Improving the arylsulfatase from *Pseudomonas aeruginosa* for anti-doping applications

A thesis submitted for the degree of Doctor of Philosophy of
The Australian National University

By

Dimanthi Roshika Uduwela

Research School of Chemistry

Australian National University

February 2017

Declaration

I do hereby declare that this thesis represents the work conducted by me during my PhD candidature between July 2013 and February 2017 under the supervision of Associate Professor Malcolm McLeod. To the best of my knowledge, the work presented in this thesis does not contain material that has been previously published, except where due acknowledgement has been made.

I also certify that this thesis contains less than 100,000 words in accordance with the guidelines of the Australian National University.

……..

Dimanthi Uduwela

February 2017

Acknowledgements

First and foremost, I wish to express my sincere thanks and gratitude to my supervisor, Associate Professor Malcolm McLeod for giving me this opportunity to study in this prestigious institution. He has been a great pillar of support right from the beginning guiding and steering me in the best direction. His doors have always been open for discussions and he is an excellent example of a mentor. The enthusiasm he has for research work is inspiring and has always kept me motivated to do more. I could not have asked for a better supervisor and I greatly appreciate for all the encouragement given and the time spent on me to overcome the difficulties and smooth sail this journey.

Next, I would like to thank Dr Bradley Stevenson, a member of my supervisory panel and my research colleague for the immense guidance given to me. I am very much grateful for him for teaching me molecular biology techniques, for the LC-MS expertise rendered and for all the advices given to me. The patience he has shown and his willingness to help as the need arose are greatly appreciated. I also wish to extend my gratitude to the rest of the supervisory panel members, Professor David Ollis and Dr. Adam Cawley for their valuable advices given to me, and specially Professor Ollis, whose lab I was working in during my candidature, for providing me with the facilities in his lab. I also thank Mr Alex Chen of Thermo Fisher, Melbourne for running my samples on the Q-Exactive instrument, and the technical staff at ANU for assisting me at various occasions.

I thank Dr. Nick Kanizaj for the dynamic atmosphere he always creates with his lively laughter and songs and also to Ms Tracy Murray for helping me with the initial laboratory issues. I also wish to thank all my colleagues from the McLeod and Ollis groups past and present whom I met during my period at ANU; Paul, Ryan, Chris, Keshav, Jacob, Natasha, Andy, Sumudu, Rikard Shu Ann, Nurul, Jo, Fizah, TK, Joleen, Adli, and Ana for sharing the good vibes, for the strength and support given and more importantly for all the enjoyable times we had together.

I would also like to thank the friends at the RSC, the Graduate House and University House communities past and present and the friends I met in Canberra for making my stay at ANU and Canberra vibrant with all the games nights, movie nights, potlucks, sing-alongs, trips, parties and so on. Last but not least, I wish to extend my sincere gratitude

and love to my parents and siblings for their continuous encouragement and strength given to me to make this thesis a reality.

Abstract

The misuse of anabolic androgenic steroids in sports remains a significant concern. They are exploited to enhance physical performance, and as a result, they are listed as prohibited substances by the world anti-doping agency (WADA). Detection of steroids is important to maintain fair play in sports and to ensure that athletes are not jeopardising their long-term health.

The human phase II steroid metabolites are mainly excreted in urine as glucuronide or sulfate conjugates. The routine screening methods require deconjugation of the phase II conjugates to produce the free steroids. The hydrolysis of steroid glucuronides is efficiently achieved by using *E. coli* β-glucuronidase enzyme. However, there is no robust method for hydrolysing steroid sulfates. Acid catalysed solvolysis is a general method of hydrolysis of steroid conjugates. However, this method is often unsuitable for routine anti-doping screening since it can degrade analytes of interest and generate a more complicated analytical matrix. Steroid sulfates are emerging as highly useful markers to detect AAS abuse, importantly as long-term markers that have the potential to improve the sensitivity and the retrospectivity of steroid screening. However, their routine detection would benefit from a robust steroid sulfatase for deconjugation, in a manner analogous to the β-glucuronidase enzyme. The arylsulfatase from *Pseudomonas aeruginosa* (PaS) is an enzyme capable of hydrolysing steroid sulfates and represents a good starting point for engineering a steroid sulfatase. However, this enzyme requires improvement in hydrolytic activity and substrate scope in order to be useful in an anti-doping context.

This study aimed to improve the catalytic rate 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. These improvements were sought by applying semi-rational design to mutate amino acid residues neighbouring the enzyme active site by taking use of the crystal structure of PaS (1HDH) and the binding modes suggested by computational ligand protein docking. Mutagenesis was implemented on single residues and multiple residue sites and finally shuffling all the beneficial mutations. Screening was performed to test the steroid sulfate hydrolysis activity of these mutant libraries by employing LC-MS.

These libraries revealed the steroid sulfate binding pocket of PaS and screening of >10000 variants resulted in three mutants that showed an improvement in the catalytic efficiency (V_{max}/ K_M) of more than 150 times that of wild type PaS for TS hydrolysis. In addition, the substrate scope of steroid sulfates for PaS enzyme was increased and a modest improvement in thermostability was observed. The mechanistic involvement of E74 as a catalytic residue was proposed with the observation of mutants created on this residue.

To apply the enzyme on real world samples, an untargeted screen was performed on pre- and post-administration testosterone propionate (TP) horse urine samples which identified three metabolites that could be potential markers for TP administration. This also revealed that PaS is compatible with urine matrices under the reaction conditions analogous to β-glucuronidase treatment in the anti-doping screens, making it a valuable steroid sulfatase for anti-doping applications.

Table of Contents

A(S)	androsterone (sulfate)
AAS	anabolic androgenic steroids
AtsA	gene encoding arylsulfatase
bp	base pairs (nucleotide)
BPER II	bacterial protein extraction reagent ®
BSA	bovine serum albumin
d_3TS	(16,16,17α-d3)-testosterone sulfate
ddMS/MS	data dependent tandem mass spectrometry
DHEA(S)	dehydroepiandrosterone (sulfate)
DNA	deoxyribonucleic acid
dNTP	deoxyribose nucleoside triphosphate
DTNB	5,5’-dithiobis-(2-nitrobenzoic acid)
E(S)	estrone (sulfate)
EA(S)	epiandrosterone (sulfate)
EC x.x.x.x	enzyme commission number
EC(S)	etiocholanolone (sulfate)
EMBL	European molecular biology laboratory
ET(S)	epitestosterone (sulfate)
FGH	formylglycine hydrate
FGly	α-formylglycine
FWHM	full width at half maximum
GC	gas chromatography
GC-CIRMS	gas chromatography – combustion isotope ratio mass spectrometry
GC-MS	gas chromatography – mass spectrometry
GC-MS/MS	gas chromatography – tandem mass spectrometry
GSH	glutathione
HARSA	human arylsulfatase A
HARSB	human arylsulfatase B
HEPES	N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]

HF	high fidelity
HPLC	high performance liquid chromatography
IPTG	isopropyl β-D-thiogalactopyranoside
IRMS	isotope ratio mass spectrometry
kb	kilo base pairs
K_M	Michaelis-Menten constant
LB	Lysogeny broth
LBA	Lysogeny broth with 100 mg L-1 ampicillin
LC	liquid chromatography
LC-MS	liquid chromatography mass spectrometry
LC-MS/MS	liquid chromatography - tandem mass spectrometry
m/z	mass to charge ratio
MES	2-(N-morpholino)ethanesulfonic acid
MMTS	minimal media testosterone sulfate
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MSM	multiple site mutagenesis
N(S)	nandrolone (sulfate)
PaS	*Pseudomonas aeruginosa* arylsulfatase
PCR	polymerase chain reaction
PDB	protein data bank
PNPS	para-nitrophenyl sulfate
QTOF	quadrupole time-of-flight
SDM	site directed mutagenesis
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIM	selected ion monitoring
SPE	solid phase extraction
SSM	site saturation mutagenesis
T(S)	testosterone (sulfate)
TB	terrific broth
TBA	terrific broth with 100 mg L-1 ampicillin

TMA	tris-mes-acetic acid
T_m^{app}	apparent melt temperature
$T_{1/2}$	half life
T/N	testosterone/nandrolone ratio
TP	testosterone propionate
Tris	tris(hydroxymethyl)aminomethane
UHPLC	ultra high performance liquid chromatography
V_{max}	maximum velocity
WADA	world anti-doping agency
WT	wild type
YENB	yeast extract and nutrient broth

Amino acids are referred to by either their standard one-letter or three-letter codes in this thesis. Amino acids with the residue number in the gene sequence are denoted in the format of X#, X#Y, or #Y, where X is the wild-type amino acid, # is the residue number, and Y is the mutant amino acid.

Chapter 1: Introduction

1.1 Anabolic androgenic steroids

Steroids are a class of natural products that share a characteristic tetracyclic fused ring system with three cyclohexane rings (rings A, B and C) and one cyclopentane ring (ring D). The steroid structure usually contains methyl groups at C-10 and C-13 positions and may contain an alkyl side chain at C-17 position [1]. Unless stated otherwise the stereochemistry of the substituents at the ring junction positions is 8β, 9α, 10β, 13β and 14α with the stereochemistry at C-5 being variable. The structure of steroids is typified by testosterone in Figure 1.1. The steroids may contain various functional groups differing in stereochemistry and position and also the degree of unsaturation.

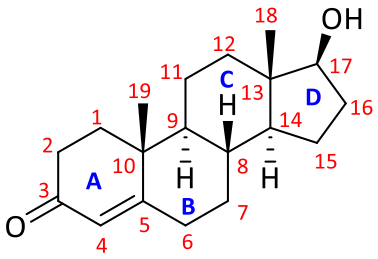


Figure 1.1: Structure of testosterone (17β-hydroxyandrost-4-en-3-one) showing the IUPAC recommended atom and ring labelling

Anabolic androgenic steroids (AASs) are substances that mimic testosterone (T), a naturally occurring steroid. They mediate their effects by binding to and activating the androgen receptors, that will facilitate muscle gain and strength (anabolic effects) and develop male sexual characteristics in both males and females (androgenic effects). The AASs are used clinically to treat wounds, burns, many forms of anaemia and steroid deficiency conditions such as delayed puberty [2, 3]. They are also used to stimulate growth in the treatment of chronic wasting conditions such as cancers and AIDS where a rapid muscle gain is desired. However, the misuse of AASs can also result in adverse effects like permanent organ damage, reduced fertility, hypertension, psychiatric and behavioural disorders [3, 4].

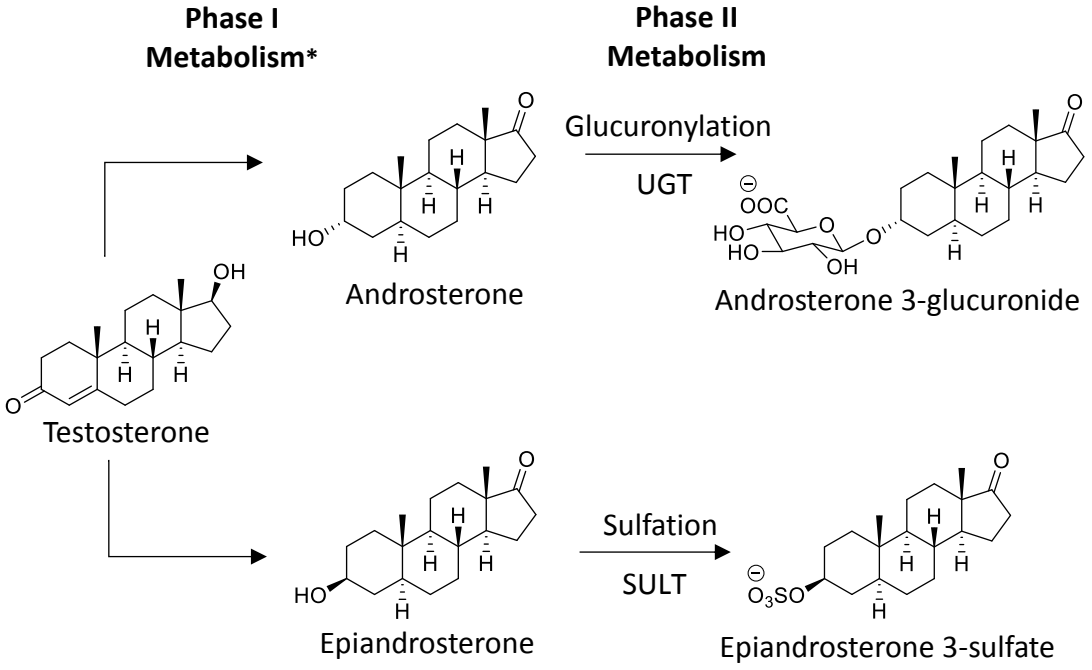
Exploitation of steroids by athletes to enhance their physical performance, commonly known as doping, is a concern that has attracted increased attention in the past few decades. Administration of steroids can build muscle mass, reduce fatigue levels, increase the rate of recovery and consequently enhance the performance of the athlete. They have been classified as prohibited substances by the World Anti-Doping Agency (WADA) [5]. The WADA classifies a substance as prohibited when it meets two of the three criteria listed as: a substance which has the potential to enhance or enhances sport performance, represents an actual or potential health risk to the athlete, or violates the spirit of sport. In addition to AASs, peptide hormones, growth factors, related substances and mimetics, hormone and metabolic modulators, diuretics and masking agents, are drug classes that could change the endogenous steroid profile or mask AAS abuse and which are also prohibited by WADA [5]. As a result AASs have been outlawed by the major sporting organizations including the International Olympic Committee (IOC) [6]. Anabolic steroids were added to the list of IOC's banned substances in 1975 as a test considered reliable was developed based on radioimmunoassay techniques,[4] and steroid testing was first conducted at the Montreal Olympics in 1976. Detection of steroid abuse in athletes frequently results in severe penalties such as retrospective disqualifications and the removal of medals or records.

1.2 Steroid metabolism

Steroids undergo a number of metabolic changes primarily in the liver but also in other tissues. These metabolic reactions are grouped into two phases, phase I and phase II, where both phases involve enzymatically controlled reactions that typically increase the polar nature of the steroid and facilitate excretion from the body. The two phases can occur in turn or independently from one another. Phase I metabolism involves oxidation, reduction or hydroxylation reactions that increase the hydrophilicity of the steroid. It is followed by phase II metabolism that involves conjugating the steroids or their metabolites with glucuronic acid, sulfate, amino acids, glutathione or methyl groups. These moieties typically serve to further increase the hydrophilicity [7, 8]. The major human phase II metabolites are glucuronate and sulfate conjugated steroids.

Figure 1.2 shows representative metabolic pathways for testosterone leading to the formation of androsterone 3-glucuronide and epiandrosterone 3-sulfate.

Glucuronylation in humans is catalysed by uridine 5'-diphosphoglucuronosyltransferases (UGTs) and uses uridine-5'-diphosphoglucuronic acid (UDPGA) as the glucoronyl donor. Sulfation reactions are catalysed by sulfotransferases (SULTs) where 3'-phosphoadenosine 5'-phosphosulfate (PAPS) is used as the sulfate donor. Conjugation typically reduces the biological activity and the toxicity of the steroid and facilitates excretion. Nevertheless, cholestatic activities, that is the stagnation of bile flow and toxicities for some D-ring glucuronide conjugates such as the glucuronides of 17β-hydroxyestrogens, testosterone and dihydrotestosterone have been reported [9, 10]. Not all the steroids and their metabolites are excreted as conjugates, but it has been reported that the unconjugated fraction of steroids is less than 3% of the total amount of steroids excreted and therefore conjugation plays an important role in steroid metabolism and excretion [11].



*Phase I metabolism reactions shown consist of enzymatically controlled multiple steps

Figure 1.2: An exemplary metabolic pathways of testosterone.

Different steroids follow different metabolic pathways. The main urinary metabolites after ingestion of T have been reported as the glucuronylated and sulfated conjugates of testosterone, androsterone, etiocholanolone, 5α-androstane-3α,17β-

diol, 5β-androstane-3α,17β-diol and epiandrosterone [12, 13]. A recent study of testosterone metabolism by Piper *et al.* [12] has been performed by administration of deuterium labelled testosterone to four male volunteers and monitoring the deuterated metabolites over a period of eight days by hydrogen isotope ratio mass spectrometry (HIRMS). These results revealed some new insights to testosterone metabolism, including an unprecedented methylation, and importantly epiandrosterone 3-sulfate was identified as a long-term marker of testosterone administration. In addition to urinary excretion products, steroid sulfates such as dehydroepiandrosterone-3-sulfate (DHEAS) can also be found in plasma and endocrine tissues [14]. Both DHEAS and DHEA are secreted in the human adrenal cortex and they act as pro-androgens involved in the biosynthesis and regulation of testosterone, a potent endogenous androgen [15, 16].

Surveillance of steroid metabolism is important in the testing of doping in sports, as the starting steroids will be excreted in a modified inactivated form. For instance, testosterone administered by an athlete will be excreted in multiple forms after metabolism, such as the conjugated forms of androsterone, etiocholanolone, androstanediol, epiandrosterone and other metabolites in addition to the starting testosterone [12, 14].

1.3 Detection of steroids

The detection of the parent compound or at least one metabolite of the synthetic steroids such as methyltestosterone, stanozolol or oxandrolone provides evidence for doping [4, 17, 18 19]. However, the situation is different for endogenous steroids such as testosterone. A number of markers to detect the illicit use of testosterone have been reported [11, 12, 20-22]. A well-known biomarker for testosterone doping is the testosterone (T) to epitestosterone (ET) ratio (T/ET) of the glucuronylated fractions in urinary excretion. The secretion of epitestosterone, an inactive epimer of testosterone, is typically unchanged by the administration of T or other AASs and hence this increases the ratio of T/ET, which is usually around 1:1 otherwise. This ratio was first introduced in 1983 [23] with a threshold at 6.0 which was later brought down to 4.0. According to the WADA criteria, a ratio of ≥ 4.0 is considered suspicious and further testing will be carried

out as this ratio is also affected by the ethnicity, sex or age of the athlete [14]. Other criteria indicating endogenous steroid abuse includes concentrations of T or ET being greater than 200 ng/mL in males and 50 ng/mL in females, or concentrations of A or EC being greater than 10 µg/mL combined with a ratio of A/EC lower than 0.4 in males (in the absence of 5α-reductase) or greater than 4 in males or females [23].

1.3.1 Athlete Biological Passport (ABP)

Recently the WADA together with medical experts introduced the concept of the Athlete Biological Passport (ABP). It is an electronic record of the athlete's biological values, composed of two modules [23]; i) the haematological module that aims to identify the enhancements of oxygen transport including erythropoiesis-stimulating agents and any form of blood transfusion or manipulation; ii) the steroid module or the steroid profile aims to identify the administration of endogenous AAS.

The testing of the steroid profile is performed on a longitudinal basis where the values from a series of steroid profiles from an individual are compared over time. The markers that compose the steroid profile are listed as, testosterone (T), epitestosterone (ET), androsterone (A), etiocholanolone (EC), 5α-androstane-3α,17β-diol (5αAdiol), 5β-androstane-3α,17β-diol (5βAdiol) and the testosterone to epitestosterone ratio (T/ET). In addition to these markers, other steroidal ratios such as A/T, A/EC, 5αAdiol/5βAdiol, 5αAdiol/E are useful in evaluating the steroid profile. These markers are detected in the free steroid form as a sum of the free steroid fraction and the hydrolysed products of the steroid glucuronides.

Where an unusual steroid profile is detected, conformation is carried out as a second step that will distinguish the endogenous steroids from the exogenous. Gas chromatography combustion isotope ratio mass spectrometry (GC-CIRMS) is used as a confirmation technique as the ratio of stable carbon isotopes $^{13}C/^{12}C$ allows discrimination between natural and synthetic steroids. Since exogenous testosterone or precursors contain less ^{13}C than their endogenous homologues, it is expected that urinary steroids with a low $^{13}C/^{12}C$ ratio originate from pharmaceutical sources [4, 14].

1.3.2 Analytical approaches

Most commonly, the urinary steroids are analysed in the anti-doping screens as urine can be acquired in noninvasively large volumes [11, 20]. Gas chromatography-mass spectrometry (GC-MS) is the method conventionally used for the detection of steroids, which are then confirmed by comparison with reference materials. In comparison to the full scan MS or single ion monitoring (SIM) techniques used for detection in GC-MS methods, the gas chromatography tandem mass spectrometry (GC-MS/MS) includes a range of scan types to identify the metabolites with greater selectivity and sensitivity. This includes full scan MS, SIM, or MS/MS techniques such as precursor ion, product ion and neutral loss scans, and selected reaction monitoring (SRM) techniques such as multiple reaction monitoring (MRM). This technique being more sensitive and specific than GC-MS typically enables detection over an increased time period post drug administration. The detection by GC-MS method requires sample processing steps such as conjugate hydrolysis, pre-concentration and derivatization to enhance stability, volatility, ionization efficiency and detection limits (Figure 1.3).

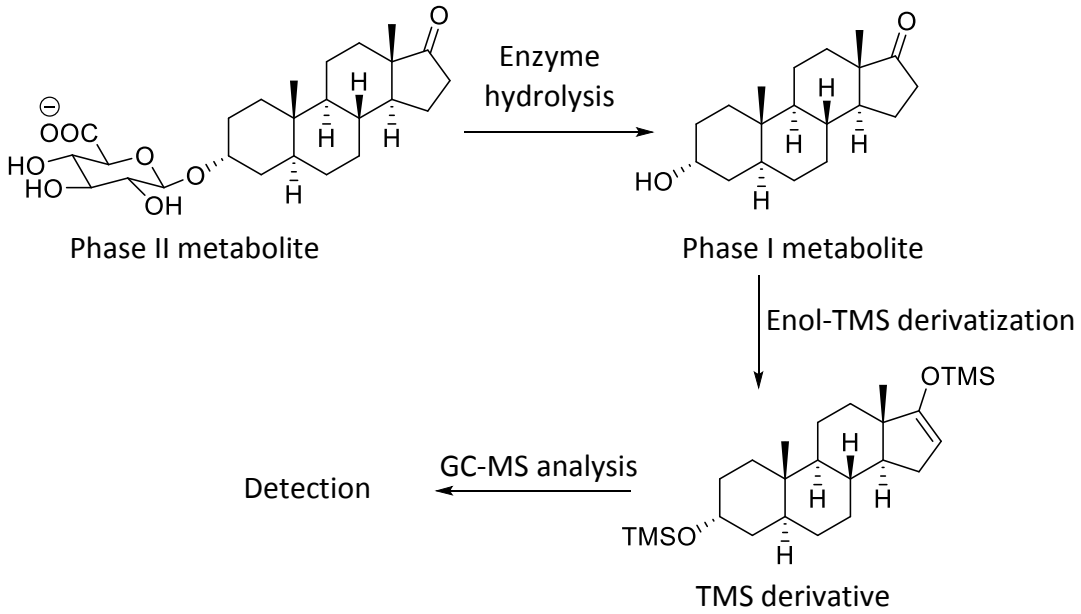


Figure 1.3: The important sample processing steps required for GC-MS detection with an example metabolite of testosterone

The hydrolysis of steroid glucuronides, a major class of human metabolites, is achieved by using *E. coli* β-glucuronidase. However, the efficiency of the hydrolysis depends on a number of factors [10, 14, 22] such as enzyme inhibitors present in the urine

matrix, the presence of contaminants in different enzyme preparations, and also the pH, temperature, and incubation time. Despite the application of optimised conditions, some steroid glucuronides such as 6β-hydroxyandrosterone glucuronide or 6β-hydroxyetiocholanolone glucuronide [10] are resistant to hydrolysis on treatment with β-glucuronidase and so remain undetected in the subsequent GC-MS screens. Nevertheless, the completeness of the hydrolysis by β-glucuronidase is tested by hydrolysing a known sample of an isotopically labelled steroid glucuronide such as androsterone glucuronide [23].

In addition to the steroid glucuronides, steroid sulfates also contribute to a significant fraction of steroid metabolites although may not be routinely analysed in the anti-doping screens, as there is no standard method for hydrolysing steroid sulfates. Solvolysis is a general method of acid-catalysed hydrolysis of steroid conjugates that has been used in equine anti-doping analysis. This method is considered unsuitable by many for routine anti-doping screening since it can degrade analytes of interest [24] and generate a more complicated matrix [10, 22, 25-28]. Hence, enzymatic methods are preferred over chemical means due to their specificity and the mild conditions used that would have a lower capacity to degrade steroid analytes. As a result, an enzyme preparation from *Helix pomatia* is used in anti-doping. However, this enzyme preparation is a crude mixture that contains alternative activities that could lead to steroid conversion, degradation and artefact formation [29-31] . The enzyme preparations from *Patella vulgate* or *Haliotis rufescens* are also occasionally used [22, 29, 32] even though they exhibit weak sulfatase activity. Failure to hydrolyse steroid sulfates leads to an incomplete steroid profile and could result in imprecise conclusions. Therefore, these enzyme preparations are not suitable for routine analysis.

In contrast to GC-MS, liquid chromatography mass spectrometry (LC-MS) enables direct analysis of steroid conjugates (Figure 1.4) and has the advantage of avoiding the sample preparatory steps so enabling higher throughput in sample analysis [25, 33]. Nevertheless, LC-MS has the difficulty in detecting some analytes such as saturated steroids and steroid diols due to poor ionization properties [8, 28].

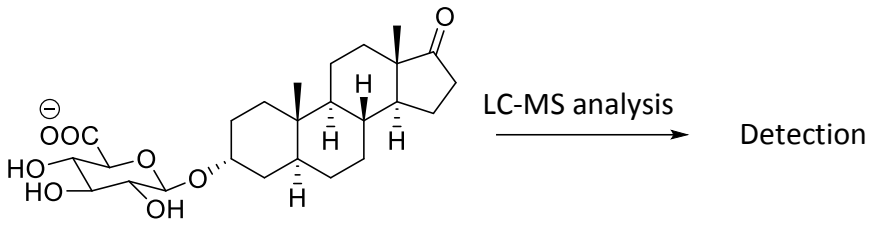


Figure 1.4: Direct detection of phase II metabolites by LC-MS

Followed by detection, confirmation is carried out by comparison with reference materials. Unfortunately, many reference materials of phase II metabolites that are required for confirmation purposes are unavailable. This makes hydrolysis of phase II metabolites necessary to liberate phase I metabolites or free steroids which can then be confirmed against the more widely available reference materials developed for anti-doping applications [28]. Therefore, hydrolysis of the steroid conjugates plays an important role in the anti-doping testing irrespective of the technique used for detection.

The recent advancements in mass spectrometry (MS) techniques such as quadrupole time-of-flight (QTOF), quadruple orbitrap (Q Exactive™) mass spectrometers used in conjunction with ultrahigh pressure liquid chromatography (UHPLC) are gaining popularity in the anti-doping arena. The UHPLC techniques provide higher chromatographic resolution while maintaining similar elution times compared to conventional LC. The QTOF, orbitrap devices enable data acquisition over a broad mass range, accurate mass measurements and high sensitivity [34].

These technological breakthroughs have assisted in implementing untargeted and open screening methods in anti-doping analysis. These methods scan for characteristic fragments or fragmentation modes of the metabolite class, instead of the individual metabolite itself. This allows detection of families of metabolites regardless of the minor structural changes they possess, hence would also be useful in identifying new metabolites derived from previously unknown substances. Metabolomics approaches enable the simultaneous detection of a broad range of small metabolites in biological samples that is performed in an untargeted mode [35, 36]. This approach is attracting

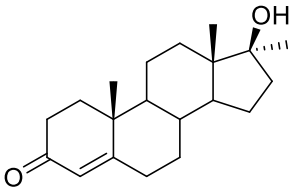
attention in doping control analysis as it is useful for the discovery of new metabolites or changes in the profile of endogenous metabolites in response to drug administration.

1.4 Importance of sulfate metabolites in anti-doping

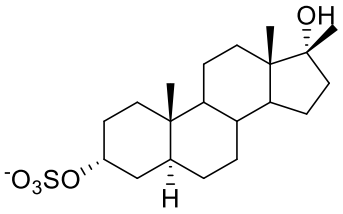
The growing importance of the detection of sulfate metabolites in anti-doping screens has been extensively discussed in previous literature [12, 18, 27, 37, 38]. The study of sulfate metabolites is important for the detection of both exogenous and endogenous steroids. Dehydroepiandrosterone (DHEA), the most abundant circulating steroid hormone in humans [39], interconverts in the forward and reverse sense with its sulfate conjugate (DHEAS), catalysed by sulfotransferase or sulfatase enzymes. Both DHEA and DHEAS are predominant precursors for androgens and estrogens where approximately 50% of androgens in men and 75% of estrogens in pre-menopausal and 100% post-menopausal women originate from DHEA and DHEAS [16, 40, 41]. Together DHEA and DHEAS form the largest steroid reservoir in the body and the DHEAS concentration is 10^3 - 10^4 times greater than estradiol and around 100 - 500 times greater than testosterone [42]. Moreover, the concentration of DHEAS in blood is about 300 times higher than free DHEA and could therefore affect the concentration of DHEA glucuronide conjugate (DHEAG). This will also be contingent on the rate of inter-conversion between DHEA and DHEAS [15, 37] compromising the effectiveness of the screening method that detects the free steroid by the deconjugation of the glucuronide fraction alone.

A study performed by Dehennin *et al.* [39] on DHEA metabolism detected the metabolites of DHEA before and after administration of an oral dose of DHEA. This study revealed that before DHEA administration, the sulfate conjugated DHEA and 3β-diols of DHEA were observed in abundance compared to their glucuronide equivalents. Whereas the other metabolites had similar levels of sulfate and glucuronide excretion or greater glucuronide conjugation. However, post-administration, the exogenous DHEA was preferentially converted into sulfate conjugated metabolites such as the sulfates of androsterone, etiocholanolone, epiandrosterone, the 3β-diols of DHEA, 5α-androstane-3α,17β-diol and 5β-androstane-3α,17β-diol. These were observed in high excess relative to the glucuronide metabolites.

Several previous studies highlight the importance of steroid sulfates as long-term markers that can be used to improve the retrospectivity of the AAS misuse. A recent WADA funded study performed by Gomez *et al*. [18] following an oral administration of methyltestosterone (MT) detected three diol mono-sulfate metabolites out of which 17β-methyl-5α-androstan-3α,17α-diol 3-sulfate (Figure 1.5) was identified as the metabolite with the longest detection window of up to 21 days after MT administration. This was an improvement in the retrospectivity of between 2 to 3 times compared to the previously reported long-term metabolite of MT, 17α-hydroxy-17β-methylandrostan-4,6-dien-3-one (up to 9 days).



