

# Structural and Chemical Profiling of the Human Cytosolic Sulfotransferases

Abdellah Allali-Hassani<sup>1</sup>, Patricia W. Pan<sup>1,2</sup>, Ludmila Dombrovski<sup>1</sup>, Rafael Najmanovich<sup>1,3</sup>, Wolfram Tempel<sup>1</sup>, Aiping Dong<sup>1</sup>, Peter Loppnau<sup>1</sup>, Fernando Martin<sup>1</sup>, Janet Thonton<sup>1,3</sup>, Aled M. Edwards<sup>1,2,4</sup>, Alexey Bochkarev<sup>1,4</sup>, Alexander N. Plotnikov<sup>1,5</sup>, Masoud Vedadi<sup>1</sup>, Cheryl H. Arrowsmith<sup>1,2,4\*</sup>

**1** Structural Genomics Consortium, University of Toronto, Toronto, Ontario, Canada, **2** Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada, **3** European Bioinformatics Institute, Cambridge, United Kingdom, **4** Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada, **5** Department of Physiology, University of Toronto, Toronto, Ontario, Canada

**The human cytosolic sulfotransferases (hSULTs) comprise a family of 12 phase II enzymes involved in the metabolism of drugs and hormones, the bioactivation of carcinogens, and the detoxification of xenobiotics. Knowledge of the structural and mechanistic basis of substrate specificity and activity is crucial for understanding steroid and hormone metabolism, drug sensitivity, pharmacogenomics, and response to environmental toxins. We have determined the crystal structures of five hSULTs for which structural information was lacking, and screened nine of the 12 hSULTs for binding and activity toward a panel of potential substrates and inhibitors, revealing unique “chemical fingerprints” for each protein. The family-wide analysis of the screening and structural data provides a comprehensive, high-level view of the determinants of substrate binding, the mechanisms of inhibition by substrates and environmental toxins, and the functions of the orphan family members SULT1C3 and SULT4A1. Evidence is provided for structural “priming” of the enzyme active site by cofactor binding, which influences the spectrum of small molecules that can bind to each enzyme. The data help explain substrate promiscuity in this family and, at the same time, reveal new similarities between hSULT family members that were previously unrecognized by sequence or structure comparison alone.**

Citation: Allali-Hassani A, Pan PW, Dombrovski L, Najmanovich R, Tempel W, et al. (2007) Structural and chemical profiling of the human cytosolic sulfotransferases. *PLoS Biol* 5(5): e97. doi:10.1371/journal.pbio.0050097

## Introduction

Cytosolic sulfotransferases (SULTs) comprise a family of enzymes that catalyze the transfer of a sulfonate group from 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to an acceptor group of the substrate (Figure 1). In doing so, SULTs modulate the activities of a large array of small endogenous and foreign chemicals, including drugs, toxic compounds, steroid hormones, and neurotransmitters. Because sulfonated molecules are highly soluble in water and easily excreted from the organism, SULTs are often referred to as enzymes of chemical defence. In some cases, however, SULTs activate certain compounds from food and the environment into mutagenic and carcinogenic metabolites [1].

To date, 13 human cytosolic sulfotransferase (hSULT) genes have been identified; they partition into four families [2,3]: SULT1, SULT2, SULT4, and SULT6. Although the family members share considerable sequence and structural similarity, they appear to have different biological functions. The SULT1 family comprises nine members divided into four subfamilies (1A1, 1A2, 1A3, and 1A4; 1C1, 1C2, and 1C3; 1B1; and 1E1). The SULT1A3 and SULT1A4 genes appear to have arisen from a segmental duplication and encode the same protein [4]. Members of the SULT1 family have been shown to sulfonate simple phenols, estradiol, and thyroid hormones, as well as environmental xenobiotics and drugs. The SULT2 family has two genes, encoding three proteins (SULT2A1, SULT2B1a, and SULT2B1b), which catalyze sulfonation of hydroxyl groups of steroids, such as androsterone, allopregnanolone, and dehydroepiandrosterone (DHEA). SULT4A1 is the only member of the SULT4 family. The fact that it is highly conserved and expressed primarily in the brain

suggests an important function; however, no activity or function has been identified for this gene [5]. Finally the SULT6B1 gene is expressed in the testis of primates, but neither the protein nor its enzymatic activity has been characterized [3].

Recent progress in the structural biology and characterization of the catalytic mechanism of hSULTs has established that many family members have distinct, but overlapping, substrate specificities and that the enzymes have a sequential catalytic mechanism that is susceptible to substrate inhibition [6,7]. Nevertheless, only a few of the human enzymes have been subjected to detailed structural and mechanistic studies [6,8–16], and there are no reports of a systematic comparison among all the hSULTs. Understanding the structural and mechanistic basis for specificity among hSULTs is essential to elucidate their role in the metabolism of regulatory hormones, drugs, and carcinogens, and may assist in chemical risk assessment and the design of more-effective therapeutics.

**Academic Editor:** Gregory A. Petsko, Brandeis University, United States of America

**Received:** September 13, 2006; **Accepted:** February 9, 2007; **Published:** April 10, 2007

**Copyright:** © 2007 Allali-Hassani et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

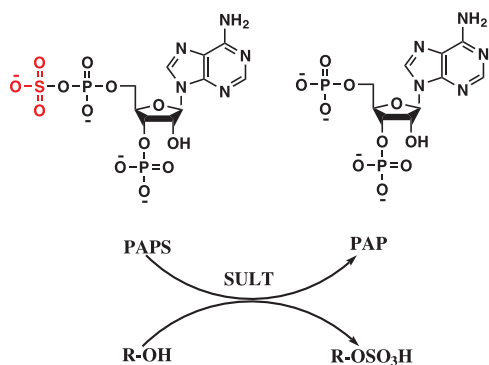
**Abbreviations:** DHEA, dehydroepiandrosterone; HPLC, high-performance liquid chromatography; hSULT, human cytosolic sulfotransferase; PAPS, PAPS or PAP; PCB, polychlorinated biphenyl; PCP, pentachlorophenol; SULT, cytosolic sulfotransferase

\* To whom correspondence should be addressed. E-mail: carrow@uhnres.utoronto.ca

## Author Summary

We metabolize many hormones, drugs, and bioactive chemicals and toxins from the environment. One family of enzymes that participate in the metabolic process consists of the cytosolic sulfotransferases, or SULTs. SULTs have a variety of mechanisms of action—sometimes they inactivate the biological activity of the chemical (e.g., in the case of estrogen). At other times, the enzymes make the chemical more toxic (e.g., for certain carcinogens). Humans have 12 distinct SULT enzymes. Determining how each of these human enzymes recognizes and distinguishes between the thousands of chemicals we confront each day is essential for understanding hormone regulation, assessing environmental risk, and eventually developing better, more-effective drugs. We have studied the human SULT family of enzymes to profile which small molecules are recognized by each enzyme. We also visualized and compared the detailed structural features that determine which enzyme interacts with which molecule. By studying the entire family, we discovered new ways in which chemicals interact with each enzyme. Furthermore, we identified new inhibitors and inhibitory mechanisms. Finally, we discovered functions for many of the human enzymes that were previously uncharacterized.

Here we report the crystal structures of five of the 12 structurally unique hSULTs. These structures, combined with those previously reported for six other hSULTs, allowed a comprehensive comparison of both global and local structural features. We further screened nine hSULTs for binding activity toward a set of 90 potential substrates and inhibitors, and eight hSULTs for enzymatic activity toward 31 potential substrates in order to better understand the relationship between binding specificity, activity, and structure within the hSULT family. These data, combined with detailed structural analysis of substrate binding sites, reveal relationships between family members not previously apparent from sequence analysis. “Chemical fingerprints” of the spectrum of small molecules that bind in the presence and absence of the cofactor product, 3′-phosphoadenosine 5′-phosphate (PAP), demonstrate a marked change in the small molecule binding profile upon PAP binding. This result, combined with the structural data, suggests PAPS has a strong influence on which compounds may bind in the substrate binding site and raises the possibility that the enzymes might be inhibited by chemically related compounds that are not productive substrates. The binding studies also provide insight into potential functions of the under-characterized SULT1C



**Figure 1.** Schematic of the Reaction Catalyzed by SULTs  
doi:10.1371/journal.pbio.0050097.g001

subfamily and of SULT4A1, an orphan member of the SULT family expressed primarily in the brain.

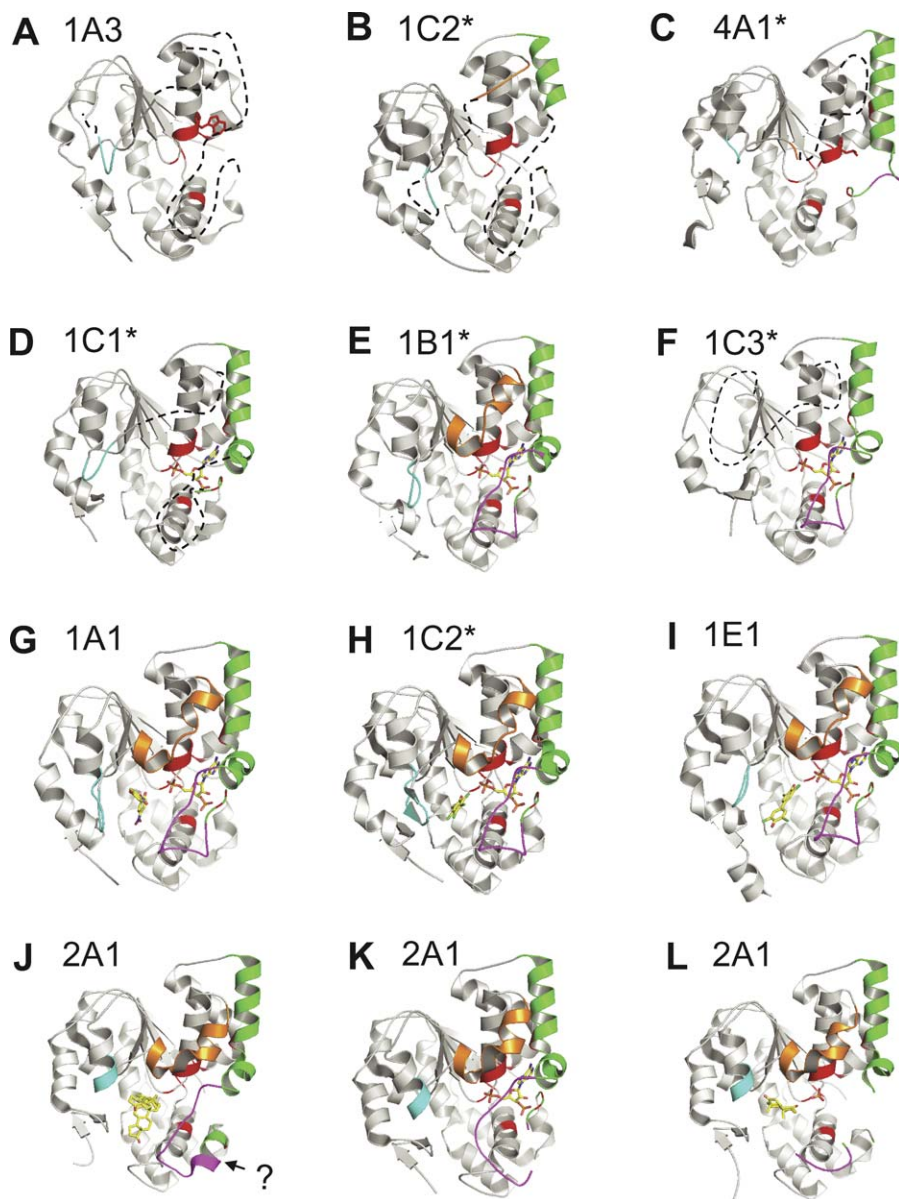
## Results

### Completion of the Structural Coverage of hSULTs

The crystal structures of SULT1C3 bound to PAP, apo SULT1C2, a ternary complex of SULT1C2 bound to PAP, and the environmental toxin, pentachlorophenol (PCP), and SULT4A1 were solved at 3.2, 2.0, 1.8, and 2.2 Å, respectively (Figure S1 and Table S1). We also recently reported the structures of SULT1B1 and SULT1C1 bound to PAP at 2.1 and 1.8 Å, respectively [17]. The structures of a single subunit of each of these normally dimeric proteins are presented in Figure 2 along with a representative structure of each of five other SULT family members previously reported in the literature [8,10–15]. Six additional SULT structures, which are available in the Protein Data Bank, are presented in Figure S2. As expected, all SULTs share the same basic fold: a central four-stranded parallel  $\beta$ -sheet surrounded by  $\alpha$ -helices and three loops that are often disordered (dashed lines) in the absence of PAP and/or substrate. These disordered segments comprise a 13-residue loop (shown in gold), a 4–10 residue loop (cyan), and a large 32–46 residue loop (green and magenta). These loops have been mapped onto the aligned protein sequences in Figure 3 using the same colouring scheme. The degree of disorder and the exact conformation of these loops vary considerably across the family, but in general, the presence of ligands (cofactor and/or substrate) is coupled with increased order, namely, the formation of helices  $\alpha$ 4– $\alpha$ 5 (gold), and  $\alpha$ 14– $\alpha$ 15 (green). In some cases, partial stabilization can be attributed to molecular packing in the crystal, as in, for example, the stabilization of  $\alpha$ 14 (green) in apo SULT1C2. The binding site for PAP or PAPS (PAP(S)) is nearly identical in all structures bound to these ligands, with highly conserved residues contributing to the binding pocket (highlighted in red in Figures 2 and 3). It is interesting to note that the SULT6B1 sequence in the protein databases lacks the N-terminal region, which encodes a  $\beta$ -sheet thought to be an important structural component of the SULT fold (Figure 3). We note that the recombinant SULT6B1 did not express in our attempts to purify it from bacteria.

