In Silico Prediction of Human Sulfotransferase 1E1 Activity Guided by Pharmacophores from Molecular Dynamics Simulations*

Received for publication, August 31, 2015, and in revised form, November 5, 2015 Published, JBC Papers in Press, November 5, 2015, DOI 10.1074/jbc.M115.685610

Christin Rakers†, Fabian Schumacher‡, Walter Meinl**, Hansruedi Glatt**, Burkhard Kleuser‡, and Gerhard Wolber**

From the †Institute of Pharmacy, Freie Universität Berlin, Königin-Luise-Str. 2/4, 14195 Berlin, the Departments of ‡Molecular Toxicology and **Nutritional Toxicology, German Institute of Human Nutrition (DIfE) Potsdam-Rehbrücke, Arthur-Scheunert-Allee 114–116, 14558 Nuthetal, the ††Department of Toxicology, Institute of Nutritional Science, University of Potsdam, Arthur-Scheunert-Allee 114–116, 14558 Nuthetal, and the ‡‡Department of Molecular Biology, University of Duisburg-Essen, Hufelandstr. 55, 45147 Essen, Germany

Acting during phase II metabolism, sulfotransferases (SULTs) serve detoxification by transforming a broad spectrum of compounds from pharmaceutical, nutritional, or environmental sources into more easily excretable metabolites. However, SULT activity has also been shown to promote formation of reactive metabolites that may have genotoxic effects. SULT subtype 1E1 (SULT1E1) was identified as a key player in estrogen homeostasis, which is involved in many physiological processes and the pathogenesis of breast and endometrial cancer. The development of an in silico prediction model for SULT1E1 ligands would therefore support the development of metabolically inert drugs and help to assess health risks related to hormonal imbalances. Here, we report on a novel approach to develop a model that enables prediction of substrates and inhibitors of SULT1E1. Molecular dynamics simulations were performed to investigate enzyme flexibility and sample protein conformational changes. Pharmacophores were developed that served as a cornerstone of the model, and machine learning techniques were applied for prediction refinement. The prediction model was used to screen the DrugBank (a database of experimental and approved drugs): 28% of the predicted hits were reported in literature as ligands of SULT1E1. From the remaining hits, a selection of nine molecules was subjected to biochemical assay validation and experimental results were in accordance with the in silico prediction of SULT1E1 inhibitors and substrates, thus affirming our prediction hypotheses.

Metabolism plays an important role in drug discovery and appropriate metabolic profiles of new drug candidates should be taken into consideration during the process of drug development. Acting in phase I metabolism, the enzyme family of cytochrome P450 is estimated to transform about 75% of marketed drugs (1) and numerous in silico approaches for the prediction of cytochrome P450-mediated metabolism have emerged to date (2, 3). Although the majority of metabolism prediction studies focuses on phase I, the significance of phase II metabolism is generally underestimated (4) and to this day, computer-based models for the prediction of phase II metabolism remain scarce (5).

Among the predominant phase II enzyme families are the soluble sulfotransferases that form a gene superfamily termed SULT. These enzymes regulate the sulfonation of smaller molecules such as endogenous hormones, neurotransmitters, and xenobiotic substances from pharmaceutical, nutritional, or environmental sources. Based on sequence similarity, functional human SULTs are divided into two main families (SULT1 and SULT2) and further into subfamilies that exhibit individual, but somewhat overlapping substrate specificities (6). Influencing the level of female sex hormones (estrogens), SULT subtype 1E1 (SULT1E1) shows a specific substrate preference for physiological estrogenic compounds ($K_m = 5 \text{ nm}$ for estradiol (7)) and has been extensively investigated in experimental studies.

In general, sulfonation reactions in which a sulfonate group from the cofactor PAPS2 is transferred to the hydroxyl group of a substrate, serve detoxification. With rare exceptions, sulfonated metabolites are charged and hydrophilic, and therefore excreted from the human body. As a consequence, sulfonation of drugs enforce their inactivation and thus reduces their efficacy (8, 9). However, sulfonation can also lead to the formation of chemically reactive or toxic metabolites (10, 11). It is now a commonly accepted concept that in some sulfonation reactions with certain molecules, e.g. alkylated polycyclic aromatic hydrocarbons or aromatic amines, the resulting sulfate group is electron withdrawing and becomes a good leaving group. Cleavage of this group is further facilitated by resonance and results in highly reactive electrophiles that cause DNA damage (10, 11).

Apart from the impact of SULTs on small molecules that act as substrates and undergo sulfonation, SULTs are in turn prone

---

* This work was supported by an Elsa-Neumann Stipendium of the state Berlin (to C. R.). The authors declare that they have no conflicts of interest with the contents of this article.

** This article contains supplemental Table S1.

1 To whom the correspondence should be addressed. E-mail: gerhard.wolber@fu-berlin.de.

2 The abbreviations used are: PAPS, 3′-phosphoadenosine-5′-phosphosulfate; CDL, concentration-dependent ligand; FN, false negative; FP, false positive; MD, molecular dynamics; SULT, sulfotransferase; PAP, 3′-phosphoadenosine-5′-phosphate; PDB, Protein Data Bank; PFS, pharmacophore fit score; r.m.s. deviation, root mean square deviation; SVM, support vector machine.
to inhibition by various endo- or exogenous substances like drugs (12, 13), food components (14–16), or environmental products (17, 18). The inhibition of SULTs decreases sulfonation rates, which disrupts homeostasis of endogenous molecules like hormones, neurotransmitters, or bile acids. Such sulfonation disorders have been linked to various diseases (19–21). Changes in SULT1E1 activity are associated with breast and endometrial cancer because estrogens can act as tumor initiators or promoters (22, 23). Thus, the inhibition of SULT1E1 by drugs or other xenobiotics may lead to increased estrogen levels, and therefore might directly promote carcinogenesis (24). Among the most notable compound classes with high inhibitory potential toward SULT1E1 are endocrine disrupting compounds, which are ubiquitous in our environment (e.g. as industrial chemicals, pesticides, or phytoestrogens). The ability to strongly inhibit SULT1E1 and the consequent danger of developing diseases stresses the importance to develop a prediction model for SULT1E1 to enable assessment of health risks associated with hormone imbalances. An in silico prediction model for SULT1E1 activation and inhibition further supports drug design by guiding the development of metabolically inert drug candidates. This might in turn decrease severe adverse events that are caused by the emergence of reactive metabolites.

Over the last decades, experimental data on SULT1E1 has constantly grown, although structure-based in silico approaches on SULT1E1 have remained scarce. Based on the enzyme structure, computational studies have investigated the mechanism of inhibition of SULT1E1 by nucleotides (25) and stereoselectivity of sulfonation via docking (26), analyzed the sulfonation reaction using QM/MM methods (27), and utilized molecular docking to predict ligand binding (28). Here, we report on a novel approach of computer-based metabolism prediction for human SULT1E1. A combination of molecular dynamics (MD) simulation, three-dimensional pharmacophores, and machine learning was used to develop a prediction model that allows identification of substrates and inhibitors. The presented in silico model was experimentally validated and allows efficient screening of large numbers of compounds.

Experimental Procedures

Molecular Modeling Methods

Molecular Dynamics Simulations—Protein Data Bank (PDB) entry 1HY3 (resolution 1.80 Å (29), chain B) was chosen as template for the structure-based study of human SULT1E1 activity as it features the cofactor PAPS in its active form (and not PAP), which is a prerequisite for sulfonation. The protein structure was analyzed and prepared using the modeling software Molecular Operating Environment 2010.12 (Chemical Computing Group Inc., Montreal, Canada). MD simulations were performed using Desmond MD package version 3.1.51.1 (30) with the OLPS-AA 2005 force field (31). PROPKA was applied for pK_a calculations and crystal water was removed beyond 5 Å from the protein (32). The orthorhombic system (maximal distance of 10 Å from the protein structure) was filled with simple point charge explicit water and ions were added to adjust to an overall neutral charge (salt concentration of 0.15 M). Minimization of the system was achieved after 2000 iterations with a convergence threshold of 1.0 kcal/mol/Å. The system was relaxed according to Desmond relaxation protocol under NPT conditions with the following alterations in relaxation times: 360 and 720 ps instead of 12 and 24 ps, respectively. Each simulation was performed for 100 ns with frames recorded every 4.8 ps on the computer cluster at the Freie Universität Berlin. Root mean square deviation (r.m.s. deviation) based clustering of the MD trajectories (excluding the first 3000 frames (∆ 14.4 ns)) was applied to the C_alpha atoms of residues 84 to 237, 237 to 259, and 142 to 150 using the g_cluster tool of GROMACS (gromos method, cutoff = 0.28 for cofactor-bound structures and cutoff = 0.25 for apo structures) (33). The center of each resulting cluster was used for subsequent ensemble docking.

