SEMMELWEIS EGYETEM
DOKTORI ISKOLA

Ph.D. értekezések

3112.

TÓTH DÁNIEL

A gyógyszerészeti tudományok korszerű kutatási irányai
című program

Programvezető: Dr. Antal István, egyetemi tanár

Témavezetők: Dr. Balog Erika, egyetemi docens
Dr. Maria Miteva, egyetemi tanár

MOLECULAR MECHANISMS OF PHASE II DRUG METABOLIZING ENZYMES SULT1 AND THEIR INTERACTIONS WITH SMALL LIGANDS MODELED THROUGH COMPUTATIONAL APPROACHES

Ph.D. thesis

Tóth Dániel

Semmelweis University Doctoral School
Pharmaceutical Sciences and Health Technologies Division

Supervisors:	Erika Balog Ph.D. Maria Miteva Ph.D.
Official reviewers:	Nathalie Lagarde, Ph.D. Zoltán Gáspári, Ph.D.
Head of the Complex Examination Committee:	Romána Zelkó, D.Sc.
Members of the Complex Examination Committee:	András Czirók Ph.D. Christian Jelsch Ph.D.

Budapest
2024

Université Paris Cité

Doctoral School MTCI 563

UMR 8038 Cibles Thérapeutiques et Conception de Médicaments CNRS – Univ. Paris Cite, MCTR Inserm U1268

Molecular mechanisms of phase II drug metabolizing enzymes SULT1 and their interactions with small ligands modeled through computational approaches

by Dániel Tóth

Doctoral Thesis in Biology, Medicine, and Health
Specialty: Bioinformatics

Directed by Dr Maria Miteva
and Dr Erika Balog

Defense scheduled for 9 December 2024

In front of the jury composed of:

Pr. Romána Zelkó	Faculty of Pharmacy, Semmelweis U.	examiner
Dr. Nathalie Lagarde	Conservatoire National des Arts et Métiers	reporter
Dr. Zoltán Gáspári	Fac. of IT and Bionics, Pázmány Péter Catholic U.	reporter
Dr. András Czirók	Department of Biological Physics, Eötvös U.	examiner
Dr. Christian Jelsch	CRM2 CNRS Univ. Lorraine	examiner
Dr. Maria Miteva	UMR 8038 CITCOM CNRS – Univ. Paris Cite, MCTR Inserm U1268	thesis director
Dr. Erika Balog	Institute of Biophysics and Radiation Biology, Semmelweis U.	co-thesis director

Table of Contents

LIST OF ABBREVIATIONS

ADME-Tox	Absorption, Distribution, Metabolism, and Excretion, Toxicity
ADR	Adverse Drug Reactions
ATP	Adenosine Triphosphate
AUC	Area Under the Curve
CYP	Cytochrome P450 enzyme
DMEs	Drug-Metabolizing Enzymes
DDI	Drug-Drug Interactions
E2	17β-Estradiol
FDA	Force Distribution Analysis
FF	Force Field
MD	Molecular Dynamics
MDeNM	Molecular Dynamics with excited Normal Modes
ML	Machine Learning
MLD	Multidrug Resistance
NMR	Nuclear Magnetic Resonance
PAPS	3'-Phosphoadenosine-5'-Phosphosulfate
PCA	Principal Component Analysis
PDB	Protein Data Bank
QM/MM	Quantum Mechanics/Molecular Mechanics
(Q)SAR	(Quantitative) Structure-Activity Relationship
Rgyr	Radius of Gyration
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SNP	Single Nucleotide Polymorphism
SULT	Cytosolic Sulfotransferase Enzyme
XRD	X-ray Diffraction

1. INTRODUCTION

1.1. Biological introduction

1.1.1. Research Background

It is well known that drug discovery is a long and costly process, often taking over a decade and billions of dollars (1). However, advancements in biotechnology and bioinformatics are speeding up the process and increasing success rates considerably. Automated processes are routinely used to identify potential drug candidates (2), often by advanced technologies like high-throughput screening, computational modeling, and even machine learning (3) to find compounds that interact mainly with protein targets, achieving the intended therapeutic effect.

The main stages in drug discovery include target and hit identification, hit-to-lead discovery, lead optimization, preclinical testing, and clinical trials. Each stage refines potential drug candidates to improve efficacy, safety, and pharmacokinetics. Understanding pharmacokinetics and pharmacodynamics, including how drugs are metabolized and interact with other medications, is one of the most complex and challenging stage, because of the many participants and complex molecular interactions. Crucial aspect are drug metabolism and transport, which include enzymes, transport proteins, co-factors, cellular components, and even external factors. Our research focuses on drug metabolizing enzymes (DMEs), thus I provide here their biological background, a complex and diverse field on its own.

1.1.2. DME Biotransformation

DMEs are part of the biotransformation process, which transforms small molecules, whether they are absorbed (xenobiotics) or produced by the body (endogenous) (4). Xenobiotics are foreign compounds to the body, such as industrial chemicals and environmental toxins, but most importantly, pharmaceuticals. If not transformed and excreted, these substances accumulate and disrupt normal cell function, commonly resulting in toxic effects. Similarly, drugs that exert therapeutic effects must be inactivated to prevent overdosing or adverse effects once they have accomplished their intended function.

Endogenous compounds, on the other hand, are naturally generated compounds within the body, including hormones and other signaling molecules. They also need to be regulated and removed when in excess or when their function has been fulfilled. The enzymes of biotransformation achieve this goal by facilitating the transfer of a chemical group between these small molecules and the enzymes' cofactors. To facilitate the elimination of these substances, biotransformation processes increase their water solubility. Many xenobiotic and endogenous compounds are lipophilic, meaning they dissolve better in lipids than in water. This property helps them cross cell membranes to exert their effects but also makes them difficult to excrete since the body primarily eliminates waste through aqueous fluids like urine and bile (5). By increasing the water solubility of these compounds, biotransformation enzymes ensure that they can be readily dissolved in bodily fluids and efficiently excreted from the body.

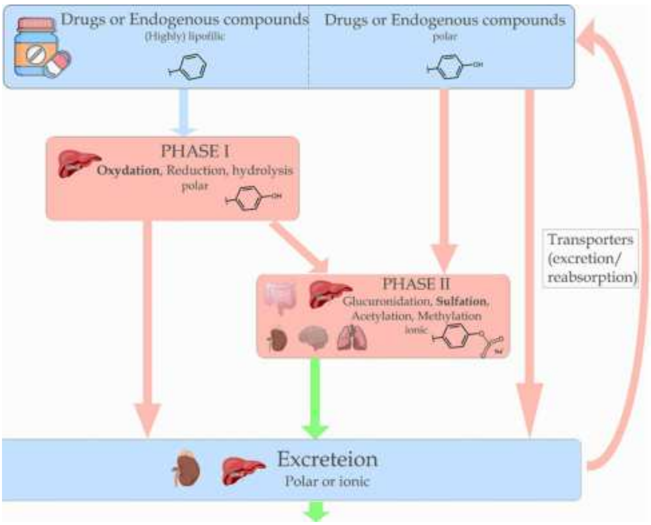

Figure 1. *Illustration of the metabolic processes of DMEs, with pictograms indicating the specific organs involved. The intracellular processes are indicated in red, while the extracellular compartments are indicated in blue. The green arrows indicate the direction of excretion*

Overall, the biotransformation system is a sophisticated mechanism that maintains homeostasis by managing and eliminating substances that are either foreign or in excess within the body. Through a series of enzymatic reactions, these molecules are chemically altered to normally render them less active and more water-soluble, thereby safeguarding the body from potential toxicity and preserving the delicate balance of internal biochemical processes. Yet, in some cases, more toxic and reactive metabolites can be created (6). Biotransformation is divided into two main phases, based on the biochemical transformation facilitated on these molecules.

Phase I Metabolism

Phase I is the so-called functionalization step, where a functional group is created on a molecule. That increases the hydrophilic character, which aids in the excretion of the compound by enhancing its solubility in aqueous bodily fluids, such as urine. However, this increased hydrophilicity generally reduces permeability across lipid membranes, making it harder for the compound to pass through without the assistance of transport mechanisms. These reactions are divided into the following groups based on the molecular transformation taking place: oxidation and reduction, where new covalent bonds are created or removed; hydration, when a water molecule is absorbed; and hydrolysis, when a water molecule is used to form two new molecules. Some other, rarer reactions include isomerization, dimerization, and decarboxylation (4).

From a pharmacological point of view, oxidation, more precisely the Cytochrome P450 (CYP450) enzyme family is the most significant (7), as it accounts for more than 75% of drug metabolism in Phase I. Therefore, considerable effort has been made to explore this diverse enzyme family (8, 9). However, there are other important reactions in metabolism. Although oxidation creates a more polarized or often ionized hydrophilic molecule, which sometimes in itself is sufficient for elimination, further transformations are often required. These additional transformations can occur as consecutive Phase I reactions or as Phase II transformation steps.

Phase II Metabolism

Also known as the conjugation phase, Phase II reactions occur when an existing functional group is extended or exchanged with a new one. This process creates products with increased water solubility and molecular weight, making them easier to excrete in urine or bile. Phase II metabolism includes reactions such as sulfation, glucuronidation, glycosidation, methylation, acetylation, condensation, and conjugation with amino acids, glutathione, or fatty acids.

The majority of DMEs in Phase II are involved in glucuronidation, where a glucose derivative is conjugated to a molecule. This is an effective way of eliminating small molecules, many of which are pharmaceuticals. The enzymes responsible for this process are known as uridine 5'-diphosphate-glucuronosyltransferases (UGTs). Another important conjugation reaction is sulfonation, where a sulfate group is transferred from the nucleotide cofactor 3'-Phosphoadenosine-5'-phosphosulfate (PAPS) to an oxygen or nitrogen atom of the molecule, resulting in a sulfonated and more polarized product. This reaction is catalyzed by sulfotransferases (SULTs) (10).

It is important to note that Phase I and Phase II reactions do not necessarily follow each other. Some molecules are conjugated without prior functionalization, and some molecules become pharmacologically active only after conjugation (so called pro-drugs). This highlights the necessity of studying all metabolizing enzymes, rather than focusing solely on the first step involving CYP450 enzymes.

1.1.3. SULT enzymes

The first historical record of sulfated metabolites dates back to 1851 when Eugene Baumann isolated phenol-sulfate from the urine of patients treated with carbolic acid, a common disinfectant at that time. However, it took a hundred years to understand the mechanism fully, with Fritz Lipmann identifying the PAPS cofactor. It was revealed that the process involved sulfonation rather than sulfatation, though the terminology remained unchanged (11, 12).

SULTs were initially identified in rat and bovine DNA. As their function became clear, screening cDNA libraries quickly led to the identification of human SULT enzymes (13). It is worth noting that another class of membrane-bound SULTs exists in Golgi vesicles, but they are not discussed in this thesis, instead we are focusing on the cytosolic SULTs.

Cytosolic SULT nomenclature was proposed in 2004, based on the cDNA sequence of the enzyme (14). Two SULT enzymes are considered members of the same family if they share at least 45% sequence identity, in which case they are denoted with the same Arabic numeral, (e.g. SULT1). If their similarity is higher, sharing 60% of the same sequence, they are further designated with the same alphabetical letter (e.g. SULT1A). Two SULTs cDNAs sharing 97% identity are delineated by a numeric title. (e.g. SULT1A1).

SULT nomenclature also includes a species prefix in parentheses, such as [HUMAN] SULT1A3, or simply the first letter of the species, like hSULT. Not all SULTs are expressed in humans, for example SULT3 and SULT5 were not identified in human DNA. Additionally, these enzymes typically have multiple alloforms with single nucleotide polymorphism (SNPs), though our research and results were based solely on the most common wild-type variants of each isoenzyme in question. As of the latest information from the HUGO Gene Nomenclature Committee (15), there are thirteen recognized SULTs, with the most important with the most important characteristics summarized in Table 1.

SULT1 (phenol sulfotransferase, PST family): The family's subdivision is based on the molecules they metabolize. The SULT1A isoforms are the primary metabolizing enzymes of phenolic compounds such as paracetamol, dopamine, and estrogens, often referred to as phenol or aryl SULTs (16). SULT1B isoforms catalyze the sulfate conjugation of larger phenolic substrates like pregnenolone and DHEA, and also exhibit high affinity for thyroid hormones (17). SULT1C is involved in sulfation of alkylated polycyclic hydrocarbons, p-nitrophenol, acetaminophen, and hydroxysteroids (18, 19). Finally, SULT1E enzymes prefer fused multiringed structures such as estrogens, estrone, and various synthetic estrogens as substrates. They also share some phenolic substrates with the SULT1A subfamily, but with much lower affinity (20).

The SULT2 family includes two subfamilies, SULT2A and SULT2B, which typically catalyze the sulfate conjugation of 3β-hydroxy groups of steroids such as DHEA, androgens, pregnenolone, bile acids, and benzylic alcohols of polycyclic aromatic hydrocarbons. While both subfamilies metabolize similar substrates, members of the SULT2B subfamily primarily act as cholesterol SULTs (21).

Much less is known of the last two enzymes. SULT4A1, primarily expressed in the brain, binds to 2-hydroxyestradiol, thyroid hormone T4, and catecholamines except dopamine. Its structural diversity, particularly in the PAPS binding site, suggests it may modulate these compounds through a unique mechanism (22). SULT6B1 specific substrates and biological roles remain under investigation, with studies suggesting activity towards thyroxine and bithionol (23). This variety of SULTs is due to genetic evolution, as cells must contend with new types of xenobiotics every day. This also means genetic differences between individuals may affect enzyme function, resulting in different responses to drugs, both in therapeutic and adverse effects, however the general function of the SULTs are similar.

1.1.4. SULT structural description

The first 3D structure of a cytosolic SULT was that of murine SULT1E1, determined using X-ray crystallography (24). Since then, numerous SULTs have been identified, and to date, the structures of 42 hSULTs have been deposited in the Protein Data Bank (25). Given their similar basic structural elements, we can provide a general overview applicable for every cytosolic SULT.

All SULTs feature a central 4-stranded parallel β-sheet surrounded by 12 to 13 α-helices, forming a rigid core for the enzyme. Several essential domains crucial to enzyme function are highly conserved, including the internal regions of the substrate binding pocket (active site), PAPS binding pocket (PSB), and the dimerization region.

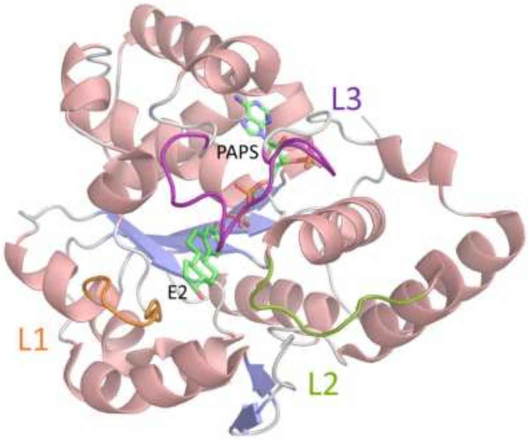

Figure 2. *Structural motifs of hSULT1A1 functional loops surrounding the binding site, L1 (orange), L2 (green) and L3 (purple). The cofactor PAPS and substrate E2 are shown as sticks. α-helices are colored salmon, while β-sheet strands are colored slate blue*

These loops serve as regulatory gates for the specificity of individual isoenzymes, although they are not universally present across all SULTs. The SULT1 family possesses three loops, SULT2 has two (L1 is absent), and SULT4 and SULT6 lack portions of two loops, retaining only a shortened version of L3 while L2 is largely unaltered.

In the SULT1 family, the focus of this thesis, Loop 1 (amino acids 81-90) is the smallest of the three, characterized by 9 highly variable amino acids among isoforms. This variability, contributes significantly to the selective binding of specific substrates. Loop 2 (amino acids 145-154) is the most conserved loop among the family. It contains several key residues important for substrate binding and the catalytic function of the enzyme. Loop 3 (amino acids 236-258) is also referred to as the cap or the lid, depending on the source. It is the largest of the three loops. It covers the outer region of both the substrate binding pocket and a portion of the PAPS binding pocket adjacent to the dimerization region.

Previous studies have highlighted the importance of these loops, attributing substrate specificity and selectivity largely to sequence variations within them (26), an observation corroborated by our own research (27). However, there are also highly conserved regions essential for sulfotransferases. One such region spans over the PAPS binding site adjacent to Loop 3 and the dimerization region. A study on single point mutations (SPM) highlighted its significance; mutations in this area notably altered the catalytic properties of SULTs (28), as changes in residues here significantly altered the catalytic properties of them, demonstrating its direct involvement in the sulfonation process. For instance, triple mutations in the GxxGxxK motif to AxxAxxA rendered the enzyme inactive. The residue K265, which is the first residue of the dimerization region, is a region highly conserved across all human SULTs. In fact, all known human SULTs are homodimers. comprising a few residues stabilizing the two monomers through strong ionic forces, with the general sequence KxxxTVxxxE (29). This efficient structure resembles a pair of hands holding each other, but with the same sequence reversely aligned, the positive amine group of K265 meeting the negative carboxyl of E274 acid to form a stable connection.

The sulfation reaction catalyzed by SULTs has been elucidated through mutation analysis ((28). In mouse SULT1E1, L47M mutation inhibits SULT1E1 activity, while L47R results in a lower k_{cat} with a similar K_m for PAPS. This observation confirms that the positive charge of the guanidinium group is essential for the sulfation reaction. Similarly, mutations of K106 and His108 have been shown to have inhibitory effects.

1.1.5. Mechanism of sulfate transfer

The precise term is sulfonation, since the product is a sulfonic acid functional group, even though a sulfonate group is transferred. Although energetically favorable, it does not occur spontaneously and requires the bond-loosening roles of two specific amino acids: H108 and K48. H108 deprotonates the hydroxyl group (or amine group) of the bound substrate, while the Lysine weakens the sulfate bond of PAPS, enabling a nucleophilic attack of the substrate on the sulfur atom of PAPS. After some bond rearrangement, the formation of the new sulfate ester is stable, while PAPS is converted to adenosine 3',5'-bisphosphate (PAP^-). The sulfated substrate is then released, and PAP^- is reprotonated by the Histidine (28).

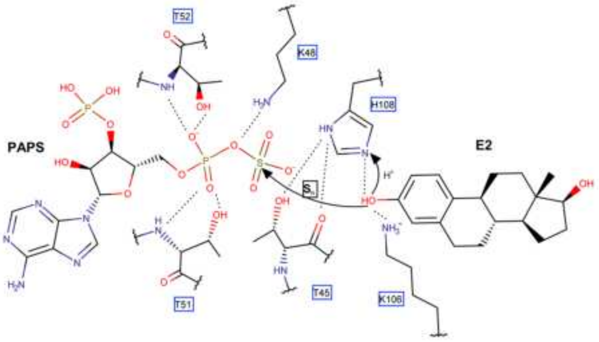

Figure 3. *Sulfate transfer from the cofactor PAPS (left) to the fenoxy group of E2 (right). Arrows indicate the electrophilic attack of the sulfur atom after the hydroxyl hydrogen has been removed by residue H108. Other important residues in bond weakening and stabilizing residues are shown with their effects indicated by dotted lines. Figure based on the article of Kakuta et al. (28)*

1.1.6. SULT substrate specificity

The traditional enzyme nomenclature, based on substrate specificities proved inadequate when the SULT enzymes were classified. As more information emerged in the '90s, it became evident that these enzymes possess unique characteristics. They often exhibit significant overlap in substrates while also having distinct sets of ligands. For instance, SULT1 and SULT2 are known as phenolic (aryl) and hydroxysteroid (alcohol) SULTs, respectively, both accepting similar substrates, especially multiple-ring compounds. Specifically, SULT1A isoforms predominantly metabolize phenolic compounds, hence their designation as phenol or aryl SULTs, whereas SULT1B isoforms show a high affinity for thyroid hormones.

