅₀ value was at least five-fold lower than that obtained for the other cation transporters). Of these four drugs (indinavir, famotidine, ritonavir, and imatinib), indinavir has so far not been shown to interact with MATE1 and thus is a potential new model inhibitor at clinically relevant plasma concentrations. Of six additional drugs studied, three (i.e., irinotecan, mitoxantrone, ondansetron) inhibited MATE1 at clinical plasma concentrations but were not selective towards MATE1 (Figs. 4-5 and Table 4-5). Interestingly, mitoxantrone and ondansetron proved to be selective for both MATE1 and MATE2-K over the OCTs, and irinotecan was a dual inhibitor for OCT2 and MATE1. The other three compounds were clinically significant inhibitors for either MATE2-K (i.e., nifekalant) or OCT2 (i.e., pantoprazole, pentamidine). Pantoprazole was truly selective for OCT2 over the MATEs, which is of interest since known DDIs in the kidney now appear to involve MATEs and questions about clinically relevant DDIs involving OCT2 have been raised.(18) Though clinical studies are needed, our data suggest that an OCT2 mediated DDI may occur between pantoprazole and metformin.

Because the above results were achieved with the probe-fluorophore ASP+, which is not relevant clinically, we used ¹⁴C-metformin to more appropriately identify the potential for DDIs. IC₅₀ values were assessed in two cellular models, HEK293 cells and polarized MDCK monolayers in which MATE1 is localized on the apical membranes (data not shown). The latter model may more accurately represent MATE1 activity in renal proximal tubule cells in vivo since it is based on polarized cells.(31,32) The K_m values of ¹⁴C-metformin obtained with the HEK293 cells and with the polarized MDCK-II model were 227 μM³³ and 114 μM (data not shown), respectively. In general, the IC₅₀ values for inhibition of ¹⁴C-metformin uptake were

similar to or lower than the ones obtained with ASP+, suggesting that the identified interactions are indeed clinically relevant (Table 4-6). In general, the two cell models yielded results that were in good agreement. Notable exceptions were ritonavir for which the IC50 was greater in the MDCK-II model compared with the HEK293 cells and pentamidine for which the IC50 was greater in the HEK293 cells compared with the MDCK-II model. Reasons for these differences may be related to differences in access of the compounds to the MATE1 binding site in the two cell lines.

DISCUSSION

To date, approximately 30 drugs have been shown to interact with MATE transporters and some of them are substrates (e.g., metformin, oxaliplatin, acyclovir).(3) A strong overlap of inhibitors (and substrates) among the organic cation transporters in the OCT family and the MATEs is inferred by many authors based on inhibitors and substrates identified so far.(3,9) However, little is known about structural features that favor inhibition of a specific human cation transporter and lead to selectivity for MATEs over OCTs. The data presented complements two recent publications that address inhibitor selectivity for OCT2 over other cation transporters(15) and for MATE1 over MATE2-K,(12) respectively. In the kidney, predicting drugs that may inhibit MATEs or OCTs is important for predicting DDIs that may potentially lead to nephrotoxic events. In particular, inhibitors of MATE1 would lead to drug accumulation in the kidney and possibly enhance nephrotoxicity, whereas inhibitors of OCT2 would be nephroprotective. The purpose of this study was to screen a prescription drug library to (i) identify novel potent inhibitors of MATE1, (ii) to compare MATE1 inhibitors with inhibitors of another important drug transporter in the kidney (OCT2), (iii) to elucidate the physicochemical properties of MATE1 inhibitors and develop a predictive model to identify MATE1-inhibitors,

and (iv) to identify MATE1 selective inhibitors that may potentially cause DDIs at clinically relevant concentrations.

Only 84 compounds were identified as hits in this large-scale library screen of MATE1 compared with 244 hits identified in our previous screen of OCT2.⁽¹⁵⁾ These findings are in agreement with publications, in which inhibitors of OCT2 and MATE1 were summarized and compared, showing that more drugs strongly inhibit OCT2 than MATE1 in in vitro assays.^(3,9,19) Furthermore charge distribution was also consistent with previously observed patterns and showed that OCT2 inhibitors were significantly enriched within both the basic compounds and the zwitterions groups, whereas among MATE1 inhibitors zwitterions were not enriched and basic drugs were much less enriched compared to OCT2 (Fig. 4-3). Paradoxically, some physicochemical properties of the MATE1 selective inhibitors, such as molecular weight, number of heavy atoms, and molecular volume, were more variable compared to those for the OCT2 selective inhibitors. For example, the median and the 25 - 75% range of molecular weights for OCT2 inhibitors were 309.33 and 269.88 - 369.41, respectively, whereas for MATE1 inhibitors they were 348.48 and 284.40 to 557.83, respectively (Fig. 4-3). These data suggest that MATE1 can be inhibited by a larger variety of compounds with more diverse chemical properties, yet fewer compounds inhibited MATE1 than OCT2. One possible explanation for this apparent inconsistency is that the structural and molecular features required for OCT2 inhibition are more common in the chemical space typically occupied by prescription drugs. This means that prescription drugs might be biased towards structural features that favor OCT2 inhibition. Another possibility is the occurrence of multiple binding sites in MATE1 and considerably fewer in OCT2. Indeed, a recent study by Astorga et al. indicates that this is a

likely assumption since the pharmacophores developed for inhibitors identified with different probes (MPP+ vs. ASP+) were markedly different.(12)

Several of the important physicochemical parameters identified in this study have been found in earlier studies with the rat and rabbit homologues of MATE1. For example, using rabbit brush border membrane vesicles (BBMV) various publications identified positive charge and, to a lesser extent, lipophilicity as important contributors to MATE1-binding.(34–36) Furthermore, studies performed in rats also identified hydrophobicity and charge as major factors for MATE1-binding.(37–40) Nevertheless, it seems unlikely that a single molecular descriptor is sufficient to appropriately describe binding to MATE1, as also noted by Astorga et al.,(12) and QSAR-models offer the advantage of combining different weighted descriptors to yield a more accurate description of the binding process.

The finding that positive charge, lipophilicity, and molecular weight are important contributors to MATE1-binding is also supported by recently developed structural models of MATE1 (41) based on the crystal structure of the NorM-homologue protein found in *Vibrio cholerae*.(42) From the proposed structural model, it becomes evident that the pore, containing the supposed binding sites, is relatively large and predominantly contains negatively charged as well as hydrophobic patches,(41) thus favoring the binding of large, lipophilic, and positively charged compounds as suggested by the results of this study and by previously published QSAR-models based on rodent MATE1-homologues as well as by the pharmacophore-model.

Several studies demonstrated the potential for drug-drug interactions (DDIs) involving MATE1. Indeed, MATE1 knockout mice had significantly increased plasma levels of metformin after intravenous administration, and the urinary excretion of the drug was markedly decreased.(43) In vitro, cimetidine has been reported to interact with metformin(19) and

fexofenadine(44) by inhibiting MATE1 mediated transport. In human volunteers, pyrimethamine, a MATE inhibitor, was reported to significantly increase the plasma concentrations of metformin and alter its renal elimination (45). With respect to in vivo studies, model inhibitors must exert their effect at clinically relevant concentrations. In this study, four drugs have been identified that selectively inhibit MATE1 at clinically relevant unbound concentrations (i.e., $C_{max,unbound} / IC_{50} \geq 0.1$), namely, famotidine, imatinib, ritonavir, and indinavir (Tables 4-4 and 4-5 as well as Fig. 4-5). To the best of our knowledge, indinavir has so far not been described as a MATE1 inhibitor. The other three compounds were all found to be MATE1 selective in previous studies. Imatinib has been previously described as potentially clinically relevant by our group.⁴⁶ This study confirms MATE1 selectivity over OCT3 in addition to MATE2-K, OCT1, and OCT2 that have been reported previously . Interestingly, while both famotidine and cimetidine have been associated with MATE1 selectivity,⁽¹⁹⁾ only the latter was examined with respect to clinical significance and was shown to probably cause the clinically observed DDI with metformin.⁽¹⁹⁾ As Tables 4-4 and 4-5 show, three more compounds were predicted to inhibit MATE1 at clinically relevant concentrations but are not selective over other cation transporters. Nonetheless, they might be of clinical use as model inhibitors and because of their effects on multiple renal transporters may result in clinically significant DDIs. For example, mitoxantrone and ondansetron selectively inhibit both MATEs over the OCTs and irinotecan might act as a dual inhibitor for MATE1 and OCT2. Of these three drugs, ondansetron is to our best knowledge the most potent inhibitor of MATEs (IC_{50} values in nanomolar range, Table 4-6), is least toxic, orally available, and may be a specific in vivo probe for studying MATE1 and MATE2-K mediated renal DDIs with basic drugs. Furthermore, three additional compounds are not clinically relevant MATE1 inhibitors, but are

otherwise of great interest as in vivo tools in preclinical studies. First, nifekalant is predicted to be a clinically significant MATE2-K inhibitor and (if dosed slightly higher) could act as a pan-inhibitor of all three renal cation transporters. Unfortunately, given its use as a class III-antiarrhythmic drug, such a higher dose might practically not be feasible due to the risk of adverse drug reactions. Second, pantoprazole selectively inhibited OCT2 at clinical concentrations. Whereas previous studies indicate that pantoprazole is selective for OCT2 over OCT1 and OCT3 (47), our data extend these findings to the MATEs and suggest that pantoprazole might thus be a useful OCT2 selective model inhibitor based on the IC50 values determined with ASP+. Finally, pentamidine inhibited OCT2 at clinically significant levels and its $C_{\text{max,unbound}}/\text{IC50}$ ratio for MATE1 (i.e., 0.06) was close to clinical significance. Pentamidine might thus be another dual inhibitor for MATE1 and OCT2 besides irinotecan. Its nephrotoxicity, however, might be a serious limiting factor for its possible use as clinical model inhibitor.

To validate the risk of DDIs associated with the inhibitors described above with ASP+ as a substrate, we determined IC50 values against the clinically relevant model substrate, metformin, which is eliminated exclusively in the kidney, using two distinct models. Metformin is among the most highly prescribed drugs in the world, and DDIs in the kidney might potentially increase the risk of its serious adverse drug reaction, lactic acidosis. OCT2, on the basolateral membrane, and MATE1 and MATE2-K on the apical membrane, appear to be the primary transporters responsible for its secretory clearance (10,48). All drugs showed similar or even lower IC50 values with metformin as a substrate compared to ASP+, underscoring the potential clinical significance of these interactions. Astorga et al. recently showed that distinct pharmacophore models of MATE1 were obtained when different substrates, i.e., ASP+ and

MPP+, were used in inhibition assays.(12) The finding that the IC50 values for metformin were different from those of ASP+ for several compounds (Table 4-6) is in agreement with the idea of substrate-dependent pharmacophore models.

Furthermore, some of the 84 HTS-hits are strong and selective MATE1 inhibitors, but are not expected to inhibit MATE1 at clinical concentrations (Table 4-1). For example, leucomycine selectively inhibits MATE1 without affecting MATE2-K or OCT2 while topotecan inhibits both MATE1 and MATE2-K and only to a lesser extent OCT2. In addition, chlorhexidine strongly inhibits all tested cation transporters. These compounds provide valuable *in vitro* tools for predicting the contribution of the various renal transporters to overall renal excretion. Reviewing the HTS-data, it became apparent that MATE1 inhibitors were not necessarily also MATE2-K inhibitors and that the constellation of physicochemical properties that determines inhibitors of MATE2-K is likely distinct from those of MATE1 and OCT2. Interestingly, MATE1 inhibitors tested in follow-up studies more frequently inhibited OCT2 mediated transport rather than affecting MATE2-K, as illustrated by gabaxate, vecuronium bromide or rimantadine. Indeed, in this study we only identified a few MATE-selective inhibitors over OCT2 (e.g., mitoxantrone, ondansetron, topotecan) when at least a five-fold difference in IC50-values was used as the criterion. Previous publications show the same trend (11,12,19). Although the study was focused on the identification of MATE1 selective inhibitors, two compounds that seem to be predominantly OCT2 selective were found. As Table 4-1 shows, ethinyl estradiol exhibits a certain selectivity for OCT2 over the MATEs.(49) The other compound, pantoprazole, has been discussed above. Additionally, we also demonstrate the usefulness of a combined *in silico* and experimental approach with iterative model improvement (i.e., development of different RF-Models based on experimental feedback). This approach proved helpful in identifying false

positive and false negative screening hits and to identify additional MATE1-inhibitors from a large *in silico* library (e.g., maraviroc). None of the inhibitors identified in the current study were tested as substrates of the transporters; therefore, direct experiments testing the uptake of the inhibitors by the relevant transporters as well as their mechanism (e.g., competitive or noncompetitive) are needed.

In summary, we tested 910 prescription drugs for their potential to inhibit MATE1 and identified 84 hits. We identified several MATE1 selective inhibitors (e.g., indinavir) as well as general inhibitors of renal transporters with less selectivity. Compounds expected to cause clinically relevant DDIs were identified including compounds that can be used as selective model inhibitors of the individual transporters. Through the development of predictive models, distinct constellations of physicochemical properties were identified that differentiated OCT2 from MATE1 inhibitors. Several interesting MATE1 inhibiting drugs, not contained in the original screening library, were predicted and then experimentally validated, including maraviroc and risperidone. This study adds new *in vitro* probes as well as possible *in vivo* model inhibitors to clarify drug excretory pathways and enhances the understanding of both OCTs and MATEs.

EXPERIMENTAL SECTION

Reagents.

The ICONIX compound screening library was obtained through the Small Molecule Discovery Center (University of California, San Francisco, CA) and contains 910 FDA-approved drugs.¹⁵ 4-(4-(Dimethylamino)styryl)-N-methylpyridinium iodide (ASP+) was purchased from Molecular Probes (Grand Island, NY). All other reagents were obtained from Sigma-Aldrich (St.

Louis, MI) and Santa Cruz Biotechnology (Santa Cruz, CA) and were of analytical grade with at least 95% purity, unless otherwise stated.

Cell Culture.

Flp-In human embryonic kidney (HEK-293-Flp-In) cell lines stably expressing human MATE1, MATE2-K, OCT1, OCT2 and OCT3 were previously established in our laboratory.^{15,50–52} Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, 100 U/mL penicillin, 100 μ g/mL streptomycin, and 200 μ g/mL hygromycin at 37 °C in a humidified atmosphere with 5% CO₂.

Madin Darby canine kidney type II (MDCK-II) cells were cultured in modified DMEM with 1 g/L glucose and 1 g/L NaHCO₃ (Caisson Labs, North Logan, UT) supplemented with 50 U/mL penicillin, 50 μ g /mL streptomycin and 10% fetal bovine serum, at 37 °C in a humidified atmosphere with 5% CO₂.

Fluorescent-Probe Uptake Assay

MATE1-overexpressing cells were seeded in black poly-D-lysine-coated 96-well plates (Greiner Bio-One, Monroe, NC) and grown for 48 hours until approximately 90% confluent. After washing twice with Hank's Buffered Saline Solution (HBSS), the cells were pre-incubated for 30 minutes in a 30 mM NH₄Cl solution in HBSS at pH 6.5. Uptake was initiated by application of HBSS at pH 7.4 containing the serially diluted fluorescent substrate ASP+ (4-(4-(dimethylamino)styryl)-N-methylpyridinium iodide). After incubating for 1.5 min at room temperature, substrate uptake was stopped by aspirating the incubation mixture and washing the cells twice with ice-cold HBSS containing 500 μ M cimetidine to stop the reaction. Similar experiments were performed using 5 μ M of ASP+ and stopping the reaction at 1, 2, 3, 5, 7, 10,

20 and 30 min. Each test condition was analyzed in triplicate. The intensity of accumulated ASP+ fluorescence was measured using an Analyst AD plate reader (Molecular Devices, Sunnyvale, CA) with excitation and emission filters tuned at 485 and 585 nm wavelength, respectively. Transport kinetics studies for MATE2-K and OCT1, 2, and 3 were performed similarly with the following modifications: No pre-loading of cells with NH₄Cl was required for ASP+ uptake by the OCT-transfected cells and the HBSS stop solution contained 200 μM spironolactone and 200 μM corticosterone for OCT1 and OCT3, respectively. Transport kinetics and Z'-values were calculated as described elsewhere (15).

High throughput Screen (HTS) of Transporter Inhibition

The high throughput screen was performed at the Small Molecule Discovery Center at the University of California, San Francisco. Assay buffers were prepared by diluting 1 mM inhibitor stock solutions in DMSO with HBSS (pH 7.4) containing ASP+ (5 μM) to a final inhibitor concentration of 20 μM (2% DMSO). Nonspecific transport was determined in separate wells on each assay plate using 500 μM cimetidine as inhibitor of MATEs. After subtraction of the nonspecific transport, residual transport rates were used for further calculations. Predicted half-maximum inhibitory concentrations (prIC₅₀) were estimated from the screening inhibition measurements as $V = (V_0 / (1 + [I] / \text{prIC}_{50}))$, where V and V₀ are the activity with and without inhibitor, respectively, and I is the inhibitor concentration of 20 μM. The estimated prIC₅₀ values were considered representative for 20% < V < 80%, where inhibition plots are linear.

Experimental IC50 Determination

Selected compounds, which reduced ASP+ uptake by 50% or more in the HTS, were subjected to experimental IC50 determination. The selection was based on literature-reported transporter selectivity (IC50 values versus MATE2-K, OCT1, OCT2, and OCT3), clinical relevance, chemical structure, and novelty. To determine IC50 values, after pre-incubation as described above, the stably transfected HEK293 cells were simultaneously exposed to ASP+ (2 μ M) and the compound of interest at various concentrations (ranging from 0 to 500 μ M) in HBSS at pH 7.4. The dilutions were prepared from 25 mM DMSO stock solutions and the DMSO concentration was kept constant at 2% in all samples. After 1.5 minutes transport was stopped by removing the uptake media and washing twice with ice-cold stop solution as described above. Then, fluorescence within the cells was determined and IC50 values were calculated using the Prism Software. Values were accepted if the R2 value of the resulting sigmoidal curve was higher than 0.98 and an acceptable standard deviation of the data-points was achieved. The uptake of ASP+ was measured in triplicate at each inhibitor-concentration. IC50 determinations for selected compounds as inhibitors of MATE2-K, OCT1, OCT2, and OCT3 were performed similarly with the following exceptions: No pre-incubation with NH4Cl is required for ASP+ uptake by OCT-transfected cells, the incubation time for OCT-transfected cells was 3 minutes, and the HBSS stop solution in the OCT1 and OCT3 experiments contained 200 μ M spironolactone and 200 μ M corticosterone, respectively.

MATE1 IC50 Determinations With 14C-Labelled Metformin in HEK293 cells

The determinations were performed as described elsewhere (46), with the modification that 28 μ M 14C-metformin (1 μ Ci/ml) were used.

MATE1 IC50 Determinations With 14C-Labelled Metformin in polarized MDCK-II cells

MDCK-II cells were seeded on 96-well porous membrane insert plates (Millipore, Carrigtwohill, Ireland) coated with rat tail type I collagen (Millipore, Temecula, CA), at a seeding density of 60K per well 24 hours prior to transfection. Fully confluent MDCK cell monolayers were transfected with pCI-MATE-1 (accession number NM_018242) or pCI-GFP control plasmid DNA at a final concentration of 30ng/ μ l, using a proprietary Opti-Expression transfection technology (Optivia Biotechnology, Menlo Park, CA). The transiently transfected MDCK cells were incubated for 48 hours to allow the cell monolayers to become polarized. The plates were washed apically and basally three times with warm HBSS (Mediatech, Manassas, VA). After the final wash, the apical sides of the MDCK monolayers were incubated in 150 μ l of HBSS-30 mM NH₄Cl buffer for 20 minutes at 37C on a 60 rpm shaking platform to allow for intracellular acidification. The basolateral side of the insert plate was blotted and left dry. After pre-incubation, the cells were washed once with warm HBSS and replaced with 10 μ M 14C-metformin dosing solution (Moravek Biochemicals, Brea, CA) containing varying concentrations of test inhibitor or 0.5% DMSO as vehicle control. As an inhibition control, 10 μ M cimetidine was added as a reference inhibitor in the presence of 10 μ M 14C-metformin. The plates were placed on a 60 rpm shaking platform for 5 minutes at 37C for 5 minutes. After this time, the cells were immediately washed both apically and basally with three washes of ice cold PBS. To assess 14C-metformin intracellular accumulation, 60 μ l of 50:50 Acetonitrile:H₂O was added to the inserts and shaken at 37C for 15 minutes. A 30 μ l sample of the solubilized cell monolayers was taken for liquid scintillation counting (MicroBeta, Perkin-Elmer, Santa Clara, CA).

***In silico* Modeling**

Three-dimensional structures of 910 compounds were generated using the Corina version 3.20 package (Molecular Networks GmbH, Erlangen, Germany). SMILES strings were used as input, hydrogen atoms were added, small fragments were removed, charges were neutralized, and 100 ring conformations were generated for each compound. The lowest energy conformations of each molecule were used to compute Dragon (Talete s.r.l., Milano, Italy) molecular descriptors. Ten compounds were removed due to computational failure in the Corina or Dragon software. Descriptors with variance near zero were removed and pair-wise correlation coefficients were computed for the remaining descriptors. One randomly chosen descriptor was removed from highly correlated descriptor pairs (correlation coefficient greater than 0.90). All descriptor values were centered and scaled. The random forest (RF) algorithm (17) was used to develop a binary classification model that differentiates between MATE1 inhibitors and non-inhibitors. Class labels were assigned to all compounds. Compounds were assigned either a class label 1 if they showed inhibition greater than 50% in the HTS (inhibitor) or a class label 0 otherwise (non-inhibitor). A training and a test set were constructed by partitioning the ICONIX data set into two non-overlapping subsets (training and test sets) each consisting of 450 compounds. The ratio of inhibitors to non-inhibitors in the subsets was kept the same as in the whole data set (approximately 1:10). Next, a RF model was tuned by the means of out-of-bag validation of the training set (30) Random forest parameters were tuned to maximize the area under the receiver operating characteristic curve (AUC). The best model from the inner loop was tested on the withheld test data set. The entire validation was repeated ten times for different partitioning of data. Descriptor importance was computed and a feature elimination algorithm was used to select the 21 most important molecular descriptors. The algorithm

iteratively removes molecular descriptors in the inner loop if the AUC does not decrease. A permutation test was performed to assess the statistical significance of the final models. This test studies the null hypothesis that the molecular descriptors and the class labels are independent, that is, that there is no difference between the classes (i.e., inhibitors and non-inhibitors). The null distribution under this hypothesis was estimated by permuting the labels of the dataset 1000 times. The results of the permutation test showed that models achieve significantly higher AUCs than expected by chance ($p < 0.0001$). RF models were used to compute a classification score in the range [0,1] for each of the 900 compounds. The score cutoff was calibrated against the classifier error in the outer validation loop and was set to 0.2 for subsequent analyses. The compounds with a score greater than or equal to 0.2 were classified as inhibitors, and compounds with a score less than 0.2 were classified as non-inhibitors. All computational modeling was performed using the caret library in R.

The kNN and PLS-DA models were constructed following the same protocol. The Student's t-test was used to assess the differences between kNN and RF as well as between PLS-DA and RF algorithms.

In silico Model Refinement

The first RF model (RF-Model-I) was refined by incorporating the validation results to yield RF-Model-II. During the refinement, compound labels were updated to reflect new experimental results and new models were developed and tested as described above. To enable comparisons between the initial and refined models, data partitioning was done with the same seed for the random number generator. A paired Student's t-test was used to assess the model improvement.

DrugBank Ligand Screening

Three-dimensional structures of 6,363 compounds were retrieved from the DrugBank database and 21 Dragon molecular descriptors (Table 4-4) were computed for each compound. The final RF model (RF-Model-II) was used to predict the MATE1 inhibition status for each of the DrugBank compounds. Compounds with a score greater than or equal to 0.2 were classified as inhibitors, and compounds with a score less than 0.2 were classified as non-inhibitors.

DrugBank similarity and substructure searches were performed using DrugBank's web interface with default options. Tanimoto coefficients of similarity were computed using the Drugbank interface.

Principal Component Analysis of Physicochemical Properties

Three-dimensional structures of 910 compounds were generated as described above (*In Silico* Modeling). The lowest energy conformations of each molecule were used to compute molecular weight, molecular volume, number of heavy atoms, number of rotatable bonds, number of hydrogen bond donors and acceptors, SLogP, topological polar surface area and charge. ChemAxon Calculator (cxcalc) version 5.11.1 was used to compute the formal charge, ApKa and BpKa at pH 7.4. The compounds were binned into acids (ApKa < 9.4), bases (BpKa > 5.4), zwitterions (ApKa < 9.4 and BpKa > 5.4), and neutral (ApKa > 9.4 and BpKa < 5.4). Compounds for which ChemAxon Calculator failed to compute ApKa and BpKa values were denoted as unknown. Principal component analysis was performed with the printcomp function in the stats library in R.

Statistical Analysis

Statistical differences in physicochemical properties of OCT2 inhibitors, MATE1 inhibitors, dual inhibitors and non-inhibitors were assessed with the Dunnett-Tukey-Kramer pairwise multiple comparison test adjusted for unequal variances and unequal sample sizes. Over- or under-representation of bases, acids, zwitterions, and neutral compounds among OCT2 inhibitors, MATE1 inhibitors, dual inhibitors and non-inhibitors was determined with the hypergeometric test. Differences between AUC values of ten independent test sets of two different models were assessed with the Student's t-test. Differences between properties of Dragon descriptors of MATE1 inhibitors and non-inhibitors were assessed using nonparametric Mann-Whitney U test.

FIGURES

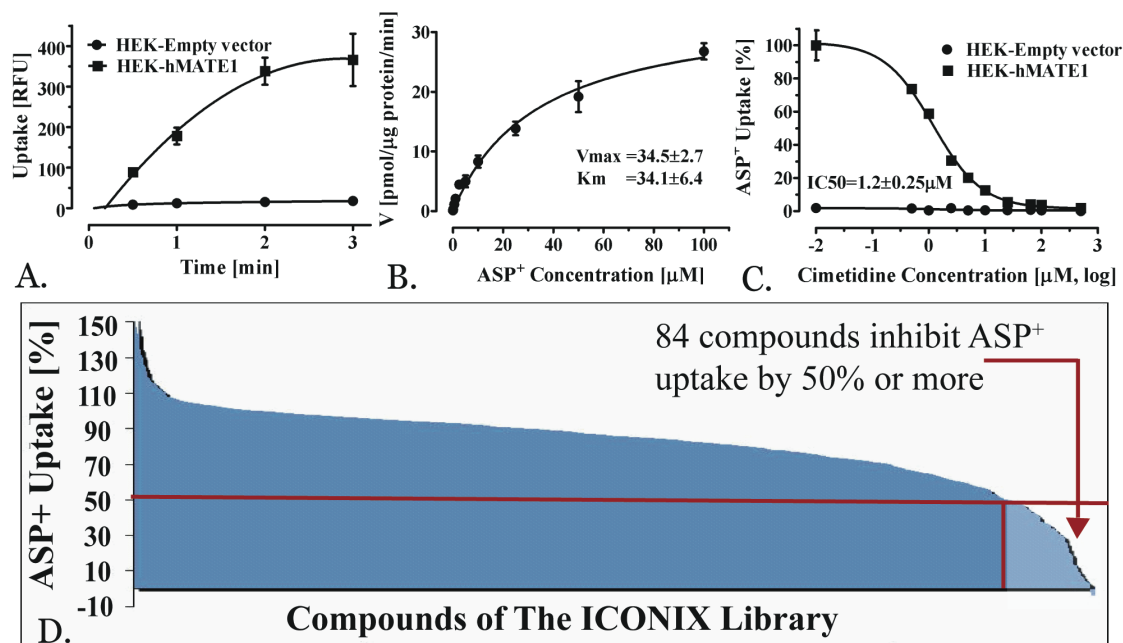


Fig. 4-1. Methods development and results of high throughput screening.

Time-course (A) and kinetics (B) of probe uptake (ASP⁺) in HEK-cells stably overexpressing MATE1. Uptake data are presented as mean \pm SEM (N=3). Inhibition of ASP⁺ uptake (C) by the model-inhibitor cimetidine. ICONIX library HTS results (D) in overview. The red line depicts the threshold of less than 50% ASP⁺ uptake at 20 μM inhibitor concentration. The red arrow marks the 84 compounds with less than 50% maximum ASP⁺ uptake; these compounds were considered as hits.

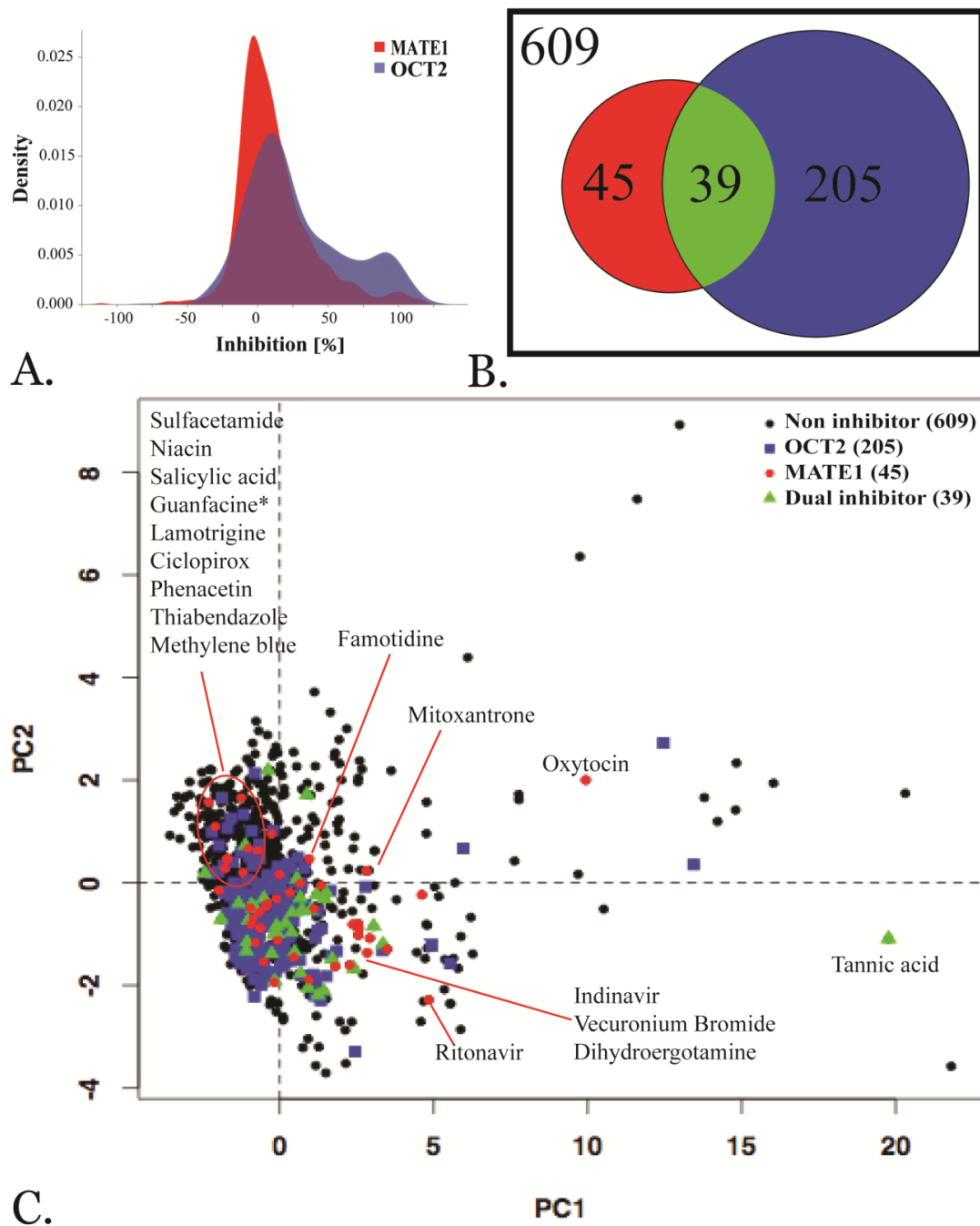
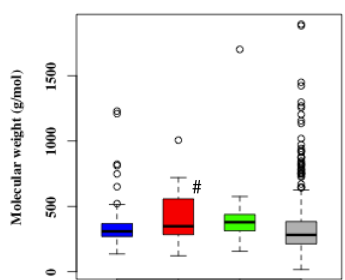


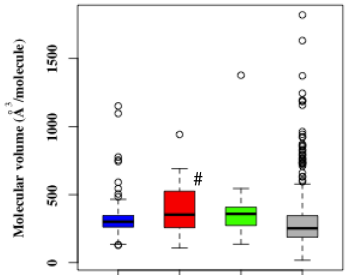
Fig. 4-2. Comparison of MATE1 and OCT2 inhibitors.

A) Density-plot of number of inhibitors per % inhibition. B) Number of MATE1 selective (red) and OCT2 selective (blue) inhibitors as well as inhibitors of both transporters (overlap; green) as

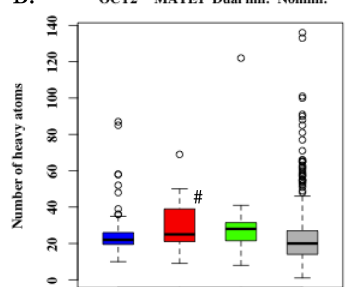
identified by screening the ICONIX library (898 compounds in total for which properties could be computed). C) Principal component analysis of the complete ICONIX library (black circles), MATE1 selective inhibitors (red circles), OCT2 selective inhibitors (blue squares), and dual inhibitors (green triangles). The first principal component is governed by polarity, size and hydrogen bonding while the second principal component is described by lipophilicity.



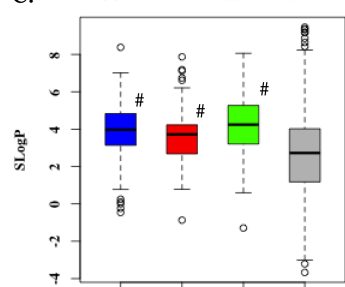
A. OCT2 MATE1 Dual inh. Noninh.



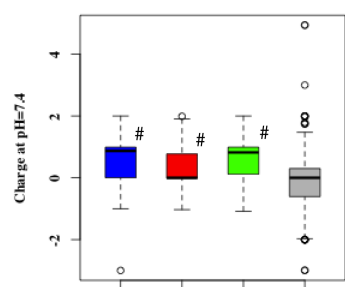
B. OCT2 MATE1 Dual inh. Noninh.



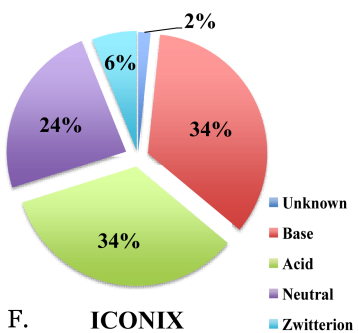
C. OCT2 MATE1 Dual inh. Noninh.



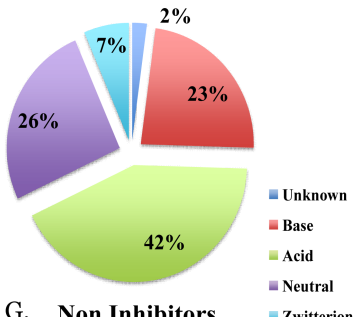
D. OCT2 MATE1 Dual inh. Noninh.



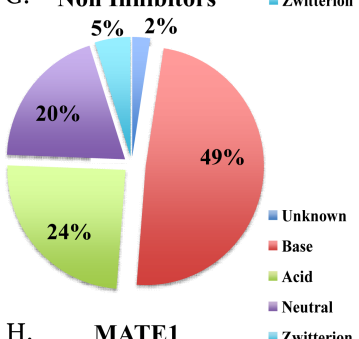
E. OCT2 MATE1 Dual inh. Noninh.



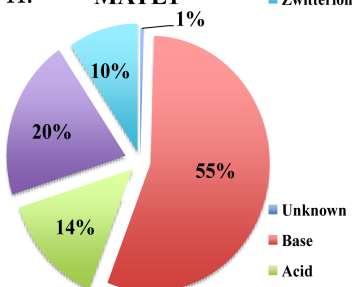
F. **ICONIX**



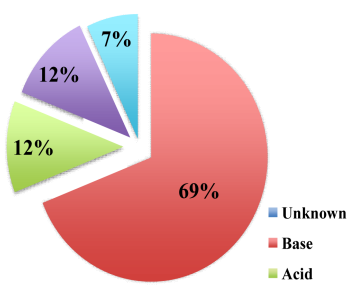
G. **Non Inhibitors**



H. **MATE1**



I. **OCT2**



J. **Dual Inhibitors**

Fig. 4-3. Comparison of physicochemical parameters for different groups of inhibitors.

A-E: Comparison of physicochemical parameters calculated for MATE1 selective inhibitors (red), OCT2 selective inhibitors (blue), dual inhibitors (OCT2/MATE1, green), and non-inhibitors (gray). # marks a statistically significant difference in comparison to the non-inhibitor group as described in the text. F-J: Distribution plots of basic ($BpK_a > 5.4$), acidic ($ApK_a < 9.4$), neutral ($ApK_a > 9.4$ and $BpK_a < 5.4$), zwitterionic ($ApK_a < 9.4$ and $BpK_a > 5.4$), and unknown drugs within the groups described above.

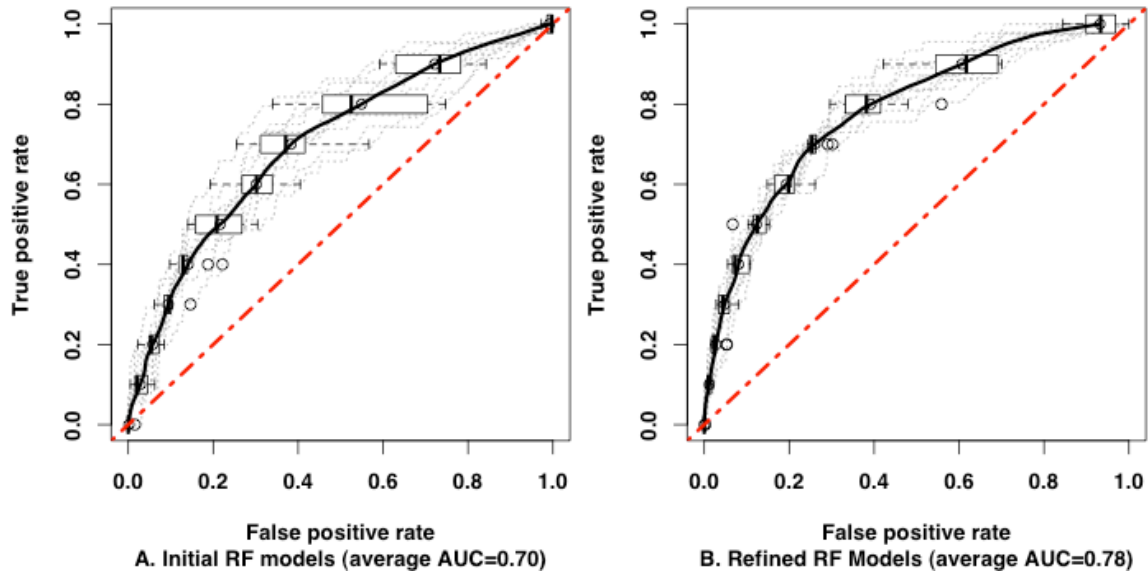


Fig. 4-4. Receiver Operating Characteristic curves for ten external data sets.

