Enhancing the Steroid Sulfatase Activity of the Arylsulfatase from *Pseudomonas aeruginosa*

Dimanthi R. Uduwela,[1] Anna Pabis,[2] Bradley J. Stevenson,[1] Shina C. L. Kamerlin,[2,#] Malcolm D. McLeod[1*]

[1] Research School of Chemistry, Australian National University, Canberra, ACT 2602, Australia

[2] Department of Cell and Molecular Biology, Uppsala University, S-751 24 Uppsala, Sweden

[#] Current address: Department of Chemistry – BMC, Uppsala University, BMC Box 576, S-751 23 Uppsala, Sweden.

[*] corresponding author: malcolm.mcleod@anu.edu.au

Abstract

Steroidal sulfate esters play a central role in many physiological processes. They serve as the reservoir for endogenous sex hormones and form a significant fraction of the steroid metabolite pool. The analysis of steroid sulfates is thus essential in fields such as medical science and sports drug testing. Although the direct detection of steroid sulfates can be readily achieved using liquid chromatography-mass spectrometry, many analytical approaches, including gas chromatography-mass spectrometry, are hampered due to the lack of suitable enzymatic or chemical methods for sulfate ester hydrolysis prior to analysis. Enhanced methods of steroid sulfate hydrolysis would expand analytical possibilities for the study of these widely occurring metabolites. The arylsulfatase from *Pseudomonas aeruginosa* (PaS) is a purified enzyme capable of hydrolysing steroid sulfates. However, this enzyme requires improvement to hydrolytic activity and substrate scope in order to be useful in analytical applications. These improvements were sought by applying semi-rational design to mutate amino acid residues neighbouring the enzyme active site. Mutagenesis was implemented on both single and multiple residue sites. Screening by UPLC-MS was performed to test the steroid sulfate hydrolysis activity of these mutant libraries against testosterone sulfate. This approach revealed the steroid sulfate binding pocket and resulted in three mutants that showed an improvement in catalytic efficiency (V_{max}/K_M) of more than 150 times that of wild-type PaS. The substrate scope of PaS was expanded and a modest increase in thermostability was observed. Finally, molecular dynamics simulations of enzyme-substrate complexes were used to provide qualitative insight into the structural origin of the observed effects.

Keywords

steroid sulfate, sulfate ester, arylsulfatase, steroid sulfatase, *Pseudomonas aeruginosa*, enzyme promiscuity, enzyme engineering, molecular dynamics

1 Introduction

Steroids and their derivatives comprise a large family of compounds with many roles in biology and medicine. Steroid compounds abound as lipids, hormones and natural products. Their importance is reflected in many therapeutic interventions that target steroidal biosynthesis or signalling, and the continued prevalence of steroid abuse as a major problem for world sport and society at large. Within the larger steroid family, steroid sulfate esters play a central role, serving as the reservoir for endogenous sex hormones and forming a significant fraction of the steroid metabolite pool.[1] Therefore, the analysis of steroid sulfates is essential in many fields of scientific endeavour including medical science and sports drug testing. Recent developments in anti-doping science include a growing emphasis on the analysis of steroid sulfate metabolites. The analysis of steroid sulfates has been shown to increase the detection times for a range of steroid agents[2-8] and can also be used to distinguish between endogenous steroids and otherwise identical exogenously administered steroids of synthetic origin.[4,9-13]

Although the direct detection of steroid sulfates can be readily achieved using liquid chromatography-mass spectrometry in combination with electrospray ionisation,[14] many analytical approaches are hampered due to the lack of suitable enzymatic or chemical methods for sulfate ester hydrolysis prior to analysis.[15] The hydrolysis of steroid sulfates is frequently required to confirm the identity of steroidal analytes against available reference materials, or to meet stringent sample preparation requirements for techniques such as gas chromatography-isotope ratio mass spectrometry. Acid catalysed solvolysis is a general method of hydrolysis for both steroid sulfate and glucuronide phase II conjugates. However, this method is often unsuitable for steroid analysis since it can degrade analytes of interest and generate a more complicated analytical matrix.[15,16] For these reasons, the mild and selective hydrolysis of steroid glucuronides is typically performed enzymatically using *E. coli* β-glucuronidase[15] and this process is mandated within the steroid module of the Athlete Biological Passport developed by the World Anti-Doping Agency.[17] In a similar manner, general methods for the mild and selective enzyme hydrolysis of sulfate esters would provide significant benefits for the analysis of steroid sulfates.

Arylsulfatase enzymes (EC 3.1.6.1) are present in all kingdoms of life and catalyse the hydrolysis of a diverse range of sulfate esters.[18] The nomenclature reflects the relative ease with which aromatic sulfate esters are hydrolysed. However, alkyl sulfate esters such as steroid sulfates are also substrates for this enzyme family. The arylsulfatases make use of an active site metal ion and a rare post-translational modification, where a cysteine or a serine residue is oxidised to formyl glycine (FGly) mediated by separate families of formyl glycine generating enzymes.[19] The signature sequence, C/S-X-P-X-R is conserved throughout the entire enzyme class and is important in directing the post-translational modification. Mechanistic proposals for sulfate ester hydrolysis[18,20,21] typically employ the post-translationally modified FGly in its hydrated form (FGH) as the catalytic nucleophile responsible for sulfate ester S-O bond cleavage. The complete catalytic cycle is proposed to involve a transesterification-elimination-hydration sequence.

Several sulfatases are commercially available, but none of them offer a broad range of activity for steroid sulfate hydrolysis.[22] The partially purified sulfatase preparation from *Helix pomatia* (HpS) offers the greatest range of steroid sulfatase activity, but does not consist of a single known sulfatase enzyme[23] and has been reported to display alternative enzyme activities.[15] In contrast, the *Pseudomonas aeruginosa* arylsulfatase (PaS) is an easily purified bacterial enzyme that has comparable steroid sulfatase activity (Figure 1) with the commercially available HpS.[22,24] However, the slow rates of the hydrolysis of β-configured steroid sulfates and almost negligible sulfatase activity with α-configured steroid sulfates indicate that PaS is not a universal steroid sulfatase suitable for analytical applications. Despite this, PaS represents a good starting point for engineering a steroid sulfatase since it is efficiently produced in *E. coli* as a highly stable enzyme[22] and has been studied by X-ray crystallography.[21] Furthermore, PaS has been reported as a catalytically promiscuous enzyme,[25-

[27] an attribute which has been argued to be important in steering the functional evolution of enzymes.[28-30]
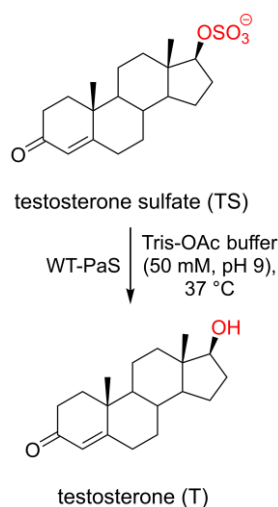

Figure 1: Hydrolysis of testosterone sulfate (TS) to testosterone (T) by wild-type *Pseudomonas aeruginosa* arylsulfatase (WT-PaS).

This study aimed to improve the catalytic activity of PaS for the hydrolysis of steroid sulfates such as testosterone sulfate (TS) and to improve the substrate scope while maintaining the stability of the enzyme. We followed a semi-rational approach where the residues were selected based on examination of the published crystal structure.[21] Five residues were initially subjected to saturation mutagenesis followed by mutating small groups of neighbouring residues simultaneously. Beneficial mutations were then shuffled to find PaS variants with improved efficiency for TS hydrolysis. Improved mutants also displayed expanded substrate scope and, in some cases, greater enzyme thermostability. Finally, molecular dynamics simulations of enzyme-substrate complexes were employed to qualitatively understand the underlying structural basis for the observed improvements. In combination, these data provide a baseline for the future engineering of PaS to enhance its activity towards other steroid sulfates.

2 Results

Our effort to generate an enzyme with improved catalytic efficiency for the hydrolysis of testosterone sulfate (TS) was based on the crystallographic structure resolved by Boltes *et al.* (PDB ID: 1HDH).[21] The catalytic formylglycine hydrate and calcium ion sit at the deepest part of a long and irregularly-shaped cleft (Figure 2). Visual inspection of this cleft revealed several residues that defined a narrow entry zone for potential steroid sulfate substrates. This included K330, one of several residues that was not fully resolved in the reported X-ray structure. Bulky residues that restricted access to the catalytic site were suggested as ideal targets for mutagenesis. To increase the chance of finding useful mutations, residues that were known to be involved in the catalytic mechanism or were highly conserved were excluded. Residue conservation was evaluated using the ConSurf web-based server to analyse a multiple sequence alignment of nearly 150 distinct sulfatase primary sequences (Supporting Information S1.1.1).[31,32] As a result of these inspections five residues spanning the active site cleft were selected for mutagenesis: R155, F328, F331, F463, and K330 (Figure 2).
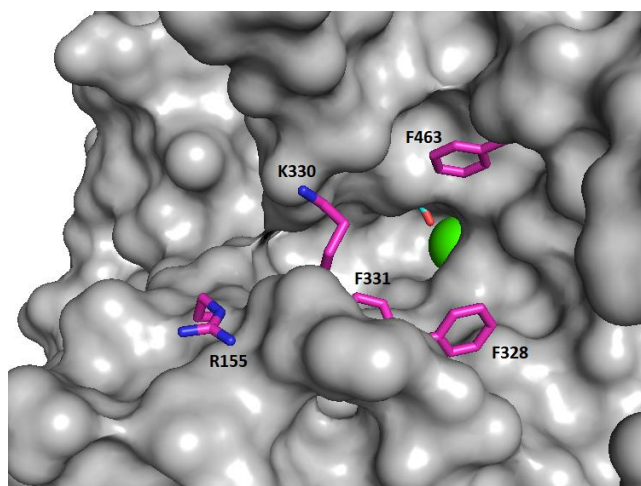
Figure 2: Molecular surface of the PaS X-ray crystal structure (PDB ID: 1HDH),[21] showing the active site cleft. The catalytic formylglycine hydrate residue (sticks, obscured) and calcium ion (sphere) are shown in green. Residues selected for mutagenesis in the active site cleft are labelled and highlighted as magenta sticks. The K330 sidechain that was not resolved in the X-ray crystal structure was introduced using the PyMOL mutagenesis wizard for visualization purposes.