Isoenzyme	Reference ligands	Localization	PDB ID	References
hSULT1A1	4-methylphenol, fulvestrant	intestines, liver	1LS6, 2D06, 3QVU, 3QVV, 3U3K, 3U3O, 3U3M, 3U3R, 3U3J, 4GRA, 1Z28	[34, 35]
hSULT1A2	naloxone, minoxidil	liver	1Z29	[35]
hSULT1A3 (hSULT1A4)	dopamine, catecholamines	intestines, brain	2A3R, 1CJM	[34] [36]
hSULT1B1	3,3',5-triiodo-L-thyronine	small intestine, kidney, liver	3CKL, 2Z5F	[35] [36]
hSULT1C2	hydroxysteroids p-nitrophenol	stomac	3BFX	[35]
hSULT1C3	hydroxy polychlorobiphenyls	GI-tract, kidney	2H8K, 2REO	[37], [38] proteinatlas.org/ ENSG00000196228-SULT1C3/tissue
hSULT1C4	-	female reproductive organs, endokrin tissues	2GWH, 2AD1	[38] ENSG00000198075-SULT1C4/tissue
hSULT1E1	Estradiol, 17β-estradiol	lung intestines, kidney, liver	4JVL, 4JVM, 1G3M, 4JVN, 1HY3	[17] [35] [39]
hSULT2A1	Dehydroepiandrosterone (DHEA), Butorphanol	endocrine tissues, liver, intestines	1OV4, 1J99, 2QP3, 2QP4, 3F3Y, 4IFB, 1EFH	[35]
hSULT2B1	-	reproductive organs, skin, lungs	1Q1Q, 1Q22, 1Q20, 1Q1Z	[17, 38] ENSG00000088002-SULT2B1/tissue,
hSULT4A1	neurosteroids and Guanidinosuccinic acid (GSA)	brain	1ZD1	[23], [40]
hSULT6B1	GSA	testis, brain	-	[23, 38] ENSG00000138068-SULT6B1/tissue

Table 1. *Summary of the cytosolic hSULT family, with representative substrates and tissue localization (41)*

Despite being classified in the SULT1 family, hSULT1E1 is highly active in estrogen sulfonation. Similarly, SULT1A3 is specific to monoamine neurotransmitters, even though its sequence is most similar to SULT1A1, one of the least selective SULTS. It was revealed, that despite their sequence similarity, hSULT1B1 and hSULT1E1 exhibit different activities. While hSULT1E1 shows high activity toward estrogens, hSULT1B1 is unable to sulfate estrogenic steroid (30). These relationships are common in the SULT family, and it illustrates that the specificity of SULT isoforms is not necessarily linked to their global sequence similarity.

One of the strangest phenomenon is the difference in substrate preference between hSULT1A1 and hSULT1A3. They share 96% of their sequences despite their very different, disparate substrates. Sequence analysis revealed a few key amino acid residues, which were later verified as the reason for the drastic change in flexibility and selectivity of the isoforms' active site, explaining the diversity of their substrates. Similarly, hSULT1E1 crystal structures reveal a narrow, slender binding pocket, while hSULT1A1 has a much wider, almost pore-like pocket, enabling it to sulfate a broader range of substrates.

There are still multiple other examples in the literature (31-33), but in general, SULTs active site size and shape correlates with their substrate size and diversity (26). However, these speculations can also be misleading, as the pocket shape and size can change in response to co-factor binding. It is well observed that the largest loop in ligand selection, Loop 3, experiences dynamic changes with PAPS binding, which can either inhibit or enhance substrate binding efficiency. Moreover, other less obvious reasons, like allosteric effect of the dimerization might play a role in the modulation of the loops that govern sulfotransferase selectivity.

1.1.7. Dimerization

All cytosolic human SULTs function as physiological homodimers, with the human SULT isoforms possessing a conserved dimerization region consisting of 10 residues with the consensus sequence KxxxTVxxxE near the C-terminus of the protein (29, 42). The biological significance of this dimerization is not entirely clear, but it has been shown to affect the catalytic efficiency and stability of the enzyme (43).

Moreover, symmetrical oligomer enzymes often function through a mechanism called "half-site reactivity," where only part of the oligomer is in a catalytically active conformation at any given time. This occurs due to a global asymmetric conformation, in which each subunit of a dimer, like SULTs, alternates between two distinct states, each with a unique ability to interact with substrates or catalyze the reaction. Although these asymmetries are not apparent in all crystal structures of half-site reactive enzymes, multiple studies have reported half-site reactivity in SULT isoforms (44-46). These results, and the proximity of the PAPS binding region(L4), the active site cap (L3) and the dimer interface suggest interaction between the two subunits that may affect enzyme activity. One theory proposed a negative cooperativity between the PAPS binding sites of dimeric subunits is likely the driving force behind half-site reactivity. This could provide the system with directionality, favoring the release of products while promoting substrate binding, similar to the functionality of ribosomes (26).

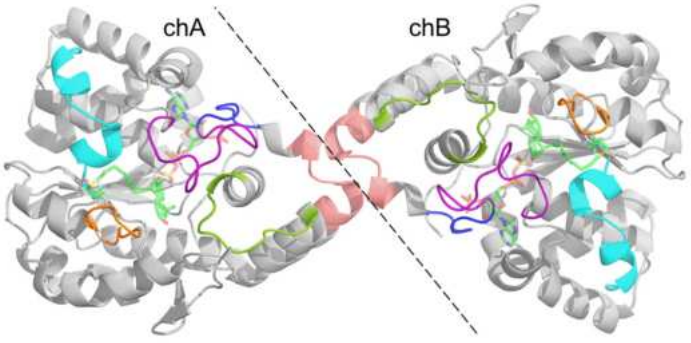

Figure 4. *SULT1A1 constructed dimer based on the crystal structure (PDBID 2D06). The cofactor PAPS and the ligand fulvestrant are shown as sticks. The major loops and domains are colored. L1(Lip): orange, L2: green, L3(Cap): magenta, L4: blue, dimerization region: pink and α3L5, a flexible domain in cyan*

1.1.8. Studied SULT enzymes

<u>*SULT1A1*</u>

This isoform, previously known as thermostable phenol sulfotransferase, is most abundant in the liver, where it reaches near-gram quantities (47), however it can also be found in the kidney, small intestine, and lung tissue as well (39). SULT1A1 has the broadest substrate specificity in the SULT enzyme superfamily, catalyzing the sulfonation of endogenous molecules like 17β-estradiol and dehydroepiandrosterone, as well as xenobiotic compounds such as 4'-hydroxy-nitrophenol and drugs like paracetamol and minoxidil with high affinity (48). Deficiency in its function have been linked to numerous diseases, including breast cancer (49), cystic fibrosis (50), diabetes, hemophilia (51) and Parkinson disease (52), indicating its diverse activity in the human body.

From a structural perspective SULT1A1 is relatively rigid. However, the enhanced flexibility of L1, coupled with the presence of numerous phenolic functional groups within the binding site, allows for a diverse range of conformations, enabling the binding of a wide variety of cyclic compounds capable of π-π stacking. Single-nucleotide polymorphisms (SNPs) have been identified with considerable frequency in SULT enzymes. Although this study does not investigate SNPs, an understanding of their behavior is informative from a structural perspective. The most prevalent SULT1A1 polymorphisms are SULT1A1*1, which is the wild type and the focus of this research project. All simulations conducted as part of this study are derived from this specific sequence. The SULT1A1*2 variant, which contains the R213H substitution becomes inactive, and the SULT1A1*3 variant, which contains the M223V substitution, has been observed to exhibit decreased activity. Additionally, SULT1A1*2 displays reduced thermostability (41).

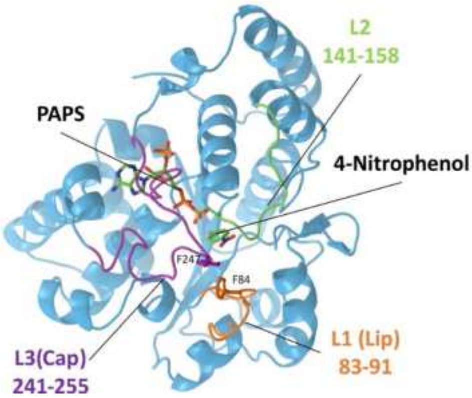

Figure 5. *Crystal structure of the monomer human SULT1A1 (PDBID 4GRA). The active co-factor PAPS, and the representative ligand, 4-Nitrophenol are shown in stick representation. The characteristic three loops gating the active sites are indicated as L1(Lip): orange L2: green and L3(Cap): magenta. Residues important for substrate recognition are labelled and represented as sticks (27)*

SULT1A3

In fact, SULT1A3/SULT1A4 constitutes an alloenzyme pair, presumably derived from gene duplication, with some SNPs influencing their metabolism of certain analgesic pharmaceuticals (53). However, due to their similar activity, we can solely focus on SULT1A3.

SULT1A3, previously named human catecholamine sulfotransferase, is highly specific for monoamine neurotransmitters. These small endogenous molecules, including dopamine, serotonin, adrenaline, and noradrenaline (54) are released into the body in response to physical or emotional stress. Unlike SULT1A1, SULT1A3 is primarily expressed in the gastrointestinal tract, with significant presence in the central nervous system, where it has also been reported to metabolize many opioid-like substances. Interest in the two isoenzymes arises from their structural similarity (93% sequence identity) despite their very different roles. Our second article focused on this phenomenon, identifying key residues that contribute to the increased flexibility of loops, modifying the gating mechanism and facilitate the recognition of monoamine functional groups in catecholamines. Several SNPs have been discovered for SULT1A3 resulting in amino acid (AA) alterations. Similar to SULT1A1, the wild type SULT1A3*1 is the most active and stable variant, therefore we conducted our research with that allotype. SULT1A3*2 is substituted in K234N, SULT1A3*3 is substituted in P101L, SULT1A3*4 is substituted in P101H and SULT1A3*5 is substituted in R144C. The SULT1A3*2 allozyme was found to be more susceptible to proteasome-mediated degradation than SULT1A3*1, which likely contributes to its reduced sulfation activity (41).

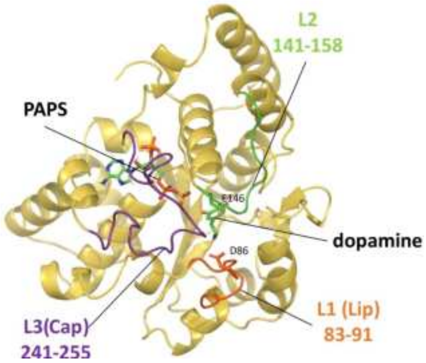

Figure 6. *The crystal structure of monomer hSULT1A3 (PDBID 2A3R). The active co-factor PAPS, and the representative ligand, dopamine are shown in stick representation. Residues important for substrate recognition are labelled and represented as sticks (27)*

1.2. Molecular Modelling Introduction

1.2.1. Computational methods in drug discovery

Effectiveness and safety of pharmaceuticals can be significantly affected by DMEs, making them essential to study for drug discovery. Understanding their interactions has led to an increasing reliance on computational methods due to the high costs and complexity of experimental studies. These computational approaches, such as molecular docking, molecular dynamics (MD), quantum mechanics/molecular mechanics (QM/MM), Coarse-Grained modeling, homology modeling and metadynamics provide information on the dynamic and structural properties of DMEs (55, 56). Ligand-based approaches like structure-activity relationship (SAR) models help identify potential substrates and inhibitors of DMEs. The incorporation of structural information about the target protein can enhance the accuracy of these methods, thereby improving the precision of the prediction of the ADME-Tox profiles of drug candidates. Screening large libraries of chemical compounds is often the first phase of early drug discovery by finding a few promising hit compounds. Mathematical representations of molecular properties are used to predict small molecular descriptors. These descriptors quantify topological, geometrical, physical, and chemical characteristics, such as molecular weight, number of functional groups, and logP. These *in silico* models can be used for virtual screening, enabling the identification of potential drug candidates by predicting their interactions with target proteins based on their molecular properties (57). In a previous study, Martiny et al. (103) employed a structure-based docking-scoring approach to predict ligand binding and interaction energies combined with machine learning methodology for SULT1A1, SULT1A3 and SULT1E1. Three different machine learning methods were employed: support vector machine, random forest, and naïve Bayesian predictors. The performance of each machine learning classification model was evaluated based on the percentage of correctly classified inhibitors for each SULT1 isoform set through leave-one-out cross-validation. The results demonstrated that this protocol effectively identifies potent binders for SULT1 isoforms. The most accurate machine learning models, which incorporated predicted protein-ligand interaction energy through docking, demonstrated a precision of 67.28%, 78.00%, and 75.46% for the SULT1A1, SULT1A3, and SULT1E1 isoforms, respectively.

An often quoted limiting factor in the past was the computational requirements. To illustrate, MD simulation of a protein consisting of approximately 50,000 atoms (such as the SULT1A1 monomer in a cubic water box) for 100 nanoseconds would have required an impractically long time at the turn of the millennium. By 2021, an affordable PC could complete the same calculation in less than a day. However, these simulations are often run on supercomputers on many nodes. The performance of nodes followed the same trend, as a single node of the newly launched Komondor supercomputer can accomplish the aforementioned task in 22 hours. Therefore, reaching the microsecond timescale for relatively large systems in MD simulations is becoming increasingly feasible, which has long been an objective in this field.

1.2.2. Theoretical background of molecular dynamics (MD)

Interatomic interactions modeled by Force Fields

In classical MD simulations the term Force Field (FF) refers to a collection of mathematical functions and parameters that describe the interaction between different types of atoms in different chemical environments (58). The total energy is defined as the sum of bonded and nonbonded interactions:

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{nonbonded}}$$

When this equation is expanded, each term describes an atomic interaction:

$$E(r)_{total} = \sum_{bonds} \frac{k_r}{2}(r - r_0)^2 + \sum_{angles} \frac{k_\theta}{2}(\theta - \theta_0)^2 + \sum_{tors.} \frac{k_\Phi^{(n)}}{2}[1 + \cos(n\Phi - \delta)] \quad + \quad \sum_{vdW} \left[\left(\frac{A_{ij}}{r_{ij}}\right)^{12} - \left(\frac{B_{ij}}{r_{ij}}\right)^{6} \right] + \sum_{electr.} \left(\frac{q_i q_j}{4\pi\varepsilon r_{ij}}\right)$$

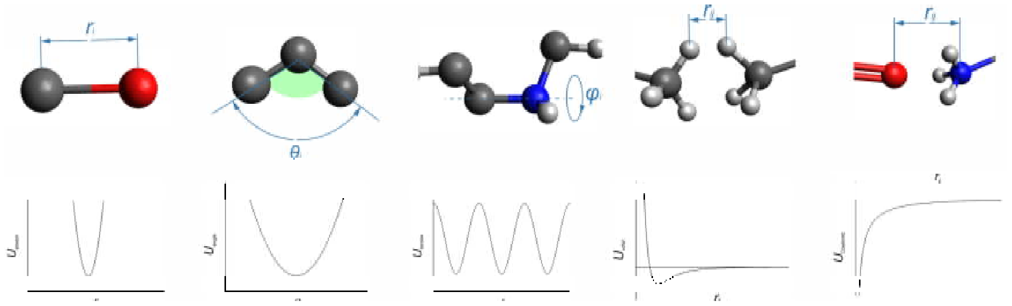

Figure 7. *Classical Energy Functions: A potential energy function with energy terms is presented, along with visual and functional shape representations of distance-force functions*

where the first three term represent the bonded, the last two the nonbonded terms. The first term represents the bond stretching energy between covalently bonded atoms, where k_r is the force constant for a bond length (r) and equilibrium bond length (r_0).

The second term, describes the energy of deformation for bond angles, with k_Θ force constants. Similarly, Θ represents the angle between two covalent bonds while Θ_0 represents the equilibrium angle.

The third term represents the energy of dihedral angles, which describes the rotational barriers around bonds. The periodicity of the potential is given by the cosine function, and δ is the phase angle. Although individually relatively small, the dihedral configurations of the backbone (axis of the Cα-N bond) are responsible for the most important conformational changes in the proteins, since they determine the stable states that produce functional loop rearrangements.

The fourth and fifth terms describe non-bonded interactions, these interactions are typically calculated using a cutoff distance, beyond which the interactions are not considered. A common practice is to use a shifted potential or switching function to smoothly reduce the interactions to zero at the cutoff distance.

The fourth term accounts for the van der Waals forces, which is modeled by the Lennard-Jones potential. A and B are constants depending on the atom types, r_{ij} being the distance between atoms i and j.

The fifth term $\left(\frac{q_i q_j}{4\pi\varepsilon r_{ij}}\right)$ describes the Coulomb forces acting between polar or ionic atoms, where q_i and q_j are the charges of atoms i and j, r_{ij} is the distance between them and the ε is the permittivity.

The parameters in of the force field the formula, such as the spring constant k and the equilibrium bond distances (r_0), are derived from ab initio calculations and experimental measurements (e.g., X-ray diffraction, IR spectroscopy).

Molecular dynamics calculations

Molecular dynamics provides insight into the dynamics, thermodynamics, and molecular changes of atomistic systems. They have been successfully used to study complex dynamic processes, such as the folding and unfolding or the functional movements (e.g., relative domain movements) of proteins (59, 60).

The electrons are not explicitly treated in classical MD simulations, but the Born-Oppenheimer approximation is used, which states that the motion of atomic nuclei and electrons can be separated due to their significant difference in mass and velocity (61). Consequently, the potential energy of the electrons depends only parametrically on the positions of the atomic nuclei. The potential energy surface, on which the nuclei move, is approximated by the force field model described in the previous chapter.

The movement of the atoms can be described using Newton's equations of motion. For a system consisting of N particles interacting via the potential energy function $U(\mathbf{r}_i)$, $i=(1,\ldots,N)$, this can be written as a second order differential equation:

$$-\nabla_{\mathbf{r}_i} U = m_i \frac{d\mathbf{v}_i(t)}{dt} = m_i \frac{d^2\mathbf{r}_i(t)}{dt^2} = \mathbf{F}_i$$

Where F_i is the force acting on atom i, m_i is the mass of the atom. The velocity v_i of i atom is the first derivative with respect to time t, while the position of the atom $\mathbf{r}_i$ is the second derivate with respect to time. U is the potential energy function, and the negative gradient of the potential energy function is $-\nabla \mathbf{r}_i$. The acting force on atom i is given by the negative gradient of the potential energy function. To obtain the state of the system for each time step, either a set of 3N second-order differential equations or an equivalent set of 6N first-order equations must be solved.

Integration Algorithms for Solving Equations of Motion

Second-order algorithms such as the Verlet (62), the Velocity Verlet (63), and the Leapfrog algorithm (64) are used as integration algorithms for MD simulations. The Verlet algorithm uses the Taylor series expansion of the position vector $\mathbf{r}(t)$. The sum of these equations, truncated at the third-order terms (the higher-order terms are neglected), gives the Verlet equation for the updated position vector:

$$\mathbf{r}_i(t+\delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\delta t) + \delta t^2 \mathbf{a}_i(t)$$

The velocity formula can be similarly determined:

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t+\delta t) - \mathbf{r}_i(t-\delta t)}{2\delta t}$$

Neglected higher order terms beyond the third order introduces truncation errors. Additionally, there are arithmetic (rounding) errors due to the finite precision of numerical calculations. These simplifications are necessary for computational efficiency and simplicity, but these errors are within acceptable limits. The accuracy of these algorithms can be validated by comparing the generated trajectories with analytical solutions of simple, solvable problems.

Time step (δt) selection is defined by the user; but it is determined by the system being modeled. If it's too large, it can cause significant energy changes, leading to instability. If it's too small, the computation time increases to impractical levels. For flexible molecules, such as proteins, the optimal time step is 1 or 2 fs, one tenth of the period of the highest frequency motion in the system, the C-H bond vibration.

1.2.3. MDeNM simulations

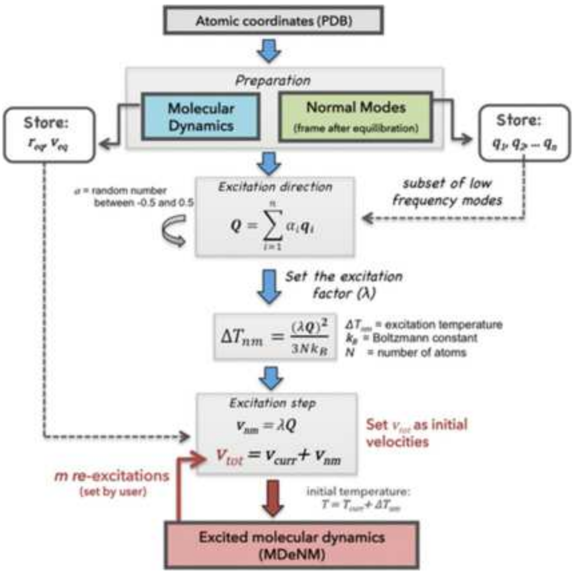

Figure 8. *The principle of MDeNM, explained in a flow chart. Figure have been published by Costa et al. (65)*

MDeNM is an enhanced MD simulation method (65) with the advantage of not altering the potential energy surface in a way that would displace a molecule from one energy minimum to another, which other accelerated methods do (66-70), revealing states and transitions that might be inaccessible. Instead, it introduces energy into the system along given directions which accelerates the transitions between energy barriers. Its working principle is the following: classical MD simulation is conducted on the system, and the normal modes (NMs) of the energy-minimized structure are calculated, with kinetic energy imparted to the system along these modes. The random linear combination of biologically relevant (typically low-frequency) modes are generated, and energy (heat) is added to the system along these vectors by introducing additional velocity in a given direction. A short classical MD simulation then performed to dissipate the injected energy and restore the system to equilibrium. The system is then subjected to a second heating process. In this way, MDeNM explores conformational wells that are not normally found in classical MD simulations due to the short timescale.

