Comparative Molecular Field Analysis of Substrates for an Aryl Sulfotransferase Based on Catalytic Mechanism and Protein Homology Modeling

Vyas Sharma and Michael W. Duffel*

Division of Medicinal and Natural Products Chemistry, College of Pharmacy, The University of Iowa, Iowa City, Iowa 52242

Received October 18, 2001

Comparative Molecular Field Analysis (CoMFA) methods were used to produce a 3D-QSAR model that correlated the catalytic efficiency of rat hepatic aryl sulfotransferase (AST) IV, expressed as $\log(k_{cat}/K_m)$, with the molecular structures of its substrates. A total of 35 substrate molecules were used to construct a CoMFA model that was evaluated on the basis of its leave-one-out cross-validated partial least-squares value (q^2) and its ability to predict the activity of six additional substrates not used in the training set. The model was constructed using substrate conformations that favored (1) proton abstraction by the catalytic histidine residue, (2) an in-line sulfuryl-group transfer mechanism, and (3) constraints imposed by the residues lining the substrate binding pocket of a homology model of AST IV. This CoMFA model had a q^2 value of 0.691, and it successfully predicted the activities of the six molecules not used in the training set. A final CoMFA model was constructed using the same methodology but with molecules from both the training set and the test set. Its q^2 value was 0.701, and it had a non-cross-validated r^2 value of 0.922. The contour coefficient map generated by this CoMFA was overlaid on the amino acids in the substrate-binding pocket of the homology model of AST IV and found to show a good fit. Additionally external validation was obtained by using the CoMFA model to design substrates that show high activities. These results establish a methodology for prediction of the substrate specificity of this sulfotransferase based on CoMFA methods that are guided by both the homology model and the catalytic mechanism of the enzyme.

Introduction

Cytosolic sulfotransferases catalyze the transfer of a sulfuryl group from 3′-phosphoadenosine 5′-phosphosulfate (PAPS) to an acceptor molecule forming adenosine 3′,5′-diphosphate (PAP) and the sulfuric acid ester conjugate of the acceptor molecule.[1−4] A significant number of cytosolic sulfotransferases have been characterized at the mRNA level in mammals and divided into several gene families based on the similarity of their amino acid sequences and catalytic properties.[5,6] Sulfation is most often a means of detoxication or inactivation of a drug or other xenobiotic substrate, but it may also serve as step toward bioactivation leading to carcinogenesis and various other toxicological responses.[3,7−10]

Aryl sulfotransferase (AST) IV is a major cytosolic sulfotransferase in the rat. Although originally given the name of AST IV,[11] the enzyme has also been named as ST1A1.[6] Similarities in its structure and function to aryl (phenol) sulfotransferases in humans (e.g., SULT1A1, SULT1A2, and SULT1A3) and other species, as well as the extensive use of the rat in many decades of studies on drug metabolism, toxicology, and chemical carcinogenesis, provide a rationale for its use in model studies. Sulfation catalyzed by aryl (phenol) sulfotransferases is a significant route of conjugation for a wide variety of drugs and their hydroxylated metabolites. Characteristic examples include the sulfation of acetaminophen,[1,12] isoproterenol,[12] 4-hydroxypropranolol,[13] minoxidil,[14,15] albuterol,[16] and many others.

AST IV shows a remarkably broad substrate specificity for phenols, benzylic alcohols, and *N*-hydroxy arylamines,[4,17] and this has further contributed to its frequent use as a representative aryl (phenol) sulfotransferase. Many catecholamines,[11] tyrosine esters,[11] peptides with *N*-terminal tyrosines,[11] hydroxamic acids,[11,18−21] oximes,[22,23] benzylic alcohols,[17,24] and nitroalkanes[14,15,25] are substrates for the enzyme. Sulfuric acid esters of several AST IV substrates, such as those derived from some arylhydroxamic acids, benzylic alcohols, and *N*-hydroxyarylamines, are known to give rise to cellular necrosis and chemical carcinogenesis.[7,8,26−29]

An important feature of the recognition of molecules by AST IV is that the ability of the enzyme to catalyze the sulfation of several different chemical classes of molecules is coupled with stereospecificity and stereoselectivity.[24] In the case of chiral benzylic alcohols, the configuration of the benzylic carbon bearing the hydroxyl group is known to play an important role in determining whether molecules are substrates of AST IV. For example, only the (*R*)-(−)-enantiomer of 1,2,3,4-tetrahydro-1-naphthol is a substrate for the enzyme, while (*S*)-(+)-1,2,3,4-tetrahydro-1-naphthol is a competitive inhibitor of the AST IV-catalyzed sulfation of 1-naphthalenemethanol.[24]

* To whom all correspondence should be addressed. Telephone: (319) 335-8840; FAX: (319) 335-8766; E-mail: michael-duffel@uiowa.edu.

10.1021/jm010481c CCC: $22.00
Published on Web 11/09/2002

The X-ray crystal structures of three cytosolic sulfotransferases and one Golgi-membrane sulfotransferase have been determined.[30,31] The sulfotransferases for which crystallographic data are available include the mouse estrogen sulfotransferase (mEST),[32] human dopamine (aryl) sulfotransferase,[33,34] the sulfotransferase domain of the human heparan sulfate *N*-deacetylase/*N*-sulfotransferase 1,[35] and human hydroxysteroid sulfotransferase.[36,37] Seen in the core of all the above structures is a central five-stranded parallel β-sheet surrounded by α-helices. Also seen along with this common α/β fold is a conserved structural motif for PAPS binding. As a result of this progress made in the determination of sulfotransferase structures and the common structural features of all sulfotransferase structures examined to date, we have conducted homology modeling and substrate/inhibitor-docking studies on AST IV.[19]

These crystal structures have also contributed to our knowledge of the sulfotransferase reaction mechanism.[38] Cocrystallization of mEST with the PAP−vanadate complex as a transition state mimic, coupled with mutational analysis, confirmed that the sulfotransferase-catalyzed reaction proceeds by an in-line sulfuryl transfer reaction from PAPS to the substrate.[38] Early kinetic studies on AST IV were found to be consistent with a rapid equilibrium Bi Bi kinetic mechanism with two dead end product inhibitor complexes.[39] More recent investigations using burst phase kinetics[40] have elaborated on the previous mechanism. These later studies revealed that AST IV has a sequential mechanism where the enzyme first binds PAPS then substrate, followed by the sulfation of the substrate, then the release of the sulfuric acid ester and PAP, respectively.[40] When these studies on catalytic function of the sulfotransferase are combined with homology modeling based upon the structural information outlined above, they provide a framework for guiding 3D-QSAR techniques.

In this paper we demonstrate the development of a 3D-QSAR model through the application of Comparative Molecular Field Analysis (CoMFA) in combination with structural constraints derived from a consideration of the enzyme mechanism and from the structural aspects of the AST IV homology model. This 3D-QSAR model aids in understanding AST IV's substrate specificity and establishes a methodology for quantitative prediction of the interaction of substrates with this enzyme.