ROC curves of each test are plotted in dotted gray lines and the average ROC curves for the ten tests, in solid black lines. The ROC curve of a random classification is plotted as a red dotted diagonal line. A: ROC curves of the initial RF models. The average AUC is 0.70. B: ROC curves of the refined RF models. The average AUC is 0.78.

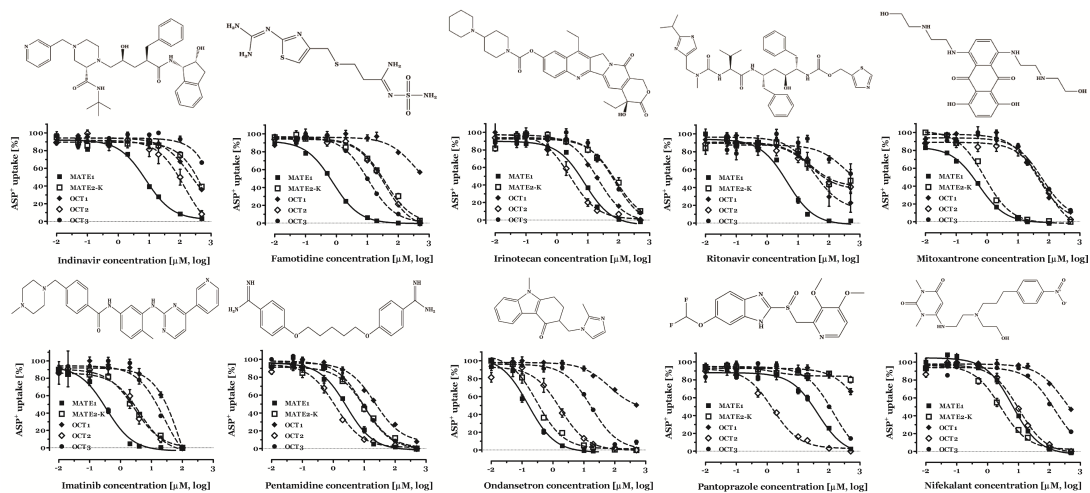


Fig. 4-5. Selectivity of putative clinically relevant inhibitors of MATE1, MATE2-K or OCT2.

IC50 values were determined for selected drugs in HEK293 cells stably over-expressing MATE1, MATE2-K, OCT1, OCT2, or OCT3. Values are presented as mean±SEM (N=3).

Clinical relevancy is defined as $C_{\text{max,unbound}}/IC_{50} \geq 0.1$.

TABLES

Table 4-1. Overview of IC50 Results for Selected HTS Hits.

<i>Compound</i>	<i>ICONIX Screen [% inhibition at 20 μM±SD]</i>	<i>p_rIC₅₀ [μM]</i>	<i>MATE1 [IC₅₀ μM, (95% CI)]</i>	<i>MATE2K [IC₅₀ μM, (95% CI)]</i>	<i>OCT2 [IC₅₀ μM, (95% CI)]</i>
Epinastine	97±3	0.6	1.1 (0.8-1.6)	29.8 (22.4-39.6)	4.3 (2.2-8.3)
Gabaxate	96±1	0.8	0.5 (0.4-0.7)	10.8 (7.9-14.7)	0.9 (0.6-1.5)
Tubocurarine	95±7	1.1	9.4 (6.2-14.4)	55.5 (36.8-83.8)	78.8 (35.4-175.5)
Bithionol	87±7	3.0	5.2 (3.1-7.2)	6.6 (3.5-12.4)	1.9 (1.3-2.8)
Chlorhexidine	86±1	3.1	0.7 (0.6-1.0)	0.5 (0.4-0.6)	0.4 (0.2-0.6)
Vecuronium bromide	85±2	3.4	1.9 (1.5-2.5)	25.2 (20.8-30.4)	3.5 (2.7-4.6)
Camostat	77±7	5.9	2.9 (2.3-3.8)	12.7 (9.1-17.7)	16.5 (13.0-21.0)
Rimantadine	74±12	7.0	7.3 (5.3-10.1)	288 (149-559)	4.4 (3.2-6.1)
Prazosin	74±1	7.0	1.6 (1.3-2.0)	38.4 (23.6-62.5)	13.6 (8.5-21.6)
Dihydroergotamine	72±2	8.0	2.8 (2.2-3.7)	12.6 (10.5-15.0)	49.9 (33.0-75.4)
Buspirone	70±9	8.4	1.7 (1.2-2.4)	46 (29.8-72.1)	12.0 (9.3-15.5)
Granisetron	67±7	9.9	5.0 (3.7-6.9)	311 (208-467)	4.3 (3.3-5.8)
Domperidone	66±2	10.2	2.3 (1.8-2.9)	14.8 (11.9-18.4)	7.9 (5.2-11.8)
Ethinyl estradiol	60±19	13.5	21.1 (13.7-32.4)	20.2 (13.6-30.1)	2.2 (0.99-4.7)
Leucomycin	56±1	16.0	11.1 (8.3-14.8)	125 (52.0-298)	182 (113-294)
Zafirlukast	52±12	19	4.3 (3.1-5.8)	7.6 (4.8-12.2)	9.7 (7.6-12.4)
Topotecan	50±6	19.7	1.3 (1.0-1.7)	8.6 (7.4-10.1)	61 (28.3-132)

Overview of screening hits that were tested based on predicted IC50 values (p_rIC50) and their expected C_{max,unbound}/IC50 as well as on their pharmacological interest and/or their supposed propensity to inhibit other cation transporters (based on published data if available). Note that these compounds are not expected to inhibit MATE1 at their clinical concentrations.

Table 4-2. Results of Random Forest-Model-I (RF-Model-I) Validation.

A. Rescued False Negative Compounds Based on QSAR-Model-I Predictions			
<i>Compound</i>	<i>Score in Computational Model-I [position out of 898]</i>	<i>ICONIX-Screen [% inhibition]</i>	<i>Actual IC50 [μM]</i>
Irinotecan	3	49 \pm 13	7.9 (5.8-10.7)
Carvedilol	10	21 \pm 8	92.4 (62.0-138)
Methyl ergonovine*	16	12 \pm 27	>250
Telmisartan	18	38 \pm 10	17.9 (14.4-22.2)
Astemizole	23	34 \pm 9	26.4 (18.8-37.1)
Noscapine	26	13.4 \pm 34	34.5 (22.3-53.6)
Amantadine	35	5 \pm 9	37.8 (20.4-70.0)
Ranitidine	118	29 \pm 14	8.3 (5.1-13.5)
B. Identification of False-Positive Compounds by QSAR-Model-I Predictions			
<i>Compound</i>	<i>Score in Computational Model-I [position out of 898]</i>	<i>ICONIX-Screen [% inhibition]</i>	<i>Actual IC50 [μM]</i>
Azacitidine	861	50 \pm 15	>500
Ciclopirox	860	52 \pm 14	>500
Sulfacetamide	842	80 \pm 4	>500
Phenacetin	806	53 \pm 7	>250
Niacin	622	54 \pm 9	>500
Thiabendazole	488	55 \pm 14	>500
C. True Negative Compounds in Both QSAR-Model-I and ICONIX Library Screen			
<i>Compound</i>	<i>Score in Computational Model-I [position out of 898]</i>	<i>ICONIX-Screen [% inhibition]</i>	<i>Actual IC50 [μM]</i>
Enalapril	355	23 \pm 18	>500
Cetirizine	376	2 \pm 5	>500
Diclofenac	445	12 \pm 5	>500
Indomethacin	454	6 \pm 6	>500
Lidocaine	556	7 \pm 13	>500
Furosemide	617	7 \pm 6	>500
Erythromycin	624	7 \pm 4	>500
Dexamethasone	862	30 \pm 4	>500
D. Mismatched Compounds Based on QSAR-Model-I Predictions			
<i>Compound</i>	<i>Computational Model-I [position out of 898]</i>	<i>ICONIX-Screen [% inhibition]</i>	<i>Actual IC50 [μM]</i>
Irbesartan	71	15 \pm 4	>500
Losartan	73	12 \pm 5	>250
Spirolactone	196	27 \pm 9	18.6 (11.8-29.3)
Levofloxacin	312	26 \pm 8	17.9 (11.5-27.9)
Quinidine	399	28 \pm 6	11.2 (7.2-17.6)
Clonidine	485	1 \pm 25	33.3 (20.9-52.9)
Cortisone	528	10 \pm 12	21.4 (10.8-42.4)
Sumatriptan	581	7 \pm 7	6.7 (3.9-11.6)

Ranks 1 to 120 were considered as likely hits based on their RF-score. A) Compounds that were negative in the screen but have been rescued based on their RF-model-I values (ranked position 1 to 120). * Methyl ergonovine was the only wrongly predicted of the tested rescue compounds.

B) Compounds that were identified as false positive by RF-Model-I (ranked lower than position 120). C) True negative-compounds that were negative in the screen and correctly predicted negative by RF-Model-I. D) Compounds that were mispredicted by the model as either positive (i.e., both sartans) or negative. Note that in the latter case the compounds were not hits in the screen either but were selected based on published data. The results shown in this Table were used to train RF-Model-I and to achieve an improved model (RF-Model-II).

Table 4-3. Selected Hits of the DrugBank Library-in silico Screening with RF-Model-II and Their Experimental IC50 Values.

<i>Compound</i>	<i>Score in Computational Model-II [position out of 6551]</i>	<i>Maximum Tc similarity to ICONIX compounds</i>	<i>MATE1 [IC₅₀ μM, (95% CI)]</i>	<i>MATE2K [IC₅₀ μM, (95% CI)]</i>	<i>OCT2 [IC₅₀ μM, (95% CI)]</i>
Dabigatran	102	0.528 (irinotecan)	8.1 (5.7-11.5)	25.3 (12.0-53.1)	4.7 (2.8-8.0)
Pimozide	126	0.724 (droperidol)	>500	>500	34.2 (23.3-52.2)
Aripiprazole	160	0.43 (mosapride)	130 (77.5-219)	>500	239 (70.1-815)
Risperidone	250	0.543 (ondansetron)	1.6 (0.98-1.9)	291 (117-719)	11.0 (7.4-16.4)
Maraviroc	579	<0.3 (dapiprazole)	17.3 (9.6-31.2)	297 (166-530)	255 (155-433)

IC50 values of compounds that were not contained in the ICONIX library but were identified by in silico screening of the DrugBank library. Compounds were chosen based on their pharmacological interest and novelty compared to previously identified inhibitors. Tadalafil and itraconazole could not be tested due to low solubility.

Table 4-4. Overview of the 21 most important descriptors used in the RF-Model-II (mean descriptor values of MATE1 inhibitors and non-inhibitors are presented).

Descriptor	MATE1 inhibitors	Non-inhibitors	p-value	Description	Category
Vindex	0.23	0.35	3.46681E-13	Balaban V index	topological
Wap	67766.85	86064.57	4.89801E-12	all-path Wiener index	topological
J	1.57	1.93	8.7592E-12	Balaban distance connectivity index	topological
D/Dr06	236.12	143.60	5.99116E-11	distance/detour ring index of order 6	topological
Jhetm	1.95	2.62	6.09E-11	Balaban-type index from mass weighted distance matrix	topological
MW	403.64	325.35	1.47811E-10	molecular weight	constitutional
MPC10	191.48	112.08	3.68301E-10	molecular path count of order 10	path and walk count
D/Dr09	88.25	23.44	6.80182E-10	distance/detour ring index of order 9	topological
nCIC	3.67	2.51	7.5614E-10	number of rings	constitutional
nR09	0.78	0.26	3.84421E-09	number of 9-membered rings	constitutional
nCIR	7.04	4.58	2.33749E-07	number of circuits	constitutional
TI2	4.31	3.32	2.86258E-06	second Mohar index	topological
Ms	2.28	2.52	7.97109E-06	mean electrotopological state	constitutional
nN	2.80	1.94	4.40216E-05	number of Nitrogen atoms	constitutional
MLOGP	2.69	1.82	0.000109292	Moriguchi octanol-water partition coefficient	molecular
C-025	1.86	1.17	0.000154955	R--CR--R atom centered fragments	atom-centered fragments
MSD	0.24	0.26	0.00380264	mean square distance index (Balaban)	topological
Me	1.00	1.01	ns	mean atomic Sanderson electronegativity (scaled on Carbon atom)	constitutional
Mv	0.62	0.62	ns	molecular volume	constitutional
ARR	0.35	0.34	ns	aromatic ratio	constitutional
nO	3.15	3.36	ns	number of Oxygen atoms	constitutional

Overview of descriptors selected using backward feature elimination algorithm. The differences between descriptor values of MATE1 inhibitors and non-inhibitors are assessed using the nonparametric Mann-Whitney U test.

Table 4-5. Summary of IC₅₀ Values for Putative Clinically Relevant Inhibitors of MATE1, MATE2-K or OCT2 in ASP+ Uptake Assay.

<i>Compound</i>	<i>MATE1 ASP⁺ [IC₅₀ μM, (95% CI)]</i>	<i>MATE2K ASP⁺ [IC₅₀ μM, (95% CI)]</i>	<i>OCT1 ASP⁺ [IC₅₀ μM, (95% CI)]</i>	<i>OCT2 ASP⁺ [IC₅₀ μM, (95% CI)]</i>	<i>OCT3 ASP⁺ [IC₅₀ μM, (95% CI)]</i>
Ritonavir	4.4 (3.3-6.0)	23.7 (7.6-74.2)	33.9 (15.9-72.2)	24.8 (12.0-51.0)	>300
Indinavir	7.8 (6.1-9.8)	>500	208 (97.6-446)	142 (59.7-335)	>500
Irinotecan	7.9 (5.8-10.7)	78.6 (48.3-127)	20.8 (14.4-30.2)	2.7 (1.9-4.0)	74.6 (47.5-117)
Imatinib	0.35 (0.27-0.44)	2.9 (2.1-3.8)	107 (36.4-315)	4.2 (1.8-9.7)	25.5 (14.0-46.3)
Mitoxantrone	0.53 (0.46-0.59)	0.83 (0.73-0.93)	43.9 (33.6-57.4)	73.3 (46.1-116)	60.5 (39.9-91.7)
Famotidine	0.76 (0.66-0.87)	36.2 (27.3-47.9)	>300	36.1 (27.2-47.9)	10.7 (8.3-13.9)
Ondansetron	0.16 (0.12-0.21)	0.30 (0.24-0.38)	>300	1.7 (1.1-2.6)	17.4 (13.1-23.2)
Nifekalant*	6.5 (5.3-8.0)	2.7 (2.0-3.5)	>300	10.3 (8.1-13.0)	146 (96.9-220)
Pantoprazole**	43.2 (29.9-62.4)	>500	>500	1.5 (1.0-2.1)	137 (87.0-217)
Pentamidine**	2.7 (2.1-3.5)	10.4 (7.9-13.7)	22.1 (17.7-27.6)	1.2 (0.9-1.5)	9.5 (7.8-11.5)

Overview of putative clinically relevant inhibitors and their IC₅₀ values against MATE1, MATE2-K, OCT1, OCT2, and OCT3. * Nifekalant does not reach a clinically significant concentration for MATE1 inhibition but does so for MATE2-K. ** Similarly, pantoprazole and pentamidine do not reach a clinically significant concentration for MATE1 inhibition but do so for OCT2.

Table 4-6. Follow-up IC50 Experiments with Metformin for Putative Clinically Relevant MATE1 Inhibitors in Different Assays.

<i>Compound</i>	<i>IC₅₀ Against MATE1 [μM, (95% CI)]</i>			<i>C_{max} [μM]</i>	<i>Protein Binding [% bound]</i>	<i>C_{max,unbound} / IC₅₀ of MATE1</i>		
	<i>ASP⁺</i>	<i>Metformin</i>	<i>Metformin Polarized Cells</i>			<i>ASP⁺</i>	<i>Metformin</i>	<i>Metformin Polarized Cells</i>
Ritonavir	4.4 (3.3-6.0)	0.08 (0.02-0.26)	2.1 (1.55- 2.71)	15.26^a	98	0.07	3.82	0.15
Indinavir	7.8 (6.1-9.8)	1.7 (1.2-2.6)	1.9 (1.03- 2.73)	9.8^a	61	0.49	2.25	2.01
Irinotecan	7.9 (5.8-10.7)	2.1 (1.1-4.2)	4.3 (2.14-6.36)	4.6^b	50	0.29	1.10	0.54
Imatinib	0.35 (0.27-0.44)	0.05 (0.03-0.08)	0.04 (0.03-0.06)	3.65^c	95	0.52	3.65	4.56
Mitoxantrone	0.53 (0.46-0.59)	0.4 (0.33-0.52)	0.19 (0.04-0.34)	0.63^b	78	0.26	0.35	0.73
Famotidine	0.76 (0.66-0.87)	0.09 (0.06-0.13)	0.16 (0.11-0.21)	0.46^b	15	0.51	4.34	2.44
Ondansetron	0.16 (0.12-0.21)	0.03 (0.02-0.05)	<0.01	0.44^b	73	0.74	3.96	>11.88
Nifekalant*	6.5 (5.3-8.0)	1.1 (0.4-2.6)	0.65 (0.30-1.01)	3.43^d	90	0.05	0.31	0.53
Pantoprazole**	43.2 (29.9-62.4)	2.8 (1.8-4.3)	2.8 (0.08-5.41)	5.45^e	98	0.003	0.12	0.12
Pentamidine**	2.7 (2.1-3.5)	2.0 (1.1-3.7)	0.41 (0.33-0.49)	0.73^b	69	0.08	0.11	0.55

Inhibitors tested against MATE1 with either ASP+ or 14C-metformin as substrate. Polarized cells are MDCK-hMATE1 in a transwell setting (see methods). The other values were generated in HEK293-cells stably expressing MATE1. *Nifekalant does not reach a clinically significant concentration for MATE1 inhibition but does so for MATE2-K. **Similarly, pantoprazole and pentamidine do not reach a clinically significant concentration for MATE1-inhibition but do so for OCT2. Values for protein binding were taken from Micromedex (53) C_{max} values were retrieved from various sources (references a-e) (54–58).

ASSOCIATED CONTENT

Supporting Information. HTS data (% inhibition at 20 μ M and SD) as well as names and SMILES codes of all ICONIX library drugs (in .xls format); Physicochemical properties of compounds in the ICONIX library; Comparison of the performance of three computational methods, kNN, PLS-DA and RF; non-significant descriptors of MATE1 inhibitors and of the other groups; development and improvement of kNN and PLS-DA models; PCA plot of the DrugBank compounds; Depiction of the importance of Dragon descriptors in RF-Model-II. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES

- (1) Drug Transporters: Molecular Characterization and Role in Drug Disposition; You, G.; Morris, M. E., Eds.; Wiley Series in Drug Discovery and Development; 1st ed.; Wiley-Interscience, Inc.: Hoboken, NJ, 2007.
- (2) Giacomini, K. M.; Huang, S.-M.; Tweedie, D. J.; Benet, L. Z.; Brouwer, K. L. R.; Chu, X.; Dahlin, A.; Evers, R.; Fischer, V.; Hillgren, K. M.; Hoffmaster, K. A.; Ishikawa, T.; Keppler, D.; Kim, R. B.; Lee, C. A.; Niemi, M.; Polli, J. W.; Sugiyama, Y.; Swaan, P. W.; Ware, J. A.; Wright, S. H.; Yee, S. W.; Zamek-Gliszczyński, M. J.; Zhang, L. Membrane transporters in drug development. *Nat Rev Drug Discov* 2010, 9, 215–236.
- (3) Nies, A. T.; Koepsell, H.; Damme, K.; Schwab, M. Organic Cation Transporters (OCTs, MATEs), In Vitro and In Vivo Evidence for the Importance in Drug Therapy. In *Drug Transporters*; Fromm, M. F.; Kim, R. B., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; Vol. 201, pp. 105–167.
- (4) Burckhardt, B. C.; Burckhardt, G. Transport of organic anions across the basolateral membrane of proximal tubule cells. *Reviews of Physiology, Biochemistry and Pharmacology*, Vol 146 2003, 146, 95–158.
- (5) Masereeuw, R.; Russel, F. G. M. Therapeutic implications of renal anionic drug transporters. *Pharmacology & Therapeutics* 2010, 126, 200–216.
- (6) Otsuka, M.; Matsumoto, T.; Morimoto, R.; Arioka, S.; Omote, H.; Moriyama, Y. A human transporter protein that mediates the final excretion step for toxic organic cations. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102, 17923–17928.
- (7) Masuda, S.; Terada, T.; Yonezawa, A.; Tanihara, Y.; Kishimoto, K.; Katsura, T.; Ogawa, O.; Inui, K. Identification and functional characterization of a new human kidney-specific

- H⁺/organic cation antiporter, kidney-specific multidrug and toxin extrusion 2. *J. Am. Soc. Nephrol.* 2006, 17, 2127–2135.
- (8) Komatsu, T.; Hiasa, M.; Miyaji, T.; Kanamoto, T.; Matsumoto, T.; Otsuka, M.; Moriyama, Y.; Omote, H. Characterization of the human MATE2 proton-coupled polyspecific organic cation exporter. *Int. J. Biochem. Cell Biol.* 2011, 43, 913–918.
 - (9) Meyer zu Schwabedissen, H. E.; Verstuyft, C.; Kroemer, H. K.; Becquemont, L.; Kim, R. B. Human multidrug and toxin extrusion 1 (MATE1/SLC47A1) transporter: functional characterization, interaction with OCT2 (SLC22A2), and single nucleotide polymorphisms. *Am. J. Physiol. Renal Physiol.* 2010, 298, F997–F1005.
 - (10) König, J.; Zolk, O.; Singer, K.; Hoffmann, C.; Fromm, M. F. Double-transfected MDCK cells expressing human OCT1/MATE1 or OCT2/MATE1: determinants of uptake and transcellular translocation of organic cations. *Br. J. Pharmacol.* 2011, 163, 546–555.
 - (11) Tanihara, Y.; Masuda, S.; Sato, T.; Katsura, T.; Ogawa, O.; Inui, K.-I. Substrate specificity of MATE1 and MATE2-K, human multidrug and toxin extrusions/H(+)-organic cation antiporters. *Biochem. Pharmacol.* 2007, 74, 359–371.
 - (12) Astorga, B.; Ekins, S.; Morales, M.; Wright, S. H. Molecular determinants of ligand selectivity for the human multidrug and toxin extruder proteins MATE1 and MATE2-K. *J. Pharmacol. Exp. Ther.* 2012, 341, 743–755.
 - (13) Ahlin, G.; Karlsson, J.; Pedersen, J. M.; Gustavsson, L.; Larsson, R.; Matsson, P.; Norinder, U.; Bergström, C. A. S.; Artursson, P. Structural requirements for drug inhibition of the liver specific human organic cation transport protein 1. *J. Med. Chem.* 2008, 51, 5932–5942.
 - (14) Badolo, L.; Rasmussen, L. M.; Hansen, H. R.; Sveigaard, C. Screening of OATP1B1/3 and OCT1 inhibitors in cryopreserved hepatocytes in suspension. *Eur J Pharm Sci* 2010, 40, 282–288.
 - (15) Kido, Y.; Matsson, P.; Giacomini, K. M. Profiling of a prescription drug library for potential renal drug-drug interactions mediated by the organic cation transporter 2. *J. Med. Chem.* 2011, 54, 4548–4558.
 - (16) Yonezawa, A.; Inui, K. Importance of the multidrug and toxin extrusion MATE/SLC47A family to pharmacokinetics, pharmacodynamics/toxicodynamics and pharmacogenomics. *Br. J. Pharmacol.* 2011, 164, 1817–1825.
 - (17) Breiman, L. Random forests. *Machine learning* 2001, 45, 5–32.
 - (18) Ito, S.; Kusuhara, H.; Yokochi, M.; Toyoshima, J.; Inoue, K.; Yuasa, H.; Sugiyama, Y. Competitive inhibition of the luminal efflux by multidrug and toxin extrusions, but not basolateral uptake by organic cation transporter 2, is the likely mechanism underlying the pharmacokinetic drug-drug interactions caused by cimetidine in the kidney. *J. Pharmacol. Exp. Ther.* 2012, 340, 393–403.
 - (19) Tsuda, M.; Terada, T.; Ueba, M.; Sato, T.; Masuda, S.; Katsura, T.; Inui, K. Involvement of human multidrug and toxin extrusion 1 in the drug interaction between cimetidine and metformin in renal epithelial cells. *J. Pharmacol. Exp. Ther.* 2009, 329, 185–191.

- (20) Karlgren, M.; Vildhede, A.; Norinder, U.; Wisniewski, J. R.; Kimoto, E.; Lai, Y.; Haglund, U.; Artursson, P. Classification of Inhibitors of Hepatic Organic Anion Transporting Polypeptides (OATPs): Influence of Protein Expression on Drug–Drug Interactions. *J Med Chem* 2012, 55, 4740–4763.
- (21) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 2006, 34, D668–672.
- (22) Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 2008, 36, D901–906.
- (23) Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y.; Eisner, R.; Guo, A. C.; Wishart, D. S. DrugBank 3.0: a comprehensive resource for “omics” research on drugs. *Nucleic Acids Res.* 2011, 39, D1035–1041.
- (24) Sheridan, R. P. Three useful dimensions for domain applicability in QSAR models using random forest. *J Chem Inf Model* 2012, 52, 814–823.
- (25) Dimitrov, S.; Dimitrova, G.; Pavlov, T.; Dimitrova, N.; Patlewicz, G.; Niemela, J.; Mekenyan, O. A stepwise approach for defining the applicability domain of SAR and QSAR models. *J Chem Inf Model* 2005, 45, 839–849.
- (26) Sahigara, F.; Mansouri, K.; Ballabio, D.; Mauri, A.; Consonni, V.; Todeschini, R. Comparison of different approaches to define the applicability domain of QSAR models. *Molecules* 2012, 17, 4791–4810.
- (27) Dragos, H.; Gilles, M.; Alexandre, V. Predicting the predictability: a unified approach to the applicability domain problem of QSAR models. *J Chem Inf Model* 2009, 49, 1762–1776.
- (28) Abel, S.; Back, D. J.; Vourvahis, M. Maraviroc: pharmacokinetics and drug interactions. *Antivir. Ther. (Lond.)* 2009, 14, 607–618.
- (29) Abel, S.; Russell, D.; Whitlock, L. A.; Ridgway, C. E.; Nedderman, A. N. R.; Walker, D. K. Assessment of the absorption, metabolism and absolute bioavailability of maraviroc in healthy male subjects. *Br J Clin Pharmacol* 2008, 65 Suppl 1, 60–67.
- (30) Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. Classification and regression trees; Wadsworth International Group: Belmont, CA, 1984.
- (31) Steele, R. E.; Preston, A. S.; Johnson, J. P.; Handler, J. S. Porous-bottom dishes for culture of polarized cells. *Am. J. Physiol.* 1986, 251, C136–139.
- (32) Handler, J. S.; Preston, A. S.; Steele, R. E. Factors affecting the differentiation of epithelial transport and responsiveness to hormones. *Fed. Proc.* 1984, 43, 2221–2224.
- (33) Chen, Y.; Teranishi, K.; Li, S.; Yee, S. W.; Hesselson, S.; Stryke, D.; Johns, S. J.; Ferrin, T. E.; Kwok, P.; Giacomini, K. M. Genetic variants in multidrug and toxic compound extrusion-1, hMATE1, alter transport function. *Pharmacogenomics J.* 2009, 9, 127–136.

- (34) Wright, S. H.; Wunz, T. M.; Wunz, T. P. Structure and interaction of inhibitors with the TEA/H⁺ exchanger of rabbit renal brush border membranes. *Pflugers Arch.* 1995, 429, 313–324.
- (35) Wright, S. H.; Wunz, T. M. Influence of substrate structure on turnover of the organic cation/H⁺ exchanger of the renal luminal membrane. *Pflugers Arch.* 1998, 436, 469–477.
- (36) Wright, S. H.; Wunz, T. M. Influence of substrate structure on substrate binding to the renal organic cation/H⁺ exchanger. *Pflugers Arch.* 1999, 437, 603–610.
- (37) Ullrich, K. J.; Papavassiliou, F.; David, C.; Rumrich, G.; Fritsch, G. Contraluminal transport of organic cations in the proximal tubule of the rat kidney. I. Kinetics of N1-methylnicotinamide and tetraethylammonium, influence of K⁺, HCO₃⁻, pH; inhibition by aliphatic primary, secondary and tertiary amines and mono- and bisquaternary compounds. *Pflugers Arch.* 1991, 419, 84–92.
- (38) Ullrich, K. J.; Rumrich, G.; Neiteler, K.; Fritsch, G. Contraluminal transport of organic cations in the proximal tubule of the rat kidney. II. Specificity: anilines, phenylalkylamines (catecholamines), heterocyclic compounds (pyridines, quinolines, acridines). *Pflugers Arch.* 1992, 420, 29–38.
- (39) David, C.; Rumrich, G.; Ullrich, K. J. Luminal transport system for H⁺/organic cations in the rat proximal tubule. Kinetics, dependence on pH; specificity as compared with the contraluminal organic cation-transport system. *Pflugers Arch.* 1995, 430, 477–492.
- (40) Ullrich, K. J.; Rumrich, G. Luminal transport system for choline⁺ in relation to the other organic cation transport systems in the rat proximal tubule. Kinetics, specificity: alkyl/aryl amines, alkyl amines with OH, O, SH, NH₂, ROCO, RSCO and H₂PO₄-groups, methylaminostyryl, rhodamine, acridine, phenanthrene and cyanine compounds. *Pflugers Arch.* 1996, 432, 471–485.
- (41) Zhang, X.; He, X.; Baker, J.; Tama, F.; Chang, G.; Wright, S. H. Twelve Transmembrane Helices Form the Functional Core of Mammalian MATE1 (Multidrug and Toxin Extruder 1) Protein. *J. Biol. Chem.* 2012, 287, 27971–27982.
- (42) He, X.; Szewczyk, P.; Karyakin, A.; Evin, M.; Hong, W.-X.; Zhang, Q.; Chang, G. Structure of a cation-bound multidrug and toxic compound extrusion transporter. *Nature* 2010, 467, 991–994.
- (43) Tsuda, M.; Terada, T.; Mizuno, T.; Katsura, T.; Shimakura, J.; Inui, K. Targeted disruption of the multidrug and toxin extrusion 1 (mate1) gene in mice reduces renal secretion of metformin. *Mol. Pharmacol.* 2009, 75, 1280–1286.
- (44) Matsushima, S.; Maeda, K.; Inoue, K.; Ohta, K.; Yuasa, H.; Kondo, T.; Nakayama, H.; Horita, S.; Kusuhara, H.; Sugiyama, Y. The inhibition of human multidrug and toxin extrusion 1 is involved in the drug-drug interaction caused by cimetidine. *Drug Metab. Dispos.* 2009, 37, 555–559.
- (45) Kusuhara, H.; Ito, S.; Kumagai, Y.; Jiang, M.; Shiroshita, T.; Moriyama, Y.; Inoue, K.; Yuasa, H.; Sugiyama, Y. Effects of a MATE protein inhibitor, pyrimethamine, on the renal elimination of metformin at oral microdose and at therapeutic dose in healthy subjects. *Clin. Pharmacol. Ther.* 2011, 89, 837–844.

- (46) Minematsu, T.; Giacomini, K. M. Interactions of Tyrosine Kinase Inhibitors with Organic Cation Transporters and Multidrug and Toxic Compound Extrusion Proteins. *Mol Cancer Ther* 2011, 10, 531–539.
- (47) Nies, A. T.; Hofmann, U.; Resch, C.; Schaeffeler, E.; Rius, M.; Schwab, M. Proton pump inhibitors inhibit metformin uptake by organic cation transporters (OCTs). *PLoS ONE* 2011, 6, e22163.
- (48) Sato, T.; Masuda, S.; Yonezawa, A.; Tanihara, Y.; Katsura, T.; Inui, K.-I. Transcellular transport of organic cations in double-transfected MDCK cells expressing human organic cation transporters hOCT1/hMATE1 and hOCT2/hMATE1. *Biochem. Pharmacol.* 2008, 76, 894–903.
- (49) Han, Y.-H.; Busler, D.; Hong, Y.; Tian, Y.; Chen, C.; Rodrigues, A. D. Transporter studies with the 3-O-sulfate conjugate of 17 α -ethinylestradiol: assessment of human kidney drug transporters. *Drug Metab. Dispos.* 2010, 38, 1064–1071.
- (50) More, S. S.; Li, S.; Yee, S. W.; Chen, L.; Xu, Z.; Jablons, D. M.; Giacomini, K. M. Organic Cation Transporters Modulate the Uptake and Cytotoxicity of Picoplatin, a Third-Generation Platinum Analogue. *Molecular Cancer Therapeutics* 2010, 9, 1058 – 1069.
- (51) Zhang, S.; Lovejoy, K. S.; Shima, J. E.; Lagpacan, L. L.; Shu, Y.; Lapuk, A.; Chen, Y.; Komori, T.; Gray, J. W.; Chen, X.; Lippard, S. J.; Giacomini, K. M. Organic Cation Transporters Are Determinants of Oxaliplatin Cytotoxicity. *Cancer Res* 2006, 66, 8847–8857.
- (52) Shu, Y.; Sheardown, S. A.; Brown, C.; Owen, R. P.; Zhang, S.; Castro, R. A.; Ianculescu, A. G.; Yue, L.; Lo, J. C.; Burchard, E. G.; Brett, C. M.; Giacomini, K. M. Effect of genetic variation in the organic cation transporter 1 (OCT1) on metformin action. *J. Clin. Invest.* 2007, 117, 1422–1431.
- (53) Home - MICROMEDEX® 2.0
http://www.thomsonhc.com/micromedex2/librarian/ND_T/evidencexpert/ND_PR/evidencexpert/CS/CDB742/ND_AppProduct/evidencexpert/DUPLICATIONSHIELDSYNC/BFA79B/ND_PG/evidencexpert/ND_B/evidencexpert/ND_P/evidencexpert/PFAActionId/pf.HomePage (accessed Jun 12, 2012).
- (54) Thummel, Kenneth E.; Shen, Danny D. Appendix II - Design and Optimization of Dosage Regimens: Pharmacokinetic Data. In *Goodman & Gilman's The Pharmacological Basis of Therapeutics*; Hardman, Joel G.; Limbird, Lee E.; Goodman Gilman, Alfred, Eds.; McGraw-Hill, Inc.: New York, NY, 2001; pp. 1917–2023.
- (55) *Clarke's Analysis of Drugs and Poisons*; Moffat, Anthony C.; Osselton, M. David; Widdop, Brian; Galichet, Laurent Y., Eds.; Third ed.; Pharmaceutical Press: London, UK, 2004; Vol. 2.
- (56) Frye, R. F.; Fitzgerald, S. M.; Lagattuta, T. F.; Hruska, M. W.; Egorin, M. J. Effect of St John's wort on imatinib mesylate pharmacokinetics. *Clin. Pharmacol. Ther.* 2004, 76, 323–329.

- (57) Nakaya, H.; Uemura, H. Electropharmacology of nifekalant, a new class III antiarrhythmic drug. *Cardiovasc. Drug Rev.* 1998, 16, 133–144.
- (58) Andersson, T.; Holmberg, J.; Röhss, K.; Walan, A. Pharmacokinetics and effect on caffeine metabolism of the proton pump inhibitors, omeprazole, lansoprazole, and pantoprazole. *Br J Clin Pharmacol* 1998, 45, 369–375.

Chapter 5: Discovery of Inhibitors of the Human OATP2B1 Transporter by Virtual Screening Against Multiple Comparative Models and Machine Learning

Natalia Khuri^{1,#}, Arik A. Zur^{2,#}, Matthias B. Wittwer², Lawrence Lin², Kathleen M.

Giacomini^{2,4,*}, and Andrej Sali^{2,3,*}

¹Graduate Group in Biophysics, University of California at San Francisco

²Department of Bioengineering and Therapeutic Sciences, University of California San Francisco

³Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences (QB3), University of California at San Francisco

⁴Institute for Human Genetics, University of California, San Francisco

[#]These authors contributed equally

^{*}Corresponding authors:

Andrej Sali, UCSF MC 2552, Byers Hall at Mission Bay, Suite 503B, University of California at San Francisco, 1700 4th Street, San Francisco, CA 94158, USA. tel: +1 415 514 4227; fax: +1 415 514 4231. E-mail: sali@salilab.org

Kathleen Giacomini, University of California San Francisco, 1550 4th Street Box 2911, San Francisco, CA 94158, USA. tel: +1 415 476 1936; fax: +1 415 514 4361; E-mail: kathy.giacomini@ucsf.edu.

ABSTRACT

Human organic anion transporters (OATPs) are vital for the uptake and efflux of drugs and endogenous compounds. Current identification of inhibitors of these transporters is based on experimental screening. Approximately 40 inhibitors have been reported for the OATP member 2B1, which is highly expressed in the intestine where it can participate in oral absorption of drugs. Here, we use machine learning based on data from experimental and virtual screening against multiple comparative protein structure models to identify OATP2B1 inhibitors in the DrugBank library of over 5000 drugs and drug-like molecules. Twenty-eight predicted compounds were tested experimentally in cells overexpressing the OATP2B1 transporter. Seventeen of these twenty-eight were validated as inhibitors of estrone 3-sulfate uptake (defined as more than 50% inhibition at 500 μ M); they include meloxicam and ticagrelor. The IC₅₀ values of ticagrelor and meloxicam suggest that they might inhibit intestinal OATP2B1 at clinically relevant concentrations and therefore modulate the actions of other concomitantly administered drugs.

INTRODUCTION

In human, the number of different membrane proteins involved in the uptake and efflux of vital substances, such as sugars, vitamins, toxins, signal molecules, and drugs, is relatively small compared to the large number of types of ligands. Thus, the structures of these proteins evolved to bind structurally diverse entities, leading to increased odds for interaction between different ligands.

Here, we focused on the human solute carrier organic anion transporter superfamily member 2B1 (OATP2B1, SLC21A9), significantly expressed in the apical membrane of human small intestine epithelia[1], as well as in other tissues of pharmacological importance, including

the blood brain barrier, liver, and tumors[2, 3]. The in vivo pharmacological relevance of OATP2B1 in drug absorption is well described and includes genetic polymorphisms[4, 5], drug-drug interactions[6-8], and drug-fruit interactions[6, 9, 10]. OATP2B1 transports a broad range of substrates including statins (e.g., fluvastatin, rosuvastatin, and atorvastatin)[11-13], estrone-3-sulfate (E3S)[14], bosantan[15], and glyburide[16]. OATP2B1 substrate specificity overlaps partially with that of other members of the OATP superfamily[14]. It has been proposed that human OATP2B1 transports its substrates in a bi-directional, pH-dependent manner[17]. Acidic extracellular pH increases transport activity of OATP2B1 and broadens its substrate specificity[18]. Studies with E3S showed biphasic uptake kinetics, indicating the presence high-affinity and low-affinity binding sites[19]. They also described different effects of inhibitors on these sites. For example, naringin, pravastatin, and taurocholic acid have been shown to selectively inhibit the high affinity site, testosterone was selective for the low affinity site, and bromsulphalein and montelukast inhibited both sites[19]. Because there is limited information about atomic structure of OATP2B1, it is not known if the two binding sites are co-localized in a single binding pocket or comprise two individual pockets in the OATP2B1 protein.

