Molecular Pharmacology Fast Forward. Published on April 9, 2013 as doi: 10.1124/mol.112.084152

Title page

Structure–based identification of OATP1B1/3 inhibitors

Tom De Bruyn, Gerard J.P. Van Westen, Adriaan P. IJzerman, Bruno Stieger, Peter de Witte, Patrick F. Augustijns and Pieter P. Annaert

Drug Delivery and Disposition, Department of Pharmaceutical and Pharmacological Sciences, KU Leuven, O&N2, Herestraat 49-bus-921, 3000 Leuven, Belgium (TDB, PFA, PPA)

Division of Medical Chemistry, Leiden/Amsterdam Center for Drug Research, Leiden, Einsteinweg 55, 2333 CC, The Netherlands (GVW, API)

Department of Clinical Pharmacology and Toxicology, University Hospital, 8091 Zürich, Switzerland (BS)

Laboratory for Molecular Biodiscovery, Department of Pharmaceutical and Pharmacological Sciences, KU Leuven, O&N2, Herestraat 49-bus-824, 3000 Leuven, Belgium (PDW)

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Running title page

Running title: Structure-based identification of OATP1B inhibitors

Corresponding Author:

Pieter Annaert, PhD
Drug Delivery and Disposition
Department of Pharmaceutical and Pharmacological Sciences
KU Leuven, O&N2
Herestraat 49 – bus 921
B-3000 Leuven, Belgium.
Tel: 32-16-330303
Fax: 32-16-330305
E-mail: pieter.annaert@pharm.kuleuven.be

Document statistics:

Number of text pages: 21
Number of tables: 2
Number of figures: 7
Number of references: 27
Number of words in Abstract: 238
Number of words in Introduction: 594
Number of words in Discussion: 1.509
Supplemental material online: 13 pages with 7 tables and 2 figures

List of non-standard abbreviations:

CHO, Chinese hamster ovary; DDI, drug-drug interactions; DMEM, Dulbecco's modified eagle's medium; FBS, fetal bovine serum; HBSS, Hank's balanced salt solution; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; OATP/Oatp, Organic anion transporting polypeptide (Human/Rat); PBS, phosphate buffered saline; QSAR, quantitative structure activity relationship; SLC, Solute-linked carrier

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Abstract

Several recent studies show that inhibition of the hepatic transport proteins organic-anion transporting polypeptide 1B1 (OATP1B1) and 1B3 (OATP1B3) can result in clinically relevant drug-drug interactions (DDI). To avoid late-stage development drug failures due to OATP1B-mediated DDI, predictive *in vitro* and *in silico* methods should be implemented at an early stage of the drug candidate evaluation process.

In the present study, we first developed a high-throughput *in vitro* transporter inhibition assay for the OATP1B subfamily. A total of 2000 compounds were tested as potential modulators of the uptake of the OATP1B substrate sodium fluorescein, in OATP1B1 or 1B3-transfected CHO cells. At an equimolar substrate-inhibitor concentration of 10 µM, 212 and 139 molecules were identified as OATP1B1 and OATP1B3 inhibitors, respectively (min 50 % inhibition). For 69 compounds, previously not identified as OATP1B inhibitors, concentration dependent inhibition was also determined, yielding *Ki* values ranging from 0.06 to 6.5 µM. Based on these *in vitro* data, we subsequently developed a proteochemometrics-based *in silico* model, which predicted OATP1B inhibitors in the test group (20 % of the dataset) with high specificity (86 %) and sensitivity (78 %). Moreover, several physicochemical compound properties and substructures related to OATP1B1/1B3 inhibition or inactivity were identified. Finally, model performance was prospectively verified with a set of 54 compounds not included in the original dataset. This validation indicated that 80 % and 74 % of the compounds were correctly classified for OATP1B1 and OATP1B3 inhibition, respectively.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Introduction

Over the past decade, increasing evidence has established the role of transport proteins in pharmacokinetic processes and drug-drug interactions. The organic anion transporting polypeptides 1B1 and 1B3, OATP1B1 (*SLCO1B1*) and OATP1B3 (*SLCO1B3*), are uptake transporters predominantly expressed at the basolateral membrane of human hepatocytes. OATP1B1 and OATP1B3 exhibit broad and overlapping substrate specificities for endogenous and exogenous compounds including thyroid hormones, bile acids, repaglinide, and several statins (Hagenbuch and Gui, 2008). The importance of these hepatic transporters has been illustrated by recent studies showing that OATP1B-mediated transport can be the rate-determining step of hepatobiliary drug clearance (Fenner et al., 2012). Moreover, OATP1B inhibition or induction may be the underlying mechanism of clinically relevant drug-drug interactions (Müller and Fromm, 2011; Shitara, 2011). Therefore, the assessment of OATP1B-mediated drug-drug interactions (DDI) has become a critical aspect of early drug development, as recognized by the International Transporter Consortium (Giacomini et al., 2010). This review acknowledged OATP1B1 and OATP1B3 as two of the seven clinically most relevant transporters and provides recommendations regarding the preferred *in vitro* methods to be used for evaluating drug interactions mediated by these transporters. In addition to their role in hepatic drug disposition and DDI, recent findings have suggested that OATPs (in particular OATP1B3) may have an important pathophysiological role in certain pancreatic (Kounnis et al., 2011) and ovarian cancers (Svoboda et al., 2011). Therefore, OATP1B modulators could form a new class of potential therapeutic agents.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Identification (and kinetic characterization) of OATP ligands early on in the drug evaluation process has thus become a prerequisite for successful drug development. In vitro methods for assessing OATP1B-mediated hepatic drug uptake based on primary human hepatocytes are generally predictive of the in vivo situation but remain resource-intensive. Hence, several studies have successfully relied on the use of higher throughput assays with transfected HEK293 or CHO cells in order to investigate OATP1B inhibition and transporter kinetics (Baldes et al., 2006; Annaert et al., 2010; Bednarczyk, 2010; Gui et al., 2010). These studies led to the identification of new OATP1B inhibitors and/or the characterization of the inhibition potential of existing drugs (e.g. HIV protease inhibitors) (Annaert et al., 2010). By supporting the evaluation of a relative large number of compounds, these in vitro approaches also enable gaining valuable information regarding the molecular structural properties of OATP ligands. Statistical Quantitative Structure Activity Relationship (QSAR) modeling is commonly applied to computationally derive chemical features shared by compounds which are active as opposed to chemical features shared by compounds which are inactive. Similarly, as reviewed by van Westen et al., (2011b), proteochemometric modeling additionally takes into account structural protein features (this can be either the full protein, or a selection of residues usually the binding pocket) (Lapinsh et al., 2001; Prusis et al., 2006; van Westen et al., 2013).

Previously, we characterized the *in vitro* hepatic uptake mechanisms of sodium fluorescein and demonstrated that sodium fluorescein can be used as a probe substrate to evaluate OATP1B1/3-mediated transport (De Bruyn et al., 2011). Based on this information, we presently report the development of a high throughput *in vitro* assay and applied it to evaluate the OATP1B1 and OATP1B3 inhibitory potential of a library of

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

2000 compounds. Next, we established an *in silico* proteochemometric model to predict OATP1B interaction potential of compounds and to identify chemical properties and substructures related to OATP1B1 and/or OATP1B3 inhibition. Our results demonstrate that this integrated *in vitro* and *in silico* approach leads to a particularly useful *in silico* algorithm for direct structure-based identification of (drug) compounds modulating OATP1B activity.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Materials and Methods

Reagents: The Spectrum Collection (2000 compounds, supplied as 10 mM Me_2SO solution) was purchased from MicroSource Discovery Systems Inc. (Gaylordsville, CT, USA). Dulbecco's modified eagle's medium (DMEM) and geneticin G418 were purchased from Invitrogen (Paisley, UK). L-glutamine, penicillin-streptomycin mixture (contains 10.000 IU potassium penicillin and 10.000 IU streptomycin sulfate per ml in 0.85% saline), fetal bovine serum (FBS), Trypsin EDTA, Hank's balanced salt solution (HBSS) and phosphate buffered saline 1x (PBS) were purchased from Lonza SPRL (Verviers, Belgium). Sodium fluorescein was obtained from UCB (Leuven, Belgium) and HEPES was purchased from MP Biochemical (Illkirch, France). Triton X-100 and sodium butyrate were purchased from Sigma-Aldrich (Schnelldorf, Germany).

In vitro assessment of OATP1B inhibitory potential.

Cell culture. Wild-type, OATP1B1- and 1B3-transfected CHO cells were cultured at passage 45 to 60, as described previously (Treiber et al., 2007). The culture medium consisted of DMEM containing 1 g/1 D-glucose, 1 mM L-glutamine, 25 mM HEPES and 110 mg/l sodium pyruvate, supplemented with 10 % FBS, 50 µg/ml L-proline, 100 IU/ml penicillin and 100 IU/ml streptomycin. The culture medium of the transfected cell lines additionally contained 500 µg/ml geneticin to maintain selection. CHO cells were grown in 75 cm^2 T-Flasks at 37°C in an atmosphere of 5% CO_2 and 90 % relative humidity and were sub-cultured every 3 days. For high-throughput uptake experiments, wild-type CHO cells were seeded in black 96-well cell culture plates with a transparent bottom (Greiner-Bio-One, Wemmel, Belgium) at a density of 8,000 cells/well. Culture medium was

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

replaced every other day and uptake experiments were performed on day 4 after seeding. One day before the experiment, cells were additionally treated with 5 mM sodium butyrate to induce gene expression.

Transport studies. Cells were washed twice with 0.2 ml/well of pre-warmed uptake buffer (HBSS with 10 mM HEPES, pH 7.4) and pre-incubated for 10 min at 37°C with 100 µL of uptake buffer. In experiments in which inhibitors were investigated, cells were pre-incubated with 100 µL of a double-concentrated inhibitor solution. Uptake assays were initiated by adding 100 µL of a double-concentrated sodium fluorescein solution. To determine the uptake kinetics in OATP1B1 or OATP1B3-transfected cells, sodium fluorescein was incubated at 8 concentrations ranging from 1 to 80 µM. The inhibitory potential of all 2000 compounds from The Spectrum Collection library was evaluated at an equimolar substrate-inhibitor concentration of 10 µM. Uptake experiments were stopped after 3 min by removing the uptake solution and rinsing the cells three times with ice-cold uptake buffer. Subsequently, cells were lysed with 100 µl of 0.5 % Triton X solution (in PBS) and placed on a plate shaker for 30 min at room temperature. Cell lysates were analyzed by fluorescence spectroscopy (ex 490 nm; em 524 nm) in a Tecan Infinite 200 plate reader (Tecan Benelux, Mechelen, Belgium). Uptake rates were normalized for protein content, which was measured using a BCA Protein assay kit (Pierce Chemical, Rockford, IL).