Materials and Methods

Substrates and Assay Reagents. 2-Naphthol, 1-naphthalenemethanol, 3-indolemethanol, 4-(1-adamantyl)phenol (Aldrich Chemical Co., Milwaukee, WI), and 4-biphenylmethanol (Fluka Chemika, Buchs, Switzerland) were recrystallized from water before their use in enzyme kinetic studies. PAPS and 2-mercaptoethanol were obtained from Sigma Chemical Co (St. Louis, MO). PAPS was purified for use in the sulfotransferase assays using a previously described procedure.[41] The synthesis and characterization of the benzylic alcohols used in this study have been reported earlier.[42,43] All other assay components and buffer reagents were from commercial sources. HPLC analyses were carried out with an Econosphere C18 column (5.4 μm, 4.6 × 250 mm) obtained from Alltech Associates Deerfield, IL.

Enzyme Assays. Rat hepatic AST IV was expressed in *Escherichia. coli*, purified, and characterized as previously described.[44] Enzyme assays were conducted by an HPLC procedure that utilizes the substrate-dependent formation of coproduct, PAP, in the determination of the progress of the AST IV-catalyzed reaction.[45] Reaction mixtures of 0.03 mL total volume contained 0.2 mM PAPS, 8.3 mM 2-mercaptoethanol, 1.0 μg of AST IV, varying concentration of the substrates, and 0.25 M potassium phosphate at pH 7.0. *p*-Alkyl benzylic alcohols, chiral benzylic alcohols, 4-(1-adamantyl)-phenol, 4-biphenylmethanol, and 3-indolemethanol were evaluated as substrates. Following incubation for 2 min at 37 °C, the reaction was initiated by the addition of enzyme, conducted for 10 min at 37 °C, and terminated by the addition of 0.03 mL of methanol. Control assays to determine the substrate-independent formation of PAP were carried out with all assay components except the substrates. Additionally, standard assays using either 2-naphthol or 1-naphthalenemethanol were conducted to ensure that the enzyme activity was retained after storage of the enzyme at −70 °C.

Homology Model. The structure of AST IV was modeled using the crystal structure of mouse estrogen sulfotransferase (mEST)[32] as a template by fitting the AST IV sequence into the electron density map of mEST.[19] Structural manipulations on the homology model and ligands were performed using the molecular modeling package SYBYL 6.5 (Tripos Associates, St. Louis, MO). The structure of a proposed mimic of the transition state intermediate of the sulfotransferase reaction catalyzed by AST IV was modeled based on coordinates that were extracted from Protein Data Bank (PDB) file 1BO6, which contained coordinates of an mEST−PAP−vanadate complex.

Alignment Rule. After the construction of the structures of the substrate molecules to be used in the CoMFA, charges were calculated using the Gasteiger−Hückel method. The resulting structures were subsequently placed into the substrate-binding pocket of AST IV. For the minimization into the homology model active site, the enzyme was given Kollman all-atom charges, and the Powel algorithm along with the Tripos force field was used; a minimum energy change of 0.01 kcal/mol or 200 iterations was set as the convergence criteria. The amino acid residues in the homology model were held fixed as were the atoms that are key elements in the mechanism of sulfuryl transfer. The alignments of minimum energy conformers obtained by this method were then used for CoMFA.

CoMFA Methods. CoMFA studies were carried out using SYBYL 6.5. CoMFA fields were generated using the Tripos standard field class for CoMFA. An sp^3 carbon atom probe with a +1 charge was used to set the steric and electrostatic field energies. Calculations of the fields were based on a smooth transition and a 30.0 kcal/mol cutoff. The biological activity of the compounds was expressed as $\log(k_{cat}/K_m)$ where k_{cat} is the turnover number of the enzyme expressed in min^{-1} and K_m is the apparent Michaelis constant of the substrate expressed in M. CoMFA equations were calculated by the partial least squares (PLS) algorithm available with the SYBYL 6.5 software. For each calculation the activity column was used as the dependent column and the column filter was set at 1.0 kcal/mol. In the leave-one-out cross validation technique for the determination of q^2, the default parameters of the PLS algorithm were used. Namely, 100 iterations were used with no bootstrapping runs or centering, and the COMFA_STD method was applied with 0.0001 as convergence criteria for preanalysis scaling using all components and cross-validation groups. The number of components with the lowest standard error of prediction values obtained by the leave-one-out cross-validation technique was selected as the optimal number of components to be used in the PLS analysis with no validation for the determination of an r^2. All other computational parameters were kept unchanged.

Results

Construction and Evaluation of the Model. Homology Model. Homology modeling of sulfotransferases is aided by the high degree of structural similarity that is present among those sulfotransferases for which crystal structures are known.[30,31] The homology

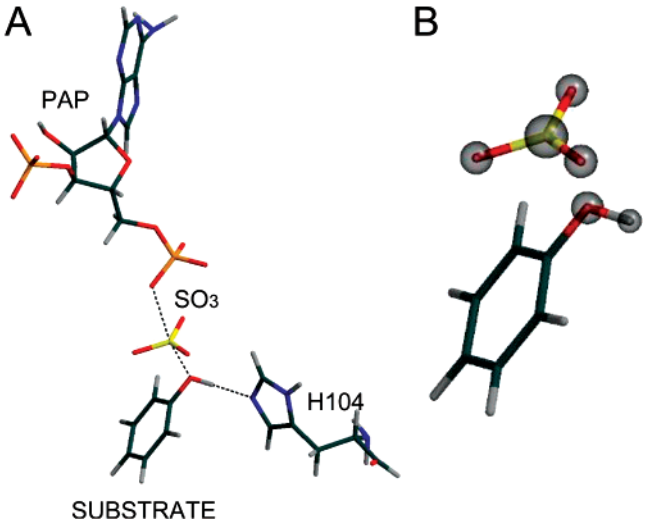

Figure 1. The mechanism of sulfotransferase-catalyzed reactions as related to the AST IV homology model. A. The atoms of phenol and PAP are in the same position with respect to the AST IV structure as the A ring of estrogen and PAP are with respect to the EST crystal structure from PDB file 1AQU. The atoms of SO_3 are in the same position with respect to the AST IV structure as are the atoms of VO_3 in the EST crystal structure PDB file 1BO6. B. The atoms that were held fixed to maintain the transition-state constraints in the ligand are shown as semitransparent spheres. These included the SO_3 group and the hydroxyl group of the substrate. These structures were drawn using MOLSCRIPT[58] and Raster3D.[59]

model of AST IV constructed for the present studies was based on the crystal structure of mouse estrogen sulfotransferase (mEST), as previously reported for analysis of *N*-hydroxy arylamine substrates.[19] This procedure included fitting the sequence of AST IV to the electron density map of mEST and using XPLOR 3.85 [46] for initial positional refinement with subsequent simulated annealing using structure factors originally obtained from X-ray diffraction data collected on mEST. The position of PAP and the acceptor substrate in AST IV were modeled as previously described[19] on the basis of the coordinates of the mEST−PAP−estradiol complex.[32] Likewise, the coordinates of a mEST−PAP−vanadate complex[38] were utilized to model the position of the sulfuryl group in AST IV by replacing the vanadium with a sulfur atom.[19] Although a short disordered region of the crystal structure of mEST (Gly60−Ala69) caused the corresponding portion of the homology model to be constructed manually, the resulting model was consistent with our previous affinity labeling of Lys65 and Cys66.[47] The overall quality of the AST IV homology model was high, with structures of estrogen sulfotransferase and AST IV exhibiting a root-mean-square deviation of 0.40 over the best-fit α carbon backbone.

