

Modeling of hSULT1A1 Metabolism

A Dissertation

Submitted in Partial Fulfilment of the Requirements for the Degree of

Doctor rerum naturalium (Dr. rer. nat.)

to the Department of Biology, Chemistry and Pharmacy of

Freie Universität Berlin (FUB)



By

Salwa Mosad Soliman

from Elmonofia, Egypt

Berlin, 2017

Declaration

The present research work was done from April 2011 till August 2016 under supervision of Prof. Dr. Gerhard Wolber at the Institute of Pharmacy, Freie Universität Berlin, Germany. Part of the work was conducted during an internship period from April 2014 to December 2015 at the School of Pharmaceutical Sciences, University of Geneva, Switzerland.

1. Gutachter: Prof. Dr. Gerhard Wolber
2. Gutachter: Prof. Dr. Leonardo Scapozza

Disputation on: 18 May 2017

Acknowledgment

Praise be to ALLAH, the most merciful and most gracious, who in his infinite mercy gave me bountiful courage and strength to pursue my research for this thesis. It has been truly a long journey, spanning more than five years, 2 different countries, different working settings, environments and languages but it's all nicely wrapped together now. I feel much in debt for a lot of people, without whom, I'd never make it to the point of writing these lines.

To Prof. Gerhard Wolber, thanks a lot for giving me the opportunity to pursue my PhD degree in your group in Berlin. It has been a life changing experience that I will remember forever. To Prof. Leonardo Scapozza who embraced me in his group during my stay in Geneva, thanks a lot for the kind hospitality and the diversified knowledge I acquired during my work in your labs.

To Dr. Jermie Mortier, Dr. Jamil Al-Asri and the whole team in Berlin, thank you for the time we spent together, the intense help and support to get things done, especially at the hard times. To Dr. Gianpaolo Chiriano, Dr. Remo Perozzo, Dr. "soon to be" Francesca Tessaro and the whole team in Geneva, thank you for all the great work we have done together, for the lengthy discussions, critical opinions, and limitless help.

To the selection committee of Erasmus Mundus, thanks a lot for the trust you put in me and for funding my project.

To my sisters in Geneva Mona AlQazzaz and Tamir De Lisser, thanks a lot for your help in revising this thesis as well as the support and care on personal levels. To my Masters supervisor and elder friend Prof. Hilde Spahn-Langguth, I'm very thankful to your help, assistance and support.

To my mother and my family in Egypt, cannot thank you enough for the endless support, love and "Doa'a".

Finally to my lovely husband Omar, It's beyond words to describe your infinite dedication, sacrifice and help. To my three little princesses and prince, Hana, Nour and Salman who just made it to our world in the middle of my PhD, although you were the hardest part in the equation, but I love you so much. Thank you for always bringing a happy smile to my face.

Salwa Mosad

Table of Contents

1	Introduction	1
1.1	Phase II metabolism.....	1
1.2	Sulfotransferases	2
1.2.1	Molecular structure of cytosolic SULTs.....	4
1.2.2	Sulfonation donor (PAPS).....	12
1.2.3	Catalytic mechanism.....	13
1.2.4	Substrate inhibition	16
1.2.5	SULTs isoforms	17
1.3	Human SULT1A1	18
1.3.1	Molecular Structure of SULT1A1.....	19
1.3.2	SULT1A1 localization.....	21
1.3.3	Physiological roles of SULT1A1	21
1.3.4	Genetic polymorphism in SULT1A1.....	24
1.3.5	SULT1A1 substrates	25
1.3.6	SULT1A1 inhibitors	27
1.4	Previous computational studies	27
2	Aim of work	33
3	Methods	35
3.1	3D Pharmacophores	35
3.2	Molecular Dynamics (MD).....	36
4	Results	39
4.1	Structural investigations	39
4.2	Pharmacophores.....	47
4.2.1	Generation of active and decoy datasets.....	47
4.2.2	Substrates 3D pharmacophores	49
4.2.3	Inhibitors 3D pharmacophores.....	55
4.2.4	Retrospective validation.....	59
4.3	Molecular Dynamics (MD) study	60
4.3.1	System Design.....	60
4.3.2	Analysis of MD results.....	61
4.3.3	Proposed hypothesis of positive cooperativity behavior	73
5	Discussion	81
5.1	Structural investigations	81

5.2	Pharmacophore generation	82
5.3	MD simulation	83
5.3.1	Effect of cofactor binding	84
5.3.2	Ligand-size induced binding site flexibility	86
6	Conclusions.....	87
7	Experimental	89
7.1	Structural investigation and pocket mapping	89
7.1.1	Multiple sequence alignment (MSA):	89
7.1.2	Molecular Interaction Fields (MIF):	89
7.1.3	Pocket-volume analysis	89
7.1.4	2D06 binding site mapping using Szmap®	90
7.2	Pharmacophore generation	90
7.2.1	Generation of PAPS containing crystal structures	90
7.2.2	Docking of EE in 2D06	91
7.2.3	3D structure- and ligand-based model generation-	91
7.3	MD simulation	92
7.3.1	System preparation	92
7.3.2	Trajectories analysis	92
8	Appendix.....	94
9	References	135

List of Figures

- Figure 1:** Simple comparison between metabolism phases and their role in drug metabolism. 2
- Figure 2:** The catalytic reaction catalyzed by SULTs from the universal sulfonate donor (PAPS) (the sulfonate group highlighted in orange) to an acceptor and producing the inactive donor (PAP) and a sulfonated product. 3
- Figure 3:** General structure of SULTs: 4- β sheet backbone (yellow), Loop 1 (magenta), Loop 2 (blue), Loop 3 (orange), dimerization domain (red), PAPS binding region (wired surface). 5
- Figure 4:** Binding of the sulfonate donor in SULTs: PAPS is accommodated by three interactions; 5' phosphosulfonate group with PSB (red dotted region), 3' phosphate group with PB loop (orange dotted region), and hydrophobic interaction of adenine ring with SLH3 motif (blue dotted region). PAPS shows tight interaction with the enzyme residue and the sulfonate group is the only exposed group. 6
- Figure 5:** Sulfonation mechanism. 1. The conserved catalytic residues Lys47 (Lys48 in SULT1A1 numbering), Ser137 (Ser138 in SULT1A1 numbering), and His107 (108 SULT1A1 numbering). 2. Histidine mediates the activation of the substrate into a strong nucleophile. 3. Transition complex formation and stabilized by Lys residue. 4. Sulfonate group transfer and the production of sulfonated product and inactive cofactor. 11
- Figure 6:** Synthesis of PAPS in human body from ATP and free sulfate. The two-step reaction occurs in PAPS synthetase enzyme that has coupled domains for both steps (ATP sulfurylase and APS kinase domain). 12
- Figure 7:** Sequential versus non-sequential kinetic mechanisms. A. Sequential addition in which both co-substrates are involved in the formation of ternary complex (enzyme-cofactor-substrate) for the reaction to take place. B. Non-sequential mechanism in which only one substrate is involved in activation of the enzyme then the second substrate binds in a double-displacement mechanism. 14
- Figure 8:** The sulfonation catalytic cycle. Random bi-bi addition mechanism involves two substrates (PAPS and Lig) and two released products (PAP and LigS). The dead-end complexes (red arrows) are formed in non-productive ternary enzyme complexes with non-interacting substrates. This diagram was modified from ⁷⁰. 16

Figure 9: SULTs isoforms. Classification of SULT isoforms according to global sequence similarity..... 17

Figure 10: The structure of SULT1A1: (PDB code: 1LS6) structure with the catalytic (Pocket I) and inhibitory (Pocket II) binding sites in red and blue spheres, respectively. The arrows indicate the substrate access channels. The PAP and relevant residues are rendered in a bond model. 20

Figure 11: SULT1A1 playing roles. a. Minoxidil activation via sulfonation. b. hydroxyl tamofen (TAM) activation into resonating free radicals. c. paracetamol metabolism vis sulfonation. d. estradiol (E2) hemostasis mediated by SULT. 23

Figure 12: Superposition of the four co-crystallized ligands in the substrate cleft of SULT1A1. The conserved crystal water found in all structures complexed with PAP except 2D06, is depicted in red CPK. The PAP molecule from PDB entry 3U3K is shown in grey ball-and-stick model. The co-crystallized ligands are 2-naphthol (2NA) (brown), estradiol (E2) (orange), 2-nitrophenol (pNP) (red), 3-cyano-7-hydroxycoumarin (3CyC) from 3U3M (grey) and 2 molecules of 3CyC from 3U3O (blue). The grey space represents the PAP binding region whereas the green space shows the extended L shape acceptor cleft. The mesh space displays the actual catalytic binding pocket..... 40

Figure 13: Comparison between the substrate pocket dimensions of SULT1A1 crystal structures..... 41

Figure 14: Catalytic distances between bound ligand (2NA in 3U3K), the catalytic residues (Lys48, Lys106, His108, and Ser138), and the co-crystallized cofactor (PAP). Yellow and blue spheres display the contact hydrophilic and lipophilic (respectively) preferences of the binding pocket..... 42

Figure 15: a) Comparison between SULT1A1 crystal structures substrate binding site volume. b) The alignment of all SULT1A1 crystal structures showing different gating loop positions. The red circle represents the substrate active site. The whole pocket is mapped according to its hydrophobicity from yellow (lipophilic) to red (hydrophilic). The colors of the gating loop are as follows: light green (1LS6), dark green (1Z28), orange (2D06), brown (3U3J), grey (3U3K), black (3U3M), yellow (3U3O), white (3UR), and violet (4GRA) (down). c) Pairwise RMSD plot for all atoms exist in the gating loop of all crystallized SULT1A1 protein. d) Pairwise RMSD plot for C α atoms of all crystallized SULT1A1 protein. 44

Figure 16: Superimposed crystal structures of the three SULT isoforms; SULT1A1 (2D06), SULT1A3 (2A3R), and SULT1E1 (4JVL). Pairwise RMSD values (in Å) for C α -atoms of

the three SULTs are indicated in blue matrix. Differences in the enzyme backbones are shown on the averaged RMSD (Å) scale versus amino acid numbering. Part of the protein backbones of SULT1A1, SULT1A3, and SULT1E1 are indicated in blue, red and yellow, respectively. 46

Figure 17: Molecular Interaction Fields (MIF) of substrate pocket of SULT1A1, SULT1E1, and SULT1A3. The molecular interaction fields are shown as grids showing H-bond acceptors (blue), H-bond donors (red), hydrophobic interactions (yellow)..... 46

Figure 18: The most plausible PAPS binding pose (cyan) compared to the binding pose of the original cofactor (PAP) crystalized in 3U3K (grey). Sulfonate group contributes three extra hydrogen bonds towards protein residues indicated in ball and stick. 50

Figure 19: Michaelis-Menten complex: Plausible PAPS binding mode in PSB of 1LS6¹⁵ shows good interaction between the sulfonate moiety of PAPS and the catalytic amino acids Lys48 and Lys106 (red hydrogen bond acceptor) and in a close proximity (3 Å) of the hydroxyl group of the acceptor to undergo sulfonate transfer reaction. 51

Figure 20: Michaelis-Menten complex of E2: Plausible E2 binding mode (silver) in 2D06-PAPS superimposed with its original binding position (gold) in 2D06¹⁷. The hydroxyl group of docked E2 shows hydrogen donating bond towards His108 (green arrow) and proper distance to the sulfonate moiety of PAPS to undergo sulfonate transfer reaction. 51

Figure 21: Structure- and Docking-based substrates pharmacophores. The five red boxes represent the five generated Michaelis-Menten complexes of 2CyC, pNP, 2NA, E2 (docked) and T2, docked in 3U3M-PAPS, 1LS6-PAPS, 3U3K-PAPS, 2D06-PAPS, and 3U3M-PAPS, respectively. The red arrows resemble the pharmacophore generation for the corresponding five pharmacophores in red boxes; SSB_PH4_2CyC, SSB_PH4_pNP, SSB_PH4_2NA, SSB_PH4_E2, and SSB_PH4_T2. The blue arrow shows the validation process that is represented by the Receiver Operating Curve (ROC). 54

Figure 22: 2D06 binding site mapping and lead optimization of E2: A) SZMAP[®] map for 2D06 showing hydrophobic (purple) and hydrophilic regions (golden) in the E2 binding site. B) 2D Gameplan map that defines the optimum binder to the binding site based on site mapping. It recommends having polar groups at the golden regions and hydrophobic ones at the purple sites to increase binding affinity to the enzyme. 56

Figure 23: Docking pose of EE: A) Plausible binding pose of EE (silver) in comparison to the original crystallized ligand E2 (golden) shown in stick mode, demonstrating hydrogen

bond donor (green arrow) to PAP. The acetylene group of EE2 shows interaction with some protein residues in the pocket (ball and stick). B) Structure-based pharmacophore derived from the docked EE in SULT 1A1 (PDB: 2D06). The pharmacophore shows three hydrophobic features (yellow spheres), aromatic feature (blue ring), hydrogen donating feature (green arrow) towards PAP and hydrogen accepting (red arrow) towards Lys48. . . 57

Figure 24: Inhibitor Structure-based and Ligand-based pharmacophores: Left: Inhibitors Structure-based pharmacophore (ISB_PH4) derived from the plausible binding pose of the docked EE (silver) in SULT 1A1 (PDB: 2D06). Right: Ligand-based pharmacophores derived from Clusters of different inhibitors scaffolds generated from flavonoids (ILB_PH4_1) and chloro-biphenyls (ILB_PH4_2) and NSAIs (ILB_PH4_3). Chemical features are color coded: H-bond acceptor (HBA) - red sphere or red arrow, H-bond donor (HBD)- green sphere or green arrow, aromatic ring (Ar)- blue circle, hydrophobic contact (HYD)- yellow spheres, and negative ionizable (NI)-red star . The models with aligned ligands are shown on the top, and the 3D pharmacophores with exclusion volumes on the bottom. 58

Figure 25: The work flow of the pharmacophore model for in silico discrimination of SULT1A1 binders. 60

Figure 26: Systems used for the MD study. 2NA: 2 naphthol, E2: estradiol, and EE: ethinyl estradiol. Systems are colored according to the presence of the cofactor into apo, PAP-containing and PAPS-containing (black, red and green respectively). 61

Figure 27: a. SULT1A1 structure showing Loop 3 with two segments; donor part (red) and acceptor part (blue) with a clear hydrophobic stacking in between (Phe255 and Tyr240). Other interactions were noted as well; molecular hinge between Asp249 and Met145 and a Lys258 salt bridge to Asp263. b. The lysine salt bridge length in different ligand-free complexes SULT1A1 (black), SULT1A1-PAP (red) and SULT1A1-PAPS (green) indicating completely broken bridge in apo enzyme. c. PAP binding interaction in PSB exhibiting hydrogen donating (green) and hydrogen accepting (red) bonds to the donor segment residue (Phe255, Arg257, Lys258 and Gly259). 64

Figure 28: Chemical structures of the ligands used in the MD simulation. 68

Figure 29: Left: Ten-residue segment that covers the PSB and stabilized by cofactor binding and big hydrophobic binders. Right: The distance between Phe255 and Tyr240 that represents the segment movement in 2NA binary system. 70

Figure 30: Plausible binding modes of high affinity binders. Left: SULT1A1 tight binding site accommodating small binder (2NA) surrounded by phenyl alanine rich cleft. Right: SULT1A1 extended binding site accommodating large binder (EE) with its ethynyl group hosted between Phe76 and Phe84..... 77

Figure 31: Planes of different binders inside SULT-1A1 ternary structures. Gating loop showed various ligand-induced positions according to the bound ligand 2NA, E2 and EE (blue, orange and grey respectively). 78

Figure 32: Binding place for small binder substrates (blue mesh), large binder inhibitors (grey mesh) and Pocket II responsible for substrate inhibition (green mesh). ... 78

List of Tables

Table 1: Comparison between Golgi (membrane) and cytosolic SULTs.	3
Table 2: List of all cytosolic SULTs crystal structures.	4
Table 3: SNPs responsible for the formation of various SULT1A1 allylic variants.	24
Table 4: Summary of the previous early computer-based techniques literatures (until 2005) published on SULTs.	29
Table 5: Summary of the most recent computer-based techniques literatures (2006- up to now) published on SULTs.	31
Table 6: RMSD values for MD simulations of ligand-free complexes given as average in Å with standard deviations in parentheses.	62
Table 7: RMSD values for MD simulations of several ligand-containing complexes given as average in Å with standard deviations in parentheses. The systems are color coded blue, orange, and grey for 2NA, E2, and EE, respectively.	68

List of Abbreviations

(In alphabetical order)

2D-QSAR	Two dimensional Quantitative structure activity relationship
2NA	2 Nitro-phenol
3CyC	3-cyano-7-hydroxycoumarin
3D-QSAR	Three dimensional Quantitative structure activity relationship
APS	adenosine 5' phosphosulfate
CoMFA	Comparative Molecular Field Analysis
CYP	Cytochrome P450
DCNP	2,6-dichloro-4-nitrophenol
DHEA	Dehydroepiandrosterone
E2	Estradiol
EE2	Ethinyl Estradiol
EE	Enrichment Efficiency
MD	Molecular Dynamics
MER	Maximum Encompassing Region
MIF	Molecular Interaction Fields
MSA	Multiple sequence Alignment
OH-PCBs	hydroxylated polychlorinated biphenyls
PAP	3'-phosphoadenosine 5'-phosphate
PAPS	3'-phosphoadenosine 5'-phosphosulphate
PAPSS	PAPS synthetase
PB	phosphate binding

PCBs	polychlorinated biphenyls
pNP	p-nitrophenol
PST	PAPS transporters
PDB	Protein Data Bank
PSB	phosphor sulfonate binding site
QSAR	Quantitative structure activity relationship
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
ROC	Receiver Operating Curve
SAR	Structure Activity Relationship
SLH3	β -strand-loop- α -helix
SNP	single nucleotide polymorphism
SULT	Sulfotransferase
TAM	Tamoxifen

1 Introduction

1.1 Phase II metabolism

Metabolism is the chemical modification of xenobiotics prior to their elimination from the body as well as to endogenous compounds, as a step of their biological regulation ¹. Metabolism transforms molecules into more hydrophilic derivatives to be easily eliminated from the body, usually via urine ². There are three prominent actions to several xenobiotics: Metabolism may (i) terminate the activity of an active moiety by significantly modifying its target affinity, it may (ii) activate prodrugs, and in some cases, it can lead to (iii) carcinogenic and toxic derivatives ³. Unwanted metabolic behavior is one of the major problems in drug discovery, often leading to drug development failure. It does not only consume time and money, but may also cause fast drug elimination from the blood stream, toxicity, carcinogenicity as well as teratogenicity for pregnant mothers ⁴. One dangerous facet to metabolism is that it can cause a potential drug-drug interaction. Co-administration of metabolic enzyme inducer or inhibitor together with another drug might alter the pharmacokinetics of the second drug to a dangerous extent. Many examples have been discovered either by extensive pharmacological studies or by death of some treated patients ⁵.

Xenobiotic metabolism in the human body takes place through phase I and/or phase II ³. Most drugs undergo Phase I metabolism through a series of oxidative reactions that unmask or form (de novo) important functional groups for further metabolism in phase II. Functional group modification is responsible for typically small changes in the drug molecular weight and solubility ⁴. In contrast, in phase II, drugs are totally changed into larger, also more soluble, excretable and hydrophilic derivative of the compound by the addition of a hydrophilic charged group (Figure 1). Such conjugating reactions enhance drug excretion and elimination out of the body. Phase I metabolizing enzymes include a large set of enzymes that can oxidize, reduce and hydrolyze functional groups. The most predominant phase I enzyme family is cytochrome P450 (CYP). It is versatile and responsible for the metabolism of more than 90% of xenobiotics ¹. Phase II metabolism is mediated by a group of enzymes that catalyze the conjugation of a charged molecule into the acceptor substrate. The catalyzed reactions include glucuronidation, sulfonation, methylation, acetylation, glutathione and amino acid conjugation ³.

Metabolism can be altered by both internal (age, gender, genetics) and external factors (smoking, medication, diet). Although phase II enzymes are generally detoxifying body from toxic materials, they attracted much less attention in clinical pharmacology than cytochromes P450 (CYP450) being a second line in metabolic defense.

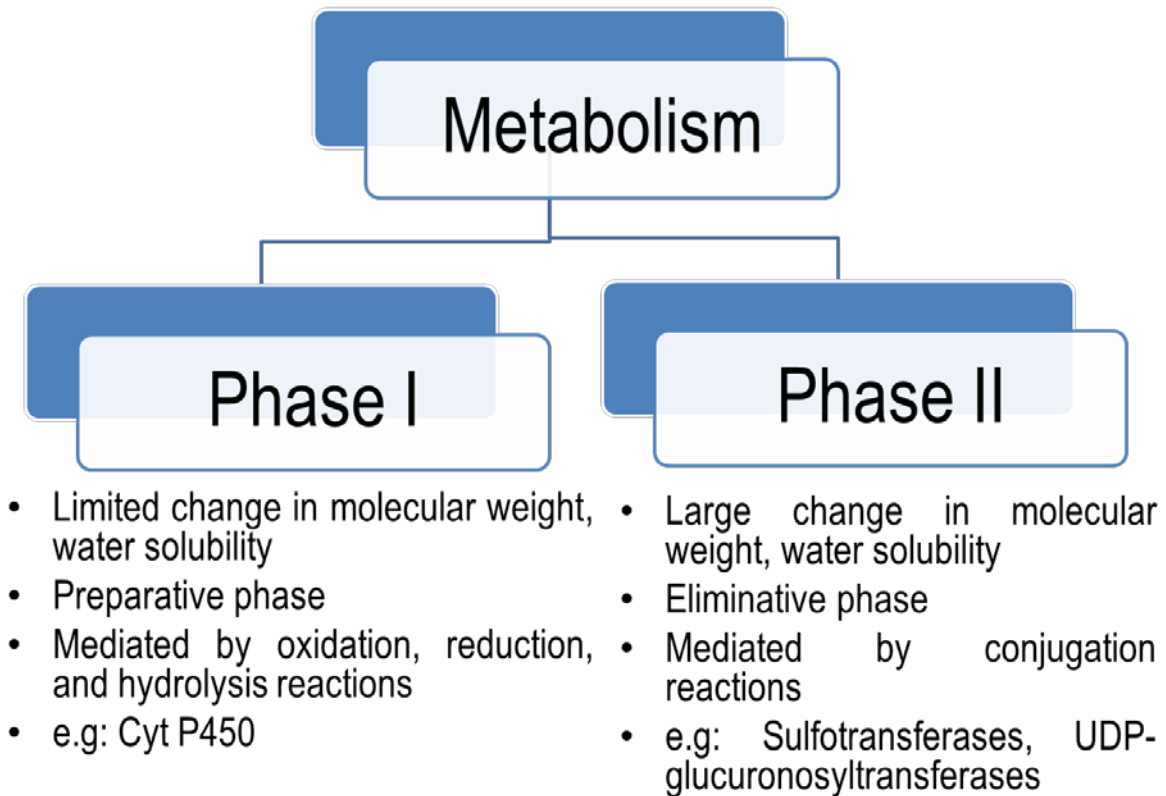


Figure 1: Simple comparison between metabolism phases and their role in drug metabolism.

1.2 Sulfotransferases

Sulfotransferases (SULTs) are the second major phase II metabolic enzyme family (after glucuronides) in adult liver and the only in fetal liver ⁶. As all metabolic enzymes, they show low substrate specificity and flexible catalytic binding sites. Despite working in coordination with glucuronosyl transferases on overlapping substrates, SULTs have high affinity and low capacity working profile ⁷. Sulfonation takes place at low substrate concentration and glucuronidation at high substrate concentration when SULTs are saturated ⁷. As illustrated in Figure 2, they catalyze the transfer of a sulfonate group from a universal donor: 3'-phosphoadenosine 5'-phosphosulphate (PAPS) to an O-, N- or S-acceptor group of an appropriate molecule (endogenous or xenobiotic) ³. The reaction generates a sulfonated product (LigS) and inactive cofactor (PAP). Basically, two general classes of SULTs, named after their cellular localization, are identified: cytosolic and

membrane-bound ⁸. As explained in Table 1, several differences can be pointed out between these two families ^{9, 10}. Membrane-associated SULTs play important roles in crucial biological processes in the body like cell communication, defense and growth. They sulfonate large biomolecules such as carbohydrates and proteins. However, they are never involved in drug metabolism because of their localization in the luminal side of the Golgi, away from exposure to xenobiotics ¹¹.

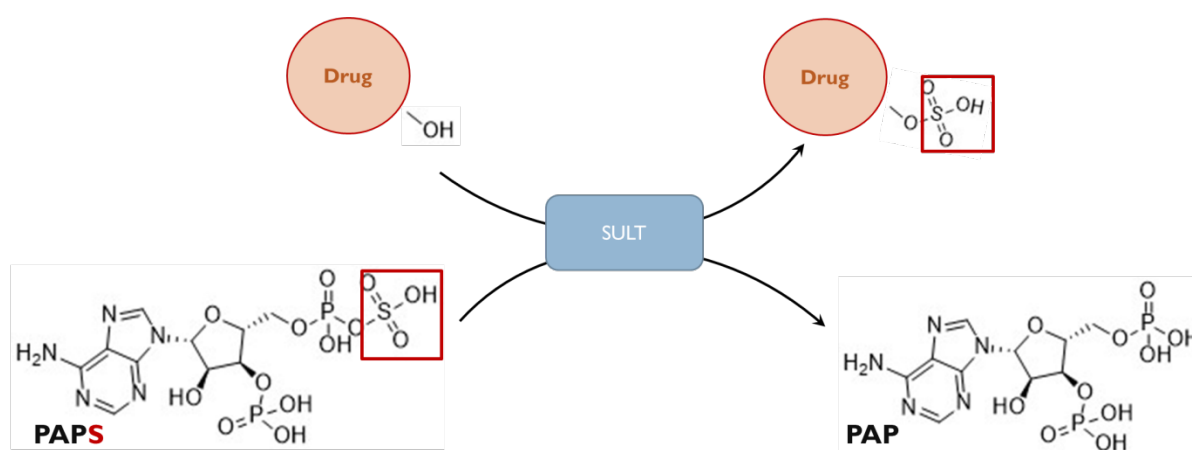


Figure 2: The catalytic reaction catalyzed by SULTs from the universal sulfonate donor (PAPS) (the sulfonate group highlighted in orange) to an acceptor and producing the inactive donor (PAP) and a sulfonated product.

Table 1: Comparison between Golgi (membrane) and cytosolic SULTs.

Difference	Membrane SULTs	Cytosolic SULTs
Significance	cell function	Metabolism and homeostasis
Localization	membrane of the Golgi	Soluble in the cytoplasm
Substrates	glycol proteins, glycopeptides	Hormones and xenobiotics
Substrate pocket	large open cleft with basic residues	deep hydrophobic pocket

Cytosolic SULTs, the focus of this work, have ubiquitous tissue distribution ³, and play various roles in phase II metabolism of xenobiotics and some endobiotic molecules such as hormones and bioamines ¹². SULTs have a defensive function. They detoxify a wide variety of drugs by increasing their hydrophilicity and making them readily excretable via urine. Moreover, the added sulfonate group modifies the receptor affinity and thereby causes termination of the drug therapeutic effect. This function is responsible for drug development failure at early stages of many therapeutics because it lowers the effective drug concentration at site of action below its therapeutic concentration. In addition, SULTs play an important role in the homeostasis of steroidal hormones oscillating between a sulfonated hormone (inactive) and the active hormone ¹³. Effective concentrations of

steroidal hormones are significantly altered when a xenobiotic is extensively sulfonated by similar SULT member. This effect may disrupt the availability of the hormone active forms and therefore the hormone homeostasis¹⁴. In addition, SULTs are responsible for the production of reactive promutagens and procarcinogens¹⁵. The inserted sulfonate group to highly conjugated systems is easily hydrolysable, leaving a resonance-stabilized charged system that has an increased reactivity against body proteins, DNA and RNA causing severe carcinogenic side effects¹⁵.

1.2.1 Molecular structure of cytosolic SULTs

Using X-ray crystallography techniques, it is possible to determine the structure of cytosolic proteins with good resolution (1-3 Å). At the time of writing this thesis, over 40 crystal structures of the various human SULT isoforms have been reported¹⁶ (Table 2).

Table 2: List of all cytosolic SULTs crystal structures.

Isoform	PDB ID	R	Year	crystallized ligands	Ref
SULT 1A1	1LS6	1.9	2003	PAP, two p-nitrophenol	15
	2D06	2.3	2005	PAP, estradiol	17
	1Z28	2.4	2010	PAP	18
	3QVU	2.5	2011	PAP, p-nitrophenol	19
	3QVV	2.35	2011	PAP, cyanoumbelliferone	19
	3U3J	2.7	2011	PAP	20
	3U3K	2.3	2011	PAP, 2-naphthol	20
	3U3M	2.3	2011	PAP, 3-cyano-7-hydroxycoumarin	20
	3U3O	2	2011	PAP, two 3-cyano-7-hydroxycoumarin	20
	3U3R	2.3	2011	PAP, two p-nitrophenol	20
4GRA	2.56	2013	PAP	21	
SULT 1A2	1Z29	2.3	2010	PAP	18
SULT 1A3	1CJM	2.4	1999	sulfate ion	22
	2A3R	2.6	2005	PAP, dopamine	23
SULT 1B1	3CKL	2	2008	PAP, resveratrol	a
SULT 1C1	2Z5F	2.1	2006	PAP	24
SULT 1C2	3BFX	1.8	2006	PAP	24
	2GWH	1.8	2007	PAP, pentachlorophenol	25
	2AD1	2	2007	no ligand	25
SULT 1C3	2REO	2.65	2007	PAP	a
	2H8K	3.5	2007	PAP	25
SULT 1E1	1AQU	1.6	1997	PAP, estradiol	26
	1AQY	1.75	1997	PAP	26
	1BO6	2.1	1998	PAP, vanadate	27
	1HY3	1.8	2002	PAPS	28
	1G3M	1.7	2003	PAP, ethandiol, 3,5,3',5' tetrachloro-biphenyl 4,4'-diol	29
	4JVL	1.94	2013	PAP, estradiol	30
	4JVM	1.99	2013	PAP, 4,4'-propane-2,2-diylbis(2,6-dibromophenol)	30
4JVN	2.05	2013	PAP, 2,6-dibromo-3-(2,4-dibromophenoxy) phenol	30	
SULT 2A1	2QP4	3	2007	etiocolanolone	31
	2QP3	2.6	2007	etiocolanolone	31
	3F3Y	2.2	2008	PAP, lithocholic acid	a

SULT 2A3	1EFH	2.4	2000	PAP	32
	1J99	1.99	2002	Dehydroepiandrosterone (DHEA)	33
	1OV4	2.7	2004	androsterone	34
	4IFB	2.3	2013	PAPS	<i>a</i>
SULT 2B1a	1Q1Q	2.91	2003	PAP, 2-[N-cyclohexyl amino] ethanesulfonic acid	35
SULT 2B1b	1Q1Z	2.4	2003	PAP	35
	1Q20	2.3	2003	PAP, pregnenolone	35
	1Q22	2.5	2003	PAP, DHEA	35
SULT 4A1	1ZD1	2.24	2007	No ligand	25

Abbreviations: PDB ID = Protein Data Bank entry number, R= resolution, PAP = 3'-phosphoadenosine-5'-phosphate, PAPS = 3'-phosphoadenosine-5'-phosphosulfate, a = to be published.

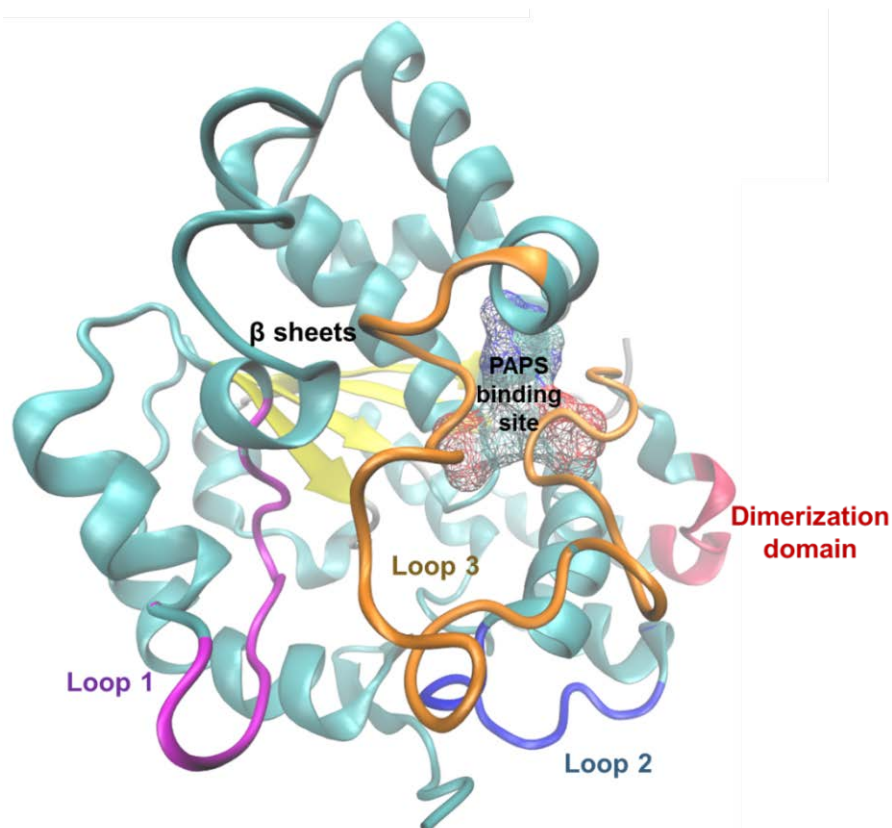


Figure 3: General structure of SULTs: 4-B sheet backbone (yellow), Loop 1 (magenta), Loop 2 (blue), Loop 3 (orange), dimerization domain (red), PAPS binding region (wired surface).

SULTs are globular dimer proteins composed of two monomeric subunits⁸. SULTs have two interconnected pockets in the core of each monomeric subunit: The substrate binding pocket and the cofactor (PAPS) binding pocket. Each monomer consists of a single alpha beta domain with characteristic four-stranded parallel β sheets (Figure 3). The beta sheets constitute the heart of the binding sites for the substrate and the cofactor (sulfonate group acceptor and donor, respectively)³⁶. The core structure of these isoforms remains remarkably preserved, keeping several domains conserved within SULTs family, including phosphosulfonate binding site (PSB) which involves PAPS binding, part of substrate pocket (Loops 1, 2 and 3), the catalytic interface and the dimerization domain¹¹ (Figure 3).

1.2.1.1 PAPS binding site

The PAPS-binding region is conserved not only at the amino acid molecular level and the 3D structure, but also on the level of charge distribution for all SULTs (membrane and cytosolic). PAPS is well accommodated in SULTs via three prominent interactions: (i) the 5'-phosphosulfonate group interacts with PSB, (ii) the 3'-phosphate group binds to phosphate binding (PB) loop, and (iii) the purine ring is stabilized by direct interaction with the β -strand-loop- α -helix (SLH3) motif³⁷ (Figure 4).

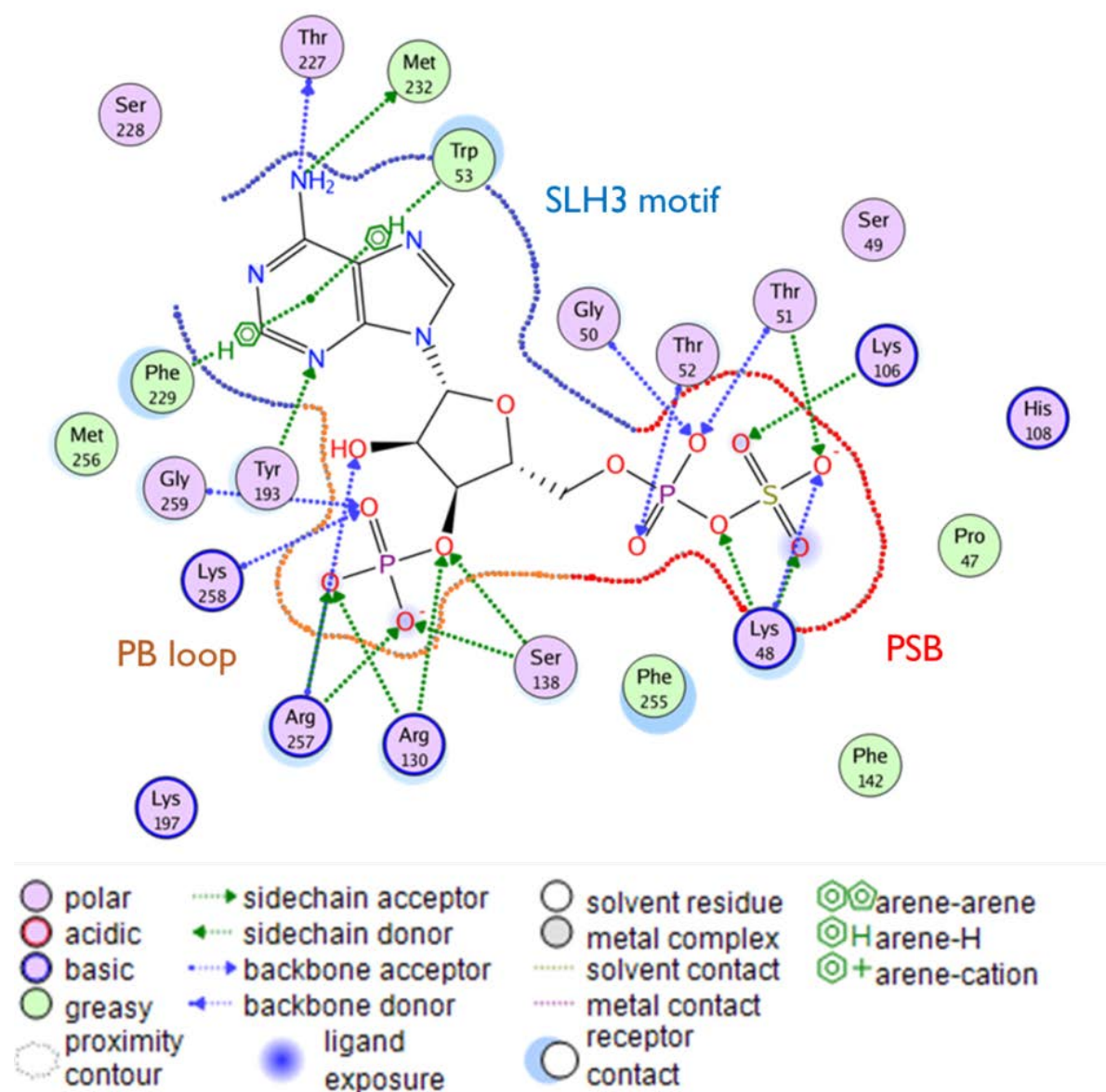


Figure 4: Binding of the sulfonate donor in SULTs: PAPS is accommodated by three interactions; 5' phosphosulfonate group with PSB (red dotted region), 3' phosphate group with PB loop (orange dotted region), and hydrophobic interaction of adenine ring with SLH3 motif (blue dotted region). PAPS shows tight interaction with the enzyme residue and the sulfonate group is the only exposed group.

PSB is defined by residues 45-51(SULT1A1 numbering). It includes residues that interact via hydrogen bonds to 5'-phosphosulfonate group (Lys48, Thr51 and Thr52)³⁷. Thr51 and Thr52 are responsible for keeping the 5'-phosphosulfonate group in proper place for catalysis and Lys48 is involved in the sulfonate group transfer to the acceptor molecule. The second interaction region, the PB loop, is formed by residues 257- 259 (part of loop 3). Loop 3 shows high flexibility in case of apoenzyme. When the sulfonate donor binds (whether in active or inactive form), it stabilizes loop 3 via hydrogen bonds formed between 3'-phosphate group of the cofactor (PAP/S) and the amide N atoms of Arg257, Lys258 and Gly259, in addition to the side chain of Arg257¹¹. In addition, Arg130 and Ser138 are also involved in hydrogen bonding stabilization of 3' phosphate group. The direct interaction of the cofactor with the backbone of loop 3 decreases its mobility and freezes it in a rigid conformation that is crystallized in all cofactor-containing SULTs crystal structures. A particular loop 3 movement subsequent to cofactor binding has recently been reported in all isoforms²¹. Finally, the third interaction site is a hydrophobic pocket consisting of Phe229 and Trp53. They appear to interact via π -stacking of the purine ring of PAPS, therefore stabilizing its orientation within the enzyme³⁷.

The only member that has a different PAPS binding site is hSULT4A1. It lacks Trp53 for adenine ring stacking. In addition, it has a short loop 3 that deprives Arg257 from interaction with 3' phosphate group. Overall, the cofactor binding region is smaller than other isoforms and that open questions if it might function without binding any cofactor³⁸.

1.2.1.2 Substrate binding site

Although the cofactor binding region is absolutely conserved in all SULTs, the substrate binding pocket is not. Sequence variations present in acceptor binding sites define a distinct (but also partly overlapping) substrate specificity for each isoform¹⁶. The acceptor pocket is flanked by three (for SULT1 subfamily), or two (for SULT2 subfamily) flexible loops (Figure 3). Loop 1 (residues 81- 90) is exclusively present in SULT1 isoforms. It shows high variability among the subfamily members in comparison to the rest s sequence identity. Loop 2 (residues 145-154) and Loop 3 (residues 235 - 263) are present in all SULTs and share a low level of conservation between isoforms⁷.

The three loops are highly flexible and therefore disordered in most of the crystal structures especially those that do not include bound ligands. Their conformations are induced and stabilized by direct interaction with various bound ligands. Their disordered flexible structures might be the reason behind the lack of crystallized apoforms³³. Loop 3, containing more than twenty residues is the largest of the three loops. It consists

of two segments: One covers the outer part of the acceptor pocket and the other overlays side of the donor pocket (Figure 3). Binding of either the sulfonate donor has been reported to significantly stabilize the entire Loop 3. This stabilization is the basis for cofactor structural priming for the conformation of substrate binding site²⁵. Recent *in vitro* testing using enzyme activity and binding assays have confirmed the Loop 3 structural shift after cofactor binding and its influence on substrate specificity by inducing active site conformation causing smaller pocket size³⁹.

For cytosolic SULT members, they all share a variable-size deep substrate pocket, which is comprised of hydrophobic residues. The distinct substrate specificity for each isoform comes from the variation in the three flexible loops. Substrate specificity was previously used for SULT classification and because of the overlapped substrate specificities between isoforms, sequence similarities are now used instead to group members in various families (See section 1.2.5)

1.2.1.3 Dimerization domain

In vivo, SULTs exist as homodimer. The dimerization domain has the preserved sequence KxxxTVxxxE, which is 100% conserved across all human SULT isoforms. Having only ten residues in total, it is considered one of the shortest protein dimerization domains recognized to date⁴⁰. The SULT dimerization domain is a symmetric interface consisting of two antiparallel monomer backbone segments. Monomers are connected with four hydrogen bonds in between and two inter-molecular salt bridges formed at its extremities, between Lys265 (SULT1A1 numbering) from one monomer and Glu274 from the other partner in the dimer and vice versa. The KTVE dimerization motif is ≈ 25 Å away from the substrate pocket and next to Loop 3 and the donor binding pocket⁴¹.

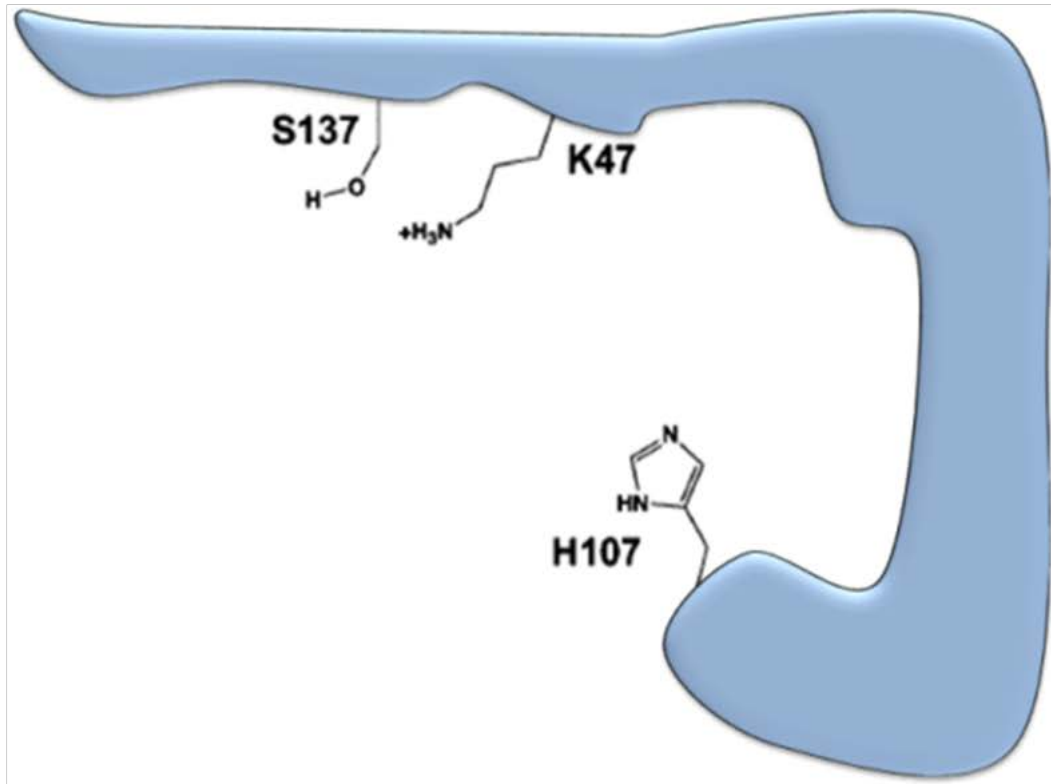
Although the SULT dimerization domain is highly conserved, its biological significance is yet unclear. Mutated SULT monomers show almost the same enzyme activity despite the decreased substrate inhibition behavior and heat stability^{42, 43}. Furthermore, SULT isoforms exhibit half-site reactivity, where only half of the catalytic pockets (in the dimeric enzyme) catalyze sulfonation at any given time⁴⁴. This phenomenon is known and also reported for other enzymes as aldehyde dehydrogenases⁴⁵. A recent study by Tibbs *et al.* has revealed that the monomers in SULT dimeric complex communicate with each other in a coordinated manner mediated by presence of PAP/S⁴⁶. The authors suggest that SULT is acting as oscillating PAP/PAPS dimer with only one active monomer at any study time⁴⁷. This study highlights another side of SULT complexity in defining substrate specificity and selectivity in the SULT family.

1.2.1.4 Catalytic interface

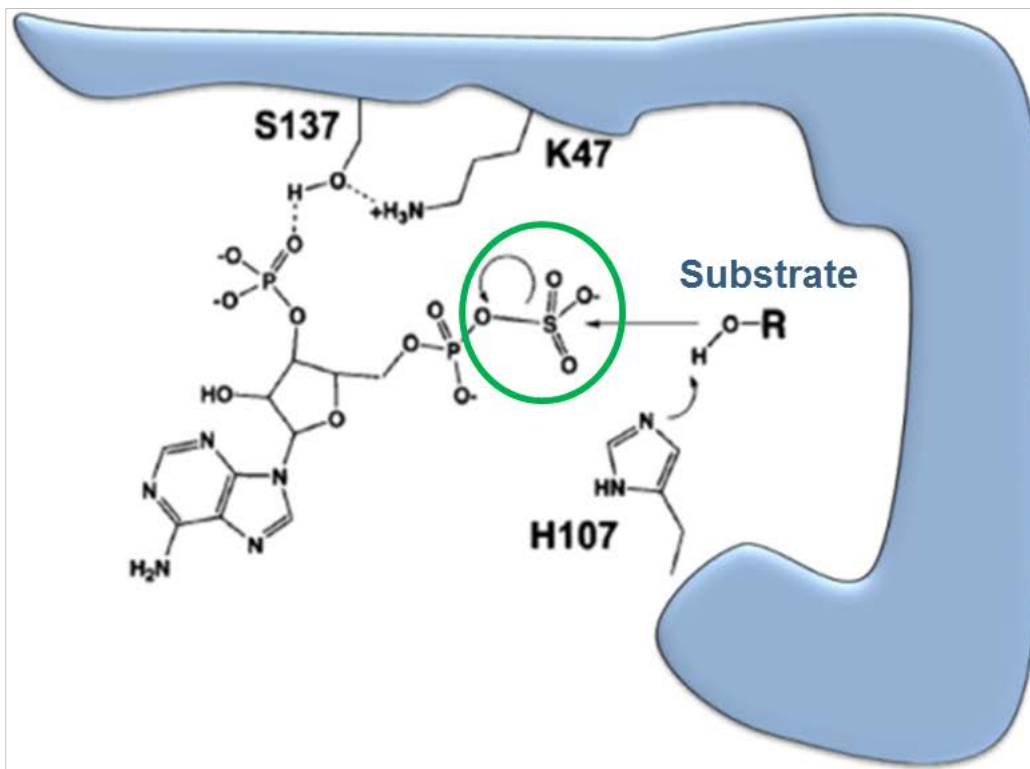
The sulfonation reaction takes place in SULTs via an in line attack of the substrate at the sulfonate group of PAPS. The conserved active site catalytic residues (His108, Lys106 and Lys48) catalyze the substrate activation, sulfonate group transfer and sulfonate product release²⁷.

As illustrated in Figure 5, His108 acts as a catalytic base that attracts a proton from the substrate acceptor group and converts it into a strong nucleophile. Once formed, the nucleophile attacks the sulfur atom of the active cofactor (PAPS) that increases a negative charge on the bridging oxygen. At this moment Lys47 supports the dissociation of sulfonate by donating its proton to the bridging oxygen. Although Ser137 is a part of the PB loop that accommodates 3' phosphate group of the cofactor, it plays a very critical role in preventing early hydrolysis of the sulfonate group of the cofactor, before sulfonate transfer to the substrate, by keeping the side chain of lys47 away from interaction with the bridging oxygen²⁷. In addition to these residues, Lys 106, only conserved in the SULT1 family, is reported to assist in the positioning of the substrate into the catalytic pocket and stabilizing the transition state through its interaction with bridging oxygen⁴⁸. Mutagenesis studies of conserved catalytic residues support the suggested transfer mechanism. His108 is obligatory for enzyme activity. In addition, mutation of the other residues (Lys47, Ser137 and Lys106) significantly reduces enzyme activity⁴⁹.

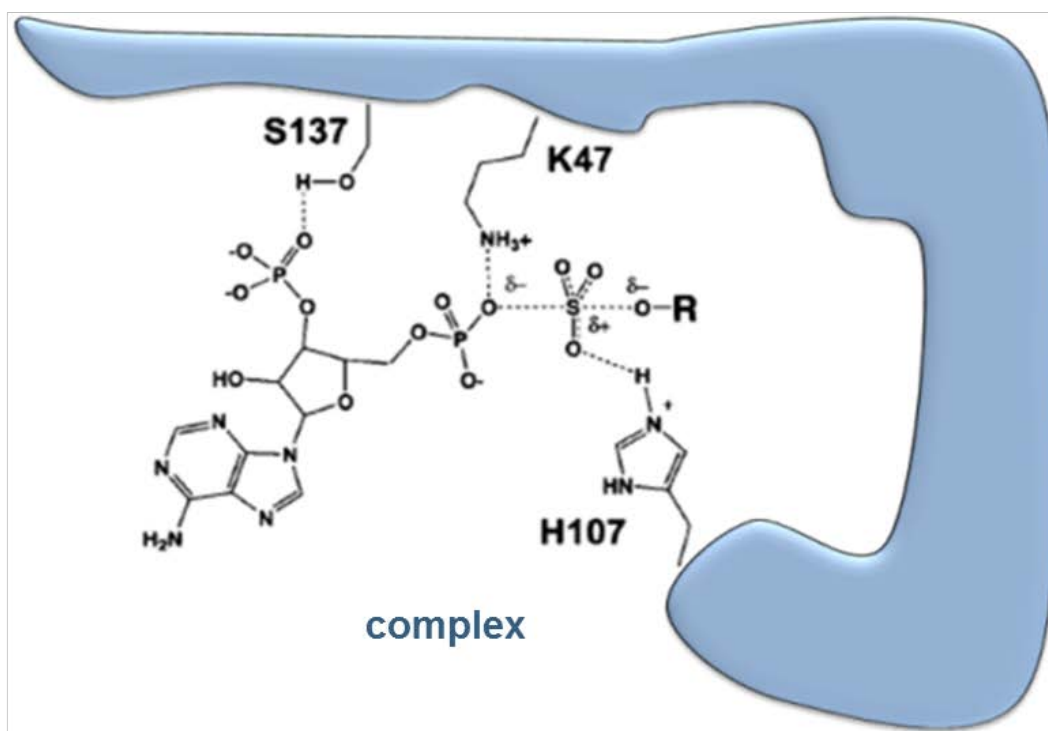
1.



2.



3.



4.