2.1 Site saturation mutagenesis (SSM) libraries

The five residues shown in Figure 2, F328, F463, F331, K330 and R155, were mutated in turn (S1.1.2) using the primers listed in the supporting information (Table S1, Table S2), to produce five site-saturation mutant libraries (SSM 1-5). For each SSM library, and subsequent libraries, variants were randomly selected to evaluate the diversity of codons by DNA sequencing. The number of mutants to be screened for TS hydrolysis from each SSM library was calculated such that each of the possible mutant codons had more than 95% probability of being tested (Table S3). To maximise the likelihood of obtaining improved mutants for steroid sulfate hydrolysis, a medium throughput 96 well plate-based UPLC-MS screening strategy was used to monitor hydrolysis of TS to testosterone (S1.2, S1.3). Alternative high-throughput screens based on the hydrolysis of chromogenic or fluorogenic substrates, such as *para*-nitrophenyl sulfate (pNPS) or 4-methylumbelliferone sulfate, would be unsuitable for finding the desired improvements. Hydrolysis of these substrates generates delocalised aromatic oxy-anions that are stable (and so more easily detected) and would not necessary require general acid catalysis during hydrolysis. Kinetic and computational evidence suggest the hydrolysis of pNPS is not mediated by a general acid.[33,34] This is contrary to steroid sulfates, that would require a general acid to generate an aliphatic alcohol leaving group for efficient hydrolysis. In this way, TS was an ideal substrate for screening efforts: it represents the most challenging steroid sulfate for wild-type PaS hydrolysis where the product can be readily quantified. This provided clear scope for the catalytic improvement of PaS.

Following LC-MS screening, two of the five SSM libraries yielded three PaS variants with improved TS hydrolysis rates: proline or threonine at position 155 (R155P or R155T) or valine at 330 (K330V). In contrast with residues at 155 and 330, mutating the three phenylalanine residues at 328, 331 and 463 did not yield any beneficial mutants.

2.2 Combined mutants

Two PaS variants were prepared (S1.1.3) to test the effect of combining the beneficial mutations, R155P/K330V (PV-PaS) and R155T/K330V (TV-PaS). Each of the single and double mutants, together with wild-type PaS (WT-PaS), were purified (S1.4) with hexa-histidine C-terminal tags by Ni-NTA affinity chromatography to study enzyme kinetics (S1.5.1, Figure S1, Table 1). Initial rate kinetic data was fitted to the Michaelis-Menten equation for a scheme involving binding and catalysis.

All improved PaS variants exhibit a significant increase in V_{max} value with the two double mutants exhibiting more than three times wild-type V_{max} value (Table 1). No improvements were observed in K_M value as a result of single mutations at 155, although K330V significantly increased K_M. Mutations of both sites together showed that their effects on TS hydrolysis kinetics were not independent; they were synergistic in accordance with their close proximity in the crystal structure.[21] The PV-PaS double mutant showed a significant improvement in substrate affinity with 62% of the wild-type K_M, whereas TV-PaS was the opposite with approximately twice the K_M value of wild-type. Overall, PV-PaS offered the best result from the first set of mutations with the highest V_{max}/K_M value: 5.6 times that of WT-PaS. Therefore, the PV-PaS double mutant was selected to be carried forward as template for the following rounds of engineering.

Table 1: Kinetic parameters for TS hydrolysis by selected PaS variants. The respective mutations, the library they were found in, template used and kinetic parameters determined over an initial TS concentration from 2.5 – 320 µM at pH 9 (50 mM Tris acetate) and 37 °C.

PaS enzyme	Mutational residue							Library	Template	Kinetic parameters[a]		
	155	156	157	158	160	325	330			K_M [µM]	V_{max} [10^{-2} µmol min^{-1} g^{-1}]	(V_{max}/K_M) [10^{-3} L g^{-1} min^{-1}]
WT-PaS	R	I	L	K	T	L	K	-	-	35 ± 5	9.7 ± 0.4	2.8 ± 0.4
V-PaS	R	I	L	K	T	L	V	SSM 4	WT-PaS	53 ± 4	16.3 ± 0.4	3.1 ± 0.3
P-PaS	P	I	L	K	T	L	K	SSM 5	WT-PaS	36 ± 4	29 ± 1	8.2 ± 1.1
T-PaS	T	I	L	K	T	L	K	SSM 5	WT-PaS	37 ± 3	24.0 ± 0.6	6.5 ± 0.5
PV-PaS	P	I	L	K	T	L	V	-	-	22 ± 1	35.3 ± 0.6	16 ± 1
TV-PaS	T	I	L	K	T	L	V	-	-	88 ± 10	47 ± 2	5.4 ± 0.6
PVFV-PaS	P	V	L	K	T	F	V	MSM 1	PV-PaS	38 ± 2	343 ± 28	91 ± 9
PVIV-PaS	P	I	V	I	T	L	V	MSM 2	PV-PaS	4.5 ± 0.2	207 ± 7	463 ± 26
VVIV-PaS	R	V	V	I	T	L	V	Shuffled	WT- or PV-PaS	9.6 ± 0.5	411 ± 19	427 ± 30
LEF-PaS	R	L	L	E	T	F	K	Shuffled	WT- or PV-PaS	17 ± 1	743 ± 35	427 ± 32

[a] These estimates are the mean ± the standard deviation from three experiments.

2.3 Multiple site mutagenesis (MSM)

The initial mutagenesis work had succeeded in delivering modest improvements by mutating residues individually. However, the greatest improvements were afforded by finding the best combination of the two mutations. The synergy between mutations is regularly observed by protein engineers and is often accommodated in experiment design. Simultaneous mutation of residues has featured in protocols such as the "combinatorial active-site saturation test"[35] or "combinatorial codon mutagenesis"[36] The success of mutagenesis lies in obtaining sufficient amino acid diversity at the targeted site so that sequence space is adequately covered. Nevertheless, due to the limitations inherent in UPLC-MS screening, it was necessary to compromise coverage to limit library size. The basis for our reduction in available combinations for MSM was to target a range of hydrophobic residues in order to increase the favourable interactions with a hydrophobic substrate (NYS codon, N=A/T/G/C, Y=C/T, S=G/C). Assuming full coverage, this degeneracy allowed for all hydrophobic residues except tryptophan and glycine, and included serine and threonine. Tryptophan was not of interest as it is a bulkier residue than the original residues. However, the option for glycine was sacrificed in order to use a simple degenerate codon for library preparation. Using this approach an acceptable 80% coverage for the amino acid diversity was predicted for primary screening of the MSM 1 and MSM 2 libraries (Table S3). In contrast >95% coverage was expected for primary screening of the MSM 3 and shuffled libraries.

2.3.1 MSM 1

The first set of residues for multiple site mutagenesis (MSM 1) were: I156, L325, and F331 (S1.1.4, Table S3). They were selected as neighbours of the successfully mutated sites (R155 and K330) and because they appeared to project into the substrate binding pocket (Figure 3A). The best mutant in hand, PV-PaS, was used as the template for this library of variants. One residue, F331, had already been mutated (SSM 3) in isolation without success, but it was hypothesised that simultaneous mutations at neighbouring residues on top of the mutations in the template PV-PaS might support a useful change.

Primary screening of 1728 colonies was performed with 100 µM TS and those that gave rise to greater testosterone release than PV-PaS were tested in the secondary screen (92 clones). Tests with purified enzymes confirmed that four beneficial PaS variants had greater activity than template PV-PaS (Table S4). The kinetic parameters of the best mutant PVFV-PaS (Table 1), containing mutations I156V and L325F in addition to R155P and K330V, showed a V_{max}/K_M value for TS that was 33 times that of WT-PaS. All beneficial mutants observed in this library retained the original F331.
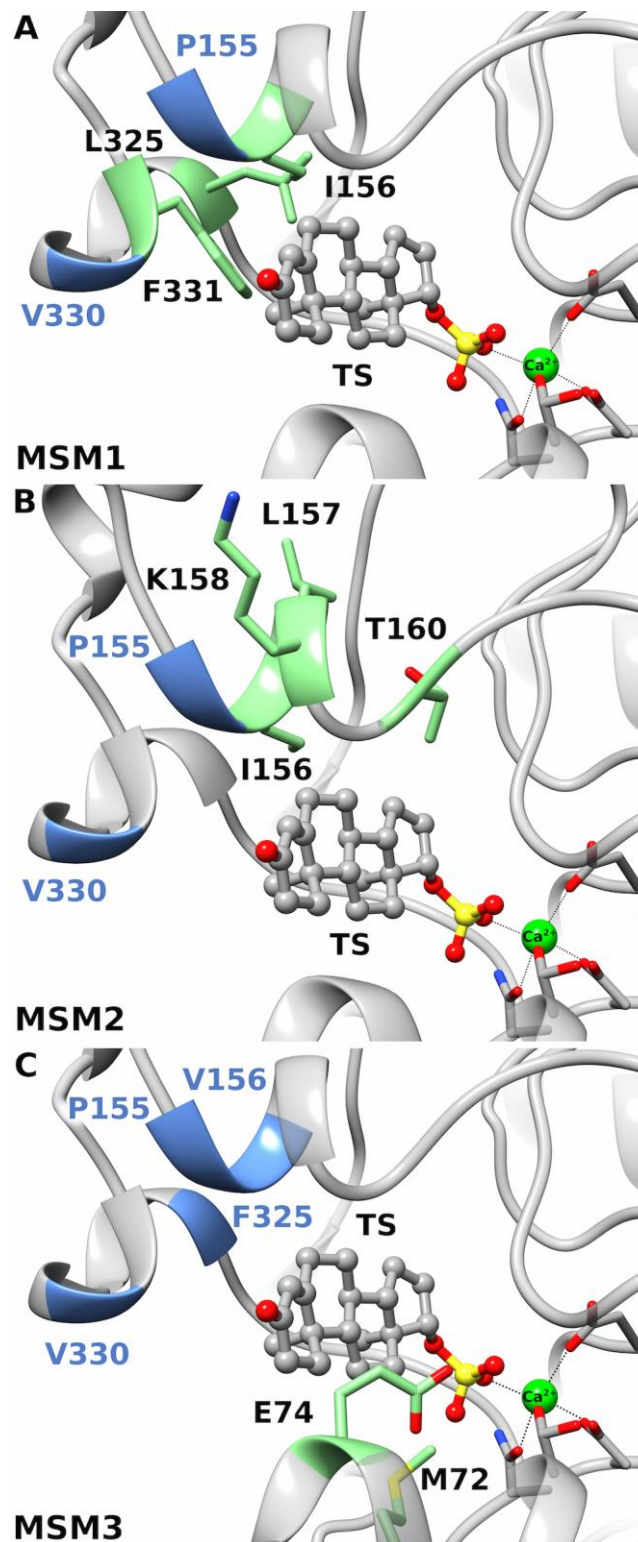
Figure 3: The residues selected for mutagenesis in each MSM library are shown with side chains as green sticks and labelled in black for MSM 1 (A), MSM 2 (B), MSM 3 (C). The substrate (TS) is shown in dark grey balls and sticks. The models are starting structures for molecular dynamics simulations and are derived from the X-ray crystal structure of wild-type PaS (PDB ID: 1HDH)[21] with TS manually placed in the active site. The active site calcium ion responsible for sulfate ester binding is shown as a green sphere. The locations of template mutations are shown and labelled in blue.