A structural representation of the mechanism of AST IV-catalyzed sulfation that is based on this homology model is shown in Figure 1A. According to this mechanism, consistent with previous kinetic studies performed on AST IV, His104 abstracts a proton from the substrate in conjunction with the transfer of a sulfuryl group from PAPS. Figure 1B shows the atoms that were held as fixed aggregates to assign transition state constraints in the alignment procedure for CoMFA.

Dataset. An initial requirement for meaningful 3D-QSAR is a dataset that possesses structural diversity and a broad distribution of activity. Compounds shown in Table 1 were used for the generation of the first CoMFA, and compounds shown in Table 2 were used for its evaluation. This set included *p*-substituted phenols, *p*-substituted benzylic alcohols, chiral benzylic alcohols, *N*-hydroxy arylamines, naphthols, and naphthalenemethanols. The catalytic efficiency of the enzyme in catalyzing the sulfation of these compounds was expressed as $\log(k_{cat}/K_m)$ where k_{cat} is the turnover number of the enzyme expressed in min^{-1}, and K_m is the apparent Michaelis constant of the substrate expressed in M. The values of k_{cat} are defined by the maximal velocity, V_{max}, divided by the total molar concentration of the enzyme. The dataset showed $\log(k_{cat}/K_m)$ values ranging from 2.04 to 6.68.

Alignment. In addition to a structurally diverse dataset, knowledge of the active conformation of the molecules in the dataset is required for 3D-QSAR techniques using molecular field analysis. From the crystal structures of mEST, the substrate, estradiol, was observed to bind in a manner characteristic for the abstraction of a proton by the catalytic histidine and the in-line transfer of a sulfuryl group from PAPS.[38] In the case of the substrates of AST IV the number of possible conformations could be limited by determining those geometries that favor both the abstraction of a proton from the phenol or benzylic alcohol and an in-line sulfuryl-group transfer. This can be accomplished by conducting a systematic search for the minimum energy conformation for each ligand in the presence of a fixed sulfuryl group while holding the position of the attacking oxygen atom and abstracted proton of the substrate fixed. Figure 2A provides a representation of the results of this alignment procedure as applied to the compounds in Table 1. In the absence of protein-structure data this would be the best method for building an alignment that takes into account the mechanism of sulfation. However the CoMFA model using this alignment methodology gave a very low cross-validated q^2 value of 0.007 for the compounds shown in Table 1. As the q^2 value provides information about the internal consistency and the predictive capacity of a model, this extremely low q^2 for a model built without the inclusion of enzyme structural constraints suggests that the substrates are most likely not binding in the conformations obtained without consideration of the environment of the substrate-binding site. Thus, a correct alignment that incorporates the active site of the enzyme is crucial to obtain a correlation between activities and substrate structures.

As it is likely that the structure of the enzyme is a major determinant of relevant substrate conformations, we devised a procedure that includes interactions with residues from the active site of an AST IV homology model. This methodology is outlined in Figure 3. In this procedure, each substrate was merged into the substrate-binding pocket of the AST IV homology model after assigning to the molecules those transition-state constraints shown in Figure 1B.[19] As amino acid side chains from the enzyme would influence catalytically relevant conformations, the structure of the docked substrate was minimized with all the atoms of the AST IV homology model held fixed. This procedure was based on the hypothesis that the structure of AST IV influences the

Table 1. Kinetic Parameters of Substrates Used to Construct the CoMFA Model[a]

compound	K_m (app)	V_{max}	k_{cat}/K_m	ref	log(k_{cat}/K_m)	CoMFA log(k_{cat}/K_m)
phenol	0.019	350	625	60	5.80	5.49
4-methylphenol	0.010	320	1080	60	6.03	5.88
4-propylphenol	0.006	427	2410	60	6.38	6.43
4-butylphenol	0.005	296	2230	60	6.35	6.64
4-heptylphenol	0.061	132	73.4	60	4.87	5.09
4-nonylphenol	0.109	91	28.3	60	4.45	4.40
N-(4-hydroxyphenyl)acetamide	0.092	181	66.7	60	4.82	4.98
2-bromo-*N*-(4-hydroxyphenyl)acetamide	0.022	95	146	60	5.17	5.18
2-chloro-4-nitrophenol	0.005	28	190	39	5.28	4.92
benzyl alcohol	0.050	30.5	20.7	b	4.32	3.50
4-methylbenzyl alcohol	0.600	508	28.7	b	4.46	4.35
4-propylbenzyl alcohol	0.040	178	151	b	5.18	5.25
4-butylbenzyl alcohol	0.030	130	147	b	5.17	5.42
4-pentylbenzyl alcohol	0.020	109	185	b	5.27	5.26
4-hexylbenzyl alcohol	0.030	96.7	109	b	5.04	4.73
4-heptylbenzyl alcohol	0.410	133	11.0	b	4.04	3.96
4-nitrobenzyl alcohol	0.199	20.6	3.5	61	3.55	3.60
3-furanmethanol	0.960	10.2	0.4	61	2.56	2.74
2-naphthol	0.008	1130	4790	44	6.68	6.91
(1*S*)-1-phenyl-1-ethanol	0.250	33.7	4.6	b	3.66	3.69
(1*S*)-1-phenyl-1-propanol	0.240	9.9	1.4	b	3.15	3.48
(1*S*)-1-phenyl-1-pentanol	0.010	30.5	103	b	5.02	4.54
(1*S*)-1-phenyl-1-hexanol	0.007	2.1	10.2	b	4.01	4.24
(1*S*)-1-phenyl-1-heptanol	0.020	5.2	8.8	b	3.95	4.16
(1*S*)-2-methyl-1-phenyl-1propanol	1.360	6.1	0.2	24	2.18	2.65
(1*R*)-1-phenyl-1-ethanol	1.730	20.6	0.4	b	2.61	2.56
N-phenylhydroxylamine	0.230	43.0	6.3	41	3.80	4.52
N-(4-chlorophenyl)hydroxylamine	0.026	35.0	45.6	17	4.66	4.41
N-methyl-*N*-phenylhydroxylamine	0.119	85.0	24.2	19	4.38	4.23
N-phenyl-*N*-propylhydroxylamine	0.161	98.6	20.8	19	4.32	3.68
1-naphthalenemethanol	0.030	26.0	29.4	44	4.47	4.69
2-naphthalenemethanol	0.026	77.6	101	61	5.01	4.68
(1*R*,2*S*)-2-(methylamino)-1-phenyl-1-propanol	6.99	22.5	0.1	24	2.04	2.22
(1*R*,2*R*)-2-(methylamino)-1-phenyl-1-propanol	11.0	40.4	0.1	24	2.09	2.31
L-tyrosine methyl ester	4.0	350	3.0	44	3.47	3.44

[a] Values for apparent K_m are expressed in mM, values for V_{max} are expressed in nmole min^{-1}(mg of AST IV)$^{-1}$, and values of k_{cat}/K_m are calculated using a relative molecular mass of 33,909 for a subunit of AST IV and expressed as $min^{-1}mM^{-1}$. Values for log(k_{cat}/K_m) were calculated after conversion of k_{cat}/K_m to units of min^{-1} M^{-1}, and corresponding log(k_{cat}/K_m) values predicted by the CoMFA model are shown in the far right column. [b] Kinetic parameters were determined as described in the Experimental Section.