Dataset Preparation—Known active ligands of SULT1E1 were collected from literature and the database BRENDA (34). The three-dimensional structures of the active compounds were processed using Corina 3.0.0 (35) and minimized using the force field MMFF94 (36). According to the literature, the chosen molecules were categorized into substrates, inhibitors, and ligands that are substrates with reported potential to cause concentration-dependent enzyme inhibition (termed CDL), which is a commonly observed phenomenon for metabolic enzymes (37, 38). With a total number of 118 active molecules, subsets for pharmacophore development and validation comprised 72 inhibitors, 36 substrates, and 10 CDL. For the development and validation of the three-dimensional pharmacophores, the inhibitor and substrate subsets were again manually split into training and test sets based on their structural similarity. Decoys (putatively inactive molecules) for pharmacophore and machine learning model validation were generated from the ZINC database (39) using an in-house KNIME workflow (40) and the online tool Directory of Useful Decoys, Enhanced (DUD-E) (41), which is available online.

Ensemble Docking, Pharmacophore Development, and Virtual Screening—Ten protein structures extracted from MD trajectories via r.m.s. deviation-based clustering served as input for ensemble docking along with the training set of active SULT1E1 ligands. GOLD suite version 5.1 (42) was used with default parameters and the Piecewise Linear Potential (ChemPLP) scoring function (43) to perform 100 docking runs per ligand and protein structure. Successful reproduction of the position of co-crystallized inhibitor 3,5,3’-tetrachlorobiphenyl-4,4’-diol in hSULT1E1 crystal structure PDB entry 1G3M (44) indicated legitimacy of the applied docking procedure. The docking results were transferred to LigandScout 3.1 (45–47) for three-dimensional pharmacophore analysis and development. Heat maps were generated for statistical evaluation using Gnu-plot 4.6. Predictive power of developed pharmacophores was assessed using the validation test set of molecules, which contained 2657 decoys (see also section Dataset preparation). Prior to pharmacophore-based screening with LigandScout, DrugBank 3.0 (48) was transferred to the LigandScout command-line tool idbgen with standard parameters (OMEGA-fast setting, 25 conformations per molecule). Pharmacophore screening was subjected to LigandScout module iscreen against the DrugBank library that comprised 6,494 drug-like molecules.
In Silico Prediction of Ligand Binding to SULT1E1

Development of Classification Models—For the development of classification models via machine learning techniques, separate datasets of active molecules were created from the total set of 118 ligands. For an inhibitor classification model, the training set contained 57 compounds (41 active and 16 inactive molecules) and the test set contained a total number of 1289 compounds (27 active, 18 inactive, and 1244 decoy molecules). Discrimination between active and inactive molecules was defined by an activity threshold of 10 μM (IC₅₀). The training set for substrate classification comprised 23 actives and 29 inactive (selected decoys), and the test set contained 23 actives and 979 decoys. LigandScout was used to calculate descriptors for all molecules including standard properties (relative topological surface area, number of rotatable bonds, acceptors, donors, rings, cLogP, MolWt, and heavy atoms) and pharmacophore fit scores (PFS) (46). The PFS is defined as follows: PFS = (10 x n) + (9 – 3 x min(r,3)), with n being the number of geometrically matched feature pairs, and r being the r.m.s. deviation of the matched feature pair distance. The molecule databases of training and test sets for inhibitors and substrates including molecular descriptors were prepared in MOE and transferred to KNIME for subsequent development of support vector machines (SVM). For further reading on the principle of SVM, see Witten and Frank (49). The legitimacy of the manual iterative selection process of descriptors was evaluated by using the WEKA KNIME node attribute selector (best fit method). The SVM model development was conducted using a polynomial kernel. The applicability domains of input molecules were assessed based on Euclidian distances.

Statistical Model Evaluation—Predictive power of the resulting models was evaluated in terms of standard parameters of the confusion matrix: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The ability to classify compounds into active and inactive was assessed by calculating the sensitivity (Equation 1), specificity (Equation 2), and accuracy (Equation 3). Matthew’s correlation coefficient was calculated based on Equation 4 to measure the quality of binary classification.

\[
Se = \frac{TP}{TP + FN} \quad \text{(Eq. 1)}
\]
\[
Sp = \frac{TN}{TN + FP} \quad \text{(Eq. 2)}
\]
\[
ACC = \frac{(TP + TN)}{(TP + FN + TN + FP)} \quad \text{(Eq. 3)}
\]
\[
MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP \times FP)(TP \times FN)(TN \times FP)(TN \times FN)}} \quad \text{(Eq. 4)}
\]

Experimental Methods

Chemicals—l–soliuritigienin (1) (CAS 961-29-5) and 17-epi­estriol (9) (CAS 1228-72-4) were obtained from Carbosynth Limited (Berkshire, UK) with a purity of >98% each. Sigma-Aldrich (Taufkirchen, Germany) supplied indazole-Cl (2) (CAS 848142-62-1), prinaberal (3) (CAS 524684-52-4), ZM241385 (4 (CAS 139180-30-6), and 2-(4-hydroxyphenylazo)benzoic acid (7) (CAS 1634-82-8) with compound purities of ≥98% as well as α-naphthol (CAS 90-15-3) with a purity of >99% and α-naphthylsulfate, potassium salt (CAS 6295-74-5). Compounds Amb1890033 (5) (CAS 341977-89-7), Amb1899186 (6) (CAS 769166-22-5), and Amb4444666 (8) (CAS 158860-23-2) were purchased from Ambinter (Orléans, France) (purities of 99, 99, and 93%, respectively). PAPS was produced using recombinant PAPS synthetase 1, expressed in Escherichia coli. After purification of PAPS by preparative anion exchange high-performance liquid chromatography (HPLC), its purity was determined as ≥99% (HPLC with UV detection).

Bacterial Strains and Cytosolic Preparations—Human SULT1E1 was expressed in Salmonella typhimurium TA1538 using the pKK233-2 expression vector as described previously (50, 51). The modified bacterial strain was grown overnight in Luria Broth medium (Roth, Karlsruhe, Germany) in the presence of ampicillin (100 μg/ml) by shaking at 37 °C for 8 h. Cytosolic fractions from bacteria were prepared as described previously (51, 52). The final protein concentration was determined to be 4.6 mg/ml using the bicinchoninic assay (Thermo Fisher Scientific, Bonn, Germany) in microtiter plates according to the manufacturer’s specifications with bovine serum albumin as a standard. Aliquots were stored at −80 °C. In comparison to purified SULT protein, Salmonella cytosol was observed to stabilize various SULT subtypes and furthermore, supports degradation of PAP, which reportedly inhibits SULT1E1 (53). Kinetic data obtained from the enzyme assay described below indicate stability of the enzyme over a time period of more than 10 h.

Enzyme Assay of SULT1E1—Time and protein dependence of α-naphthol sulfonation by SULT1E1 were investigated to optimize assay conditions. The incubation time of SULT1E1 with substrate was set to 15 min and protein concentration with linear formation of α-naphthylsulfate was set to 2.3 μg/sample. Standard sample mixtures of the assay contained 50 mM potassium phosphate buffer (pH 7.4), 5 mM MgCl₂, 50 μM of the cofactor PAPS, and 2.3 μg of protein (100 μl total volume). The samples were incubated for 2 min at 37 °C prior to the addition of the substrate. To determine enzyme kinetics of α-naphthol sulfonation, the enzyme reaction was initiated by adding α-naphthol in 10 different concentrations (ranging from 0.1 to 30 μM) and samples were incubated at 37 °C for 15 min under gentle shaking. Enzymatic activity was terminated by heat inactivation at 95 °C for 2 min under shaking. Denatured protein was precipitated by incubation on ice for 10 min and subsequent centrifugation at 21,800 × g at 4 °C for 10 min. The supernatant was stored at −20 °C for HPLC analysis.