In vitro data analysis. Net uptake values were obtained by subtracting uptake in wild-type CHO cells from total uptake at 37°C in transfected cells. Cellular fluorescein uptake kinetics were determined by fitting the Michaelis-Menten equation to net uptake values:

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

$$v = \frac{V_{max} \cdot C}{K_m + C}$$

For inhibition experiments, data were expressed as uptake of sodium fluorescein in presence of the inhibitor compared to the control cellular uptake (no inhibition).

The *Inhibitory Effect Emax*-model was used to describe the concentration-dependent inhibitory effect by various inhibitors in transfected CHO cells:

$$E = E_{max} - \left[(E_{max} - E_0) \cdot \frac{C^{\,n}}{C^{\,n} + IC_{50}{}^{n}} \right]$$

with E representing the cellular substrate uptake, E_{max} the substrate uptake without inhibitor (≤ 100 %), E_0 the substrate uptake at the maximum inhibitory effect of inhibitor (≥ 0 %) and ($E_{max} - E_0$) the maximum inhibitory effect. The best fits of the above equations to the individual uptake and inhibition data sets were obtained by non-linear regression analysis using the Solver tool in Microsoft Excel 2007.

For the estimation of K_i from IC_{50}, assuming competitive inhibition, the Cheng-Prusoff equation was used:

$$K_i = \frac{IC_{50}}{1 + \frac{S}{K_m}}$$

with IC_{50} representing the concentration of the inhibitor producing a 50 % inhibition, S the substrate concentration and K_m the Michaelis constant of the substrate.

The data analysis tool in Microsoft Excel 2007 ($p < 0.001$) was used to evaluate statistical differences of the correlation between remaining sodium fluorescein uptake and the computationally predicted inactivity correlation value for all compounds in the prospective validation set.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

***In silico* modeling for the prediction of OATP1B inhibition.**

Dataset. The dataset consisted of *in* vitro OATP modulation data for 2000 compounds (The Spectrum Collection, MicroSource Discovery Systems, Inc.). For all test compounds, the assay activity value was expressed as a percent of the control sodium fluorescein uptake (after correction for uptake in wild-type CHO cells). When the residual activity was lower than 50 %, compounds were deemed 'Active' (able to inhibit the uptake of sodium fluorescein) and *vice versa*.

The predictivity of the *in silico* model was prospectively verified with an external test set of 54 compounds that were not included in The Spectrum Collection.

Molecular and Protein Descriptors. The compounds were standardized and their ionization state was determined at pH 7.4 (identical to assay conditions). Furthermore, hydrogens were added and 2-dimensional coordinates were calculated. Subsequently, compounds were converted to circular fingerprints (FCFP_6) and a number of numeric descriptors were calculated. These included molecular weight, number of hydrogen bond donors/acceptors, and polar surface area (Supplemental Table 1). Standardization and descriptor calculation steps were performed in Pipeline Pilot 8.5 (Accelrys Software Inc, San Diego, CA).

Proteins were described based on their sequence (Supplemental Table 2). Sequences of OATP1B1, OATP1B3, OATP1A2, and OATP2B1 were aligned with ClustalW (Slow alignment, Gap Open 4, Gap Extend 4, Supplemental Information Alignment 1). The latter two OATP isoforms were included to increase the quality of the alignment. From this alignment, amino acids previously shown to be important for OATP activity (Meier-Abt et al., 2005; Mandery et al., 2011) were selected and translated to a descriptor based

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

on Z-scales or a feature-based ProtFP descriptor (Lapinsh et al., 2002; van Westen et al., 2011a). Supplemental Figure 1 shows the similarity of the proteins as described by the Z-scales descriptors in a two dimensional plot.

Proteochemometric Modeling (PCM). Pipeline Pilot 8.5 (Accelrys Software Inc, San Diego, CA) was used for PCM modeling. Random forest models were trained using the R package randomForest as described previously (Liaw and Wiener, 2002). The optimal number of trees was determined to be 25 and for each tree, the optimal number of allowed descriptors per split was set at a fraction of 0.5 from the total number of descriptors. Model performance was estimated with the out-of-bag (oob) data, external validation, and prospective validation.

For all trained classification models, performance was estimated using the sensitivity, specificity, the Matthews correlation coefficient (Matthews, 1975), and the fraction correctly predicted compounds (Supplemental Table 3).

Two final models were selected, i.e. a 2-class model (distinguishing between compounds either 'Active' or 'Inactive'; model 8 in Supplemental Table 3) and a 4-class model (distinguishing between 'Active_1B1', 'Active_1B3', 'Inactive_1B1', or 'Inactive_1B3'; model 17 in Supplemental Table 3).

Descriptor interpretation. In order to interpret the random forest models, descriptors used in the model training were converted. The numerical compound descriptors were binned into 4 classes based on their distribution in the dataset. The 'low' bin contained the lower 20 % of the frequency histogram for each descriptor, the 'high' bin contained the higher 20 % of the frequency histogram, and the remaining 60 % was equally

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

distributed between ‘medium_low’ and ‘medium_high’. For further details per descriptor see Supplemental Table 4.

For the circular fingerprints, the ‘Fingerprints to properties’ component from Pipeline Pilot 8.5 was used to make a selection; from a total of 16,341 chemical features present, 1,308 relevant features were encoded for presence or absence. We selected the features that were most equally/unequally distributed between actives and inactives (512/512 features), supplemented with a random selection of 2 % of all features (297). Thirteen features overlapped and were removed.

For the protein descriptors, each Z-scale (representing a physicochemical property) was binned similar to the molecular descriptors. As a consequence, every amino acid obtained a score for Hydrophobicity (Z1), Size/Polarizability (Z2), Polarity (Z3), Electronegativity (Z4), and Electrophilicity (Z5). Presence or absence of each of those scores was encoded (Supplemental Table 5). The bins were selected such that each amino acid was uniquely identifiable via these binned values.

To estimate the significance of each descriptor, the model provided 2 importance values: Accuracy and Gini. The accuracy value was calculated by omitting one descriptor and calculating the deterioration of the model’s ability to correctly classify compounds. The Gini measure for a descriptor estimates the ability to differentiate between activity classes when the data set is split using this descriptor. Finally, the correlation of each descriptor with each activity class allowed interpretation to see whether a descriptor is positively or negatively correlated with OATP1B activity.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Prospective validation. A total of 54 compounds (not present in the original compound library) were selected based on in-house availability and solubility, and computationally evaluated for OATP1B1 and OATP1B3 inhibition with the final random forest model (model 8). Using the 2 class *in silico* model, a total of 30 and 23 compounds were predicted to be inhibitors of OATP1B1 and OATP1B3, respectively. Activity/inactivity correlation values, reflecting the potency of the prediction, were directly obtained from the Pipeline Pilot software output. Next, we experimentally determined the OATP1B1 and OATP1B3 inhibitory potential of all 54 compounds, following the described *in vitro* OATP1B inhibition assay and assessed the relationship between residual sodium fluorescein uptake and the inactivity correlation value.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Results

Uptake kinetics of sodium fluorescein in OATP1B1- and OATP1B3-transfected CHO cells

Concentration-dependent uptake of sodium fluorescein showed Michaelis-Menten type kinetics in both OATP1B1- and OATP1B3-transfected CHO cells (Figure 1). Kinetic parameters for net active uptake were 12.2 ± 4.9 μM and 33.2 ± 3.4 μM for K_m and 30.1 ± 3.8 and 195.6 ± 16.6 pmol/mg protein/min for V_{max}, respectively (n = 4). This corresponds to uptake clearance values of 2.5 and 5.9 μl/mg protein/min for OATP1B1- and OATP1B3- transfected CHO cells, respectively.

Inhibition of sodium fluorescein uptake by established OATP1B inhibitors in transfected CHO cells

The high-throughput *in vitro* screening method was validated by assessing the inhibitory potential of two established OATP1B inhibitors, rifampicin and darunavir (Annaert et al., 2010). Both inhibitors decreased the uptake of sodium fluorescein in a concentration-dependent manner with *Ki* values of 0.30 ± 0.05 μM and 0.22 ± 0.06 μM for rifampicin and 3.6 ± 1.3 μM and 6.7 ± 2.8 μM for darunavir in OATP1B1- and OATP1B3-transfected CHO cells, respectively (Figure 2).

High-throughput *in vitro* screening assay

A total of 2000 compounds were evaluated as potential inhibitors of the uptake of sodium fluorescein in OATP1B1 and OATP1B3-transfected CHO cells (Figure 3). At an equimolar substrate-inhibitor concentration of 10 μM, 212 and 139 compounds inhibited

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

the uptake of sodium fluorescein by at least 50 % in OATP1B1 and OATP1B-transfected CHO cells, respectively (Table 1 and Supplemental data). Some of these compounds (e.g. cyclosporin, rosuvastatin, sildenafil, and paclitaxel) had been previously reported to inhibit the transport of OATP substrates, while others were demonstrated to be OATP1B1 and/or OATP1B3 inhibitors for the first time (e.g. utilin). Secondly, we measured the inhibition of sodium fluorescein uptake by 69 of the most potent inhibitors at concentrations ranging from 0.1 to 10 µM. All selected compounds inhibited the uptake of sodium fluorescein (10 µM) in a concentration-dependent manner. The corresponding *Ki* values are listed in Table 1.

Proteochemometric modeling

Proteochemometric Models. The final 2-class model had an out-of-bag ROC (receiver operating characteristic) score of 0.86, a sensitivity of 0.78, a specificity of 0.86, and a percentage correctly classified compounds of 89.2 % (Supplemental Figure 2A and Supplemental Table 3). Identical models were trained per OATP leaving out the protein descriptor (and hence creating a QSAR model) to compare the performance to the state of the art as was done previously. The QSAR model reached a percentage correctly classified of 84.6 % on the same set. In an external validation experiment, leaving out 30 % of the compounds rather than a percentage of the bioactivity data points, similar results were seen with the PCM model reaching 84.2 % and the QSAR models reaching 85.2%. Due to the increased interpretability of the PCM models we preferred PCM over QSAR.
The final 4-class model had an ROC score of 0.94 (Supplemental Figure 2B), a sensitivity of 0.75, a specificity of 0.85 and a percentage correctly classified compounds

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

of 88.5 % (Supplemental Figure 2B and Supplemental Table 3). Sensitivity and specificity were calculated using the classification active or inactive (regardless of the protein isoform).