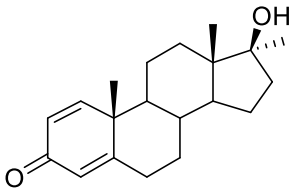
methyltestosterone



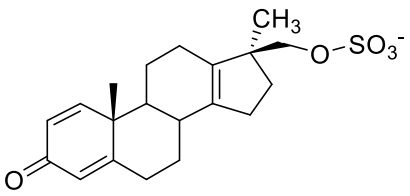
17β-methyl-5α-androstan-3α,17α-diol 3-sulfate

Figure 1.5: Methyltestosterone and its long-term sulfate metabolite

A separate study [43] conducted by the same group on the oral administration of metandienone detected seven sulfate metabolites where 18-nor-17β-hydroxymethyl-17α-methylandrost-1,4,13-triene-3-one sulfate (Figure 1.6) was detected up to 26 days after metandienone administration. This improved the detection window compared to the previously reported long-term metabolite, 18-nor-17β-hydroxymethyl-17α-methylandrost-1,4,13-triene-3-one glucuronide (14 days) that is used to identify metandienone misuse.



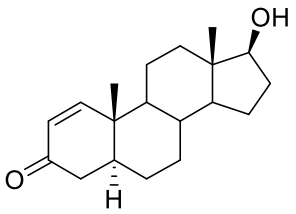
metandienone



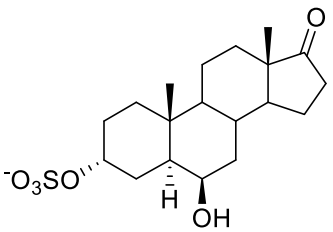
18-nor-17β-hydroxymethyl-17α-methylandrost-1,4,13-triene-3-one sulfate

Figure 1.6: Metandienone and its long-term sulfate metabolite

A similar study by Liu *et al.* [44] on oral administration of 1-testosterone identified a sulfate metabolite, 3α,6β-dihydroxy-5α-androstan-17-one 3-sulfate (Figure 1.7), as a long-term marker that could be detected up to 15 days after administration of 1-testosterone. This provided five times greater retrospectivity compared to the alternative long-term metabolites, 1-testosterone (<2 days), 5α-androst-1-ene-3α,17β-diol (up to 3 days) and 5α-androst-1-ene-3,17-dione (up to 3 days) identified as unconjugated metabolites in this study.



1-testosterone



3α,6β-dihydroxy-5α-androstan-17-one 3-sulfate

Figure 1.7: 1-testosterone and its long-term sulfate metabolite

Studies on 19-nor-steroids [27, 45-47] (Figure 1.8) also suggest that the sulfate conjugated metabolites are eliminated at a slower rate than the glucuronide conjugates during excretion. Once the excretion of glucuronides declines, the sulfate metabolites can become predominant. Further, the high levels of 19-nor-androsterone sulfate observed compared to the glucuronide counterpart over a longer period of time makes it a potential target for discrimination between metabolites of an endogenous and exogenous nature by the IRMS technique [46].

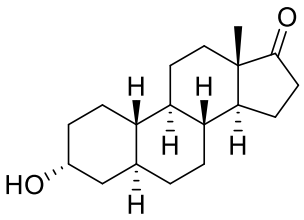
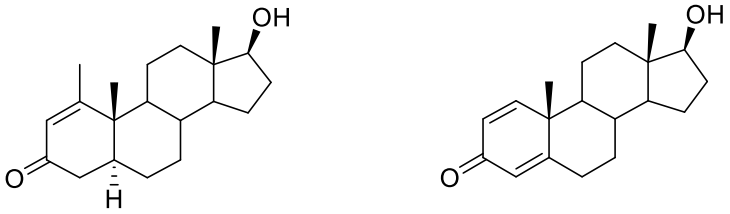


Figure 1.8: 19-nor-androsterone sulfate

A recent study performed by Piper *et al.* [12] on testosterone metabolism showed that the sulfated conjugates of the main urinary testosterone metabolites, have longer excretion times than their glucuronylated counterparts. While the metabolites of

glucuronide conjugates were observed up to two days after drug ingestion, only the sulfate conjugates were observed after eight days. Therefore, the sulfate metabolites can be used as long-term markers in the anti-doping screens. Moreover, this study shows that epiandrosterone, that is only excreted in its sulfated form, may serve as a useful target for IRMS to detect the administration of exogenous testosterone.

A recent WADA funded study of oral administration of methenolone [48] (Figure 1.9) identified four major sulfate metabolites including methenolone 17-sulfate and 3α-hydroxy-1-methylidene-5α-androstan-17-one 3-sulfate that had detection windows comparable with the hydrolysed glucuronide analogues. Another WADA funded study of the oral administration of boldenone [49] (Figure 1.9) identified two sulfate metabolites, boldenone sulfate and epiboldenone sulfate in minor concentrations. These sulfate metabolites were detected mainly after exogenous administration of boldenone. Hence these sulfate metabolites can be used as additional markers to identify the exogenous origin of boldenone that could reduce the requirement to analyse samples by IRMS.



methenolone boldenone

Figure 1.9: Methenolone and boldenone

An untargeted study by Boccard *et al.* [50] on steroidomics using UHPLC-QTOF-MS upon testosterone undecanoate administration revealed that the performance indices of the metabolites such as area under the curve and sensitivity are notably higher for androsterone sulfate (AS) and etiocholanolone sulfate (ECS), than for their glucuronide counterparts. Hence, they suggested that AS, ECS and an unidentified steroid sulfate, an isomer of ECS, AS or dihydrotestosterone sulfate, possibly epiandrosterone sulfate, could be used as additional markers for the detection of testosterone administration.

Therefore, including the sulfate metabolites is important to circumvent false negative results in the steroid profiles of the athletes and also to increase the retrospectivity for the detection of the misuse of some AASs. However, harsh acidic conditions used to deconjugate the sulfates is not suitable for routine analysis. Further,

confirmation techniques such as GC-CIRMS that differentiates the endogenous form exogenous steroids require careful sample clean-up prior to analysis. This could be challenging for sulfate metabolites that require severe acidic conditions for hydrolysis. Due to the emerging importance of sulfate metabolites in anti-doping and the need for mild methods for sulfate hydrolysis, attention is drawn toward suitable sulfatase enzymes as candidates for steroid sulfate hydrolysis.

1.5 Sulfatase family (EC 3.1.6)

The sulfatases catalyse the hydrolysis of a diverse range of sulfate ester substrates ranging from small steroids such as estrogen sulfates to complex cell surface carbohydrates such as glycosoaminoglycans. They are divided into different enzyme sub-classes [51] according to the type of substrate they can hydrolyse or sequence similarity to other known sulfates enzymes. For example; EC 3.1.6.1: arylsulfatases; EC 3.1.6.2: steryl-sulfatases (arylsulfatase C); EC 3.1.6.8: cerebroside-sulfatase (arylsulfatase A and E); EC 3.1.6.12: *N*-acetylgalactosamine-4-sulfatase (arylsulfatase B). Arylsulfatases (EC 3.1.6.1) are present in all kingdoms of life and catalyse the hydrolysis of sulfate esters. The historical nomenclature likely reflects the relative ease with which aromatic sulfate esters are hydrolysed. However, alkyl sulfate esters such as saturated steroid sulfates are also substrates for this enzyme family.

The bacterial and human sulfatases share a high degree of sequence and structure homology which consist of a globular monomer with a mixed α/β topology, however the human sulfatases are membrane anchored. The core polypeptide motif C/S-X-P-X-R is conserved throughout the entire enzyme class. It directs a post-translational modification of the conserved cysteine or serine residue to α-formyglycine (FGly). This residue is a unique feature observed in the sulfatases and some closely related enzyme families that is essential for the catalytic activity [52, 53].

Seventeen human sulfatase proteins and their corresponding genes have been identified [54], some of which were uncovered through the study of diseases [52, 55, 56] . The biological importance of sulfatases has been reported in relation to human diseases such as: leukodystrophy that can contribute to movement problems and developmental

delay, ichthyosis that results in dark scaly skin or scoliosis that results in abnormal curvature of spine due to a condition known as multiple sulfatase deficiency (MSD). The MSD is a rare autosomal disorder occurring in humans due to a mutation in the sulfatase modifying factor 1 (*SUMF1*), the gene that encodes the formylglycine generating enzyme (FGE).

1.5.1 Post-translational modification

A unique type of post-translational modification of a cysteine (Cys-type sulfatases in eukaryotes and some prokaryotes) or a serine (Ser-type sulfatases in some prokaryotes) to formylglycine (FGly) is a crucial functional feature for the catalytic activity of sulfatases. The cysteine or serine residue included in the signature sequence, C/S-X-P-X-R [52, 57-60] forms the recognition sequence for the generation of FGly in sulfatases. The Cys-type modification takes place in the cytoplasm whereas the Ser-type modification takes place in the periplasm. In eukaryotes, FGly generation occurs during or immediately after the translocation of sulfatase into the endoplasmic reticulum. Conversion of serine to FGly is merely an oxidation while that of cysteine goes through a presumed thio-aldehyde intermediate that requires an additional hydrolytic step to form FGly [53] (Figure 1.10).

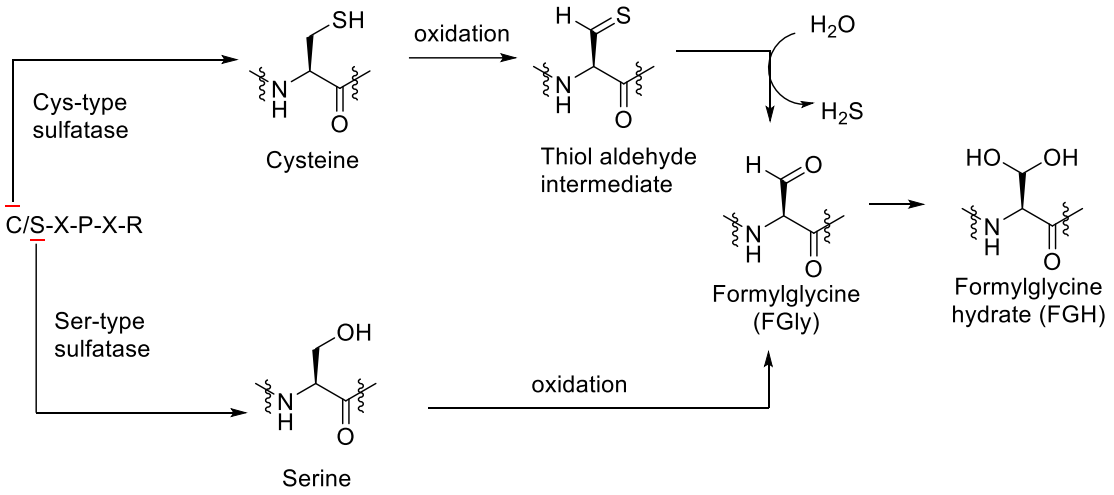


Figure 1.10: Post-translational modification in Cys-type and Ser-type sulfatases

The modification of cysteine to FGly by the formyl glycine generating enzyme (FGE) has been reported in previous literature [38, 52, 54]. It uses molecular oxygen and is a

cofactor independent catalysis. The FGly modification of the Ser-type sulfatases in prokaryotes is mediated by AtsB gene which is an iron-sulfur protein [61]. Studies by Dierks *et al.* [60, 61] reveal that it is unlikely that the Cys-type sulfatases undergo hydrolysis to serine prior to oxidation, as mutation of the critical cysteine to a serine in *Pseudomonas aeruginosa* arylsulfatase abolished the FGly generation. These studies also suggest that the modifying machinery for generation of FGly from cysteine and serine is distinct and highly specific for the respective residue. Nevertheless, having an essentially identical sequence neighbouring FGly makes it intriguing that the post-translational modification for FGly in bacteria has two separate modification machineries.

1.6 The arylsulfatase from *Pseudomonas aeruginosa* (PaS)

Pseudomonas aeruginosa is a common Gram-negative pathogenic bacterium found in many natural and artificial environments, not only in normal atmospheres but also in low oxygen atmospheres. The arylsulfatase from *P. aeruginosa* (PaS) (EC 3.1.6.1) is one of the most widely studied enzymes in the sulfatase family. Suppression of arylsulfatase expression in *P. aeruginosa* is observed in the presence of good sulfate sources such as inorganic sulfate, and its expression under sulfate starved conditions is reported [53, 62]. This shows its important role in bacterial growth under different sulfur regulated conditions. It has been suggested that the substrate specificity of such enzymes expressed in response to sulfur starved conditions could be quite broad, allowing flexibility in scavenging sulfur from the environment [53].

The promiscuity of PaS has been broadly discussed in previous literature [63-65] which is an important feature in steering enzyme evolution. Although the native substrate of PaS is unknown, the enzyme has been studied in relation to its hydrolytic activity of phosphate and phosphonate esters and small aromatic sulfates such as *para*-nitrophenyl sulfate (PNPS), *para*-nitrocatechol sulfate (PNCS) to show its catalytic promiscuity. However, the hydrolytic activity of PaS in relation to steroid sulfates had not been discussed widely. Studies by Stevenson *et al.* [32] revealed that WT-PaS could hydrolyse sulfate groups with a β configuration such as testosterone 17-sulfate, epiandrosterone 3-sulfate etc., however, under the screening conditions employed it

did not show detectable activity with α-configured steroid sulfates such as epitestosterone 17-sulfate, androsterone 3-sulfate and etiocholanolone 3-sulfate.

1.6.1 Structure of PaS

The crystal structure of native PaS has been determined at 1.3 Å by Boltes *et al.* [66]. It essentially comprises of two domains with a monomeric structure of 536 amino acids. The overall dimensions of the globular PaS are 55 x 55 x 50 Å with a molecular weight of 59 kDa. The C-terminal domain is the smaller of the two domains and contains a four stranded anti-parallel β sheet and a solvent exposed terminal α helix (Figure 1.11).

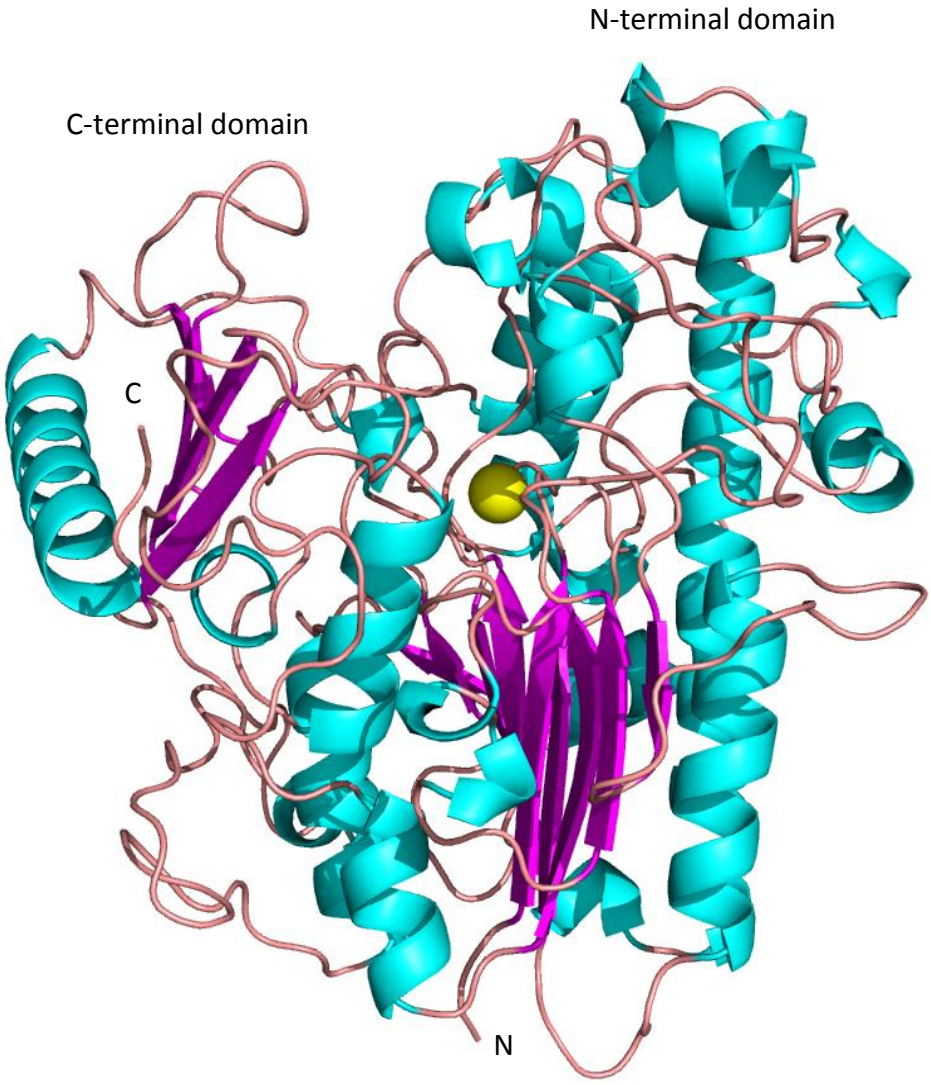


Figure 1.11: Crystal structure of PaS (PDB Code: 1HDH [66]) (blue cylinders: α helices, magenta arrows: β-sheets) characterized by two domains. Calcium ion (light green) is shown that is located in the substrate binding pocket.

The N-terminal domain has an α/β topology that contains a ten stranded β sheet flanked by several α helices. The active site is located at the end of a narrow cleft between the C terminal β-sheet and a number of N terminal α-helices [52, 66]. The structure contains no disulfide bridges and no cysteine residues as the only one expressed (Cys51) is post-translationally modified to FGly which is present in its hydrated form (FGH). A metal cation was located in the crystal structure in the vicinity of FGH and that was determined to be calcium (by comparison of B values and bond valence calculations). The crystal structure contained a sulfate anion in the active site that was non-covalently bound to the protein and eight more sulfate anions located on the surface of the protein originated from the precipitation reagent ($(NH_4)_2SO_4$) used for crystallization.

Three other x-ray crystal structures of sulfatases have been solved to date. They are: human arylsulfatase A (HARSA, cerebroside-sulfatase: EC 3.1.6.8) at 2.1 Å resolution [67], human arylsulfatase B (HARSB, N-acetylgalactosamine-4-sulfatase: EC 3.1.6.12) at 2.5 Å resolution [68], arylsulfatase C (HARSC, steryl-sulfatase: EC 3.1.6.2) at 2.6 Å resolution [69]. All these sulfatases show strikingly similar structure with a superimposable catalytic domain and a high degree of sequence similarity [52, 70]. However, none of the structures were solved with a native substrate bound.

1.6.2 The functional residues and the sulfate ester hydrolysis mechanism

The substrate binding pocket is made up of ten highly interconnected residues and a Ca^{2+} ion (Figures 1.12 and 1.13). The ten residues in PaS comprise five basic residues: Lys113, Lys375, His115, His211, Arg55; three acidic Asp residues: Asp317, Asp14, Asp13; and two polar residues: Asn318 and the FGly51. It was observed that FGly was present in its hydrated form (FGH), which is believed to be essential for sulfatase activity. The residues Lys113, Lys375, His211 and the divalent calcium ion are involved in binding the anionic sulfate and by extension, sulfate ester substrates. These interactions of the positively charged residues with sulfate group make the sulfur center more electrophilic by withdrawing the electron density. These residues are also suggested to orient the tetrahedral sulfate ester for an in-line nucleophilic substitution by FGH.

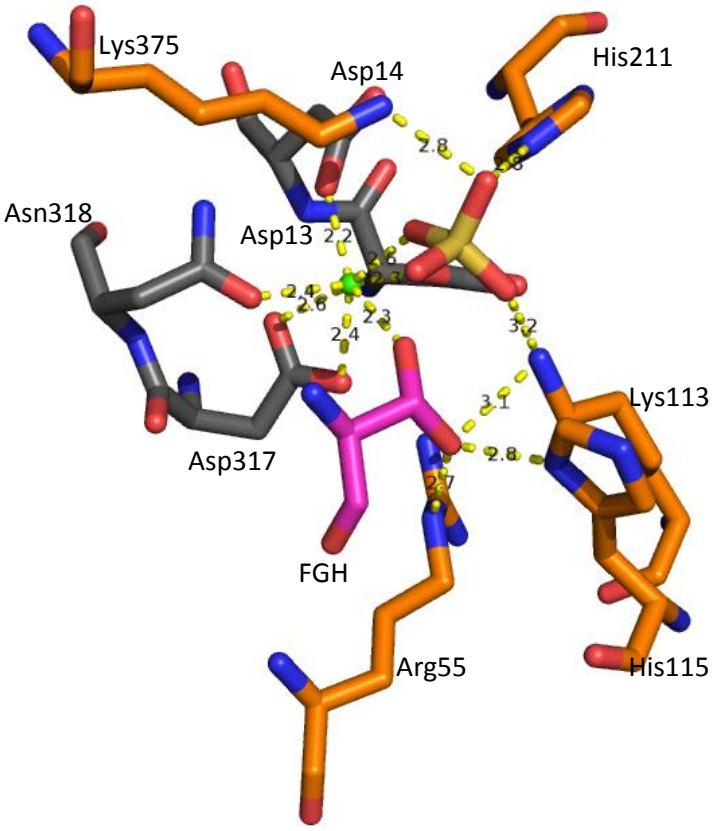


Figure 1.12: The functional residues of PaS enzyme as observed in the crystal structure (1HDH) with the bound sulfate, the FGH residue (pink) and Ca^{2+} ion (green). The residues interacting with sulfate and/or FGH are coloured in orange and the interactions are shown by yellow dashed lines.

The Ca^{2+} ion plays an important role in the active site by coordinating to both FGH and the substrate sulfate group and allowing sufficient proximity for the nucleophilic substitution to take place. The Ca^{2+} ion is coordinated by seven-fold ligands including the substrate, and holds a distorted pentagonal bipyramid with the Asp13 and Asn318 coordinating in the axial directions while the oxygen atoms of Asp 317, Asp14, FGH and sulfate coordinating in the equatorial directions. Besides FGH, the three aspartate residues and the asparagine residue are bound to the metal cation, which helps define the catalytic pocket. In addition to the Ca^{2+} ion, the residues Lys113, Arg55, His115 are involved in FGH stabilization.

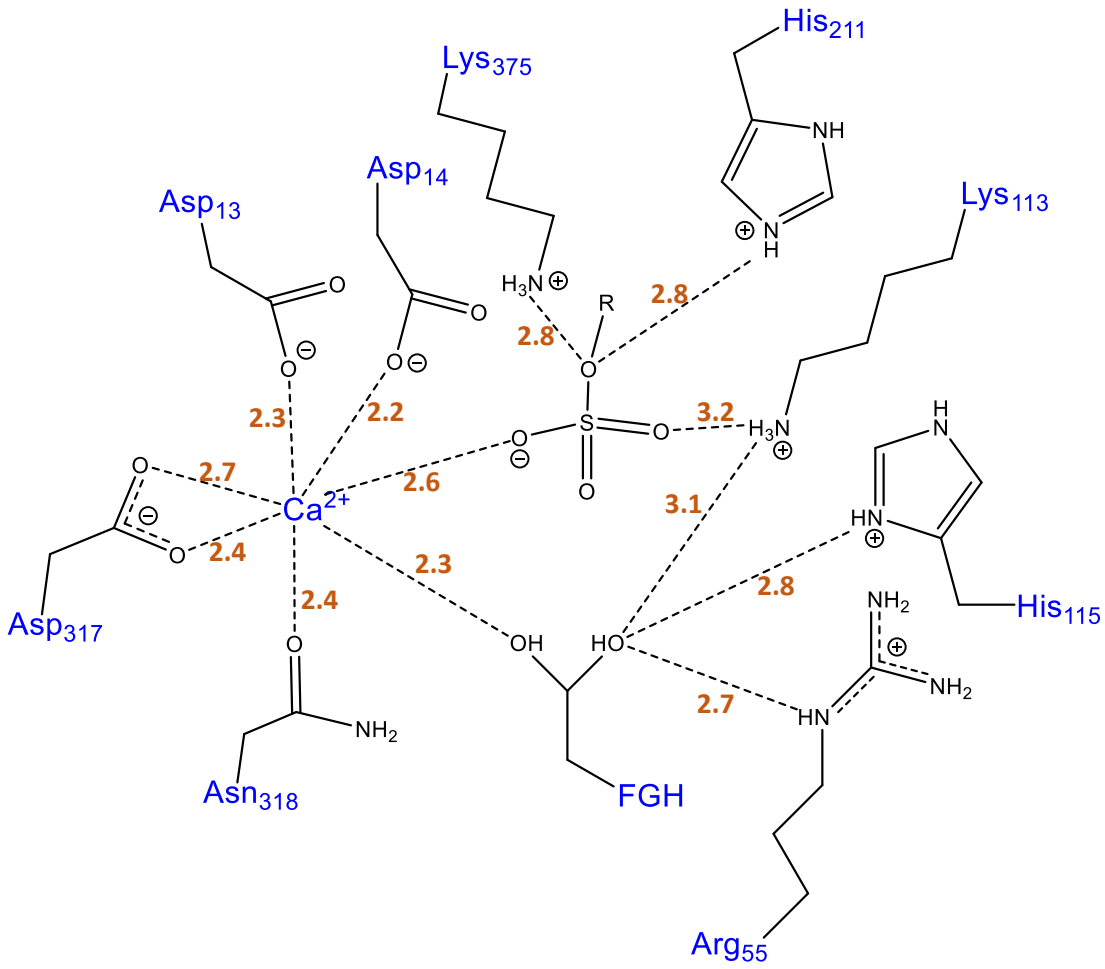


Figure 1.13: A 2D view of the active site residues of PaS enzyme with the interactions (dashed lines) with the substrate, Ca^{2+} ion and FGH. The distance measurements are displayed in Å (brown).

Two mechanisms for sulfate ester hydrolysis have been proposed [52, 67, 68]. The first involves an addition-hydrolysis (AH) mechanism and was proposed based on the HARSB crystal structure [68] that contains a covalently bound sulfate unit. The negatively charged sulfate ester oxygen adds to the electrophilic aldehyde group of FGly to form a sulfate di-ester, which is then hydrolysed through nucleophilic substitution by a water molecule to release the product. It is followed by the regeneration of the aldehyde group, releasing sulfate and completing the catalytic cycle (Figure 1.14). The crystal structure of HARSB assumed the FGH-sulfate ester to be the resting state that led to the proposal of the AH mechanism.

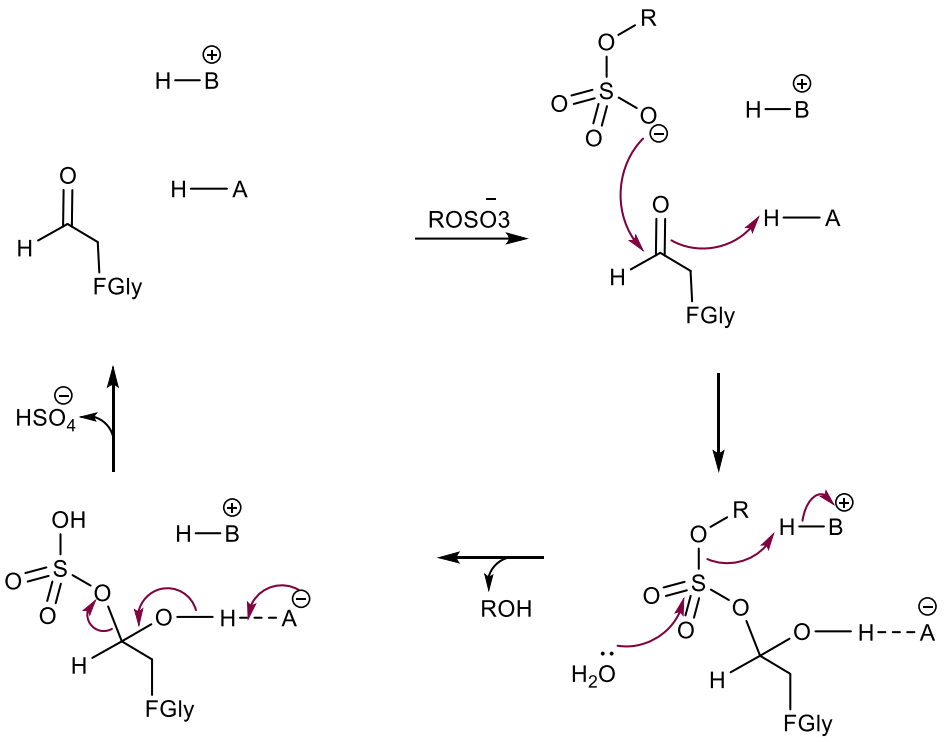


Figure 1.14: Addition hydrolysis mechanism of sulfate esters by FGly in the form of an aldehyde.

The second mechanism involves a transesterification-elimination-hydration (TEH) [67] sequence that was proposed based on the HARSA crystal structure. In the crystal structure, the post-translationally modified cysteine residue was observed as formylglycine hydrate (FGH) in the resting state of the enzyme that led to proposing TEH mechanism. This mechanism is illustrated in Figure 1.15 in relation to the residue numbering in PaS enzyme. The residues Lys375, His211, His115, Asp13 together with the Ca^{2+} help in coordinating to FGH, sulfate moiety or both. This helps orient FGH and substrate for an in-line nucleophilic substitution. The four residues also assist in the proton exchange acting as general acids or bases. In this mechanism, the FGly is first hydrated to produce a geminal-diol (FGH). The Asp13 residue acts as a general base where it abstracts a proton from one of the hydroxyl groups from the geminal-diol of FGH. This leads to the nucleophilic substitution where the oxygen of FGH acts as a nucleophile and substitutes (S_N2) at the sulfur atom of the sulfate ester. This results in the hydrolysis of the sulfate ester S-OR bond, releasing the free alcohol, which abstracts a proton from His211 that acts as a general acid. Next His115 abstracts a proton from

the second hydroxyl group of the sulfated FGH intermediate that helps to reform FGly by eliminating the sulfate completing the catalytic cycle [52] (Figure 1.15).