### Structural Comparison Supports a Role for PAPS in Priming the Conformation of Substrate Binding Loops

It is generally agreed that sulfonation takes place via a sequential mechanism in which a ternary enzyme complex is first formed, followed by reaction and release of products [7]. However, both random and ordered binding of the substrate and cofactor molecules have been reported, and the detailed kinetic mechanism (or mechanisms) of the sulfonate transfer reaction is the subject of continuing research (reviewed in [7]). Comparison of all the available structural data provides insight into the order of substrate and cofactor binding. The structures provide evidence for both binary complexes (enzyme/substrate and enzyme/cofactor) consistent with a random bi-bi mechanism and ruling out an ordered mechanism in which binding of substrate requires binding of cofactor (or vice versa). This is in agreement with a detailed kinetic analysis for SULT1E1 [18]. However, a closer inspection of the structures also suggests that binding of



**Figure 2.** Representative Structures for Each of the hSULTs Grouped According to Similar Global Structural Features

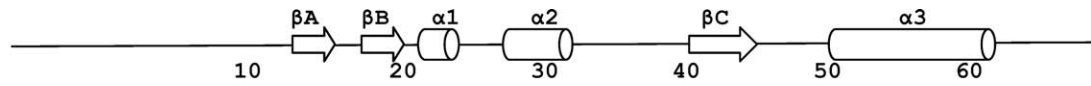
Apo proteins with considerable disorder in several loops: (A) SULT1A3 (1CJM), (B) SULT1C2 (2AD1), and (C) SULT4A1 (1ZD1). Structures with increased order due to binding of PAP (green helices, purple loop): (D) SULT1C1 (2ETG), (E) SULT1B1 (1XV1), and (F) SULT1C3 (2H8K). Ternary complexes display further ordering of substrate binding loops: (G) SULT1A1 with PAP and *p*-nitrophenol (1LS6), (H) SULT1C2 with PAP and PCP (2GWH), and (I) 1E1 with PAP and 3,5,3',5'-tetrachloro-biphenyl-4,4'-diol (1G3M). Structures with unusual features that likely reflect catalytically unproductive proteins: (J) SULT2A1 bound to DHEA, but without PAP (1J99); compare to (K) structure of the same protein with PAP (1EFH), and (L) with androsterone (1OV4). The structures we solved are labelled with an asterisk. The question mark (?) indicates a helix formation, which leads to a non-productive conformation. The Protein Data Bank code for each structure is shown in parentheses. The proteins are represented by a ribbon model; PAP and substrate are shown as stick models coloured per element (carbon in yellow, oxygen in red, nitrogen in blue, and phosphate in magenta). The loops are coloured as discussed in the text (gold, cyan, green, and purple).

doi:10.1371/journal.pbio.0050097.g002

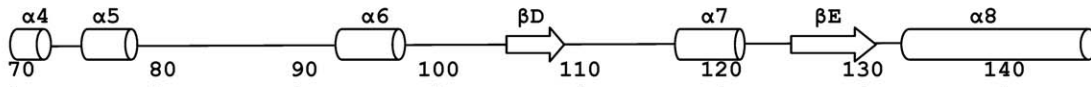
substrates may not be completely uncoupled from binding of the cofactor. In all the structures with the co-factor product, PAP,  $\alpha$ 14- $\alpha$ 15, and the C-terminal segment of the largest flexible loop (green in Figure 2) are ordered. This region contributes three absolutely conserved residues necessary for PAPS binding, T228, R258, and G260 (SULT1A1 numbering and red in Figure 2). Importantly, although the other loops (cyan, gold, and magenta) do not contribute directly to PAPS binding, they are more likely to be partially ordered in the presence of PAP(S). The PAP(S)-induced ordering of  $\alpha$ 14- $\alpha$ 15

and residues 256–262 (green and red) may also restrict the conformations available to the intervening substrate-binding magenta loop when PAP is bound. Thus, the structural data suggest that PAPS binding tends to prime the cyan, gold, and magenta loops for binding to the substrate.

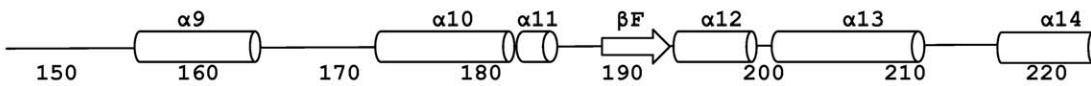
On the other hand, the structure of SULT2A1 bound to androsterone [13] (Figure 2L) hints that binding of substrates does not prime the PAPS binding loops. In this structure, the substrate-binding cyan and gold loops are ordered, but the magenta loop and adjacent PAPS binding residues (green and



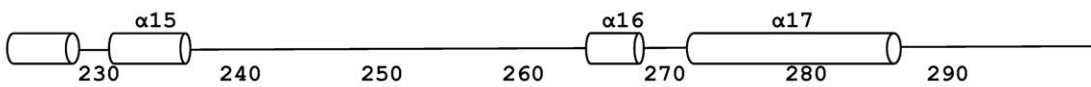
|         |   |    |
|---------|---|----|
| SULT1A1 | MELIQDTSRPP-LEYVKGVPLIKYFAEALGPLQSFQARPDDLLISTYPKSGTTWVSQILDMIYQGGDLEK          | 69 |
| SULT1A2 | MELIQDISRPP-LEYVKGVPLIKYFAEALGPLQSFQARPDDLLISTYPKSGTTWVSQILDMIYQGGDLEK          | 69 |
| SULT1A3 | MELIQDTSRPP-LEYVKGVPLIKYFAEALGPLQSFQARPDDLLINTYPKSGTTWVSQILDMIYQGGDLEK          | 69 |
| SULT1B1 | MLSPKDI LRKD-LKLVHGYPMTCAFASNWEKIEQFHSRPDDIVIATYPKSGTTWVSEIIDMILNDGDI EK        | 69 |
| SULT1C1 | MALTSDLGKQIK-LKEVEGTLLOPATVDNWSQIQSFEAKPDDLLICTYPKAGTTWVQIIVDMIEQNGDVEK         | 70 |
| SULT1C2 | MALHEMEDFTFDGTRKLS-VNYVKGILQPTDTCDIWDKIWNFQAKPDDLLISTYPKAGTTWVQIIVDMIEQNGDVEK   | 76 |
| SULT1C3 | MAKIEKNAPTMEKKPELFNIMEVDGVP TLLILSKEWWEKVCNFQAKPDDLLILATYPKSGTTWVHEIIDMILNDGVEK | 77 |
| SULT1E1 | MNSELDYYEK--FEEVHGILMYKDFVKYWDNVEAFQARPDDLV IATYPKSGTTWVSEIVYMIYKEGDVEK         | 68 |
| SULT2A1 | MSDDFLWFGEGIAFPTMGFRSETLRKVRDEFVIRDEDV IILTYPKSGTNLAEILCLMHSKGD AKW             | 65 |
| SULT2B1 | MASPPPFHS-QKLPGEYFRYKGVPPFVGLYSLESI SLAENTQDVRDDDFIITYPKSGTTWVHEIICLILKEGDPSW   | 76 |
| SULT4A1 | MAESEAETPSTPGEFESK-YFEFHGVRLLPFCRGKME EIANFVPRPSDVWIVTYPKSGTSLIQEVVYLV SQGADPDE | 76 |
| SULT6B1 | MCTSETFQALDTFEARHDDIVLASYPKCGSNWILHIVSELIYAVSKKK                                | 48 |



|         |  |     |
|---------|--|-----|
| SULT1A1 | CHRAPIFMRVPFLEFKAPGIP-SG METLKDTPAPRLLKTHLPLALLPQTLLDQKVKV VVVAANAKDVAVSYHFYHM | 145 |
| SULT1A2 | CHRAPIFMRVPFLEFKVPGIP-SG METLKNTPAPRLLKTHLPLALLPQTLLDQKVKV VVVAANAKDVAVSYHFYHM | 145 |
| SULT1A3 | CNRAPIYVRVFPVNDPGE P-SGLETLKDTPPRLIKSHLPLALLPQTLLDQKVKV VVVAANPKDVAVSYHFHRM    | 145 |
| SULT1B1 | CKRGFITEKVPMLEMTLPGLR TSGIEQLEKNPSPRIVKTHLPTDLLPKSFWENNCKMIYLA NAKDVSVSYHFDLM  | 146 |
| SULT1C1 | CQRAIQHRHPFIEWARPPQP-SG VEKAKAMSPRILKTHLSTQLLPPSFWENNCKFLYVA NAKDCMVSYHFQRM    | 146 |
| SULT1C2 | SKRAPTHQRFPFLEMKIPSLG-SGLEQAHAMSPRILKTHL PPHLLPPSLEKNCKIIYVAANPKDNMVS YHFQRM   | 152 |
| SULT1C3 | CKRAQTLDRHAFLELKFPHKEK PDLFVLEMSSPQLIKTHLPSHLIPPSIWKENCKIVYVA NPKDCLVSYHFHRM   | 154 |
| SULT1E1 | CKEDVIFNRIPFLECRKENLMN-GVKQLDEMNSPRIVKTHL PPELLPASFWEKDKCIIYLCNNAKDVAVSEYFFFLM | 144 |
| SULT2A1 | IQSVPIWERSPWVESE-----IGYALSETESPR LFSHLPQLFPKSFSSKAKVIYLMNPRDVLVGYFFWKN        | 136 |
| SULT2B1 | IRSVPIWERAPWCE TI-----VGA FSLPDQYSPRLMSHLPQLFPTKAFSSKAKVIYMG NPRDVVLYHYSKI     | 147 |
| SULT4A1 | IGLMNIDEQLPVLEYPQPG-----LDI IKELTSPRLIKSHLPHYRFLPSDLHNGDSKVIYMA NPKDLVSYQFHRS  | 148 |
| SULT6B1 | YK----YPEFPVLECGDSEKY-----QRMKGFSPRILATHLHYDKLPGSIFENKAKILVIFNPKDTAVSEFLHFHND  | 117 |



|         |   |     |
|---------|---|-----|
| SULT1A1 | AKVHPEPGTWD SFLEKFMVGEVSYGSWYQHVQEWELSR THPVLVLYLFYEDMKENPKREIQKILEFVGRSLPEETVDF  | 223 |
| SULT1A2 | AKVYPHPGTWESFLEKFMAGEVSYGSWYQHVQEWELSR THPVLVLYLFYEDMKENPKREIQKILEFVGRSLPEETVDL   | 223 |
| SULT1A3 | EKAHPEPGTWD SFLEKFMAGEVSYGSWYQHVQEWELSR THPVLVLYLFYEDMKENPKREIQKILEFVGRSLPEETMDF  | 223 |
| SULT1B1 | NNLQPPFGTWEEYLEKFLTGKVAYGSWFTHVKNWKKKEHPI LFLYEDMKENPKKEIKKIRFLEKNLND EILDR       | 223 |
| SULT1C1 | NHMLDPFGTWEEYFETFINGKVVWGSWFDHVKGWWEKDRHQ ILLFLYEDIKRDPKHEIRKVMKGVDETVL DK        | 224 |
| SULT1C2 | NKALPAGPTWEEYFETFLAGKVCWGSWFHDHVKGWWEAKDHR ILLYLFYEDMKENPKHEIQKLAEIFGKLLDDKVL DK  | 229 |
| SULT1C3 | ASFMPDPQNLEEFYEKFMGKVVGGSWFDHVKGWWA AKDMHRILYLFYEDIKDPKREIEKILKFL EKDISEEILNK     | 231 |
| SULT1E1 | VAGHPNPGSFPEFVEKFMQGQVPYGSWKVHKVSWWEKGS PRVLFYEDLKEDIRKEVIKLIHFLEKRPSELVDR        | 221 |
| SULT2A1 | MKFIKPKSWEEYFEWFQCQGTVLVYGSWFDH IHGWMMPREKFNLLLSYEELKQDTGRTIEKICQFLGKTLEPEELNL    | 213 |
| SULT2B1 | AGQLKDPGTPDQFLRDLFKGEVQFGSWFDH IKGWLRMKGDNF LFITYEELQDDLQGSVERICGFLGRPLGKEALGS    | 224 |
| SULT4A1 | LRTMSYRGTFQEFCCR FEMNDKLGYSWFVEHVQEFWEHRMDS NVLFLKYEDMHRDLVTMVEQLARFLGVSCDKAQL EA | 225 |
| SULT6B1 | VDPIDPSYGSWDEFRRQFMKGQVSWGRYFDFAINWNKHL DGDNVKFI LYEDLKENLAAGIKQIAEFLGFFLTGEQIQ T | 194 |



|         |   |     |
|---------|---|-----|
| SULT1A1 | MVQHTSFKEMKKNPMTNYTTVPQEFMDHSISP FMRKGMAGDWKTTFTVAQNERFDADYAKMAGCS-LTFRSEL    | 296 |
| SULT1A2 | MVEHTSFKEMKKNPMTNYTTVRRFMDHSISP FMRKGMAGDWKTTFTVAQNERFDADYAEKMAGCS-LSFRSEL    | 296 |
| SULT1A3 | MVQHTSFKEMKKNPMTNYTTVPQELMDHSISP FMRKGMAGDWKTTFTVAQNERFDADYAEKMAGCS-LSFRSEL   | 296 |
| SULT1B1 | IIHHTSFEVMDKNPLVNYTHLPTVMDHSKSP FMRKGTAGDWKNYFTVAQNEKFD AIYETEMSKTA-LQFRTEI   | 296 |
| SULT1C1 | IVQHTSFEKMKENPMTNRSTVSKSILDQSISS FMRKGTVDWKNHFTVAQNERFDEIYRRKMEGTS-INF CMEL   | 297 |
| SULT1C2 | IVHYTSFDVMKQNP MANYSSIPAEIMDHSISP FMRKGA VGDWKKHFTVAQNERFDEDYKKKMTDTR-LTFHFQF | 302 |
| SULT1C3 | IIYHTSFDVMKQNPMTNYTLLPTSIMDHSISP FMRKGM PGDWKNYFTVAQNEEFDKDYQKMGAST-LTFRTEI   | 304 |
| SULT1E1 | IIHHTSFEQEMKNPSTNYTLLPDEIMNQKLS PFMKGIITGDWKNHFTVALNEKFDKHYEQQKMEST-LKFRTEI   | 294 |
| SULT2A1 | ILKNSFQSMKENKMSNYSLLSVDYVVDK-AQL LRKGVSGDWKNHFTVAQAEFDKLFQEK MADLPRELFPWE     | 285 |
| SULT2B1 | VVAHSTFSAMKANTMSNYTLLPPSLLDHRRGAF LKRGVCGDWKNHFTVAQSEAFDRAYRQMRGMP--TFPWEDPE  | 229 |
| SULT4A1 | LTEHCHQLVDQCCNA-----EALPVG R-GRVGLWFDIFTVSMNEKFDLVYKQKMGKCD-LTFDFYL           | 284 |
| SULT6B1 | ISVQSTFQAMRAKSQDTHGAVG-----PFLFRKGEVGDWKNL FSEIQNQEMDEKFKCELAGTSLGAKLKYESYC   | 263 |

**Figure 3.** Sequence Alignment of hSULTs Showing Structural Features with Colour Coding Corresponding to That of the Structures in Figure 2  
doi:10.1371/journal.pbio.0050097.g003

red portion of the loop) are disordered. Thus, although in this case, substrate and PAP(S) molecules can each bind independently to the enzyme as in a random bi-bi mechanism, there may be some degree of cooperativity between substrate binding and prior cofactor binding, but not vice versa. Given that the estimated cellular concentration of PAPS is well above that of most substrates, this may be relevant to the catalytic mechanism.