1.2.4. Molecular docking

Often referred to simply as docking, this method is widely used and increasingly popular due to its due to its simplicity, computational cost-effectiveness, and capacity to generate and profile a vast amount of results with reliability. However, this robustness is accompanied by certain limitations, as the energy profiling and conformational search processes rely on a number of approximations and assumptions. It is essential to clearly define these in order to draw meaningful conclusions.

We can distinguish between blind and site-specific molecular docking. In the case of blind docking, the entire surface area and potential binding sites of the protein are considered without prior knowledge of the ligand's binding location. Blind docking requires multiple iterations and energy calculations to identify a favorable protein-ligand complex pose. However, the exact number of iterations and energy evaluations needed for a successful blind docking run is unpredictable due to the stochastic nature of the search process. Therefore, reducing search complexity is required (71).

Targeted or site-specific docking predefines the area of the docking site (72, 73). A docking grid, which is a three-dimensional array is created in order to define potential positions and orientations for the ligand that will be searched. However, it is a boundary, that the ligands cannot exceed. Most docking approaches accounts for the flexibility of the ligand, and some methods can also consider protein flexibility (74). However, the docking process itself can be defined as an optimization problem where the objective is to identify the optimal orientation of a ligand within an active site, while also considering internal changes of ligands (and proteins), for example torsion angle rotations. A variety of search algorithms are available for the purpose of generating different conformations of ligands. These include systematic, stochastic, and deterministic algorithms.

Search algorithms

The systematic search algorithms explore all degrees of freedom of the ligand in question. However, the number of generated conformations needs to be limited. One potential solution to this issue is to halt the exploration process when a specific criterion is met. This process can be accelerated by cutting the ligand at the flexible bonds, treating each rigid part separately. Two examples of software utilising systematic algorithms are DOCK (75) and FRED (76).

The stochastic search algorithm generates many random conformations based on a probability function, varying the rotations and translations to explore the conformational space. Monte Carlo uses a predefined scoring function to determine if a new randomly generated conformation is sufficiently distinct to be accepted or rejected until the desired number of configurations is reached. A genetic algorithm models the genetic evolution through the use of state variables, such as parameters describing translation, rotation, and the conformation which are represented as genes within the algorithm. New generations are created from the combination of these variables. One example of swarm optimization is the Ant Colony Optimization (ACO) algorithm, which treats each degree of freedom as a 'pheromone'. Virtual ants contribute to pheromone deposition by choosing the fastest path to a 'food', i.e. conformations with the highest pheromone value. The PLANTS package (77) uses this algorithm. Software such as GOLD (78), AutoDock (79), Vina (80), and ICM (81) are also using stochastic algorithms.

Generally, ligand binding to a protein involves adaptation of both the ligand and the protein to achieve the most favorable interaction. This process can be modeled by a process known as induced-fit. Some docking software is capable of considering protein flexibility through induced-fit approaches. This process involves MD simulations to further optimize protein-ligand complexes, until the most favorable one (at local minimum) is found. This process can involve using deterministic search algorithms to ensure thorough exploration of the conformational space. Software packages for NAMD (82) and GROMACS (83) can be used for such simulations. However, due the computational costs, this method is rarely used.

Docking scoring

Regardless of the algorithm employed during the docking process a large amount of conformation has to be evaluated. These conformations, referred to as "poses" or "modes," are generated for the ligand within the protein. The scoring process enables the ranking of these poses based on an approximated binding free-energy (BE), derived from the Gibbs free energy:

$$\Delta G=\Delta H-T\Delta S$$

Where ΔH and ΔS is the change in enthalpy and entropy respectively upon complex forming, and T is the temperature in Kelvin. Assuming thermodynamic equilibrium conditions for the protein-ligand complex formation scoring functions will give an approximation on the free energy. However, as different algorithms consider different functions in some cases multiple different scoring function are combined in a weighed manner to improve the reliability of the final prediction, and ranking. In force-field-based scoring, the binding free energy is predicted by summing the non-bonding force field potentials, Lennard-Jones and Coulomb potentials, torsional forces of ligands, and in some cases, solvent molecules and solvation functions (84). This score can also be referred to as interaction energy (IE). Most notably, UCSF DOCK (85) and AutoDock (With the AMBER FF (83)) use this type of scoring.

$$\Delta G=\Delta G_{\mathrm{vdW}}+\Delta G_{\mathrm{elec}}+\Delta G_{\mathrm{desolv}}+\Delta G_{\mathrm{tor}}$$

The main limitation of these approximations is the lack of precision of the scoring functions themselves, for example, the explicit consideration of water molecules (73, 86). It is possible to calculate interaction free energy with greater precision using quantum mechanics (QM).

Empirical scoring functions incorporate several simplified energy terms that approximate various interactions, including ionic, hydrogen bond, and lipophilic contacts, in order to estimate the binding free energy of receptor-ligand complexes. The weights assigned to these terms are derived from experimental observations and data from known receptor-ligand complexes. AutoDock Vina (Vina) (80) employs this type of scoring function. In this case, BE is approximated based on dissociation constant (K_d), using the following formula:

$$\Delta G^0 = -RTlnK_d \quad \text{where } K_d = \frac{[L][R]}{[LR]}$$

where $K_d = \frac{[L][R]}{[LR]}$, with [R] representing the concentration of the free receptor (which could be an enzyme, protein, or any binding site), [L] the concentration of the free ligand, and [LR] the concentration of the ligand-receptor complex (87). This relationship is based on the principle that the dissociation constant can be empirically quantified. Empirical scoring functions, often rely on simple terms to approximate the binding free energy.

Knowledge-based scoring functions are derived from statistical analysis of known protein-ligand complexes. These functions operate on the assumption that experimental results represent the most accurate placement of ligands, and scoring prioritize over geometries that align with this experimental knowledge. In other words, the pose with the highest ranking should be the one that is most similar to the experimental structure. In a computational context, knowledge-based scoring functions use pairwise atomic potentials, which means the interaction energy is calculated based on the probability distributions of interatomic distances between atoms of the protein and the ligand. The DrugScore suite (87) is an example of a collection of tools that employs this type of scoring function.

2. OBJECTIVES

The broad objective of this research is to elucidate the molecular mechanisms governing substrate specificity, binding mechanisms and selectivity of human sulfotransferases (SULTs).

Objective1: Exploration of Structural Dynamics of SULT1A1 and SULT1A3 monomers

We aim to clarify the structural and dynamic differences of SULT1A1 and SULT1A3 monomers, with a particular focus on their conformational flexibility and substrate specificity using MD and MDeNM simulations.

Objective2: Structural bases of substrate specificity and selectivity

We aim is to understand the impact of the structural variations of these isoenzymes, influencing the binding of different substrates and inhibitors, explaining their selectivity and specificity. A digital repository of substrates, inhibitors, selective substrates and selective inhibitors are gathered for docking into the conformation ensembles generated based on dynamics simulations.

Objective3: Impact of Dimerization on SULT1A1 Structure and Function

We aim to elucidate the effect of dimerization on cofactor and substrate biding of SULT1A1. SULTs have been studied as monomers *in silico*, but their *in vivo* dimerization suggests functional importance. Therefore, we compare the behavior of SULT1A1 in monomeric and dimeric forms, while examining it at different occupancy states of the PAPS and ligand in the dimer chains.

3. METHODS

3.1. Structure preparation

Initial X-ray crystal structures for both SULT1A1 and SULT1A3 monomers were obtained from the Protein Data Bank (25) with PDBID 4GRA (88) and PDBID 2A3R (89), respectively. The structure of SULT1A1 contains the co-factor PAP, while SULT1A3 also contains PAP in addition to dopamine, which is a selective substrate of the latter isoenzyme. For both the SULT1A1 and SULT1A3, an identical preparation protocol was followed: PAP was replaced with PAPS by overlapping the shared atoms, and PAPS was subsequently parameterized using CGenFF 2.2.0 (90). The protonation states of the proteins were determined using PROPKA (91). Dopamine was also parametrized with CGenFF.

All residues showed normal pK values, and histidine residues in both isoenzymes were singly protonated at either the delta or epsilon nitrogen, with assignments made based on the local environment.

For the cytosolic SULT1 enzymes, a rectangular solvent box of water molecules was constructed around the proteins. The CHARMM-GUI (92, 93) web tool was used, with a similar 82 Å × 82 Å × 82 Å cube, the boundaries of which was set at least 14 Å beyond the protein to prevent self-interaction across the periodic boundaries. Solvation was executed with TIP3 water molecules while the NaCl concentration was adjusted to 0.15 mol/L.

For energy minimization we performed steepest descent (94), and Newton-Raphson (95) which follow gradients downhill on the potential energy surface until a local minimum is reached. Usually several cycles of equilibration are performed, using at least two different methods for better results.

Energy minimization for all SULT1 systems was performed using a series of harmonic restraints that were progressively reduced, starting with the steepest descent (SD) method. The harmonic force constant was decreased every 100 steps with values of 50, 10, 1, and 0.1 kcal/mol/Å^2. Following the release of restraints, three cycles of 250 steps each of SD and adopted basis Newton–Raphson (ABNR) minimizations were carried out, followed by a final cycle of 500 steps. Minimization was executed using CHARMM software (96) with the CHARMM C36m force field (97).

Dimer preparation

The previously used crystal structure, 4GRA, exhibited dimerization in a different region, an artifact previously identified as a result of crystal packing, in the isoenzyme SULT1B1 (26). For this reason, the 2D06 (98) structure was taken which contains the canonical dimerization region, in the sequence H213R mutation was applied to restore the wide-type sequence. The RMSD between one chain of 4GRA and 2D06 is 0.38 Å showing a high similarity. 2D06 also includes the inactive cofactor PAP and estradiol in both chains, these molecules were treated similarly as in the monomer preparation.

The online web tool CHARMM-GUI (92, 93) was used to generate a solvent box of TIP3 water molecules around the protein, the boundaries of which was at least 12 Å beyond the most distal part of the protein to prevent self-interaction across the periodic boundaries (129 Å × 129 Å × 129 Å). NaCl concentration was set to 0.15 mol/L. The previously described protocol was used for energy minimization, heating and equilibration. Furthermore, a relatively large substrate of SULT1A1, fulvestrant have been chosen for additional MD simulations, based on the results of SULT1A1+PAPS simulations. The parameters of fulvestrant were determined by CGenFF.

3.2. Parameters for MD and MDeNM simulations

MD

All systems were heated and equilibrated in an NVT ensemble at 300 K for 100 ps, followed by a 5 ns NPT simulation at 1 atm pressure with random velocities assigned according to the Maxwell-Boltzmann distribution at 300 K using NAMD (99) with CHARMM C36m force field. Temperature was controlled via Langevin dynamics with a damping coefficient of 1 ps^{-1}, while pressure was regulated using the Nose–Hoover method with a piston oscillation period of 50 fs and a decay time of 25 fs. The integration time step was 2 fs. For energy calculations, a dielectric constant of 1 was used. Electrostatic interactions were calculated with the particle mesh Ewald (PME) method, utilizing a grid spacing of 1 Å or less and an order of 6. The real-space summation cutoff was set at 12.0 Å, and the Gaussian distribution width was set to 0.34 $Å^{-1}$. Van der Waals interactions were smoothly truncated using a ‘switch’ function between 10.0 and 12.0 Å.

For the production runs three parallel simulations were carried out with different starting velocities. Three 1 μs-long MD simulations were conducted for the SULT1A1+PAPS monomer, each starting from the equilibrated structure and. A time step of 2 fs was used, with the coordinates saved at 10 ps intervals. The length of the simulation for different systems are presented in Table 2.

Enzyme	Type	Simulation time
SULT1A1	MD	3×1 μs
SULT1A1+PAPS	MD MDeNM	3×0.2/0.5/1 μs 240 x 0.2 ns
SULT1A1+PAPS+fulvestrant	MD	3×1 μs
SULT1A3+PAPS	MD	3×0.5 μs
SULT1A3+PAPS+dopamine	MD	3×0.5 μs
2SULT1A1	MD	3×1 μs
2SULT1A1+1PAPS	MD	3×1 μs
2SULT1A1+2PAPS	MD	3×1 μs
2SULT1A1+2PAPS+2fulvestrant	MD	3×1 μs
2SULT1A1+2PAPS+1fulvestrant	MD	3×1 μs

Table 2. *Summary of performed simulations*

MDeNM

Since SULT1A1 exhibited large rigidity during classical MD simulations MDeNM simulation was carried out to comprehensively map its conformational surface. The normal modes of the energy-minimized structure were calculated using the VIBRAN module of CHARMM (100). Subsequently, three low-frequency normal modes that contributed the most to the atomic displacements (RMSF) were selected, and random linear combinations of these modes were generated. This resulted in RMSD differences from 0.3 Å to 1 Å of the displaced structural models along the combined NM directions. In total, 240 distinct directions were generated. For each of the 240 directions, MD simulations were conducted to promote the motion described by the combined NM vector kinetically. This was accomplished by adding an additional velocity in the direction of the combined NM vector to the current MD velocities, resulting in an overall 2 K increase in the system temperature. A series of 50 consecutive excitations was achieved at 4 ps intervals throughout the course of the MD simulation. Consequently, the total duration of the MDeNM simulation was

48 ns, comprising 240 x 50 x 4 ps. The remaining MD parameters were identical to those described in the previous paragraph.

3.3. Conformational Clustering

A clustering approach was implemented to create a conformational ensemble, which represents the entirety of the mapped conformational space while significantly reducing the number of conformations for docking. All three parallel MD and a representative set of the MDeNM simulations (101) of SULT1A1 were clustered with the inclusion of the cofactor PAPS. For SULT1A3, MD simulations of SULT1A3+PAPS+dopamine system was clustered, however dopamine was removed from the trajectories prior to clustering. The quality threshold (QT) (100) algorithm was applied, as implemented in Visual Molecular Dynamics (VMD) software (102). The RMSD difference calculated for the binding pocket heavy atoms (defined in the Supplementary Materials) was used to determine the maximum cluster distance, which was set to 1.3 Å. This resulted in 94 MD and 86 MDeNM centroid conformations for SULT1A1, while for SULT1A3 the centroids of the first 104 clusters were selected for docking. These clusters covered 80% of all the frames from the three MD trajectories.

3.4. Ligand collections and preparation

The ligands of SULT1A1 had been previously collected by Martiny et al. (103). A classification system based on the experimental activities of the ligands was developed. It consists of the following groups: substrates, inhibitors, selective substrates, and selective inhibitors. Ligand structures were converted to PDBQT format using Autodock Tools (79), and Gasteiger charges were assigned. A new collection of ligands was conducted for SULT1A3 during thesis work. The resulting chemical library consists of 143 substrates and inhibitors with documented biological activity (K_m K_i or IC), sourced from the following databases: PubChem (104), DrugBank (105), and Reaxys (106). Classification and ligand preparation followed the same protocol used in the case of the ligands of SULT1A1.

3.5. Ensemble docking

Docking experiments were performed using AutoDock Vina (80). AutoDock Vina employs gradient-based conformational docking and an empirical scoring function that predicts the interaction energy between a protein and a ligand (in kcal/mol). For both SULT1A1 and SULT1A3, we performed docking simulations on the selected centroid protein conformations. They were pre-processed using AutoDockTools (79), removing solvents, merging non-polar hydrogens, and assigning Gasteiger charges. The maximum number of binding modes was set to 20, and the exhaustiveness of the global search was set to 10. During the docking process of SULT1A1, the binding site residues K106 and F247 which exhibited rapid side-chain conformational changes during the MD and MDeNM simulations were handled flexibly; the remainder of the protein and the co-factor were maintained in a rigid state. In the case of SULT1A3, only the ligands were treated flexibly, while the protein and the co-factor were kept rigid. In both cases, a grid box was centered in the binding pocket with dimensions of 24 Å × 24 Å × 24 Å, and the spacing was set to 1 Å. Filtering was performed to ensure that the distances between the substrate acceptor hydroxyl or primary amino functional group and the sulfate group of the co-factor PAPS and the catalytic residue H108 fell within 5 Å of all the substrates docked into SULT1A1 and SULT1A3.

3.6. Substrate clustering

The common substrates of the two isoenzymes were identified, and a diversity clustering was performed using FCFP_4 with a Tanimoto similarity criterion of 0.6, as implemented in Biovia Pipeline Pilot (107). Ligands comprising a single aromatic ring and those with polycyclic or multiple-ringed structures were classified as distinct categories, designated as "small ligands" and "large ligands," respectively. A total of 14 clusters were obtained for large common substrates, and 6 for small common substrates.

3.7. Quantities used for analysis

Root Mean Square Deviation (RMSD)

It measures the average deviation of the atomic positions of a molecule from a reference structure over time. The square root of the distances between all matching atom pairs for the given conformation and the reference structures is given as:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{r}_i^a - \mathbf{r}_i^b)^2}$$

This equation explains the global difference between two structures, *a* and *b*. $\mathbf{r}_i^a$ and $\mathbf{r}_i^b$ denote the position vectors of the atom *i* in and *N* is the number of atoms in the system. This metric allows us to characterize the global difference between two structures. If $\mathbf{r}_i^b$ is the reference structure, and $\mathbf{r}_i^a$ is stepped along the trajectory, we can plot the deviation of a system as a function of time.

Root Mean Square Fluctuation(RMSF)

The average fluctuation of each atom around its mean position provides a statistical measure for identifying rigid and flexible regions within a protein, independent of time. The formula for calculating the RMSF for a residue in the protein is:

$$RMSF_i = \sqrt{\sum_{t_j=1}^{T} \frac{\left(x_i(t_j) - \overline{x_i}\right)^2}{T}}$$

where x_i denotes the atoms of the protein, but for practical reasons and to reduce noise in the fluctuation measurements, only the C_α atom of each amino acid residue is typically considered. *T* represents the total number of frames in the trajectory, and t_j= (1, 2, 3 .. *T*).

Principal Component Analysis (PCA)

This statistical technique used to reduce the complexity of high-dimensional data by transforming them into lower-dimensional space. In the context of MD trajectories, this means that a reduced number of coordinates that captures the essential features of the conformational ensemble can be presented.

In essence, principal components (PCs) are obtained by constructing a covariance matrix from the linear combinations of the Cartesian coordinates of the atoms. The first principal component (PC1), represented by the largest eigenvalue corresponds to the direction of the largest variance in the data. The second principal component (PC2) captures the second largest variance and is orthogonal to the first, and so on. The conformational ensemble can then be projected onto the reduced space spanned by the PC(s) (commonly represented in 1D, 2D, and 3D coordinate maps), helping to visualize and interpret the conformational changes that dominate the dynamics of the system (108, 109). It was carried out on the simulations data using CHARMM (96).

All three runs were considered together in the PCA analysis for each system. After calculating the PCs, the structures of the simulation were projected onto the reduced space spanned by the first two PCs with the largest variances. Furthermore, the free energy landscapes (FEL) of the PCA projections were also calculated based on the 2D population distributions. The most populated state was used as a reference for calculating free energy differences. The free energy difference (ΔG_α) of a given state α was determined by considering the probability of the occurrence of the two states $P(q_\alpha)$ and $P_{max}(q)$, given by the following equation:

$$\Delta G_\alpha = -k_B T \ln\left[\frac{P(q_\alpha)}{P_{max}(q)}\right]$$

where k_B is the Boltzmann constant, T is the temperature of the simulation, and $P(q_\alpha)$ is an estimate of the probability density function obtained from the bi-dimensional histogram of the conformations. $P_{max}(q)$ is the probability of the most populated state.