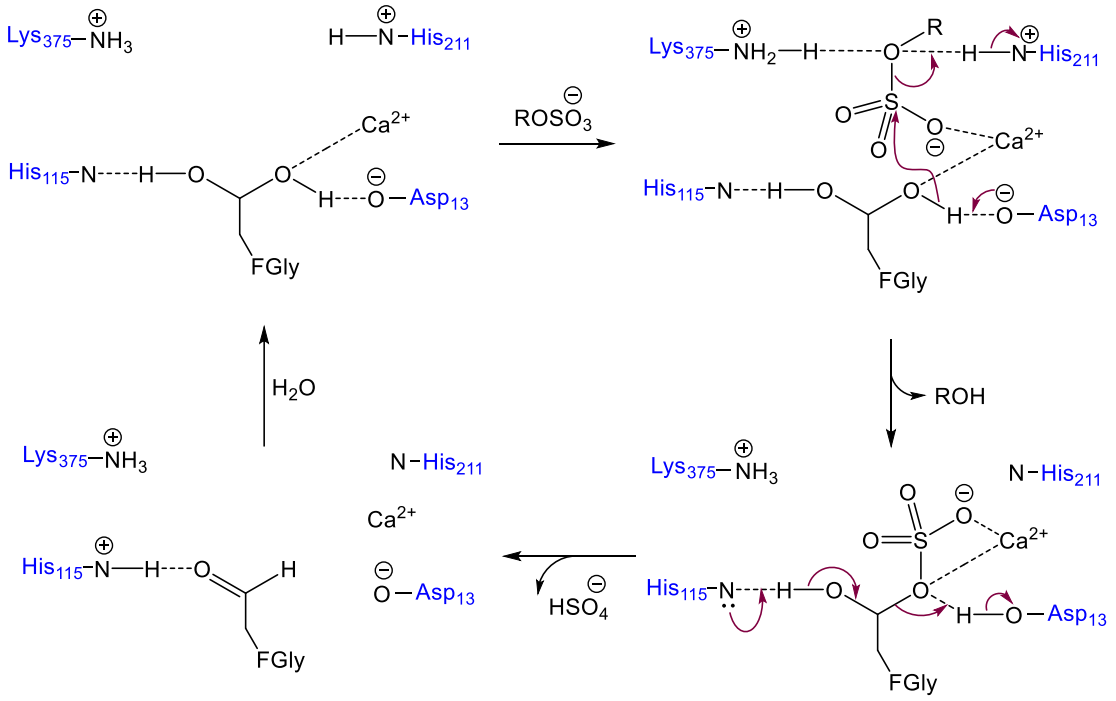


Figure 1.15: Transesterification elimination mechanism of sulfate esters by FGly in the form of an aldehyde hydrate.

Evidence in support of the second mechanism includes a study on a mutational change performed on the crucial cysteine residue to a serine that blocks the FGly generation[71]. This progressed the reaction as far as the elimination of ROH to form a serine sulfate ester intermediate. However, it does not release HSO_4^- to complete the catalytic cycle. This study reveals the importance of the second hydroxyl group of FGH to eliminate the sulfate group. A similar mutational study [72] performed with [^{35}S]-*p*-nitrocatechol sulfate revealed that the serine mutants had the radioactive sulfate group trapped in the active site while the unmutated enzyme did not show traces of radioactivity. These studies also discourage the addition-hydrolysis mechanism, as it is highly unlikely a nucleophilic addition occurs at the saturated serine residue by the oxygen atom of sulfate. Moreover, the crystallographic structure of PaS has revealed that the FGly residue is present in its hydrated form rather than the unsaturated FGly form in the resting state that is near to but independent from the sulfate moiety. These observations support the transesterification elimination mechanism.

1.7 Previous work

The PaS enzyme has been previously investigated in the McLeod laboratory at the Australian National University. The hydrolytic activity of WT-PaS had been compared to four other commercially available crude enzyme preparations over a range of substrates representing a diversity of structure and stereochemistry (Figure 1.16).

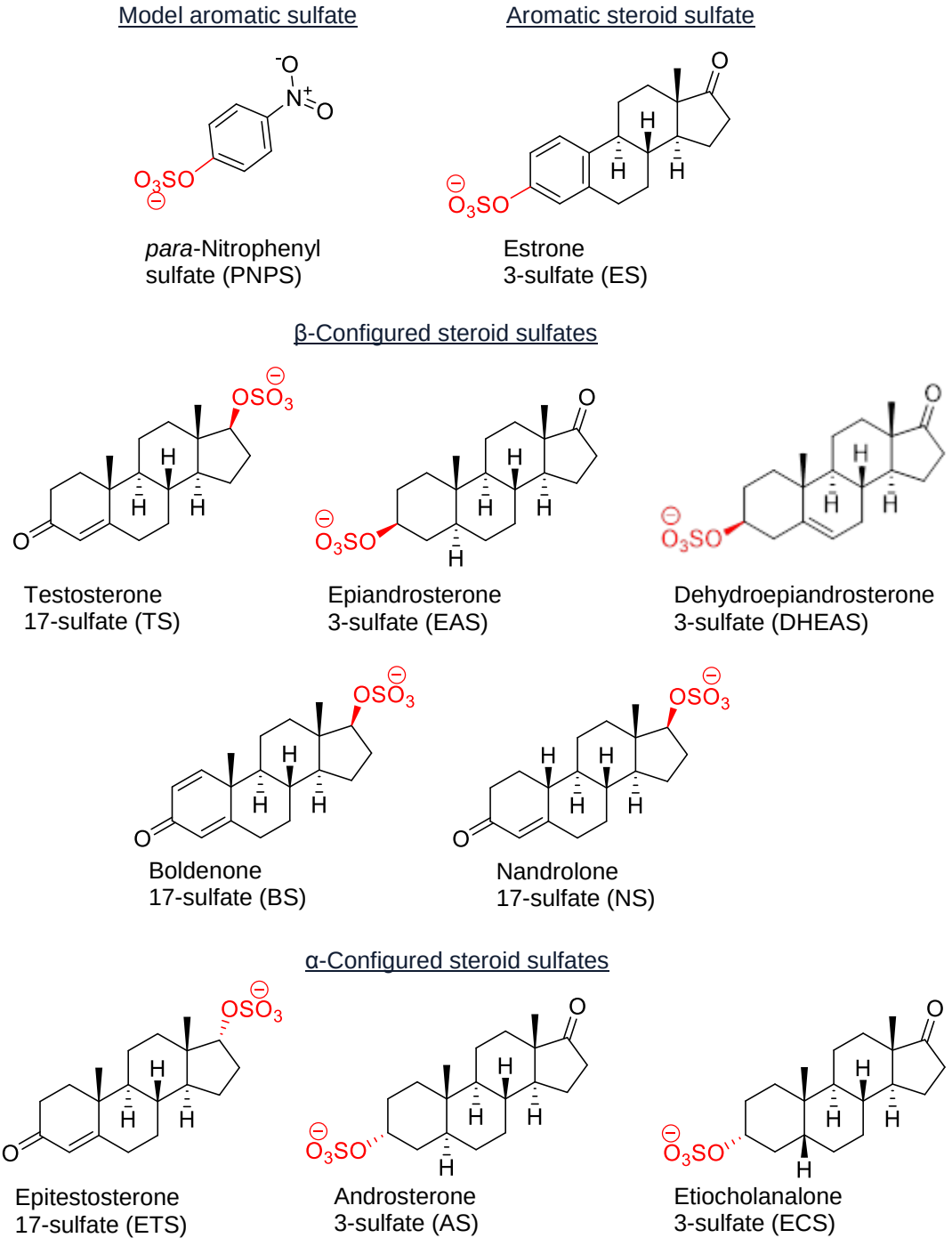


Figure 1.16: Substrates tested on native PaS

This study reported by Stevenson *et al.* [32] shows that PaS has comparable activity with the commercially available sulfatase from *Helix pomatia* (HpS). However, the sulfatase from *Helix pomatia* is a crude enzyme preparation that can also exhibit additional activities such as oxidase, reductase or glucuronidase activity. The other commercial sulfatases studied: sulfatases from *Haliotis rufescens* (HrS), *Klebsiella pneumonia* (KpS), or *Patella vulgata* (PvS), have shown a narrower substrate range and lower activity per gram of protein. Of these sulfatase preparations, only KpS has a known gene sequence but none has structural information. In this study, PaS was found to show activity towards the hydrolysis of model aryl sulfates, such as *para*-nitrophenyl sulfate (PNPS), used for its chromogenic properties, and some steroid sulfates, such as estrone 3-sulfate (ES), epiandrosterone 3-sulfate (EAS) and testosterone 17-sulfate (TS). However, the maximum hydrolytic activity (V_{max}) for testosterone 17-sulfate is five orders of magnitude lower than that observed for PNPS. The enzyme kinetics of native PaS for PNPS, ES, EAS and TS were measured and compared with that of the crude enzyme preparation of HpS.

The previous studies demonstrate that PaS requires further improvements to be used in an anti-doping context. However, engineering efforts are not straightforward because the orientation of steroid sulfates in the cleft leading to the catalytic site is unknown. Crystallographic studies to determine substrate binding would be challenging due to potential substrate turnover, and the high concentration of sulfate used in known crystallization conditions that would be difficult to displace. Therefore, analysis of PaS with specific mutations could be of use to determine important residues for substrate binding and thereby to identify the binding site. Nevertheless, the promiscuity of PaS makes it a good candidate for improvement and this is possible since the sequence is known and the gene can be readily modified. The enzyme can be expressed from a recombinant gene with a His-tag for affinity purification therefore can be easily purified, in a stable and soluble form. The observed expression of native PaS under sulfur starved conditions makes it a better candidate as it is likely to have evolved for a broad substrate range. In addition, the crystal structure of PaS has been determined by X-ray crystallography [66]. For these reasons, PaS was selected as the suitable candidate for protein engineering.

1.8 References

[1] Moss, G. P. (1989) The nomenclature of steroids - recommendations 1989, *Eur. J. Biochem. 186*, 429-458.

[2] Athanasiadou, I., Angelis, Y. S., Lyris, E., Georgakopoulos, C., Athanasiadou, I., and Georgakopoulos, C. (2013) Chemical derivatization to enhance ionization of anabolic steroids in LC-MS for doping-control analysis, *TrAC-Trend Anal. Chem. 42*, 137-156.

[3] Saha, S., Mandal, M. K., Nonami, H., and Hiraoka, K. (2014) Direct analysis of anabolic steroids in urine using Leidenfrost phenomenon assisted thermal desorption-dielectric barrier discharge ionization mass spectrometry, *Anal. Chim. Acta 839*, 1-7.

[4] Saudan, C., Baume, N., Robinson, N., Avois, L., Mangin, P., and Saugy, M. (2006) Testosterone and doping control, *Brit. J. Sport Med. 40*, 21-24.

[5] World anti-doping agency: 2017 prohibited list, http://www.usada.org/substances/prohibited-list/ (Accessed 05.12.2016).

[6] Prendergast, H. M., Bannen, T., Erickson, T. B., and Honore, K. R. (2003) The toxic torch of the modern olympic games, *Vet. Hum. Toxicol. 45*, 97-102.

[7] Schanzer, W. (1996) Metabolism of anabolic androgenic steroids, *Clin. Chem. 42*, 1001-1020.

[8] McKinney, A. R., Cawley, A. T., Young, E. B., Kerwick, C. M., Cunnington, K., Stewart, R. T., Ambrus, J. I., Willis, A. C., and McLeod, M. D. (2013) The metabolism of anabolic-androgenic steroids in the greyhound, *Bioanalysis 5*, 769-781.

[9] Vore, M., and Slikker, W. (1985) Steroid D-ring glucuronides - a new class of cholestatic agents, *Trends Pharmacol. Sci. 6*, 256-259.

[10] Fabregat, A., Pozo, O. J., Marcos, J., Segura, J., and Ventura, R. (2013) Use of LC-MS/MS for the open detection of steroid metabolites conjugated with glucuronic acid, *Anal. Chem. 85*, 5005-5014.

[11] Dehennin, L., and Matsumoto, A. M. (1993) Long-term administration of testosterone enanthate to normal men - alterations of the urinary profile of androgen metabolites potentially useful for detection of testosterone misuse in sport, *J. Steroid Biochem. 44*, 179-189.

[12] Piper, T., Schanzer, W., and Thevis, M. (2016) Genotype-dependant metabolism of exogenous testosterone - new biomarkers result in prolonged detectability, *Drug Test. Anal*. *8*, 1163-1173.

[13] Baulieu, E. E., and Mauvais-Jarvis, P. (1964) Studies on testosterone metabolism. Ii. metabolism of testosterone-4-14c and androst-4-ene-3, 17-dione-1,2-3h, *J. Biol. Chem. 239*, 1578-1584.

[14] Gomes, R. L., Meredith, W., Snape, C. E., and Sephton, M. A. (2009) Conjugated steroids: analytical approaches and applications, *Anal. Bioanal. Chem. 393*, 453-458.

[15] Endoh, A., Kristiansen, S. B., Casson, P. R., Buster, J. E., and Hornsby, P. J. (1996) The zona reticularis is the site of biosynthesis of dehydroepiandrosterone and dehydroepiandrosterone sulfate in the adult human adrenal cortex resulting from its low expression of 3 beta-hydroxysteroid dehydrogenase, *J. Clin. Endocr. Metab. 81*, 3558-3565.

[16] Burger, H. G. (2002) Androgen production in women, *Fertil. Steril. 77*, S3-S5.

[17] Thevis, M., Fussholler, G., Geyer, H., Rodchenkov, G., Mareck, U., Sigmund, G., Koch, A., Thomas, A., and Schanzer, W. (2006) Detection of stanozolol and its major metabolites in human urine by liquid chromatography-tandem mass spectrometry, *Chromatographia 64*, 441-446.

[18] Gomez, C., Pozo, O. J., Marcos, J., Segura, J., and Ventura, R. (2013) Alternative long-term markers for the detection of methyltestosterone misuse, *Steroids 78*, 44-52.

[19] Marcos, J., and Pozo, O. J. (2016) Current LC-MS methods and procedures applied to the identification of new steroid metabolites, *J. Steroid Biochem. 162*, 41-56.
[20] Perry, P. J., MacIndoe, J. H., Yates, W. R., Scott, S. D., and Holman, T. L. (1997) Detection of anabolic steroid administration: ratio of urinary testosterone to epitestosterone vs the ratio of urinary testosterone to luteinizing hormone, *Clin. Chem. 43*, 731-735.
[21] Palonek, E., Gottlieb, C., Garle, M., Bjorkhem, I., and Carlstrom, K. (1995) Serum and urinary markers of exogenous testosterone administration, *J. Steroid Biochem. 55*, 121-127.
[22] Gomes, R. L., Meredith, W., Snape, C. E., and Sephton, M. A. (2009) Analysis of conjugated steroid androgens: deconjugation, derivatisation and associated issues, *J. Pharmaceut. Biomed. 49*, 1133-1140.
[23] World anti-doping agency: athlete biological passport operating guidelines version 4.0, https://www.wada-ama.org/en/resources/athlete-biological-passport/athlete-biological-passport-abp-operating-guidelines (Accessed 20.12.2016).
[24] Pedersen, M., Frandsen, H. L., and Andersen, J. H. (2017) Optimised deconjugation of androgenic steroid conjugates in bovine urine, *Food Addit. Contam. A, DOI: 10.1080/19440049.2016.1276637*
[25] Robards, K., and Towers, P. (1990) Chromatography as a reference technique for the determination of clinically important steroids, *Biomed. Chromatogr. 4*, 1-19.
[26] Jacobsohn, G. M., and Lieberman, S. (1962) Studies on chemical cleavage of urinary glucuronosides of 17-ketosteroids, *J. Biol. Chem. 237*, 1469-1475.
[27] Piper, T., Schanzer, W., and Thevis, M. (2016) Revisiting the metabolism of 19-nortestosterone using isotope ratio and high resolution/high accuracy mass spectrometry, *J. Steroid Biochem. 162*, 80-91.
[28] Scarth, J. P., Teale, P., and Kuuranne, T. (2011) Drug metabolism in the horse: a review, *Drug. Test. Anal. 3*, 19-53.
[29] Gomez, C., Fabregat, A., Pozo, O. J., Marcos, J., Segura, J., and Ventura, R. (2014) Analytical strategies based on mass spectrometric techniques for the study of steroid metabolism, *Trac-Trend Anal. Chem. 53*, 106-116.
[30] Andersen, D. W., and Linnet, K. (2014) Screening for anabolic steroids in urine of forensic cases using fully automated solid phase extraction and LC-MS-MS, *J. Anal. Toxicol. 38*, 637-644.
[31] Hauser, B., Schulz, D., Boesch, C., and Deschner, T. (2008) Measuring urinary testosterone levels of the great apes - Problems with enzymatic hydrolysis using *Helix pomatia* juice, *Gen. Comp. Endocr. 158*, 77-86.
[32] Stevenson, B. J., Waller, C. C., Ma, P., Li, K., Cawley, A. T., Ollis, D. L., and McLeod, M. D. (2015) *Pseudomonas aeruginosa* arylsulfatase: a purified enzyme for the mild hydrolysis of steroid sulfates, *Drug Test. Anal. 7,* 903-911.
[33] Galuska, C. E., Hartmann, M. F., Sanchez-Guijo, A., Bakhaus, K., Geyer, J., Schuler, G., Zimmer, K. P., and Wudy, S. A. (2013) Profiling intact steroid sulfates and unconjugated steroids in biological fluids by liquid chromatography-tandem mass spectrometry (LC-MS-MS), *Analyst 138*, 3792-3801.
[34] Boccard, J., Badoud, F., Jan, N., Nicoli, R., Schweizer, C., Pralong, F., Veuthey, J. L., Baume, N., Rudaz, S., and Saugy, M. (2014) Untargeted profiling of urinary steroid metabolites after testosterone ingestion: opening new perspectives for antidoping testing, *Bioanalysis 6*, 2523-2536.
[35] Raro, M., Ibanez, M., Gil, R., Fabregat, A., Tudela, E., Deventer, K., Ventura, R., Segura, J., Marcos, J., Kotronoulas, A., Joglar, J., Farre, M., Yang, S., Xing, Y., Van Eenoo, P., Pitarch, E., Hernandez, F., Sancho, J. V., and Pozo, O. J. (2015) Untargeted metabolomics in doping control: detection of new markers of testosterone misuse by ultrahigh performance liquid chromatography coupled to high-resolution mass spectrometry, *Anal. Chem. 87*, 8373-8380.

[36] Dervilly-Pinel, G., Weigel, S., Lommen, A., Chereau, S., Rambaud, L., Essers, M., Antignac, J. P., Nielen, M. W., and Le Bizec, B. (2011) Assessment of two complementary liquid chromatography coupled to high resolution mass spectrometry metabolomics strategies for the screening of anabolic steroid treatment in calves, *Anal. Chim. Acta 700*, 144-154.

[37] Cawley, A. T., Hine, E. R., Trout, G. J., George, A. V., and Kazlauskas, R. (2004) Searching for new markers of endogenous steroid administration in athletes: "looking outside the metabolic box", *Forensic Sci. Int*. *143*, 103-114.

[38] Balcells, G., Gomez, C., Garrostas, L., Pozo, O. J., and Ventura, R. (2016) Sulfate metabolites as alternative markers for the detection of 4-chlorometandienone misuse in doping control, *Drug Test. Anal*. *DOI: 10.1002/dta.2101*

[39] Dehennin, L., Ferry, M., Lafarge, P., Peres, G., and Lafarge, J. P. (1998) Oral administration of dehydroepiandrosterone to healthy men: Alteration of the urinary androgen profile and consequences for the detection of abuse in sport by gas chromatography mass spectrometry, *Steroids 63*, 80-87.

[40] Labrie, F., Belanger, A., Cusan, L., Gomez, J. L., and Candas, B. (1997) Marked decline in serum concentrations of adrenal C19 sex steroid precursors and conjugated androgen metabolites during aging, *J. Clin. Endocr. Metab. 82*, 2396-2402.

[41] Labrie, F. (1991) Intracrinology, *Mol. Cell. Endocrinol. 78*, C113-C118.

[42] Labrie, F., Luu-The, V., Belanger, A., Lin, S. X., Simard, J., Pelletier, G., and Labrie, C. (2005) Is dehydroepiandrosterone a hormone?, *J. Endocrinol. 187*, 169-196.

[43] Gomez, C., Pozo, O. J., Garrostas, L., Segura, J., and Ventura, R. (2013) A new sulphate metabolite as a long-term marker of metandienone misuse, *Steroids 78*, 1245-1253.

[44] Liu, Y., Lu, J. H., Yang, S., Xu, Y. X., and Wang, X. B. (2015) A new potential biomarker for 1-testosterone misuse in human urine by liquid chromatography quadruple time-of-flight mass spectrometry, *Anal. Methods-UK 7*, 4486-4492.

[45] Tseng, Y. L., Sun, C. Y., and Kuo, F. H. (2006) Detection and quantification of glucuro- and sulfoconjugated metabolites in human urine following oral administration of xenobiotic 19-norsteroids, *Steroids 71*, 817-827.

[46] Strahm, E., Baume, N., Mangin, P., Saugy, M., Ayotte, C., and Saudan, C. (2009) Profiling of 19-norandrosterone sulfate and glucuronide in human urine: implications in athlete's drug testing, *Steroids 74*, 359-364.

[47] Torrado, S., Roig, M., Farre, M., Segura, J., and Ventura, R. (2008) Urinary metabolic profile of 19-norsteroids in humans: glucuronide and sulphate conjugates after oral administration of 19-nor-4-androstenediol, *Rapid Commun. Mass Sp. 22*, 3035-3042.

[48] Fragkaki, A. G., Angelis, Y. S., Kiousi, P., Georgakopoulos, C. G., and Lyris, E. (2015) Comparison of sulfo-conjugated and gluco-conjugated urinary metabolites for detection of methenolone misuse in doping control by LC-HRMS, GC-MS and GC-HRMS, *J. Mass Spectrom*. *50*, 740-748.

[49] Gomez, C., Pozo, O. J., Geyer, H., Marcos, J., Thevis, M., Schanzer, W., Segura, J., and Ventura, R. (2012) New potential markers for the detection of boldenone misuse, *J. Steroid Biochem. 132*, 239-246.

[50] Boccard, J., Badoud, F., Grata, E., Ouertani, S., Hanafi, M., Mazerolles, G., Lanteri, P., Veuthey, J. L., Saugy, M., and Rudaz, S. (2011) A steroidomic approach for biomarkers discovery in doping control, *Forensic Sci. Int. 213*, 85-94.

[51] BRENDA the comprehensive enzyme information system, http://www.brenda-enzymes.info/ (Accessed 26.10.2016)

[52] Hanson, S. R., Best, M. D., and Wong, C. H. (2004) Sulfatases: Structure, mechanism, biological activity, inhibition, and synthetic utility, *Angew. Chem. Int. Edit. 43*, 5736-5763.

[53] Kertesz, M. A. (2000) Riding the sulfur cycle - metabolism of sulfonates and sulfate esters in Gram-negative bacteria, *Fems Microbiol. Rev. 24*, 135-175.

[54] Bojarova, P., and Williams, S. J. (2008) Sulfotransferases, sulfatases and formylglycine-generating enzymes: a sulfation fascination, *Curr. Opin. Chem. Biol. 12*, 573-581.
[55] Diez-Roux, G., and Ballabio, A. (2005) Sulfatases and human disease, *Annu. Rev. Genom. Hum. G. 6*, 355-379.
[56] Stoll, C., and Eyer, D. (1999) A syndrome of congenital ichthyosis, hypogonadism, small stature, facial dysmorphism, scoliosis and myogenic dystrophy, *Ann. Genet.-Paris 42*, 45-50.
[57] Knaust, A., Schmidt, B., Dierks, T., von Bulow, R., and von Figura, K. (1998) Residues critical for formylglycine formation and/or catalytic activity of arylsulfatase A, *Biochemistry-US 37*, 13941-13946.
[58] Dierks, T., Lecca, M. R., Schlotterhose, P., Schmidt, B., and von Figura, K. (1999) Sequence determinants directing conversion of cysteine to formylglycine in eukaryotic sulfatases, *Embo J. 18*, 2084-2091.
[59] Dierks, T., Lecca, M. R., Schmidt, B., and von Figura, K. (1998) Conversion of cysteine to formylglycine in eukaryotic sulfatases occurs by a common mechanism in the endoplasmic reticulum, *FEBS lett. 423*, 61-65.
[60] Dierks, T., Miech, C., Hummerjohann, J., Schmidt, B., Kertesz, M. A., and von Figura, K. (1998) Posttranslational formation of formylglycine in prokaryotic sulfatases by modification of either cysteine or serine, *J. Biol. Chem. 273*, 25560-25564.
[61] Marquordt, C., Fang, Q. H., Will, E., Peng, J. H., von Figura, K., and Dierks, T. (2003) Posttranslational modification of serine to formylglycine in bacterial sulfatases - recognition of the modification motif by the iron-sulfur protein AtsB, *J. Biol. Chem. 278*, 2212-2218.
[62] Hummerjohann, J., Kuttel, E., Quadroni, M., Ragaller, J., Leisinger, T., and Kertesz, M. A. (1998) Regulation of the sulfate starvation response in *Pseudomonas aeruginosa*: role of cysteine biosynthetic intermediates, *Microbiol.-Sgm 144*, 1375-1386.
[63] Luo, J., van Loo, B., and Kamerlin, S. C. (2012) Examining the promiscuous phosphatase activity of *Pseudomonas aeruginosa* arylsulfatase: a comparison to analogous phosphatases, *Proteins 80*, 1211-1226.
[64] Barrozo, A., Duarte, F., Bauer, P., Carvalho, A. T., and Kamerlin, S. C. (2015) Cooperative electrostatic interactions drive functional evolution in the alkaline phosphatase superfamily, *J. Am. Chem. Soc. 137*, 9061-9076.
[65] Luo, J., van Loo, B., and Kamerlin, S. C. (2012) Catalytic promiscuity in *Pseudomonas aeruginosa* arylsulfatase as an example of chemistry-driven protein evolution, *FEBS lett. 586*, 1622-1630.
[66] Boltes, I., Czapinska, H., Kahnert, A., von Bulow, R., Dierks, T., Schmidt, B., von Figura, K., Kertesz, M. A., and Uson, I. (2001) 1.3 angstrom structure of arylsulfatase from *Pseudomonas aeruginosa* establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family, *Structure 9*, 483-491.
[67] Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieselmann, V., von Figura, K., and Saenger, W. (1998) 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-US 37*, 3654-3664.
[68] Bond, C. S., Clements, P. R., Ashby, S. J., Collyer, C. A., Harrop, S. J., Hopwood, J. J., and Guss, J. M. (1997) Structure of a human lysosomal sulfatase, *Structure 5*, 277-289.
[69] Hernandez-Guzman, F. G., Higashiyama, T., Pangborn, W., Osawa, Y., and Ghosh, D. (2003) Structure of human estrone sulfatase suggests functional roles of membrane association, *J. Biol. Chem. 278*, 22989-22997.
[70] Stressler, T., Seitl, I., Kuhn, A., and Fischer, L. (2016) Detection, production, and application of microbial arylsulfatases, *Appl. Microbiol. Biot. 100*, 9053-9067.

[71] Williams, S. J., Denehy, E., and Krenske, E. H. (2014) 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. 79*, 1995-2005.
[72] Recksiek, R., Selmer, T., Dierks, T., Schmidt, B., and von Figura, K. (1998) Sulfatases, trapping of the sulfated enzyme intermediate by substituting the active site formylglycine, *J. Biol. Chem. 273*, 6096-6103.

Chapter 2: Engineering of PaS for improved TS hydrolysis, and characterization of the PaS mutants

2.1 Introduction

The growing importance of a robust steroid sulfatase for anti-doping applications was discussed in Chapter 1. The arylsulfatase from *Pseudomonas aeruginosa* (PaS) was selected as a suitable candidate as it exhibited some necessary activity. However, due to the limitations discussed, PaS requires modifications to be used in an anti-doping context. This chapter aims to improve PaS by enzyme engineering that would bring together protein chemistry, molecular biology, enzymology and analytical chemistry.

2.1.1 Enzyme engineering

Enzymes are biological catalysts that can dramatically increase the rate of specific chemical reactions, and their properties and characteristics have been studied for well over a century. Subtilisin, a type of alkaline protease is an early example [1] of an enzyme that was isolated and characterized (1967) [2]. Some enzymes such as the glucose dehydrogenase from *Thermoproteus* species [3] or dipeptidyl peptidase from *Porphyromonas gingivalis* [4] have evolved with apparent strict specificity for their natural substrates while some enzymes accept a broader range of substrates such as the arylsulfatase from *Pseudomonas aeruginosa* [5]. Modification of these enzymes or enzyme engineering to accomplish faster reaction rates for alternative substrates or to enhance the stability has great significance, mainly for industrial and synthetic applications [6, 7]. The first commercialized enzyme expressed in a genetically modified organism was a lipase called LipolaseR introduced for detergents in 1988 [1].

Evolution of enzymes can follow several pathways. Directed evolution [8-11] involves random mutagenesis of the whole gene, where in-depth understanding of the structure-

function relationships is not crucial. This process mimics natural selection, which involves iterative rounds of randomly distributed mutations introduced into the gene of interest, selection for mutants that contain the desired property and amplification of the gene as the template to be used in the next round. Error prone PCR and DNA shuffling are the methods commonly used for library generation in directed evolution. This technique requires screening of large libraries and hence developing a high-throughput method to isolate the desired property is often a major constraint in directed evolution [12-14].

Rational design on the other hand uses prior detailed knowledge to predict the best amino acid substitution that results in a desired property [14, 15]. Where available, comprehensive analysis of structure-function relationships of the catalytic residues can lead to mutating targeted residues to a single or a few selected amino acids of interest (site directed mutagenesis, SDM). Despite this technique being less laborious in terms of screening, the significant complexity and incomplete understanding of the enzyme structure-function relationships is a limiting factor for rational design. Hence semi-rational approaches are more often used by protein engineers upon the availability of enzyme structure [6, 15].

Semi-rational approaches are combinations of random mutagenesis and rational design where some prior knowledge on structure and function is required. The balance of random and rational approaches depends on the amount of information available regarding the structure-function relationship [15]. If structural details are available with vague functional information, then the selection of residues to be mutated could be rational while the amino acids investigated could be chosen randomly. In such instances, the targeted residues can be saturated (site saturation mutagenesis, SSM) to all possibilities with selection of the best variants. Whereas, if there is limited information on both structure and function, this could lead to random mutagenesis on a few residues by saturating the sites simultaneously [15-17] (combinatorial active-site saturation testing, CASTing) or in turn. If the required functionality (such as hydrophilicity or aromaticity) is known, then the amino acid alphabets could be restricted to a smaller group to achieve the desired property in a single site or multiple sites (multiple site mutagenesis, MSM). Such restricted amino acid libraries reduce the level of screening required to fully investigate the mutants generated.

The work described in this chapter takes the semi-rational approach and employed two techniques for mutagenesis. First, all possible residues were introduced at a single position by site saturation mutagenesis (SSM). Second, multiple positions were altered simultaneously with only a subset of the possible residues by multiple site mutagenesis (MSM).