Because co-administered drugs may affect OATP2B1 activity, characterization of its ligands is important for optimization of oral drug delivery and the prediction of drug-drug interactions. Transporter inhibition assays are used with increased throughput and productivity due to the availability of fluorescent substrates and overexpressing cell lines. To date, more than 10 transporter inhibition screens have been published[20-37], including a small inhibition screen of 225 compounds against OATP2B1 that identified 45 inhibitors of E3S uptake[38]. In silico methods, either alone or in combination with a high-throughput screen (HTS), can be utilized to rapidly examine large virtual libraries of compounds. Current methods for predicting ligand-

transporter interactions are either structure- or ligand-based. In structure-based modeling, large virtual libraries of compounds can be docked against comparative models of transporter proteins to rank the compounds based on their physicochemical complementarity to the putative binding site on the transporter[39-42]. These methods require accurate homology models and known binding sites. The usability of the model can be tested by discriminating between known ligands and decoy molecules[43, 44]. In ligand-based modeling, 2D and 3D physicochemical properties or pharmacophores important for the ligand's interaction with the transporter are identified and used to develop classification or regression models, correlating in vitro activity with molecular properties[29].

The primary goal of this study was to identify new inhibitors of E3S transport by human OATP2B1 transporter by utilizing predictions from both structure- and ligand-based methods (Figure 1). Mechanistically, there are three types of such inhibitors. First, a ligand that is transported by OATP2B1, thus competing with E3S (competitive substrate). Second, an orthosteric ligand that binds to any of the substrate-binding sites, but is not transported (orthosteric or competitive inhibitor). Third, an allosteric ligand that binds to another site on OATP2B1 and inhibits binding and/or transport of E3S (allosteric inhibitor); alternatively, an allosteric ligand may also activate the transport of E3S. Transporter inhibition assays, including ours, do not distinguish between these three types of inhibitors. While the ligand-based prediction may identify inhibitors of any of the three types, the structure-based approach is likely to predict only inhibitors of the first two types. Therefore, the second aim of this study was to determine physicochemical properties of inhibitors predicted by the two different approaches. Moreover, we aimed to increase the accuracy of structure-based approach by docking against comparative models of OATP2B1 in various conformations. Instead of evaluating docking by

ranking the known inhibitors and their decoys[43, 44], we utilized docking results in an unsupervised learning procedure that first identified groups of compounds with similar docking outcomes across homology models in different conformations and then determined several clusters of compounds interacting favorably with one or more of these models. Finally, we performed a prospective validation of our method by testing 28 predicted hits in vitro using inhibition of E3S uptake assay at two inhibitor concentrations, confirming 17 predictions.

RESULTS

Selection of informative molecular descriptors of known OATP2B1 inhibitors and noninhibitors

We used the data set of inhibitors and noninhibitors[38] of E3S transport as a training set (n=225). The data set was well diversified with the mean pairwise Tanimoto similarity coefficient (Tc) of 0.45 (Supplemental figure S1A). The reported inhibition values ranged from 100% (strong inhibitors) to -400% (activators) (Supplemental Figure S1B). We computed 219 molecular descriptors for all compounds in the data set and selected the 14 features with the highest weighted sum of “information gain”[45] and correlation coefficient between the feature and the inhibition values (Supplemental Table S1). Among the selected descriptors, thirteen correlated positively with reported inhibition values. In contrast, chemical diversity descriptor (BCUTc-11) had a negative correlation (Figure 2A), implying that known inhibitors are chemically more similar to each other than noninhibitors. Interestingly, the algorithm selected two different estimates of compound’s lipophilicity, GrippenLogP and XLogP. These octanol/water partition coefficients positively correlated with inhibition, indicating that OATP2B1 inhibitors are more hydrophobic than noninhibitors. Although charge moderately

correlated with inhibition (correlation coefficient $r=-0.37$; Supplemental Table S1), we did not include it into the final selection because its computed information gain was zero.

Development and validation of binary classification models using machine learning

We used repeated retrospective five-fold cross-validation to find the best performing binary classifier based on the 14 selected features, as follows. First, we partitioned the database into training ($n=180$) and test sets ($n=45$) by sampling known inhibitors and noninhibitors without replacement. Second, binary classifiers were optimized by five-fold cross-validation using the training set. The predictive power of each classifier was assessed on the withheld test set. We tested three widely used machine learning algorithms, k-nearest neighbors (kNN)[46], Support Vector Machine (SVM)[47], and Random forest (RF)[48]. The kNN algorithm is a simple similarity search technique and was used as a baseline for the comparison between algorithms. One hundred repeated iterations of data partitioning, model building, and model testing were performed for each algorithm; model testing relied on the area under the Receiver Operating Characteristic (ROC) curves (auROCs). While all algorithms produced relatively accurate models, the most accurate classifiers (as judged by auROC) were built with the RF algorithm. The average auROC for RF classifiers was 0.85 ± 0.09 , significantly higher than the average auROC values obtained with the kNN (auROC= 0.81 ± 0.11 , p -value= 0.003961) and SVM (auROC= 0.81 ± 0.12 , p -value= 0.008367) algorithms (Figure 2B and Supplemental Figure S2). Additionally, the distribution of the auROC values for RF classifiers had a shorter tail, indicating that the accuracy of the RF algorithm is easier to predict than those of the kNN and SVM algorithms (Figures 2B and 2C).

Given the heuristic nature of feature selection, we tested its robustness by demonstrating that the 14 selected features performed as well as all 219 features (Supplemental Table S2). The

kNN and RF models with the 14 features were more accurate than models built with all features, while the average auROC of the SVM algorithm slightly increased from 0.81 to 0.84. These differences are not statistically significant, thus justifying the use of the selected subset of molecular descriptors.

Binary classifiers output a score between 0 and 1 that predicts the probability of inhibition. A cutoff value for classification must be chosen to classify each compound as an inhibitor or noninhibitor. When a training set contains equal numbers of inhibitors and noninhibitors, a cutoff of 0.5 is typically used[49]. Because our training data set contained 45 inhibitors and 180 noninhibitors (Supplemental Figure S1B), we determined the cutoff by maximizing the classification accuracy (i.e. percent of correctly classified test compounds; Figure 2D). For the RF algorithm, the highest classification accuracy of 84.9% is achieved incidentally at the cutoff of 0.5. A less stringent cutoff of 0.3 decreased the classification accuracy to 78.0%. For ligand-based virtual screening, a less stringent cutoff value would allow for identification of chemically novel inhibitors at the expense of a greater number of false positives, while a stringent cutoff would favor compounds similar to those in the training set and fewer false positives[49].

Virtual screening of the library with the RF classifier

In contrast to the validation of the RF classifier based on retrospective predictions, we trained the RF classifier for prospective predictions using all 225 compounds from the training data set. Next, we applied the resulting classifier to predict the inhibition scores for 5,192 compounds in the DrugBank library[50]. We predicted 92 inhibitors using a stringent cutoff of 0.5, and 509 inhibitors with a less stringent cutoff of 0.3 (Figure 2E). Interestingly, predicted inhibitors clustered into two groups: compounds structurally similar to the training compounds

(Tc greater or equal to 0.6) and putative novel inhibitors, i.e., compounds with low similarity to the training set (Tc less than 0.6; Figure 2F). We selected 21 compounds from four distinct groups for in vitro validation: (1) compounds with high similarity to the training set and high predicted score of inhibition (greater than or equal to 0.5), (2) compounds with high similarity to the training set and scores between 0.3 and 0.5, (3) compounds with low similarity to the training set and scores greater than 0.5, and (4) compounds with low similarity to the training set and scores between 0.3 and 0.5.

In vitro validation of inhibitors predicted with RF model

Predicted inhibitors were tested in the CHO cells stably overexpressing the human OATP2B1 protein. The uptake of the probe substrate estrone-3-sulfate was tested in the presence of 500 μM or 20 μM of the putative inhibitors (Figure 3A). Uptake values were normalized to a 0-100 scale using no inhibition and full inhibition controls, corresponding to 1% DMSO and 50 μM ritonavir, respectively. Protein kinase inhibitors (PKIs) were tested separately; the tested inhibitor concentrations were 8 μM and 200 μM . A test compound was considered an inhibitor if it inhibited the E3S uptake either (1) by more than 50% at the lower concentration (8 μM , 20 μM) or (2) by more than 50% at the higher concentration (200 μM , 500 μM) accompanied by at least 40% decrease in uptake compared to the lower concentration.

In group 1, erlotinib (positive control) inhibited the uptake completely at both concentrations, suggesting IC₅₀ in the nanomolar range. Dabigatran presented high variability at the 500 μM concentration, but met the criteria for inhibition. Doxycycline did not show inhibition across tested concentrations despite high similarity to reported inhibitor tetracycline[38] (Tc=0.68). Compounds in group 2 included a very potent inhibitor (i.e., meloxicam), a weak inhibitor (i.e., nilutamide), a noninhibitor (i.e., ceftriaxone), and non-log-

linear inhibitors (i.e., leucovorin and maraviroc). Non-log linear inhibitors are defined as compounds that showed more than 50% inhibition at the high concentration, but the increase in inhibition was not greater than 40% compared to the lower concentration. Uptake difference of less than 40% suggests that the IC₅₀ curve of an inhibitor does not fit the Hill slope of about -1 and extrapolation of inhibition requires further measurements. Both PKIs (lapatinib and nilotinib) from group 3 were potent inhibitors, despite low similarity to the compounds used to train the model, suggesting expansion of the chemical space by the RF classifier. Interestingly, for the inhibitor lapatinib (over 50% inhibition at 8 μ M), an activation of the E3S uptake appeared at high concentration (200 μ M). Group 4 included 6 inhibitors (vancomycin, tocopherol, amiodarone, ticagrelor, deferoxamine, and sorafenib), 2 non-log-linear inhibitors (bicalutamide and plerixafor), and 2 noninhibitors (sitagliptin and pimozide). In summary, 12 out of the 21 (57%) inhibitors predicted with the RF model were confirmed in vitro.

Comparative structure modeling of human OATP2B1 transporter

Classifiers built with supervised machine-learning algorithms often assign low scores to compounds that are dissimilar to the training compounds,[51, 52] thus labeling these compounds as noninhibitors in our case. For example, among the compounds below the cutoff of 0.3 (Figure 2F), there might be several inhibitors as well as noninhibitors. Therefore, we decided to apply an orthogonal, structure-based approach, to predict OATP2B1 inhibitors missed by the RF classification. Using automated homology modeling by ModWeb[53], we built eighteen comparative models based on experimentally determined atomic structures of homologous Major Facilitator Superfamily (MFS) proteins from prokaryotic and eukaryotic organisms, manually selected seven models, and refined them with Modeller-9v13[54, 55] using the SOAP-Protein statistical potential[56]. Sequence identity between the templates and human OATP2B1 ranged

from 11 to 20% (Supplemental data). Five models were based on inward-facing template structures and two models on outward-facing templates (Supplementary Figure S3), thus allowing us to probe ligand binding on both the intracellular and extracellular sides of the OATP2B1 transporter. We identified several putative binding sites for each comparative model using a fast fragment docking protocol[57] (Figures 4A and 4B). For final docking, the putative binding site with the highest density of docked fragments was selected (Supplemental Data). The positively charged binding sites of the seven comparative models were structurally similar and included several backbone and side chain variations. Several residues in the predicted binding site might be important for interactions with ligands (e.g., Lys70, Glu95, His579, Arg607, and Trp611; Figure 4C). Positively charged residues in the binding site may contribute to the substrate specificity of OATP2B1 because its substrates are anionic. For example, the protonation state of His579 may depend on the pH of the extracellular environment, because it is exposed to the solvent in the predicted outward-facing conformation[58], as previously noted.

Docking and clustering of docked compounds

We docked 5,192 DrugBank compounds against the seven comparative models, followed by unsupervised machine learning to partition all molecules into nine clusters with similar docking profiles. The largest cluster (n=4,826) consisted of compounds that did not dock favorably against any comparative model (Figure 5 and Supplemental Figure S4). For in vitro validation, we selected compounds from clusters enriched with good scoring docking poses, as follows. Three clusters, II (n=263), III (n=26), and VIII (n=44), contained compounds with favorable docking scores against all models, and we hypothesized that these clusters are enriched with OATP2B1 inhibitors (Figure 5, Supplemental Figure S4, and Supplemental Tables S4, S5, and S7). Seven inhibitors from Clusters II, III, and VIII were predicted by the RF model and

already validated in vitro (Figures 3A and 3B). We selected several compounds with the RF scores below 0.3, such as cholecalciferol, finasteride, gentamicin, sulindac, and sunitinib from Cluster III and gefitinib from Cluster VIII. Sunitinib, gentamycin, and finasteride showed some inhibition at the lower concentration and complete inhibition at the higher concentration, while sulindac exhibited no inhibition at the lower concentration and complete inhibition at 500 μM (Figure 3B). Cholecalciferol inhibited E3S uptake by more than 50% at both concentrations, but the inhibition ratio was not log-linear. Gefitinib, an inhibitor of the epidermal growth factor receptor (EGFR) from Cluster VIII, completely inhibited E3S uptake at both concentrations (8 and 200 μM).

As a negative control, we also selected one compound with the RF score below 0.3 from Cluster IV, eplerenone, which did not dock favorably against any comparative model (Supplemental Table S6). As expected, eplerenone did not inhibit E3S uptake at tested concentrations (20 μM , 500 μM). Surprisingly, however, we saw a 4-fold increase in uptake of E3S at the 500 μM eplerenone concentration, indicating that eplerenone may be an allosteric activator of the E3S transport by OATP2B1.

Determination of IC₅₀ values for selected compounds

Half maximal inhibitory concentration (IC₅₀) was determined for 9 validated inhibitors (Figure 6 and Supplemental Table S8). Ticagrelor, meloxicam, sunitinib, and bicalutamide, in particular, were tested to estimate the clinical relevance of OATP2B1 inhibition in the intestine and eplerenone was tested to evaluate the dose dependency of the increased uptake observed in the 2-concentration experiment (Figure 3B). Activation by eplerenone was calculated as a percentage of the difference between E3S uptakes in the absence and presence of 50 μM

OATP2B1 inhibitor ritonavir. Activation of 250% and 900% was achieved for respectively 80 nM and 250 μ M eplerenone.

To estimate whether the IC₅₀ values for ticagrelor, meloxicam, sunitinib, and bicalutamide (2.0 μ M, 6.9 μ M, 15.0 μ M, and 2.4 μ M) might be relevant clinically, we calculated the [I₂]/IC₅₀ ratio of perpetrator drug concentration versus inhibition potency. The 2012 draft US Food and Drug Administration guidance (<http://www.fda.gov>) recommends a cutoff ratio value of higher than 10 for intestinal drug-drug interaction, where I₂ is defined as the drug concentration in the intestinal milieu after oral administration of a single standard dose (90 mg, 15 mg, 50 mg, and 50 mg respectively) divided by a volume of 250 ml[59]. The resulting concentrations for ticagrelor, meloxicam, sunitinib, and bicalutamide were 690 μ M, 170 μ M, 375 μ M, and 464 μ M respectively and the [I₂]/IC₅₀ ratios are 345, 24, 25, and 193 respectively. Therefore, all four compounds could lead to significant inhibition of OATP2B1 transport, potentially causing drug-drug interactions.

Physicochemical properties of validated hits

Finally, we examined physicochemical properties and structural diversity of 17 inhibitors and 12 noninhibitors validated in this work. Overall, the inhibitors were larger and more hydrophobic than noninhibitors, as evidenced by the larger molecular weight, molecular volume, number of heavy atoms, and SLogP (Supplemental Figure S5). OATP2B1 inhibitors also had a significantly larger number of rotatable bonds than noninhibitors (p-value < 0.05), indicating greater structural flexibility, which might be required for binding within a translocation pore. Surprisingly, several OATP2B1 inhibitors were positively charged drugs (pH=7.4), such as antibiotics gentamicin and vancomycin, antiretroviral maraviroc, chelating agent deferoxamine, antiarrhythmic drug amiodarone, and PKI sunitinib. Nonsteroidal anti-inflammatory drug

meloxicam was the only validated anionic inhibitor. Interestingly, positively charged sunitinib, amiodarone, and gentamicin were grouped in Cluster III enriched with known OATP inhibitors (Supplemental Table S5). Overall, half of the validated inhibitors are in Cluster III, suggesting similar physicochemical interactions with OATP2B1 transporter. Despite the high density of validated inhibitors in Cluster III, they were structurally diverse with pairwise 2D and 3D similarity coefficients less than 0.68 and 0.83, respectively (Supplemental Figure S6).

Interestingly, the 5 inhibitors uniquely predicted by the RF model were on average larger and more hydrophobic than the 7 inhibitors identified by both methods or the 5 inhibitors predicted by docking only, although the differences were not statistically significant (Table 1). This observation supports the hypothesis that the two approaches predict compounds with different physicochemical properties, which in turn may result in interaction with different binding sites and/or different binding modes.

DISCUSSION

In human, the organic anion transporting proteins are members of the OATP superfamily. Members of this superfamily mediate sodium-independent transport of structurally diverse amphipathic compounds, such as hormones, bile salts, toxins, and different drugs[60]. In comparison to the hepatic superfamily members, OATP1B1 and OATP1B3, the chemical space of compounds that interact with the intestinal OATP2B1 remains relatively unexplored[14, 38]. Understanding the molecular features that regulate the flux of molecules through OATP2B1 is likely to facilitate design of orally available drugs. A key step toward describing the transport mechanism and substrate specificity is the identification of OATP2B1 ligands. Our goal was to identify inhibitors of OATP2B1-mediated E3S transport in a library of drugs and drug-like molecules. We used two computational approaches, ligand-based and structure-based, to achieve

this goal. While the ligand-based approach may identify inhibitors of any of the three types (competitive substrates, orthosteric inhibitors, and allosteric inhibitors), the structure-based approach predicts only ligands of a defined binding site, thus likely revealing only competitive substrates and orthosteric inhibitors.

Four key findings emerge from this work. First, inhibitors of OATP2B1-mediated transport of E3S can be accurately predicted. Second, four predicted inhibitors ticagrelor, meloxicam, sunitinib, and bicalutamide could inhibit OATP2B1 at clinically relevant concentrations. Third, in addition to predicting inhibitors structurally similar to compounds used in training, statistical models developed with machine learning algorithms can be used to predict new inhibitors that are structurally somewhat dissimilar from compounds used in training. Fourth, despite low sequence identity with bacterial homologues (Supplemental Data), comparative models of the OATP2B1 transporter in different conformations can be used to predict inhibitors missed by the ligand-based approach. We will now address each of these three findings in turn.

The first finding is that OATP2B1 inhibitors of E3S transport can be accurately predicted. We predicted 28 inhibitors of the human OATP2B1 transporter among the 5,192 compounds in the DrugBank library. Seven of these twenty-eight were identified using ligand-based approach, further seven were predicted by docking against comparative models and fourteen were predicted by both methods (Figure 1). We experimentally validated 17 predicted inhibitors (Figure 3); 57% of predicted inhibitors were correctly predicted by the ligand-based approach, 57% by the structure-based approach, and 61% by either method (Figures 3A and 3B).

There has been an increasing interest in combining ligand-based and structure-based approaches, with the aim of maximizing the accuracy, number, and diversity of predicted

ligands[52]. Consensus ranking selects compounds that are top scoring by both methods, while complementary selection includes compounds that are top scoring by either method[61]. The consensus ranking has the advantage of computational efficiency, because computationally inexpensive ligand-based algorithms can propose a subset of promising putative ligands for further examination by the computationally expensive structure-based methods. In this work, complementary selection was more accurate than the consensus ranking, as commonly observed[62]. Additionally, for a specific class of drugs, protein kinase inhibitors, we showed that complementary selection was 100% accurate, while consensus ranking was only 40% accurate. Additionally, the accuracy of our predictions of PKIs as inhibitors of OATP2B1 (100%) was better than the accuracy of the recently reported empirical screen of 13 protein kinase inhibitors (46%)[63].

The second key finding is that we identified putative clinically relevant inhibitors of OATP2B1: ticagrelor, meloxicam, sunitinib, and bicalutamide (Figure 6). These inhibitors were predicted by both ligand- and structure-based methods; meloxicam has a high similarity to peroxicam ($T_c=0.78$), a known inhibitor used in training, and ticagrelor is a structurally novel hit ($T_c=0$; Supplementary Table S3). Four compounds inhibited OATP2B1 with IC_{50} of 2.0 μM , 6.9 μM , 15.0 μM , and 2.4 μM for ticagrelor, meloxicam, sunitinib, and bicalutamide, respectively. Meloxicam is a non-steroidal anti-inflammatory drug used to treat arthritis, a chronic disease with high prevalence in geriatric patients who also take other medications. For example, when meloxicam is co-administered with OATP2B1 substrates, such as statins, it may decrease the absorption and efficacy of statins. Ticagrelor is an adenosine diphosphate (ADP) receptor antagonist used as anticoagulant in case of coronary thrombosis and many of the patients with coronary heart disease also require statin treatment, thus potentially suffering from

a drug-drug interactions. Sunitinib and bicalutamide are mostly used as anti cancer treatment where co-medication profile varies and the risk of a DDI should be evaluated individually.

The third key finding is that statistical models developed with machine learning algorithms can be used to predict new inhibitors, both similar as well as somewhat structurally dissimilar from compounds used in training. Such models have been previously utilized in retrospective modeling of the quantitative structure-activity relationship for human transporters[28-30, 38]. The accuracy of a machine learning method depends on several factors, such as structural diversity of the training set, the ability of the method to learn from a training set with different numbers of active and inactive compounds, and information content of the input features. Here, we show that optimal classifiers can also be used for prospective prediction of inhibitors (Figure 2F). To find a most accurate machine learning algorithm, we evaluated three different algorithms (i.e., kNN, SVM, RF) in repeated retrospective five-fold crossvalidation (Figure 2B). Because the three algorithms vary in their computational complexity and efficiency, we selected the most simple and rapid similarity search method, kNN, as the reference for evaluation of the SVM and RF algorithms. The RF algorithm (auROC=0.85) was significantly more accurate than the SVM (0.81) and kNN algorithms (0.81). To avoid overfitting, we selected 14 molecular descriptors from the total of 219 features computed for the training set (Figure 2A). Among the selected molecular descriptors, we confirmed previously reported dependency between the inhibition strength and inhibitor hydrophobicity[38]. In contrast, we did not identify charge and molecular weight as features with high information content, presumably because noninhibitors in the training set included all compounds that did not inhibit E3S uptake by more than 50%, even the putative activators. We validated 12 out of predicted 21 OATP2B1 inhibitors (57%) from the DrugBank library of over 5000 compounds.

Although ligand-based approaches rely on the training set of known ligands, our results show that they can also predict somewhat chemically novel inhibitors. For example, 62% of predicted structurally novel inhibitors of E3S transport were confirmed experimentally (Figure 2E and Supplemental Table S3).

The fourth key finding is that docking against multiple comparative models of OATP2B1 transporter in different conformations predicts inhibitors from a different subset of physicochemical space, thus identifying inhibitors missed with the ligand-based approach. Compounds predicted to interact favorably with some or all OATP2B1 comparative models are clustered in three out of the 9 observed clusters (Figure 5 and Supplemental Figure S4). For experimental validation, we selected 21 predicted inhibitors from these three groups; 12 of them were validated in vitro (Figure 3B). Seven of the 12 inhibitors were identified by both approaches, ligand- and structure-based. Five inhibitors (cholecalciferol, finasteride, gefitinib, gentamicin, and sunitinib) were predicted by structure-based method only; they are structurally dissimilar (pairwise Tc less than 0.6) from inhibitors identified using ligand-based approach (Supplemental Figure S6). This finding is not surprising, as it appears that OATP2B1 transporter binds different ligands at different sites within the translocation pore[19].

The predicted positively charged pore of OATP2B1 is surrounded by transmembrane helices (TM) 2, 8, 9, and 10 (Figure 5). Using fragment-based docking, we identified several residues that might be important for the interaction with OATP2B1 inhibitors: Lys70, Glu95, His579, Arg607, and Trp611. Among the 39 members of the OATP superfamily, histidine residue is only found in OATP2B1 (Supplemental Data), potentially rationalizing its unique localization in gut[1, 58]. Glu95 in the predicted TM2 is highly conserved among OATP superfamily members and the corresponding glutamic acid residue (Glu74) in the hepatic

homologue, OATP1B1, was critical for low concentration E3S uptake identified by alanine-scanning and site-directed mutagenesis[64]. The function of this residue in OATP2B1 is still unknown; it has been suggested that positively and negatively charged residues in the binding site of the OATP transporters form salt bridges to stabilize the binding site[65]. Our results show that several inhibitors predicted using docking are positively charged, suggesting that it is also possible that Glu95 is directly involved in interactions with OATP2B1 inhibitors. Both positively charged residues in predicted binding sites, Lys70 and Arg607, participate in predicted interactions between OATP2B1 and favorably docked compounds; while Arg607 is involved in contacts with all predicted inhibitors, Lys70 is less frequently required for favorable docking. Both Lys70 and Arg607 are conserved among all members of OATP superfamily (Supplementary data). In OATP1B3, mutations of corresponding arginine residue (Arg580) for alanine and lysine resulted in reduced transport of sulphobromophthalein (BSP) and pravastatin[66]. Further experimental studies are needed to elucidate the role of these two residues in inhibition of OATP2B1-mediated transport of E3S.

In summary, we demonstrated that new inhibitors of OATP2B1 transporter can be accurately predicted by ligand- and structure-based methods. The physicochemical properties of ligands predicted with these two methods differ, suggesting different modes of interaction with OATP2B1. Among the predicted inhibitors, several compounds were structurally novel. The IC50 values of ticagrelor and meloxicam suggest that they might inhibit intestinal OATP2B1 at clinically relevant concentrations and therefore modulate the actions of other concomitantly administered drugs.

MATERIALS AND METHODS

Pairwise compound similarity computation and clustering

Two-dimensional MACCS fingerprints and pairwise Tanimoto similarity coefficients were computed using MayaChemTools (<http://www.mayachemtools.org>). Three-dimensional shape/feature similarity coefficients (Shape ST-Optimized, 10 configurations) were computed using on-line PubChem Chemical Structure Clustering Tool[67] with default settings. Hierarchical compound clustering was performed in R using the hclust function of the stats library[68].

Selection of informative molecular descriptors and machine learning

Compound names were collected from ref.[38] and used to retrieve molecular structure files from the PubChem database[67]. PaDEL software[69] and cxcalc program in the ChemAxon (<http://www.chemaxon.com>) package were used to compute for each molecule 863 descriptors and molecular charge, respectively. Experimentally determined percent inhibition values[38] were converted to binary class labels for compounds, with '1' encoding percent inhibition value greater than or equal to 50% and '0', otherwise. All data analyses and computational modeling were performed using R[68]. Molecular descriptors with near zero variance were removed. Pairwise descriptor correlations were computed to identify highly correlated descriptor pairs (correlation coefficient > 0.95) and one randomly chosen descriptor was removed from each highly correlated pair. The cfs filtering algorithm in the FSelector package[68] was used to find most informative descriptors. Three machine learning algorithms from the caret package in R, k-nearest Neighbors, Support Vector Machine, and Random forest, were employed to build binary classifiers with selected features. A double loop five-fold cross-

validation[30] protocol was used to access the predictive power of each classifier. Double loop cross-validation was repeated 100 times and the average area under the Receiver Operating Characteristic curve (auROC)[70] was computed to evaluate the accuracy of the three algorithms. Differences between the distributions of auROC values were quantified with the Student's t-test in R. The algorithm with the highest average auROC (Random forest) was selected for further modeling. The final binary classifier for virtual screening was built by utilizing all compounds from the training set and optimizing algorithm's parameters in a five-fold cross-validation. To show that classifiers' performance does not degrade when only a subset of features is utilized, final binary classifier was evaluated using all 219 uncorrelated features with non-zero variance in 100 double loop cross-validation experiments.

Virtual screening of the DrugBank library using Random forest classifier

Molecular structure files of 5,805 compounds were downloaded from the DrugBank 3.0 database[50]. Fourteen molecular descriptors selected in the previous step were successfully computed with PaDEL[69] for 5,192 compounds; the remaining 613 were ignored in further analyses. The final Random forest classifier was employed to compute a classification score for each DrugBank compound; the classifier's score is a number from zero to one, which is interpreted as a probability of inhibition. Compounds with the predicted inhibition probability higher than a user-defined cutoff were labeled as inhibitors. Two cutoff values were utilized in this work, 0.3 and 0.5. To prioritize candidate hits for in vitro validation, all DrugBank compounds were ranked using their predicted inhibition scores and the results were manually inspected. Chemical novelty of predicted inhibitors was assessed by computing MACCS fingerprints and pairwise Tanimoto similarity coefficients between test and training compounds with MayaChemTools (<http://www.mayachemtools.org>).

Comparative modeling of the 3D structure of the human OATP2B1 and fragment-based identification of putative binding sites

Search for suitable templates, multiple sequence alignment, and initial comparative modeling was performed using our on-line server ModWeb[53]. Comparative models were manually examined and seven models (template PDB codes: 1PV7, 1PW4, 3WOD, 4GBZ, 4LDS, 4IKY, 4J05) were selected for further refinement. Models with charged helical residues facing the membrane were removed from further analysis. For each retained homology model, one hundred additional homology models were generated using the default automodel protocol of MODELLER-9v13[54, 55] and assessed by using the SOAP-Protein score[56]. Predicted intracellular and extracellular loops were not modeled. For each template, the model with the lowest SOAP-Protein score was used in subsequent analyses. Putative binding sites within the central substrate-translocation pore of the seven comparative models of OATP2B1 were mapped using the FT-MAP server[57] with default settings. The best scoring predicted binding site was used in further analyses.

Virtual screening of the DrugBank library against the comparative models of OATP2B1

Compounds from the DrugBank library[50] were docked against each predicted binding site by using a semi-automated docking protocol[44, 71] implemented in DOCK 3.5.54[72]. For each compound, a raw docking score was computed for each binding pose as a sum of three terms, van der Waals, Poisson–Boltzmann finite-difference electrostatics, and ligand desolvation penalty; the lowest scoring binding pose was retained for each ligand. The largest docking score of all docked compounds (1000) was assigned to compounds that failed docking. For each compound, average docking score was computed as a mean value of seven raw DOCK scores and 5,192 docking profiles (i.e., a profile is a vector of docking scores of a compound against

each of the seven comparative models) were clustered using bi-clustering heatmap function implemented in the stats library in R. Additionally, docking profiles were clustered into nine clusters using unsupervised expectation-maximization algorithm implemented in the mclust library in R. Densities of average raw DOCK scores were estimated using the density function in R. Compounds from three clusters (III, IV, and VIII) were ranked using their average DOCK scores. Docking poses against each model were manually inspected to prioritize hits for in vitro validation.

Reagents

All reagents were obtained from Sigma-Aldrich (St. Louis, MI) and Santa Cruz Biotechnology (Santa Cruz, CA) and were of analytical grade with at least 95% purity, unless otherwise stated. All cell culture media and supplements were obtained from the Cell Culture Facility at the University of California, San Francisco, CA.

Cell culture

Chinese hamster ovary (CHO) cells stably expressing OATP2B1 were a kind gift from Optivia Biotechnology, Inc. (Menlo Park, CA). CHO cells were grown in Dubelcco's Modified Eagle's Medium (DMEM, Cell Culture Facility, UCSF) substituted with 10% Foetal Bovine Serum (FBS, Cell Culture Facility, UCSF), 100 U/mL penicillin, 100 μ g/mL streptomycin, 1% non-essential amino acids (NEAA), 2mM L-glutamine and 500 μ g/mL geneticin at 37 °C in a humidified atmosphere with 5% CO₂.

Inhibition experiments with CHO-OATP2B1

CHO cells stably transfected with OATP2B1 were seeded at a density of 100,000 cells/ml in 24-well plates approximately 48 h prior to the experiments. On the day of the experiments, cells were washed with 0.5 ml Hank's buffered salt solution (HBSS) per well and then pre-incubated for 10 minutes in 0.5 ml HBSS per well. To assess inhibition, the cells were exposed to uptake buffer (0.1 μ M unlabeled and 20 nM [3H]-labeled estrone sulfate in HBSS) containing either 8 μ M or 200 μ M inhibitor (for tyrosine kinase inhibitors) or 20 μ M or 500 μ M of inhibitor (all other compounds). Uptake was stopped after 2 minutes by washing twice with ice-cold HBSS. The cells were lysed in 800 μ L lysis buffer (0.1 N NaOH and 0.1% SDS in bi-distilled water) per well while shaking for 2 hours. 650 μ L of the lysate were then added to 3 mL EcoLite scintillation fluid (MP Bio) and the radioactivity was determined on a LS6500 Scintillation Counter (Beckman Coulter, Pasadena, CA). Values were corrected for protein concentration as determined with a BCA assay kit (Thermo Scientific, Rockford, IL). All values were determined in triplicate and final values are expressed as % uptake relative to negative controls (2% DMSO).

Determination of IC50 values

CHO-OATP2B1 were seeded as described above. On the day of the experiments, cells were washed with 0.5 ml Hank's buffered salt solution (HBSS) per well and then pre-incubated for 10 minutes in 0.5 ml HBSS per well. To assess inhibition, the cells were exposed to uptake buffer (0.1 μ M unlabeled and 20 nM [3H]-labeled estrone sulfate in HBSS) containing raising concentrations of inhibitor from 0.08 to 250 μ M. Uptake was stopped after 2 minutes by washing twice with ice-cold HBSS. Radioactivity of samples was determined as described above. All values were determined in triplicate and uptake was compared to positive (50 μ M ritonavir) and negative (1% DMSO) controls.

ACKNOWLEDGEMENTS

This work has been supported by NIH grants (GM071790, GM093342, and GM008284; R01 GM054762 to A.S.). We are also grateful for computing hardware gifts from Mike Homer, Ron Conway, NetApp, IBM, Hewlett Packard and Intel.

REFERENCES

1. Kobayashi, D., et al., Involvement of human organic anion transporting polypeptide OATP-B (SLC21A9) in pH-dependent transport across intestinal apical membrane. *J Pharmacol Exp Ther*, 2003. **306**(2): p. 703-8.
2. Banerjee, N., C. Allen, and R. Bendayan, Differential role of organic anion-transporting polypeptides in estrone-3-sulphate uptake by breast epithelial cells and breast cancer cells. *J Pharmacol Exp Ther*, 2012. **342**(2): p. 510-9.
3. Obaidat, A., M. Roth, and B. Hagenbuch, The expression and function of organic anion transporting polypeptides in normal tissues and in cancer. *Annu Rev Pharmacol Toxicol*, 2012. **52**: p. 135-51.
4. Imanaga, J., et al., The effects of the SLCO2B1 c.1457C > T polymorphism and apple juice on the pharmacokinetics of fexofenadine and midazolam in humans. *Pharmacogenet Genomics*, 2011. **21**(2): p. 84-93.
5. Nakanishi, T. and I. Tamai, Genetic polymorphisms of OATP transporters and their impact on intestinal absorption and hepatic disposition of drugs. *Drug Metab Pharmacokinet*, 2012. **27**(1): p. 106-21.
6. Shirasaka, Y., et al., Intestinal absorption of HMG-CoA reductase inhibitor pravastatin mediated by organic anion transporting polypeptide. *Pharm Res*, 2010. **27**(10): p. 2141-9.
7. Tamai, I., Oral drug delivery utilizing intestinal OATP transporters. *Adv Drug Deliv Rev*, 2012. **64**(6): p. 508-14.
8. Tamai, I. and T. Nakanishi, OATP transporter-mediated drug absorption and interaction. *Curr Opin Pharmacol*, 2013. **13**(6): p. 859-63.
9. Shirasaka, Y., et al., Long-lasting inhibitory effect of apple and orange juices, but not grapefruit juice, on OATP2B1-mediated drug absorption. *Drug Metab Dispos*, 2013. **41**(3): p. 615-21.
10. Shirasaka, Y., et al., Differential effect of grapefruit juice on intestinal absorption of statins due to inhibition of organic anion transporting polypeptide and/or P-glycoprotein. *J Pharm Sci*, 2011. **100**(9): p. 3843-53.

11. Grube, M., et al., Organic anion transporting polypeptide 2B1 is a high-affinity transporter for atorvastatin and is expressed in the human heart. *Clin Pharmacol Ther*, 2006. **80**(6): p. 607-20.
12. Noe, J., et al., Substrate-dependent drug-drug interactions between gemfibrozil, fluvastatin and other organic anion-transporting peptide (OATP) substrates on OATP1B1, OATP2B1, and OATP1B3. *Drug Metab Dispos*, 2007. **35**(8): p. 1308-14.
13. Kitamura, S., et al., Involvement of multiple transporters in the hepatobiliary transport of rosuvastatin. *Drug Metab Dispos*, 2008. **36**(10): p. 2014-23.
14. Kullak-Ublick, G.A., et al., Organic anion-transporting polypeptide B (OATP-B) and its functional comparison with three other OATPs of human liver. *Gastroenterology*, 2001. **120**(2): p. 525-33.
15. Treiber, A., et al., Bosentan is a substrate of human OATP1B1 and OATP1B3: inhibition of hepatic uptake as the common mechanism of its interactions with cyclosporin A, rifampicin, and sildenafil. *Drug Metab Dispos*, 2007. **35**(8): p. 1400-7.
16. Satoh, H., et al., Citrus juices inhibit the function of human organic anion-transporting polypeptide OATP-B. *Drug Metab Dispos*, 2005. **33**(4): p. 518-23.
17. Nozawa, T., et al., Functional characterization of pH-sensitive organic anion transporting polypeptide OATP-B in human. *J Pharmacol Exp Ther*, 2004. **308**(2): p. 438-45.
18. Varma, M.V., et al., pH-sensitive interaction of HMG-CoA reductase inhibitors (statins) with organic anion transporting polypeptide 2B1. *Mol Pharm*, 2011. **8**(4): p. 1303-13.
19. Shirasaka, Y., et al., Functional pleiotropy of organic anion transporting polypeptide OATP2B1 due to multiple binding sites. *Drug Metab Pharmacokinet*, 2012. **27**(3): p. 360-4.
20. Hu, G., et al., New fluorescent substrate enables quantitative and high-throughput examination of vesicular monoamine transporter 2 (VMAT2). *ACS Chem Biol*, 2013. **8**(9): p. 1947-54.
21. Zhou, W., et al., Development and validation of a high-throughput screening assay for human long-chain fatty acid transport proteins 4 and 5. *J Biomol Screen*, 2010. **15**(5): p. 488-97.
22. Sindelar, M. and K.T. Wanner, Library screening by means of mass spectrometry (MS) binding assays-exemplarily demonstrated for a pseudostatic library addressing gamma-aminobutyric acid (GABA) transporter 1 (GAT1). *ChemMedChem*, 2012. **7**(9): p. 1678-90.
23. Zhang, L., et al., A high-throughput screen for chemical inhibitors of exocytic transport in yeast. *Chembiochem*, 2010. **11**(9): p. 1291-301.
24. Susa, M., et al., Multidrug resistance reversal agent, NSC77037, identified with a cell-based screening assay. *J Biomol Screen*, 2010. **15**(3): p. 287-96.
25. Badolo, L., et al., Screening of OATP1B1/3 and OCT1 inhibitors in cryopreserved hepatocytes in suspension. *Eur J Pharm Sci*, 2010. **40**(4): p. 282-8.