Table 2. Kinetic Parameters of Substrates Used to Validate the CoMFA Model[a]

compound	K_m (app)	V_{max}	k_{cat}/K_m	ref	actual log(k_{cat}/K_m)	predicted log(k_{cat}/K_m)
4-ethylphenol	0.010	410	1390	60	6.14	5.89
4-pentylphenol	0.009	208	784	60	5.89	6.64
4-ethylbenzyl alcohol	0.060	182	103	b	5.01	4.58
(1*S*)-1-phenyl-1-butanol	0.360	8.5	0.8	b	2.90	3.94
N-ethyl-*N*-phenyl-hydroxylamine	0.650	273	14.2	19	4.15	3.91
(1*R*)-1,2,3,4-tetrahydro-1-naphthol	0.030	28.7	32.4	24	4.51	4.85

[a] Values for apparent K_m are expressed in mM, values for V_{max} are expressed in nmole min^{-1} (mg of AST IV)$^{-1}$, and values of k_{cat}/K_m were calculated using a relative molecular mass of 33,909 for a subunit of AST IV and are expressed as min^{-1} mM^{-1}. Values for log(k_{cat}/K_m) were calculated after conversion of k_{cat}/K_m to units of min^{-1} M^{-1}. [b] Kinetic parameters were determined as described in the Experimental Section.

active conformation of its substrate. The alignment resulting from this methodology is shown in Figure 2B. Since this procedure takes into consideration both the sulfotransferase mechanism and the influence of amino acid residues at the active site it represented a dramatic improvement over the model that did not take into account both of these factors. The q^2 value was found to be 0.691. Values of log(k_{cat}/K_m) calculated for the CoMFA model are shown in Table 1, and a comparison of the actual and CoMFA-predicted values from the leave-one-out cross validation is graphically presented in Figure 4A. The predictive ability of this model with compounds not used in the training set (i.e., those listed in Table 2) is shown in Figure 4B. The closest agreement between the experimentally determined value of log(k_{cat}/K_m) and the CoMFA model was obtained with 4-pentylbenzyl alcohol, and the least agreement between actual and predicted values was with benzyl alcohol. Examination of the data indicated that there was no clear relationship between the type of functional group in the sulfuryl acceptor (e.g., phenol, benzylic alcohol, or *N*-hydroxy arylamine) and agreement between experimental and calculated values.

Next, all the molecules of Tables 1 and 2 were merged to build a complete CoMFA model. The q^2 value for this model was 0.701. As we wanted to evaluate the contour coefficient maps generated from this model we ran the PLS analysis with no validation and the number of components with the lowest standard error of prediction, namely five. A correlation between the optimum number of components and number of structural features of the compounds used in the study was not readily apparent. Using this PLS analysis we obtained an r^2 value of 0.922. The contour coefficient map with *p*-propylphenol

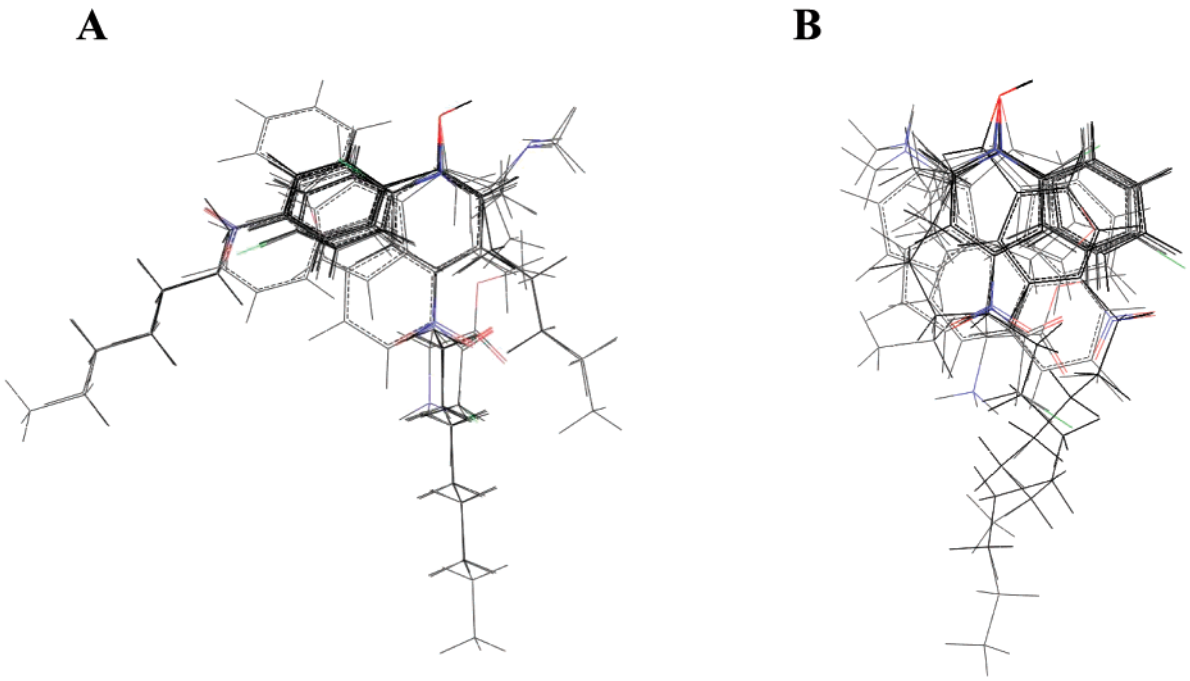

Figure 2. Overlay of conformational alignments for the CoMFA model. A. The molecules listed in Table 1 were aligned by pharmacophore overlay with the position of the oxygen and hydrogen as described in Figure 1, but the conformation was then minimized without any constraints imposed by the active site of the enzyme. B. Overlay of the conformations of molecules in Table 1 determined with the complete mechanistic and active site constraints described in Figure 3.