2.1.2 Previous mutagenesis of PaS enzyme

A related mutagenesis study on PaS performed by Li *et al.* [18] revealed two beneficial mutants with greater V_{max} for TS and NS hydrolysis. The two mutations performed based on site directed mutagenesis were arginine at position 155 mutated to glycine (R155G) and lysine at position 330 mutated to alanine (K330A).

Table 2.1: Mutations found by Li *et al.* [18] in comparison with wild type (WT)-PaS

Mutation	Kinetic parameter	TS hydrolysis	NS hydrolysis
R155G	V_{max}	1.71 x WT	1.83 x WT
	K_M	2.75 x WT	1.89 x WT
	V_{max}/K_M	0.62 x WT	0.97 x WT
K330A	V_{max}	0.96 x WT	1.29 x WT
	K_M	3.10 x WT	2.06 x WT
	V_{max}/K_M	0.68 x WT	0.62 x WT

These two residues were selected based on the hypothesis that bulky and charged residues near the proposed substrate binding pocket may restrict access for steroid sulfates. These two mutations show that replacing the bulky charged arginine and lysine to glycine and alanine respectively resulted in an improvement of activity (V_{max}) as shown in Table 2.1. Particularly the R155G mutant showed a ~2-fold improvement for TS hydrolysis. However, the catalytic efficiency (V_{max}/K_M) of the hydrolysis of above substrates have slightly decreased compared to WT-PAS. Yet, these two sites provide scope for further improvement by performing site saturation mutagenesis.

2.1.3 Project goals and hypothesis

In view of the long-standing interest in a sulfatase for anti-doping applications, PaS is a possible solution that requires some improvement. Testosterone sulfate (TS) that represents a challenging substrate for PaS, was selected as the preferred substrate for the engineering of PaS. Further, TS plays a vital role in anti-doping and starting the enzyme engineering with a substrate that shows at least a little activity is important. The primary goals of this project were to: identify the steroid binding site of PaS, improve the catalytic efficiency of PaS for the hydrolysis of testosterone sulfate by 10-100 fold, characterise the mutants, improve the substrate scope of PaS, and to develop the enzyme for the hydrolysis to take place at 37 °C, pH 7.5. These hydrolysis conditions are analogous to those used with β-glucuronidase from *Escherichia coli* bacteria for the hydrolysis of steroid glucuronide metabolites in anti-doping laboratories.

The hypothesis generated is that the native activity of PaS can be improved by protein engineering: mutating bulky and/or charged residues around the substrate binding pocket to smaller hydrophobic residues that will facilitate substrate entry to, and product release from, the binding pocket by opening it up. Such an approach could also lower the substrate specificity providing a broader substrate scope.

2.2 Engineering of PaS enzyme for improved TS activity

In the course of engineering PaS for improved TS activity, site saturation mutagenesis, was performed on single residues and then small groups of residues were mutated simultaneously. Finally, the beneficial mutations from all libraries were shuffled to search for further improvements.

2.2.1 The recombinant *E. coli*

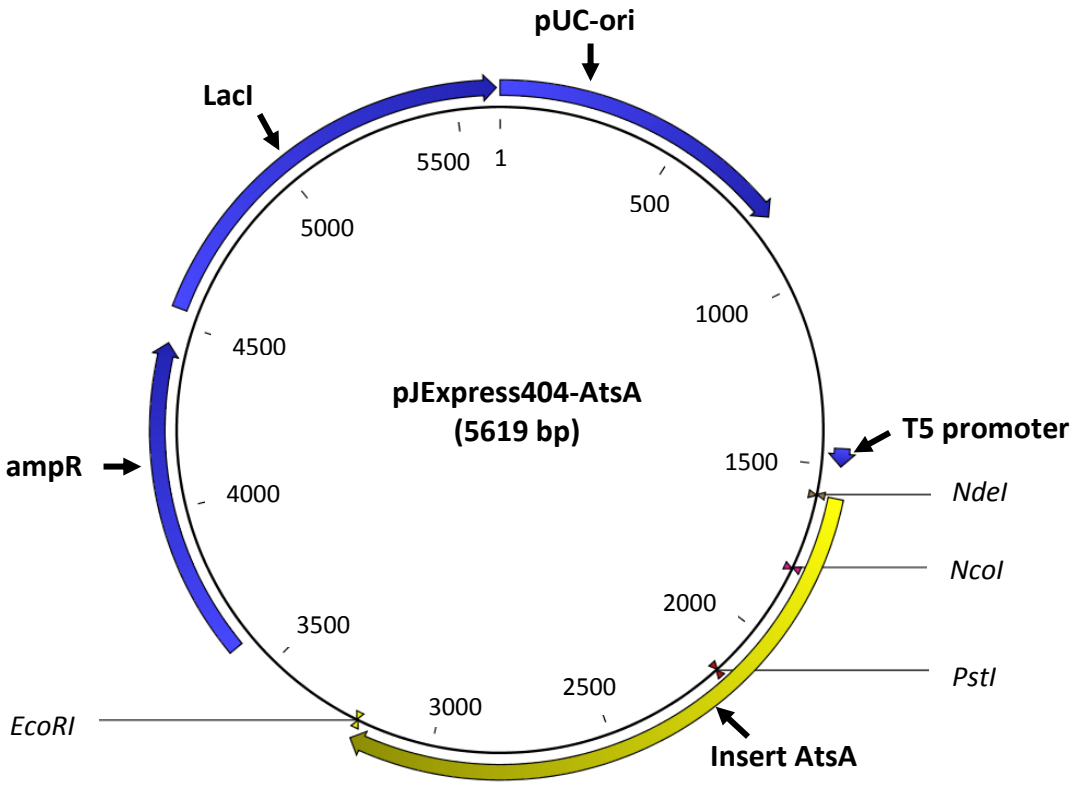


Figure 2.1: The pJExpress404 vector map with the AtsA gene inserted at the *NdeI*, *EcoRI* restriction sites.

The recombinant AtsA gene with the C-terminal hexa-histidine tag was expressed in the pJExpress404 vector (Figure 2.1) that carries the ampicillin resistance gene (ampR) and also the T5 promotor and *LacI* repressor. The AtsA gene was sub cloned to the pJExpress404 vector at the *NdeI* and *EcoRI* restriction sites. This expression system was used in all aspects of work in this project hosted in the *E. coli* DH5α strain.

2.2.2 Medium throughput screen

The sorting step is probably the most labour intensive yet the most crucial step in enzyme engineering [19] that identifies the desired mutants from a large mutant library. This could be of two types: selection and screening. Selection identifies mutants based on a phenotype such as antibiotic resistance, visual inspection of growth on minimal media or zones around colonies [14] that will select the variants with the desired property.

In occasions where the desired property cannot be linked with a phenotype, screening is employed. Screening involves a random selection of the variants that are usually cultured and tested in a 96-well or 384-well format and typically rely on a chromophore released from a chromogenic substrate. Despite selection offering a higher throughput in a shorter time, it is often performed on solid media plates and can be hard to set-up with optimized conditions to favour a particular phenotype. In contrast, the random screening systems blindly select mutants but require a discriminating assay [14, 19]. Due to the difficulties in generating a distinguishable selection method, random screening was selected as the method of choice for the sorting of the variants.

Mutating the whole gene of PaS by error-prone PCR with an error rate of two nucleotide mutations per sequence would require screening a library of a magnitude of 10^7 [20]. A library of this size would most suitably be best screened by a higher throughput colourimetric assay with a commercially available model substrate like *para*-nitrophenyl sulfate (PNPS) which releases a yellow chromophore upon hydrolysis by PaS enzyme (Figure 2.2).

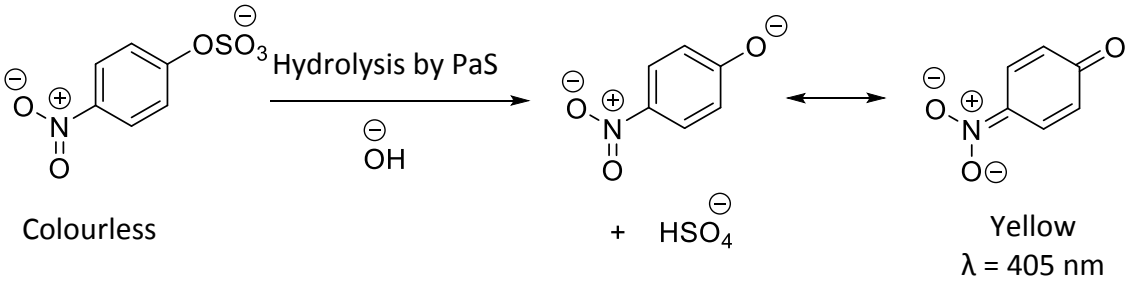


Figure 2.2: Hydrolysis of PNPS by PaS enzyme at alkaline pH

The colourimetric assay with PNPS, however, is not appropriate as a primary assay to engineer improved steroid sulfate hydrolysis, as the hydrolytic activities displayed by the PaS mutants could be contrasting for the two different substrates and could even follow different enzyme mechanisms. Therefore, performing a colourimetric screen with PNPS was not adopted as it was thought unlikely to lead to improved mutants for steroid sulfate hydrolysis. Hence, a medium throughput LC-MS assay that directly monitored testosterone sulfate hydrolysis was adopted for this study. Instead of directed evolution of the PaS gene with error-prone PCR, a few sites neighbouring the catalytic site were selected for mutagenesis that was performed in turn or simultaneously.

The practical limitations of LC-MS such as the long run time per sample (approximately 5 hours per 96-well plate compared to 10 seconds on a plate reader), necessity of a periodic system clean-up at least every 200 samples, the use of HPLC grade solvents, and the extra steps required to ensure no cell debris pass through the system make the process time consuming and costly. These reasons limited the library size to a maximum of a few thousands a library. This was achieved by targeting a few residues for mutagenesis at a time and trimming the library size by mutating them to selected amino acids of interest. However, screening of such focused libraries has previously been reported to have resulted in the discovery of beneficial mutations. In a study performed by Reetz *et al.* [21] to improve the enatioselectivity of a Baeyer-Villiger monooxigenase, a loop of four residues were selected as targets for mutagenesis. Using greatly reduced amino acid alphabets at each position significantly reduced the required screening effort to 2587 transformants to achieve 95% coverage as compared to over three million transformants in saturation mutagenesis. Yet, the screen was arbitrarily further restricted to 1700 clones even in which outstanding results were obtained. More studies have shown that focused libraries have resulted in improved mutants while reducing the screening effort [22, 23].

The primary screening of the mutagenesis libraries involved several hundreds to several thousands of samples and was performed by LC-MS in 96-well microtiter plates. A microtiter plate based screening was employed as it is simple and easy for liquid handling when coupled with multi-channel pipetting. However, the main problem is that the uneven environment it provides to samples. This is often observed as an 'edge effect' where greater aeration during culture growth or evaporation during reactions occurs for wells at the edges of the plate. This can result in false positive 'hits' arising from elevated production of enzyme during culturing or increased concentration of reagents during reaction.

Precautions were taken to minimize errors on sample treatment, edge effects and in selection of the active hits. Culturing, setting up reactions and screening for the beneficial mutants were the three steps involved and at most eight plates were subjected to these steps in a single session. Sample pipetting was performed in an orderly staggered manner so as to minimize variations in the time provided between samples for reaction. All steps were performed under tape sealed (culturing and

reactions) or thin film sealed (screening) conditions to reduce evaporation. Each plate included one or more control samples (WT-PaS for the single SSM libraries and an improved PaS mutant for MSM and shuffling libraries) for comparison of clones between plates. Instead of the product (testosterone, T) alone, the product/substrate (T/TS) response ratio was considered for the selection of hits of the initial libraries (libraries 1-6) to overcome the difficulties encountered in selection due to sample evaporation, as the responses for both substrate and product would be enhanced due to sample evaporation. Each half plate was considered separately by placing control samples in each half and also in an edge well. However, the T/TS ratio does not result in a precise quantification as the changes in the ratios may not reflect the actual improvement or drop of activity. While TS was analysed in negative mode T was analysed in positive mode and the efficiency of ionization and detection did not appear to be well correlated. In addition, because TS eluted very early, it was prone to matrix effects that T was not exposed to. Further, the TS concentration being roughly two orders of magnitude greater than the testosterone produced, the T/TS ratio could lead to less precise interpretations. For these reasons, an internal standard was introduced in the latter libraries that would be detected in positive mode, would be eluted after the water soluble components and would not interfere with the reaction or product analysis. Nandrolone (N) was the internal standard of choice and was included in the buffer of the reaction where the response ratio between the product/internal standard (T/N) was taken into consideration. Since the nandrolone concentration is more comparable with the testosterone concentration, the T/N ratio would be an accurate measure for quantification. This ratio of each sample was plotted plate wise and was compared with the activity of the control on each plate. Depending on the distribution of the plot from the primary screen, samples that displayed improved activity from each library were selected for a secondary screen.

The secondary screens were performed in biological triplicates. The culturing and the reactions of the secondary screens of the SSM libraries and the tertiary screens of the MSM libraries were performed in 1.5 mL Eppendorf tubes and screening in closed HPLC vials. This helped in treating all samples in an equal environment, thus minimizing the errors due to sample evaporation.

2.2.3 Site saturation mutagenesis libraries (libraries 1-5)

The NNS (N=A/T/G/C, S=C/G) codon degeneracy was used in the primer design to randomly alter the sites of interest to any of the other possible 19 amino acids with 1/32 probability of introducing a stop codon. However, multiple codons translating the same amino acid increase the probability of those amino acids being over-represented in a library.

2.2.3.1 Overview of library generation and screening

With the crystallographic structure in hand, a semi-rational approach was followed to create PaS variants. The prudently selected residues were mutated by site saturation mutagenesis (SSM) and the libraries were expressed in *E. coli.* The choice of *E. coli* strain used throughout the course of this work was DH5α, as it is known for its high transformation efficiency, relatively simple preparation, and easy storage.

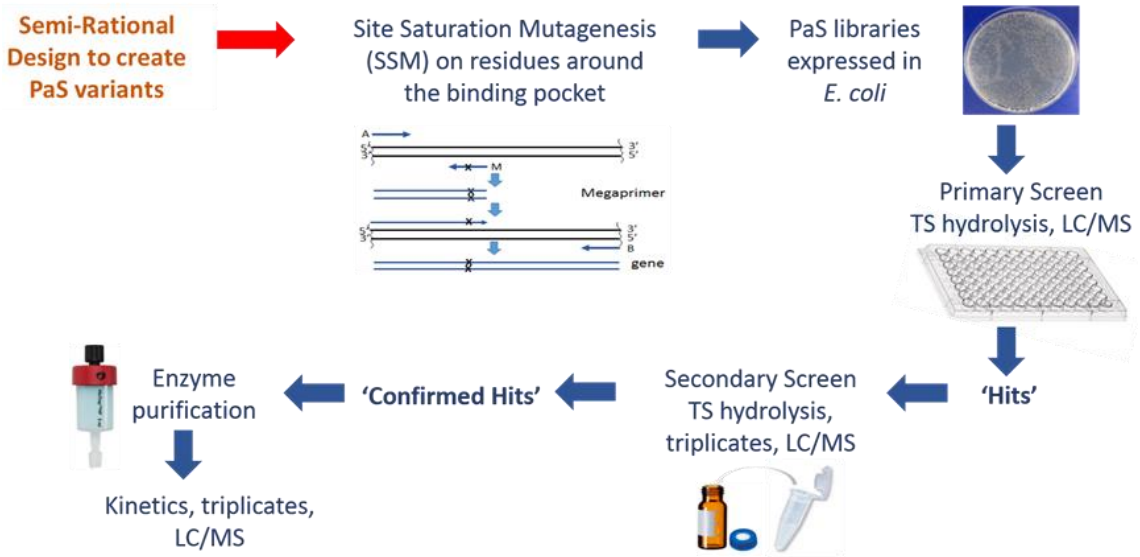


Figure 2.3: An overview of the library generation and screening of the site saturation mutagenesis libraries

The selected clones were screened for testosterone sulfate (TS) hydrolysis. The primary screen was conducted in 96-well plates with a pre-determined library size to screen all possible variants with more than 95% probability. The beneficial mutants from

the primary screen or the 'hits' were carried forward for a secondary screen where thorough sealed conditions were used for culture growth, TS reactions and screening that was performed as biological triplicates. When they were confirmed as hits, the expressed proteins were purified for further characterization. All hydrolysis reactions were monitored by liquid chromatography-mass spectrometry (LC-MS).

2.2.3.2 Selection of residues

The efforts to generate an enzyme with improved catalytic efficiency for the hydrolysis of testosterone sulfate (TS) were based on the crystallographic structure resolved by Boltes *et al.* [24]. Two possible binding modes of TS were identified *in silico* by Dr. Thomas Balle (Faculty of Pharmacy, University of Sydney). The modelling was preliminary and introduced simplifications such as using a carboxylate in place of a formylglycine hydrate (FGH) nevertheless provided useful insights.

Inspection of each predicted pose (Figure 2.4) indicates several residues that could be restricting catalytic site access for bulky substrates. As depicted in pose I (Figure 2.4A)), the three phenylalanine residues surrounding the binding pocket at positions 328, 331 and 463 could be hindering the substrate entry to, and the product release from, the binding pocket. For pose II (Figure 2.4B)), the positively charged arginine and lysine at 155 and 330 respectively could be disfavouring the interactions with the hydrophobic steroid skeleton. The two residues (R155 and K330) that obstruct the binding pocket depicted in pose II have been previously investigated by Li *et al.* [18] with modest improvements to V_{max} on hydrolysis of some steroid sulfates.

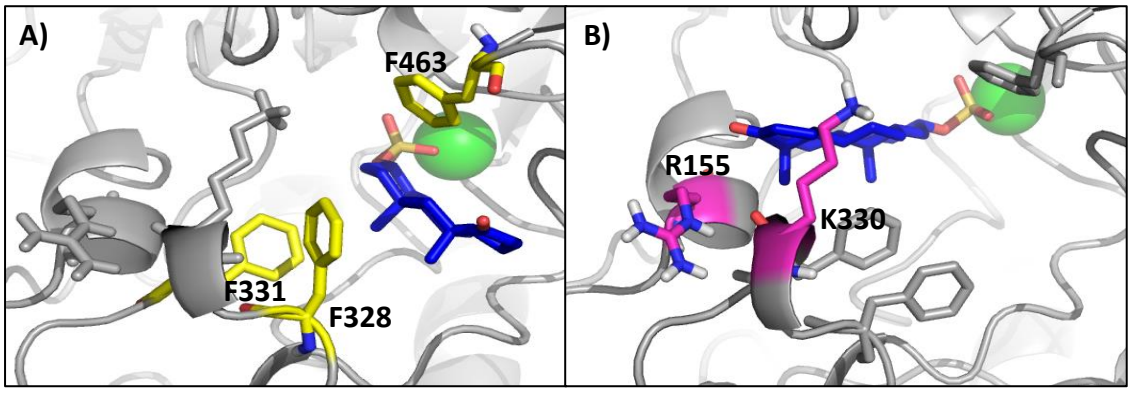


Figure 2.4: The two binding modes suggested by computational ligand protein docking. The substrate (DHEAS) is shown in dark blue. A) Pose I. B) Pose II

The substrate docking in pose I aligns well for an inline nucleophilic substitution to take place with an angle of 164.0° between the nucleophilic oxygen of FGH, and the S—O bond of the sulfate ester. In contrast, this angle in pose II is 113.6° making the pose II poorly arranged for the proposed inline nucleophilic substitution (Figure 2.5).

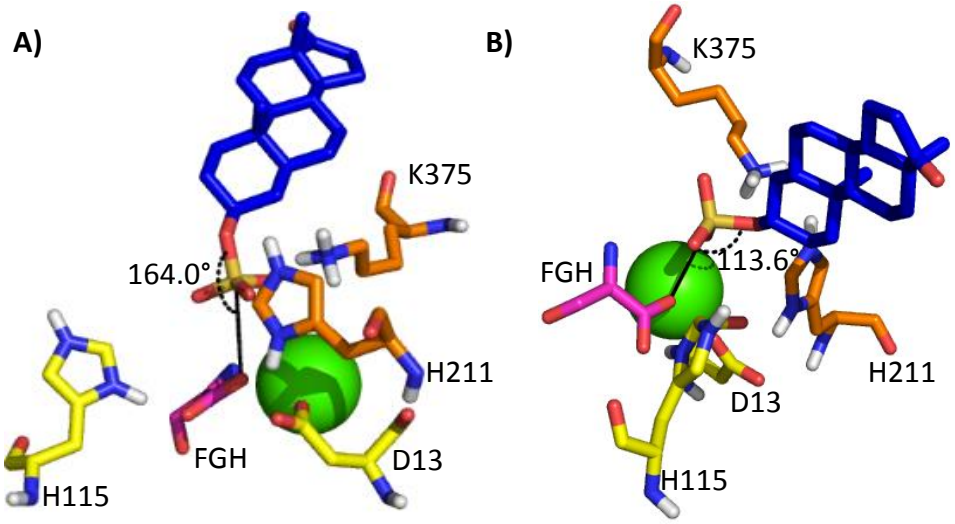


Figure 2.5: FGH (magenta) and docked DHEAS substrate (dark blue) showing the angle between the nucleophilic oxygen of FGH, and the S—O bond of the sulfate ester in A) pose I and B) pose II. Selected catalytic residues in PaS active site are shown: residues activating and stabilizing FGH (yellow), residues that facilitate substrate binding, activation and leaving group protonation (orange).

Hence, the residues corresponding to pose I were targeted first for mutagenesis despite the improvements observed for steroid sulfate hydrolysis with respect to the mutagenesis of the corresponding residues of pose II by Li *et al.*

2.2.3.3 Library generation and screening

Site saturation mutagenesis was first performed using the Stratagene QuickChange™ PCR method [25] with the wild type gene as the template. This method employs a set of complementary primers containing the mutation which is introduced in a single PCR. Although a DpnI digestion was carried out to digest the methylated sites of the parent DNA, observing 80-90% of wild type contamination was one problem encountered. This method also included introducing silent mutations to incorporate restriction enzyme cleavage to reduce wild type contamination when generating the

SSM libraries at 328 (BpmI site) and 463 (SphI site) positions. However, due to inefficient digestion, this technique was not successful.

Therefore, an alternative method, the megaprimer (MP) approach [26] (see section 2.6.5.1) was employed to overcome this problem. This is a two-stage process where the first PCR introduces the mutation and the product of it serves as a megaprimer for the second PCR to create the full-length gene. A subsequent gel purification of the final product diminished the wild type contamination to 10-20%. A schematic representation of the library generation is shown in Figure 2.6.

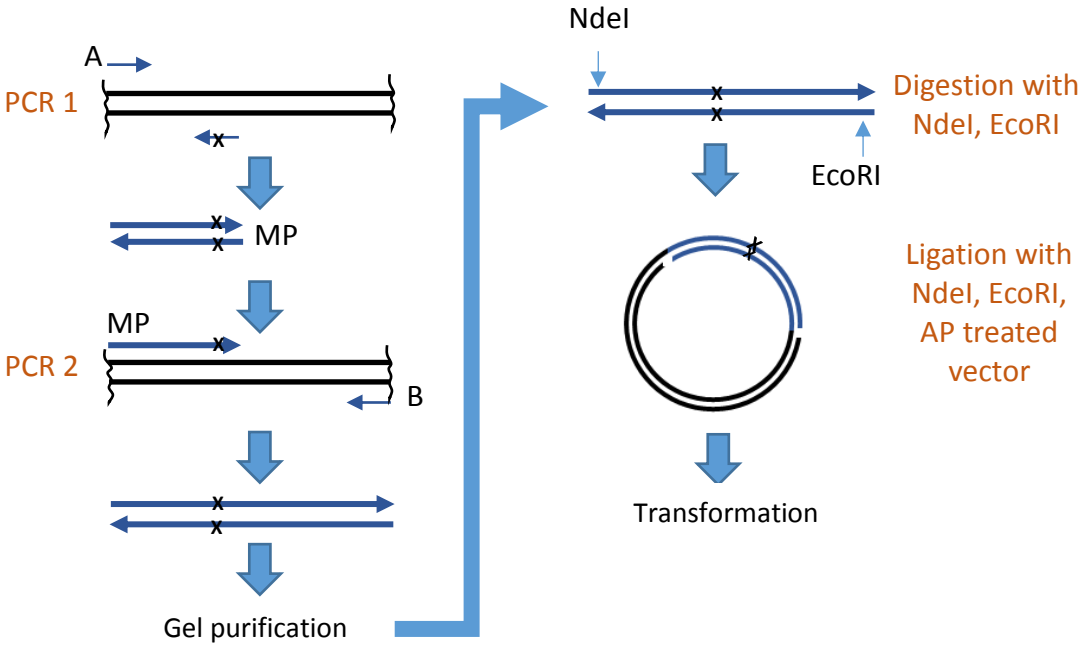


Figure 2.6: Reaction scheme for library generation by the megaprimer PCR method.

The three phenylalanine residues shown in Figure 2.4 A), F328, F463, F331, were mutated in turn as described in section 2.6.5.1 using the primers listed in Table A1 (see Appendix), to produce three site-saturation mutant libraries (SSM 1,2,3 respectively).

Before performing the primary screening, a small-scale library (10-15 colonies) for each site was subjected to a colony PCR to check for the presence of the correct size of the insert, and the samples with the correct insert were DNA sequenced to check for wild type contamination, after which the library size was determined. The number of mutants to be screened or the library size for TS hydrolysis from each library was calculated such that each of the possible mutant codons had more than 95% probability

of being tested by substituting to the equation used by Patrick *et al.*[20], L = -V ln (-ln Pc / V); where L= library size; V= possible sequence variants; Pc= probability every variant is represented. The activity profiles of the primary screen of SSM 1, SSM 2 and SSM 3 are shown in Figure 2.7.

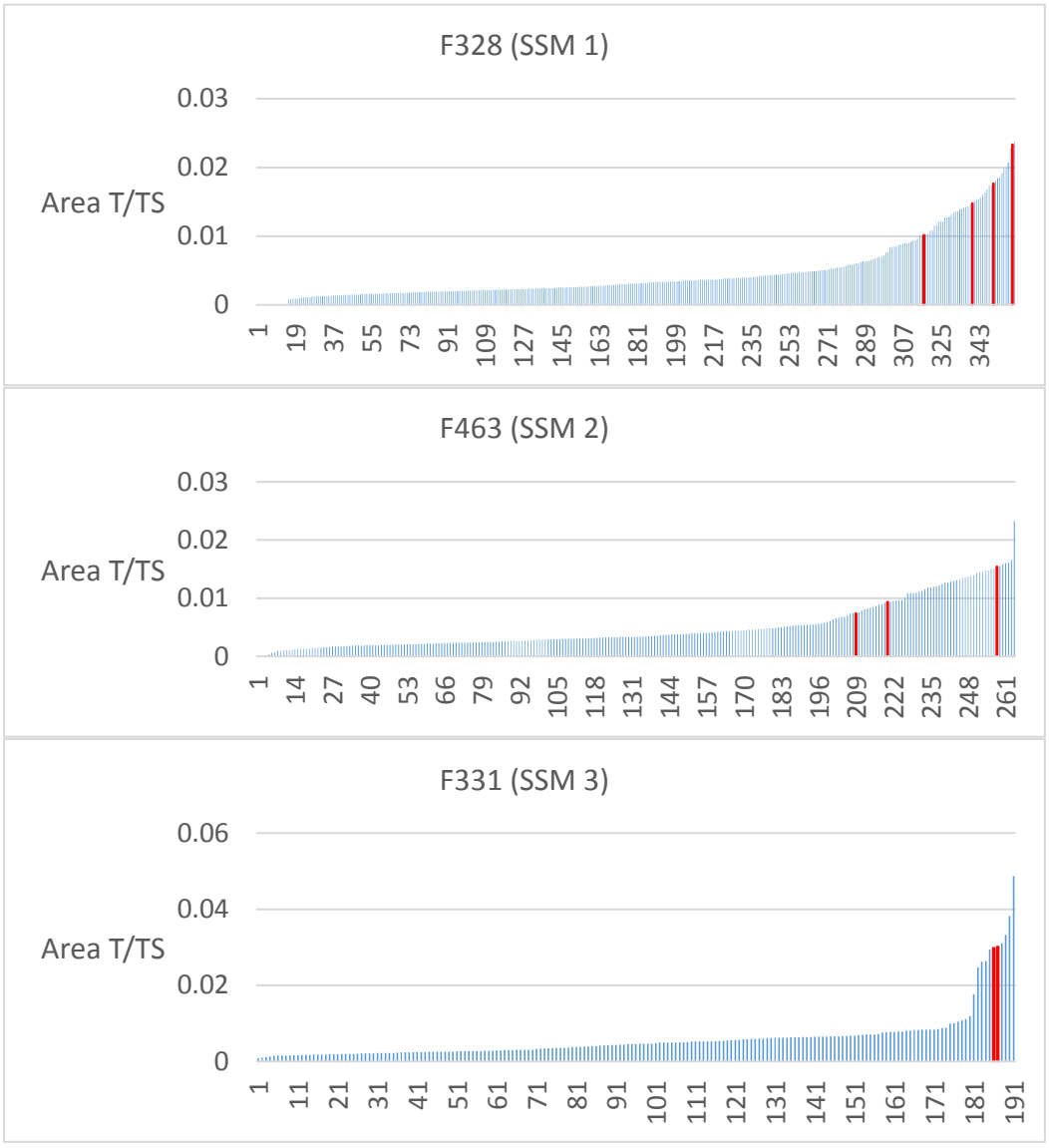


Figure 2.7: The activity profiles of the primary screens of SSM 1, 2 and 3 ordered by increasing activity from left to right. The red lines represent WT controls and the blue lines represent the test samples. The x-axis represents the clone number screened.

Mutating the three Phe residues did not result in beneficial mutations. The samples that showed better activity in the primary screen were tested with DNA sequencing and resulted as wild type. Therefore, the residues corresponding to binding mode II were investigated despite the angle measured from substrate docking being

poor for an in-line nucleophilic substitution. As a result, K330 and R155 were mutated and screened similarly to the three phenylalanine residues, which resulted in two libraries, SSM 4 and 5 respectively.