26. Gui, C., et al., Development of a cell-based high-throughput assay to screen for inhibitors of organic anion transporting polypeptides 1B1 and 1B3. *Curr Chem Genomics*, 2010. **4**: p. 1-8.
27. Karaki, F., et al., Structure-activity relationship study of non-steroidal NPC1L1 ligands identified through cell-based assay using pharmacological chaperone effect as a readout. *Bioorg Med Chem*, 2014. **22**(14): p. 3587-609.
28. De Bruyn, T., et al., Structure-based identification of OATP1B1/3 inhibitors. *Mol Pharmacol*, 2013. **83**(6): p. 1257-67.
29. Wittwer, M.B., et al., Discovery of potent, selective multidrug and toxin extrusion transporter 1 (MATE1, SLC47A1) inhibitors through prescription drug profiling and computational modeling. *J Med Chem*, 2013. **56**(3): p. 781-95.
30. Kido, Y., P. Matsson, and K.M. Giacomini, Profiling of a prescription drug library for potential renal drug-drug interactions mediated by the organic cation transporter 2. *J Med Chem*, 2011. **54**(13): p. 4548-58.
31. Esteva-Font, C., et al., A small molecule screen identifies selective inhibitors of urea transporter UT-A. *Chem Biol*, 2013. **20**(10): p. 1235-44.
32. Strouse, J.J., et al., A selective ATP-binding cassette subfamily G member 2 efflux inhibitor revealed via high-throughput flow cytometry. *J Biomol Screen*, 2013. **18**(1): p. 26-38.
33. Mohamed, T.M., et al., Optimisation and validation of a high throughput screening compatible assay to identify inhibitors of the plasma membrane calcium ATPase pump--a novel therapeutic target for contraception and malaria. *J Pharm Pharm Sci*, 2013. **16**(2): p. 217-30.
34. Colton, C.K., et al., Identification of translational activators of glial glutamate transporter EAAT2 through cell-based high-throughput screening: an approach to prevent excitotoxicity. *J Biomol Screen*, 2010. **15**(6): p. 653-62.
35. Zhang, Y., et al., Identification of inhibitors of ABCG2 by a bioluminescence imaging-based high-throughput assay. *Cancer Res*, 2009. **69**(14): p. 5867-75.
36. Gao, J., et al., Identification of upregulators of human ATP-binding cassette transporter A1 via high-throughput screening of a synthetic and natural compound library. *J Biomol Screen*, 2008. **13**(7): p. 648-56.
37. 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.
38. Karlgren, M., et al., Classification of inhibitors of hepatic organic anion transporting polypeptides (OATPs): influence of protein expression on drug-drug interactions. *J Med Chem*, 2012. **55**(10): p. 4740-63.
39. Schlessinger, A., et al., High selectivity of the gamma-aminobutyric acid transporter 2 (GAT-2, SLC6A13) revealed by structure-based approach. *J Biol Chem*, 2012. **287**(45): p. 37745-56.

40. Geier, E.G., et al., Structure-based ligand discovery for the Large-neutral Amino Acid Transporter 1, LAT-1. *Proc Natl Acad Sci U S A*, 2013. **110**(14): p. 5480-5.
41. Schlessinger, A., et al., Molecular modeling and ligand docking for solute carrier (SLC) transporters. *Curr Top Med Chem*, 2013. **13**(7): p. 843-56.
42. Schlessinger, A., et al., Structure-based discovery of prescription drugs that interact with the norepinephrine transporter, NET. *Proc Natl Acad Sci U S A*, 2011. **108**(38): p. 15810-5.
43. Fan, H., J.J. Irwin, and A. Sali, Virtual ligand screening against comparative protein structure models. *Methods Mol Biol*, 2012. **819**: p. 105-26.
44. Fan, H., et al., Molecular docking screens using comparative models of proteins. *J Chem Inf Model*, 2009. **49**(11): p. 2512-27.
45. Kullback, S. and R.A. Leibler, On Information and Sufficiency. 1951: p. 79-86.
46. Altman, N.S., An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician*, 1992. **46**(3): p. 175-185.
47. Cortes, C. and V. Vapnik, Support-vector networks. *Machine Learning*, 1995. **20**(3): p. 273-297.
48. Breiman, L., Random Forests. *Machine Learning*, 2001. **45**(1): p. 5-32.
49. Freeman, E.A. and G.G. Moisen, A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling*, 2008. **217**(1-2): p. 48-58.
50. Knox, C., et al., DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res*, 2011. **39**(Database issue): p. D1035-41.
51. Balfer, J., M. Vogt, and J. Bajorath, Searching for closely related ligands with different mechanisms of action using machine learning and mapping algorithms. *J Chem Inf Model*, 2013. **53**(9): p. 2252-74.
52. Lavecchia, A. and C. Di Giovanni, Virtual screening strategies in drug discovery: a critical review. *Curr Med Chem*, 2013. **20**(23): p. 2839-60.
53. Pieper, U., et al., ModBase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res*, 2014. **42**(Database issue): p. D336-46.
54. Webb, B. and A. Sali, Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics*, 2014. **47**: p. 5 6 1-5 6 32.
55. Sali, A. and T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 1993. **234**(3): p. 779-815.
56. Dong, G.Q., et al., Optimized atomic statistical potentials: assessment of protein interfaces and loops. *Bioinformatics*, 2013. **29**(24): p. 3158-66.
57. Brenke, R., et al., Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics*, 2009. **25**(5): p. 621-7.

58. Leuthold, S., et al., Mechanisms of pH-gradient driven transport mediated by organic anion polypeptide transporters. Vol. 296. 2009. C570-C582.
59. Agarwal, S., V. Arya, and L. Zhang, Review of P-gp inhibition data in recently approved new drug applications: utility of the proposed [I(1)]/IC(50) and [I(2)]/IC(50) criteria in the P-gp decision tree. *J Clin Pharmacol*, 2013. **53**(2): p. 228-33.
60. Shitara, Y., et al., Clinical significance of organic anion transporting polypeptides (OATPs) in drug disposition: their roles in hepatic clearance and intestinal absorption. *Biopharm Drug Dispos*, 2013. **34**(1): p. 45-78.
61. Tan, L., et al., Integrating structure- and ligand-based virtual screening: comparison of individual, parallel, and fused molecular docking and similarity search calculations on multiple targets. *ChemMedChem*, 2008. **3**(10): p. 1566-71.
62. Bissantz, C., et al., Focused library design in GPCR projects on the example of 5-HT(2c) agonists: comparison of structure-based virtual screening with ligand-based search methods. *Proteins*, 2005. **61**(4): p. 938-52.
63. Johnston, R.A., et al., Selective inhibition of human solute carrier transporters by multikinase inhibitors. *Drug Metab Dispos*, 2014. **42**(11): p. 1851-7.
64. Li, N., et al., Identification of amino acids essential for estrone-3-sulfate transport within transmembrane domain 2 of organic anion transporting polypeptide 1B1. *PLoS One*, 2012. **7**(5): p. e36647.
65. Rizwan, A.N., W. Krick, and G. Burckhardt, The chloride dependence of the human organic anion transporter 1 (hOAT1) is blunted by mutation of a single amino acid. *J Biol Chem*, 2007. **282**(18): p. 13402-9.
66. Glaeser, H., et al., Relevance of conserved lysine and arginine residues in transmembrane helices for the transport activity of organic anion transporting polypeptide 1B3. *Br J Pharmacol*, 2010. **159**(3): p. 698-708.
67. Kaiser, J., Science resources. Chemists want NIH to curtail database. *Science*, 2005. **308**(5723): p. 774.
68. Team, R.C., R: A language and environment for statistical computing. 2012.
69. Yap, C.W., PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem*, 2011. **32**(7): p. 1466-74.
70. Hanley, J.A. and B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 1982. **143**(1): p. 29-36.
71. Huang, N., B.K. Shoichet, and J.J. Irwin, Benchmarking sets for molecular docking. *J Med Chem*, 2006. **49**(23): p. 6789-801.
72. Mysinger, M.M. and B.K. Shoichet, Rapid context-dependent ligand desolvation in molecular docking. *J Chem Inf Model*, 2010. **50**(9): p. 1561-73.

TABLES

Table 5-1. Physicochemical properties of OATP2B1 inhibitors identified with different computational methods.

MW: molecular weight, HA: number of heavy atoms, MV: molecular volume, RB: number of rotatable bonds, HBD: hydrogen bond donors, HBA: hydrogen bond acceptors, TPSA: total polar surface area.

Method	MW	HA	MV	RB	HBD	HBA	SLogP	TPSA	Charge
RF (n=5)	746.48	52.60	672.31	14.60	6.00	14.60	5.36	213.61	0.60
RF+docking (n=7)	476.78	31.29	381.52	7.71	2.00	7.00	5.15	93.24	0.00
Docking (n=5)	416.03	29.60	406.76	6.00	3.00	6.00	4.29	86.07	1.20
All (n=17)	538.23	37.06	474.47	9.24	3.47	8.94	4.96	126.53	0.53

FIGURES

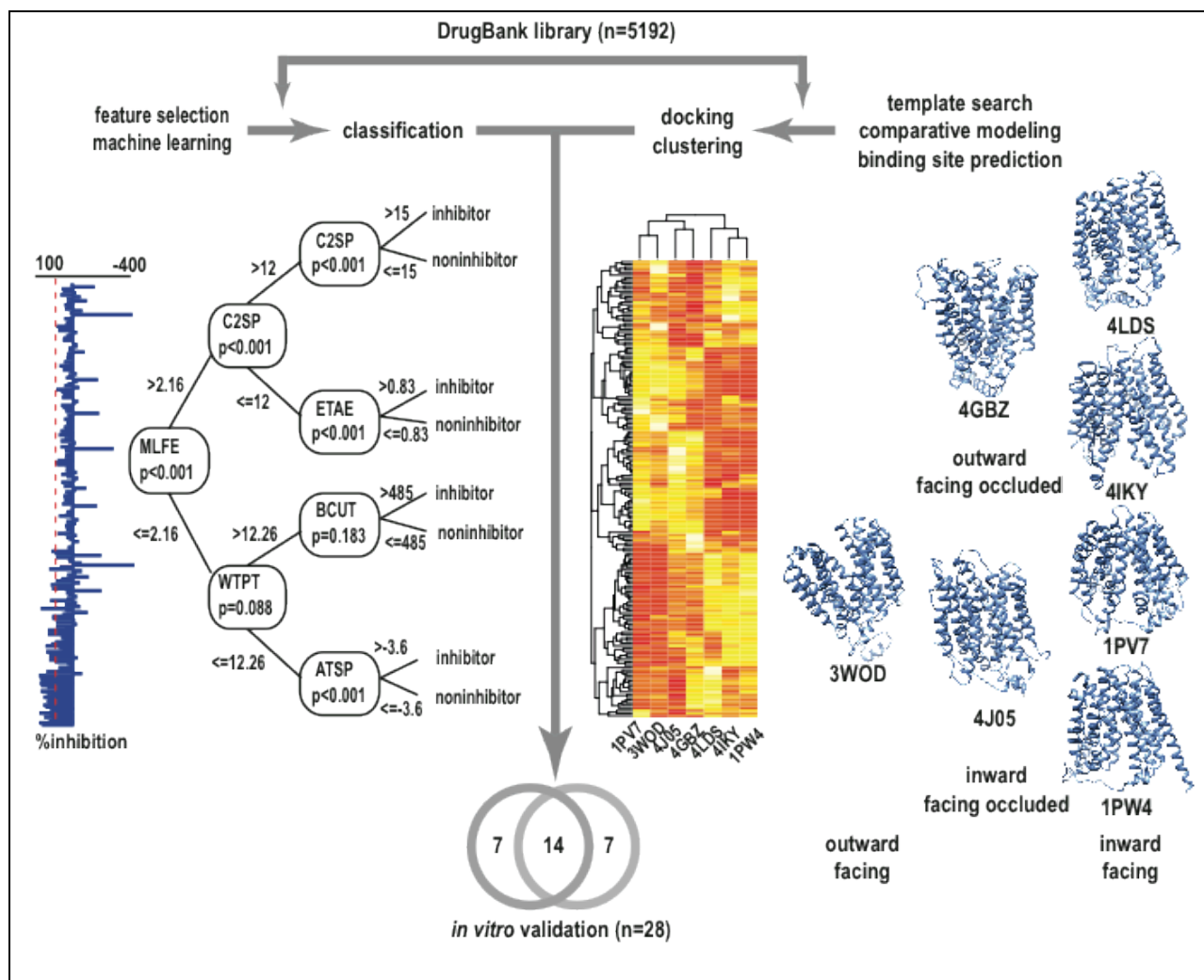


Fig. 5-1. In silico OATP2B1 inhibitor discovery pipeline.

Putative OATP2B1 inhibitors were predicted by docking DrugBank library of 5,192 compounds against seven comparative models of human OATP2B1 transporter in three different conformations and by classifying DrugBank compounds with Random forest algorithm trained with known inhibitors and noninhibitors. Twenty-eight predicted inhibitors were selected for *in vitro* validation.

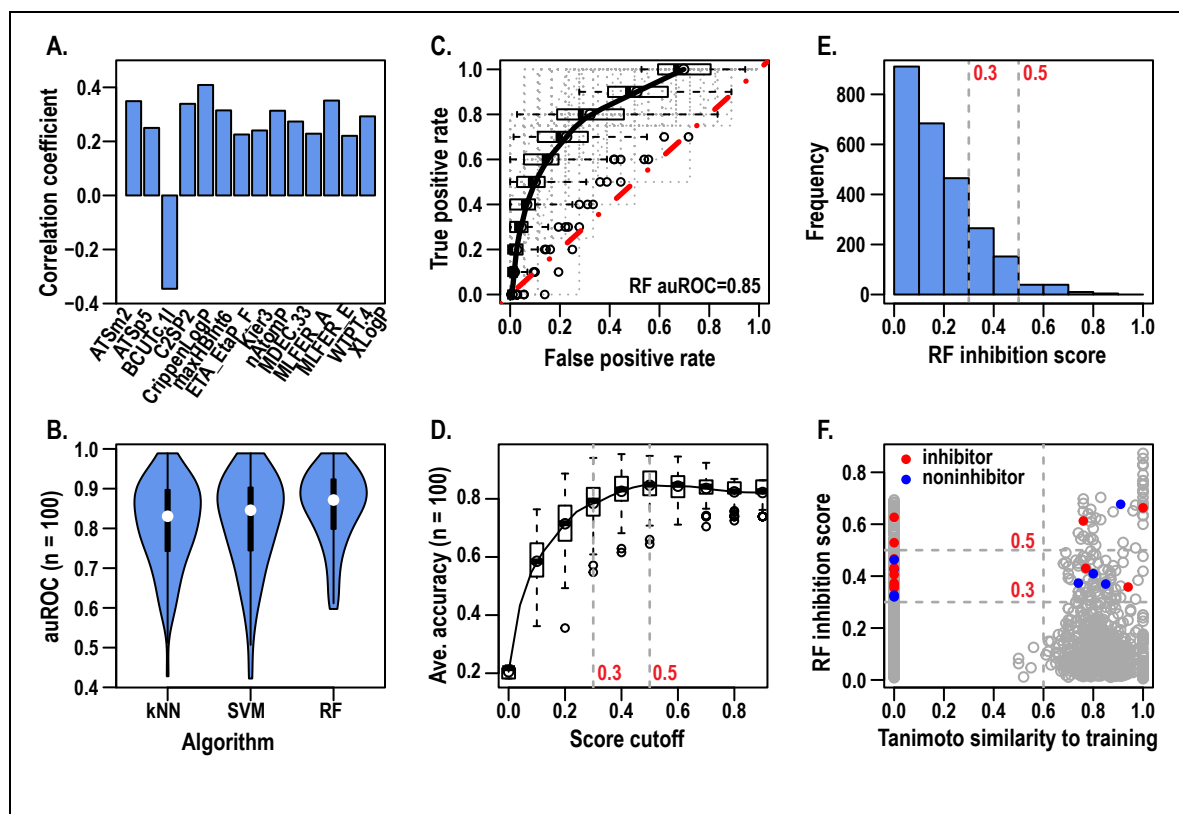


Fig. 5-2. Results of feature selection, ligand-based modeling, and virtual screening of DrugBank library.

A. Fourteen most informative features derived from training data and their correlation with the inhibition values: Moreau-Broto autocorrelation descriptors using atomic weight (ATSm2) and polarizability (ATSp5), eigenvalue based descriptor noted for its utility in chemical diversity (BCUTc-11), carbon connectivity in terms of hybridization (C2SP2), atom-based LogP calculated using Crippen's approach (CrippenLogP), atom type electrotopological state (MaxHBint6), extended topochemical atoms (ETA_EtaP_F), Kappa shape index (Kier3), largest Pi system (nAtomP), molecular distance edge (MDEC.33), molecular linear free energy relations (MLFER_A and MLFER_E), weighted path (WTPT.4), and XLogP. B. Distribution of average auROCs for models built using three machine learning algorithms, k-nearest neighbors (kNN),

Support Vector Machine (SVM), and Random forest (RF). C. Receiver Operating Characteristic (ROC) curve of 100 independent tests of binary RF classifiers. Average ROC curve is shown in black. ROC for a random classification is shown as a red dotted line. D. Distribution of accuracies in 100 independent tests of binary RF classifiers as a function of cutoff used in classification. Average accuracy is shown as a black solid line. Vertical gray dotted lines are drawn at cutoffs 0.3 and 0.5, respectively. E. Histogram of the number of predicted OATP2B1 inhibitors at different classification cutoffs. Vertical dotted lines are drawn at cutoffs 0.3 and 0.5, respectively. F. Plot of predicted inhibition scores for 5,192 DrugBank compounds versus highest pairwise 2D Tanimoto similarity to training set. Each compound is shown as circle. Compounds tested for inhibition in vitro are shown as filled circles: validated inhibitors are red and noninhibitors blue. Horizontal dotted lines are classification cutoffs (0.3 and 0.5) and the vertical dotted line is a pairwise 2D Tanimoto similarity cutoff (0.6).

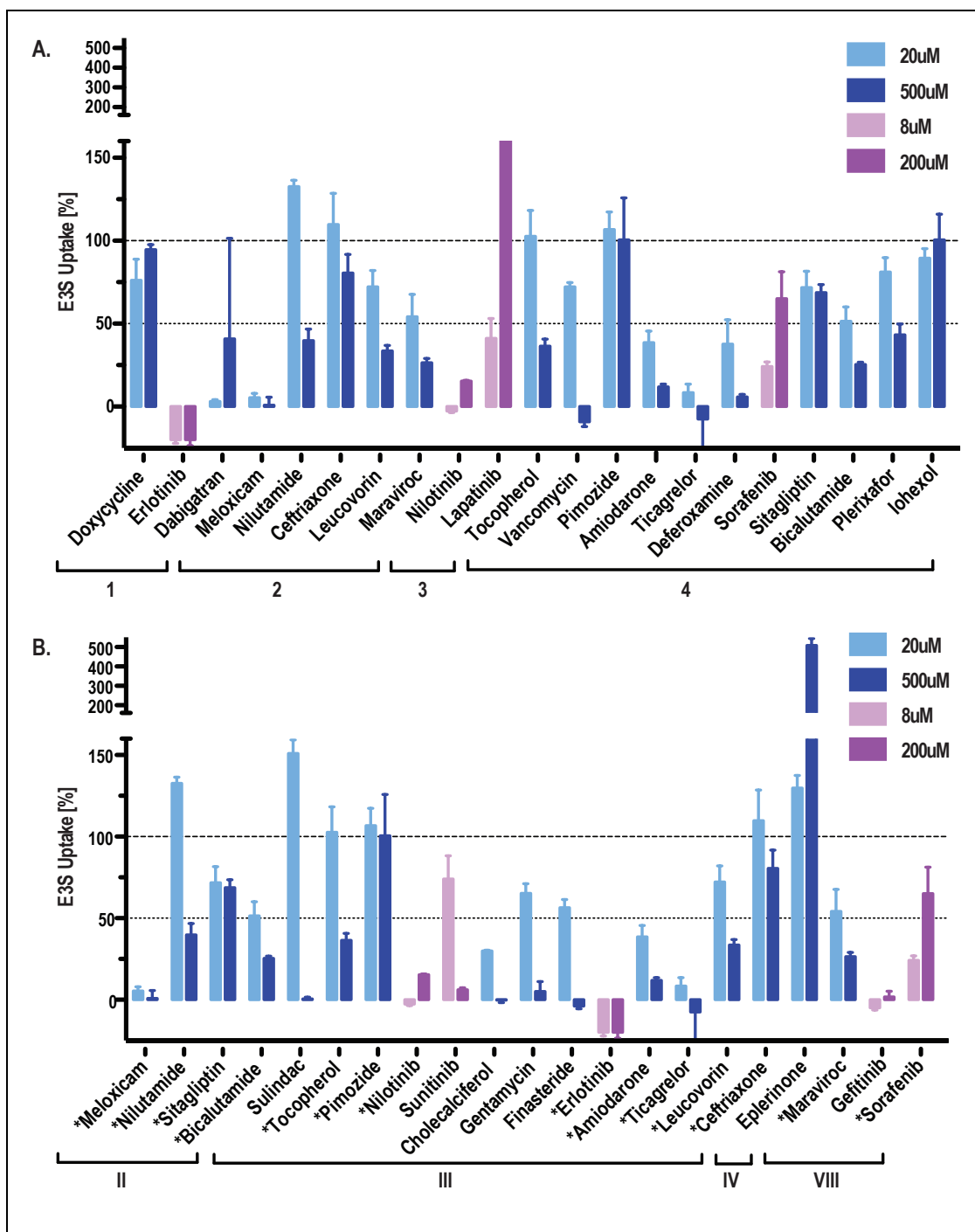


Fig. 5-3. 2-concentration inhibition of E3S uptake in human OATP2B1 over expressing cells.

Twenty-eight compounds were tested for inhibition of 3H-E3S uptake (0.1µM) in CHO-hOATP2B1 cells. Twenty-one compounds were predicted with ligand-based approach (Panel A),

and twenty-one compounds with structure-based approach (represented by roman numbers, Panel B). Data is presented as percent uptake relative to controls (1% DMSO is 100% uptake and 50 μ M ritonavir is 0% uptake controls). Inhibitor concentrations tested were 20 μ M (light blue) and 500 μ M (dark blue) and 8 μ M (pink) and 200 μ M (purple) for protein kinase inhibitors. Data is presented as mean \pm standard deviation.

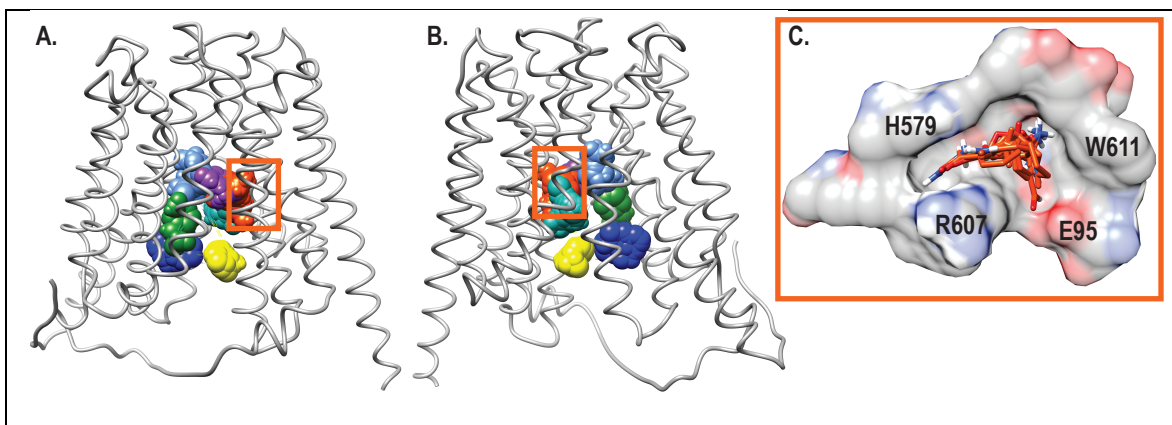


Fig. 5-4. Results of binding site prediction for comparative models of the human OATP2B1 transporter.

A. Several predicted binding sites are shown as colored spheres on the representative comparative model built using template structure in inward-facing conformation (PDBID: 1PW4). Best scoring predicted binding site is shown in orange. B. 180° rotation of inward-facing comparative model. C. Surface representation of the best scoring predicted binding site in inward-facing conformation. Fragments docked using FTMap software are shown as orange sticks.

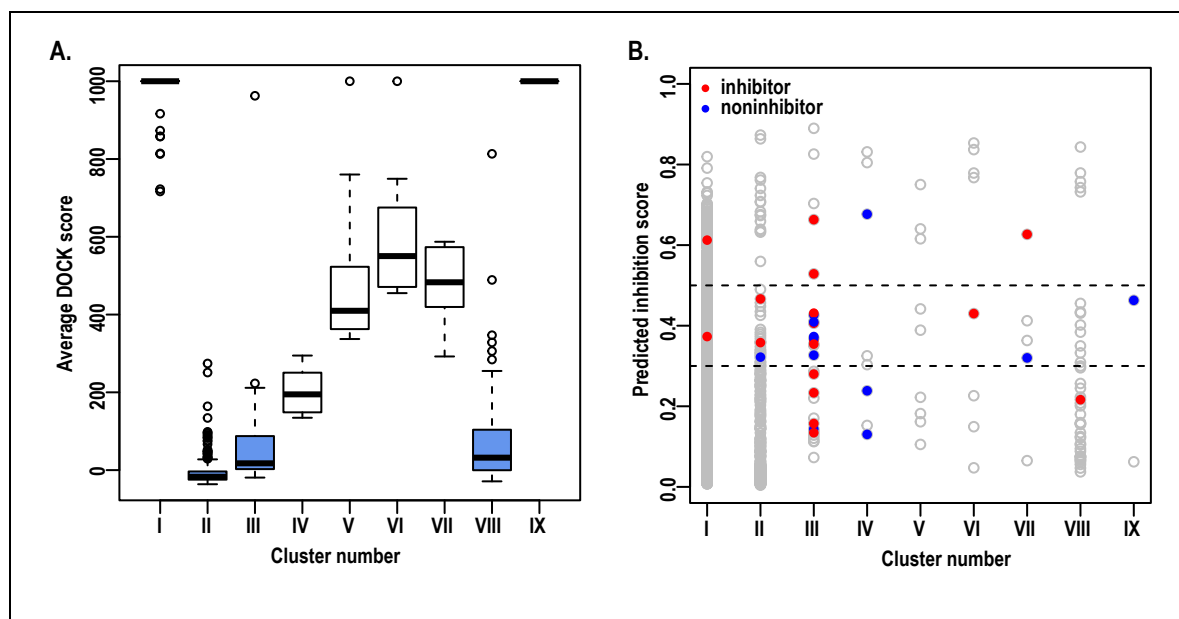


Fig. 5-5. Results of the structure-based virtual screening of DrugBank library.

A. Distribution of average DOCK scores in nine clusters. Compounds with good docking scores against each of the seven comparative models have low average DOCK scores. Compounds in clusters III, IV, VIII (shown in blue) were manually prioritized for in vitro validation. B. Plot of RF inhibition scores of compounds in nine clusters. Each DrugBank compound is shown as a grey circle. Red filled circles show confirmed inhibitors and blue filled circles noninhibitors. Horizontal grey dotted lines show RF cutoffs of 0.3 and 0.5, respectively.

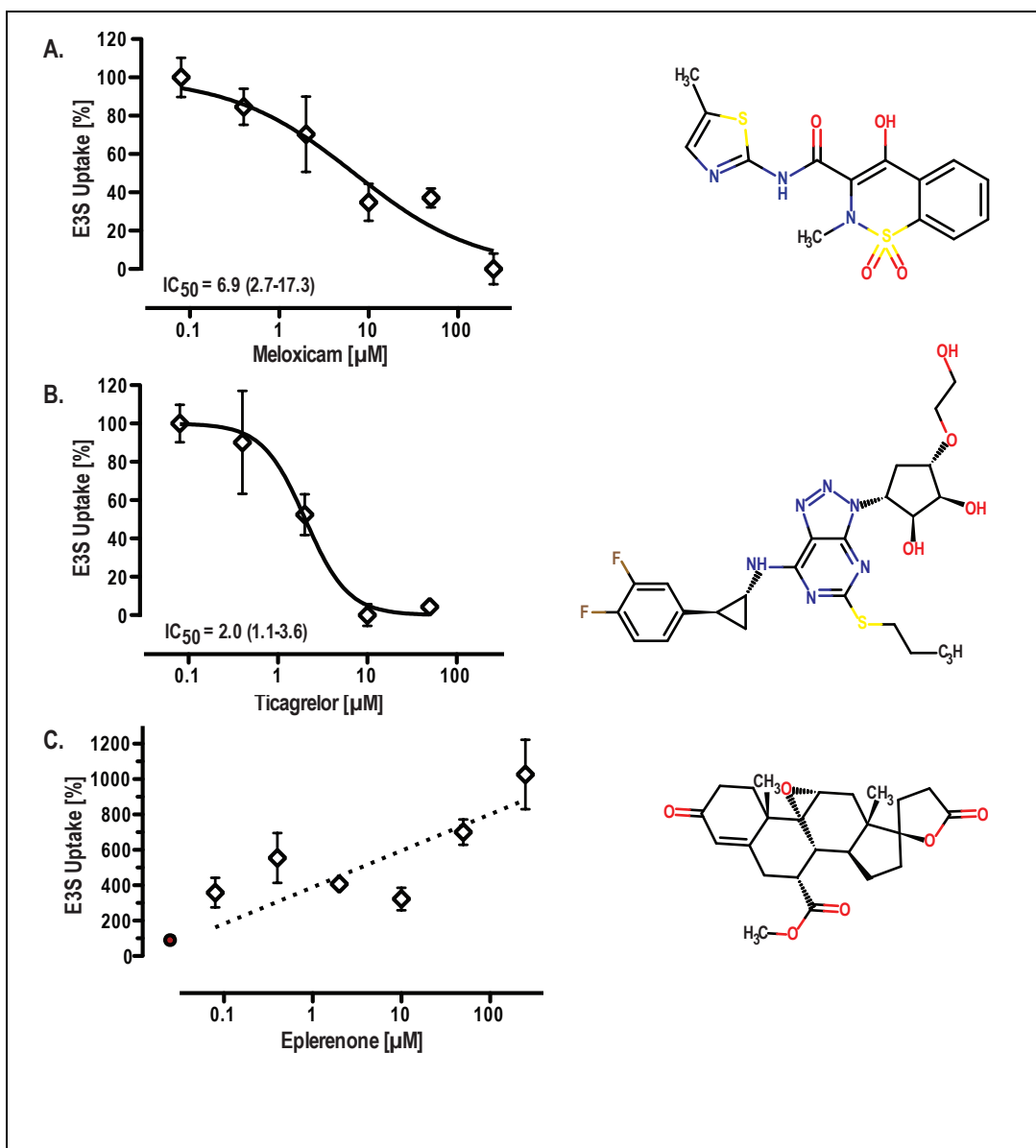


Fig. 5-6. Results of IC₅₀ determination for putative clinically relevant compounds.

Inhibition of E3S uptake in CHO-hOATP2B1 cells was tested across a concentration range of 80 nM to 250 μM. Data is presented as percent uptake relative to controls (1% DMSO is 100% uptake and 50 μM ritonavir is 0% uptake controls). Data for each concentration tested is marked with a diamond and presented as mean ± standard deviation. Solid lines are non-linear regression

with a variable slope used to calculate the IC50. Dashed line depicts linear trend of induction starting from 1% DMSO control value (red dot).

SUPPLEMENTARY INFORMATION

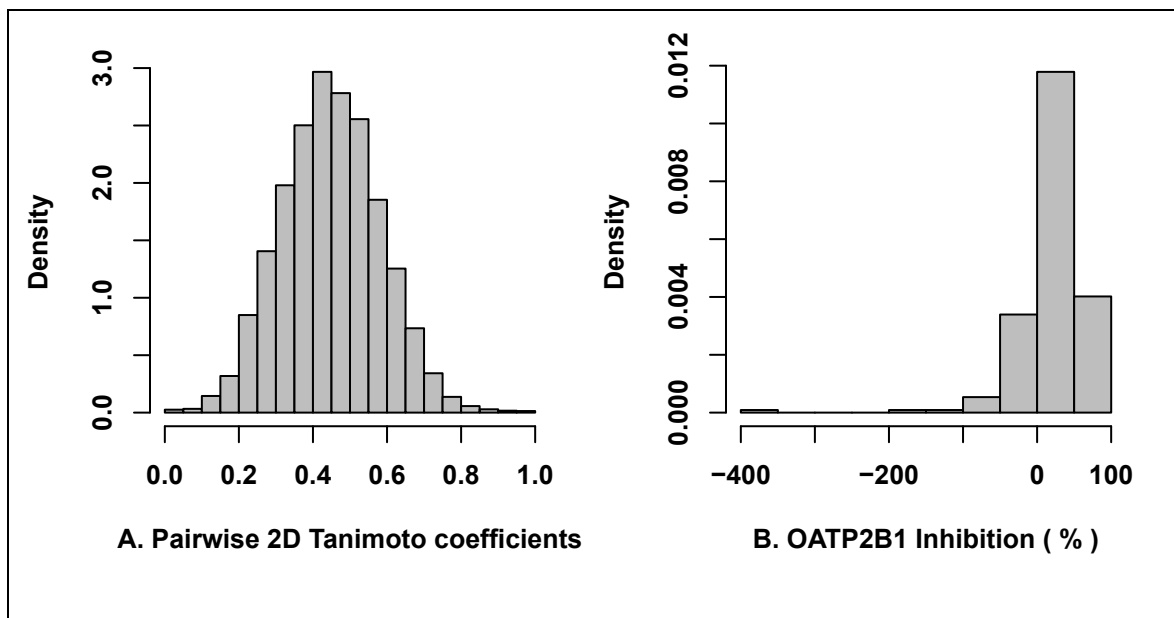


Fig. 5-S1. Training set diversity and inhibition activity.

A. Distribution of pairwise 2D Tanimoto similarity coefficients between compounds in the training set. B. Distribution of inhibition values reported for compounds in the training set. Compounds with percent inhibition greater than or equal than 50 were labeled as inhibitors of OATP2B1.

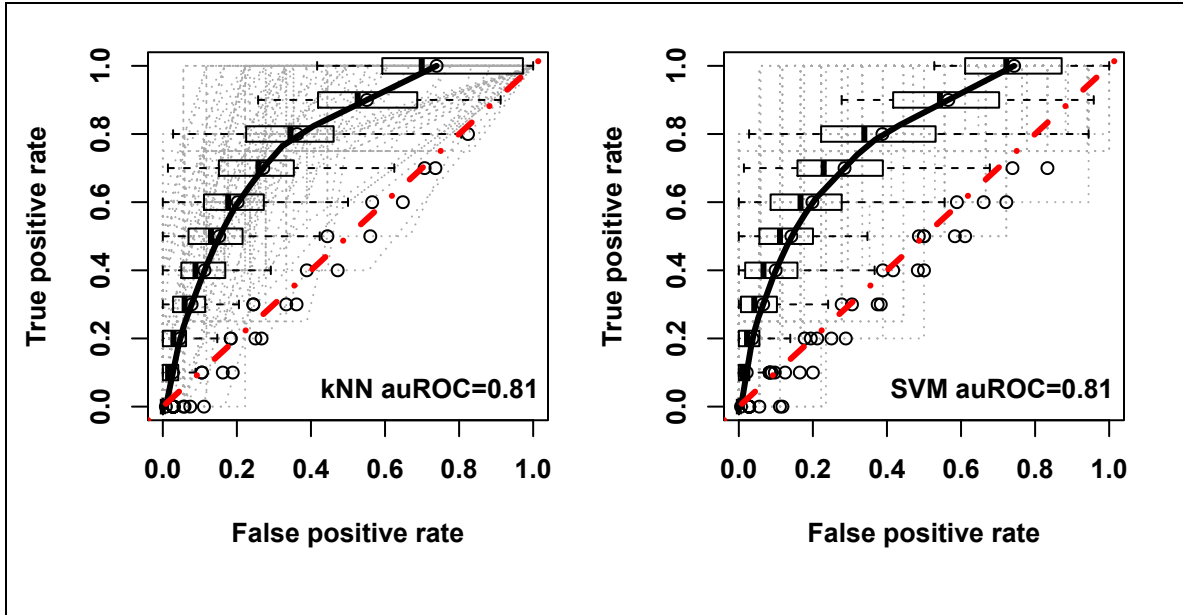


Fig. 5-S2. Receiver operating characteristic curves for 100 independent tests using kNN and SVM algorithms.

Random classifier ROCs are shown as the dotted red lines

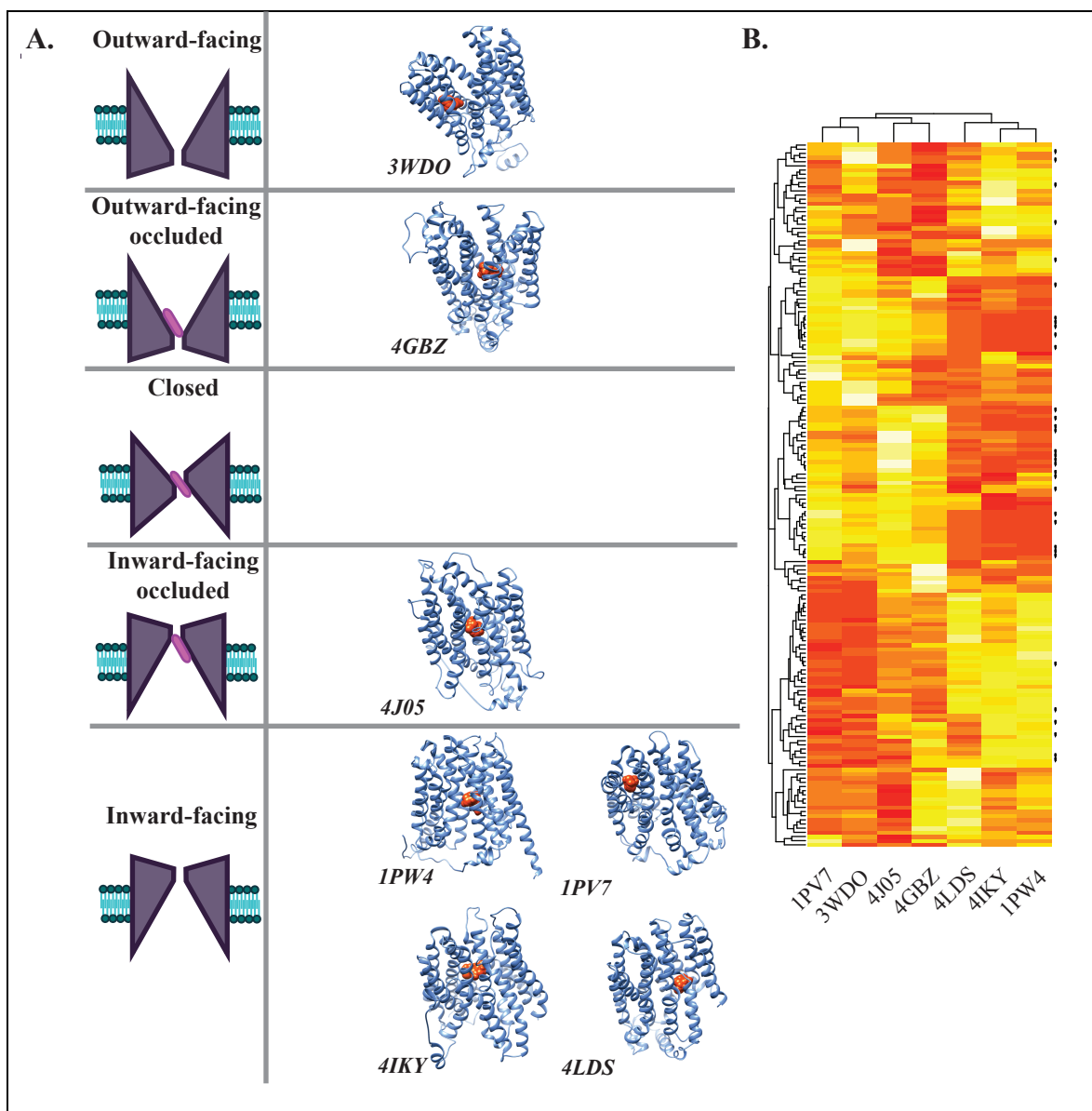


Fig. 5-S3. Seven templates utilized for comparative structure modeling and two-way clustering of docking against the seven comparative models.