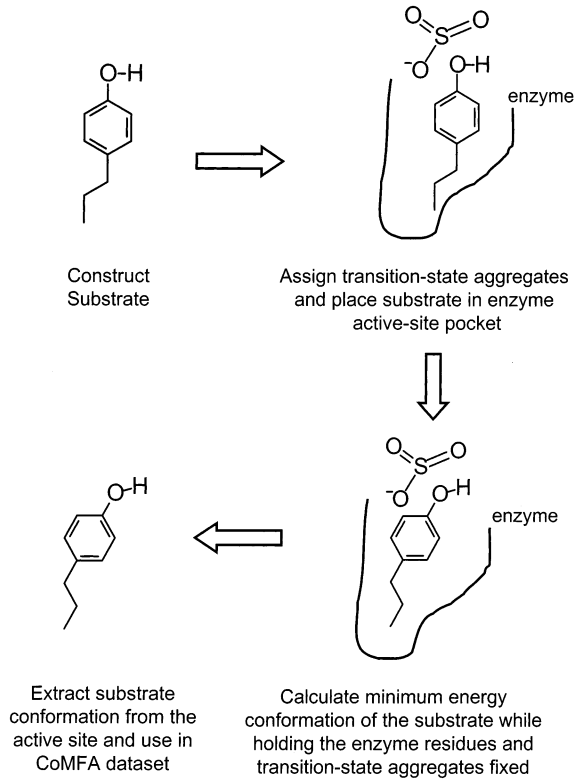

Figure 3. Outline of the alignment methodology for CoMFA.

and the amino acids in the AST IV model that correspond to the map are shown in Figure 5. The amino acids that correspond to regions of unfavorable steric bulk for substrates are Pro43, Lys44, Phe77, Phe80, Tyr135, Phe138 Tyr236, Met244 and Phe251. The regions of favorable steric bulk for substrates are seen as cavities in the active site of the homology model lined by the hydrophobic amino acids Phe20, Phe80, Ala142,

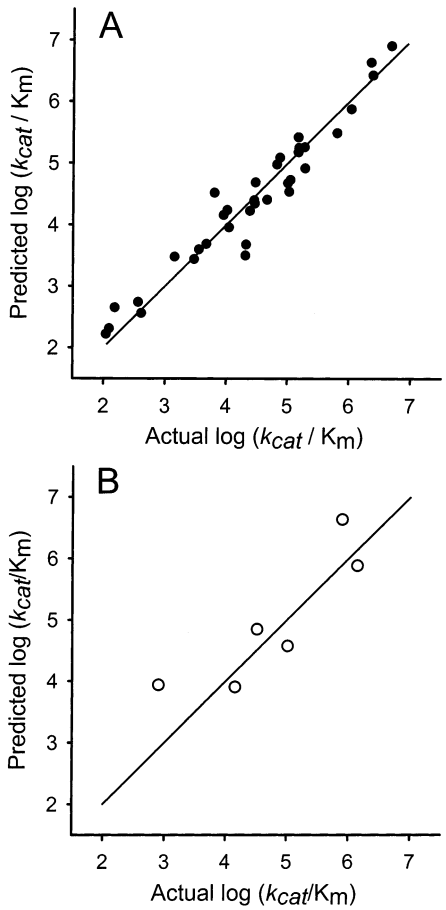

Figure 4. Quantitative results from CoMFA. Lines in both plots indicate expected results from a perfectly predicting model and are not based on regression analysis. A. Predicted verses actual values for compounds used to build the CoMFA model. B. Predicted verses actual values for compounds used to test the CoMFA model.

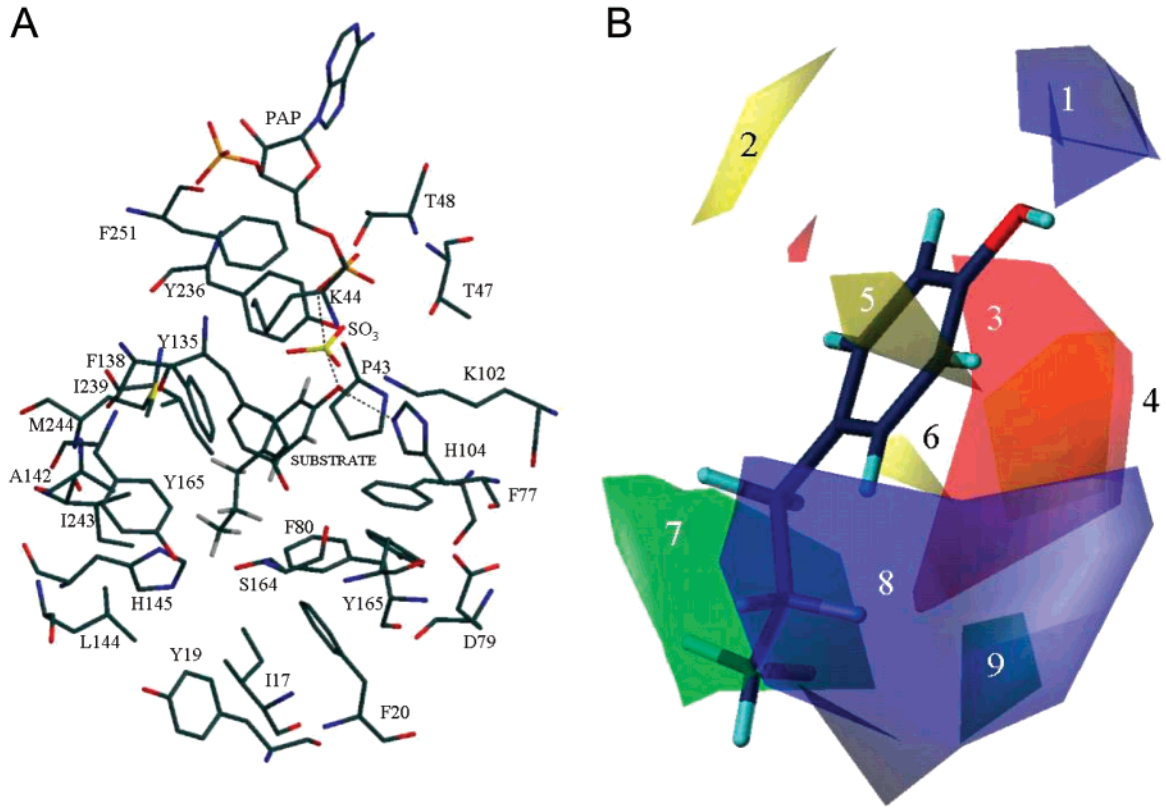

Figure 5. A. The conformation for *p*-propylphenol obtained after alignment in the substrate-binding pocket**.** This structure was drawn using MOLSCRIPT[58] and Raster3D.[59] B. Contour coefficient map of CoMFA. Sterically favored regions for substrates (numbered 7 and 9; shown in green) occupy cavities in the AST IV active site, while certain sterically unfavorable regions (numbered 2, 4, 5, and 6; shown in yellow) overlap with enzyme residues. The region favoring a positive charge on the substrate (numbered 3; shown in red) occupies an area in the vicinity of the charged residues, as do regions favoring a negative charge (numbered 1 and 8; shown in blue).

Table 3. Evaluation of Kinetic Parameters Predicted by the CoMFA Model[a]

compound	K_m (app)	V_{max}	k_{cat}/K_m	actual log(k_{cat}/K_m)	predicted log(k_{cat}/K_m)
4-(1-adamantyl)phenol	0.001	47.2	1600	6.20	6.48
4-biphenylmethanol	0.091	212	79.1	4.90	5.61
3-indolemethanol	0.029	33.2	38.9	4.60	5.11

[a] Values for apparent K_m are expressed in mM, and values for V_{max} are expressed in nmol min^{-1}(mg of AST IV)$^{-1}$. Values of k_{cat}/K_m were calculated using a relative molecular mass of 33909 for a subunit of AST IV and are expressed as min^{-1} mM^{-1}. Values for log(k_{cat}/K_m) were calculated after conversion of k_{cat}/K_m to units of min^{-1} M^{-1}.