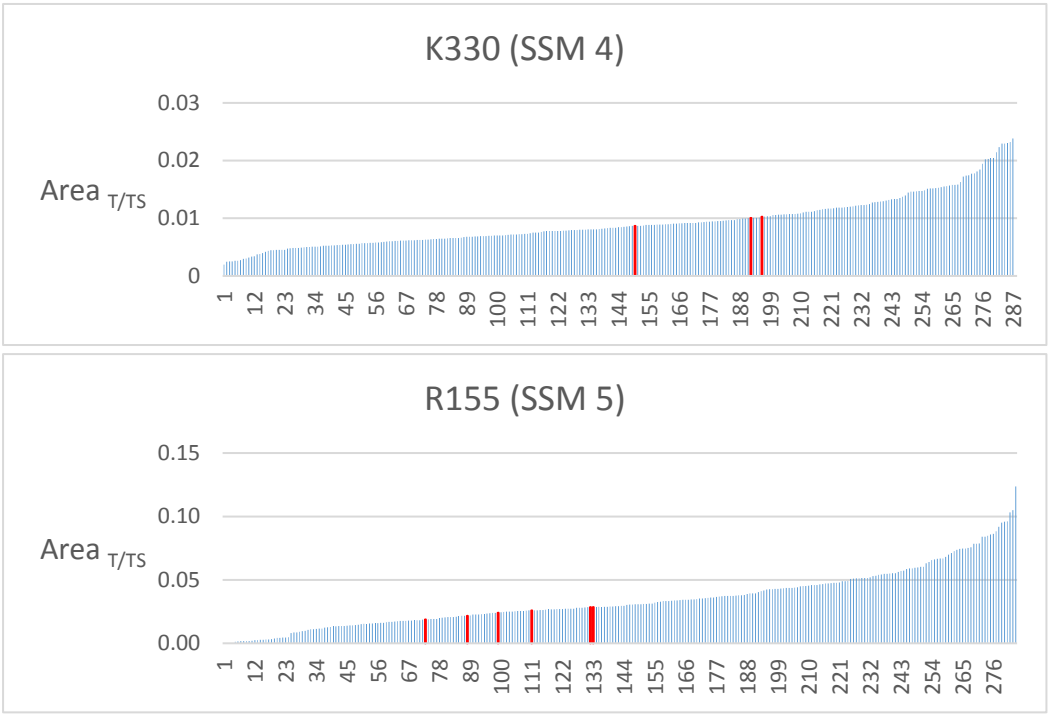


Figure 2.8: The activity profiles of the primary screens of SSM 4 and 5 ordered by increasing activity from left to right. The red lines represent WT controls and the blue lines represent the test samples. The x-axis represents the clone number screened.

The activity profiles of SSM 4 and SSM 5 (Figure 2.8) are quite dissimilar when compared to that of SSM 1-3 with more variants resulting in better activity than the wild type control used. Upon subsequent secondary screening and DNA sequencing of all variants selected for the secondary screen (12 and 18 variants from SSM 4 and SSM 5 libraries respectively), these two libraries resulted in three improved mutants: proline or threonine at position 155 (R155P or R155T) or valine at 330 (K330V). In contrast to the residues at 155 and 330, mutating the phenylalanine residues did not yield any beneficial mutants.

Each of the improved mutants (R155P/T, K330V) discovered in the site saturation mutagenesis (SSM) introduced smaller residues that would help accommodate bulky steroid sulfate substrates more easily. It is interesting to note that the crystal structure of PaS [24] has several regions that cannot be refined due to a lack of electron density and

presumably conformational flexibility. The side chain of K330 is one of these regions and could be impeding the substrate entry or product release. Valine at this position should reduce the obstruction. Additionally, R155P not only allows a larger opening around the putative entrance to the catalytic site, but also restricts movement of the protein backbone. This occurs where a single turn of alpha-helix is flanked by the proline at 154. This proline-proline unit forms a *trans* conformation during folding and provides rigidity due to the conformationally restricted five membered rings.

2.2.3.4 Comparison of TS and PNPS hydrolysis

The TS and PNPS hydrolysis were tested with ten randomly picked colonies in two libraries representing the two binding modes as described in section 2.6.7.3. A wild-type colony was used as a control and the results shown in Figure 2.9 are normalized with respect to the WT control. The test was not replicated as it aimed to compare the mutants of two libraries against a model substrate, PNPS, and a steroid sulfate, TS, and to test the aptness of using PNPS as the substrate when engineering PaS for TS.

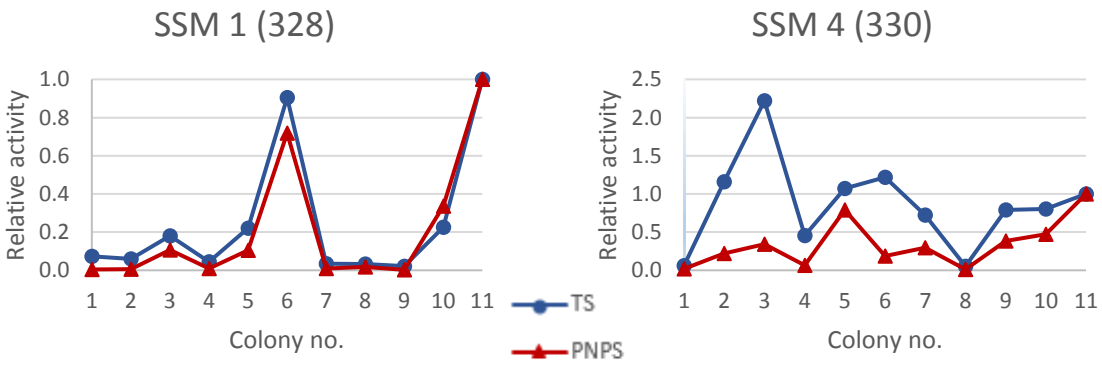


Figure 2.9: The activity profiles for 100 µM TS and 100 µM PNPS hydrolysis by crude lysates of a 10-colony random library in SSM 1 and SSM 4. The connected lines between discrete data points is for clarity.

The comparative loss of function for both substrates by different mutations at 328 position is similar; this could reflect other enzyme characteristics such as stability, and expression. It could also be due to the mutations on the 328 residue having minimal impact on substrate specificity. The colony 6 in SSM 1 was confirmed as WT after sequencing. In contrast, the variations at the 330 residue affects the hydrolytic activity

of the two substrates differently and reflects genuine improvement or detrimental effects in the catalytic activity.

The colony 3 in the SSM 4 library represents an example for improved activity for TS but suppressed activity for PNPS. This validates the use of TS as the substrate in screening despite its longer reaction time and more extensive sample preparation compared to the use of PNPS as the substrate.

2.2.4 Combined mutants

Site saturation mutagenesis identified two sites that can be mutated for improved TS hydrolysis. The combination of mutations at these sites were then investigated to determine if the benefits were additive. Two PaS variants were prepared by combining the beneficial mutations, R155P/K330V (PV-PaS) and R155T/K330V (TV-PaS) by double digestion of the three single mutants from the two sites followed by ligation (Figure 2.10)

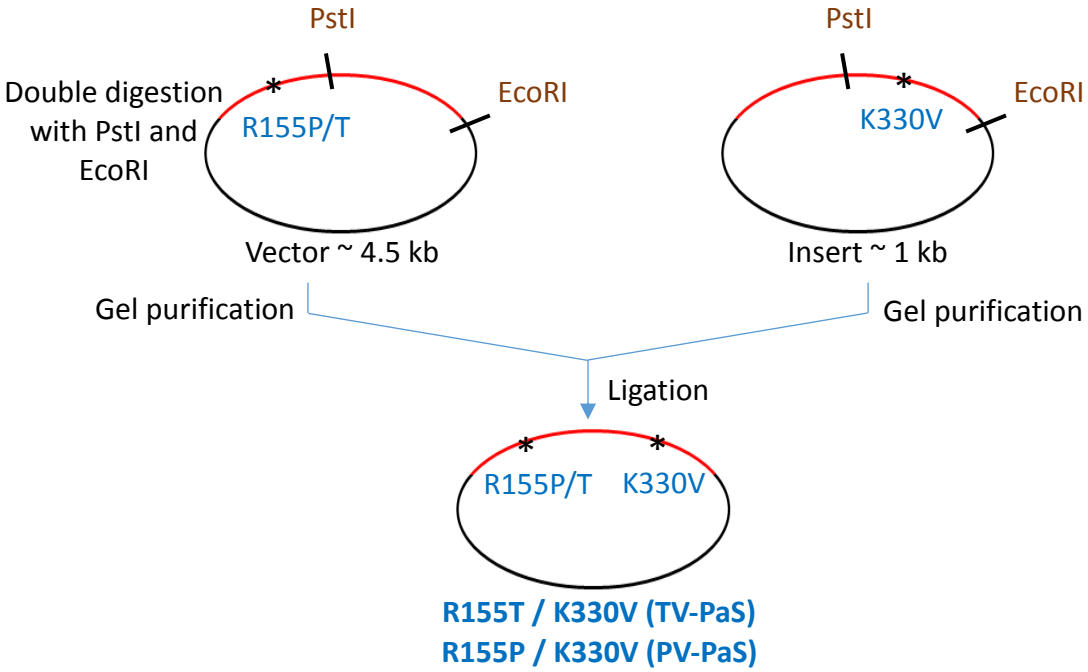


Figure 2.10: Combination of the three single mutants to produce two double mutants

The resulting double mutants were sequence verified, purified and used for further characterization.

2.2.4.1 Enzyme kinetics of the beneficial mutants

Enzyme kinetics of the purified proteins of the beneficial mutants obtained from the SSM libraries and the combined mutants together with wild type PaS (WT-PaS) were studied for the hydrolysis of TS as described in section 2.6.8.1. The final concentration of the enzymes used was 0.4 µM. The results were fitted to the Michaelis Menten model (Figure 2.11).

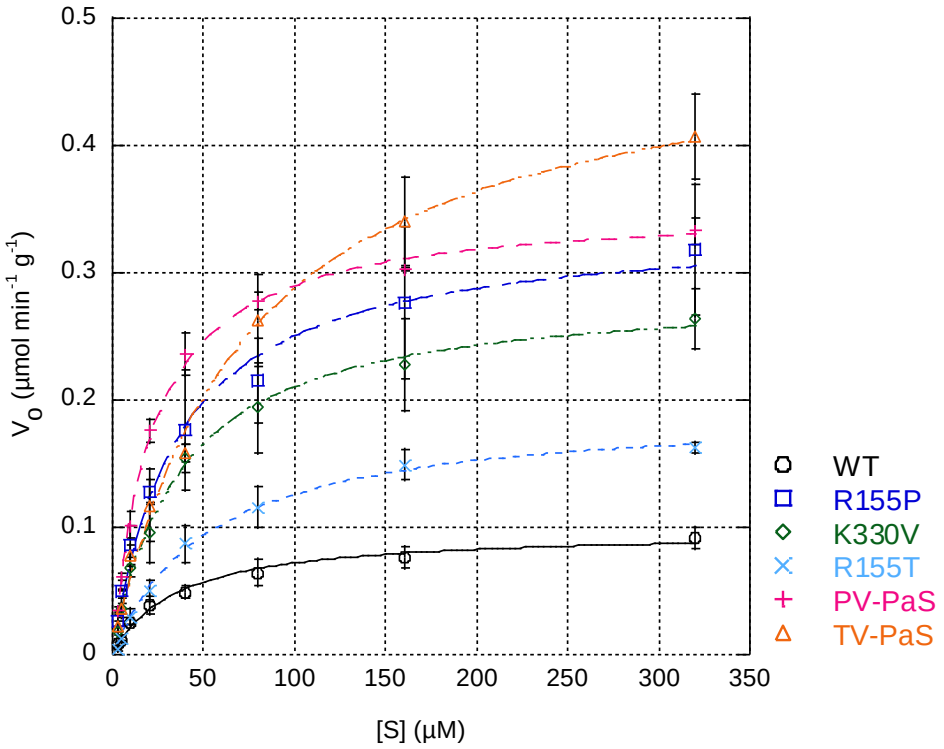


Figure 2.11: Substrate saturation kinetics, at 37 °C, pH 9 (50 mM Tris acetate), fitted to the Michaelis-Menten model: $V_o = V_{max}*[S]/(K_M + [S])$ by using KaleidaGraph 4.5 software. Initial velocity (V_o, µmol min^{-1} (g protein)$^{-1}$) of PaS mutants, 0.4 µM, is plotted against TS concentration ([S], µM). The error bars correspond to the standard deviation between triplicates.

All PaS variants exhibit a significant increase in V_{max} with the two double mutants exhibiting more than three times wild type V_{max} (Table 2.2). In terms of substrate affinity, there was no significant change in K_M as a result of single mutations at 155, whereas K330V significantly increased K_M. Mutations of both sites together showed that their effects on TS hydrolysis kinetics were not independent; they are synergistic in accordance with their close proximity in the crystal structure [24]. The PV-PaS mutant shows a significant improvement in substrate affinity with 62% of the wild-type K_M,

whereas TV-PaS is the opposite with about twice the K_M of wild-type. Overall, PV-PaS offered the best result from the first set of mutations with the highest V_{max}/K_M: 5.6 times that of wild-type. Therefore, PV-PaS double mutant was selected as the best mutant to be carried forward for the next rounds of engineering.

Table 2.2: Kinetic parameters of the beneficial PaS variants from SSM libraries and of combined mutants over an initial TS concentration from 2.5 – 320 µM, pH 9 (50 mM Tris acetate), 37 °C. See section 2.6.8.1.

PaS enzyme	Mutation	Library	Kinetic parameters*		
			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	-	-	35.1 ± 4.6	9.72 ± 0.40	2.77 ± 0.38
V-PaS	K330V	SSM 4	52.5 ± 4.1	16.3 ± 0.4	3.10 ± 0.25
P-PaS	R155P	SSM 5	35.6 ± 4.4	29.3 ± 1.1	8.23 ± 1.16
T-PaS	R155T	SSM 5	36.7 ± 2.7	24.0 ± 0.6	6.54 ± 0.51
PV-PaS	R155P/K330V	-	21.7 ± 1.4	35.3 ± 0.6	15.6 ± 1.0
TV-PaS	R155T/K330V	-	87.5 ± 10.0	47.1 ± 1.7	5.38 ± 0.64

*The errors are the standard deviation between triplicates.

With regard to enzyme substrate interactions, if pose I reflects the steroid sulfate binding site then the mutational analysis suggests Phe is optimal at each position. Given this, improvements arising from mutation at R155 and K330 must operate at a distance to improve affinity and catalytic activity. Conversely, if pose II is the steroid sulfate binding site then the R155P/T and K330V mutations introduce smaller uncharged residues that would help to accommodate the bulky substrates. In this case, mutation to each of the three Phe residues perturb enzyme performance through attenuated substrate binding, reduced enzyme activity, lower PaS stability or a combination of these factors. Thus, this early evidence suggested pose II as the likely steroid sulfate binding site. The pose II showing a poor geometry for an in-line nucleophilic substitution could therefore be an artifact of the modelling, as in the enzyme, the substrate can presumably adopt a more productive binding mode. Despite the modelling performed being preliminary it provided the initial guidance to target the residues to mutate by introducing the binding poses. However, a comprehensive modelling is currently being

carried out by Professor Lynn Kamerlin and her team (University of Uppsala, Sweden) to further characterize enzyme-substrate interactions. Irrespective of the precise location of substrate binding site, next it was sought to investigate additional mutations on and around the PV-PaS template in order to further improve PaS.

2.2.5 Multiple site mutagenesis (MSM)

Further efforts to improve PaS for TS hydrolysis aimed to build on finding optimal combinations of mutations. The success of mutagenesis lies in obtaining and screening sufficient amino acid diversity at the targeted site so that sequence space is adequately covered. Therefore, for the MSM libraries, reduced amino acid alphabets were employed (NYS randomization, N=A/T/G/C, Y=C/T, S=G/C) to limit the library size for screening. Assuming full coverage, this degeneracy allowed for all hydrophobic residues except tryptophan and glycine. 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.

The residues for MSM 1 and MSM 2 were selected based on an online tool for 'hot spot' identification called HotSpot Wizard reported by Pavelka *et al.* [27]. This web server helps in identifying residues of a protein that could be mutated for improved enzymatic properties such as catalytic activity, substrate specificity and enantioselectiviy. When the PDB code or the structure is uploaded, the server predicts the hot spots by integration of several bioinformatics databases such as RCSB PDB, UniProt, PDBSWS, Catalytic Site Atlas and nr NCBI and computational tools such as CASTp, CAVER, BLAST, CD-HIT, MUSCLE and Rate4Site. The mutability of each amino acid is determined by their level of conservation within the enzyme family. The server first extracts the residue annotations from UniProt fields that determine the amino acids crucial for the catalytic function and collect information about mutagenesis and naturally occurring variants. Other databases such as Catalytic Site Atlas and PDBSWS are also used to acquire information about catalytic site residues. The CASTp tool helps in finding all the pockets in the structure and with the available information on catalytic site residues, the active site pockets are selected. Then, a point in the active site is calculated such that a sphere around it does not intersect any atom in the structure and the distance to the centre of

the structure is minimal. Next, the CAVER tool is used to identify tunnels connecting this point with the outside solvent. Finally, a BLAST search is carried out to obtain similar sequences to the query. They are clustered with CD-HIT and aligned with MUSCLE computational tools. The evolutionary rates obtained for each amino acid position from Rate4Site program are converted to a conservation score which is used to assign mutability to individual residues of the query enzyme and outputs an annotated structure with the mutable residues.

Apart from the residues previously mutated by SSM, the HotSpot Wizard assisted in identifying more mutable residues. Amongst these, the residues forming the short alpha helix flanked by R155 are predicted to accommodate mutations (Figure 2.12).

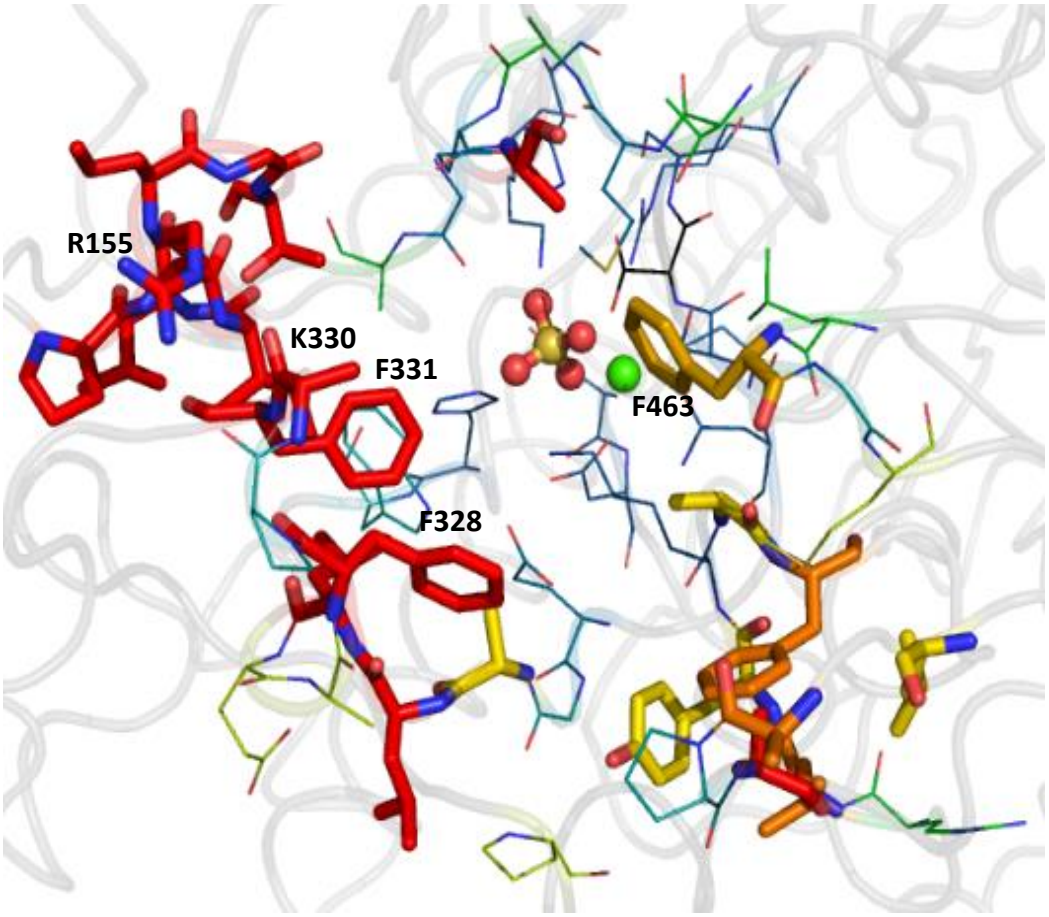


Figure 2.12: The functional residues and the neighbouring residues of the PaS catalytic pocket. The mutable residues or the hot spots (sticks) and functional residues (lines) are coloured by mutability according to the rainbow spectrum; from red having the highest mutability to violet having the lowest mutability. The initially mutated residues by SSM are labelled.

2.2.5.1 Multiple site mutagenesis 1 (MSM1)

The first set of residues selected for multiple site mutagenesis (MSM1) were: I156, L325, and F331. They were selected as neighbours of the successfully mutated sites (R155 and K330), because they appear to project into the substrate binding pocket (Figure 2.13) and have a high mutability as depicted by the red sticks in Figure 2.12.

These traits suggested that they would have some influence over substrate specificity. The best mutant in hand, PV-PaS, was used as the template for this library of variants. One residue, F331, had already been mutated in isolation without success and it was hypothesised that simultaneous mutations at neighbouring residues and the mutations in PV-PaS might support a useful change that would otherwise be unstable.

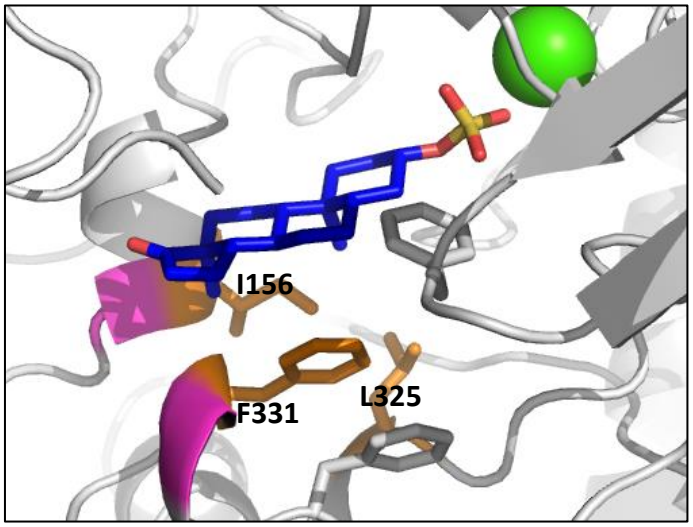


Figure 2.13: Pose II and the residues selected for mutagenesis in MSM 1 (orange sticks). The mutational locations of the template, PV-PaS (in pink), the Ca^{2+} ion (green) and DHEAS docked (blue) are shown.

The primary screening of 1728 clones was performed with 100 µM TS and cultures that gave rise to greater testosterone release than PV-PaS were selected for further testing. A typical activity profile of a single plate (Figure 2.14) contained approximately 5 % beneficial mutants compared to the template (PV-PaS) used. A total of 92 clones from the primary screen were subjected to a secondary screen to confirm the activity by comparing with that of PV-PaS. The secondary screen was performed in 96-well plates in biological triplicates.

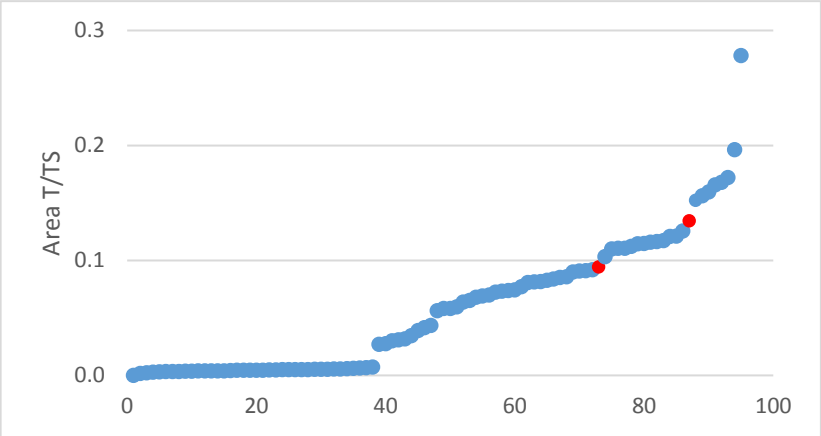


Figure 2.14: Activity profile of crude lysates of a single plate of the primary screen of MSM 1 with the tested samples (blue) in comparison to the PV-PaS control samples (red). See 2.6.7.1.

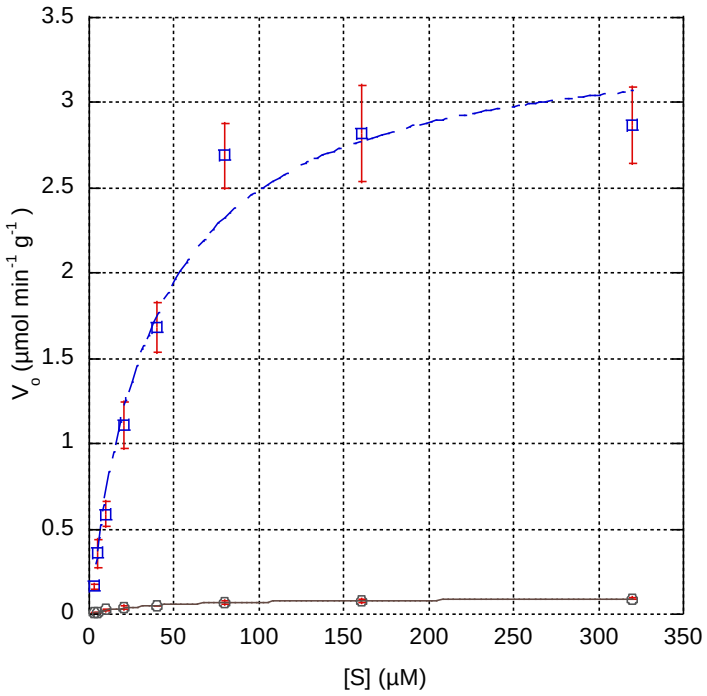
The tests with purified PaS variants confirmed that four beneficial PaS variants had greater activity than PV-PaS, as outlined in Table 2.3. All beneficial mutants observed in this library had the F331 unmutated.

Table 2.3: The beneficial mutants found in MSM 1 library and the initial hydrolysis rates of TS (20 µM) hydrolysis by the purified mutants (0.4 µM) relative to PV-PaS (0.4 µM).

No.	Mutations	Initial hydrolysis rate relative to PV-PaS[a]
1	I156/**L325V**/F331	1.1 ± 0.05
2	**I156L**/**L325V**/F331	1.7 ± 0.15
3	I156/**L325F**/F331	2.1 ± 0.07
4	**I156V**/**L325F**/F331 **(PVFV)**	3.0 ± 0.23

[a] The initial hydrolysis rates measured for 1 hr reactions at 37 °C, pH 9 (50 mM Tris acetate). The errors represent the standard deviation between triplicates.

The kinetic parameters of the best mutant (No. 4 or PVFV-PaS, Table 2.3) show a significant improvement in TS hydrolysis compared to the template used, PV-PaS (Figure 2.15). The improvement in catalytic efficiency, V_{max}/K_M, of PVFV-PaS compared to WT-PaS is approximately 33 times greater (Table 2.8). The PVFV-PaS mutant has two additional mutations (I156V and L325F). The mutation at 156 essentially removes a single methyl group from the sidechain and may create more space for the substrate.



WT-PaS	
V_{max}	0.0972 ± 0.0039
K_M	35.1 ± 4.6
χ^2	7.5673e-5
R^2	0.99419

PVFV-PaS	
V_{max}	3.43 ± 0.21
K_M	37.7 ± 7.4
χ^2	0.20532
R^2	0.9887

Figure 2.15: Substrate saturation kinetics, at 37 °C, pH 9 (50 mM Tris acetate), fitted to the Michaelis-Menten model: $V_o = V_{max}*[S]/(K_M + [S])$ by using KaleidaGraph 4.5 software. Initial velocity (V_o, µmol min^{-1} (g protein)$^{-1}$) of PVFV-PaS mutant, 0.4 µM (blue) and WT-PaS, 0.4 µM (black) are plotted against TS concentration ([S], µM). The error bars correspond to the standard deviation between triplicates. The kinetic parameters for each plot are tabulated beside the chart.

2.2.5.2 Multiple site mutagenesis 2 (MSM 2)

The second round of MSM targeted I156, L157, K158 and T160 (Figure 2.16) with PV-PaS as the template. These residues are on the short helix where two other beneficial mutations (R155P and I156V) were found and they are offered as highly mutable residues (Figure 2.12). When the residues of the short α-helix were considered, the G159 was disregarded to limit the library size, as it is originally a small hydrophobic residue. The PVFV-PaS was not selected as the template in the MSM 2 because it was not confirmed as the best mutant from the MSM 1 library by kinetic analysis by the time MSM 2 was commenced. Hence, the three residues at 156 position (isoleucine, valine and leucine) identified as beneficial from the secondary screen of the purified proteins in the MSM 1 library (Table 2.3) were included for randomization in the MSM 2 library using VTT degenerate codon at this position and NYS at other positions.

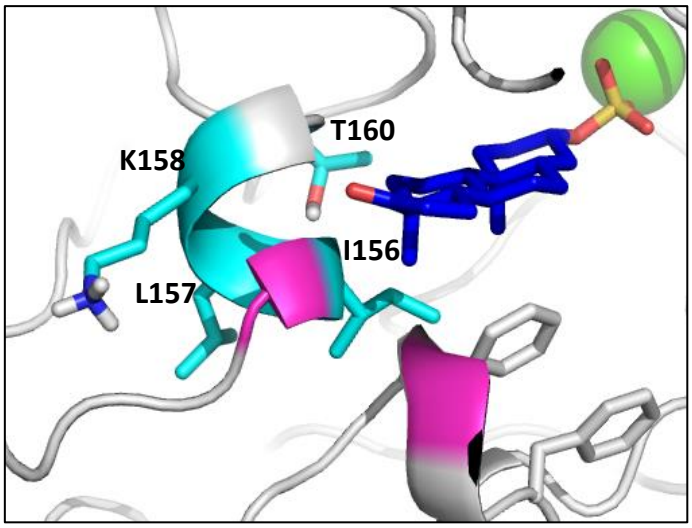


Figure 2.16: Pose II and the residues selected for mutagenesis in MSM 2 (cyan sticks). The mutational locations of the template, PV-PaS (in pink), the Ca^{2+} ion (green) and DHEAS docked (blue) are shown.

Having a mutant with ~33-fold improvement in hand, more 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 so as to discover mutants with greater affinity for TS; necessary for the samples encountered in the anti-doping screens. A typical activity profile of a single plate is presented in Figure 2.17 and on average a plate contained approximately 10 % beneficial mutants compared to the template (PV-PaS) used.