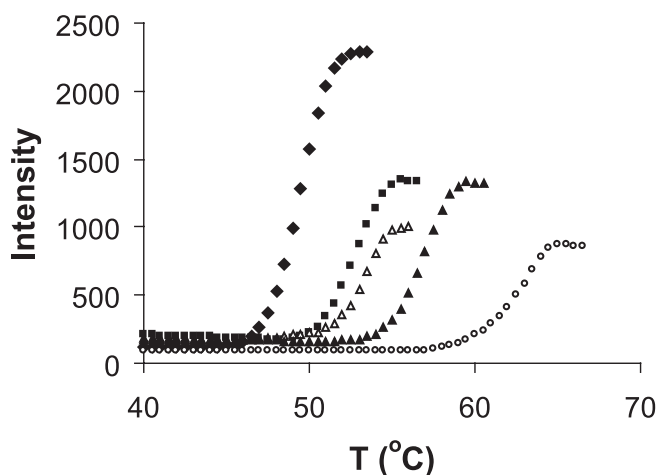
The family-wide structural comparison also suggests an additional or alternative explanation for the well-documented substrate inhibitory effect. Previously reported cases of substrate inhibition have been attributed either to two substrate molecules occupying the active site at the same time [8,19] or to the ability of substrates to bind in unproductive orientations at higher concentrations [6,14]. Examination of the structures in Figure 2 suggests a third or alternative mechanism; at high concentrations, substrates may bind in a mode in which the binding loops are incompatible with PAPS binding. This case is exemplified by the structure of SULT2A1 with DHEA [14]. As shown in Figure 2J, this structure can accommodate two substrate orientations at roughly 30° to one another. Comparison with other hSULT structures strongly suggests that SULT2A1 in this structure adopts a non-productive conformation. A portion of the green-and-magenta loop that contributes two residues for PAP(S) binding is folded into a helix, orienting the crucial PAPS binding residues away from the cofactor binding pocket. This helix conformation is not an intrinsic feature of SULT2A1, because in the SULT2A1-PAP complex, this region adopts a conformation similar to that in other SULT-PAP structures (compare Figure 2K and 2J with this region highlighted by a question mark). Thus, it appears that the structure adopted by SULT2A1 with two molecules of DHEA is incompatible with PAP(S) binding and that this conformation is induced by the substrate. This is further evidence of “communication” between the substrate binding site and the PAPS binding site.

#### Ligand Binding and Activity Profiles Reveal Enzyme-Specific Chemical Fingerprints

In order to predict and understand the fate of xenobiotics and drug candidates in humans, it is essential to better understand the selectivity and specificity of binding and activity within the hSULT family. Although detailed analyses of individual structures have been very informative in this regard [6,8–12,16], we sought to compare all active sites relative to the spectrum of small molecules that can bind to each site. However, several of the proteins whose structures were solved in this study have not been previously characterized, and it was difficult to directly compare data from the literature due to differences in experimental conditions. Therefore, in order to evaluate specificity and selectivity in a consistent manner, nine purified, recombinant hSULTs were screened for binding to a library of 90 small molecules (Table S2) that comprised known substrates, inhibitors, related hormones, bioamines, and drugs [20,21]. In order to profile the entire hSULT family, we made use of the well-known fact that equilibrium binding of a ligand increases the thermal stability of a protein in a manner proportional to the concentration and binding affinity of the ligand [20,21]. In a

multi-well format, the thermal stability of each hSULT was monitored as a function of temperature and in the presence or absence of compounds (Figure 4). In the absence of compounds, well-behaved, sigmoidal thermo-denaturation/aggregation profiles were obtained for hSULTs 1A1, 1A3, 1B1, 1E1, 1C1, 1C2, 1C3, 2A1, and 4A1. SULT2B1b did not denature within the range of temperature used for this type of analysis (up to 80 °C). In this screening format, compounds that stabilize a protein by more than 2° C are scored as positives (Table 1). It was not possible to assay binding to ligands in the presence of the sulfonated co-factor, PAPS, because the sulfonate transfer reaction would have taken place. However, except for SULT4A1, PAPS and PAP had equivalent stabilizing effects on all hSULTs. Thus, PAP was used as a substitute for PAPS in considering the effect of cofactor upon substrate binding, and screens for the binding of ligands were performed in the absence and presence of a saturating amount of PAP.

Based on these binding results, a set of 20 compounds that bound to at least one hSULT plus 11 additional related compounds or known substrates were used as a pool of potential substrates for enzymatic activity of hSULTs 1A1, 1A3, 1B1, 1C1, 1C2, 1C3, 1E1, and 2A1 (Table 2). We monitored the conversion of PAPS to PAP by high-performance liquid chromatography (HPLC) as a tractable method of screening multiple proteins against multiple substrates (eight proteins and 31 substrates in this study), and the results serve as a convenient first approximation of enzymatic activities. We note that due to the relatively low sensitivity of this method, we were not able to reliably assay substrates at nanomolar concentrations, and therefore, some of the results may be complicated by substrate-mediated inhibition. Indeed, inspection of Table 2 shows that in some cases, the highest levels of activity were observed at lower substrate



**Figure 4.** Thermo-denaturation/Aggregation Profile of SULT2A1

The intensity of scattered light is plotted as a function of temperature ( $T$ ) for SULT2A1 in the absence of any ligand (filled diamond), and in the presence of AMP-PNP (filled square), lithocholic acid (open triangle), PAP (filled triangle), and both PAP and dehydroisoandrosterone 3-sulfate (open circle). The inflection point of the denaturation curve is taken as  $T_{agg}$  as reported in Table 1.

doi:10.1371/journal.pbio.0050097.g004

Table 1. Thermostability Screen for Ligand Binding to hSULTs

| Category                               | Compound                             | 1C1   |       | 1C2   |       | 1C3   |       | 4A1   |       | 1B1   |       | 1A1   |       | 1A3   |       | 2A1   |       | 1 E1  |       |
|--|--------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|  |                                      | - PAP | + PAP | - PAP | + PAP | - PAP | + PAP | - PAP | + PAP | - PAP | + PAP | - PAP | + PAP | - PAP | + PAP | - PAP | + PAP | - PAP | + PAP |
| <b>T<sub>agg</sub> of protein (°C)</b> |                                      | 48.4  | 57    | 44.5  | 52.5  | 39.2  | 48.3  | 63.4  | 62.9  | 47.8  | 58.2  | 49.3  | 59    | 47.1  | 55.6  | 48.9  | 59.4  | 46.8  | 56.8  |
| <b>Inhibitors</b>                      | PAP                                  | 4.5   | NA    | 6.4   | NA    | 6.5   | NA    | 0     | NA    | 7.8   | NA    | 6.7   | NA    | 6.6   | NA    | 7.4   | NA    | 6.4   | NA    |
|  | DBHD                                 | 10.8  | 2.3   | 0     | 0     | 8     | 0     | 6     | 8.3   | 11.4  | 0     | 5.4   | 4.1   | 3.5   | 2.5   | 8.4   | 3     | 3.9   | 0     |
|  | DBHM                                 | 9.3   | 0     | 0     | 0     | 2.8   | 0     | 3     | 0     | 7.4   | 0     | 3.6   | 5.6   | 0     | 2.2   | 4.9   | 2.9   | 2.4   | 0     |
|  | Pyridoxal 5-Phosphate                | 8.7   | 0     | 4.5   | 0     | 5.7   | 0     | 6     | 6.3   | 5.4   | 0     | 4.3   | 0     | 6.6   | 0     | 9     | 0     | 7.9   | 0     |
|  | AMP-PNP                              | 4.5   | 0     | 5.2   | 0     | 6.4   | 0     | 2.8   | 4     | 4.5   | 0     | 4.4   | 0     | 5.6   | 0     | 3.4   | 0     | 6.1   | 0     |
|  | Quercetin dihydrate                  | 31    | 23    | 4.6   | 0     | 13.5  | 2.3   | 18    | 18    | 8.4   | 0     | 6.5   | 0     | 33    | 0     | 9.1   | 20    | 0     | 3.9   |
| <b>Catecholamines</b>                  | Dopamine                             | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 2.3   | 0     | 0     | 0     | 0     |
|  | Epinephrine                          | 0     | 0     | 0     | 0     | 6.9   | 0     | 16    | 5.9   | 4.6   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
|  | Norepinephrine                       | 0     | 0     | 0     | 0     | 5.1   | 0     | 2     | 7     | 4.5   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
|  | Isoprenaline                         | 0     | 0     | 0     | 0     | 6.4   | 0     | 18    | 18    | 5     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| <b>Steroid-/large/fused ring</b>       | 2-Hydroxyestradiol                   | 0     | 0     | 0     | 0     | 2     | 0     | 9     | 18    | 0     | 0     | 0     | 0     | 0     | 0     | 3     | 0     | 3.1   | 0     |
|  | T4 (3,3',5,5'-tetraiodo-L-thyronine) | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 2.3   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 3.5   | 5.2   | 0     |
|  | 17 $\alpha$ -Ethinylestradiol        | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 3.4   | 0     | 0     | 0     | 4.2   | 0     | 4.3   |
|  | Lithocholic acid                     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 4.4   | 0     | 2.4   | 0     |
|  | Dehydroisandrosterone 3-sulfate      | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 2.8   | 3.1   | 2.2   | 0     |
|  | $\alpha$ -Zearalenol                 | 0     | 0     | 0     | 0     | 0     | 2.8   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 3.5   | 0     | 2.6   |
|  | Adenosyl cobalamin                   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 2     | 0     | 0     |
|  | Estrone                              | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 2.2   |
| <b>Phenols</b>                         | Apomorphine                          | 0     | 0     | 0     | 3.2   | 17.6  | 0     | 18    | 18    | 12.6  | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
|  | <i>p</i> -Cresol                     | 0     | 0     | 0     | 0     | 7.2   | 0     | 7     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
|  | 4-Aminophenol                        | 0     | 0     | 0     | 0     | 8.3   | 4     | 0     | 0     | 3.3   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
|  | 1-Naphthol                           | 0     | 0     | 0     | 2.9   | 0     | 2.2   | 0     | 0     | 0     | 2.5   | 0     | 4.5   | 0     | 0     | 0     | 0     | 0     | 0     |
|  | 2-Naphthol                           | 0     | 0     | 0     | 4.1   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 5     | 0     | 0     | 0     | 0     | 0     |       |
|  | 2-Ethylphenol                        | 0     | 0     | 0     | 0     | 0     | 2.1   | 0     | 0     | 0     | 0     | 0     | 3     | 0     | 0     | 0     | 0     | 0     |       |
|  | 2-n-Propylphenol                     | 0     | 0     | 0     | 0     | 0     | 2.6   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |       |
|  | 2-sec-Butylphenol                    | 0     | 0     | 0     | 2     | 0     | 2.2   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |       |
|  | Vanillin                             | 0     | 0     | 0     | 2.9   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 4.7   | 0     | 0     | 0     | 0     | 0     |       |
|  | Pentachlorophenol                    | 0     | 0     | 0     | 5.1   | 0     | 0     | 0     | 0     | 3.7   | 8.9   | 0     | 10.2  | 0     | 0     | 0     | 0     | 0     |       |
|  | 2, 6-Dichloro-4-nitrophenol          | 0     | 0     | 0     | 3.5   | 0     | 0     | 0     | 0     | 3     | 0     | 0     | 8.7   | 0     | 0     | 0     | 0     | 2.8   |       |
| <b>Acids</b>                           | Resveratrol                          | 6.6   | 0     | 0     | 0     | 7.4   | 0     | 18    | 18    | 8.6   | 0     | 0     | 0     | 0     | 0     | 3.8   | 0     | 4.5   | 0.0   |
|  | Salicylic acid                       | 0     | 0     | 0     | 2.2   | 0     | 0     | 0     | 0     | 2.4   | 0     | 0     | 2.5   | 0     | 0     | 0     | 0     | 0     | 0     |
|  | Mefenamic acid                       | 0     | 0     | 0     | 3.6   | 0     | 0     | 0     | 0     | 2.9   | 0     | 0     | 9.6   | 0     | 0     | 0     | 0     | 0     | 4.7   |

The thermostability of each protein (0.4 mg/ml) in the presence and absence of potential ligands (1 mM) was measured as described in Material and Methods. Effect of PAP was also assessed at 1 mM concentration (second row). Except for the numbers in the first row, all the values in this table are  $\Delta T_{agg}$  of the protein in the presence of the compound minus  $T_{agg}$  in the absence of the compound [21].  $\Delta T_{agg}$  value below 2 °C is considered within the error limit of the technique and is shown as zero (0). The values in the first row are for  $T_{agg}$  of each protein in the presence or absence of 10 mM PAP. Over a concentration range of 0.1 to 5 mM, both PAPS and PAP had equivalent stabilizing effects on all SULTs except SULT4A1. Thus, PAP was used as a substitute for PAPS in considering the effect of cofactor upon substrate binding. The  $\Delta T_{agg}$  values that are less than 4 °C, between 4 °C and 10 °C, and above 10 °C are highlighted by green, blue, and red, respectively.