2.3.2 MSM 2

The second round of MSM targeted I156, L157, K158 and T160 (Figure 3B) with PV-PaS as the template (S1.1.5, Table S3). These residues were on the short helix where two other beneficial mutations (155P and 156V) were found. It was reasoned that further optimisation of this helix may be possible by finding the best combination of residues. To minimise library size, residue 159 was not targeted for mutagenesis since this wild-type glycine residue directly faced the active site and an enforced change to larger residues was thus expected to be deleterious. The MSM 2 library was prepared by using the NYS degenerate codon at positions 157, 158, and 160, whereas position 156 was targeted to the three residues found in the best MSM 1 mutants (Table S4). The degenerate codon at position 156 was VTT (where V = A/G/C) to allow for isoleucine, valine or leucine. Greater stringency was applied for the screening of the MSM 2 library by using 20 µM TS; one fifth the previous concentration. This brought the substrate concentration below the wild-type K_M value so as to discover mutants with greater affinity for TS; necessary for the samples of low concentration typically encountered in analytical applications. With primary screening of 4800 colonies and subsequent secondary (118 clones) and tertiary (20 clones) screens, three mutants were identified with greater activity than the PV-PaS template (Table S4). The best mutant, PVIV-PaS, which contained the template mutations together with I156, L157V and K158I, exhibited a V_{max}/K_M value that was 167 times greater than the WT-PaS (Table 1).

2.3.3 MSM 3

The targets for MSM 3 focused on two residues on the opposite side of the binding pocket compared with previous sites: M72 and E74 (Figure 3C). These residues were in close proximity to the catalytic residues and were of interest because they provide some bulk to the substrate binding pocket. In this case PVFV-PaS (from MSM 1) was used as the template and included screening of the possible variations with NYS degeneracy and a > 95% probability of all variants being tested (Table S3). Mutating these two residues (S1.1.5) resulted in a library of inactive PaS variants, suggesting that these residues may directly or indirectly be involved in enzyme catalysis. Importantly, the NYS degenerate codon allowed for methionine, but not glutamate residues, suggesting that it was mutation of E74 that led to the dramatic loss of activity.

2.4 Shuffled library

The beneficial mutations found in the improved PaS variants (Table S4) were shuffled to test for further improvements in catalytic activity (S1.1.6). This involved mutations at seven residues (Table S5), with WT-PaS or PV-PaS as the template. Suitable degenerate codons were available at all positions with the exception of position 158. Here, the RWA degenerate codon included glutamate (E) in addition to the targeted amino acids. From a primary screen of 1536 clones, six PaS variants were discovered with greater initial hydrolysis rates for TS hydrolysis that exceeded PVIV-PaS (Table S4). Surprisingly, two of the top mutants incorporated a K158E substitution. Two mutants, one without K158E substitution, VVIV-PaS, and one with K158E substitution, LEF-PaS, were selected for further characterization. Kinetics studies were performed on these mutants and revealed that both V_{max} and K_M values had increased such that their V_{max}/K_M value was similar to that of PVIV (Table 1).

2.5 The pH dependence

The optimum pH value for TS hydrolysis by WT-PaS was determined over a range from pH 5 to 9 as described in section S1.5.2. The pH profile for TS hydrolysis shows an optimum V_{max}/K_M of pH 7.1 as shown in Figure 4. The K_M for TS showed a broad minimum from pH 6.5 to 8.0 whereas the V_{max} reached an optimum at pH 6.0 to 6.5 (Figure S2).
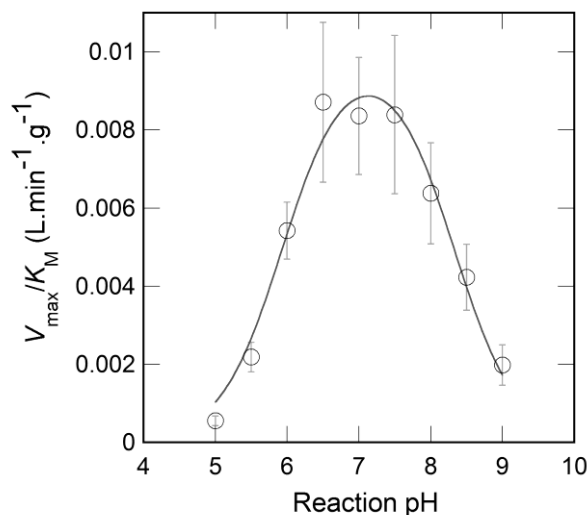
Figure 4: The pH profile of V_{max}/K_M for TS hydrolysis by WT-PaS at 20 °C. Points at each pH show the estimate from fitting the Michaelis-Menten equation to three replicates of eight TS concentrations from 2.5 to 320 µM. The error bars correspond to the combined standard error from estimates of V_{max} and K_M. The curve represents non-linear regression based on a model with two ionisable residues where productive substrate binding occurs when only one group is protonated.[37] The ionisable residue pK_a values are estimated as 5.9 ± 0.1 and 8.3 ± 0.1. The pH profiles for V_{max} and K_M are provided in the supporting information (Figure S2).

2.6 Substrate scope at pH 7.5

The hydrolytic activities of the best mutants from each level of mutagenesis were tested over four steroid sulfates, with the sulfate ester varying in stereochemistry and position (S1.5.3). The four substrates tested were TS, epiandrosterone sulfate (EAS), etiocholanolone sulfate (ECS) and epitestosterone sulfate (ETS) as shown in Figure 5 and free steroids were quantified against two calibrators containing the free steroids. They were tested at pH 7.5, compatible with the conditions typically used by analytical laboratories for enzyme hydrolysis of the glucuronide fraction with *E. coli* β-glucuronidase, and as the enzymes exhibit faster turnover closer to neutral pH for the hydrolysis of steroid sulfates (Figure 4). For this analysis, the enzyme concentration, substrate concentration and the reaction time used for the hydrolysis of α-configured steroid sulfates ECS and ETS was greater, given the low turnover of these substrates by PaS.
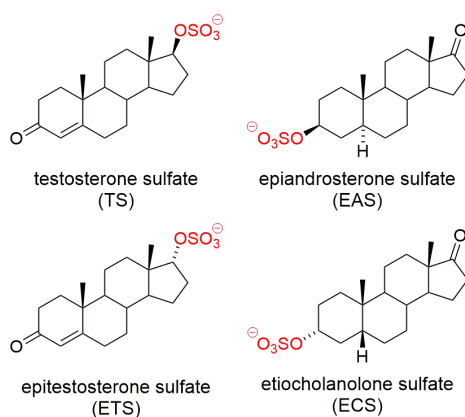

Figure 5: Range of substrates tested.

All mutants exhibited improved hydrolytic activity for TS, the substrate used in screening to uncover improved activity in this the enzyme engineering study (Table 1, Figure 6). The best mutant for TS may

not necessarily be the best for the other steroidal sulfate substrates. The WT-PaS enzyme hydrolyses one order of magnitude more EAS than TS which agrees well with the published results.[22] Considering the low turnover of α-configured steroid sulfates, higher steroid sulfate and enzyme concentrations, and longer reaction times were required to observe detectable activity. As a result, the relative activities of α- and β-configured steroid sulfates may not be directly compared in Figure 6. Apart from TS hydrolysis showing remarkable improvements, ECS hydrolysis too showed enhanced activity with all mutants compared to WT-PaS. The large errors associated with ETS hydrolysis for all but PVFV-PaS (0.094 µM free steroid produced) indicate that the signals are within the level of noise. Taken together the results show that PVFV-PaS displays improved activity for the four substrates tested. Subsequent rounds of mutagenesis led to further improvement but also growing selectivity for the TS substrate used in screening.
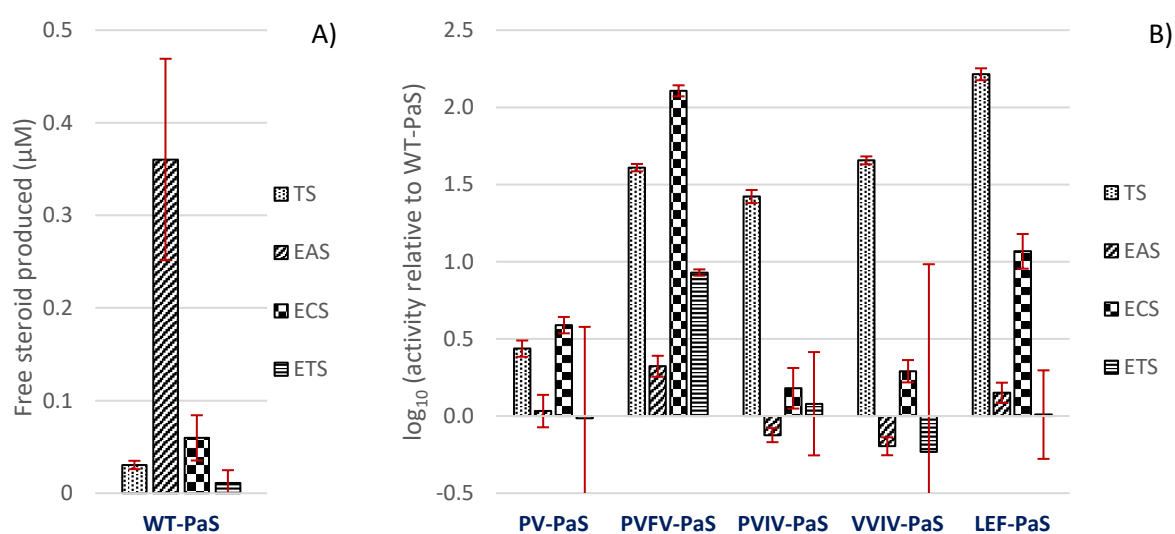

Figure 6: The activity profiles of WT-PaS and PaS mutants (S1.5.3). A) Free steroid produced by WT-PaS. B) The activity profiles for the PaS mutants showing the free steroid produced for each substrate, normalized with respect to WT-PaS and presented on a logarithmic scale. The negative bars represent mutant activities less than the corresponding WT activity. The error bars correspond to the standard deviation from triplicates.

2.7 Effect of temperature on enzyme activity and stability

The thermostability of WT-PaS and the best variants from each library was tested in two ways: determining the temperature that causes 50% irreversible inactivation after 5 min (apparent melt temperature, T_m^{app}) at pH 7.5 or pH 9, or by measuring the half-life at 50 °C ($T_{1/2}^{50\,°C}$) at pH 9 (Table 2, S1.5.4, Figures S3 and S4). All six versions of PaS show higher T_m^{app} at pH 9 compared to that at pH 7.5. PVFV-PaS and LEF-PaS have c. a. 3.5 °C increase over WT-PaS for T_m^{app} at pH 9, but only PVFV-PaS retains improved thermostability at pH 7.5 (c. a. 3 °C higher than WT-PaS). As expected from these results, the PVFV-PaS mutant also displayed a longer half-life at 50 °C relative to WT-PaS (Table 2).