Leu144, Tyr165, Ile239, Ile243, and Met244. Amino acids that are near regions that favor a negative charge on the substrate are Ile17, Tyr19, Phe20, Pro43, Lys44, Thr47, Thr48, Glu79, Lys102, His104, His145, Ser164, and Tyr165, while those that are near regions favoring a positive charge on the substrate are Phe77, Glu79, Phe80, Lys102, and Tyr236.

External Validation of the Model. While cross-validation, as represented by the q^2 values, provides information about the internal consistency and thus the predictive capacity of a model, and the contour coefficient maps correlate the homology model with the substrate profile, the ultimate test for a model is to challenge it by using it to design molecules that would show improvements in activity. After examining the CoMFA model, we decided to exploit region 7 (Figure 5) where steric bulk is preferred. Three molecules, 4-(1-adamantyl)phenol, 4-biphenylmethanol, and 3-indolemethanol, were selected on the basis of the CoMFA prediction that they would exhibit very high activities in relation to other phenols and benzylic alcohols, respectively. Assays were conducted on these molecules to test this application of the CoMFA model. The results of this analysis are represented in Table 3, wherein it is evident that these molecules had experimental kinetic constants that were within 0.3−0.8 log units of the log(k_{cat}/K_m) values predicted by the CoMFA model.

Discussion

Although this is the first report of a 3D-QSAR for a sulfotransferase, two QSAR models of sulfotransferases have been previously reported.[33,48] The most recent of these was performed on human dopamine (aryl) sulfotransferase in conjunction with determination of its crystal structure.[33] Physicochemical descriptors for lipophilic and hydrogen bonding groups were correlated with the apparent Michaelis constant for each substrate. Although this method, as well as the earlier QSAR model, demonstrated an ability to predict apparent Michaelis constants, the influences of active ligand conformations on activity were not taken into account. These previous methods thus meet with the same restrictions commonly seen with fragmented QSAR

approaches, the inability to include the influence of conformations and chirality on activity.

CoMFA provides a means to account for the influence of active conformations and chirality on the biological activity of a molecule and to obtain a quantitative 3D QSAR based upon the steric and electrostatic fields surrounding it. From its early development,[49] CoMFA has proven to be a powerful tool to relate these molecular properties to ligand-binding and/or catalytic function of a protein. Nevertheless, a difficulty in the application of CoMFA comes with the critical first step in which the conformations of molecules that are responsible for biological activity must be aligned. Alignment procedures have been classified as either structure-based (e.g., alignment based on pharmacophore analogy to a known active compound) or protein-based (e.g., docking molecules into the active site of a protein).[50]

While not previously applied to sulfotransferases, various studies on other proteins have employed methods for using structural information about the enzyme or receptor to guide the selection of alignments for CoMFA. For example, docking procedures have been recently described for the use of homology models of proteins such as photosystem II,[51] human pancreatic phospholipase A_2,[52] and a cytochrome P450 [53] to guide molecular alignments in CoMFA. However, these docking procedures are not easily applied to enzymes where an interacting molecule could assume many possible conformations in the active site. In one solution to this problem, molecular docking procedures utilizing anchor points with X-ray structures have been recently employed to guide the alignment of inhibitors of acetylcholinesterase[54] and nonpancreatic secretory phospholipase A_2 [55] for CoMFA. The latter CoMFA utilized initial anchor-points defined by analogy to the X-ray structure of a transition state analogue inhibitor bound to the enzyme.[55]

Our current studies have further advanced these approaches for CoMFA alignment by using the combination of conformational constraints imposed by a homology model with modeling constraints based on the mechanism of the enzymatic reaction to correlate 3D-structures with catalytic efficiency. It must be noted, however, that by holding the amino acids of the AST IV homology model fixed, we are biasing our model to reflect a single conformation of the active site. If we were to consider an entirely flexible active site, we might expect to produce a good CoMFA model by using conformations of the substrates based solely on transition state constraints devoid of homology model restrictions. Interestingly, a q^2 value of only 0.007 was observed for a CoMFA model obtained from such an alignment, while a significantly higher q^2 value of 0.691 was obtained after including homology model constraints in the alignment procedure. Thus, our assumption of a single mechanistically relevant conformation of the active site provides a valid starting point for analysis. Although the assumption of a single active site conformation in these studies yields a CoMFA that is highly predictive of catalytic efficiency of the enzyme in relation to substrate structure, the enzyme may show conformational changes upon binding the substrate. Thus, even though a highly flexible active site does not provide useful alignments for CoMFA, it is possible that future investigations using structural alignments based on some limited flexibility in the active site may provide additional refinement of the CoMFA model.

It is important to note that k_{cat}/K_m is the most relevant kinetic parameter for development of the CoMFA model for AST IV with a diverse array of sulfuryl acceptor substrates. Attempts to apply CoMFA to either V_m or K_m data alone were unsuccessful for such a diverse array of substrate structures. Values for k_{cat}/K_m, or V_m/K_m under conditions where the total concentration of enzyme is the same for each assay, have long been used as a measure of catalytic efficiency for comparison of substrate specificity and isotope effects.[56] More recently, k_{cat}/K_m and V_m/K_m have been described as apparent rate constants for the capture of a substrate into an enzyme complex that will at some later time produce product(s).[57] This definition, based on a capture of substrate into a catalytically relevant complex, seems particularly germane to our CoMFA alignment based on knowledge of the mechanism of the sulfation reaction and structural components of the active site.

Finally, we anticipate that this general methodology for CoMFA based on a dataset alignment that employs aspects of the enzymatic mechanism coupled with active site structural constraints for prediction of catalytic efficiency may be applicable to other related enzymes. For example, similarities in the catalytic mechanisms of sulfuryl and phosphoryl transfer catalyzed by sulfotransferases and kinases, respectively,[30,38] are consistent with the possibility that an extension of the alignment procedures developed for CoMFA of AST IV substrates might be highly valuable in the CoMFA of kinase substrates. In a more direct application of the alignment methods outlined here, we have successfully applied this CoMFA alignment procedure to substrates of hydroxysteroid sulfotransferase, STa, to obtain encouraging preliminary results. The direct application of these CoMFA methods to human sulfotransferase substrates is also in progress, and this will undoubtedly aid in understanding and predicting the roles of these enzymes in drug metabolism, toxicology, and chemical carcinogenesis.

Acknowledgment. This investigation was supported by U.S. Public Health Service Grant CA38683, awarded by the National Cancer Institute, Department of Health and Human Services.

References

(1) Mulder, G. J.; Jakoby, W. B. Sulfation. *Conjugation reactions in drug metabolism*; Taylor & Francis: New York, 1990; pp 107−161.
(2) Weinshilboum, R.; Otterness, D. Sulfotransferase enzymes. *Handbook of Experimental Pharmacology*; Springer-Verlag: Berlin, 1994; pp 45−78.
(3) Duffel, M. W. Sulfotransferases. In Volume 3. Biotransformation (Guengerich, F. P., vol. Ed.). In *Comprehensive Toxicology*; Sipes, I. G., McQueen, C. A., Gandolfi, A. J., Eds.; Elsevier: Oxford, 1997; pp 365−383.
(4) Duffel, M. W.; Marshall, A. D.; McPhie, P.; Sharma, V.; Jakoby, W. B. Enzymatic aspects of the phenol (aryl) sulfotransferases. *Drug Metab. Rev.* **2001**, *33*, 369−395.
(5) Weinshilboum, R. M.; Otterness, D. M.; Aksoy, I. A.; Wood, T. C.; Her, C.; Raftogianis, R. B. Sulfation and sulfotransferases 1: Sulfotransferase molecular biology: cDNAs and genes. *FASEB J.* **1997**, *11*, 3−14.
(6) Nagata, K.; Yamazoe, Y. Pharmacogenetics of sulfotransferase. *Annu. Rev. Pharmacol. Toxicol.* **2000**, *40*, 159−176.