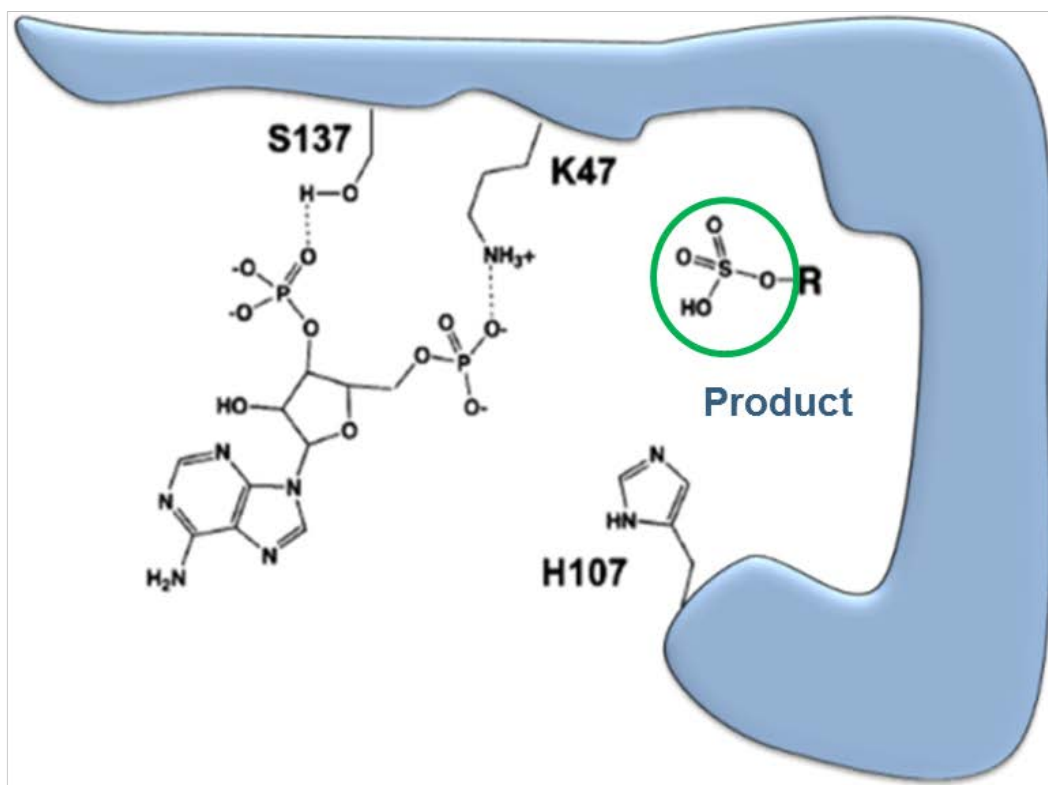


Figure 5: Sulfonation mechanism. 1. The conserved catalytic residues Lys47 (Lys48 in SULT1A1 numbering), Ser137 (Ser138 in SULT1A1 numbering), and His107 (108 SULT1A1 numbering). 2. Histidine mediates the activation of the substrate into a strong nucleophile. 3. Transition complex formation and stabilized by Lys residue. 4. Sulfonate group transfer and the production of sulfonated product and inactive cofactor.

1.2.2 Sulfonation donor (PAPS)

PAPS, the active sulfonated cofactor acts as sulfonate donor in all sulfonation reactions all over the body. Therefore, the level of possible sulfonation is limited by the availability of PAPS or its production precursors. It is produced by a two-step reaction that utilizes ATP and sulfate⁵⁰. As illustrated in Figure 6, ATP is first transformed to adenosine 5' phosphosulfate (APS), and then converted to PAPS. In the human body, both reactions are coupled in the two-domain PAPS synthetase (PAPSS) enzyme. Two isoforms of PAPSS were recognized showing different activity and body distribution; PAPSS1 and PAPSS2⁵¹. Although PAPSS1 is ubiquitously expressed in the human body and responsible for maintaining basal PAPS levels, it exhibits about 10 folds less activity than PAPSS2⁵². Having high catalytic efficiency and tissue specific expression, PAPSS2 is the predominant isoform in human liver. PAPS synthesis depends as well on the availability of inorganic sulfates. The decrease in sulfate concentration due to high glutathione activity or sulfur-deficient diet significantly affects physiologic PAPS levels⁵³.

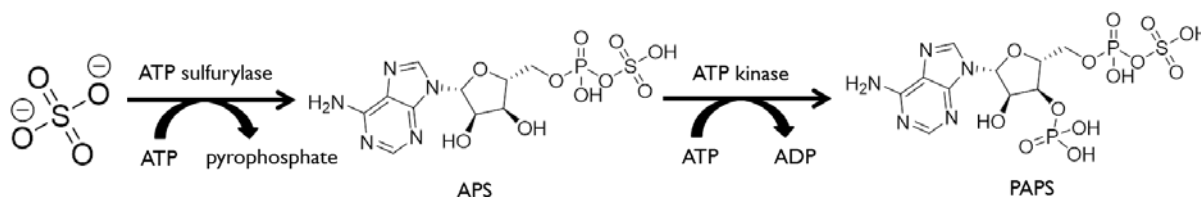


Figure 6: Synthesis of PAPS in human body from ATP and free sulfate. The two-step reaction occurs in PAPS synthetase enzyme that has coupled domains for both steps (ATP sulfurylase and APS kinase domain).

After PAPS synthesis, it is transported by PAPS transporters (PST) from the nucleus (the major site of production) to the Golgi apparatus to be consumed by membrane-SULTs, or to the cytosol to be exploited by cytosolic SULTs⁵⁴. Two PST isoforms (PST-1, PST-2) are identified to transport PAPS from the cytosol to the Golgi but similar transporters that mediate the transport of PAPS out of the nucleus into the cytosol are not yet characterized⁵⁵.

SULT capacity therefore depends on the local concentration of PAPS. The empty apoenzyme SULT has a big binding site that can adopt both large and small substrates. Binding of one PAPS to one monomer of the dimeric enzyme restricts the space and opening, making it selective for only small substrates⁵⁶. On the contrary, saturation of both monomers with PAPS greatly enhances the catalytic efficiency towards larger substrates up to two orders of magnitude⁵⁷. Knowing that the PAPS level varies from one tissue to another explains why SULT reactivity and substrate specificity are highly tissue dependent⁷. The liver as the primary metabolism organ, shows the highest PAPS levels (\approx 25 nmol/g tissue) and shows the double occupancy state to support the defensive role of

hepatic SULTs⁵⁸. The influence of PAPS as an activity regulator to SULTs adds to the complexity of the whole system and makes it very difficult for *in vivo* sulfonation prediction.

1.2.3 Catalytic mechanism

Crystal structures contributed to our understanding of the reaction mechanisms by revealing catalytically positioned acceptors in the catalytic interface²⁷. Although experimental results are crucial to reveal mechanistic aspects of enzyme mechanisms, reports targeting SULTs-catalyzed reactions could not confirm the reaction type of the sulfonation process. However, many studies using enzyme binding tests and structural crystallography helped massively to clarify the picture.

As described before, sulfonation takes place by an in-line attack of the nucleophile (acceptor) on the sulfonate group of PAPS. His108, the catalytic residue in active site is globally conserved. Its role in the sulfonation mechanism is either to attract the nucleophile proton to initiate its activity (SN2), or attack the sulfonate group itself to form an unstable protein sulfonate complex (SN1). The earliest reports support the dissociative (SN1-like) reaction by the formation of loose transition state with minimal involvement of the nucleophilic substrate⁵⁹. However, solved crystal structure of the enzyme-cofactor-ligand complex turned this hypothesis upside down. As illustrated in Figure 5-4, an associative (SN2-like) mechanism was proposed suggesting an involvement of the acceptor together with PAPS in forming the transition complex. The complex represents a properly oriented acceptor ready for a nucleophilic attack on the sulfur atom of PAPS with a fully intact bridging oxygen bond⁴⁸. More studies on the SULT reaction mechanism are required to further rebat this issue and decide if the reaction mechanism is pure SN1, SN2 or hybrid SN1/SN2 as it is in some other catalytic enzymes (e.g. yeast protein farnesyltransferase)^{60, 61}.

The SULT enzymatic mechanism is named “bi-bi” as the reaction involves two co-substrates (PAPS and compound), and releases two products (PAP and sulfonated compound)¹¹. The order of addition is different in both cytosolic and membrane SULTs. There are two basic enzyme well-defined addition mechanisms for dual-substrate enzymes (such as the SULTs); sequential and non-sequential mechanisms (Figure 7). Cytosolic SULTs perform sulfonation via a sequential mechanism. In this type of mechanism, a ternary complex has to be formed by a simultaneous binding of both substrates to the enzyme, subsequent to the catalysis and the release of two products (Figure 7-A)⁶². On the other side, the membrane SULT follows the non-sequential mechanism. It requires only one

substrate to bind to the enzyme forming an activated intermediate, for the first product to be released. Now the complex is ready to accept a second substrate, catalyze the reaction, and release the final product in a double-displacement mechanism, also known as the “ping-pong” mechanism (Figure 7-B) ⁶³.

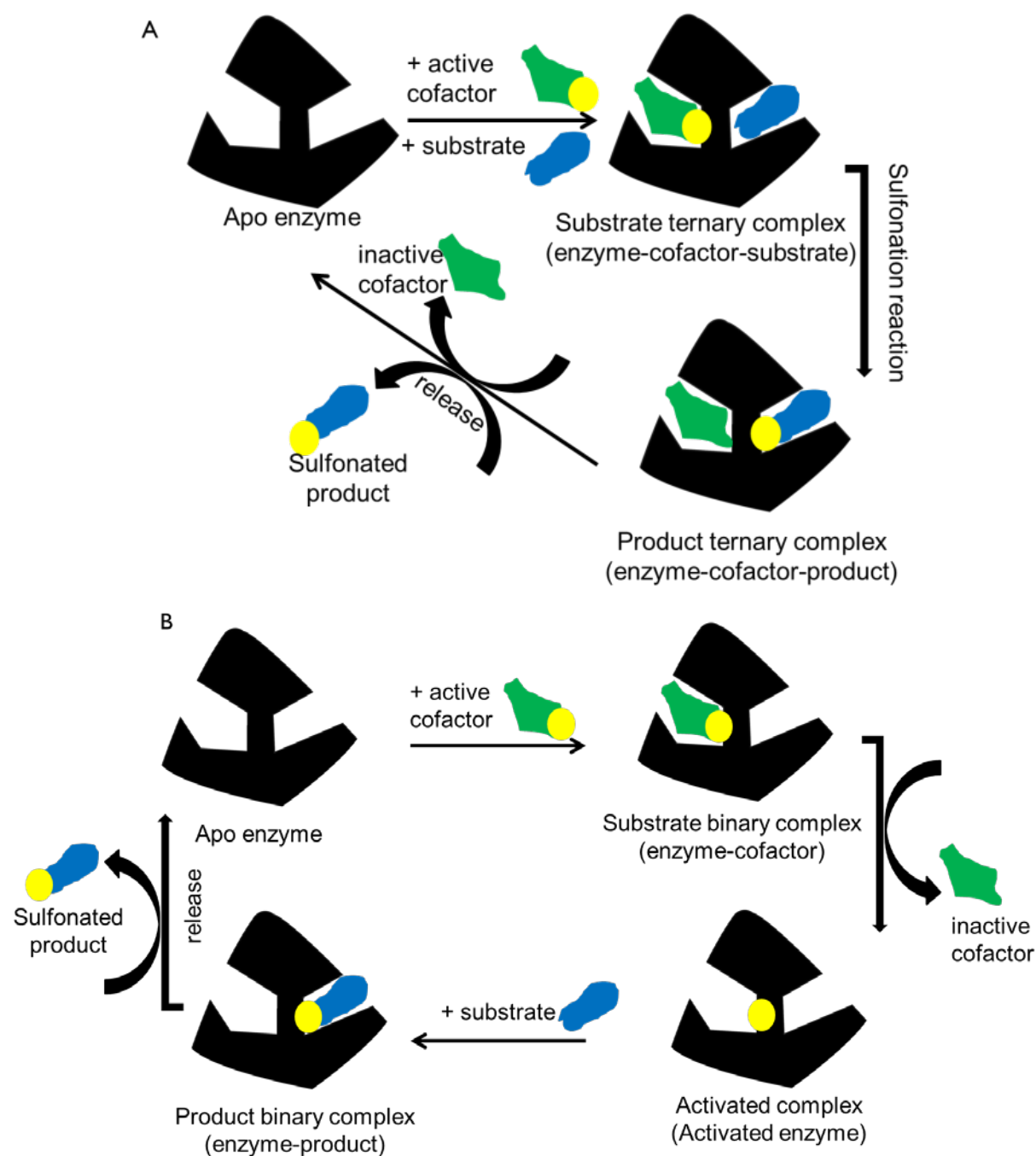


Figure 7: Sequential versus non-sequential kinetic mechanisms. A. Sequential addition in which both co-substrates are involved in the formation of ternary complex (enzyme-cofactor-substrate) for the reaction to take place. B. Non-sequential mechanism in which only one substrate is involved in activation of the enzyme then the second substrate binds in a double-displacement mechanism.

Within the sequential mechanism there are two different kinds defining the dependence of both substrates on each other: (i) Random and (ii) ordered. In the random-fashioned sequential mechanism, each substrate binding is completely independent with no obligation for specific addition order. Alternatively, a substrates binding order is mandatory for the formation of a productive ternary complex in an ordered-fashioned sequential mechanism. Determination of the type of addition order is a controversial and debatable subject. Some early studies, suggested an ordered mechanism for the SULT isoforms in which the active cofactor (PAPS) is obligated to bind first^{64, 65, 66}. Other literature supports that SULTs proceeds by a random mechanism^{67, 68}.

The origin of discrepancy is that SULT behavior cannot be fully explained by a solely ordered or random pattern. The formation of dead-end complex (E-PAP-substrate) which is well known to be formed by all SULTs, is incapacitated by the ordered addition⁶⁴. On the other side, the random-fashioned addition mechanism falls short in explaining the positive or negative synergy that happens to some SULTs substrates binding when they bind consequent to the cofactor^{39, 56, 69}. Actually, the real mechanism might be in between these two extremes depending on the kind of attacking substrate. Further experimental studies on the SULT reaction addition mechanism targeting several substrates are required before announcing this conclusion.

A recent paper by Wang et al.⁷⁰ suggests the standard sulfonation cycle of SULTs depending on the random pattern sequential mechanism to proceed as illustrated in Figure 8. In this cycle, the enzyme (E) has equal chances to initially bind the acceptor (Lig) or donor (PAPS) and form a binary complex. Afterwards, the ternary complex is formed by the addition of the second substrate. If the addition mechanism is purely random, the formed ternary complex will be productive regardless of the order of addition. Actually most SULT substrates exhibit this kind of addition, showing neutral synergy when bind to cofactor-containing complexes. Others show enhanced binding affinity (positive synergy) if bind after the cofactor⁶⁹. A third category of SULTs substrates describe a lower affinity towards cofactor-containing compounds⁵⁶. The two last categories show a degree of "ordered" addition mechanism in forming the productive ternary complex. The binding order seems to be important for some substrates and may play a role in substrate selection of SULTs.

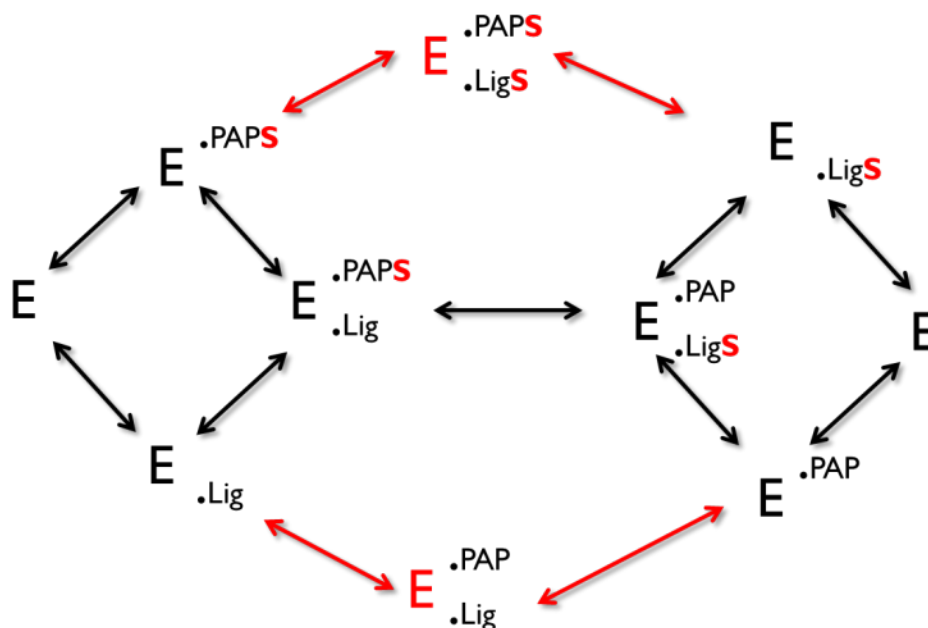


Figure 8: The sulfonation catalytic cycle. Random bi-bi addition mechanism involves two substrates (PAPS and Lig) and two released products (PAP and LigS). The dead-end complexes (red arrows) are formed in non-productive ternary enzyme complexes with non-interacting substrates. This diagram was modified from ⁷⁰.

1.2.4 Substrate inhibition

SULTs, like many other metabolic enzymes, are inhibited by their own substrates, showing a velocity declination as the substrate concentration increases beyond the maximum velocity (V_{max}) instead of leveling off ⁷¹. Substrate inhibition is a regulation phenomenon in enzyme kinetics with important biological significance. It is widely spread in the enzyme families controlling diverse biological functions including synthesis stability and preventing rapid degradation ⁷². Although many cases of SULT substrate inhibition have been reported, its biological importance is still to be clarified. SULT substrate inhibition might be a way of natural regulation for substrate selectivity and specificity of various isoforms.

Inhibition of SULT activity by their own substrate can take place via various mechanisms ⁷¹. First, an additional substrate in the active site or in an inhibitory binding site is responsible for decreasing the overall catalytic efficiency of sulfonate transfer. SULT1A1 is the only isoform that show crystal structures complexed with two molecules of the same substrate (e.g., p-nitrophenol (pNP) ¹⁵ or 3-cyano-7-hydroxycoumarin (3CyC) ²⁰. Moreover, other members (i.e. SULT1A3) can be modeled with two substrates as well ⁷³. Second, ternary dead-end enzyme complexes [E.PAP.Substrate] might be connected to substrate inhibition in metabolic enzymes by hindering the completion of the catalytic cycle. This enzyme complex is completely non-productive and its turnover takes time in which the enzyme is fully inhibited. A structural evidence for this mechanism is

crystallized also in other isoforms. In SULT1A1, estradiol (E2) shows a non-catalytically competent binding mode interacting with the cofactor and different from its competent mode crystallized in SULT1E1¹⁷. Another example is crystallized in SULT2A1: DHEA binds in a non-productive manner that might evolve by hydrogen bond towards PAP to a dead-end complex³³. Another third mechanism is explained with no structural evidence but emerged from comparing SULT2A1 crystal structures with and without cofactor; binding of substrate before the cofactor induces conformational changes in the cofactor binding pocket that disfavor cofactor binding and cause substrate inhibition²⁵.

1.2.5 SULTs isoforms

X-ray crystallography has significantly contributed to our knowledge by revealing the molecular structure of similar and heterogeneous isoforms. Therefore, a new nomenclature was approved based on protein sequence similarities⁷⁴. SULTs known up to date can be divided into three main families and collectively account for thirteen distinct members shown in Figure 9⁷⁴: SULT1—A1, A2, A3, A4, B1, C2, C4, E1; SULT2—A1 and B1 (SULT2B1_v1 and SULT2B1_v2); SULT4A1 (SULT4A1_v1 and SULT4A1_v2). The isoforms sharing at least 45% of amino acid sequence identity are grouped in the same family, whereas members with homology up to 60 % are classified also in the same subfamily¹¹.

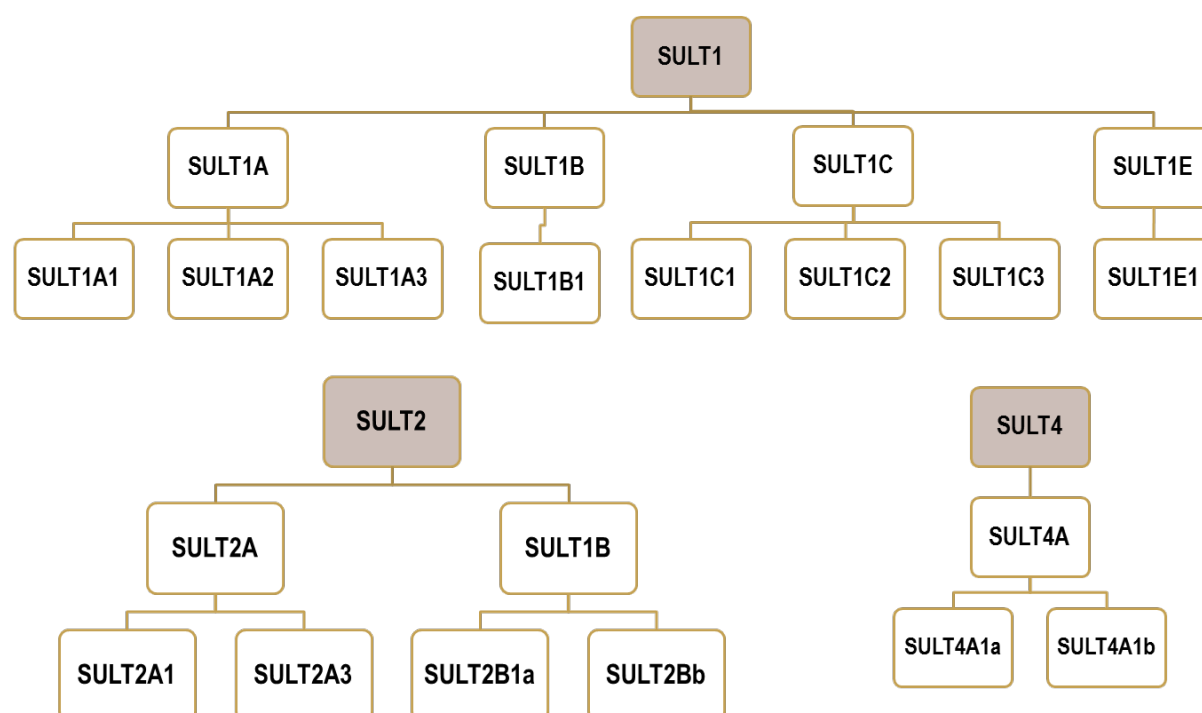


Figure 9: SULTs isoforms. Classification of SULT isoforms according to global sequence similarity.

Although SULTs share a high level of conservation, each SULT isoform has recognizable features that differentiate it from the others. Considering substrate specificity in general, SULT isoforms sharing global high sequence similarity, are expected to catalyze similar substrates. This assumption is correct enough to give a general activity profile to each SULT family. For example, all SULT1 family members have higher efficiency towards phenolic compounds, whereas members of the SULT2 family exhibit higher catalysis for steroids. However, lots of exceptions were detected that highlight the importance of considering the local sequence similarities on the level of the binding site. For example, SULT1E1 mediates the metabolism of estradiol hormone rather than the typical small phenolic compounds metabolized by other family SULT1 isoforms²⁵. Another prominent example is SULT1A3 that has more than 96% global sequence similarity to SULT1A1. However it exhibits less activity towards acidic xenobiotics (opposite to SULT1A1) and higher activity for catechol sulfonation⁷⁵. It is worth mentioning that SULT1A3 local sequence alterations from SULT1A1 are 8 residues located in substrates binding loops (Loop 1, 2, and 3)²⁵. Although the overlapping substrate specificities amongst SULT isoforms can be explained by a high sequence homology, the opposite is not. In other words, the enzyme global sequence similarity does not translate a similar activity profile if the binding site local similarities are ignored. Therefore, correlating SULTs structures to their activity profiles are still not well defined.

SULTs are widely distributed all over the human body including the lungs⁷⁶, the stomach lining, adult and even fetal liver⁷⁷, duodenum⁷⁸, jejunum and intestine¹⁶. Some isoforms have single localization (SULT4 is present mainly in the human brain). Others have ubiquitous localization in many tissues and organs, i.e. SULT1 is vastly spread in metabolism organs (liver, intestine and kidneys) and other peripheral tissues (lungs, prostate and breast)⁷⁹. Despite their structural conservation, each SULT isoform displays a unique activity profile and specific tissue localization that is pivotal for its biological function.

1.3 Human SULT1A1

hSULT1A1 is the progenitor of all SULTs and the first identified member^{80, 81}. In the old nomenclature, it was termed phenol sulfotransferase (PST) because of its metabolic activity towards phenol substrates⁷⁵. In addition, because of its thermal stability over other members, name was especially given to SULT1A1 to differentiate it from its high sequence similarity sibling: thermostable sulfotransferase (TS PST 1). Beside high specificity for small phenolic compounds, human SULT1A1 exhibits a wide substrate

catalytic profile⁸². Its cDNA was first isolated from the liver cDNA library and characterized by its high affinity towards p-nitrophenol and inhibition by 2,6-dichloro-4-nitrophenol (DCNP)⁷⁵. It is a 32 kDa protein with about 295 residues. The SULT1A1 gene is located on chromosome 16, very close to another gene of a related isoform SULT1A2⁸³.

1.3.1 Molecular Structure of SULT1A1

As revealed by the first SULT1A1 crystal structure, SULT1A1 shares the same framework of all other cytosolic SULT: Four central parallel stranded β sheets surrounded by 12 α helices¹⁵. Like others, SULT1A1 has two connected binding sites in close proximity to each other. The first accommodates the sulfonate-group donor²¹ and the second accommodates the sulfonate acceptor²⁰. The three less conserved loops (Loop 1, 2, and 3) contribute to the unique substrate binding pocket of SULT1A1. It has a deep, hydrophobic L-shaped and phenylalanine rich substrate binding site. SULT1A1 is the SULT isoform with the most X-ray crystallography data available in the Protein Data Bank (PDB); 11 various crystal structures show a flexible pocket bound to variety of small phenyl rings⁸⁴ up to large multi-rings¹⁷.

The acceptor binding site has two overlapping pockets as previously described in literature²⁰. The first pocket (Pocket I) is surrounded by Phe24, Phe81, Lys106, His108, Val148 and Phe247 and responsible of binding the acceptor in the catalytically competent site facing catalytic residues (blue sphere in Figure 10). The two residues Phe81 and Phe142 are in charge of stacking the acceptor's phenol ring and fitting it in front of the catalytic residues His108 and Lys48. The other pocket (Pocket II) formed by Phe76, Phe84, Ile89, Tyr240, and Phe247 is very close to the former and binds the second molecule that decreases the overall catalytic capability of the enzyme (red sphere in Figure 10)²⁰. Similarly, amino acids Phe84 and Phe76 play a crucial role in holding the second molecule in place. A so-called 'molecular clamp' was lately described, showing two key phenyl residues (Phe81 and Phe84) undergoing repositioning and greatly enhancing catalytic efficiency of the enzyme towards high affinity binders⁶⁹.

Structural evidence is supporting the hypothesis that substrates can access the binding cleft through 2 channels which are very approximate to each other: Channel I and channel II⁸⁵. It was not completely clear which is the main responsible channel for a substrate. Channel I (blue arrow in Figure 10) is characterized by the distance between Val148 (from Loop 2) and Phe247 (from Loop3) and is assumed to be the entrance for the main substrate. It might affect the size of the substrate entering to the substrate binding site¹⁸. On the other hand, Channel II (red arrow in Figure 10), represented by the distance

between Ile89 and Phe76, is presumably responsible for the second substrate entrance ⁸⁵. Ile89 as well represents a part of the gating loop of the enzyme (residues from 86-90), which is claimed to be responsible for the access of nonsulfonated substrate and the release of sulfonated product ²⁰.

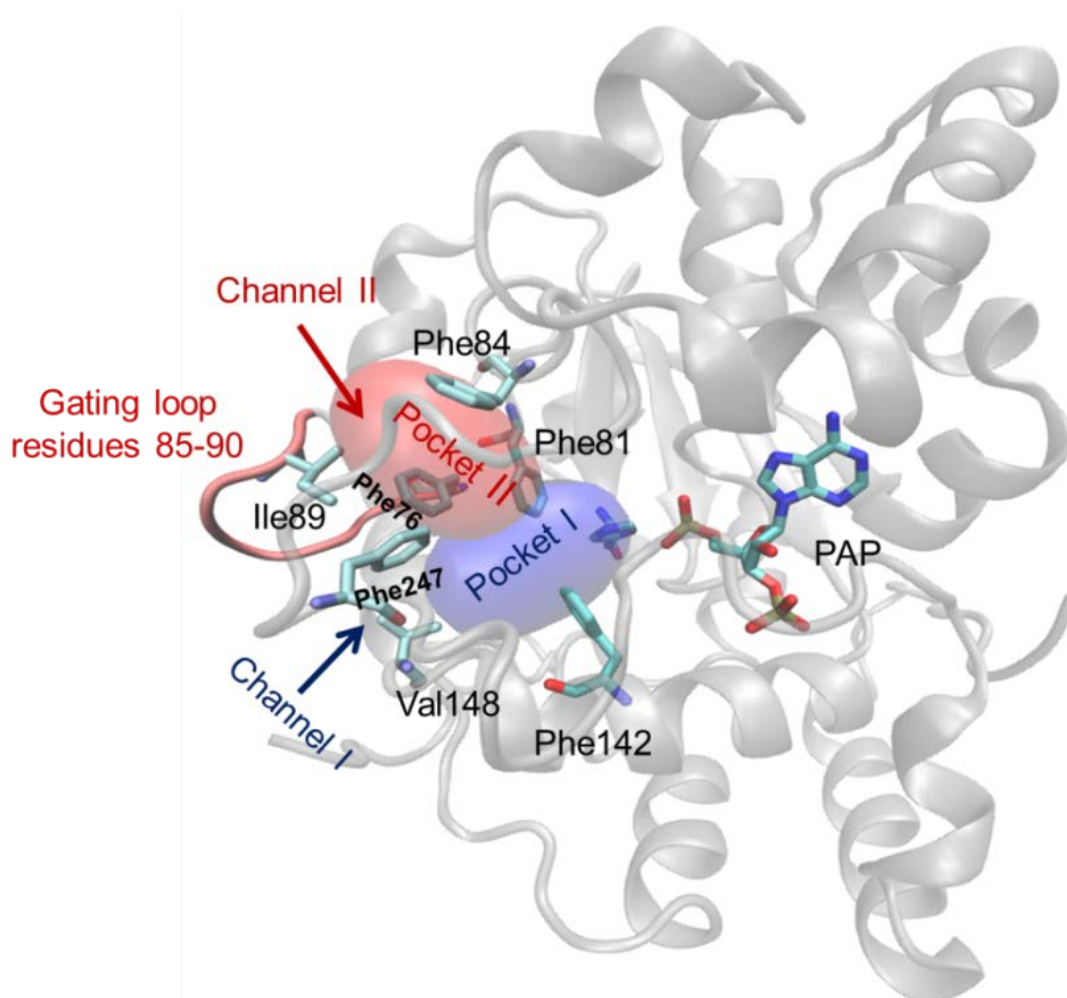


Figure 10: The structure of SUL1A1: (PDB code: 1LS6) structure with the catalytic (Pocket I) and inhibitory (Pocket II) binding sites in red and blue spheres, respectively. The arrows indicate the substrate access channels. The PAP and relevant residues are rendered in a bond model.

Considerable information about the enzyme structure and important residues have been assessed by mutational experimental studies. Point mutations contribute to the understanding of the importance of each residue in SUL1A1 isoform and the structural differences among related siblings. Mutating Phe247 affects the enzyme response to substrate inhibition ⁷³. Furthermore, I89A mutation deteriorates the binding affinity of high affinity small substrates ⁶⁹. Another mutation in the dimerization domain V270E results in formation of a monomeric enzyme instead of the homodimer ⁴². The SUL1A1 mutant A146E shows higher affinity to catecholamine's like its sibling SUL1A3 and more than 400-fold less affinity for its small substrates ⁸⁶.

Structural comparisons of the binding sites for SULT1A1 and various SULT isoforms reveal steric similarities with SULT1B1 and 1C2 but not with SULT1E1. The absence of Loop 1 in SULT2A1 and 2B1 and presence of Loop 1 in β strand in SULT1E1 increase the space of the substrate binding site⁸⁵. SULT1A2 and 1A1 share about 96% of sequence identity, having only one different residue (Tyr149 instead of His) in its substrate binding pocket. This one residue difference results in different catalytic profile, decreasing activity towards pNP, but increasing activities towards hydroxylamine xenobiotics⁸⁷. SULT1A3 also shows high homology with SULT1A1 (95%), with more than 7 different residues different in the substrate pocket⁸⁸. SULT1A3 has more acidic residues in the substrate cleft (Asp86 and Glu146) instead of hydrophobic residues in SULT1A1 (Ala86 and Ala146), which renders its pocket negatively charged disfavoring hydrophobic compounds binding and favoring accommodation of amino-containing catecholamines²³. Although SULT1A1 does not show high sequence similarity with SULT1B1 and 1C2 (~50%), it shows overlapping substrate specificity. This may be due to similarities in geometry and volume of the substrate-binding sites²⁵.

1.3.2 SULT1A1 localization

Previous studies have employed substrate-specific sulfonation, metabolic probes and immunohistochemistry procedures to show that SULT1A1 enzyme has ubiquitous distribution throughout the body. Expression of SULT1A1 has been reported in various tissues such as gastrointestinal tissues, platelets, kidney, brain, prostate, skin, lungs, and liver^{66, 89, 90}. In the major metabolizing organ, the liver, SULT1A1 is the major sulfonating isoform (>50%), followed by SULT2A1, SULT1B1, and SULT1E1. Contrarily, it constitutes a minor fraction (19%) of the total amount of SULT in the small intestine after SULT1B1 and SULT1A3. Despite low levels of overall SULTs in the kidney, SULT1A1 is the most prominently occurring isoform in this organ. In lung tissue, it comprises 20% of the total sulfonating isoforms⁸⁹. In addition, SULT1A1 is found at low levels in mammary glands and epithelial tissue⁷⁹.

1.3.3 Physiological roles of SULT1A1

Due to its wide-spread distribution, SULT1A1 plays two important roles in the human body (Figure 11): The first function is activation of a wide variety of substrates either to a pharmacologically active forms (prodrugs to drugs) or to a harmful carcinogens. The second function is a deactivation task, through which SULT1A1 performs a combined hemostatic/defensive function towards many endobiotics and xenobiotics.

Sulfonation is the metabolic fate of various substrates, commonly associated with chemical inactivation. The chemical reaction renders xenobiotic and endobiotic substrates into biologically inactive moieties that are more water soluble. Xenobiotics are removed from human body through physiological fluids (via the kidney) (Figure 11- c). As illustrated in Figure 11- d endogenous hormones as thyroid hormone ⁹¹ and estradiol ⁷ active forms are kept under physiological hemostasis by the cooperative contribution of SULTs with sulfatase enzyme.

However, few compounds are biologically activated by sulfonation. Minoxidil, an antihypertensive drug that promotes hair growth, is activated via sulfonation to the active metabolite minoxidil sulfate (Figure 11-a). The N,O-sulfate of minoxidil is 10 times more active vasodilator and hair-growth promoter than minoxidil itself ⁹². Many SULTs are involved in its activation, including; SULT2A1, SULT1A3, and SULT1A1 ^{93, 94}. Not all compounds are activated to a pharmacologically favorable species. Some highly conjugated xenobiotics are activated by SULTs into harmful carcinogens. Tamoxifen (TAM) is an interesting example for the role of SULT1A1 in bio activation. The active metabolite, 4-hydroxy TAM, produced after an initial phase I metabolism for TAM, is further metabolized by sulfonation. The electron-withdrawing sulfonate group supports the heterolytic cleavage leaving behind a stable resonating intermediate that is highly reactive. The reactive species can form DNA, RNA, and protein adducts resulting in lots of carcinogenic changes ⁹⁵. Many dietary compounds can be activated to electrophilic mutagens by SULT1A1 ^{83, 96}. Significant differences in bio activation are noticed between the various SULT isoforms and different races, and also among the same SULT from different species ⁹⁶.

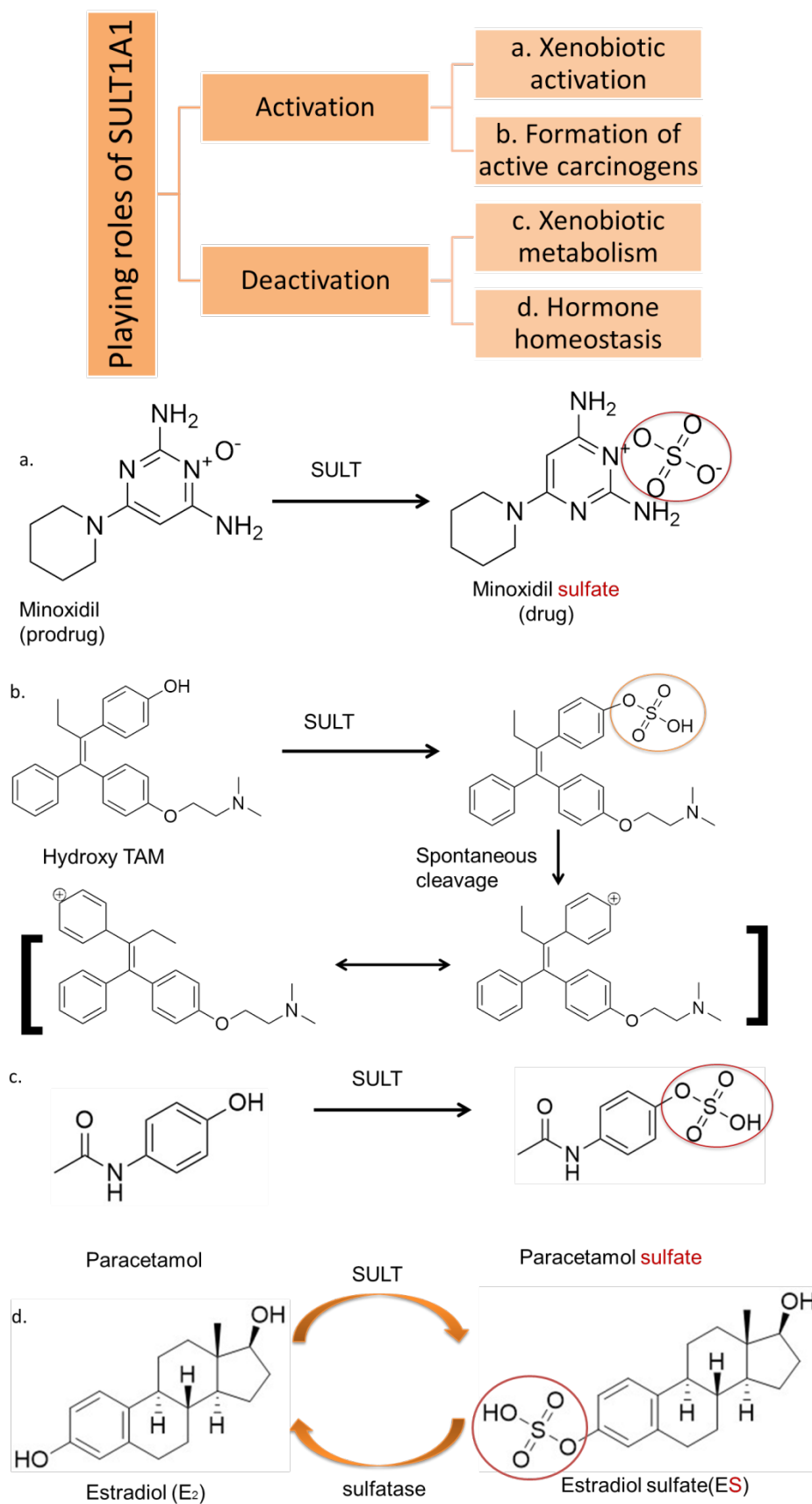


Figure 11: SULT1A1 playing roles. a. Minoxidil activation via sulfonation. b. hydroxyl tamofen (TAM) activation into resonating free radicals. c. paracetamol metabolism vis sulfonation. d. estradiol (E₂) hemostasis mediated by SULT.

1.3.4 Genetic polymorphism in SULT1A1

SULT1A1 shows few number of single nucleotide polymorphisms (SNPs) that contribute to sulfonation variability among individuals. This genetic polymorphism is a change in the gene nucleotide sequence in only one point that results in one amino acid change in the protein sequence. People with different SULT SNPs may have varying levels for SULT activity towards the same xenobiotic. A variety of SNPs in SULT1A1 have been recognized that have a significant influence on its activity⁹⁷. The most common SNP is at Arg213, that can be histidine amino acid in the enzyme allyl SULT1A1*2. This mutation has high allelic frequency (30%) and show significant reduction in sulfonation activity⁹⁸. Another less common SULT1A1 allyl is SULT1A1*3, a M233V mutation with less frequent allelic frequency (10% in African Americans). It is worth mentioning that SULT1A1*3 crystal structure was solved 6 years ago¹⁸. SULT1A1*3 is mostly found very related to the SULT1A1*2 mutation and has been reported to minimize SULT1A1 sulfonation activity⁹⁹. As shown in Table 3, other SULT1A1 allozymes are reported to have altered enzyme activity and thermal stability¹⁰⁰. Several publications have demonstrated the pharmacogenetics of several SULT1A1 allelic variants. The SULT polymorphism might have an effect on developing several cancers such as lung carcinoma¹⁰¹, brain cancer¹⁰², urothelial tumor¹⁰³. Genetic polymorphism is also demonstrated for other SULT isoforms such as SULT1A2, SULT1A3, SULT1C2, SULT2A1, SULT2A3, SULT2B1¹⁰⁴.

Table 3: SNPs responsible for the formation of various SULT1A1 allylic variants.

Allozyme	37	213	223	90	243	282	290	146	181
SULT1A1*1	R	R	M	P	V	Q	S	A	Q
SULT1A1*2	R	H	M	P	V	Q	S	A	Q
SULT1A1*3	R	R	V	P	V	Q	S	A	Q
SULT1A1*4	G	R	M	P	V	Q	S	A	Q
SULT1A1*5	R	R	M	L	A	Q	S	A	Q
SULT1A1*6	R	R	M	P	V	K	T	A	Q
SULT1A1*7	R	H	M	P	V	Q	S	T	G

1.3.5 SULT1A1 substrates

SULT1A1 is the major “phenol” metabolizing SULT, mainly targeting phenol substrates. Other SULT1A1 acceptors are the hydroxyls and primary amines of hundreds of small molecules, including endogenous metabolites, drugs, and other xenobiotics ²¹. Some substrates of SULT1A1 can be generated by Phase I metabolic unmasking, such as some chlorinated environmental pollutants after its oxygenation with Cytochrome P450 ¹⁰⁵. Others are directly metabolized by SULT1A1 without prior Phase I activation.

Describing high affinity SULT1A1 substrates is very challenging. A precise and indicative descriptor is needed to combine the strong binding towards the enzyme with the fast catalytic reaction. There are lots of parameters that can be used to describe the substrate catalysis by an enzyme; such as Michaelis-Menten constant (K_m), maximum velocity (V_{max}), turnover rate (k_{cat}) and catalytic efficiency (V_{max}/k_{cat}) ¹⁰⁶. Also the binding constant (K_d) is a parameter that indicates how strong a molecule binds to an enzyme. However, this parameter does not reflect the speed of the catalytic reaction. Some very strong binders show very low sulfonation or even act as inhibitors ⁸⁴. To consider both factors (binding and catalysis), the catalytic efficiency parameter (k_{cat}/K_m) is the most relevant kinetic parameter to compare the relative rates of enzyme activity on diverse substrates ¹⁰⁷ and has been already used in most SULT modeling studies ¹⁰⁷. k_{cat} gives a direct measure of the catalytic turnover rate, whereas K_m measures the substrate affinity to the enzyme. k_{cat} stands for the catalytic constant and represents the number of substrate molecules turned over per enzyme molecule per second. K_m stands for the Michaelis constant and is defined as the substrate concentration at one-half of the maximum velocity. Altogether, the term (k_{cat}/K_m) represents the perfect descriptor that defines strong substrates having high enzyme affinity and fast turnover.

Mostly all SULT1A1 substrates are planar aromatic small molecules ²⁵. Appendix Table A-1 lists some of the reported substrates to SULT1A1 showing data discrepancy resulting from the experimental data variability generated in several labs, utilizing different biochemical setups (assay conditions and experimental setup). In addition to catalytic efficiency, physiological concentration has to be considered when assaying SULT1A1 substrates. E2 in *in-vitro* studies has been detected to be sulfonated by SULT1A1 and its K_m has been determined to be in the micromolar concentration range ⁹⁹. However, physiological concentration of E2 is less than 1nM ⁶⁷ and clearly, SULT1A1 is not sulfonating E2 *in-vivo* in relevant amounts ¹⁰⁸.

Another way to find the chemical profile of SULT1A1 is to measure the increase in thermal stability due to strong substrate binding. In this experiment, the unequal affinity towards various complexes of the enzyme is introduced. Most SULT1A1 substrates have the ability to stabilize the cofactor-containing complex forming a stable ternary complex and not the apoenzyme. This result suggests the necessity of cofactor binding for proper (tighter) binding of the substrate which is called the cofactor priming²⁵. *In vivo*, enzyme complexes of SULT1A1 occur in all possible combinations of cofactors and binders⁷⁰; namely apoenzyme (SULT1A1), binary complexes (SULT1A1-PAP, SULT1A1-PAPS and SULT1A1-Lig), ternary complexes (SULT1A1-PAP-Lig and SULT1A1-PAPS-Lig) and non-sulfonated cofactor in binary and ternary complexes (SULT1A1-PAP and SULT1A1-PAP-Lig) (Figure 8).

SULT substrates have recently been classified according to their binding affinity towards various complexes into three categories: neutral, positive, and negative synergy substrates. The term describes the dependency of substrate binding on the cofactor binding to be completely independent (neutral synergy), increasingly dependent (positive synergy) or decreasingly dependent (negative synergy). The neutral synergy substrates are compounds that follow a pure random-fashioned binding mechanism with comparably equal binding affinities to binary complexes. If the compound preferentially shows higher or lower affinity towards binary complexes over the apoenzyme, it is described as positive and negative synergy substrate, respectively. Many examples of the three categories are present in literature explaining the effect of donor priming on the binding affinity of the acceptor^{39, 56, 69}.

SULT1A1 substrates vary between neutral, positive and negative synergy substrates⁶⁹. Although several publications describe co-crystallized SULT1A1 ligands^{20, 85, 109}, the intermolecular interactions between the enzyme and its ligands responsible for positive synergy behavior are not well investigated. An explanation to the positive synergy phenomenon has been proposed recently: The previously mentioned molecular clamp undergoes a conformational change of two residues (Phe81 and 84) repositioning the substrate in the binding pocket facing catalytic residues prior to sulfonation. Subsequent to this particular binding pose, positive synergy substrates exhibit twenty fold enhanced binding affinity towards the cofactor-containing complex and a three orders of magnitude improved catalytic efficiency⁶⁹. Although this explanation is supported by a mutational study for relevant residues, the described molecular clamp has never been crystallized. In contrast to this explanation, some positive synergy binders show potent inhibitory behavior rather than high SULT1A1 catalytic efficiency⁸⁴.

1.3.6 SULT1A1 inhibitors

The concept behind certain ligands acting as inhibitors for SULT1A1 is yet unclear. Lots of ligands have been identified as inhibitors all over last decade from natural sources^{110, 111}, environmental pollutants^{29, 112, 113} or even from drugs on the market^{84, 109}. The mechanism of inhibition of inhibitors was determined by biochemical experiments mostly to be (i) competitive or (ii) noncompetitive inhibition¹¹² as described below. In some cases the inhibitor is crystallized in the substrate binding site²⁹.

SULT competitive inhibitor competes for either (i) the cofactor or (ii) the substrate binding site. The first class includes inhibitors that often analogues of the cofactor^{114, 115, 116}. However, targeting the PSB binding site might inhibit all SULTs present in the whole body including cytosolic and membrane isoforms. The second class of competitive inhibition, takes place when a molecule competes for the acceptor binding site. The inhibitor might either undergo a sulfonation reaction with slower kinetics (lower K_{cat} and V_{max} values) or is not sulfonated at all. Ethinyl estradiol (EE), an E2 analogue, is a potent inhibitor that shows a lower V_{max} and positive synergy to SULT1A1⁸⁴. Other inhibitors are derivatives from confirmed substrates or can be metabolized to substrates. The binding features for this kind of inhibitors mimic those of substrates but sometimes lack the group that should undergo sulfonation¹⁰⁹. Another form of inhibition mechanism is the noncompetitive inhibition; in which inhibitor binds to an allosteric binding site that decreases the overall activity of the enzyme. Two separate, allosteric binding pockets are recently confirmed to be present on the surface of SULT1A1¹¹⁷.

Despite several SULT1A1 inhibitors from various categories documented in previous literature (Appendix Table A-2), little is known about the chemical features of strong inhibitors. To date, no inhibitor has been co-crystallized with SULT1A1 and their binding pose is not yet discovered. Furthermore, borders between substrates and inhibitors, and the relationship to positive synergy behavior remain extremely unclear. Inhibition of SULT1A1 causes the lowering in the detoxification capacity by decreasing the xenobiotic sulfonation rate. The inhibition is considered as well a reason for drug/drug or drug/hormone interactions, which can have clinically relevant consequences.

1.4 Previous computational studies

Due to the several recently solved SULT crystal structures, several valuable starting points for structure-based molecular modeling approaches have become available. More than two thirds of the total number of crystal structures have been resolved in the last 10

years (Table 2). Furthermore, this development has caused with considerable progress in *in silico* modelling approaches. Previous computational studies on SULT can be divided into two phases: i) an early phase mainly focusing on discovering Structure Activity Relationship (SAR) between different substrates and SULT isoforms often using ligand-based techniques, and ii) late phase mainly interested in finding mechanistic information for various SULT isoforms and its relationship to their unique chemical profile depending on combined structural and ligand-based procedures (Table 4).

Early SAR-focused studies have been based on QSAR models, either conventional 2D^{118, 119, 120, 121} or more advanced 3D models^{106, 122, 123}. Table 4 shows that the first 2D-QSAR study on phenol SULT from rat and human origin was reported in 1987¹¹⁸. The next study has been conducted more than 10 years later on human SULT1A3¹¹⁹. The major objective of both studies has been the identification of structural descriptors that affect substrate specificity and can be correlated to high K_m values. Using the same technique, the potency of SULT1E1 inhibition by polychlorinated biphenyls (PCBs) has been modeled¹²⁰. A similar study has been reported one year later combining physico-chemical descriptors together with experimental sulfonation data to find structure-conjugation relationships of various catechols¹²¹.

Because of crystal structure unavailability at that time, adding a third dimension (the protein structure) has been achieved by either homology modelling^{106, 123} or resolving new crystal structures¹²² (Table 4). Comparative Molecular Field Analysis (CoMFA) technique could then be applied for SULT1A1 and SULT2A3 from rat origin to construct 3D-QSAR models to correlate the catalytic efficiency ($\log(k_{cat}/K_m)$) of rat hepatic aryl sulfotransferase to its substrates. The homology models of the needed enzymes have been developed based on templates from existing crystal structures.^{106, 123} In 2003, the first SULT1A3 crystal structure was published and allowed the development of a 3D-QSAR model that predicts K_m values of phenolic compounds using CoMFA¹²¹.

Table 4: Summary of the previous early computer-based techniques literatures (until 2005) published on SULTs.

SULT	Year	Ref	Computations			objectives
			SAR		Homology modeling	
			2D-QSAR	3D-QSAR		
SULT1A1	1987	118				Quantitative SAR (QSAR) model to predict K_m values of phenolic substrates for SULT1A1 based on experimental data
SULT1A3	1999	119				QSAR model for to predict SULT1A3 enzyme specificity for 4-substituted phenols and catechols.
rSULT1A1	2002	106				Comparative Molecular Field Analysis (CoMFA) models to correlate the catalytic efficiency ($\log(k_{cat}/K_m)$) of rat hepatic aryl sulfotransferase to its substrates.
SULT1E1	2002	120				QSAR analysis of hydroxylated polychlorinated biphenyls (OH-PCBs) inhibition potency towards SULT1E1
SULT1A3	2003	122				CoMFA models to evaluate the K_m values for phenolic substrates of SULT1A3.
SULT1A3	2003	121				QSAR predictive models for the structure-conjugation relationships of catechols.
rSULT1A1, 2A3	2005	123				CoMFA predicting model for the catalytic efficiencies (k_{cat}/K_m) of two SULTs.

The second phase of computational work has started ten years ago (2006-now) with the rapid publication of several of crystal structures in the Protein Data Bank (PDB) and recent developments in *in silico* techniques (Table 5). Computational studies on SULTs in this period had two interests: i) prediction of SULT activity and/or ii) structural investigations¹²⁴.

Molecular Dynamics (MD) has helped to explain many experimental observations in the SULT field [46]. High accuracy *in silico* models have been developed by docking SULT1A1 and SULT2A substrates and inhibitors in MD generated protein conformations¹⁰⁹. Additionally, 3D pharmacophores generated from MD simulations have been employed for the first time to predict novel SULT1E1 substrates and inhibitors¹²⁵. Furthermore, MD-guided QSAR models have been generated for some SULT isoforms (SULT1A1, SULT1A3, and SULT1E1) and the chemical profile for each isoform has been defined¹²⁶. Also, MD-combined techniques have been used in previous literature to generate different conformations, to investigate the binding of some ligands^{84, 127, 128} and to demonstrate enzyme flexibility due to cofactor binding^{21, 56}. While the use of modern molecular modeling techniques like MD has filled a part of the information gap¹²⁹, its limits cannot be disregarded. MD is time consuming, computationally expensive, limited in sampling capacity and cannot be used in simulating bond breaking or formation since it is based on force fields using molecular mechanics.

In addition, docking has significantly contributed to finding plausible binding poses and evaluating favorable binding interactions of various substrates and inhibitors in SULTs^{8, 56, 130, 131}. SULTs have been combined with number of other enzymes in virtual screening studies to identify biological targets of clinical importance¹³², predict small molecules potential toxicity and side effects¹³³, and evaluate the endocrine-disrupting potencies of some brominated flame retardants¹³⁴. Over a quarter of a decade, plenty of publications have emerged addressing substrate binding modes^{8, 56, 130, 131}, defining chemical profiles for various SULT isoforms^{109, 126}, evaluating the effect of cofactor binding on substrate size^{39, 56 21}, and investigating the mechanistic characteristics of protein dimers⁴⁶.

Due to its significant role in modulating the activity of a multitude of compounds, the sulfotransferase progenitor SULT1A1 has been selected to be the focus of this study. Despite having two reported structure-based *in silico* predictive models^{109, 126}, the chemical profile for SULT1A1 substrates and inhibitors is not yet well defined and the positive cooperativity behavior is not fully explained. Moreover, empty substrate binding site was always modeled via MD to investigate the conformational flexibility of the sulfotransferase substrate pocket despite the known ligand-dependent induced flexibility. Further, only one ligand-based QSAR study has been reported¹¹⁸. In addition, number of docking studies have been conducted to predict the plausible binding pose of some relevant carcinogens⁸, substrates¹²⁸ and inhibitors⁸⁴.