Table 2: Thermal parameters for WT and mutant PaS variants

	WT-PaS	PV-PaS	PVFV-PaS	PVIV-PaS	VVIV-PaS	LEF-PaS
T_m^{app}, pH 7.5 (°C)	51.2 ± 0.3	45.4 ± 0.7	54.3 ± 0.3	48.8 ± 1.4	47.9 ± 0.9	51.4 ± 0.2
T_m^{app}, pH 9 (°C)	56.0 ± 0.1	50.2 ± 0.2	59.4 ± 0.2	53.3 ± 0.2	52.7 ± 0.2	59.5 ± 0.2
$T_{1/2}^{50\,°C}$, pH 9 (h)	0.67 ± 0.04	0.12 ± 0.01	5.8 ± 0.2	0.6 ± 0.2	0.30 ± 0.04	1.2 ± 0.1

2.8 Hydrolysis of 16,16,17-d₃-testosterone sulfate (d₃-TS) in spiked urine

The best TS-hydrolysing PaS variant, LEF-PaS, was tested in comparison with WT-PaS in a urine matrix. Human urine was spiked with 100, 200, or 300 ng mL^{-1} of d₃-TS and incubated for 16 hours at pH 7.5 and 37 °C with 10 µg mL^{-1} of the PaS variants (S1.5.5). The LEF-PaS mutant reached complete conversion to hydrolysed d₃-testosterone whereas WT-PaS conversion was < 2% at all concentrations (Figure 7).
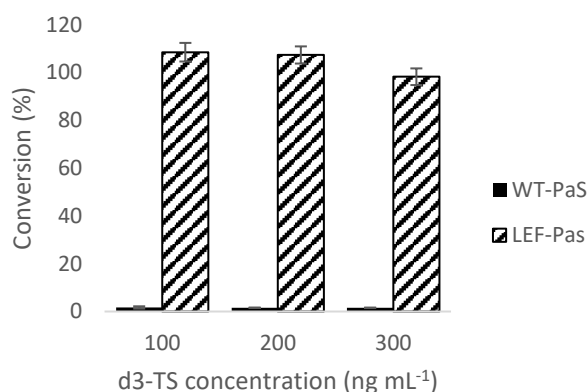

Figure 7: The proportion of d₃TS hydrolysed by WT-PaS or LEF-PaS (S1.5.5). The mean of triplicates is presented with error bars representing standard deviation.

2.9 Molecular dynamics simulations

Molecular dynamics simulations of wild-type and mutant PaS variants in complex with TS provided qualitative insight into the structural origins of the observed changes in activity (S1.6). One challenge when performing simulations of PaS is correctly assigning the protonation states of the key active site residues K113, H115, H211 and K375, as well as the active site FGH nucleophile (Figure S5). The FGH residue has been previously modelled in its anionic form in the related manganese-dependent phosphonate monoester hydrolases from *Burkholderia caryophylli* (BcPMH) and *Rhizobium leguminosarum* (RlPMH),[38] although in the case of these enzymes, the structurally equivalent residues to K113 and H115 in PaS are a Tyr (Y105) and a Thr (T107), respectively, and so the issue of correctly assigning their protonation states does not arise. On the basis of semi-macroscopic pK_a value calculations,[33] a depressed pK_a value for the K113 sidechain had been previously predicted, such that this residue was deprotonated in the resting state of the PaS Michaelis complex with pNPS. A similar prediction was made based on empirical pK_a screening using PROPKA 3.1,[39] using, the high-resolution (1.3 Å) structure of native PaS in complex with SO₄$^{2-}$ (PDB ID: 1HDH).[21] This calculation suggested a slightly elevated pKa value of 7.8 for H115, and a correspondingly depressed pK_a value of 8.4 for K113. When taking into account previous empirical valence bond (EVB) calculations on this system,[33] K113 was proposed as the general base that deprotonates the FGH catalytic nucleophile in the active site. In addition, while H211 has been suggested to act as a general-acid during sulfate transesterification by PaS,[24,25] the estimated pK_a value for this residue was low (0.94). While this pKa prediction is most likely an exaggeration due to the empirical nature of the calculation, nevertheless, a depressed pK_a value is chemically logical given that this residue was located between both a positively charged lysine (K375) and the catalytic calcium ion. Therefore, it was proposed that K375 served as a general acid during the transesterification step, while H211 was present in its mono-protonated form, helping position the general acid K375, as in previous EVB models for both PaS[33] as well as for BcPMH and RlPMH.[38]

To validate these assumptions, 30 ns molecular dynamics simulations were performed of WT-PaS in complex with both a SO₄$^{2-}$ di-anion as in the crystal structure as well as the mono-anionic TS substrate, which was manually built onto the SO₄$^{2-}$ ion in the initial crystal structure. The resulting average values

of S-O_{nuc} distance, O_{nuc}-S-O_{lg} angle, and other key interactions are shown in Table S6. From this table, it can be seen that the models that yield the most stable Michaelis complexes with the closest to in-line geometries for nucleophilic substitution involve deprotonated FGH and H115 side chains, and protonated K113 and K375 side chains. In the case of H211, this residue appears to be preferentially di-protonated in the presence of the SO_4^{2-} di-anion in order to offset the additional negative charge, but preferentially mono-protonated at π-N in the presence of the mono-anionic TS substrate, consistent with the depressed pK_a value of this residue predicted by PROPKA.[39] Based on these molecular dynamics simulations, as well as the insights from previous EVB simulations,[33] a revised mechanism shown in Figure 8 was proposed, and used as a model for subsequent molecular dynamics simulations with TS. Notably, this preferred model (deprotonated FGH/H115/H211, protonated K113/K375) was an extreme representation of the mechanism shown in Figure 8, in which the FGH catalytic nucleophile was fully deprotonated by the K113 general base. This model, involving a single proton transfer between mechanistically coupled FGH and K113 catalytic residues was adopted as it yielded much more computationally stable Michaelis complexes (the average S-O_{nuc} distances and O_{nuc}-S-O_{lg} angles, as well as other key interactions, are shown in Table S6). Once satisfied with the model for the Michaelis complex between PaS and the TS substrate, long-timescale molecular dynamics simulations of the Michaelis complexes of the wild-type enzyme, as well as the TV, PV, PVFV, LEF, PVIV and VVIV, variants were performed. A comparison of per-residue root mean square fluctuations (RMSF) of all backbone heavy atoms in our simulations of each enzyme variant is shown in Figures S6 and S7, and an analysis of all protein-substrate contacts is shown in Figures S8 and S9. Representative structures from each simulation have been uploaded to Dryad (available for download at DOI: To be provided after publication or prior to publication on request), so that interested readers can view the associated structural changes in 3D.
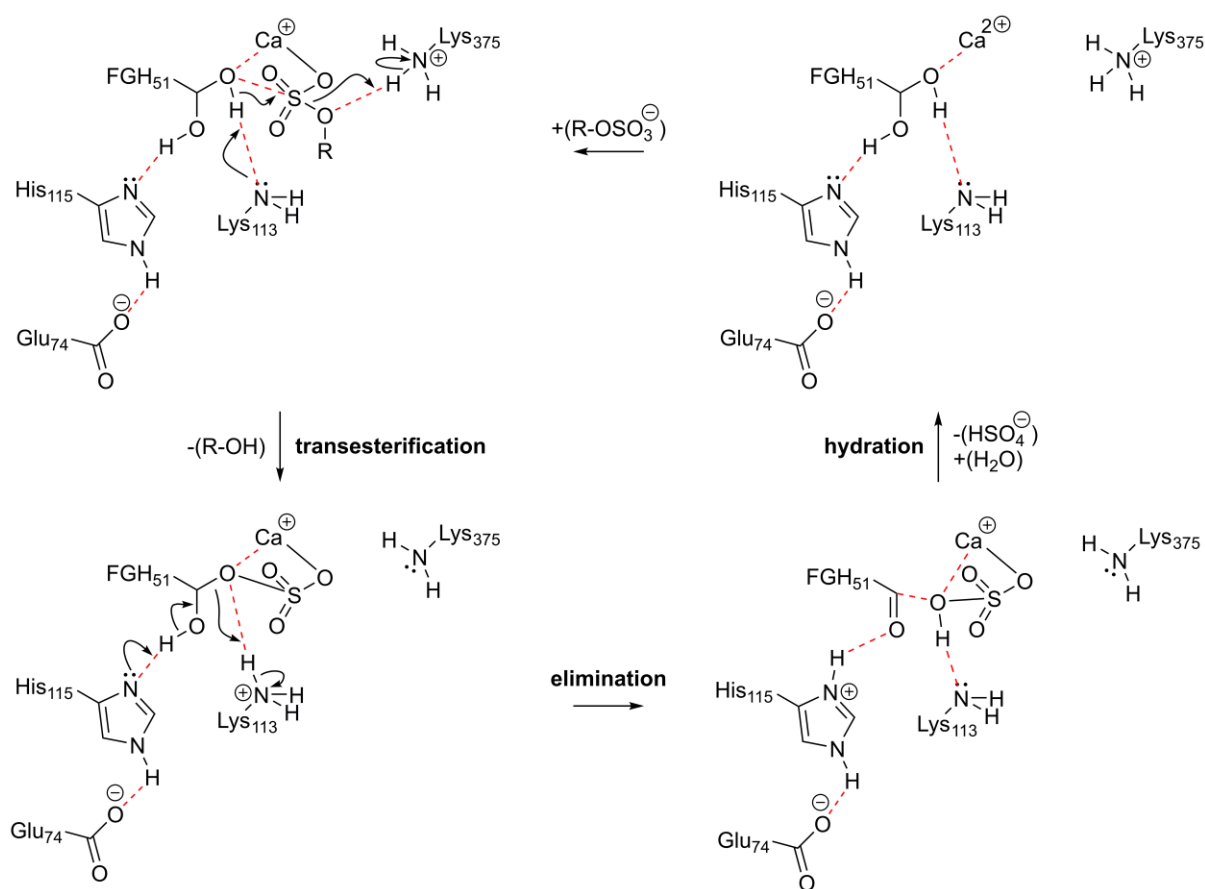

Figure 8: The proposed transesterification-elimination-hydration of PaS, highlighting the role of E74 and H115 during elimination of the sulfated FGH intermediate.

2.9.1 TV-PaS, PV-PaS and WT-PaS

For simplicity, the different variants were analyzed in pairs reflecting the step-wise evolutionary pathway associated with their discovery. The first such comparison involved TV-PaS and PV-PaS derived from combination of the beneficial mutations identified from site saturation mutagenesis (section 2.2). From the RMSF analysis, it was observed that TV-PaS was quite similar in flexibility to the wild-type enzyme (Figures S6 and S7). As shown in Table 1, TV-PaS showed an increase in the V_{max} value, that is largely offset by an increase in the K_M value. The lower affinity could, in part, be due to the loss of favorable enzyme-substrate interactions with K330 in the K330V mutation.