To investigate the inhibitory potential of the purchased compounds, standard sample mixtures (conditions see above) were preincubated with varying molecule concentrations at 37 °C for 2 min under shaking. Based on the results of preliminary experiments, five concentrations of each inhibitor were used in the final experiment. They were applied in the following ranges: 1 to 10 μM for 1, 0.5 to 8 μM for 2, 0.1 to 1 μM for 3, 30 to 140 μM for 4, 50 to 800 nm for 5 and 6, 0.25 to 4 nm for 7, 5 to 80 μM for 8, and 75 to 300 nm for 9. Then, 10 μM α-naphthol was added and the reaction was treated as described above.

To investigate potential sulfonation of purchased compounds numbers 1 to 4, and 7 to 9, samples were incubated with
each compound in concentrations close to their IC\textsubscript{50} values in the absence of the substrate α-naphthol. Standard sample mixtures were incubated at 37 °C for 3 h under gentle shaking with the following compound concentrations: 8 μM of 1, 4 μM of 2, 1 μM of 3, 200 μM of 4, 1 mM of 7, 30 μM of 8, and 200 nM of 9. After incubation, reactions were terminated by heat inactivation at 95 °C and samples were prepared for subsequent liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis as described above.

All samples were determined in triplicate except for the above mentioned preliminary tests that were performed to refine the concentration range for inhibition. GraphPad Prism 5 from GraphPad Software (La Jolla, CA) was used for data analysis. \( K_m \) and \( V_{max} \) for α-naphthol sulfonation were determined to be 2.82 ± 0.49 μM and 1.43 ± 0.08 nmol min\(^{-1}\) mg\(^{-1}\), respectively, via the Michaelis-Menten model in Prism. IC\textsubscript{50} values were determined by nonlinear regression (four parametric logistic standard curve analysis function).

**HPLC Analysis**—Aliquots from the enzyme assay were transferred into HPLC glass vials and sealed with septa. α-Naphthol and α-naphthylsulfate were separated by HPLC ( Dionex, Idstein, Germany) using a NovaPak C18 column (4 μm, 150 × 3.9 mm) from Waters (Eschborn, Germany) at 30 °C isocratically with 0.1 M KH\textsubscript{2}PO\textsubscript{4} containing (v/v) 0.1% acetic acid, 0.75% isopropyl alcohol, and 4% methanol at a flow rate of 0.7 ml/min. Both analytes were detected at 280 nm using an ultraviolet diode array detector and α-naphthylsulfate was detected using fluorescence detection (\( \lambda_{ex} = 280 \text{ nm}, \lambda_{em} = 340 \text{ nm} \)). Calibration curves of α-naphthylsulfate were prepared for quantification with standards ranging from 1 to 1,000 nM.

**Identification of Sulfonated Products by LC-MS/MS**—None of the sulfonated products of compounds 1–4 and 7–9 were commercially available (SciFinder search) and, therefore, LC-MS/MS parameters could not be optimized for each individual substance due to the lack of standard material. Thus, identical chromatographic parameters and instrumental settings of the tandem-mass spectrometer, based on a method of Engst et al. (54), were used for identification of all sulfo-conjugates investigated. Analyses were conducted with an Agilent 1200 Infinity LC system coupled to an Agilent 6490 triple quadrupole-mass spectrometer (both from Waldbronn, Germany) interfaced with an electrospray ion source operating in the negative ion mode (ESI−). Chromatographic separation was carried out using an Agilent Poroshell 120 EC-C18 column (2.7 μm, 3 × 50 mm). 10 mM Ammonium acetate/methanol (90:10, v/v) and acetonitrile/methanol (95:5, v/v) were used as eluents A and B, respectively. Samples of the SULT1E1 substrate assay were centrifuged (21,800 \( \times g \) at 4 °C for 10 min) and 10 to 20 μl of the supernatants were injected into a mobile phase consisting of 90% eluent A. Sulfates were eluted from the column, which was tempered at 30 °C, with a 3-min linear gradient to 30% eluent A at a flow rate of 0.4 ml/min. The total run time for one analysis was 7 min, including re-equilibration of the LC system. The following settings of the ESI source were used: drying gas temperature = 120 °C, drying gas flow = 11 liters/min of nitrogen, sheath gas temperature = 400 °C, sheath gas flow = 12 liters/min of nitrogen, nebulizer pressure = 40 psi, capillary voltage = 3000 V, nozzle voltage = 1500 V. To identify sulfonated products of the seven compounds investigated, the following MS and MS/MS experiments were conducted. First, samples were screened for appropriate precursor ions by full scan MS mode (m/z 100 to 500). Second, identified precursor ions were fragmented in the collision cell of the mass spectrometer and characteristic product ions were detected using product ion scans (low mass cutoff: m/z 50). Different collision energies were tested (0 to 70 V) to obtain optimal product ion mass spectra. The optimized collision energy for each sulfonation product is given in the legend of Fig. 9.

**Results**

**Prediction Model Development in a Nutshell**—To develop a computer-based prediction model for SULT1E1 activity, a specific in silico workflow was developed. The first step of the model development was to perform MD simulations of the apo and cofactor-bound structures of SULT1E1 to address the degree of structural flexibility and therefore the broad substrate spectrum of the enzyme. Analyzing the MD trajectories, the average distance between atoms can be determined via the r.m.s. deviation. Geometry- or r.m.s. deviation-based clustering of the MD trajectories enabled extraction of diverse enzyme conformations. These conformations were used as templates in a subsequent ensemble docking step with active SULT1E1 ligands that were collected from literature beforehand. Based on selected docking results, eight specific three-dimensional pharmacophores were created that enable identification of active ligands of SULT1E1. These pharmacophores serve as a key element of the prediction model and allow efficient virtual screening of any given molecule set during the prediction procedure. For further refinement of the pharmacophore prediction, additional post-filtering steps and SVM were developed to increase the overall predictive power. Individual steps of our prediction approach and results will be discussed in the following sections.

**Molecular Dynamics Simulations and Structural Investigations**—In general, metabolic enzymes like SULTs show broad substrate spectra to cover a wide range of chemically different endo- or exogenous molecules. It has been shown that SULTs show a highly flexible active site that is surrounded by three flexible loops (Fig. 1): loop 1 (amino acids 85 to 89), loop 2 (amino acids 144 to 149), and loop 3 (amino acids 234 to 262) (55). The former one is only found in the SULT1 family and not in SULT2 (55). The variable sequence identities among the different SULT subtypes and the high degree of flexibility are key factors for the broad and overlapping substrate specificities of SULTs (6). Loop 3 is simultaneously covering substrate- and cofactor-binding sites, and is divided by a hinge region (56). It was shown that cofactor binding to the enzyme causes alterations in the substrate-binding site mediated by loop 3, which subsequently influences ligand selectivity (57, 58). The flexibility of SULTs and the cooperativity between cofactor- and ligand-binding events pose a challenge for developing accurate prediction models. Because MD simulations are apt to mechanistically understand the dynamic nature of proteins (59), MD simulations were performed using the apo and cofactor-bound monomeric structure to sample the conformational space and address the protein flexibility that is responsible for the broad
range of substrates. During simulations, the loop regions surrounding the substrate binding site showed increased movement in the C\textalpha\,-atoms compared with the rest of the protein (Fig. 1), whereas the area surrounding the cofactor PAPS including amino acids His-107 and Lys-105, which are important for the catalytic reaction, remains less flexible. Comparing the extent of observed movement between the three loops, loop 1 is twice as flexible as loops 2 and 3 with fluctuations of 6.8 Å in loop 1 and fluctuations of 3.6 Å and 3.3 Å in loops 2 and 3, respectively. The pliability of loop 1 can significantly modulate the shape of the active site and therefore influence binding events. In addition, amino acid Lys-85 on loop 1 was identified as an essential element. Its inward flip causes blockage of the active site entry, which constrains potential ligand binding to smaller molecules. We surmise that the side chain of Lys-85 also operates as a lid that entraps a molecule once it found its ways into the active site (Fig. 2B). Measurements of the distances between the nitrogen of Lys-85 and sulfur of the cofactor PAPS show distance variations of up to 13 Å, which is significant compared with relatively constant distances like the one between the nitrogen of His-107 and sulfur of cofactor PAPS with a distance range of 3.1 Å. The reported division of loop 3 into two parts that cover the cofactor and substrate binding site (56) also became apparent during simulations. The part that covers the PAPS-binding site stays fixed compared with the loop segment that covers the active site, which is able to move independently and fluctuates between an open and closed conformation when PAPS is absent (see also Fig. 3, left side). This suggests that, once PAPS is bound to the enzyme, the energy barrier to release the cofactor might urge the enzyme to remain in the cofactor-bound state. Furthermore, we assume that this rigid lid formed by loop 3 promotes the formation of catalytically incompetent dead-end complexes (PAP-enzyme-ligand complexes) that are a commonly observed phenomenon of SULTs. The flexibility observed during MD simulations was compared with the five available crystal structures of SULT1E1, namely 1G3M (44), 1HY3 (29), 4JVL (60), 4JVM (60), and 4JVN (60). The alignment of crystal structures indicates almost identical conformations (Fig. 2A) with only minor differences in loop 3. Furthermore amino acid residue Lys-85 did not exhibit the magnitude of flexibility that was observed during simulations.