Important chemical properties of the 2-class model. The interpretation of the 2-class model revealed several physicochemical properties and substructures that correlate with OATP inhibition (Figure 4 and 5 and Supplemental Table 6). Based on descriptor importance 'gini', the most important compound property was a LogD between 3.4 and 7.5, followed by the absence of positively charged atoms and a number of ringbonds between 18 and 32. Furthermore, the presence of an anionic group (e.g. carboxylate group), a large molecular volume (> 200 $Å^3$) and a high number of hydrogen bond acceptors (> 5) are correlated with increased likelihood of OATP inhibition.

Several physiochemical properties and substructures are highly correlated with OATP1B inactivity (Figure 4 and 5 and Supplemental Table 6). The strongest negative correlations were found with substructures that are or can be positively charged at pH 7.4. Furthermore, small hydrophilic compounds with positively charged atoms and without negatively charged atoms are favored as inactives. More specifically, a total number of heavy atoms lower than 20, a molecular volume less than 200 $Å^3$, a molecular weight under 300, a calculated aqueous solubility > 0.1mmol/L, and number of rings < 4 were correlated with inactivity.

Important chemical properties of the 4-class model. In general, the 4-class model provides similar results to the 2-class model when considering the importance of

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

chemical features for activity or inactivity on both proteins. The most important feature for activity also included a group that could be negatively charged (based on descriptor importance 'gini'). Important features for inactivity were substructures that are or can be positively charged (at pH 7.4).

In addition, the 4-class model provides information regarding the selectivity for either OATP1B1 or OATP1B3. A low number of aromatic bonds (< 7) was found to be positively correlated with OATP1B1 activity and negatively with OATP1B3 inhibition. Conversely, a high LogD value (> 7.5) and a medium low number of hydrogen bond donors (3-4) were positively correlated with OATP1B3 activity.

Finally, we did identify some substructures favoring inhibition of OATP1B1 over OATP1B3 and *vice versa* (Figure 5C and D). A complete list of the important chemical properties is available electronically as Supplemental Table 6.

Important protein properties. As we constructed a PCM model on only two targets, only limited interpretation of the model with respect to the target can be performed (Supplemental Table 7). However, a follow-up of the current work should include a larger group of protein targets. Previously, we have shown that incorporating a larger number of proteins (8, 14 or even > 10.000 proteins) leads to improved performance and model interpretability, hence a similar approach could be followed here (van Westen et al., 2011a, 2012, 2013). A good starting point would be the ChEMBL database that lists bioactivity data for 22 OATP isoforms. Nine of these are human isoforms and include the four isoforms we used to construct our multiple sequence alignment (Gaulton et al., 2012).

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Prospective validation

The results of the prospective validation are depicted in Table 2 and Figure 6 and are uploaded as supplemental data. For OATP1B1, 22 out of 30 computationally predicted active compounds inhibited OATP1B1-mediated sodium fluorescein uptake by more than 50 %; 21 out of 24 compounds predicted negative, i.e. did not inhibit OATP1B1 by more than 50 % (Figure 6, panel A). For OATP1B3, 13 out of 23 computationally predicted active compounds inhibited OATP1B3-mediated sodium fluorescein by more than 50 % and 27 out of 31 negatives did not inhibit OATP1B3 (Figure 6, panel B). This corresponds to a sensitivity of 0.88 and 0.77 and a specificity of 0.72 and 0.73 for OATP1B1 and OATP1B3, respectively.

The correlations between the residual OATP1B1- and OATP1B3-mediated uptake of sodium fluorescein and the computationally generated inactivity correlation values for all 54 compounds of the prospective validation are represented in Figure 7A ($R^2 = 0.45$) and Figure 7B ($R^2 = 0.51$), respectively.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Discussion

We previously demonstrated that sodium fluorescein can be used as a probe substrate to identify xenobiotics modulating OATP1B-mediated transport in fluorescence-based *in vitro* transporter assays (De Bruyn et al., 2011). In the present study, we optimized this concept into a high-throughput *in vitro* assay that can be implemented in an early drug development setting to assess the risk for OATP1B-mediated DDI of new drug candidates. In addition, the extensive size and structural diversity of our *in vitro* data set allowed us to develop an accurate *in silico* proteochemometric model. In a final prospective validation, we confirmed the utility of this model to computationally predict OATP1B-mediated inhibition and to identify molecular characteristics important for OATP1B inhibition.

Fluorescence-based *in vitro* screening assays provide rapid and sensitive approaches for studying drug-transporter interactions (Bednarczyk, 2010; De Waart et al, 2010; Gui et al., 2010). The developed high-throughput assay here was used to assess the inhibitory potential of a very extensive and chemically diverse compound dataset. All 2000 test compounds were evaluated at an equimolar inhibitor-substrate concentration of 10 µM. In addition, compounds were only classified as inhibitors when the uptake of sodium fluorescein decreased by more than 50 % of the control value. Following this approach, only relatively potent (and more likely clinically relevant) OATP1B inhibitors are identified. In this respect, the experimental conditions used in our in vitro model differ from previously reported methodologies using an excess of inhibitor. Applying a high inhibitor/substrate concentration ratio might indeed lead to the selection of also those compounds with a relatively lower OATP inhibition potency (Badolo et al., 2010;

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Karlgren et al., 2012a). In the present study, rifampicin and darunavir inhibited sodium fluorescein uptake by more than 80 and 60 % of the control value (Figure 2), confirming that established OATP inhibitors are indeed detected with our in vitro assay.

Different computational approaches are possible given the available data. Structure-based approaches (docking) rely on a crystal structure, but this information is not available (Rognan, 2013). Hence, we opted for a proteochemometric approach that relies on (2D) structural sequence information for a given protein. Based on in vitro OATP inhibition data, we optimized a 2-class computational model to predict OATP1B inhibition. Although *Ki* values were determined for 69 potent OATP1B inhibitors (Table 1), lack of accurate parameter values (such as Ki) for non-inhibitors, precluded the use of these data for in silico model development. Therefore, classification data (active versus inactive) of all 2000 compounds were used for model building. In contrast to a regression model (based on *Ki* values), this classification approach allowed to include both active and inactive compounds and therefore to determine the characteristics of compounds either inhibiting or not inhibiting OATP1B.

The predictivity of the model was first estimated based on the out-of-bag error for which the original data set was divided in a training set (80 % of the dataset) and a 20 % test set. Using this approach, 89 % of the compounds in the test set were correctly classified (with QSAR models reaching 84 %). In an external validation experiment the PCM model classified 84% of the compounds correctly (85 % for QSAR models). Arguably, the only unbiased assessment of the reliability of any *in silico* model is a prospective validation, i.e. on application of the in silico model on a selection of compounds unknown to the model. Therefore, we composed a validation set of 54 compounds which were not present

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

in the original dataset. After computationally predicting the OATP1B inhibition potential, experimentally obtained *in vitro* results revealed that 80 % and 74 % of the compounds were correctly classified for OATP1B1 and OATP1B3, respectively (Table 2 and Figure 6). This prospective validation confirms the reliability and applicability of the computational model. Further, there appears to be a reasonable correlation (R^2 = 0.45-0.51) between the residual values for OATP1B-mediated uptake of sodium fluorescein and the computationally generated correlation values with inactivity for all 54 compounds. This suggests that this correlation value could be reliably used for *in silico* estimation of the residual sodium fluorescein uptake in the presence of a given putative inhibitor with known structure and physicochemical properties.

The *in silico* model was further applied to 51 compounds that were not present in the original data set and for which an OATP1B1 inhibition value was published by Karlgren et al. (2012a). Our model correctly classified 26 of the 35 OATP1B1 inhibitors (74 %) and 14 of the 16 non-inhibitors (88 %). The relative lower prediction for inhibitors confirmed the hypothesis that an excess of inhibitor compared to the substrate concentration is reflected in a higher number of compounds classified as (weakly) active which are not detected by our *in silico* model. The moderate sensitivity but high specificity of our *in silico* model towards published inhibition data results in a limited number of false positive results and thus, the model favors compounds with a potent OATP1B inhibitory potency. This could be an important advantage when searching for the most potent OATP1B inhibitors among a (very) large set of test compounds. However, to further improve the predictive value of our model, the data set supporting this model can further be expanded. The ChEMBL database, a large public database of

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

bioactivity data, also contains OATP bioactivity data for 9 human OATP proteins (including the 4 isoforms we used to generate our multiple sequence alignment). Because inclusion of more protein targets has previously been shown to lead to better prospective qualities (van Westen et al., 2012), we encourage a follow-up study to create a second generation OATP PCM model (Gaulton et al., 2012; Kramer et al., 2012).

The 2-class model was further expanded to a 4-class model for the descriptor interpretation. However, predictive information regarding compound properties extracted from both models was essentially comparable. The 4-class model additionally offered the advantage of pinpointing compound and protein properties important for inhibition of only one of the OATP1B isoforms.

The key descriptors important for an increase in OATP1B inhibition potency were the increase in lipophilicity and polar surface, the presence of anionic atoms or several hydrogen bond acceptors and the absence of cationic atoms. These findings are consistent with previous results from Chang et al. (2005) who constructed a pharmacophore model for OATP1B1 and concluded that lipophilicity and hydrogen bond-acceptor strength are important properties for OATP1B1 inhibition. Our results also confirm the conclusions drawn by Badolo et al. (2010) who determined that lipophilicity and the number of hydrogen bond acceptors (next to polarity and pKa) are the key properties for inhibiting the OATP1B1 probe substrate estradiol-17β-glucuronide in human hepatocytes. The key descriptors identified with our computational model are consistent with the data published by Karlgren et al. (2012b), who concluded that lipophilicity and polar surface area are key molecular features for inhibition of several OATP1B isoforms. More recently, Soars et al. (2012) developed an OATP1B1-HEK inhibition assay and also

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

concluded that the prime molecular descriptors for the prediction of OATP1B1 inhibition are maximal hydrogen bonding strength followed by lipophilicity.