In contrast, the PV-PaS variant was more flexible between residues 155-165 and 320-350, which included the R155P and K330V mutations (Figures S6 and S7). The importance of flexibility in enzyme evolution has been demonstrated in a number of previous computational studies,[38,40] and this increased flexibility could here again be linked to the lower K_M and higher V_{max} values of this mutant, as it allows for better adaptation of the active site to the bulky steroid substrate. More substantial structural changes were also observed for PV-PaS involving substrate displacement and reshaping of the active site. Slightly increased enzyme-substrate contacts were observed between TS and K375 and A376. However, the most notable change in PV-PaS, compared to the wild-type, involved residues L324 and L325. Neither of these residues made contact with the substrate directly in WT-PaS or TV-PaS (Figures S8 and S9), but in PV-PaS there were structural adaptations that allow for interactions with L324 and L325. This led to the formation of a "hydrophobic wall" on one side of the substrate, which blocked solvent penetration into the active site, as can be seen from comparison of the water densities and number of water molecules around the substrate (Figure 9, Figure S10). This could explain the reduction in K_M value shown in Table 1 for PV-PaS compared to WT-PaS. In addition, the increased flexibility of PV-PaS mentioned above was accompanied by generally higher structural variability of the active site residues and substrate position, and could be linked to the increased catalytic efficiency observed.
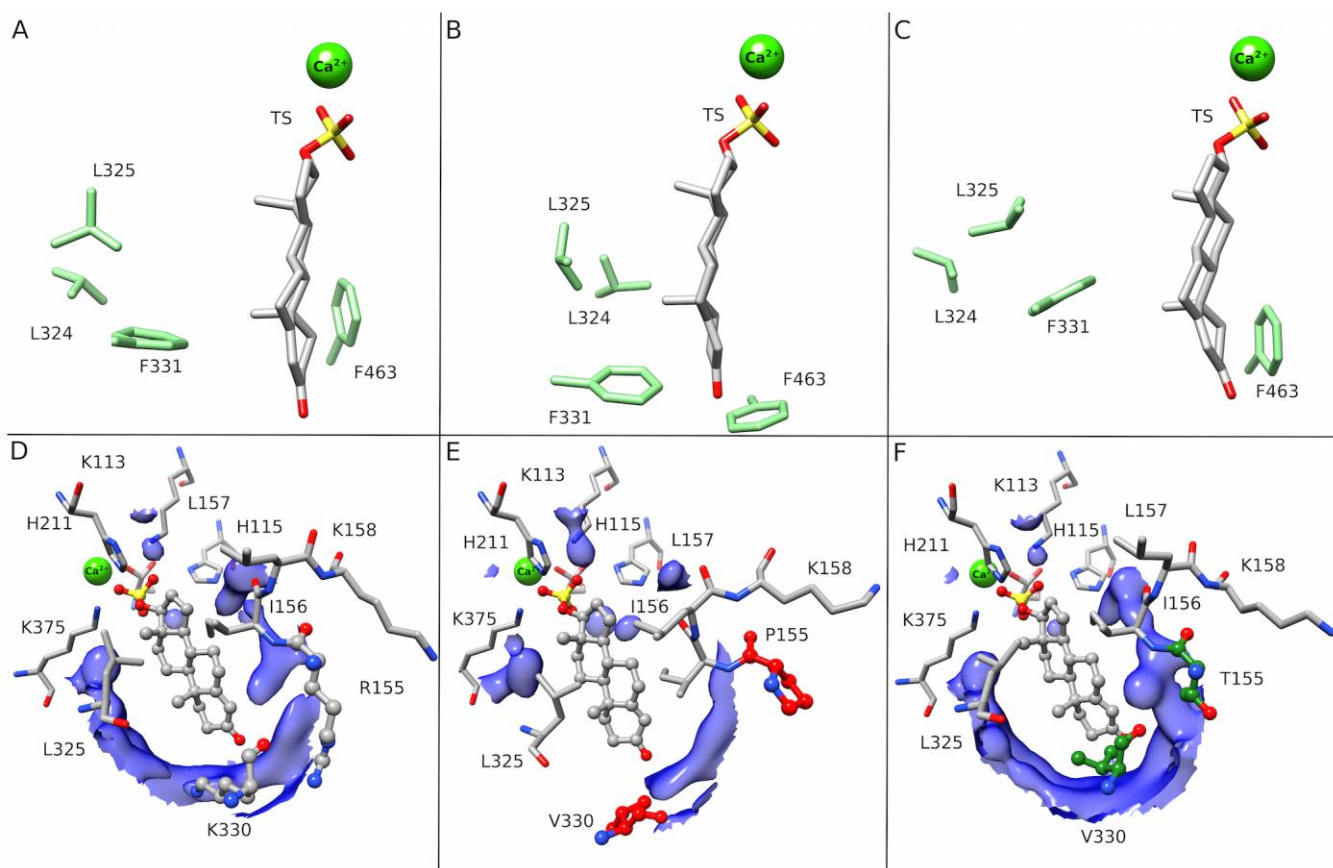
Figure 9: Upper panel: key structural differences between the active sites of wild-type PaS (A), PV-PaS (B) and TV-PaS (C) observed in the MD simulations of the three variants. Bottom panel: water density within 5Å of the substrate molecule (blue surface) in wild-type PaS (D), PV-PaS (E) and TV-PaS (F) showing a hydrophobic wall formed in PV-PaS as a result of rearrangement of L324, L325 and F331 shown in the upper panel.

2.9.2 PVFV-PaS, PVIV-PaS and PV-PaS

Following from this, the PVFV and PVIV mutants were compared to PV-PaS, as these variants were constructed by MSM based on the PV-PaS template (section 2.3.1 and 2.3.2). These two variants displayed significant differences in activity, with PVFV showing 5.8-times higher V_{max}/K_M value, and PVIV-PaS showing 30-times higher V_{max}/K_M value, than PV-PaS (Table 1). Note however that these improvements were underpinned by changes in K_M, which was slightly increased in PVFV-PaS, and more significantly decreased in PVIV-PaS, compared to PV-PaS. A comparison of the overall flexibility of the protein (Figures S6 and S7) shows, once again, changes in residues 155-165 and 320-350. Compared with PV-PaS, the residues in the 320-350 region become more flexible in PVFV-PaS and both regions become more rigid in PVIV-PaS. Interestingly, a comparison of the active site volumes of the different variants (Table S7 and Figure S11) showed that both PVFV-PaS and PVIV-PaS had larger active site volumes than PV-PaS, but all three variants had smaller active site volumes than WT-PaS. Notably, the effect of the "hydrophobic wall" observed in PV-PaS (Figure 9) had been lost or diminished in both mutants, and in the case of PVFV this could explain why the K_M value increased to values closer to that of WT-PaS (Table 1). In terms of active site flexibility (as expressed by the standard deviations of the calculated volumes), PV-PaS and PVFV-PaS had a more flexible active site than PVIV-PaS, and both variants were more flexible than WT-PaS (Figure S11 and Table S7).

In PVFV-PaS the I156V mutation introduced a smaller amino acid sidechain and allowed the short helical region from 155-159 to approach the substrate more closely (Figures S8 and S9). The L325F mutation provided additional hydrophobic contacts with neighboring residues W212, W358 and the

newly introduced L156V (Figure 10). The PVFV-PaS mutant retained the high mobility of the PV-PaS parent (Figures S6 and S7) but was also (along with LEF PaS below) associated with higher thermostability (Table 2), in contrast with modest decreases in thermostability for PV-PaS, PVIV-PaS, and other variants. Thus, the L325F mutation provided additional stabilization for the flexible PVFV-mutant. It should be noted that an inverse relationship between flexibility and thermostability may seem counterintuitive; however, both experimental and computational studies on several protein systems indicate that high stability and enhanced conformational flexibility are not necessarily mutually exclusive.[41] The PVFV mutant also displayed the broadest substrate scope on the evolutionary pathway, with activity for the hydrolysis of α-configured steroid sulfates ESC and ETS (Figures 5 and 6) that was not observed for WT-PaS. There appears to be a correlation between active site volume, polar residues in the active site and active site flexibility, and the number of observed activities in alkaline phosphatases.[38] It is therefore plausible that this increased flexibility leads to the broader substrate scope observed for the PVFV variant of PaS, that allows it, for example, to accommodate ECS (Figures 5 and 6).
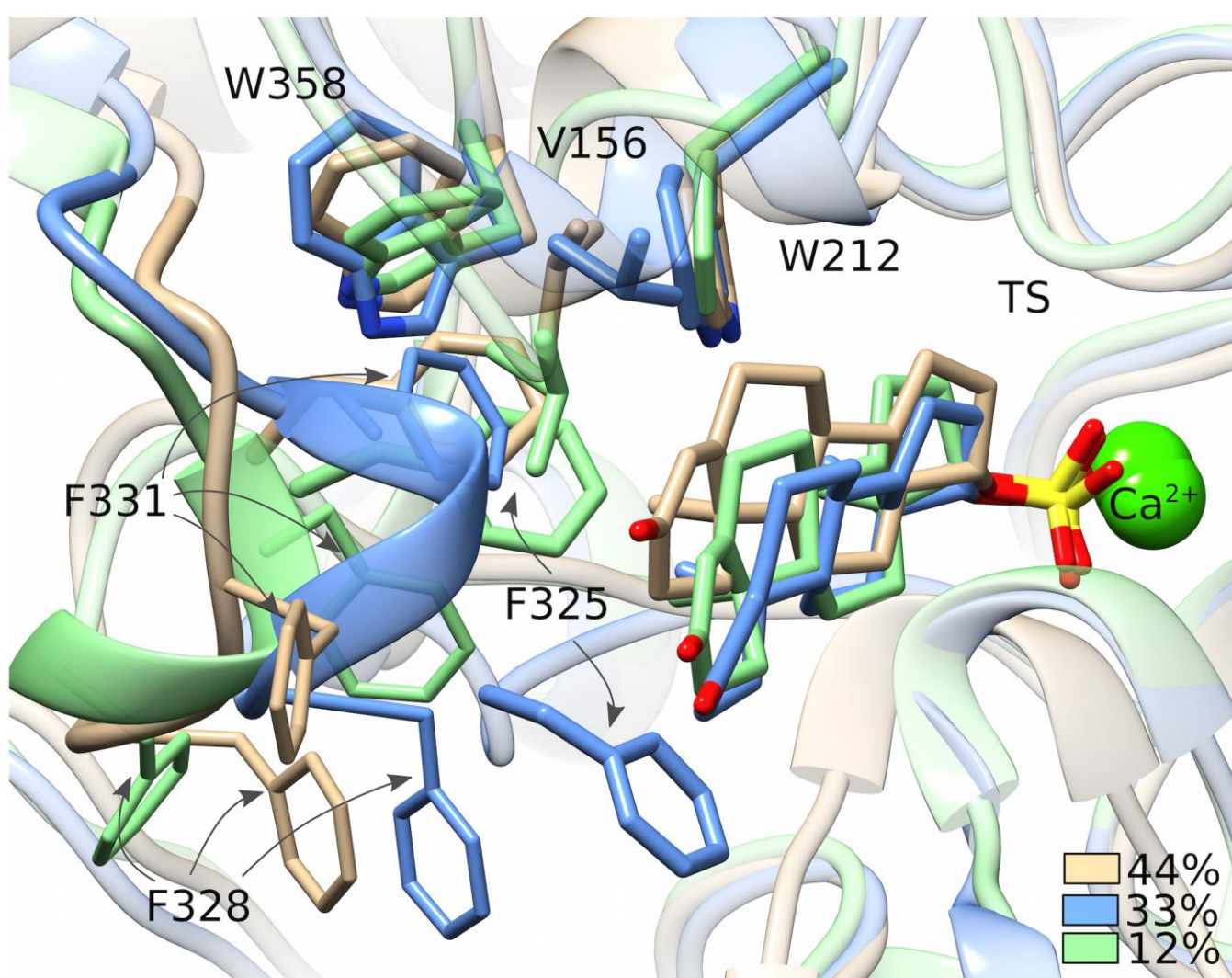

Figure 10: Overlay of three representative structures from the molecular dynamics simulation of PVFV-PaS in complex with TS. The structures depict the location of L325F in PVFV-PaS and its interactions with the neighbouring aromatic residues, specifically W212 and W358 observed in the majority of the simulation. The percentage values indicate the fraction of the simulation for which each structure is representative, as determined through RMSD-based clustering with average-linkage algorithm.