To develop a prediction model that addresses target flexibility, the MD trajectories were clustered based on active site loop fluctuations to extract representative protein conformations (Fig. 3). Five apo and five cofactor-bound conformations were obtained that differ from the original MD template (PDB entry 1HY3 (29)). These cluster centers were used as a basis for the subsequent ensemble docking approach.

**Ensemble Docking and Three-dimensional Pharmacophore Development**—Because three-dimensional pharmacophores represent a valuable tool for metabolism prediction (2), our goal was to create a collection of pharmacophores that would allow identification of substrates and inhibitors of SULT1E1. Generally, a three-dimensional pharmacophore is defined as a three-dimensional arrangement of features (e.g. hydrogen bonds, electrostatic interactions, or hydrophobic contacts) that are essential for ligand-target interaction. To create a basis for the pharmacophore development, the clustered protein structures from MD simulations were used for ensemble docking with a collection of active SULT1E1 ligands. These active ligands were collected from literature and contained inhibitors, substrates, and CDLs. These CDLs were defined as substrates that are able to inhibit an enzyme at higher concentrations via substrate inhibition, which is a commonly observed phenomenon among many metabolic enzymes (37, 38). During ensemble docking, 64,000 individual ligand-protein complexes were generated and the question arose whether preferences in terms of protein conformation were observed.
In Silico Prediction of Ligand Binding to SULT1E1

Conformation emerged during docking. Therefore, the number of docking events between each protein and ligand was calculated (ranging between 0 and 100 docking events) and heat maps were created for visual analysis. The resulting heat maps are shown in Fig. 4 (left side). These maps show that most ligands were docked into apo-structures 3 and 5, and cofactor-bound structure 3, indicating that these structures are more receptive to ligand binding than other conformations. Additionally, during the ensemble docking process the selection of the most relevant structures is inherently implemented in the employed genetic algorithm used for this protocol (42). Docked ligand conformations are evaluated based on steric and energetic factors and protein preferences are established on this basis. Careful visual inspection of the ensemble docking results showed that the entry of the active site of cofactor-bound structure 3 was more open than in structures 1, 2, 4, and 5. Especially loop 1 contributed to a larger active site volume, which facilitates ligand binding. Furthermore, we were interested in the extent of interaction that was allowed in each ligand-protein complex. The extent of ligand-target interaction was evaluated by calculating the amount of geometrically plausible interaction pharmacophore features (e.g. hydrogen bonds or hydrophobic contacts) between each docked ligand and its target conformation. The maximum number of pharmacophore features that was determined in each of the complexes was also visualized using heat maps (Fig. 4, right side). Based on these

FIGURE 2. Comparison of structural flexibility observed in crystal structures and MD simulations of SULT1E1. A, all available crystal structures of SULT1E1 (PDB codes 1G3M (44), 1HY3 (29), 4JVL (60), 4JVM (60), and 4JVN (60)) were aligned and the experimentally observed flexibility of loops 1, 2, and 3 was assessed (distances in Å). B, mobility of amino acid residue Lys-85 on loop 1 during MD simulations (gray protein backbone) was compared with the crystal structure of SULT1E1 (yellow protein backbone), which was used as a template for the simulation (PDB code 1HY3). The residue movement was described in reference to the sulfur atom of PAPS and residue His-107 (which stay relatively stable during MD simulations). The range of movement is given as value of Δ Å and additionally as r.m.s. deviation plot.

FIGURE 3. Superposition of extracted enzyme conformations from MD trajectories via clustering in comparison to the PDB template. On the left side, the cofactor-binding site is shown for apo (green protein backbone) and cofactor-bound conformations (blue protein backbone) that were derived from MD simulations. The superposed conformations depicted in the middle and on the right illustrate the diversity of active site shapes caused by loop movements. The crystal structure of SULT1E1 (PDB entry 1HY3) is highlighted in yellow and cofactor PAPS is represented as ball-and-stick model. The overall r.m.s. deviation differences between the five apo or cofactor-bound structures and the PDB entry are given as a correlation map.
In Silico Prediction of Ligand Binding to SULT1E1

FIGURE 4. Statistical visualization of the ensemble docking results of SULT1E1 ligands and the 10 protein conformations (apo and cofactor-bound) extracted from MD simulations. On the left side, heat maps indicate the number of docking events per enzyme conformation and ligand (color scale ranging from black to yellow, maximum number of docking events = 100). On the right, heat maps indicate the maximum number of observed pharmacophore features per protein-ligand complex (black-green-yellow color scale). Pharmacophore features include hydrogen bonds, hydrophobic areas, ionic, or aromatic interactions. Lighter colors imply that the docking conformation of a ligand into one of the 10 conformations enabled increased pharmacophore feature formation.

results, three protein conformations (apo conformations 3 and 5, and cofactor-bound conformation 3) were determined that form most interactions with docked ligands, thus representing structures that are likely to facilitate inhibitor binding. A connection was observed between protein preferences during docking (i.e., protein conformations that were used as a basis for most docking events) and the ability to allow extensive ligand-target interactions.

After quantitative analysis via heat maps, the next step was to inspect individual ligand-protein complexes qualitatively. Due to known issues of unreliability of docking scoring functions (61), the pharmacophore fit of each individual ligand-protein complex complemented visual inspection of the ligand conformations and allowed guided choices of plausible binding modes. During extensive and careful visual analysis of the docking complexes of the apo structures, ligands were frequently found in inappropriate positions, e.g., sliding into the PAPS-binding site. Therefore, many docking solutions that involved apo structures were not further evaluated for pharmacophore development. As a starting point for the development of three-dimensional pharmacophores that would serve prediction of SULT1E1 ligands, the focus was narrowed down to PAPS-bound docking conformations because this is the most relevant initial situation for ligand binding (compared with an empty and therefore catalytically incompetent enzyme). Protein-ligand conformations that are concordant with the catalytic mechanism of sulfonation are an important criterion for valid pharmacophores and the composition of the Michaelis complex was considered during model development (Fig. 5). This state is initiated by residue His-107, which deprotonates the substrate hydroxyl group and thus enables the nucleophilic oxygen to attack the sulfur atom of PAPS. Assisted by residues Lys-105 and Lys-47, the sulfonate group is hydrolyzed from PAP and transferred to the substrate. This transition state renders hydrogen bonds from the hydroxyl group of a potential substrate to histidine and lysine important for pharmacophore creation. Additionally, the active site of SULT1E1 is occupied by a series of lipophilic amino acid residues (Tyr-20, Phe-23, Phe-75, Phe-80, Phe-138, Phe-141, Val-145, Tyr-168, Tyr-239, Leu-242, Ile-246, Met-247, and Phe-254) in a barrel-like manner. Under consideration of these attributes, eight three-dimensional pharmacophores were created that represent different binding modes between the enzyme and specific substrates, inhibitors, or CDLs (Fig. 6). Using a test set of active ligands and 2657 decoys, these pharmacophores were validated showing a total sensitivity of 60% and individual sensitivities for inhibitors, substrates, and CDLs of 56, 33, and 80% respectively.

The total amount of decoys (i.e., presumably inactive molecules) was 2%, indicating high specificity of the eight pharmacophores.