Table 5: Summary of the most recent computer-based techniques literatures (2006- up to now) published on SULTs.

SULT	Year	Ref	Purpose of the study		Computational techniques				Objectives
			SULT activity	structural info	QSAR	docking	MD	others	
SULT1A1, 1A2, 1A3	2006	8							The plausible binding mode of N-hydroxy aromatic and heterocyclic amines.
SULT1E1	2007	135							Predictive QSAR for the endocrine disruptive potencies of the brominated flame retardants.
SULT2A1	2008	130							Plausible allosteric binding mode of celecoxib and nimesulide in SULT2A1
SULT2A1	2010	39							Cofactor-dependent structural rearrangement of SULT2A1
SULT2A1	2011	136							SAR model to evaluate the inhibition power of hydroxylated polychlorinated biphenyls for the dehydroepiandrosterone sulfonation
SULT1A3	2012	131							Regiospecific sulfonation of dietary flavonoids
SULT1A1	2012	84							The inhibitory binding pose of ethinyl estradiol (EE) in SULT1A1.
SULT2A1	2012	56							Cofactor-gated pore restricts large substrates
SULT1A1	2013	128							The plausible binding mode of halogenated organic substrates in SULT1A1.
SULT1A1, 2A1	2013	21							Structural investigation for nucleotide-induced gate formation that controls substrate size
SULT1A1, 1A3, 1E1	2013	126							<i>In silico</i> structure-based QSAR to predict ligand binding to various SULTs.
SULT1A1, 2A1	2013	109							Predicting <i>in silico</i> models that used to identify substrates and inhibitors from the DrugBank.
SULT2A1	2015	137							The effect of cofactor/ligand binding on the thermal stability and flexibility of the protein.
SULT1B1	2015	46							cofactor-dependent communication between monomeric subunits in dimeric SULT1B1
SULT2A1	2016	138							QSAR study using mono-fluorinated analogues of poly-halogenated biphenyl substrates
SULT1E1	2016	125							MD-guided pharmacophores for prediction of SULT1E1 substrates and inhibitors from the DrugBank.

2 Aim of work

Despite the fact that SULT1A1 shows the broadest sulfonation profile among all SULTs, it exerts exceedingly high catalytic efficiency towards a particular group of substrates: positive cooperativity substrates. This group of compounds exert higher binding affinity towards cofactor containing complexes than the open enzyme. Although several publications describe the co-crystallized SULT1A1 ligands, the intermolecular enzyme-ligand interactions responsible for the positive cooperativity behavior have not been well defined. Furthermore, some positive cooperative binders act as potent inhibitors rather than undergoing enhanced sulfonation. Therefore, borders between substrate and inhibitors and the relationship to positive cooperativity behavior remain unclear.

Research activities in this dissertation focus on the positive cooperativity binders of SULT1A1 including high affinity substrates and potent inhibitors. The main objective is to provide a better explanation to the positive cooperativity phenomenon and the key intermolecular interactions responsible for it. This understanding will be helpful to better describe the chemical features of the high affinity SULT1A1 binders (substrates and inhibitors). The study is conducted in the following steps:

1. Structural comparison:

This part was conducted to explore the structural differences between all available PDB crystal structures of SULT1A1 hosting variety of different size ligands including positive cooperativity binders. Furthermore, various SULT isoforms (SULT1A3 and SULT1E1 that share a substrate profile for catecholamine and estradiol, respectively with SULT1A1) were structurally investigated to define the major structural features that distinguish SULT enzyme isoforms. The objectives of the structural comparison are to explore the substrate inducible regions within SULT1A1 crystal structure, discover the characteristic structural features and pocket environment that enable the unique substrate profile for each isoform, and elements that enable broad SULT1A1 broad substrate specificity.

2. Pharmacophores development:

The main goal of this part of study was to identify the key chemical features for SULT1A1 protein binding, especially needed for the positive cooperativity behavior. 3D pharmacophore models can be used as discriminative filters to differentiate SULT1A1 binders (inhibitors and substrates) of diverse chemical scaffolds. In order to

achieve this goal, a group of pharmacophore filters were constructed. First, structure-based 3D pharmacophore models were derived from relevant crystal structures. Second, to increase diversity of information, high affinity positive cooperativity SULT1A1 inhibitors and substrates were docked into the SULT1A1 enzyme. Third, the collected information from potent inhibitors were used to generate ligand-based pharmacophores for SULT1A1 inhibitors. The specificity and accuracy for all pharmacophores were determined using the predictive power of the model to discriminate binders into substrates and inhibitors.

3. Molecular Dynamic study (MD):

The main aim of this part was to unveil the positive cooperativity binding mode for both substrates and inhibitors to offer insights towards relevant binding patterns for various binders. Also, the influence of the binding of one or more ligands (PAP/S and/or substrate or inhibitor) on the protein flexibility. In addition, investigating the enzyme dynamic behavior aimed to provide valuable information about the stabilization of “flexible regions” within various SULT1A1 complexes, by positive cooperativity binders. To reach these aims, key complexes in the sulfonation cycle of SULT1A1 for which no structural information is available were modeled. Then, dynamics of the different complexes involving SULT1A1, with or without cofactor, as well as high affinity positive cooperativity ligands (including substrates and inhibitors), were investigated using MD.

In summary, better understanding of SULT1A1 positive cooperativity behavior will contribute in the overall understanding of SULT specificity and selectivity, and therefore opens the door for better prediction of enzyme-drug interactions and metabolic patterns.

3 Methods

Computational *in silico* approaches are generally categorized depending on the employed data into structure-based and ligand-based¹³⁹. While structure-based approaches utilize the target structural data (homology, NMR, or crystallography) to generate computational models from target/ligand interaction information, ligand-based approaches employ solely ligand data to find similarities needed for activity. Although no superiority has been confirmed for an approach over the other¹⁴⁰, protein structural information is an asset that makes structure-based approaches more appealing. Moreover, combining both computational methods can be beneficial for studying enzyme-substrate complex¹²⁵.

This chapter gives a quick overview on the computational procedures conducted in this study: 3D-structure- and ligand-based pharmacophores, and molecular dynamics (MD) simulations.

3.1 3D Pharmacophores

Since its first emergence by Ehrlich in early 20th century, the term “pharmacophore” describes the spatial framework that holds the key features (phoros) essential for a drug (pharmacon) biological activity¹⁴¹. However, late in the same century a more concrete definition has been proposed by IUPAC; “*A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response*”¹⁴². Simply, a pharmacophore does not describe a functional group or chemical fragment but an abstract conception for the mandatory chemical features (such as hydrogen bonds, charges, aromatic rings and hydrophobic areas) for a ligand molecular interaction with a particular protein¹⁴³. This fact enables pharmacophores to recognize various actives having the same chemical functionality of diverse bioisosteric groups and contribute to scaffold hopping¹⁴⁴. In addition, 3D pharmacophore is a powerful and fast computational tool for virtual screening. Also, pharmacophores are understandable, modifiable versatile tools enabling researchers to convey their expertise in a clear and informative way¹⁴⁵. Due to all previously mentioned advantages, 3D pharmacophore modeling has been identified as a valuable method in drug discovery.

Describing the interaction between two entities (a protein and a chemical compound), a pharmacophore can be generated from the information of both of them (structure-based) or solely the latter (ligand-based). In case of a known target of crystallized structure, the structure-based pharmacophore is created by finding common interaction points between the protein and a bound molecule. If target information is missing (unknown target, or unavailable structure), a set of structurally related actives, assumed to bind to the same protein binding site and have the same binding mode, can be exploited to derive a pharmacophore, from common chemical functionalities believed to be mandatory for binding ¹⁴⁶.

To date, several pharmacophore elucidation software packages have been published in literature such as: LigandScout ^{147, 148}, Catalyst ^{149, 150}, Phase ^{151, 152}, and MOE ¹⁵³. In this study, LigandScout was used to generate both structure-based and ligand-based models of SULT1A1. For structure-based models, the software has the ability to retrieve the crystal structure from PDB, assign the proper hybridization states and bond types, and automatically generate the pharmacophore. For ligand-based models, LigandScout aligns the set of training ligands considering their conformational flexibility, in a way that enhances the accuracy and prediction power of the pharmacophore. Using the same software, large molecule libraries can be virtually screened through the generated pharmacophores using its pattern-matching algorithm to identify novel lead scaffolds ¹⁴⁷.

3.2 Molecular Dynamics (MD)

Molecular Dynamics (MD) simulations are a well-established *in silico* tool that plays a major role in drug design ¹⁵⁴. The procedure was initially developed by theoretical physicists in the last century and used for biological studies 15 years later ¹⁵⁵. Since then, MD has been used as an important explorative tool to investigate the dynamics, structure, and function of macromolecules ¹⁵⁶. Lots of problems have been tackled by MD starting from simple structure refinement, explanation of the experimental results up to the complex study of the possibility of some structural and conformational events (various protein-protein interactions) ^{154, 157}.

MD simulations deal with the molecular system as a group of particles (atoms) connected by springs (bonds). Applying Newton's equations of motion, MD produces a series of configurations as a function of time starting from an initial configuration. This computational method provides invaluable structural information about for the dynamic evolution of the protein showing movements and binding events. However, the precision

of a MD simulation relays on many factors including: i) accuracy of the used force field, ii) initial conformation used to simulate the actual macromolecule at given conditions (temperature, pressure, etc.), and iii) simulation time ¹⁵⁶.

To start an MD simulation, a starting protein structure should be available from imaging sources (magnetic resonance (NMR), crystallography), or homology modelling. This initial structure provides an initial source for particle coordinates in the protein system. To reproduce the actual dynamics of the protein structure, the forces acting on each element should be carefully parameterized. Force fields are used to predict inter- and intramolecular interactions in the protein system as well as to define the parameters set for each particle (atom types, chemical bonds, dihedral angles. etc.). Afterwards, Newton's laws of motion are applied to calculate the new coordinates of the atoms in the system with progression of time. Up to date, many available open-source and commercial MD software packages are available including GROMACS ¹⁵⁸, CHARMM ¹⁵⁹ and AMBER ¹⁶⁰. The latest software; AMBER was used to conduct all MD simulations presented in this study.

4 Results

Since this work aims at investigating and explaining the positive cooperativity behavior of SULT1A1 binders (substrates and inhibitors), three consecutive tasks were conducted starting with (i) structural investigation of SULT1A1, followed by (ii) the development of 3D pharmacophore models for SULT1A1 binders from available crystal structures and experimental data and finally (iii) an MD study of the key complexes of SULT1A1 in different combinations of cofactor and high affinity positive cooperativity ligands.

4.1 Structural investigations

The main goal of this part was to explore the flexible structure (ligand-inducible parts) of SULT1A1 that is the reason behind the broad enzyme specificity. All available SULT1A1 crystal structures, with and without ligands, were analyzed and differences were carefully elaborated. The SULT1A1 structure was compared to two other SULT isoforms, SULT1A3 and SULT1E1 in order to determine the structural reasons behind having distinct chemical profile different from SULT1A1.

Three SULT1A1 crystal structures were available in the PDB when this project started. One represents the SULT1A1*3 variant (PDB 1Z28¹⁸) and the other two represent SULT1A1*2 (PDB entries 1LS6¹⁵ and 2D06¹⁷) (Table 2). The substrate active site is empty in case of 1Z28, while it contains substrates of different sizes in the other two crystal structures 1LS6 and 2D06. Early after the start of this work, (in 2011), more crystal structures were resolved from two different groups (PDB IDs 3QVU, 3QVV¹⁹, and 3U3J, 3U3K, 3U3R, 3U3M, 3U3O²⁰) (Table 2). Most structures have been co-crystallized with PAP and a ligand, except 3U3R and 3U3J that do not have any ligand bound. Because crystal structures 3QVU and 3QVV represent mutated SULT1A1 with 8 different mutations, they were excluded from the structural comparison¹⁹. As well, 3U3R was excluded from the comparison as it displays the D249G mutant of SULT1A1 with altered structure, sulfonation activity and thermal stability²⁰. 3U3J shows the typical architecture of ligand-free PAP-containing SULT1A1. Other crystal structures are co-crystallized with high affinity substrates; 2NA (3U3K), 3-cyano-7-hydroxycoumarin (3CyC) (3U3M), or two molecules of 3CyC (3U3O). However, they are not resembling the wild type allylic variant but SULT1A1*2 allyl having His213 instead of Arg. During the year 2013, the only structure of SULT1A1*1 has been reported co-crystallized also with PAP and no ligands²¹.

The crystal structures of all mutant-free SULT1A1 in PDB, with and without ligands were compared by superimposition of the nine crystal structures (1LS6, 2D06, 1Z28, 3U3J, 3U3K, 3U3M, 3U3O, 3U3R, and 4GRA) (Table 2). If dimers were present in the unit cell of a crystal structure, only chain A was superimposed due to their structural similarity (applied to 2D06, 3U3J, 3U3K, and 4GRA). PAP binding mode and interactions were found identical in all structures with all stabilizing interactions towards Loop 3 donor segment (Figure 4). Except in 2D06, interesting crystal water is conserved in almost all SULT crystal structures, which might resemble the hydrolyzed sulfonate group. The crystal water forms a hydrogen bond between the hydroxyl groups of substrates and cofactor. At the same time, the hydroxyl groups of all substrates (except E2) are oriented towards the catalytic residue His108, with an average distance of 2.35 Å between the oxygen atoms of the substrates' hydroxyl groups and the nitrogen group of His108. In 2D06, E2 molecule invades deep in the protein structure resulting in the displacement of the crystal water and direct interaction between the hydroxyl group of E2 and the cofactor. The interaction with His108 is lost in 2D06 showing a non-catalytic binding mode (dead-end complex) (Figure 12).

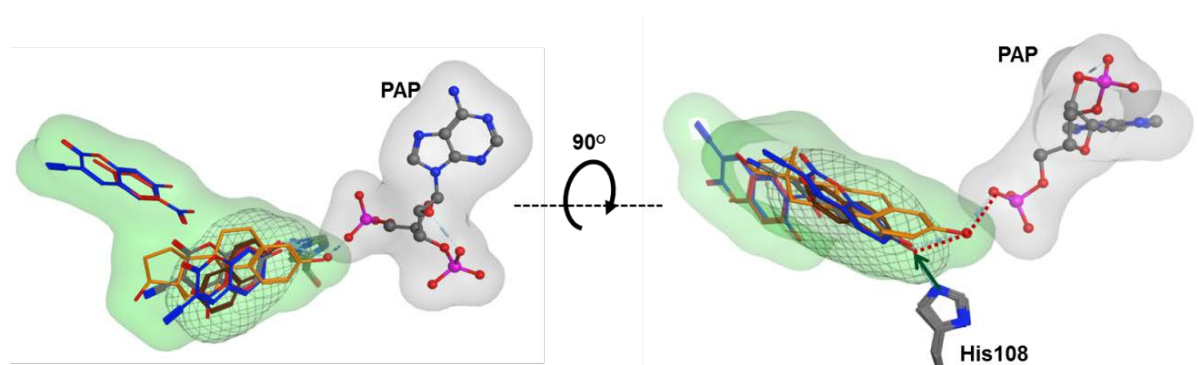


Figure 12: Superposition of the four co-crystallized ligands in the substrate cleft of SULT1A1. The conserved crystal water found in all structures complexed with PAP except 2D06, is depicted in red CPK. The PAP molecule from PDB entry 3U3K is shown in grey ball-and-stick model. The co-crystallized ligands are 2-naphthol (2NA) (brown), estradiol (E2) (orange), 2-nitrophenol (pNP) (red), 3-cyano-7-hydroxycoumarin (3CyC) from 3U3M (grey) and 2 molecules of 3CyC from 3U3O (blue). The grey space represents the PAP binding region whereas the green space shows the extended L shape acceptor cleft. The mesh space displays the actual catalytic binding pocket.

Based on this structural superimposition, all catalytically competent ligands are oriented in the same way, but occupy different areas depending on the ligand size (all ligands except E2 and the second molecules of pNP and 2CyC). The dimensions of the substrate binding pockets have been measured and compared in order to get more insights into the pocket flexible residues. The aromatic rings of substrates are stacked between the side chains of Phe81 and Phe142 that define Pocket I. The distance between the

aromatic rings of this residue pair is constant in all crystal structures (7.5 Å) (Figure 13). Similarly, for Pocket II that contains the second molecule, the distance between the aromatic rings of Phe84 and 76 was measured. The diameter of pocket II ranges from 6.8 to 8 Å. The diameter shows its lowest value when pocket II it is not filled (4GRA, 2D06, 3U3J, and 3U3K) (Figure 13), whereas it shows higher diameters when a second molecule is stacked between the residue pair (3U3R, 1LS6, and 3U3O). For channel I (Figure 13) that might represent the entrance for the first molecule, the range amidst Val148 and Phe274 was considered. The channel varies in diameter (6.1 - 7.2 Å) giving the highest diameter when a large binder occupies Pocket I. Moreover, for channel II (Figure 13) thought to be the gate for the second molecule, the distance between Ile89 and Phe76 was taken into account. This channel exhibits the maximum range of flexibility (5.8 - 8.5 Å) mainly increasing in crystal structures that contain two molecules of the substrate. In order to study SULT1A1 substrate binding pocket plasticity, the cavity volumes for all crystal structures were calculated and compared. The variability of the cavity volume among crystal structures confirmed the high plasticity of SULT1A1 enzyme (numerical values are provided in the appendix Table A-3). The substrate binding site volume showed great variability between crystal structures ranging from 500 - 900 Å³. The tightest pocket is an empty pocket or accommodating small binders (4GRA and 3U3K, respectively), while larger binders induce pocket expansion when adopted.

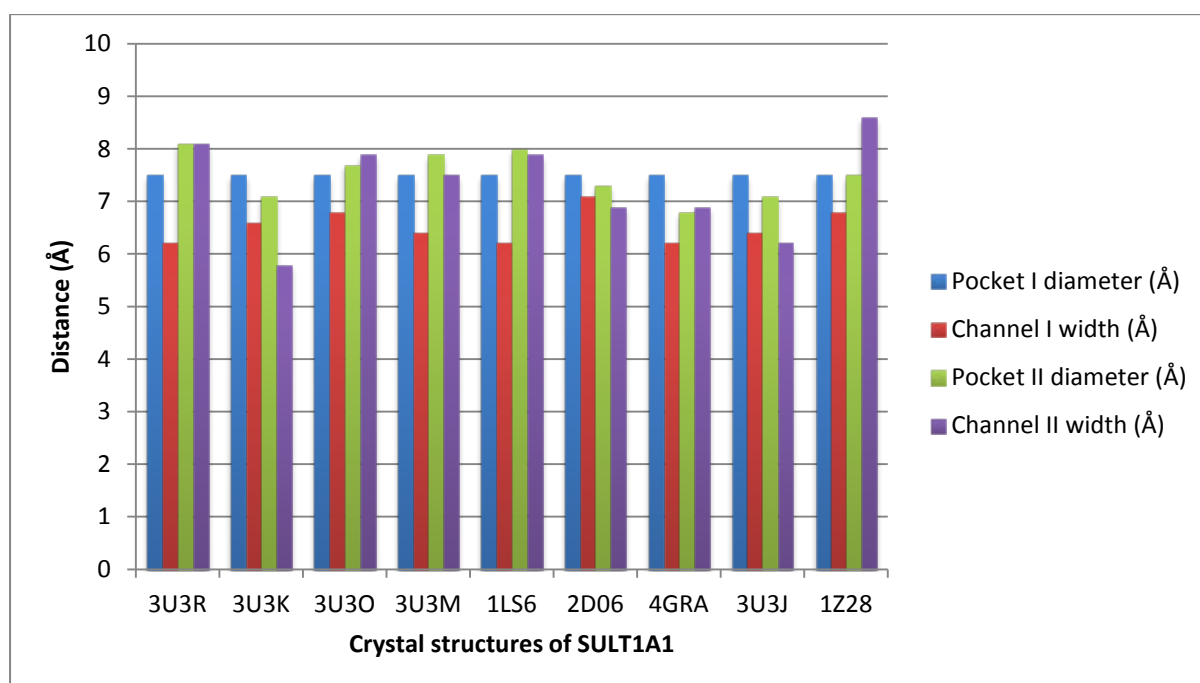


Figure 13: Comparison between the substrate pocket dimensions of SULT1A1 crystal structures.

Distances between catalytic residues, the cofactor and binders were carefully chosen to identify plausible hydrogen bonds between residues. These measured distances show a high degree of conservation between various crystal structures (except E2 in 2D06). As illustrated in Figure 14, hydroxyl groups of all bound ligands (except E2 in 2D06) form hydrogen bonds to the catalytic residues Lys106 and His108 with distances of 2.6 and 2.3 Å, respectively. The same hydroxyl group is located far from the 5' PO4 group of the cofactor (5.4 Å), preserving a space for the sulfonate group of the active sulfonated cofactor (PAPS). For 2D06, the orientation of the aromatic ring is 27° away from other planes of aromatic rings. The hydroxyl group is far from Lys106 (5.3 Å) and near to the 5' phosphate group of the cofactor (3.7 Å) and therefore catalytically non-competent for sulfonate group transfer. Its distance towards His108 is reasonable (3.2 Å). Our result is in agreement with previously published results that estradiol shows a non-catalytically competent binding mode, resulting in a substrate inhibition¹⁷. In all crystal structures, other catalytic residues (Lys48 and Ser138) show reasonable hydrogen bond distance (2.4 and 2.9 Å respectively). These residues stabilize the cofactor and show a key role for the sulfonation transfer reaction²⁸.

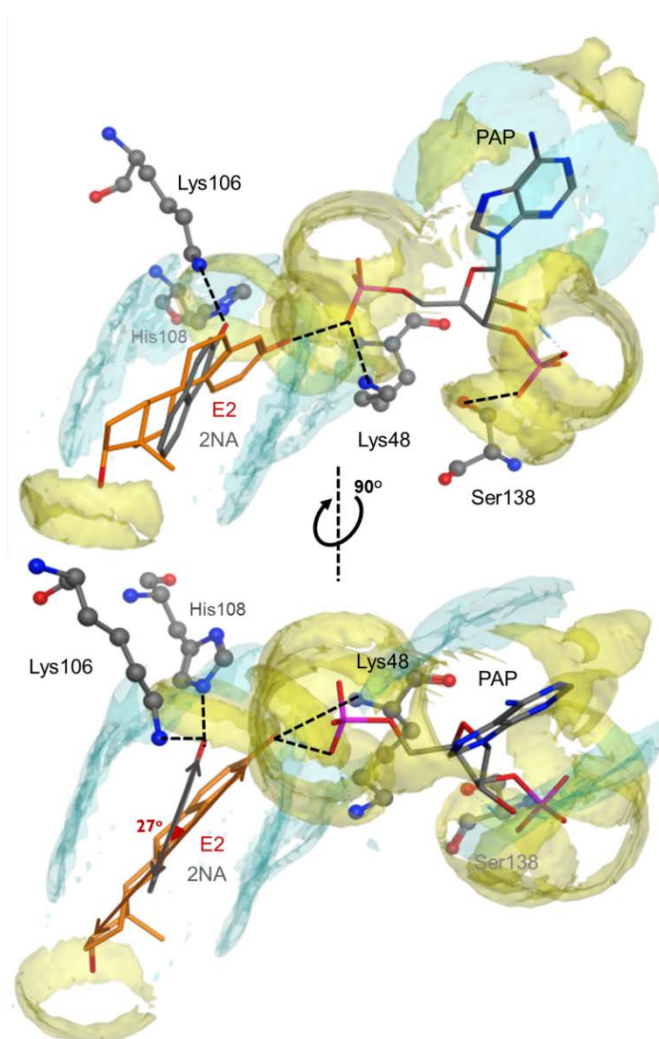
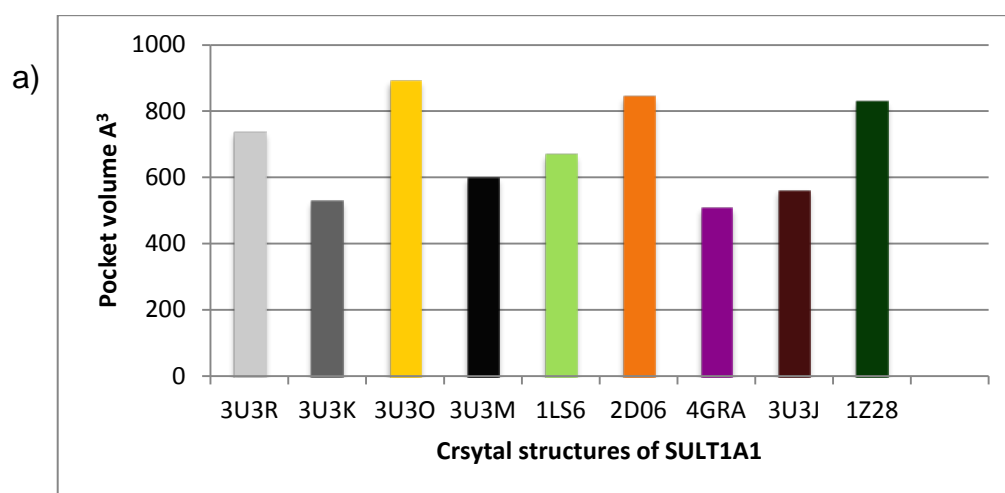
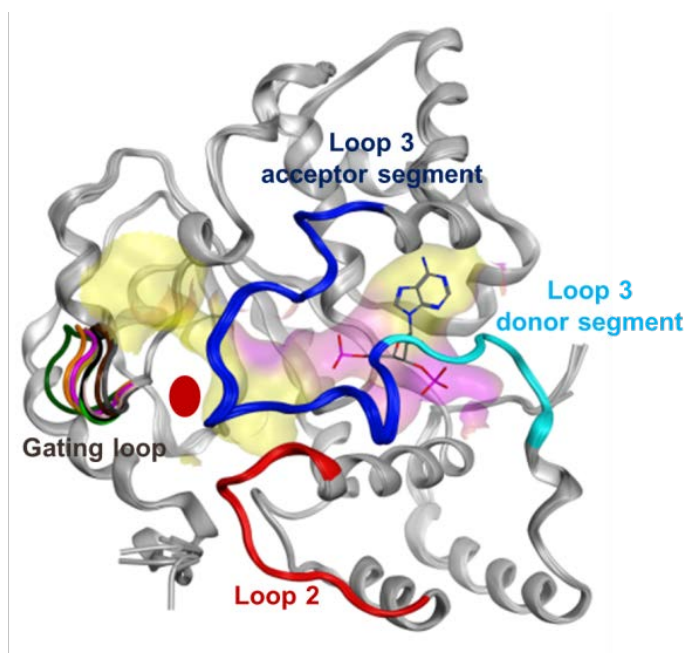


Figure 14: Catalytic distances between bound ligand (2NA in 3U3K), the catalytic residues (Lys48, Lys106, His108, and Ser138), and the co-crystallized cofactor (PAP). Yellow and blue spheres display the contact hydrophilic and lipophilic (respectively) preferences of the binding pocket.

In order to investigate the structural flexibility that led to the high degree of variation in substrate pocket volume, all available crystal structures were aligned. The volume of the hydrophobic cavity is surrounded by three loops; (Loop 1, Loop 2, and the acceptor segment of Loop 3). The average RMSD between the C α of SULT1A1 structures is rather small with a value of 0.365 Å. However, considerable alterations between different conformations can be observed for the gating loop (Figure 15-a). Substrate binding site volume calculations revealed a direct correlation between the gating loop withdrawal and the increase in the cavity volume (Figure 15-b), illustrating the plasticity of the cavity to adopt diverse binders. Similar results correlating gating loop position to the substrate cavity volume have been reported previously²⁰.



b)



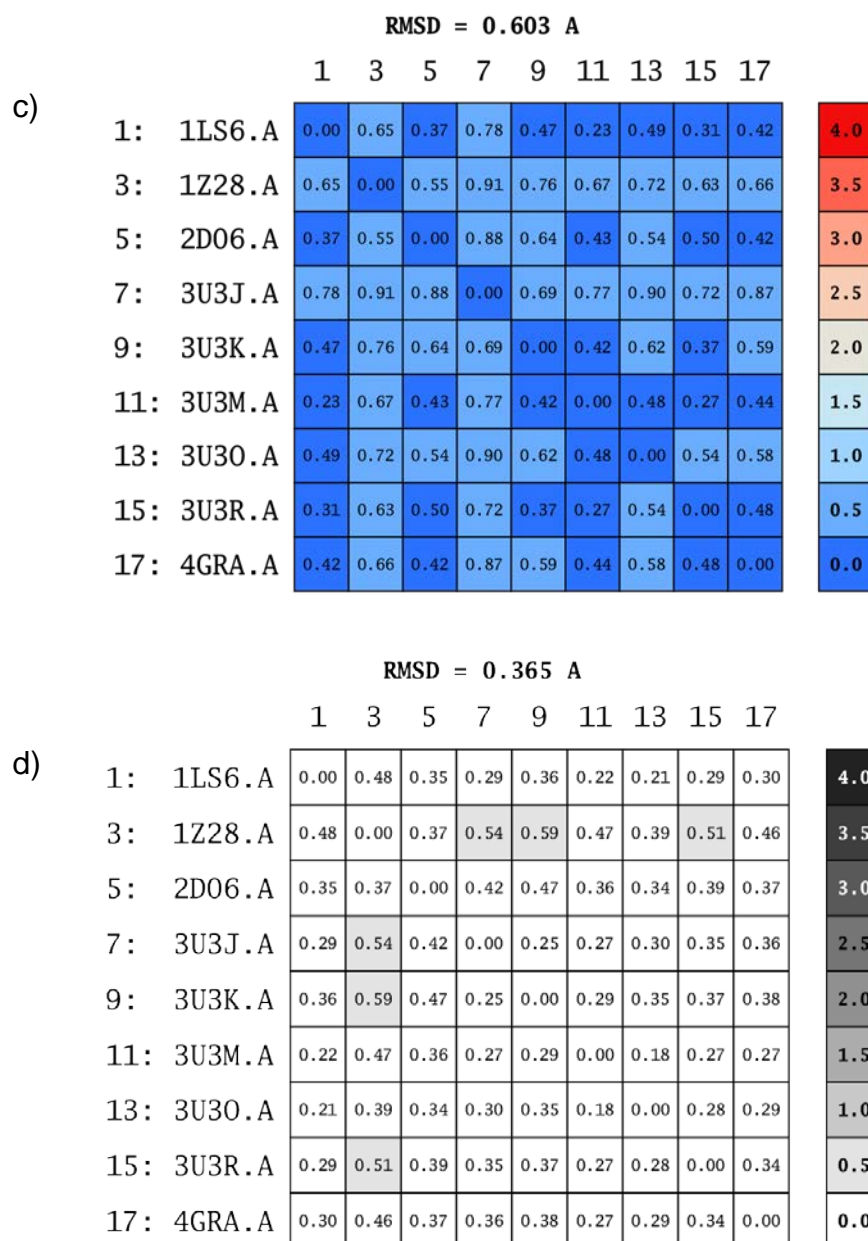


Figure 15: a) Comparison between SULT1A1 crystal structures substrate binding site volume. b) The alignment of all SULT1A1 crystal structures showing different gating loop positions. The red circle represents the substrate active site. The whole pocket is mapped according to its hydrophobicity from yellow (lipophilic) to red (hydrophilic). The colors of the gating loop are as follows: light green (1LS6), dark green (1Z28), orange (2D06), brown (3U3J), grey (3U3K), black (3U3M), yellow (3U3O), white (3UR), and violet (4GRA) (down). c) Pairwise RMSD plot for all atoms exist in the gating loop of all crystallized SULT1A1 protein. d) Pairwise RMSD plot for C α atoms of all crystallized SULT1A1 protein.

The secondary structure and pocket residues of SULT1A1 were compared to other highly related SULT isoforms with overlapped substrate specificity, SULT1A3 and SULT1E1. Initially, a multiple sequence alignment was performed on the three isoforms. Then, 3D structural superimpositions were performed in order to assess the pocket residues variations responsible for distinct substrate specificity of each isoform, and finally the nature of each substrate pocket was investigated to explore important chemical features

for proper binding, using Molecular Interaction Fields (MIF) ¹⁶¹ calculated for crystal structures 2D06, 2A3R, and 4JVL. Three main probes were used; "N:" to resemble the H-accepting groups, "OH2" that identifies hydrogen bond donor interactions and "DRY" that identifies hydrophobic areas.

The investigation of the residue composition in the three loops of SULT1A1 and SULT1A3 indicated few variations in amino acid sequences of all loops residues (the substrate binding site residues are underlined in the multiple sequence alignment detailed in the appendix Figure A-1). Having the highest degree of overall homology (global) and binding site similarity (local) to SULT1A1 ²⁵, the overall 3D structural of both enzymes (SULT1A1 and SULT1A3) is identical with RMSD value of 0.3 (illustrated in Figure 16). SULT1A3 binding site displayed only eight different amino acids; all are located in Loops 1, 2, and 3. These eight residues in the binding site vicinity are responsible for creating an electronic environment different from SULT1A1. MIF charts (Figure 17) display a highly hydrophobic SULT1A1 pocket, which is formed by hydrophobic side chains of residues Phe81, 142, Ala146, Ile89 and Val148. On the other hand, the MIF map for SULT1A3 pocket shows a more hydrophilic nature due to high proportion of acidic and basic residues.

Despite being from the same family, SULT1E1 shows higher global sequence variations in comparison to SULT1A3. In addition, most of the pocket residues are different from the other isoforms (the substrate binding site residues are underlined in the multiple sequence alignment detailed in the appendix Figure A-1). These sequence differences affect the secondary structure of the enzyme showing higher RMSD value of C α -atoms (1.1 Å) between the SULT1E1 and the two other isoforms (Figure 16). Not only for the conformation of the three above-mentioned loops (Loop 1, Loop 2, and Loop 3), but SULT1E1 has two other regions (between Loop 2 and Loop 3, shown in Figure 16) that also display sequence variance and different arrangements in comparison to SULT1A1 and SULT1A3. As illustrated in Figure 16, the regions of altered average RMSD highlight the participation of all loops in structural difference among crystal structures. MIF map of SULT1E1 (Figure 17) displayed a hydrophilic environment resulted from the presence of hydrophilic residues that helps to give more insights into the substrate interaction pattern of its unique substrate binding profile.

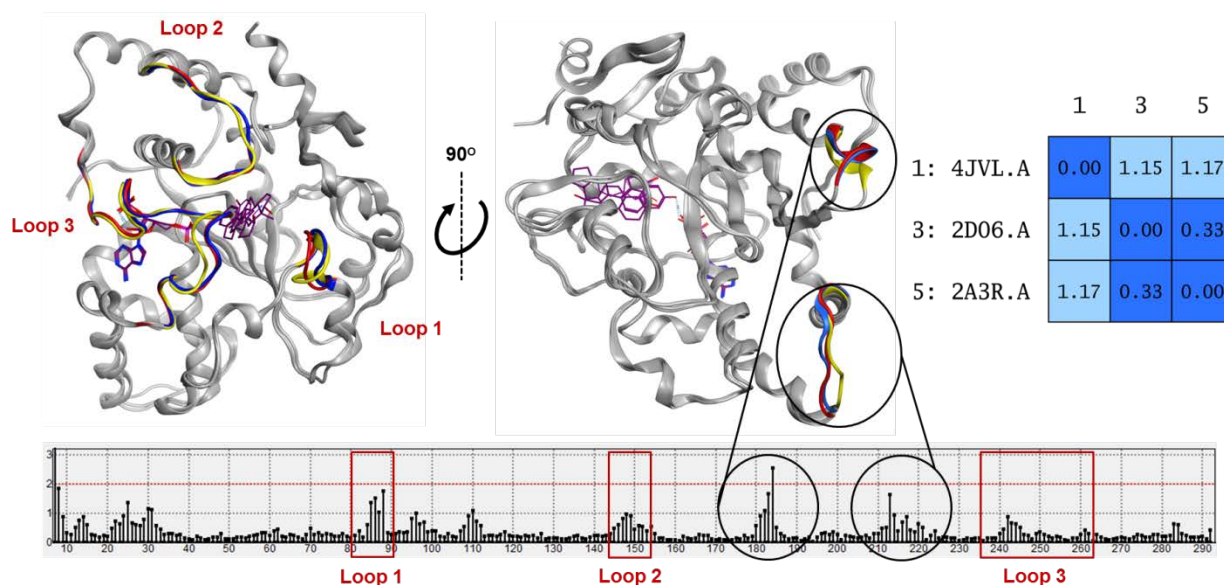


Figure 16: Superimposed crystal structures of the three SULT isoforms; SULT1A1 (2D06), SULT1A3 (2A3R), and SULT1E1 (4JVL). Pairwise RMSD values (in Å) for C α -atoms of the three SULTs are indicated in blue matrix. Differences in the SULT1A1 and SULT1E1 enzyme backbones are shown on the averaged RMSD (Å) scale versus amino acid numbering. Part of the protein backbones of SULT1A1, SULT1A3, and SULT1E1 are indicated in blue, red and yellow, respectively.

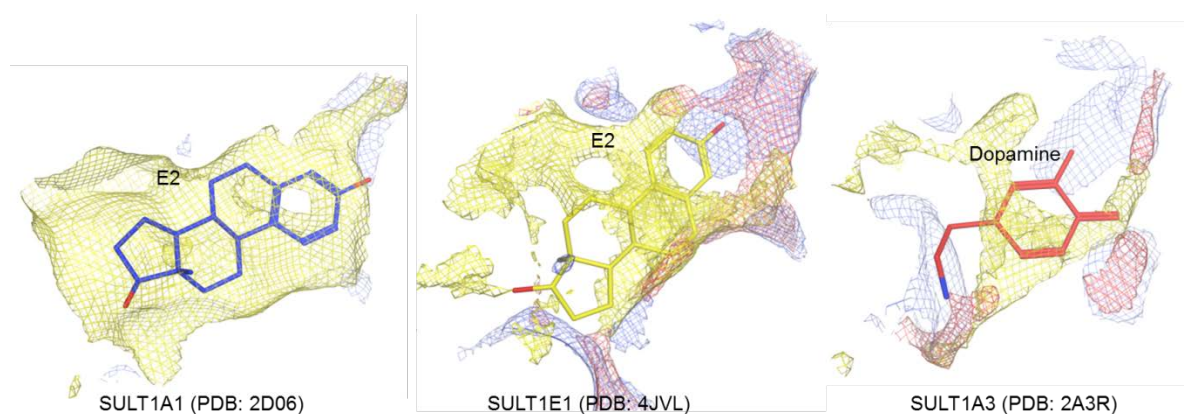


Figure 17: Molecular Interaction Fields (MIF) of substrate pocket of SULT1A1, SULT1E1, and SULT1A3. The molecular interaction fields are shown as grids showing H-bond acceptors (blue), H-bond donors (red), hydrophobic interactions (yellow).

In summary, findings of the structural investigations across SULT1A1 crystal structures revealed high degree of the enzyme binding site plasticity. The size of the crystallized ligand directly affects the dimensions of the substrate binding pockets through affecting different conformations of the gating loop (Loop 1). The calculated substrate binding site volumes of various SULT1A1 crystal structures adopting diverse binders clearly demonstrated a direct correlation between the gating loop withdrawal and the increase in the cavity volume. Results of the structural investigations of SULT1A1 and other highly related SULT isoforms with overlapped substrate specificity, represented binding site

sequence and structural homology of high (SULT1A3) and low (SULT1E1) degree. However, both siblings showed variations in the substrate pocket residues that are responsible for creating the electronic environment of the substrate cleft which reflects the unique substrate specificity of each isoform. Investigation of SULT1A1 structures and other highly related siblings support the contribution of SULT1A1 pocket characteristics (ligand-induced size, ligand-induced geometry, and hydrophobic nature) in defining its broad substrate specificity.

4.2 Pharmacophores

In order to elucidate the key features for SULT1A1 binding, 3D pharmacophores were generated from all available data (structure-based and ligand-based) for various binders. All ligand-containing crystal structures were utilized as templates for the 3D structure-based pharmacophore generation in order to account for enzyme flexibility and conformational alteration. At the same time, a binder database of high affinity SULT1A1 binders was generated including substrates and inhibitors. This dataset was separated into a training set used to generate ligand-based 3D models and test set utilized to validate the performance of all pharmacophores. More structure-based pharmacophores were produced from the docking pose of high affinity non-crystallized binders of both kinds of activity (substrates and inhibitors). The development of the 3D pharmacophores and assessment of their performance are explained in more detail in the following part.

4.2.1 Generation of active and decoy datasets

The dataset containing high affinity substrates for SULT1A1 was created by literature search. The collected data were ambiguous, generated in several labs, utilizing different biochemical setups (assay conditions and experimental setup), and measuring various biochemical markers. In order to categorize high affinity substrates, several biochemical parameters were considered. All substrates exhibiting Michaelis-Menten constant (K_m) value $\leq 5 \mu\text{M}$ were considered. In case of multiple data for the same substrate, a substance was considered as a substrate if at least one K_m value $\leq 5 \mu\text{M}$. In addition, all positive synergy substrates showing very low binding affinity (K_d) towards binary complexes were counted in the substrate list.

In addition to K_m and K_d , other biochemical values were considered for enriching the substrate dataset, wherever available. Substrates with high turnover numbers (k_{cat}) values (more than 50 sec^{-1}) were included in the substrate list. Since k_{cat} is not a common value in

SULTs biochemical analysis, V_{\max} was taken into account because it is directly related to k_{cat} ¹⁶². Substrates that display high V_{\max} ($> 100 \text{ nM min}^{-1} \text{ mg}^{-1}$) were inserted into the list. Another useful biochemical descriptor is the catalytic efficiency value (k_{cat}/K_m). Yet, the substrate list included substrates with reported high SULT1A1 catalytic efficiency values (greater than $10^3 \text{ M}^{-1} \text{ sec}^{-1}$).

In summary, all substrates showing at least one of the following biochemical descriptors were included in the high affinity substrate dataset: low K_m , low K_d , high k_{cat} , high V_{\max} , and/or high k_{cat}/K_m . The complete list of all included substrates (35 high affinity substrates) together with their reported biochemical data is given in the appendix Table A-1.

For the inhibitors list, the biochemical parameters which were considered were; K_i and IC_{50} . All inhibitors that show low K_i and/or low IC_{50} (less than $1\mu\text{M}$) were included in the list, whether they show sulfonation activity or not. The sulfonation activity data for some inhibitors confirms their competitive inhibition through tight binding into the substrate binding site. However, some inhibitors compete for the substrate pocket without having a sulfonation activity due to the absence of the functional group liable to sulfonation. This class of inhibitors is represented by substrate analogues¹⁰⁹ and flavonoids that do not have a hydroxyl group at the liable sulfonation position (7'OH)¹⁶³. On the other hand, non-competitive inhibitors are also identified for SULT. Catechins and nonsteroidal anti-inflammatory drugs (NSAIDs) are confirmed to bind in an allosteric binding site other than the substrate binding site through which they exert their inhibition¹¹⁷. The inhibitors list (26 pure inhibitors and 22 mixed inhibitor substrates); together with their reported biochemical inhibition data and the sulfonation activity are provided in the appendix Table A-2.

For the substrate decoy list, random selection of compounds with substrate similarities (structural and physicochemical properties) was carried from ZINC database¹⁶⁴. A decoy inhibitor dataset was generated by the same procedure using the inhibitor dataset. In addition, confirmed high affinity substrates for other SULT isoforms (SULT1A3, SULT1E1, and SULT2A1) were included in the decoy list (appendix Table A-4). Finally, inhibitors that do not show any sulfonation activity (pure inhibitors) were added to the SULT1A1 decoy list only for the validation of the substrate pharmacophores.

4.2.2 Substrates 3D pharmacophores

In all available SULT1A1 crystal structures, the cofactor PAP is co-crystallized in the same cavity²⁰. Yet, no crystal structure of SULT1A1 was published with the active sulfate donor PAPS^{15, 20}. It is substantial to model the Michaelis-Menten complex (SULT1A1-PAPS combined with a substrate) to illustrate the reactive state prior to its sulfonation. Studying such a complex would help in recognizing the catalytically competent binding mode, identifying the cofactor's role in substrate binding, and give more insights to SULT1A1 mechanism of sulfonation. PAPS was previously reported to bind with exactly the same mode and conformation as PAP to human SULT1E1²⁸, Therefore, the first step in this process was to dock the cofactor PAPS in SULT1A1's PAP binding site (PSB) to generate Michaelis-Menten complexes.

To take into account the low selectivity and flexibility of SULT1A1 substrate pocket, all SULT1A1-PAP crystal structures that were co-crystallized with substrates (PDBs: 1LS6¹⁵, 3U3M, 3U3O and 3U3K²⁰) were included in pharmacophore generation. As confirmed by the previous structural investigation in section 4.1, these crystal structures show diverse binding site conformations (size and geometry), all to be considered to cover the binding features of different types of acceptors. Because of its non-catalytically competent binding pose, E2 (PDB = 2D06¹⁷) was excluded from this study and used to generate a structure-based model by a different technique that is described later in section 4.2.3.1.

First, PAP molecules were extracted from the four crystal structures, leaving an empty cofactor binding pocket ready for PAPS docking. The docking pose for each crystal structure with the lowest RMSD to the co-crystallized PAP was selected. Adopting this strategy, interactions stabilizing the PAP in the crystal structures were recovered by PAPS when docked in the PSB (Figure 18). The sulfate moiety in SULT-PAPS generated structures is oriented towards Lys48 and Lys106, and stabilized by key interactions with these residues that are believed to play pivotal role in sulfonate transfer reaction²⁸.

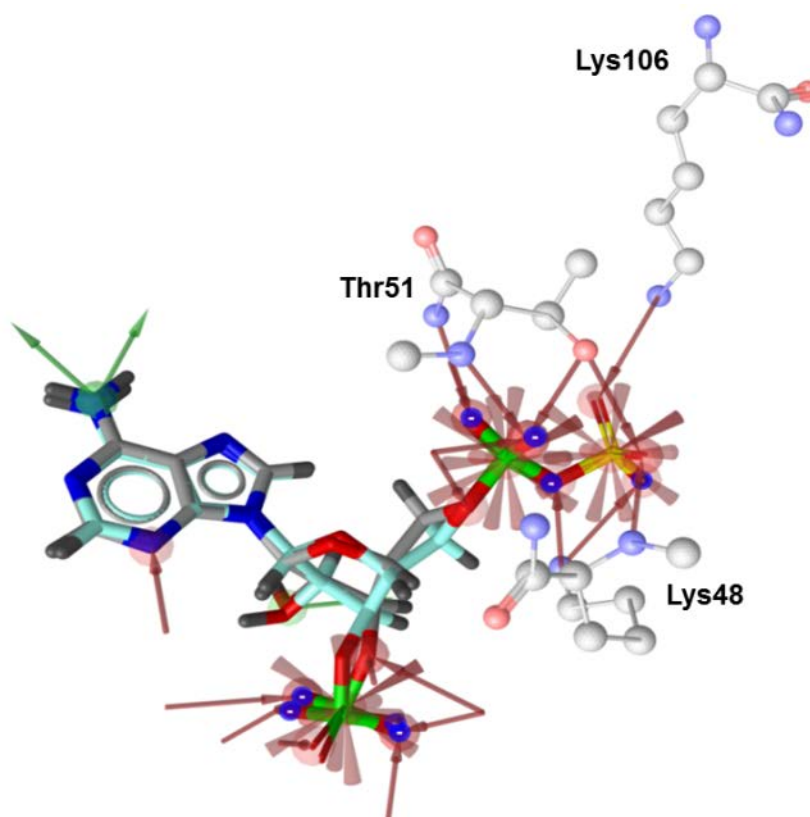


Figure 18: The most plausible PAPS binding pose (cyan) compared to the binding pose of the original cofactor (PAP) crystalized in 3U3K (grey). Sulfonate group contributes three extra hydrogen bonds towards protein residues indicated in ball and stick.

To reveal the plausible binding mode of substrates in sulfonated active enzymes, each SULT1A1-PAPS docking result was superimposed on its corresponding SULT1A1-PAP crystal structure (Figure 19). The distance between sulfonate group of the cofactor PAPS and the hydroxyl group of the co-crystallized ligand was measured to ensure its proximity for sulfonate transfer reaction. In all generated docking poses, the sulfur atom was located at about 3 Å from the acceptor hydroxyl group, which is consistent with the in-line transfer reaction mechanism^{48, 106}.

For the preparation of the final ternary complex from PDB crystal structure 2D06, PAPS was docked as described for other structures. However, when the docking result was superimposed on the 2D06-PAP crystal structure (Figure 20), the distance between the sulfur atom of PAPS and the ligand's acceptor group (3-phenol of estradiol, E2) was too short (1.5 Å) to be catalytically competent for sulfonate group transfer. In addition, the O-3 hydroxyl group was not positioned correctly for catalysis with respect to His108. Therefore, this structure cannot be considered the Michaelis-Menten complex for E2 in SULT1A1.

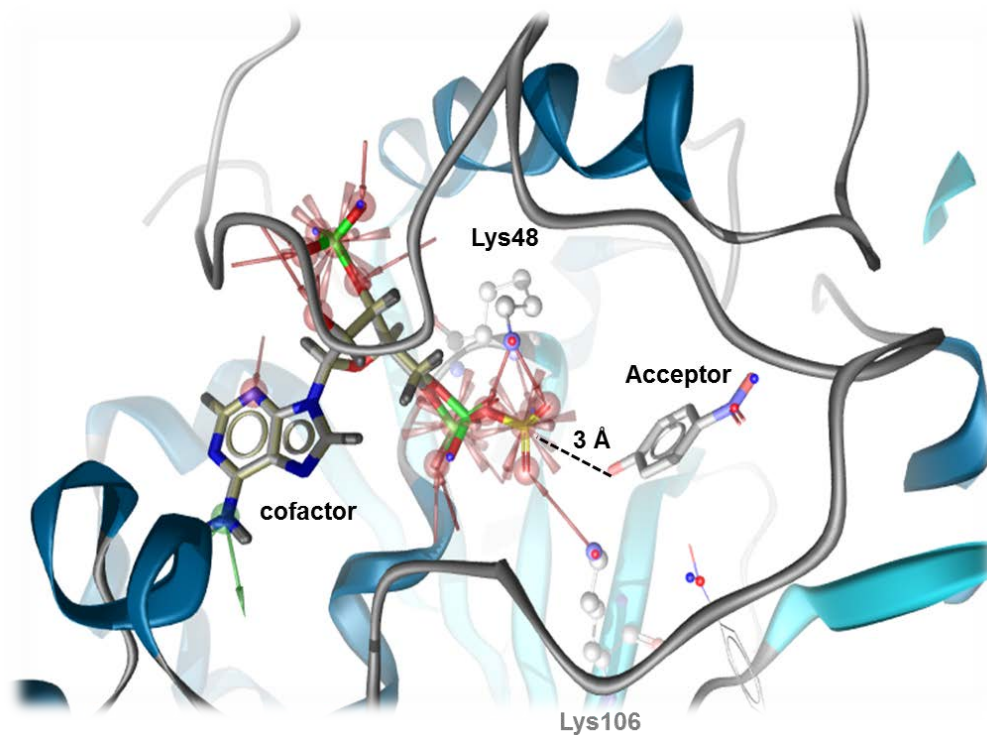


Figure 19: Michaelis-Menten complex: Plausible PAPS binding mode in PSB of 1LS6¹⁵ shows good interaction between the sulfonate moiety of PAPS and the catalytic amino acids Lys48 and Lys106 (red hydrogen bond acceptor) and in a close proximity (3 Å) of the hydroxyl group of the acceptor to undergo sulfonate transfer reaction.

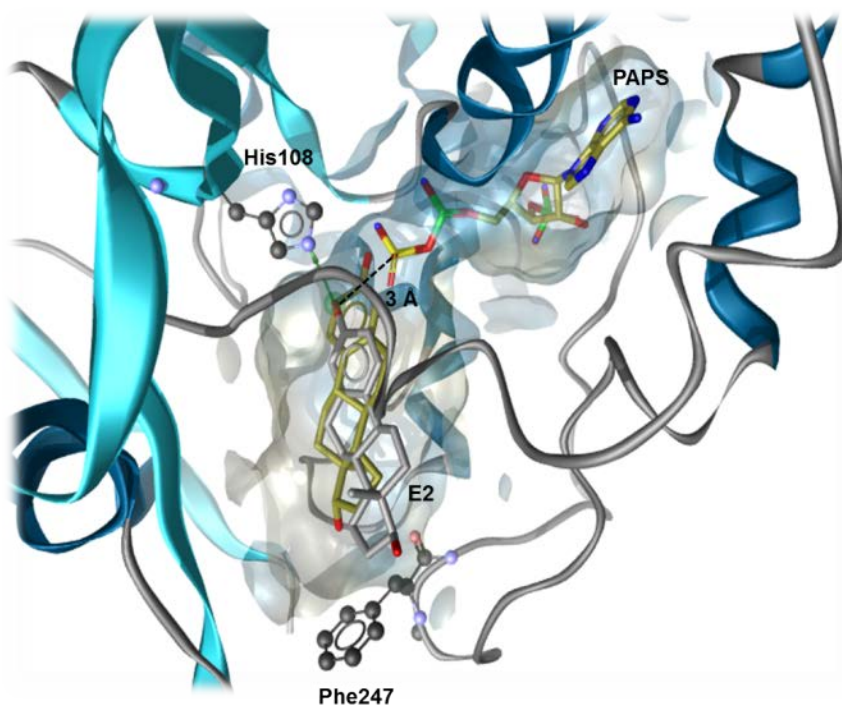


Figure 20: Michaelis-Menten complex of E2: Plausible E2 binding mode (silver) in 2D06-PAPS superimposed with its original binding position (gold) in 2D06¹⁷. The hydroxyl group of docked E2 shows hydrogen donating bond towards His108 (green arrow) and proper distance to the sulfonate moiety of PAPS to undergo sulfonate transfer reaction.

Being a substrate, E2 is believed to bind to SULT1A1, in the presence of PAPS, in a productive mode that allows transfer of sulfonate group¹⁷. To manifest the Michaelis-Menten complex for the substrate E2 in a sulfonated active enzyme prior to sulfonate transfer reaction, E2 was docked into the modeled sulfonated enzyme (2D06-PAPS). The most plausible pose was chosen to be in a reasonable interactive distance ($\approx 3 \text{ \AA}$) from PAPS sulfonate group and catalytic His108. The docked E2 (Figure 20) had a binding mode similar to that observed in the SULT1E1 crystal structure (both were 2.9-3.0 \AA from the catalytic His108)¹⁶⁵.