doi:10.1371/journal.pbio.0050097.t001

**Table 2.** Sulfotransferase Activity Screen of hSULTs

| Category                         | Substrate                            | Sulfotransferase Activity (nmol/min/mg) |     |     |           |      |      |           |     |     |
|----------------------------------|--------------------------------------|---|-----|-----|-----------|------|------|-----------|-----|-----|
|                                  |                                      | SULT1C1                                 |     |     | SULT1C2   |      |      | SULT1C3   |     |     |
| Substrate concentration (μM)     |                                      | 10                                      | 25  | 100 | 10        | 25   | 100  | 10        | 25  | 100 |
| <b>Catecholamines</b>            | Dopamine                             | N                                       | N   | N   | N         | N    | N    | N         | N   | N   |
|                                  | Epinephrine                          | N                                       | 0.1 | 0.1 | N         | N    | 0.8  | N         | N   | N   |
|                                  | Norepinephrine                       | N                                       | 0.1 | 0.1 | N         | N    | N    | N         | N   | N   |
|                                  | Isoprenaline                         | N                                       | N   | N   | N         | N    | 0.8  | N         | N   | N   |
| <b>Steroid-/large/fused ring</b> | 17α-Ethinylestradiol                 | N                                       | N   | N   | N         | N    | N    | N         | N   | N   |
|                                  | 2-Hydroxyestradiol                   | N                                       | N   | 0.1 | N         | N    | N    | N         | N   | N   |
|                                  | Lithocholic acid                     | N                                       | N   | N   | N         | N    | N    | 0.6       | 0.8 | 1.1 |
|                                  | α-Zearalenol                         | 0.3                                     | 0.6 | 1.5 | 2.0       | 8.1  | 9.3  | 1.1       | 2.7 | 4.1 |
|                                  | T2 (3,5-diiodo-L-thyronine)          | N                                       | N   | 0.1 | N         | N    | N    | N         | N   | N   |
|                                  | T3 (3,3',5-triiodo-L-thyronine)      | 0.4                                     | 0.6 | 1.1 | N         | N    | 2.9  | N         | N   | N   |
|                                  | T4 (3,3',5,5'-tetraiodo-L-thyronine) | 0.3                                     | 0.5 | 0.8 | N         | N    | N    | N         | N   | N   |
|                                  | Cholesterol                          | N                                       | N   | N   | N         | N    | 0.7  | N         | N   | 0.4 |
|                                  | Estrone                              | 0.1                                     | N   | N   | N         | N    | 0.8  | N         | N   | N   |
|                                  | Minoxidil                            | N                                       | N   | N   | N         | N    | N    | N         | N   | N   |
|                                  | Salbutamol                           | N                                       | N   | N   | N         | N    | N    | N         | N   | N   |
| <b>Phenols</b>                   | 4-Aminophenol                        | N                                       | N   | N   | 1.3       | 1.2  | 2.6  | N         | N   | N   |
|                                  | 1-Naphthol                           | 0.1                                     | 0.1 | 0.6 | 4.6       | 11.7 | 31.5 | 0.7       | 1.2 | 1.8 |
|                                  | 2-Naphthol                           | 0.2                                     | 0.3 | 1.1 | 3.3       | 16.2 | 11.7 | N         | N   | N   |
|                                  | 2-Ethylphenol                        | N                                       | N   | N   | 1.4       | 4.8  | 25.8 | 0.9       | 1.4 | 2.2 |
|                                  | 4-Ethylphenol                        | N                                       | N   | 0.1 | 1.8       | 9.9  | 39.5 | N         | N   | N   |
|                                  | 2-n-Propylphenol                     | N                                       | N   | 0.1 | 0.8       | 3.4  | 15.5 | N         | 0.7 | 0.9 |
|                                  | 2-sec-Butylphenol                    | 0.3                                     | N   | N   | 1.2       | N    | N    | N         | 1.0 | 2.1 |
|                                  | 4-n-Amylphenol                       | 0.1                                     | 0.1 | 0.3 | N         | 6.4  | 17.2 | N         | N   | 0.3 |
|                                  | 4-n-Heptylphenol                     | N                                       | N   | N   | N         | N    | N    | N         | N   | N   |
|                                  | 4-Nitrophenol                        | 0.3                                     | 0.5 | 1.9 | N         | 8.5  | 33.4 | N         | N   | 0.6 |
|                                  | Vanillin                             | 0.3                                     | 0.8 | 2.5 | 5.2       | 15.8 | 23.0 | N         | N   | 0.8 |
|                                  | 1,2,3,4-Tetrahydro-1-Naphthol        | N                                       | N   | N   | N         | N    | N    | N         | 0.3 | 0.8 |
|                                  | p-Cresol                             | N                                       | N   | 0.1 | 4.9       | 17.7 | 40.7 | N         | N   | N   |
|                                  | Acetaminophen (Paracetamol)          | N                                       | N   | 0.1 | 4.8       | 8.6  | 22.6 | N         | N   | N   |
| Tyramine                         | N                                    | N                                       | N   | 3.1 | 9.0       | 38.2 | N    | N         | N   |     |
| Resveratrol                      | 0.3                                  | 0.5                                     | 0.3 | 3.4 | 15.0      | 5.2  | N    | N         | N   |     |
| <b>Background</b>                |                                      | 0.25 ± 0.01                             |     |     | 6.2 ± 0.6 |      |      | 0.9 ± 0.2 |     |     |

Enzymatic activities of the eight hSULTs were assayed at concentrations of 10, 25, and 100 μM of each substrate as described in Material and Methods. The first row indicates the concentration of each substrate in μM in each set of experiments. The last row indicates the background rate of conversion of PAPS to PAP in the presence of enzyme, but absence of any substrate. The error in the background rate is the standard deviation of three measurements. All other numbers in the table are the activity in nmol/min/mg of protein after subtracting the background rate. For a given enzyme/substrate pair, the substrate concentration that exhibited the highest activity is highlighted in blue. Activity values between one and two standard deviations of the background rate are highlighted in red. Those above two standard deviations are in green.

N, not detected.

doi:10.1371/journal.pbio.0050097.t002

concentrations. This is especially true of SULT1A1, an enzyme for which significant substrate inhibition has been noted previously [8].

The combined ligand binding and activity screens revealed a unique “chemical fingerprint” for each hSULT (Tables 1 and 2). First, as expected from previous studies, there was considerable overlap in the substrate specificity for enzymatic activity. For example, all hSULTs assayed here were able to sulfonate a number of phenolic compounds such as naphthols and/or alkylphenols. However, within each substrate profile, there were also elements of specificity. For example, SULT1A1 and SULT1A3 were the only two hSULTs that showed significant activity toward catecholamines compared to other substrates, with SULT1A3 being more specific for dopamine, as expected from previous studies [22–26].

SULT1A3 was also the only protein to bind dopamine in the binding assays, consistent with its designation as human dopamine sulfotransferase [25]. It is interesting to note that in the past, SULT1A1 and 1A3 have been distinguished from one another in tissue fractions by the higher sensitivity of SULT1A1 to inhibition by 2,6-dichloro-4-nitrophenol (DCNP) [27]. Although we did not measure inhibition by this compound, we note that SULT1A1 bound DCNP in the presence of PAP, whereas SULT1A3 did not (Table 1).

Six hSULTs (1C1, 1C2, 1E1, 1B1, 1A1, and 1A3) had enzymatic activity toward resveratrol, a polyphenolic compound present in grapes and wine, with possible anticancer and cardioprotective activities [28]. The activity profiles for resveratrol also displayed evidence of substrate inhibition by this compound for SULTs 1C1, 1C2, 1A1, and 1E1. Acet-

Table 2. Extended.

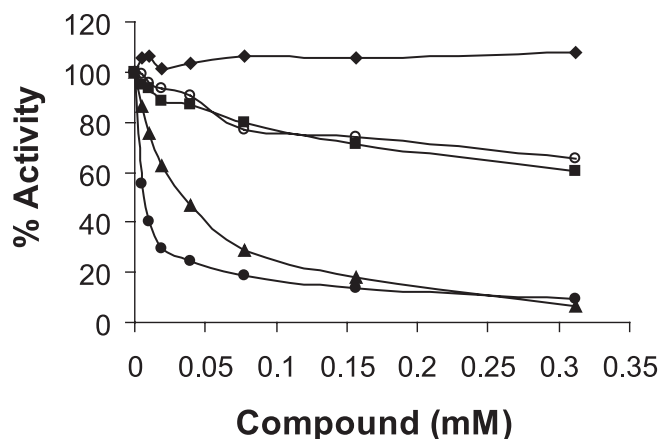
| Sulfotransferase Activity (nmol/min/mg) |     |      |           |      |      |           |      |      |           |     |      |           |      |      |
|---|-----|------|-----------|------|------|-----------|------|------|-----------|-----|------|-----------|------|------|
| SULT1B1                                 |     |      | SULT1A1   |      |      | SULT1A3   |      |      | SULT2A1   |     |      | SULT1E1   |      |      |
| 10                                      | 25  | 100  | 10        | 25   | 100  | 10        | 25   | 100  | 10        | 25  | 100  | 10        | 25   | 100  |
| N                                       | N   | 0.6  | 2.3       | 5.2  | 13.4 | 7.4       | 25.7 | 74.4 | N         | N   | N    | N         | N    | 0.7  |
| N                                       | N   | 1.2  | 1.8       | 2.9  | 8.0  | N         | 0.5  | 6.0  | N         | N   | N    | N         | N    | N    |
| N                                       | N   | 0.5  | 1.7       | 1.7  | 4.2  | 0.7       | 2.7  | 16.9 | N         | N   | N    | N         | N    | 0.7  |
| N                                       | 0.6 | 1.1  | 2.7       | 6.1  | 13.6 | 5.1       | 16.6 | 54.0 | N         | N   | N    | N         | N    | 0.8  |
| N                                       | N   | N    | 2.5       | 7.2  | N    | N         | N    | 1.4  | 1.4       | 1.6 | 1.3  | 1.2       | 4.3  | 13.3 |
| N                                       | N   | 0.5  | 1.0       | 1.8  | 3.1  | N         | N    | N    | 1.0       | 1.5 | 3.1  | N         | N    | N    |
| N                                       | N   | N    | N         | N    | N    | N         | N    | N    | 2.7       | 8.3 | 11.4 | N         | N    | N    |
| 0.8                                     | 1.5 | 1.1  | 6.3       | 16.9 | 25.9 | 0.8       | 2.9  | 5.5  | N         | N   | N    | 2.5       | 7.1  | 13.0 |
| N                                       | N   | N    | 1.0       | N    | N    | N         | N    | N    | 0.3       | 0.6 | 1.2  | N         | N    | N    |
| N                                       | N   | 0.9  | 1.1       | 1.2  | N    | N         | 0.5  | 1.5  | N         | N   | 0.3  | 2.0       | 2.2  | 4.4  |
| N                                       | N   | 2.5  | N         | N    | N    | N         | N    | 0.6  | N         | N   | 1.6  | N         | 1.2  | 2.1  |
| N                                       | N   | N    | 0.8       | N    | N    | N         | N    | N    | N         | N   | N    | N         | N    | N    |
| N                                       | N   | N    | 5.2       | 4.9  | 5.0  | N         | N    | N    | 0.3       | 1.1 | 1.1  | N         | 0.8  | 3.4  |
| N                                       | N   | 1.6  | 1.6       | N    | N    | N         | N    | N    | N         | N   | N    | N         | 0.9  | 1.9  |
| N                                       | N   | N    | 0.9       | 1.3  | N    | 1.2       | 2.7  | 12.7 | N         | 0.3 | N    | N         | N    | N    |
| N                                       | 0.7 | 1.7  | 9.2       | 16.4 | 4.7  | N         | N    | N    | N         | N   | N    | N         | N    | 1.2  |
| 2.4                                     | 5.9 | 15.3 | 8.8       | 6.7  | 4.1  | 9.5       | 24.7 | 68.5 | N         | 0.4 | 1.0  | 5.0       | 9.6  | 26.0 |
| 2.9                                     | 8.9 | 16.8 | 11.1      | 16.8 | 10.0 | 10.0      | 20.3 | 42.6 | N         | 0.2 | 0.8  | 16.6      | 41.5 | 57.9 |
| 2.1                                     | 5.5 | 16.5 | 10.3      | 10.4 | 4.9  | 1.2       | 4.7  | 15.2 | N         | N   | 0.2  | 2.5       | 5.3  | 15.7 |
| 1.7                                     | 4.0 | 9.9  | 4.8       | 63.0 | 25.3 | 0.5       | 1.3  | 4.3  | N         | N   | N    | 9.4       | 21.9 | 38.2 |
| 2.1                                     | 5.2 | 14.2 | 10.0      | 15.4 | 8.2  | 1.0       | 2.8  | 11.0 | N         | N   | N    | 1.8       | 4.6  | 12.7 |
| N                                       | N   | 1.1  | 1.0       | 1.2  | 1.2  | N         | 1.4  | 1.2  | N         | N   | N    | N         | N    | N    |
| 0.7                                     | 2.2 | 6.2  | 5.4       | 12.5 | 24.3 | 1.5       | 2.4  | 4.7  | N         | 0.4 | 1.3  | 2.6       | 5.7  | 20.0 |
| N                                       | N   | 0.5  | 0.7       | 2.6  | 1.6  | N         | N    | 0.8  | N         | N   | 0.2  | N         | N    | 1.0  |
| 1.6                                     | 5.3 | 13.2 | 4.3       | 21.4 | 7.2  | N         | 1.7  | 4.1  | N         | 0.2 | 0.3  | 2.4       | 5.3  | 16.8 |
| 2.6                                     | 8.2 | 19.0 | N         | N    | N    | 11.7      | 29.4 | 64.0 | N         | 0.3 | 0.5  | 5.4       | 12.2 | 25.1 |
|   | 1.6 | 3.4  | 1.3       | 1.1  | N    | N         | N    | N    | N         | N   | 0.9  | N         | N    | N    |
| 0.9                                     |     |      |           |      |      |           |      |      |           |     |      |           |      |      |
| 3.4                                     | 9.2 | 16.3 | 11.3      | 28.5 | 16.7 | 1.3       | 3.4  | 9.5  | N         | N   | 0.3  | 4.0       | 9.2  | 26.8 |
| N                                       | 0.6 | 1.1  | 3.0       | 5.4  | 14.1 | N         | 0.7  | 3.8  | N         | N   | N    | 1.5       | 3.6  | 11.6 |
| 2.1                                     | 5.7 | 13.0 | 0.9       | 0.9  | 1.2  | 1.0       | 1.9  | 7.5  | N         | N   | N    | 1.4       | 4.5  | 12.2 |
| 2.9                                     | 6.9 | 18.4 | 7.4       | 3.6  | 1.0  | 5.4       | 16.6 | 26.1 | N         | N   | N    | 15.8      | 36.6 | 10.3 |
| 1.3 ± 0.4                               |     |      | 5.4 ± 0.3 |      |      | 6.8 ± 0.4 |      |      | 2.1 ± 0.1 |     |      | 1.7 ± 0.5 |      |      |