Force Distribution Analysis (FDA)

This analytical method is used to calculate forces that arise between two atoms or residues (110). With this method, we could examine the perturbation caused by the ligands in the dimer structure, where systems containing a specific ligand are called perturbed (*pert*), while in the reference state (*ref*), the ligand is absent. Subtracting the forces of both states for each pair of interacting residues, parts under mechanical strain become visible.

$$\Delta F_{uv} = F_{uv}^{pert} - F_{uv}^{ref}$$

To visualize these results, we can use the absolute sum of force differences, sensed by a residue, which is called punctual stress (111):

$$\Delta F_u = \sum_v |\Delta F_{uv}|$$

The FDA code is implemented as an add-on for GROMACS 2020.4, enabling the calculation of forces, thereby enabling the calculation of forces, F_{uv}, between each residue. The forces are obtained by summing the atom-pair forces F_{ij} for all pairs of residues (u and v) where atoms $i \epsilon u$, $j \epsilon v$ (112). Further information can be found at the website github.com/HITS-MBM/gromacs-fda (113). Forces were recorded for every tenth frame during MD simulations. To achieve converged averages, the forces from each simulation (initiated with random velocities) were averaged over, which resulted in three equilibrium trajectories for each system. The differences between averaged pairwise forces ΔF_{ij} of the apo and complexed forms of the SULT1A1 dimer were calculated in order to describe the perturbation caused by substrate binding. The term "atomic punctual stress" has previously been defined as an accessible measure for assessing perturbations resulting from ligand binding (111). The same method is employed here; however, since the pairwise forces are calculated at the residue level, the resulting punctual stress is also residue-based.

Interaction Analysis

Examining specific interactions such as the forming of salt bridges, π-stacking, hydrogen bonds, hydrophobic contacts and other interactions helps to understand the key features that drive binding. Tools like the PLIP online profiler (114) which offers a variety of tools for post-docking analysis of the docking results. One objective of such research can be to identify important sidechains that influence ligand recognition or enzyme specificity, especially in cases where multiple isoenzymes recognize different ligands, as demonstrated in the comparison of SULT enzymes (27). This ligand-protein interaction profile provides valuable information for identifying potential active compounds (hit). Additionally, docking suites can be utilized to calculate the binding energies associated with the most stable conformations of drug-receptor interactions.

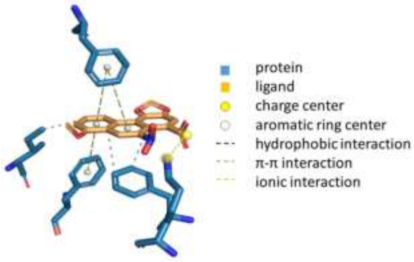

Figure 9. *Interaction calculation by the online web tool PLIP (114), on SULT1A1 and one of its selective substrate aristolochic acid, visualized by PyMol graphical tool (115)*

4. RESULTS

4.1. The role of conformational dynamics in substrate specificity of SULT1A1and SULT1A3 monomers

Structural analysis

The structural dynamics of the isoenzymes SULT1A1 and SULT1A3 were investigated through a series of all-atom MD and MDeNM simulations. These simulations aimed to shed light on the differences in flexibility and conformational behavior between the two isoenzymes, particularly in relation to their binding pockets and overall protein structures.

To quantify their conformational flexibility, in Figure 10 we present the relative frequency of the RMSD values of SULT1A1 and SULT1A3 MD simulations. The values were calculated for the heavy atoms of the backbone after removing the global translational and rotational motions by superimposing to the first structure of the production run. The RMSD data reveals substantial differences between the two isoenzymes. SULT1A1 exhibits a relatively rigid structure (light blue), with the RMSD values for the entire protein and the binding pocket confined within narrow ranges, indicative of limited conformational fluctuations around a single energy minimum. Specifically, the RMSD for the whole protein is centered around 1.22 Å, for the binding pocket. Since SULT1A1 exhibits such a rigid structure, the parallel MD simulations were run in this study for 200 ns. Instead, we performed MDeNM simulations to wider explore its conformational surface. Indeed, the MDeNM conformations distribution (dark blue) was more dispersed, even reaching conformations with a binding pocket deviating up to 2.25 Å from the crystal structure.

In contrast, SULT1A3 demonstrated a broader range of conformational flexibility, particularly in the case of the dopamine-less structure. Longer, 500ns MD simulations were run for SULT1A3, to thoroughly explore the conformational changes exhibited by this isoenzyme. The RMSD values varied significantly, ranging from 0.6 Å to 3.5 Å for the binding pocket. These results clearly indicate the higher conformational flexibility of SULT1A3 compared to SULT1A1. The presence of dopamine decreased these values, varying between 0.4 Å and 2.5 Å (Figure 10F). This reduction in variability suggests that dopamine plays a stabilizing role in the SULT1A3 enzyme, constraining its conformational space to fewer energy minima.

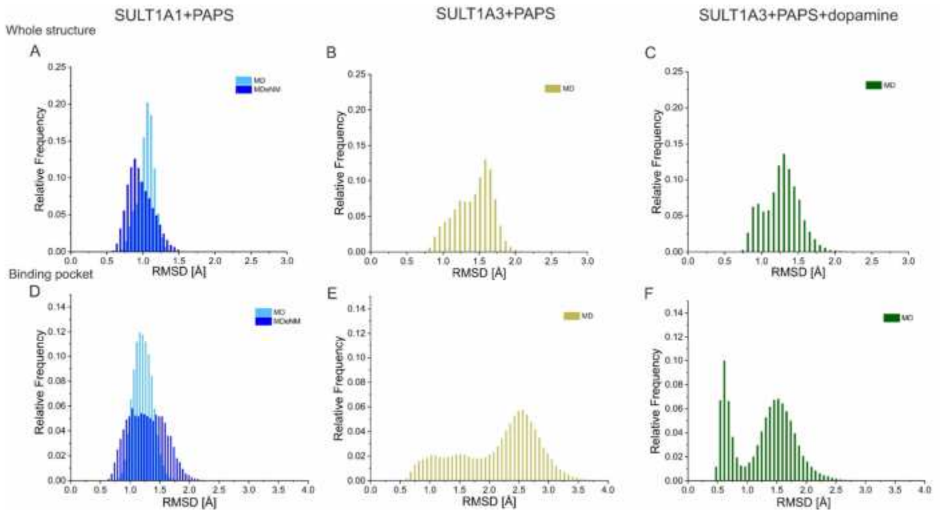

__Figure 10.__ The root mean square deviations (RMSD) of MD simulations for SULT1A1 and SULT1A3 with respect to the corresponding crystal structures. RMSD are calculated on the backbone of the whole protein (A–C) and on the binding pocket heavy atoms (D-F). For SULT1A1+PAPS (A,D), light blue represents the MD, while dark blue represents the MDeNM-generated data. (B,E) represent MD simulations of SULT1A3+PAPS. (C,F) represent MD simulations of SULT1A3+PAPS+dopamine (27)

To identify and characterize the structural elements responsible for the structural elements responsible for the RMSD variations that contribute to the observed differences in flexibility, the RMSF values for the Cα atoms of both enzymes are presented in Figure 11. The results show that the functional loops L2 and L3 of SULT1A3 (gold) exhibits significantly higher fluctuations than those in SULT1A1 (blue), indicating that these regions are the primary contributors to SULT1A3's overall flexibility. As a comparison, the residue A148 (in loop L2) of SULT1A3 shows an RMSF value of 4.0 Å, while the corresponding residue V148 of SULT1A1 shows a very low RMSF value of 0.9 Å; the residue H250 of L3 loop (SULT1A3) shows an RMSF value of 3.3 Å, whereas I252 (SULT1A1) exhibits RMSF of 1.1 Å.

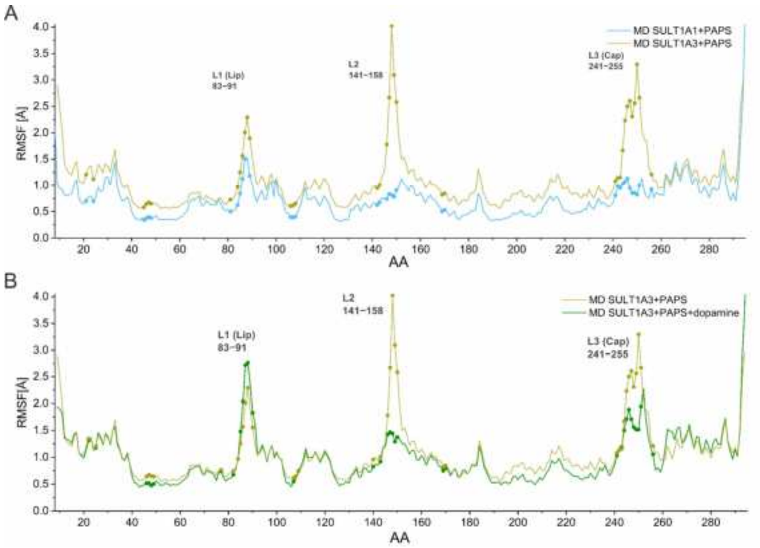

Figure 11. *RMSF of the Cα atoms of the amino acid residues (AA) for the MD simulations performed on SULT1A1 and SULT1A3. (A). RMSF of SULT1A1+PAPS (blue) and SULT1A3+PAPS (gold). (B). RMSF of SULT1A3 without (gold) and with bound dopamine (green). Residues of the binding pocket are indicated with dots (27)*

Figure 11B shows the presence of dopamine (green) had a marked effect on the flexibility of loops L2 and L3 in SULT1A3, compared to the structure without dopamine (gold). Specifically, the RMSF value for residue A148 in loop L2 decreased from 4.0 Å in the dopamine-free form to 1.4 Å when dopamine was bound. Similarly, the RMSF for the L3 loop also decreased in the presence of dopamine. This reduction in loop flexibility suggests that dopamine binding stabilizes the L2 and L3 loops by restricting the conformational space available for substrate binding.

To gain a deeper understanding of the collective motions and conformational transitions in SULT1A1 and SULT1A3, we performed PCA on all three MD simulation trajectories (see Figure 12). PCA allowed us to identify the major conformational changes that occurred during the simulations by projecting the protein conformations onto the subspace defined by the first two principal components (ranked by the corresponding variances). Directions of PC1 and PC2 are represented so that more positive values correspond to more open conformations.

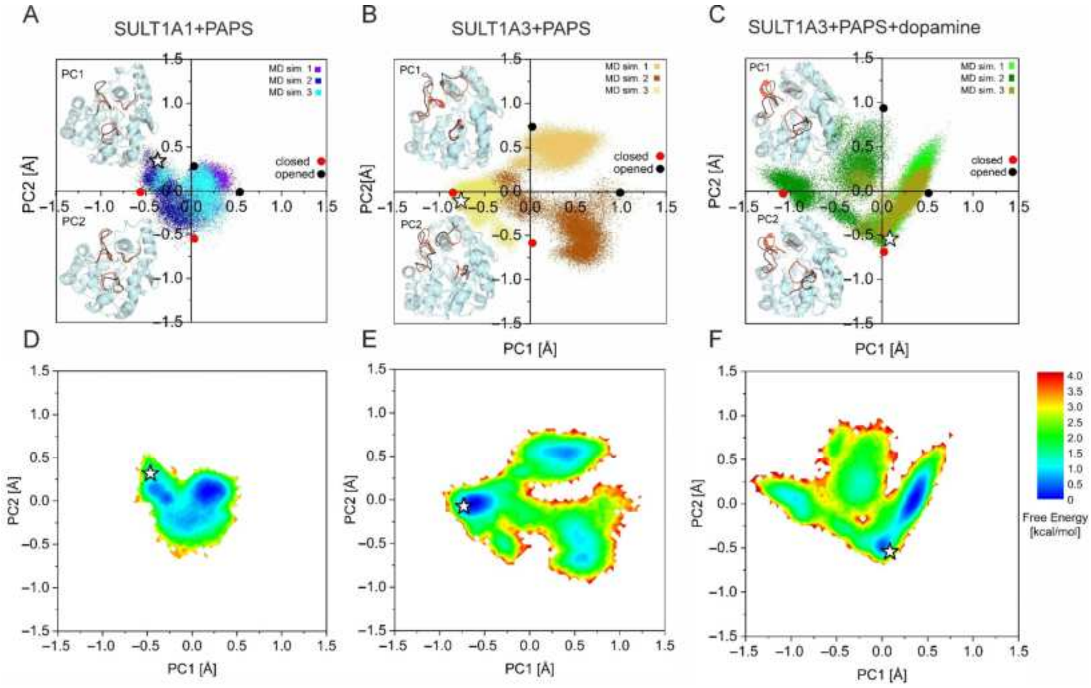

Figure 12. *PCA analysis. Population distributions of conformers generated by MD of SULT1A1+PAPS (A), SULT1A3+PAPS (B) and SULT1A3+PAPS+dopamine (C), projected onto the subspace covered by PC1 and PC2. In the insets, structures corresponding to the extremities of PC1 and PC2 are represented, and their location on the diagram is indicated by circles. The functionally important loops L1, L2, and L3 are color-coded: open-loop conformations are depicted in black, while closed ones are depicted in red. Free energy landscapes (FELs) of SULT1A1+PAPS (D), SULT1A3+PAPS (E), and SULT1A3+PAPS+dopamine (F) are in the space defined by PC1 and PC2. The initial structures are denoted by white stars (27)*

SULT1A1 exhibited very constrained fluctuations along both PC1 and PC2, with values ranging between −0.5 and 0.5 Å. This limited sampling along the principal components further supports the notion that SULT1A1 operates within a restricted conformational space, consistent with its observed rigidity in the RMSD and RMSF analyses.

In contrast, SULT1A3, both in its ligand-free and dopamine-bound forms, displayed much broader conformational sampling. The variance along PC1 and PC2 was approximately 4.5 times higher for SULT1A3 compared to SULT1A1. This broader sampling suggests that SULT1A3 can adopt a wider range of conformations, potentially enabling it to interact with a more diverse array of substrates. However, the presence of dopamine reduced the variance along both PC1 and PC2, further emphasizing dopamine's role in stabilizing specific conformational states within SULT1A3. For both ligand-free and dopamine-bound

SULT1A3, the projections of the conformations onto PCs vary in the same magnitude (from −1 to 1 Å along PC1, and from −1 to 0.75 Å along PC2 for SULT1A3+PAPS; from −1.25 to 0.5 along PC1, and from −0.75 to 1 along PC 2 for SULT1A3+PAPS+dopamine). However, the variance along both PC1 and PC2 is considerably higher for the dopamine-free system, further emphasizing the stabilizing effect of the presence of dopamine in the binding pocket. The projections for the dopamine-bound SULT1A3 conformations were more focused and directional, suggesting a more constrained dynamic behavior.

By mapping the extremities of PC1 and PC2 onto the 3D structures of SULT1A1 and SULT1A3, we identified that the conformational changes associated with these principal components were primarily localized to the functionally important loops L1, L2, and L3, and the directionality of the PCs correspond to the opening and closing of the loops. For SULT1A1+PAPS, the most significant changes were observed in loop L1, consistent with its flexibility. In contrast, SULT1A3+PAPS exhibited larger amplitude motions in all three loops, particularly in L2, which contributed to the enzyme's broader conformational sampling, also in agreement with the RMSF results. The binding of dopamine stabilized loop L2 in a closed conformation, while loops L1 and L3 remained flexible, highlighting the nuanced role of dopamine in modulating SULT1A3's dynamics.

To complement the PCA, we also calculated the distribution of conformations described by PCA in the free energy landscapes (FELs) representation (shown in Figure 12 D–F). The FEL of SULT1A1 revealed a shallow energy minimum, consistent with the enzyme's limited conformational flexibility. This shallow landscape suggests that SULT1A1 exists primarily in a single, stable conformational state, with low-energy barriers separating closely related conformations. In contrast, the FELs for both forms of SULT1A3 (with and without dopamine) displayed distinct energy minima separated by relatively higher energy barriers.

Clustering and Ensemble Docking

To generate representative conformational ensembles for SULT1A1 and SULT1A3, we performed conformational clustering based on the RMSD values of the heavy atoms within the binding pocket (see details in the Methods section). For SULT1A1+PAPS, clustering was executed separately on the MD and MDeNM-generated conformations, resulting in a total of 722 clusters, with 94 MD clusters and 87 MDeNM clusters retained for further analysis. In the case of SULT1A3, the clustering was carried out on the MD simulations of SULT1A3+PAPS+dopamine, using the same protocol, and finally retaining 104 MD clusters for ensemble docking. without bound ligand dopamine the large opening of the binding pocket made it inappropriate for ensemble docking (103).

The radius of gyration (RGYR) of the active site residues was used to measure the expanse of the binding pocket across the different clusters. For SULT1A1, the RGYR of the MD centroids ranged from 8.5 Å to 8.9 Å, with larger values indicating a more open binding pocket. The MDeNM-derived centroids exhibited an even wider range of RGYR values, from 8.6 Å to 9.1 Å, reflecting the broader conformational sampling achieved through this enhanced simulation technique. In comparison, SULT1A3 exhibited a much more open binding pocket, with RGYR values ranging from 9.1 Å to 9.9 Å.

For SULT1A1, 131 previously collected substrates and inhibitors were docked into the binding pockets of the selected centroids. Docking scores are summarized in Figure 13. These scores for SULT1A1 revealed a strong correlation between the RGYR of the binding pocket and the interaction energy of larger ligands. Clusters with RGYR values greater than 8.8 Å, particularly those derived from MDeNM simulations, exhibited significantly better docking scores, indicating that the more extended conformations of SULT1A1 are better suited for accommodating larger ligands. This finding underscores the utility of MDeNM in generating a more diverse set of conformations that can better capture the binding potential of large and complex substrates.

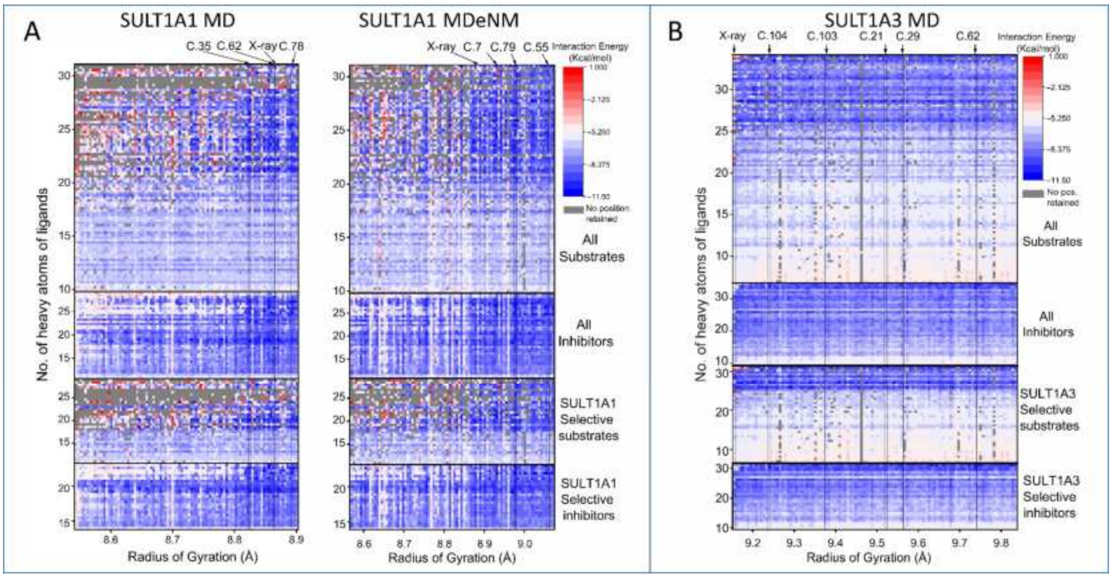

Figure 13. *Docking score map of SULT1 substrates and inhibitors into protein cluster centroids. (A) Docking scores of SULT1A1on the MD and MDeNM centroids, ordered by the radius of gyration (RGYR) of the binding pocket. (B) Docking scores for SULT1A3 on the MD centroids. Ligands in the ordinate axes are ordered by the number of their heavy atoms. Selective ligands indicate ligands not interacting with the other isoform. Blue to white colors indicate good predicted scores (for interaction energy), while red colors represent poor scores. Gray color indicates that docking has not produced a satisfactory pose of the substrate for the catalytic reaction. The results of the top 5 well-performing centroids and the X-ray conformation are marked with arrows (27)*

For SULT1A3, 143 substrates and inhibitors were docked. The docking scores are summarized in Figure 13. Interestingly, SULT1A3 shows considerably more open states of the binding pocket, ranging approximately from 9.1 to 9.9 Å. As expected, small ligands of SULT1A3 show worse IE than SULT1A1 due to the wider binding pockets generated. This confirms that the widely extended binding pocket is less efficient for accommodating small ligands. Being extremely flexible, it would be more difficult for SULT1A3 to establish strong interactions with small substrates.