In order to give more diversity to the pharmacophores, one of the non-crystallized SULT1A1 substrates was docked into the protein and the substrate Michaelis-Menten complex was used to generate a docking-based pharmacophore. The criterion for the selection of this substrate was to be structurally different from the other small substrates already resolved in SULT1A1 crystal structures. The endogenous substrate 3,3-diiodothyronine (T2) is a low K_m SULT1A1 substrate^{100, 166}. One of the benefits of docking T2 was to identify the binding mode of flexible multi-ring substrates in SULT1A1. In order to account for the enzyme substrate pocket plasticity, T2 was docked into all PAPS-containing models simultaneously via ensemble docking. Its Michaelis-Menten complex of T2 was at its least energy in 3U3M, due to wide pocket geometry compatible with the substrate large size. The two phenyl rings of T2 were adopted in the L-shaped pocket via L-shaped conformation.

After the generation of Michaelis-Menten complexes for various substrates, 3D structure-based pharmacophores were built from the 2NA, 2CyC, p-NP, E2 (docked), and T2 (docked) in 3U3K, 3U3M, 1LS6, 2D06 and 3U3M, respectively. Key substrate binding features of different substrates were investigated, and optimized using our previous knowledge about SULT1A1. The generated pharmacophores are represented in Figure 21. Finally, these pharmacophores were simultaneously validated to set a model that could demonstrate the essential interactions and retrieve other substrates.

The structure-based pharmacophores generated by LigandScout¹⁴⁷ for the substrates in 2D06 describe the key interactions for ligand binding and the geometric shape of the cavity. The structure-based pharmacophore for p-NP in 1LS6 (SSB_PH4_pNP) consists of five chemical features showing two groups of dependent features. First, the hydroxyl group of p-NP translated three dependent features; H-bond donor (HBD) to the nitrogen group of His108, H-bond acceptor (HBA) towards the sulfonate group of PAPS, and another H-bond acceptor (HBA) to the terminal amino group of Lys106. Second, the aromatic ring

was interpreted to two dependent features; an aromatic feature (Ar) for the ring and a hydrophobic site (HYD). Interestingly, the nitro group was supported by a hydrogen accepting interaction to a crystal water in the hydrophobic pocket. This crystal water is not present in other crystal structures. Furthermore, the structure-based pharmacophore for p-NP in 1LS6 (SSB_PH4_pNP) consists of five features showing two groups of dependent features. The pharmacophore for 2CyC in 3U3M (SSB_PH4_2CyC) exhibited similar interactions as pNP. The other three pharmacophores for 2NA, E2, and T2 (SSB_PH4_2NA, SSB_PH4_E2, and SSB_PH4_T2, respectively), display an additional hydrophobic feature for the extended structure in the lipophilic substrate cleft (Figure 21). A complete exclusion volume coat was generated for each substrate to draw boundaries that determine the capacity of the binding site.

Pharmacophores of substrates of different sizes suggested a better substrate pocket binding (the increase in pharmacophore fit score) for the extended large molecules such as T2 and E2. In reality, this hypothesis proved to be wrong. SULT1A1 sulfonates small substrates efficiently with low K_m in comparison to larger substrates⁶⁹. Although the binding of large substrates in SULT1A1 is well accommodated by hydrophobic interactions, it was not translated to enhanced sulfonation activity. As suggested by a previous study, binding of an extended multi-ring substrate needs the expansion of pocket size and a conformational rearrangement to the gating loop which is energetically expensive¹⁵. This hypothesis guarantees the high sulfonation activity of SULT1A1 towards smaller substrates. More studies are needed to better understand substrate binding and its relationship to sulfonation activity.

In summary, the generated pharmacophores share some features. First, the repeated aromatic feature (Ar) in all substrate pharmacophores ensures the importance of aromaticity in SULT1A1 binding. Second, The (HYD) displays a gate-like restriction between Phe142 and Phe81 that identifies only planar compounds to be hosted between these two residues. Third, the hydrogen donating feature towards His108 and hydrogen accepting to PAPS resembles the catalytically-competent binding mode pivotal for substrates. For large substrates that contain more than one aromatic ring or extended hydrophobic system, an extra (HYD) is located close to Ala146, Met248 and Val148, resembling another site for hydrophobic interaction. Finally, the generated pharmacophores were simultaneously validated. The ability of pharmacophores to retrieve 35 actives from more than 360 decoys was examined. The model performance was assessed by the use of a receiver operating characteristic (ROC) curve. The curve evaluates the discriminative power of a model by plotting the rate of true positives, or the

sensitivity, against the rate of false positives, calculated as $(1 - \text{specificity})$ into a graphical presentation. The ROC shows a model with high selectivity and specificity that can retrieve about 90% of the actives with no false positives.

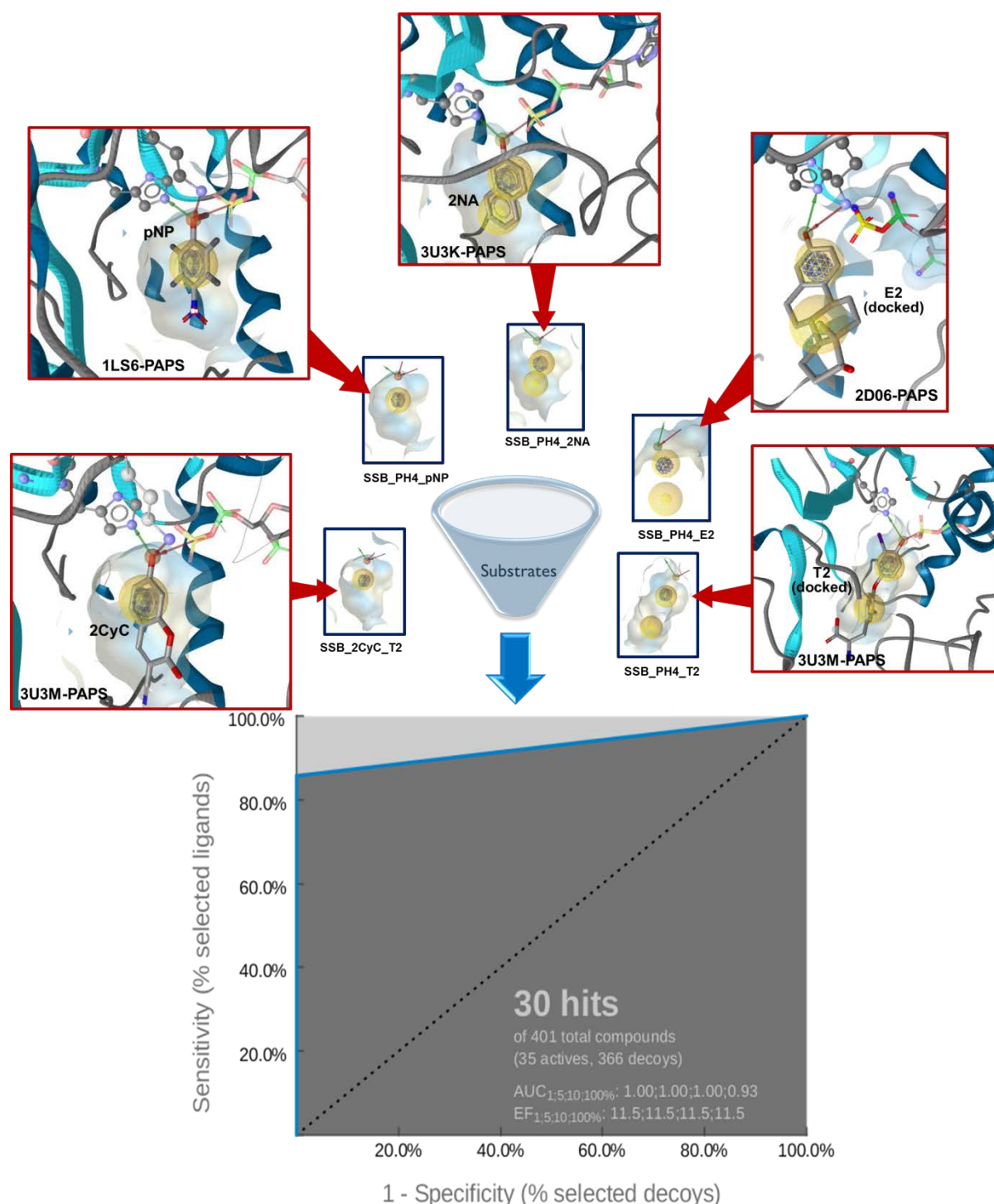


Figure 21: Structure- and Docking-based substrates pharmacophores. The five red boxes represent the five generated Michaelis-Menten complexes of 2CyC, pNP, 2NA, E2 (docked) and T2, docked in 3U3M-PAPS, 1LS6-PAPS, 3U3K-PAPS, 2D06-PAPS, and 3U3M-PAPS, respectively. The red arrows resemble the pharmacophore generation for the corresponding five pharmacophores in red boxes; SSB_PH4_2CyC, SSB_PH4_pNP, SSB_PH4_2NA, SSB_PH4_E2, and SSB_PH4_T2. The blue arrow shows the validation process that is represented by the Receiver Operating Curve (ROC).

4.2.3 Inhibitors 3D pharmacophores

4.2.3.1 Structure-based inhibitors pharmacophores

E2 was crystallized in a mode completely different from the binding mode of other crystallized substrates¹⁷. This mode has been described in literature as a non-productive position that shows strong substrate inhibition¹⁷. Furthermore, one estradiol derivative, namely ethinyl estradiol (EE) has been recently reported as potent SULT1A1 inhibitor⁸⁴. It is informative to know how bad substrates act as potent inhibitors. Because of its structural similarity to EE, E2 binding pose in SULT1A1, described earlier, was considered as inhibitor binding mode for the 3D design of structure-based inhibitor pharmacophore. Also, EE was docked in SULT1A1 to investigate its plausible binding conformation responsible for the superior inhibition.

The MIF data illustrated in Figure 17 using the binding mode of E2 in 2D06 reveals a highly hydrophobic pocket. The aliphatic hydroxyl group of E2 is surrounded with hydrophobic side chains (Val243, Val148, and Phe247). The lipophilic environment may disfavor E2 hydroxyl group binding and lead to a destabilizing interaction that decrease its binding affinity and therefore its inhibitory potency. Designing a better inhibitor could involve removal of the hydrophilic function group or adding an extra hydrophobic group that interacts and binds better with the lipophilic binding cleft. This pocket nature might explain the superior inhibitor effect of ethinyl estradiol (EE) over E2⁸⁴, having an acetylene additional group.

To explore E2 structural modifications for better SULT1A1 binding, SZMAP software was used¹⁶⁷. SZMAP is a binding site mapping tool that uses a probe of water molecule, and provides insights on critical features of a binding site. It uses semi-continuum solvation¹⁶⁸ to predict how and where neighboring waters can impact binding of a ligand, and by using related software (Gameplan)¹⁶⁹, it suggests ligand modification hypotheses designed to better exploit specific regions on a binding site. SZMAP maps highlight hydrophilic sites in the binding cleft where the water molecule is tightly bound and its structural orientations, as well as hydrophobic zones where crystal water is loosely bond and disordered. SZMAP analysis for E2 in 2D06 (Figure 22-A) showed a similar map to MIF map in the distribution of hydrophobic (purple) and hydrophilic regions (gold). A hydrophobic zone was highlighted close to the aliphatic hydroxyl group of E2 to show that it is not favored and surrounded with loosely disordered water molecules. The map explained that a water molecule at the aliphatic hydroxyl group (that can resemble protein interaction with E2

aliphatic hydroxyl group) has a destabilizing effect against E2 binding to SULT1A1 pocket. GAMEPLAN[®] analyzed SZMAP[®] results and suggested modifications for better binding. The proposed change was to replace the polar hydrophilic OH group with a hydrophobic one to increase binding affinity to the enzyme (Figure 22-B). Obviously, adding an acetylene group at the aliphatic site (in case of EE2) increases the hydrophobic interaction in the binding cleft⁸⁴. Again, that improved our understanding of the superior inhibitory potency of ethinyl estradiol (EE) over E2.

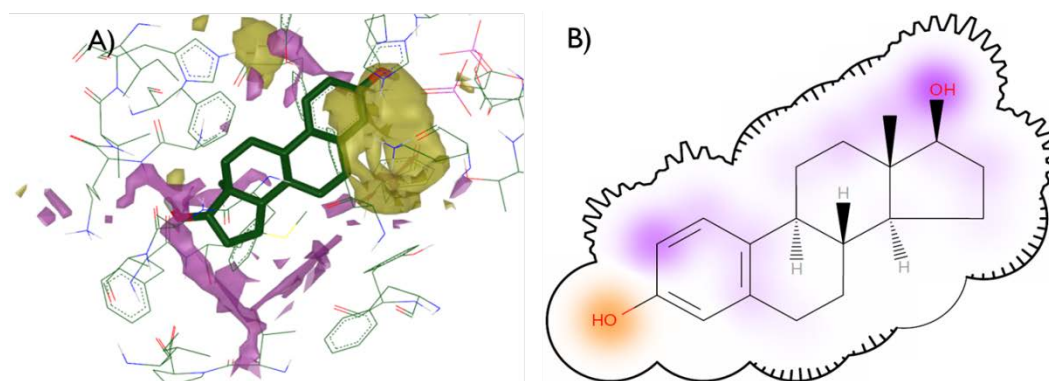


Figure 22: 2D06 binding site mapping and lead optimization of E2: A) SZMAP[®] map for 2D06 showing hydrophobic (purple) and hydrophilic regions (golden) in the E2 binding site. B) 2D Gameplan map that defines the optimum binder to the binding site based on site mapping. It recommends having polar groups at the golden regions and hydrophobic ones at the purple sites to increase binding affinity to the enzyme.

In order to better understand the key features of EE binding pose resulting in a superior inhibitory activity, a 3-D structure-based pharmacophore was developed from the plausible docking pose of EE in 2D06. In this docking study, the inhibition binding mode of E2 was taken as a reference in choosing the plausible binding position (Figure 23). The produced plausible binding pose of EE displayed different orientation in 2D06. EE showed the exact same position for the aromatic ring and hydroxyl group but with plane rotation of EE2 skeleton by 17.3° deeper in the pocket. EE interacted similarly to E2 with PAP showing the previously described E2 inhibitory binding interaction. The hydrophobic nature of acetylene group was responsible for the creation of additional hydrophobic interactions with Phe76, Phe84 and Ile89.

The structure-based pharmacophore generated by LigandScout for EE in 2D06¹⁴⁷ compiled the key interactions for ligand binding and the geometric shape of the cavity. The pharmacophore (SB_PH4) is made of five features showing a H-bond donor (HBD) to the carbonyl oxygen of PAP, a H-bond acceptor (HBA) to the amide nitrogen of Lys48, an aromatic feature and two hydrophobic sites (HYDs). The Hydrogen donating feature towards PAP and away from His108 was believed to be essential to control the inhibitor

conformation and position. The first hydrophobic feature (HYD1) represents the hydrophobic stacking with Phe142 and Phe81, whereas the second (HYD2) was located close to Ala146, Met248 and Val148, resembling the hydrophobic extension that restricts the inhibitor detection for large compounds (aliphatic or aromatic) which do not exceed a certain volume governed by the position of Phe274. In order to translate the superior inhibition of EE over E2 into a binding feature, an additional (HYD3) was included in the model and situated in the area of Phe76, Ile89 and Phe84. It is an exclusive feature for EE that might explain the better inhibition of EE over E2. A complete exclusion volume coat was generated to draw boundaries that determine the capacity of the binding site.

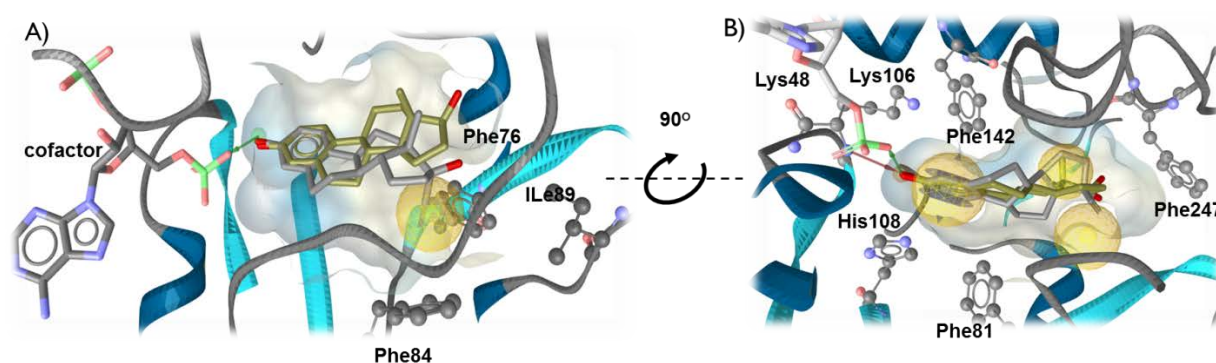


Figure 23: Docking pose of EE: A) Plausible binding pose of EE (silver) in comparison to the original crystallized ligand E2 (golden) shown in stick mode, demonstrating hydrogen bond donor (green arrow) to PAP. The acetylene group of EE2 shows interaction with some protein residues in the pocket (ball and stick). B) Structure-based pharmacophore derived from the docked EE in SULT 1A1 (PDB: 2D06). The pharmacophore shows three hydrophobic features (yellow spheres), aromatic feature (blue ring), hydrogen donating feature (green arrow) towards PAP and hydrogen accepting (red arrow) towards Lys48.

The performance of the pharmacophore was validated against a list of highly active inhibitors (mixed substrates/inhibitors and pure inhibitors). The inhibitor structure-based pharmacophore, and all pharmacophores showed a sensitivity and an early Enrichment Factor (EF) of 45 %, 62%, 5.3 and 8, respectively. The model can retrieve 50% of the compounds, as well as 2 decoys that are structurally related to E2 with no evidence in literature of their SULT1A1 inhibition activity. The ROC curve of the process validation is provided in Appendix Figure A-2.

4.2.3.2 Ligand-based inhibitor pharmacophores

Looking to the fact that the SULT1A1 binding site is flexible and can accommodate a wide variety of compounds, and some SULT1A1 inhibitors bind in an allosteric binding site, the assumption can be made that one pharmacophore is not enough to cover all possible chemical scaffolds and allosteric sites. Using inhibitors collected from literature, ligand-

based pharmacophores were generated from different clusters of highly active inhibitors (Figure 24) using the Espresso function in LigandScout¹⁷⁰. Three ligand-based 3D pharmacophore models were generated representing diverse compounds inhibiting SULT1A1. The first pharmacophore ILB_PH4_1 (Figure 24), generated from a flavonoid cluster, comprises six features: A H-bond donor (HBD) and a H-bond acceptor (HBA) very characteristic to the binding of an inhibitor of PAP and surrounding catalytic residues, two hydrophobic features (HYD1 and HYD2) that resemble binding to a hydrophobic cleft with an important aromatic ring (Ar) that insures planarity at this site of the molecule, and an extra (HBA) that is distinctive to the binding of flavonoids. The second (ILB_PH4_2, Figure 24), created from chlorobiphenyl cluster, contains six features as well: a H-bond donor (HBD) and a H-bond acceptor (HBA) for the binding to the cofactor and surrounding catalytic amino acids, three hydrophobic features (HYD1, HYD2 and HYD3) and an aromatic plane feature (Ar). The third (ILB_PH4_3, Figure 24), produced from non-steroidal anti-inflammatory (NSAI) cluster, contains seven features: two H-bond donors (HBD) and a negative ionizable group (NI) for the binding of the special acidic group of NSAI, two hydrophobic features (HYD1 and HYD2) and two aromatic rings (Ar).

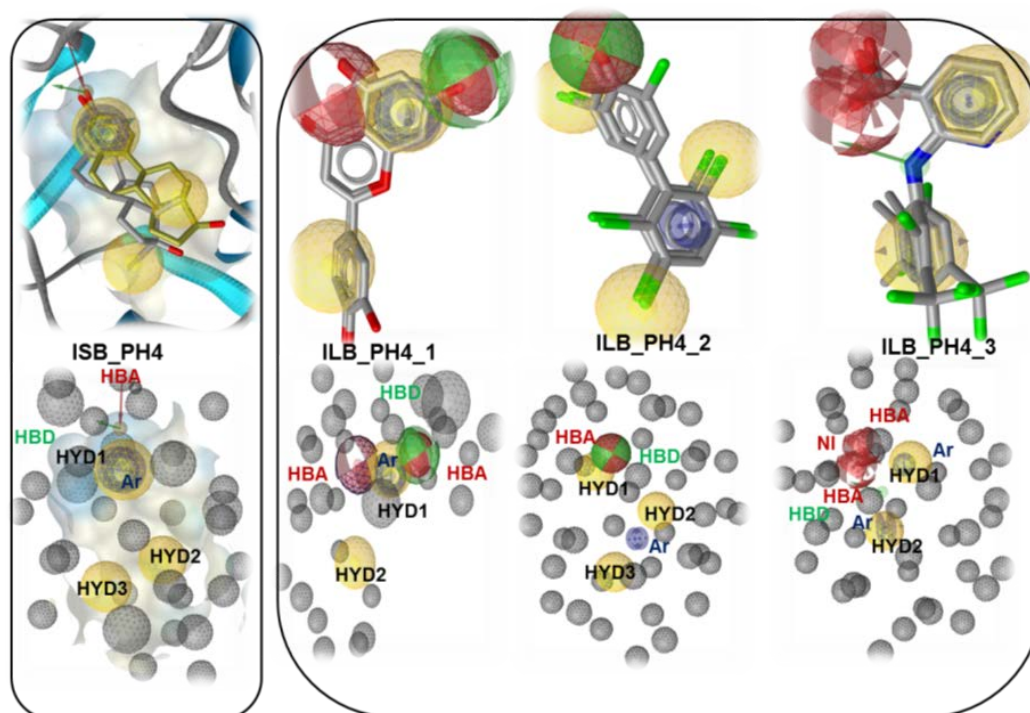


Figure 24: Inhibitor Structure-based and Ligand-based pharmacophores: Left: Inhibitors Structure-based pharmacophore (ISB_PH4) derived from the plausible binding pose of the docked EE (silver) in SULT 1A1 (PDB: 2D06). Right: Ligand-based pharmacophores derived from Clusters of different inhibitors scaffolds generated from flavonoids (ILB_PH4_1) and chloro-biphenyls (ILB_PH4_2) and NSAI (ILB_PH4_3). Chemical features are color coded: H-bond acceptor (HBA) - red sphere or red arrow, H-bond donor (HBD)- green sphere or green arrow, aromatic ring (Ar)- blue circle, hydrophobic contact (HYD)- yellow spheres, and negative ionizable (NI)- red star . The models with aligned ligands are shown on the top, and the 3D pharmacophores with exclusion volumes on the bottom.

The three clusters (flavonoids, biphenyls, and NSAIs) were found to be distant from each other in terms of pharmacophoric features. Thus, the generation of meaningful shared or merged models for these compounds was not possible, as a different binding mode had to be assumed, taking into account that they all had been characterized as SULT1A1 inhibitors. Each pharmacophore could retrieve some structurally related compounds in the validation process. Simultaneous validation of all inhibitor pharmacophores gave a ROC of high selectivity and specificity and retrieved about 70% of the inhibitors (shown in Appendix Figure A-2).

4.2.4 Retrospective validation

About 400 compounds (substrates, mixed substrates/inhibitors, inhibitors and decoys) have been screened through the pharmacophore filters to evaluate their ability to specifically retrieve the desired compounds. A substrate (that binds to SULT1A1 and undergoes sulfonation reaction) was recognized when it fits in one of the substrate pharmacophores and not in the inhibitor pharmacophores. Furthermore, a mixed substrate/inhibitor (that binds to the enzyme, undergoes sulfonation reaction and inhibits binding of other substrates) was defined to fit in one of the inhibitor pharmacophores and in substrate pharmacophores as well. However, an inhibitor was determined if it fits in one of inhibitor pharmacophores and not in substrate pharmacophores.

SULT1A1 binders covering several activity classes and different chemical scaffolds (35 substrates, 26 mixed substrate/inhibitor, and 46 inhibitor) were successfully retrieved by virtual screening through the generated pharmacophores (Figure 25). The generated model showed good discriminative power to differentiate between inhibitors, substrates, and mixed substrates/inhibitors describing the essential features for enzyme binding (accuracy of 80%, 70% and 60% for substrates, mixed substrate/inhibitors and inhibitors, respectively). However, no mechanistic explanation for positive cooperativity behavior could be found using these methods. Accordingly, it was necessary to mechanistically investigate binding behavior in a more detailed way.

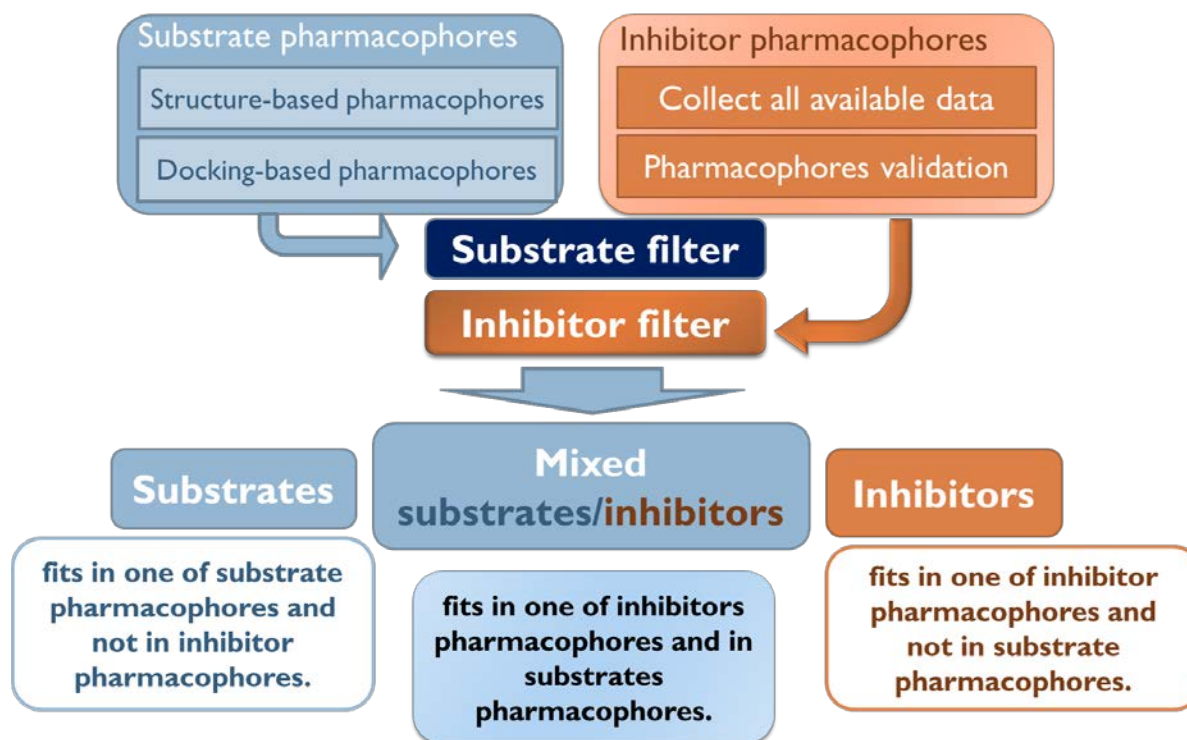


Figure 25: The work flow of the pharmacophore model for in silico discrimination of SULT1A1 binders.

4.3 Molecular Dynamics (MD) study

4.3.1 System Design

To gain insight into the different SULT1A1 complex states and compare their conformations and dynamics, a model of the crystallized and non-crystallized complexes was first constructed *in silico*. To date, the only PAP-containing SULT1A1 binary complex reported in the PDB that shows no mutations has the code 4GRA²¹. The starting coordinates of PAP and SULT1A1 protein used in this study are those of the 4GRA. The apoenzyme was prepared by excision of the cofactor from PSB. For the SULT1A1-PAPS binary complex, PAPS was docked in the same binding site as PAP in the 4GRA structure, after removal of the original cofactor.

To study the SULT1A1 ternary complex dynamics, complexes of different states of cofactor and ligand binding with different ligands were prepared. Ligand-containing SULT1A1 crystal structures present in PDB show varying substrate binding site volume and geometry, which is induced by the presence of several substrates. Therefore, the selection of crystal structures containing the studied ligand already co-crystallized in SULT1A1 was one of the selection criteria. Because of their availability, 2NA, EE and E2 were selected for testing. The first two substrates are positive synergy substrates showing high binding

affinity towards the cofactor-containing complexes^{69, 84}, whereas the latter is neutral synergy substrate showing similar binding affinities towards the cofactor-containing and free complexes⁶⁹. The crystal structure of SULT1A1 bound to the high affinity substrate 2NA was selected to study the first ligand (PDB entry 3U3K)²⁰. For the second and third ligand, the co-crystal of SULT1A1-estradiol (E2) was used, as E2 is an analogue of the high affinity inhibitor EE (PDB entry 2D06)¹⁷. The ligand coordinates from these two structures were extracted to build all possible SULT1A1 complex combinations with ligands (Figure 26). At this stage, twelve complexes for four different systems involving all possible combination of SULT1A1, PAP(S) and a ligand were minimized and used as starting conformation for the MD simulations.

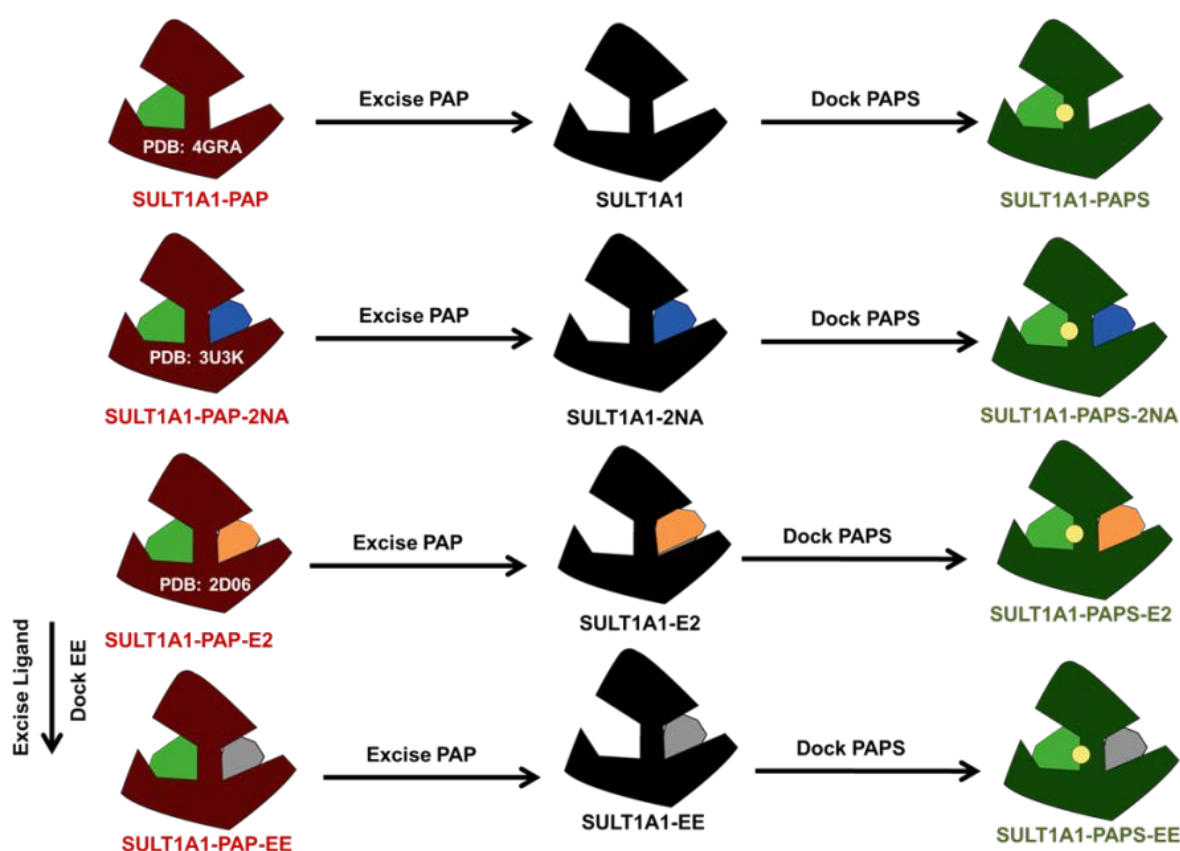


Figure 26: Systems used for the MD study. 2NA: 2 naphthol, E2: estradiol, and EE: ethinyl estradiol. Systems are colored according to the presence of the cofactor into apo, PAP-containing and PAPS-containing (black, red and green respectively).

4.3.2 Analysis of MD results

4.3.2.1 Effect of cofactor binding to the apoenzyme

Ligand-free complexes were simulated in duplicates for 100 ns in explicit water using the apo (SULT1A1) and cofactor-bound conformations (SULT1A1-PAP, and SULT1A1-PAPS).

The analysis of the resulting trajectories showed stable protein over the time frame of 100 ns in all conformations (RMSD and RMSF plots are provided in the appendix). To further focus on the flexibility and/or the stability and substrate binding site volume and geometry after cofactor binding to an empty enzyme; trajectories of the three ligand-free systems were analyzed in terms of structural fluctuation and the most dynamic parts were thoroughly studied.

After equilibration, all ligand-free complexes had a degree of relative stability (average RMSD about 1.5 Å) on the level of C_α of the enzyme (Table 6). However, the substrate binding site showed the highest flexibility when PAPS was bound. After 15 ns of simulation, PAPS appeared to increase the substrate binding site flexibility (average RMSD for C_α pocket residues = 3 Å) in comparison to PAP and apoenzyme (1.5 Å) (Appendix Figure A-3).

Table 6: RMSD values for MD simulations of ligand-free complexes given as average in Å with standard deviations in parentheses.

system		RMSD (C _α protein)	RMSD (C _α substrate pocket)	RMSD (cofactor atoms)	RMSD (ligand atoms)
SULT1A1	1#	1.6 (0.2)	1.6 (0.2)	_____	_____
	2#	1.7 (0.2)	1.7 (0.2)	_____	_____
SULT1A1-PAP	1#	1.5 (0.2)	1.7 (0.2)	1.2 (0.2)	_____
	2#	1.4 (0.2)	1.6 (0.2)	1.1 (0.2)	_____
SULT1A1-PAPS	1#	1.6 (0.2)	2.7 (0.7)	0.8 (0.2)	_____
	2#	1.6 (0.2)	2.6 (0.6)	0.8 (0.2)	_____

The root mean square fluctuations (RMSF) for the complexes (Appendix Figure A-4) backbones reveals two main regions showing the most dynamic residues that respond to the presence of the cofactor in binary complexes: Loop 1 and Loop 3. A higher degree of mobility of Loop 1 was noted by direct comparison of RMSF lines. The loop RMSD value is doubled (RMSD = 2.5 Å) after binding of PAPS in comparison to the non-sulfonated and apoenzymes. The other region of significant alteration observed upon PAP and PAPS binding is Loop 3 which is composed of two segments: Cofactor and ligand segments (Figure 27-a). As illustrated in the RMSF (Appendix Figure A-4), the binding of cofactors induce the stabilization of the cofactor part of Loop 3 residues that covers the cofactor binding site (Phe255- Asp263) over the apocomplex (Figure 27-b). Analysis of MD simulations provided visual indication on the cofactor stable structure in the cofactor active site. Cofactor 3'PO4 exhibits stabilizing interactions to Loop 3 residues (Phe255,

Arg257, Lys258 and Gly259), which are stable allover simulation study (Figure 27-c). The sulfonate moiety in SULT1A1-PAPS is oriented towards Lys48 and Lys106 and stabilized by key interactions with these residues that are believed to play pivotal role in sulfonate transfer reaction. However, Loop 3 acceptor segment that covers the substrate binding site (Asn235 - Phe254) shows different responses to the binding of PAP and PAPS. The presence of PAPS even induces the flexibility of substrate segment of Loop 3. At the same time, it results in the most stabilization for the donor segment of the same loop. A well-defined trough is seen in all enzyme complexes at Phe255, which is considered the junction point between the two segments. Phe255 separates the substrate binding domain from the cofactor binding domain of Loop 3 and shows hydrophobic stacking with Tyr240 in all systems. Another trough is detected at Asp249. This residue displays a molecular hinge interacting with Met145 and stabilizing Loop 3 at its middle. These two linkages are detected in all ligand-free systems. However, direct comparison of RMSF lines showed that the presence of the cofactor decreases the flexibility of residue Asp263. Structure analysis highlighted another salt bridge (with distance less than 4 Å) is found between this residue and another residue present in donor segment (Lys258). This lysine salt bridge is only found in cofactor bound complexes (Figure 27-b) and completely broken in the apoenzyme.

Analysis of MD simulations provided visual indication on the cofactor stable structure in the active site. The cofactor exhibited stabilizing interactions to Loop 3 residues (Phe255, Arg257, Lys258 and Gly259) which were stable allover simulation study. The sulfonate moiety in SULT1A1-PAPS was oriented towards Lys48 and Lys106 and stabilized by key interactions with these residues that are believed to play crucial role in sulfonate transfer reaction.

To measure the stability of the cofactor through the simulation, average RMSD for the PAP/S cofactor atoms was calculated. Overall, PAP and PAPS showed a low degree of flexibility (average RMSD less than 1.5 Å) during the MD study (Appendix Figure A-5 up) which can be attributed to good interaction of the flexible single bonds of the phosphate arms in PSB. The adenine ring RMSD showed stability of the hetero aromatic ring (average RMSD less than 0.1 Å) well sandwiched in donor cleft (Appendix Figure A-5 down). In comparison to PAP, PAPS was less flexible in the PSB due to extra interactions observed between the sulfonate group and Lys48 and Lys106 that stabilized the overall structure of the cofactor.

Figure 27: a. SULT1A1 structure showing Loop 3 with two segments; donor part (red) and acceptor part (blue) with a clear hydrophobic stacking in between (Phe255 and Tyr240). Other interactions were noted as well; molecular hinge between Asp249 and Met145 and a Lys258 salt bridge to Asp263. b. The lysine salt bridge length in different ligand-free complexes SULT1A1 (black), SULT1A1-PAP (red) and SULT1A1-PAPS (green) indicating completely broken bridge in apoenzyme. c. PAP binding interaction in PSB exhibiting hydrogen donating (green) and hydrogen accepting (red) bonds to the donor segment residue (Phe255, Arg257, Lys258 and Gly259).

The substrate binding site dimensions together with its volume were significantly affected by cofactor binding. The acceptor cleft is flanked by the three flexible loops (Loop 1, 2, and 3) and any change of their position directly influences the pocket size and geometry. The substrate binding pocket volume showed many variations in different complexes. The flexible binding cleft exhibited different volumes for various complexes (Appendix Figure A-6). It reached its smaller volume ($\approx 300 \text{ \AA}^3$) when PAP is in place. The cofactor binding structured the substrate binding site by direct interaction with Loop 3 and decreased 40% of its volume tighter than the apocomplex. Similarly, binding of the sulfonated donor (PAPS) gave the same result as PAP in the beginning of the simulation (first 15 ns), then the Loop 1 moved to enlarge the substrate pocket even 50% wider ($\approx 750 \text{ \AA}^3$) than the apoenzyme. One explanation for this observation can be that the sulfonate group is not well stabilized in the hydrophobic environment of substrate pocket. The residue contribution showed that the movement of Phe84 (part of Loop 1) was the most contributor for PAPS-related substrate pocket-volume expansion. In SULT1A1-PAPS complex, the position of the gating loop showed 5 Å withdrawal towards the solvent which resulted in increase in the pocket volume.

The substrate binding site dimensions were carefully studied for all ligand-free complexes. Pocket I and II were not affected by the absence of the cofactor or the flexibility of the loops. They were stable in all complexes (Appendix Figure A-7 a, b) keeping the same distance as the crystal structure ($\approx 10 \text{ \AA}$) which seemed to be ready to engulf aromatic or hydrophobic parts of different binders. These results explain the ability of some substrates to bind to the Apo form²¹. Positions of the pocket residues (Phe81, Phe142, Phe84 and Phe76) ensure the planarity and aromaticity of the hosted substrates through π - π stacking. Indeed, almost all known SULT1A1 binders are of aromatic nature²⁵. On the other hand, the presence of Phe84 as a part of the molecular clamp resulted in varying distances in different unliganded complexes. In the apoenzyme, the two residues kept close π - π stacking at a distance of 5 Å like the parent crystal structure. The binding of PAP did not greatly change the molecular clamp (6 Å) as it did not affect Loop 1 position and hence Phe84. For SULT1A1-PAPS, the complex shows an even much wider clamp (8 Å). Channel I was stable in all complexes with diameter around 6 Å. On the

contrary, Channel II (having Ile89 from Loop 1) showed great alterations between SULT1A1 complexes. Channel II was open wide in case of SULT1A1-PAPS (12 Å), moderately open in case of the free SULT1A1 (8 Å), and reached the narrowest diameter in SULT1A1-PAP (4 Å). The effect of PAPS sulfonate group bound in the enzyme was a conformational change in Loop 1 that caused tripling of Channel II diameter compared to PAP ligand-free complex.

SULT1A1 catalyzes sulfonate transfer with the aid of catalytic residues i.e. Lys48, Ser138. Positions of these residues in relation to cofactor and substrate are pivotal for sulfonation to occur. It was interesting to observe the evolution happens to these catalytic residues during simulation to evaluate the complexes functionality. The cofactor 3'PO₃ group is stabilized in all complexes via hydrogen bond towards Ser138. This H-bond stabilization of 3' phosphate group by the hydroxyl group of Ser138 is preserved in all cofactor-containing crystal structures¹⁰⁹. In addition, Lys48 was found to interact with 5'PO₃ of the cofactor only in PAPS-containing system. This residue is critical for stabilizing the bridging oxygen (the oxygen between sulfonate and phosphate groups) to support the hydrolysis of the sulfonate from donor to acceptor²⁸. In PAP-containing complex, Lys48 unexpectedly shows a hydrogen bond towards Ser138 to keep it away from the catalytic interface.

The MD results reveal that the presence of cofactor greatly affects the overall 3D structure of the enzyme complexes and the geometry of the substrate binding site. The modeled Apo enzyme shows highly flexible loops (Loop1, 2, and 3) greatly affecting the vicinity of cofactor and substrate binding sites. This flexibility can explain the difficulty to crystalize empty SULT1A1²⁰. The dynamic nature of Loop 3 has significant influence of on the overall 3D structure of the apocomplex. Loop 3 flexibility in the apocomplex hinders the formation of a salt bridge between its residue Lys258 and Asp263, located in the adjacent cofactor segment. This lysine salt bridge is uniquely formed in all SULT1A1 crystal structures that contain a bound cofactor. Asp263 is directly adjacent to the dimerization domain and might be responsible for conveying the information of cofactor binding across SULT dimer subunits^{16, 46}. It is worth mentioning that those two residues (Lys258 and Asp263) are conserved throughout the whole SULT family [6]. However, the flexibility of Loop 3 does not affect its stabilizing interaction with Loop 2 at the level of Asp249. Residue Asp249 displays two stabilizing hydrogen bonds towards two Loop 2 adjacent residues (Met145 and Lys147) that stabilize Loop 2 in all SULT1A1 crystal structures^{69, 171}. Both interactions were observed in all complexes, suggesting that they are not involved in cofactor binding.

Because Loop 1 and 3, together with Loop 2, shape the substrate binding site, binding of PAPS directly increases the flexibility and volume of the substrate cleft through increasing the flexibility of Loop 1 and the substrate segment of Loop 3. The present MD simulations provide a dynamic perspective on the substrate binding site in different states in order to understand the conformational consequences that led to substrate binding site expansion in PAPS-containing complex. Simulation results confirm the involvement of the gating loop movement in expanding the volume of substrate binding site. Presence of PAP in the cofactor binding region tightens the substrate binding site up to a volume of 300 \AA^3 (40% tighter than the apo form) by direct interaction with the cofactor segment of Loop 3. Having the same skeleton, PAPS showed the same interactions and stabilization to the cofactor segment of Loop 3 and resulted in substrate binding site reduction at the early stage of the simulation (first 15 ns). After this simulation time, the gating loop restructured to double the substrate binding site volume induced by the presence of the sulfonate group. The gating loop showed 5 \AA -withdrawal movement towards the solvent, increasing SULT1A1-PAPS substrate binding site to 750 \AA^3 . The residue contribution results showed that Phe84 was the main contributor to this volume expansion. Knowing that Phe84 belongs to Loop 1 and just precedes the gating loop indicates that this loop has strong influence on the volume change as reported before²⁰.

Although all the ligand-free complexes were generated from the same crystal structure, various complexes evolved differently through the simulation depending on presence/absence of PAP/PAPS. Some evolutionary events were noticed including Loop 1 withdrawal, Channel I and II opening and closure (Appendix Figure A-8). In general, the ligand-free complexes behaved differently. SULT-PAPS underwent a major conformational change after 15 ns (Appendix Figure A-9). Analysis of the RMSF indicated that Loop 1 and Loop 3 were the most flexible regions and directly related to the substrate binding site volume.

4.3.2.2 Effect of binding different ligands

To investigate the ligand-induced changes in the SULT1A1 binding site in different cofactor-ligand combinations, three different-sized and/or affinity ligands were selected for the MD study. These ligands were 2NA, E2 and EE (Figure 28). 2NA and EE are positive cooperativity small substrate and large inhibitor, respectively; whereas E2 is a neutral cooperativity large size mixed substrate/inhibitor. The analysis of the RMSD and RMSF of the various complexes of both systems shed some light on the influence of ligand(s) binding on the dynamic of the enzyme. The RMSD of the backbone of 2NA complexes

showed a degree of flexibility of the enzyme when the cofactor was not bound (≈ 2.5 Å) (Table 7). However, on the level of the substrate pocket, SULT1A1-PAPS-2NA showed the highest flexible binding site (≈ 3.5 Å) (Table 7) over other systems which can be interpreted as a way to prepare the system to eject the sulfonated product out of the enzyme. E2 showed the same contribution for all systems on the level of the C_{α} of the enzyme (RMSD 1.5 Å). However, SULT1A1-PAPS-E2 demonstrated more stability than its other complexes. The third ligand complexes, EE exhibited double the flexibility in the free enzyme (≈ 2 Å) more than the cofactor-bound complexes (≈ 1 Å). While on the level of the substrate binding pocket, the RMSD represented reduction in the binding site flexibility of all its complexes that could be attributed to the stabilizing effect of the hydrophobic rigid compound and might be a mechanism for inhibition.

Direct comparison of the RMSF of the various complexes of both systems showed the stabilization of the donor segment of Loop 3 in all cofactor-containing complexes (Appendix Figure A-10). Similar to the ligand-free complexes, the cofactor segment exhibited mobility when the binary complex was cofactor-free. In case of 2NA cofactor-free complex, donor part of Loop 3 presented maximum mobility with 50% increase (about 1.5 Å more) in the RMSF value of the trough at Phe255. This event was attributed to broken hydrophobic stacking between Tyr240 and Phe255 (Figure 29) suggesting destabilization effect induced by 2NA. All other binary and ternary complexes exhibited intact stacking suggesting an exclusive destabilization behavior for small-sized substrates towards cofactor-free complex.

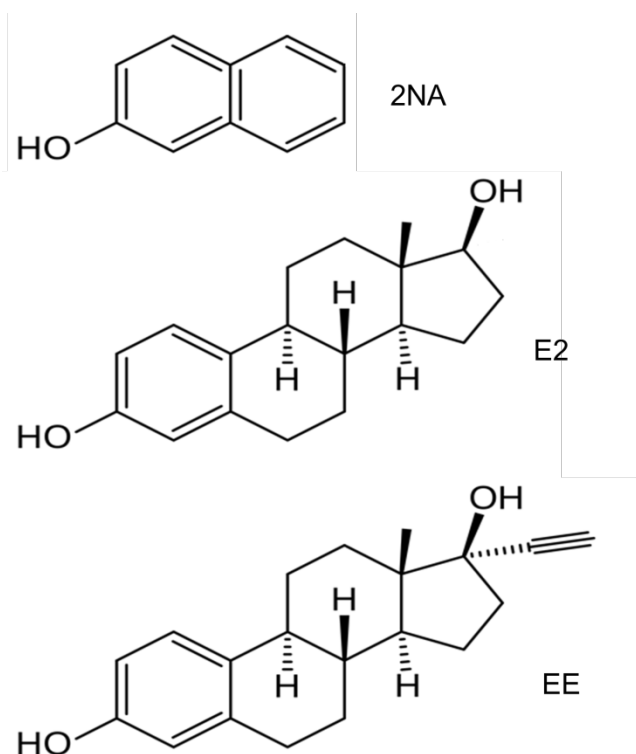


Figure 28: Chemical structures of the ligands used in the MD simulation.

Table 7: RMSD values for MD simulations of several ligand-containing complexes given as average in Å with standard deviations in parentheses. The systems are color coded blue, orange, and grey for 2NA, E2, and EE, respectively.

system		RMSD (C _α protein)	RMSD (C _α substrate pocket)	RMSD (cofactor atoms)	RMSD (ligand atoms)	
SULT1A1	2NA	1#	1.8 (0.4)	2.1 (0.4)	————	0.12 (0.03)
		2#	2.1 (0.4)	2.5 (0.4)	————	0.17 (0.03)
	PAP-2NA	1#	1.2 (0.14)	1.2 (0.2)	1.3 (0.2)	0.12 (0.02)
		2#	1.1 (0.2)	0.9 (0.1)	0.6 (0.2)	0.2 (0.02)
	PAPS-2NA	1#	1.6 (0.2)	2.9 (0.6)	0.7 (0.2)	0.2 (0.03)
		2#	1.3 (0.2)	1.5 (0.2)	0.5 (0.15)	0.2 (0.02)
E2	1#	1.6 (0.2)	1.7 (0.4)	————	0.3 (0.05)	

	2#	1.5 (0.2)	1.6 (0.3)	—————	0.3 (0.05)
SULT1A1-PAP-E2	1#	1.4 (0.1)	1.8 (0.2)	0.6 (0.2)	0.2 (0.06)
	2#	1.4 (0.1)	1.8 (0.2)	0.6 (0.2)	0.3 (0.06)
SULT1A1-PAPS-E2	1#	1.4 (0.2)	1.2 (0.2)	1.4 (0.1)	0.2 (0.06)
	2#	1.5 (0.2)	1.1 (0.2)	1.4 (0.1)	0.3 (0.05)
SULT1A1-EE	1#	1.9 (0.2)	1.8 (0.2)	—————	1.5 (0.2)
	2#	2.0 (0.2)	1.9 (0.3)	—————	1.5 (0.2)
SULT1A1-PAP-EE	1#	1.2 (0.2)	1.2 (0.2)	0.8 (0.4)	1.1 (0.6)
	2#	1.3 (0.1)	1.3 (0.1)	0.8 (0.3)	1.2 (0.5)
SULT1A1-PAPS-EE	1#	1.1 (0.08)	1.4 (0.1)	1.3 (0.2)	1.3 (0.4)
	2#	1.0 (0.07)	1.3 (0.2)	1.3 (0.1)	1.3 (0.5)

The volume of the substrate binding site was found to be ligand-inducible; in other words, substrate binding site volume depends on the size of the hosted ligand (Appendix Figure A-11, 12). SULT1A1-PAP-2NA shows a tight small binding site volume ($\approx 250 \text{ \AA}^3$) over the whole simulation. Larger volume was measured for the PAPS-containing 2NA system (500 \AA^3), while 2NA binary enzyme complex displayed the largest volume calculated in this study ($\approx 1500 \text{ \AA}^3$). However, larger compounds, i.e. EE and E2, induced the enzyme to widen its substrate binding site to be able to adopt them. EE system did not show substrate binding site volume variations among different complexes ($\approx 750 \text{ \AA}^3$ each) presenting a stabilized substrate binding site volume all over the simulation time. E2 system displayed different substrate binding site volumes among different complexes. SULT1A1-PAPS-E2 had the least E2 binding site volume of 400 \AA^3 in comparison to the other complexes ($\approx 750 \text{ \AA}^3$ each).

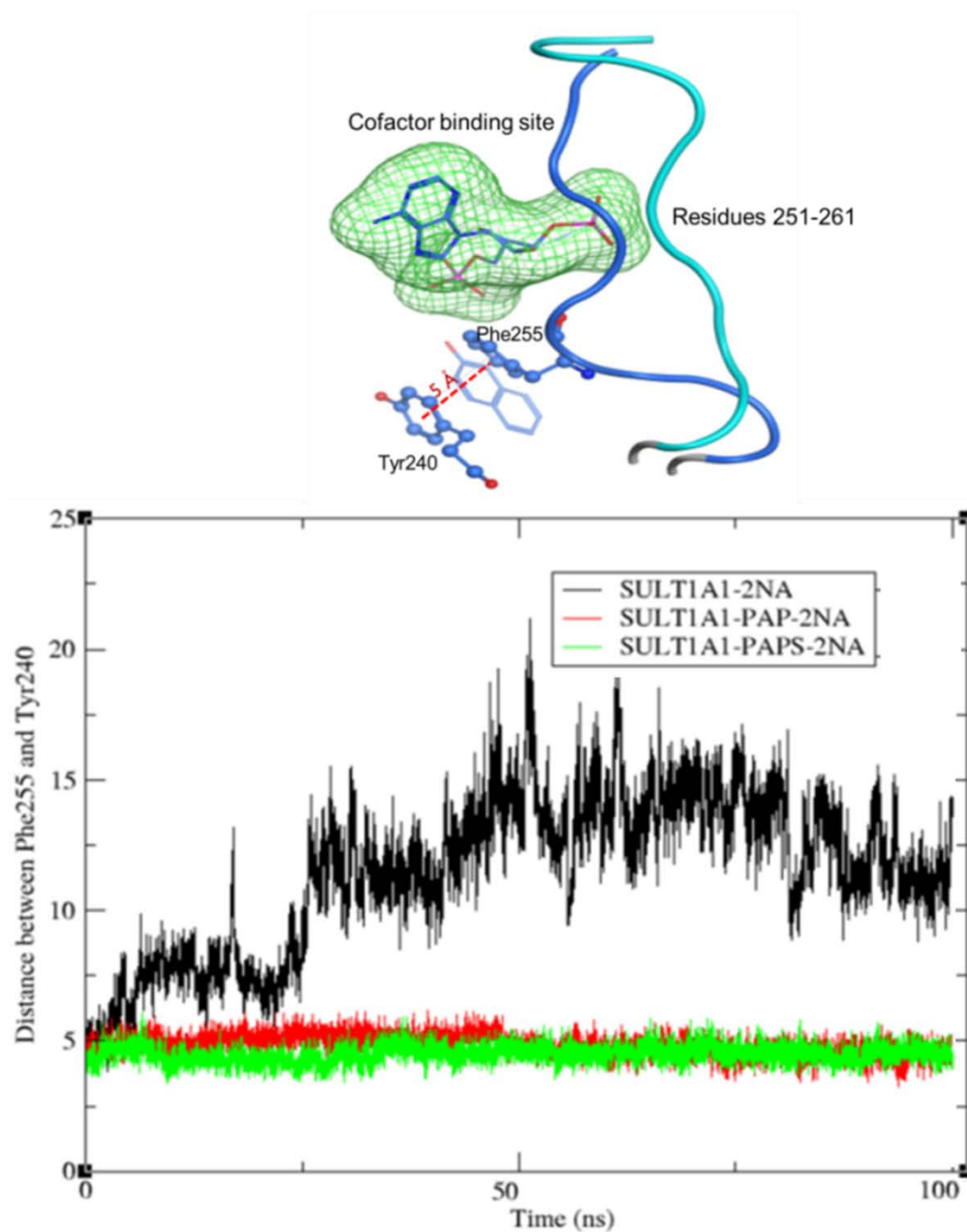


Figure 29: Left: Ten-residue segment that covers the PSB and stabilized by cofactor binding and big hydrophobic binders. Right: The distance between Phe255 and Tyr240 that represents the segment movement in 2NA binary system.