(7) Miller, J. A.; Surh, Y.-J. Sulfonation in chemical carcinogenesis. *Handbook of Experimental Pharmacology*; Springer-Verlag: Berlin, 1994; pp 429−457.
(8) Beland, F. A.; Kadlubar, F. F. Metabolic activation and DNA adducts of aromatic amines and nitroaromatic hydrocarbons. *Handbook of Experimental Pharmacology*, *Vol. 94/I*; Springer-Verlag: Heidelberg, 1990; pp 267−325.
(9) Glatt, H. R. An overview of bioactivation of chemical carcinogens. *Biochem. Soc. Trans.* **2000**, *28*, 1−6.
(10) Glatt, H.; Engelke, C. E.; Pabel, U.; Teubner, W.; Jones, A. L.; Coughtrie, M. W.; Andrae, U.; Falany, C. N.; Meinl, W. Sulfotransferases: genetics and role in toxicology. *Toxicol. Lett.* **2000**, *112*−*113*, 341−8.
(11) Sekura, R. D.; Jakoby, W. B. Aryl sulfotransferase IV from rat liver. *Arch Biochem. Biophys.* **1981**, *211*, 352−359.
(12) Lewis, A. J.; Kelly, M. M.; Walle, U. K.; Eaton, E. A.; Falany, C. N.; Walle, T. Improved bacterial expression of the human P form phenolsulfotransferase: applications to drug metabolism. *Drug Metab. Dispos.* **1996**, *24*, 1180−1185.
(13) Walle, T.; Walle, U. K. Stereoselective sulfate conjugation of racemic 4-hydroxypropranolol by human and rat liver cytosol. *Drug Metab. Dispos.* **1991**, *19*, 448−453.
(14) Johnson, G. A.; Barshun, K. J.; McCall, J. M. Sulfation of minoxidil by liver sulfotransferase. *Biochem. Pharmacol.* **1982**, *31*, 2949−2954.
(15) Singer, S. S. The same enzymes catalyze sulfation of minoxidil, minoxidil analogues and catecholamines. *Chem. Biol. Interact.* **1994**, *92*, 33−45.
(16) Hartman, A. P.; Wilson, A. A.; Wilson, H. M.; Aberg, G.; Falany, C. N.; Walle, T. Enantioselective sulfation of b2-receptor agonists by the human intestine and the recombinant M-form phenol-sulfotransferase. *Chirality* **1998**, *10*, 800−803.
(17) Duffel, M. W. Molecular specificity of aryl sulfotransferase IV (tyrosine-ester sulfotransferase) for xenobiotic substrates and inhibitors. *Chem. Biol. Interact.* **1994**, *92*, 3−14.
(18) Duffel, M. W.; Modi, R. B.; King, R. Interactions of a Primary N−Hydroxy Arylamine with Rat Hepatic Aryl Sulfotransferase IV. *Drug Metab. Dispos.* **1992**, *20*, 339−342.
(19) King, R. S.; Sharma, V.; Pedersen, L. C.; Kakuta, Y.; Negishi, M.; Duffel, M. W. Structure−function modeling of the interactions of N-alkyl-N-hydroxyanilines with rat hepatic aryl sulfotransferase IV. *Chem. Res. Toxicol.* **2000**, *13*, 1251−1258.
(20) King, R. S.; Teitel, C. H.; Kadlubar, F. F. In Vitro bioactivation of *N*-hydroxy-2-amino-α-carboline. *Carcinogenesis* **2000**, *21*, 1347−1354.
(21) Meerman, J. H.; Ringer, D. P.; Coughtrie, M. W.; Bamforth, K. J.; Gilissen, R. A. Sulfation of carcinogenic aromatic hydroxylamines and hydroxamic acids by rat and human sulfotransferases: substrate specificity, developmental aspects and sex differences. *Chem. Biol. Interact.* **1994**, *92*, 321−328.
(22) Mangold, J. B.; Spina, A.; McCann, D. J. Sulfation of mono- and diaryl oximes by aryl sulfotransferase isozymes. *Biochim. Biophys. Acta* **1989**, *991*, 453−458.
(23) Mangold, J. B.; McCann, D. J.; Spina, A. Aryl sulfotransferase-IV-catalyzed sulfation of aryl oximes: steric and substituent effects. *Biochim. Biophys. Acta* **1993**, *1163*, 217−222.
(24) Rao, S. I.; Duffel, M. W. Benzylic alcohols as stereospecific substrates and inhibitors for aryl sulfotransferase. *Chirality* **1991**, *3*, 104−111.
(25) Sodum, R. S.; Sohn, O. S.; Nie, G.; Fiala, E. S. Activation of the liver carcinogen 2-nitropropane by aryl sulfotransferase. *Chem. Res. Toxicol.* **1994**, *7*, 344−351.
(26) Borchert, P.; Miller, J. A.; Miller, E. C.; Shires, T. K. 1′-Hydroxysafrole, a proximate carcinogenic metabolite of safrole in the rat and mouse. *Cancer Res.* **1973**, *33*, 590−600.
(27) Miller, E. C.; Swanson, A. B.; Phillips, D. H.; Fletcher, T. L.; Liem, A.; Miller, J. A. Structure−activity studies of the carcinogenicities in the mouse and rat of some naturally occurring and synthetic alkenylbenzene derivatives related to safrole and estragole. *Cancer Res.* **1983**, *43*, 1124−1134.
(28) Kato, R.; Yamazoe, Y. Metabolic activation of N-hydroxylated metabolites of carcinogenic and mutagenic arylamines and arylamides by esterification. *Drug Metab. Rev.* **1994**, *26*, 413−429.
(29) Fiala, E. S.; Sodum, R. S.; Hussain, N. S.; Rivenson, A.; Dolan, L. Secondary nitroalkanes: induction of DNA repair in rat hepatocytes, activation by aryl sulfotransferase and hepatocarcinogenicity of 2-nitrobutane and 3-nitropentane in male F344 rats. *Toxicology* **1995**, *99*, 89−97.
(30) Yoshinari, K.; Petrotchenko, E. V.; Pedersen, L. C.; Negishi, M. Crystal structure-based studies of cytosolic sulfotransferase. *J. Biochem. Mol. Toxicol.* **2001**, *15*, 67−75.
(31) Negishi, M.; Pedersen, L. G.; Petrotchenko, E.; Shevtsov, S.; Gorokhov, A.; Kakuta, Y.; Pedersen, L. C. Structure and function of sulfotransferases. *Arch. Biochem. Biophys.* **2001**, *390*, 149−157.
(32) Kakuta, Y.; Pedersen, L. G.; Carter, C. W.; Negishi, M.; Pedersen, L. C. Crystal structure of estrogen sulphotransferase. *Nat. Struct. Biol.* **1997**, *4*, 904−908.
(33) Dajani, R.; Cleasby, A.; Neu, M.; Wonacott, A. J.; Jhoti, H.; Hood, A. M.; Modi, S.; Hersey, A.; Taskinen, J.; Cooke, R. M.; Manchee, G. R.; Coughtrie, M. W. X-ray crystal structure of human dopamine sulfotransferase, SULT1A3. Molecular modeling and quantitative structure−activity relationship analysis demonstrate a molecular basis for sulfotransferase substrate specificity. *J. Biol. Chem.* **1999**, *274*, 37862−37868.