Besides these previously reported key descriptors, we identified several new compound properties (e.g. high number of ring bonds) and substructures important for OATP1B inhibition (Figure 5 and Supplemental Table 6). Moreover, defined ranges for these physicochemical compound properties were specified rather than just a qualitative interpretation. To our knowledge, physicochemical compound properties along with compound substructures important for the OATP1B isoform selectivity are reported here for the first time (Figure 5). Finally, due to the nature of our proteochemometric models, we were also able to identify residues in both OATP1B1 and OATP1B3 that are important for OATP1B1 inhibition.

In conclusion, we developed an in vitro OATP1B inhibition assay that can immediately be implemented in drug discovery settings for enhanced throughput evaluation of the OATP1B inhibition potency of large numbers of drug candidates. For a total of 69 compounds we generated for the first time Ki values regarding their inhibition potential towards OATP1B1 and or OATP1B3. Based on our extensive amount of *in vitro* data we developed a proteochemometric model that accurately predicts OATP1B1 and OATP1B3 inhibition for a given new chemical entity based on its structure. We conclude that a high lipophilicity and polar surface, the presence of anionic atoms or several hydrogen bond acceptors and the absence of cationic atoms are key features for OATP1B inhibition. In addition, our proteochemometric approach allowed identifying residues in both OATP1B1 and OATP1B3 that are important for OATP1B1 inhibition.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

The in silico models developed in the present study are included in the Supplemental Methods and Supplemental Figure 3 and can be used within pipeline pilot 8.5 (Accelrys Software Inc, San Diego, CA) to predict the ability of any compound to inhibit OATP1B1 or OATP1B3.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Authorship Contributions

Participated in research design: TDB, GVW, PFA, PPA.

Conducted experiments: TDB, GVW.

Contributed new reagents or analytic tools: BS, PFA, PDW, API.

Performed data analysis: TDB, GVW, PPA

Wrote or contributed to the writing of the manuscript: TDB, GVW, PFA, PPA, API, BS, PDW.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

References:

Annaert P, Ye ZW, Stieger B, and Augustijns P (2010) Interaction of HIV protease inhibitors with OATP1B1, 1B3, and 2B1. *Xenobiotica* **40**: 163–176.

Badolo L, Rasmussen LM, Hansen HR, and Sveigaard C (2010) Screening of OATP1B1/3 and OCT1 inhibitors in cryopreserved hepatocytes in suspension. *Eur J Pharm Sci* **40**: 282–288.

Baldes C, Koenig P, Neumann D, Lenhof H-P, Kohlbacher O, and Lehr C-M (2006) Development of a fluorescence-based assay for screening of modulators of human Organic Anion Transporter 1B3 (OATP1B3). *European Journal of Pharmaceutics and Biopharmaceutics* **62**: 39–43.

Bednarczyk D (2010) Fluorescence-based assays for the assessment of drug interaction with the human transporters OATP1B1 and OATP1B3. *Analytical Biochemistry* **405**: 50–58.

De Bruyn T, Fattah S, Stieger B, Augustijns P, and Annaert P (2011) Sodium fluorescein is a probe substrate for hepatic drug transport mediated by OATP1B1 and OATP1B3. *J Pharm Sci* **100**: 5018–5030.

Chang C, Pang KS, Swaan PW, and Ekins S (2005) Comparative pharmacophore modeling of organic anion transporting polypeptides: a meta-analysis of rat Oatp1a1 and human OATP1B1. *J. Pharmacol. Exp. Ther*. **314**: 533–541.

Fenner KS, Jones HM, Ullah M, Kempshall S, Dickins M, Lai Y, Morgan P, and Barton HA (2012) The evolution of the OATP hepatic uptake transport protein family in DMPK sciences: from obscure liver transporters to key determinants of hepatobiliary clearance. *Xenobiotica* **42**: 28–45.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, and Overington JP (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**: D1100–1107.

Giacomini KM, Huang S-M, Tweedie DJ, Benet LZ, Brouwer KLR, Chu X, Dahlin A, Evers R, Fischer V, Hillgren KM, Hoffmaster KA, Ishikawa T, Keppler D, Kim RB, Lee CA, Niemi M, Polli JW, Sugiyama Y, Swaan PW, Ware JA, Wright SH, Yee SW, Zamek-Gliszczynski MJ, and Zhang L (2010) Membrane transporters in drug development. *Nat Rev Drug Discov* **9**: 215–236.

Gui C, Obaidat A, Chaguturu R, and Hagenbuch B (2010) Development of a cell-based high-throughput assay to screen for inhibitors of organic anion transporting polypeptides 1B1 and 1B3. *Curr Chem Genomics* **4**: 1–8.

Hagenbuch B and Gui C (2008) Xenobiotic transporters of the human organic anion transporting polypeptides (OATP) family. *Xenobiotica* **38**: 778–801.

Karlgren M, Ahlin G, Bergström CAS, Svensson R, Palm J, and Artursson P (2012a) In vitro and in silico strategies to identify OATP1B1 inhibitors and predict clinical drug-drug interactions. *Pharm. Res.* **29**: 411–426.

Karlgren M, Vildhede A, Norinder U, Wisniewski JR, Kimoto E, Lai Y, Haglund U, and Artursson P (2012b) Classification of inhibitors of hepatic organic anion transporting polypeptides (OATPs): influence of protein expression on drug-drug interactions. *J. Med. Chem.* **55**: 4740–4763.

Kounnis V, Ioachim E, Martin Svoboda, Andreas Tzakos, Ioannis Sainis, thalhammer theresia, Steiner G, and Briasoulis E (2011) Expression of organic anion-

transporting polypeptides 1B3, 1B1, and 1A2 in human pancreatic cancer reveals a new class of potential therapeutic targets. *OncoTargets and Therapy* 27.

Kramer C, Kalliokoski T, Gedeck P, and Vulpetti A (2012) The Experimental Uncertainty of Heterogeneous Public Ki Data. *Journal of Medicinal Chemistry* **55**: 5165–5173.

Lapinsh M, Prusis P, Gutcaits A, Lundstedt T, and Wikberg JE (2001) Development of proteo-chemometrics: a novel technology for the analysis of drug-receptor interactions. *Biochim. Biophys. Acta* **1525**: 180–190.

Lapinsh M, Prusis P, Lundstedt T, and Wikberg JES (2002) Proteochemometrics modeling of the interaction of amine G-protein coupled receptors with a diverse set of ligands. *Mol. Pharmacol.* **61**: 1465–1475.

Liaw A and Wiener M (2002) Classification and Regression by randomForest. *R news* **2**: 18–22.

Mandery K, Sticht H, Bujok K, Schmidt I, Fahrmayr C, Balk B, Fromm MF, and Glaeser H (2011) Functional and structural relevance of conserved positively charged lysine residues in organic anion transporting polypeptide 1B3. *Mol. Pharmacol.* **80**: 400–406.

Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **405**: 442–451.

Meier-Abt F, Mokrab Y, and Mizuguchi K (2005) Organic anion transporting polypeptides of the OATP/SLCO superfamily: identification of new members in nonmammalian species, comparative modeling and a potential transport mode. *J. Membr. Biol.* **208**: 213–227.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Müller F and Fromm MF (2011) Transporter-mediated drug-drug interactions. *Pharmacogenomics* **12**: 1017–1037.

Prusis P, Uhlén S, Petrovska R, Lapinsh M, and Wikberg JES (2006) Prediction of indirect interactions in proteins. *BMC Bioinformatics* **7**: 167.

Rognan D Proteome-scale docking: myth and reality. *Drug Discovery Today: Technologies*

Shitara Y (2011) Clinical importance of OATP1B1 and OATP1B3 in drug-drug interactions. *Drug Metab. Pharmacokinet.* **26**: 220–227.

Soars MG, Barton P, Ismair M, Jupp R, and Riley RJ (2012) The development, characterization, and application of an OATP1B1 inhibition assay in drug discovery. *Drug Metab. Dispos.* **40**: 1641–1648.

Svoboda M, Wlcek K, Taferner B, Hering S, Stieger B, Tong D, Zeillinger R, Thalhammer T, and Jäger W (2011) Expression of organic anion-transporting polypeptides 1B1 and 1B3 in ovarian cancer cells: relevance for paclitaxel transport. *Biomed. Pharmacother.* **65**: 417–426.

Treiber A, Schneiter R, Häusler S, and Stieger B (2007) 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* **35**: 1400–1407.

de Waart DR, Häusler S, Vlaming MLH, Kunne C, Hänggi E, Gruss H-J, Oude Elferink RPJ, and Stieger B (2010) Hepatic transport mechanisms of cholyl-L-lysyl-fluorescein. *J. Pharmacol. Exp. Ther.* **334**: 78–86.

van Westen GJP, Hendriks A, Wegner JK, IJzerman AP, van Vlijmen HWT, and Bender A (2013) Significantly Improved HIV Inhibitor Efficacy Prediction Employing Proteochemometric Models Generated From Antivirogram Data. *PLoS Comput Biol* **9**: e1002899.

van Westen GJP, van den Hoven OO, van der Pijl R, Mulder-Krieger T, de Vries H, Wegner JK, IJzerman AP, van Vlijmen HWT, and Bender A (2012) Identifying Novel Adenosine Receptor Ligands by Simultaneous Proteochemometric Modeling of Rat and Human Bioactivity Data. *J. Med. Chem.* **55**: 7010–7020.

van Westen GJP, Wegner JK, Geluykens P, Kwanten L, Vereycken I, Peeters A, Ijzerman AP, van Vlijmen HWT, and Bender A (2011a) Which compound to select in lead optimization? Prospectively validated proteochemometric models guide preclinical development. *PLoS ONE* **6**: e27518.

van Westen GJP, Wegner JK, IJzerman AP, van Vlijmen HWT, and Bender A (2011b) Proteochemometric modeling as a tool to design selective compounds and for extrapolating to novel targets. *MedChemComm* **2**: 16.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Footnotes

TDB received a PhD scholarship [093306] from the Agency for Innovation by Science and Technology, Flanders. GVW received funding from Tibotec BVBA. This study was supported by grants from 'Fonds voor Wetenschappelijk Onderzoek', Flanders [G.0662.09N] and 'Onderzoeksfonds' of the KU Leuven, Belgium.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Legends for figures:

Figure 1

Concentration-dependent uptake of sodium fluorescein in OATP1B1- (squares) and OATP1B3- (triangles) transfected CHO cells (n = 4). CHO cells were incubated with 1 to 80 μM sodium fluorescein. OATP-mediated sodium fluorescein uptake was obtained by subtracting uptake in wild-type CHO cells from total uptake in OATP-transfected cells. Points represent mean (± SD) of triplicate determinations and lines represent best fit to experimental data according to the Michaelis-Menten equation.