In order to explore structural differences behind the fluctuation in the substrate binding site volume of various systems, the active site dimensions were investigated. Except for SULT1A1-PAPS-2NA, Pocket I shows good stability in all systems with width about 10 Å similar to ligand-free enzyme complexes. In the complex showing the smallest substrate binding site, SULT1A1-PAP-2NA shows a collapsed Pocket II (diameter of 5 Å in comparison to 8 and 10 Å in other 2NA and EE complexes, respectively). The small and hydrophobic ligand (2NA) was only accommodated in Pocket I and exerted interactions

with Pocket II residues to shrink it completely (Appendix Figure A-11). As a result of this volume reduction, the binding site of SULT1A1-PAP-2NA not only collapsed Pocket II, but also narrowed Channel II and the molecular clamp (6 Å each) (Appendix Figure A-13). In contrast, the large substrate binding site of the 2NA binary complex (1500 Å³) was attributed to the induced flexibility of Loop 3 (Appendix Figure A-14). The presence of 2NA in the cofactor-free enzyme destabilized the cofactor segment of Loop 3 by disrupting the π stacking between Tyr240 and Phe255. For SULT-PAPS-2NA system, the sulfonate group induced substrate binding site expansion. Pocket I diameter of this system continued to increase during simulation time until it reached 16 Å that was coordinated with the opening of Channel I (up to 10 Å). Pocket I was destabilized due to presence of sulfonate group in 2NA system.

E2 only occupies Pocket I in SULT1A1-PAPS-E2 showing enlarged Pocket I volume (400 Å³). In this complex, Pocket II showed the “collapsed” form as SULT1A1-PAP-2NA. E2 does not contain hydrophobic groups that could be accommodated in pocket II and restores its structure. For the other complexes, SULT1A1-PAP-E2 and SULT1A1-E2, E2 occupied Pocket I as well with frequent opening of Pocket II that was indicated by fluctuation in Pocket II diameter. The diameters of Pocket II of various systems are provided in the Appendix Figure A-11.

For EE system, Pocket II was wide and stable in all EE complexes (9 Å). The pocket was accommodating the ethynyl group of EE via CH π -stacking in all its complexes presenting new binding pose different from the crystallized E2. EE was accommodated in both pockets, unlike 2NA, which only occupied Pocket I. The larger ligand size might not be the only reason for the new binding position, as EE and E2 more or less have the same size. E2 in our studies was accommodated as well in an “enlarged” Pocket I. Opening of Pocket II might depend also on the hydrophobicity of the terminal functional group. If the terminal functional group is hydrophobic enough (like ethynyl group in EE), it might interact with the hydrophobic residues to induce binding site growing.

The molecular clamp behavior was highly related to Pocket II having a residue (Phe84) in common. It had smaller diameter (5 Å) when Pocket II was collapsed (2NA and E2 systems) and a wider diameter (8 Å) when Pocket II was open (in EE system). Moreover, Channel I showed closed conformation in majority of the complexes with a width around 5 Å. The only complex that deviated was SULT1A1-PAPS-2NA that showed very wide Channel I (up to 10 Å) through which 2NA could exit. It was clearly demonstrated in the MD that 2NA almost got out through Channel I after 30 ns of the simulation (Appendix Figure A-15).

Furthermore, Channel II was dependable on Pocket II being a part of it. The channel showed small width if Pocket II is in the collapse form that was showed in 2NA and E2 systems. On the contrary, Channel II expanded to engulf the extended hydrophobic part of the ligand in the case of SULT1A1-EE systems.

2NA binding conformations in different complexes were investigated and compared to each other, and to the 3U3K crystal structure. In the dead-end complex (SULT1A1-PAP-2NA), the ligand conformation occupying Pocket I was similar to the crystallized pose, with its hydroxyl group forming H bonds with the cofactor and the two catalytic residues Lys106 and His108 (Appendix Figure A-13). The complex had some structural differences than the crystal structure i) a small substrate binding site, induced by 2NA interaction with aromatic/hydrophobic residues, ii) the diameters for Pocket II and the molecular clamp got narrower forming a collapsed part of the active site, iii) Channels I and II diameters displayed narrower distances as well (5, and 6 Å, respectively), iv) Lys48 interacted with Ser138 keeping it away from the catalytic interface, and iv) the aromatic ring invaded deeper to interact directly with PAP (dead-end interaction). All the previous changes were translated to a significant reduction of the binding site volume (250 Å³).

The SULT1A1-PAPS-2NA enzyme complex shows great differences compared to the dead-end complex (SULT1A1-PAP-2NA), in terms of active site structure and binding evolution scenario (Appendix Figure A-15). In the beginning, and for no longer than 10 ns, it exhibited a similar structure to SULT1A1-PAP-2NA. Only three minor differences were observed: (i) the distance between the cofactor and substrate was longer to withstand the extra sulfonate group of the cofactor, (ii) Pocket II was a bit wider (7Å), and (iii) the orientation of Lys48 was different, stabilizing the bridging oxygen in the cofactor rather than Ser138. These differences are most probably the evolution of the binding site from a dead-end complex to an "active" Michaelis-Menten complex that would undergo sulfonation. Furthermore, after 10 ns, the substrate binding site showed similar conformational movement to SULT1A1-PAPS. Pocket I and II got larger with increased mobility. At 30 ns, opening of Channel I was observed and the molecule exited through it after 40 ns.

In the SULT1A1-PAPS-E2 complex (Appendix Figure A-16), Lys48 interacts with Lys138 during most of the simulation time with low probability of interaction with bridging oxygen (at 5ns and 95ns). Although the E2 hydroxyl group shows a typical interaction distance from His108 and the cofactor in all complexes, the catalytic interaction with Lys106 was distorted showing far distances and improper hydrogen bonding angles. E2 was invading

deep into the binding pocket interacting with the amide group of Lys48 instead of the cofactor. The hydroxyl group of E2 in the pocket was oriented away from Lys106, which is essential for sulfonation. Pocket II and channel II remain in the closed form.

The simulation of EE in the SULT1A1 crystal structure with PBD entry 2D06 revealed a new binding conformation different from the co-crystallized E2. EE in the starting conformation had the same binding position of E2 occupying Pocket I only (Appendix Figure A-17). Furthermore, Pocket II, and Channel II diameters of the original crystal structure were moderately wide (7 Å each). Throughout the simulation, Pocket II and Channel II diameters of EE complexes became wider (9, and 10 Å, respectively) and EE invaded deeper in the substrate binding site. The hydrophobic acetylene group was accommodated in Pocket II (the inhibitory pocket) by C-H π interaction with Phe76 and Phe84. Overall, this complex shows a conformation capable of sulfonation, where Lys48 stabilizes the bridging oxygen. The hydroxyl group forms hydrogen bonds to His108, Lys106 and the active cofactor. The Michaelis-Menten complex appeared to be catalytically well positioned ready for the sulfonation by SULT1A1 that explains its reported sulfonation activity⁸⁴.

4.3.3 Proposed hypothesis of positive cooperativity behavior

Although various sources mention that SULT1A1 is active towards small phenolic compounds, its binding site was co-crystallized with larger binders¹⁷. However, SULT1A1 exhibits much higher sulfonation activity towards small-sized ligands that show high binding affinity⁶⁹. The catalytic efficiency of an enzyme is not solely linked to the ligand high binding affinity but also to a fast turnover rate. Some mutations, which increase enzyme affinity to substrates, decrease catalytic efficiency by hindering product release¹⁷². Strong substrates show a balance between high binding affinities and their rate of release out of the enzyme after sulfonation. For small aromatic rings, the attachment of an ionized sulfonate group is enough to sweep the sulfonated product out of the binding cleft. On the contrary, bulky ligands tend to bind more tightly in the substrate binding site via more interactions features. The hydrophilic sulfonate group is not enough to expulse the well accommodated compounds out of the enzyme and to let another molecule get in. Therefore, larger compounds show a slow turnover rate despite their high binding affinity. These results greatly support the notion that size affects the molecule activity towards SULT1A1, to be a substrate or an inhibitor. Similar results are published recently, showing the high catalytic efficiency that exceeds the diffusion limit for some small aromatic

binders⁶⁹. The present MD study reveals that ligand size can define the type of ligand activity, favoring small binders to be substrates and large binders to be inhibitors.

The 2NA binding pose in the PAP-containing crystal structure has been captured before and reported in a previous study as a typical pose for substrates²⁰. 2NA in the crystal structure is positioned in a catalytically competent manner in which the acceptor group (hydroxyl) is positioned about 2.5 Å from the catalytic amine group of His108 and about 3.5 Å from Lys106 as depicted in PDB 3U3K²⁰. The acceptor location is stabilized by π -stacking interactions in Pocket I by hydrophobic residues (Phe81 and 142), as well as by hydrogen bonds with a crystal water that resembles the hydrolyzed sulfonate group²⁰. Its binding mode is exactly the same like other crystallized substrates; however, the substrate pocket shows a smaller volume and narrower Channel II. The high affinity and high catalytic efficiency behavior of 2NA cannot be correlated to its binding mode. However, its positive cooperativity behavior was explained in a recent report by a conformational change of two molecular clamp residues (Phe81 and 84) that translocate to fix the substrate in the binding pocket facing catalytic residues competent to sulfonation⁶⁹. Nevertheless, the unavailability of structural evidence to this molecular clamp conformation decreases the credibility of this explanation: This conformation has been used as a starting point for a 1ns simulation after docking various acceptors in 4GRA ligand-free crystal structure. The previous results might be biased by docking of the acceptor in a ligand-free enzyme crystal structure that may need longer time to show ligand-induced conformational changes. Surprisingly, this position of an aromatic ring positioned between Phe81 and 84 couldn't be sampled by our MD study. Instead, the molecular clamp showed very narrow distance in between much related to the collapse of the second binding pocket.

The analysis of enzyme complexes with 2NA (SULT1A1-2NA, SULT1A1-PAP-2NA and SULT1A1-PAPS-2NA) gave an explanation for the positive cooperativity and high catalytic efficiency behavior of small binders. 2NA showed a stable dead-end complex (PAP-containing) on the level of α carbon of the whole enzyme and binding site. Binding of the small hydrophobic molecule induced the closure of Channels I and II, and the collapse of Pocket II that significantly reduced the substrate binding site volume (250 Å³). This reduction might be responsible for 2NA two prominent behaviors. First, the binding site reduction increased the substrate hydrophobic interaction with more residues (Pocket II residues (Phe84 and 76) together with channel residues (Ile89), enhancing binding affinity and thus favoring a positive cooperativity behavior. Second, the acceptor hydroxyl group is brought in a good position for hydrogen bonding with the cofactor and two catalytic

residues (Lys106 and His108). This interaction is pivotal for sulfonate group transfer if sulfonate group and Lys48 are involved⁴⁸ and might be the reason behind substrate enhanced catalytic activity. Previous experimental findings support the formation of a strong dead-end complex of small binders^{20, 69}.

Despite the stable PAP-containing ternary complex of 2NA, the presence of PAPS instead of PAP in the enzyme complex drastically changes the flexibility of the enzyme complex. Simulation showed the possibility of Channel I opening in response to the presence of sulfonate group in the binding cleft. This releasing event may be the reason for the high catalytic efficiency and fast turnover of small binders induced by the presence of the sulfonate group. The tight hydrophobic interactions of small binders can be completely distorted due to the presence of the sulfonate group that will drive the sulfonated compound out of the substrate cleft. The dead-end complex is considered previously as a mechanism of substrate inhibition by stabilizing the “non-productive” ternary complex⁷⁰. This hypothesis may be not be valid if SULT1A1 is considered a coordinated dimer in which each monomeric partner is communicating with the other. In the enzyme dimer, the presence of PAP significantly enhances PAPS binding to the other monomer. Subsequently, the bound PAPS stimulates PAP release from the dimer subunit providing oscillating half-site reactive system⁴⁷. Indeed, some SULT isoforms are found to exhibit half-site reactivity with only half of the catalytic binding sites (in the dimeric enzyme) catalyzing sulfonation at any given time^{44, 70}. This half-site reactivity supported by oscillatory cycle has a pivotal biological significance in maintaining SULT1A1 high catalytic efficiency towards small phenyl binders. The most similar simulated complex to the 3U3K crystallized protein²⁰ is SULT1A1-PAPS-2NA at the beginning of the performed simulation. It might be true that the crystallized complex was just a picture of the enzyme prior to sulfonation with the sulfonate group replaced by crystal water in the crystallization process. This crystal structure can give crystallographic proof of the tight binding site hypothesis proposed here. Previous mutational results confirmed the importance of all protein residues involved in this tight hydrophobic interaction for the positive cooperativity behavior of small substrates. For example, after Y250C mutation, SULT1A1 shows less affinity to small phenolic compounds despite enhanced catalytic velocity²⁰. Mutation in Phe255 into alanine residue is proved to decrease SULT1A1 catalytic efficiency towards positive cooperativity substrates⁶⁹. Mutations of F255A and I89A are proved to transform positive cooperativity substrates into neutral cooperativity substrates⁶⁹.

For the apoenzyme containing 2NA, the molecule was too small to provide the needed interactions for Loop 3 stabilization in absence of the cofactor which is responsible for low binding affinity of 2NA towards the open enzyme. Furthermore, the binding of 2NA to the apoenzyme destabilized the cofactor segment of Loop 3 by disrupting the interaction between Tyr240 and Phe255. This destabilization of the closed form of the complex may result in an open gate that is more capable of binding the cofactor, producing an “active” ternary complex. This hypothesis requires further experimental proof.

EE is the only known positive cooperativity inhibitor⁸⁴. The binding affinity has not been measured for other inhibitors that show comparable K_i values. However, many other inhibitors may exist. In comparison to its structural analogue E2, EE has potent inhibitory properties and lower binding affinity (by two orders of magnitude) towards the cofactor-containing complexes. An explanation to this behavior is proposed by a previous docking study that claimed that EE is able to catalytically bind the free enzyme in a catalytically competent orientation with proper positioning of hydroxyl group towards His108, but in a non-competent orientation via formation of a charge interaction with Tyr169 if PAPS is bound first⁸⁴. Although, this explains the superior inhibition of EE, it cannot rationalize the positive cooperativity behavior of EE. The molecular clamp theory falls short here as well. If EE would show a molecular clamp conformation to achieve high binding affinity to the cofactor-containing complexes, it would enhance its sulfonation rate, which is not the case. The data discrepancy in these regards opens lots of questions about the relationship between size, binding affinity and ligand activity.

In the light of the simulations presented in this work, another explanation is proposed for the EE behavior. Indeed, no inhibitor-Tyr169 interaction was observed throughout 100ns of MD. Simulation of EE in different enzyme complexes revealed a favorable binding pose that might be the reason for the enhanced binding affinity and, at the same time, superior inhibition behavior. The ethinyl group of EE significantly increased the binder interaction with the hydrophobic binding cleft more than E2. As illustrated in Figure 30, the ethinyl group was positioned directly in Pocket II, well interacting with its hydrophobic residues (Phe76 and 84). As shown previously, Pocket II is involved in a second molecule binding that is considered a mechanism of substrate inhibition. Occupying Pocket II with a second molecule considerably decreases the overall catalytic efficiency of SULT1A1^{15, 20}. Therefore, occupying both pockets might be a way of inhibition for bulky binders as EE.

Not only size matters. The MD study for E2 and EE revealed another important characteristic for a good binder: Hydrophobicity. Having approximately the same volume,

EE and E2 are accommodated differently according to their terminal hydrophobicity. The hydrophobic terminal ethynyl group of EE was responsible to open a new pocket between Phe84 and 76. In addition, the study showed that binding of EE highly stabilized all complexes of the enzyme. PAPS-containing complex did not show any releasing event. Some ligand-induced changes in the binding site that accompanied the entrance of EE could be noticed including the opening of phenyl alanine rich binding site in the vicinity of the enzyme to accommodate the large inhibitor. This conformational change might distort or delay the ability of the enzyme to release the sulfonated product.

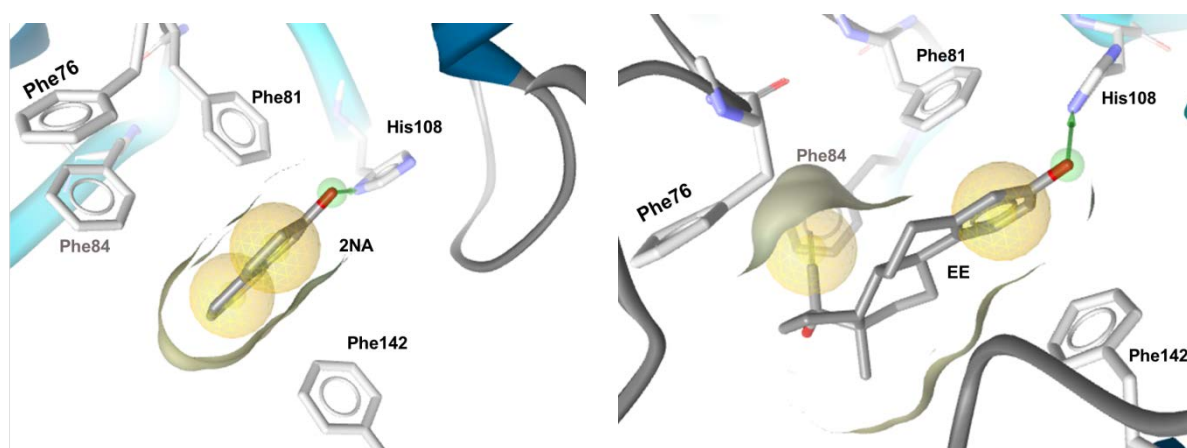


Figure 30: Plausible binding modes of high affinity binders. Left: SULT1A1 tight binding site accommodating small binder (2NA) surrounded by phenyl alanine rich cleft. Right: SULT1A1 extended binding site accommodating large binder (EE) with its ethynyl group hosted between Phe76 and Phe84.

MD results explained more information about E2 neutral cooperativity behavior that it might show different binding position than EE. In all enzyme complexes, the hydrophobicity of E2 could not enforce its accommodation in pocket II. Pocket II was mostly collapsed and not occupied. Moreover, E2, 2NA and EE showed diverse planes inside the binding pocket. If we take the plane of 2NA as reference considering it the highest binding affinity and substrate catalytic efficiency, E2 and EE showed 35° to 55° different plane angle respectively (Figure 31). It was reported previously that changing the ligand binding plane inside SULT2A1 pocket is a way of substrate inhibition³³. It seemed that both EE and E2 are disoriented in the binding cleft that gives a degree of inhibition to the catalytic activity of SULT1A1.

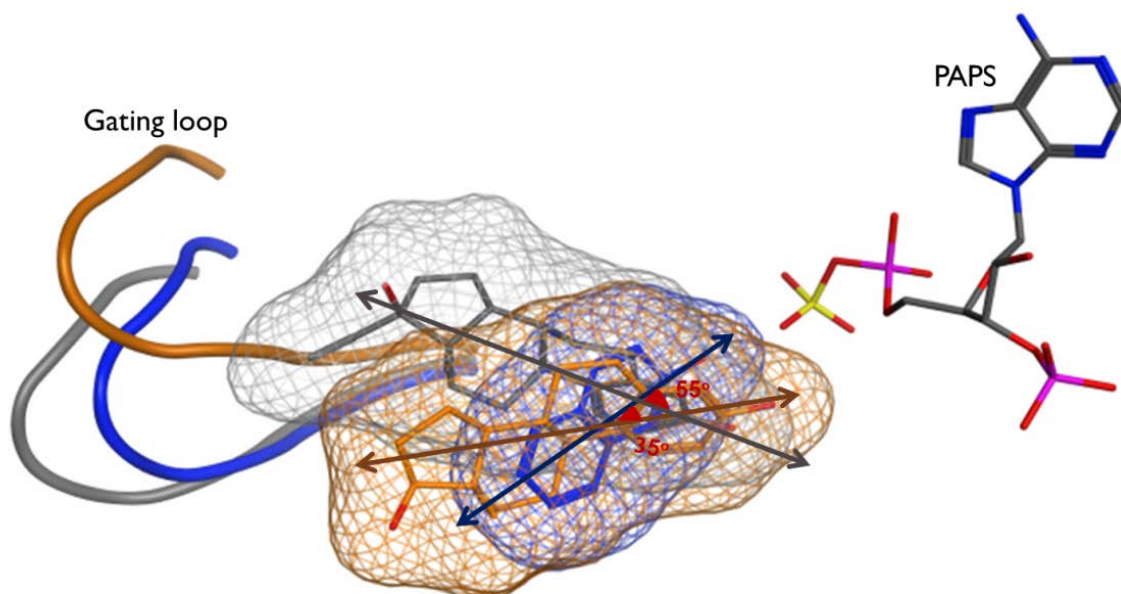


Figure 31: Planes of different binders inside SULT-1A1 ternary structures. Gating loop showed various ligand-induced positions according to the bound ligand 2NA, E2 and EE (blue, orange and grey respectively).

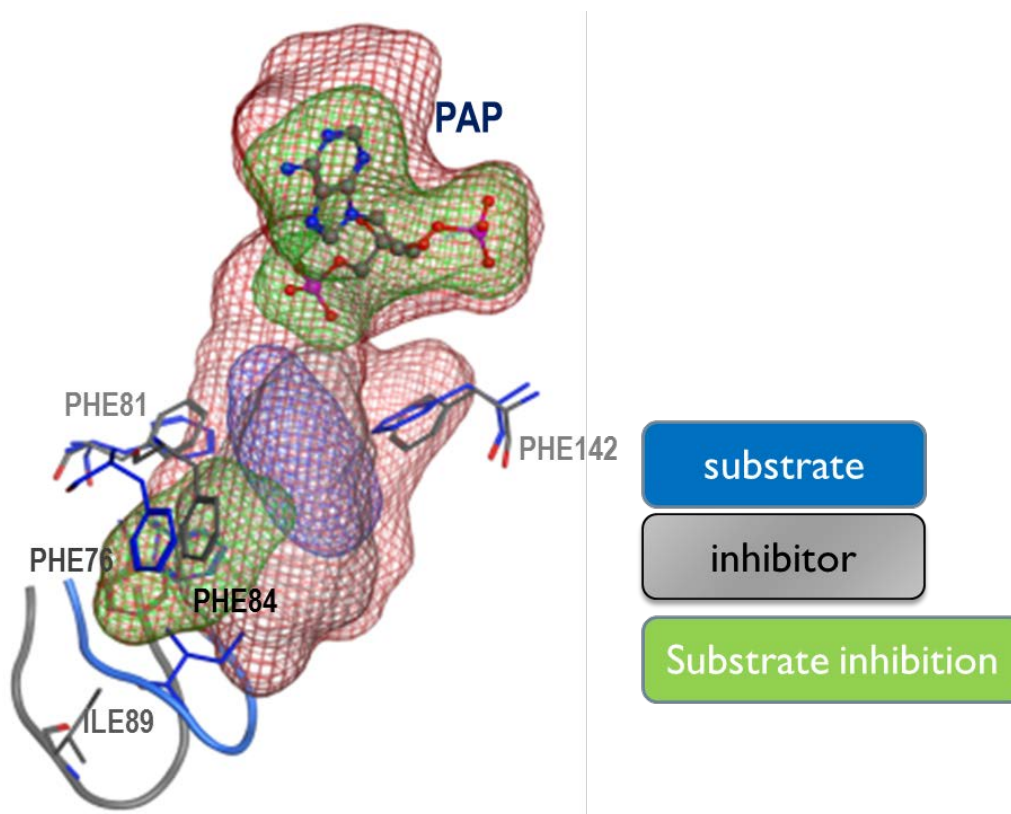


Figure 32: Binding place for small binder substrates (blue mesh), large binder inhibitors (grey mesh) and Pocket II responsible for substrate inhibition (green mesh).

In conclusion, a positive cooperative SULT1A1 substrate should be small sized and aromatic to interact with all aromatic residues in the Pocket I (Phe81, Phe142) as well as collapse Pocket II (Phe84, Phe76) (blue mesh in Figure 32). On the other hand, larger high affinity binders would mostly act as inhibitors. They bind in the substrate binding site occupying both Pocket I and II spaces. They should show a degree of hydrophobicity at both poles of the compound to fit exactly in the extended pocket (grey mesh in Figure 32).

5 Discussion

5.1 Structural investigations

Structural investigations of available SULT1A1 crystallographic data aimed to identify catalytically important parameters and structural descriptors defining the substrate selectivity and specificity of SULT1A1 enzyme. Investigations revealed fundamental SULT1A1 features relevant to its broad specificity. These features include i) the substrate plastic pocket that has been co-crystallized with substrates of different sizes and sometimes two molecules of the same substrate, ii) its L shape conformation lined with hydrophobic and/or aromatic residues defining a lipophilic environment, and iii) the catalytic interface consisting of key catalytic residues (Lys106 and His108) together with the cofactor.

Except for 2D06 crystal structure, all other co-crystallized ligands display exactly the same binding position and orientation in which the hydroxyl groups of all substrates (except E2) orient towards the catalytic residue His108 with average distances 2.2 - 2.5 Å. In 2D06, E2 molecule invades deep in the protein structure with no interaction towards His108, showing a non-catalytic binding mode (dead-end complex). The amino acid residue His108 is a key residue in the sulfonate transfer reaction as it acts as a base for unleashing nucleophilic power of the hydroxyl group via deprotonation. Therefore, a proper distance of (2.2 - 2.5 Å) between the hydroxyl group of a substrate and His108 is considered to be substantial for the catalytic competent binding mode.

Structural superposition revealed another parameter: the pocket volume is induced by ligand size. The dimensions of the substrate binding pockets vary in response to the change in the conformation of gating loop located in Loop 1. Plasticity of the three loops (Loop 1, 2, and 3) has been reported to be responsible for defining the architecture and geometry of the substrate cleft¹⁶. However, only Loop 1 flexibility was observed across SULT1A1 crystal structures. Loop 1 showed different conformations in various crystal structures co-crystallized with different-sized substrates. The loop withdrawal towards the solvent was correlated to the expansion of the pocket volume under the influence of large substrates.

Structural comparison of SULT1A1 to its siblings SULT1A3 and SULT1E1 revealed alterations in the position and residue composition of Loop 1, Loop 2, and Loop 3. In

addition to the three loops, SULT1E1 has two other regions that also display different conformation and sequence in comparison to the two other isoforms. Variations in the substrate pocket sequence between various isoform were found to create distinct electronic environment that is responsible for the unique substrate binding profile. SULT1A1 has a highly hydrophobic substrate pocket with minimal regions of hydrogen accepting or donating residues. This unique lipophilic pocket originates from hydrophobic side chains oriented into the pocket (Phe81, 142, Ala146, Ile89 and Val148). In summary, structural investigations support the contribution of SULT1A1 pocket parameters (size, geometry, nature) in defining its broad substrate specificity and understanding its substrate binding profile.

5.2 Pharmacophore generation

3D pharmacophore generation for substrates and inhibitors was carried out next, in an attempt to distinguish the 3D pharmacophoric features for SULT1A1 substrates and inhibitors and try to define the boundaries between sulfonation and inhibition activities. The generated pharmacophores were validated by virtually screening a substrate/inhibitor dataset in presence of many decoys. The power of the pharmacophores to retrieve true positives was evaluated by ROC.

The generated substrate pharmacophores reflected the essential features for good substrate binding. The pharmacophore consisted of 2 or 3 independent features (depending on the size of the substrate); i) aromatic (Ar) and hydrophobic interaction (HYD) characteristic for the aromatic ring of all substrates, ii) hydrogen donating (HBD) and hydrogen accepting (HBA) features of the group liable to sulfonation (OH or NH₂), and iii) an extra (HYD) interaction that facilitates the accommodation of the large compounds (rigid or flexible) in the proper orientation inside the substrate pocket. The low number of features present in the substrate pharmacophores reflects the broad specificity of SULT1A1 enzyme that sulfonates any aromatic hydroxyl or amino containing compound that can be accommodated in the flexible pocket. Upon simultaneous validation using substrate dataset, pharmacophores showed high selectivity and specificity to retrieve most of the substrates (90%) from more than 360 decoys. However, the substrate model falls short to explain substrate activity of large substrates. The increase in pharmacophore fit score of large substrates cannot be correlated to their diminished sulfonation activity in reality ⁶⁹. Binding of an extended multi-ring substrate needs a conformational rearrangement to the gating loop that leads to pocket size expansion. The expansion of

the pocket volume might occur at an energy cost which restricts high sulfonation activity of SULT1A1 towards exclusively small phenolics¹⁵.

Considering the binding mode of inhibitors in SULT1A1, E2 co-crystallized binding pose in SULT1A1 (2D06), showing non-catalytically competent orientation and strong substrate inhibition, was considered as inhibitor binding mode. Therefore, the 3D structure-based pharmacophore describing inhibitors mode was built from the plausible binding pose of the potent inhibitor and estradiol derivative EE in 2D06⁸⁴. The binding orientation of EE in 2D06 was 17.3° deeper in the hydrophobic pocket. Rotation resulted from the hydrophobic nature of acetylene group that exerted additional hydrophobic interactions with Phe76, Phe84 and Ile89 although the aromatic ring and hydroxyl group occupied the exact same position. The generated pharmacophore could retrieve only 40% from various inhibitors. Regarding the fact that some inhibitors do not bind in the same binding site as modeled, ligand-based pharmacophores were beneficial in covering these inhibitor classes. The inhibitor dataset was clustered into 3 different big classes and each one was used to generate a structure-based pharmacophore. The derived structure-based pharmacophores could recognize some structurally related compounds similar to those used in their construction. The simultaneous validation of all pharmacophores (structure- and ligand-based), displayed good specificity and selectivity with an early enrichment Factor (EF) of 8 and could retrieve about 60% of the inhibitors.

Although constructed 3D models have good ability to discriminate inhibitors from substrates and decoys, their applicability to virtually screen of large datasets remains limited. The minimum number of features in a pharmacophore to be a useful tool for virtual screening is 4-5 independent features¹⁷³. The few number of independent features present in substrate- and inhibitor- models reflected the broad specificity nature of SULT1A1. In addition, the rationale behind ligand classification, defining the chemical profile for each class is still not clear. Therefore, another modeling technique was needed to investigate the induced changes in the substrate binding pocket due to binding of various binders.

5.3 MD simulation

Hypothetically, SULT1A1 may show several *in vivo* enzyme complexes representing all possible combinations of cofactors and binders. However, only one complex of SULT1A1 was crystallized: PAP-containing enzyme with or without ligand^{15, 17, 20, 21}. Other enzyme complexes might be present in the cytoplasm competing for the acceptor. The apoenzyme

that contains neither the cofactor nor the substrate might exist in a period of enzyme life. Indeed, the concentration of the active cofactor (PAPS) in the cytoplasm - which is very tissue dependent⁷ - greatly affects the observed proportion and life time of the SULT-PAPS complex. PAPS binds to the apoform giving an "active" holoenzyme that sulfonates different acceptors. By progression of the sulfonation reaction, the non-activated cofactor (PAP) accumulates in the medium. In cellular cytoplasm, PAP competes with PAPS for SULTs binding to exactly the same place; i.e. PAP is a potent inhibitor to all SULTs³⁶. Even though it was more feasible to crystallize PAP containing SULTs, it is not the active species performing sulfonation. Therefore, it is important to study the dynamics and stability of the missing key complexes of SULT (with and without cofactor and/or ligands) in order to unveil the molecular mechanisms of SULT substrate recognition and reach better understanding of SULT1A1 specificity and selectivity. At this stage of the work, 12 different SULT1A1 complexes were modeled *in silico* and simulated for 100 ns covering all possibilities of ligand and cofactor binding.

5.3.1 Effect of cofactor binding

Cofactor binding is a crucial step in the sulfonation process, regardless of binding before or after the ligand. The dual substrate property was confirmed more than twenty years ago⁶², but which substrate will bind first is still a controversial issue⁶⁷. Binding of the sulfonate donor has been reported to significantly stabilize the cofactor segment of Loop 3. The cofactor interacts with some residues of this segment causing its structural rearrangement into a rigid form [24, 61]. The cofactor-dependent stabilization greatly restricts the entrance of the second co-substrate (ligand). This gating hypothesis was supported by mutational and MD studies [39, 41]. In SULT1A1, some *in vitro* and *in silico* experiments showed the possibility of a similar phenomenon [9]. Although Loop 3 stabilization is the basis for cofactor priming to the substrate binding site conformation [26], the influence of cofactor binding on the volume and flexibility of substrate binding site has not been addressed yet.

MD analysis confirmed that cofactors (PAP/S) are not equivalent, affecting the dynamic of SULT1A1 in a different manner. Radically different substrate binding site volume and conformation were observed for SULT1A1-PAP and SULT1A1-PAPS, which might affect differently the binding affinity of various substrates. In contrast to PAP, PAPS binding did not stabilize the substrate part of the loop. SULT1A1-PAPS showed the highest flexibility at the level of the hydrophobic substrate segment although, at the same time, it also showed the best stabilization of the cofactor segment. The additional sulfonate group

of the cofactor PAPS had a prominent effect on the flexibility of the substrate segment. Additionally, another difference was found in Loop 1 (Lys81–Pro90), a region only found in the SULT1 family which contains the gating loop (Lys85–Pro90) at its end. This region displayed the highest dynamic differences in RMSF between the two cofactor-containing ligand-free complexes. It is believed to be the main entrance for ligands binding to the substrate binding site⁸⁵ and it has an integral role in substrate inhibition by allowing the entry of a second molecule¹⁵.

In the presented work, it was observed that PAPS binding significantly increases the flexibility of this loop. Even though sulfonate group is the only variable between the two structures, our results indicate that its presence in the cavity can confer significant conformational changes to the gating loop, resulting in an increase in Channel II diameter and an expansion of the binding site volume. Our results indicate that the presence of the sulfonate group is responsible for all conformational changes happening to SULT1A1-PAPS. The reason may be the destabilizing effect of the sulfonate group in the tight hydrophobic substrate binding site, which is not favorable for the sulfonate group leading to expansion and flooding with water molecules to stabilize the electronic environment of the whole system. SULT1A1-PAPS shows the widest distance between Phe81 and Phe84, the broadest Channel II, and therefore enlargement of the binding site volume to its maximum between all ligand-free complexes.

Considering SULT as a dynamic enzyme, with a changing binding pocket volume throughout the enzyme cycle, can give an insight on the size and geometry of the probable binders for each complex. If a substrate in the cytoplasm is exposed to all enzyme complexes, it will have a preferential binding to one of them depending on the substrate size. Small substrates (less than 300 Å³) will have equal chances to sink to all enzyme complexes. However, larger substrates (300 - 750 Å³) will be restricted to open and PAPS-containing complexes only. Presence of the SULT enzyme as an oscillating PAP/PAPS dimer, combining two enzyme complexes together in one system is still under investigation⁴⁷. If confirmed, this additional information will add even more complexity to the investigation of SULT dynamics. Noteworthy, our results reveal that the apoenzyme does not respond evenly to the binding of PAP and PAPS. Significant differences in substrate binding site volume, and flexibility were observed. Consequently, enzyme complexes can be considered as distinct conformational entities in equilibrium rather than equivalent forms.

5.3.2 Ligand-size induced binding site flexibility

In order to define a relationship between size, binding affinity and ligand activity, different-sized binders were studied in this work, covering diverse binding behaviors; a positive cooperativity small-sized substrate (2NA), a neutral cooperativity large-sized substrate (E2) and a positive cooperativity large-sized inhibitor (EE). The described MD study revealed some new chemical features of SULT1A1 binders. This work suggests that binder dimensions have a prominent effect on the binding site volume and geometry by inducing various conformational changes leading to different favorable binding poses that affect the enzyme specificity and selectivity.

Based on MD results, a novel explanation can be given for the positive synergy behavior for both substrates and inhibitors. Small-sized substrates can reduce volume of the substrate binding site and make intimate interactions with hydrophobic/aromatic residues that lead to a proper pose, directed towards catalytic residues, only in cofactor-containing complexes. Therefore, cofactor-containing complexes will have higher tendencies to be filled first by substrates (positive cooperativity behavior). If the concentration of the substrate increases to reach the saturation concentration, it may compete for binding the apoform as well. Their binding before the cofactor destabilizes the closed form leaving the gate open for the cofactor binding and formation of the "productive" ternary complex. An experimental study is needed to measure the binding affinity of the cofactor, and if it is enhanced by binding of small substrates to confirm our hypothesis.

For the positive cooperativity inhibitors, simulation of EE in different enzyme complexes revealed a favorable binding pose that might be the reason for the enhanced binding affinity, and at the same time, superior inhibition behavior. EE hydrophobic terminal ethinyl group was responsible to open an extra pocket between Phe84 and 76 extending the hydrophobic system, allowing stacking interactions with the ligand and triplicating the binding site volume. This extra pocket opening is responsible to partially lock large hydrophobic binders inside even after sulfonation and might be responsible for their inhibitory effect. Furthermore, the MD study for E2 and EE revealed another important characteristic for a good binder: Terminal hydrophobicity. Having the same skeleton, EE and E2 are accommodated differently according to their terminal functionality. If the terminal functional group is hydrophobic, it interacts with Pocket II residues increasing the ligand binding affinity.

6 Conclusions

Binding site information is important to define the general boundaries for binders' specificities to a given enzyme. However, better understanding of the dynamic substrate recognition process is the corner stone for accurate prediction of substrate binding profile to various isoforms. This research aimed to explore several angles of the positive cooperativity behavior of SULT1A1 substrates and inhibitors.

Findings of the structural investigations across SULT1A1 crystal structures significantly reveal the substrate binding site flexibility. The binding site is induced to occupy larger or tighter space by the size of the binders. Furthermore, structural investigation of other highly related siblings clears the characteristic hydrophobic nature of the substrate binding site of SULT1A1.

Structure-based and ligand-based 3D pharmacophores defined the essential features for SULT1A1 binding. The pharmacophores successfully differentiated more than 100 SULT1A1 binders of various chemical scaffolds and different activity classes. The generated model discriminated well inhibitors, substrates and mixed substrates/inhibitors (70% and 60% for substrates and inhibitors, respectively). However, the positive cooperativity substrates and inhibitors did not result in a high pharmacophore fit score. Therefore, another *in silico* modelling technique was necessary to explore the enzyme flexibility.

An MD study was conducted to assess the stability of SULT1A1 in its apo form, together with other complexes bound to different states of cofactor and a variety of different binders. The analysis revealed that SULT1A1 binding site is flexible and its size varies not only according to binder size but also its hydrophobicity. Small phenyl binders can tighten the binding pocket up to 250 Å³ by close hydrophobic interactions with Phe81, Phe142, Phe76, Phe84, and Ile89 in cofactor-containing complexes. In contrast, binding of small-sized substrates to apoenzyme distorted the stacking between Tyr240 and Phe255 giving more flexibility to the cofactor part of Loop 3. In the case of bigger binders, Phe76 and Phe81 flexibility contributes to opening an extra pocket and extend the hydrophobic system allowing stacking interactions, if the ligand is hydrophobic enough. Since all SULT members are structurally highly conserved, information obtained by studying SULT1A1 could be eventually extended to other enzymes of the family after experimental testing.

7 Experimental

7.1 Structural investigation and pocket mapping

7.1.1 Multiple sequence alignment (MSA):

The crystal structure of SULT1A1 (PDB: 2D06¹⁷), SULT1A3 (PDB 2A3R²³), and SULT1E1 (PDB 4JLV³⁰) were aligned to investigate the structural and sequential differences between the isoforms using Clustal Omega offered from the EMBL-EBI¹⁷⁴. The analysis tool web service conducts Multiple Sequence Alignment (MSA) via a guide tree produced from progressive alignment of various sequences¹⁷⁵. The three FASTA sequences were fed to the web-based software and aligned using the default parameters.

7.1.2 Molecular Interaction Fields (MIF):

This part was done to identify electronic nature of SULT1A1 binding pocket. Using MOE (Chemical Computing Group, Montreal, CA)¹⁶¹, the molecular interaction fields (MIF) of the protein residues lining the substrate pocket of SULT1A1 (PDB: 2D06¹⁷), SULT1A3 (PDB 2A3R²³), and SULT1E1 (PDB 4JLV³⁰) were generated. Three main probes were used; "N:" to resemble H-accepting groups, "OH2" to identify H-donating interactions, and "DRY" to calculate the hydrophobic interaction energies. The interaction potentials have been manually adjusted to provide a precise mapping of the binding pocket, relating spots with favorable interaction to specific protein residues.

7.1.3 Pocket-volume analysis

The volume of the binding pocket for SULT1A1 published crystal structures (1LS6, 3U3K, 3U3M, 3U3O, 3U3R, 3U3J, 2D06, 1Z28, and 4GRA) was calculated from their respective crystal structures using Epock[®]¹⁷⁶. The pocket was characterized as maximum encompassing region (MER) in a configuration file by the aid of inclusion spheres that cover the vicinity of the known binding cleft. In case of ligand-containing crystal structures (1LS6, 3U3K, 3U3M, 3U3O, 3U3R, and 2D06), the coordinates of the original ligands were used as centers for Epock inclusion spheres. The software plugin for VMD was used to visualize the input files and results. For the ligand-free crystal structures (3U3J, 1Z28, and 4GRA), the cavity was described as big inclusion spheres around pocket key residues (His108 and Phe247). Grid points spaced 0.5 Å were generated by the software in

the pocket using inclusion spheres of 9 Å radius around the atom positions. The contiguous option was turned “true” with cutoff of 4 Å during all volume calculation. The contiguous seed spheres were centered at the same coordinates as the inclusion spheres.

7.1.4 2D06 binding site mapping using Szmap®

Szmap software is used to assay the binding cleft using a molecule of water. The software uses semi-continuum solvation [23] to predict how and where neighboring waters can impact binding of a ligand, and by using related software (Gameplan) it suggests ligand modification hypotheses designed to better exploit specific regions on a binding site. Primarily, protein (PDB: 2D06) and ligand (E2) were prepared; hydrogens were added and partial charges were calculated using MOE. Then, the protein containing PAP was separated in one file different from that one that contains ligand (E2) using pch application provided from Szmap, version 1.1.1, openeye scientific software, Inc., Santa Fe, New Mexico, USA ¹⁶⁷. Next, using the same software, calculations for complex, apo, and ligand grids were tested and then further processed to yield (complex - apo - ligand) stabilization grids, explaining where water is stabilizing or destabilizing the binding reaction. The application Gameplan ¹⁶⁹ provided by Szmap, was used to analyze ways to modify E2 chemistry based on water structure and energetics in its immediate environment.

7.2 Pharmacophore generation

7.2.1 Generation of PAPS containing crystal structures

Interestingly, no published crystal structures of SULT1A1 contain PAPS as a cofactor ¹⁵, ²⁰. PAPS-containing models were generated by docking PAPS after removal of PAP from phosphor sulfonate binding site (PSB). Six PAP containing crystal structures were obtained from the Protein Data Bank ¹⁸ (PDBs: 1LS6 ¹⁵, 2D06 ¹⁷, 4GRA ²¹, 3U3M, 3U3O and 3U3K ²⁰) and used as matrix to generate six PAPS containing structures through three steps. The first step was preparation of PAPS and protein crystal structures for the docking study. The PAPS chemical structure was obtained from PDB: 1HY3²⁸. It was primarily protonated using Protonate 3D ¹⁷⁷ and minimized using the MMFF94x force field in MOE (Chemical Computing Group, Montreal, CA) ¹⁶¹. Next, the crystal structures were prepared by GOLD, V 5.1 (CCDC, Cambridge, UK) ¹⁷⁸; Hydrogen atoms were added. The second step was validation of the docking protocol. The co-crystallized PAP was removed from the structure and the PAPS binding site (PSB) was defined by choosing all residues around the original cofactor in

a sphere of 7 Å. All docking runs were performed in triplicates using standard default variables and GoldScore. The PAP original binding pose could be retrieved using default parameters. The third step was docking of the active sulfonated cofactor (PAPS). The same protocol succeeded previously in PAP redocking, was then used for docking PAPS in the binding site of PAP. The PAPS binding mode has been reported previously for the human SULT1E1²⁸, highlighting an orientation identical to crystallized PAP. Therefore, the pose with least RMSD to the original conformation of PAP in the original crystal structure was the one considered to be most plausible.

7.2.2 Docking of EE in 2D06

The protein crystal structure (2D06)¹⁷ was used for the EE docking study. The previously described protocol was used in ligand and protein preparation and protocol validation. Original E2 was excised from the structure and its binding site was defined exactly like the previous docking. In order to model SULT1A1-PAP-EE enzyme complex, the most plausible pose for EE was chosen having the least RMSD to the original E2 ligand and at the same time achieving important hydrogen bond interactions towards key residues in the binding cleft. Furthermore, to demonstrate the complex for EE in PAPS-containing enzyme prior to sulfonyl transfer, EE was docked into the sulfonated enzyme (2D06-PAPS). For this system, the catalytically competent binding mode of E2 observed in SULT1E1 crystal structure¹⁶⁵ was used to choose the most plausible pose of EE with the least RMSD value and reasonable interactive distance (≈ 3 Å) to PAPS sulfonate group and catalytic His108.

7.2.3 3D structure- and ligand-based model generation-

The software package LigandScout 3.1¹⁴⁷ was used to generate 3D structure-based pharmacophores from SULT1A1 crystalized or docked ligands in crystal structures (EE (docked) in 2D06 for the inhibitor model and pNP in 1LS6, 3CyC in 3U3M, and 2NA in 3U3K for the substrate models) using default options. Similarly, the most active inhibitors have been clustered to different scaffolds and used for generation of ligand-based models. These pharmacophores were used to virtually screen different binders.

7.3 MD simulation

7.3.1 System preparation

For the ligand-free enzyme PDB entry 4GRA²¹ was used to study SULT1A1-PAP, SULT1A1-PAPS after docking of PAPS instead of PAP in PSB, and SULT1A1 after excision of PAP from PSB using the same coordinates of PAP and cofactor-free SULT1A1. The three complexes of ligand-free enzyme (SULT1A1, SULT1A1-PAP and SULT1A1-PAPS) were then minimized and used in the MD. For the enzyme complexes that contain ligands, 2NA, E2 and EE were used in our systems in various combinations with cofactors as shown in Figure 26. 3U3K²⁰ was used to study the dynamics of its bound ligand 2NA whereas PDB: 2D06 was employed to study the dynamics of its co-crystallized ligand E2 and another analogue EE¹⁷. For each starting structure, three different systems were constructed: SULT1A1-Lig (ligand bound but no cofactor), SULT1A1-PAP-Lig (ligand with PAP cofactor) and SULT1A1-PAPS-Lig (ligand with PAPS cofactor).

Ligands (2NA/E2/EE) and cofactors (PAP/PAPS) were parameterized by antechamber^{160, 179}, one of the AMBER tools¹⁶⁰. Tleap¹⁶⁰ was used to solvate the complexes in a TIP3P water box and to neutralize them by adding Na⁺ counter ions. MD simulations were carried out for the twelve systems, starting with minimization using SANDER¹⁶⁰. The systems were minimized then subjected to gradual heating up process to 300⁰ K during a short initial MD simulation (20 ps) with weak descending restraints on the whole system. This was followed by a production phase for (100 ns) performed under the NPT ensemble using software AMBER¹⁶⁰ and the ff12SB¹⁸⁰ force field with a time step of 2 fs. Trajectories were saved at 2 ps intervals.

7.3.2 Trajectories analysis

All MD simulations were analyzed by the software cpptraj¹⁸¹, to calculate the root mean squared deviation (RMSD) value for the internal atomic coordinates to the reference molecule coordinates (the minimized structure). As well, cpptraj was used to compute the average RMSDs for the C_α of the substrate pocket. Additionally, the same software was used to track the distance change between the protein residues in the substrate binding site and the catalytic interface.

7.3.2.1 Distance analysis

Catalytic distances:

The distances between catalytic residues, the cofactor and binders were carefully chosen to identify plausible hydrogen bonds between residues. The distances were selected between least flexible species not to be affected by bond rotation through dynamic simulation. To resemble H-bond formed between the Ser138 and the cofactor, the distance between the terminal O of Ser138 and 3' PO₄ of the cofactor was measured. For the hydrogen bond that takes place between Lys48 and Ser138, the space between the terminal N of Lys48 and hydroxyl O of Ser137 was computed. Moreover, the hydrogen bond between amino group of Lys48 and the cofactor was assessed by measuring the distance between Lys48 amino N and 5' PO₄ of the cofactor. For the possible distance between the donor and acceptor, the distance between 5' PO₄ (in case of PAP-containing systems) or S (in case of PAPS-containing systems) and the O atom of the acceptor hydroxyl group was counted. Finally, the distance between His108 and cofactor was measured by minding the space between 5' PO₄ (in case of PAP-containing systems) or S (in case of PAPS-containing systems) and the cyclic N in His108.

Binding distances:

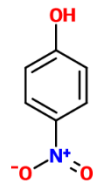
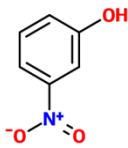
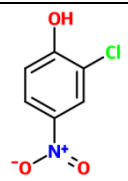
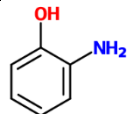
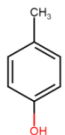
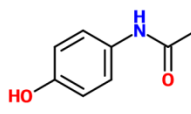
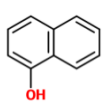
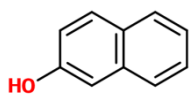
To assess the binding distances in the hydrophobic pocket, the distances between the centers of masses of the hydrophobic side chains were measured and traced through simulation. For Pocket I, Pocket II, Channel I, Channel II, and molecular clamp, the space between the two pairs Phe81 and 142, Phe84 and 76, Val148 and Phe274, Ile89 and Phe76 and Phe81 and 84, were predicted respectively and monitored throughout the simulation course.

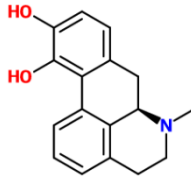
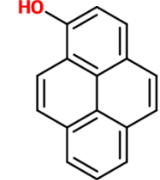
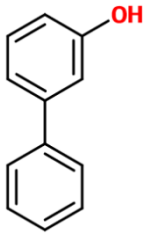
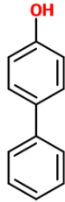
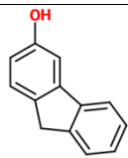
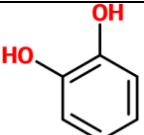
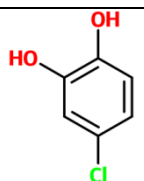
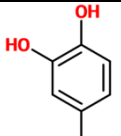
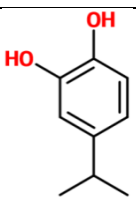
7.3.2.2 Pocket-volume analysis of the whole trajectories

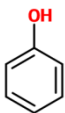
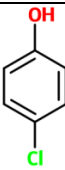
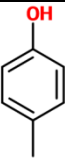
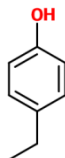
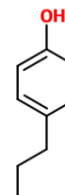
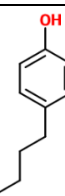
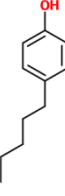
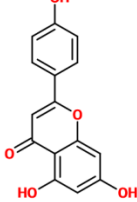
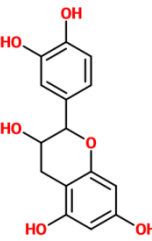
The volume of the binding pocket over simulation time was calculated from the MD trajectories using Epock[®] 176. The pocket was characterized by the same procedure described in 7.1.3. The extracted frames (one per five = every 10 ps) were superimposed onto the starting coordinates of the minimized system that was used for cavity determination. Residue contribution in pocket volume was also calculated using the same software.

8 Appendix

Table A-1: Substrates used in pharmacophore validation and their reported biochemical values.