To identify the key residues involved in ligand binding within the active sites of SULT1A1 and SULT1A3, we focused on protein conformations that accommodated a large number of ligands with good interaction energies. These key conformations were selected based on their ability to dock the majority of ligands with interaction energies lower than −5 kcal/mol. three MD centroids and three MDeNM centroids were selected, docking 129 out of the 131 ligands with IE lower than −5 kcal/mol. For SULT1A3, five centroids were selected, docking 142 out of the 143 ligands with IE lower than −5 kcal/mol.

Identifying key residues

For the analysis of the interactions with particular residues in the binding pocket, we need to look only at the substrates, because their "good" positions can be selected considering the knowledge of metabolites, whereas inhibitors can bind and work in different ways (116, 117). The top-ranked competent docking poses of all substrates in the best centroids of SULT1A1 and SULT1A3, and in the crystal structures (PDBID 4GRA and (PDBID 2A3R), the residues F24, F81, K106, H108, F142, H149, Y169, Y240, M248, and F255 are in contact with the docked substrates. For SULT1A3, the docking analysis highlighted the importance of residues D86 and E146, which possess carboxylic groups that stabilize the binding of ligands with positively charged groups or hydrogen-bond donors.

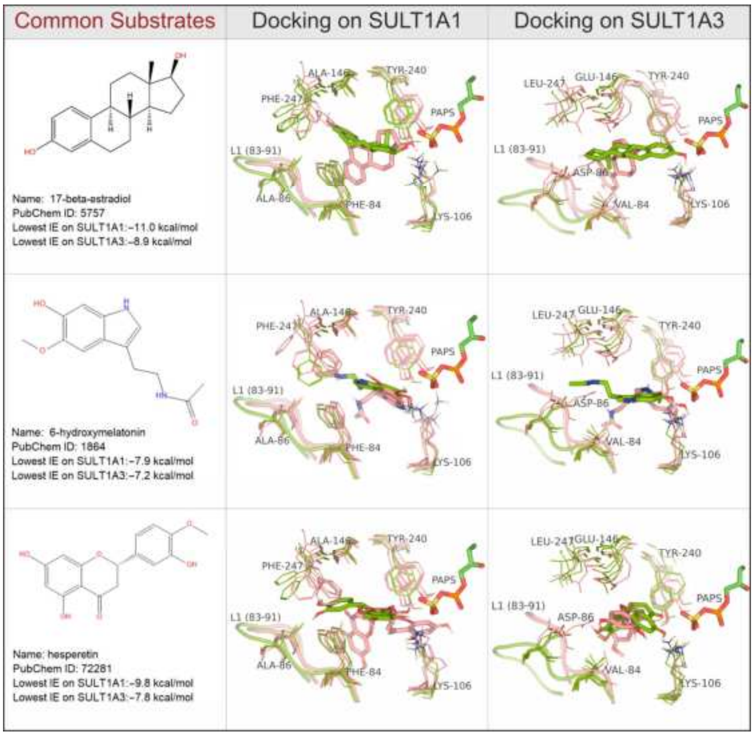

Figure 14. *Key interactions of common substrates of SULT1A1 and SULT1A3. The selected best protein centroids noted in Figure 14 and the X-ray structures are shown. Favorable and unfavorable docking poses and the corresponding protein conformations are shown in green and salmon, respectively. Favorable docking poses are competent with the catalytic reaction. Highly flexible amino acid sidechains are depicted as lines, while the substrates and PAPS are depicted as sticks. Loop L1 is shown as a cartoon (27)*

In SULT1A1, residues such as F76, F84, A146, and F247 were found to play pivotal roles in binding selective substrates. These residues contribute to the enzyme's ability to accommodate a diverse range of substrates, particularly through aromatic stacking interactions. Differently, SULT1A3 possesses two carboxylic groups of D86 and E146 that are well suited to stabilizing the binding of ligands containing positively charged groups, or containing hydrogen-bond donors, thus narrowing the range of accepted substrates.

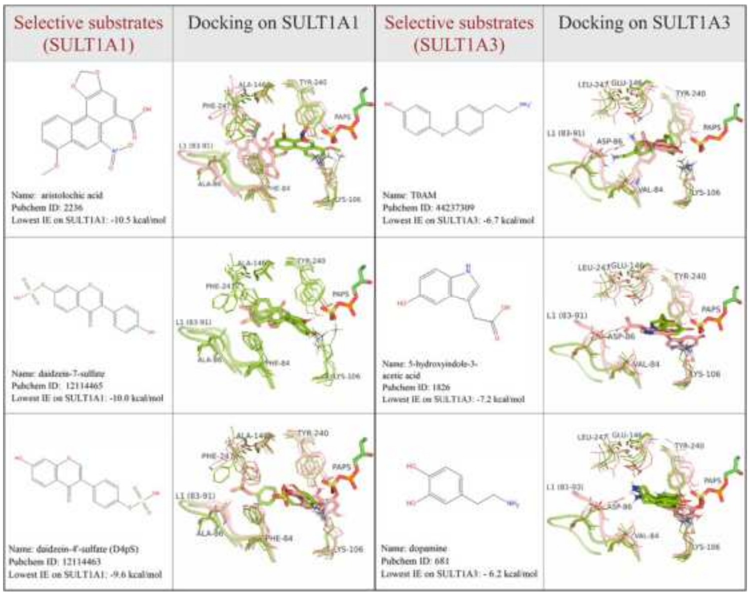

Figure 15. *Key interactions of selective substrates of SULT1A1 and SULT1A3. The selected best protein centroids noted in Figure 14 and the X-ray structures are shown. Favorable and unfavorable docking poses and the corresponding protein conformations are shown in green and salmon, respectively. Favorable docking poses are competent within the catalytic reaction. Highly flexible amino acid sidechains are depicted as lines, while the substrates and PAPS are depicted as sticks. Loop L1 is shown as a cartoon (27)*

4.2. Dimerization effects on SULT1A1 structural dynamics and ligand interactions

Dimerization effect

To follow the effect of dimerization on the dynamical behavior of SULT1A1 we compare the apo-, the PAPS containing, and the PAPS+fulvestrant containing forms of the enzyme in the monomer and dimer forms. In this study all parallel simulations were run for 1 μs. The average RMSD values and their standard deviation of the three parallel simulations for each system are represented in Figure 16. The RMSD values were calculated after removing the global translational and rotational motions by superimposing the backbone heavy atoms of each chain constituting the dimer separately to the first structure of the production run.

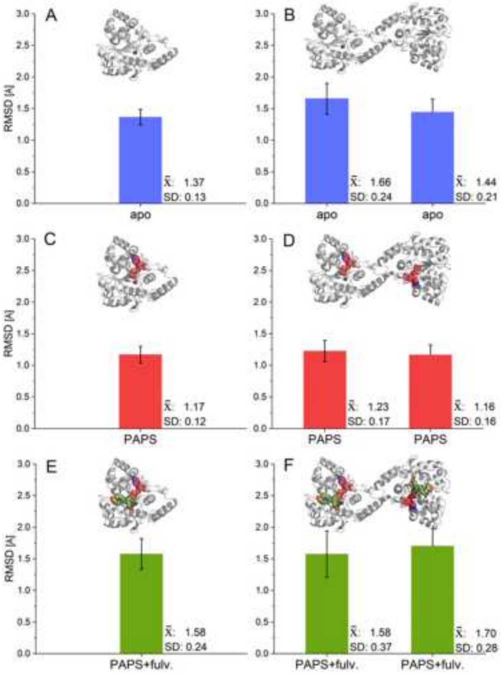

Figure 16. *RMSD of the MD simulations of SULT1A1 monomer and dimer apoenzme (A,B) with PAPS (C,D) and PAPS+fulvestrant (E,F) respectively, calculated on the whole protein backbone heavy atoms*

As panel A and B of the Figure shows, the overall RMSD shows an increase and becomes asymmetric upon dimerization of the apo enzyme changing from 1.37±0.13 Å for the monomer, to 1.66±0.24 Å and 1.44±0.21 Å for the two chains of the dimer.

Binding of the PAPS cofactor significantly rigidifies both the monomer and the dimer with a value of 1.17±0.12 Å for the monomer, not showing any significant change upon dimerization (1.23±0.17 Å and 1.16±0.16 Å - panels C and D). Furthermore, in the case of PAPS+fulvestrant containing enzyme the RMSDs indicate a flexible enzyme in both forms (panels E and F), with increased RMSD values (1.58±0.24 Å for the monomer) and higher standard deviation but not significant difference upon dimerization (1.58±0.37 Å and 1.70±0.28 Å).

To localize the regions responsible for varying RMSD values, we calculated residue based RMSF of the Cα atoms for each MD simulation. These results show that the increase of the fluctuation of the apo form upon dimerization (Figure 17 A and B) is mostly concentrated to the helix-loop region formed by helix α14 and the loop located to the N-terminus of L3 - we will denote this region by α14L3’ (formed by the residues 228-240); to the loop over PAPS (residues 256-263 - that will be named L4) and another helix-loop region of residues 66-78 named α3L5, which is sterically close to L3. A limited increase in fluctuation is also observed in L2 (Figure 17A and B).

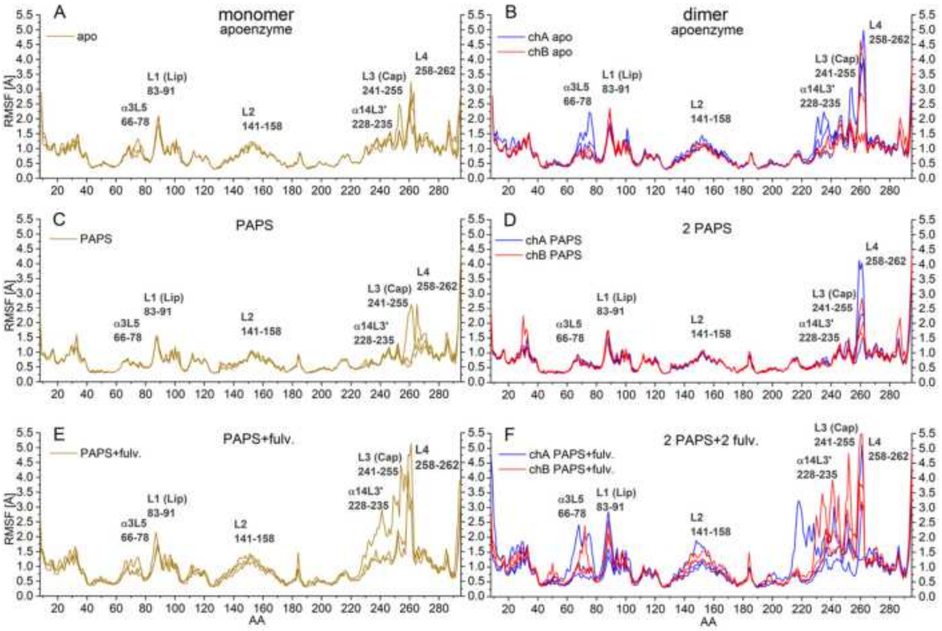

Figure 17. *RMSF of the Cα atoms of the amino acid residues (AA) for the MD simulations performed on SULT1A1 monomer and dimer with PAPS (C,D) and PAPS+fulvestrant (E,F) respectively. The monomers (panel A, C, E) are represented in brown color with the three MD simulation replicas plotted on the same panel. Chain A of the dimers are shown in blue and chain B in red (panel B, D, F)*

The comparison of RMSFs for the PAPS containing monomer and dimer show that there is a marginal increase of the L4 RMSF upon dimerization. Both in the monomer but strikingly in the PAPS+fulvestrant dimer (Figure 23 E and F) a substantial increase of the α14L3’ RMSF can be seen, as well as an increased RMSF can be noticed at L1, L2, L4 and α3L5. These results are in agreement with the RMSD, indicating that the fulvestrant binding strongly increases the flexibility of the enzyme. Furthermore, one simulation exhibits a high RMSF between residues 213-228 due to a helix to loop transformation of α13.

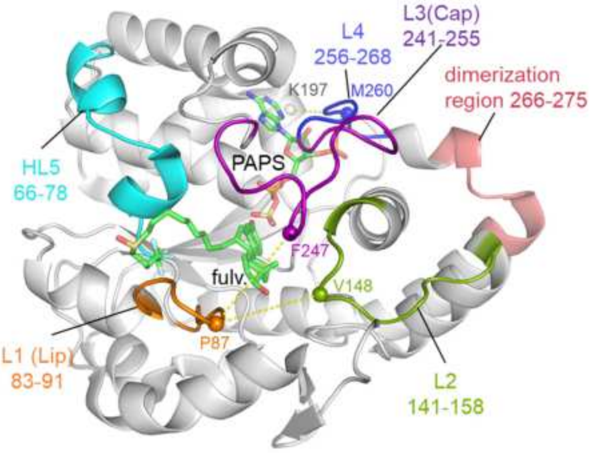

Figure 18. *SULT1A1 representation, with the large ligand fulvestrant and the active co-factor PAPS, shown in sticks. The three loops gating the active sites are indicated L1(Lip) in orange, L2 in green, L3(Cap) in magenta, L4 in blue, α3L5 in cyan, α14L3', and the dimerization region in salmon. Residues for distance calculation are labelled and represented by vdW spheres*

Functional loop characterization with distance calculation

In order to see if the increase of the loop fluctuations can be directly connected to the opening of the gate on either the ligand binding or the PAPS binding site, we calculated distances that characterize the opening of these clefts. To describe the openness of the gate of the ligand binding site we used the distances defined in our previous article, namely distance L1-L2 is defined by the distance between the Cα atoms of P87 (the tip of L1) and V148 (the tip of L2); and L1-L3 defined between the Cα of P87 (L1) and F247 (L3). Furthermore, we characterize the openness of the nucleotide binding site with the distance defined between Cα of K197 (α15) and M260 (L4). These residues are denoted as vdW spheres in Figure 18.

By comparing the openness of loop L4 for the monomer and dimer of the apo enzyme (Figure 19B-C), we notice that the B chain of the dimer exhibits similar behavior to the monomer with their population centered around 7 Å. However, chain A of the dimer behaves significantly differently, showing a widely spread population corresponding to more open conformations, ranging between 6-19Å. Such, in general we see that the open conformations of the nucleotide covering loop L4 are more populated upon dimerization.

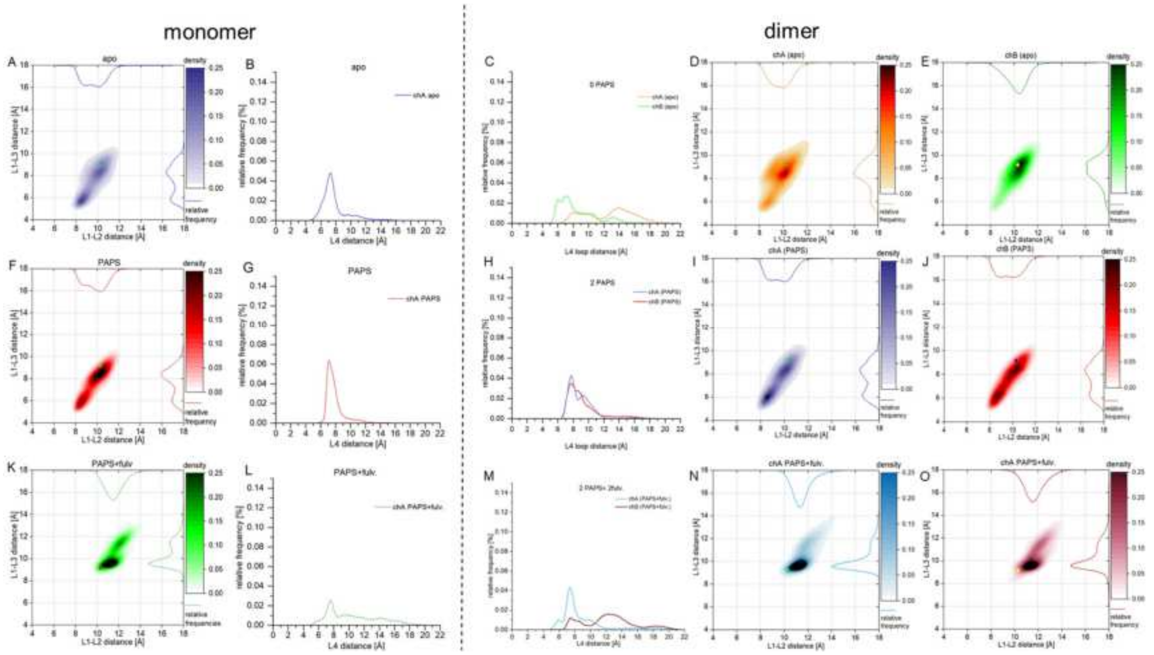

Figure 19. *Distance plots of the distribution of L1-L2 and L1-L3 distances for the MD generated structures. For SULT1A1 apo monomer (A) and dimer (D,E) with PAPS monomer (F) and dimer (I,J) PAPS+fulvestrant monomer (K) and dimer (N,O). The relative frequency of distance L4 during the MD simulations for the apo monomer (B) and dimer (C), with PAPS monomer (G) and dimer (H) and PAPS+fulvestrant monomer (L) and dimer (M). For the dimer systems, the same colors represent the same chain. The distances of the initial crystal structure (PDBID 4GRA) is denoted in each graph with a colored diamond symbol*

Comparing the L1-L2 distance of chain A of the dimer (Figure 19D) shows essentially similar behavior as the monomer (Figure 19A), while the L1-L3 distance opens up, with the population corresponding to a 6 Å opening being diminished and a 9 Å opening becoming more populated in the dimer chain A. The chain B exhibits opened conformations in both L1-L2 and L1-L3 distances (Figure 19E), Such, the gate opening increases in both chains upon dimerization showing an asymmetric behavior, a greater opening in chain B.

The populations of the gate- and of L4-opening are strikingly similar for the PAPS containing enzyme in monomer and dimer form with a slightly higher population of the L4 open conformations for the dimer (Figure 19G-H). The gate opening distances (Figure 19, F, I-J), L1-L2 and L1-L3, exhibit similar behavior in both monomer and dimer forms.

In the PAPS+fulvestrant containing monomer and dimer, the L4 loop opening exhibits clear asymmetric behavior in the dimer (Figure 19M). Chain A shows a closer conformation similar to the monomer (Figure 19L), while chain B, although still exploring similarly closed conformations, also reaches more widely open states around 13 and 19 Å. The gate opening distances (Figure 19N and O), L1-L2 and L1-L3, remain similar in both monomer and dimer forms, which is not surprising, due to the presence of the large ligand, fulvestrant.

The effect of PAPS binding on the dimer

In order to follow the effect of PAPS binding to the dimer, we compare the behavior of three different complexes: the apo dimer (without any ligands), a dimer with one PAPS molecule in chain A (1PAPS dimer), and a dimer with PAPS bound to both chains (2PAPS dimer). Figure 20 shows how the RMSD changes upon PAPS binding. The RMSD for chain A of the dimer, where PAPS binds significantly decreases (1.21±0.15 Å) compared to both chains of the apo dimer. The PAPS-less B chain maintaining its RMSD values similar to the apo dimer chain. The 2PAPS dimer showing a decreased RMSD value in both chains.

The RMSF values also shown in Figure 20 represent these changes. In the 1PAPS dimer (Figure 20 D), chain A with PAPS bound exhibiting lower fluctuation at all the biologically important loop as well as stabilizes the α3L5 and α14L3’ helix-loop regions, compared to both chain B and the apo enzyme (Figure 20 B). In the 2PAPS dimer (Figure 20F), both domains are rigid, while interestingly L4 retaining high fluctuation reaching 4 Å.