Figure 2

Concentration-dependent inhibition of sodium fluorescein (10 μM) uptake by the known OATP1B inhibitors rifampicin (Panel A) and darunavir (Panel B) in CHO cells. Diamonds represent OATP1B1-expressing CHO cells while squares represent OATP1B3-expressing CHO cells. Net uptake values were obtained by subtracting uptake in wild-type CHO cells from total uptake in transfected cells. Lines represent best fit to data as described in the method section.

Figure 3

The inhibitory effect of 2000 compounds (10 μM) on the uptake of sodium fluorescein (10 μM) in OATP1B1- (Panel A) and OATP1B3-transfected CHO cells. Relative sodium fluorescein uptake was obtained by dividing net sodium fluorescein uptake in presence of the potential inhibitor by the net uptake of sodium fluorescein in the control condition for each isoform. Compounds are ranked from potent inhibitor to potent activator of sodium fluorescein uptake.

Molecular Pharmacology Fast Forward. Published on April 9, 2013 as DOI: 10.1124/mol.112.084152
This article has not been copyedited and formatted. The final version may differ from this version.

Figure 4

The most important compound properties that positively or negatively correlate with OATP1B inhibition.

Figure 5

Selection of substructures that correlate with an activity class based on the 4-class model. (A) Substructures that lead to activity on both OATP1B1 and OATP1B3. (B) Substructures that lead to inactivity on both OATP1B1 and OATP1B3. (C) Substructures specific for activity on OATP1B1. (D) Substructures specific for activity on OATP1B3.

Figure 6

Prospective validation of the *in silico* model. A total number of 54 compounds were computationally evaluated for OATP1B1 (Panel A) and OATP1B3 (Panel B) inhibition with the final random forest model. Compounds that are computationally predicted to be inhibitors and non-inhibitors are represented with 1 and -1, respectively. The OATP1B1 and OATP1B3 inhibitory potential of all 54 compounds was determined following the described *in vitro* OATP1B inhibition assay with sodium fluorescein as a substrate.

Figure 7

Relationship between the experimentally determined inhibitory potential (remaining sodium fluorescein uptake) and the computationally predicted inactivity correlation value for all 54 compounds evaluated in the prospective validation. Panel A and panel B represent data for OATP1B1- and OATP1B3-mediated inhibition, respectively.

Table 1. OATP1B Inhibition data for all test compounds that inhibited sodium fluorescein uptake by at least 50 %.

Compound	Sodium Fluorescein (% of control)		Ki Value (μM)		Compound	Sodium Fluorescein (% of control)		Ki Value (μM)	
	OATP1B1	OATP1B3	OATP1B1	OATP1B3		OATP1B1	OATP1B3	OATP1B1	OATP1B3
Utilin	-3.8	2.7	0.38	0.60	3β-hydroxydeoxydesacetoxy-7-oxogedunin	10.8	22.2	1.33	3.81
Entandrophragmin	-0.4	0.6			Norstictic acid	10.9	1.8	0.63	0.47
Erythromycin ethylsuccinate	3.6	30.5	1.67	1.61	β-Escin	11.0	11.7	0.78	2.61
Rosolic Acid	4.1	20.4	1.43	3.15	Oligomycin A	11.3	3.0	0.25	0.44
Gangaleoidin	4.1	15.0	0.06	2.02	Dicloxacillin Sodium	11.4	9.8	0.37	1.58
Cyclosporin	4.6	4.2			6-Acetoxyangolensic acid methyl ester	12.0	35.0	0.80	2.94
3α-hydroxygedinin	6.1	23.2	1.21	2.89	Hydroxyzine pamoate	12.2	13.6	0.45	0.94
Lithocholic acid	6.6	60.2			Tetrahydrogambogic acid	12.5	4.4	1.10	2.15
Chukrasin, methyl ether	6.8	3.8	0.42	0.56	1,2α-epoxydeacetoxydihyrogedunin	12.7	21.3		
Erythromycin estolate	6.9	7.4	2.41	2.35	Decahydrogambogic acid	12.8	18.2	0.62	1.74
Bussein	7.3	5.4	0.38	0.53	Theaflavin digallate	13.1	9.4	1.51	3.42
Abietic acid	7.4	33.8	1.47	4.08	Ethinyl estradiol	13.1	61.2		
Hexachlorophene	7.6	3.1			Scandenin	13.8	9.0	0.59	0.55
Leoidin	7.8	14.6	0.08	1.84	3β-acetoxydeoxyangolensic acid	14.0	21.9	0.51	3.47
Rhodomyrtoxin B	8.1	37.5			Swietenine	14.0	69.9		
Deoxygedunin	8.7	22.7	0.83	2.89	Cephaloridine	14.2	32.0		
Chlormadinone acetate	9.7	24.4			Dihydrogambogic acid	14.2	17.3	1.23	2.24
β-Carotene	10.0	13.4	0.34	2.26	7-Deacetylkhivorin	14.3	9.5	0.54	2.27
Lasalocid sodium	10.3	6.7	0.26	0.59	Bergenin	14.6	94.5		
Acetyl isogambogic acid	10.4	30.3	2.24	4.17	3-deoxo-3β-acetoxydeoxydihydrogedunin	15.0	20.7	0.41	2.62
Candesartan cilextil	10.6	6.0	0.40	1.90	Bilirubin	15.3	32.1	0.72	2.63
Olmesartan medoxomil	15.4	9.7	0.54	0.75	Diffratic acid	22.7	12.7		

Carbenoxolone sodium	15.4	14.5	1.07	1.48
19-hydroxytotarol	15.9	20.5		
3α-hydroxy-3-deoxyangolensic acid	16.0	17.5	0.68	3.06
Paclitaxel	16.4	52.9		
Lorglumide sodium	16.5	32.8	2.01	2.71
Lobaric acid	18.3	7.6	0.40	1.25
α-dihydrogedunol	18.3	32.2		
Isogedunin	18.6	46.9		
Tannic acid	18.7	10.6	1.46	3.23
3α-Acetoxydihydrodeoxygedunin	18.9	8.0	0.44	1.08
Retinol	19.0	32.0		
Retinyl acetate	19.4	23.5	1.22	3.89
11-oxoursolic acid	19.8	5.2		
Pentachlorophenol	20.0	40.7		
Clobetasol propionate	20.2	21.3		
Angolensic acid	20.3	33.7		
Anthothecol	20.4	12.4	1.65	2.64
β-Dihydrogedunol	20.5	23.8	0.93	2.5
11α-Acetoxykhivorin	20.6	46.1	0.48	3.05
Phenolphthalein	20.6	38.0		
Suramin	20.8	30.8		
Carapin	20.9	66.5		
Carminic acid	21.6	121.9		
Salinomycin sodium	21.7	17.0	2.03	3.31
Humilin A	22.3	34.4		
Xylocarpus A	22.5	9.8		
Fissinolide	27.4	51.8		
Violastyrene	27.5	47.9		
Prieurianin	22.8	25.5	1.28	2.63
Dihydrodeoxygedunin	22.8	54.7	1.94	5.99
Baeomycesic acid	23.0	46.8		
Dimethyl Gambogate	23.2	62.1	1.26	3.94
Losartan	23.2	3.8	0.16	0.88
Rotenone	23.2	43.9		
Deoxygedunol acetate	23.3	49.2		
Arachidonic acid	23.4	5.4	0.33	0.8
6-hydroxyangolensic acid	23.7	61.1		
Lithocholenic acid	23.7	112.0		
Morin	24.2	45.4		
Dihydrocaryophyllenone	24.2	39.1		
Gossypol	24.5	15.8	1.78	3.41
Chrysanthellin A	25.1	17.4	1.82	4.02
Teniposide	25.4	32.6		
Betulinic acid	25.9	38.6		
Khayanthone	26.0	57.6		
Cholic acid, methyl ester	26.0	40.9	4.58	2.82
Carnosine	26.4	98.9		
Lonchocarpic acid	26.5	36.3		
Umbelliferone	26.7	76.5		
Diltiazem Hydrochloride	26.7	80.8		
Tretinon	26.8	55.5		
Bromo-3-hydroxy-4-succinylcaryolane	26.8	50.1		
Dihydroergotamine mesylate	27.1	10.1		
3-deoxy-3β-hydroxymexicanolide 16-enol ether	27.2	106.5		
Isotretinon	32.2	51.5		
N-(9-Fluorenylmethoxycarbonyl)-L-Leucine	32.5	48.3		

Deoxyandirobin	27.7	42.8		
Androsterone sodium sulfate	27.8	40.8		
Prieuranin acetate	28.2	54.4		
8β-hydroxycarapin	28.5	49.6		
Euphol acetate	29.2	40.2		
Deoxyandirobin lactone	29.5	53.1		
2-methylene-5-dioxotetrahydrofuranyl-6-oxodimethylbicycloundecane	29.8	47.6		
Dihydromunduletone	30.3	58.7		
Cyproterone	30.6	112.2		
Garcinolic acid	30.6	3.6	0.65	1.73
1,2α-epoxydeoxydihydrogedunin	30.8	38.6	0.86	2.28
Usnic acid	31.1	69.1		
2,2-bisepigallocatechin monogallate	31.1	18.8	2.05	6.50
1,3-dideacetyldeoxykhivorin	31.1	50.8	2.96	6.49
Iopanic acid	31.1	59.6		
Rhodomyrtoxin	31.1	66.2		
Mundulone	31.2	45.1		
Pyrantel pamoate	31.4	34.8		
3β-Hydroxydeoxodihydrodeoxygedunin	31.5	25.9	1.49	2.39
Ergosterol	31.7	52.3		
Antimycin A	31.9	28.6	2.20	3.82
Canrenone	32.0	78.7		
Isoxsuprine HCl	38.4	77.5		
Isoreserpine	38.5	42.2		
6-oxoprogesterone	38.5	75.5		
Dihydrogeduninic acid, methyl ester	32.7	97.1		
Nonoxynol	32.7	28.0	1.92	2.87
Dicumarol	33.1	19.3		
Tetrac	33.4	71.4		
1,7-Dideacetoxy-1,7-dioxo-3-deacetylkhivorin	33.5	61.5		
Rifampicin	33.6	22.0		
Mometasone furoate	34.0	25.6	1.62	0.87
Citrinin	34.4	67.3		
6,4’-Dimethoxyflavone	34.4	110.4		
Ajmaline	34.6	91.5		
Gangleoidin acetate	34.6	54.6		
Dihydrogedunin	35.5	62.4		
Enoxolone	35.6	4.3	0.29	0.41
Amcinonide	25.8	56.6		
Atorvastatin Calcium	36.1	32.1	0.45	2.58
Bithionol	36.1	37.2		
Isoginkgetin	36.2	63.5		
Proadifen HCl	36.2	53.5		
Fludrocortisone acetate	36.3	71.2		
Oleanolic acid acetate	37.0	24.7		
Chenodiol	37.9	71.6		
α-hyodeoxycholic acid	38.2	89.1		
Clomipramine HCl	43.6	78.8		
Budesonide	43.8	73.0		
Mexicanolide	44.1	122.7		