A. Seven comparative models of OATP2B1 built using six prokaryotic and one eukaryotic atomic structures (3WDO: Escherichia coli YajR transporter at 3.15 Å resolution in an outward-facing conformation, 4GBZ: Escherichia coli Xyle transporter at 2.8 Å resolution in outward-facing, partly occluded conformation, 4J05: Piriformospora indica PipT transporter at 2.6 Å

resolution in inward-facing occluded conformation, 1PW4: Escherichia coli GlpT transporter at 3.3 Å resolution in inward-facing conformation, 1PV7: Escherichia coli LacY transporter at 3.5 Å resolution in inward-facing conformation, 4IKY: Geobacillus kaustophilus POT transporter at 1.9 Å resolution in inward-facing conformation, and 4LDS: Staphylococcus epidermidis GlcP transporter at 3.2 Å resolution in inward-facing conformation). Predicted binding site used for docking is shown as orange spheres. B. Heatmap of docking scores of 225 known inhibitors and noninhibitors against the seven comparative models. Rows are clustered based on the similarity between docking scores of compounds. Columns are clustered based on similarity between all docked scores against each model. Red denotes favorable docking score and yellow unfavorable. Black dots on the right of the heatmap denote docking scores of known inhibitors.

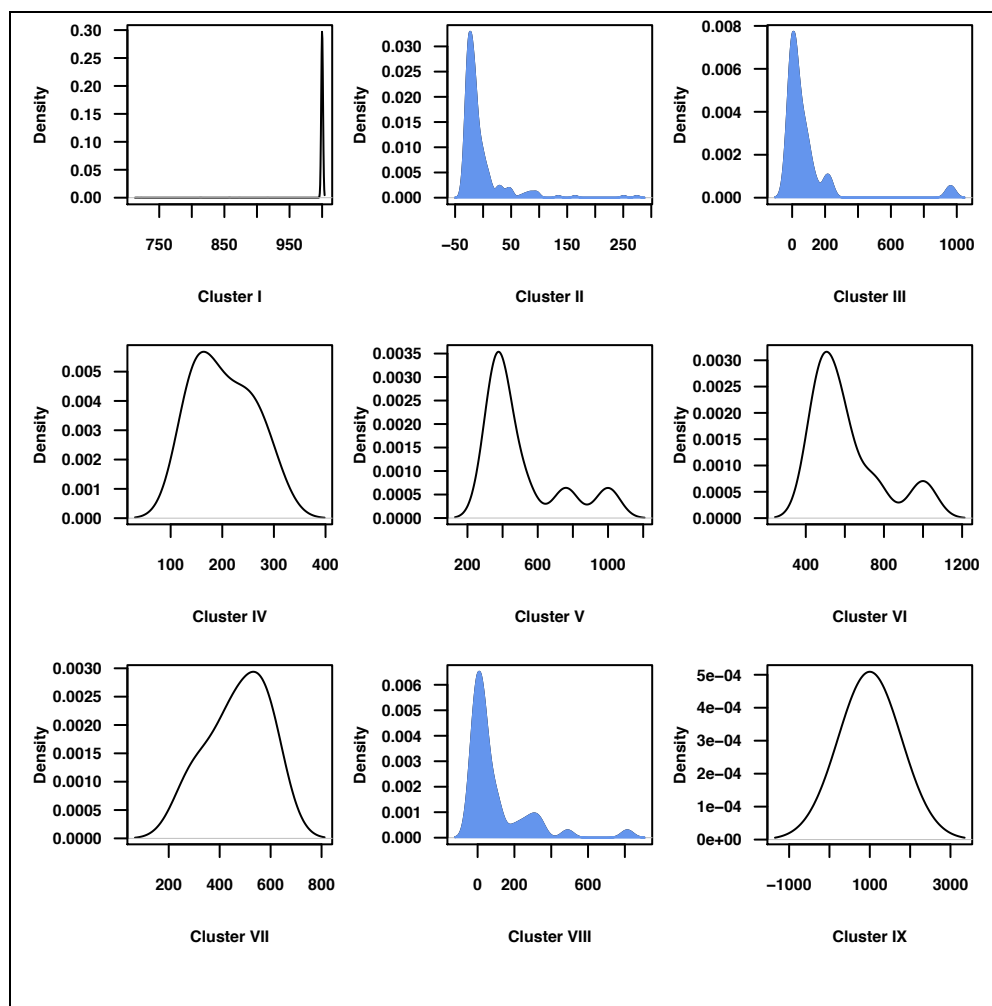


Fig. 5-S4. Density of DOCK scores in each cluster.

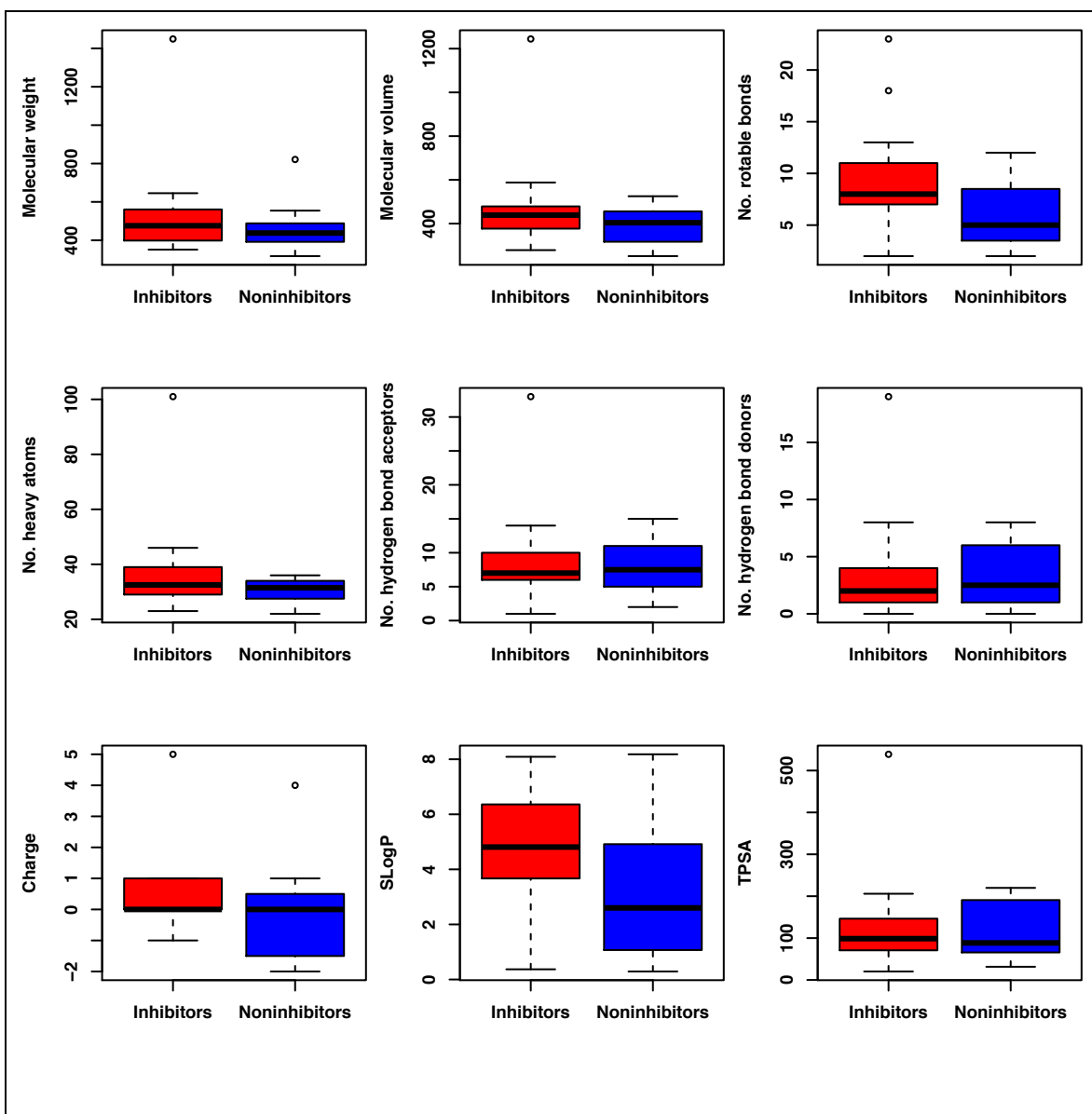


Fig. 5-S5. Box plots of physicochemical properties of OATP2B1 inhibitors and noninhibitors identified in this study.

OATP2B1 inhibitors have significantly greater number of rotatable bonds than noninhibitors (Student's t-test p-value < 0.05). The differences between other molecular properties are not statistically significant.

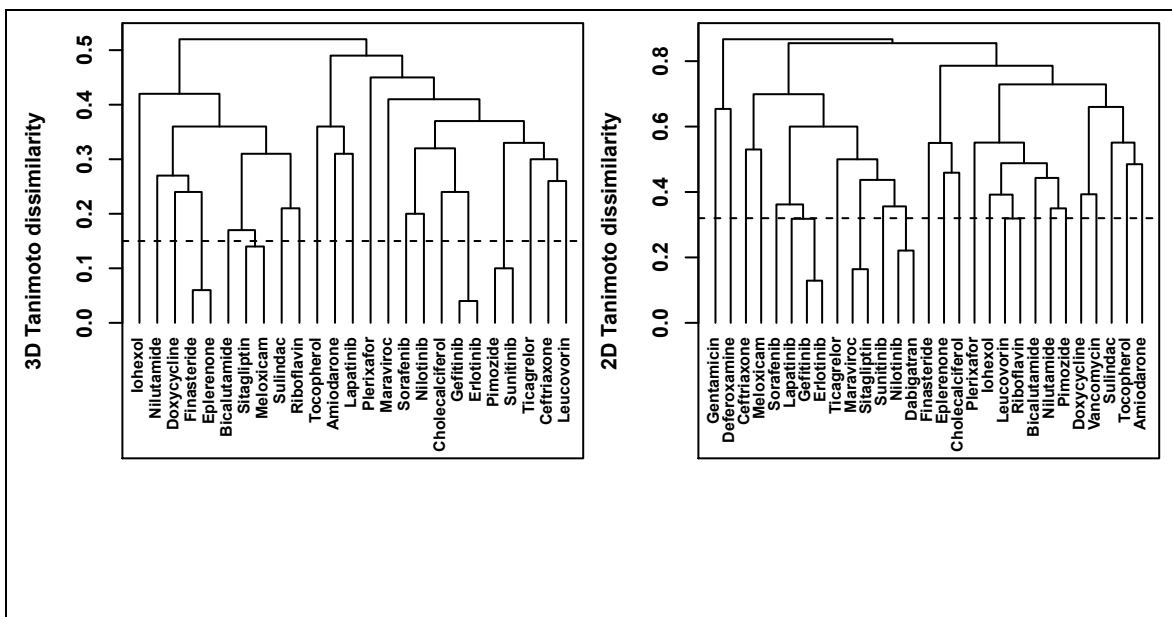


Fig. 5-S6. Structural similarity of predicted OATP2B1 inhibitors.

Hierarchical clustering of pairwise 2D Tanimoto dissimilarity coefficients. Compounds clustering below the dissimilarity cutoff of 0.32 are considered structurally similar at 95% confidence level. Hierarchical clustering of pairwise 3D Tanimoto dissimilarity coefficients (right panel); clusters below the dissimilarity cutoff of 0.15 consist of similar compounds (at 95% confidence level).

SUPPLEMENTARY ON-LINE DATA FILE

<http://salilab.org/~nkhuri/supplements/khuri-oatp2b1-2014.zip> (available after publication).

Multiple sequence alignment of OATP2B1 and template sequences, PDB files of comparative models with docked fragments.

Chapter 6: ModBase, a database of annotated comparative protein structure models, and associated resources

Ursula Pieper¹, Benjamin M. Webb¹, Guang Qiang Dong¹, Dina Schneidman-Duhovny¹, Hao Fan¹, Seung Joong Kim¹, Natalia Khuri^{1,2}, Yannick Spill^{3,4}, Patrick Weinkam¹, Michal Hammel⁵, John A. Tainer^{6,7}, Michael Nilges³, and Andrej Sali^{1§}

¹Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences, Byers Hall at Mission Bay, Office 503B, University of California at San Francisco, 1700 4th Street, San Francisco, CA 94158, USA. tel: +1 415 514 4227; fax: +1 415 514 4231. ²Graduate Group in Biophysics, University of California at San Francisco. ³Unité de Bioinformatique Structurale, Institut Pasteur, 25 rue du Docteur Roux, 75015 Paris, France. ⁴Université Paris Diderot - Paris 7, Paris Rive Gauche, 5 rue Thomas Mann, 75013 Paris, France. ⁵Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA. ⁶Department of Molecular Biology, Skaggs Institute of Chemical Biology, The Scripps Research Institute, La Jolla, California, USA. ⁷Life Sciences Division, Department of Molecular Biology, Lawrence Berkeley National Laboratory, Berkeley, California, USA.

§Corresponding author

Email sali@salilab.org; URL <http://salilab.org/>

September 15, 2013

Nucleic Acids Research

ABSTRACT

ModBase (<http://salilab.org/modbase>) is a database of annotated comparative protein structure models. The models are calculated by ModPipe, an automated modeling pipeline that relies primarily on Modeller for fold assignment, sequence-structure alignment, model building, and model assessment (<http://salilab.org/modeller/>). ModBase currently contains almost 30 million reliable models for domains in 4.7 million unique protein sequences. ModBase allows users to compute or update comparative models on demand, through an interface to the ModWeb modeling server (<http://salilab.org/modweb>). ModBase models are also available through the Protein Model Portal (<http://www.proteinmodelportal.org/>). Recently developed associated resources include the AllosMod server for modeling ligand-induced protein dynamics (<http://salilab.org/allosmod>), the AllosMod-FoXS server for predicting a structural ensemble that fits a SAXS profile (<http://salilab.org/allosmod-foxs>), the FoXSDock server for protein-protein docking filtered by a SAXS profile (<http://salilab.org/foxsdock>), the SAXS Merge server for automatic merging of SAXS profiles (<http://salilab.org/saxsmerge>), and the Pose & Rank server for scoring protein-ligand complexes (<http://salilab.org/poseandrank>). In this update, we also highlight two applications of ModBase, a PSI:BiologY initiative to maximize the structural coverage of the human α -helical transmembrane proteome and a determination of structural determinants of HIV-1 protease specificity.

INTRODUCTION

The genome sequencing efforts are providing us with the complete genetic blueprints of thousands of organisms, including many eukaryotic genomes. We are now faced with the challenge of assigning, investigating, and modifying the functions of proteins encoded by these genomes. This task is generally facilitated by the knowledge of the 3D protein structures, which

are best determined by experimental methods such as X-ray crystallography and NMR-spectroscopy. While the number of experimentally determined structures deposited in the Protein Data Bank (PDB)[259] increased by nearly 40% to approximately 93,000 in the last 3 years (September 2013), the number of sequences in the comprehensive sequence databases, such as UniProtKB[260] and GenPept[261], continues to grow even more rapidly; for example, the number of sequences in UniProtKB has now reached over 41 million, compared to 12 million only three years ago. Therefore, protein structure prediction is essential to bridge this gap. The need for accurate models can frequently be met by homology or comparative modeling[262-271]. Comparative modeling is carried out in four sequential steps: identifying known structures (templates) related to the sequence to be modeled (target), aligning the target sequence with the templates, building models, and assessing the models. For this reason, comparative modeling is only applicable when the target sequence is detectably related to a known protein structure.

As more proteins are modeled, web-accessible resources that assist biologists in evaluating and analyzing models become increasingly useful. Here, we describe the current state of the ModBase database of comparative protein structure models, the ModWeb comparative modeling web-server, and several new associated resources, including web-servers that utilize SAXS data in the context of comparative modeling: The AllosMod server for modeling ligand-induced protein dynamics (<http://salilab.org/allosmod>)[272], the AllosMod-FoXS server for predicting the ensemble of conformations that best fit a given SAXS profile (<http://salilab.org/allosmod-foxs>) (Weinkam et al, in preparation), the FoXSDock server that performs protein-protein docking filtered by a SAXS profile (<http://salilab.org/foxsdock>)[273], the SAXS Merge server for merging SAXS profiles (<http://salilab.org/saxsmerge>) (Spill et al, in preparation), and the Pose & Rank server for scoring protein-ligand complexes based on a

statistical potential (<http://salilab.org/poseandrank>)[274]. Finally, we highlight applications of ModBase models to maximizing the structural coverage of the human α -helical transmembrane proteome in a PSI:BiologY effort and to an analysis of structural determinants of HIV-1 protease specificity.

METHODS

Model generation by comparative modeling (Modeller and modpipe)

Models in ModBase are calculated using our automated software pipeline for comparative protein structure modeling, ModPipe[275]. ModPipe relies mostly on modules of Modeller[243] as well as fold assignment and sequence-structure alignment by PSI-BLAST[276] and the HHSuite modules HHBlits[277] and HHSearch[278]. To be able to process a large number of sequences, it is implemented on a Linux cluster.

ModPipe uses sequence-sequence[279], sequence-profile[276, 280], and profile-profile[263, 281] methods for fold assignment and target-template alignment, using a promiscuous E-value threshold of 1.0 to increase the likelihood of identifying the best available template structure. In addition to the previously implemented profile methods (Modeller's Build-Profile and PPScan, and PSI-BLAST), we recently added an option to use HHBlits and HHSearch. These will be included in the next public release of ModPipe (2.3.0, expected December 2013). Alignments created by any of these methods can cover the complete target sequence, or only a segment of it, depending on the availability of suitable PDB templates. With the added functionality of HHBlits and HHSearch, some ModPipe models are now based on multiple templates.

To increase efficiency, the available target-template alignments are filtered by sequence identity (ModPipe template option: TOP): If the highest target-template sequence identity is 40% or lower, ModPipe selects alignments for all detected templates. Otherwise, the selection only contains alignments for each target-template alignment that is created in a 20% sequence identity window starting from the highest sequence identity. For each selected target-template alignment, 10 models are calculated[243] and the model with the best value of the DOPE statistical potential[282] is selected and then evaluated by several additional quality criteria: (i) target-template sequence identity, (ii) GA341 score[283], (iii) Z-DOPE score[282], (iv) MPQS score (ModPipe Quality Score)[284], and (v)TSVMod score[285]. The models that score best with at least one of these quality criteria are selected for further filtering. If more than 30 residues of a target sequence are not covered by a selected model, additional models are selected even if they don't score best with at least one of the quality criteria. Finally, only the models with quality criteria values above specified thresholds or with an E-value smaller than 10^{-4} are included in the final model set.

A key feature of the pipeline is that the validity of sequence–structure relationships is not prejudged at the fold-assignment stage; instead, sequence–structure matches are assessed after the construction of the models and their evaluation. This approach enables a thorough exploration of fold assignments, sequence–structure alignments and conformations, with the aim of finding the model with the best evaluation score, at the expense of increasing the computational time significantly; for some sequences, a few thousand models can be calculated. For sequences with high quality templates, the optional “TOP” keyword can reduce the amount of computer time by up to 60%.

The source code for ModPipe is freely accessible under the Gnu Public license (<http://salilab.org/modpipe>). The binary code for Modeller is also available freely to academics for a number of different operating systems (<http://salilab.org/modeller>).

SOAP statistical potentials: optimized statistical potentials for assessing protein interfaces and loops

Both loop modeling and protein-protein docking require accurate scoring functions for selecting the most accurate sampled models. SOAP-PP and SOAP-Loop are atomic statistical potentials for assessing protein interfaces and loops, respectively (<http://salilab.org/SOAP>, also available in Modeller)[286]. They were derived using a Bayesian framework for inferring Statistically Optimized Atomic Potentials (SOAP). When using SOAP-PP for scoring protein-protein docking models, a near-native model is within the top 10 scoring models in 52% of the PatchDock decoys[287], compared to 23% and 27% for the state-of-the-art ZRANK[288] and FireDock[289] scoring functions, respectively. Similarly, for modeling 12-residue loops in the PLOP benchmark[290], the average main-chain RMSD of the best scored conformations by SOAP-Loop is 1.5Å, close to the average RMSD of the best sampled conformations (1.2Å) and significantly better than that selected by the Rosetta[291] (2.1Å), DFIRE[292] (2.3Å), DOPE (2.5Å)[282], and PLOP scoring functions (3.0Å). The SOAP-PP score is utilized by our AllosMod-FoXS server (below). We are incorporating SOAP scores into the modeling and model assessment modules of ModPipe.

ModBase model sets

Models in ModBase are organized in datasets. Because of the rapid growth of the public sequence databases, we concentrate our efforts on adding datasets that are useful for specific projects, rather than attempt to model all known protein sequences based on all detectably related

known structures. Currently, ModBase includes a model dataset for each of 65 complete genomes, as well as datasets for all sequences in the Structure Function Linkage Database (SFLD)[293], and for the complete SwissProt/TrEMBL database as of 2005 (<http://salilab.org/modbase/statistics>). Additionally, available models for new SFLD sequences are added weekly. Together with other project-oriented datasets, ModBase currently contains approximately 29 million reliable models for domains in 4.7 million unique sequences. The “Nominate a modelome!” feature allows community users to request modeling of additional complete genomes as our computational resources allow. This feature has been used, for example, to support the Tropical Disease Initiative (<http://tropicaldisease.org>)[294-297]

ModWeb: Comparative modeling web-server

The ModWeb comparative modeling web-server is an integral module of ModBase (<http://salilab.org/modweb>)[275]. In the default mode, ModWeb accepts one or more sequences in the FASTA format, followed by calculating and evaluating their models using ModPipe based on the best available templates from the PDB. Alternatively, ModWeb also accepts a protein structure as input (template-based calculation), calculates a multiple sequence profile and identifies all homologous sequences in the UniProtKB database, followed by modeling these homologs based on the user-provided structure. This alternative protocol is a useful tool for measuring the impact of new structures, such as those generated by structural genomics efforts[298]. Moreover, new members of sequence superfamilies with at least one known structure can be identified[299].

In addition to anonymous access, registered users get unified access to all their ModWeb datasets and can submit template-based calculations.

Associated Resources

A number of web-services are associated with ModBase. Some of these are tightly integrated with ModBase, while others contain data that are derived through ModBase (eg, single nucleotide polymorphism (SNP) annotations created by LS-SNP[300]). We already described the interactions of ModBase with the ModLoop server for loop modeling in protein structures (<http://salilab.org/modloop>)[301], the PIBASE database of protein-protein interaction (<http://salilab.org/pibase>)[302], the DBAli database of structural alignments (<http://salilab.org/dbali>)[303, 304], the LS-SNP database of structural annotations of human non-synonymous single-nucleotide polymorphisms (<http://salilab.org/LS-SNP>)[300, 305, 306], the SALIGN server for multiple sequence and structure alignment (<http://salilab.org/salign>)[284, 307], the ModEval server for predicting the accuracy of protein structure models (<http://salilab.org/modeval>)[284], the PCSS server for predicting which peptides bind to a given protein (<http://salilab.org/pcss>)[284], and the FoXS server for calculating and fitting Small Angle X-ray Scattering profiles (<http://salilab.org/foxs>)[284, 308]. Here, we describe several new servers that interact with ModBase.

AllosMod: a web-server for modeling ligand-induced protein dynamics

Conformational transitions of biomolecules are key to many aspects of biology. These dynamic changes span a broad range of time and size scales, and include protein folding, aggregation, induced fit, and allostery.

The AllosMod web-server (<http://salilab.org/allosmod>) predicts conformational changes that occur in the native ensemble, such as allosteric conformational transitions. The input is one or more macromolecular coordinate files (including DNA, RNA, and sugar molecules) and the corresponding sequence(s). The output is a set of molecular dynamics trajectories based on a

simplified energy landscape. The documentation includes analysis examples to help the user in interpreting the expected output. Carefully designed energy landscapes allow efficient molecular dynamics sampling at constant temperatures, thereby providing ergodic sampling of conformational space. AllosMod energy landscapes are constructed using contacts in crystal structure(s) to define the energetic minima. This model is referred to as a structure-based or Gō model[309-311]. The energy landscapes are sampled using many short, constant temperature molecular dynamics simulations. Sampling occurs quickly, even for large systems with up to 10,000 residues, because the simplified landscapes can be stored in memory. The user can also download Python scripts necessary to run and modify the simulations, which are performed using Modeller[243].

The capabilities of the AllosMod server have been demonstrated in a study of allosteric systems with known effector bound and unbound crystal structures[272, 312]. Effector bound and unbound simulations are performed using a landscape with a single minimum for the interactions in the effector binding site, corresponding to the bound or unbound structure, and dual minima for interactions in the rest of the protein, corresponding to the bound and unbound structures. AllosMod can also be used to predict coupling (i.e., $\Delta\Delta G$) between a mutation site and the effector binding site.

A family of web-servers for computation and application of SAXS profiles

SAXS is a common technique for low-resolution structural characterization of molecules in solution[313, 314]. SAXS experiments determine the scattering intensity of a molecule as a function of spatial frequency, resulting in a SAXS profile that can be easily converted into the approximate distribution of atomic distances in the measured system. The experiments can be performed with the protein sample in solution, and usually take only a few minutes on a well-

equipped synchrotron beamline[314]. Here, we describe new features of the FoXS server for calculating and fitting SAXS profiles, the AllosMod-FoXS server that predicts the structural ensemble that best fits a given SAXS profile, the FoXSDock server that performs protein-protein docking filtered by a SAXS profile, and the SAXS Merge server for merging SAXS profiles measured at different concentrations and exposure times.

FoXS (<http://salilab.org/foxs>) is a rapid and accurate server for calculating a SAXS profile of a given molecular structure[308]. The input is one or more macromolecular coordinate files or PDB codes and an experimental profile. The output is a calculated SAXS profile for each input structure, fitted onto the experimental profile. The method explicitly computes all inter-atomic distances, and models the first solvation layer based on solvent accessibility. FoXS was tested on 11 protein, one DNA, and two RNA structures, revealing superior accuracy and speed versus CRY SOL[315], AquaSAXS[316], the Zernike polynomials-based method[317], and Fast-SAXS-pro[318]. In addition, we demonstrated a significant correlation of the SAXS score with the accuracy of a structural model[319]. We have recently updated the server to an interactive user interface; profiles are displayed via an HTML5 canvas element and structures are shown in a Jmol window (Fig. 6.6-1). If the user uploads multiple structures, the server automatically performs the minimal ensemble computation with MES[320].

AllosMod-FoXS (<http://salilab.org/allosmod-foxs>) is a server that predicts the structural ensemble that best fits a given SAXS profile. The input is one or more macromolecular coordinate files, the corresponding sequence(s), and an “experimental” SAXS profile. The output is the structural ensemble that best fits the input SAXS profile. The server relies on AllosMod conformational sampling[272], FoXS calculations of theoretical SAXS profiles, minimal ensemble computation with MES[320], and the SOAP-PP score. The server was motivated to

describe conformational changes in proteins, such as the allostery, based on both modeling considerations (as represented by AllosMod) and experimental SAXS data (as represented by FoXS).

The AllosMod-FoXS server utilizes various sampling algorithms in AllosMod to generate structures that are directly entered into FoXS. Because FoXS explicitly computes all inter-atomic distances and models the first solvation layer based on solvent accessibility, it can be used to score the similarity of the experimental SAXS profile to the predicted SAXS profiles corresponding to structures from the AllosMod simulations. In addition to the FoXS score, each conformation is assessed for structural quality, using the SOAP-PP score. These two scores are combined to predict structures that collectively best explain the experimental SAXS profile.

FoXSDock (<http://salilab.org/foxsdock>) is a web-server that utilizes SAXS profiles to filter the models produced by protein-protein docking. It accepts as input structures of two docked proteins and an experimental SAXS profile of their complex. The output is a set of docking models and their calculated SAXS profiles fitted onto the experimental profile. While many structures of single protein components are becoming available, structural characterization of their complexes remains challenging. While general, protein-protein docking methods suffer from large errors, due to protein flexibility and inaccurate scoring functions. However, when additional information, such as a SAXS profile, is available, it is possible to significantly increase the accuracy of the computational docking.

FoXSDock combines rigid global docking by PatchDock, filtering of the models based on the SAXS profile, and interface refinement by FireDock[273]. The approach was benchmarked on 176 protein complexes with simulated SAXS profiles, as well as on 7 complexes with experimentally determined SAXS profiles[287]. When induced fit is less than

1.5Å interface Cα RMSD and the fraction of residues missing from the component structures is less than 3%, FoXSDock can find a model close to the native structure within the top 10 predictions in 77% of the cases; in comparison, docking alone succeeds in only 34% of the cases.

SAXS Merge (<http://salilab.org/saxsmerge>) is a web-server that uses automated statistical methods to merge SAXS profiles determined at different concentrations and exposure times. The input file consists only of the buffer-subtracted SAXS profiles in a common three-column text format. The output comprises (i) a list of individual q points with associated source profiles, (ii) an estimate of the mean profile, along with a 95% Bayesian credible interval and (iii) the most suitable parametric mean function for the resulting profile, an estimate of the noise level in the pooled dataset. The output is visualized interactively through the web-browser, and can also be downloaded. High-throughput SAXS data collection requires robust, accurate, and automated tools for data processing and merging[321, 322]. However, SAXS data are generally processed highly subjectively, often manually with the aid of the PRIMUS software package[323]. The operation requires an experienced user who can manually inspect each profile to be merged and decide where the SAXS profiles agree or not.

The SAXS Merge web-server alleviates user intervention through an automated and statistically principled merging procedure based on a Bayesian approach (Spill et al, submitted). The SAXS Merge web-server was successfully validated on a benchmark of 16 SAXS data sets.

Pose & Rank: a web-server for scoring protein-ligand complexes

Molecular recognition between proteins and ligands plays an important role in many biological processes. Predicting the structures of protein-ligand complexes and finding ligands by virtual screening of small molecule databases are two long-standing goals in molecular

biophysics and medicinal chemistry. Solving both problems requires the development of an accurate and efficient scoring function to assess protein-ligand interactions.

The Pose & Rank web-server (<http://salilab.org/poseandrank>)[274] provides access to two atomic distance-dependent statistical scoring functions based on probability theory that can be used in protein-ligand docking: The PoseScore was optimized for recognizing native binding geometries of ligands from other poses, and the RankScore was optimized for distinguishing ligands from non-binding molecules. The server accepts as input a coordinate file of the target protein structure in the PDB format and docking poses of small molecules. The output is a list of scores for each protein-small molecule complex. PoseScore ranks a near-native binding pose the best, top 5, and top 10 for 88%, 97%, and 99% of targets, respectively. RankScore improves the overall ligand enrichment (logAUC) and early enrichment (EF1) scores computed by DOCK 3.6[324] for 68% and 74% of targets, respectively. The Pose & Rank resource can contribute to many applications, such as selecting ligand candidates from virtual screening for experimental testing, predicting the binding geometries for known ligands, and suggesting binding site mutations that alter the ligand binding properties and consequently protein functions.

APPLICATION EXAMPLES

Coordinating the impact of structural genomics on the human α -helical transmembrane proteome

With the recent successes in determining membrane protein structures, we explored the tractability of determining representatives for the entire human transmembrane proteome[325]. This proteome contains 2,925 unique integral α -helical transmembrane domain sequences that cluster into 1,201 families sharing more than 25% sequence identity. We assessed the modeling

coverage by processing all sequences through ModPipe, and analyzing the resulting ModBase dataset. We then clustered all sequences (BlastClust[326]), annotated them with cluster size, modeling coverage, and number of predicted transmembrane helices. Finally, we explored several target selection strategies. Structures of 100 optimally selected targets would increase the fraction of modelable human α -helical transmembrane domains from 26% to 58%, thus providing structure/function information not otherwise available.

To leverage the results of this study, the PSI:Biological Network (http://www.nigms.nih.gov/Research/FeaturedPrograms/PSI/psi_biology/psibiology_networkorg.htm), including high-throughput and membrane PSI centers as well as the Structural Genomics Consortium, is attempting to express nearly 100 human transmembrane proteins using their standard high-throughput methods. The goal of this survey is to determine which methods best express certain classes of transmembrane proteins. The sequences of our previous analysis were further annotated by fraction of predicted disordered regions[327, 328], number of glycosylation sites[260, 329, 330], clone availability[331-333], HUGO annotations[334], sequence length, and several additional metrics. 86 targets were hand-picked from the largest clusters to represent a diverse selection of human membrane proteins with maximum coverage of the transmembrane proteome. Cloning, expression, and solubility experiments of these targets using the pipelines of the ten participating research groups are currently in progress. Participants also use shared and individual sets of six controls. A standard method will be used by all to visualize the protein bands to quantify yield. A final full comparison will determine the most successful methods for each representative transmembrane protein. Progress of the survey is cataloged by the portal of the Protein Structure Initiative Structural Biology Knowledgebase (PSI SBKB[335];

<http://hmpps.sbkb.org/>) and will be accessible to the public after the conclusion of the experiment. A final publication will summarize the survey's findings.

Structural determinants of Human Immunodeficiency Virus-1 (HIV-1) protease

The maturation of the HIV virion is facilitated by the cleavage of the Gag and Pol polyproteins[19]. A homodimeric aspartic protease (HIV-1 protease) catalyzes these processing events at ten non-homologous sites and is the target of some of the most effective antiretroviral drugs[24-26]. These sites are eight amino acid residues in length; the cleavage occurs between the third and fourth residues[20-23]. In addition to processing viral proteins, HIV-1 protease cleaves several human proteins during infection, such as the eukaryotic translation initiation factor 3 subunit D (eIF3D)[27-30].

To predict cleavage sites in human proteins, we began by examining sequence and structural features of over 120 cellular substrates of HIV-1 protease that were recently identified *in vitro*[31] (for an example, see Fig. 6-2). First, every residue of the cleaved and non-cleaved octapeptides was encoded using over 512 physicochemical amino acid indices[336, 337]. To account for cooperativity between residues in different positions of the octapeptide, frequencies of dipeptides and gapped dipeptides (i.e. two specific residues separated by any residue) were also used to train machine-learning algorithms for binary classification. Second, a greedy feature selection procedure was applied to determine features of octapeptides important for protease activity. Interestingly, while features encoding known viral cleavage motif ELLE were important for classification, most discriminating features encode structural preferences of amino acid residues in the second and fifth positions of the octapeptide. Therefore, we created a ModBase dataset of 405 models for 118 human proteins cleaved *in vitro*. PSI-Pred[338] was used to predict secondary structure elements for protein regions without templates. Analysis of the

structural models showed the enrichment of alpha+beta protein class (SCOP ID=53931) among cleaved proteins and coiled secondary structure (~41%) among cleaved sites. We added structure-based descriptors of cleaved and non-cleaved sites to the sequence-based features and assessed classifiers' performance in a five-fold cross-validation procedure. The average area under the receiver operating characteristic (ROC) curve for the classifier trained with the Random Forest algorithm[339] was 0.965 (72% sensitivity and 98% specificity) and the entire human proteome was scanned for putative human substrates of the HIV-1 protease. We are currently experimentally validating several of the predicted cleavage sites.

ACCESS AND INTERFACE

Direct access

The main access to ModBase is through its web interface at <http://salilab.org/modbase>, by querying with UniprotKB[260, 261] and GI[340] identifiers, gene names, annotation keywords, PDB[259] codes, dataset names, organism names, sequence similarity to the modeled sequences (BLAST[276]), and model-specific criteria such as model reliability, model size, and target-template sequence identity. Additionally, it is possible to retrieve coordinate files and alignment files of all models for a specific sequence as text files. Meta-data for all current ModBase models (updated weekly), all genome datasets, and several additional project specific datasets, are also available from our FTP server (<ftp://salilab.org/databases/modbase/projects>).

The output of a search is displayed on pages with varying amounts of information about the modeled sequences, template structures, alignments, and functional annotations. Output examples from a search resulting in one model are shown in Fig. 6-3. A ribbon diagram of the model with the highest target-template sequence identity is displayed by default, together with

some details of the modeling calculation. Ribbon thumbprints of additional models for this sequence link to corresponding pages with more information. Ribbon diagrams are generated on the fly using Molscript[341] and Raster3D[342]. A pull-down menu provides links to additional functionalities: the SNP module; retrieval of coordinate and alignment files; molecular visualization by UCSF Chimera that allows the user to display template and model coordinates together with their alignment; and Chimera visualization of predicted cavities (ConCavity[343]). If mutation information is available for a protein sequence, links to the details are provided in the cross-references section. Additionally, cross-references to various other databases, including PDB[344], UniProtKB[345], the UCSC Genome Browser[346], EBI's InterPro[347], PharmGKB[348], and SFLD[293] are given. Other ModBase pages provide overviews of more than one sequence or structure. All ModBase pages are interconnected to facilitate easy navigation between different views.

The Protein Model Portal

The Protein Model Portal (PMP) has become a valuable option for accessing ModBase models (<http://proteinmodelportal.org>)[349]. The PMP is a single point of entry for accessing protein structure models from a number of different databases. PMP queries all participating source model databases and serves the user with the model coordinates, alignments, and quality criteria from a central location. It has been developed as a module of the Protein Structure Initiative Knowledgebase (PSI KB)[335, 350]. The PMP provides a flexible search interface for all deposited models, quality estimation, cross-links to other sequence and structure databases, annotations of sequences and their models, a central point of entry to comparative modeling servers (including ModWeb) and quality estimation servers (including ModEval), and detailed

tutorials on all aspects of comparative modeling. Currently the PMP retrieves approximately 450,000 ModBase model coordinate files each week from ModBase.

A sister web-service to PMP, CAMEO (<http://cameo3d.org>)[349] continuously evaluates the accuracy and reliability of several comparative protein structure prediction servers in a fully automated manner. The ModWeb server currently participates in the testing mode, and is expected to move into the production mode in the first quarter of 2014.

Access through external databases

ModBase models in academic and public datasets are also directly accessible from several databases, including the Protein Model Portal[349], UniProtKB[351], PIR's iProClass[345], EBI's InterPro[347], the UCSC Genome Browser[346], PubMed (LinkOut)[352], PharmGKB[348], and SFLD[293].