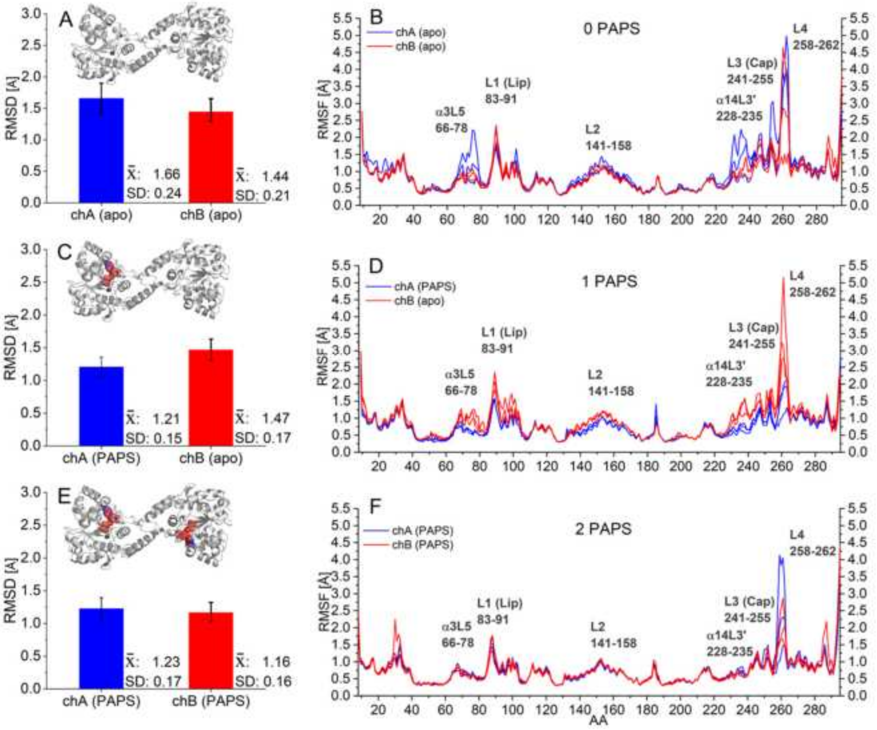

Figure 20. *PAPS binding effect on dimer systems. RMSD of the apo (A) 1PAPS (C) and 2PAPS (E) was calculated on the protein backbone, with respect to the crystal structure. RMSF of the Cα atoms for each MD simulations with the parallel MD simulations plotted on the same panel. Chain A (blue) and Chain B (red) of the dimers are plotted on apo (B), 1PAPS (D) and 2PAPS (F)*

Figure 21 compares distance values upon PAPS binding. 21A and 21D compare the loop opening of L4, showing that PAPS binding to Chain A restricts the opening of L4 between 6-12 Å. Without the cofactor, this value varies between 5-19 Å. In the 2PAPS dimer, where PAPS is bound to both chains, the L4 opening is similarly constrained (Figure 21G) as in the PAPS-bound chain of the 1PAPS system (Figure 21D). The gate opening distances in the 1PAPS dimer (Figures 21B-C and 21E-F) shows PAPS binding constrains the gate opening in Chain A, while chain B remains flexible, both at the cofactor and the ligand binding sites (Figure 21E and F). Figures 21H and 21I show that the gate opening is rigidified upon binding of the second PAPS to the dimer which is manifested on one hand by the more confined population distribution, and on the other hand, by the solid appearance of the L1-L2 closed conformations peaked around 9 Å and of L1-L3 at 6 Å.

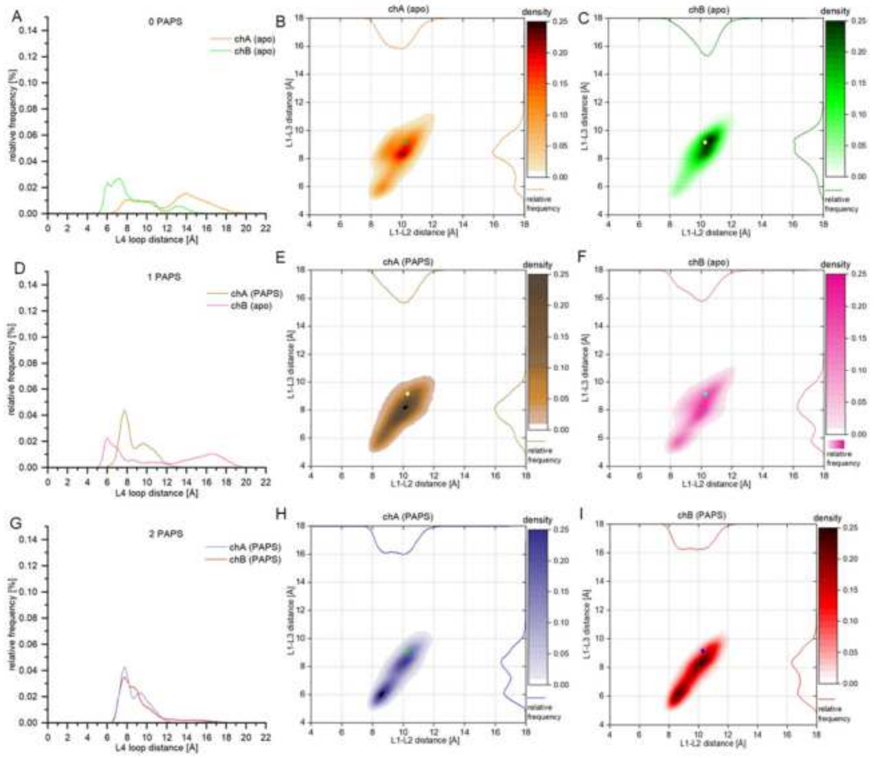

Figure 21. *Distance plots of the distribution of L1-L2 and L1-L3 distances for the MD generated structures. For SULT1A1 apo (B,C) with PAPS (E,F) PAPS+fulvestrant (H,I). The relative frequency of distance L4 during the MD simulations for the apo (A), with PAPS (D) and PAPS+fulvestrant (G). The same colors represent the same chain. The distances of the initial crystal structure (PDBID 4GRA) is denoted in each graph with a colored diamond symbol*

Based on the previous data we can see that the 1PAPS binding from one hand affects even remote parts of the ligand binding pocket if its own chain and on the other hand influences the behavior of the PAPS-less chain of the dimer. To follow this allosteric effect in more detail we performed FDA calculations to closely examine how the PAPS binding to chain A affects the dimer. Differences in atomic pairwise forces between the apo and the 1PAPS bound dimer were calculated and summed for each atom to determine atomic punctual stresses (more details in methods). These stresses were then summed up by residue and mapped onto the 3D structure of the dimer in a color-coded format, as shown in Figure 22.

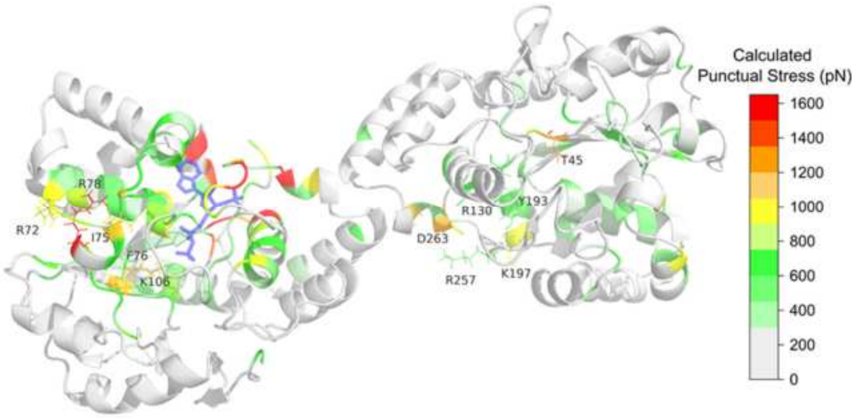

Figure 22. *Residue based punctual stress. Color coded representation of residue based punctual stress between (2SULT1A1+1PAPS) - (2SULT1A1), mapped (as a β factor value) on the 3D structure of the dimer, Colors range from grey (minimum value) to red (maximum). Residues with high value are labeled and shown as wire. PAPS, the is the perturbation effect is denoted by blue sticks. Ligands are presented as sticks, the ligand effecting the examined perturbation is colored white*

The Figure shows that increased punctual stresses are observed in both chains of the dimer, showing straight forward allosteric effect. Besides the high stress values of PAPS binding residues of chain A, which contains the bound PAPS, increased stress can be seen at more remote parts of chain A at residues of α3L5 region: R72, I75, R78. It is important to note, that F76 and K106 also exhibits considerable punctual stress, which are part of the ligand binding pocket, exactly where the RMSF shows a decrease in fluctuation upon PAPS binding.

The FDA results also show how the PAPS binding in chain A allosterically affects the PAPS binding site of chain B, showing increased stress values at residues T45, R130, Y193 constituting the binding pocket, and further more K197, D263 which in turn interacts with R257 (26, 118) stabilizing the movement of L4.

The effect of fulvestrant binding on the dimer

To follow the fulvestrant binding effects on the dimer we compare the behavior of the 2PAPS containing dimer with 2PAPS+1fulvestrant (with fulvestrant in chain A) and 2PAPS+2fulvestrant containing systems. By comparing the RMSD values of these systems we noted that binding of 1 fulvestrant loosens its own chain, not changing the other fulvestrant-less and PAPS containing chain's RMSD significantly (figure 23 A and C), while upon binding of the second fulvestrant the RMSD of the second chain also increases.

The RMSF values upon fulvestrant binding (Figure 23 B, D) show, that one fulvestrant significantly increases the flexibility of L1, L3, α14L3, L4 and α3L5 on the same chain. With two fulvestrant bound, all biologically important loops exhibit high flexibility.

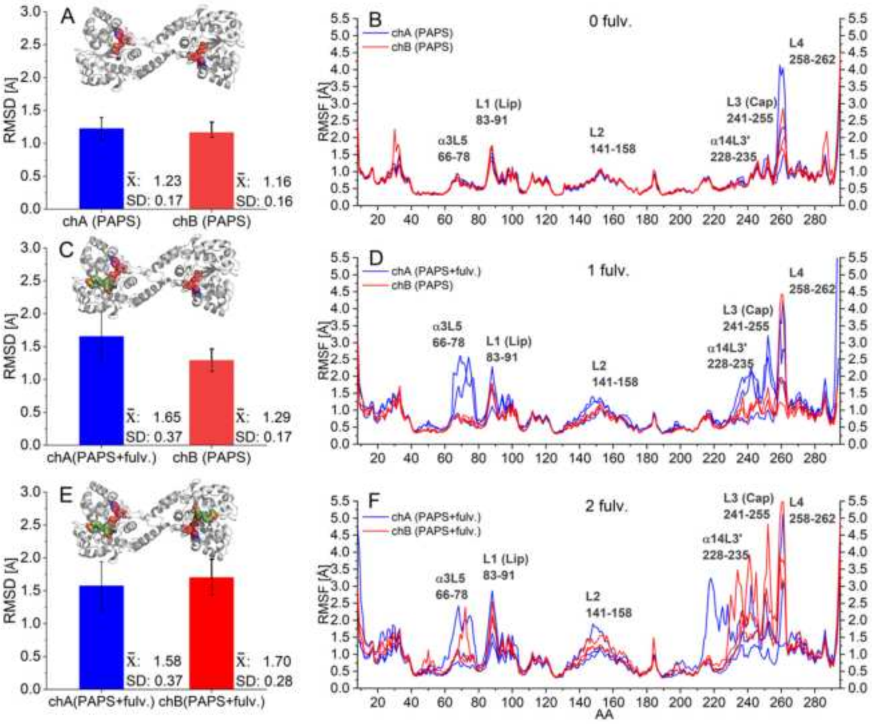

Figure 23. *Fulvestrant binding effect on dimer systems. RMSD of the 0 fulv. (A) 1 fulv. (C) and 2 fulv. (E) was calculated on the protein backbone, with respect to the crystal structure. RMSF of the Cα atoms for each MD simulations with the parallel MD simulations plotted on the same panel. Chain A (blue) and Chain B (red) of the dimers are plotted on 0 fulv. (B), 1 fulv. (D) and 2 fulv. (F)*

In Figure 24, we present the extent of loop openings upon fulvestrant binding. The comparison of part A and D of the Figure binding of one fulvestrant introduces a new, more open population of L4 conformations around 12 Å. Binding the second fulvestrant results in an asymmetric L4 opening, with chain A showing rather closed while chain B showing widely open conformations as well. As we discussed before, the gate opening in the 2PAPS systems is initially constrained to two conformations, but binding of one fulvestrant to chain A eliminates the closed gate conformation, leading to a single, well-defined one state distribution spreading towards the open conformations. The fulvestrant-less B chain of the system retains its rather closed conformation (Panel F and C), but exploring a wider conformational space compared to the fulvestrant-less dimer. Upon binding of the second fulvestrant, similarly to the fulvestrant containing chain of the 1fulvestrant system, both gates show a one state population spreading towards the open gate conformations.

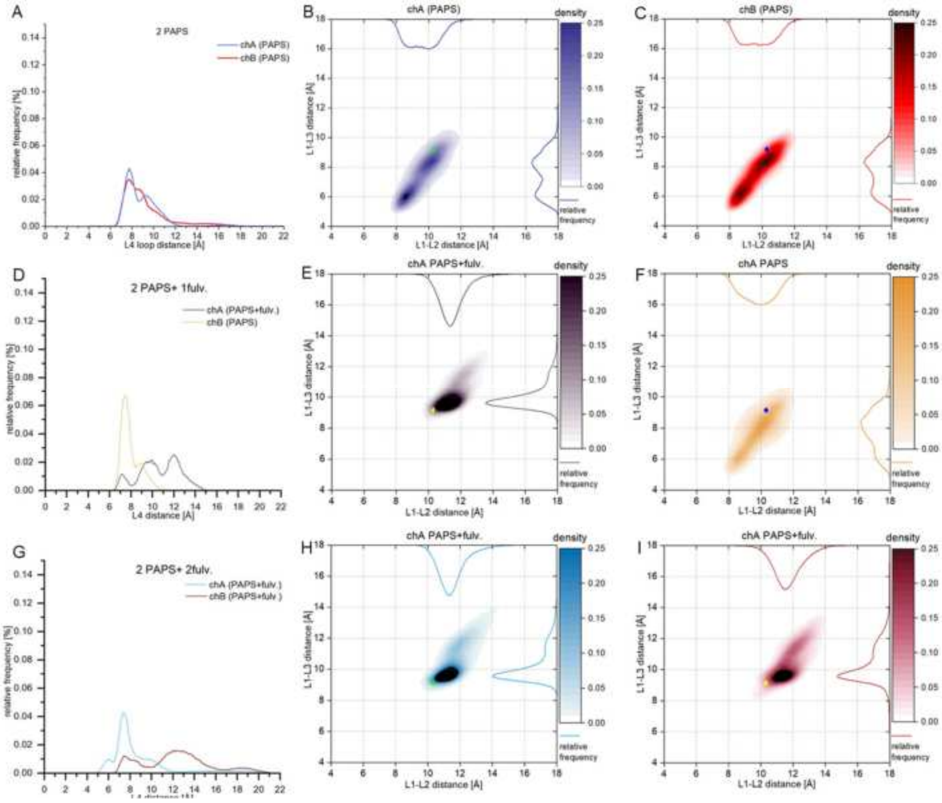

Figure 24. *Distance plots of the distribution of L1-L2 and L1-L3 distances for the MD generated structures. For SULT1A1 0 fulv. (B,C) with 1 fulv. (E,F) 2 fulv. (H,I). The relative frequency of distance L4 during the MD simulations for the 0 fulv. (A), 1 fulv. (D) and 2 fulv. (G). The same colors represent the same chain. The distances of the initial crystal structure (PDBID 4GRA) is denoted in each graph with a colored diamond symbol*

FDA calculation was also used to have a closer view of the allosteric effect of how fulvestrant binding to chain A affects gate opening of the ligand binding site of the fulvestrant-less chain B. Differences in atomic pairwise forces between the 2PAPS and 2PAPS+1fulvestrant bound dimer were calculated. Not surprisingly, Figure 25 shows increased stress values at the fulvestrant binging site of chain A extending to residues of L3, and α3L5, but the residues involved in PAPS binding of chain A also show increased punctual stress. It is important to note that these are the same residues sensed by FDA calculations upon PAPS binding discussed in the previous section. Namely D263 and R257 (their interaction stabilizing L4) and K197. A clear allosteric effect can be seen in the fulvestrant-less B chain as well, showing an increased punctual stress at the PAPS binding residue R130 and of the ligand binding residues R72, R78.

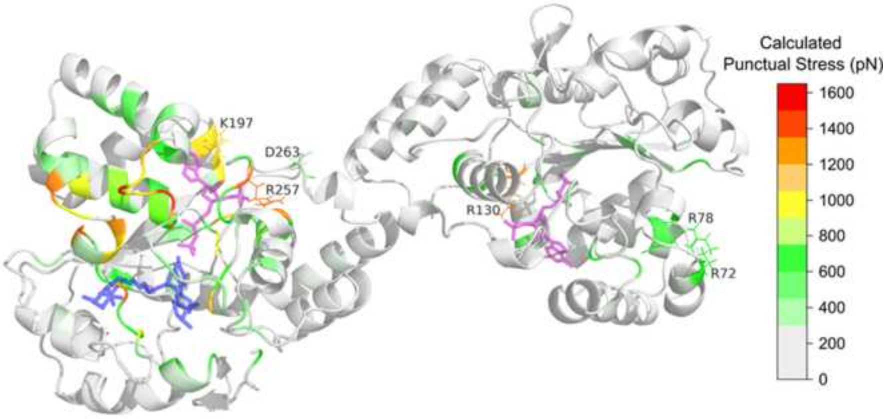

Figure 25. *FDA punctual stress calculated on the systems (SULT1A1+2PAPS+1fulv)-(2SULT1A1+2PAPS). On the 3D structure of the dimer colors range from grey (minimum value) to red (maximum). Residues with high value are labeled and shown as wire. Ligands are presented a sticks, PAPS colored purple while fulvestrant, the presence or absence of which we examine, is colored blue.*

5. DISCUSSION

5.1. The selectivity of SULT1A1 and SULT1A3

5.1.1. Efficiently mapping the conformational space of SULT1A1 by MDeNM

MD simulations have been instrumental to elucidate many mechanisms of DMEs. Regarding SULT1A1, Cook et al. used MD to elucidate the mechanisms of SULTs explaining their selectivity (88) where they described the gate flexibility that governs sulfotransferase selectivity of SULT1A1 and SULT2A1 in response to nucleotide binding, thereby regulating access to large substrates. In their study, they proposed that the presence of PAPS restricts the binding of large substrates like fulvestrant, while small substrates remain unaffected. This hypothesis suggests that at saturating PAPS levels, the concentration of the open form of the enzyme drops, weakening affinity for large substrates but not small ones. They also described a "molecular clamp" mechanism where the usual phenolic substrates induced residue rearrangement around the phenolic moiety, vastly increasing the affinity of SULT1A1 enzymes to these substrates (119).

The inhibition mechanism of SULT1A1 by certain NSAIDs like mefenamic acid has been proven, with explanation provided based on molecular dynamics, equilibrium and pre-steady-state ligand-binding studies (120). Their results suggest an indirect (helix-mediated) stabilization of the closed form that inhibits SULT1A1 turnover. Isvoran et al. used MD simulations to examine the wild type and allelic variants SULT1A1*2 (R213H) and SULT1A1*3 (M223V). Their study aimed to understand how amino acid substitutions affect enzyme dynamics in both apo and holo states. They reported increased flexibility in the loop regions surrounding the substrate-binding site (41).

In order to elucidate the molecular mechanisms involved in sulfate transfer catalyzed by SULT enzymes, in our first project, we employed classical MD simulations, an advanced method known as MDeNM, and ensemble docking (101). Our results highlight the significance of using MDeNM to achieve exhaustive conformational sampling of SULT1A1. Compared to classical MD simulations, MDeNM generated a more extensive range of conformations, including "open-like" states that are essential for the accommodation of large substrates such as fulvestrant. This expanded conformational sampling has been achieved even in the presence of the active co-factor PAPS. Indeed, crystal structures without a bound PAPS cofactor have been observed to undergo significant unfolding, thereby demonstrating its importance (22). Nevertheless, classical MD simulations demonstrated a more constrained fluctuation of the loops L1-L3, primarily constrained around a single energy minimum, as evidenced by the narrower RMSD distribution. However, the combination of MD and MDeNM provided a more comprehensive picture, revealing that the binding pocket of SULT1A1 can adopt conformations with RMSDs up to 2.25 Å from the crystal structure, even in the presence of PAPS, which could be able to accommodate large ligands. Previous investigations on SULT1A3 were less numerous. In one study, QSAR analysis have been carried out with a series of phenols and catechol (121), determining the high selectivity of this isoenzyme, and using site-directed mutagenesis (E146A) to confirm the pivotal role of this residue. In the research that published the first complete, high-resolution X-ray crystallographic image of SULT1A3 (PDBID 2A3R, (89)), Lu et al. complemented with docking analysis of D-Dopa (a compound slightly different from dopamine containing an extra carboxylic group, making it chiral). Their results suggested the important selective role of D86 and E146 in comparison to SULT1A1 which lacks these groups. However, the gating and ligand selection/binding were not explained in these studies.