Prediction Model Refinement and Machine Learning Approach—Testing the eight pharmacophores in a trial virtual screening first results indicated overlaps in hit identification, meaning that, for example, in some cases a hit molecule was identified simultaneously as substrate and inhibitor (see also Fig. 7). To improve ligand identification, we developed additional elements for the final prediction model involving machine learning models and a compound filtering step. SVM for substrate and inhibitor classification were developed (SVM-S and SVM-I) and the quality of the models was assessed in terms of sensitivity (Se), specificity (Sp), accuracy (ACC), and Matthews’s correlation coefficient (given in Table 1). For model development, physicochemical properties and pharmacophore fit scores served as descriptors. The best performing models were derived via manual, iterative selection processes and were based on a set of descriptors encoding the lipophilicity (cLogP), the topological surface area, the number of hydrogen bond donors, and the PFS (description provided in the methodological section). The substrate classification model (SVM-S) correctly predicted 91% of the test set molecules and the Matthew’s correlation coefficient of 0.38 indicates a fairly robust classification performance. With a sensitivity of 0.87 and a specificity of 0.91, the SVM-S model allows relatively reliable identification of true positives and negatives. The inhibitor classification model (SVM-I) enables correct classification of true inhibitors (sensitivity of 1), and solidly predicts true negatives (specificity of 0.84). In total, the SVM-I model correctly predicted 85% of the test set molecules and the Matthew’s correlation coefficient of 0.32 indicates above average quality of prediction (a value of 0 indicates random and 1 perfect prediction).

The general prediction procedure is illustrated in Fig. 7 and operates as follows. The eight specific three-dimensional pharmacophores are used to screen a given database and depending on the pharmacophore, hits are accumulated into the categories of substrates, inhibitors, and CDLs. If a hit shows ambiguous prediction (e.g., simultaneously as substrate and inhibitor),
it is temporarily integrated into the group of CDLs. Pharmacophore hits that were predicted to be substrates would undergo OH-filtering to ensure the molecules feature a hydroxyl group (OH) and are competent to undergo potential sulfonation. Only hits that were classified as "active" during pharmacophore screening as well as SVM classification were considered substrates of SULT1E1. Similarly for inhibitor classification, predicted inhibitors from the pharmacophore screening would undergo SVM-I classification and the prediction would only be considered valid if both steps were in agreement. In the case of CDL-categorized hits, these molecules first undergo OH filtering to ensure catalytic competency. Second, these compounds are submitted to both SVM models (for substrate and inhibitor classification). In case both SVM models declare a hit to be active, the compound is considered as valid CDL (i.e. the compound has the potential to be a substrate and simultaneously an inhibitor in a concentration-dependent manner). If only the SVM-S or -I classification step indicate activity, the compound is considered a substrate or inhibitor, respectively. These SVM models ensure clarification of the pharmacophore-predicted identity of SULT1E1 molecules into substrates, inhibitors, and CDLs and serve as a refinement of the prediction procedure.

Application of the Prediction Model to the DrugBank—The prediction model was used to screen the DrugBank (48) that includes marketed and experimental drugs (6,494 molecules). The preliminary pharmacophore screening resulted in 131 hits (94 hits without overlapping prediction). After filtering and SVM classification a total number of 68 molecules were predicted to be active including 23 substrates, 12 inhibitors, and 33 CDLs.

FIGURE 5. The process of three-dimensional pharmacophore development based on ensemble docking conformations of the cofactor-bound enzyme. During the catalytic mechanism of sulfonation, the Michaelis complex is formed, in which the sulfonate moiety of PAPS is transferred to a substrate that features a hydroxyl group. His-107 and Lys-105 play a central role during this process and the distance and feature formation between a potential ligand and these residues was taken into account during pharmacophore development. Depending on the nature of the ligand (substrate or inhibitor), specific three-dimensional pharmacophore models were developed that enable identification of active molecules. The pharmacophore features include hydrophobic areas (yellow spheres), and hydrogen bond donors/acceptors (arrows in green/red).

FIGURE 6. Docking conformations of active SULT1E1 ligands and their associated, validated three-dimensional pharmacophores. The docking conformations of a substrate (1), two CDLs (2 and 3), and five inhibitors (4–8) were chosen as templates for three-dimensional pharmacophore development. The three-dimensional pharmacophores were developed and iteratively refined to achieve high predictive power based on the training data set, and were validated using a validation test set (see "Experimental Procedures"). Exclusion volumes, which were created based on the active site shape to incorporate geometric restrictions, are not shown for reasons of clarity. The three-dimensional pharmacophore features include hydrophobic areas (yellow spheres), hydrogen bond donors/acceptors (arrows or spheres in green/red), and aromatic areas (blue disks). Ligands used were: 1, 2-(4-dimethylaminophenyl)-1,3-benzothiazol-6-ol; 2, kaempferol; 3, 17-β-estradiol; 4, 2-OH-7,8-dichlorodibenzo-p-dioxin; 5, 4-OH-2,3,5,2’4’,5’-hexachlorobiphenyl; 6, 2-OH-1,3,7,8-tetrachlorodibenzo-p-dioxin; 7, daidzein-4-sulfate; 8, 4-OH-2,2’,4’,6-tetrachlorobiphenyl.
reported interaction with SULTs (35.3% of hits) including 19 compounds interacting with SULT1E1 (27.9%). Of the remaining hits, 29 were commercially not available (42.7% of the total hit collection). Analyzing the last 15 compounds, two were found to be unstable under normal conditions (DB entries 00162 and 02699) (62). Based on structural diversity, 9 of the remaining 13 molecules were selected and purchased (Fig. 8).

Experimental Validation of the Predicted Inhibitory Molecules—A total of nine selected compounds was chosen for experimental validation of predicted substrates, CDLs, and inhibitors. Due to the nature of the pharmacophores, molecules were assumed to fit and bind the active site of SULT1E1 and thus inhibition would take place in a competitive manner. First, inhibition assays were performed and all molecules were shown to inhibit SULT1E1 (Table 2 and Fig. 8). The two predicted inhibitors 5 and 6 showed IC₅₀ values of 310 and 230 nM, respectively. Three of the CDL molecules showed IC₅₀ values in the low μM range (IC₅₀ of 5.33, 3.15, and 0.52 μM for compounds 1, 2, and 3, respectively) and one in the mid-micromolar range (IC₅₀ of 89.3 μM for compound 4). The predicted substrates showed a fairly wide IC₅₀ range between 1.30 mM (compound 7), 21.3 μM (compound 8), and 90 nM (compound 9). The prediction of inhibitors and CDLs was in agreement with the experimental results because all these compounds were able to inhibit sulfo-conjugation of α-naphthol by SULT1E1.

Confirmation of Predicted Substrates by Identification of Corresponding Sulfonation Products Using LC-MS/MS—The next step was to determine sulfonation of the predicted molecules (1 to 4 and 7 to 9). The mass spectrometry-based detection of sulfonated substrates was performed qualitatively due to lack of standards. Initially, samples of the SULT1E1 substrate assay, in which compounds 1 to 4 and 7 to 9 were incubated with PAPS and cytosolic protein of bacteria expressing human SULT1E1, were screened for sulfonation products using full scan MS. For each compound a corresponding precursor ion ([M-H]⁻/(m/z): 335.0, 338.8, 349.9, 416.2, 320.8, 371.0, and 367.2 for compounds 1, 2, 3, 4, 7, 8, and 9, respectively) representing a singly sulfonated metabolite could be identified. These signals were not detected when negative control samples (without addition of PAPS) were analyzed. Sulfates of compounds 1 to 4 and 8 and 9 eluted from the separation column with retention times in the range of 2.9 to 3.6 min. The sulfate of compound 7 eluted earlier at 1.2 min. It should be noted that compounds 1 and 2 showed multiple isomeric LC peaks for the singly sulfonated metabolite could be identified. These signals were not detected when negative control samples (without addition of PAPS) were analyzed. Sulfates of compounds 1 to 4 and 8 and 9 eluted from the separation column with retention times in the range of 2.9 to 3.6 min. The sulfate of compound 7 eluted earlier at 1.2 min. It should be noted that compounds 1 and 2 showed multiple isomeric LC peaks for the singly sulfonated metabolite could be identified. These signals were not detected when negative control samples (without addition of PAPS) were analyzed. Sulfates of compounds 1 to 4 and 8 and 9 eluted from the separation column with retention times in the range of 2.9 to 3.6 min.