Future Directions

ModBase will grow by adding models calculated on demand by external users (using ModWeb) as well as our own calculations of model datasets that are needed for our research projects (using ModPipe, ModWeb, or Modeller). These updates will reflect improvements in the methods and software used for calculating the models as well as new template structures in the PDB and new sequences in UniProtKB. In the future, we expect that most of the users will access ModBase models through the Protein Model Portal.

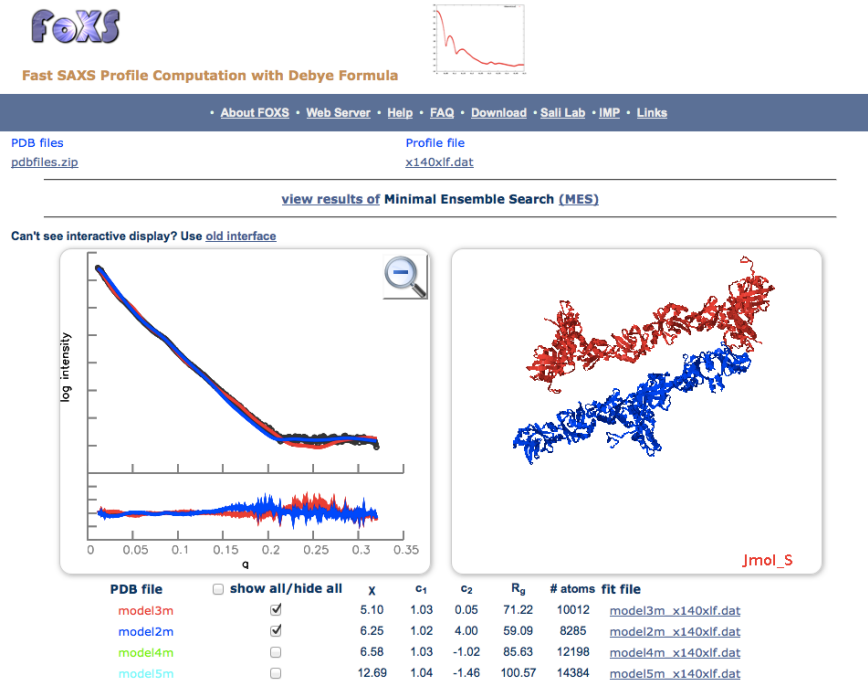


Fig. 6-1: The computed profiles for filament models of the XLF-XRCC4 complex[353] are fitted to the experimental SAXS profile with FoXS.

The interactive user interface displays the profiles in the left panel and the models in the right panel using the same color for each model/profile pair. The table below the panels displays the fit parameters and includes buttons to simultaneously show or hide each model/profile pair. Clicking on Minimal Ensemble Search (MES) results (above the display panel) takes the user to the MES output page.

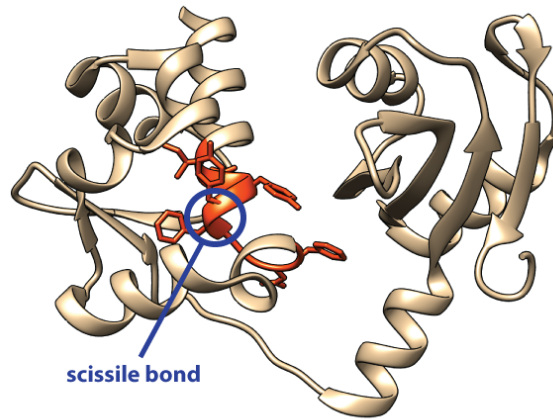


Fig. 6-2. Cleavage of human proteins by the HIV-1 protease: Crystal structure of the N-terminal domain of human Lupus La protein [354].

Residues of the cleavage site (Ile-Asp-Tyr-Tyr-Phe-Gly-Glu-Phe) are shown in orange. Scissile bond between Tyr and Phe in the α -helix is cleaved by the HIV-1 protease in vitro.

Search Form

Model Details - Sketch

Update and Remodel

Chimera Visualization

Model Details Options

Quality Criteria

Model Overview

Chimera Cavity View

Fig. 6-3. ModBase Interface Elements: Search Form: search options are available through the pull-down menu.

A quick overview of the available representations is displayed below the search form. Model Details Sketch: The Model details page provides information for all models of a given sequences. The sketch is comprised of two parts: The model coverage sketch, that indicates the sequence coverage by all models (top line) and the sequence coverage by the current model (second line), and a ribbon diagram of the current model. Other models are available via

thumbprints. Update and Remodel: This box show the date of the last modeling calculation for the current sequence, and allows the user to request an update. Chimera Visualization: The visualization includes the model and template structures, and the alignment. Cross-references: Links to the PMP, UniProtKB, Genbank, UCSC Genome Browser, and other databases. Model Details Options: The pull-down menu switches between representations and allows downloads of coordinate and alignment files. Quality Criteria: red indicates unreliable, green reliable. Model Overview: A different representation for several sequences gives a quick overview on modeling coverage and quality. Chimera Cavity View: visualizes cavities predicted by ConCavity.

Chapter 7: Hemoglobin S Polymerization Studied by Rapid Brownian Dynamics Simulations

INTRODUCTION

The phase or configuration space of physical systems can be explored with simulations of their dynamics (REFS). The resulting samples of states can be used to calculate expectation values of various observables. The resolution of the underlying system representation guides the selection of a simulation method used to propagate the modeled system in space and time (e.g., quantum mechanics methods, atomistic molecular dynamics, and various simplified simulation methods, such as Brownian dynamics). In molecular dynamics simulations, physicochemical properties of molecules and information from experiments determine how intermolecular interactions are modeled; these include interactions between solvated molecules, solvent- and membrane-molecule interactions. Brownian dynamics (BD) [32] is a simulation method that, in addition to modeling intermolecular interactions between solute molecules, accounts for interactions between solute and solvent described by their average action. Individual interactions between solute and solvent in BD simulations are reduced to statistical fluctuations induced by thermal noise, allowing for separation of time scales. Thus, solute-solvent interactions in BD simulations are modeled by their collective action using diffusion tensors derived from statistical and transport properties of the solvent.

Considering that most cellular processes involve diffusion-mediated steps, there is a growing interest in simulating molecular systems composed of a few to thousands species with the highest possible structural resolution and long time scales. BD simulations have been applied to model different physical and biological systems [355-369] and several computer programs are

available for running these simulations [370-373]. However, simulating biological systems at atomic resolution remains difficult, due to significant computational overhead. Thus, acceleration of molecular simulations has been attempted with supercomputers [374], computer clusters [375], graphical processing units (GPUs) [376], and custom hardware [377],[378],379].

In BD simulations, the most time-consuming step is computation of intermolecular forces and torques acting on solute molecules for all but the simplest of intermolecular potentials, making physiologically relevant time scales inaccessible. For example, give what time scales some seminal BD simulations achieved (Elcock, Skolnick, Jacobson). It has been previously shown that pre-computed inter-particle potentials and forces can accelerate molecular dynamics (MD) simulations by as much as a factor of four for Lennard-Jones potential functions [380], albeit at a loss of accuracy in the resulting trajectories and increased use of computer memory. In this work, we use pre-computed pairwise forces and torques to allow us to simulate biological processes involving high concentrations of molecules of a single or few types. We illustrate the method by simulating the polymerization of sickle cell hemoglobin molecules represented with rigid bodies at physiologically relevant concentrations and time scales.

Adult wild-type human hemoglobin (HbA) is a globular protein consisting of four chains, including two alpha and two beta chains [381]. A single amino acid residue substitution of a hydrophobic valine for a charged glutamic acid in the sixth position on the beta chain is associated in homozygous form with sickle cell disease [33, 34]. In comparison to HbA, sickle cell hemoglobin (HbS) variant has two altered properties: low oxygen affinity and propensity to polymerize in deoxygenated form at high concentration (Fig. 7-1) [382]. In vivo, the HbS polymers associate laterally into fibers and bundles of fibers distorting the shape of the red blood cells [36, 382-386]. As a result, distorted cells increase blood's viscosity and hinder its flow

through the capillaries. Delayed passage through the capillaries increases the deoxygenation and produces an even greater number of rigid cells, which can completely block blood vessels during a sickle-cell crisis. While the damage to the membrane of the red blood cells by long hemoglobin fibers causes premature cell death and anemia, decreased oxygen delivery to tissues is compensated by an enhanced oxygen release due to low oxygen affinity of HbS [33, 34, 382, 387]. Thus, it is not anemia, but the impedance to blood flow that is the most important factor in the sickle cell disease. The inhibition or delay of polymerization could, in principle, alleviate patient's symptoms. Therefore, rational design of inhibitors of hemoglobin sickling necessitates not only the knowledge of the structure of the HbS polymer but also the dynamical properties of HbS polymerization.

Next, we describe the simulation protocol, including preparation of macromolecular structures, tabulation of forces, and propagation of the system with our BD algorithm. Next, we present the methods for analyzing the output of simulations and conclude by discussing future directions.

METHODS AND RESULTS

Preparation of the Macromolecular Structures for Simulation

The X-ray structure of the deoxy-hemoglobin S tetramer (chains A-D) at 2.05 Å resolution was retrieved from the Protein Data Bank (PDB code 2HBS) and prepared by following a previously published protocol [370]. In summary, the PDB2PQR program [388] was used to add hydrogen atoms and compute partial radii and charges with AMBER 94 force field [389]. Electrostatic potential grid was computed with the APBS software [388] by solving the non-linear Poisson-Boltzmann (PB) equation. In these PB calculations, we used previously

described parameters [364]: a $200 \times 200 \times 200 \text{ \AA}^3$ grid with a 1 \AA grid spacing and default solvent dielectric constant of 78.5, protein dielectric constant of 2.0, and ionic strength of 150 mM. The electrostatic desolvation, nonpolar desolvation, and soft-core repulsion grids were set to $120 \times 120 \times 120 \text{ \AA}^3$, $100 \times 100 \times 100 \text{ \AA}^3$, and $100 \times 100 \times 100 \text{ \AA}^3$, respectively, and the grid spacing of 1 \AA ; they were computed using the SDA program [370]. Effective charges for the hemoglobin S tetramer were computed using the SDA software [370].

Tabulation of Protein-Protein Interaction Energies and Forces

Pairwise protein-protein interaction energies, forces, and torques were computed by exhaustively enumerating relative orientations and rotations of two HbS tetramers. This enumeration was performed on a grid by rotating and translating one of the tetramers around the second stationary tetramer (Fig. 7-2). The size of the grid was set to $140 \times 140 \times 140 \text{ \AA}^3$ and the grid spacing to 1 \AA . At each grid point, the three Euler angles were uniformly sampled [390] with 60 relative orientations, thus resulting in approximately 1.71×10^8 grid points. The grid size was chosen to ensure that all energies and forces at the grid boundary evaluate to zero. Translational and rotational grid cell size as well as the simulation time step were parameterized by ensuring that the effective translation of HbS tetramers [391] is reproduced in 100 independent $1 \mu\text{s}$ simulations of the equilibrated system; a simulation reproduces the experiment if the computed diffusion coefficient is within 10% error. We used the energy function, implemented in the SDA software, consisting of electrostatic interaction energy, electrostatic, and non-polar desolvation energies as well as soft sphere repulsion terms[370]:

$$\begin{aligned}
\Delta G^{1-2} = & \frac{1}{2} \sum_{i_2} \Phi_{el_1}(r_{i_2}) \times q_{i_2} + \frac{1}{2} \sum_{j_1} \Phi_{el_2}(r_{j_1}) \times q_{j_1} \\
& + \sum_{i_2} \Phi_{edesolv_1}(r_{i_2}) \times q_{i_2}^2 + \sum_{j_1} \Phi_{edesolv_2}(r_{j_1}) \times q_{j_1}^2 \\
& + \sum_{m_2} \Phi_{npdesolv_1}(r_{m_2}) \times SASA_{m_2} + \sum_{n_1} \Phi_{npdesolv_2}(r_{n_1}) \times SASA_{n_1} \\
& + \sum_{m_2} E_{softcore_1}(r_{m_2}) + \sum_{n_1} E_{softcore_2}(r_{n_1})
\end{aligned}$$

The SDA software was modified to output interaction energies, forces, and torques computed for each sampled orientation and rotation to a file on a hard disk. The forces were computed as finite-difference derivatives of the pairwise free energies of interaction between proteins [370]. Only forces, torques and energies that are significantly different from zero and infinity are stored, requiring approximately 16GB of memory for each type of a pairwise protein-protein interaction.

Brownian Dynamics Simulations of Polymerization

We first evaluated different BD propagation schemes (Fig. 7-3) to select the most efficient algorithm as measured by the computer time for simulating a system of 512 spheres with radii of 30 Å for 1000 time steps (time step = 123 ps). Ermak-McCammon propagation scheme is equivalent to the first-order Euler algorithm with the Brownian drift term [32] and the Iniesta and de la Torre algorithm is a two step predictor-corrector algorithm [392]. The second order algorithm is considered more accurate but less efficient. We also evaluated two different schemes for storing neighbor lists, the Verlet list[393] and kd-trees[394], and two types of simulation time step, constant and adaptive [395]. Our results indicate that the most efficient BD scheme is based on the Ermak-McCammon algorithm with a constant time step and a neighbor list stored in a kd-tree data structure with the cutoff distance of 120 Å for the separation between

two centers of mass of interacting molecules (Fig. 7-3). The cutoff was determined by the size of the pre-computed grid; all intermolecular forces and energies beyond this cutoff are set to zero.

We implemented the optimal protocol in our open-source Integrative Modeling Platform (IMP) software (<http://salilab.org/imp>) [35]. Each hemoglobin S tetramer diffused as a rigid body represented at atomic resolution. We used the experimentally determined translational diffusion coefficient of 3.4×10^{-8} cm²/sec [391]. Infinite rotational diffusion coefficients were computed using the Einstein–Smoluchowski equations [396] for a sphere that bounds hemoglobin S tetramer atoms (radius of 30 Å) and the solvent viscosity of 0.89cP (corresponding to the viscosity of pure water at 25 C) and scaled by a factor of 100.

At each time step of the simulation, relative pair-wise distances and orientations for the neighboring rigid bodies were determined and used as a lookup key for the pre-computed table of pair-wise forces and torques. Forces and torques for the exact distances and orientations were computed by the nearest neighbor interpolation between the pre-computed values. These values were then added to the total force and torque for each rigid body, followed by their application to the centers of mass of rigid bodies to propagate the molecules according to the BD algorithm.

Each BD simulation was initiated from an ensemble of non-overlapping randomly positioned and oriented molecules in a simulation box. These initial conditions correspond to initiating the polymerization by a change from an infinitely dilute solution to the volume fraction determined by the number of molecules in the volume of the simulation box. Simulations were performed at two volume fractions (densities) of 0.25 and 0.33 by varying the size of the simulation box under periodic boundary conditions. The volume fraction of 0.25 is physiologically relevant [391]. For each volume fraction, 100 simulations with different initial configurations were performed. The simulations were carried out at 300 K. Trajectories of

molecules were recorded every 1 ns. The time step (Δt) was set to 123 ps. The time step in a BD simulation is chosen according to two criteria. It should be small enough to ensure that the forces and gradients of the diffusion tensor are constant and large enough to ensure that solute-solvent interactions can be averaged in the diffusion tensor. Practically, time step is selected such that Δt is larger than $m/6\pi\eta a$, where η is the solvent viscosity and m and a are the mass and radius of the solute molecule, respectively [397].

Analysis of time courses

To monitor polymerization as a function of time, we compute the following three order parameters: 1) degree of polymerization as the number of hemoglobin tetramers in aggregated form, 2) total number of contact residues in polymerized species, and 3) the size of the largest polymer. We define a polymer as a cluster of HbS tetramers in contact with each other and the contact as the distance of at most 6 Å between any two non-hydrogen atoms of the neighboring tetramers. Thus, the number of individual tetramers in a cluster determines the size of the polymer. Our results show that all HbS tetramers polymerize within 10 μs , forming over 1800 total contacts (Fig. 7-4). The average size of the largest oligomer observed in 100 independent simulations is 18 (Fig. 7-4).

We computed probabilities of polymers of various sizes at different volume fractions (Fig. 7-5) in 100 independent simulations of 10 μs . While monomers were predominant species in physiologically relevant concentrations (density = 0.25), the distribution of polymer sizes changed at higher concentrations; we observed similar densities of polymers of sizes 1 (monomers), 2 (dimers), 3, and 4.

Analysis of the HbS Dimers

Because the dimers of HbS tetramers are the major building block of the HbS fiber[383-386], we performed several analyses of dimer interfaces formed in our BD simulations.

First, we used the self-organizing map (SOM) algorithm[398] to cluster pairwise configurations from the BD simulations. Pairwise conformations were encoded by the 7-dimensional vectors of the orientations of one molecule relative to the second molecule and used to train the SOM using the *kohonen* library in R [255]. We trained a periodic Euclidean SOM containing 10 x 10 neurons as follows. First, SOM was initialized with a random uniform distribution covering the range of values of the input vectors of size $7N$, where N is the number of dimers. Second, model parameters were tuned by exponentially decreasing learning parameter from 0.25 to 0 and Gaussian radius from 0.01 to 0 during 100 iterations of random presentation of input vectors to the SOM training algorithm. Finally, the final SOM was used to compute several quantities, such as the number of conformations within each cluster, the mean distance between clusters, and the unified distance matrices, defined as the sum of the distances to all immediate neighboring clusters.

Several observables, such as dimers' interaction energy and frequencies of dimer configurations were projected onto the map (Fig. 7-6). We identified several highly populated clusters of HbS dimers. Several of these clusters contained dimer interfaces found in the crystals used for structure determination of Hb by X-ray crystallography. Gratifyingly, configurations with the known interface between Val6 in the beta chain of one tetramer and Leu88 in the beta chain of the other tetramer were among densely populated clusters. Additionally, the mean pairwise interaction energy of these configurations was more favorable than the mean energies of configurations in the other clusters. We also, identified several previously undescribed interfaces

between HbS tetramers, such as, for example, the interaction between the mutated valine and Leu83 in the alpha chain of the other molecule. Closer examination of residues surrounding Leu88 in the alpha chain revealed an environment similar to the Leu88 neighborhood in the beta chain. It is conceivable that this interaction also occurs *in vivo*. This possibility is not surprising as the two chains are quite similar in their primary sequence and structure.

We collected statistics about all pairwise center-of-mass (COM) distances in the simulation (Fig. 7-7). While the largest density of distances was observed at a distance of approximately 60 Å, two additional peaks in the density plot were seen. The peak at a distance of approximately 90 Å corresponds to several clusters (polymers) of HbS molecules. The second peak at approximately 57 Å corresponds to the distance of the known lateral dimer contact between Val6 of the beta chain of one molecule and Leu88 of the beta chain of another [399].

In addition to the simulations of deoxygenated HbS, we performed similar simulations of the wild type HbA in the deoxygenated form and the sickle cell mutant in its oxygenated form, followed by examining the differences in contacts made at dimer interfaces. Remarkably, the distribution of the number of oxy-HbS tetramers in contact with other oxy-HbS tetramers was flatter than the distribution for the deoxy-HbS contacts. The same outcome was observed for deoxy-HbA. In contrast, in deoxy-HbS simulations, the contacts between Val6 in the beta chain and Leu88 in the beta chain dominate and these contacts comprise 40% of all contacts involving Val6 residue.

We also examined the residue in the alpha chain contributing to the highest peak in Fig. 7-8. Approximately 17% of all contacts involving Val6 in the beta chain were with the alpha chain of the other tetramer. We identified Asp47 as the major residue contributing to this interaction; α Asp47 is a solvent exposed loop residue located near the alpha-beta interface (Fig.

7-9). The alpha-beta interface is involved in the transition between oxy and deoxy states. The importance of this finding is underscored by the report that mutation of Asp47 for histidine has a destabilizing effect on polymerization *in vitro* [400]. In the wild-type HbA, there is a negatively charged residue in place of Val6, which is predicted to have a repulsive effect, preventing the formation of this interface. The observed destabilization of polymerization for the histidine mutant is most likely, due to the shift to the higher affinity oxygenated conformation of the hemoglobin tetramer. Because only deoxygenated HbS polymerizes, lowering its concentration would decrease polymerization [382, 387]. While this finding did not provide us with a dimer interface that remains unaltered in the oxy and deoxy state, interface between Asp47 and Val6 could be explored for the discovery of small molecules capable of binding to the area surrounding Asp47 in the alpha chain, altering HbS conformation and concentration of HbS in higher affinity state, and consequently, slowing the polymerization process.

SUMMARY AND FUTURE DIRECTIONS

Pathways of assembling macromolecular complexes exhibit multiple space and time dependencies, inaccessible to most molecular simulations. We developed a variant of the BD simulations that relies on pre-computed forces between rigid macromolecules of a small number of different types, allowing us to study systems of thousands of macromolecules on the time scales of seconds. We applied the method to hemoglobin S polymerization, recapitulating known axial and lateral protein-protein interaction interfaces as well as predicting several previously undescribed interactions.

In the future, we will calculate the density of states, i.e., the number of all possible configurations for an energy level of the system, by a modified Wang-Landau (WL)

algorithm[402]. This mapping will allow us to compute thermodynamic observables of the system.

FIGURES

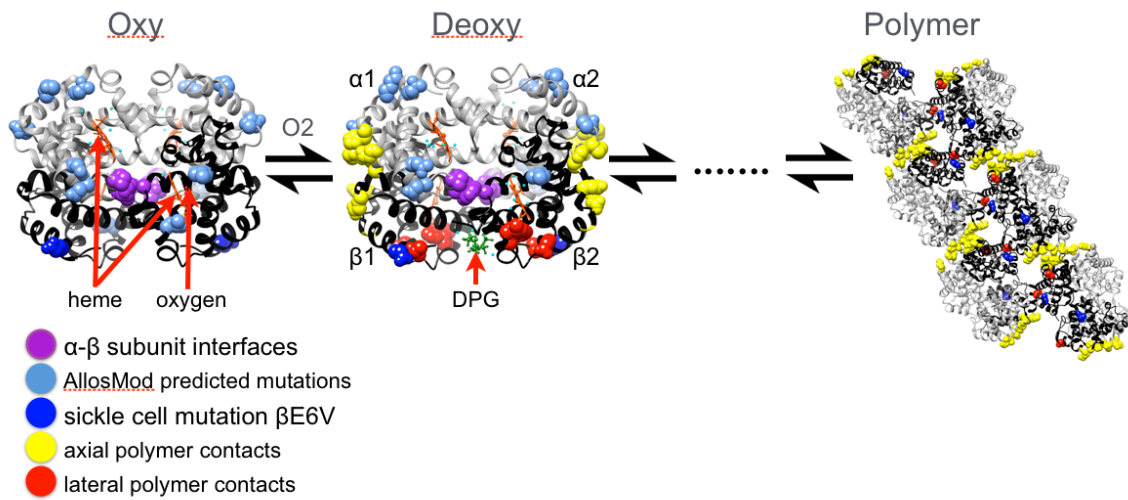


Fig. 7-1. Sickle cell polymerization process.

Oxygen unbinding changes conformation of the hemoglobin tetramer, resulting in reduced solubility. HbS molecules polymerize into long double-stranded fibers. Alpha subunits are colored in grey, beta subunits in black. Heme, oxygen, and DPG binding pockets are indicated with red arrows. Residues at the alpha-beta interface are shown as purple spheres and mutations predicted with AllosMod software[401] are shown as cornflower blue spheres. Residues at position 6 in the beta chain are shown as blue spheres and residues interacting with them are shown as red spheres. Axial polymer contacts are shown in yellow.

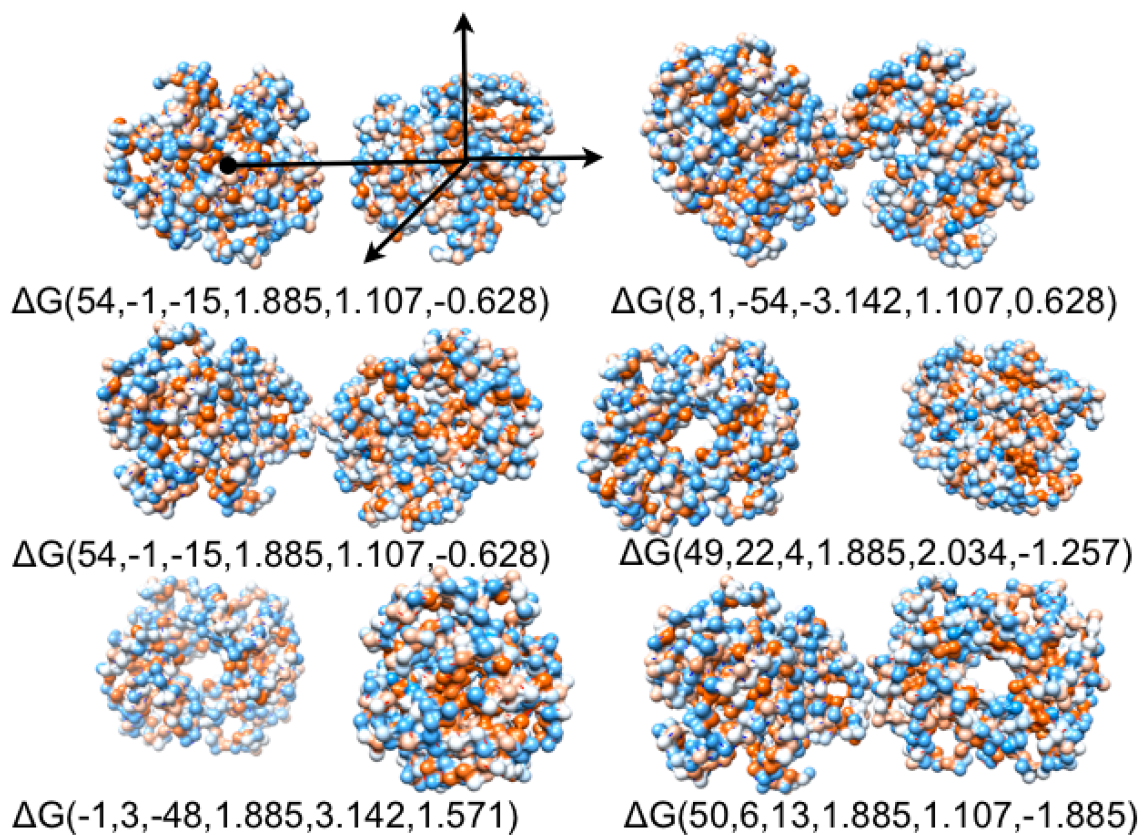


Fig. 7-2. Pre-computing pairwise forces and torques.

Pairwise interaction energies are computed by uniform sampling of relative orientations of HbS tetramers. Forces and torques are then computed as described in the text.

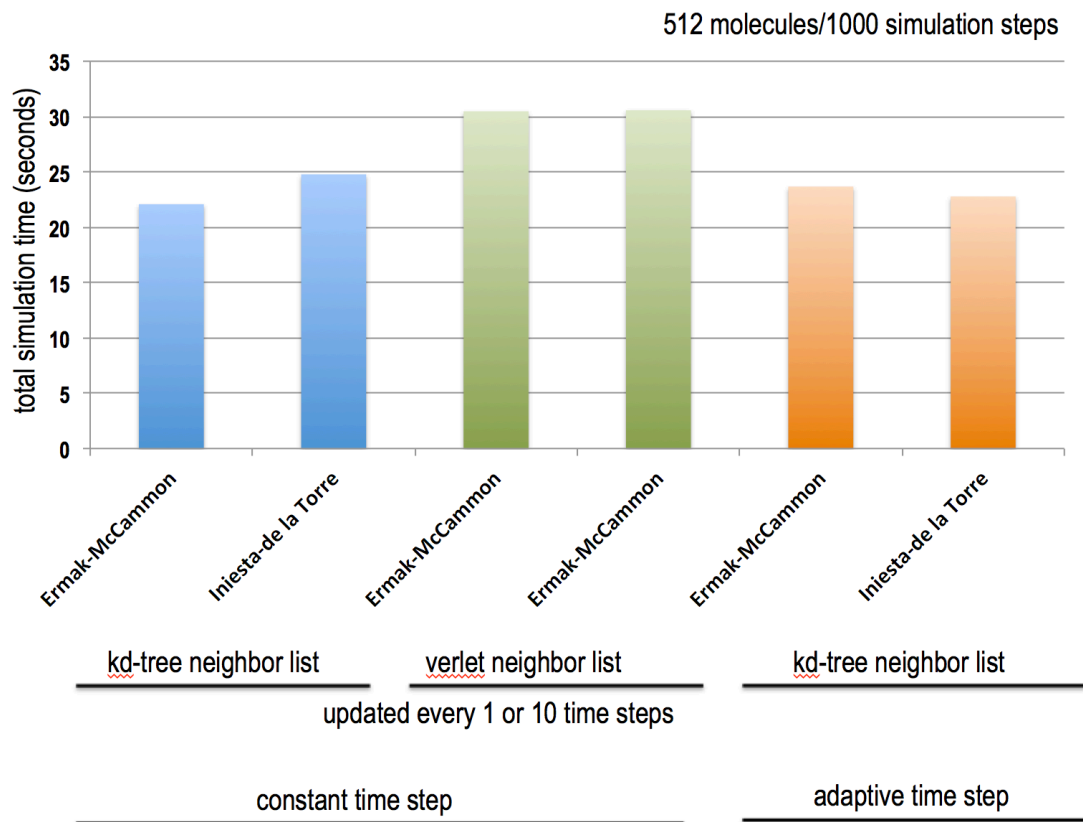


Fig. 7-3. Benchmarking time step and neighbor list algorithm.

Short BD simulations of 1000 time steps were run using different BD propagation algorithms, a neighbor list data structure, and time step selection. Total simulation time in seconds for a system of 512 hemoglobin S tetramers at volume fraction = 0.17 is shown on the vertical axis.

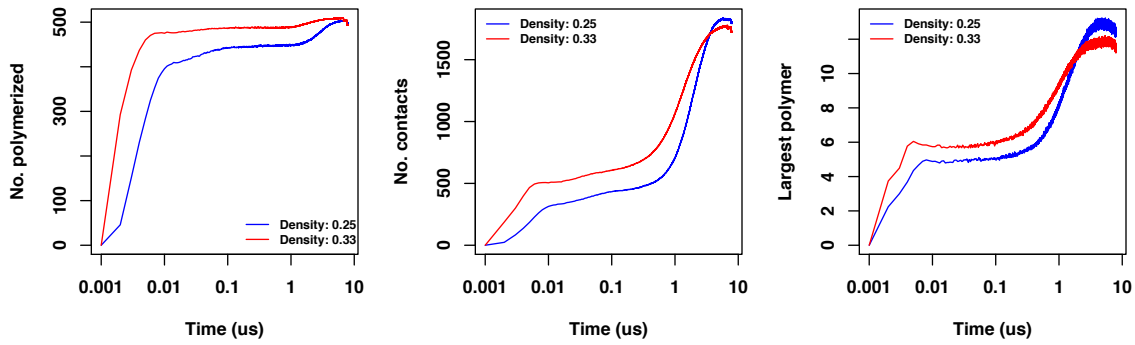


Fig. 7-4. Results of the time-course analyses for 100 independent simulations at two different volume fractions (0.25 and 0.33).

HbS polymerization is plotted as a function of time. Total number of polymerized HbS tetramers, total number of contacts between HbS tetramers, and the size of the largest polymer are shown on the vertical axis. Time (log scale) is shown on the horizontal axis.

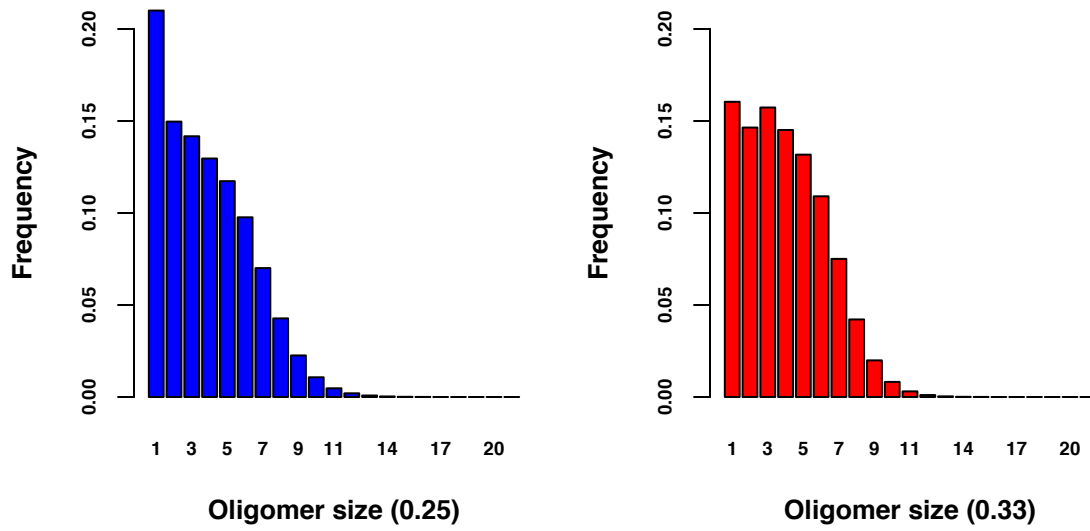


Fig. 7-5. Frequency of HbS polymers in 100 independent simulations at two volume fractions (0.25 and 0.33).

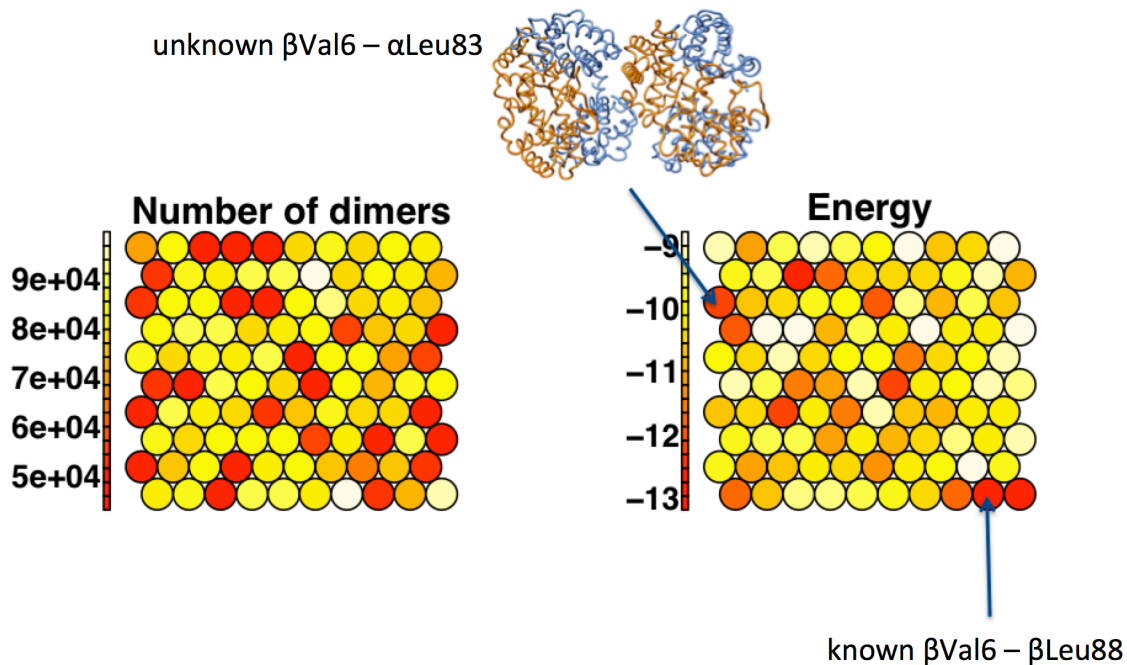


Fig. 7-6. Analysis of HbS dimers using self-organizing maps (SOMs).

Relative pairwise conformations of two HbS in a dimer are clustered using SOMs. Each circle denotes a cluster of dimer conformations. On the left, the color of the circle denotes the total number of dimer conformations within corresponding cluster. On the right, the color of the circle represents the mean pairwise energy of HbS pairs within the cluster.

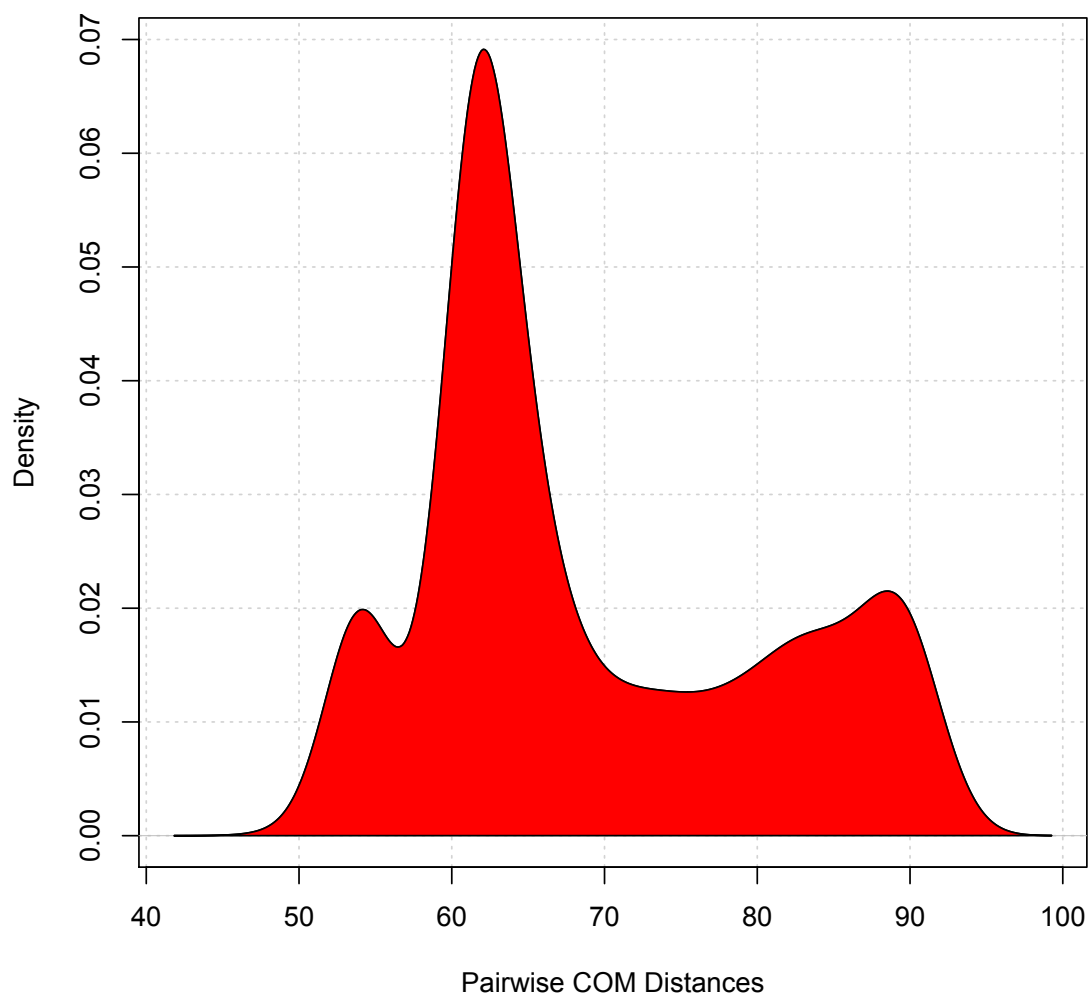


Fig. 7-7. Density of all pairwise center-of-mass distances (\AA) in 100 independent simulations at volume fraction of 0.25.