5.1.2. Comparative Analysis of SULT1A1 and SULT1A3

Comparison of SULT1A1 and SULT1A3, as well as the simulation of SULT1A3 with the selective substrate dopamine yielded new insight into on the dynamic behavior of the enzymes due to their structural specificities. The PAPS-bound SULT1A3 displayed significantly higher flexibility than SULT1A1, while the selective substrate dopamine stabilized the enzyme. The inclusion of dopamine in the SULT1A3 system was particularly interesting, as it provided insight into how substrate binding affects the enzyme's dynamic behavior. Our results regarding the high degree of flexibility, observed in L3 of SULT1A3 corroborate previously reported results on mutational analysis (53, 122). To further elucidate this flexibility principal component analysis (PCA) with an approximated Free energy landscape (FEL) was calculated, confirming the different behavior of the isoenzymes.

The differences in the energy landscapes between SULT1A1 and SULT1A3 have significant implications for their respective enzymatic functions. The shallow energy minimum of SULT1A1 suggests that despite its relatively rigid structure, it can accommodate a large variety of ligands, making it more adaptable to different substrates. The energy landscape of SULT1A1 is indicative of a protein that operates within a narrow range of conformational states, which may facilitate the retention of diverse ligands once they are accommodated within the binding pocket.

In contrast to SULT1A1, SULT1A3 forms multiple distinct energy minima which are separated by higher energy barriers. The presence of dopamine stabilized specific conformations of SULT1A3, reducing the RMSD and restricting the conformational space to two subpopulations. The rugged landscape of SULT1A3 with multiple energy minima may allow it to be more selective in substrate binding, with each conformational state potentially corresponding to a different substrate or inhibitor binding mode. The presence of dopamine, which stabilizes specific conformational states, further suggests that SULT1A3 might utilize its flexibility to fine-tune its substrate specificity through induced fit mechanisms. The energy landscape of SULT1A3 landscape indicates that SULT1A3 can exist in multiple stable conformational states, each associated with different conformational subpopulations. (Figure 13 E-F). The higher energy barriers suggest that transitions between these states are less frequent, particularly in the presence of dopamine, which further constrains the conformational space by stabilizing specific energy minima. Experimental studies on the flexibility of the L3 loop in SULT1A3 have shown that the absence of PAPS in the crystal structures of SULT1A3 leads to a disordered L3 loop, further supporting the idea that ligand binding is crucial for stabilizing the active site and ensuring proper substrate recognition. The RMSF and FDA analysis thus provided insights into the dynamic nature of SULT1A3 and underscored the importance of loop flexibility in the enzyme's function. Dynamic study therefore indicates that the relative rigidity of SULT1A1 may contribute to the retention of diverse ligands within the binding pockets once they are accommodated. In contrast, the highly substrate-specific SULT1A3 employs an additional mechanism for recognizing its specific ligands. Additionally, the formation of multiple subpopulations of conformations was observed, as illustrated in Figure12 C, F.

To further elucidate the reasons of selectivity, we have employed molecular docking on conformational ensembles. Interestingly, in this ensemble docking, inhibitors showed better interaction energies than the substrates across both isoforms, which might be due to the fact that in the IEs calculated here, the reactivity of the substrates is not taken into account. In fact, the substrate reactivity is key for the catalytic reaction, and sometimes it makes a stronger contribution than the binding energy itself (123). Similar trends can be observed when comparing selective substrates to selective inhibitors, namely inhibitors showing stronger IE than substrates. In addition, the selective substrates of SULT1A3 show a clear division of IE depending on their size.

Regarding the selectivity of SULT1A3, our modeling results clearly show that the ligand position is strongly dependent on the distance between the sidechains of D86 and E146. On the contrary, the small sidechains of A86 and A146 in SULT1A1 allow more diverse ligands to be accommodated in the active site.

The substrate specificity of SULT1A3 has been explored in several site-directed mutagenesis studies, demonstrating that the variants E146A, D86A, and E89I of SULT1A3 exhibited the characteristics of SULT1A1, reducing Km for p-nitrophenol and increasing it for dopamine by two magnitudes (53),(122). Although F247 has been previously suggested to participate in allosteric binding (124), our results indicate its importance also for the binding of substrates in the active site. The phenolic groups of F81 in SULT1A1 and F84 in SULT1A3 have been previously noted in the literature to be strong attractors of phenolic compounds, thus acting as a molecular clamp (119). Interestingly, in our results, V84 of SULT1A3 has shown some contacts only with the common substrates. Thus, we could conclude that residues F84, K106, and Y240 are important for the interactions in the two isoforms, and thus contribute to their promiscuity.

5.1.3. Gating Mechanism and Substrate Recognition

To build upon the substrate recognition, first explored in our first article (101), significant insights into the gating mechanism and substrate recognition of SULT enzymes have been made (101). Conformational ensembles generated by MD and MDeNM simulations for SULT1A1 and the docking of known substrates provided virtual screening information (43, 125-128). However, it should be noted that the comparison between experimental free binding energies and docking scores can only be qualitative. It is also noteworthy, that the opening along L1-L3 has been particularly important in the binding of larger ligands, a phenomenon that has been further elucidated in our second publication.

To this end, a larger set of ligands has been collected and classified according to their functional characteristics. These ligands have been docked on both isoenzymes. As anticipated, the higher flexibility and broader conformational space of SULT1A3 facilitated an overall lower interaction energy of substrates. However, the interaction energy of smaller, unspecific ligands was less favorable than expected. Subsequent analysis revealed that the binding pocket of SULT1A1, with its non-specific aromatic residues (F247 and F84) and the functionally important residues (Y240 and K106), are capable of accommodating a diverse range of ligands. This result is in accordance with earlier studies also recognizing the role of these residues (41, 119). The docking results indicated that SULT1A3 is less efficient for accommodating small diverse ligands, which aligns with its higher substrate specificity compared to SULT1A1. Additionally, the binding site lacks aromatic residues but instead contains carboxylic groups, D86 and E146, which specifically stabilize compounds if the ionic interaction matches, explaining high selectivity it shows for primer amines, like dopamine. However, the previously discussed approximate nature of docking/scoring making definitive statements unreliable. Although these studies yielded comprehensive results, it is necessary to consider the limitations of the research. For example, all the results discussed above were obtained in the presence of the active co factor PAPS in our simulations. In the crystal structures, the inactive cofactor PAP is present. In previous studies, it has been proposed that the recognition of large substrates depends on co-factor isomerization (PAP/PAPS) (26). However, this was not investigated in the present study, as the focus was on ligand recognition and binding, which is assumed to occur while the active cofactor is bound.

5.2. Discussion on SULT dimerization

The dimerization effect of SULT isoenzymes have been studied particularly on SULT1B1 (118). Using site-directed mutagenesis the role of the key amino acids K266 and E275 were demonstrated which are part of the small dimerization motif (KxxxTVxxxE) of every SULT. However, the role and importance of the dimerization was debated over the years. Lu et al (43) observed no significant impact on enzyme activity, they proposed that dimerization serves as a stabilizing factor, enabling the enzyme to withstand higher temperatures and urea concentrations without unfolding. In contrast, other studies have suggested half-site reactivity, indicating that subunit communication through the dimerization region is dependent on cofactor binding in the case of SULT1B1 (118). Similar effects have been observed for SULT1E1 (45) and SULT1A1 (125).

When examining the apo enzyme, we observe that the fluctuation and the opening of functional loops increases upon dimerization, in this way increasing the possibility to binding the cofactor and the ligand. Furthermore, the apo dimer shows an asymmetric flexibility of its two chains one being more flexible, mapping more open conformations than its counterpart, such being more capable to bind the cofactor. (Figure 17 B-D).

5.2.1. Effect of PAPS binding

Following the behavior of the dimer upon PAPS binding in general we can state that the enzyme becomes less flexible, upon binding the cofactor. Our results show that 1PAPS-binding rigidifies particularly the loop over the cofactor binding site (L4); but furthermore, on one hand affects the gate opening of the PAPS containing chain confining it towards the more populated open conformations and on the other hand L4 of the PAPS-less chain exhibits a large fluctuation, mapping from the very closed to the very open conformations topologically preparing the PAPS binding. The reduced fluctuation upon PAPS binding also shows that the cofactor binding has a stabilization effect on the 3D structure of the enzyme (Figure 20).

Our FDA results also corroborate the fluctuation data showing punctual stress at F76 which is part of the ligand binding pocket and K106, a catalytically important residue of the PAPS interacting chain. This can be interpreted as PAPS binding alters/prepares the ligand binding. All these results point out an intra-chain allosteric effect of the PAPS binding (i.e. within the PAPS biding chain) by confining the gate opening and shifting it towards the more open conformations with influencing the ligand-binding residues. Furthermore, besides the intra-chain, an inter-chain allosteric effect is also present by the PAPS-less chain showing an increased fluctuation and punctual stress on PAPS binding residues, such preparing for PAPS to the other chain (Figure 22).

These results are only partly in line with the findings of Wang et al. (46) that the binding of the first nucleotide induces the closure of the ligand binding site on the same subunit. This might suggest a cooperative binding of the cofactor and substrate, while stabilizing the ligand binding site of the adjacent subunit in an open position. Our finding suggests that upon 1 PAPS binding the gate of the same subunit does not close, but its fluctuation is reduced around the open conformation (Figure 21 E). Upon binding of the second PAPS molecule the fluctuation L4 and the gate opening in both chains are reduced (Figure 21 H-I), the later showing two well defined minima one with a closed, one with a more open conformation possibly being efficient of binding smaller ligands. Our results would also fine-tune the findings of Wang et al. (46), showing that besides the open conformation of the gate opening a more closed conformation is also present. All these findings suggest that the structural changes which dictate the ligand and cofactor bindings are dependent on PAPS occupancy and subunit communication.

5.2.2. Effect of fulvestrant binding

One fulvestrant binding to the 2PAPS containing dimer not only affects the ligand-binding gate but also induces significant changes in the flexibility of L4 in the same chain shifting it towards the more open conformations (Figure 24 D). This effect is also sensed by FDA calculations showing increase punctual stress at the PAPS binding residues. These data indicate that the intra-chain allostery seems to be a two-way-street communication between the cofactor- and the ligand- binding site: if PAPS binds, the ligand binding site is affected by it, and vice versa: if a ligand binds, the PAPS binding site is going to detect it.

As for the inter-chain allostery we can note that the population of the ligand binding gate of the fulvestrant-less chain is shifted towards the more open conformations, which can be interpreted as the preparation for binding of the second fulvestrant. This is also being supported by the FDA calculations showing an increased punctual stress at the ligand binding residues and at the PAPS binding residues as well. The detected inter-chain allosteric effects that alter the structural dynamics leading to an asymmetric behavior of the two chains of the enzyme will play a crucial role in in ligand binding and enzyme activity, which is described as the SULT enzyme half-site reaction, a phenomenon long discussed in the literature of the field (44, 46).

6. CONCLUSIONS

In this comprehensive study we explored the dynamic behaviors of the sulfotransferase isoenzymes SULT1A1 and SULT1A3. By employing MD and MDeNM simulations of both monomer and dimers, and combining these results with ensemble docking of a collection of ligands for both isoenzymes.

The advanced MDeNM simulations demonstrated better performance compared to classical MD simulations in generating a broader range of conformations, including "open-like" states of PAPS-bound SULT1A1. These results demonstrated that SULT1A1 is capable of accommodating larger substrates, such as fulvestrant, regardless of co-factor occupancy. This broader conformational sampling was necessary for clarifying the dynamic flexibility required for this transition.

The comparative analysis revealed that SULT1A3 exhibits greater structural flexibility than SULT1A1, particularly of the functional loops L2 and L3. The relative rigidity of SULT1A1 would suggest higher selectivity, but it is evident that this is not the case here. One may speculate that a more complex mechanism for the broader substrate range is involved for SULT1A1. For instance, the application of its flexible L1 (participating in the gating) would be of great importance. The distinctive characteristics within the binding site are crucial for their function. Ensemble docking demonstrated the key residues for SULT1A1 are F247 and F84 are responsible for the interactions with aromatic groups of substrates, while the carboxylic groups of D86 and E146 of SULT1A3 are responsible for its substrate specificity.

The application of MD simulations with ligands on the dimeric SULT1A1 revealed previously unidentified loop fluctuations. Furthermore, the overall flexibility of the enzyme has increased upon dimerization. The presence of the large ligand further enhanced this flexibility, while the cofactor stabilized the enzyme. Asymmetric behavior has been noted between the subunit chains. These results suggest an allosteric behavior between both the active site accommodation, and the neighboring subunit communication on ligand binding, especially for the binding of the cofactor PAPS, which might explain the experimentally observed but poorly understood behavior and role of dimerization in this enzyme family.

7. SUMMARY

A major challenge associated with identifying promising drug candidates is to find a good balance between the required efficacy, selectivity and affinity against its intended therapeutic target while also showing an appropriate absorption, distribution, metabolism, excretion and toxicity (ADME-Tox) profile. A high percentage of drug candidate failures are due to toxicity or undesirable drug-drug interactions (DDIs), many of these are due to the inhibition of Drug Metabolizing Enzymes (DMEs). Sulfotransferases (SULTs) are one such family with a wide variety of endogenous compounds and drugs. Although SULT1A1 and SULT1A3 share 93% identity, SULT1A1, exhibits a broad substrate range, while SULT1A3 displays a high affinity toward monoamine neurotransmitters like dopamine. The first aim of the thesis was to elucidate the selectivity of two major SULT1 isoforms, SULT1A1 and SULT1A3, by combining different *in silico* methods. To understand the factors determining the substrate specificity of the SULT1 isoenzymes, we studied the dynamic behavior and structural specificities of SULT1A1 and SULT1A3 by using molecular dynamics (MD) simulations and an enhanced MD approach, MD with excited Normal Modes (MDeNM). Additionally, we performed ensemble docking of common and specific substrates of the two isoforms. These results helped us to elucidate molecular mechanisms guiding the recognition of diverse substrates and inhibitors of SULT1. We identified key protein residues strongly involved in the recognition of different substrates for the two isoforms. Our analyses indicated that being more specific and more flexible, the structure of SULT1A3 has particularities in the binding site, which are crucial for its substrate selectivity. Our second main objective was to better understand how the dimerization is involved in their specificity. Performing comparative MD simulations for SULT1A1 as a monomer and as a dimer, with bound cofactor PAPS and the bulky substrate fulvestrant, revealed unexpected interactions within the dimer, and a clear influence of the dimerization on the conformational behavior of the enzyme, which is strongly related to substrate binding. Our results shed new light on the molecular mechanisms involved in the specificity of SULT1 family important for the metabolism of endogenous compounds, drugs, and DDIs.

8. REFERENCES

1. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ. 2016;47:20-33.
2. Naga D, Parrott N, Ecker GF, Olivares-Morales A. Evaluation of the Success of High-Throughput Physiologically Based Pharmacokinetic (HT-PBPK) Modeling Predictions to Inform Early Drug Discovery. Mol Pharm. 2022;19(7):2203-16.
3. Dara S, Dhamercherla S, Jadav SS, Babu CM, Ahsan MJ. Machine Learning in Drug Discovery: A Review. Artif Intell Rev. 2022;55(3):1947-99.
4. Meyer UA. Overview of enzymes of drug metabolism. J Pharmacokinet Biopharm. 1996;24(5):449-59.
5. Bojić M, Debeljak Ž, Guengerich FP. Principles of Xenobiotic Metabolism (Biotransformation). In: Primorac D, Höppner W, Bach-Rojecky L, editors. Pharmacogenomics in Clinical Practice. Cham: Springer International Publishing; 2023. p. 13-33.
6. Isin EM. Unusual Biotransformation Reactions of Drugs and Drug Candidates. Drug Metab Dispos. 2023;51(4):413-26.
7. Esteves F, Rueff J, Kranendonk M. The Central Role of Cytochrome P450 in Xenobiotic Metabolism-A Brief Review on a Fascinating Enzyme Family. J Xenobiot. 2021;11(3):94-114.
8. Guengerich FP. Cytochrome p450 and chemical toxicology. Chem Res Toxicol. 2008;21(1):70-83.
9. Martiny VY, Miteva MA. Advances in molecular modeling of human cytochrome P450 polymorphism. J Mol Biol. 2013;425(21):3978-92.
10. Brand W, Boersma MG, Bik H, Hoek-van den Hil EF, Vervoort J, Barron D, et al. Phase II metabolism of hesperetin by individual UDP-glucuronosyltransferases and sulfotransferases and rat and human tissue samples. Drug Metab Dispos. 2010;38(4):617-25.
11. Rheinberger H-Jr. Toward a history of epistemic things : synthesizing proteins in the test tube. Stanford, Calif.: Stanford University Press; 1997. x, 325 p. p.
12. Morange M. A history of molecular biology. Cambridge, Mass.: Harvard University Press; 1998. 336 p. p.
13. Rheinberger HJ. [Cultures of experiment]. Ber Wiss. 2007;30(2):135-44.
14. Blanchard RL, Freimuth RR, Buck J, Weinshilboum RM, Coughtrie MW. A proposed nomenclature system for the cytosolic sulfotransferase (SULT) superfamily. Pharmacogenetics. 2004;14(3):199-211.
15. Seal RL, Braschi B, Gray K, Jones TEM, Tweedie S, Haim-Vilmovsky L, et al. Genenames.org: the HGNC resources in 2023. Nucleic Acids Res. 2023;51(D1):D1003-D9.
16. Jakoby WB, Ziegler DM. The enzymes of detoxication. J Biol Chem. 1990;265(34):20715-8.
17. Wang J, Falany JL, Falany CN. Expression and characterization of a novel thyroid hormone-sulfating form of cytosolic sulfotransferase from human liver. Mol Pharmacol. 1998;53(2):274-82.
18. Meinl W, Donath C, Schneider H, Sommer Y, Glatt H. SULT1C3, an orphan sequence of the human genome, encodes an enzyme activating various promutagens. Food Chem Toxicol. 2008;46(4):1249-56.
19. Runge-Morris M, Kocarek TA. Expression of the sulfotransferase 1C family: implications for xenobiotic toxicity. Drug Metab Rev. 2013;45(4):450-9.

20. Pedersen LC, Petrotchenko E, Shevtsov S, Negishi M. Crystal structure of the human estrogen sulfotransferase-PAPS complex: evidence for catalytic role of Ser137 in the sulfuryl transfer reaction. J Biol Chem. 2002;277(20):17928-32.

21. Javitt NB, Lee YC, Shimizu C, Fuda H, Strott CA. Cholesterol and hydroxycholesterol sulfotransferases: identification, distinction from dehydroepiandrosterone sulfotransferase, and differential tissue expression. Endocrinology. 2001;142(7):2978-84.

22. Allali-Hassani A, Pan PW, Dombrovski L, Najmanovich R, Tempel W, Dong A, et al. Structural and chemical profiling of the human cytosolic sulfotransferases. PLoS Biol. 2007;5(5):e97.

23. Sun Y, Machalz D, Wolber G, Parr MK, Bureik M. Functional Expression of All Human Sulfotransferases in Fission Yeast, Assay Development, and Structural Models for Isoforms SULT4A1 and SULT6B1. Biomolecules. 2020;10(11).

24. Kakuta Y, Pedersen LG, Carter CW, Negishi M, Pedersen LC. Crystal structure of estrogen sulphotransferase. Nat Struct Biol. 1997;4(11):904-8.

25. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-42.

26. Tibbs ZE, Rohn-Glowacki KJ, Crittenden F, Guidry AL, Falany CN. Structural plasticity in the human cytosolic sulfotransferase dimer and its role in substrate selectivity and catalysis. Drug Metab Pharmacokinet. 2015;30(1):3-20.