aminophen was a substrate for SULTs 1A1, 1E1, 1A3, 1C2, and 1B1, but not SULT1C3 and SULT2A1. Substrates for SULT1C3 have not been reported previously. Our data indicate that this recently identified member of the hSULT family is able to sulfonate *p*-nitrophenol, 1-naphthol, 2-ethylphenol, 2-*n*-propylphenol, and 2-*sec*-butylphenol, as well as the steroid-related compounds,  $\alpha$ -zearalenol and lithocholic acid. SULT1C3 appeared to be most active with  $\alpha$ -zearalenol (4.1 nmol/min/mg) and 2-ethylphenol (2.2 nmol/min/mg). These data suggest SULT1C3 may contribute to the metabolism of steroid and phenolic compounds. Finally, SULTs 2A1 and 1E1, which are reported to metabolize steroids [29–32], both bound to and sulfonated multiple steroid and steroid-like compounds with different apparent specificities. These data show that despite the limitations of our rapid screening method, the enzymatic activity data in Table 2 reflects, to a first approximation, the expected relative substrate activities reported in the literature and reveal new activities toward pharmacologically important compounds. The binding data, on the other hand, suggest a more complicated situation.

### PAP Is Able to Alter Ligand Binding Profiles

The ligand binding profiles were remarkably different in the presence and absence of PAP (with the exception of SULT4A1, which will be discussed separately below). Some known substrates for the well-characterized SULTs appeared to bind only in the presence of PAP—for example, dopamine for SULT1A3 [24,25], and 1-naphthol for SULT1B1 [33]—whereas many previously unreported compounds bound to these and other family members in the absence of PAP. It is interesting to note that not all known substrates, nor all of those with reactivity in our activity screens, were found to *bind* to the enzyme in the presence of PAP. For example, resveratrol was a substrate for SULT1A3 and SULT1C2, but did not stabilize either protein in the binding assays, either in the presence or absence of PAP. There are several reasons that may account for these observations. First, the enzymatic screen is likely more sensitive than the ligand binding assay in which compounds with  $K_m$  or  $K_d$  values in the high micromolar range are less likely to be detected [20,21]. Second, some ternary complexes may not be significantly stabilized relative to the PAP–enzyme complex (especially at elevated





**Figure 5** Inhibitory Effect of Isoprenaline, PLP, Quercetin, DBHD, and PCP on Sulfotransferase Activity of SULT1B1 with 1-Naphthol as Substrate

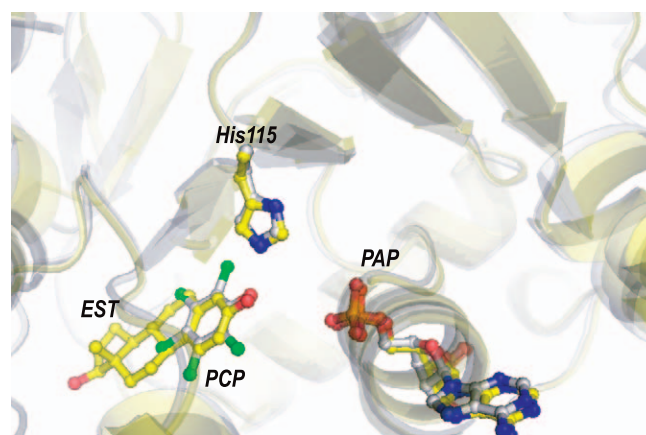
Sulfotransferase activity of SULT1B1 at different concentrations of each compound (5 to 312  $\mu$ M) has been assayed as described in Materials and Methods. Isoprenaline is indicated by a filled diamond, PLP by an open circle, quercetin by a filled square, DBHD by a filled triangle, and PCP by a filled circle.

doi:10.1371/journal.pbio.0050097.g005

temperatures). Finally, the presence of the sulfonate group of PAPS may also contribute to binding of substrates, and these cases may not be detected in our binding assay. Interestingly, binding of SULT1C1, 1B1, and 1E1 to resveratrol was only observed in the absence of PAP, but all are active toward this substrate. Nevertheless, the radically different binding profiles observed in the presence and absence of PAP are consistent with the structure-based mechanisms proposed above. Specifically, PAP (and presumably PAPS) appears to prime the substrate binding loops for subsequent binding to certain substrates, whereas in the absence of cofactor, the loops are free to bind alternative ligands (perhaps only at high concentrations), or non-productive ligand-bound conformations may exist. This priming of the substrate binding loops is likely made possible by flexibility of the binding loops observed in the structure. This structural plasticity may allow a reconfiguration of substrate binding loops in the absence of PAP(S) in order to bind a different chemical class of compound. For example, both SULT1C3 and SULT1B1 were stabilized by catecholamines in the absence of PAP, but neither showed significant activity toward this class of compounds. This raises the possibility that certain endogenous and/or exogenous compounds, such as those that bind in the absence of PAP, may act as competitive inhibitors of SULTs by occupying the substrate binding pocket and preventing a productive PAPS binding conformation, as for SULT2A1 (Figure 2J).

### Screening and Structural Analysis Reveals Novel Mechanism of Inhibition

Excluding SULT4A1, three compounds bound to all hSULTs: adenosine 5'-( $\beta,\gamma$ -imido) triphosphate (AMP-PNP), a non-hydrolysable ATP analog; pyridoxal 5-phosphate (PLP), a competitive inhibitor for sulfotransferases [34]; and quercetin, a potent inhibitor of SULT1A1 and SULT1E1 [35]. These compounds had been known to inhibit one or more sulfotransferases, but our data suggest that they may be universal SULT inhibitors. AMP-PNP and PLP bound only in



**Figure 6.** The Superimposition of the Active Sites of the SULT1E1-PAP-17 $\beta$ -estradiol Complex (1AQU) and the SULT1C2-PAP-PCP Complex (2GWH)

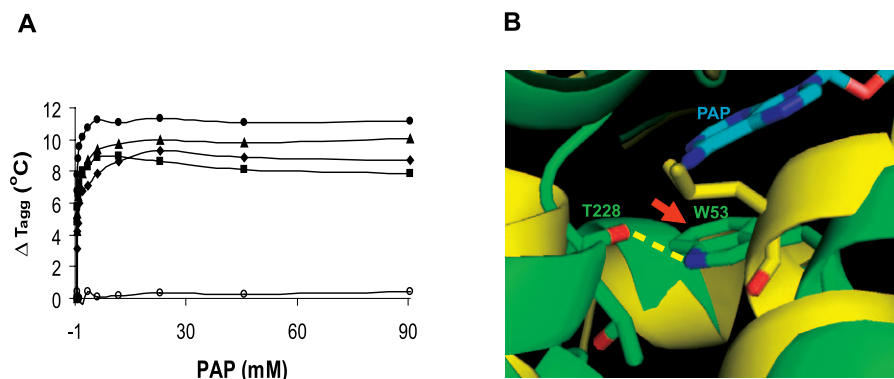
The SULT1E1-PAP-17 $\beta$ -estradiol complex is shown in yellow, and the SULT1C2-PAP-PCP complex is shown in grey. The PCP (chlorine atoms in green) and 17 $\beta$ -estradiol (EST) molecules are highlighted along with the catalytic His115 and PAP. See text for discussion.

doi:10.1371/journal.pbio.0050097.g006

the absence of saturating amounts of PAP, suggesting that they occupy the PAP binding site, as might be expected from their structural similarity to PAP. Quercetin, found in many fresh fruits and vegetables, is a flavonoid with anti-tumour and anti-inflammatory activities. It is possible that some of its favourable physiological effects may be related to inhibition of hSULT activity. Additional inhibitors bound to only a subset of the hSULTs, including 3,5-dibromo-4-hydroxybenzoic acid (6,8-dichloro-4-oxo-4*H*-chromen-3-ylmethylene) hydrazide (DBHD), 3,5-dibromo-4-hydroxybenzoic acid (6-chloro-4-oxo-4*H*-chromen-3-ylmethylene)-hydrazide (DBHM), and PCP.

The binding profiles in Table 1 raise the possibility that compounds that bind to hSULTs only in the absence of PAP may inhibit hSULT activity. In order to investigate this possibility, we assayed the activity of SULT1B1 in the presence of increasing concentrations of five compounds found in our screens. These five include known inhibitors (quercetin, DBHD, PLP, and PCP), as well as isoprenaline, which binds to SULT1B1 only in the absence of PAP and is a poor substrate for this enzyme (Tables 1 and 2). As shown in Figure 5, PCP and DBHD were strong inhibitors of SULT1B1, whereas PLP and quercetin had intermediate effects. Isoprenaline, however, had no inhibitory effect on the activity of SULT1B1 with 1-naphthol, and it was sulfonated by SULT1B1 (Table 2), indicating that not all compounds that bind in the absence of PAP are necessarily inhibitors.

PCP is a significant environmental toxin due to its common use as a wood preservative and its use in the pulp and paper industry. Its chemical structure is related to hydroxylated metabolites of polychlorinated biphenols (OH-PCBs) whose endocrine-disruptive properties may be related to the inhibition of estradiol sulfonation by SULT1E1 [36]. The mechanism of inhibition of SULT1E1 by OH-PCBs and related compounds has been proposed to take place via both allosteric [36,37] and competitive [12,37] mechanisms. Our binding data suggest that PCP may be a competitive inhibitor of SULTs 1B1, 1C2, and 1A1 (Table 1). For 1C2 and 1A1, PAP



**Figure 7.** SULT4A1 Does Not Bind PAP

(A) Thermal aggregation temperatures for SULT1B1 (filled circle), SULT1C3 (filled triangle), SULT1C1 (filled diamond), SULT1C2 (filled square), and SULT4A1 (open circle) as a function of PAP concentration. PAP provides no stabilizing effect to SULT4A1.

(B) Superimposition of the PAP binding site of SULT1C1 (green) and SULT4A1 (yellow). A conserved Trp residue (W53 of SULT1C1; green ring in the middle of figure, indicated by a red arrow) normally forms  $\pi$ - $\pi$  stacking interactions with the adenine ring of PAP (blue) and hydrogen bonds to a neighbouring conserved Thr (T228 in SULT1C1). SULT4A1 is the only SULT to have a Leu instead of Trp at this position. The Leu residue cannot make stabilizing interactions with Thr or the adenine ring, and its position would cause a severe steric clash with PAP and PAPS. doi:10.1371/journal.pbio.0050097.g007

was required for binding; and for 1B1, PCP binds much better in the presence of PAP, suggesting that PCP binds in a substrate-like conformation facilitated by PAP(S). In order to better understand the mechanism of inhibition by PCP, we determined the structure of SULT1C2 bound to PAP and PCP (Figure S3). The structure reveals that the protein undergoes a disorder-order transition upon PCP and PAP binding. Helices  $\alpha 4$ ,  $\alpha 5$ , and  $\alpha 15$ , and loops  $\alpha 5$ - $\alpha 6$  and  $\alpha 15$ - $\alpha 16$  are ordered only in the ternary complex, but not in the apo SULT1C2 structure (Figure 2B and 2H). PCP is found in the substrate binding pocket and therefore appears to be a competitive inhibitor, consistent with crystallographic analysis of SULT1E1 bound to PAP and 3,5,3',5'-tetrachlorobiphenol [12]. Comparison of the SULT1C2-PAP-PCP and SULT1E1-PAP-estradiol structures (Figure 6) revealed two structural features that may be relevant in explaining the mechanism of PCP inhibition. In the co-crystal structures, the phenol moieties of PCP and estradiol share the same relative position and orientation, positioning the phenolic OH within hydrogen-bond distance of the catalytic histidine. This histidine is thought to be deprotonated and made catalytically competent to accept the phenolic hydrogen from estradiol, facilitating nucleophilic attack on the sulfonate of PAPS. Our structure shows that PCP appears to be in a catalytically competent conformation. However, PCP and estradiol differ dramatically in the acidity of their hydroxyl groups; the estimated pKa for estradiol is approximately 15, whereas that for PCP is 4.5. Thus, although PCP appears to bind in a catalytically competent conformation, the phenolic oxygen of PCP may be too weak a nucleophile to attack the sulfonate of PAPS.