In PVIV-PaS, the introduction of the β-branched residues, L157V and K158I mutations, led to reduced flexibility of the 155-165 and 320-350 regions (Figures S6 and S7), in the presence of the flexibility-inducing R155P change. In addition, the K158I mutation eliminated a salt bridge between K158 and E151 located on the loop adjacent the alpha-helix containing the 158 residue. Associated with this, a slight repositioning of K330V and F331, leading to enhanced interactions with the substrate (Figures S8 and S9) that could in turn lead to the reduction in K_M value observed for this variant. This mutant displayed a dramatic increase in the V_{max} value and a significant decrease in K_M value for TS hydrolysis (Table 1). However, in comparison to the PV-PaS template, the PVIV-PaS mutant was more specialized and showed reduced EAS and ECS hydrolysis activity (Figures 5 and 6).

2.9.3 LEF-PaS and VVIV-PaS

A final comparison was between VVIV-PaS and LEF-PaS, which were identified from a shuffled library of all beneficial mutants identified in earlier rounds (Section 2.4). Both mutants showed similar efficiency (~150 times WT-PaS) for TS hydrolysis to PVIV-PaS with modest increases in V_{max} value offset by increases in K_M value (Table 1). The two mutants were distinguished by different backbone flexibility, with VVIV-PaS showing a similar RMSF profile to WT-PaS and PVIV-PaS (Figures S6 and S7). In contrast, LEF-PaS showed slightly increased flexibility than WT-PaS in residues 155-165 and 320-350, a pattern that was similar to, but attenuated from, that observed for PV-PaS and PVFV-PaS. Greater protein flexibility was also reflected in slightly higher variability of the active site volume for LEF-PaS compared to VVIV-PaS (Table S7 and Figure S11). Both proteins showed similar average active site volumes that were equal or slightly smaller than WT-PaS. Interestingly, the original combined mutation R155P reverted to R155 in both VVIV-PaS and LEF-PaS leading to increased interactions of this residue with the TS substrate in LEF-PaS (Figures S8 and S9). The LEF-PaS mutant was also observed to lose the K330V change introduced in the combined mutant.

For LEF-PaS the I156L change was observed as beneficial in the MSM 1 and MSM 2 libraries, and the L325F mutation appeared in PVFV-PaS. Both positions were closely associated and packed together in a hydrophobic region of the protein adjacent to the active site. The L325F mutation (Figure 10) was previously associated with higher thermostability for PVFV-PaS that is also observed for the LEF-PaS variant (Table 2). Interestingly, LEF-PaS included the K158E mutation that was not previously identified as beneficial but arose by serendipity through the adoption of the RWA degenerate codon at this position. The K158E mutation eliminated the salt bridge between K158 and E151 located on the loop above the alpha-helix containing the 158 residue, resulting in the movement of the alpha-helix toward the substrate. The K158E residue was also observed (along with E74) to interact with R155, thereby forming a lid over the active site of the enzyme. Overall, the LEF-PaS mutant most closely resembled PVFV-PaS in terms of structure, flexibility and activity profile. In comparison to PVFV-PaS that showed the broadest substrate range, LEF-PaS showed greater specialization with enhanced TS hydrolysis but lower activity towards other substrates (Figures 5 and 6). This change in substrate profile can possibly be traced to the attenuated flexibility of LEF-PaS relative to PVFV-PaS (Figures S6 and S7).

The VVIV-PaS shuffled mutant retained the L157V, K158I and K330V mutations of PVIV-PaS and varied only in the reversion of R155P to R155 discussed above and the introduction of an I156V change. The I156V mutation reduced contact of this residue with the TS substrate (Figures S8 and S9). In common with PVIV-PaS, the K158I mutation resulted in removal of the salt bridge between K158 and E151 located on the loop adjacent the alpha-helix containing the 158 residue, which is accompanied by a repositioning of V330 and F331 residues towards the substrate (Figures S8 and S9). Overall, the VVIV-PaS mutant most closely resembled PVIV-PaS in terms of structure, flexibility and activity profile. Both VVIV-PaS and PVIV-PaS mutants displayed similar activity profiles with enhanced TS hydrolysis but much smaller changes for EAS and ECS substrates (Figures 5 and 6).

3 Discussion

The primary objective of this study was to engineer an improved enzyme for steroid sulfate hydrolysis. The PaS enzyme was selected as the candidate as it had recently been reported to display hydrolytic activity towards steroid sulfates.[22] We applied semi-rational design for the process of protein engineering. Five rounds of SSM, three rounds of MSM and finally a shuffled library were created in the process of engineering PaS to discover improvements in the hydrolysis of testosterone sulfate. Together, three mutants were identified with ~150 times greater V_{max}/K_M value than WT-Pas (Table 1) and accompanied by a reduction in K_M value. In addition, increases in substrate scope (Figure 5 and 6) and an improvement in thermostability were observed for some variants (Table 2). One of the leading mutants, LEF-PaS, exhibited improved performance relative to WT-PaS in tests performed in spiked human urine instead of a simple buffer, indicating that PaS variants are minimally affected by the inhibitory substances in a urine matrix (Figure 7).

As a part of this study, molecular dynamics simulations of enzyme-substrate complexes were used to provide insight into the structural origin of the observed improvements. These simulations were based on well-established methods to model the catalytic metal ion, the FGH post-translational modification, and included new studies to evaluate the protonation state of key active site residues. Based on these simulations, the mutational pathway followed in this study can be briefly summarised as: the PV-PaS mutant introduced significant flexibility into the system and from this point two mutational trajectories emerged: (1) PV to PVFV to LEF, which was characterised by higher flexibility, higher V_{max} value, higher thermostability and generally broader substrate scope; and (2) PV to PVIV to VVIV, which was characterised by lower flexibility, lower K_M value and greater substrate specialization. From this analysis, it was clear that subtle changes in active site flexibility and the positioning of key residues could have a major impact on enzymatic activity and substrate scope. Significantly, previous experimental[25,26,34,42,43] and computational[33,44,45] studies have focused primarily on smaller model substrates, like pNPS, with activated leaving groups, whereas the present work examines larger steroid substrates of applied interest. Such substrates are notable for two reasons: they incorporate unstabilised leaving groups and so have a strict requirement for general acid catalysis during hydrolysis, and they are large and so more completely fill the PaS active site.

Of the intermediates along this mutational pathway PVFV-PaS showed the broadest substrate range with significant increases in EAS, ECS and ETS hydrolysis (Figure 5 and 6). This correlated with the greatest enzyme flexibility and was important, as previous studies of WT-PaS showed no activity for ECS or ETS hydrolysis and commercially available HpS showed no activity for ETS.[22] The substrate range improved early in the selection process, but continued screening for a single substrate proved detrimental to substrate range, as stated by You and Arnold[46] 'you get what you screen for'. This has often been observed with protein selection or directed evolution. The development of PVFV-PaS in this work that can hydrolyse the 17α-configured ETS is highly significant and provides a platform for further engineering efforts towards a universal steroid sulfatase enzyme for analytical applications.

Although sequential rounds of mutagenesis typically afforded improvements in catalytic efficiency, the MSM 3 library provided an exception (section 2.3.3). The design of the MSM 3 library was inspired by the study of Kintses $et. al.$,[47] where PaS was evolved for hydrolysis of a bulky aryl phosphonate ester. Several mutations to the 72-75 region were identified, in close proximity with the active site, that displayed improved hydrolytic activity or expression.[47] In the current study, the M72 and E74 residues were selected for randomization. The use of NYS codon degeneracy and screening to achieve >95% coverage of available amino acid diversity in the library ensured that mutants containing M72 should be sampled, in contrast to E74 that is not included in the NYS codon. Unlike the positive results obtained by Kintses $et. al.$[47] a library of inactive PaS variants was obtained for TS hydrolysis suggesting that the steroid sulfate hydrolysis presumably proceeds via a different catalytic pathway or has different requirements to the phosphonate ester hydrolysis by PaS. Based on inspection of the native PaS X-ray structure,[21] it is possible that the negative charge on E74 stabilizes the adjacent H115 and

the two residues form an in-line interaction that facilitates deprotonation of the FGH for the elimination of the sulfated FGH intermediate, an arrangement reminiscent of the serine protease catalytic triad.[48,49] Strongly attenuated pNPS hydrolysis activity was also observed for several independently prepared WT-PaS E74A/D/Q mutants (data not shown). The FGH residue had been observed in the X-ray structure of native PaS,[21] resolving earlier uncertainty of surrounding key active site residues and supporting the proposed transesterification-elimination-hydration mechanism that pinpointed H115 as a general base responsible for the elimination step. This work implicates E74 as a key residue directing and activating H115 in this elimination sequence.[18,21] The requirement for a charged residue such as E74 to interact with the general base for the elimination step does not appear to be universal for the sulfatase family. The X-ray crystal structures of the human arylsulfatase A[20] or B[50] enzymes do not contain a charged glutamate in this role and instead the histidine residue interacts with the backbone carbonyl of L92 or I114, respectively. An alignment of similar bacterial or fungal sulfatase genes (from UniRef 90 sequences) using ConSurf (S1.1.1) revealed that glutamate is common at residue 74 (corresponding to PaS residue numbering). ConSurf scored E74 in the top 40% of conserved residues. By comparison, the residues critical for metal ion coordination or catalysis are scored in the top 10%. The phylogenetic analysis of this alignment revealed that all PaS homologs from *Mycobacterium*, *Pseudomonas* or fungal species had a residue corresponding to E74.