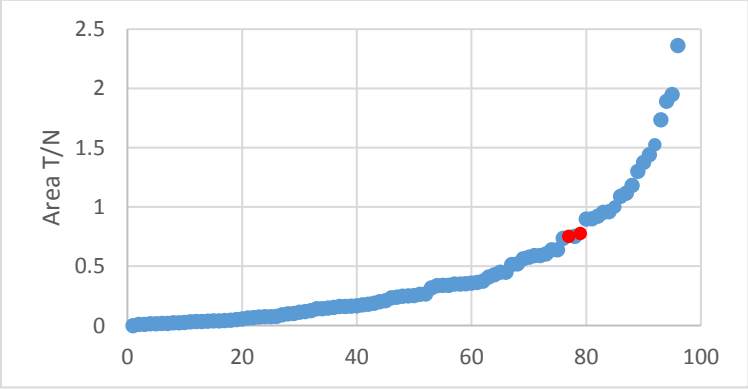


Figure 2.17: Activity profile crude lysates of a single plate of the primary screen of MSM 2 with the tested samples (blue) in comparison to the PV-PaS control samples (red)

From the primary screen of 4800 clones, 118 clones that resulted in greater testosterone release than PV-PaS were tested in the secondary screen and a subsequent

tertiary screen (20 clones). The secondary and the tertiary screens were performed in biological triplicates. The 20 clones were confirmed as 13 distinct mutants (some being replicated) from gene sequencing out of which three mutants that showed standout activity were purified and their initial TS hydrolysis rates were tested (Table 2.4).

Table 2.4: The beneficial mutants found in MSM 2 library and the initial hydrolysis rates of TS (20 µM) hydrolysis by the purified mutants (0.4 µM) relative to PV-PaS (0.4 µM).

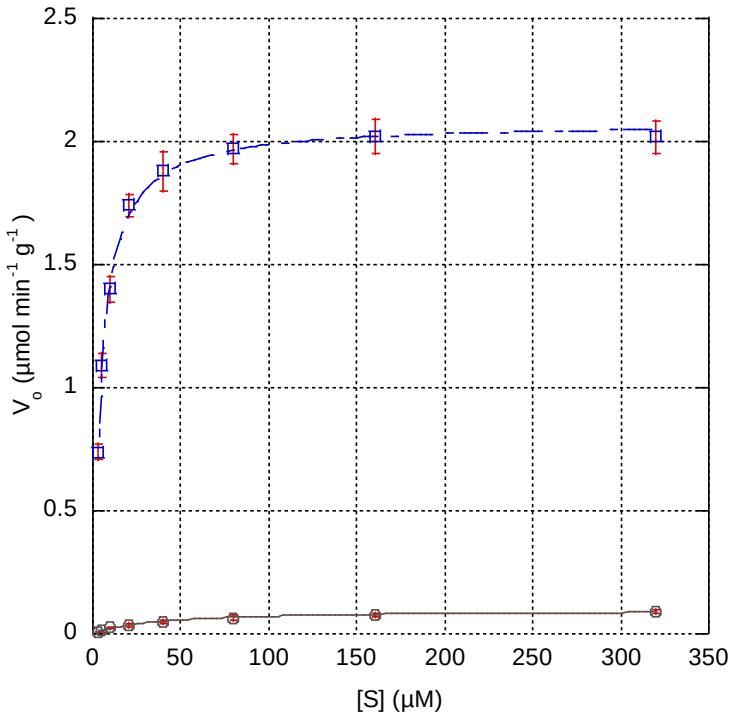
No.	Mutations	Initial hydrolysis rate relative to PV-PaS[a]
1	**I156V/L157A/K158V/T160P**	2.0 ± 0.06
2	**I156L/L157V/K158V/T160S (PLVVSV)**	5.7 ± 0.17
3	I156/**L157V/K158I**/T160 **(PVIV)**	8.5 ± 0.28

[a] The initial hydrolysis rates measured under 1 hr reactions at 37 °C, pH 9 (50 mM Tris acetate). The errors represent the standard deviation between triplicates.

Enzyme kinetics was performed on the best mutant discovered from the initial hydrolysis rates (Table 2.4: No 3 or PVIV-PaS). This mutant shows a remarkable improvement in the catalytic efficiency for TS hydrolysis compared to the previously discovered mutants (Figure 2.18). The improvement of V_{max}/K_M is approximately 167 times greater than that of WT-PaS (Table 2.8).

Discovering a mutant with remarkably improved substrate affinity in MSM 2 library was achieved by modifying the screening conditions. As lower substrate concentrations are more relevant to the low concentration samples encountered in the anti-doping screens, a substrate concentration lower than the K_M of WT-PaS was deliberately used that resulted in sighting PVIV-PaS mutant with a K_M of 4.5 µM compared to 35 µM for WT-PaS.

The two additional mutations in PVIV-PaS distinct to PV-PaS template are L157V and K158I. Presumably the loss of the positive charge and the introduction of a smaller hydrophobic residue in the K158I mutation accounts for the enhanced substrate affinity resulting from the increased interactions with the hydrophobic steroid skeleton. Alternatively, both mutations to smaller residues ease interactions between the α-helix and adjacent loops potentially leading to an opening of the pocket.



WT-PaS	
V_{max}	0.0972 ± 0.0039
K_M	35.1 ± 4.6
χ^2	7.5673e-5
R^2	0.99419

PVIV-PaS	
V_{max}	2.07 ± 0.01
K_M	4.5 ± 0.2
χ^2	0.0040229
R^2	0.99876

Figure 2.18: Substrate saturation kinetics, at 37 °C, pH 9 (50 mM Tris acetate), fitted to the Michaelis-Menten model: $V_o = V_{max}*[S]/(K_M + [S])$ by using KaleidaGraph 4.5 software. Specific activity (V_o, µmol min^{-1} (g protein)$^{-1}$) of PVIV-PaS mutant, 0.08 µM (blue) and WT-PaS, 0.4 µM (black) are plotted against TS concentration ([S], µM). The error bars correspond to standard deviation between triplicates. The kinetic parameters for each plot are tabulated beside the chart.

2.2.5.3 Multiple site mutagenesis 3 (MSM 3)

An additional site, MSM 3 focused on two residues, M72 and E74 (Figure 2.19), on the opposite side of the binding pocket compared with previous sites. Research by Kintses *et al.* [28], used directed evolution to improve PaS for hydrolysis of a bulky phosphonate ester. This resulted in beneficial mutations of residues 72-75 that are in close proximity to the binding site. The results of this study were taken into consideration when the residues for mutagenesis of MSM 3 were selected. These residues are 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 of >95% probability being tested.

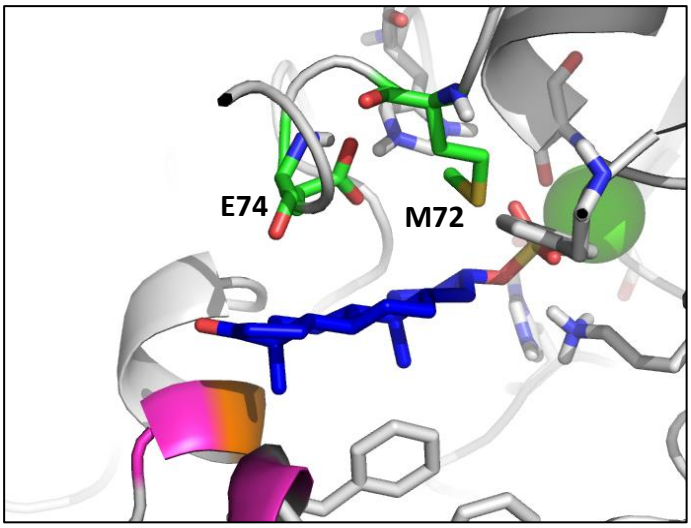


Figure 2.19: Pose II and the residues selected for mutagenesis in MSM 3 (green sticks). The mutational locations of the template, PVFV-PaS are shown in pink and orange. The Ca^{2+} ion (green) and DHEAS docked (blue) are shown.

Unlike the positive results obtained by Kintses *et al.* here a library of inactive PaS variants was obtained for TS hydrolysis suggesting that the steroid sulfate hydrolysis could proceed via a different catalytic pathway to the phosphonate ester hydrolysis by PaS. Importantly, as the NYS degenerate codon allowed methionine to be coded, but not glutamate, suggesting the deleterious mutation must occur at E74 rather than at M72. The results suggest that E74 is involved in the catalytic process. A typical activity profile of a single plate is shown below (Figure 2.20), which profiles the substantially reduced activity of the mutants.

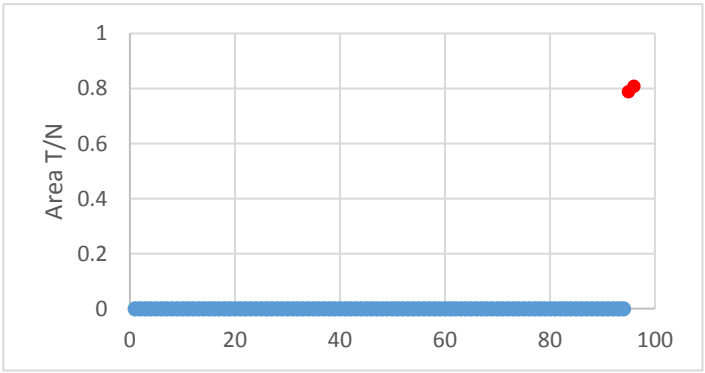


Figure 2.20: Activity profile of crude lysates of a single plate of MSM 3 with the tested samples (blue) in comparison to the PVFV-PaS control samples (red).

2.2.6 Shuffled library

The beneficial mutations found from all the libraries were recombined (shuffled) to test for further improvements in catalytic activity. In addition to the mutations from the best mutants discovered from each library, the mutations from PLVVSV-PaS mutant from MSM 2 library was also included for shuffling as the initial hydrolysis rate of this mutant was greater than that of PV-PaS or PVFV-PaS and also because the mutant provides a range of additional mutations.

Table 2.5: Randomization for the shuffling library

PaS variant	155	156	157	158	160	325	330
WT	R	I	L	K	T	L	K
PV	P	-	-	-	-	-	V
PVFV	P	V	-	-	-	F	V
PVIV	P	-	V	I	-	-	V
PLVVSV	P	L	V	V	S	-	V
Degenerate codon*	CST	VTT	STG	RWA	WCC	TTS	-
Expected AA's	R/P	I/V/L	L/V	K/I/V/E	T/S	L/F	K/V

*S=C/G; V=A/C/G; R=A/G; W=A/T

The shuffle involved mutations at seven residues, as outlined in Table 2.5, with WT-PaS or PV-PaS used as the template. Suitable degenerate codons were available at all positions with the exception of position 158. Here, in addition to the targeted amino acids, a glutamate (E) residue was included for randomization in order to use a simple degenerate codon (RWA) for library preparation.

Variation at codon 330 was not introduced by oligonucleotide primers, instead two templates (WT-PaS and PV-PaS) were used to obtain the required residues (K or V) at this position while randomizing the other positions with the designed primers. An equal number of clones with both templates was screened with the same low TS concentration (20 µM) used in MSM 2. From a primary screen of 1152 clones, 30 samples were selected for the secondary screen that was performed in triplicates out of which 22 samples that had comparable activity were sequenced resulting in 14 unique mutants (Table 2.6). The entries 1 and 10 were found three times and entries 3, 6, 13 and 14 were found two times in the 22 samples identified by the secondary screen.

Table 2.6: The mutations of the beneficial mutants observed at each mutated position.

Sample no.	155 R/P	156 I/V/L	157 L/V	158 K/V/I/E	160 T/S	325 L/F	330 K/V	Frequency observed
1	R	I	L	K	T	F	K	3
2	R	I	V	I	T	L	K	1
3	R	L	L	E	S	F	K	2
4	R	L	L	E	T	F	K	1
5	R	V	V	I	T	L	K	1
6	P	I	V	V	T	L	V	2
7	P	L	V	I	S	L	V	1
8	P	V	V	I	T	L	V	1
9	R	I	V	I	S	L	V	1
10	R	I	V	I	T	L	V	3
11	R	I	V	V	T	L	V	1
12	R	L	V	I	T	L	V	1
13	R	V	V	I	T	L	V	2
14	R	V	V	V	T	L	V	2

The expected residues at each position is shown (blue) underneath the residue number.
The mutations are highlighted in grey.

Interesting mutational patterns were observed among the beneficial mutations. From the four possible combinations with residues 157 and 325, L157V/L325 or L157/L325F were the only combinations observed in all the beneficial mutants while L157V/L325 was dominant (79 %). The L157V/L325 combination always occurred with the mutants that contained K330V while the L157/L325F combination was present in three of the five unique mutants with K330. Furthermore, R155 was always unmutated when K330 was also unmutated while the K330V mutation was present with both R155 and R155P variants. The wild-type residues were more frequently observed than the mutations at the 155, 156, 160 and 325 positions occurring in 11, 6, 11 and 11 mutants respectively. In contrast, the 158 position was unmutated only in a single mutant. This suggests that 158 residue could be mutated by SSM to test the optimum residue at this position. However, the combination of R155P/K330V mutations used as the template for the generation of MSM 1 and MSM 2 libraries was observed only in three out of 14 mutants. These observations support the importance of incorporating the wild-type residues into the shuffle.

Given the rare occurrence of the R155P/K330V combination, the previously identified best mutants, PVFV, PVIV or PLVVSV based on this template were not

observed in the shuffled library. However, entry No 7 has a similar sequence to PLVVSV with a K158I mutation instead of K158V and entries No 6 and 10 are similar to PVIV with a single change at 158 and 155 positions respectively. Interestingly and unexpectedly, K158E mutation was observed in two beneficial mutants that always occurred with R155, I156L, L157, L325F and K330.

Six PaS variants that showed the greatest crude lysate activity were purified and the initial hydrolysis rates for TS hydrolysis were tested (Table 2.7) for confirmation.

Table 2.7: The beneficial mutants found in shuffled library and the initial hydrolysis rates of TS (20 µM) hydrolysis by the purified mutants (0.04 µM) relative to PV-PaS (0.4 µM).

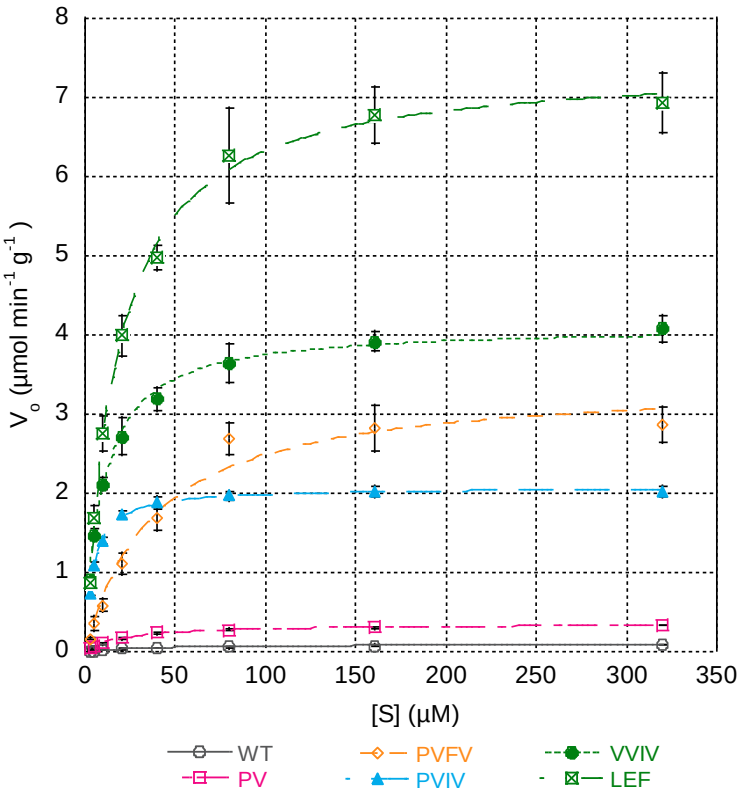
No.*	Mutations	Initial hydrolysis rate relative to PV-PaS[a]
11	R155/I156/**L157V**/**K158V**/T160/L325/**K330V**	11.1 ± 0.49
8	**R155P**/**I156V**/**L157V**/**K158I**/T160/L325/**K330V**	11.9 ± 0.06
10	R155/I156/**L157V**/**K158I**/T160/L325/**K330V**	13.6 ± 0.14
13	R155/**I156V**/**L157V**/**K158I**/T160/L325/**K330V** **(VVIV)**	15.3 ± 0.27
4	R155/**I156L**/L158/**K158E**/T160/**L325F**/K330 **(LEF)**	30.6 ± 1.19
3	R155/**I156L**/L158/**K158E**/**T160S**/**L325F**/K330	31.5 ± 1.11

* Mutant numbers from table 2.6

[a] The initial hydrolysis rates measured under 1 hr reactions at 37 °C, pH 9 (50 mM Tris acetate). The errors correspond to the standard deviation between triplicates.

Surprisingly, the two mutants incorporated the unexpected K158E substitution (No. 3 and No. 4) exhibited the greatest initial rates for TS hydrolysis. These two mutants were found as equally beneficial for TS hydrolysis having similar rates with T or S at 160 position.

Two mutants, one without K158E substitution VVIV-PaS (No. 13) and one with K158E substitution LEF-PaS (No. 4) were selected for further characterization. Kinetics studies were performed on these mutants at pH 9 and both showed a similar improvement in V_{max}/K_M of ~154 times (Table 2.8) with a compensation between the V_{max} and K_M. Figure 2.21 shows the substrate saturation kinetics fitted to the Michaelis-Menten model of the two shuffled mutants in comparison to the previously identified best mutants in each round and WT-PaS.



LEF-PaS	
V_{max}	7.43 ± 0.10
K_M	17.4 ± 0.9
χ^2	0.0923
R^2	0.9988

VVIV-PaS	
V_{max}	4.11 ± 0.05
K_M	9.6 ± 0.5
χ^2	0.037238
R^2	0.99805

Figure 2.21: Substrate saturation kinetics, at 37 °C, pH 9 (50 mM Tris acetate), fitted to the Michaelis-Menten model: $V_o = V_{max}*[S]/(K_M + [S])$ by using KaleidaGraph 4.5 software. Initial velocity (V_o, µmol min^{-1} (g protein)$^{-1}$) of PaS variants (WT, PV and PVFV at 0.4 µM, PVIV, VVIV and LEF at 0.08 µM) is plotted against TS concentration ([S], µM). The error bars correspond to the standard deviation between triplicates. The kinetic parameters for LEF-PaS and VVIV-PaS are tabulated beside the chart

The discovery of K158E mutation in LEF-PaS was serendipitous. The original positive charge at the residue is converted to a negative charge upon mutation. This mutation may allow a salt bridge to form with R155; the α-carbons of residues 155 and 158 are 4.9 Å apart (Figure 2.22). Since both residues are solvent exposed they are expected to be conformationally flexible, but a salt bridge could restrict flexibility and alter the opening to the substrate binding pocket.

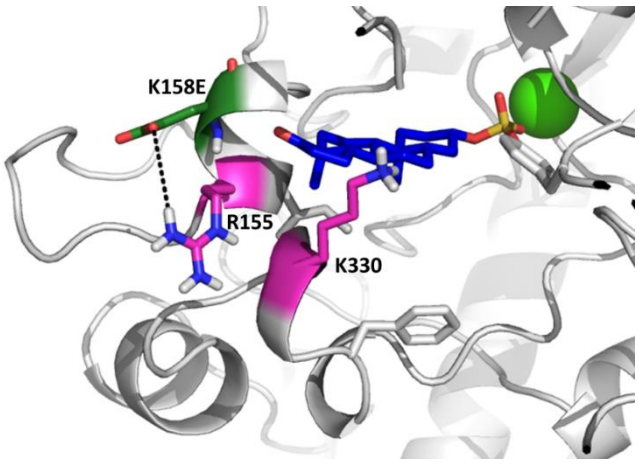


Figure 2.22: Pose II and proposed salt bridge formation between R155 and the K158E mutation imposed on the model derived from the crystal structure (1HDH) [24].

The K158I mutation observed in the PVIV-PaS (K_M of ~4.5 µM) of MSM 2 library recurred in the VVIV-PaS mutant discovered from the shuffled library that also has a significant improvement in substrate affinity with a K_M of ~9.6 µM compared to WT-PaS (K_M of ~35 µM). The loss of the positive charge and the introduction of a small hydrophobic residue could be among reasons for the greater TS affinity observed. In addition, K158 had a mutation in all the beneficial mutants that were purified and tested in the shuffled library (Table 2.7), suggests that this position is worth being mutated by SSM.

2.2.7 Summary of enzyme kinetics of the beneficial mutants

The enzyme kinetic parameters of the best mutants identified from each round of enzyme engineering in comparison to WT-PaS are summarized in Table 2.8. The final concentrations of the enzymes used to study the enzyme kinetics were 0.4 µM for WT-, PV-, PVFV-PaS and 0.08 µM for PVIV-, VVIV- and LEF-PaS.

Table 2.8: The respective mutations, the library they were found in, template used, and kinetic parameters of beneficial mutants over an initial TS concentration from 2.5 – 320 µM, pH 9 (50 mM Tris acetate), 37 °C. See section 2.6.8.1.

PaS enzyme	Library	Template	Kinetic parameters*		
			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	-	-	35.1 ± 4.6	9.72 ± 0.40	2.77 ± 0.38
PV-PaS	-	-	21.7 ± 1.4	35.3 ± 0.60	15.6 ± 1.0
PVFV-PaS	MSM 1 (No. 3)	PV-PaS	37.7 ± 2.4	343 ± 28	91.0 ± 9.4
PVIV-PaS	MSM 2 (No. 7)	PV-PaS	4.5 ± 0.2	207 ± 6.8	463 ± 25.6
VVIV-PaS	Shuffling (No. 11)	WT- or PV-PaS	9.6 ± 0.5	411 ± 19	427 ± 29.7
LEF-PaS	Shuffling (No. 13)	WT- or PV-PaS	17.4 ± 1.0	743 ± 35	427 ± 31.7

*The errors correspond to standard deviation between triplicates.

2.2.8 Alterations in the screening protocols

During the enzyme engineering work, several variations to the evolution protocol were investigated in an effort to improve the screening process. Although these proved ultimately fruitless, they are briefly described below.

2.2.8.1 Minimal media selection

To reduce the amount of screening, a minimal media selection was attempted in the MSM 2 prior to screening as described in section 2.6.11. This protocol was developed by Dr. Bradley Stevenson of McLeod group, ANU and then deployed in the current work. The *E. coli* was grown on minimal growth media (MMTS) without any other sulfates but testosterone sulfate. In theory, considerable growth of colonies should be observed if they could release sulfate by the hydrolysis of testosterone sulfate, which is the only source of sulfur available in the media. Agarose was used instead of agar due to the presence of sulfur in agar. Glycerol was used as a carbon source as it has previously been shown in the laboratory that glycerol affords higher sulfatase expression in comparison with glucose. The chloride salts of Mg^{2+} and Ca^{2+} were used to avoid sulfates from the original recipe. Despite IPTG being a sulfur-containing additive, it is not metabolized by *E. coli* but it induces the expression of the sulfatase via the lac promoter.

Pin-prick colonies were observed in 3 days on incubation at 30 °C and the plates were transferred to 37 °C. Distinguishable sizes of clones for PVFV-PaS and PV-PaS used

as controls were identified within a week of plating. However, the sizes of library clones were not discernible for an appropriate selection even after 10 days of growth. From a challenging selection, 10 clones were re-grown on LBA, out of which eight did not grow, as they were false positives, or could have been dead clones due to prolonged incubation. Hence, the usual screening protocol was continued in place of selection on minimal media plates, as this technique needs further optimization to be used for PaS selection for steroid sulfate hydrolysis.

2.2.8.2 Four-substrate screen

The need of PaS to be evolved for other steroid sulfates led to efforts to perform the simultaneous evolution on different steroid sulfates of interest. As a first step the variants selected from the primary screen of MSM 1 as beneficial for TS hydrolysis were subjected to a secondary screen that included a mixture of β-configured steroid sulfates (TS and NS) and α-configured steroid sulfates (ECS and ETS).

An LC-MS method was developed to detect the steroids and their sulfates by changing the eluting solvent. An isocratic solvent system was used with 58 % MeOH to achieve a better separation with a run time of 8 minutes (Figure 2.23).

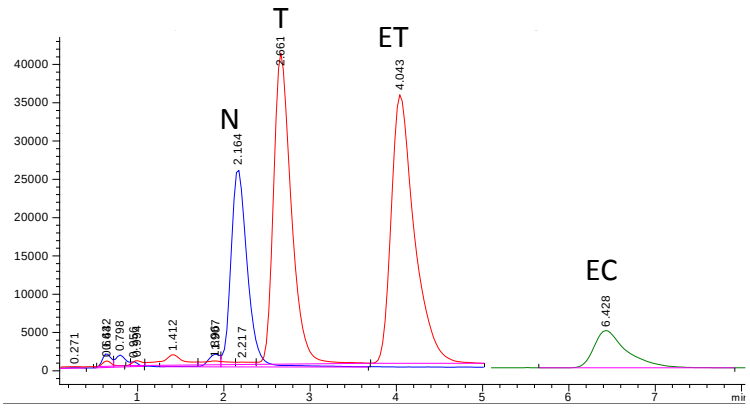


Figure 2.23: The overlay of the chromatograms from the positive mode ESI single ion monitoring (SIM) of the four expected product peaks with three different *m/z* values.

The reaction times were extended to obtain measurable hydrolytic activity with the α-configured steroid sulfates (ET and EC). Incubating the crude cell lysates with the four substrates in one pot for 21 hours made the variants less distinguishable for the β-

configured steroid sulfates and yet no detectable activity was observed for the hydrolysis of α-configured steroid sulfates. Therefore, the secondary screen was moved back to screening and confirming the mutants against TS hydrolysis only.

2.2.9 Validation of libraries

The completeness and the coverage of the screened libraries were validated theoretically and experimentally immediately after preparation.

2.2.9.1 Completeness of the libraries

Table 2.9: Completeness of the libraries in DNA and amino acid diversities.

Library	Codon degeneracy	Screened library size	DNA completeness (%) (GLUE)	Amino acid completeness (%) (GLUE-IT)
SSM 1	NNS	355	100	100
SSM 2	NNS	361	100	100
SSM 3	NNS	189	100	100
SSM 4	NNS	278	100	100
SSM 5	NNS	284	100	100
MSM 1	NYS	1728	34	81
MSM 2	NYS, 156-VTT	4800	32	79
MSM 3	NYS	768	95	99
Shuffled	See Table 2.5	1536	98	98

The completeness of the libraries screened was calculated (Table 2.9) as described by Firth *et al.*[29, 30]. The two online tools used to estimate the completeness were GLUE for DNA completeness and GLUE-IT for amino acid completeness. The GLUE tool helps in finding the expected DNA sequence completeness when the number of equally probable variants within the screened library size is used as input. The GLUE tool is applicable when the gene variants have an equal probability of occurring in the library as with targeted mutagenesis. The GLUE-IT (GLUE - Including Translation) tools calculate the amino acid completeness of the library when the codon degeneracy and the screened library size is used as input.

The expected completeness of the SSM libraries, the MSM 3 library and the shuffled library have achieved a complete coverage (≥ 95 %) whereas one third completeness is achieved by the MSM 1 and MSM 2 libraries in DNA level. At the amino acid level considering degenerate codons for the same amino acid, the SSM, MSM 3 and shuffled libraries have ≥ 98 % completeness. The completeness of the MSM 1 and MSM 2 libraries improves to ~ 80 %.

2.2.9.2 Quality control of libraries

Three SSM libraries (SSM 1 (F328), SSM 4 (K330) and SSM 5 (R155)), all the MSM libraries and the shuffled library were validated experimentally for completeness. Several variants were randomly picked (small-scale library) and their diversities were evaluated by DNA sequencing as a check of adequate coverage of sequence space. The DNA of 32 randomly picked clones from three SSM libraries were sequenced and the presence of nucleotide bases is depicted in the pie charts in Figure 2.24.

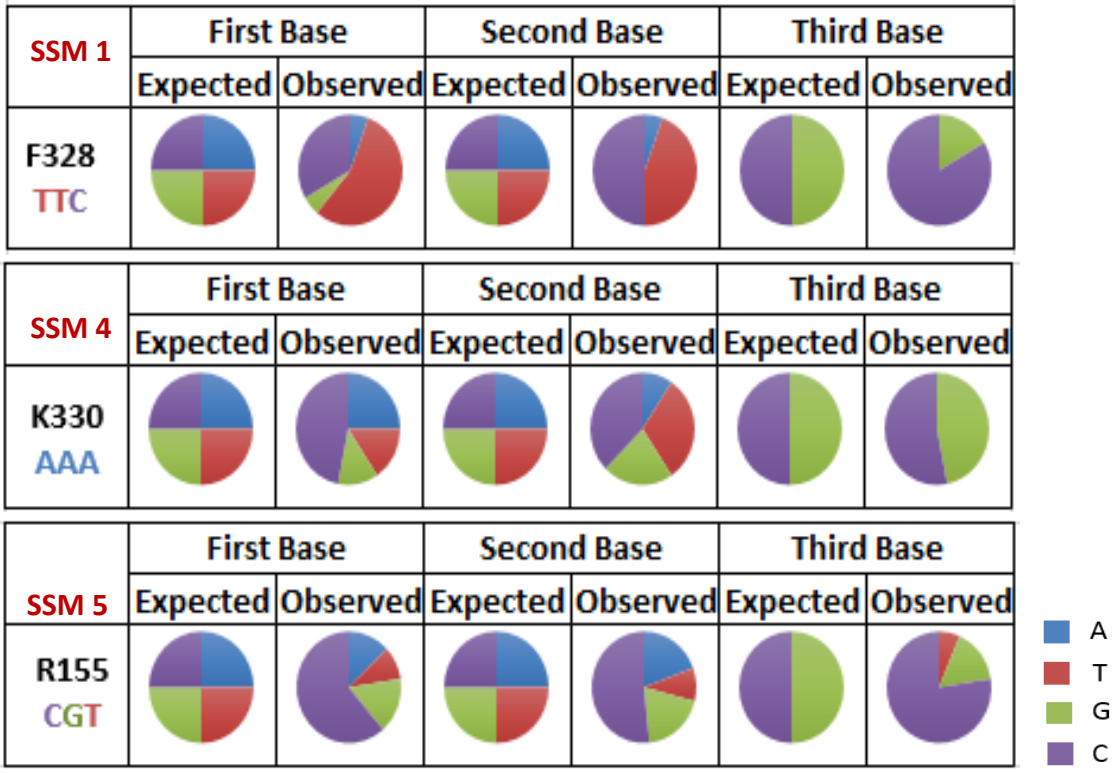


Figure 2.24: Degeneracies of small-scale libraries of SSM 1 (328), SSM 4 (330) and SSM 5 (155). The four colours represent the four nucleotide bases: A (adenine): blue; T (thiamine): red; G (guanine): green; C (cytosine): purple

Although 32 clones from each library were sequenced, the numbers of complete sequences obtained were 18, 32 and 31 samples from the SSM 1, SSM 4 and SSM 5 libraries respectively. The presence of thiamine (*T*) as the third base in SSM 5 is an unexpected observation, which could be a result of wild-type contamination. Whereas the absence of guanine (*G*) as the second base in the SSM 1 library would result in the absence of cysteine, tryptophan, arginine, and glycine that are only coded by a *G* as the second base. However, the complete SSM 1 library being approximately 20 times the small-scale library evaluated for completeness has a greater chance of these residues being tested. The amino acid residues observed in each of the small-scale libraries are shown in Table 2.10. Assuming a 100 % coverage of all 20 amino acids from 32 clones, the representation of amino acid coverage in the small-scale libraries of SSM 1, SSM 4, and SSM 5 are 62 %, 65 % and 67 % respectively.