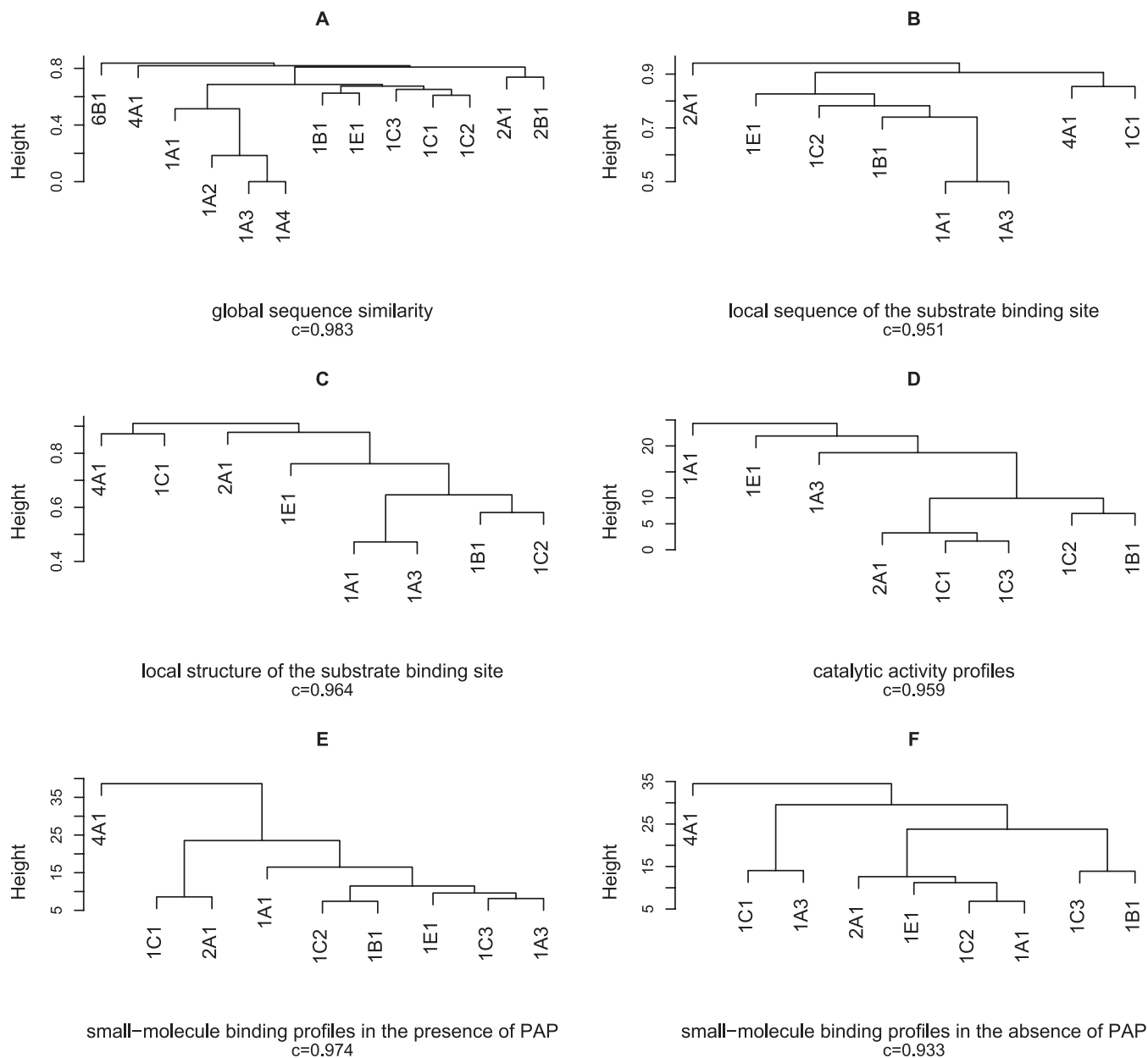
The inhibitory effects of OH-PCBs and PCP have been interpreted previously in terms of variations in the bound conformation of the halogenated compounds relative to that of estradiol [12,36]. Although steric and conformational factors clearly play a role, our structure of SULT1C2 with PCP and PAP suggests a key role for the electronic nature of the halogenated phenols. We examined the calculated pKa values for the series of 21 4-hydroxyl-substituted PCBs for which Kester et al. [36] reported 50% inhibitory concen-

tration ( $IC_{50}$ ) values. The results show a strong correlation between the calculated acidity of these phenols with  $IC_{50}$  values, with the most acidic compounds having the strongest inhibitory effect (Figure S4).

#### SULT4A1 Is an Atypical SULT with an Atypical Structure

Under the conditions of our screens, PAP bound to all hSULTs except for SULT4A1. In order to rule out the possibility that SULT4A1 simply has a much weaker affinity for PAP, we performed titration experiments for several of the proteins with increasing concentrations of PAP ranging from 90  $\mu$ M to 90 mM (Figure 7A). SULT1C1, SULT1C2, SULT1C3, and SULT1B1 showed similar saturation binding curves, reaching saturation at about 100  $\mu$ M. However, PAP when added at concentrations as high as 90 mM did not stabilize SULT4A1. SULT4A1 is one of the hSULTs that is most divergent in sequence, and examination of the binding pocket revealed two significant differences that are predicted to affect PAP binding. First, a Trp in  $\alpha 3$  that is conserved in all other hSULTs and stacks with the adenine ring of PAPS (Trp53 in SULT1A1) is replaced with a Leu in SULT4A1 (Figure 3). Second, the magenta PAP binding loop is much shorter in SULT4A1 than in the other hSULTs, and lacks the conserved Lys residue that separates the key PAPS binding residues Arg258 and Gly260 (SULT1A1 numbering), which results in these residues being out of register. Taken together, SULT4A1 has a slightly smaller PAPS binding pocket that is predicted to be unable to accommodate the cofactor. Interestingly, some residual electron density was observed in the PAP binding pocket of recombinant SULT4A1. Presumably this derives from a bound small molecule that was co-crystallized, although we were not able to identify it either by modelling atoms into the electron density or using mass spectrometry.

To test the possibility that SULT4A1 might use an alternate sulfonate donor, we tested several potential alternates such as adenosine phosphosulfate, 4-nitrocatechol sulfate, 4-acetylphenyl sulfate, estrone 3-sulfate, indoxyl sulfate, and 4-methylumbelliferyl sulfate. None of these compounds stabilized SULT4A1 against thermal aggregation (unpublished



**Figure 8.** Average Linkage Hierarchical Clustering of SULTs

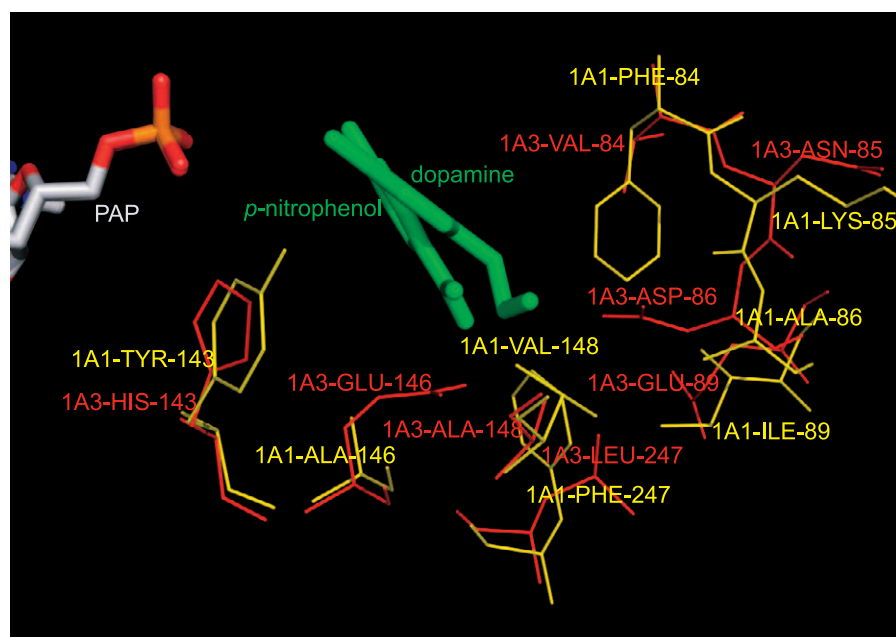
Clustering is shown according to (A) global sequence similarity, (B) local sequence of the substrate binding site, (C) local structure of the substrate binding site, (D) catalytic activity profiles, (E) small molecule binding profiles in the presence of PAP, and (F) small molecule binding profiles in the absence of PAP. The value under each clustering (c) is the correlation between the original data matrix and the cophenetic matrix—a matrix whose elements represent the height at which these elements first meet in the tree. The higher this correlation, the more accurate the tree represents the original data. See Materials and Methods for further details.

doi:10.1371/journal.pbio.0050097.g008

data). These data strongly suggest that SULT4A1 may not have significant catalytic activity *in vivo*. Indeed, although very weak activity has been reported in one case [38], other groups have failed to observe activity for human SULT4A1 [5,39]. Significantly, this protein, which is expressed primarily in the brain, binds to 2-hydroxylestradiol, thyroid hormone, T4 (3,3',5,5'-tetraiodo-L-thyronine), and the catecholamines norepinephrine, epinephrine, and isoprenaline (but not dopamine), suggesting that SULT4A1 may modulate the bioactivity of these compounds via a mechanism distinct from sulfonation. Of note, SULT4A1 did not bind any simple phenolic compounds under the conditions tested here.

### Binding Profiles Suggest Alternative Classifications of hSULTs

Examination of the chemical fingerprints reflected in Tables 1 and 2 suggests that subsets of the hSULTs can be clustered based on the chemical properties of the compounds that they bind—relationships that are not evident from global sequence comparison. To explore alternative activity or structure-based classifications of hSULTs in more detail, we performed average hierarchical clustering on the experimental data in an attempt to identify correlations between the *local* sequence or structural features of the substrate binding pockets and activity profiles among the hSULTs. Figure 8



**Figure 9.** Binding Site Comparison of SULT1A1 Bound to PAP and *p*-Nitrophenol (1LS6) and SULT1A3 Bound to PAP and Dopamine (2A3R). The side chains of the eight binding-site residues that differ between these two proteins are shown in yellow (SULT1A1) and red (SULT1A3). doi:10.1371/journal.pbio.0050097.g009

shows the clustering of similarity matrices for each parameter, viewed using trees. Considering only global sequence similarity, the hSULTs cluster according to their nomenclature and phylogenetic relationships (Figure 8A). Considering only the nine proteins for which we have binding or enzymatic data, SULTs 1A1 and 1A3 are most closely related, with a global sequence identity of 95%. The three SULT1C proteins cluster with average sequence identities close to 55%, as do SULT1E1 and 1B1. SULT2, -4, and -6 subfamilies are relative outliers with sequence identities to all other SULTs considered here around 35% or less. It is well known that related enzymes with sequence identities below 40% have often evolved to have different substrate specificities [40], and therefore, we would expect most of the SULTs to sulfonate different substrates, except perhaps 1A1 and 1A3. In keeping with this concept, the closely related SULTs 1A1 and 1A3 are clearly the most similar within the substrate binding site, as measured by both local sequence and structure comparisons (Figure 8B and 8C). However, the clustering of the other more distantly related SULTs is sufficiently different at the level of local sequence and structure, such that the SULT1C proteins are no longer clustered together and the outliers are different. These comparisons show that the local sequence and structures of the substrate binding sites do not correspond to the global sequence relationships.

These results also illustrate why members of the same subfamily do not metabolize the same classes of compounds. Although the clustering of hSULTs presented here is likely influenced by the limited subset of compounds used in our binding and activity assays, the results provide an initial view of the family-wide activity-based classifications. The trees that cluster the eight or nine proteins according to their binding and activity profiles (Figure 8D–8F) show a remarkably different clustering from the sequence- and structure-based

trees. First, 1A3 and 1A1 no longer cluster together, despite their strong global sequence similarity (95% identity). Inspection of Table 1 shows that 1A1 strongly binds most of the phenols and acidic compounds, once PAP is bound, whereas 1A3 shows absolutely no binding of these substrates. These SULTs are also distinguished by their differential reactivity toward catecholamines, with SULT1A3 able to bind dopamine and having higher relative activity toward catecholamines compared to phenols, as previously noted [28]. Comparison of the residues in the substrate binding loops of SULT1A1 and SULT1A3 revealed that all residues are identical (and in identical positions in the structures) except for the eight residues shown in Figure 9. These changes map to two loops (residues 84–89 and 143–148) and residue 247 (SULT1A1 numbering). Importantly, SULT1A3, which does not bind acidic compounds, has acidic instead of hydrophobic residues at three of these positions. The net result is a much more negatively charged pocket for SULT1A3 compared to SULT1A1. This difference would disfavor binding of compounds with a net negative charge to SULT1A3 and favour interactions with the amino group of catecholamines. Thus, the strong local sequence and structure similarities in 1A1 and 1A3 are manifested in their similar ability to bind similar inhibitors and sulfonate catecholamines (as a class), but small, local sequence changes in the substrate binding site have enhanced the ability of SULT1A3 to bind catecholamines such as dopamine [28], and completely changed its ability to bind acidic compounds. The influence of residue 146 on the specificity of SULT1A1 compared to SULT1A3 (Ala vs. Glu) has been previously noted [28]; however, the results presented here suggest that additional differences in the binding loops also contribute to specificity.

SULT1B1 and 1C2 cluster together in the trees of local structure, activity, and binding in the presence of PAP. These groupings reflect their common ability to bind acidic

compounds (and the acidic phenols PCP and 2,6-dichloro-4-nitrophenol), with promiscuous activity profiles toward phenols. Thus, SULT1B1 and 1C2 appear to be much more closely related in terms of structure and activity than their global sequences would indicate.

Finally, Figure 8F shows that the clustering of hSULTs again differs when considering the binding of compounds in the absence of PAP. In this case, none of the previously noted similarities is evident. Many of the compounds in Table 1 are inhibitors of hSULT activity. The differential clustering in the absence of PAP may reflect possible unrelated configurations of binding-site residues when bound to these compounds or the tendency for more disorder in the absence of PAP. The clustering of both binding profiles (Figure 8E and 8F) places SULT4A1 as the furthest outlier, analogous to its position in the global sequence comparison. Although this is consistent with the inability of SULT4A1 to bind PAP and catalyze sulfonation, it is interesting to note that this is not due to a radical difference in local binding-site sequence or structure, because SULT4A1 clusters with SULT1C1 when considering these local factors (Figure 8B and 8C). As noted above, apparently small differences of one to two residues in the PAP binding site are likely responsible for this behaviour.