With respect to aryl phosphonate ester hydrolysis, the results of Kintses *et al.*[47] identified a quadruple mutant M72L, E74G, M110I, K193Q with 6 times improvement in activity suggesting less sensitivity to the E74 residue for efficient catalysis. This observation may reflect different mechanistic pathways for phosphonate and sulfate ester hydrolysis. The accepted mechanism for the related alkaline phosphatase enzyme involves a metal coordinated serine as the catalytic nucleophile in the transesterification step.[51,52] Enzyme turnover is then achieved by direct nucleophilic substitution of the serine phosphate intermediate by water. In contrast, the sulfated FGH intermediate in PaS is believed to hydrolyse by cleavage of the sulfate ester C-O bond with generation of FGly that then undergoes subsequent hydration to regenerate FGH. It has been observed that substituting the FGH residue with serine in human arylsulfatase A and B leads to mutants capable of transesterification to form a sulfated serine intermediate, but that this is stable at pH 5 and is hydrolysed only slowly at higher pH, so limiting enzyme turnover.[53] Simple quantum mechanical computational models have identified the elimination of sulfated FGH as offering a lower energy pathway for sulfate ester hydrolysis than direct substitution at sulfur by an activated water molecule.[34] These computational models also reveal that direct hydrolysis by an activated water molecule provides a competitive pathway for hydrolysis in the case of the FGH phosphate esters.[34] Based on our experimental and computational analysis, a modified catalytic cycle for this enzyme is proposed (Figure 8), in which a catalytic nucleophile substitutes at the sulfate ester, with leaving group departure assisted by general acid catalysis from the K375 side chain. A mono-protonated H115 side chain then promotes elimination of the covalent sulfated FGH intermediate though hemiacetal cleavage. The mutational studies in this work serve to highlight the role of E74 in activating or pre-organizing the H115 general base responsible for elimination and so efficient sulfatase activity.

At a practical level, the pH optimum (V_{max}/K_M) for TS hydrolysis by WT-PaS is near 7.1 (Figure 4) indicating that catalytic activity is determined by the ionisation of two active-site residues of pKa values of 5.9 ± 0.1 and 8.3 ± 0.1. This contrasts with the reported pH optimum (k_{cat}/K_M) for the hydrolysis of the model substrate *para*-methoxyphenyl sulfate of 7.9 with ionisable residue pKa values of 7.1 ± 0.2 and 8.7 ± 0.1.[54] More significant differences in pH optimum were observed for the maximum initial hydrolysis rate (V_{max}) by WT-PaS: greater than 9[54] for *para*-methoxyphenyl sulfate or 8.9[55] for *para*-nitrocatechol sulfate, compared with a value near 6.0 for TS hydrolysis (Figure S2). The significant changes in pH optima may reflect the greater leaving group stability inherent in phenolic sulfate substrates. Computational studies on the hydrolysis of pNPS by WT-PaS have suggested that general acid catalysis is not involved in the key transesterification step.[33] Further, kinetic studies for a range of different phenolic sulfates revealed little influence of leaving group on the rate of hydrolysis

(V_{max}) suggesting the rate determining step of catalysis did not involve the leaving group.[34] This points to elimination (or possibly hydration) as the rate determining step for these aromatic substrates such that the pH optimum observed would reflect the residues central to this process. The hydrolysis of steroidal sulfates that involve a non-stabilised saturated alcohol as leaving group would have a strict requirement for general acid catalysis. The large change in pH optimum observed for steroidal substrates may reflect a change in the rate determining step and so a change in the residues involved in rate limiting catalysis.

Simultaneous to the development of this manuscript, two reports[11,13] of work conducted within the World Anti-Doping Agency-accredited laboratory based at the Institute of Biochemistry, German Sport University, Cologne employed the PVFV-PaS mutant described for the first time in this work for sample preparation prior to Gas Chromatography-Carbon Isotope Ratio-Mass Spectrometry (GC-CIR-MS). In this complementary study, PVFV-PaS-mediated sample preparation enabled measurements of epiandrosterone liberated from epiandrosterone sulfate, a metabolite of testosterone and related androgenic anabolic steroids.[12,13] The GC-CIR-MS technique provides an avenue to distinguish between endogenous steroid metabolites and those derived from otherwise identical exogenously administered steroids of synthetic origin. One conclusion of this study was that epiandrosterone sulfate served as a suitable marker for testosterone administration over a longer period than other reported markers suggesting that this method of analysis facilitated by PaS variants could improve the detection capabilities of anti-doping laboratories. Concordant with the results of this study, the PVFV-PaS variant, with greater thermostability and activity for EAS hydrolysis (Figures 5 and 6), provided higher yields of free steroid during sample preparation than WT-PaS, highlighting the real world benefits of this mutagenesis study.[11,13]

4 Conclusions

The arylsulfatase from *Pseudomonas aeruginosa* was engineered by a semi-rational approach to discover mutants with improved catalytic activity for testosterone sulfate (TS) hydrolysis and to improve the substrate scope. After screening eight libraries for TS hydrolysis, mutants with significant improvements in catalytic activity, substrate affinity or both were obtained. The LEF-PaS exhibits the highest catalytic activity (V_{max}) for TS hydrolysis with a catalytic efficiency (V_{max}/K_M) more than 150 times than WT-PaS. However, the PVFV-PaS intermediate showed the broadest substrate range. Importantly, the detection of significant hydrolytic activity for the α-configured steroid sulfates derived from etiocholanolone and epitestosterone was a breakthrough. The PVFV-PaS mutant hydrolyses ECS at a rate two orders of magnitude higher than that of WT-PaS and represents a starting point for engineering a universal steroid sulfatase. In addition, the importance of E74 residue in the hydrolysis of steroid sulfates was uncovered and provides an avenue for further mechanistic investigation. As PaS variants are compatible with urine matrices under reaction conditions and pH similar to those employed during *E. coli* β-glucuronidase treatment, this study advances PaS as a valuable steroid sulfatase for analytical applications.

Supporting Information Available

Materials and methods for mutational studies, enzyme purification and characterisation, molecular dynamics simulations, supporting information tables and figures. This information is available free of charge on the ACS Publications website.

Acknowledgements

References

(1) Mueller, J. W.; Gilligan, L. C.; Idkowiak, J.; Arlt, W.; Foster, P. A. The Regulation of Steroid Action by Sulfation and Desulfation. *Endocr. Rev.* **2015**, *36*, 526–563.

(2) Gómez, C.; Pozo, O. J.; Marcos, J.; Segura, J.; Ventura, R. Alternative Long-Term Markers for the Detection of Methyltestosterone Misuse. *Steroids* **2013**, *78*, 44–52.

(3) Gómez, C.; Pozo, O. J.; Garrostas, L.; Segura, J.; Ventura, R. A New Sulphate Metabolite as a Long-Term Marker of Metandienone Misuse. *Steroids* **2013**, *78*, 1245–1253.

(4) Strahm, E.; Baume, N.; Mangin, P.; Saugy, M.; Ayotte, C.; Saudan, C. Profiling of 19-Norandrosterone Sulfate and Glucuronide in Human Urine: Implications in Athlete's Drug Testing. *Steroids* **2009**, *74*, 359–364.

(5) Torrado, S.; Roig, M.; Farré, M.; Segura, J.; Ventura, R. Urinary Metabolic Profile of 19-Norsteroids in Humans: Glucuronide and Sulphate Conjugates after Oral Administration of 19-nor-4-Androstenediol. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3035–3042.

(6) Gomez, C.; Fabregat, A.; Pozo, Ó. J.; Marcos, J.; Segura, J.; Ventura, R. Analytical Strategies Based on Mass Spectrometric Techniques for the Study of Steroid Metabolism. *TrAC Trends Anal. Chem.* **2014**, *53*, 106–116.

(7) Thevis, M.; Kuuranne, T.; Walpurgis, K.; Geyer, H.; Schänzer, W. Annual Banned-Substance Review: Analytical Approaches in Human Sports Drug Testing. *Drug Test. Anal.* **2016**, *8*, 7–29.

(8) Liu, Y.; Lu, J.; Yang, S.; Xu, Y.; Wang, X. A New Potential Biomarker for 1-Testosterone Misuse in Human Urine by Liquid Chromatography Quadruple Time-of-Flight Mass Spectrometry. *Anal. Methods* **2015**, *7*, 4486–4492.

(9) Gómez, C.; Pozo, O. J.; Geyer, H.; Marcos, J.; Thevis, M.; Schänzer, W.; Segura, J.; Ventura, R. New Potential Markers for the Detection of Boldenone Misuse. *J. Steroid Biochem. Mol. Biol.* **2012**, *132*, 239–246.

(10) Boccard, J.; Badoud, F.; Grata, E.; Ouertani, S.; Hanafi, M.; Mazerolles, G.; Lantéri, P.; Veuthey, J.-L.; Saugy, M.; Rudaz, S. A Steroidomic Approach for Biomarkers Discovery in Doping Control. *Forensic Sci. Int.* **2011**, *213*, 85–94.

(11) Piper, T.; Schänzer, W.; Thevis, M. Genotype-Dependent Metabolism of Exogenous Testosterone – New Biomarkers Result in Prolonged Detectability. *Drug Test. Anal.* **2016**, *8*, 1163–1173.

(12) Piper, T.; Putz, M.; Schänzer, W.; Pop, V.; McLeod, M. D.; Uduwela, D. R.; Stevenson, B. J.; Thevis, M. Epiandrosterone Sulfate Prolongs the Detectability of Testosterone, 4-Androstenedione, and Dihydrotestosterone Misuse by Means of Carbon Isotope Ratio Mass Spectrometry. *Drug Test. Anal.* **2017**, *9*, 1695–1703.

(13) Putz, M.; Piper, T.; Casilli, A.; Radler de Aquino Neto, F.; Pigozzo, F.; Thevis, M. Development and Validation of a Multidimensional Gas

Chromatography/Combustion/Isotope Ratio Mass Spectrometry-Based Test Method for Analyzing Urinary Steroids in Doping Controls. *Anal. Chim. Acta* **2018**, *1030*, 105–114.

(14) Marcos, J.; Pozo, O. J. Current LC–MS Methods and Procedures Applied to the Identification of New Steroid Metabolites. *J. Steroid Biochem. Mol. Biol.* **2016**, *162*, 41–56.

(15) Gomes, R. L.; Meredith, W.; Snape, C. E.; Sephton, M. A. Analysis of Conjugated Steroid Androgens: Deconjugation, Derivatisation and Associated Issues. *J. Pharm. Biomed. Anal.* **2009**, *49*, 1133–1140.

(16) Pedersen, M.; Frandsen, H. L.; Andersen, J. H. Optimised Deconjugation of Androgenic Steroid Conjugates in Bovine Urine. *Food Addit. Contam. Part A* **2017**, *34*, 482–488.