Table 2.10: Mutations observed from the small-scale libraries of SSM 1, SSM 4 and SSM 5. The amino acids are represented by their single letter code.

	G	A	V	L	I	M	W	F	P	S	T	C	Y	N	Q	D	E	K	R	H
SSM 1			x	x				x	x	x	x		x							
SSM 4		x	x	x	x	x		x	x	x	x		x	x	x				x	
SSM 5		x	x	x	x				x	x	x	x	x			x		x	x	x

x: Presence of mutation
Grey highlights: wild-type residue

Over-representation of some codons in the libraries was observed. However, most of the small hydrophobic residues were observed in the small-scale library. The primary screen that was 8-20 times larger than this sample library has a greater probability of having screened all the mutations of interest. The lesser number of mutations observed in the SSM 1 reflects the low number of successful sequencing results acquired in the particular library.

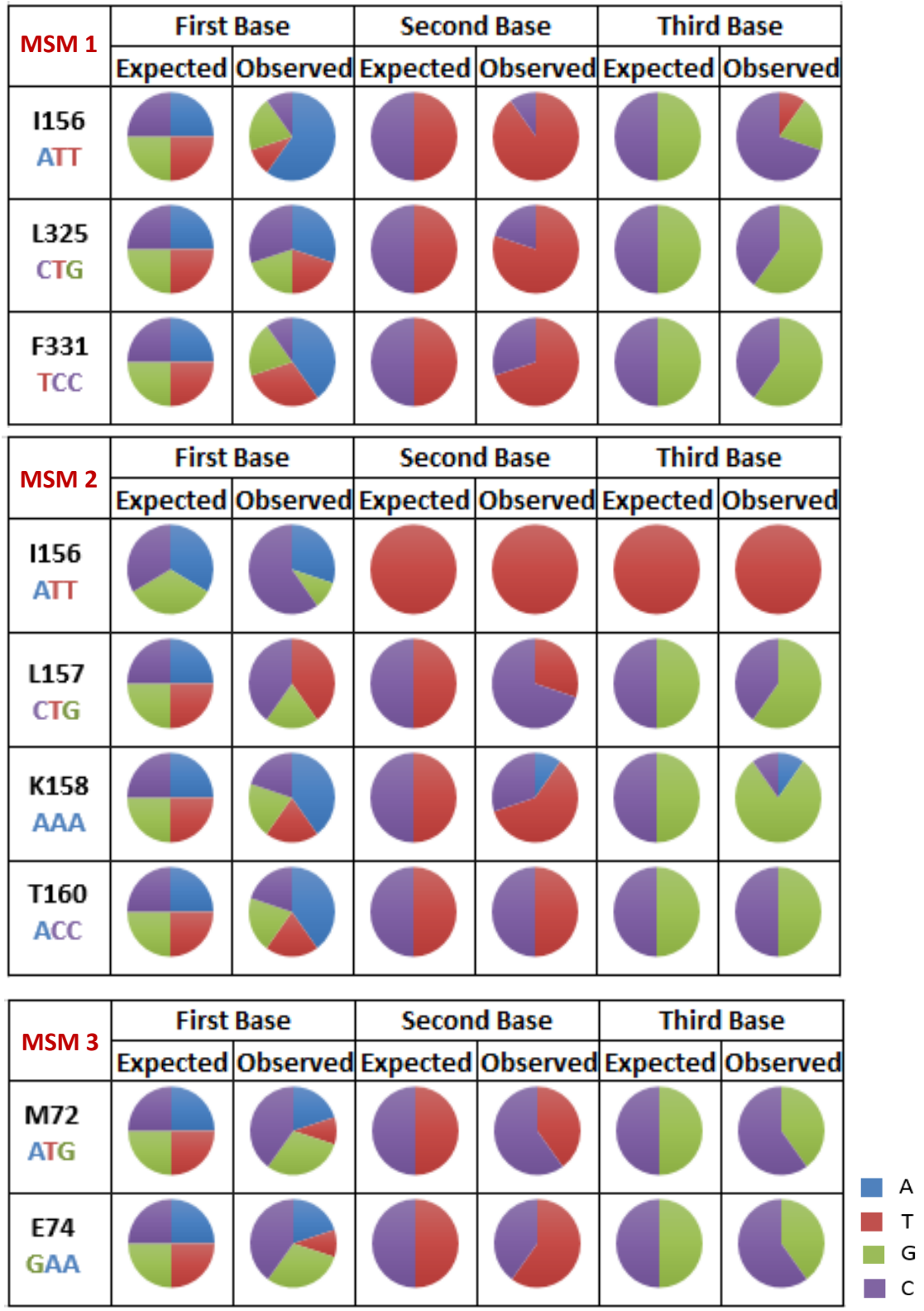


Figure 2.25: Degeneracies of small-scale libraries of MSM 1, MSM 2 and MSM 3. The four colours represent the four nucleotide bases: A (adenine): blue; T (thiamine): red; G (guanine): green; C (cytosine): purple.

For each of the MSM libraries and each half of the shuffled library (with WT- or PV-PaS template) 10 clones were randomly selected for sequencing. The presence of the nucleotide bases of the small-scale MSM libraries (Figure 2.25) and the presence of the amino acids in the shuffled library are presented in pie charts (Figure 2.26).

The MSM 2 and MSM 3 libraries introduced the mutations through synthetic duplex oligonucleotide inserts with the required codon degeneracy whereas the introduction of mutations in the SSM libraries and MSM 1 was PCR based. Hence, a biasing of the mutagenesis towards the template codon in the SSM libraries was observed in certain occasions such as in the first and third bases of SSM 1, the first base of SSM 5 and the first and second bases at the 156 position of MSM 1 library. However, a critical bias based on the template codon was not observed in the MSM 2 and MSM 3 libraries. Template contamination of a single colony each was observed in the MSM 1 and MSM 2 libraries. The SSM 4 and MSM 3 small-scale libraries exhibit a reasonable representation of the codon degeneracy.

Table 2.11: Mutations observed from the small scale MSM libraries. The amino acids are represented by their single letter code.

		A	V	L	I	M	F	P	S	T
MSM 1	156		x	x	x		x			x
	325		x	x		x	x		x	x
	331		x	x	x	x	x		x	x
MSM 2	156		x	x	x					
	157	x		x			x	x	x	
	158		x	x		x		x		x
	160	x	x		x	x	x	x	x	x
MSM 3	72	x	x	x				x		x
	74	x	x	x	x			x	x	

Out of the nine amino acids that are expected to be observed in the MSM libraries with the NYS codon randomization, isoleucine, methionine, and phenylalanine have the least probability of being observed (1/16) while leucine has the highest probability (3/16). Valine, serine, proline, threonine, and alanine have equal probability of being observed (1/8). The absence of alanine (codon: GCN) or proline (codon: CCN) in the small scale library of MSM 1 is due to the under-representation of a cytosine (C) as the second base. However, the mutations that were absent in the small scale library of the MSM 2

library were observed in the beneficial mutants of this library such as V, I and M at 157 position, A, I and F at position 158. This indicates the probable presence of all amino acids in the complete library despite the absence in the small-scale library.

The presence of the mutations in the small-scale shuffled library shows a good representation at all positions except the 325 position where leucine is over-represented exhibiting 90 % presence (Figure 2.26). The dominant representation of L325 in the beneficial mutants observed (Table 2.6) therefore could possibly be reflecting the library rather than the fitness. The degenerate codon used to randomize leucine and phenylalanine was TTS (S=C/G), selected because TTG and TTC both show high expression levels in *E. coli* [31]. However, there could have been bias towards primers with codons that are more similar to the WT codon (CTG for 325) during the annealing phase of PCR. Leucine's codon (TTG) causes only one mismatch with the template DNA whereas phenylalanine's (TTC) causes two. Using the TTW degenerate codon could have avoided this bias.

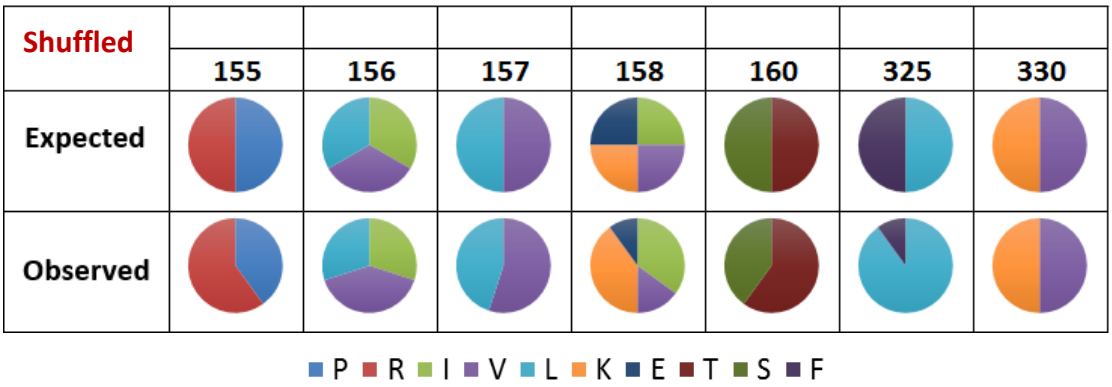


Figure 2.26: Expected and observed amino acids in the shuffled library at each mutated position. The legend below the chart represents the codes for different amino acids.

Overall, the expected quality was observed from the SSM 4, MSM 3 and shuffled libraries while SSM 5, MSM 1 and MSM 2 exhibit some template contamination. The SSM 1, SSM 5, MSM 1 and MSM 2 show considerable bias towards some nucleotides. Importantly, a fair representation of amino acids was observed in all libraries as expected in the theoretical calculations (section 2.2.9.1). However, with the actual library screens being 8-20 times in the SSM libraries and 77-480 in the remaining libraries, there would be a higher chance of observing the completeness specified in Table 2.9.

2.3 Characterization of the PaS variants

2.3.1 The pH dependence

The optimum pH for PaS hydrolysis was determined with respect to two different substrates, TS and PNPS, in a pH range as described in section 2.6.8.2. The test was performed at two different substrate concentrations (Figure 2.27).

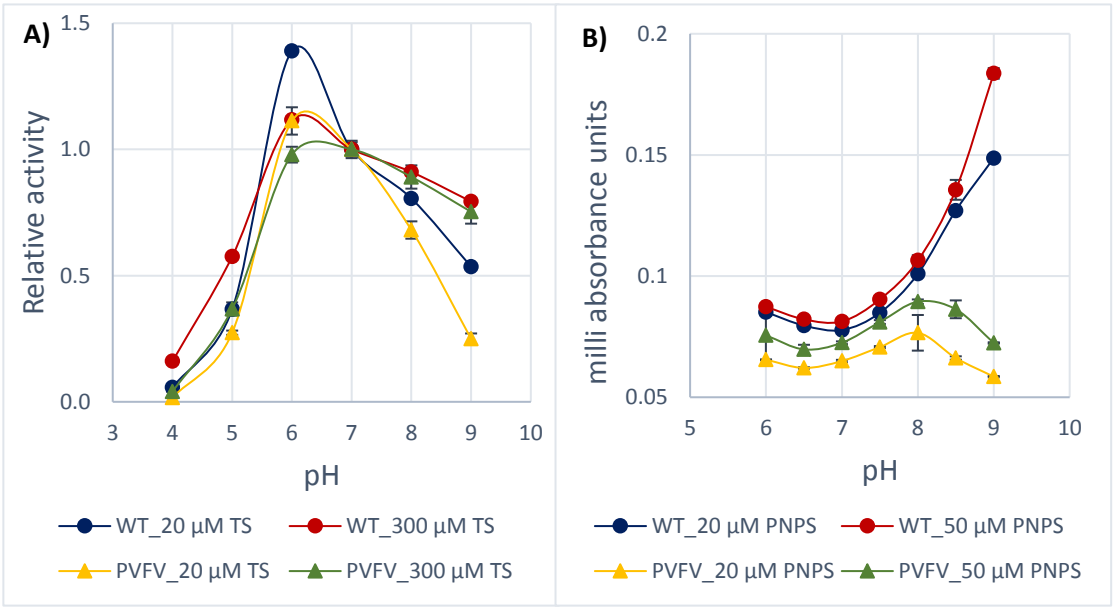


Figure 2.27: The pH profiles of PVFV-PaS and WT-PaS for hydrolysis of two different substrates performed in TMA buffer (50 mM Tris, 25 mM MES and 25 mM acetic acid) at 37 °C. The error bars correspond to the standard deviation between triplicates. A) Hydrolysis of 20 µM or 300 µM TS by 0.4 µM enzyme concentration, from pH 4-9 at increments of 0.5 pH units. The results are presented as relative activities with respect to the activities observed at pH 7. B) Hydrolysis of 20 µM or 50 µM PNPS by 10 nM enzyme concentration from pH 6-9 at every half pH, the data series is not normalized.

The pH profile for TS hydrolysis appears almost bell shaped. The pH optimum for TS hydrolysis, an alkyl steroid sulfate, at approximately at pH 6 could be explained in terms of the local ionisable residues in the active site that are responsible for enzyme catalysis. The general acid catalysed protonation of the departing steroidal oxygen (Figure 2.28) is required for the release of the product. According to the previously reported [32] transesterification-elimination-hydration sequence (Figure 2.28), the sulfate ester oxygen (O1) forms hydrogen bonding with the protonated forms of H211 and

K375. Upon hydrolysis, oxygen will abstract a proton from H211 that acts as a general acid as the sulfate ester substitution occurs. The pKa of the ionisable side chain of histidine is near 6 and would be expected to exist in its deprotonated form under basic conditions. Hence at high pH, impaired general acid catalysis and lower rates of catalysis would be expected. Nevertheless, the bell shaped dependence is more prominent with the hydrolysis at the low substrate concentration (20 µM) profiles as at higher TS concentration (300 µM) the release of product does not decrease as drastically as pH increases.

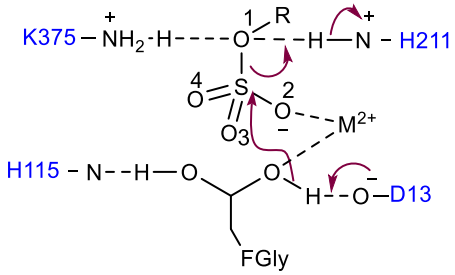
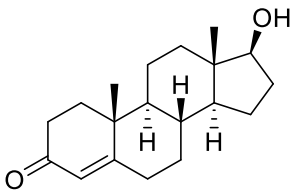
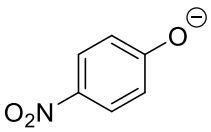


Figure 2.28: The proposed mechanism of transesterification step in the transesterification-elimination-hydration sequence.

The activity drop in the acidic pH range could be due to the D13 (pKa = 3.65) or H115 (pKa = 6) being protonated and the inability to activate the FGH (formyl glycine hydrate) by abstracting the proton from a –OH group for the nucleophilic substitution at sulfur to occur. Hence a decrease of activity is also observed when the pH is lowered.



Testosterone leaving group



para-Nitrophenolate leaving group

Figure 2.29: The leaving groups following TS and PNPS hydrolysis.

The pH profiles of PNPS hydrolysis by PaS enzymes, however, follows a different pattern to TS hydrolysis. The release of a phenolate ion (Figure 2.29) does not essentially require a general acid catalyst for protonation [5] as the *para*-nitrophenolate ion is stabilized by delocalization. Hence alkaline pHs would be best suitable for the hydrolysis

of PNPS. The activity is expected to plateau at high pH, however, the pH dependence was not tested above pH 9 as other factors such as enzyme instability would become then prominent. Further, since PNPS was not the target substrate this study was not pursued on a broader pH range. The shift in the acidic limb to a higher pH for PNPS hydrolysis suggest a change in the role of the general base or even a different base involved in the hydrolysis of PNPS, possibly due to a change in the rate determining step for this substrate.

2.3.1.1 PaS hydrolysis in a range of buffers

The hydrolytic activity of WT-PaS and PVFV-PaS was tested on a range of buffers (50 mM) at two pH values for each buffer differing by one pH unit. The buffers and their corresponding pHs (measured at 50 °C) are tabulated (Table 2.12).

Table 2.12: The buffers used with their corresponding pHs measured at 50 °C

Buffer	Abbreviation	pH at 50 °C
Ammonium acetate	A	8.1
		7.2
Bis-tris propane-HCl	B	8.4
		7.4
Sodium HEPES	H	7.7
		6.7
Sodium phosphate	P	8.0
		7.0
Tris-HCl	T	8.3
		7.3

The hydrolytic activity was tested against TS at 37 °C or 50 °C. It should be noted that the pH values are recorded at 50 °C, hence the actual pH at 37 °C could be slightly varied for some buffers. The results presented are the average of duplicates (Figure 2.30).

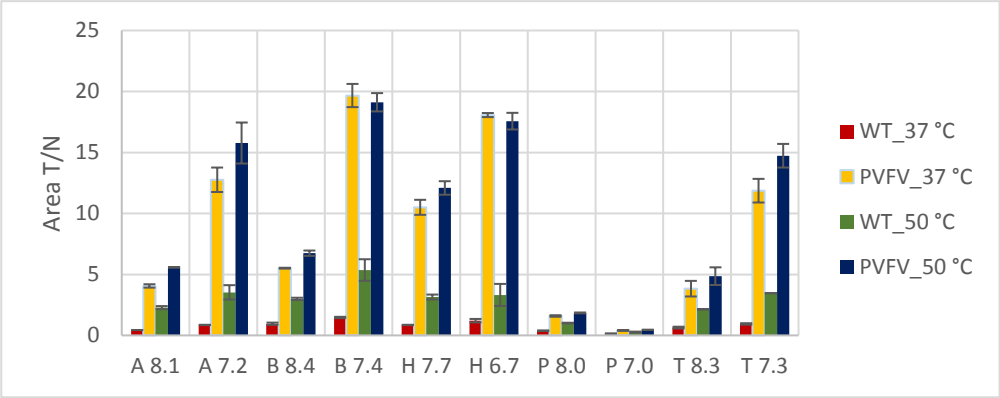


Figure 2.30: The activity profiles by WT-PaS (0.4 µM) and PVFV-PaS (0.4 µM) on TS (20 µM) hydrolysis in different buffers (50 mM) at different pHs. The error bars represent standard deviation of the duplicate measurements.

The results exhibit a general trend of a greater release of testosterone at the lower pH for each buffer across the range that is consistent with the pH optimum previously observed in TMA buffer. The phosphate buffer (pH 5.8 – 8.0) is exceptional showing low activity for PaS and lower activity at pH 7 compared to pH 8 and is not recommended for future studies. Despite Tris (pH 7.5 – 9.0) buffer used in most of the experimental work in this thesis, bis-tris propane (pH 6.3 – 9.5) shows comparatively greater hydrolysis rates and has been used in the previous PaS related studies [33] hence can be deemed suitable for PaS enzyme studies. Ammonium acetate (pH 6.0 – 8.0) and HEPES (pH 6.8 – 8.2) buffers with the useful pH range close to pH 7 will be suitable for the studies performed close to neutral pH.

2.3.2 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.

2.3.2.1 The determination of apparent melt temperatures

The T_m^{app} of the PaS variants in comparison to WT-PaS was determined by incubation of the enzymes for 5 minutes at temperatures ranging from 30 °C - 80 °C at pH 9, and 20 °C – 65 °C at pH 7.5 as described in section 2.6.8.3. Figure 2.31 shows the residual activity of the enzymes as a function of temperature.

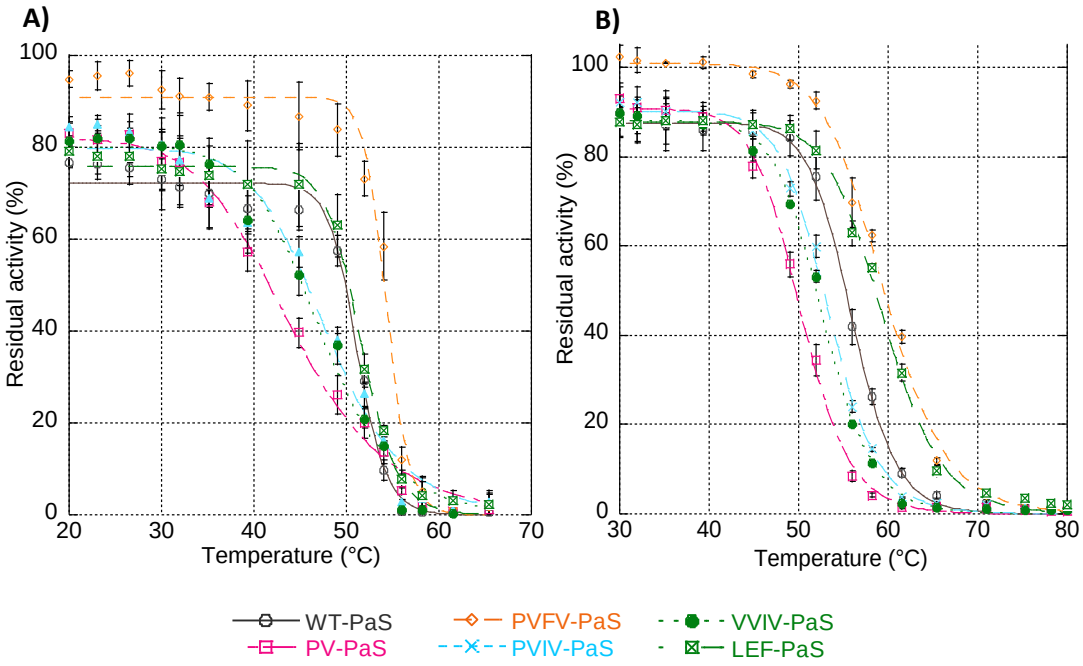


Figure 2.31: Effect of temperature on enzyme stability fitted to a modified form of the Hill equation: $a = a_0 - (a_0.T^h)/(T_{1/2}^h + T^h)$, where the residual activity (a) of the heat treated samples are compared to the activity (a_0) of the non-heat treated samples and plotted against temperature (T) using KaleidaGraph 4.5 software. The temperature ($T_{1/2}$) at which PaS had 50% of its initial activity is considered the apparent melting temperature and h is the Hill coefficient. The figure shows the melting curves of the PaS mutants (20 µM and PVFV-PaS at 80 µM) compared to WT-PaS (20 µM) at A) pH 7.5 (Tris HCl, 50 mM); B) pH 9 (Tris acetate, 50 mM). The error bars represent the standard deviation of biological triplicates.

The PaS mutants and WT-PaS show higher T_m^{app} at pH 9 compared to that at pH 7.5. Two mutants, PVFV-PaS and LEF-PaS, show greater thermostability compared to that of WT-PaS with approximately 4 °C improvement in T_m^{app} at pH 9. However, only PVFV-PaS retains improved thermostability at pH 7.5 (approximately 3 °C higher than WT-PaS).

2.3.2.2 The determination of half-life at 50 °C

The half-life of the PaS variants in comparison to WT-PaS was determined by incubation of the enzymes at 50 °C for 0 – 9.5 hours and measuring the enzyme activity at different time points as described in section 2.6.8.3.

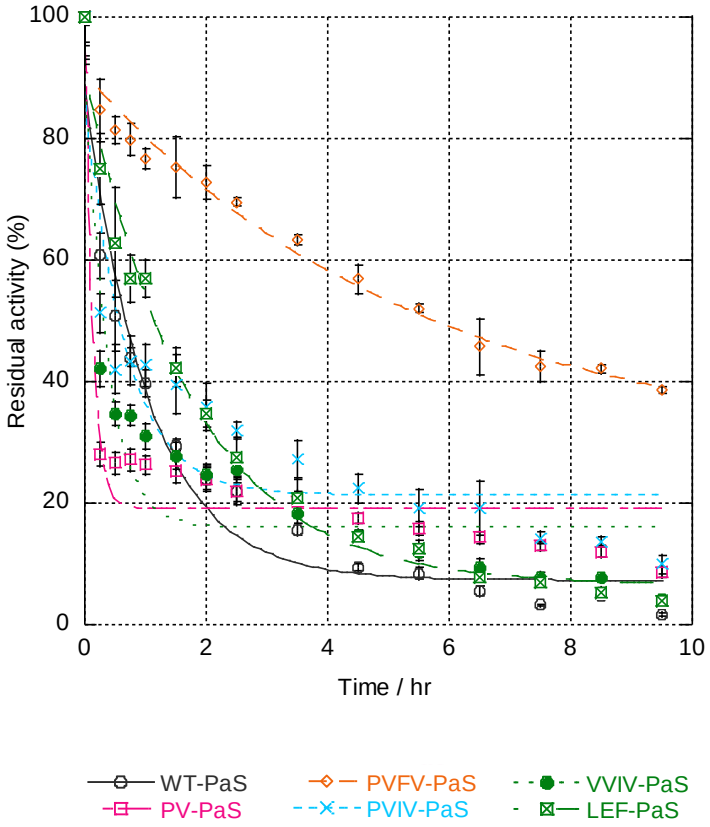


Figure 2.32: Half-life of PaS variants at 50 °C fitted to exponential decay equation: a=100*(m1+(m2*exp(-t*m3))), where the residual activity (a) of the heat treated samples are plotted against the heat treated time (t) using KaleidaGraph 4.5 software. The time at which PaS had 50% of its initial activity (where a=50%) is considered the half-life ($t_{1/2}$) and is calculated using the final value (m1), initial value (m2) and the initial slope (m3). The figure shows the half-life curves of the PaS mutants (20 µM and PVFV-PaS at 80 µM) compared to WT-PaS (20 µM) at 50 °C, pH 9 (Tris acetate, 50 mM). The error bars represent the standard deviation of biological triplicates.

The half-life curves (Figure 2.32) of the PaS variants show that PVFV-PaS has a remarkably greater half-life of ~5.8 hours compared to the other variants and WT-PaS (Table 2.13).

Table 2.13: Thermal parameters for PaS mutants

	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 (hrs)	0.67 ± 0.04	0.12 ± 0.01	5.76 ± 0.14	0.57 ± 0.22	0.30 ± 0.04	1.15 ± 0.13

The PVFV-PaS and LEF-PaS mutants exhibit greater thermostability compared to WT-PaS. The mutation of leucine to phenylalanine at 325 position introduces an aromatic cluster with the neighboring aromatic residues, W212, F328 and F331 as depicted in Figure 2.33. The increased aromatic pi-stacking interactions upon the introduction of L325F could lead to the improvement in thermostability [34-36] of the PVFV-PaS and likewise in the LEF-PaS mutants as outlined in Table 2.13. The improved stability of these mutants could be a feature credited to the improved activity in long reactions.

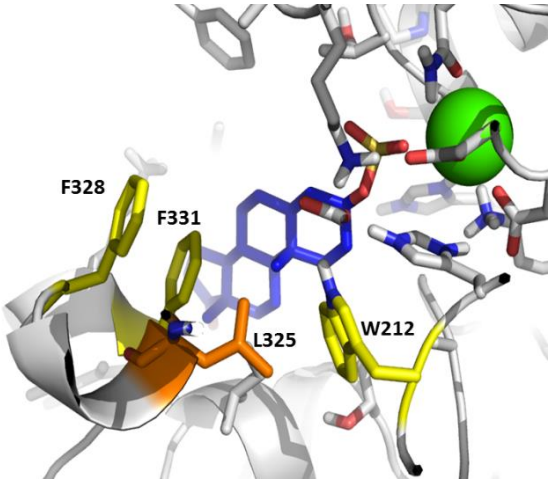
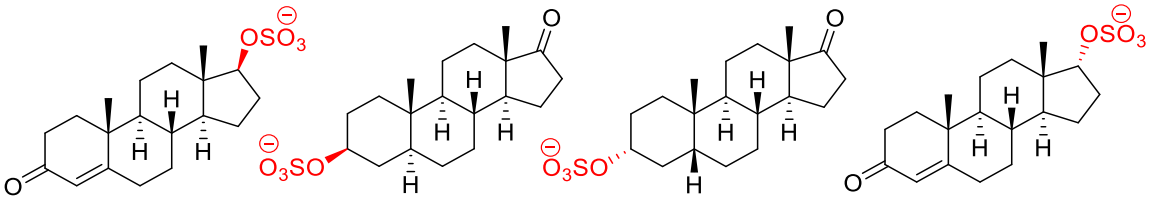


Figure 2.33: The location of L325 (orange) with the neighbouring aromatic residues

However, the K158E mutation of the LEF-PaS could be lowering the thermostability compared to the original lysine. Because, the adjacent loop containing D150 and E151 could form salt bridges with K158 that would enhance the thermostability [37], are removed upon the introduction of K158E mutation. The PV-PaS mutants is the least thermostable variant tested. The reduced thermostability of this mutant could be due to the hydrophobic mutations introduced to solvent exposed residues thereby decreasing the polar surface area [36].

2.3.3 Substrate scope at pH 7.5

The hydrolytic activity of the best mutants from each level of mutagenesis were tested over four steroid sulfates, with the sulfate group varying in stereochemistry and position. The four substrates tested were TS, EAS, ECS and ETS (Figure 2.34). They were tested at pH 7.5 as the enzymes exhibit faster turnover closer to neutral pH for the hydrolysis of steroid sulfates (Figure 2.27), and were quantified against the standards of free steroids. The reaction conditions used are described in 2.2.1. It is noteworthy that the enzyme concentration, substrate concentration and the reaction time used for the hydrolysis of α-configured steroid sulfates is greater, considering the low turnover of these substrates by PaS enzyme.