## Discussion

An important challenge in structural biology, chemical biology, and drug discovery is to relate changes in local sequence and structure to the binding and activity profiles of homologous enzymes in order to better predict or explain substrate and inhibitor specificity within an enzyme family. In the case of the hSULTs, this is particularly desirable in order to predict the fate of xenobiotics, hormones, and drug candidates in humans. Like other phase II detoxification enzymes, hSULTs are known to have broad and overlapping substrate specificities. We and others have shown that this promiscuity derives from the considerable flexibility or plasticity of the hSULT binding sites and therefore, a full understanding of specificity will require multiple three-dimensional structures for each hSULT in complex with substrates and inhibitors, as well as knowledge of the full spectrum of small molecules that bind in both productive and non-productive conformations. Our structural and chemical profiling data prepare the foundation for such detailed studies.

Here we also reveal a previously unrecognized structural role for the cofactor PAP(S); namely priming of the often disordered substrate binding loops for interaction with substrates. The “magenta loop” (Figure 2), which contributes to both PAPS and substrate binding pockets, can also, in some instances, adopt an inactive conformation and may explain substrate-induced inhibition and/or inhibition by compounds that bind in the absence of PAPS, as well as by known inhibitors. The flexibility of the substrate binding loops in hSULTs likely contributes to the wide repertoire of compounds that can be accommodated in the substrate binding pocket, only a subset of which lead to productive sulfonation.

Our results provide insight into mechanisms of inhibition of hSULTs. In addition to structural mechanisms of substrate inhibition (above), we identified three compounds (PLP, AMP-PNP, and quercetin) that appear to be broad-spectrum hSULT inhibitors, and may also inhibit other non-cytosolic

sulfotransferases. We have also provided insight into how PCP and possibly other polychlorinated phenolic compounds can inhibit hSULTs. These compounds, which are known endocrine disruptors, appear to bind in a manner very similar to other productive substrates, but are unreactive, at least in part, due to their weakly acidic properties.

Our analyses have also provided insight into the functions of the less well-characterized SULT1C subfamily and SULT4A1. We have identified a number of novel substrates, inhibitors, and compounds that bind to these SULTs in the absence of PAP. These data combined with activity assays revealed that SULT1C3 can bind catecholamines and phenolic compounds, but only the latter are substrates. SULT4A1 is inactive as an enzyme in our experiments, likely due to its inability to bind PAPS or other sulfonate donors. This orphan SULT likely has an important function in the brain, nevertheless, because it is highly conserved and binds well to the neurotransmitters epinephrine and norepinephrine.

The approach outlined here in which simple, medium throughput binding and activity screens can be used to profile properties of purified enzymes has proven extremely useful for identification of novel substrates and analysis of specificity across a human protein family. We demonstrate that the relationship between sequence/structure and function within this small family is remarkably complex, and differences in activity can reflect just a few amino acid changes at critical locations within the protein's active site. Global sequence/structure comparisons provide good clues for broad functional classification, but cannot simply define an enzyme's cognate substrate or class of substrates. For the sulfotransferases, which have a large and flexible binding site, it is clearly necessary to perform much more detailed studies to understand both the binding and catalytic activities in terms of local structure. The actual cellular activity of hSULTs will depend on the spectrum of compounds available to a given enzyme and their relative concentrations. The tissue-specific and developmental variation in both hSULT expression and the cellular milieu of small molecules complicates further attempts to *predict* activities. Ultimately, detailed enzymatic characterization of all purified hSULTs, as well as cellular assays, will be needed to fully understand this family. The data presented here form a basis for further detailed biochemical and structural studies of both active and inactive enzyme–small molecule complexes in order to fully understand the role of hSULTs in the metabolic fate of endogenous substrates, as well as drugs and toxic compounds.

## Materials and Methods

**Protein purification and crystallization.** The SULT1B1, SULT1C1, SULT1C2, SULT1C3, and SULT4A1 genes were amplified by PCR from the Mammalian Gene Collection clones and subcloned into a modified pET28a-LIC vector. Expression and purification of recombinant proteins was as described by Dombrowski et al. [17]. Purified recombinant proteins contained an additional Gly-Ser dipeptide at the N-terminus. Additional details are provided at <http://www.sgc.utoronto.ca> in the structure gallery for each protein.

Purified SULT1B1, SULT1C1, and SULT1C3 were crystallized in the presence of 2 mM PAP using the hanging drop method at 20 °C by mixing: for SULT1B1—2  $\mu$ l of the protein solution with 2  $\mu$ l of the reservoir solution containing 0.1 M Bis-Tris (pH 6.5), 0.2 M ammonium sulfate, and 16%–20% polyethylene glycol 4000; for SULT1C1—2  $\mu$ l of the protein solution with 2  $\mu$ l of the reservoir solution containing 0.1 M  $K_2HPO_4$  and 12%–16% polyethylene glycol 3350; and for SULT1C3—2  $\mu$ l of the protein solution with 2  $\mu$ l of the reservoir solution containing 18% polyethylene glycol 3350, 0.2 M

ammonium formate, 0.1 M Bis-Tris (pH 6.5). To obtain crystals of SULT1C2-PAP-PCP ternary complex, 10 mg/ml of purified SULT1C2 was mixed with 2 mM PAP and 2 mM PCP in 20 mM MES-NaOH buffer (pH 6.5), and incubated on ice for 30 min. SULT1C2-PAP-PCP complex was crystallized using the sitting drop method at 20 °C by mixing 0.8  $\mu$ l of the protein-cofactor-inhibitor mix with 0.8  $\mu$ l of the reservoir solution containing 25% polyethylene glycol 3350, 0.2 M lithium sulfate, 0.1 M Bis-Tris (pH 6.5). SULT4A1 and SULT1C2 crystals were obtained by using the hanging drop method at 20 °C by mixing 2  $\mu$ l of the protein solution with 2  $\mu$ l of the reservoir solution containing 20% polyethylene glycol 4000, 0.2 M ammonium tartrate, and 14%–20% polyethylene glycol 3350, 0.2 M lithium citrate, 0.1 M sodium citrate (pH 4.6), respectively.

**Chemical library preparation.** A library of 90 compounds was created for screening sulfotransferases. These compounds were known substrates, products, and inhibitors of sulfotransferases, their analogs, and compounds with high similarity to known inhibitors identified from the literature and public databases (<http://www.rcsb.org> and <http://www.brenda.uni-koeln.de>). Certain substrates, such as controlled substances, were not included, and some additional compounds were selected through chemical similarity to known SULT substrates and inhibitors using the ChemNavigator search engine (<http://www.chemnavigator.com/>). The compounds were dissolved in 100% DMSO at 100 mM concentration and subsequently diluted stepwise to 10 mM and 1 mM in Hepes buffer (100 mM Hepes, 150 mM NaCl [pH 7.5]). The full list of compounds in the library is included in Table S2.

**Ligand binding screens.** Screening for ligand binding was performed in 50- $\mu$ l volume with a final concentration of 1 mM of compound per well, in 384-well plates. The concentration of protein was the same for all wells at 0.4 mg/ml. Ligand binding was detected by monitoring the increase in thermostability of proteins in the presence of ligands. Protein thermostability at pH 7.5 was studied using StarGazer technology that monitors protein stability by its aggregation properties [20,21]. Protein samples at 0.4 mg/ml were heated from 27 °C to 80 °C at the rate of 0.5 °C per min in clear-bottom 384-well plates (Nunc, <http://www.nuncbrand.com/>) in 50  $\mu$ l of 100 mM Hepes (pH 7.5) and 150 mM NaCl. Protein aggregation was monitored by capturing images of scattered light every 30 s with a CCD (Charge-Coupled Device) camera. The pixel intensities in a pre-selected region of each well were integrated to generate a value representative of the total amount of scattered light in that region. These total intensities were then plotted against temperature for each sample well and fit to the Boltzman equation by nonlinear regression. The point of inflection of each “denaturation” curve was identified as  $T_{agg}$  (aggregation temperature). The increase in stability of the protein in the presence of a ligand is shown as  $\Delta T_{agg}$ .

**Sulfotransferase activity screens.** Enzyme assays were performed using a HPLC-based method that we developed for sulfotransferase activity assay by modifying the protocol previously used to monitor ADP production and ATP hydrolysis by a purified bacterial ATPase [41]. SULTs at 1–5  $\mu$ M were assayed in the presence of 0.1 to 0.5 mM PAPS and different concentrations of each substrate in 100 mM Hepes (pH 7.5) by incubating the reaction at 37 °C for a period of time from 15 to 120 min depending on how fast PAPS was converted to PAP. The  $K_m$  values for characterized sulfotransferases are in the range of nanomolar to millimolar concentrations [24,42], with a significant variation in catalytic efficiency and substrate specificity. Based on these observations and considering possible substrate inhibition [19,43], we tested all sulfotransferases at substrate concentrations of 10, 25, and 100  $\mu$ M. The reactions were stopped by adding two volumes of urea (final concentration of 5.3 M), and the mixture was filtered through a 5-kDa molecular weight cutoff Amicon Ultrafree-MC filter (Millipore, Bedford, Massachusetts, United States) to remove the protein. The ratio of PAP and PAPS was determined after separating them on HPLC using a 4.5 mm  $\times$  50 mm WP QUAT, a strong ion-exchange column (J. T. Baker, Phillipsburg, New Jersey, United States), using a gradient of triethylamine bicarbonate from 20 to 500 mM applied at 2 ml/min for 7 min. The progress of the reaction was monitored by reading the absorbance at 259 nm, and the amount of PAP produced was determined by integration of the resolved peaks using the HPLC software (Waters, <http://www.waters.com/>). All values in Table 2 were corrected for the rate of conversion of PAPS to PAP in the presence of enzyme, but no substrate. This background activity is reported for each SULT in the last row of Table 2, and the values are the average of three independent measurements.

**Sequence, structure, and data clustering analysis.** Sequences used to generate Figures 3 and 8 are as follows: SULT1A1, SULT1A2, SULT1A3, SULT1A4, SULT1B1, SULT1C1, SULT1C2, SULT1C3, SULT1E1, SULT2A1, SULT2B1, SULT4A1, and SULT6B1. We

created a multiple sequence alignment of the above-mentioned sequences using HMMer [44] and the pfam [45] Sulfotransferase\_1 (PF00685.15) Hidden Markov Model. Sequence similarity is measured using the Tanimoto coefficient of residues in common in the HMM-based alignment. The calculation of local sequence similarities involves the detection of binding site residues [46], and their subsequent mapping onto the HMM-based alignment. Cofactor (PAP/PAPS) binding residues were also mapped onto the alignment and excluded from all pairwise comparisons and similarity calculations. Substrate binding-site structural similarities were detected using a two-stage graph-matching process providing a one-to-one chemical and spatial correspondence between atoms in clefts. The method considers all non-hydrogen atoms and can use large sets of atoms as input allowing larger, over-predicted and apo-form binding sites to be analyzed. In the first stage, the two clefts are superimposed [47] via the detection of the largest clique [48] in a  $C_q$  association graph corresponding to the largest subset of identical residues in equivalent spatial locations. The first stage is used to constrain the construction of the second-stage all-atom association graph. The second-stage graph-matching results in the detection of the largest subset of heavy atoms of equivalent atoms types [49], and spatial positions. Pairwise local structural similarity was calculated as a Tanimoto coefficient based on the size of the largest clique in the second graph-matching stage. Dissimilarity matrices were derived from the similarity measures described above. Pairwise experimental catalytic and binding-profile dissimilarity matrices were calculated as the  $L_2$  distance of the vectors with the corresponding experimental measurements. Hierarchical clustering was used to create the clustering trees shown in Figure 8. The correlation between the cophenetic matrix and the original dissimilarity matrix was used to choose the linkage method that results in the most accurate representation of the original data [50]. Average linkage was found to be the clustering method of choice in all instances.

**pKa calculations of polyphenol hydroxyl moiety.** To assess the extent to which the acidity of the hydroxyl moiety of hydroxylated polychlorinated biphenols is related to their inhibitory strength, we computed the pKa values of the 4-hydroxyl group for a series of hydroxylated PCB analogs. This group of compounds and their inhibitory effect on SULT1E1 were previously reported [36]. To this end, we used the pKa calculator in the PC stand-alone version of the ACD (<http://www.acdlabs.com>) suite of programs. Two clusters of compounds, namely 4-OH-(2,3,4,5,6)Cl and 4-OH-(3,5)Cl, are identified as outliers upon computing a linear regression on the relationship ( $R^2 = 0.57$ ).

## Supporting Information

**Figure S1.** Stereo Images of Representative Model-Phased Electron Density

Found at doi:10.1371/journal.pbio.0050097.sg001 (3.1 MB DOC).

**Figure S2.** Additional Structures of Human SULTs from the Protein Data Bank

Found at doi:10.1371/journal.pbio.0050097.sg002 (1.3 MB DOC).

**Figure S3.** Stereo Image of Model-Phased 2Fo-Fc and Anomalous Difference Fourier Maps for the SULT1C2:PAP:PCP Complex

Found at doi:10.1371/journal.pbio.0050097.sg003 (700 KB DOC).

**Figure S4.** Correlation between Computed pKa Values and Published IC<sub>50</sub> Values of Hydroxylated Polychlorinated Biphenyls on hSULT1E1

Found at doi:10.1371/journal.pbio.0050097.sg004 (46 KB DOC).

**Table S1.** Data Collection and Refinement Statistics

Found at doi:10.1371/journal.pbio.0050097.st001 (95 KB DOC).

**Table S2.** List of 90 Compounds in SULT Compound Library

Found at doi:10.1371/journal.pbio.0050097.st002 (102 KB DOC).

## Accession Numbers

The National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>) Reference Sequence (RefSeq) accession numbers for the proteins referred to in this paper are SULT1A1 (NP\_803880), SULT1A2 (NP\_001045), SULT1A3 (NP\_003157), SULT1A4 (NP\_001017389), SULT1B1 (NP\_055280), SULT1C1 (NP\_789795), SULT1C2 (NP\_006579), SULT1C3 (NP\_001008743), SULT1E1

(NP\_005411), SULT2A1 (NP\_003158), SULT2B1 (NP\_814444), SULT4A1 (NP\_055166), and SULT6B1 (NP\_001027549).