### Table 1

SVM models for substrate (SVM-S) and inhibitor (SVM-I) classification of pharmacophore screening hits

| Model | TP    | TN    | FP    | FN    | Se  | Sp  | ACC  | MCC  |
|-------|-------|-------|-------|-------|-----|-----|------|------|
| SVM-S | Final model* | 20 | 890  | 89   | 3   | 0.87| 0.91| 0.91 | 0.38|
|       | Two-dimensional descriptors only | 21 | 858  | 121  | 2   | 0.91| 0.88| 0.88 | 0.34|
|       | PFS descriptors only | 21 | 824  | 155  | 2   | 0.91| 0.84| 0.84 | 0.30|
| SVM-I | Final model* | 27 | 1063 | 199  | 0   | 1   | 0.94| 0.85 | 0.32|
|       | Two-dimensional descriptors only | 26 | 901  | 361  | 1   | 0.96| 0.71| 0.72 | 0.21|
|       | PFS descriptors only | 26 | 537  | 725  | 1   | 0.96| 0.43| 0.44 | 0.11|

* Selected descriptors: PFS-A1, -A2, -I3, -I5, -S, cLogP, and Rel. TPSA, #Don.

* Selected descriptors: PFS-A1, -A2, -I3, -S, cLogP, Rel. TPSA, and #Don.
not for compound 9. Product ion scans were performed to identify characteristic fragments of the singly sulfonated products. The product ion mass spectra of all sulfonated compounds (Fig. 9) possessed an abundant fragment ion that is formed by cleavage of a sulfo group and, consequently, represents a characteristic \([\text{M-H-SO}_3^-]\) product ion \((m/z: 255.0, 258.9, 269.9, 336.2, 240.9, 291.3,\) and 287.3 for compounds 1, 2, 3, 4, 7, 8, and 9, respectively). Moreover, spectra of three sulfates (compounds 2, 4, and 9) additionally showed the accordant counterpart ion, \(\text{SO}_3^-\) \((m/z 80)\). To conclude, the mass spectrometric detection of deprotonated molecular ions \([\text{M-H}]^-\), \([\text{M-2H}]^{2-}\) and characteristic fragment ions \([\text{M-H-SO}_3^-], \text{SO}_3^-\) proved the presence of sulfonation products of compounds 1 to 4 and 7 to 9 and, consequently, confirmed the in silico prediction.

**Discussion**

To date, only two in silico prediction models have been reported that utilize a structure-based approach including protein flexibility (28, 63) and one of them proposed activity prediction for SULT subtype 1E1 (28). Both employ the strategy of performing MD simulations to sample the conformational space of different SULT subtypes as a basis for their prediction model. The study on SULT1E1 (28), although not providing experimental data for validation of their prediction, used inter-

---

**TABLE 2**

**Overview on in silico prediction and experimental results of selected molecules**

| Compound number | DB entry | In silico prediction | IC_{50} | Sulfonation |
|-----------------|----------|----------------------|--------|------------|
| 1               | DB03285  | CDL                  | 5.33 ± 0.45 | +          |
| 2               | DB07708  | CDL                  | 3.15 ± 0.51 | +          |
| 3               | DB06832  | CDL                  | 0.52 ± 0.04 | +          |
| 4               | DB08770  | CDL                  | 89.3 ± 3.2  | +          |
| 5               | DB08950  | Inhibitor            | 0.31 ± 0.05 | NA*        |
| 6               | DB00785  | Inhibitor            | 0.23 ± 0.05 | NA*        |
| 7               | DB07800  | Substrate            | 1.298 ± 140 | +          |
| 8               | DB08252  | Substrate            | 21.3 ± 1.2  | +          |
| 9               | DB07702  | Substrate            | 0.09 ± 0.01 | +          |

* NA, not applicable.

---

The in silico prediction model addressed in this article is the first experimentally validated model for SULT1E1 activity. It uses a combination of three-dimensional pharmacophores and machine learning models as cornerstones of the prediction process. During model development, flexibility of the active site of the enzyme was taken into consideration as it was shown to be a key factor for substrate selectivity (64). Areas of high flexibility at the binding site, namely the three loops, restruc- ture the shape of the binding pocket and therefore allow structurally different molecules to be accommodated. At the beginning of this study, only two SULT1E1 crystals were available (Protein Data Bank entries 1HY3 (29) and 1G3M (44)) and both structures showed very similar conformations. Because a single enzyme conformation hardly reflects its nature appropriately, we utilized MD simulations to sample its conformational space and used cluster centers of the trajectory for subsequent ensemble docking. Although there is no guarantee that the extracted structures are valid representatives of the conformational space due to the fact that during simulations low-energy states of a protein might be separated by high-energy barriers that cannot be crossed in time ranges that were applied here. Nevertheless, MD has proven beneficial in delivering physiologically relevant states (65) and led to a useful model in our case.

The conformations extracted from MD simulations were used for ensemble docking with known active ligands of SULT1E1. Docking conformations were prioritized according to two basic criteria: first, ligand conformations were analyzed based on catalytic competency, i.e. distance to catalytically important residues. Despite considerable differences between the subtypes (6, 66), ligand-cofactor distance has been used as a common theme in most docking studies on SULTs as it is a prerequisite for the sulfonation reaction. In this study, the distance criterion was applied not only to substrates but also inhibitors of SULT1E1. This approach is supported by the observation that the co-crystallized ligand (3,5,3'-5'-tetrachloro-biphenyl-4,4'-diol) in crystal structure PDB code 1G3M (44), which is reported as a subnanomolar inhibitor, mimics the
In Silico Prediction of Ligand Binding to SULT1E1

FIGURE 9. Product ion mass spectra of singly sulfonated metabolites of (1) isoliquiritigenin, (2) indazole-Cl, (3) prinaberel, (4) ZM241385, (7) 2-(4-hydroxyphenylazo)benzoic acid, (8) Amb44464666, and (9) 17-epiestriol. The spectra were numbered according to the studied compounds (see “Chemicals” section). Black rhombs indicate the deprotonated precursor ions ([M-H]⁻) that were selected to generate the accordant product ions. The predicted structural formula of each sulfonation product is given as the inset in the corresponding spectrum. For compounds 1 and 2 the in silico simulation predicted further possible singly sulfonated metabolites referred to as 1* and 2*, respectively, that are depicted in the down-most right panel. It has not been elucidated which sulfate isomer (1 or 1*, 2 or 2*) is represented by the mass spectrum given for the sulfonation product of isoliquiritigenin and indazole-Cl, respectively. MS/MS parameters: ESI–, collision energy: (1) 20 V, (2) 20 V, (3) 15 V, (4) 25 V, (7) 10 V, (8) 10 V, and (9) 35 V.
binding of substrate estradiol and is in close proximity to the catalytic residue histidine 107 (44). Second, the docking results were also investigated based on three-dimensional pharmacophore features such as hydrogen bonds, hydrophobic contacts, or aromatic areas. To the best of our knowledge, this approach has rarely been applied to SULT docking studies because primarily hydrogen bonds have been considered to evaluate docked ligand conformations (26, 67–69). Due to the nature of the active site of SULT1E1 (a barrel-like structure lined with lipophilic and/or aromatic amino acids), the consideration of hydrophobic areas and aromatic contacts bears advantages as shown by the predictive power of the presented in silico model for ligands of SULT1E1.

The resulting eight pharmacophores were based on cofactor-bound conformations because the starting point of model development was the catalytically competent state of the enzyme (i.e. bound to PAPS). The prediction model proposed in this article does not address the reported phenomenon of dead-end complexes (PAP-enzyme-ligand) yet. In vivo concentrations of PAPS can significantly vary depending on the tissue and depending on sulfonation rates of other heavily sulfonated molecules present in the cytosol (55). Sulfonation has been shown to be a high-affinity, low-capacity conjugation reaction (70, 71) and PAPS availability might influence the level of catalytic activity (72). Also, PAPS release was determined to be a rate-limiting step during sulfonation reactions (72) and “trapped,” un-sulfonated cofactor (PAP) could promote the formation of catalytically incompetent dead-end complexes (PAP-enzyme-ligand). These fluctuations in cytosol composition, molecule concentrations, and also the dynamics of cofactor release remain challenging to reproduce in a computer-based prediction model. Therefore only the two physiologically important processes of (i) direct inhibitor binding leading to decreased enzyme activity, and (ii) substrate binding leading to sulfonation were considered here. The final three-dimensional pharmacophores represent characteristic states of inhibition and sulfonation and allow efficient screening of any given database to identify potential ligands of SULT1E1.

The hit identification via pharmacophore screening was further refined by SVM models that enable classification of hits into substrates, inhibitors, and CDLs. To date, the incorporation of the pharmacophore fit score (calculated in LigandScout) as molecular descriptor into a predictive model is a novelty and solidly served our purpose to create post-filtering steps to refine pharmacophore screening results. The SVM models created here are not supposed to serve as stand-alone prediction tools (but rather serve refinement of hit identification). Therefore the choice of molecular descriptors could be kept relatively simple by including the pharmacophore fit score and physicochemical properties.