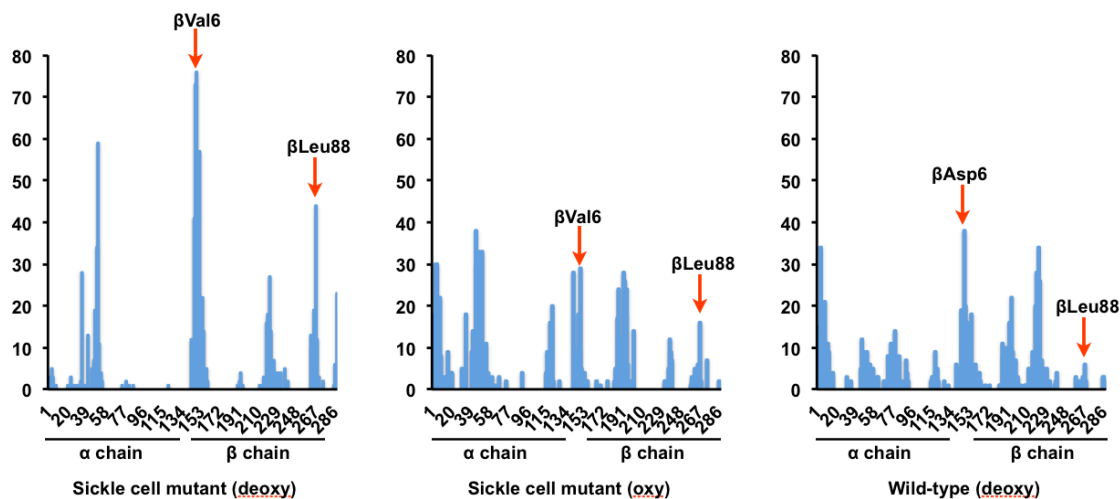


Fig. 7-8. Contact frequencies of residues in hemoglobin dimers.

Residue numbers of alpha chain (1-140) and beta chain (141-286) are shown on the horizontal axis. The number of times each residue was encountered at the dimer interface is shown on the vertical axis. The frequencies are mean values of 100 independent simulation runs at volume fraction = 0.25.

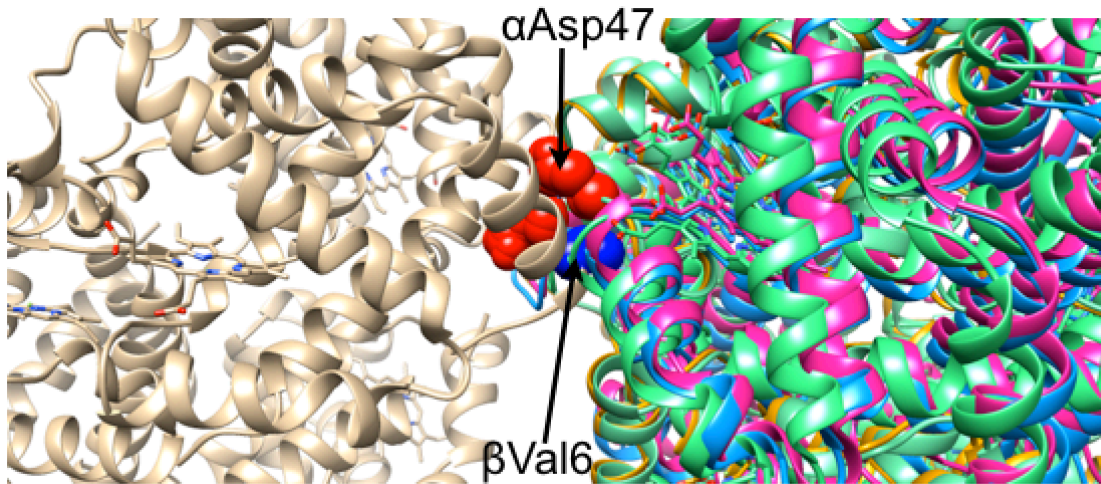


Fig. 7-9. Second most frequent interface involving the sickle cell mutation.

For clarity, only one 3D structure is shown on the left. Several arrangements of interacting HbS tetramers involving this contact are possible. Valine 6 in the beta chain is shown as blue spheres and Asp47 in the alpha chain is shown as red spheres.

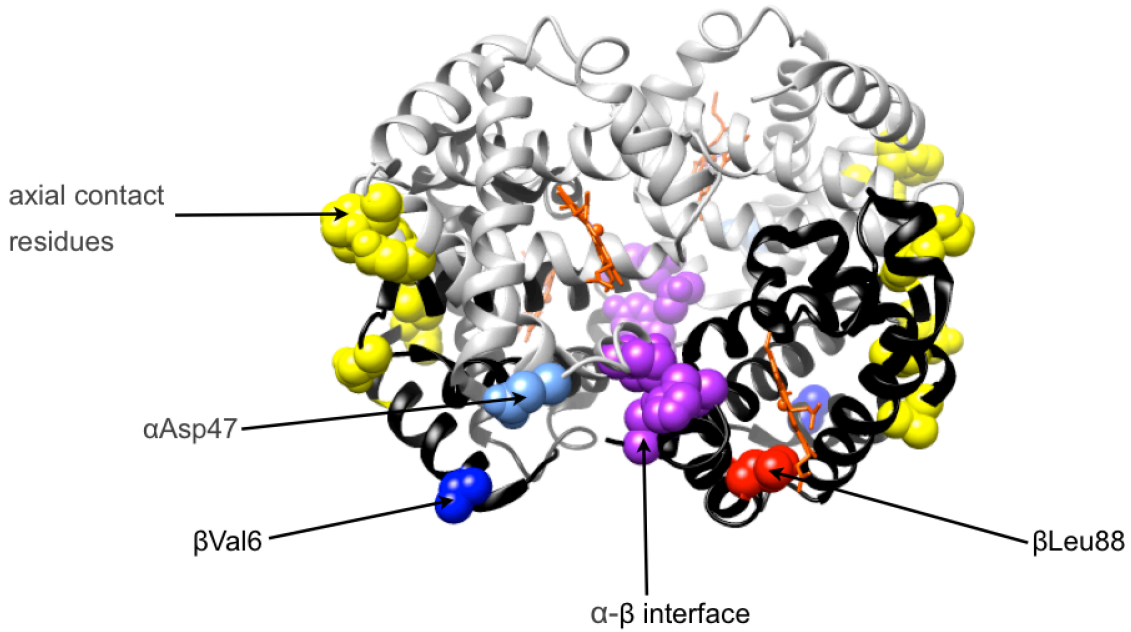


Fig. 7-10. Detailed representation of Asp47 in the structure of deoxy-HbS.

Alpha chains of HbS are shown in grey, beta chains in black. The residue involved in axial contacts in hemoglobin polymers are shown as yellow spheres; mutated valine 6 in blue, Leu88 in red. Residues at the alpha-beta interface of HbS tetramer are shown as purple spheres and Asp47 in cornflower blue.

REFERENCES

1. Giacomini, K.M. and S.M. Huang, *Transporters in drug development and clinical pharmacology*. Clin Pharmacol Ther, 2013. **94**(1): p. 3-9.
2. Lemstrova, R., et al., *Role of solute carrier transporters in pancreatic cancer: a review*. Pharmacogenomics, 2014. **15**(8): p. 1133-45.
3. Nagamori, S. and Y. Kanai, [*Amino acid transporters in cancer*]. Seikagaku, 2014. **86**(3): p. 338-44.
4. DeGorter, M.K., et al., *Drug transporters in drug efficacy and toxicity*. Annu Rev Pharmacol Toxicol, 2012. **52**: p. 249-73.
5. Klaassen, C.D. and L.M. Aleksunes, *Xenobiotic, bile acid, and cholesterol transporters: function and regulation*. Pharmacol Rev, 2010. **62**(1): p. 1-96.
6. Nakanishi, T. and I. Tamai, *Genetic polymorphisms of OATP transporters and their impact on intestinal absorption and hepatic disposition of drugs*. Drug Metab Pharmacokinet, 2012. **27**(1): p. 106-21.
7. Dror, O., et al., *Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design*. Curr Med Chem, 2004. **11**(1): p. 71-90.
8. Jahn, A., et al., *Optimal assignment methods for ligand-based virtual screening*. J Cheminform, 2009. **1**: p. 14.
9. Villoutreix, B.O., et al., *Free resources to assist structure-based virtual ligand screening experiments*. Curr Protein Pept Sci, 2007. **8**(4): p. 381-411.
10. Fan, H., J.J. Irwin, and A. Sali, *Virtual ligand screening against comparative protein structure models*. Methods Mol Biol, 2012. **819**: p. 105-26.
11. Schlessinger, A., et al., *SLC classification: an update*. Clin Pharmacol Ther, 2013. **94**(1): p. 19-23.
12. Breiman, L., *Random Forests*. Machine Learning, 2001. **45**(1): p. 5-32.
13. Knox, C., et al., *DrugBank 3.0: a comprehensive resource for 'omics' research on drugs*. Nucleic Acids Res, 2011. **39**(Database issue): p. D1035-41.
14. Dimitrov, S., et al., *A stepwise approach for defining the applicability domain of SAR and QSAR models*. J Chem Inf Model, 2005. **45**(4): p. 839-49.
15. Dragos, H., M. Gilles, and V. Alexandre, *Predicting the predictability: a unified approach to the applicability domain problem of QSAR models*. J Chem Inf Model, 2009. **49**(7): p. 1762-76.
16. Sahigara, F., et al., *Comparison of different approaches to define the applicability domain of QSAR models*. Molecules, 2012. **17**(5): p. 4791-810.
17. Sheridan, R.P., *Three useful dimensions for domain applicability in QSAR models using random forest*. J Chem Inf Model, 2012. **52**(3): p. 814-23.

18. Brenke, R., et al., *Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques*. *Bioinformatics*, 2009. **25**(5): p. 621-7.
19. Freed, E.O., *HIV-1 replication*. *Somatic cell and molecular genetics*, 2001. **26**(1-6): p. 13-33.
20. Kohl, N.E., et al., *Active human immunodeficiency virus protease is required for viral infectivity*. *Proceedings of the National Academy of Sciences of the United States of America*, 1988. **85**(13): p. 4686-90.
21. Murthy, K.H., et al., *The crystal structures at 2.2-Å resolution of hydroxyethylene-based inhibitors bound to human immunodeficiency virus type 1 protease show that the inhibitors are present in two distinct orientations*. *The Journal of biological chemistry*, 1992. **267**(32): p. 22770-8.
22. Prabu-Jeyabalan, M., E. Nalivaika, and C.A. Schiffer, *How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease*. *Journal of molecular biology*, 2000. **301**(5): p. 1207-20.
23. Mahalingam, B., et al., *Structural implications of drug-resistant mutants of HIV-1 protease: high-resolution crystal structures of the mutant protease/substrate analogue complexes*. *Proteins*, 2001. **43**(4): p. 455-64.
24. McDonald, C.K. and D.R. Kuritzkes, *Human immunodeficiency virus type 1 protease inhibitors*. *Archives of internal medicine*, 1997. **157**(9): p. 951-9.
25. Drag, M. and G.S. Salvesen, *Emerging principles in protease-based drug discovery*. *Nature reviews. Drug discovery*, 2010. **9**(9): p. 690-701.
26. Flexner, C., *HIV-protease inhibitors*. *The New England journal of medicine*, 1998. **338**(18): p. 1281-92.
27. Nie, Z., et al., *HIV-1 protease processes procaspase 8 to cause mitochondrial release of cytochrome c, caspase cleavage and nuclear fragmentation*. *Cell death and differentiation*, 2002. **9**(11): p. 1172-84.
28. Algeciras-Schimnich, A., et al., *Analysis of HIV Protease Killing Through Caspase 8 Reveals a Novel Interaction Between Caspase 8 and Mitochondria*. *The open virology journal*, 2007. **1**: p. 39-46.
29. Nie, Z., et al., *HIV Protease Cleavage of Procaspase 8 is Necessary for Death of HIV-Infected Cells*. *The open virology journal*, 2008. **2**: p. 1-7.
30. Castello, A., et al., *HIV-1 protease inhibits Cap- and poly(A)-dependent translation upon eIF4GI and PABP cleavage*. *PloS one*, 2009. **4**(11): p. e7997.
31. Impens, F., et al., *A catalogue of putative HIV-1 protease host cell substrates*. *Biological chemistry*, 2012. **393**(9): p. 915-31.
32. Ermak, D.L. and J.A. McCammon, *Brownian dynamics with hydrodynamic interactions*. *The Journal of Chemical Physics*, 1978. **69**(4): p. 1352-1360.
33. Perutz, M.F. and J.M. Mitchison, *State of haemoglobin in sickle-cell anaemia*. *Nature*, 1950. **166**(4225): p. 677-9.

34. Perutz, R.R., A.M. Liquori, and F. Eirich, *X-ray and solubility studies of the haemoglobin of sickle-cell anaemia patients*. *Nature*, 1951. **167**(4258): p. 929-31.
35. Webb, B., et al., *Modeling of proteins and their assemblies with the Integrative Modeling Platform*. *Methods Mol Biol*, 2014. **1091**: p. 277-95.
36. Nagel, R.L., et al., *Beta-chain contact sites in the haemoglobin S polymer*. *Nature*, 1980. **283**(5750): p. 832-4.
37. Hediger, M.A., et al., *The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins* Introduction. *Pflugers Arch*, 2004. **447**(5): p. 465-8.
38. Povey, S., et al., *The HUGO Gene Nomenclature Committee (HGNC)*. *Hum Genet*, 2001. **109**(6): p. 678-80.
39. Saier, M.H., Jr., *A functional-phylogenetic classification system for transmembrane solute transporters*. *Microbiol Mol Biol Rev*, 2000. **64**(2): p. 354-411.
40. Saier, M.H., Jr., et al., *The Transporter Classification Database: recent advances*. *Nucleic Acids Res*, 2009. **37**(Database issue): p. D274-8.
41. Forrest, L.R. and G. Rudnick, *The rocking bundle: a mechanism for ion-coupled solute flux by symmetrical transporters*. *Physiology (Bethesda)*, 2009. **24**: p. 377-86.
42. Forrest, L.R., R. Kramer, and C. Ziegler, *The structural basis of secondary active transport mechanisms*. *Biochimica et biophysica acta*, 2011. **1807**(2): p. 167-88.
43. Jardetzky, O., *Simple allosteric model for membrane pumps*. *Nature*, 1966. **211**(5052): p. 969-70.
44. Gether, U., et al., *Neurotransmitter transporters: molecular function of important drug targets*. *Trends Pharmacol Sci*, 2006. **27**(7): p. 375-83.
45. Blakely, R.D. and L.J. DeFelice, *All aglow about presynaptic receptor regulation of neurotransmitter transporters*. *Mol Pharmacol*, 2007. **71**(5): p. 1206-8.
46. Chen, N.H., M.E. Reith, and M.W. Quick, *Synaptic uptake and beyond: the sodium- and chloride-dependent neurotransmitter transporter family SLC6*. *Pflugers Arch*, 2004. **447**(5): p. 519-31.
47. You, G., Morris, M. E., Wang, B., *Drug Transporters: Molecular Characterization and Role in Drug Disposition*, ed. B. Wang. 2007, Hoboken: Wiley. 897.
48. Otsuka, M., et al., *A human transporter protein that mediates the final excretion step for toxic organic cations*. *Proceedings of the National Academy of Sciences of the United States of America*, 2005. **102**(50): p. 17923-8.
49. Meyer zu Schwabedissen, H.E., et al., *Human multidrug and toxin extrusion 1 (MATE1/SLC47A1) transporter: functional characterization, interaction with OCT2 (SLC22A2), and single nucleotide polymorphisms*. *American journal of physiology. Renal physiology*, 2010. **298**(4): p. F997-F1005.

50. Leabman, M.K., et al., *Natural variation in human membrane transporter genes reveals evolutionary and functional constraints*. Proc Natl Acad Sci U S A, 2003. **100**(10): p. 5896-901.
51. Koepsell, H., B.M. Schmitt, and V. Gorboulev, *Organic cation transporters*. Rev Physiol Biochem Pharmacol, 2003. **150**: p. 36-90.
52. Cropp, C.D., S.W. Yee, and K.M. Giacomini, *Genetic variation in drug transporters in ethnic populations*. Clinical pharmacology and therapeutics, 2008. **84**(3): p. 412-6.
53. Giacomini, K.M., et al., *Membrane transporters in drug development*. Nature reviews. Drug discovery, 2010. **9**(3): p. 215-36.
54. Shu, Y., et al., *Effect of genetic variation in the organic cation transporter 1, OCT1, on metformin pharmacokinetics*. Clin Pharmacol Ther, 2008. **83**(2): p. 273-80.
55. Shu, Y., et al., *Effect of genetic variation in the organic cation transporter 1 (OCT1) on metformin action*. J Clin Invest, 2007. **117**(5): p. 1422-31.
56. Shu, Y., et al., *Evolutionary conservation predicts function of variants of the human organic cation transporter, OCT1*. Proc Natl Acad Sci U S A, 2003. **100**(10): p. 5902-7.
57. Dresser, M.J., A.T. Gray, and K.M. Giacomini, *Kinetic and selectivity differences between rodent, rabbit, and human organic cation transporters (OCT1)*. J Pharmacol Exp Ther, 2000. **292**(3): p. 1146-52.
58. Fredriksson, R., et al., *The solute carrier (SLC) complement of the human genome: phylogenetic classification reveals four major families*. FEBS Lett, 2008. **582**(27): p. 3811-6.
59. Gruswitz, F., Chaudhary, S., Ho, J. D., Schlessinger, A., Pezeshki, B., Ho, C. M., Sali, A., Westhoff, C. M., Stroud, R.M., *Renal Ammonia Transport and the Structure of the Human Rhesus Glycoprotein RhCG at 2.1 Å*. submitted, 2009.
60. Karpowich, N.K. and D.N. Wang, *Structural biology. Symmetric transporters for asymmetric transport*. Science, 2008. **321**(5890): p. 781-2.
61. Forrest, L.R., et al., *Mechanism for alternating access in neurotransmitter transporters*. Proc Natl Acad Sci U S A, 2008. **105**(30): p. 10338-43.
62. Gao, X., et al., *Structure and mechanism of an amino acid antiporter*. Science, 2009. **324**(5934): p. 1565-8.
63. Fang, Y., et al., *Structure of a prokaryotic virtual proton pump at 3.2 Å resolution*. Nature, 2009. **460**(7258): p. 1040-3.
64. Shaffer, P.L., et al., *Structure and mechanism of a Na⁺-independent amino acid transporter*. Science, 2009. **325**(5943): p. 1010-4.
65. Lu, F., et al., *Structure and mechanism of the uracil transporter UraA*. Nature, 2011. **472**(7342): p. 243-6.
66. Weyand, S., et al., *Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter*. Science, 2008. **322**(5902): p. 709-13.

67. Ressler, S., et al., *Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP*. Nature, 2009. **458**(7234): p. 47-52.
68. Ma, D., et al., *Structure and mechanism of a glutamate-GABA antiporter*. Nature, 2012. **483**(7391): p. 632-6.
69. Schulze, S., et al., *Structural basis of Na(+)-independent and cooperative substrate/product antiport in CaiT*. Nature, 2010. **467**(7312): p. 233-6.
70. Lomize, M.A., et al., *OPM: orientations of proteins in membranes database*. Bioinformatics, 2006. **22**(5): p. 623-5.
71. Krishnamurthy, H., C.L. Piscitelli, and E. Gouaux, *Unlocking the molecular secrets of sodium-coupled transporters*. Nature, 2009. **459**(7245): p. 347-55.
72. Abramson, J. and E.M. Wright, *Structure and function of Na(+)-symporters with inverted repeats*. Current opinion in structural biology, 2009. **19**(4): p. 425-32.
73. Khafizov, K., et al., *A study of the evolution of inverted-topology repeats from LeuT-fold transporters using AlignMe*. Biochemistry, 2010. **49**(50): p. 10702-13.
74. Kanner, B.I. and E. Zomot, *Sodium-coupled neurotransmitter transporters*. Chem Rev, 2008. **108**(5): p. 1654-68.
75. Guan, L. and H.R. Kaback, *Lessons from lactose permease*. Annu Rev Biophys Biomol Struct, 2006. **35**: p. 67-91.
76. Krishnamurthy, H. and E. Gouaux, *X-ray structures of LeuT in substrate-free outward-open and apo inward-open states*. Nature, 2012. **481**(7382): p. 469-74.
77. Nyola, A., et al., *Substrate and drug binding sites in LeuT*. Curr Opin Struct Biol, 2010. **20**(4): p. 415-22.
78. Singh, S.K., et al., *A competitive inhibitor traps LeuT in an open-to-out conformation*. Science, 2008. **322**(5908): p. 1655-61.
79. Shi, L., et al., *The mechanism of a neurotransmitter:sodium symporter--inward release of Na+ and substrate is triggered by substrate in a second binding site*. Mol Cell, 2008. **30**(6): p. 667-77.
80. Zhao, Y., et al., *Single-molecule dynamics of gating in a neurotransmitter transporter homologue*. Nature, 2010. **465**(7295): p. 188-93.
81. Zhao, Y., et al., *Substrate-modulated gating dynamics in a Na(+)-coupled neurotransmitter transporter homologue*. Nature, 2011.
82. Nyola, A., et al., *Substrate and drug binding sites in LeuT*. Current opinion in structural biology, 2010. **20**(4): p. 415-22.
83. Claxton, D.P., et al., *Ion/substrate-dependent conformational dynamics of a bacterial homolog of neurotransmitter:sodium symporters*. Nature structural & molecular biology, 2010. **17**(7): p. 822-9.
84. Shimamura, T., et al., *Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1*. Science, 2010. **328**(5977): p. 470-3.

85. Gao, X., et al., *Mechanism of substrate recognition and transport by an amino acid antiporter*. Nature, 2010. **463**(7282): p. 828-32.
86. Radestock, S. and L.R. Forrest, *The alternating-access mechanism of MFS transporters arises from inverted-topology repeats*. Journal of molecular biology, 2011. **407**(5): p. 698-715.
87. Faraldo-Gomez, J.D. and L.R. Forrest, *Modeling and simulation of ion-coupled and ATP-driven membrane proteins*. Current opinion in structural biology, 2011. **21**(2): p. 173-9.
88. Schlessinger, A., et al., *Comparison of human solute carriers*. Protein science : a publication of the Protein Society, 2010. **19**(3): p. 412-28.
89. Rost, B., et al., *Automatic prediction of protein function*. Cellular and molecular life sciences : CMLS, 2003. **60**(12): p. 2637-50.
90. Sun, L., et al., *Crystal structure of a bacterial homologue of glucose transporters GLUT1-4*. Nature, 2012. **490**(7420): p. 361-6.
91. Abramson, J., et al., *Structure and mechanism of the lactose permease of Escherichia coli*. Science, 2003. **301**(5633): p. 610-5.
92. Newstead, S., et al., *Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2*. The EMBO journal, 2011. **30**(2): p. 417-26.
93. Punta, M., et al., *Membrane protein prediction methods*. Methods, 2007. **41**(4): p. 460-74.
94. Nugent, T. and D.T. Jones, *Membrane protein structural bioinformatics*. Journal of structural biology, 2012. **179**(3): p. 327-37.
95. Nugent, T. and D.T. Jones, *Transmembrane protein topology prediction using support vector machines*. BMC bioinformatics, 2009. **10**: p. 159.
96. Soding, J., A. Biegert, and A.N. Lupas, *The HHpred interactive server for protein homology detection and structure prediction*. Nucleic acids research, 2005. **33**(Web Server issue): p. W244-8.
97. Pei, J., B.H. Kim, and N.V. Grishin, *PROMALS3D: a tool for multiple protein sequence and structure alignments*. Nucleic acids research, 2008. **36**(7): p. 2295-300.
98. Albers, T., et al., *Defining substrate and blocker activity of alanine-serine-cysteine transporter 2 (ASCT2) Ligands with Novel Serine Analogs*. Molecular pharmacology, 2012. **81**(3): p. 356-65.
99. Beuming, T., et al., *A comprehensive structure-based alignment of prokaryotic and eukaryotic neurotransmitter/Na⁺ symporters (NSS) aids in the use of the LeuT structure to probe NSS structure and function*. Mol Pharmacol, 2006. **70**(5): p. 1630-42.
100. Atkinson, H.J., et al., *Using sequence similarity networks for visualization of relationships across diverse protein superfamilies*. PLoS ONE, 2009. **4**(2): p. e4345.

101. Brown, S.D. and P.C. Babbitt, *Inference of functional properties from large-scale analysis of enzyme superfamilies*. The Journal of biological chemistry, 2012. **287**(1): p. 35-42.
102. Shoichet, B.K., et al., *Lead discovery using molecular docking*. Curr Opin Chem Biol, 2002. **6**(4): p. 439-46.
103. Shoichet, B.K., *Virtual screening of chemical libraries*. Nature, 2004. **432**(7019): p. 862-5.
104. Jacobson, M. and A. Sali, *Comparative Protein Structure Modeling and its Applications to Drug Discovery*, in *Annual Reports in Medicinal Chemistry*. 2004, Academic Press. p. 259-276.
105. Shoichet, B.K., et al., *Structure-based discovery of inhibitors of thymidylate synthase*. Science, 1993. **259**(5100): p. 1445-50.
106. Mysinger, M.M. and B.K. Shoichet, *Rapid context-dependent ligand desolvation in molecular docking*. Journal of chemical information and modeling, 2010. **50**(9): p. 1561-73.
107. Kuntz, I.D., *Structure-based strategies for drug design and discovery*. Science, 1992. **257**(5073): p. 1078-82.
108. Huang, N., B.K. Shoichet, and J.J. Irwin, *Benchmarking sets for molecular docking*. J Med Chem, 2006. **49**(23): p. 6789-801.
109. Fan, H., et al., *Molecular docking screens using comparative models of proteins*. J Chem Inf Model, 2009. **49**(11): p. 2512-27.
110. Schlessinger, A., et al., *Structure-based discovery of prescription drugs that interact with the norepinephrine transporter, NET*. Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(38): p. 15810-5.
111. Evers, A., H. Gohlke, and G. Klebe, *Ligand-supported homology modelling of protein binding-sites using knowledge-based potentials*. Journal of molecular biology, 2003. **334**(2): p. 327-45.
112. Cavasotto, C.N., et al., *Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening*. Journal of medicinal chemistry, 2008. **51**(3): p. 581-8.
113. Carlsson, J., et al., *Ligand discovery from a dopamine D3 receptor homology model and crystal structure*. Nature chemical biology, 2011. **7**(11): p. 769-78.
114. Schlessinger, A., et al., *High Selectivity of the gamma-Aminobutyric Acid Transporter 2 (GAT-2, SLC6A13) Revealed by Structure-based Approach*. The Journal of biological chemistry, 2012. **287**(45): p. 37745-56.
115. Shoichet, B.K. and B.K. Kobilka, *Structure-based drug screening for G-protein-coupled receptors*. Trends in pharmacological sciences, 2012. **33**(5): p. 268-72.
116. Kufareva, I., et al., *Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment*. Structure, 2011. **19**(8): p. 1108-26.

117. Schapira, M., R. Abagyan, and M. Totrov, *Nuclear hormone receptor targeted virtual screening*. Journal of medicinal chemistry, 2003. **46**(14): p. 3045-59.
118. Diller, D.J. and R. Li, *Kinases, homology models, and high throughput docking*. Journal of medicinal chemistry, 2003. **46**(22): p. 4638-47.
119. Song, L., et al., *Prediction and assignment of function for a divergent N-succinyl amino acid racemase*. Nature chemical biology, 2007. **3**(8): p. 486-91.
120. Kalyanaraman, C., et al., *Discovery of a dipeptide epimerase enzymatic function guided by homology modeling and virtual screening*. Structure, 2008. **16**(11): p. 1668-77.
121. Astorga, B., et al., *Molecular determinants of ligand selectivity for the human multidrug and toxin extruder proteins MATE1 and MATE2-K*. The Journal of pharmacology and experimental therapeutics, 2012. **341**(3): p. 743-55.
122. Andersen, J., et al., *Recent advances in the understanding of the interaction of antidepressant drugs with serotonin and norepinephrine transporters*. Chem Commun (Camb), 2009(25): p. 3677-92.
123. Ekins, S., et al., *Future directions for drug transporter modelling*. Xenobiotica; the fate of foreign compounds in biological systems, 2007. **37**(10-11): p. 1152-70.
124. Chang, C. and P.W. Swaan, *Computational approaches to modeling drug transporters*. European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences, 2006. **27**(5): p. 411-24.
125. Eyre, T.A., L. Partridge, and J.M. Thornton, *Computational analysis of alpha-helical membrane protein structure: implications for the prediction of 3D structural models*. Protein engineering, design & selection : PEDS, 2004. **17**(8): p. 613-24.
126. Hahn, M.K. and R.D. Blakely, *The functional impact of SLC6 transporter genetic variation*. Annu Rev Pharmacol Toxicol, 2007. **47**: p. 401-41.
127. Pacholczyk, T., R.D. Blakely, and S.G. Amara, *Expression cloning of a cocaine- and antidepressant-sensitive human noradrenaline transporter*. Nature, 1991. **350**(6316): p. 350-4.
128. Celik, L., et al., *Binding of serotonin to the human serotonin transporter. Molecular modeling and experimental validation*. J Am Chem Soc, 2008. **130**(12): p. 3853-65.
129. Beuming, T., et al., *The binding sites for cocaine and dopamine in the dopamine transporter overlap*. Nat Neurosci, 2008. **11**(7): p. 780-9.
130. Severinsen, K., et al., *Binding of the Amphetamine-like 1-Phenyl-piperazine to Monoamine Transporters*. ACS chemical neuroscience, 2012. **3**(9): p. 693-705.
131. Daylight Chemical Information Systems, Inc.: Laguna Niguel, CA 92677.
132. Adamus, W.S., J.P. Leonard, and W. Troger, *Phase I clinical trials with WAL 2014, a new muscarinic agonist for the treatment of Alzheimer's disease*. Life sciences, 1995. **56**(11-12): p. 883-90.

133. Madsen, K.K., H.S. White, and A. Schousboe, *Neuronal and non-neuronal GABA transporters as targets for antiepileptic drugs*. *Pharmacology & therapeutics*, 2010. **125**(3): p. 394-401.
134. Christiansen, B., et al., *Cloning and characterization of a functional human gamma-aminobutyric acid (GABA) transporter, human GAT-2*. *The Journal of biological chemistry*, 2007. **282**(27): p. 19331-41.
135. Nakashita, M., et al., *Effects of tricyclic and tetracyclic antidepressants on the three subtypes of GABA transporter*. *Neuroscience research*, 1997. **29**(1): p. 87-91.
136. Melamed, N. and B.I. Kanner, *Transmembrane domains I and II of the gamma-aminobutyric acid transporter GAT-4 contain molecular determinants of substrate specificity*. *Molecular pharmacology*, 2004. **65**(6): p. 1452-61.
137. Kanner, B.I., *Transmembrane domain I of the gamma-aminobutyric acid transporter GAT-1 plays a crucial role in the transition between cation leak and transport modes*. *The Journal of biological chemistry*, 2003. **278**(6): p. 3705-12.
138. Irwin, J.J. and B.K. Shoichet, *ZINC--a free database of commercially available compounds for virtual screening*. *J Chem Inf Model*, 2005. **45**(1): p. 177-82.
139. Albers, J.W. and J.K. Fink, *Porphyric neuropathy*. *Muscle & nerve*, 2004. **30**(4): p. 410-22.
140. Lindberg, R.L., et al., *Motor neuropathy in porphobilinogen deaminase-deficient mice imitates the peripheral neuropathy of human acute porphyria*. *The Journal of clinical investigation*, 1999. **103**(8): p. 1127-34.
141. Musiol, R., M. Serda, and J. Polanski, *Prodrugs in photodynamic anticancer therapy*. *Current pharmaceutical design*, 2011. **17**(32): p. 3548-59.
142. Stummer, W., et al., *Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial*. *The lancet oncology*, 2006. **7**(5): p. 392-401.
143. Fukuda, H., A. Casas, and A. Batlle, *Aminolevulinic acid: from its unique biological function to its star role in photodynamic therapy*. *The international journal of biochemistry & cell biology*, 2005. **37**(2): p. 272-6.
144. Kanai, Y., et al., *Expression cloning and characterization of a transporter for large neutral amino acids activated by the heavy chain of 4F2 antigen (CD98)*. *The Journal of biological chemistry*, 1998. **273**(37): p. 23629-32.
145. Roberts, L.M., et al., *Subcellular localization of transporters along the rat blood-brain barrier and blood-cerebral-spinal fluid barrier by in vivo biotinylation*. *Neuroscience*, 2008. **155**(2): p. 423-38.
146. Alexander, G.M., et al., *Effect of plasma levels of large neutral amino acids and degree of parkinsonism on the blood-to-brain transport of levodopa in naive and MPTP parkinsonian monkeys*. *Neurology*, 1994. **44**(8): p. 1491-9.

147. Wang, Y. and D.F. Welty, *The simultaneous estimation of the influx and efflux blood-brain barrier permeabilities of gabapentin using a microdialysis-pharmacokinetic approach*. *Pharmaceutical research*, 1996. **13**(3): p. 398-403.
148. Kaira, K., et al., *Prognostic significance of L-type amino acid transporter 1 expression in resectable stage I-III nonsmall cell lung cancer*. *British journal of cancer*, 2008. **98**(4): p. 742-8.
149. Kobayashi, K., et al., *Enhanced tumor growth elicited by L-type amino acid transporter 1 in human malignant glioma cells*. *Neurosurgery*, 2008. **62**(2): p. 493-503; discussion 503-4.
150. Kroemer, G. and J. Pouyssegur, *Tumor cell metabolism: cancer's Achilles' heel*. *Cancer Cell*, 2008. **13**(6): p. 472-82.
151. Nicklin, P., et al., *Bidirectional transport of amino acids regulates mTOR and autophagy*. *Cell*, 2009. **136**(3): p. 521-34.
152. Verrey, F., et al., *CATs and HATs: the SLC7 family of amino acid transporters*. *Pflügers Archiv : European journal of physiology*, 2004. **447**(5): p. 532-42.
153. Geier, E.G., Schlessinger, A., Fan, H., Gable, J. E., Irwin, J. J., Sali, A., Giacomini, K. M., *Structure-based ligand discovery for the Large-neutral Amino Acid Transporter 1, LAT-1*. Submitted.
154. Hidalgo, M., et al., *A Phase I and pharmacological study of the glutamine antagonist acivicin with the amino acid solution aminosyn in patients with advanced solid malignancies*. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 1998. **4**(11): p. 2763-70.
155. Dolgih, E., et al., *Predicting binding to p-glycoprotein by flexible receptor docking*. *PLoS computational biology*, 2011. **7**(6): p. e1002083.
156. Masuda, S., et al., *Identification and functional characterization of a new human kidney-specific H⁺/organic cation antiporter, kidney-specific multidrug and toxin extrusion 2*. *Journal of the American Society of Nephrology : JASN*, 2006. **17**(8): p. 2127-35.
157. Otsuka, M., et al., *Identification of essential amino acid residues of the NorM Na⁺/multidrug antiporter in Vibrio parahaemolyticus*. *Journal of bacteriology*, 2005. **187**(5): p. 1552-8.
158. Becker, M.L., et al., *Genetic variation in the organic cation transporter 1 is associated with metformin response in patients with diabetes mellitus*. *The pharmacogenomics journal*, 2009. **9**(4): p. 242-7.
159. Jablonski, K.A., et al., *Common variants in 40 genes assessed for diabetes incidence and response to metformin and lifestyle intervention in the diabetes prevention program*. *Diabetes*, 2010. **59**(10): p. 2672-81.
160. Nies, A.T., et al., *Organic cation transporters (OCTs, MATEs), in vitro and in vivo evidence for the importance in drug therapy*. *Handbook of experimental pharmacology*, 2011(201): p. 105-67.

161. Damme, K., et al., *Mammalian MATE (SLC47A) transport proteins: impact on efflux of endogenous substrates and xenobiotics*. Drug metabolism reviews, 2011. **43**(4): p. 499-523.
162. Kido, Y., P. Matsson, and K.M. Giacomini, *Profiling of a prescription drug library for potential renal drug-drug interactions mediated by the organic cation transporter 2*. Journal of medicinal chemistry, 2011. **54**(13): p. 4548-58.
163. Wittwer, M.B., Zur, A. A., Khuri, N., Kido, Y., Kosaka A., Zhang, X., Morrissey K. M., Sali A., Huang Y., Giacomini K. M., *Discovery of Potent, Selective Multidrug And Toxin Extrusion Transporter 1 (MATE1, SLC47A1) Inhibitors Through Prescription Drug Profiling and Computational Modeling*. Journal of Medicinal Chemistry, In press.
164. Wishart, D.S., et al., *DrugBank: a comprehensive resource for in silico drug discovery and exploration*. Nucleic Acids Res, 2006. **34**(Database issue): p. D668-72.
165. He, X., et al., *Structure of a cation-bound multidrug and toxic compound extrusion transporter*. Nature, 2010. **467**(7318): p. 991-4.
166. Zhang, X. and S.H. Wright, *MATE1 has an external COOH terminus, consistent with a 13-helix topology*. American journal of physiology. Renal physiology, 2009. **297**(2): p. F263-71.
167. Zhang, X., et al., *Twelve transmembrane helices form the functional core of mammalian MATE1 (multidrug and toxin extruder 1) protein*. The Journal of biological chemistry, 2012. **287**(33): p. 27971-82.
168. Kajiwara, M., et al., *Identification of multidrug and toxin extrusion (MATE1 and MATE2-K) variants with complete loss of transport activity*. Journal of human genetics, 2009. **54**(1): p. 40-6.
169. Choi, J.H., et al., *A Common 5'-UTR Variant in MATE2-K Is Associated With Poor Response to Metformin*. Clinical pharmacology and therapeutics, 2011.
170. Stroud, R.M., et al., *2007 Annual progress report synopsis of the Center for Structures of Membrane Proteins*. J Struct Funct Genomics, 2009.
171. Love, J., et al., *The New York Consortium on Membrane Protein Structure (NYCOMPS): a high-throughput platform for structural genomics of integral membrane proteins*. Journal of structural and functional genomics, 2010. **11**(3): p. 191-9.
172. Mysinger, M.M., et al., *Structure-based ligand discovery for the protein-protein interface of chemokine receptor CXCR4*. Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(14): p. 5517-22.
173. Fan, H., et al., *Statistical potential for modeling and ranking of protein-ligand interactions*. Journal of chemical information and modeling, 2011. **51**(12): p. 3078-92.
174. Zamek-Gliszczyński, M.J., et al., *Highlights from the international transporter consortium second workshop*. Clinical pharmacology and therapeutics, 2012. **92**(5): p. 553-6.
175. Boudker, O., et al., *Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter*. Nature, 2007. **445**(7126): p. 387-93.