Compound	Structure	Reported biochemical parameters						Ref	Remarks
		K_m (μM)	V_{max} ($\text{nM min}^{-1} \text{mg}^{-1}$)	k_{cat} (sec^{-1})	k_{cat}/K_m ($\mu\text{M}^{-1} \text{sec}^{-1}$)	K_d (μM)			
						E	E.C		
P-NP		1	580					88	SI ¹⁵
		4	74					182	
		0.36	9.2					183	
m-NP		0.7	364					88	
Cl-NP		5	28		0.003			106	
O-AP		9	121					15	SI
P-cre		0.5	315					88	
Ace		0.02		100	4500	0.5	0.018	69	PS
		430	22					184	
1NA		0.08		100	1300	1.3	0.07	69	PS
		13.7	1220		0.1			185	
2NA		0.06		110	1700	1.5	0.06	69	PS
		12	1380					185	
		8	1130		0.07			106	

apo		0.05		58	1100	0.8	0.05	69	PS
		4.3	81.8					185	
1HY		0.02		120	6600*	5.5	0.02	69	PS
3-phenP		1.3	2410		0.02			185	
4-phenP		56	3630		2			185	
2-Fluor		23	1450		0.07			185	
C		3.9	40					119	
4-CIC		1.4	16					119	
4-MethC		3.7	52					119	
4-IPC		4.4	79					119	

Phenol		4.5	151				119	
		19	350		0.01		106	
4-CIP		6	487				119	
4-MethP		7.6	572				119	
		10	320		0.02		106	
4-EthP		10	410		0.02		106	
		1.4	296				88	
4-PropP		6	427		0.04		106	
4-ButP		5	296		0.03		106	
4-PenP		9	208		0.01		106	
Api		1.7	5.8				186	
Epic		241	3.8				186	
		1.8	10.5				187	

Chry		2.5	8.4					186	
Que		2	2.8					186	
Res		2.2	6.8					186	NS
		1.6		1.5	0.93	1.1	0.85	69	
3,3' T2		0.12	0.5					166	
		0.63						100	
AdP		1	47		0.02			106	
BPC		0.06	1.6					188	

Abbreviations: E= apoenzyme, EC = enzyme cofactor binary structure, p-NP = para-nitro-phenol, Cl-NP2 = chloro-4-nitrophenol, O-AP = 2-Aminophenol, p-cre = P-cresol, Ace = Acetaminophen, 1NA = 1-naphthol, 2NA = 2-naphthol, apo = apomorphine, 1HY = 1-hydroxypyrene, 3-phenP = 3-phenylphenol, 4-phenP = 4-phenylphenol, 2-Fluor = 2-Fluoreno, C = catechol, 4-CIC = 4-Chlorocatechol, 4-MethC = 4-Methylcatechol, 4-IPC = 4-Isopropylcatechol, 4-CIP = 4-Clorophenol, 4-MethP = 4-Methylphenol, 4-EthP = 4-eylphenol, 4-PropP = 4-propylphenol, 4-ButP = 4-butylphenol, 4-PenP = 4-pentylphenol, Api = Apigenin, Epi = Epicatechin, Chry = Chrysin, Que = quercetin, Res = Resveratrol, 3,3' T2 = 3,3'-Diiodothyronine, AdP = 4-(1-adamanty)phenol, BPC = Benzo[α]pyrene-7,8-catechol, asterisk (*) indicates that the value was not correctly calculated in the original publication, SI = substrate inhibition, NS = neutral synergy, and PS = positive synergy.

Table A-2: Inhibitors used in pharmacophore validation and their reported inhibition information. Abbreviations: No ev. = no evidence is reported in literature for being sulfonated.

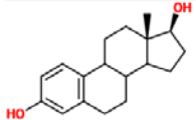
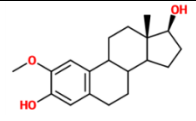
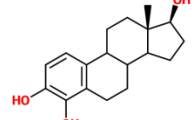
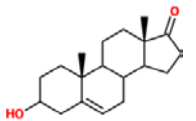
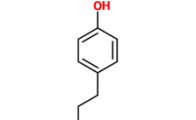
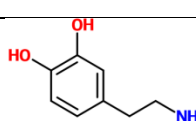
	Compound	Inhibition data (μM)		Substrate?	Ref
		IC ₅₀	K _i		
1	Epigallocatechin gallate		0.04	No	110
2	Epicatechin gallate		0.06	No	110
3	Mefenamic acid	0.02		No	189
4	Tolfenamic acid	0.12		No	189
5	Niflumic acid	0.28		No	189
6	Meclofenamic acid	0.87		No	189
7	Flufenamic acid	1.5		No	189
8	3-hydroxyflavone	1		No	163
9	5- hydroxyflavone	0.06		No	163
10	3',4'-Dihydroxyflavone		0.001	No	190
11	3,6-Dihydroxyflavone		1	No	190
12	S-(+)-1,2,3,4-tetrahydro-1-naphthol		0.05	No	191
13	(1R,2R)-(-naphthalene-1,2-dihydrodiol		1.1	No	192
14	(1s,2s)-(-naphthalene-1,2-dihydrodiol		1.7	No	192
15	(5S,6S)pentacyclo[10.6.2.0p2,7.0p]9,19.0p16,20]icosa-1,3,7,9(19),10,12(20),13,15,17-nonaene-5,6-diol]		0.004	No	192
16	(5R,6R)pentacyclo[10.6.2.0p2,7.0p]9,19.0p16,20]icosa-1,3,7,9(19),10,12(20),13,15,17-nonaene-5,6-diol]		0.004	No	192
17	Benzoic acid		0.28	No	192
18	Phenylacetic acid		1.1	No	192
19	1-Naphthoic acid		0.02	No	192
20	1-Naphthylacetic acid		0.2	No	192
21	2-Naphthoic acid		0.041	No	192
22	2-Naphthylacetic acid		0.41	No	192
23	Salicylic acid		0.07	No	192
24	Naproxen		0.13	No	192
25	1-naphthaldehyde		0.03	No	192
26	2-naphthaldehyde		0.9	No	192
27	4-chlorobiphenyl-3'-ol	0.26		Yes	193

28	2,4 ,6, 3` -tetrachlorobiphenyl-4` -ol	1		Yes	193
29	2,5 ,3` , 5` -tetrachlorobiphenyl-4` -ol	0.41		Yes	193
30	2,3 ,5 , 6, 3` -pentachlorobiphenyl-4` -ol	0.54		Yes	193
31	2,3 ,5 , 6, 3` , 5` -hexachlorobiphenyl-4` -ol	0.51		Yes	193
32	Curcumin	0.014		Yes	111
33	Quercetin	0.01		Yes	186
34	Kaempferol	0.4		No ev.	194
35	Fisetin	0.1		Yes	194
36	Myricetin	0.2		No ev.	194
37	Apigenin	0.06		Yes	186
38	Galangin	0.06		Yes	194
39	Chrysin	0.3		Yes	194
40	Deidzein	0.4		Yes	195
41	Genistein	0.5		Yes	195
42	3,7-Dihydroxyflavone		0.2	Yes	190
43	Baicalein		0.03	Yes	190
44	Luteolin		0.003	Yes	190
45	Morin		0.8	Yes	190
46	Formononetin		0.5	Yes	190
47	3',4',7-Trihydroxyisoflavone		0.02	Yes	190
48	17 α -Ethinylestradiol		0.01	Yes	84
49	2,6-dichloro-4-nitrophenol	1		Yes	196

Table A-3: Comparison between substrate clefts dimensions of all SULT1A1 crystal structures.

PDB ID	3U3R	3U3K	3U3O	3U3M	1LS6	2D06	4GRA	3U3J	1Z28
Cavity volume (Å³)	738	531	893	599	671	845	510	562	830
Pocket I diameter (Å)	7.5	7.5	7.5	7.5	7.5	7.5	7.5	7.5	7.5
Channel I width (Å)	6.2	6.6	6.8	6.4	6.2	7.1	6.2	6.4	6.8
Pocket II diameter (Å)	8.1	7.1	7.7	7.9	8	7.3	6.8	7.1	7.5
Channel II width (Å)	8.1	5.8	7.9	7.5	7.9	6.9	6.9	6.2	8.6

Table A-4: Some high affinity substrates for other SULT isoforms. The molecules were added to the decoy list used in pharmacophore validation. Red color refers to SULT1E1 substrates, yellow to SULT2A1 substrates, and green to SULT1A3 substrates. Abbreviations: E = apoenzyme, EC = enzyme cofactor binary structure, E2 = Estradiol, DHEA: Dehydroepiandrosterone, EE = Ethinyl estradiol, Tyr = Tyramine, dopa = dopamine, ME2 = 2-methoxyestradiol

Compound	Structure	Reported biochemical parameters						Ref	Remarks
		K _m (μ M)	V _{max} (nM min ⁻¹ mg ⁻¹)	k _{cat} (sec ⁻¹)	k _{cat} /K _m (μ M ⁻¹ s ⁻¹)	K _d (μ M)			
						E	E.C		
E2		1		1.2	1.2	1	0.9	69	NS, SI 17
		2.4	16			2.3	1.2	84	
ME2		0.9	3.6					197	
OHE2		2.5						198	
DHEA		0.77		0.75	0.97	0.7	0.7	69	NS
Tyr		6483	197					88	
Dopa		88	4					183	

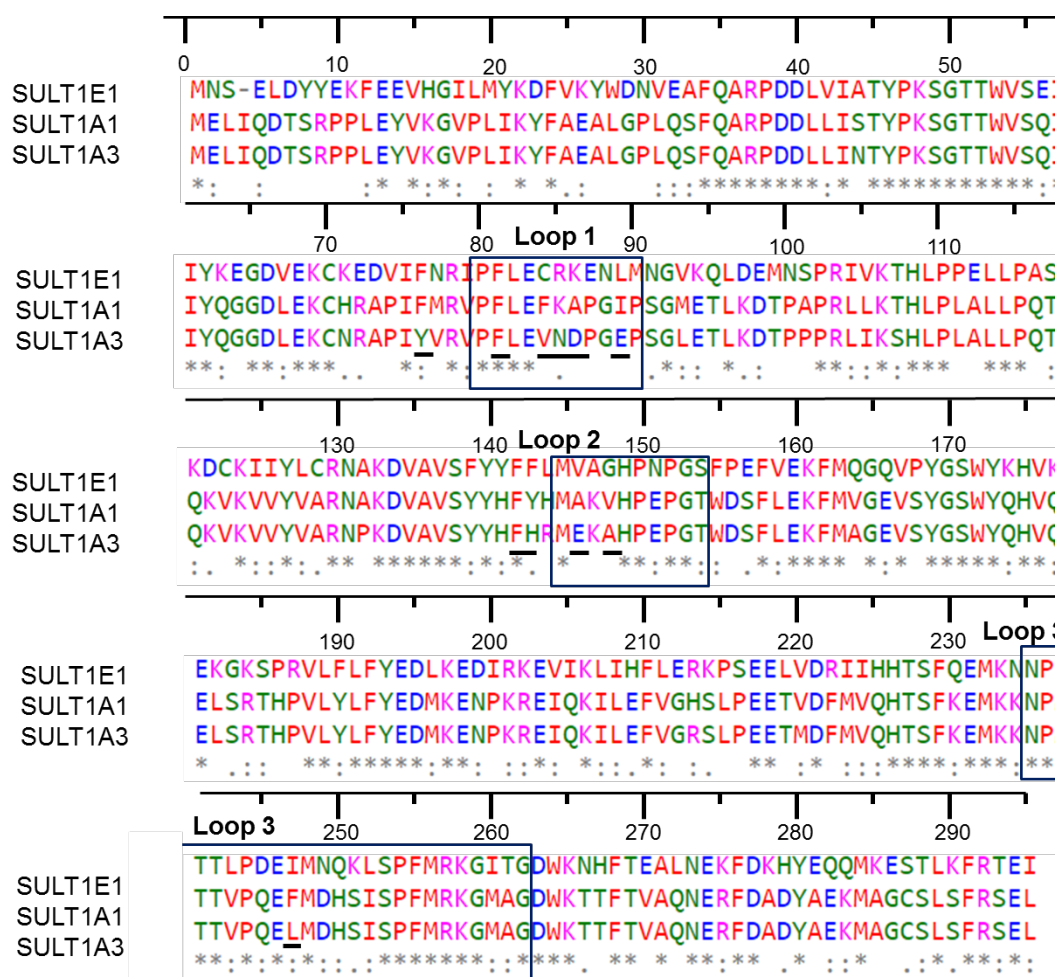


Figure A-1: Multiple Sequence Alignment (MSA) of SULT isoforms SULT1A1 (PDB 2D06), SULT1E1 (PDB 1AQU) and SULT1A3 (PDB 2A3R). Residues of Loops 1, 2 and 3 shaping the substrate pocket of SULTs are shown in black boxes. The important pocket residues are underlined. The amino acids are colored as follows: red for small and hydrophobic (A, V, F, P, M, I, L, W, Y), blue for acidic residues basic (D, and E), magenta for basic residues (R, K, excluding H), and green for hydroxyl, sulfhydryl, amine and G (S, T, Y, H, C, N, G, and Q). The Symbols under the alignment indicates the following: an asterisk (*) resembles positions of fully conserved residue, a colon (:) displays preservation between groups of strongly similar properties (scoring > 0.5 in the Gonnet PAM 250 matrix), and a period (.) shows conservation between groups of weakly similar properties (scoring =< 0.5 in the Gonnet PAM 250 matrix).

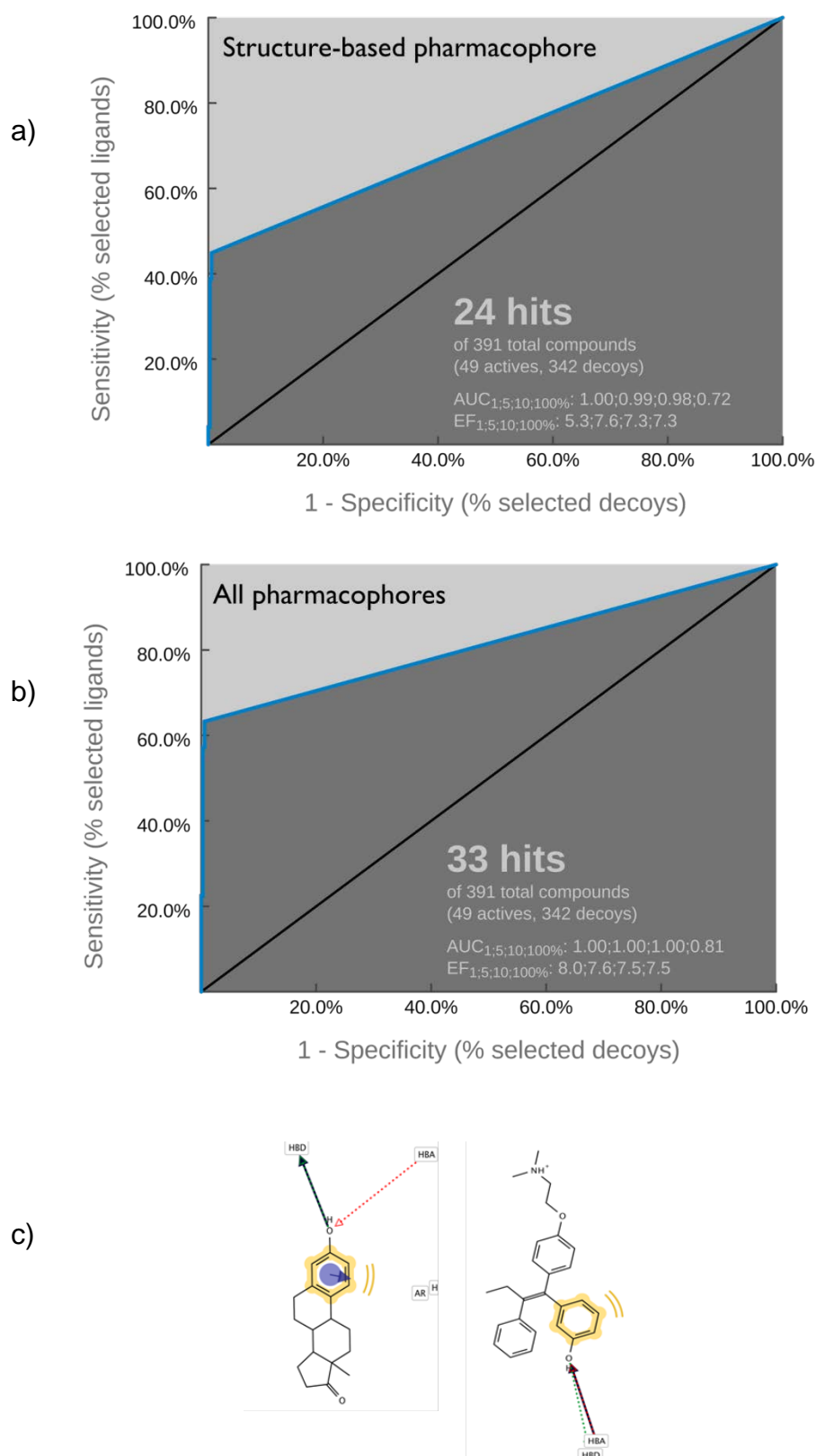


Figure A-2: Receiver operating curve (ROC) for the validation of SULT1A1 inhibitor structure-based pharmacophore (Left) and all pharmacophores simultaneously (middle); showing good early enrichment Factor (EF) of 5.3 and 8 respectively. The pharmacophore could retrieve about 40% and 60% respectively of the inhibitors and two decoys. Right: the two decoys retrieved by the pharmacophores are structurally related to E2 and there is no evidence in the literature not to be SULT1A1 binders.

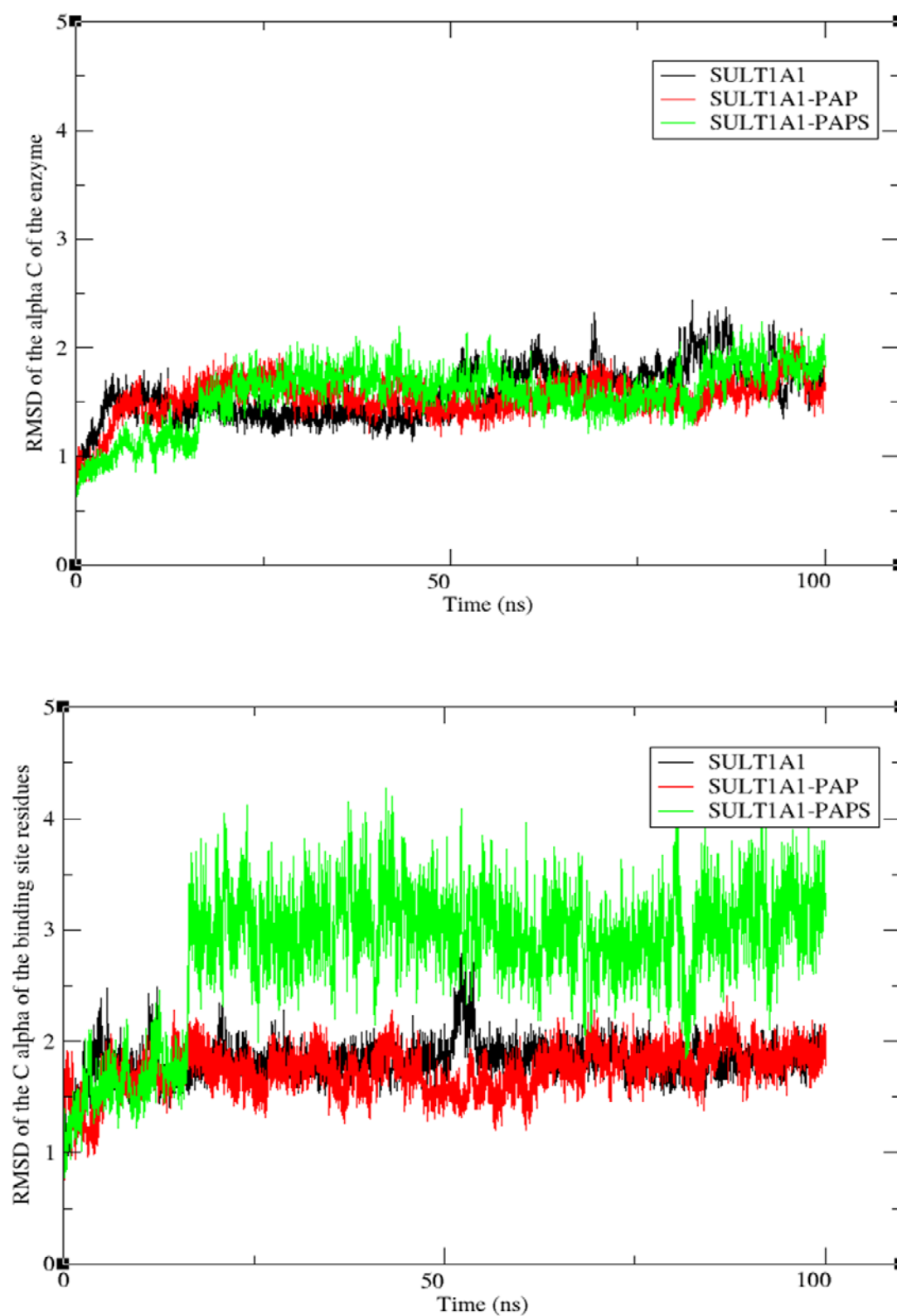


Figure A-3. Top: RMSD of $C\alpha$ of the enzyme residues (\AA), Down: RMSD of $C\alpha$ of substrate binding site residues over 100 ns simulation time from MD simulations of ligand-free enzyme complexes. System color codes are black for SULT1A1, green for SULT1A1-PAPS and red for SULT1A1-PAP.

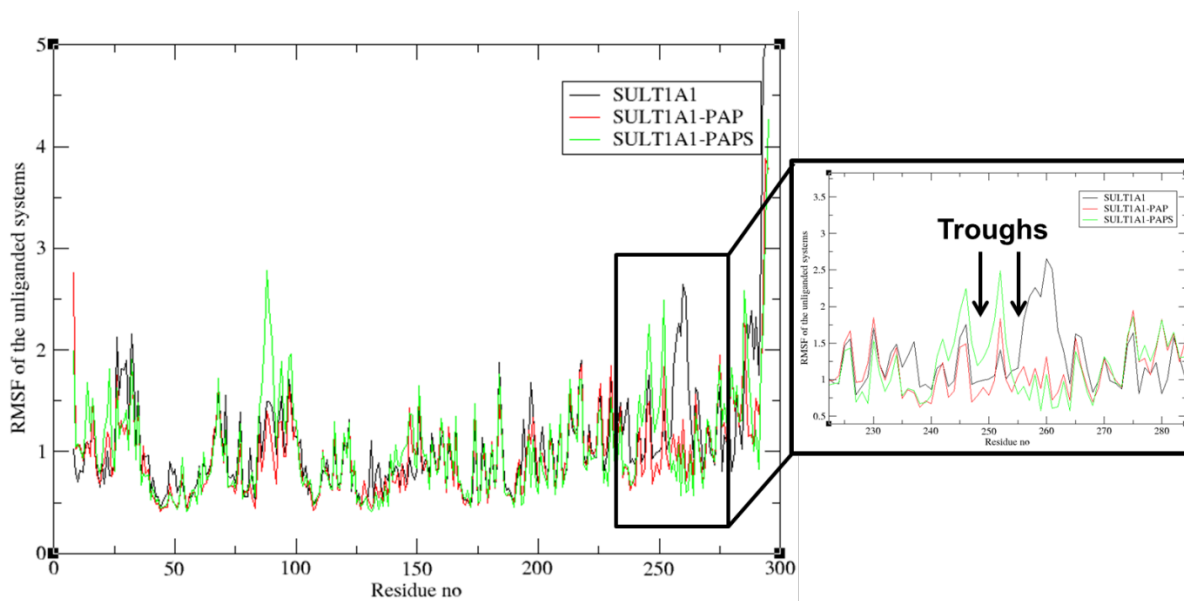


Figure A-4: Plot per-residue Root mean square fluctuation (RMSF) value (Å) of SULT1A1 (black), SULT1A1-PAP (red) and SULT1A1-PAPS (green). RMSF magnified part is showing troughs at residues Asp249 and Phe255. The total simulation time was 100 ns for each run.

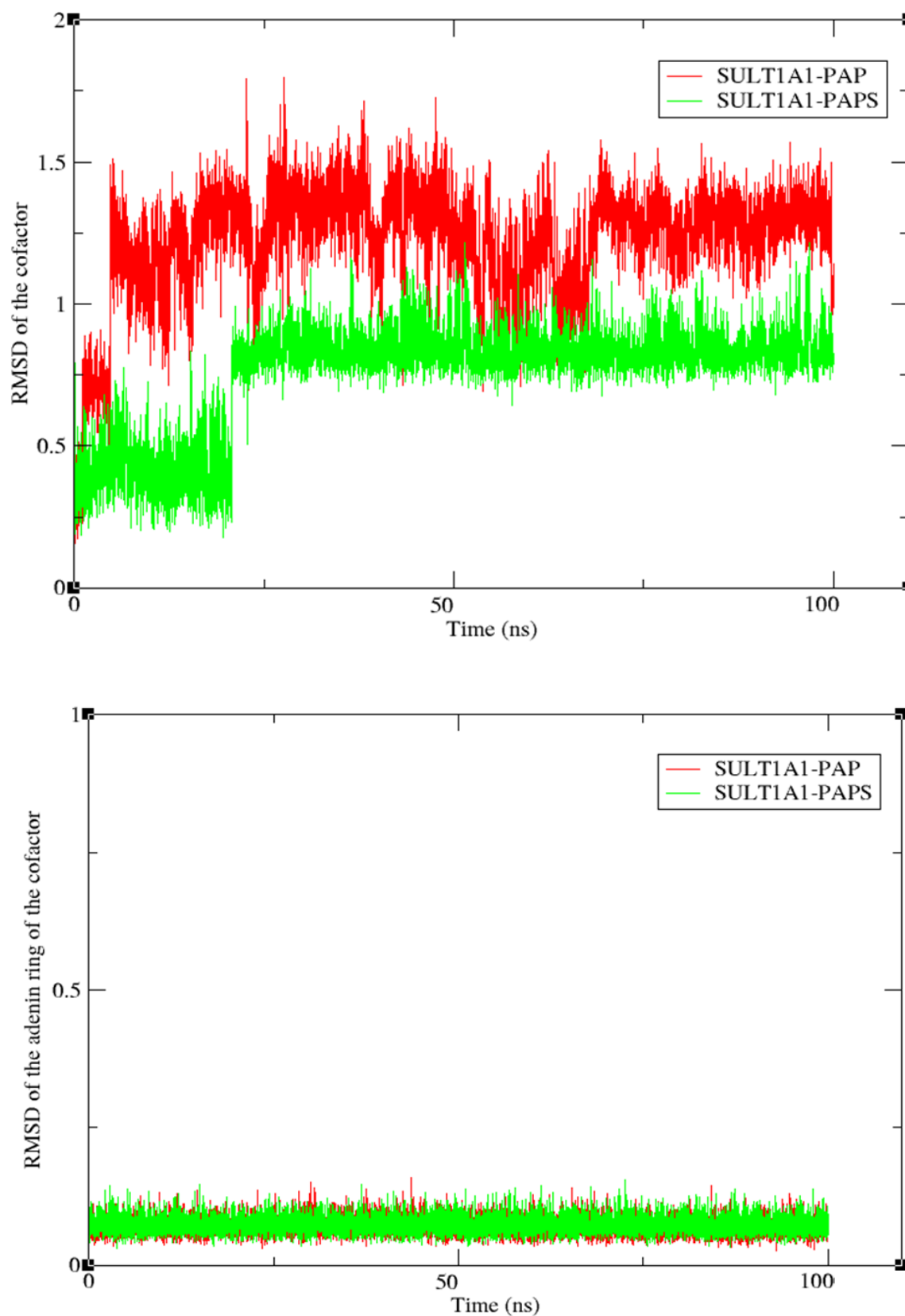


Figure A-5: Up: The RMSD of the cofactor atoms. Down: The RMSD of the adenine group of the cofactor in SULT1A1-PAPS (green) and PAP SULT1A1-PAP (red) over 100 ns study.

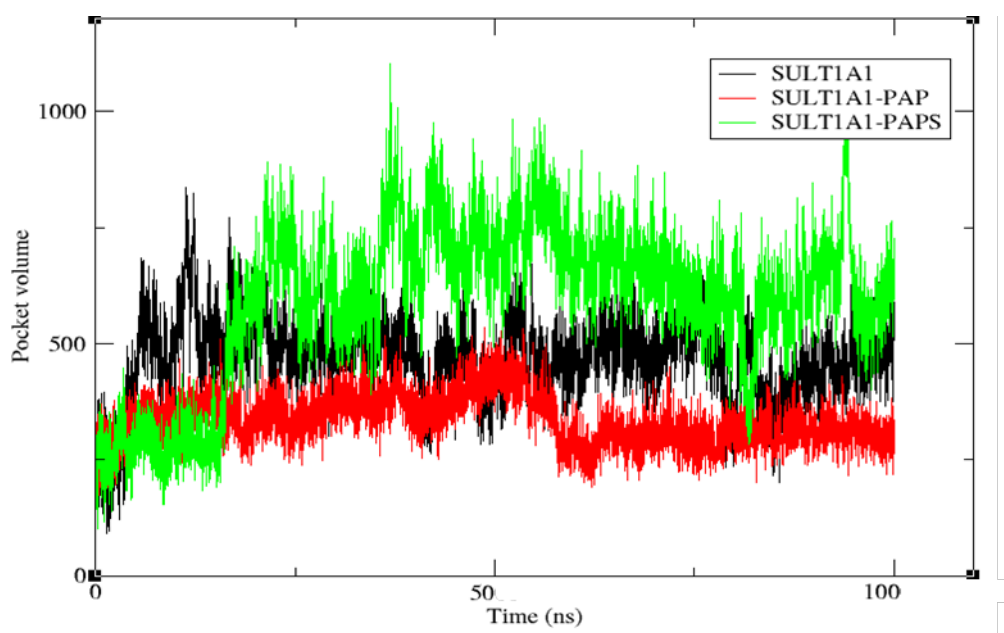
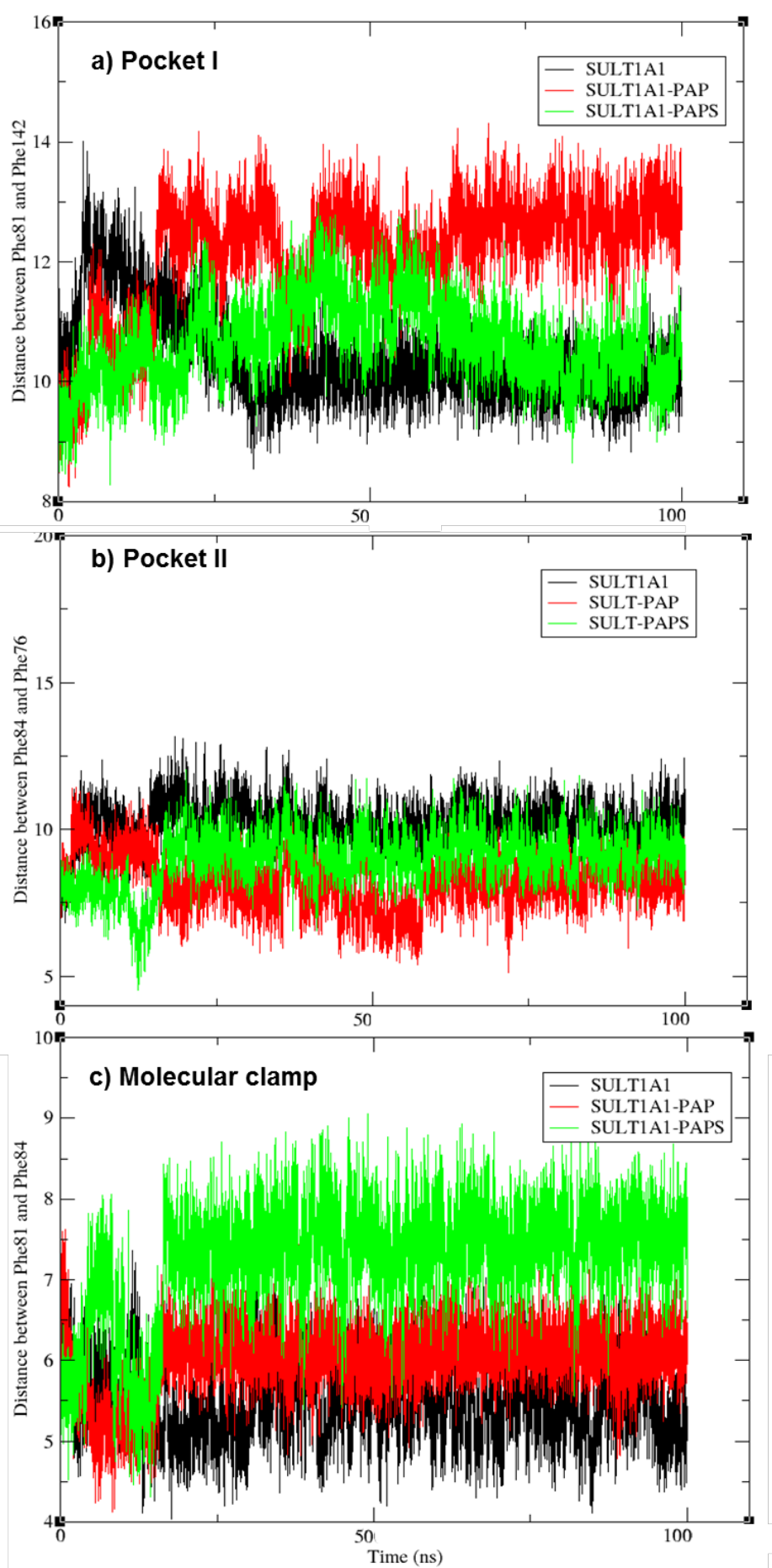
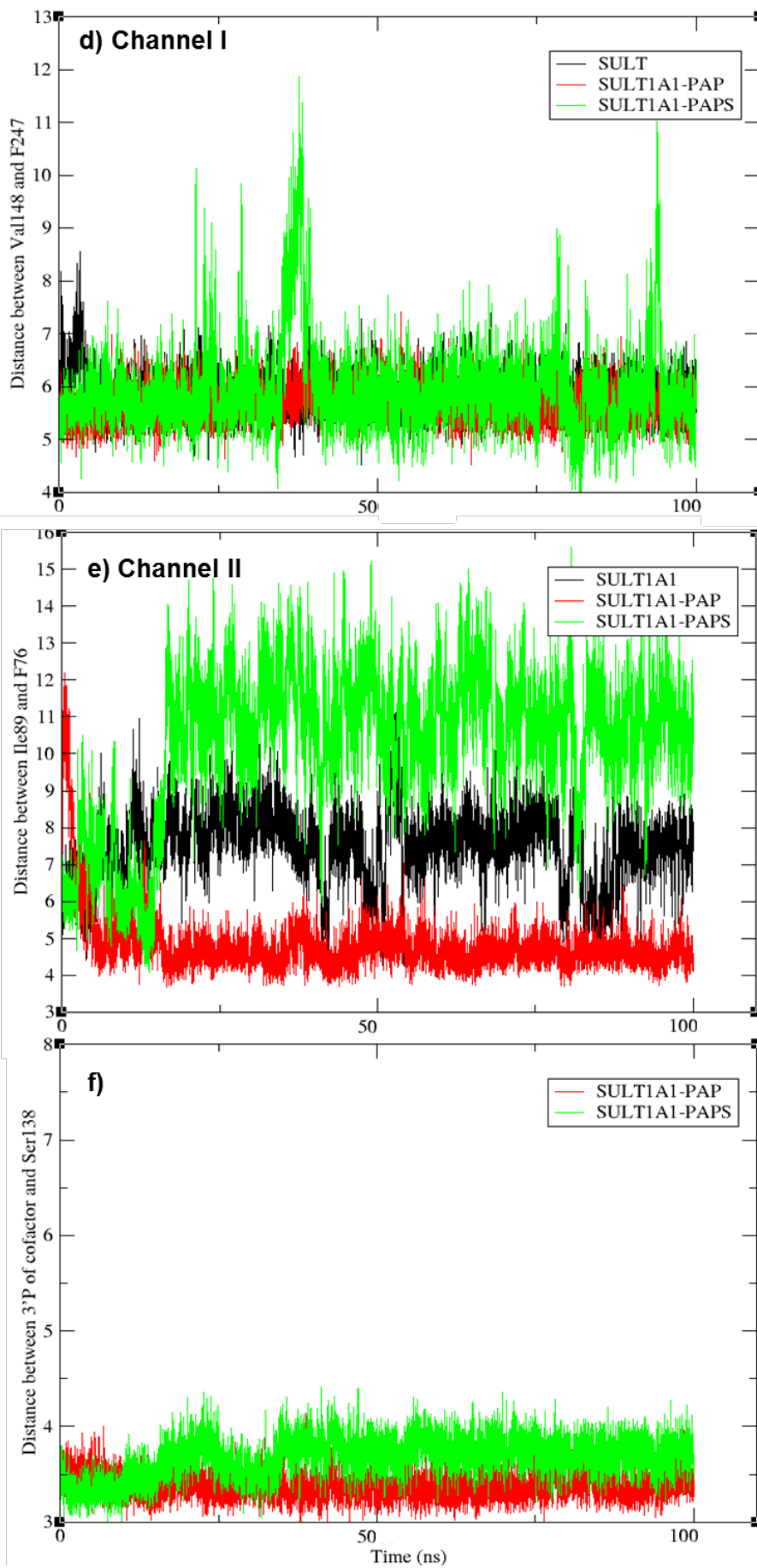


Figure A-6: Pocket volume over 100 ns MD study time for SULT1A1 (black), SULT1A1-PAP (red) and SULT1A1-PAPS (green).





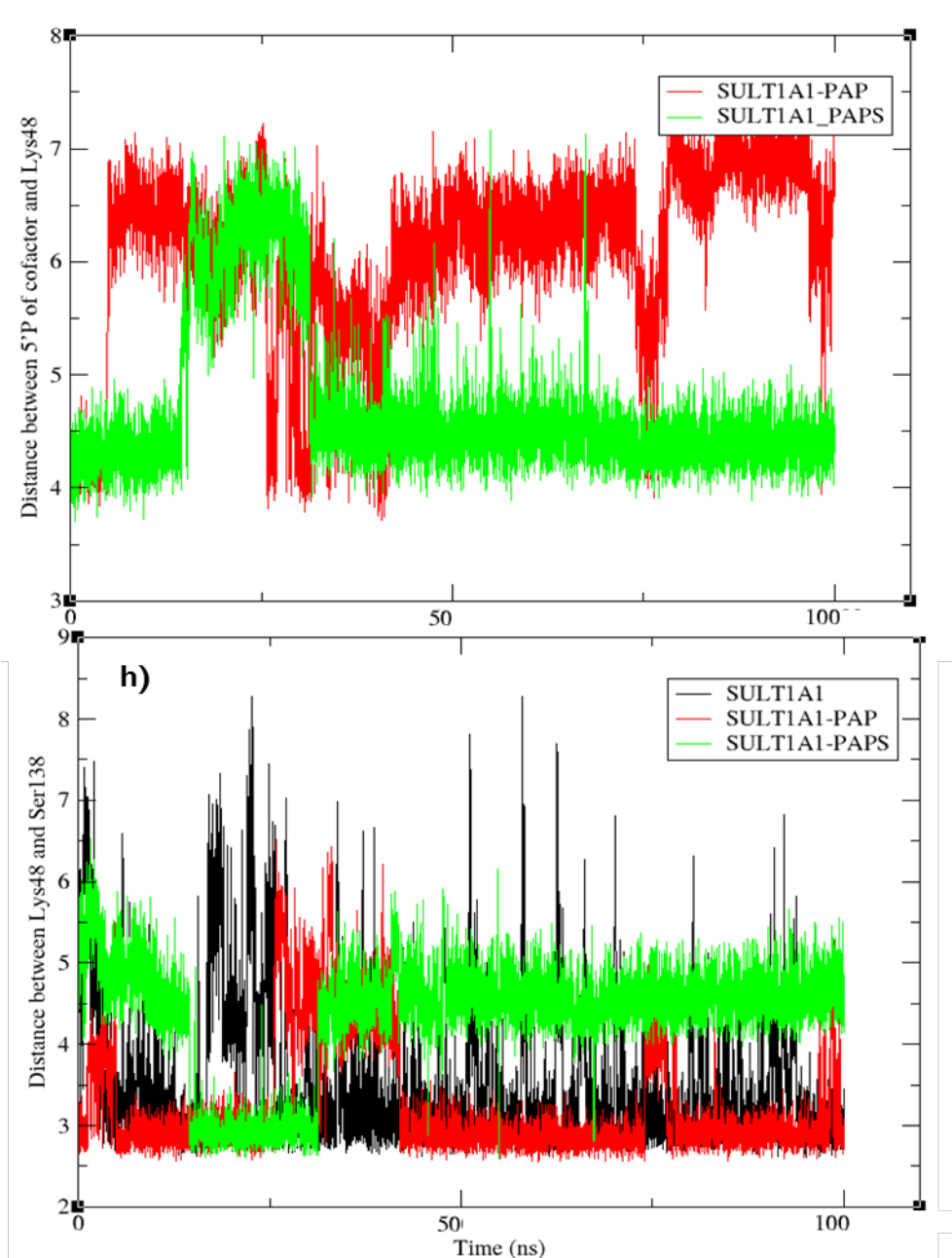


Figure A-7: Distances between center of masses of the aromatic rings of a) Phe81 and Phe142 (Pocket I), b) Phe84 and Phe76 (Pocket II), c) Phe84 and Phe81 (Molecular Clamp), the side branches of d) Val148 and Phe247 (Channel I), e) Ile89 and Phe76 (Channel II), the functional groups f) NH₂ of Lys48 and 5' PO₄ of the cofactor, g) OH of Ser138 and NH₂ of Lys48, and h) OH of Ser138 and NH₂ of Lys48 for SULT1A1 (black), SULT1A1-PAP (red) and SULT1A1-PAPS (green) for SULT1A1 (black), SULT1A1-PAP (red) and SULT1A1-PAPS (green).

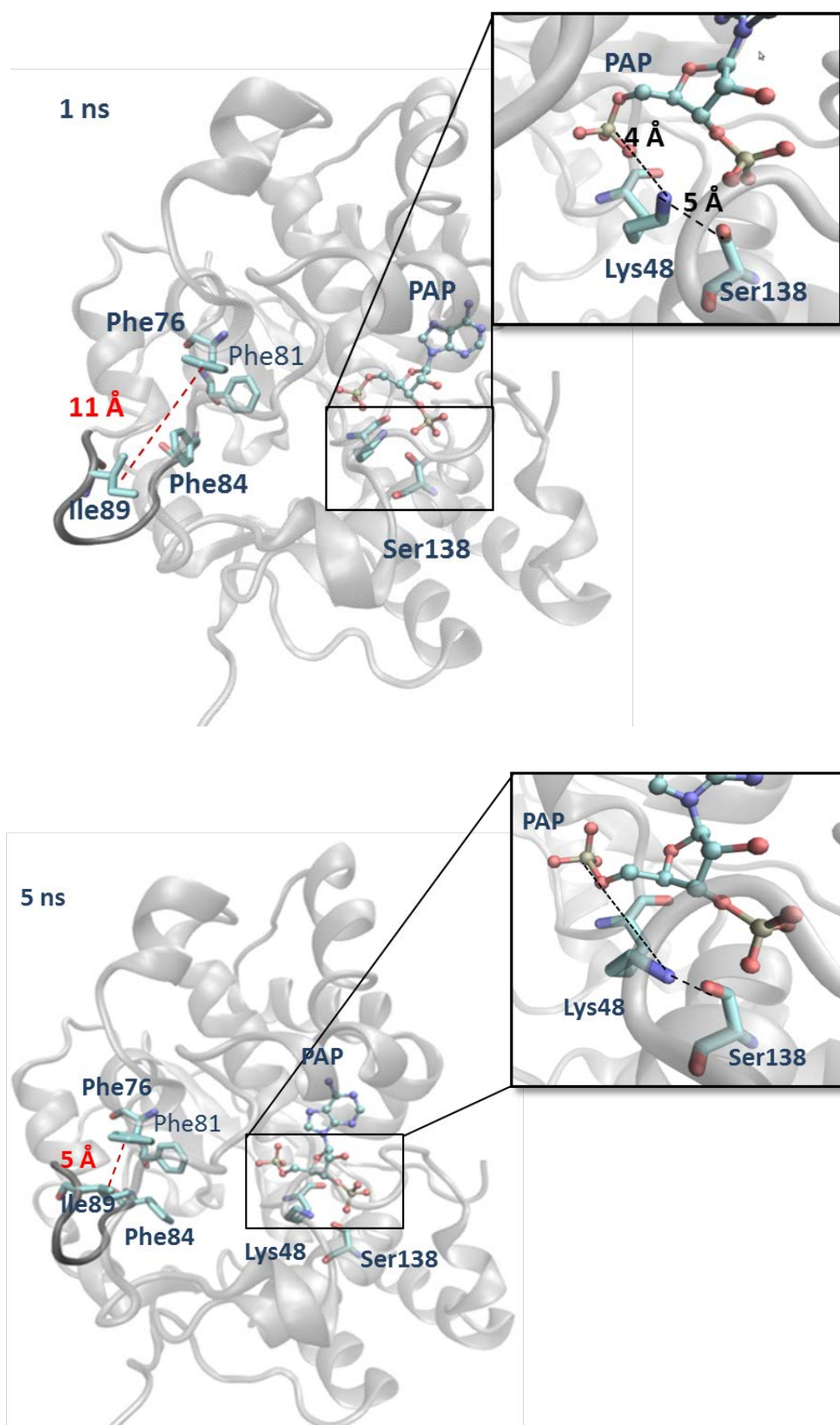
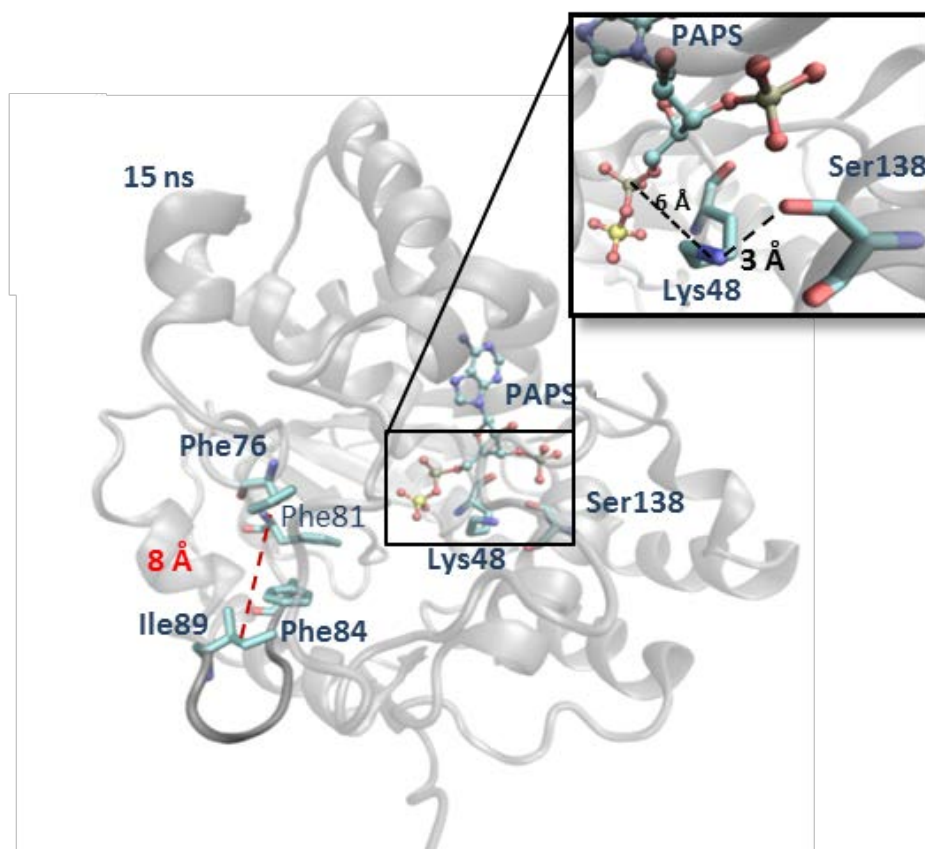
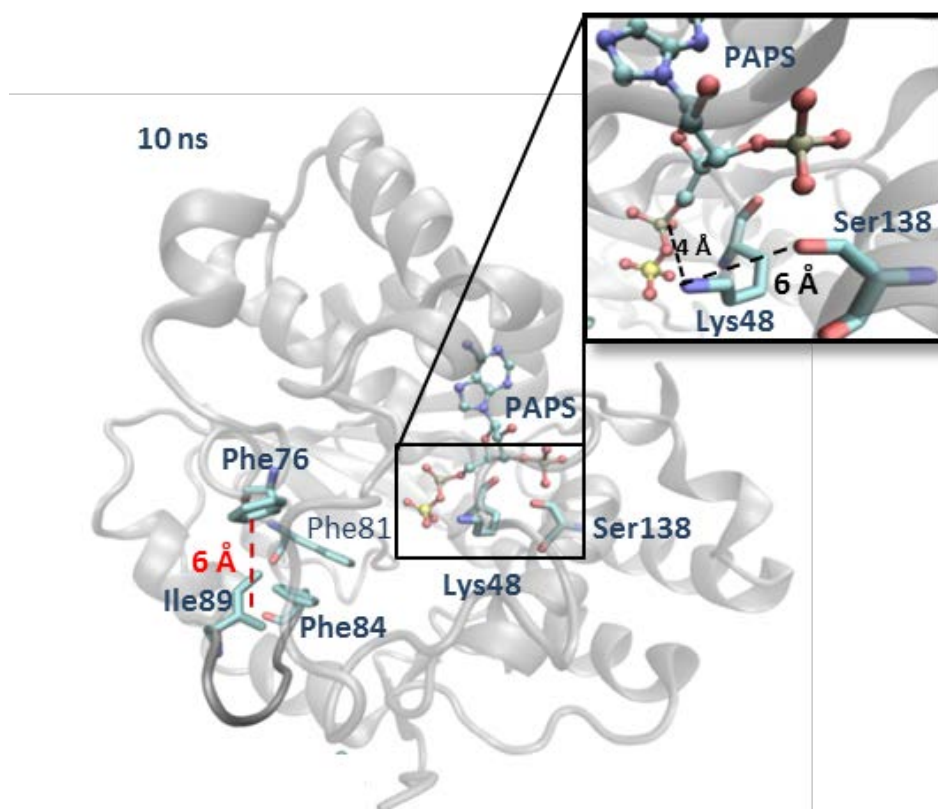


Figure A-8: Observational changes in Channel II and catalytic interface with progression of the simulation time in SULT1A1-PAP system. Channel II closure and catalytic interface rearrangement with progression of the simulation time in SULT1A1-PAP system. Lys48 moved away after 5 ns from cofactor stabilization “ready” state to be stabilized by Ser138 “stable” state for the rest of simulation life. The movement of Lys48 was coordinated with Channel II closure and pocket volume reduction.



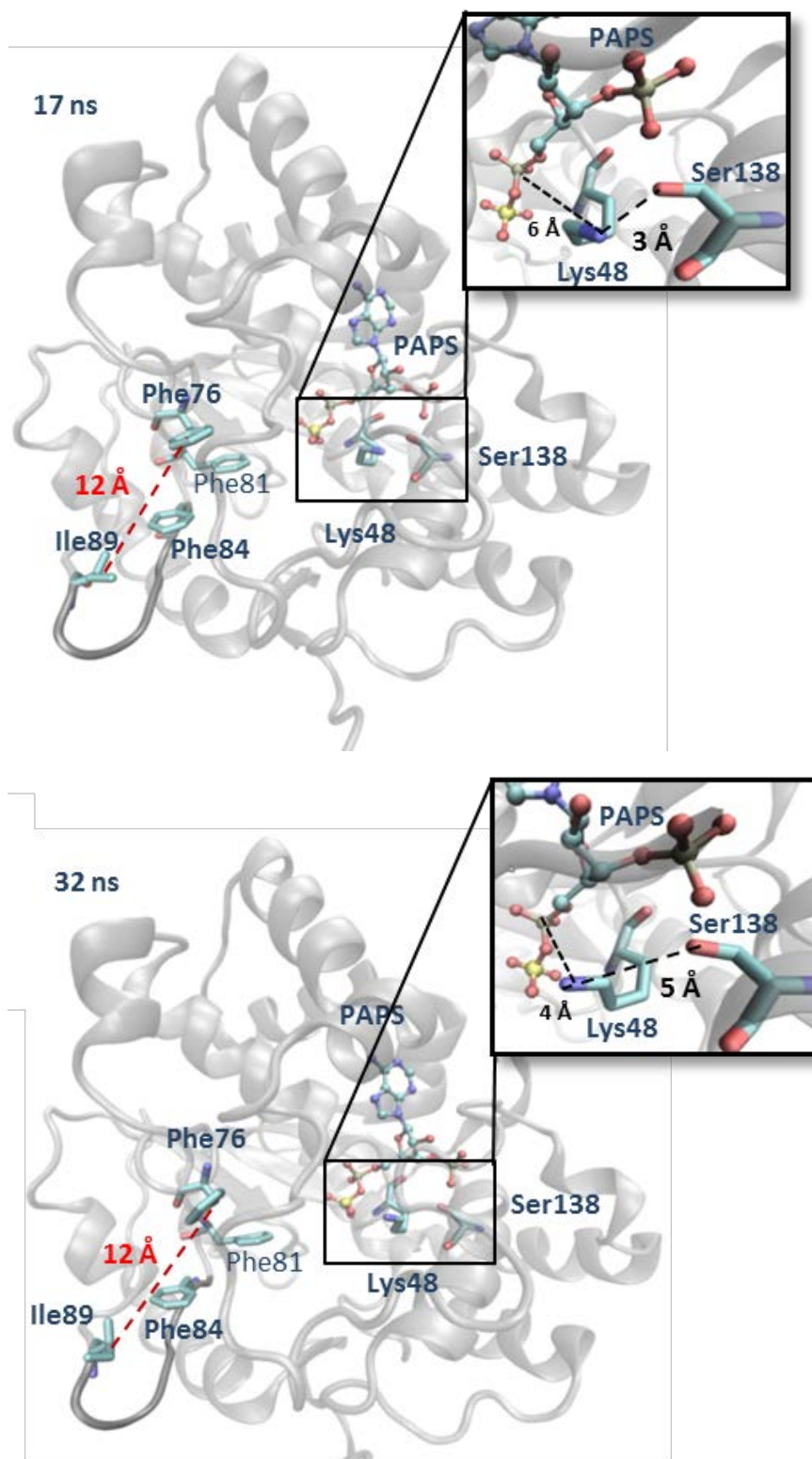


Figure A-9: Observational changes in Channel II and catalytic cleft with progression of the simulation time in SULT1A1-PAPS system. Lys48 demonstrated similar movement like PAP system with coordination with Channel II opening at 15 ns. After 30 ns, Lys48 moved back to regain the favorable interaction with the cofactor.