(34) Bidwell, L. M.; McManus, M. E.; Gaedigk, A.; Kakuta, Y.; Negishi, M.; Pedersen, L.; Martin, J. L. Crystal structure of human catecholamine sulfotransferase. *J. Mol. Biol.* **1999**, *293*, 521−530.
(35) Kakuta, Y.; Sueyoshi, T.; Negishi, M.; Pedersen, L. C. Crystal structure of the sulfotransferase domain of human heparan sulfate N-deacetylase/N-sulfotransferase 1. *J. Biol. Chem.* **1999**, *274*, 10673−10676.
(36) Pedersen, L. C.; Petrotchenko, E. V.; Negishi, M. Crystal structure of SULT2A3, human hydroxysteroid sulfotransferase. *FEBS Lett.* **2000**, *475*, 61−64.
(37) Petrotchenko, E. V.; Pedersen, L. C.; Borchers, C. H.; Tomer, K. B.; Negishi, M. The dimerization motif of cytosolic sulfotransferases. *FEBS Lett.* **2001**, *490*, 39−43.
(38) Kakuta, Y.; Petrotchenko, E. V.; Pedersen, L. C.; Negishi, M. The sulfuryl transfer mechanism. Crystal structure of a vanadate complex of estrogen sulfotransferase and mutational analysis. *J. Biol. Chem.* **1998**, *273*, 27325−27330.
(39) Duffel, M. W.; Jakoby, W. B. On the mechanism of aryl sulfotransferase. *J. Biol. Chem.* **1981**, *256*, 11123−11127.
(40) Marshall, A. D.; McPhie, P.; Jakoby, W. B. Redox control of aryl sulfotransferase specificity. *Arch. Biochem. Biophys.* **2000**, *382*, 95−104.
(41) King, R. S.; Duffel, M. W. Oxidation-dependent inactivation of aryl sulfotransferase IV by primary N-hydroxy arylamines during in vitro assays. *Carcinogenesis* **1997**, *18*, 843−849.
(42) Chen, G.; Banoglu, E.; Duffel, M. W. Influence of substrate structure on the catalytic efficiency of hydroxysteroid sulfotransferase STa in the sulfation of alcohols. *Chem. Res. Toxicol.* **1996**, *9*, 67−74.
(43) Banoglu, E.; Duffel, M. W. Studies on the interactions of chiral secondary alcohols with rat hydroxysteroid sulfotransferase STa. *Drug Metab. Dispos.* **1997**, *25*, 1304−1310.
(44) Chen, X.; Yang, Y. S.; Zheng, Y.; Martin, B. M.; Duffel, M. W.; Jakoby, W. B. Tyrosine-ester sulfotransferase from rat liver: bacterial expression and identification. *Protein Expr. Purif.* **1992**, *3*, 421−426.
(45) Duffel, M. W.; Binder, T. P.; Rao, S. I. Assay of purified aryl sulfotransferase suitable for reactions yielding unstable sulfuric acid esters. *Anal. Biochem.* **1989**, *183*, 320−324.
(46) Brunger, A. T.; Kuriyan, J.; Karplus, M. Crystallographic R-factor refinement by molecular dynamics. *Science* **1987**, *235*, 458−460.
(47) Zheng, Y.; Bergold, A.; Duffel, M. Affinity labeling of aryl sulfotransferase IV: identification of a peptide sequence at the binding site for 3′-phosphoadenosine-5′-phosphosulfate. *J. Biol. Chem.* **1994**, *269*, 30313−30319.
(48) Campbell, N. R.; Van Loon, J. A.; Sundaram, R. S.; Ames, M. M.; Hansch, C.; Weinshilboum, R. Human and rat liver phenol sulfotransferase: structure−activity relationships for phenolic substrates. *Mol. Pharmacol.* **1987**, *32*, 813−819.
(49) Cramer, R. D. I.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(50) Golbraikh, A.; Bernard, P.; Chretien, J. R. Validation of protein-based alignment in 3D quantitative structure−activity relationships with CoMFA models. *Eur. J. Med. Chem.* **2000**, *35*, 123−136.
(51) Jalaie, M.; Erickson, J. A. Homology model directed alignment selection for comparative molecular field analysis: application to photosystem II inhibitors. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 181−197.
(52) Pintore, M.; Bernard, P.; Berthon, J.-Y.; Chretien, J. R. Protein-based alignment in 3D QSAR of 26 indole inhibitors of human pancreatic phospholipase A_2. *Eur. J. Med. Chem.* **2001**, *36*, 21−30.
(53) Cavalli, A.; Greco, G.; Novellino, E.; Recanatini, M. Linking CoMFA and protein homology models of enzyme−inhibitor interactions: an application to nonsteroidal aromatase inhibitors. *Bioorg. Med. Chem.* **2000**, *8*, 2771−2780.
(54) Bernard, P.; Kireev, D. B.; Chretien, J. R.; Fortier, P.-L.; Coppet, L. Automated docking of 82 *N*-benzylpiperidine derivatives to mouse acetylcholinesterase and comparative molecular field analysis with "natural" alignment. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 355−371.

(55) Bernard, P.; Pintore, M.; Berthon, J.-Y.; Chretien, J. R. A molecular modeling and 3D QSAR study of a large series of indole inhibitors of human nonpancreatic secretory phospholipase A_2. *Eur. J. Med. Chem.* **2001**, *36*, 1−19.
(56) Cleland, W. W. What limits the rate of an enzyme-catalyzed reaction? *Acc. Chem. Res.* **1975**, *8*, 145−151.
(57) Northrup, D. B. Rethinking fundamentals of enzyme action. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1999**, *73*, 25−55.
(58) Kraulis, P. J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Applied Crystallography* **1991**, *24*, 946−950.
(59) Merrit, E. A.; Bacon, D. J. Raster3D: Photorealistic molecular graphics. *Methods Enzymol* **1997**, *77*, 505−524.
(60) Duffel, M. W.; Chen, G.; Sharma, V. Studies on an affinity label for the sulfuryl acceptor binding site in an aryl sulfotransferase. *Chem. Biol. Interact.* **1998**, *109*, 81−92.
(61) Binder, T. P.; Duffel, M. W. Sulfation of benzylic alcohols catalyzed by aryl sulfotransferase IV. *Mol. Pharmacol.* **1988**, *33*, 477−479.

JM010481C