The application of the prediction model to the DrugBank resulted in 68 hits and literature search confirmed 28% to be active SULT1E1 ligands. The remaining compounds were investigated for commercial availability and a selection of nine molecules was experimentally tested. All in silico predictions were experimentally confirmed via sulfonation and inhibition assays. Two of the predicted substrates showed fairly low IC_{50} values and the question might arise why these compounds were not predicted CDLs (i.e. substrates with inhibitory potency at increasing concentrations) instead of substrates. This is probably due to the fact that development of the prediction model was based on experimental data from the literature and molecules that built the basis of the substrate prediction model were often not tested for inhibition. This implies that even though a hit from a given database is predicted to be a substrate does not preclude the possibility of having the potential to inhibit the enzyme. Therefore the model as described here predicts substrates of SULT1E1, which are competent to undergo sulfonation, whereas simultaneously not excluding the possibility of potential enzyme inhibition at higher concentrations.

The proposed computer-based prediction model enables screening of large databases and efficient identification of potential substrates and inhibitors of SULT1E1. The presented workflow has the potential to be transferred to other SULT subtypes for prediction model development and could lead to the creation of an automated SULT activity prediction pipeline that may facilitate future drug design and could also assist in risk assessment.

**Author Contributions**—C. R. performed the theoretical modeling under supervision of G. W., worked on the experiments together with F. S. under supervision of W. M. and H. G. B. K. and G. W. coordinated the study and contributed to the interpretation of the results. All authors reviewed the results and approved the final version of the manuscript.

**Acknowledgments**—We thank Martina Scholtyseck (Department of Molecular Toxicology, German Institute of Human Nutrition (DIfE Potsdam-Rehbrücke) for outstanding technical assistance and Jérémie Mortier (Institute of Pharmacy, Freie Universität Berlin) for critical reading of the manuscript. Special acknowledgments go to Susanne Dupré for exceptional computational assistance. We thank the computing center of the FU Berlin (ZEDAT) for providing the compute cluster SOROBAN for molecular dynamics calculations.

**References**

1. Guengerich, F. P. (2008) Cytochrome P450 and chemical toxicology. *Chem. Res. Toxicol.* 21, 70–83
2. Kirchmair, J., Williamson, M. J., Tyzack, J. D., Tan, L., Bond, P. J., Bender, A., and Glen, R. C. (2012) Computational prediction of metabolism: sites, products, SAR, P450 enzyme dynamics, and mechanisms. *J. Chem. Inf. Model.* 52, 617–648
3. Kirchmair, J., Göller, A. H., Lang, D., Kunze, J., Testa, B., Wilson, I. D., Glen, R. C., and Schneider, G. (2015) Predicting drug metabolism: experiment and/or computation? *Nat. Rev. Drug Discov.* 14, 387–404
4. Testa, B., Pedretti, A., and Vistoli, G. (2012) Reactions and enzymes in the metabolism of drugs and other xenobiotics. *Drug Discov. Today* 17, 549–560
5. Moroy, G., Martiny, V. Y., Vayer, P., Villoutreix, B. O., and Miteva, M. A. (2012) Toward in silico structure-based ADMET prediction in drug discovery. *Drug Discov. Today* 17, 44–55
6. Allali-Hassani, A., Pan, P. W., Dombrowski, L., Najmanovich, R., Tempel, W., Dong, A., Loppnau, P., Martin, F., Thornton, J., Thonton, J., Edwards, A. M., Bochkarev, A., Plotnikov, A. N., Vedadi, M., and Arrowsmith, C. H. (2007) Structural and chemical profiling of the human cytosolic sulfotransferases. *PLos Biol.* 5, e97
7. Zhang, H., Varlamova, O., Vargas, F. M., Falany, C. N., Leyh, T. S., and Varmalova, O. (1998) Sulfuryl transfer: the catalytic mechanism of human estrogen sulfotransferase. *J. Biol. Chem.* 273, 10888–10892
8. Kauffman, F. C. (2004) Sulfonation in pharmacology and toxicology. *Drug
In Silico Prediction of Ligand Binding to SULT1E1

9. Jancova, P., Anzenbacher, P., and Anzenbacherova, E. (2010) Phase II drug metabolizing enzymes. Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czech. Repub. 154, 103–116

10. Glatt, H. (2000) Sulfotransferases in the bioactivation of xenobiotics. Chem. Biol. Interact. 129, 141–170

11. Reinen, J., and Vermeulen, N. P. (2015) Biotransformation of endocrine disrupting compounds by selected phase I and phase II enzymes: formation of estrogenic and chemically reactive metabolites by cytochromes P450 and sulfotransferases. Curr. Med. Chem. 22, 500–527

12. Bamforth, K. J., Jones, A. L., Roberts, R. C., and Coughtrie, M. W. (1993) Estrogen sulfotransferases in breast and endometrial cancers. Int. J. Mol. Med. 2, 553–560

13. Pasqualini, J. R. (2009) Estrogen sulfotransferases in breast and endometrial cancers. Int. J. Mol. Med. 23, 453–461

14. Bamforth, K. J., Jones, A. L., Roberts, R. C., and Coughtrie, M. W. (1993) Common food additives are potent inhibitors of human liver 17β-estrogen-lyoestradiol and dopamine sulfotransferases. Biochem. Pharmacol. 46, 1713–1720

15. Coughtrie, M. W., and Johnston, L. E. (2001) Interactions between dietary chemicals and human sulfotransferases: molecular mechanisms and clinical significance. Drug Metab. Dispos. 29, 522–528

16. Kester, M. H., Buldak, S., Tibboel, D., Meinl, W., Glatt, H., Falany, C. N., Coughtrie, M. W., Bergman, A., Safe, S. H., Kuiper, G. G., Schuur, A. G., Brouwer, A., and Visser, T. J. (2000) Potent inhibition of estrogen sulfotransferase by hydroxylated PCP metabolites: a novel pathway explaining the estrogenic activity of PCBs. Endocrinology 141, 1897–1900

17. Okimoto, K., Sakakibara, Y., Suido, M., Yoshikawa, H., Liu, M.-C., and Tamura, H. (2005) Biocides, tributyltin and triphenyltin, as possible inhibitors of the human sulfotransferase involved in the estrogen homeostasis. Pestic. Biochem. Physiol. 81, 32–38

18. Steventon, G. B., Heafield, M. T., Waring, R. H., and Williams, A. C. (1989) The role of Ser137 in the sulfuryl transfer reaction. Crystal structure of the human estrogen sulfotransferase-PAPS complex: evidence for catalytic role of Ser137 in the sulfuryl transfer reaction. J. Biol. Chem. 277, 17928–17932

36. Halgren, T. A. (1996) Merck molecular force field: I. basis, form, scope, parameterization, and performance of MMFF94. J. Comput. Chem. 17, 490–519

37. Reed, M. C., Lieb, A., and Nijhout, H. F. (2010) The biological significance of substrate inhibition: a mechanism with diverse functions. Bioessays 32, 422–429

38. Wu, B. (2011) Substrate inhibition kinetics in drug metabolism reactions. Drug Metab. Rev. 43, 440–456

39. Irwin, J. J., and Shoichet, B. K. (2005) ZINC—a free database of commercially available compounds for virtual screening. J. Chem. Inf. Model. 45, 177–182

40. Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Ko¨tter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., and Wiswedel, B. (2008) KNIME: Konstanz information miner. in Data analysis, machine learning and applications, pp. 319–326, Springer, New York

41. Mysinger, M. S., Carchia, M., Irwin, J. J., and Shoichet, B. K. (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J. Med. Chem. 55, 6582–6594

42. Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol. 267, 727–748

43. Korh, O., Stützle, T., and Exner, T. E. (2009) Empirical scoring functions for advanced protein-ligand docking with PLANTS. J. Chem. Inf. Model. 49, 84–96

44. Shevtsov, S., Petrotnchenko, E. V., Pedersen, L. C., and Negishi, M. (2003) Crystallographic analysis of a hydroxylated polychlorinated biphenyl (OH-PCB) bound to the catalytic estrogen binding site of human estrogen sulfotransferase. Environ. Health Perspect. 111, 884–888

45. Wolber, G., and Langer, T. (2005) LigandScount: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J. Chem. Inf. Model. 45, 160–169