(17) Athlete Biological Passport (ABP) Operating Guidelines https://www.wada-ama.org/en/resources/athlete-biological-passport/athlete-biological-passport-abp-operating-guidelines (accessed Aug 15, 2017).

(18) Hanson, S. R.; Best, M. D.; Wong, C. Sulfatases: Structure, Mechanism, Biological Activity, Inhibition, and Synthetic Utility. *Angew. Chem. Int. Ed.* **2004**, *43*, 5736–5763.

(19) Appel, M. J.; Bertozzi, C. R. Formylglycine, a Post-Translationally Generated Residue with Unique Catalytic Capabilities and Biotechnology Applications. *ACS Chem. Biol.* **2015**, *10*, 72–84.

(20) Lukatela, G.; Krauss, N.; Theis, K.; Selmer, T.; Gieselmann, V.; von Figura, K.; Saenger, W. Crystal Structure of Human Arylsulfatase A: The Aldehyde Function and the Metal Ion at the Active Site Suggest a Novel Mechanism for Sulfate Ester Hydrolysis,. *Biochemistry* **1998**, *37*, 3654–3664.

(21) Boltes, I.; Czapinska, H.; Kahnert, A.; von Bülow, R.; Dierks, T.; Schmidt, B.; von Figura, K.; Kertesz, M. A.; Usón, I. 1.3 Å Structure of Arylsulfatase from Pseudomonas Aeruginosa Establishes the Catalytic Mechanism of Sulfate Ester Cleavage in the Sulfatase Family. *Structure* **2001**, *9*, 483–491.

(22) Stevenson, B. J.; Waller, C. C.; Ma, P.; Li, K.; Cawley, A. T.; Ollis, D. L.; McLeod, M. D. Pseudomonas Aeruginosa Arylsulfatase: A Purified Enzyme for the Mild Hydrolysis of Steroid Sulfates. *Drug Test. Anal.* **2015**, *7*, 903–911.

(23) Dodgson, K. S.; Powell, G. M. Studies on Sulphatases. 27. The Partial Purification and Properties of the Arylsulphatase of the Digestive Gland of Helix Pomatia. *Biochem. J.* **1959**, *73*, 672–679.

(24) Waller, C. C.; Cawley, A. T.; Suann, C. J.; Ma, P.; McLeod, M. D. In Vivo and in Vitro Metabolism of the Designer Anabolic Steroid Furazadrol in Thoroughbred Racehorses. *J. Pharm. Biomed. Anal.* **2016**, *124*, 198–206.

(25) Olguin, L. F.; Askew, S. E.; O'Donoghue, A. C.; Hollfelder, F. Efficient Catalytic Promiscuity in an Enzyme Superfamily: An Arylsulfatase Shows a Rate Acceleration of 1013 for Phosphate Monoester Hydrolysis. *J. Am. Chem. Soc.* **2008**, *130*, 16547–16555.

(26) Babtie, A. C.; Bandyopadhyay, S.; Olguin, L. F.; Hollfelder, F. Efficient Catalytic Promiscuity for Chemically Distinct Reactions. *Angew. Chem. Int. Ed.* **2009**, *48*, 3692–3694.

(27) Pabis, A.; Duarte, F.; Kamerlin, S. C. L. Promiscuity in the Enzymatic Catalysis of Phosphate and Sulfate Transfer. *Biochemistry* **2016**, *55*, 3061–3081.

(28) Jensen, R. A. Enzyme Recruitment in Evolution of New Function. *Annu. Rev. Microbiol.* **1976**, *30*, 409–425.

(29) O'Brien, P. J.; Herschlag, D. Catalytic Promiscuity and the Evolution of New Enzymatic Activities. *Chem. Biol.* **1999**, *6*, R91–R105.

(30) Khersonsky, O.; Tawfik, D. S. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annu. Rev. Biochem.* **2010**, *79*, 471–505.

(31) Ashkenazy, H.; Erez, E.; Martz, E.; Pupko, T.; Ben-Tal, N. ConSurf 2010: Calculating Evolutionary Conservation in Sequence and Structure of Proteins and Nucleic Acids. *Nucleic Acids Res.* **2010**, *38*, W529–W533.

(32) Celniker, G.; Nimrod, G.; Ashkenazy, H.; Glaser, F.; Martz, E.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Isr. J. Chem.* **2013**, *53*, 199–206.

(33) Luo, J.; van Loo, B.; Kamerlin, S. C. L. Catalytic Promiscuity in Pseudomonas Aeruginosa Arylsulfatase as an Example of Chemistry-Driven Protein Evolution. *FEBS Lett.* **2012**, *586*, 1622–1630.

(34) Williams, S. J.; Denehy, E.; Krenske, E. H. Experimental and Theoretical Insights into the Mechanisms of Sulfate and Sulfamate Ester Hydrolysis and the End Products of Type I Sulfatase Inactivation by Aryl Sulfamates. *J. Org. Chem.* **2014**, *79*, 1995–2005.

(35) Reetz, M. T.; Bocola, M.; Carballeira, J. D.; Zha, D.; Vogel, A. Expanding the Range of Substrate Acceptance of Enzymes: Combinatorial Active-Site Saturation Test. *Angew. Chem. Int. Ed.* **2005**, *44*, 4192–4196.

(36) Belsare, K. D.; Andorfer, M. C.; Cardenas, F. S.; Chael, J. R.; Park, H. J.; Lewis, J. C. A Simple Combinatorial Codon Mutagenesis Method for Targeted Protein Engineering. *ACS Synth. Biol.* **2017**, *6*, 416–420.

(37) Cornish-Bowden, A. *Fundamentals of Enzyme Kinetics, Third Ed.*; Portland Press: London, 2004.

(38) Barrozo, A.; Duarte, F.; Bauer, P.; Carvalho, A. T. P.; Kamerlin, S. C. L. Cooperative Electrostatic Interactions Drive Functional Evolution in the Alkaline Phosphatase Superfamily. *J. Am. Chem. Soc.* **2015**, *137*, 9061–9076.

(39) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of PKa Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.

(40) Ma, H.; Szeler, K.; Kamerlin, S. C. L.; Widersten, M. Linking Coupled Motions and Entropic Effects to the Catalytic Activity of 2-Deoxyribose-5-Phosphate Aldolase (DERA). *Chem. Sci.* **2016**, *7*, 1415–1421.

(41) Karshikoff, A.; Nilsson, L.; Ladenstein, R. Rigidity versus Flexibility: The Dilemma of Understanding Protein Thermal Stability. *FEBS J.* **2015**, *282*, 3899–3917.

(42) Schober, M.; Toesch, M.; Knaus, T.; Strohmeier, G. A.; van Loo, B.; Fuchs, M.; Hollfelder, F.; Macheroux, P.; Faber, K. One-Pot Deracemization of Sec-Alcohols: Enantioconvergent Enzymatic Hydrolysis of Alkyl Sulfates Using Stereocomplementary Sulfatases. *Angew. Chem. Int. Ed.* **2013**, *52*, 3277–3279.

(43) van Loo, B.; Schober, M.; Valkov, E.; Heberlein, M.; Bornberg-Bauer, E.; Faber, K.; Hyvönen, M.; Hollfelder, F. Structural and Mechanistic Analysis of the Choline Sulfatase from Sinorhizobium Melliloti: A Class I Sulfatase Specific for an Alkyl Sulfate Ester. *J. Mol. Biol.* **2018**, *430*, 1004–1023.

(44) Luo, J.; van Loo, B.; Kamerlin, S. C. L. Examining the Promiscuous Phosphatase Activity of Pseudomonas Aeruginosa Arylsulfatase: A Comparison to Analogous Phosphatases. *Proteins Struct. Funct. Bioinforma.* **2012**, *80*, 1211–1226.

(45) Marino, T.; Russo, N.; Toscano, M. Catalytic Mechanism of the Arylsulfatase Promiscuous Enzyme from Pseudomonas Aeruginosa. *Chem. Eur. J.* **2013**, *19*, 2185–2192.

(46) You, L.; Arnold, F. H. Directed Evolution of Subtilisin E in Bacillus Subtilis to Enhance Total Activity in Aqueous Dimethylformamide. *Protein Eng. Des. Sel.* **1996**, *9*, 77–83.

(47) Kintses, B.; Hein, C.; Mohamed, M. F.; Fischlechner, M.; Courtois, F.; Lainé, C.; Hollfelder, F. Picoliter Cell Lysate Assays in Microfluidic Droplet Compartments for Directed Enzyme Evolution. *Chem. Biol.* **2012**, *19*, 1001–1009.

(48) Kraut, J. Serine Proteases: Structure and Mechanism of Catalysis. *Annu. Rev. Biochem.* **1977**, *46*, 331–358.

(49) Frey, P. A.; Whitt, S. A.; Tobin, J. B. A Low-Barrier Hydrogen Bond in the Catalytic Triad of Serine Proteases. *Science* **1994**, *264*, 1927–1930.

(50) Bond, C. S.; Clements, P. R.; Ashby, S. J.; Collyer, C. A.; Harrop, S. J.; Hopwood, J. J.; Guss, J. M. Structure of a Human Lysosomal Sulfatase. *Structure* **1997**, *5*, 277–289.

(51) Kim, E. E.; Wyckoff, H. W. Reaction Mechanism of Alkaline Phosphatase Based on Crystal Structures. *J. Mol. Biol.* **1991**, *218*, 449–464.

(52) Holtz, K. M.; Kantrowitz, E. R. The Mechanism of the Alkaline Phosphatase Reaction: Insights from NMR, Crystallography and Site-Specific Mutagenesis. *FEBS Lett.* **1999**, *462*, 7–11.

(53) Recksiek, M.; Selmer, T.; Dierks, T.; Schmidt, B.; Figura, K. von. Sulfatases, Trapping of the Sulfated Enzyme Intermediate by Substituting the Active Site Formylglycine. *J. Biol. Chem.* **1998**, *273*, 6096–6103.

(54) Bojarová, P.; Denehy, E.; Walker, I.; Loft, K.; De Souza, D. P.; Woo, L. W. L.; Potter, B. V. L.; McConville, M. J.; Williams, S. J. Direct Evidence for ArO⬚S Bond Cleavage upon Inactivation of Pseudomonas Aeruginosa Arylsulfatase by Aryl Sulfamates. *ChemBioChem* **2008**, *9*, 613–623.

(55) Beil, S.; Kehrli, H.; James, P.; Staudenmann, W.; Cook, A. M.; Leisinger, T.; Kertesz, M. A. Purification and Characterization of the Arylsulfatase Synthesized by Pseudomonas Aeruginosa PAO During Growth in Sulfate-Free Medium and Cloning of the Arylsulfatase Gene (AtsA). *Eur. J. Biochem.* **1995**, *229*, 385–394.

TOC Graphic
For table of contents only.