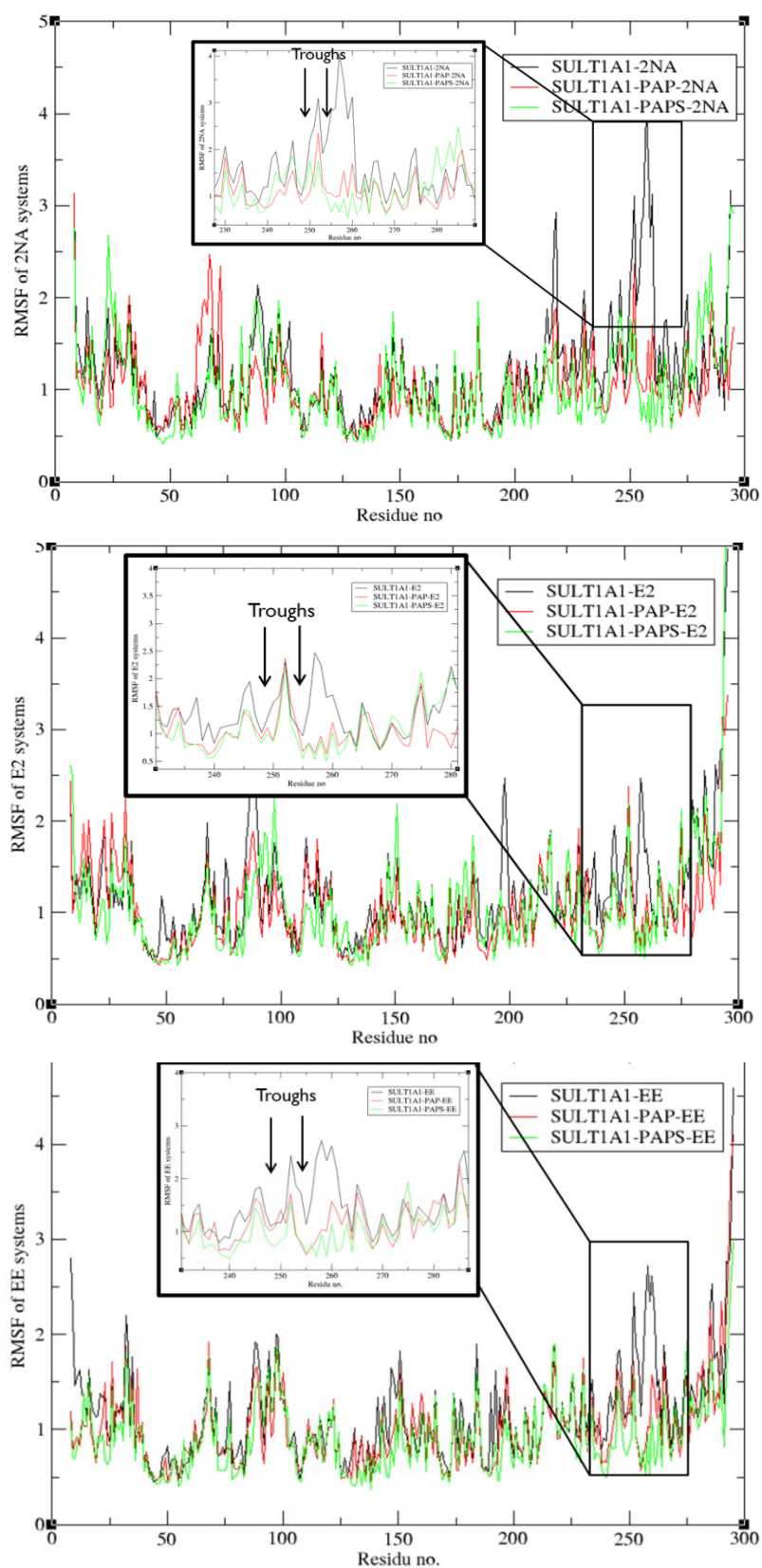
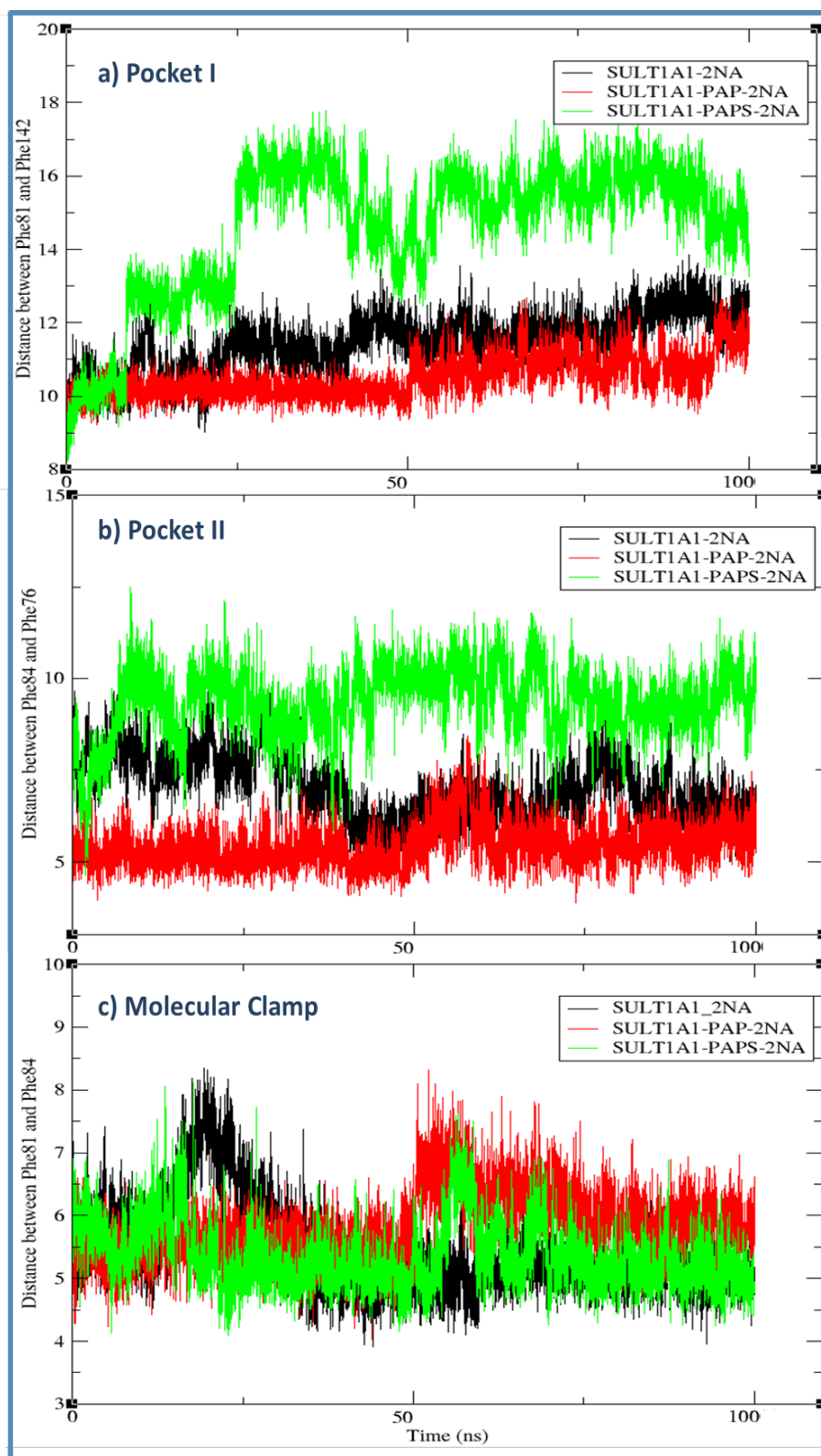
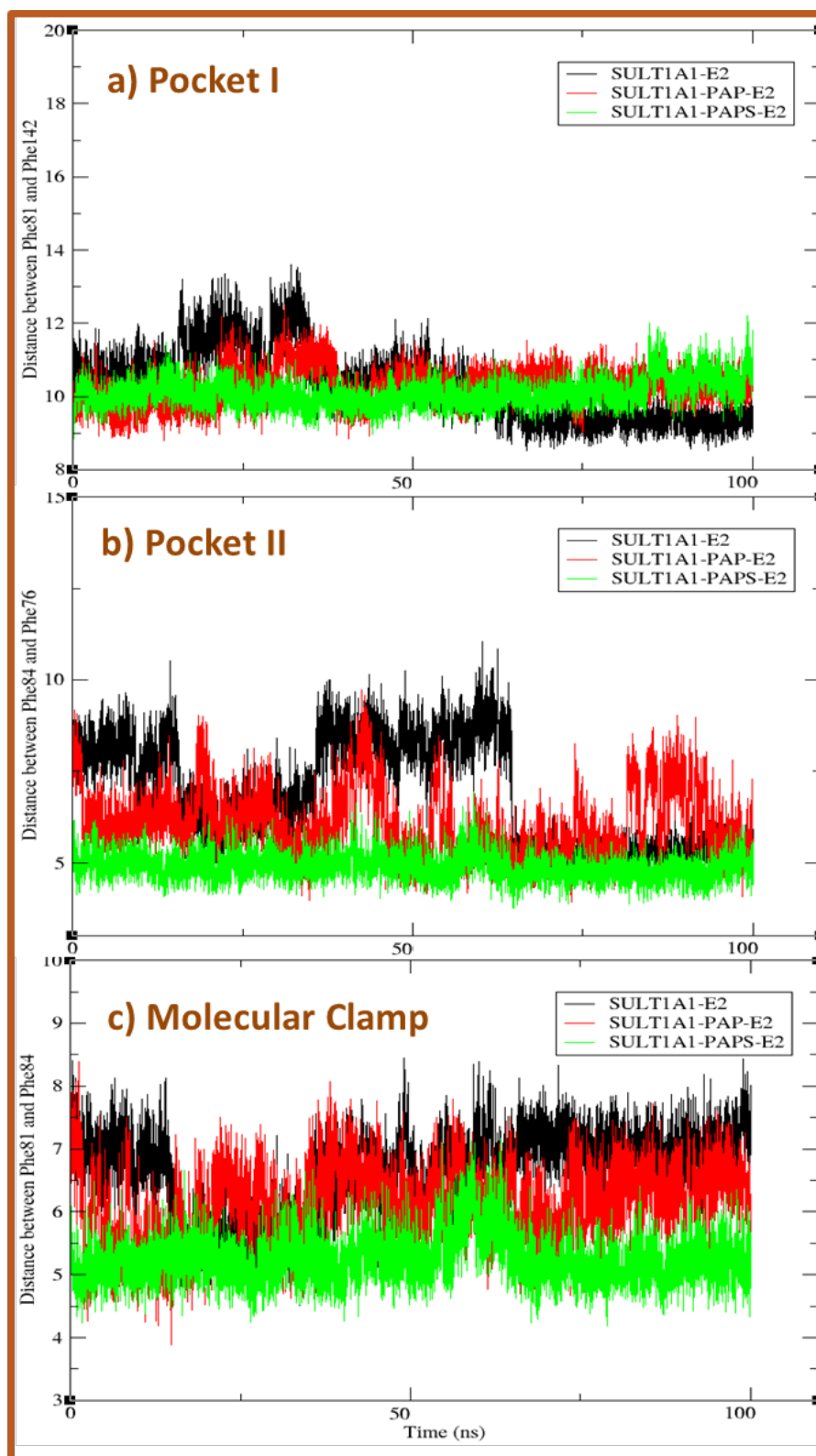


Figure A-10: RMSF plots of the enzyme residues SULT1A1 (black), SULT1A1-PAPS (green) and SULT1A1-PAP (red) for several ligand-bound enzymes. RMSF magnified part is showing the trough at residue Phe255.





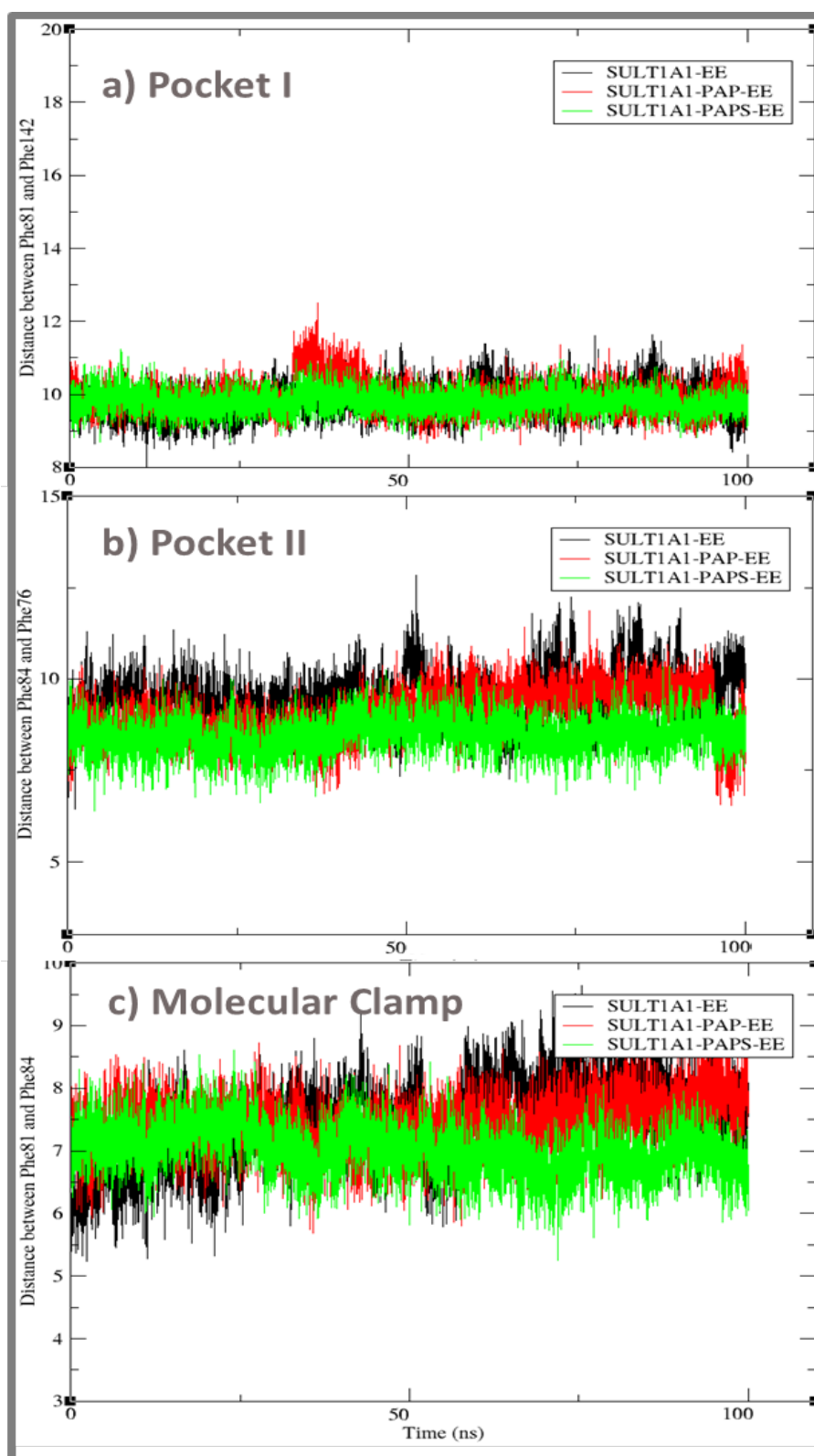
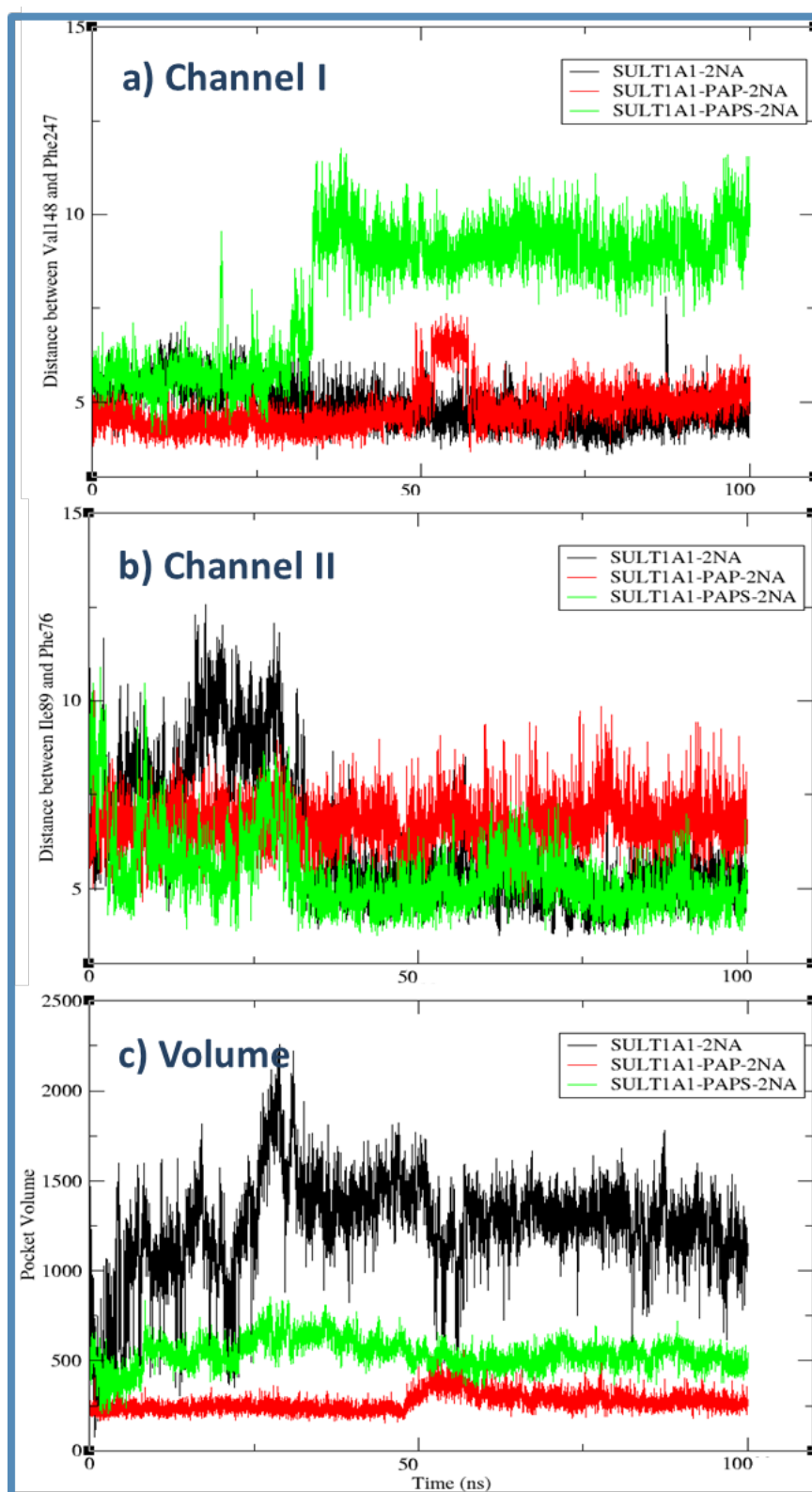
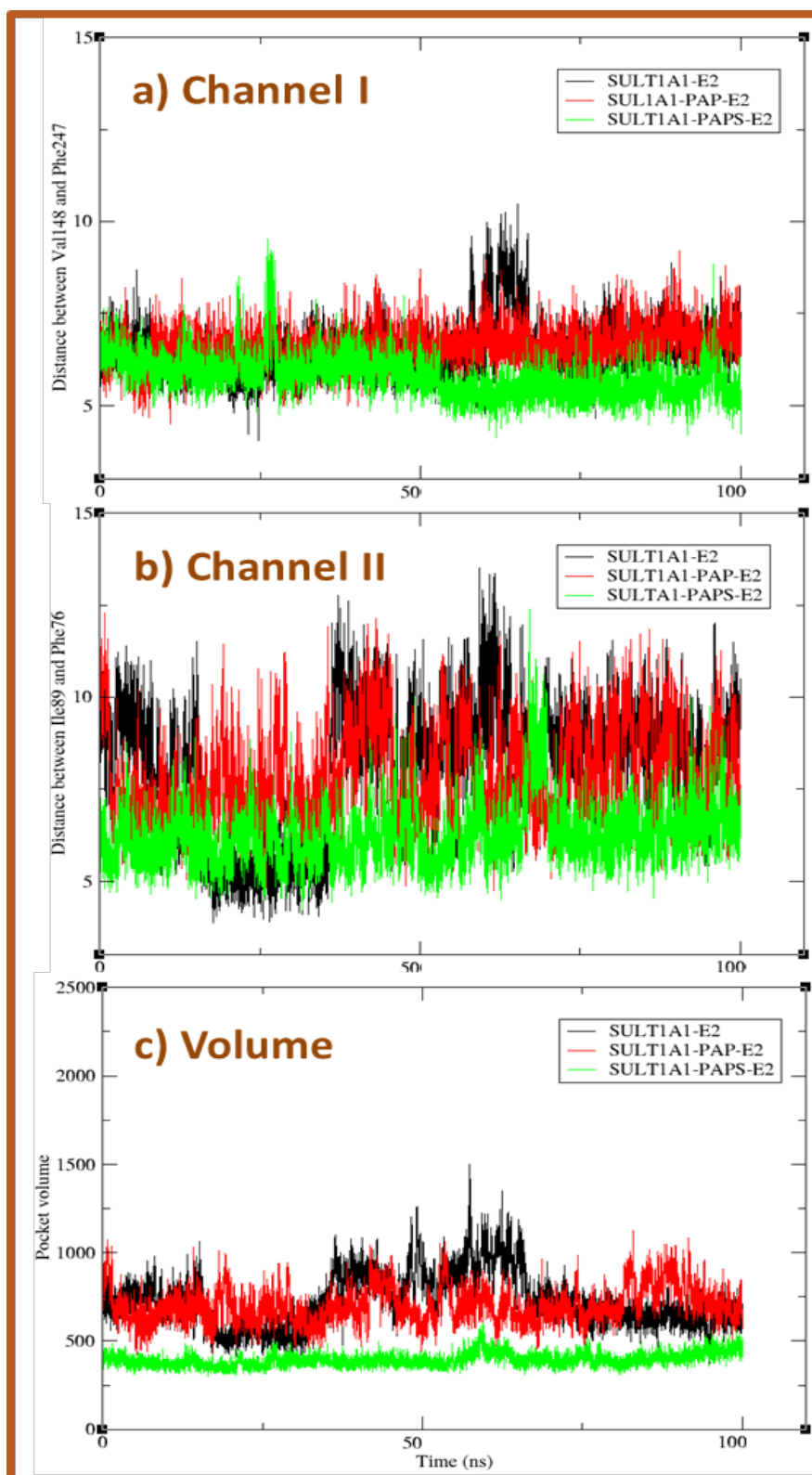


Figure A-11: a) Distance between center of masses of the aromatic rings of Phe81 and Phe142 that resembles Pocket I. b) Distance between center of masses of the aromatic rings of Phe84 and Phe76 that resembles Pocket II. c) Distance between center of masses of the aromatic rings of Phe84 and Phe81 that resembles Molecular Clamp over 100 ns MD study time for SULT1A1 (black), SULT1A1-PAP (red) and SULT1A1-PAPS (green) of 2NA (Blue), E2 (Brown) and EE (Grey) systems.





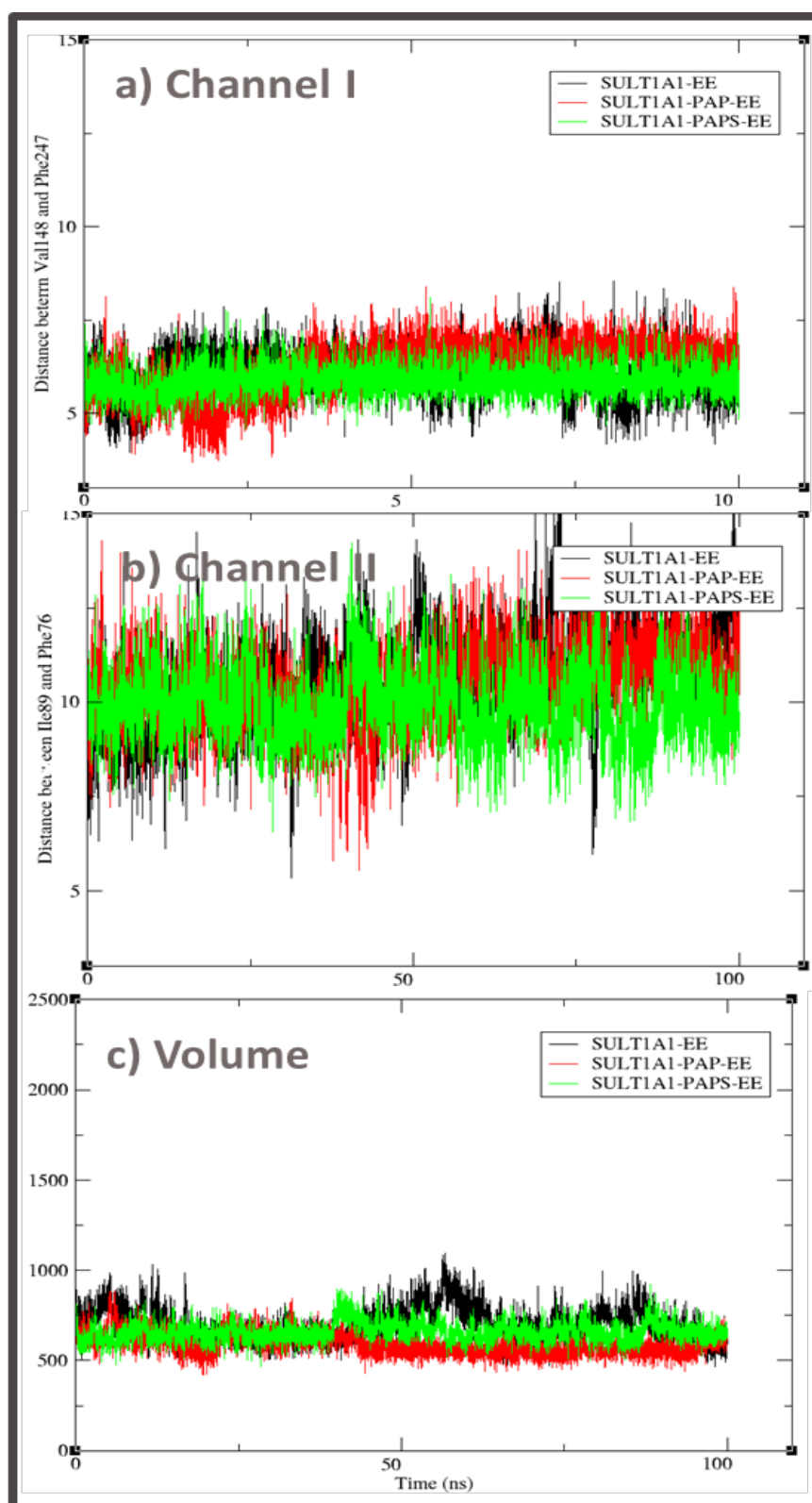


Figure A-12: a) Distance between center of masses of the side branches of Val148 and Phe247 that resembles Channel I. b) Distance between center of masses of the side branches of Ile89 and Phe76 that resembles Channel II. c) Pocket volume over 100 ns MD study time for SULT1A1 (black), SULT1A1-PAP (red) and SULT1A1-PAPS (green) of 2NA (Blue), E2 (Brown) and EE (Grey) systems.

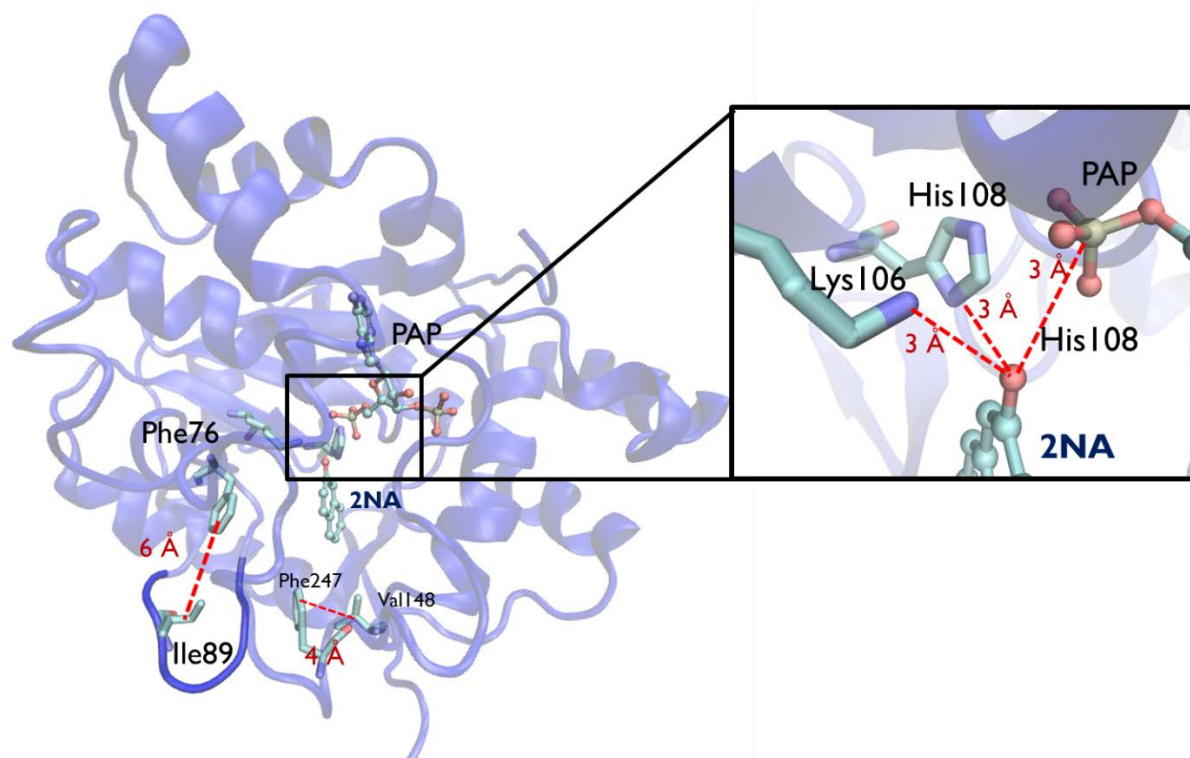


Figure A-13: SULT1A1-PAP-2NA average structure showing the dead-end complex with closed Channel I and II. The catalytic interface is enlarged for clarity in the box showing well-established hydrogen bonds with the cofactor and the catalytic residues.

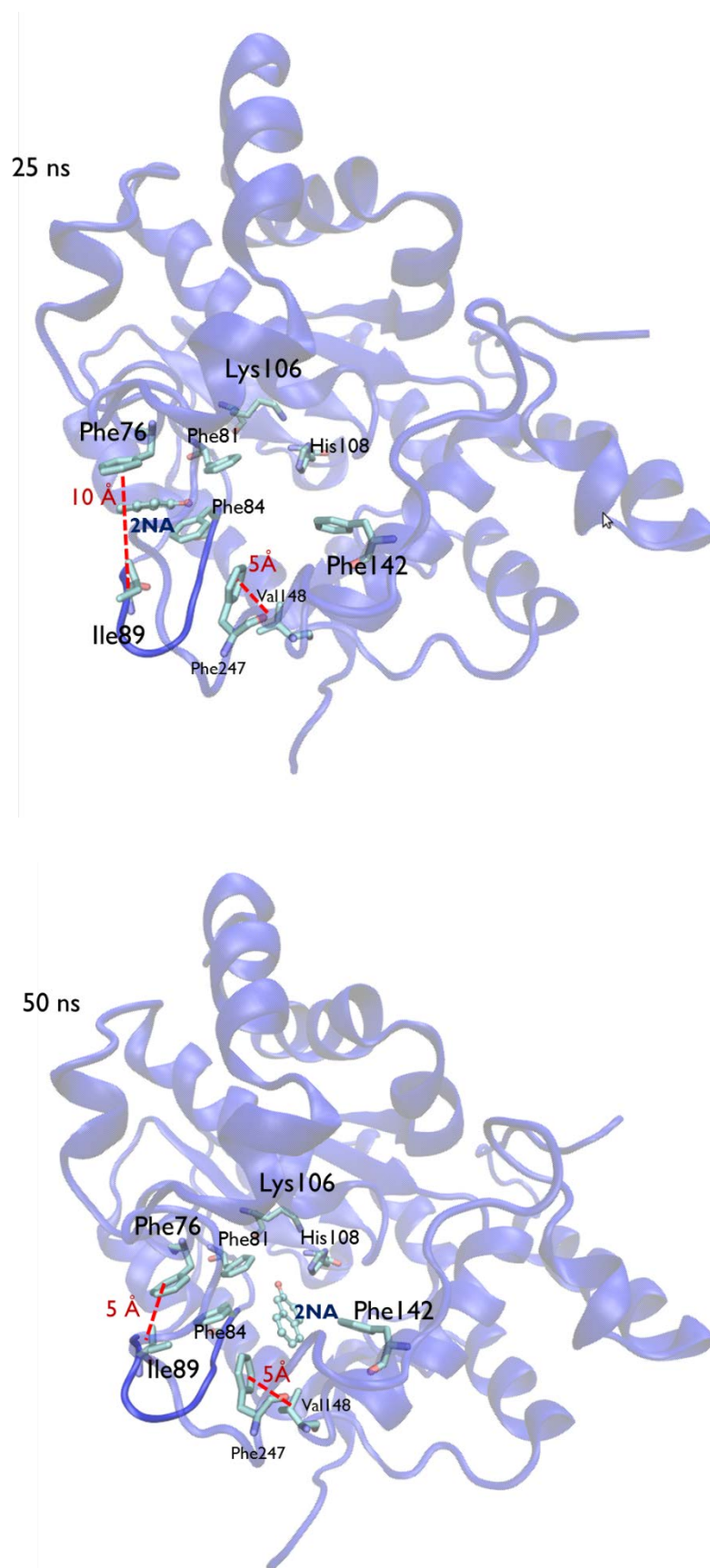


Figure A-14: Channel II observational changes illustrating opening event at 25ns of the simulation study of SULT1A1-2NA system.

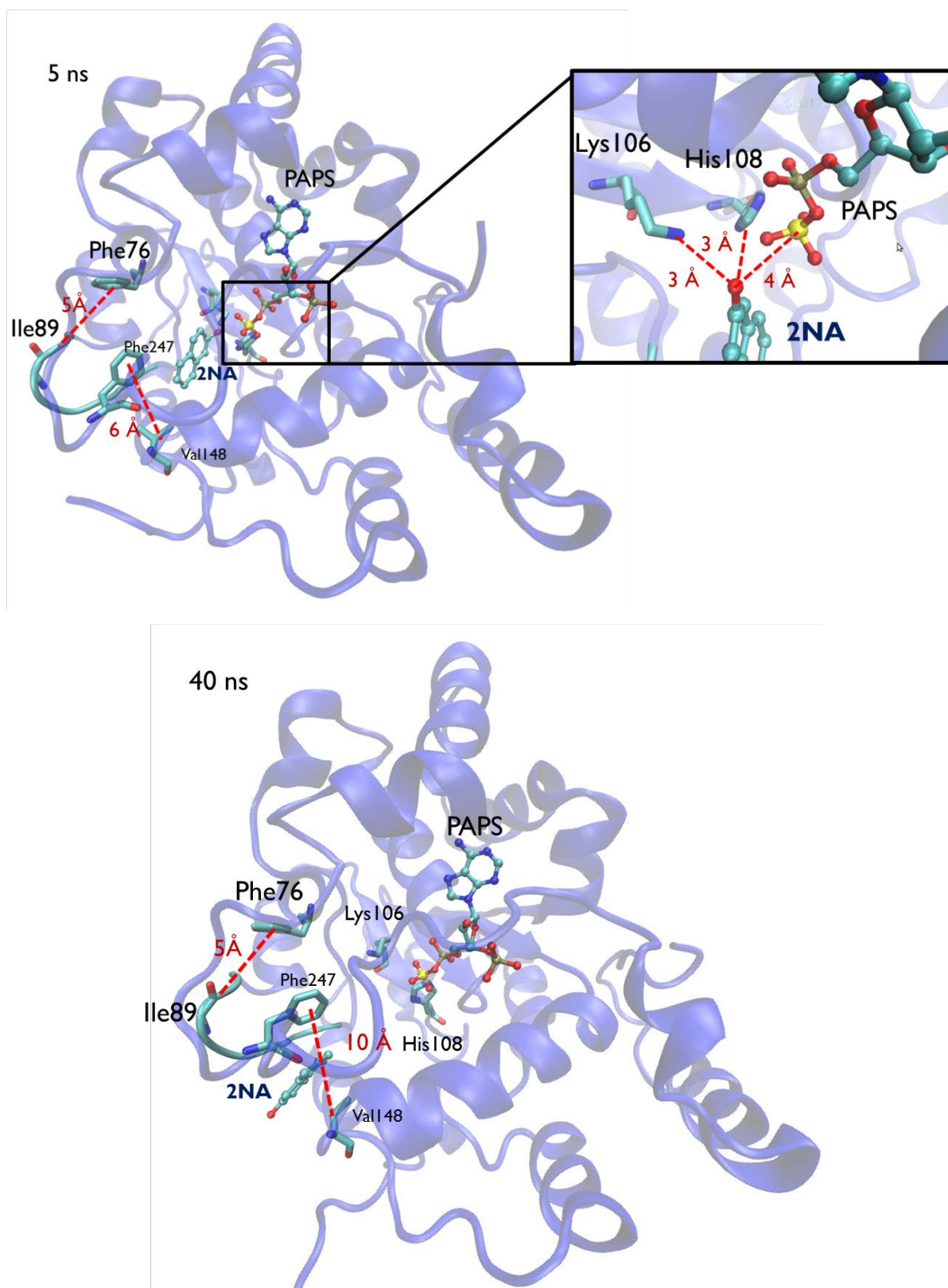


Figure A-15: SULT1A1-PAPS-2NA evolutionary events showing binding site evolution during simulation. The catalytic interface is enlarged for clarity in the box showing well-established hydrogen with the cofactor and the catalytic residues. At 40 ns Channel I was opened and the molecule exited through it.

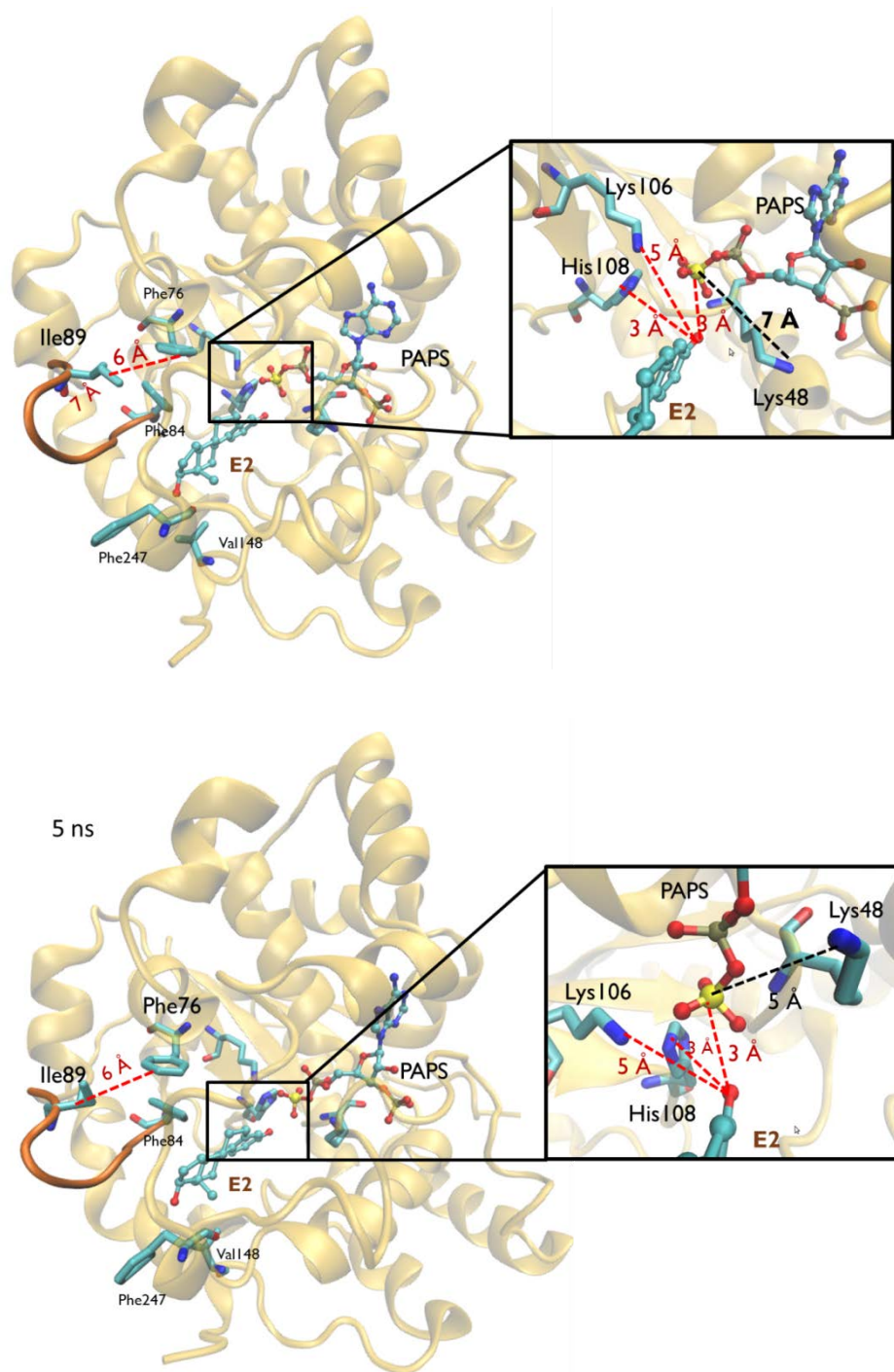


Figure A-16: Up: average structure of SULT1A1-PAPS-E2 presenting closed pocket II and channel II. Lys48 shows a steady conformation away from the bridging oxygen. Down: Observational event of Lys48 ready conformation happened at 5 ns and 95 ns.

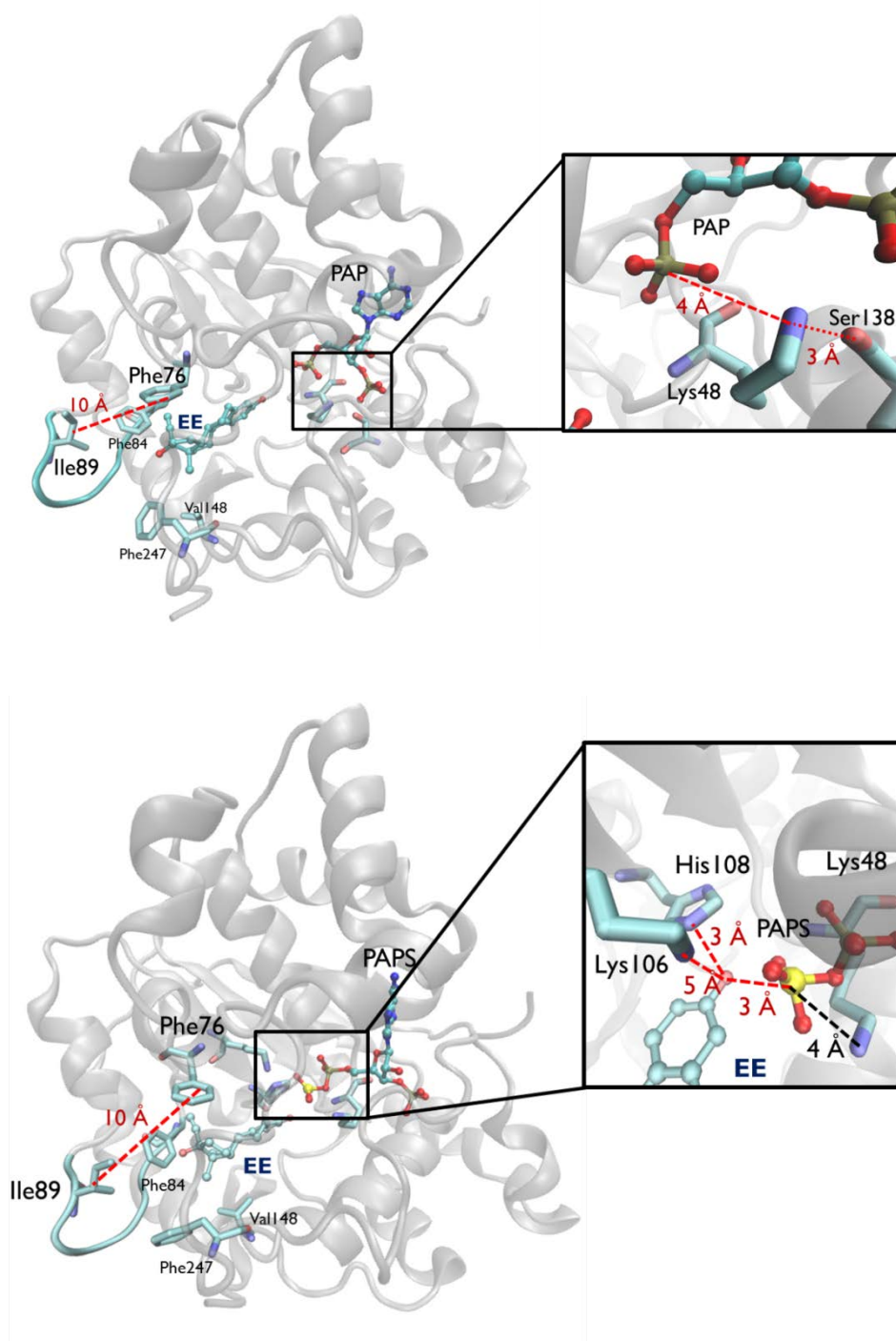


Figure A-17: Left: SULT1A1-PAP-EE average structure showing ethynyl group interaction with Pocket II residues Phe76 and 84. The catalytic interface was explained in an enlarged box showing the mid-way position of Lys48. Right: SULT1A1-PAPS-EE average structure showing Michaelis-Menten complex of EE. The catalytic interface is enlarged for clarity in the box showing a well-established tetrahedral hydrogen bond interaction.

Summary

Human cytosolic sulfotransferase 1A1 (SULT1A1), having the broadest specificity among sulfotransferase family, mediates the metabolism of myriads of small phenolic endobiotics and xenobiotics e.g. para nitrophenol. It catalyzes a sulfonate group transfer from a universal sulfonate donor: 3'-phosphoadenosine 5'-phosphosulphate (PAPS) to an acceptor (Lig), leaving behind the non-sulfonated cofactor (PAP) and a sulfonated product (LigS). Having two binding sites randomly filled, SULT1A1 can be found *in vivo* in many complexes of different combinations of cofactor and ligand; apoform (SULT1A1), binary complexes (SULT1A1-PAP, SULT1A1-PAPS and SULT1A1-Lig) and ternary complexes (SULT1A1-PAP-Lig and SULT1A1-PAPS-Lig). SULT substrates are classified into three categories based on their binding affinities to various complexes: Neutral, positive and negative cooperativity substrates. Positive cooperativity substrates are ligands that have a high binding affinity for complexes containing the cofactor. Yet, the intermolecular enzyme-ligand interactions responsible for this positive cooperativity behavior are still unclear. In addition, there are no structural studies on complexes of different cofactor and ligand combinations and most of them are not crystallized.

The aim of this work was to understand the interactions responsible for positive cooperativity behavior of SULT1A1 binders. To reach this goal, a combination of *in silico* tools was used. First, all crystallographic data of SULT1A1 were collected and compared to each other and to the structural data of other highly related enzyme members of overlapped specificity with SULT1A1. Second, 3D pharmacophores were employed to rationalize the positive cooperativity behavior of crystalized ligands. 3D models were generated from the structural data (structure-based) and from binder information collected from literature (ligand-based). Third, key complexes in the sulfonation cycle of SULT1A1, for which no structural information was available, were modeled and used for an MD study.

The first step of the study, the structural investigation of all available SULT1A1 crystallographic data, was conducted to highlight the flexible regions. Structural analysis and calculated substrate binding site volumes of various SULT1A1 crystal structures hosting various binders confirmed the ligand-size dependent nature of the enzyme-substrate binding site, with the direct involvement of the gating loop (residues 85 - 90) conformation determining substrate binding site shape and volume. Structural investigation of SULT1A1 structure and other highly related siblings with overlapped

substrate specificity and of the same family to identify the substrate binding site characteristics that are responsible for substrate specificity. Despite having binding site sequence and structural homology of high (SULT1A3) and low (SULT1E1) degree to SULT1A1, the MIF maps revealed completely different electronic environment of the substrate cleft which reflects the unique substrate specificity of each isoform. Data support that SULT1A1 pocket characteristics (ligand-induced size, ligand-induced geometry, and hydrophobic nature) defines its broad substrate specificity.

The second step was the construction of a combination of structure-based, ligand-based, and docking-based 3D pharmacophores to explore the binding features of SULT1A1 ligands. Structure-based pharmacophores were derived from structural data of strong binders, whereas only substrates were available as co-crystallized ligands. Next, ligand-based pharmacophores were generated from experimentally determined SULT1A1 potent inhibitors. Lastly, one of the highly active substrates (3,3-diiodothyronine (T2)) and inhibitors (Ethinyl estradiol (EE)), with no available crystallographic data, were docked into the enzyme and subsequently docking-based 3D pharmacophores were developed for the most plausible poses. As retrospective validation, about 400 compounds (substrates, mixed substrates/inhibitors, inhibitors and decoys) have been screened through the pharmacophore filters to evaluate their ability to specifically retrieve the desired compounds. 107 SULT1A1 binders covering several activity classes and different chemical scaffolds (35 substrates, 26 mixed substrate/inhibitor, and 46 inhibitor) were successfully retrieved by virtual screening through the generated pharmacophores. The generated model showed good discriminative power to differentiate between inhibitors (that fit in one of inhibitors pharmacophores), substrates (that fit in one of substrates pharmacophores and not in inhibitors pharmacophores), and mixed substrates/inhibitors (that fit in one of inhibitors pharmacophores and in substrates pharmacophores) describing the essential features for enzyme binding (70% and 60% for substrates and inhibitors, respectively). However, no mechanistic explanation for positive cooperativity behavior could be found using these methods. Accordingly, it was necessary to mechanistically investigate binding behavior in a more detailed way.

In order to consider the flexibility of the ligand-induced substrate binding site, molecular dynamics (MD) were employed in the third part of this study to get more insights about the dynamics of the missing complexes of the sulfonation cycle. Previously published studies on SULT used MD to sample the conformational space of an empty substrate binding site in cofactor-free and cofactor-containing complexes^{21, 125, 171}. Despite the importance of data provided by those studies, the induced volume and architecture of

the substrate binding site can never be practically sampled without the existence of a ligand. Therefore, 100 ns molecular dynamic (MD) simulations were conducted for SULT complexes in the presence of neutral and positive cooperativity substrates. These simulations were useful to assess the stability of SULT1A1 complexes in different states of cofactor and ligand binding as well as bound to variety of different binders. MD combined with substrate binding site volume analysis suggests a relation between the size of the ligand, pocket dimensions and hence molecule substrate/inhibitor characteristics. Small hydrophobic binders (such as 2-naphthol (2NA)) were found to induce conformational reduction to the substrate binding site volume to reach 250 Å³. The positive cooperativity binding can be explained by the tight hydrophobic interactions with Phe81, Phe142, Phe76, Phe84, and Ile89 in cofactor-containing complexes. On the contrary, binding of small-sized substrates to apoenzyme destabilized the cofactor part of Loop 3 by distorting aromatic π -stacking between Tyr240 and Phe255. These findings rationalize the lower binding affinity of positive cooperativity substrates towards cofactor-free complex. When bigger positive cooperativity binders such as ethinyl estradiol (EE) are accommodated, Phe76 and Phe84 flexibility contribute to the opening of Pocket II and extend the hydrophobic system, allowing stacking interactions with the ligand and triplicating the pocket volume. This extra pocket opening partially “locks” large hydrophobic binders inside, even after sulfonation and might be responsible for their inhibitory effect.

In summary, substrate binding site flexibility is responsible for the ligand-induced volume and shape alterations. Our work contributes to the general understanding of SULT1A1 broad specificity and low selectivity. Since all SULT members are structurally conserved to a great extent, insights obtained by studying SULT1A1 can be eventually extended to other family isoforms.

Zusammenfassung

Die menschliche zytosolische Sulfotransferase 1A1 (SULT1A1) hat die breiteste Substratspezifität in der Familie der Sulfotransferasen und metabolisiert Myriaden von niedermolekularen phenolischen Endobiotika sowie Xenobiotika, z.B. p-Nitrophenol. Sie katalysiert den Transfer einer Sulfonat-Gruppe von einem universellen Sulfonat-Donor, dem 3'-Phosphoadenosine-5'-phosphosulphat (PAPS), auf einen Akzeptor (Lig) unter Bildung von nichtsulfonyliertem Kofaktor (PAP) und dem sulfonylierten Produkt (LigS). Aufgrund des Vorhandenseins von 2 Bindungsstellen, die besetzt oder nichtbesetzt sein können, kann SULT1A1 *in vivo* während der Reaktion in verschiedenen Kombinationen mit dem Kofaktor und dem Liganden vorliegen: Apoform (SULT1A1), binäre Komplexe (SULT1A1-PAP, SULT1A1-PAPS und SULT1A1-Lig) sowie ternäre Komplexe (SULT1A1-PAP-Lig und SULT1A1-PAPS-Lig). Substrate für SULT können in 3 Kategorien unterteilt werden, neutrale Substrate sowie solche mit positiver und negativer Kooperativität. Substrate mit positiver Kooperativität sind diejenigen, bei denen eine hohe Affinität zu kofaktorhaltigen Komplexen besteht. Allerdings sind die intermolekularen Enzym-Liganden-Interaktionen bislang nicht ausreichend charakterisiert. Zudem gibt es keine strukturellen Untersuchungen zu Komplexen mit unterschiedlichen Kombinationen von Kofaktoren und Liganden, die meisten liegen zudem nicht kristallisiert vor.