27. Toth D, Dudas B, Miteva MA, Balog E. Role of Conformational Dynamics of Sulfotransferases SULT1A1 and SULT1A3 in Substrate Specificity. Int J Mol Sci. 2023;24(23).

28. Kakuta Y, Petrotchenko EV, Pedersen LC, Negishi M. The sulfuryl transfer mechanism. Crystal structure of a vanadate complex of estrogen sulfotransferase and mutational analysis. J Biol Chem. 1998;273(42):27325-30.

29. Petrotchenko EV, Pedersen LC, Borchers CH, Tomer KB, Negishi M. The dimerization motif of cytosolic sulfotransferases. FEBS Lett. 2001;490(1-2):39-43.

30. Falany CN, Krasnykh V, Falany JL. Bacterial expression and characterization of a cDNA for human liver estrogen sulfotransferase. J Steroid Biochem Mol Biol. 1995;52(6):529-39.

31. Fuda H, Lee YC, Shimizu C, Javitt NB, Strott CA. Mutational analysis of human hydroxysteroid sulfotransferase SULT2B1 isoforms reveals that exon 1B of the SULT2B1 gene produces cholesterol sulfotransferase, whereas exon 1A yields pregnenolone sulfotransferase. J Biol Chem. 2002;277(39):36161-6.

32. Lee KA, Fuda H, Lee YC, Negishi M, Strott CA, Pedersen LC. 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. 2003;278(45):44593-9.

33. Petrotchenko EV, Doerflein ME, Kakuta Y, Pedersen LC, Negishi M. Substrate gating confers steroid specificity to estrogen sulfotransferase. J Biol Chem. 1999;274(42):30019-22.

34. Salman ED, Kadlubar SA, Falany CN. Expression and localization of cytosolic sulfotransferase (SULT) 1A1 and SULT1A3 in normal human brain. Drug Metab Dispos. 2009;37(4):706-9.

35. Teubner W, Meinl W, Florian S, Kretzschmar M, Glatt H. Identification and localization of soluble sulfotransferases in the human gastrointestinal tract. Biochem J. 2007;404(2):207-15.

36. He D, Frost AR, Falany CN. Identification and immunohistochemical localization of Sulfotransferase 2B1b (SULT2B1b) in human lung. Biochim Biophys Acta. 2005;1724(1-2):119-26.

37. Kurogi K, Shimohira T, Kouriki-Nagatomo H, Zhang G, Miller ER, Sakakibara Y, et al. Human Cytosolic Sulphotransferase SULT1C3: genomic analysis and functional characterization of splice variant SULT1C3a and SULT1C3d. J Biochem. 2017;162(6):403-14.

38. Karlsson M, Zhang C, Mear L, Zhong W, Digre A, Katona B, et al. A single-cell type transcriptomics map of human tissues. Sci Adv. 2021;7(31).

39. Riches Z, Stanley EL, Bloomer JC, Coughtrie MW. Quantitative evaluation of the expression and activity of five major sulfotransferases (SULTs) in human tissues: the SULT "pie". Drug Metab Dispos. 2009;37(11):2255-61.

40. Suiko M, Kurogi K, Hashiguchi T, Sakakibara Y, Liu MC. Updated perspectives on the cytosolic sulfotransferases (SULTs) and SULT-mediated sulfation. Biosci Biotechnol Biochem. 2017;81(1):63-72.

41. Isvoran A, Peng Y, Ceauranu S, Schmidt L, Nicot AB, Miteva MA. Pharmacogenetics of human sulfotransferases and impact of amino acid exchange on Phase II drug metabolism. Drug Discov Today. 2022;27(11):103349.

42. Weitzner B, Meehan T, Xu Q, Dunbrack RL, Jr. An unusually small dimer interface is observed in all available crystal structures of cytosolic sulfotransferases. Proteins. 2009;75(2):289-95.

43. Lu LY, Chiang HP, Chen WT, Yang YS. Dimerization is responsible for the structural stability of human sulfotransferase 1A1. Drug Metab Dispos. 2009;37(5):1083-8.

44. Beckmann JD, Burkett RJ, Sharpe M, Giannunzio L, Johnston D, Abbey S, et al. Spectrofluorimetric analysis of 7-hydroxycoumarin binding to bovine phenol sulfotransferase. Biochim Biophys Acta. 2003;1648(1-2):134-9.

45. Sun M, Leyh TS. The human estrogen sulfotransferase: a half-site reactive enzyme. Biochemistry. 2010;49(23):4779-85.

46. Wang T, Cook I, Falany CN, Leyh TS. Paradigms of sulfotransferase catalysis: the mechanism of SULT2A1. J Biol Chem. 2014;289(38):26474-80.

47. Coughtrie MWH. Function and organization of the human cytosolic sulfotransferase (SULT) family. Chem Biol Interact. 2016;259(Pt A):2-7.

48. Hebbring SJ, Adjei AA, Baer JL, Jenkins GD, Zhang J, Cunningham JM, et al. Human SULT1A1 gene: copy number differences and functional implications. Hum Mol Genet. 2007;16(5):463-70.

49. Falany JL, Macrina N, Falany CN. Regulation of MCF-7 breast cancer cell growth by beta-estradiol sulfation. Breast Cancer Res Treat. 2002;74(2):167-76.

50. Li L, Falany CN. Elevated hepatic SULT1E1 activity in mouse models of cystic fibrosis alters the regulation of estrogen responsive proteins. J Cyst Fibros. 2007;6(1):23-30.

51. Moore KL. The biology and enzymology of protein tyrosine O-sulfation. J Biol Chem. 2003;278(27):24243-6.

52. Steventon GB, Heafield MT, Waring RH, Williams AC. Xenobiotic metabolism in Parkinson's disease. Neurology. 1989;39(7):883-7.

53. Brix LA, Barnett AC, Duggleby RG, Leggett B, McManus ME. Analysis of the substrate specificity of human sulfotransferases SULT1A1 and SULT1A3: site-directed mutagenesis and kinetic studies. Biochemistry. 1999;38(32):10474-9.

54. Bairam AF, Rasool MI, Alherz FA, Abunnaja MS, El Daibani AA, Kurogi K, et al. Effects of human SULT1A3/SULT1A4 genetic polymorphisms on the sulfation of acetaminophen and opioid drugs by the cytosolic sulfotransferase SULT1A3. Arch Biochem Biophys. 2018;648:44-52.

55. Leelananda SP, Lindert S. Computational methods in drug discovery. Beilstein J Org Chem. 2016;12:2694-718.

56. Shaker B, Ahmad S, Lee J, Jung C, Na D. In silico methods and tools for drug discovery. Comput Biol Med. 2021;137:104851.

57. Djuris J, Cvijic S, Djekic L. Model-Informed Drug Development: In Silico Assessment of Drug Bioperformance following Oral and Percutaneous Administration. Pharmaceuticals (Basel). 2024;17(2).

58. Hug S. Classical molecular dynamics in a nutshell. Methods Mol Biol. 2013;924:127-52.

59. Allen MP, Tildesley DJ. Computer simulation of liquids. New York: Oxford University Press Inc.; 1987.

60. Berg JM, Tymoczko JL, Stryer L. Biochemistry. New York: W.H.Freeman & Co Ltd; 2002.

61. Leach AR. Molecular modelling principles and applications. 2. ed. XXIII, 744 S. p.

62. Verlet L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. Physical Review. 1967;159(1):98-103.

63. Swope WC, Andersen HC, Berens PH, Wilson KR. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. The Journal of Chemical Physics. 1982;76(1):637-49.

64. Hockney RW, Goel SP, Eastwood JW. Quiet high-resolution computer models of a plasma. Journal of Computational Physics. 1974;14(2):148-58.

65. Costa MG, Batista PR, Bisch PM, Perahia D. Exploring free energy landscapes of large conformational changes: molecular dynamics with excited normal modes. J Chem Theory Comput. 2015;11(6):2755-67.

66. Voter AF. A method for accelerating the molecular dynamics simulation of infrequent events. The Journal of Chemical Physics. 1997;106(11):4665-77.

67. Tribello GA, Ceriotti M, Parrinello M. A self-learning algorithm for biased molecular dynamics. Proc Natl Acad Sci U S A. 2010;107(41):17509-14.

68. So/rensen MR, Voter AF. Temperature-accelerated dynamics for simulation of infrequent events. The Journal of Chemical Physics. 2000;112(21):9599-606.

69. Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G. String method in collective variables: minimum free energy paths and isocommittor surfaces. J Chem Phys. 2006;125(2):24106.

70. Bolhuis PG, Chandler D, Dellago C, Geissler PL. Transition path sampling: throwing ropes over rough mountain passes, in the dark. Annu Rev Phys Chem. 2002;53:291-318.

71. Hassan NM, Alhossary AA, Mu Y, Kwoh CK. Protein-Ligand Blind Docking Using QuickVina-W With Inter-Process Spatio-Temporal Integration. Sci Rep. 2017;7(1):15451.

72. Lionta E, Spyrou G, Vassilatis DK, Cournia Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. Curr Top Med Chem. 2014;14(16):1923-38.

73. Moroy G, Martiny VY, Vayer P, Villoutreix BO, Miteva MA. Toward in silico structure-based ADMET prediction in drug discovery. Drug Discov Today. 2012;17(1-2):44-55.

74. Lexa KW, Carlson HA. Protein flexibility in docking and surface mapping. Q Rev Biophys. 2012;45(3):301-43.

75. Moustakas DT, Lang PT, Pegg S, Pettersen E, Kuntz ID, Brooijmans N, et al. Development and validation of a modular, extensible docking program: DOCK 5. J Comput Aided Mol Des. 2006;20(10-11):601-19.

76. Kelley BP, Brown SP, Warren GL, Muchmore SW. POSIT: Flexible Shape-Guided Docking For Pose Prediction. J Chem Inf Model. 2015;55(8):1771-80.

77. Korb O, Stützle T, Exner TE, editors. PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. Ant Colony Optimization and Swarm Intelligence; 2006 2006//; Berlin, Heidelberg: Springer Berlin Heidelberg.

78. Lin ZP, Al Zouabi NN, Xu ML, Bowen NE, Wu TL, Lavi ES, et al. In silico screening identifies a novel small molecule inhibitor that counteracts PARP inhibitor resistance in ovarian cancer. Sci Rep. 2021;11(1):8042.

79. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009;30(16):2785-91.

80. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455-61.

81. Neves MA, Totrov M, Abagyan R. Docking and scoring with ICM: the benchmarking results and strategies for improvement. J Comput Aided Mol Des. 2012;26(6):675-86.

82. Sellers MS, Hurley MM. XPairIt Docking Protocol for peptide docking and analysis. Molecular Simulation. 2016;42(2):149-61.

83. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. WIREs Computational Molecular Science. 2013;3(2):198-210.

84. Hill AD, Reilly PJ. Scoring functions for AutoDock. Methods Mol Biol. 2015;1273:467-74.

85. Sauton N, Lagorce D, Villoutreix BO, Miteva MA. MS-DOCK: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. BMC Bioinformatics. 2008;9:184.

86. Miteva MA. In Silico Lead Discovery. Sharjah, UAE: Bentham Science Publishers; 2011.

87. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol. 2000;295(2):337-56.

88. Cook I, Wang T, Almo SC, Kim J, Falany CN, Leyh TS. The gate that governs sulfotransferase selectivity. Biochemistry. 2013;52(2):415-24.

89. Lu JH, Li HT, Liu MC, Zhang JP, Li M, An XM, et al. Crystal structure of human sulfotransferase SULT1A3 in complex with dopamine and 3'-phosphoadenosine 5'-phosphate. Biochem Biophys Res Commun. 2005;335(2):417-23.

90. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem. 2010;31(4):671-90.

91. Sondergaard CR, Olsson MH, Rostkowski M, Jensen JH. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. J Chem Theory Comput. 2011;7(7):2284-95.

92. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem. 2008;29(11):1859-65.

93. Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. J Chem Theory Comput. 2016;12(1):405-13.

94. Snyman J, Wilke N. Practical Mathematical Optimization Basic Optimization Theory and Gradient-Based Algorithms
Preface to the second edition. Springer Optim Appl. 2018;133:Vii-+.

95. Ben-Israel A. A Newton-Raphson method for the solution of systems of equations. Journal of Mathematical analysis and applications. 1966;15(2):243-52.

96. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: The Biomolecular Simulation Program. Journal of Computational Chemistry. 2009;30(10):1545-614.

97. Huang J, MacKerell AD, Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J Comput Chem. 2013;34(25):2135-45.

98. Gamage NU, Tsvetanov S, Duggleby RG, McManus ME, Martin JL. The structure of human SULT1A1 crystallized with estradiol. An insight into active site plasticity and substrate inhibition with multi-ring substrates. J Biol Chem. 2005;280(50):41482-6.
99. Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JV, Bernardi RC, et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. J Chem Phys. 2020;153(4):044130.
100. Heyer LJ, Kruglyak S, Yooseph S. Exploring expression data: identification and analysis of coexpressed genes. Genome Res. 1999;9(11):1106-15.
101. Dudas B, Toth D, Perahia D, Nicot AB, Balog E, Miteva MA. Insights into the substrate binding mechanism of SULT1A1 through molecular dynamics with excited normal modes simulations. Scientific Reports. 2021;11(1):13129.
102. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14(1):33-8, 27-8.
103. Martiny VY, Carbonell P, Lagorce D, Villoutreix BO, Moroy G, Miteva MA. In silico mechanistic profiling to probe small molecule binding to sulfotransferases. PLoS One. 2013;8(9):e73587.
104. National Institute of Health (NIH). PubChem Database. [Accessed: 2022-04-13]. Available from: https://pubchem.ncbi.nlm.nih.gov/
105. The Metabolomics Innovation Center (TMIC). DrugBankOnline. [Accessed: 2022-04-13]. Available from: https://go.drugbank.com/
106. Elsevier Information Systems GmbH. Reaxys database. [Accessed: 2022-04-15]. Available from: https://www.reaxys.com/
107. Systèmes D. BIOVIA Pipeline Pilot. 2024. Available from: www.3ds.com/products/biovia/pipeline-pilot
108. Amadei A, Linssen AB, Berendsen HJ. Essential dynamics of proteins. Proteins. 1993;17(4):412-25.
109. More-Adate P, Lokhande KB, Shrivastava A, Doiphode S, Nagar S, Singh A, et al. Pharmacoinformatics approach for the screening of Kovidra (Bauhinia variegata) phytoconstituents against tumor suppressor protein in triple negative breast cancer. J Biomol Struct Dyn. 2023:1-20.
110. Stacklies W, Seifert C, Graeter F. Implementation of force distribution analysis for molecular dynamics simulations. BMC Bioinformatics. 2011;12:101.
111. Costescu BI, Grater F. Time-resolved force distribution analysis. BMC Biophys. 2013;6(1):5.
112. Palmai Z, Seifert C, Grater F, Balog E. An allosteric signaling pathway of human 3-phosphoglycerate kinase from force distribution analysis. PLoS Comput Biol. 2014;10(1):e1003444.
113. Bogdan I Costescu FG. GROMACS-FDA: GitHub Repository. 2024. [Accessed:2024. 03. 14.] Available from: https://github.com/HITS-MBM/gromacs-fda
114. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein-ligand interaction profiler. Nucleic Acids Res. 2015;43(W1):W443-7.
115. DeLano WL. The PyMOL Molecular Graphics System. San Carlos, CA: DeLano Scientific; 2002.
116. Coughtrie MW, Johnston LE. Interactions between dietary chemicals and human sulfotransferases-molecular mechanisms and clinical significance. Drug Metab Dispos. 2001;29(4 Pt 2):522-8.
117. Cook I, Wang T, Falany CN, Leyh TS. The allosteric binding sites of sulfotransferase 1A1. Drug Metab Dispos. 2015;43(3):418-23.
118. Tibbs ZE, Falany CN. Dimeric human sulfotransferase 1B1 displays cofactor-dependent subunit communication. Pharmacol Res Perspect. 2015;3(3):e00147.

119. Cook I, Wang T, Leyh TS. Sulfotransferase 1A1 Substrate Selectivity: A Molecular Clamp Mechanism. Biochemistry. 2015;54(39):6114-22.

120. Wang T, Cook I, Leyh TS. The NSAID allosteric site of human cytosolic sulfotransferases. Journal of Biological Chemistry. 2017;292(49):20305-12.

121. Dajani R, Cleasby A, Neu M, Wonacott AJ, Jhoti H, Hood AM, et al. X-ray crystal structure of human dopamine sulfotransferase, SULT1A3. Molecular modeling and quantitative structure-activity relationship analysis demonstrate a molecular basis for sulfotransferase substrate specificity. J Biol Chem. 1999;274(53):37862-8.

122. Liu MC, Suiko M, Sakakibara Y. Mutational analysis of the substrate binding/catalytic domains of human M form and P form phenol sulfotransferases. J Biol Chem. 2000;275(18):13460-4.

123. Robinson PK. Enzymes: principles and biotechnological applications. Essays Biochem. 2015;59:1-41.

124. Cook I, Wang T, Girvin M, Leyh TS. The structure of the catechin-binding site of human sulfotransferase 1A1. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(50):14312-7.

125. Wang T, Cook I, Leyh TS. 3’-Phosphoadenosine 5’-Phosphosulfate Allosterically Regulates Sulfotransferase Turnover. Biochemistry. 2014;53(44):6893-900.

126. Thomas NL, Coughtrie MW. Sulfation of apomorphine by human sulfotransferases: evidence of a major role for the polymorphic phenol sulfotransferase, SULT1A1. Xenobiotica. 2003;33(11):1139-48.

127. Rohn KJ, Cook IT, Leyh TS, Kadlubar SA, Falany CN. 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. 2012;40(8):1588-95.

128. Rakers C, Schumacher F, Meinl W, Glatt H, Kleuser B, Wolber G. In Silico Prediction of Human Sulfotransferase 1E1 Activity Guided by Pharmacophores from Molecular Dynamics Simulations. Journal of Biological Chemistry. 2016;291(1):58-71.

9. BIBLIOGRAPHY OF THE CANDIDATE’S PUBLICATIONS

B. Dudas#, D. Toth #, D. Perahia, A. B. Nicot, E Balog, M. A. Miteva. Insights into the Substrate Binding Mechanism of SULT1A1 through Molecular Dynamics with Excited Normal Modes Simulations. *Scientific Reports* 11, no. 1 (2021): 13129.
# contributed equally
D. Toth, B. Dudas, M. A. Miteva, E. Balog. Role of Conformational Dynamics of Sulfotransferases SULT1A1 and SULT1A3 in Substrate Specificity. Int J Mol Sci 24, no. 23 (2023).

In preparation:
D. Toth, B. Dudas, A. B. Nicot, M. A. Miteva, E. Balog. Dimerisation effects on SULT1A1 structural dynamics and ligand interactions.

10. ACKNOWLEDGEMENTS

I would like to start by thanking my wonderful supervisors, Erika Balog and Maria Miteva, for all their valuable advice and support throughout my Ph.D. studies. Although this journey has not been without its difficulties, they have shown me how to focus on this topic and how to conduct research in a thorough and detailed manner. Erika, I am grateful for the long discussions that went from the smallest details to the big picture. I learned a lot from you, not only as a scientist, but also as a responsible person. Maria, I am grateful for your hospitality and professional advice, which I could always ask about any aspect of my research topic. I will never forget the advice and help I received from you, both professionally and personally.

I would like to thank all my colleagues in both the Biophysics Institute and in the Miteva laboratory. Bálint Dudás, for being the forerunner in my research, for showing me the approaches used in an easily understandable way, with such energy that I believed I could do it as easily as you did. Janka Czigléczki, I am grateful for the discussions that went into the smallest details of this complex field, when I needed them the most. Youcef Bagdad, thank you for welcoming me into the lab and helping me with both molecular docking and food selection methods.

I would like to thank the Therapeutic Targets and Drug Design (CiTCoM) group of the Faculty of Pharmacy in Paris for accepting me, and I am equally grateful to the Biophysics Institute and Miklós Kellermayer for supporting me and helping me to pursue this professional career. I have been honored to work in this environment, both in Hungary and in France. I am thankful to the Campus France and the colleagues of the Hungarian Institute Française, for their commitment and financial support.

Finally, I would like to thank my family, my mom and dad, who have supported me through the best and worst of times, never (loudly) questioning my judgment to try to become a researcher instead of continuing our family businesses they have worked so hard on. I have no words to express my gratitude for this incredible support. Thank you so much!