The Mammalian Gene Collection (<http://mgc.nci.nih.gov>) clones discussed in this paper are SULT1B1 (gi: 29550928), SULT1C1 (gi: 4507305), SULT1C2 (gi: 28830308), SULT1C3 (gi: 56847626), and SULT4A1 (gi: 7657633).

The Protein Data Bank (<http://www.pdb.org>) IDs for the proteins discussed in this paper are SULT1A3 (1CJM), SULT1C2 (2AD1), SULT4A1 (1ZD1), SULT1C1 with PAP (2ETG), SULT1B1 with PAP (1XV1), SULT1C3 with PAP (2H8K), SULT1A1 with PAP and *p*-nitrophenol (1LS6), SULT2B1b with PAP and pregnenolone (1Q20), SULT1E1 with PAP and 3,5,3',5'-tetrachloro-biphenyl-4,4'-diol (1G3M), SULT2A1 with DHEA (1J99), SULT2A1 with PAP (1EFH), SULT1C2 with PAP and PCP (2GWH), SULT1A3 with PAP and dopamine (2A3R), SULT1A1 with PAP and estradiol (2D06), SULT2A1 with androstosterone (1OV4), SULT2B1a with PAP and 2-[*N*-cyclohexylamino]ethane sulfonic acid (1Q1Q), and SULT2B1b with PAP and DHEA (1Q22).

## Acknowledgments

The authors are grateful to the reviewers of this manuscript for their excellent and insightful suggestions and corrections. We also thank R. Weinsilboum and M. Coughtrie for fruitful discussions.

**Author contributions.** AAH performed all binding assays, developed the SULT activity assay, and performed some of the activity assays. PWP performed many of the activity assays, purified the proteins for activity and binding assays, and crystallized SULT1C2

with PAP and PCP. LD purified and crystallized all other proteins whose structures are reported here. RN performed structure, binding, and activity profile analysis. WT solved the structure of SULT1C2 with PAP and PCP, and performed structural analysis. AD helped with X-ray data collection and crystallization, and solved the structures of SULTs 1B1, 1C1, 1C2, and 4A1. PL generated all the expression clones. FM performed computational analysis of SULT4A1 and SULT1C2 with PAP and PCP, and PCBs. JT contributed to data analysis and the manuscript. AME contributed to overall project design and the manuscript. AB led the crystallographic part of the project and contributed to data analysis and the manuscript. ANP led the structural genomics part of the project and contributed to structure analysis and the manuscript. MV led the screening part of the project and contributed to data analysis and the manuscript. CHA contributed to overall project design and data analysis, and led the preparation of the manuscript. AAH, PWP, LD, RN, ANP, and MV contributed equally to this paper.

**Funding.** This research was supported by the Structural Genomics Consortium with funds from Genome Canada through the Ontario Genomics Institute, the Canadian Institutes for Health Research, the Canada Foundation for Innovation, the Ontario Challenge Fund, the Ontario Innovation Trust, the Wellcome Trust, GlaxoSmithKline, the Knut and Alice Wallenberg Foundation, Vinnova, and Swedish Foundation for Strategic Research.

**Competing interests.** The authors have declared that no competing interests exist.

## References

- Glatt H (2000) Sulfotransferases in the bioactivation of xenobiotics. *Chem Biol Interact* 129: 141–170.
- Blanchard RL, Freimuth RR, Buck J, Weinsilboum RM, Coughtrie MW (2004) A proposed nomenclature system for the cytosolic sulfotransferase (SULT) superfamily. *Pharmacogenetics* 14: 199–211.
- Freimuth RR, Wierpert M, Chute CG, Wieben ED, Weinsilboum RM (2004) Human cytosolic sulfotransferase database mining: Identification of seven novel genes and pseudogenes. *Pharmacogenomics* 4: 54–65.
- Hildebrandt MA, Salavaggione OE, Martin YN, Flynn HC, Jalal S, et al. (2004) Human SULT1A3 pharmacogenetics: gene duplication and functional genomic studies. *Biochem Biophys Res Commun* 321: 870–878.
- Falany CN, Xie X, Wang J, Ferrer J, Falany JL (2000) Molecular cloning and expression of novel sulphotransferase-like cDNAs from human and rat brain. *Biochem J* 346: 857–864.
- Gamage NU, Tsvetanov S, Duggleby RG, McManus ME, Martin JL (2005) The structure of human SULT1A1 crystallized with estradiol. An insight into active site plasticity and substrate inhibition with multi-ring substrates. *J Biol Chem* 280: 41482–41486.
- Chapman E, Best MD, Hanson SR, Wong CH (2004) Sulfotransferases: Structure, mechanism, biological activity, inhibition, and synthetic utility. *Angew Chem Int Ed Engl* 43: 3526–3548.
- Gamage NU, Duggleby RG, Barnett AC, Tresillian M, Latham CF, et al. (2003) Structure of a human carcinogen-converting enzyme, SULT1A1. Structural and kinetic implications of substrate inhibition. *J Biol Chem* 278: 7655–7662.
- Lu JH, Li HT, Liu MC, Zhang JP, Li M, et al. (2005) Crystal structure of human sulfotransferase SULT1A3 in complex with dopamine and 3'-phosphoadenosine 5'-phosphate. *Biochem Biophys Res Commun* 335: 417–423.
- Bidwell LM, McManus ME, Gaedigk A, Kakuta Y, Negishi M, et al. (1999) Crystal structure of human catecholamine sulfotransferase. *J Mol Biol* 293: 521–530.
- Pedersen LC, Petrotchenko E, Shevtsov S, Negishi M (2002) Crystal structure of the human estrogen sulfotransferase-PAPS complex: Evidence for catalytic role of Ser137 in the sulfonyl transfer reaction. *J Biol Chem* 277: 17928–17932.
- Shevtsov S, Petrotchenko EV, Pedersen LC, Negishi M (2003) Crystallographic analysis of a hydroxylated polychlorinated biphenyl (OH-PCB) bound to the catalytic estrogen binding site of human estrogen sulfotransferase. *Environ Health Perspect* 111: 884–888.
- Chang HJ, Shi R, Rehse P, Lin SX (2004) Identifying androstosterone (ADT) as a cognate substrate for human dehydroepiandrosterone sulfotransferase (DHEA-ST) important for steroid homeostasis: Structure of the enzyme-ADT complex. *J Biol Chem* 279: 2689–2696.
- Rehse PH, Zhou M, Lin SX (2002) Crystal structure of human dehydroepiandrosterone sulphotransferase in complex with substrate. *Biochem J* 364: 165–171.
- Pedersen LC, Petrotchenko EV, Negishi M (2000) Crystal structure of SULT2A3, human hydroxysteroid sulfotransferase. *FEBS Lett* 475: 61–64.
- Lee KA, Fuda H, Lee YC, Negishi M, Strott CA, et al. (2003) Crystal structure of human cholesterol sulfotransferase (SULT2B1b) in the presence of pregnenolone and 3'-phosphoadenosine 5'-phosphate. Rationale for specificity differences between prototypical SULT2A1 and the SULT2B1 isoforms. *J Biol Chem* 278: 44593–44599.
- Dombrowski L, Dong A, Bochkarev A, Plotnikov AN (2006) Crystal structures of human sulfotransferases SULT1B1 and SULT1C1 complexed with the cofactor product adenosine-3'-5'-diphosphate (PAP). *Proteins* 64: 1091–1094.
- Hoff RH, Czuryca PG, Sun M, Leyh TS, Hengge AC (2006) Transition state of the sulfonyl transfer reaction of estrogen sulfotransferase. *J Biol Chem* 281: 30645–30649.
- Barnett AC, Tsvetanov S, Gamage N, Martin JL, Duggleby RG, et al. (2004) Active site mutations and substrate inhibition in human sulfotransferase 1A1 and 1A3. *J Biol Chem* 279: 18799–18805.
- Senisterra GA, Markin E, Yamazaki K, Hui R, Vedadi M, et al. (2006) Screening for ligands using a generic and high-throughput light-scattering-based assay. *J Biomol Screen* 11: 940–948.
- Vedadi M, Niesen FH, Allali-Hassani A, Fedorov OY, Finerty PJ Jr, et al. (2006) Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A* 103: 15835–15840.
- Dooley TP (1998) Cloning of the human phenol sulfotransferase gene family: Three genes implicated in the metabolism of catecholamines, thyroid hormones and drugs. *Chem Biol Interact* 109: 29–41.
- Coughtrie MW (1998) Catecholamine sulfation in health and disease. *Adv Pharmacol* 42: 339–342.
- Brix LA, Barnett AC, Duggleby RG, Leggett B, McManus ME (1999) Analysis of the substrate specificity of human sulfotransferases SULT1A1 and SULT1A3: site-directed mutagenesis and kinetic studies. *Biochemistry* 38: 10474–10479.
- Dajani R, Cleasby A, Neu M, Wonacott AJ, Jhoti H, et al. (1999) X-ray crystal structure of human dopamine sulfotransferase, SULT1A3. Molecular modeling and quantitative structure-activity relationship analysis demonstrate a molecular basis for sulfotransferase substrate specificity. *J Biol Chem* 274: 37862–37868.
- Hempel N, Barnett A, Gamage N, Duggleby RG, Windmill KF, et al. (2005) Human SULT1A SULTs. In: Pacifici GM, Coughtrie MW, editors. *Human cytosolic sulfotransferases*. Boca Raton (Florida): CRC Press. pp. 179–230.
- Veronese ME, Burgess W, Zhu X, McManus ME (1994) Functional characterization of two human sulphotransferase cDNAs that encode monoamine- and phenol-sulphating forms of phenol sulphotransferase: Substrate kinetics, thermal-stability and inhibitor-sensitivity studies. *Biochem J* 302: (Pt 2): 497–502.
- Baur JA, Sinclair DA (2006) Therapeutic potential of resveratrol: The in vivo evidence. *Nat Rev Drug Discov* 5: 493–506.
- Falany CN, Comer KA, Dooley TP, Glatt H (1995) Human dehydroepiandrosterone sulfotransferase. Purification, molecular cloning, and characterization. *Ann N Y Acad Sci* 774: 59–72.
- Falany CN, Krasnykh V, Falany JL (1995) Bacterial expression and characterization of a cDNA for human liver estrogen sulfotransferase. *J Steroid Biochem Mol Biol* 52: 529–539.
- Adjei AA, Weinsilboum RM (2002) Catecholestrogen sulfation: Possible role in carcinogenesis. *Biochem Biophys Res Commun* 292: 402–408.
- Comer KA, Falany JL, Falany CN (1993) Cloning and expression of human

- liver dehydroepiandrosterone sulphotransferase. *Biochem J* 289: (Pt 1): 233–240.
33. Wang J, Falany JL, Falany CN (1998) Expression and characterization of a novel thyroid hormone-sulfating form of cytosolic sulfotransferase from human liver. *Mol Pharmacol* 53: 274–282.
  34. Kamio K, Honke K, Makita A (1995) Pyridoxal 5'-phosphate binds to a lysine residue in the adenosine 3'-phosphate 5'-phosphosulfate recognition site of glycolipid sulfotransferase from human renal cancer cells. *Glycoconj J* 12: 762–766.
  35. Pacifici GM (2004) Inhibition of human liver and duodenum sulfotransferases by drugs and dietary chemicals: A review of the literature. *Int J Clin Pharmacol Ther* 42: 488–495.
  36. Kester MH, Bulduk S, Tibboel D, Meinel W, Glatt H, et al. (2000) Potent inhibition of estrogen sulfotransferase by hydroxylated PCB metabolites: A novel pathway explaining the estrogenic activity of PCBs. *Endocrinology* 141: 1897–1900.
  37. Kester MH, Bulduk S, van Toor H, Tibboel D, Meinel W, et al. (2002) Potent inhibition of estrogen sulfotransferase by hydroxylated metabolites of polyhalogenated aromatic hydrocarbons reveals alternative mechanism for estrogenic activity of endocrine disrupters. *J Clin Endocrinol Metab* 87: 1142–1150.
  38. Sakakibara Y, Suiko M, Pai TG, Nakayama T, Takami Y, et al. (2002) Highly conserved mouse and human brain sulfotransferases: molecular cloning, expression, and functional characterization. *Gene* 285: 39–47.
  39. Glatt H, Meinel W (2005) Sulfotransferases and acetyltransferases in mutagenicity testing: Technical aspects. *Methods Enzymol* 400: 230–249.
  40. Todd AE, Orengo CA, Thornton JM (2001) Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol* 307: 1113–1143.
  41. Allali-Hassani A, Campbell TL, Ho A, Schertzer JW, Brown ED (2004) Probing the active site of YjeE: A vital *Escherichia coli* protein of unknown function. *Biochem J* 384: 577–584.
  42. Schrag ML, Cui D, Rushmore TH, Shou M, Ma B, et al. (2004) Sulfotransferase 1E1 is a low Km isoform mediating the 3-O-sulfation of ethinyl estradiol. *Drug Metab Dispos* 32: 1299–1303.
  43. Gamage N, Barnett A, Hempel N, Duggleby RG, Windmill KF, et al. (2006) Human sulfotransferases and their role in chemical metabolism. *Toxicol Sci* 90: 5–22.
  44. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14: 755–766.
  45. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, et al. (2004) The Pfam protein families database. *Nucleic Acids Res* 32: D138–141.
  46. Laskowski RA (1995) Surfnet—A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph* 13: 323–330.
  47. Arun KS, Huang TS, Blostein SD (1987) Least-squares fitting of 2 3-D point sets. *IEEE Trans Patt Anal Mach Intell* 9: 699–700.
  48. Bron C, Kerbosch J (1973) Algorithm 457: Finding all cliques of an undirected graph. *Commun ACM* 16: 575–577.
  49. Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics* 15: 327–332.
  50. Everitt B (2001) Cluster analysis of subjects, hierarchical methods. In: Armitage P, Colton T, editors. *Encyclopaedia of biostatistics*: John Wiley & Sons. pp. 713–731.