176. Yamashita, A., et al., *Crystal structure of a bacterial homologue of Na⁺/Cl⁻-dependent neurotransmitter transporters*. Nature, 2005. **437**(7056): p. 215-23.
177. Hu, N.J., et al., *Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT*. Nature, 2011. **478**(7369): p. 408-11.
178. Pebay-Peyroula, E., et al., *Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside*. Nature, 2003. **426**(6962): p. 39-44.
179. Johnson, Z.L., C.G. Cheong, and S.Y. Lee, *Crystal structure of a concentrative nucleoside transporter from Vibrio cholerae at 2.4 Å*. Nature, 2012. **483**(7390): p. 489-93.
180. Gruswitz, F., et al., *Function of human Rh based on structure of RhCG at 2.1 Å*. Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(21): p. 9638-43.
181. DeLano, W.L., *The PyMOL Molecular Graphics System*, 2002: San Carlos, CA, USA.
182. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks*. Genome Res, 2003. **13**(11): p. 2498-504.
183. Madhusudhan, M.S., et al., *Alignment of multiple protein structures based on sequence and structure features*. Protein Eng Des Sel, 2009. **22**(9): p. 569-74.
184. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis*. J Comput Chem, 2004. **25**(13): p. 1605-12.
185. Enright, A.J., S. Van Dongen, and C.A. Ouzounis, *An efficient algorithm for large-scale detection of protein families*. Nucleic acids research, 2002. **30**(7): p. 1575-84.
186. Shen, M.Y. and A. Sali, *Statistical potential for assessment and prediction of protein structures*. Protein Sci, 2006. **15**(11): p. 2507-24.
187. Krivov, G.G., M.V. Shapovalov, and R.L. Dunbrack, Jr., *Improved prediction of protein side-chain conformations with SCWRL4*. Proteins, 2009. **77**(4): p. 778-95.
188. Hess, B., et al., *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*. Journal of chemical theory and computation, 2008. **4**(3): p. 435-447.
189. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field*. Proteins, 2010. **78**(8): p. 1950-8.
190. Shaw, D.E., et al., *Atomic-level characterization of the structural dynamics of proteins*. Science, 2010. **330**(6002): p. 341-6.
191. Hess, B., *P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation*. Journal of chemical theory and computation, 2007. **4**(1): p. 116-122.
192. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*. The Journal of chemical physics, 2007. **126**(1): p. 014101.
193. Tusnady, G.E., L. Kalmar, and I. Simon, *TOPDB: topology data bank of transmembrane proteins*. Nucleic acids research, 2008. **36**(Database issue): p. D234-9.

194. Kobayashi, D., et al., *Involvement of human organic anion transporting polypeptide OATP-B (SLC21A9) in pH-dependent transport across intestinal apical membrane*. J Pharmacol Exp Ther, 2003. **306**(2): p. 703-8.
195. Banerjee, N., C. Allen, and R. Bendayan, *Differential role of organic anion-transporting polypeptides in estrone-3-sulphate uptake by breast epithelial cells and breast cancer cells*. J Pharmacol Exp Ther, 2012. **342**(2): p. 510-9.
196. Obaidat, A., M. Roth, and B. Hagenbuch, *The expression and function of organic anion transporting polypeptides in normal tissues and in cancer*. Annu Rev Pharmacol Toxicol, 2012. **52**: p. 135-51.
197. Imanaga, J., et al., *The effects of the SLCO2B1 c.1457C > T polymorphism and apple juice on the pharmacokinetics of fexofenadine and midazolam in humans*. Pharmacogenet Genomics, 2011. **21**(2): p. 84-93.
198. Shirasaka, Y., et al., *Intestinal absorption of HMG-CoA reductase inhibitor pravastatin mediated by organic anion transporting polypeptide*. Pharm Res, 2010. **27**(10): p. 2141-9.
199. Tamai, I., *Oral drug delivery utilizing intestinal OATP transporters*. Adv Drug Deliv Rev, 2012. **64**(6): p. 508-14.
200. Tamai, I. and T. Nakanishi, *OATP transporter-mediated drug absorption and interaction*. Curr Opin Pharmacol, 2013. **13**(6): p. 859-63.
201. Shirasaka, Y., et al., *Long-lasting inhibitory effect of apple and orange juices, but not grapefruit juice, on OATP2B1-mediated drug absorption*. Drug Metab Dispos, 2013. **41**(3): p. 615-21.
202. Shirasaka, Y., et al., *Differential effect of grapefruit juice on intestinal absorption of statins due to inhibition of organic anion transporting polypeptide and/or P-glycoprotein*. J Pharm Sci, 2011. **100**(9): p. 3843-53.
203. Grube, M., et al., *Organic anion transporting polypeptide 2B1 is a high-affinity transporter for atorvastatin and is expressed in the human heart*. Clin Pharmacol Ther, 2006. **80**(6): p. 607-20.
204. Noe, J., et al., *Substrate-dependent drug-drug interactions between gemfibrozil, fluvastatin and other organic anion-transporting peptide (OATP) substrates on OATP1B1, OATP2B1, and OATP1B3*. Drug Metab Dispos, 2007. **35**(8): p. 1308-14.
205. Kitamura, S., et al., *Involvement of multiple transporters in the hepatobiliary transport of rosuvastatin*. Drug Metab Dispos, 2008. **36**(10): p. 2014-23.
206. Kullak-Ublick, G.A., et al., *Organic anion-transporting polypeptide B (OATP-B) and its functional comparison with three other OATPs of human liver*. Gastroenterology, 2001. **120**(2): p. 525-33.
207. Treiber, A., et al., *Bosentan is a substrate of human OATP1B1 and OATP1B3: inhibition of hepatic uptake as the common mechanism of its interactions with cyclosporin A, rifampicin, and sildenafil*. Drug Metab Dispos, 2007. **35**(8): p. 1400-7.

208. Satoh, H., et al., *Citrus juices inhibit the function of human organic anion-transporting polypeptide OATP-B*. Drug Metab Dispos, 2005. **33**(4): p. 518-23.
209. Nozawa, T., et al., *Functional characterization of pH-sensitive organic anion transporting polypeptide OATP-B in human*. J Pharmacol Exp Ther, 2004. **308**(2): p. 438-45.
210. Varma, M.V., et al., *pH-sensitive interaction of HMG-CoA reductase inhibitors (statins) with organic anion transporting polypeptide 2B1*. Mol Pharm, 2011. **8**(4): p. 1303-13.
211. Shirasaka, Y., et al., *Functional pleiotropy of organic anion transporting polypeptide OATP2B1 due to multiple binding sites*. Drug Metab Pharmacokinet, 2012. **27**(3): p. 360-4.
212. Hu, G., et al., *New fluorescent substrate enables quantitative and high-throughput examination of vesicular monoamine transporter 2 (VMAT2)*. ACS Chem Biol, 2013. **8**(9): p. 1947-54.
213. Zhou, W., et al., *Development and validation of a high-throughput screening assay for human long-chain fatty acid transport proteins 4 and 5*. J Biomol Screen, 2010. **15**(5): p. 488-97.
214. Sindelar, M. and K.T. Wanner, *Library screening by means of mass spectrometry (MS) binding assays-exemplarily demonstrated for a pseudostatic library addressing gamma-aminobutyric acid (GABA) transporter 1 (GAT1)*. ChemMedChem, 2012. **7**(9): p. 1678-90.
215. Zhang, L., et al., *A high-throughput screen for chemical inhibitors of exocytic transport in yeast*. Chembiochem, 2010. **11**(9): p. 1291-301.
216. Susa, M., et al., *Multidrug resistance reversal agent, NSC77037, identified with a cell-based screening assay*. J Biomol Screen, 2010. **15**(3): p. 287-96.
217. Badolo, L., et al., *Screening of OATP1B1/3 and OCT1 inhibitors in cryopreserved hepatocytes in suspension*. Eur J Pharm Sci, 2010. **40**(4): p. 282-8.
218. Gui, C., et al., *Development of a cell-based high-throughput assay to screen for inhibitors of organic anion transporting polypeptides 1B1 and 1B3*. Curr Chem Genomics, 2010. **4**: p. 1-8.
219. Karaki, F., et al., *Structure-activity relationship study of non-steroidal NPC1L1 ligands identified through cell-based assay using pharmacological chaperone effect as a readout*. Bioorg Med Chem, 2014. **22**(14): p. 3587-609.
220. De Bruyn, T., et al., *Structure-based identification of OATP1B1/3 inhibitors*. Mol Pharmacol, 2013. **83**(6): p. 1257-67.
221. Wittwer, M.B., et al., *Discovery of potent, selective multidrug and toxin extrusion transporter 1 (MATE1, SLC47A1) inhibitors through prescription drug profiling and computational modeling*. J Med Chem, 2013. **56**(3): p. 781-95.
222. Kido, Y., P. Matsson, and K.M. Giacomini, *Profiling of a prescription drug library for potential renal drug-drug interactions mediated by the organic cation transporter 2*. J Med Chem, 2011. **54**(13): p. 4548-58.

223. Esteva-Font, C., et al., *A small molecule screen identifies selective inhibitors of urea transporter UT-A*. Chem Biol, 2013. **20**(10): p. 1235-44.
224. Strouse, J.J., et al., *A selective ATP-binding cassette subfamily G member 2 efflux inhibitor revealed via high-throughput flow cytometry*. J Biomol Screen, 2013. **18**(1): p. 26-38.
225. Mohamed, T.M., et al., *Optimisation and validation of a high throughput screening compatible assay to identify inhibitors of the plasma membrane calcium ATPase pump--a novel therapeutic target for contraception and malaria*. J Pharm Pharm Sci, 2013. **16**(2): p. 217-30.
226. Colton, C.K., et al., *Identification of translational activators of glial glutamate transporter EAAT2 through cell-based high-throughput screening: an approach to prevent excitotoxicity*. J Biomol Screen, 2010. **15**(6): p. 653-62.
227. Zhang, Y., et al., *Identification of inhibitors of ABCG2 by a bioluminescence imaging-based high-throughput assay*. Cancer Res, 2009. **69**(14): p. 5867-75.
228. Gao, J., et al., *Identification of upregulators of human ATP-binding cassette transporter A1 via high-throughput screening of a synthetic and natural compound library*. J Biomol Screen, 2008. **13**(7): p. 648-56.
229. 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.
230. Karlgren, M., et al., *Classification of inhibitors of hepatic organic anion transporting polypeptides (OATPs): influence of protein expression on drug-drug interactions*. J Med Chem, 2012. **55**(10): p. 4740-63.
231. Schlessinger, A., et al., *High selectivity of the gamma-aminobutyric acid transporter 2 (GAT-2, SLC6A13) revealed by structure-based approach*. J Biol Chem, 2012. **287**(45): p. 37745-56.
232. Geier, E.G., et al., *Structure-based ligand discovery for the Large-neutral Amino Acid Transporter 1, LAT-1*. Proc Natl Acad Sci U S A, 2013. **110**(14): p. 5480-5.
233. Schlessinger, A., et al., *Molecular modeling and ligand docking for solute carrier (SLC) transporters*. Curr Top Med Chem, 2013. **13**(7): p. 843-56.
234. Schlessinger, A., et al., *Structure-based discovery of prescription drugs that interact with the norepinephrine transporter, NET*. Proc Natl Acad Sci U S A, 2011. **108**(38): p. 15810-5.
235. Kullback, S. and R.A. Leibler, *On Information and Sufficiency*. 1951: p. 79-86.
236. Altman, N.S., *An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression*. The American Statistician, 1992. **46**(3): p. 175-185.
237. Cortes, C. and V. Vapnik, *Support-vector networks*. Machine Learning, 1995. **20**(3): p. 273-297.

238. Freeman, E.A. and G.G. Moisen, *A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa*. *Ecological Modelling*, 2008. **217**(1–2): p. 48-58.
239. Balfer, J., M. Vogt, and J. Bajorath, *Searching for closely related ligands with different mechanisms of action using machine learning and mapping algorithms*. *J Chem Inf Model*, 2013. **53**(9): p. 2252-74.
240. Lavecchia, A. and C. Di Giovanni, *Virtual screening strategies in drug discovery: a critical review*. *Curr Med Chem*, 2013. **20**(23): p. 2839-60.
241. Pieper, U., et al., *ModBase, a database of annotated comparative protein structure models and associated resources*. *Nucleic Acids Res*, 2014. **42**(Database issue): p. D336-46.
242. Webb, B. and A. Sali, *Comparative Protein Structure Modeling Using MODELLER*. *Curr Protoc Bioinformatics*, 2014. **47**: p. 5 6 1-5 6 32.
243. Sali, A. and T.L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints*. *J Mol Biol*, 1993. **234**(3): p. 779-815.
244. Dong, G.Q., et al., *Optimized atomic statistical potentials: assessment of protein interfaces and loops*. *Bioinformatics*, 2013. **29**(24): p. 3158-66.
245. Leuthold, S., et al., *Mechanisms of pH-gradient driven transport mediated by organic anion polypeptide transporters*. Vol. 296. 2009. C570-C582.
246. Agarwal, S., V. Arya, and L. Zhang, *Review of P-gp inhibition data in recently approved new drug applications: utility of the proposed [I(1)]/IC(50) and [I(2)]/IC(50) criteria in the P-gp decision tree*. *J Clin Pharmacol*, 2013. **53**(2): p. 228-33.
247. Shitara, Y., et al., *Clinical significance of organic anion transporting polypeptides (OATPs) in drug disposition: their roles in hepatic clearance and intestinal absorption*. *Biopharm Drug Dispos*, 2013. **34**(1): p. 45-78.
248. Tan, L., et al., *Integrating structure- and ligand-based virtual screening: comparison of individual, parallel, and fused molecular docking and similarity search calculations on multiple targets*. *ChemMedChem*, 2008. **3**(10): p. 1566-71.
249. Bissantz, C., et al., *Focused library design in GPCR projects on the example of 5-HT(2c) agonists: comparison of structure-based virtual screening with ligand-based search methods*. *Proteins*, 2005. **61**(4): p. 938-52.
250. Johnston, R.A., et al., *Selective inhibition of human solute carrier transporters by multikinase inhibitors*. *Drug Metab Dispos*, 2014. **42**(11): p. 1851-7.
251. Li, N., et al., *Identification of amino acids essential for estrone-3-sulfate transport within transmembrane domain 2 of organic anion transporting polypeptide 1B1*. *PLoS One*, 2012. **7**(5): p. e36647.
252. Rizwan, A.N., W. Krick, and G. Burekhardt, *The chloride dependence of the human organic anion transporter 1 (hOAT1) is blunted by mutation of a single amino acid*. *J Biol Chem*, 2007. **282**(18): p. 13402-9.

253. Glaeser, H., et al., *Relevance of conserved lysine and arginine residues in transmembrane helices for the transport activity of organic anion transporting polypeptide 1B3*. Br J Pharmacol, 2010. **159**(3): p. 698-708.
254. Kaiser, J., *Science resources. Chemists want NIH to curtail database*. Science, 2005. **308**(5723): p. 774.
255. Team, R.C., *R: A language and environment for statistical computing*. 2012.
256. Yap, C.W., *PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints*. J Comput Chem, 2011. **32**(7): p. 1466-74.
257. Hanley, J.A. and B.J. McNeil, *The meaning and use of the area under a receiver operating characteristic (ROC) curve*. Radiology, 1982. **143**(1): p. 29-36.
258. Mysinger, M.M. and B.K. Shoichet, *Rapid context-dependent ligand desolvation in molecular docking*. J Chem Inf Model, 2010. **50**(9): p. 1561-73.
259. Rose, P.W., et al., *The RCSB Protein Data Bank: new resources for research and education*. Nucleic acids research, 2013. **41**(Database issue): p. D475-82.
260. Consortium, U., *Ongoing and future developments at the Universal Protein Resource*. Nucleic acids research, 2011. **39**(Database issue): p. D214-9.
261. Benson, D.A., et al., *GenBank*. Nucleic acids research, 2012. **40**(Database issue): p. D48-53.
262. Baker, D. and A. Sali, *Protein structure prediction and structural genomics*. Science, 2001. **294**(5540): p. 93-96.
263. Eswar, N., et al., *Comparative protein structure modeling using MODELLER*. Curr Protoc Bioinformatics, 2006. **Chapter 5**: p. Unit 5.6.1-5.6.30.
264. Eswar, N., et al., *Protein structure modeling with MODELLER*. Methods Mol Biol, 2008. **426**: p. 145-159.
265. Schwede, T., et al., *Protein Structure Modeling*, in *Computational Structural Biology*, T. Schwede and M.C. Peitsch, Editors. 2008, World Scientific Publishing Ltd.: Singapore. p. 3-35.
266. Forrest, L.R., C.L. Tang, and B. Honig, *On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins*. Biophysical journal, 2006. **91**(2): p. 508-17.
267. Liu, T., G.W. Tang, and E. Capriotti, *Comparative modeling: the state of the art and protein drug target structure prediction*. Combinatorial chemistry & high throughput screening, 2011. **14**(6): p. 532-47.
268. Fiser, A., *Template-based protein structure modeling*. Methods in molecular biology, 2010. **673**: p. 73-94.
269. Daga, P.R., R.Y. Patel, and R.J. Doerksen, *Template-based protein modeling: recent methodological advances*. Current topics in medicinal chemistry, 2010. **10**(1): p. 84-94.
270. Hillisch, A., L.F. Pineda, and R. Hilgenfeld, *Utility of homology models in the drug discovery process*. Drug discovery today, 2004. **9**(15): p. 659-69.

271. Wallner, B. and A. Elofsson, *All are not equal: a benchmark of different homology modeling programs*. Protein science : a publication of the Protein Society, 2005. **14**(5): p. 1315-27.
272. Weinkam, P., J. Pons, and A. Sali, *Structure-based Model of Allostery Predicts Coupling Between Distant Sites*. Proc Natl Acad Sci USA, 2012. **109**: p. 4875-4880.
273. Schneidman-Duhovny, D., M. Hammel, and A. Sali, *Macromolecular docking restrained by a small angle X-ray scattering profile*. J Struct Biol, 2011. **3**: p. 461-471.
274. Fan, H., et al., *Statistical Potential for Modeling and Ranking Protein-Ligand Interactions*. J Chem Inf Model, 2011. **51**: p. 3078-3092.
275. Eswar, N., et al., *Tools for comparative protein structure modeling and analysis*. Nucleic Acids Res, 2003. **31**(13): p. 3375-3380.
276. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. Nucleic Acids Res, 1997. **25**(17): p. 3389-402.
277. Remmert, M., et al., *HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment*. Nature methods, 2012. **9**(2): p. 173-5.
278. Soding, J., *Protein homology detection by HMM-HMM comparison*. Bioinformatics, 2005. **21**(7): p. 951-60.
279. Smith, T.F. and M.S. Waterman, *Identification of common molecular subsequences*. J Mol Biol, 1981. **147**(1): p. 195-7.
280. Eswar, N., et al., *Comparative protein structure modeling using Modeller*. Current Protocols in Bioinformatics, 2006. **Chapter 5**: p. Unit 5.6.
281. Marti-Renom, M.A., M.S. Madhusudhan, and A. Sali, *Alignment of protein sequences by their profiles*. Protein Sci, 2004. **13**(4): p. 1071-1087.
282. Shen, M.Y. and A. Sali, *Statistical potential for assessment and prediction of protein structures*. Protein Sci, 2006. **15**(11): p. 2507-2524.
283. Melo, F. and A. Sali, *Fold assessment for comparative protein structure modeling*. Protein Sci, 2007. **16**(11): p. 2412-2426.
284. Pieper, U., et al., *ModBase, a database of annotated comparative protein structure models, and associated resources*. Nucleic Acids Research, 2011. **39**: p. 465-474.
285. Eramian, D., et al., *How well can the accuracy of comparative protein structure models be predicted?* Protein Sci, 2008. **17**(11): p. 1881-1893.
286. Dong, G.Q., et al., *Optimized atomic statistical potentials: Assessment of protein interfaces and loops*. submitted.
287. Schneidman-Duhovny, D., et al., *A Method for Integrative Structure Determination of Protein-Protein Complexes*. Bioinformatics, 2012. **28**: p. 3282-3289.
288. Pierce, B. and Z. Weng, *ZRANK: reranking protein docking predictions with an optimized energy function*. Proteins, 2007. **67**(4): p. 1078-86.

289. Andrusier, N., R. Nussinov, and H.J. Wolfson, *FireDock: fast interaction refinement in molecular docking*. Proteins, 2007. **69**(1): p. 139-59.
290. Jacobson, M.P., et al., *A hierarchical approach to all-atom protein loop prediction*. Proteins, 2004. **55**(2): p. 351-67.
291. Gront, D., et al., *Generalized fragment picking in Rosetta: design, protocols and applications*. PLoS one, 2011. **6**(8): p. e23294.
292. Zhang, C., S. Liu, and Y. Zhou, *Accurate and efficient loop selections by the DFIRE-based all-atom statistical potential*. Protein science : a publication of the Protein Society, 2004. **13**(2): p. 391-9.
293. Pegg, S.C., et al., *Leveraging enzyme structure-function relationships for functional inference and experimental design: the structure-function linkage database*. Biochemistry, 2006. **45**(8): p. 2545-55.
294. Maurer, S.M., A. Rai, and A. Sali, *Finding Cures for Tropical Diseases: Is Open Source an Answer?* PLoS Medicine, 2004. **1**(3): p. e56.
295. Orti, L., et al., *A kernel for open source drug discovery in tropical diseases*. PLoS Negl Trop Dis, 2009. **3**(4): p. e418.
296. Aguerro, F., et al., *Genomic-scale prioritization of drug targets: the TDR Targets database*. Nat Rev Drug Discov, 2008. **7**(11): p. 900-907.
297. Martinez-Jimenez, F., et al., *Target prediction for an open access set of compounds active against Mycobacterium tuberculosis*. PLoS Comp Biol, 2013. **9**(10): p. e1003253.
298. Sampathkumar, P., et al., *Structure of a putative BenF-like porin from Pseudomonas fluorescens Pf-5 at 2.6 Å resolution*. Proteins:Struct Funct Bioinform, 2010. **78**: p. 3056-3062.
299. Pieper, U., et al., *Target selection and annotation for the structural genomics of the amidohydrolase and enolase superfamilies*. J Struct Funct Genom, 2009. **10**: p. 107-125.
300. Karchin, R., et al., *LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources*. Bioinformatics, 2005. **21**(12): p. 2814-2820.
301. Fiser, A. and A. Sali, *ModLoop: automated modeling of loops in protein structures*. Bioinformatics, 2003. **19**(18): p. 2500-2501.
302. Davis, F. and A. Sali, *PIBASE: a comprehensive database of structurally defined protein interfaces*. Bioinformatics, 2005. **21**(9): p. 1901-1907.
303. Marti-Renom, M.A., et al., *DBAli tools: mining the protein structure space*. Nucleic Acids Res, 2007. **35**(Web Server issue): p. W393-397.
304. Marti-Renom, M.A., V.A. Ilyin, and A. Sali, *DBAli: a database of protein structure alignments*. Bioinformatics, 2001. **17**(8): p. 746-747.
305. Pieper, U., et al., *MODBASE, a database of annotated comparative protein structure models and associated resources*. Nucleic Acids Res, 2009. **37**(Database issue): p. D347-354.

306. Pieper, U., et al., *MODBASE: a database of annotated comparative protein structure models and associated resources*. Nucleic Acids Res, 2006. **34**(Database issue): p. D291-295.
307. Braberg, H., et al., *SALIGN: A webserver for alignment of multiple protein sequences and structures*. Bioinformatics, 2012. **15**: p. 2072-2073.
308. Schneidman-Duhovny, D., M. Hammel, and A. Sali, *FoXS: A Web Server for Rapid Computation and Fitting of SAXS Profiles*. Nucleic Acids Res, 2010. **38**: p. 541-544.
309. Ueda, Y., H. Taketomi, and N. Go, *Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three dimensional lattice model of lysozyme. .* Biopolymers, 1978. **17**: p. 1531-1548.
310. Okazaki, K., et al., *Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(32): p. 11844-9.
311. Whitford, P.C., et al., *An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields*. Proteins, 2009. **75**(2): p. 430-41.
312. Weinkam, P., et al., *Impact of mutations on the allosteric conformational equilibrium*. J Mol Biol, 2013. **425**: p. 647-661.
313. Petoukhov, M.V. and D.I. Svergun, *Analysis of X-ray and neutron scattering from biomacromolecular solutions*. Curr Opin Struct Biol, 2007. **17**(5): p. 562-71.
314. Hura, G.L., et al., *Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS)*. Nat Methods, 2009. **6**(8): p. 606-12.
315. Svergun, D., C. Barberato, and M.H.J. Koch, *CRY SOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates*. Journal of Applied Crystallography, 1995. **28**: p. 768-773.
316. Poitevin, F., et al., *AquaSAXS: a web server for computation and fitting of SAXS profiles with non-uniformly hydrated atomic models*. Nucleic acids research, 2011. **39**(Web Server issue): p. W184-9.
317. Liu, H., et al., *Computation of small-angle scattering profiles with three-dimensional Zernike polynomials*. Acta crystallographica. Section A, Foundations of crystallography, 2012. **68**(Pt 2): p. 278-85.
318. Ravikumar, K.M., W. Huang, and S. Yang, *Fast-SAXS-pro: a unified approach to computing SAXS profiles of DNA, RNA, protein, and their complexes*. The Journal of chemical physics, 2013. **138**(2): p. 024112.
319. Schneidman-Duhovny, D., et al., *Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments*. Biophysical journal, 2013. **105**(4): p. 962-74.
320. Pelikan, M., G.L. Hura, and M. Hammel, *Structure and flexibility within proteins as identified through small angle X-ray scattering*. General physiology and biophysics, 2009. **28**(2): p. 174-89.

321. Hura, G.L., et al., *Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS)*. Nature methods, 2009. **6**(8): p. 606-12.
322. Grant, T.D., et al., *Small angle X-ray scattering as a complementary tool for high-throughput structural studies*. Biopolymers, 2011. **95**(8): p. 517-30.
323. Konarev, P.V., et al., *PRIMUS: a Windows PC-based system for small-angle scattering data analysis*. Journal of Applied Crystallography, 2003. **36**: p. 1277-1282.
324. Shoichet, B.K. and I.D. Kuntz, *Protein docking and complementarity*. Journal of molecular biology, 1991. **221**(1): p. 327-46.
325. Pieper, U., et al., *Coordinating the impact of structural genomics on the human α -helical transmembrane proteome*. Nat Struct Mol Biol, 2013. **20**: p. 135-138.
326. Johnson, M., et al., *NCBI BLAST: a better web interface*. Nucleic acids research, 2008. **36**(Web Server issue): p. W5-9.
327. Ward, J.J., et al., *The DISOPRED server for the prediction of protein disorder*. Bioinformatics, 2004. **20**(13): p. 2138-9.
328. Dosztanyi, Z., et al., *IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content*. Bioinformatics, 2005. **21**(16): p. 3433-4.
329. Steentoft, C., et al., *Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology*. The EMBO journal, 2013. **32**(10): p. 1478-88.
330. Gupta, R., et al., *Scanning the available Dictyostelium discoideum proteome for O-linked GlcNAc glycosylation sites using neural networks*. Glycobiology, 1999. **9**(10): p. 1009-22.
331. Cormier, C.Y., et al., *PSI: Biology-materials repository: a biologist's resource for protein expression plasmids*. Journal of structural and functional genomics, 2011. **12**(2): p. 55-62.
332. Lamesch, P., et al., *hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes*. Genomics, 2007. **89**(3): p. 307-15.
333. Temple, G., et al., *The completion of the Mammalian Gene Collection (MGC)*. Genome research, 2009. **19**(12): p. 2324-33.
334. Seal, R.L., et al., *genenames.org: the HGNC resources in 2011*. Nucleic acids research, 2011. **39**(Database issue): p. D514-9.
335. Gifford, L.K., et al., *The Protein Structure Initiative Structural Biology Knowledgebase Technology Portal: a structural biology web resource*. Journal of structural and functional genomics, 2012. **13**(2): p. 57-62.
336. Nakai, K., A. Kidera, and M. Kanehisa, *Cluster analysis of amino acid indices for prediction of protein structure and function*. Protein engineering, 1988. **2**(2): p. 93-100.
337. Tomii, K. and M. Kanehisa, *Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins*. Protein engineering, 1996. **9**(1): p. 27-36.

338. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices*. Journal of molecular biology, 1999. **292**(2): p. 195-202.
339. Breiman, L. and E. Schapire, *Random forests*. Machine Learning, 2001. **45**(1): p. 5-32.
340. Benson, D.A., et al., *GenBank*. Nucleic Acids Res, 2010. **38**(Database issue): p. D46-51.
341. Kraulis, P.J., *MOLSCRIPT: A Program to produce both detailed and schematic plots of Protein Structures*. Journal of Applied Crystallography, 1991. **24**: p. 946-950.
342. Merritt, E.A. and D.J. Bacon, *Raster3D: Photorealistic Molecular Graphics*. Methods in Enzymology, 1997. **277**: p. 505-524.
343. Capra, J.A., et al., *Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure*. PLoS computational biology, 2009. **5**(12): p. e1000585.
344. Henrick, K., et al., *Remediation of the protein data bank archive*. Nucleic Acids Res, 2008. **36**(Database issue): p. D426-33.
345. Wu, C.H., et al., *The Universal Protein Resource (UniProt): an expanding universe of protein information*. Nucleic Acids Res, 2006. **34**(Database issue): p. D187-91.
346. Rhead, B., et al., *The UCSC Genome Browser database: update 2010*. Nucleic Acids Res, 2010. **38**(Database issue): p. D613-9.
347. Hunter, S., et al., *InterPro: the integrative protein signature database*. Nucleic Acids Res, 2009. **37**(Database issue): p. D211-5.
348. Klein, T.E., et al., *Integrating genotype and phenotype information: an overview of the PharmGKB project*. Pharmacogenetics Research Network and Knowledge Base. Pharmacogenomics J, 2001. **1**(3): p. 167-70.
349. Haas, J., et al., *The Protein Model Portal--a comprehensive resource for protein structure and model information*. Database : the journal of biological databases and curation, 2013. **2013**: p. bat031.
350. Schwede, T., et al., *Outcome of a workshop on applications of protein models in biomedical research*. Structure, 2009. **17**(2): p. 151-159.
351. Bairoch, A., et al., *The Universal Protein Resource (UniProt)*. Nucleic Acids Res, 2005. **33**(Database issue): p. D154-9.
352. Giglia, E., *New year, new PubMed*. Eur J Phys Rehabil Med, 2009. **45**(1): p. 155-9.
353. Hammel, M., et al., *XRCC4 protein interactions with XRCC4-like factor (XLF) create an extended grooved scaffold for DNA ligation and double strand break repair*. The Journal of biological chemistry, 2011. **286**(37): p. 32638-50.
354. Kotik-Kogan, O., et al., *Structural analysis reveals conformational plasticity in the recognition of RNA 3' ends by the human La protein*. Structure, 2008. **16**(6): p. 852-62.
355. Balbo, J., et al., *The shape of protein crowders is a major determinant of protein diffusion*. Biophys J, 2013. **104**(7): p. 1576-84.

356. Elcock, A.H., et al., *Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin*. J Mol Biol, 1999. **291**(1): p. 149-62.
357. Feldman-Salit, A., et al., *A mechanistic model of the cysteine synthase complex*. J Mol Biol, 2009. **386**(1): p. 37-59.
358. Gabdouliline, R.R., et al., *Concerted simulations reveal how peroxidase compound III formation results in cellular oscillations*. Biophys J, 2003. **85**(3): p. 1421-8.
359. Gabdouliline, R.R. and R.C. Wade, *Protein-protein association: investigation of factors influencing association rates by brownian dynamics simulations*. J Mol Biol, 2001. **306**(5): p. 1139-55.
360. Mereghetti, P., R.R. Gabdouliline, and R.C. Wade, *Brownian dynamics simulation of protein solutions: structural and dynamical properties*. Biophys J, 2010. **99**(11): p. 3782-91.
361. Mereghetti, P., et al., *Diffusion and association processes in biological systems: theory, computation and experiment*. BMC Biophys, 2011. **4**: p. 2.
362. Mereghetti, P., M. Martinez, and R.C. Wade, *Long range Debye-Huckel correction for computation of grid-based electrostatic forces between biomacromolecules*. BMC Biophys, 2014. **7**: p. 4.
363. Mereghetti, P. and R.C. Wade, *Diffusion of hydrophobin proteins in solution and interactions with a graphite surface*. BMC Biophys, 2011. **4**: p. 9.
364. Mereghetti, P. and R.C. Wade, *Atomic detail brownian dynamics simulations of concentrated protein solutions with a mean field treatment of hydrodynamic interactions*. J Phys Chem B, 2012. **116**(29): p. 8523-33.
365. Motiejunas, D., et al., *Protein-protein docking by simulating the process of association subject to biochemical constraints*. Proteins, 2008. **71**(4): p. 1955-69.
366. Spaar, A., et al., *Diffusional encounter of barnase and barstar*. Biophys J, 2006. **90**(6): p. 1913-24.
367. Elcock, A.H., *A molecule-centered method for accelerating the calculation of hydrodynamic interactions in Brownian dynamics simulations containing many flexible biomolecules*. J Chem Theory Comput, 2013. **9**(7): p. 3224-3239.
368. Frembgen-Kesner, T. and A.H. Elcock, *Absolute protein-protein association rate constants from flexible, coarse-grained Brownian dynamics simulations: the role of intermolecular hydrodynamic interactions in barnase-barstar association*. Biophys J, 2010. **99**(9): p. L75-7.
369. McGuffee, S.R. and A.H. Elcock, *Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm*. PLoS Comput Biol, 2010. **6**(3): p. e1000694.
370. Gabdouliline, R.R. and R.C. Wade, *Brownian dynamics simulation of protein-protein diffusional encounter*. Methods, 1998. **14**(3): p. 329-41.
371. Dlugosz, M., P. Zielinski, and J. Trylska, *Brownian dynamics simulations on CPU and GPU with BD_BOX*. J Comput Chem, 2011. **32**(12): p. 2734-44.

372. Greives, N. and H.X. Zhou, *BDflex: a method for efficient treatment of molecular flexibility in calculating protein-ligand binding rate constants from brownian dynamics simulations*. J Chem Phys, 2012. **137**(13): p. 135105.
373. Huber, G.A. and J.A. McCammon, *Browndye: A Software Package for Brownian Dynamics*. Comput Phys Commun, 2010. **181**(11): p. 1896-1905.
374. Fitch, B.G., et al. *Blue Matter: Approaching the Limits of Concurrency for Classical Molecular Dynamics*. in *SC 2006 Conference, Proceedings of the ACM/IEEE*. 2006.
375. Bowers, K.J., et al. *Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters*. in *SC 2006 Conference, Proceedings of the ACM/IEEE*. 2006.
376. Rodrigues, C.I., et al., *GPU acceleration of cutoff pair potentials for molecular modeling applications*, in *Proceedings of the 5th conference on Computing frontiers2008*, ACM: Ischia, Italy. p. 273-282.
377. Komeiji, Y., et al., *Fast and accurate molecular dynamics simulation of a protein using a special-purpose computer*. Journal of Computational Chemistry, 1997. **18**(12): p. 1546-1563.
378. Shaw, D.E., et al., *Anton, a special-purpose machine for molecular dynamics simulation*, in *Proceedings of the 34th annual international symposium on Computer architecture2007*, ACM: San Diego, California, USA. p. 1-12.
379. Taiji, M., et al. *Protein Explorer: A Petaflops Special-Purpose Computer System for Molecular Dynamics Simulations*. in *Supercomputing, 2003 ACM/IEEE Conference*. 2003.
380. Wolff, D. and W.G. Rudd, *Tabulated potentials in molecular dynamics simulations*. Computer Physics Communications, 1999. **120**(1): p. 20-32.
381. Perutz, M.F., *X-ray analysis of hemoglobin*. Science, 1963. **140**(3569): p. 863-9.
382. Eaton, W.A. and J. Hofrichter, *Sickle cell hemoglobin polymerization*. Adv Protein Chem, 1990. **40**: p. 63-279.
383. Bluemke, D.A., et al., *Structural analysis of polymers of sickle cell hemoglobin. II. Sickle hemoglobin macrofibers*. J Mol Biol, 1988. **199**(2): p. 333-48.
384. Bluemke, D.A., et al., *The three-dimensional structure of sickle hemoglobin macrofibers*. Prog Clin Biol Res, 1987. **240**: p. 31-46.
385. Carragher, B., et al., *Structural analysis of polymers of sickle cell hemoglobin. I. Sickle hemoglobin fibers*. J Mol Biol, 1988. **199**(2): p. 315-31.
386. Potel, M.J., et al., *Macrofiber structure and the dynamics of sickle cell hemoglobin crystallization*. J Mol Biol, 1984. **177**(4): p. 819-39.
387. Wang, Z., et al., *Polymerization of deoxy-sickle cell hemoglobin in high-phosphate buffer*. J Struct Biol, 2000. **131**(3): p. 197-209.
388. Unni, S., et al., *Web servers and services for electrostatics calculations with APBS and PDB2PQR*. J Comput Chem, 2011. **32**(7): p. 1488-91.

389. Cornell, W.D., et al., *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules*. Journal of the American Chemical Society, 1995. **117**(19): p. 5179-5197.
390. Yershova, A., et al., *Generating Uniform Incremental Grids on $SO(3)$ Using the Hopf Fibration*. The International Journal of Robotics Research, 2010. **29**(7): p. 801-812.
391. Adams, L.R. and I. Fatt, *The diffusion coefficient of human hemoglobin at high concentrations*. Respiration Physiology, 1967. **2**(3): p. 293-301.
392. Iniesta, A. and J. García de la Torre, *A second-order algorithm for the simulation of the Brownian dynamics of macromolecular models*. The Journal of Chemical Physics, 1990. **92**(3): p. 2015-2018.
393. Gonnet, P., *Pairwise verlet lists: combining cell lists and verlet lists to improve memory locality and parallelism*. J Comput Chem, 2012. **33**(1): p. 76-81.
394. Bentley, J.L., *Multidimensional binary search trees used for associative searching*. Commun. ACM, 1975. **18**(9): p. 509-517.
395. Franklin, J. and S. Doniach, *Adaptive time stepping in biomolecular dynamics*. J Chem Phys, 2005. **123**(12): p. 124909.
396. Knowles, P.F., *Biophysical chemistry: Part II 'Techniques for the study of biological structure and function': By CR Cantor and PR Schimmel. pp 503. WH Freeman and Co, Oxford. 1980. £20.70/£10.60 (paperback) ISBN 0-7167-1189-3/0-7167-1190-7 (paperback)*. Biochemical Education, 1981. **9**(4): p. 157-157.
397. Ando, T., E. Chow, and J. Skolnick, *Dynamic simulation of concentrated macromolecular solutions with screened long-range hydrodynamic interactions: algorithm and limitations*. J Chem Phys, 2013. **139**(12): p. 121922.
398. Kohonen, T., *Self-organized formation of topologically correct feature maps*. Biological Cybernetics, 1982. **43**(1): p. 59-69.
399. Mu, X.Q., L. Makowski, and B. Magdoff-Fairchild, *Analysis of the stability of hemoglobin S double strands*. Biophys J, 1998. **74**(1): p. 655-68.
400. Benesch, R.E., et al., *alpha Chain mutations with opposite effects on the gelation of hemoglobin S*. J Biol Chem, 1979. **254**(17): p. 8169-72.
401. Weinkam, P. and A. Sali, *Mapping polymerization and allostery of hemoglobin S using point mutations*. J Phys Chem B, 2013. **117**(42): p. 13058-68.
402. Wang, F. and D.P. Landau, *Efficient, multiple-range random walk algorithm to calculate the density of states*. Phys Rev Lett, 2001. **86**(10): p. 2050-3.

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature Natalia Khuri Date 12/19/2014