Betamethasone valerate	38.6	10.2		
3,5-Dinitrocatechol	38.6	75.4		
N-Formylmethionylphenylalanine	28.8	104.4		
Pectolinarin	39.1	155.8		
2-Methyl gramine	39.2	80.1		
Sildenafil	39.3	107.1		
Fusidic acid	39.5	58.3		
3β-acetoxydeoxodihydrogedunin	39.6	104.7		
5-Nitro-2-phenylpropylaminobenzoic acid	40.1	33.4		
Hederagenin	40.2	18.6	2.11	0.68
Deacetoxy-7-oxogedunin	40.3	85.1		
Chol-11-enic acid	40.4	203.4		
Troleandomycin	40.5	42.0		
Hydroxyprogesterone caproate	40.6	55.8		
Allopregnanolone	41.5	89.7		
Beclomethasone dipropionate	41.5	14.8		
Hecogenin	41.8	105.1		
Oxyphenbutazone	42.0	83.2		
Verapamil HCl	42.1	89.2		
4’-Demethylepipodophyllotoxin	42.5	69.6		
Emetine	42.6	65.6		
Pyrvinium pamoate	42.6	32.7		
Azelastine HCl	43.4	134.9		
Chloroxine	43.5	62.2		
Larixol	48.6	78.3		
Nicolsamide	48.7	47.1		
Simvastatin	44.7	60.2		
Cedrelone	44.7	60.7		
Haematoporphyrin dihydrochloride	45.0	40.1		
Cephalothin sodium	45.1	66.3		
Thyroxine	45.2	67.3		
Epicatechin monogallate	46.4	25.3	2.07	3.4
Ginkgetin	46.4	46.7		
Bacampicillin HCl	46.4	80.6		
Hexestrol	46.6	43.2		
18α-Glycyrrhetinic acid	47.1	59.1		
Quercetin	47.1	122.7		
Cloxacillin sodium	47.3	63.2		
12-hydroxy-4,4’-bisnor-podocarpatetraen-one	47.5	82.2		
Dihydrocelastryl diacetate	47.6	13.7		
Glycyrrhizic acid	47.6	39.9		
Carprofen	47.7	77.8		
Boldine	47.9	100.1		
Tridesacetoxykhivorin	48.0	84.2		
Methyl-7-deshydroxypyrogallin-4-carboxylate	48.2	80.2		
1,3-dideacetyl-7-deacetoxy-7-oxokhivorin	48.2	93.9		
Solidagenone	48.3	78.8		
Epiafzelechin	48.3	26.7		
Alverine citrate	48.3	74.4		
Aloin	48.5	66.8		
2,3-diacetoxy-7,8-epoxy-24,29-dinor-1,3,5-friedelatriene-20-carboxylic acid	63.6	37.7		
Nimodipine	64.8	37.8		

Dienestrol	48.7	61.9		
Asarylaldehyde	48.7	59.2		
Acetylglucosamine	49.0	83.0		
Epigallocatechin-3-monogallate	49.0	45.9		
Guaiazulene	49.1	50.3		
Loganic acid	49.2	99.5		
Nystatin	49.2	76.2		
Protoporphynyrinix	49.6	94.6		
Suxibuzone	49.6	84.7		
Digydrogedunic acid, methyl ester	49.7	99.1		
Coniferyl alcohol	49.7	66.1		
Dextromethorphan HBr	49.8	78.2		
Rosuvastatin	52.1	5.7	2.43	3.61
Glucosaminic acid	63.5	25.6		
Smilagenin	50.8	25.7	0.13	0.08
Ergocalciferol	81.7	28.8		
Isopimpinellin	84.5	29.0		
Echinocystic acid	69.3	30.5		
Ezetimibe	51.4	30.6		
Stictic acid	94.3	32.3		
Madecassic acid	69.3	33.8		
Irigenin, 7-benzyl ether	60.7	37.0		
Genistein	95.6	39.1	5.00	2.96
Dihydrocelasterol	70.5	40.4		
Telmisartan	77.6	41.9	1.03	0.96
Sulfinpyrazone	55.2	43.9		
Diethylstilbestrol	63.7	45.6		
Silibinin	54.8	46.9	3.37	3.2
Novobiocin sodium	56.2	47.2		
Irbesartan	76.0	47.2		
Dihydro-7-desacetyldeoxygedunin	54.8	48.3	3.44	2.5
Cosmosiin hexaacetate	95.2	48.4		
Troxerutin	61.9	49.5		

Table 2. Prospective Validation

Protein Isoform	True positives	False positives	True negatives	False negatives	Sensitivity	Specificity	Matthews	Percent correctly classified
OATP1B1	22	8	21	3	0.88	0.72	0.60	79.6 %
OATP1B3	13	10	27	4	0.77	0.73	0.46	74.1 %

Figure 1

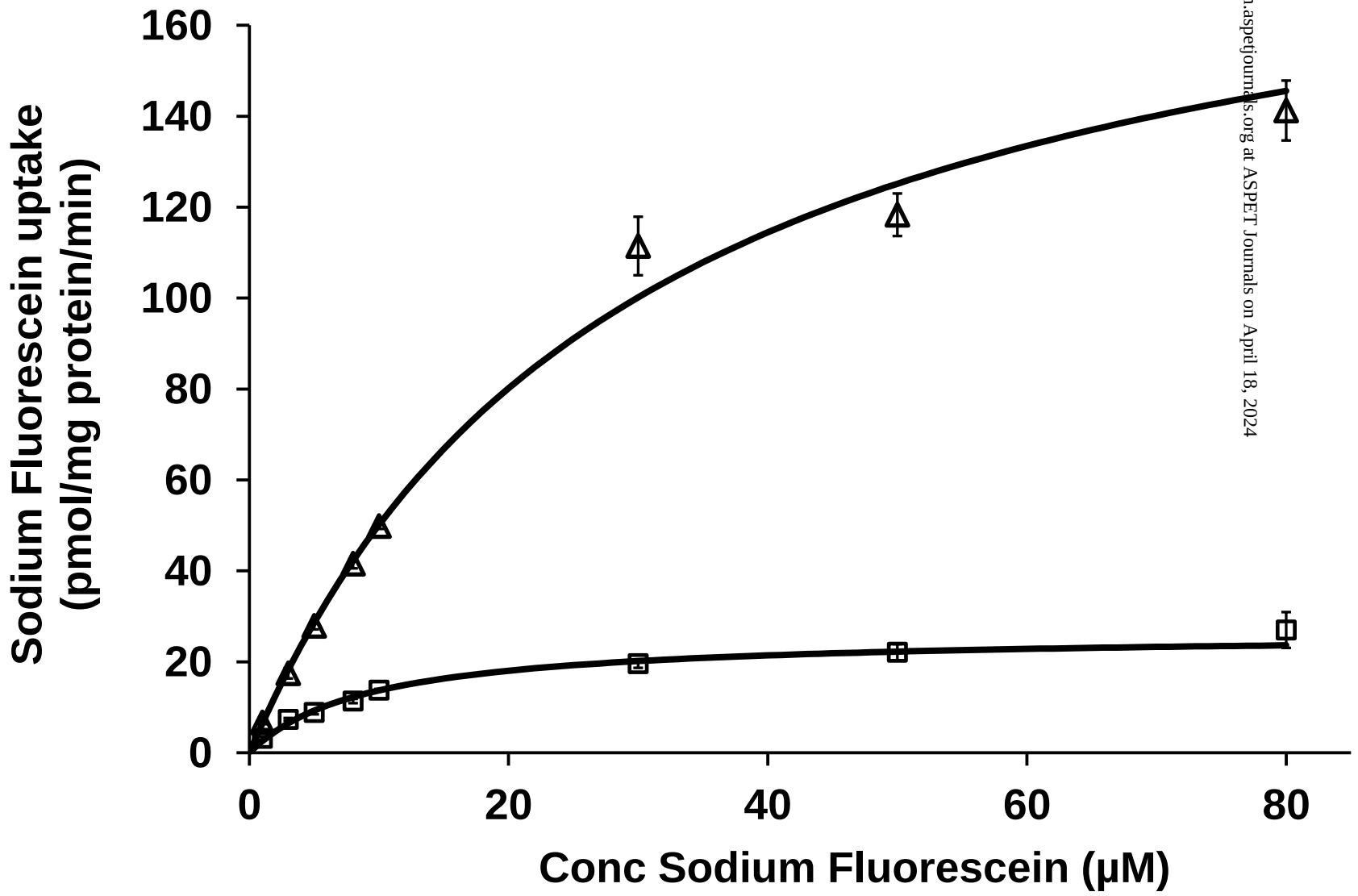
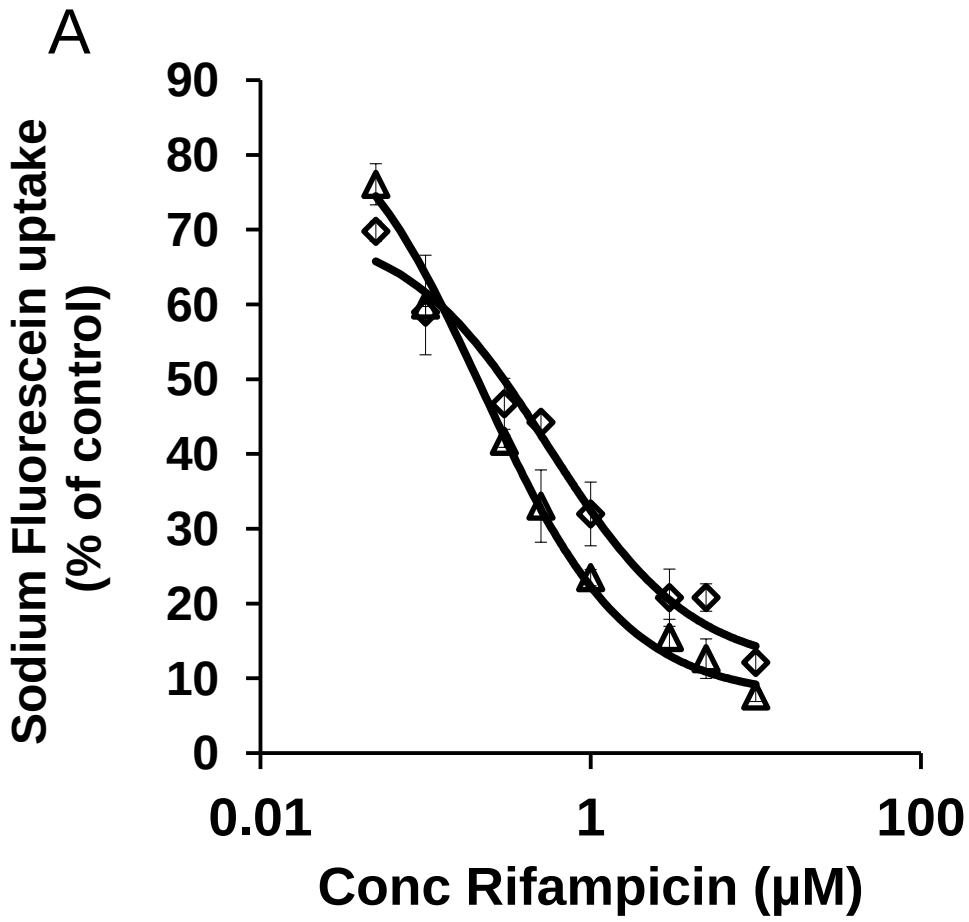