46. Wolber, G., and Dornhofer, A. A., and Langer, T. (2006) Efficient overlay of small organic molecules using 3D pharmacophores. J. Comput. Aided Mol. Des. 20, 773–788

47. Wolber, G., Seidel, T., Bendix, F., and Langer, T. (2008) Molecular-pharmacophore superpositioning and pattern matching in computational drug design. Drug Discov. Today 13, 23–29

48. Witten, I. H., and Frank, E. (2005) Data Mining: Practical Machine Learning Tools and Techniques, pp. 65–69 and 214, Morgan Kaufmann Publishers, Burlington, Massachusetts

49. Hagen, M., Pabel, U., Landsiedel, R., Bartsch, L., Falany, C. N., and Glatt, H. (1998) Expression of human estrogen sulfotransferase in Salmonella typhimurium: differences between hHST and hEST in the enantioselective activation of 1-hydroxythiophyrene to a mutagen. Chem. Biol. Interact. 109, 249–253

11. Reinen, J., and Vermeulen, N. P. (2015) Engineering sulfotransferases.” Maturitas 84–96

12. Bamforth, K. J., Jones, A. L., Roberts, R. C., and Coughtrie, M. W. (1993) Estrogen sulfotransferases in breast and endometrial cancers. Int. J. Mol. Med. 2, 553–560

13. Pasqualini, J. R. (2009) Estrogen sulfotransferases in breast and endometrial cancers. Ann. N. Y. Acad. Sci. 1155, 88–98

14. Iyer, L. V., Ramamoorthy, A., Rutkowski, E., Furismey, A. M., Tang, L., Catz, P., Green, C. E., Jozaik, K., and Wainer, I. W. (2012) The stereoselective sulfite conjugation of 4′-methoxynitrosodiesterosomers by sulfotransferase enzymes. Chirality 24, 796–803

15. Lin, P., Yang, W., Pedersen, L. C., Negishi, M., and Pedersen, L. G. (2006) Searching for the minimum energy path in the sulfonyl transfer reaction catalyzed by human estrogen sulfotransferase: role of enzyme dynamics. Int. J. Quantum Chem. 106, 2981–2998

16. Martin, Y. V., Carbonell, P., Lagorce, D., Villoitreux, B. O., Moroy, G., and Mitrovic, M. A. (2013) In silico mechanistic profiling to probe small molecule binding to sulfotransferases. PLoS ONE 8, e73587

17. Pedersen, L. C., Petrotnchenko, E., Shevtsov, S., and Negishi, M. (2002) Crystal structure of the human estrogen sulfotransferase-PAPS complex: evidence for catalytic role of Ser137 in the sulfonyl transfer reaction. J. Biol. Chem. 277, 17928–17932
51. Meinl, W., Meerman, J. H., and Glatt, H. (2002) Differential activation of promutagens by alloenzymes of human sulfotransferase 1A2 expressed in Salmonella typhimurium. Pharmacogenetics 12, 677–689.

52. Meinl, W., Pabel, U., Osterloh-Quiroz, M., Hengstler, J. G., and Glatt, H. (2006) Human sulfotransferases are involved in the activation of aristolochic acids and are expressed in renal target tissue. Int. J. Cancer 118, 1090–1097.

53. Reinen, J., Vriese, E., Glatt, H., and Vermeulen, N. P. (2006) Development and validation of a fluorescence HPLC-based screening assay for inhibition of human estrogensulfotransferase. Anal. Biochem. 357, 85–92.

54. Engst, W., Pabel, U., and Glatt, H. (2002) Conjugation of 4-nitrophenol and 4-hydroxylonazolacin V79-derived cellsexpressingindividualforms of humansulphotransferases. Environ. Toxicol. Pharmacol. 11, 243–250.

55. Tibbs, Z. E., Rohn-Glowacki, K. J., Crittenden, F., Guidry, A. L., and Falany, C. N. (2015) Structural plasticity in the human cytosolic sulfotransferase dimer and its role in substrate selectivity and catalysis. Drug Metab. Pharmacokinet. 30, 3–20.

56. Cook, I., Wang, T., Almo, S. C., Kim, J., Falany, C. N., and Leyh, T. S. (2013) The gate that governs sulfotransferase selectivity. Biochemistry 52, 415–424.

57. Cook, I., Wang, T., Almo, S. C., Kim, J., Falany, C. N., and Leyh, T. S. (2013) Testing the sulfotransferase molecular pore hypothesis. J. Biol. Chem. 288, 8619–8626.

58. Wang, T., Cook, I., and Leyh, T. S. (2014) In Silico Prediction of Ligand Binding to SULT1E1. Biochemistry 53, 6893–6900.

59. Mortier, J., Rakers, C., Bermudez, M., Murgueitio, M. S., Riniker, S., and Wolber, G. (2015) The impact of molecular dynamics on drug design: applications for the characterization of ligand-macromolecule complexes. Drug Discov. Today. 20, 686–702.

60. Gosavi, R. A., Knudsen, G. A., Birnbaum, L. S., and Pedersen, L. C. (2013) Mimicking of estradiol binding by flame retardants and their metabolites: a crystallographic analysis. Environ. Health Perspect. 121, 1194–1199.

61. Warren, G. L., Andrews, C. W., Capelli, A.-M., Clarke, B., LaLonde, J., Lambert, M. H., Lindvall, M., Nevins, N., Semus, S. F., Senger, S., Tedesco, G., Wall, I. D., Woolven, J. M., Peishoff, C. E., and Head, M. S. (2006) A critical assessment of docking programs and scoring functions. J. Med. Chem. 49, 5912–5931.

62. Blomhoff, R., and Blomhoff, H. K. (2006) Overview of retinoid metabolism and function. J. Neurobiol. 66, 606–630.

63. Cook, I., Wang, T., Falany, C. N., and Leyh, T. S. (2013) High accuracy in silico sulfotransferase models. J. Biol. Chem. 288, 34494–34501.

64. Cook, I. T., Leyh, T. S., Kadlubar, S. A., and Falany, C. N. (2010) Structural rearrangement of SULT2A1: effects on dehydroepiandrosterone and raloxifene sulfation. Horm. Mol. Biol. Clin. Investig. 1, 81–87.

65. Feixas, F., Lindert, S., Sinko, W., and McCammon, J. A. (2014) Exploring the role of receptor flexibility in structure-based drug discovery. Biophys. Chem. 186, 31–45.

66. Najmanovich, R. J., Allali-Hassani, A., Morris, R. J., Dombrovsky, L., Pan, P. W., Vedadi, M., Plotnikov, A. N., Edwards, A., Arrowsmith, C., and Thornton, J. M. (2007) Analysis of binding site similarity, small-molecule similarity and experimental binding profiles in the human cytosolic sulfotransferase family. Bioinformatics 23, e104–109.

67. Meng, S., Wu, B., Singh, R., Yin, T., Morrow, J. K., Zhang, S., and Hu, M. (2012) SULT1A3-mediated regiospecific 7-O-sulfation of flavonoids in Caco-2 cells can be explained by the relevant molecular docking studies. Mol. Pharm. 9, 862–873.

68. Butt, C. M., and Stapleton, H. M. (2013) Inhibition of thyroid hormone sulfotransferase activity by brominated flame retardants and halogenated phenolcs. Chem. Res. Toxicol. 26, 1692–1702.

69. Tian, X., Huo, X., Dong, P., Wu, B., Wang, X., Wang, C., Liu, K., and Ma, X. (2015) Sulfation of melatonin: enzymatic characterization, differences of organs, species and genders, and bioactivity variation. Biochem. Pharmacol. 94, 282–296.

70. Kim, H. J., Cho, J. H., and Klaassen, C. D. (1995) Depletion of hepatic 3′-phosphoadenosine 5′-phosphosulfate (PAPS) and sulfate in rats by ex-nobiotics that are sulfated. J. Pharmacol. Exp. Ther. 275, 654–658.

71. Klaassen, C. D., and Boles, J. W. (1997) Sulfation and sulfotransferases 5: the importance of 3′-phosphoadenosine 5′-phosphosulfate (PAPS) in the regulation of sulfation. FASEB J. 11, 404–418.

72. Wang, T., Cook, I., Falany, C. N., and Leyh, T. S. (2014) Paradigms of sulfotransferase catalysis: the mechanism of SULT2A1. J. Biol. Chem. 289, 26474–26480.