Ziel der vorliegenden Arbeit war, Untersuchungen durchzuführen, die zu einem besseren Verständnis der positiven Kooperativität bei Liganden für SULT1A1 führen. Dazu wurden *in silico*-Methoden eingesetzt. Zunächst wurden verfügbare kristallographische Daten recherchiert und gesammelt, miteinander verglichen und dann mit Strukturdaten zu anderen, nah verwandten Enzymen mit SULT1A1-überlappender Substratspezifität verglichen. Im nächsten Schritt wurden 3D-Pharmakophore generiert, um in Bezug auf die positive Kooperativität von (kristallisierten) untersuchten Liganden relevante Charakteristika herauszuarbeiten. 3D-Modelle wurden generiert aus strukturellen Daten (Struktur-basiert; wwPDB) und aus Information zu Liganden aus der verfügbaren Literatur (Liganden-basiert). Im darauffolgenden Schritt wurden Schlüssel-Komplexe des Sulfonylierungszyklus bei SULT1A1, für die bislang keine Strukturdaten existieren, modelliert und in einer MD(Molekulardynamik)-Studie modelliert.

Im ersten Teil der Untersuchungen wurden Strukturuntersuchungen durchgeführt, in die alle verfügbaren kristallographischen Daten zu SULT1A1 miteinbezogen wurden, und die flexiblen Bereiche identifiziert. Durch die Strukturanalyse wurde bestätigt, dass die Enzym-Substrat-Bindungsstelle (bzw. -tasche) von der Größe des Liganden abhängig ist und

dass über eine AS-Schleife im Zugangsbereich des Bindungsareals (AS 85-90) Form und Volumen des Substrats als Determinante miteingehen. Zudem wurde das Enzym verglichen mit anderen Isoformen aus der betreffenden Enzymfamilie, um Substrat-Bindungsstellen-Charakteristika zu identifizieren, die für die Substratspezifität relevant sind. Es wurde gefunden, dass die Substrat-Bindungsstelle am Enzym sehr hydrophob ist und aromatische und / oder -allgemein- hydrophobe Reste diese auskleiden, wobei hydrophile Reste hier kaum vorhanden sind. Die Ergebnisse aus den Strukturuntersuchungen bestätigen, dass Parameter der SULT1A1-Bindungsstelle (räumliche Parameter (Größe, Geometrie) und Lipophilie) die Substratspezifität bestimmen und das Profil des Bindungsverhaltens von Substraten an diesem Enzym erklären.

Im zweiten Teil der Studien lag der Focus auf der Kombinierung von Struktur-basierten, Liganden-basierten und Docking-basierten Informationen, um 3D-Pharmakophore zu konstruieren, die der Erkundung des Bindungsverhaltens von SULT1A1-Liganden dienen sollten. Pharmakophor-Struktur-Charakteristika (für hochaffine, stark bindende Liganden; nur Substrat-Daten für kokristallisierte Liganden) für den Struktur-basierten Approach entstanden mithilfe von Informationen aus der wwPDB. Danach wurden Liganden-basierte Pharmakophore über experimentell festgestellte potente SULT1A1-Inhibitoren konstruiert. Zu guter Letzt wurden die Interaktionen eines sehr aktiven/affinen Substrates (3,3-Diiodthyronin) und eines Inhibitors (Ethinylestradiol (EE)), bei denen jeweils keine kristallographischen Daten vorlagen, mit dem Enzym untersucht und für die plausibelsten Konformationen Docking-basierte 3D-Pharmakophore entwickelt. Mit den verfügbaren Pharmakophor-Filtern wurden retrospektiv über 400 Substanzen gescreent (Substrat-, Substrat+Inhibitor-Mix-, Inhibitor- und Decoy-Moleküle), um die Selektivität des Modells zu testen bzw. zu bestätigen. Über 100 Substanzen aus dem Datenpool konnten durch 'virtual screening' über die entwickelten Pharmacophore identifiziert werden als SULT1A1-Liganden, wobei sie in Bezug auf ihre chemische Struktur unterschiedlich waren und auch unterschiedlichen Affinitäten zeigten (107 SULT1A1-Liganden; davon 35 Substrate, 26 gemischte Substrate+Inhibitoren, und 46 Inhibitoren). Mit dem entwickelten Modell liessen sich mit hoher Genauigkeit Liganden in Inhibitoren (in ein Inhibitor-Pharmacophor-Modell passend), Substrate (in ein Substrat-Pharmacophor-Modell und nicht in ein Inhibitor-Pharmacophor-Modell passend) sowie Substanzen mit sowohl Substrat- als auch Inhibitor-Eigenschaften (in ein Inhibitor- und ein Substrat-Pharmacophor-Modell passend) diskriminieren bzw. die für eine Enzymbindung essentiellen Merkmale beschreiben (70% für Substrate und 60% für Inhibitoren). Allerdings konnte keine mechanistische Erklärungsmöglichkeit für positive Kooperativität gefunden werden. Daher war es erforderlich, das Bindungsverhalten mechanistisch in anderer Weise zu untersuchen.

Zur Untersuchung der Flexibilität der Liganden-induzierten Substratbindungsstelle wurden Molekulardynamik-Untersuchungen durchgeführt, die den dritten Part darstellen und Einsichten in die Dynamik der bislang nicht charakterisierten Komplexe des Sulfonierungszyklus gewähren sollten. Auch vorher publizierte Untersuchungen zu SULT implizierten den MD-Approach und hatten zum Inhalt, den konformativen Raum einer unbesetzten Substratbindungsstelle in Kofaktor-freien und Kofaktor-haltigen Komplexen "auszuleuchten". Allerdings wurde bei diesen Untersuchungen noch nicht berücksichtigt, dass das induzierte Volumen und die Architektur der Substratbindungsstelle nie in relevanter Weise in der Detailprofilierung erfasst werden kann, ohne dass ein Ligand präsent ist. Aus diesem Grund wurden für SULT-Komplexe 100ns-MD-Simulationen durchgeführt und dabei die Präsenz von solchen Substraten miteinbezogen, die sich entweder als neutral erwiesen hatten oder für die eine positive Kooperativität gefunden worden war. Diese Simulationen erwiesen sich als außerordentlich nützlich, um die Stabilität von SULT1A1-Komplexen in unterschiedlichen Phasen der Kofaktor- und Liganden-Bindung bei Bindung unterschiedlicher Substrate zu definieren. Die molekulardynamischen Studien und die Analyse des Volumens der Substratbindungsstelle weisen auf eine Beziehung zwischen der Größe des entsprechenden Liganden, den Dimensionen der Bindungstasche demzufolge den Substrat-/Inhibitor-Eigenschaften hin. Bei kleinen hydrophoben Liganden (wie beispielsweise 2-Naphthol (2NA) wurde in Bezug auf die Konformation an der Substratbindungsstelle gefunden, dass sie eine 50%ige Volumenreduktion (= auf 250 \AA^3) induzieren. Positive Kooperativität bei der Bindung kann erklärt werden durch enge hydrophobe Interaktionen mit Phe81, Phe142, Phe76, Phe84 und Ile89 in den Enzym-Substrat-Komplexen, die kofaktorhaltig sind. Im Gegensatz dazu destabilisierte die Bindung von kleinen bzw. niedermolekularen Substraten an das kofaktorfreie Apoenzym den Kofaktor-bindenden Teil in Schleife 3 durch Distorsion des aromatischen π -Stackings bei Tyr240 und Phe255. Diese Ergebnisse geben eine Erklärung für die beschriebene geringere Bindungsaffinität von Substraten mit positiver Kooperativität im kofaktorfreien Enzym-Komplex. Interagieren größere Moleküle, die Substrate mit positiver Kooperativität darstellen (z.B. Ethinylestradiol (EE)), trägt die Flexibilität von Phe76 und Phe84 dazu bei, dass sich die Bindungstasche II öffnet und sich das hydrophobe System aufweitet, was wiederum dazu führt, dass der Ligand mit den Bindungsarealen interagieren kann und das Volumen der Bindungstasche sich verdreifacht. Diese zusätzliche Öffnung der Bindungstasche führt zu einem festen Einschluss größerer hydrophober Moleküle in der Tasche, auch noch nach der Sulfonierung, und kann die Ursache für einen inhibitorischen Effekt sein.

Zusammenfassend kann festgestellt werden, dass die Flexibilität der Substratbindungsstelle Liganden-induzierte Veränderungen in Form und Volumen bedingt. Die Ergebnisse der vorliegenden Arbeit leisten einen wichtigen Beitrag zum Verständnis der breiten Substratspezifität und geringen Selektivität bei SULT1A1. Und da alle Mitglieder der SULT-Enzymfamilie eine hohe Übereinstimmung in ihrer chemischen Struktur aufweisen, kann der hier für SULT1A1 gewählte Ansatz und können auch Hypothesen und Schlussfolgerungen ggf. auf andere Isoformen aus dieser Enzymfamilie übertragen werden.

9 References

1. Raunio H, Kuusisto M, Juvonen RO, Pentikainen OT. 2015. Modeling of interactions between xenobiotics and cytochrome P450 (CYP) enzymes. *Front Pharmacol* 6123.
2. Moroy G, Martiny VY, Vayer P, Villoutreix BO, Miteva MA. 2012. Toward in silico structure-based ADMET prediction in drug discovery. *Drug Discov Today* 17(1-2): 44-55.
3. Jancova P, Anzenbacher P, Anzenbacherova E. 2010. Phase II drug metabolizing enzymes. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 154(2): 103-16.
4. Anzenbacher P, Anzenbacherova E. 2001. Cytochromes P450 and metabolism of xenobiotics. *Cell Mol Life Sci* 58(5-6): 737-47.
5. Tanaka E. 1998. Clinically important pharmacokinetic drug-drug interactions: role of cytochrome P450 enzymes. *J Clin Pharm Ther* 23(6): 403-416.
6. Ring JA, Ghabrial H, Ching MS, Smallwood RA, Morgan DJ. 1999. Fetal hepatic drug elimination. *Pharmacol Ther* 84(3): 429-45.
7. Klaassen CD, Boles JW. 1997. Sulfation and sulfotransferases 5: the importance of 3'-phosphoadenosine 5'-phosphosulfate (PAPS) in the regulation of sulfation. *FASEB J* 11(6): 404-18.
8. Gamage N, Barnett A, Hempel N, Duggleby RG, Windmill KF, Martin JL, McManus ME. 2006. Human sulfotransferases and their role in chemical metabolism. *Toxicol Sci* 90(1): 5-22.
9. Bistrup A, Bhakta S, Lee JK, Belov YY, Gunn MD, Zuo FR, Huang CC, Kannagi R, Rosen SD, Hemmerich S. 1999. Sulfotransferases of two specificities function in the reconstitution of high endothelial cell ligands for L-selectin. *J Cell Biol* 145(4): 899-910.
10. Grunwell JR, Bertozzi CR. 2002. Carbohydrate sulfotransferases of the GalNAc/Gal/GlcNAc6ST family. *Biochemistry* 41(44): 13117-26.
11. Chapman E, Best MD, Hanson SR, Wong CH. 2004. Sulfotransferases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angew Chem Int Ed Engl* 43(27): 3526-48.
12. Paul P, Suwan J, Liu J, Dordick JS, Linhardt RJ. 2012. Recent advances in sulfotransferase enzyme activity assays. *Anal Bioanal Chem* 403(6): 1491-500.
13. Kohalmy K, Vrzal R. 2011. Regulation of phase II biotransformation enzymes by steroid hormones. *Curr Drug Metab* 12(2): 104-23.
14. Wang LQ, James MO. 2006. Inhibition of sulfotransferases by xenobiotics. *Curr Drug Metab* 7(1): 83-104.
15. Gamage NU, Duggleby RG, Barnett AC, Tresillian M, Latham CF, Liyou NE, McManus ME, Martin JL. 2003. Structure of a human carcinogen-converting enzyme, SULT1A1. Structural and kinetic implications of substrate inhibition. *J Biol Chem* 278(9): 7655-62.

16. Tibbs ZE, Rohn-Glowacki KJ, Crittenden F, Guidry AL, Falany CN. 2015. Structural plasticity in the human cytosolic sulfotransferase dimer and its role in substrate selectivity and catalysis. *Drug Metabolism and Pharmacokinetics* 30(1): 3-20.
17. Gamage NU, Tsvetanov S, Duggleby RG, McManus ME, Martin JL. 2005. The structure of human SULT1A1 crystallized with estradiol. An insight into active site plasticity and substrate inhibition with multi-ring substrates. *J Biol Chem* 280(50): 41482-6.
18. Lu J, Li H, Zhang J, Li M, Liu MY, An X, Liu MC, Chang W. 2010. Crystal structures of SULT1A2 and SULT1A1 *3: insights into the substrate inhibition and the role of Tyr149 in SULT1A2. *Biochem Biophys Res Commun* 396(2): 429-34.
19. Alcolombri U, Elias M, Tawfik DS. 2011. Directed evolution of sulfotransferases and paraoxonases by ancestral libraries. *J Mol Biol* 411(4): 837-53.
20. Berger I, Guttman C, Amar D, Zarivach R, Aharoni A. 2011. The molecular basis for the broad substrate specificity of human sulfotransferase 1A1. *PLoS One* 6(11): e26794.
21. Cook I, Wang T, Almo SC, Kim J, Falany CN, Leyh TS. 2013. The gate that governs sulfotransferase selectivity. *Biochemistry* 52(2): 415-24.
22. Bidwell LM, McManus ME, Gaedigk A, Kakuta Y, Negishi M, Pedersen L, Martin JL. 1999. Crystal structure of human catecholamine sulfotransferase. *J Mol Biol* 293(3): 521-30.
23. Lu JH, Li HT, Liu MC, Zhang JP, Li M, An XM, Chang WR. 2005. Crystal structure of human sulfotransferase SULT1A3 in complex with dopamine and 3'-phosphoadenosine 5'-phosphate. *Biochem Biophys Res Commun* 335(2): 417-23.
24. Dombrovski L, Dong A, Bochkarev A, Plotnikov AN. 2006. Crystal structures of human sulfotransferases SULT1B1 and SULT1C1 complexed with the cofactor product adenosine-3'-5'-diphosphate (PAP). *Proteins* 64(4): 1091-4.
25. Allali-Hassani A, Pan PW, Dombrovski L, Najmanovich R, Tempel W, Dong A, Loppnau P, Martin F, Thornton J, Edwards AM, Bochkarev A, Plotnikov AN, Vedadi M, Arrowsmith CH. 2007. Structural and chemical profiling of the human cytosolic sulfotransferases. *PLoS Biol* 5(5): e97.
26. Kakuta Y Fau - Pedersen LG, Pedersen Lg Fau - Carter CW, Carter Cw Fau - Negishi M, Negishi M Fau - Pedersen LC, Pedersen LC. Crystal structure of estrogen sulphotransferase. (1072-8368 (Print)).
27. Kakuta Y, Petrotchenko EV, Pedersen LC, Negishi M. 1998. The sulfuryl transfer mechanism. Crystal structure of a vanadate complex of estrogen sulfotransferase and mutational analysis. *J Biol Chem* 273(42): 27325-30.
28. Pedersen LC, Petrotchenko E, Shevtsov S, Negishi M. 2002. Crystal structure of the human estrogen sulfotransferase-PAPS complex: evidence for catalytic role of Ser137 in the sulfuryl transfer reaction. *J Biol Chem* 277(20): 17928-32.
29. Shevtsov S, Petrotchenko EV, Pedersen LC, 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(7): 884-8.

30. Gosavi RA, Knudsen Ga Fau - Birnbaum LS, Birnbaum Ls Fau - Pedersen LC, Pedersen LC. Mimicking of estradiol binding by flame retardants and their metabolites: a crystallographic analysis. (1552-9924 (Electronic)).
31. Lu LY, Hsieh YC, Liu MY, Lin YH, Chen CJ, Yang YS. 2008. Identification and characterization of two amino acids critical for the substrate inhibition of human dehydroepiandrosterone sulfotransferase (SULT2A1). *Mol Pharmacol* 73(3): 660-8.
32. Pedersen LC, Petrotchenko EV, Negishi M. 2000. Crystal structure of SULT2A3, human hydroxysteroid sulfotransferase. *FEBS Lett* 475(1): 61-4.
33. Rehse PH, Zhou M, Lin SX. 2002. Crystal structure of human dehydroepiandrosterone sulphotransferase in complex with substrate. *Biochem J* 364(Pt 1): 165-71.
34. Chang HJ, Shi R, Rehse P, Lin SX. 2004. Identifying androsterone (ADT) as a cognate substrate for human dehydroepiandrosterone sulfotransferase (DHEA-ST) important for steroid homeostasis: structure of the enzyme-ADT complex. *J Biol Chem* 279(4): 2689-96.
35. Lee KA, Fuda H, Lee YC, Negishi M, Strott CA, Pedersen LC. 2003. Crystal structure of human cholesterol sulfotransferase (SULT2B1b) in the presence of pregnenolone and 3'-phosphoadenosine 5'-phosphate. Rationale for specificity differences between prototypical SULT2A1 and the SULT2BG1 isoforms. *J Biol Chem* 278(45): 44593-9.
36. Matsui M, Homma H. 1994. Biochemistry and molecular biology of drug-metabolizing sulfotransferase. *Int J Biochem* 26(10-11): 1237-47.
37. Raman R, Myette J, Venkataraman G, Sasisekharan V, Sasisekharan R. 2002. Identification of structural motifs and amino acids within the structure of human heparan sulfate 3-O-sulfotransferase that mediate enzymatic function. *Biochem Biophys Res Commun* 290(4): 1214-9.
38. Falany CN, Xie X, Wang J, Ferrer J, Falany JL. 2000. Molecular cloning and expression of novel sulphotransferase-like cDNAs from human and rat brain. *Biochem J* 346 Pt 3857-64.
39. Cook IT, Leyh TS, Kadlubar SA, Falany CN. 2010. Structural rearrangement of SULT2A1: effects on dehydroepiandrosterone and raloxifene sulfation. *Horm Mol Biol Clin Investig* 1(2): 81-87.
40. Petrotchenko EV, Pedersen LC, Borchers CH, Tomer KB, Negishi M. 2001. The dimerization motif of cytosolic sulfotransferases. *FEBS Lett* 490(1-2): 39-43.
41. Weitzner B, Meehan T, Xu Q, Dunbrack RL, Jr. 2009. An unusually small dimer interface is observed in all available crystal structures of cytosolic sulfotransferases. *Proteins* 75(2): 289-95.
42. Lu LY, Chiang HP, Chen WT, Yang YS. 2009. Dimerization is responsible for the structural stability of human sulfotransferase 1A1. *Drug Metab Dispos* 37(5): 1083-8.
43. Cook IT, Leyh TS, Kadlubar SA, Falany CN. 2010. Lack of substrate inhibition in a monomeric form of human cytosolic SULT2A1. *Horm Mol Biol Clin Investig* 3(1): 357-366.
44. Sun M, Leyh TS. The human estrogen sulfotransferase: a half-site reactive enzyme. *Biochemistry* 49(23): 4779-85.

45. Yoval-Sanchez B, Pardo Jp Fau - Rodriguez-Zavala JS, Rodriguez-Zavala JS. New insights into the half-of-the-sites reactivity of human aldehyde dehydrogenase 1A1. (1097-0134 (Electronic)).
46. Tibbs ZE, Falany CN. 2015. Dimeric human sulfotransferase 1B1 displays cofactor-dependent subunit communication. *Pharmacology Research & Perspectives* 3(3): e00147.
47. Tibbs Z, Falany C. 2015. Cofactor Dependent Cooperation Between Dimeric hSULT1B1 Subunits. *The FASEB Journal* 29(1 Supplement).
48. Teramoto T, Sakakibara Y, Liu MC, Suiko M, Kimura M, Kakuta Y. 2009. Snapshot of a Michaelis complex in a sulfuryl transfer reaction: Crystal structure of a mouse sulfotransferase, mSULT1D1, complexed with donor substrate and acceptor substrate. *Biochem Biophys Res Commun* 383(1): 83-7.
49. Ong E, Yeh JC, Ding Y, Hindsgaul O, Pedersen LC, Negishi M, Fukuda M. 1999. Structure and function of HNK-1 sulfotransferase. Identification of donor and acceptor binding sites by site-directed mutagenesis. *J Biol Chem* 274(36): 25608-12.
50. Robbins PW, Lipmann F. 1958. Enzymatic synthesis of adenosine-5'-phosphosulfate. *J Biol Chem* 233(3): 686-90.
51. Venkatachalam KV, Akita H, Strott CA. 1998. Molecular cloning, expression, and characterization of human bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase and its functional domains. *J Biol Chem* 273(30): 19311-20.
52. Xu Z, Wood TC, Adjei AA, Weinshilboum RM. 2001. Human 3'-phosphoadenosine 5'-phosphosulfate synthetase: radiochemical enzymatic assay, biochemical properties, and hepatic variation. *Drug Metab Dispos* 29(2): 172-8.
53. Sendelbach LE, White CA, Howell S, Gregus Z, Klaassen CD. 1990. Effect of sulfhydryl-deficient diets on hepatic metallothionein, glutathione, and adenosine 3'-phosphate 5'-phosphosulfate (PAPS) levels in rats. *Toxicol Appl Pharmacol* 102(2): 259-67.
54. Bhattacharya R, Townley Ra Fau - Berry KL, Berry KI Fau - Bulow HE, Bulow HE. The PAPS transporter PST-1 is required for heparan sulfation and is essential for viability and neural development in *C. elegans*. (1477-9137 (Electronic)).
55. Kamiyama S, Sasaki N, Goda E, Ui-Tei K, Saigo K, Narimatsu H, Jigami Y, Kannagi R, Irimura T, Nishihara S. 2006. Molecular Cloning and Characterization of a Novel 3'-Phosphoadenosine 5'-Phosphosulfate Transporter, PAPST2. *Journal of Biological Chemistry* 281(16): 10945-10953.
56. Cook I, Wang T, Falany CN, Leyh TS. 2012. A nucleotide-gated molecular pore selects sulfotransferase substrates. *Biochemistry* 51(28): 5674-83.
57. Wang T, Cook I, Leyh TS. 2014. 3'-Phosphoadenosine 5'-phosphosulfate allosterically regulates sulfotransferase turnover. *Biochemistry* 53(44): 6893-900.
58. Cappiello M, Franchi M, Rane A, Pacifici GM. 1990. Sulphotransferase and its substrate: adenosine-3'-phosphate-5'-phosphosulphate in human fetal liver and placenta. *Dev Pharmacol Ther* 14(1): 62-5.
59. Hoff RH, Czyryca PG, Sun M, Leyh TS, Hengge AC. 2006. Transition state of the sulfuryl transfer reaction of estrogen sulfotransferase. *J Biol Chem* 281(41): 30645-9.

-
60. Lenevich S, Xu J, Hosokawa A, Cramer CJ, Distefano MD. 2007. Transition state analysis of model and enzymatic prenylation reactions. *J Am Chem Soc* 129(18): 5796-7.
 61. Dolence JM, Poulter CD. 1995. A mechanism for posttranslational modifications of proteins by yeast protein farnesyltransferase. *Proc Natl Acad Sci U S A* 92(11): 5008-11.
 62. Leyh TS. 1993. The physical biochemistry and molecular genetics of sulfate activation. *Crit Rev Biochem Mol Biol* 28(6): 515-42.
 63. Danan LM, Yu Z, Ludden PJ, Jia W, Moore KL, Leary JA. Catalytic mechanism of Golgi-resident human tyrosylprotein sulfotransferase-2: a mass spectrometry approach. *J Am Soc Mass Spectrom* 21(9): 1633-42.
 64. Tyapochkin E, Cook PF, Chen G. 2008. Isotope exchange at equilibrium indicates a steady state ordered kinetic mechanism for human sulfotransferase. *Biochemistry* 47(45): 11894-9.
 65. Banerjee RK, Roy AB. 1968. Kinetic studies of the phenol sulphotransferase reaction. *Biochim Biophys Acta* 151(3): 573-86.
 66. Whittemore RM, Pearce LB, Roth JA. 1986. Purification and kinetic characterization of a phenol-sulfating form of phenol sulfotransferase from human brain. *Arch Biochem Biophys* 249(2): 464-71.
 67. Zhang H, Varlamova O, Vargas FM, Falany CN, Leyh TS. 1998. Sulfuryl transfer: the catalytic mechanism of human estrogen sulfotransferase. *J Biol Chem* 273(18): 10888-92.
 68. Cook PF, Cleland WW. 1981. Mechanistic deductions from isotope effects in multireactant enzyme mechanisms. *Biochemistry* 20(7): 1790-6.
 69. Cook I, Wang T, Leyh TS. 2015. Sulfotransferase 1A1 Substrate Selectivity: A Molecular Clamp Mechanism. *Biochemistry* 54(39): 6114-22.
 70. Wang T, Cook I, Falany CN, Leyh TS. 2014. Paradigms of sulfotransferase catalysis: the mechanism of SULT2A1. *J Biol Chem* 289(38): 26474-80.
 71. Wu B. 2011. Substrate inhibition kinetics in drug metabolism reactions. *Drug Metab Rev* 43(4): 440-56.
 72. Reed MC, Lieb A, Nijhout HF. 2010. The biological significance of substrate inhibition: a mechanism with diverse functions. *Bioessays* 32(5): 422-9.
 73. Barnett AC, Tsvetanov S, Gamage N, Martin JL, Duggleby RG, McManus ME. 2004. Active site mutations and substrate inhibition in human sulfotransferase 1A1 and 1A3. *J Biol Chem* 279(18): 18799-805.
 74. Blanchard RL, Freimuth RR, Buck J, Weinshilboum RM, Coughtrie MW. 2004. A proposed nomenclature system for the cytosolic sulfotransferase (SULT) superfamily. *Pharmacogenetics* 14(3): 199-211.
 75. Veronese ME, Burgess W, Zhu X, McManus ME. 1994. Functional characterization of two human sulphotransferase cDNAs that encode monoamine- and phenol-sulphating forms of phenol sulphotransferase: substrate kinetics, thermal-stability and inhibitor-sensitivity studies. *Biochem J* 302 (Pt 2)497-502.

-
76. He D, Frost AR, Falany CN. 2005. Identification and immunohistochemical localization of Sulfotransferase 2B1b (SULT2B1b) in human lung. *Biochim Biophys Acta* 1724(1-2): 119-26.
77. Stanley EL, Hume R, Coughtrie MW. 2005. Expression profiling of human fetal cytosolic sulfotransferases involved in steroid and thyroid hormone metabolism and in detoxification. *Mol Cell Endocrinol* 240(1-2): 32-42.
78. Teubner W, Meinl W, Florian S, Kretzschmar M, Glatt H. 2007. Identification and localization of soluble sulfotransferases in the human gastrointestinal tract. *Biochem J* 404(2): 207-15.
79. Dooley TP, Haldeman-Cahill R, Joiner J, Wilborn TW. 2000. Expression profiling of human sulfotransferase and sulfatase gene superfamilies in epithelial tissues and cultured cells. *Biochem Biophys Res Commun* 277(1): 236-45.
80. Heroux JA, Roth JA. 1988. Physical characterization of a monoamine-sulfating form of phenol sulfotransferase from human platelets. *Mol Pharmacol* 34(2): 194-9.
81. Rein G, Glover V, Sandler M. 1982. Multiple forms of phenolsulphotransferase in human tissues: selective inhibition by dichloronitrophenol. *Biochem Pharmacol* 31(10): 1893-7.
82. Hempel N, Gamage N, Martin JL, McManus ME. 2007. Human cytosolic sulfotransferase SULT1A1. *Int J Biochem Cell Biol* 39(4): 685-9.
83. Glatt H, Boeing H, Engelke CE, Ma L, Kuhlow A, Pabel U, Pomplun D, Teubner W, Meinl W. 2001. Human cytosolic sulphotransferases: genetics, characteristics, toxicological aspects. *Mutat Res* 482(1-2): 27-40.
84. Rohn KJ, Cook IT, Leyh TS, Kadlubar SA, Falany CN. 2012. Potent inhibition of human sulfotransferase 1A1 by 17alpha-ethinylestradiol: role of 3'-phosphoadenosine 5'-phosphosulfate binding and structural rearrangements in regulating inhibition and activity. *Drug Metab Dispos* 40(8): 1588-95.
85. Dong D, Ako R, Wu B. 2012. Crystal structures of human sulfotransferases: insights into the mechanisms of action and substrate selectivity. *Expert Opin Drug Metab Toxicol* 8(6): 635-46.
86. Dajani R, Hood AM, Coughtrie MW. 1998. A single amino acid, glu146, governs the substrate specificity of a human dopamine sulfotransferase, SULT1A3. *Mol Pharmacol* 54(6): 942-8.
87. Meinl W, Meerman JH, Glatt H. 2002. Differential activation of promutagens by alloenzymes of human sulfotransferase 1A2 expressed in *Salmonella typhimurium*. *Pharmacogenetics* 12(9): 677-89.
88. Brix LA, Barnett AC, Duggleby RG, Leggett B, McManus ME. 1999. Analysis of the substrate specificity of human sulfotransferases SULT1A1 and SULT1A3: site-directed mutagenesis and kinetic studies. *Biochemistry* 38(32): 10474-9.
89. Riches Z, Stanley EL, Bloomer JC, Coughtrie MW. 2009. Quantitative evaluation of the expression and activity of five major sulfotransferases (SULTs) in human tissues: the SULT "pie". *Drug Metab Dispos* 37(11): 2255-61.

-
90. Heroux JA, Falany CN, Roth JA. 1989. Immunological characterization of human phenol sulfotransferase. *Mol Pharmacol* 36(1): 29-33.
91. Richard K, Hume R, Kaptein E, Stanley EL, Visser TJ, Coughtrie MW. 2001. Sulfation of thyroid hormone and dopamine during human development: ontogeny of phenol sulfotransferases and arylsulfatase in liver, lung, and brain. *J Clin Endocrinol Metab* 86(6): 2734-42.
92. Buhl AE, Waldon DJ, Fau - Baker CA, Baker CA, Fau - Johnson GA, Johnson GA. Minoxidil sulfate is the active metabolite that stimulates hair follicles. (0022-202X (Print)).
93. Falany CN, Kerl EA. 1990. Sulfation of minoxidil by human liver phenol sulfotransferase. *Biochem Pharmacol* 40(5): 1027-32.
94. Anderson RJ, Kudlacek PE, Clemens DL. 1998. Sulfation of minoxidil by multiple human cytosolic sulfotransferases. *Chem Biol Interact* 109(1-3): 53-67.
95. Glatt H. 2000. Sulfotransferases in the bioactivation of xenobiotics. *Chem Biol Interact* 129(1-2): 141-70.
96. Coughtrie MW, Gilissen RA, Shek B, Strange RC, Fryer AA, Jones PW, Bamber DE. 1999. Phenol sulphotransferase SULT1A1 polymorphism: molecular diagnosis and allele frequencies in Caucasian and African populations. *Biochem J* 337 (Pt 1)45-9.
97. Raftogianis RB, Wood TC, Weinshilboum RM. 1999. Human phenol sulfotransferases SULT1A2 and SULT1A1: genetic polymorphisms, allozyme properties, and human liver genotype-phenotype correlations. *Biochem Pharmacol* 58(4): 605-16.
98. Arlt VM, Glatt H, Muckel E, Pabel U, Sorg BL, Schmeiser HH, Phillips DH. 2002. Metabolic activation of the environmental contaminant 3-nitrobenzanthrone by human acetyltransferases and sulfotransferase. *Carcinogenesis* 23(11): 1937-45.
99. Nagar S, Walther S, Blanchard RL. 2006. Sulfotransferase (SULT) 1A1 polymorphic variants *1, *2, and *3 are associated with altered enzymatic activity, cellular phenotype, and protein degradation. *Mol Pharmacol* 69(6): 2084-92.
100. Li X, Clemens DL, Cole JR, Anderson RJ. 2001. Characterization of human liver thermostable phenol sulfotransferase (SULT1A1) allozymes with 3,3',5-triiodothyronine as the substrate. *J Endocrinol* 171(3): 525-32.
101. Arslan S, Silig Y, Fau - Pinarbasi H, Pinarbasi H. An investigation of the relationship between SULT1A1 Arg(213)His polymorphism and lung cancer susceptibility in a Turkish population. (1099-0844 (Electronic)).
102. Bardakci F, Arslan S, Bardakci S, Binatli AO, Budak M. 2008. Sulfotransferase 1A1 (SULT1A1) polymorphism and susceptibility to primary brain tumors. *J Cancer Res Clin Oncol* 134(1): 109-14.
103. Huang SK, Chiu AW, Pu YS, Huang YK, Chung CJ, Tsai HJ, Yang MH, Chen CJ, Hsueh YM. 2009. Arsenic methylation capability, myeloperoxidase and sulfotransferase genetic polymorphisms, and the stage and grade of urothelial carcinoma. *Urol Int* 82(2): 227-34.
104. Lindsay J, Wang LL, Li Y, Zhou SF. 2008. Structure, function and polymorphism of human cytosolic sulfotransferases. *Curr Drug Metab* 9(2): 99-105.

105. Liu Y, Apak TI, Lehmler HJ, Robertson LW, Duffel MW. 2006. Hydroxylated polychlorinated biphenyls are substrates and inhibitors of human hydroxysteroid sulfotransferase SULT2A1. *Chem Res Toxicol* 19(11): 1420-5.
106. Sharma V, Duffel MW. 2002. Comparative molecular field analysis of substrates for an aryl sulfotransferase based on catalytic mechanism and protein homology modeling. *J Med Chem* 45(25): 5514-22.
107. Eisenthal R, Danson MJ, Hough DW. 2007. Catalytic efficiency and k_{cat}/K_M : a useful comparator? *Trends Biotechnol* 25(6): 247-9.
108. Furimsky AM, Green CE, Sharp LE, Catz P, Adjei AA, Parman T, Kapetanovic IM, Weinshilboum RM, Iyer LV. 2008. Effect of resveratrol on 17 β -estradiol sulfation by human hepatic and jejunal S9 and recombinant sulfotransferase 1E1. *Drug Metab Dispos* 36(1): 129-36.
109. Cook I, Wang T, Falany CN, Leyh TS. 2013. High accuracy in silico sulfotransferase models. *J Biol Chem* 288(48): 34494-501.
110. Coughtrie MW, Johnston LE. 2001. Interactions between dietary chemicals and human sulfotransferases-molecular mechanisms and clinical significance. *Drug Metab Dispos* 29(4 Pt 2): 522-8.
111. Vietri M, Pietrabissa A, Mosca F, Spisni R, Pacifici GM. 2003. Curcumin is a potent inhibitor of phenol sulfotransferase (SULT1A1) in human liver and extrahepatic tissues. *Xenobiotica* 33(4): 357-63.
112. Kester MH, Bulduk S, Tibboel D, Meini W, Glatt H, Falany CN, Coughtrie MW, Bergman A, Safe SH, Kuiper GG, Schuur AG, Brouwer A, Visser TJ. 2000. Potent inhibition of estrogen sulfotransferase by hydroxylated PCB metabolites: a novel pathway explaining the estrogenic activity of PCBs. *Endocrinology* 141(5): 1897-900.
113. Wang LQ, Lehmler HJ, Robertson LW, Falany CN, James MO. 2005. In vitro inhibition of human hepatic and cDNA-expressed sulfotransferase activity with 3-hydroxybenzo[a]pyrene by polychlorobiphenyls. *Environ Health Perspect* 113(6): 680-7.
114. Nimmagadda D, Cherala G, Ghatta S. 2006. Cytosolic sulfotransferases. *Indian J Exp Biol* 44(3): 171-82.
115. Verdugo DE, Cancilla MT, Ge X, Gray NS, Chang YT, Schultz PG, Negishi M, Leary JA, Bertozzi CR. 2001. Discovery of estrogen sulfotransferase inhibitors from a purine library screen. *J Med Chem* 44(17): 2683-6.
116. Horwitz JP, Misra RS, Rozhin J, Neenan JP, Huo A, Godefroi VC, Philips KD, Chung HL, Butke G, Brooks SC. 1978. IV. Synthesis and assay of analogs of adenosine 3',5'-diphosphate as inhibitors of bovine adrenal estrogen sulfotransferase. *Biochimica et biophysica acta* 525(2): 364-72.
117. Cook I, Wang T, Falany CN, Leyh TS. 2015. The allosteric binding sites of sulfotransferase 1A1. *Drug Metab Dispos* 43(3): 418-23.
118. Campbell NR, Van Loon JA, Sundaram RS, Ames MM, Hansch C, Weinshilboum R. 1987. Human and rat liver phenol sulfotransferase: structure-activity relationships for phenolic substrates. *Mol Pharmacol* 32(6): 813-9.

119. Dajani R, Cleasby A, Neu M, Wonacott AJ, Jhoti H, Hood AM, Modi S, Hersey A, Taskinen J, Cooke RM, Manchee GR, Coughtrie MW. 1999. 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* 274(53): 37862-8.
120. Heimstad ES, Andersson PL. 2002. Docking and QSAR studies of an indirect estrogenic effect of hydroxylated PCBs. *Quantitative Structure-Activity Relationships* 21(3): 257-266.
121. Taskinen J, Ethell BT, Pihlavisto P, Hood AM, Burchell B, Coughtrie MW. 2003. Conjugation of catechols by recombinant human sulfotransferases, UDP-glucuronosyltransferases, and soluble catechol O-methyltransferase: structure-conjugation relationships and predictive models. *Drug Metab Dispos* 31(9): 1187-97.
122. Sipila J, Hood AM, Coughtrie MW, Taskinen J. 2003. CoMFA modeling of enzyme kinetics: $K(m)$ values for sulfation of diverse phenolic substrates by human catecholamine sulfotransferase SULT1A3. *J Chem Inf Comput Sci* 43(5): 1563-9.
123. Sharma V, Duffel MW. 2005. A comparative molecular field analysis-based approach to prediction of sulfotransferase catalytic specificity. *Methods Enzymol* 400:249-63.
124. Rakers C, Wolber G. Prediction of sulfotransferase specificity for risk assessment in drug design. Freie Universität Berlin Berlin, Germany, 2016.
125. Rakers C, Schumacher F, Meinel W, Glatt H, Kleuser B, Wolber G. 2016. In silico prediction of human sulfotransferase 1E1 activity guided by pharmacophores from molecular dynamics simulations. *J Biol Chem* 291(1): 58-71.
126. Martiny VY, Carbonell P, Lagorce D, Villoutreix BO, Moroy G, Miteva MA. 2013. In silico mechanistic profiling to probe small molecule binding to sulfotransferases. *PLoS One* 8(9): e73587.
127. Falany JL, Pilloff DE, Leyh TS, Falany CN. 2006. Sulfation of raloxifene and 4-hydroxytamoxifen by human cytosolic sulfotransferases. *Drug Metab Dispos* 34(3): 361-8.
128. Butt CM, Stapleton HM. Inhibition of thyroid hormone sulfotransferase activity by brominated flame retardants and halogenated phenolics. (1520-5010 (Electronic)).
129. Nair PC, Miners JO. 2014. Molecular dynamics simulations: from structure function relationships to drug discovery. *In Silico Pharmacol* 24.
130. Yalcin EB, Struzik SM, King RS. 2008. Allosteric modulation of SULT2A1 by celecoxib and nimesulide: computational analyses. *Drug Metab Lett* 2(3): 198-204.
131. Meng S, Wu B, Singh R, Yin T, Morrow JK, Zhang S, Hu M. 2012. SULT1A3-mediated regiospecific 7-O-sulfation of flavonoids in Caco-2 cells can be explained by the relevant molecular docking studies. *Molecular Pharmaceutics* 9(4): 862-873.
132. Campagna-Slater V, Schapira M. 2009. Evaluation of virtual screening as a tool for chemical genetic applications. *J Chem Inf Model* 49(9): 2082-91.
133. Chen YZ, Ung CY. 2001. Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand-protein inverse docking approach. *Journal of Molecular Graphics and Modelling* 20(3): 199-218.

134. Hamers T, Kamstra JH, Sonneveld E, Murk AJ, Kester MH, Andersson PL, Legler J, Brouwer A. 2006. In vitro profiling of the endocrine-disrupting potency of brominated flame retardants. *Toxicol Sci* 92(1): 157-73.
135. Harju M, Hamers T, Kamstra JH, Sonneveld E, Boon JP, Tysklind M, Andersson PL. 2007. Quantitative structure-activity relationship modeling on in vitro endocrine effects and metabolic stability involving 26 selected brominated flame retardants. *Environ Toxicol Chem* 26(4): 816-26.
136. Ekuase EJ, Liu Y, Lehmler HJ, Robertson LW, Duffel MW. 2011. Structure-activity relationships for hydroxylated polychlorinated biphenyls as inhibitors of the sulfation of dehydroepiandrosterone catalyzed by human hydroxysteroid sulfotransferase SULT2A1. *Chem Res Toxicol* 24(10): 1720-8.
137. Zhang PP, Zhao L, Long SY, Tian P. 2015. The effect of ligands on the thermal stability of sulfotransferases: a molecular dynamics simulation study. *J Mol Model* 21(4): 72.
138. Ekuase EJ, van 't Erve TJ, Rahaman A, Robertson LW, Duffel MW, Luthe G. Mechanistic insights into the specificity of human cytosolic sulfotransferase 2A1 (hSULT2A1) for hydroxylated polychlorinated biphenyls through the use of fluoro-tagged probes. (1614-7499 (Electronic)).
139. Basak SC. 2013. Recent developments and future directions at Current Computer Aided Drug Design. *Curr Comput Aided Drug Des* 9(1): 1.
140. Hawkins PCD, Skillman AG, Nicholls A. 2007. Comparison of shape-matching and docking as virtual screening tools. *Journal of Medicinal Chemistry* 50(1): 74-82.
141. Ehrlich P. 1909. Über den jetzigen Stand der Chemotherapie. *Berichte der deutschen chemischen Gesellschaft* 42(1): 17-47.
142. Fulton BS, Glossary of terms used in medicinal chemistry. In *Drug Discovery for the Treatment of Addiction*, John Wiley & Sons, Inc.: 2014; pp 271-289.
143. Leach AR, Gillet VJ, Lewis RA, Taylor R. 2010. Three-dimensional pharmacophore methods in drug discovery. *J Med Chem* 53(2): 539-58.
144. McGaughey GB, Sheridan RP, Bayly CI, Culberson JC, Kretsoulas C, Lindsley S, Maiorov V, Truchon J-F, Cornell WD. 2007. Comparison of topological, shape, and docking methods in virtual screening. *Journal of Chemical Information and Modeling* 47(4): 1504-1519.
145. Murgueitio MS, Bermudez M, Mortier J, Wolber G. 2012. In silico virtual screening approaches for anti-viral drug discovery. *Drug Discovery Today: Technologies* 9(3): e219-e225.
146. Yang SY. 2010. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov Today* 15(11-12): 444-50.
147. Wolber G, Langer T. 2005. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model* 45(1): 160-9.
148. Wolber G, Dornhofer AA, Langer T. 2006. Efficient overlay of small organic molecules using 3D pharmacophores. *J Comput Aided Mol Des* 20(12): 773-88.

149. Kurogi Y, Guner OF. 2001. Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Curr Med Chem* 8(9): 1035-55.
150. Guner O, Clement O, Kurogi Y. 2004. Pharmacophore modeling and three dimensional database searching for drug design using catalyst: recent advances. *Curr Med Chem* 11(22): 2991-3005.
151. Dixon SL, Smondyrev AM, Rao SN. 2006. PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des* 67(5): 370-2.
152. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA. 2006. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aided Mol Des* 20(10-11): 647-71.
153. Labute P, Williams C, Feher M, Sourial E, Schmidt JM. 2001. Flexible alignment of small molecules. *J Med Chem* 44(10): 1483-90.
154. Mortier J, Rakers C, Bermudez M, Murgueitio MS, Riniker S, Wolber G. 2015. The impact of molecular dynamics on drug design: applications for the characterization of ligand-macromolecule complexes. *Drug Discov Today* 20(6): 686-702.
155. McCammon JA, Gelin BR, Karplus M. 1977. Dynamics of folded proteins. *Nature* 267(5612): 585-90.
156. van Gunsteren WF. 1993. Molecular dynamics studies of proteins. *Current Opinion in Structural Biology* 3(2): 277-281.
157. Rakers C, Bermudez M, Keller BG, Mortier J, Wolber G. 2015. Computational close up on protein-protein interactions: how to unravel the invisible using molecular dynamics simulations? *Wiley Interdisciplinary Reviews: Computational Molecular Science* 5(5): 345-359.
158. Berendsen HJC, van der Spoel D, van Drunen R. 1995. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* 91(1): 43-56.
159. Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. 2009. CHARMM: the biomolecular simulation program. *J Comput Chem* 30(10): 1545-614.
160. Case DA, Cheatham TE, 3rd, Darden T, Gohlke H, Luo R, Merz KM, Jr., Onufriev A, Simmerling C, Wang B, Woods RJ. 2005. The Amber biomolecular simulation programs. *J Comput Chem* 26(16): 1668-88.
161. Labute P. 2008. The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area. *J Comput Chem* 29(10): 1693-8.
162. Copeland RA, Kinetics of single-substrate enzyme reactions. In *Enzymes*, John Wiley & Sons, Inc.: 2002; pp 109-145.

163. Vietri M, Pietrabissa A Fau - Spisni R, Spisni R Fau - Mosca F, Mosca F Fau - Pacifici GM, Pacifici GM. 7-OH-flavone is sulfated in the human liver and duodenum, whereas 5-OH-flavone and 3-OH-flavone are potent inhibitors of SULT1A1 activity and 7-OH-flavone sulfation rate. (0049-8254 (Print)).
164. Irwin JJ, Shoichet BK. 2005. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45(1): 177-82.
165. Kakuta Y, Pedersen LG, Carter CW, Negishi M, Pedersen LC. 1997. Crystal structure of estrogen sulphotransferase. *Nat Struct Biol* 4(11): 904-8.
166. Visser TJ, Kaptein E, Glatt H, Bartsch I, Hagen M, Coughtrie MW. 1998. Characterization of thyroid hormone sulfotransferases. *Chem Biol Interact* 109(1-3): 279-91.
167. 2015. SZMAP 1.2.1.4: OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>.
168. Tanger JC, Pitzer KS. 1989. Calculation of the thermodynamic properties of aqueous electrolytes to 1000.degree.C and 5000 bar from a semicontinuum model for ion hydration. *Journal of Physical Chemistry* 93(12): 4941-4951.
169. 2015. GAMEPLAN, version 1.2.1, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, www.eyesopen.com.
170. Wolber G, Seidel T, Bendix F, Langer T. 2008. Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov Today* 13(1-2): 23-9.
171. Cook I, Wang T, Almo SC, Kim J, Falany CN, Leyh TS. 2013. Testing the sulfotransferase molecular pore hypothesis. *J Biol Chem* 288(12): 8619-26.
172. Chen G, Rabjohn PA, York JL, Wooldridge C, Zhang D, Falany CN, Radominska-Pandya A. 2000. Carboxyl residues in the active site of human phenol sulfotransferase (SULT1A1). *Biochemistry* 39(51): 16000-7.
173. Tanrikulu Y, Kruger B, Proschak E. 2013. The holistic integration of virtual screening in drug discovery. *Drug Discov Today* 18(7-8): 358-64.
174. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7(1).
175. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. 2013. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res* 41(Web Server issue): W597-600.
176. Laurent B, Chavent M, Cragolini T, Dahl AC, Pasquali S, Derreumaux P, Sansom MS, Baaden M. 2015. Epock: rapid analysis of protein pocket dynamics. *Bioinformatics* 31(9): 1478-80.
177. Labute P. 2009. Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* 75(1): 187-205.

178. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. 2003. Improved protein-ligand docking using GOLD. *Proteins* 52(4): 609-23.
179. Kaus JW, Pierce LT, Walker RC, McCammont JA. 2013. Improving the efficiency of free energy calculations in the Amber molecular dynamics package. *J Chem Theory Comput* 9(9).
180. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65(3): 712-25.
181. Roe DR, Cheatham TE, 3rd. 2013. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* 9(7): 3084-95.
182. Riches Z, Bloomer JC, Coughtrie MW. 2007. Comparison of 2-aminophenol and 4-nitrophenol as in vitro probe substrates for the major human hepatic sulfotransferase, SULT1A1, demonstrates improved selectivity with 2-aminophenol. *Biochem Pharmacol* 74(2): 352-8.
183. Sakakibara Y, Takami Y, Nakayama T, Suiko M, Liu MC. 1998. Localization and functional analysis of the substrate specificity/catalytic domains of human M-form and P-form phenol sulfotransferases. *J Biol Chem* 273(11): 6242-7.
184. Sugahara T, Pai TG, Suiko M, Sakakibara Y, Liu MC. 2003. Differential roles of human monoamine (M)-form and simple phenol (P)-form phenol sulfotransferases in drug metabolism. *J Biochem* 133(2): 259-62.
185. Parker MH, McCann DJ, Mangold JB. 1994. Sulfation of di- and tricyclic phenols by rat liver aryl sulfotransferase isozymes. *Arch Biochem Biophys* 310(2): 325-31.
186. Ung D, Nagar S. 2007. Variable sulfation of dietary polyphenols by recombinant human sulfotransferase (SULT) 1A1 genetic variants and SULT1E1. *Drug Metab Dispos* 35(5): 740-6.
187. Vaidyanathan JB, Walle T. 2002. Glucuronidation and sulfation of the tea flavonoid (-)-epicatechin by the human and rat enzymes. *Drug Metab Dispos* 30(8): 897-903.
188. Zhang L, Huang M, Blair IA, Penning TM. 2012. Detoxication of benzo[a]pyrene-7,8-dione by sulfotransferases (SULTs) in human lung cells. *J Biol Chem* 287(35): 29909-20.
189. Vietri M, De Santi C, Pietrabissa A, Mosca F, Pacifici GM. 2000. Fenamates and the potent inhibition of human liver phenol sulphotransferase. *Xenobiotica* 30(2): 111-6.
190. Harris RM, Wood DM, Bottomley L, Blagg S, Owen K, Hughes PJ, Waring RH, Kirk CJ. 2004. Phytoestrogens are potent inhibitors of estrogen sulfation: implications for breast cancer risk and treatment. *J Clin Endocrinol Metab* 89(4): 1779-87.
191. Rao SI, Duffel MW. 1991. Benzylic alcohols as stereospecific substrates and inhibitors for aryl sulfotransferase. *Chirality* 3(2): 104-11.
192. Duffel MW. 1994. Molecular specificity of aryl sulfotransferase IV (tyrosine-ester sulfotransferase) for xenobiotic substrates and inhibitors. *Chem Biol Interact* 92(1-3): 3-14.

-
193. Wang LQ, Lehmler HJ, Robertson LW, James MO. 2006. Polychlorobiphenyls are selective inhibitors of human phenol sulfotransferase 1A1 with 4-nitrophenol as a substrate. *Chem Biol Interact* 159(3): 235-46.
194. Eaton EA, Walle UK, Lewis AJ, Hudson T, Wilson AA, Walle T. 1996. Flavonoids, potent inhibitors of the human P-form phenolsulfotransferase. Potential role in drug metabolism and chemoprevention. *Drug Metab Dispos* 24(2): 232-7.
195. Mesia-Vela S, Kauffman FC. 2003. Inhibition of rat liver sulfotransferases SULT1A1 and SULT2A1 and glucuronosyltransferase by dietary flavonoids. *Xenobiotica* 33(12): 1211-20.
196. Banoglu E, King RS. 2002. Sulfation of indoxyl by human and rat aryl (phenol) sulfotransferases to form indoxyl sulfate. *Eur J Drug Metab Pharmacokinet* 27(2): 135-40.
197. Spink BC, Katz BH, Hussain MM, Pang S, Connor SP, Aldous KM, Gierthy JF, Spink DC. 2000. SULT1A1 catalyzes 2-methoxyestradiol sulfonation in MCF-7 breast cancer cells. *Carcinogenesis* 21(11): 1947-57.
198. Adjei AA, Weinshilboum RM. 2002. Catecholesterogen sulfation: possible role in carcinogenesis. *Biochem Biophys Res Commun* 292(2): 402-8.

Published articles

1. Design, synthesis and molecular docking study of novel quinoxalin-2(1H)-ones as anti-tumor active agents with inhibition of tyrosine kinase receptor and studying their cyclooxygenase-2 activity. Galal SA, Khairat SH, Ragab FA, Abdelsamie AS, Ali MM, Soliman SM, Mortier J, Wolber G, El Diwani HI. *Eur J Med Chem* (2014) 15;86:122-32.
2. Synthesis, Biological Evaluation, and Docking Studies of New 2-Furylbenzimidazoles as Anti-Angiogenic Agents: Part II. Ahmed Temirak, Yasser M. Shaker, Fatma A. F. Ragab, Mamdouh M. Ali, Salwa M. Soliman, Jeremie Mortier, Gerhard Wolber, Hamed I. Ali, Hoda I. El Diwani. *Arch. Pharm. Chem. Life Sci.* 347 (2014), 1-14.
3. Design, Synthesis and Structure-Activity Relationship of Novel Quinoxaline Derivatives as Cancer Chemopreventive Agent by Inhibition of Tyrosine Kinase Receptor. S. Galal, A. Abdelsamie, S. Soliman, J. Mortier, G. Wolber, M. Ali, H. Tokuda, N. Suzuki, A. Lida, R. Ramadan, H. Diwani. *European Journal of Medicinal Chemistry* 69 (2013) 115-124
4. Baclofen ester and carbamate prodrug candidates: A simultaneous chromatographic assay, resolution optimized with DryLab®. R. Hanafi , S. Mosad , K. Abouzid, R. Nieß, H. Spahn-Langguth. *Journal of Pharmaceutical and Biomedical Analysis* 56 (2011) 569-576

Article in preparation

1. Ligand-size dependent substrate binding site plasticity of human sulfotransferase affects binder activity: a molecular dynamic (MD) study of SULT1A1 with high affinity binders. Salwa M. Solimana,b, Remo Perozzo, Gianpaolo Chiriano**b**, Zachary Tibbs, Jérémie Mortier **a**, Charles Falany, Leonardo Scapozza**b** and Gerhard Wolber**a**