Figure 2



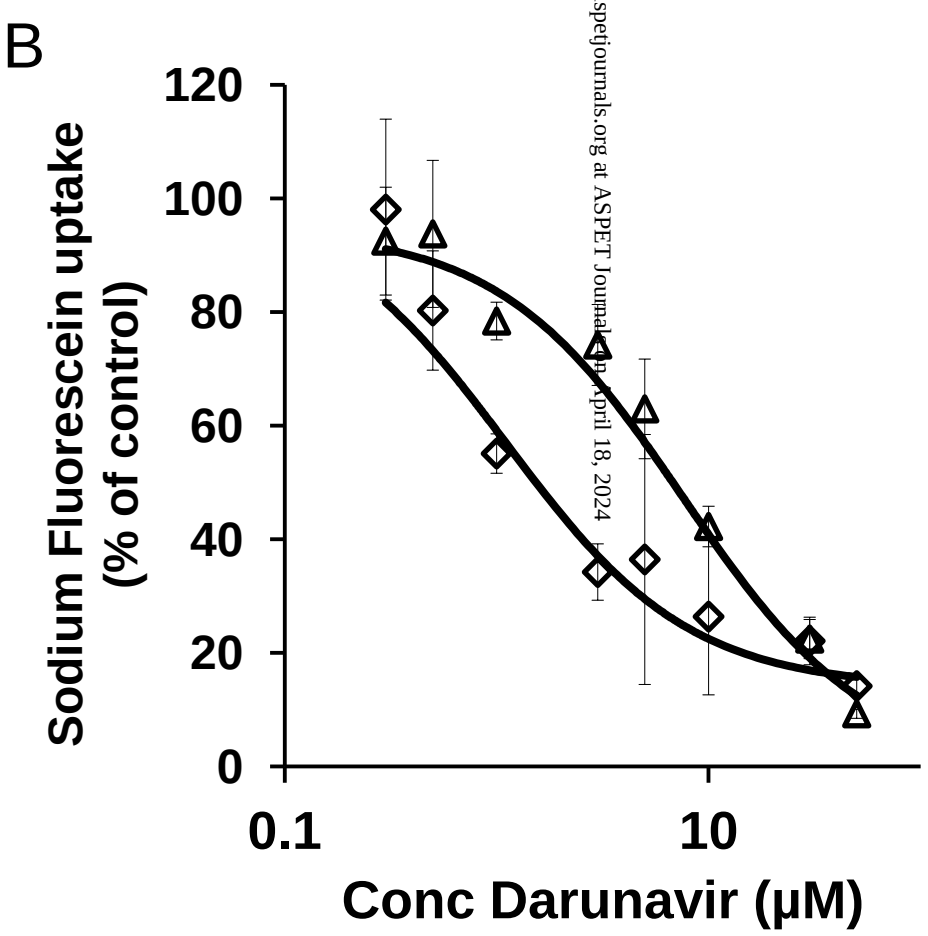


Figure 3

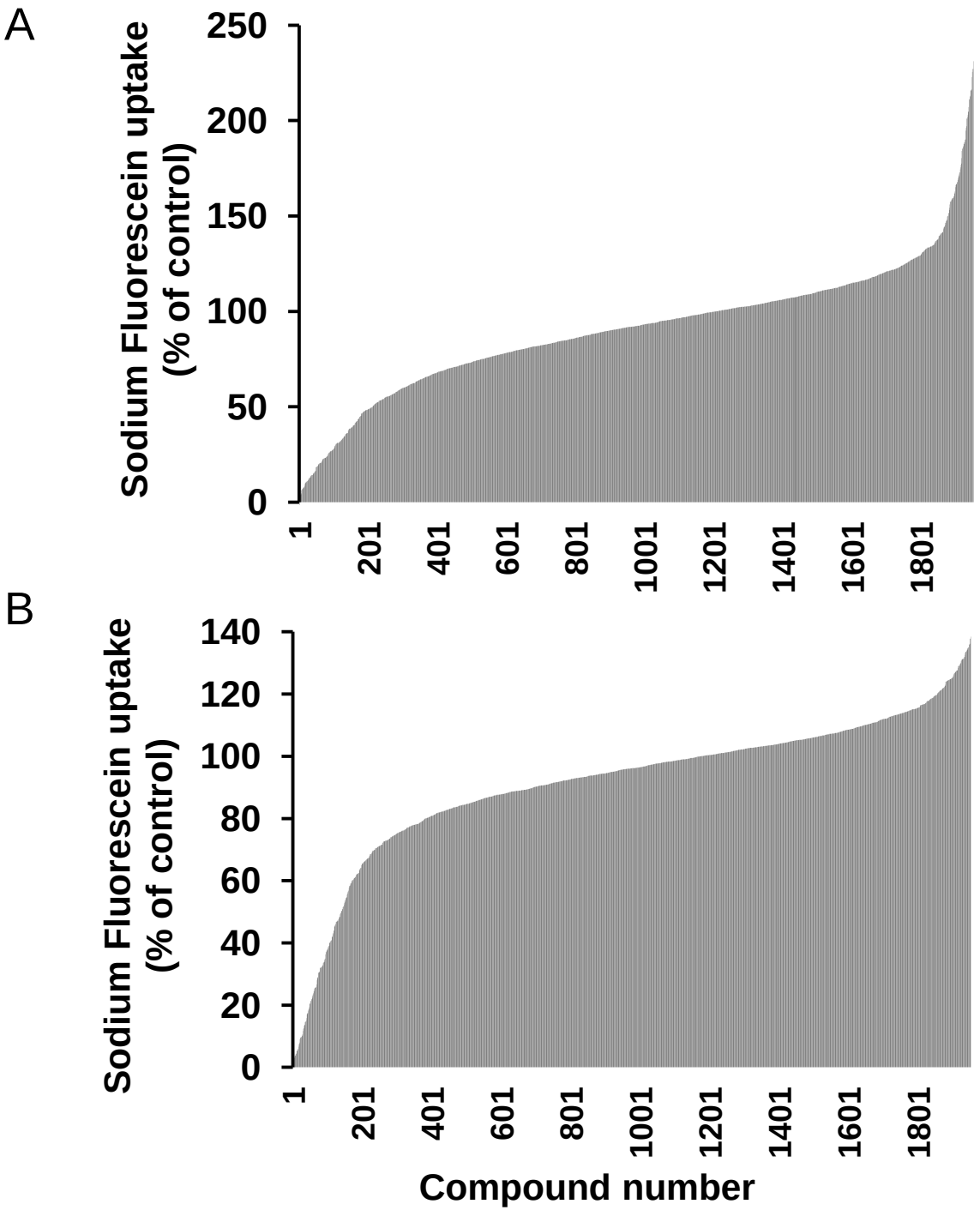


Figure 4

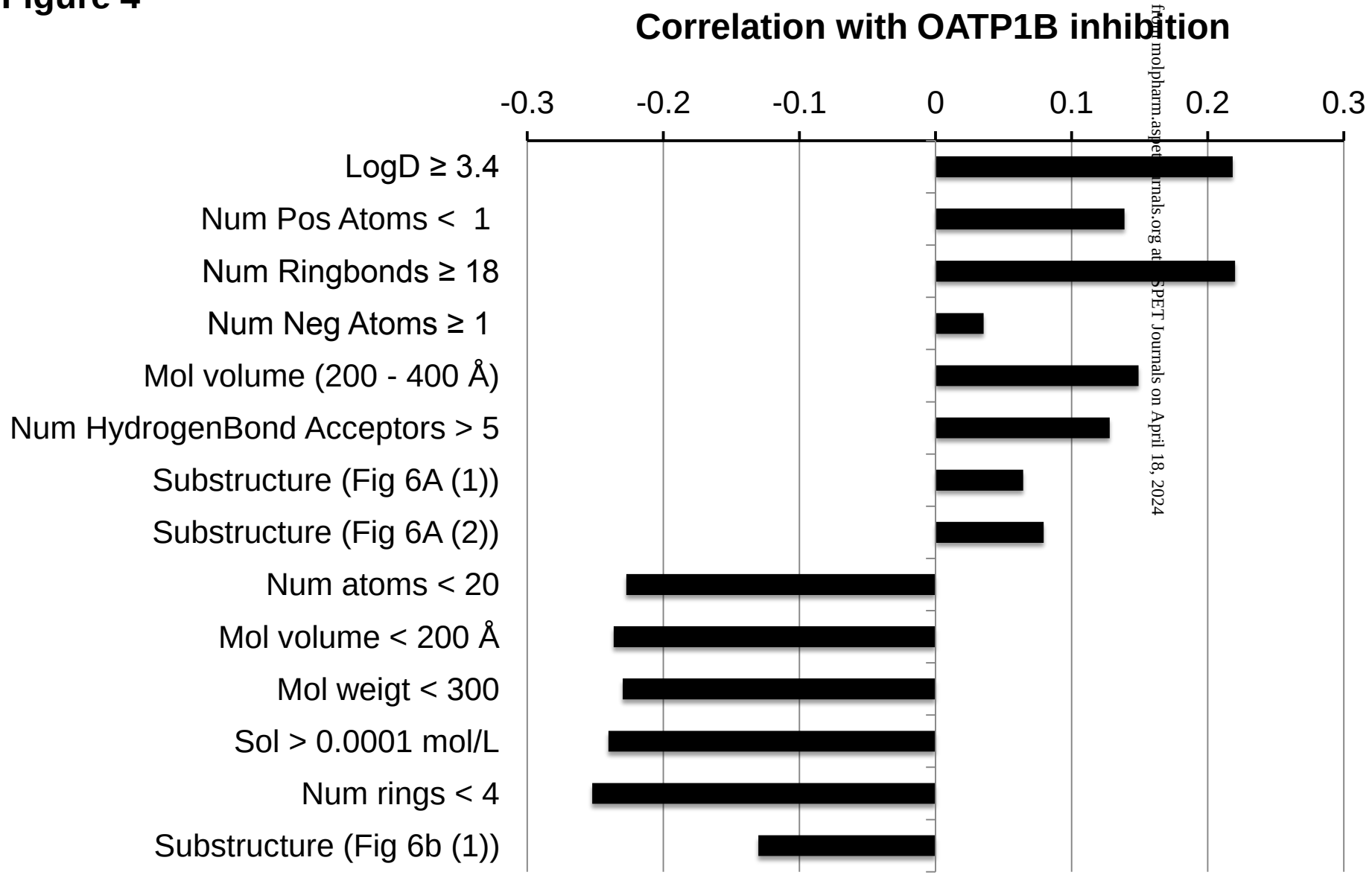


Figure 5

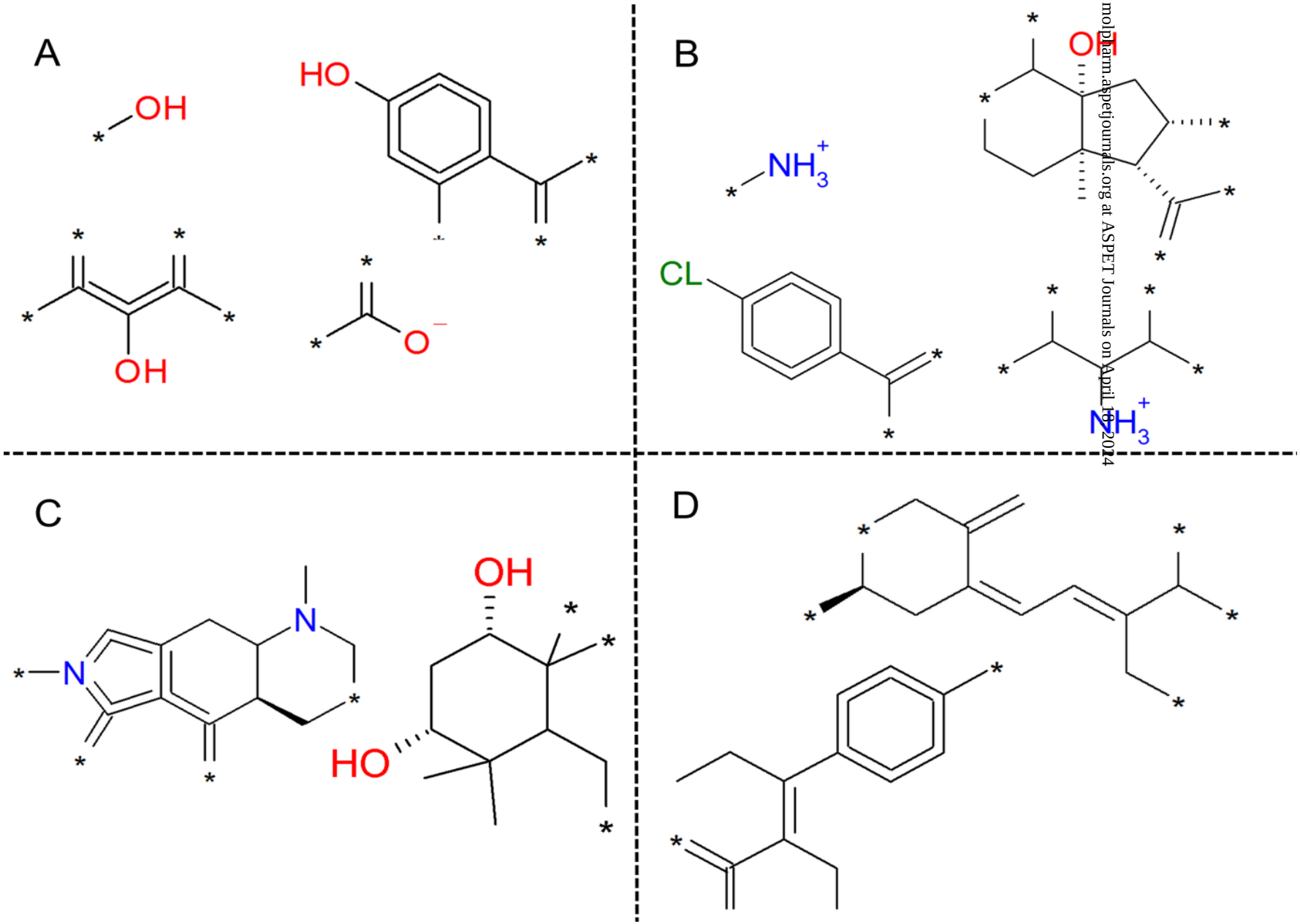


Figure 6

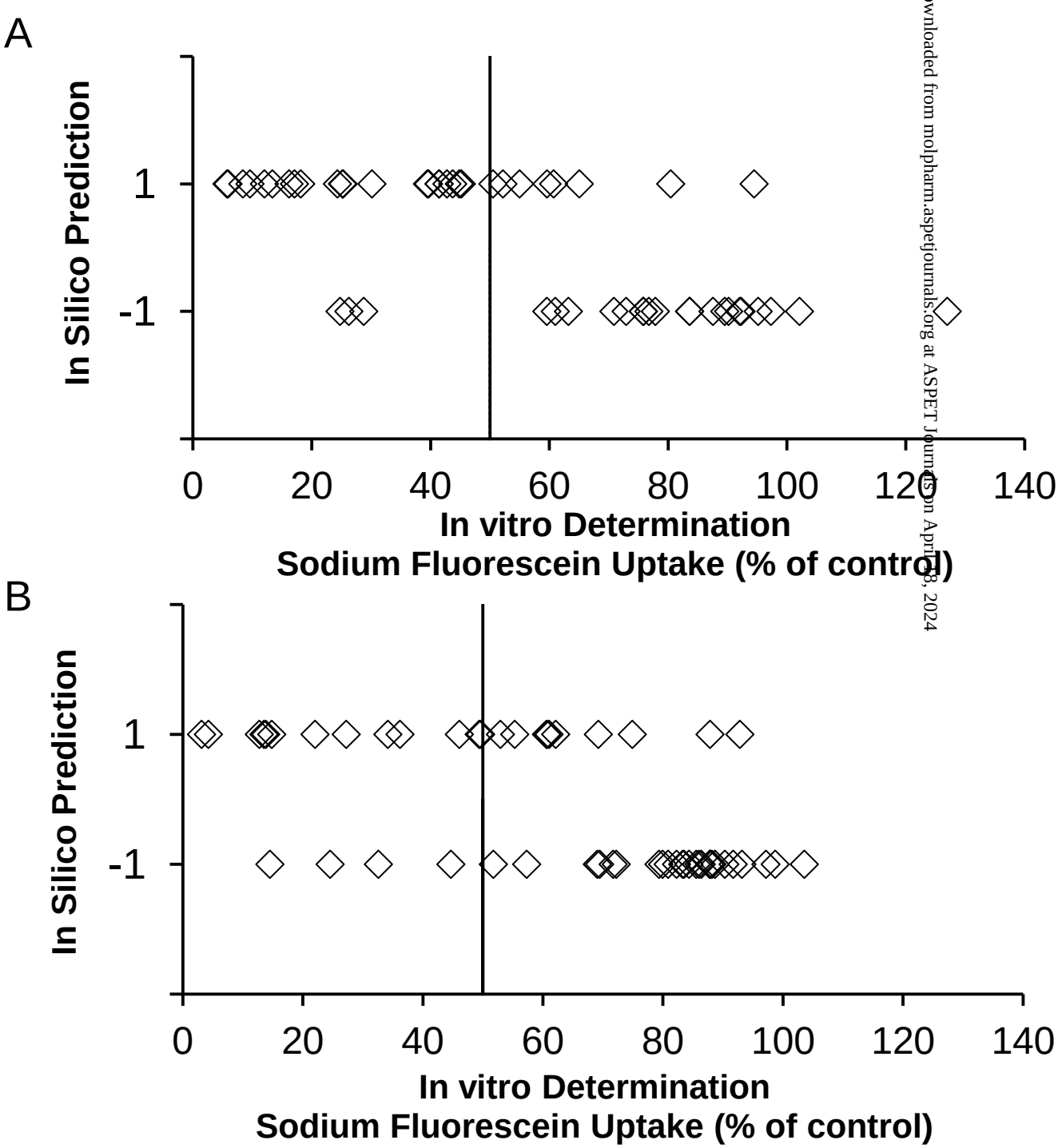


Figure 7