Testosterone 17-sulfate (TS) | Epiandrosterone 3-sulfate (EAS) | Etiocholanalone 3-sulfate (ECS) | Epitestosterone 17-sulfate

Figure 2.34: range of substrates tested

All mutants exhibit improved hydrolytic activity for TS, the substrate employed in enzyme engineering for improved activity. However, the best mutant for TS may not necessarily be the best for the other substrates. WT-PaS hydrolyses EAS one order of magnitude greater than TS which agrees well with the published results [35]. The EAS hydrolysis varies but does not dramatically change over the course of protein engineering. Considering the low turnover of α-configured steroid sulfates, higher enzyme and steroid sulfate concentrations 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 2.35. Apart from TS hydrolysis showing remarkable improvements, ECS hydrolysis too shows enhanced activity with all mutants compared to WT-PaS particularly PVFV-PaS. The large errors associated with ETS hydrolysis for all but PVFV-PaS suggest that the signals are within the level of noise.

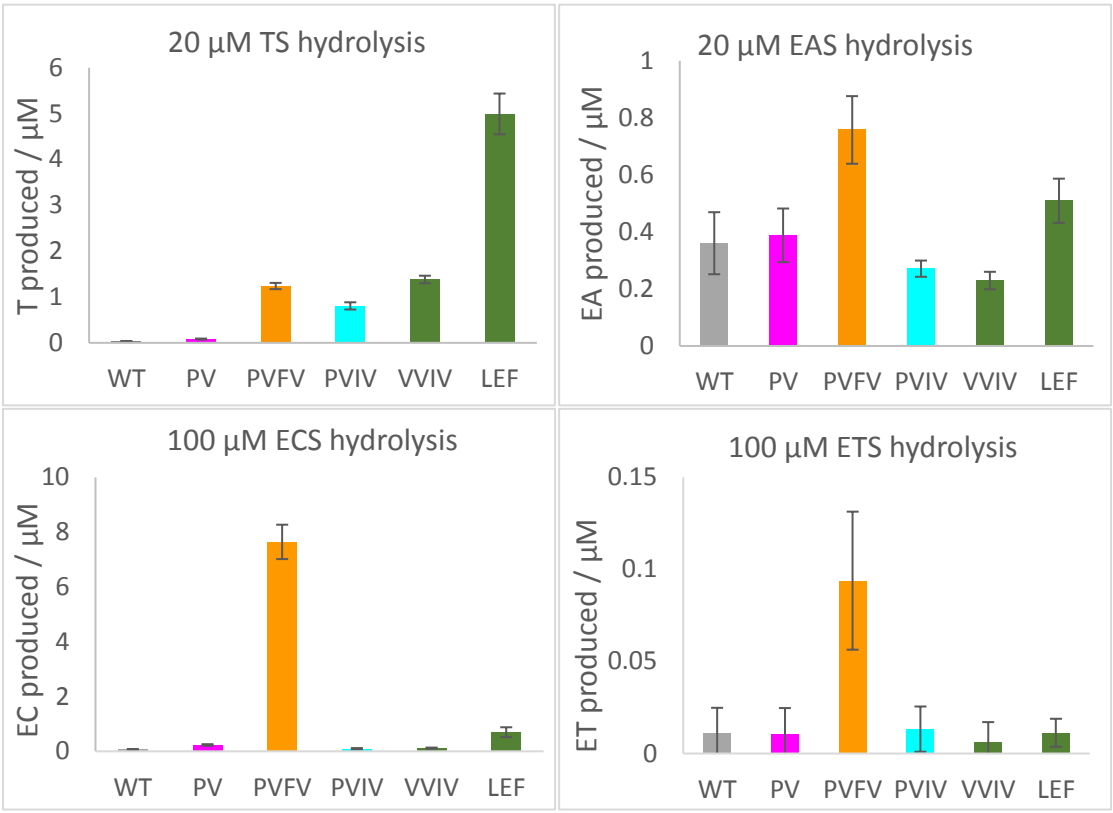


Figure 2.35: The activity profiles of WT-PaS and PaS mutants. The error bars correspond to standard deviation between triplicates.

Amongst the beneficial mutants discovered for TS hydrolysis, LEF-PaS exhibits the highest activity (V_{max}) for TS hydrolysis and PVFV-PaS shows the broadest substrate range. The substrate range improved early in the screening process, but continued screening for a single substrate was detrimental to substrate range, as stated by You and Arnold [39] 'you get what you select for'. This has often been observed in directed evolution.

2.3.4 Level of post-translational modification by Ellman's assay

In PaS, the only cysteine at 51 position (C51) is post-translationally modified to formylglycine (FGly). The level of post-translational modification for each enzyme was tested by performing Ellman's assay that assesses the amount of remaining cysteine in the enzyme that has not undergone post-translational modification to FGly. A similar study performed by Bojarova *et al.* [33] reported that 97% of the cysteine residues of PaS is modified to FGly. This is a simple spectrophotometric assay that is used for rapid thiol quantification that uses 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) commonly known as

Ellman's reagent. The Ellman's reagent is an aromatic disulfide that reacts with thiol groups to form a mixed disulfide and 2-nitro-5-thiobenzoate (TNB^{2-}) that produces an intense yellow colour and a strong absorbance at 412 nm. One mole of TNB^{2-} is produced per one mole of protein sulfhydryl group (Figure 2.36).

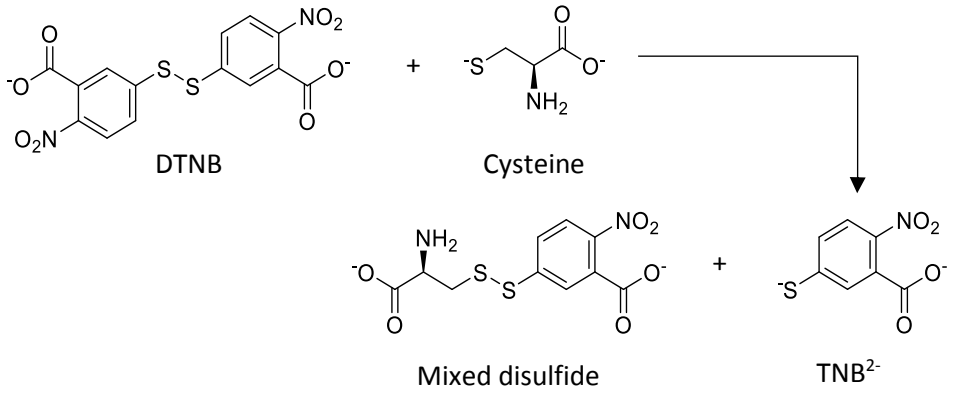


Figure 2.36: Scheme for the reaction of cysteine with Ellman's reagent to produce the dianionic yellow species (TNB^{2-}).

The preparation of standards and protein samples and the absorbance measurements were performed as described in section 2.6.8.5. The path length was considered as 0.57 cm considering the total volume of a well as 350 µL and the volume of a reaction as 200 µL. The extinction coefficient was calculated by dividing the slope of the calibration plot (Figure 2.37) by the path length and was 12295 M^{-1} cm^{-1}. The published values for the molar extinction coefficient of TNB^{2-} varies between 11400 to 14150 M^{-1} cm^{-1} [40].

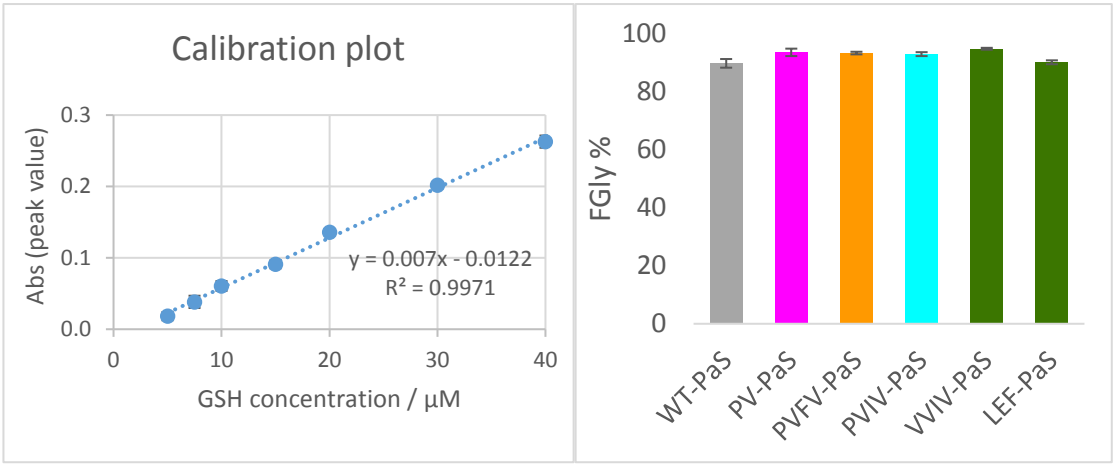


Figure 2.37: A) The calibration plot for the standard series of TNB^{2-}. B) The levels of post-translational modification as a percentage. The error bars correspond to standard deviation between triplicates.

By subtracting the amount of cysteine present, the amount of FGly formed in each enzyme is presented as a percentage. The amount of FGly present in each enzyme is nearly the same with a value ranging from 90-95% for all the enzymes. The range observed could be due to systematic errors in the experiment rather than incomplete post-translational modification. These results suggest that the mutations introduced in each of the PaS variants do not impact on the level of post-translational modification and the improvements observed in hydrolysis activity cannot be attributed to changes in post-translational modification.

2.4 Testing the involvement of E74 residue in catalysis

In the MSM 3 library (2.2.5.3), it was observed that mutating the E74 residue led to a dramatic loss in PaS activity. It is possible that the negative charge on E74 is polarizing H115 where the two residues form an in-line interaction that facilitates activation of the formyl glycine hydrate (FGH) for the nucleophilic substitution (Figure 2.38A).

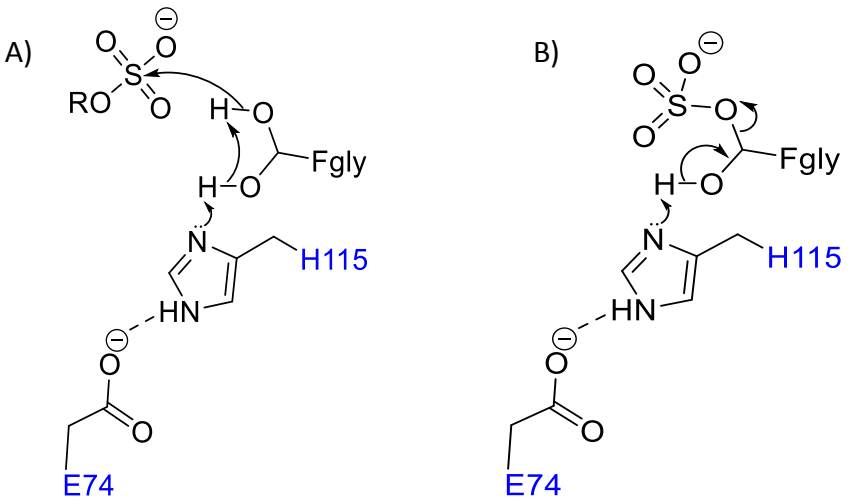


Figure 2.38: Two distinct possibilities for mechanistic involvement of H115 residue as a general base polarized by E74. A) Activation of FGH to facilitate the nucleophilic substitution. B) Promoting elimination of the sulfated FGH intermediate.

This forms a catalytic triad that is reminiscent to the charge relay system seen in seine proteases [41] with FGH as the catalytic nucleophile in PaS. However, this activation is indirect as the hydroxyl group directly activated by H115 is the one further from the

sulfate. The involvement of E74 residue in the catalytic cycle reflects the same transesterification-elimination mechanism proceeding via a different residue (H115) for the activation of the nucleophile (FGH) than that depicted in Figure 2.28.

Alternatively, the polarized H115 may serve as a general base to promote elimination of the sulfated FGH intermediate, thus regenerating the enzyme (Figure 2.38B). At first glance, the latter is the more likely occurrence as no examples are known for H-bonding between the geminal hydroxyl groups of FGH-like diols in small molecule crystals, examined in the Cambridge structural database (CSD) as implied by Figure 2.38A.

To test the hypothesis that the E74 residue is involved in catalysis, site directed mutagenesis was performed to produce three mutants at the 74 position: E74Q, E74D or E74A (Figure 2.39) as described in section 2.6.9.1. Glutamine would test the necessity of the negative charge functionality without changing the size drastically while aspartate reduces the bulk while maintaining the functionality whereas alanine removing both the charge and bulk.

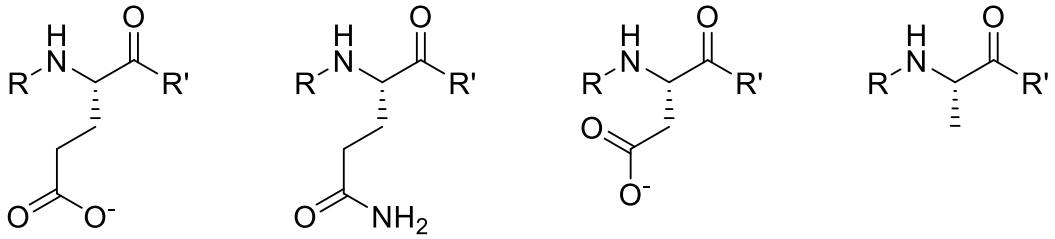


Figure 2.39: The mutations introduced at the 74 position with the wild-type residue (Glutamate), where R and R' represent the linked peptide chains.

Kinetics for PNPS hydrolysis was studied with the purified E74 variants and WT-PaS and the kinetic parameters were determined (Table 2.14) by fitting the data to the Michaelis-Menten model.

Table 2.14: Kinetic parameters of PNPS hydrolysis by WT-PaS and the E74 variants at pH 9 (50 mM Tris acetate), 37 °C. See section 2.6.9.2.

Kinetic parameter	WT-PaS	E74Q	E74D	E74A
V_{max} / 10^5 µmol min^{-1} g^{-1}	75 ± 2.0	0.9 ± 0.02	4.0 ± 0.07	2.9 ± 0.11
K_M / µM	8.12 ± 0.5	29.2 ± 3.1	143 ± 6.3	164 ± 24
(V_{max}/K_M) / 10^4 L g^{-1} min^{-1}	92 ± 6	0.31 ± 0.03	0.28 ± 0.01	0.18 ± 0.03

In agreement with MSM 3 results, the study of PNPS kinetics (Table 2.14) of these mutants revealed that all mutants show detrimental effects on PNPS hydrolysis when compared to WT-PaS where the catalytic efficiency (V_{max}/K_M) is 300 fold lower. The K_M of E74Q is 3.6 times greater than WT-PaS and that of E74D and E74A are >17 times greater than WT-PaS. The shape of E and Q are similar with just a terminal amide group replacing a carboxylate (Figure 2.39), hence substrate binding as seen in K_M measurements is not dramatically perturbed, but V_{max} is reduced approximately 83 times indicating the importance of the carboxylate functionality. Out of the mutants, the E74D mutant has the highest V_{max} (28% higher than E74A and 78% higher than E74Q) that supports the hypothesis that the negative charge functionality is important for turnover. This also indicates that not only the functionality, but the proximity of the charge to polarize the H115 residue is also important as observed by the significantly reduced V_{max} of E74D mutant.

Testing the kinetics of EAS/TS was attempted with the E74 variants. Due to impaired activity, assessing of the kinetic parameters was not possible. However, it would be an interesting residue to further investigate with an easier steroid sulfate such as EAS.

2.5 Conclusions and future directions

In the course of engineering, a suitable sulfatase for the hydrolysis of steroid sulfates, PaS was selected as the candidate of choice. By investigation of the active site and the two proposed binding poses five residues were mutated in turn by SSM and the results indicated pose II as the preferred binding mode. This was supported by the results of the latter libraries that targeted multiple residues around the opening of the binding pocket of pose II. By the selection of 12 residues to improve for TS hydrolysis by mutagenesis in five SSM libraries and three MSM libraries and by shuffling the beneficial mutations three mutants (PVIV-, VVIV- and LEF-PaS) were identified with a catalytic efficiency (V_{max}/K_M) > 150 times than WT-PaS.

The enzyme engineering also broadened the substrate scope for PaS. Observing hydrolysis of ECS and ETS, which are α-configured steroid sulfates, was an important breakthrough. The PVFV-PaS exhibited the greatest substrate range with an ability to hydrolyse ECS and ETS with rates two orders and an order of magnitude greater than WT-PaS, respectively. Although the substrate range improved at the early stages of enzyme engineering, continued screening for a single substrate was unfavourable for the substrate range in the latter rounds.

The optimum pH for TS (an alkyl sulfate) lies approximately at pH 6 which is at least 2 pH units lower than the optimum for PNPS hydrolysis (an aryl sulfate). This is explained by the requirement of a general acid to protonate the leaving group of TS whereas it is not required for the leaving group of PNPS. In addition, the study of PaS in a range of buffers indicates that phosphate buffer is inhibitory to the enzyme, while bis-tris propane with a broader pH range is recommended for future work.

The studies on thermostability of the PaS variants indicated that PVFV-PaS possesses the greatest thermostability and is a moderate improvement of ~3 °C in melting temperature. In addition, the half-life of PVFV-PaS (5.8 hr) is significantly higher than WT-PaS (0.7 hr) at 50 °C. The improvements in thermostability would combine with catalytic improvement to give further benefit for PVFV-PaS. The level of post-translational modification observed in all PaS variants range from 90-95% and the differences could be accountable for the systematic errors associated, hence the

improvements in activity observed in the PaS mutants are not due to differences in post-translational modification.

In addition, the importance of the E74 residue was discovered by the observed inactive MSM 3 library. It is proposed that the E74 is engaged in the catalytic mechanism by polarizing H115, which will activate FGH either for nucleophilic substitution or for the elimination of the sulfate group. This opens up an avenue for future investigation. Performing enzyme kinetics for the hydrolysis of a steroid sulfate by the produced E74 mutants and molecular modelling simulations would provide further insights to the involvement of the E74 in the catalytic mechanism.

Molecular modelling of the PaS variants with TS docked, or crystallization of the PaS variants with the substrate will be useful to further understand, how changes to the binding pocket influence enzyme stability and activity. This would also be helpful in further engineering the enzyme for an improved sulfatase.

With the improved substrate scope, PVFV-PaS would be a good starting point to evolve for improved variants for ECS and ETS hydrolysis. The PVFV-PaS could be used as the template and introduce new mutations to the residues mutated in MSM 1 and MSM 2 in small groups. Alternatively introduce mutations by error-prone PCR to discover improved variants for the hydrolysis of α-configured steroid sulfates. In addition, K158 is an important residue that could be mutated by SSM to test the optimum residue at this position as this residue was mostly mutated in the beneficial mutants observed in the shuffled library (Table 2.6). With a version of PaS available for hydrolysis of TS and related steroid sulfates, addition of WT-PaS and subsequent versions would improve the range. Since one variant may not be the best for all substrates as discussed, a cocktail of enzymes could be used for steroid sulfate hydrolysis in anti-doping.

2.6 Experimental

Unless otherwise specified, enzymes used for cloning and restriction digestion enzymes were purchased from New England Biolabs and chemicals were purchased from Sigma Aldrich. Phusion DNA polymerase was used for PCR throughout except in colony PCR. Dehydroepiandrosterone (DHEA) was obtained from BDH (Poole, UK) while the other free steroids were purchased from Steraloids (Rhode Island, USA). Steroid sulfates synthesised by Christopher Waller [42] from McLeod laboratory were used for experimental work. The AtsA gene for *P. aeruginosa* sulfatase was submitted to DNA 2.0 (California) for *E. coli* codon optimization, synthesis and cloning into pJExpress404. Purification of PCR products and gel extractions were done using Promega® PCR purification kit. DNA extractions were done with QIAGEN® DNA miniprep purification kit. Sanger DNA sequencing was performed by the Biomolecular Resource Facility (BRF) in the John Curtin Medical School of Research, Australian National University. MilliQ water was used in all aqueous solutions and HPLC grade methanol was used for LC-MS related work. Unless otherwise stated, amounts and times used in the PCRs, digestions and ligations were according to the respective protocols from the enzyme supplier.

2.6.1 DNA manipulations

This section includes the general protocols for DNA manipulations performed during enzyme engineering.

2.6.1.1 Oligonucleotides

The primer design was facilitated by CLC sequence viewer (version 6.9) software that was used for sequence analysis and restriction site analysis, and was also assisted by Integrated DNA technologies (IDT) OligoAnalyzer 3.1 online tool [43] for estimation of melting temperatures and secondary structure analysis. The primers used for library creation were purchased from Integrated DNA Technologies, USA. The primers were dissolved in 10 mM Tris-Cl at pH 8.5 (Qiagen elution buffer) to yield a 100 µM concentration and stored at -20 °C.

2.6.1.2 Polymerase chain reaction (PCR)

The PCRs were performed in either Veriti-96 well thermocycler or BioRad C1000^{TM} thermocycler. All PCRs were performed in PCR microtubes with a typical reaction containing mini-prep DNA or a purified PCR product as the template (Table 2.15).

Table 2.15: Sample preparation PCR

Component	Volume (µL)
DNA template (50 – 100 ng)	*a*
Forward primer (10 µM)*	1.25
Reverse primer (10 µM)*	1.25
dNTPs (10 mM each)	0.5
Phusion HF buffer (5 X)	5
Phusion HF polymerase (2 U μL^{-1})	0.25
Water	16.75 - *a*
Total	**25**

*Primers are listed in Table A1 (See Appendix).

A typical thermocycle contained an initial denaturation step (98 °C, 1 min), 25 or 30 cycles of amplification step (denaturation: 98 °C, 15s; annealing: 55 °C – 65 °C (Varied depending on the melt temperatures of the primers), 30 s; extension: 72 °C, 30 s – 2 min) and a final extension step (72 °C, 5 – 10 min).

2.6.1.3 DNA separation by gel electrophoresis

The DNA gel electrophoresis was performed to check for the presence and the purity of the PCR products, digested products and ligated products. Most often 0.8 % agarose gels were used with a high range reference marker (0.5 – 10 kb) and less often 1.5 % agarose gels were used with a low range reference marker (25 – 766 bp). The gels were prepared in 1x SB buffer (see Appendix) using RedSafe^{TM} (iNtRON Biotechnology) nucleic acid staining solution (5 µL / 100 mL) and solidified on glass plates with a thickness of ~0.5 cm. the separation distance was ~6 cm. The wells contained a volume of approximately 0.5 x 0.1 x 0.5 cm^3 that could load ~20 µL of sample.

The samples were prepared by mixing the DNA sample (1 µL) with 6x loading dye (1 µL) followed by dilution with water (4 µL). The reference marker (New England

Biolabs) was prepared in a similar way and loaded with the samples. The gel electrophoresis was performed in a Wide mini sub™ gel tank with a voltage of 120 V supplied by BioRad 200/20 power supply. The run time was ~30 minutes and the bands were visualized under UV (254 nm) radiation or blue LED (when required to extract the DNA from the gel).

2.6.1.4 Restriction enzyme digestion

The DNA was digested with 10 units of the appropriate restriction enzyme (New England Biolabs) per 1 µg of DNA in a reaction volume of 50 µL with the manufacturer's recommended buffer. Unless otherwise specified the digestion mixtures were incubated at 37 °C (Qualtex solidstat incubators) for 1 hour. If the digestion was carried out on a PCR product of an amplified gene, then the digested product was PCR purified, whereas if the digestion was carried out on vector DNA, then calf intestinal alkaline phosphatase (CIP, 10 units) treatment was performed along with digestion followed by gel purification.

2.6.1.5 Ligation

The calf intestinal alkaline phosphatase (CIP) treated pJExpress404 vector (0.025 pmol) was ligated to insert in a 1:3 molar ratio unless otherwise specified. The ligation reaction was performed using T4 DNA ligase (1 µL) 10 x T4 ligase buffer (2 µL) diluted up to 20 µL with water. The reaction mixture was incubated at room temperature for 20 minutes followed by DNA purification.

2.6.1.6 DNA purification

DNA purification was performed for several types of DNA. The DNA from PCR products, digested or ligated products were purified by the Promega® PCR purification kit. Gel purification was performed by first subjecting the DNA sample to gel electrophoresis (see 2.6.1.3). The band of the expected DNA was cut from the gel by visualizing under blue light with an orange filter followed by DNA extraction from the gel fragment according to the protocol of the Promega® purification kit. Plasmid DNA

was extracted and purified from overnight cultures (3 mL) using the Qiagen Miniprep ® kit.

2.6.1.7 DNA quantification

The DNA concentration was determined using NanoDrop ND-1000 (Thermo Scientific) spectrophotometer by measuring the absorption at 260 nm. The absorption at 280 nm was also measured as an indication of sample purity. A ratio of absorbance at 260:280 nm greater than 1.8 indicates nucleic acid without protein contamination.

2.6.1.8 Bacterial transformation

Plasmid DNA was transformed into bacterial cells by electroporation using a BioRad MicroPulser according to the manufacturer's instructions. Then 50-70 ng of construct DNA of the ligation product (in water) was transformed into 50 µL of electro-competent *E. coli* DH5α cells. The cells were recovered in 0.5 mL LB medium and were incubated in 1.5 mL tubes at 37 °C, 180 rpm for 1 hour to develop ampicillin resistance. Transformants were plated on LB-agar 100 mg/L ampicillin (LBA) in 90 mm petri dishes and incubated overnight at 37 °C, for colony growth.

2.6.1.9 Colony PCR

Colony PCRs were performed to check for the presence and correct size of the inserts and for DNA amplification for sequencing purposes using the *Taq* DNA polymerase enzyme (New England Biolabs). First, a single colony was resuspended in water (5 µL). Then reaction components were mixed in two tubes: one with *Taq* DNA polymerase, the other with nucleic acid components (Table 2.16).

Table 2.16: Sample preparation for colony PCR

Component	Volume (µL)
A) Polymerase buffer (10 x) without $MgCl_2$	5
$MgCl_2$ (25 mM)	3
Taq DNA polymerase (5 U μL^{-1})	0.25
Water	16.75
B) dNTPs (10 mM each)	1
Forward primer, P30-f (10 µM)*	1
Reverse primer, P31-r (10 µM)*	1
Water	21
Colony resuspension	1
Total	**50**

*Primers are listed in Table A1 (See Appendix).

The B) component was first pipetted into a PCR tube and mixed with 1 µL of the colony suspension followed by the mixing of 25 µL of A. The thermocycling conditions used were, 95 °C for 5 minutes followed by 25 cycles of 95 °C for 20 s, 55 °C for 30 s and 68 °C for 1 minute and a final 5-minute extension at 68 °C.

To check for the presence of the correct size of the insert, 1 µL of the colony PCR product was run on a 0.8 % agarose gel (see 2.6.1.3) with a band at ~2 kb confirming the correct size of the insert. The colony PCR product was purified and was used as the template for sequencing.

2.6.1.10 DNA sequencing

The DNA sequencing was performed by the Biomolecular Resource Facility (BRF) in the John Curtin School of Medical Research, Australian National University. Sequencing was performed on two different types of DNA templates; templates from plasmid purification (see 2.6.1.6) or templates from colony PCR (see 2.6.1.9). The DNA amplification was performed by PCR by using BigDye® terminator (Thermo Fisher Scientific) as recommended by the facility. The samples were prepared in a reaction volume of 20 µL (Table 2.17)

Table 2.17: Sample preparation for DNA sequencing

Component	Volume (µL)
Plasmid DNA (150-300 ng) or Colony PCR DNA (20 – 50 ng)	*a*
BigDye® terminator version 3.1	0.5
Sequencing buffer 400 mM Tris at pH 9 with 10 mM $MgCl_2$	3.5
Primer (3.2 µM)*	1
Nuclease free water	15 - *a*
Total	**20**

*P6-f or P23-f or P24-r see Table A1, Appendix

The thermocycling conditions used were, 96 °C for 5 minutes followed by 50 cycles of 96 °C for 10 s, 50 °C for 5 s and 60 °C for 4 minutes. The PCR amplified DNA samples were prepared for analysis by a clean-up protocol. First, the DNA was mixed with a precipitating solution containing ethanol (62.5 µL), sodium acetate (3M, pH 4.6, 3 µL) and water (14.5 µL). The mixture was incubated at room temperature for 15 minutes followed by centrifugation for 30 minutes at 3700 rpm (Eppendorf centrifuge 5804). The supernatant was removed by an inverted quick spin (300 rpm, 1 minute). The pellet was washed with 70 % (v/v) ethanol (200 µL) and centrifuged for a further 15 minutes (3700 rpm). The supernatant was removed as above. Finally, the pellet was air dried for ~30 minutes before submission for analysis.

2.6.2 Preparation of *E. coli* electro-competent cells

The *E. coli* strain of choice for all forms of this work was DH5α, which is known for its high transformation efficiency, relatively simple preparation and easy storage. All solutions, culture media, water, 1.5 mL tubes were pre-sterilized by autoclaving and the centrifuge bottles were sterilized by washing with 10% bleach, plenty of sterile water and 70% ethanol before drying in a sterile hood. All steps in the preparation was performed under sterile conditions. A single colony of an overnight grown DH5α strain streaked on a LB-agar plate was inoculated into 10 mL of YenB (yeast extract and nutrient broth) medium in a falcon tube. Following overnight growth of the starter culture at 37 °C it was inoculated into 990 mL of YenB medium and incubated at 37 °C shaking at 180 rpm until the OD reached ~0.8. Next, the culture was chilled on ice for 5

minutes and was transferred to the centrifuge bottles stored at 4 °C and centrifuged for 10 minutes at 4000*g*, 4 °C. The supernatant was discarded and the cell pellet was washed with 100 mL ice-cold sterile 10% glycerol and centrifuged as before. This step was repeated. Next, the cells were resuspended in 20 mL of ice-cold 10% glycerol and centrifuged as before. The supernatant was discarded and the cells were resuspended with 200 µL of 10% glycerol for a final volume of ~ 2 mL. The cells were pipetted in 50 µL aliquots into 1.5 mL Eppendorf tubes, snap frozen on dry ice and stored at -80 °C until required.

2.6.3 Protein analysis

This section describes the protocols for the visualization and quantification of the proteins required to validate the purity and the concentration before evaluation of the proteins.

2.6.3.1 Protein visualization by SDS-PAGE

Polyacrylamide gels were used to evaluate the purity of the fractions collected during enzyme purification. The gels were prepared based on a protocol by Laemmli *et al.* [44] by using 15 % separating region (w/v) at pH 8.8 and a 4.5 % stacking gel at pH 6.8 both containing 37.5:1 ratio of acrylamide to *N*,*N*-methylenebisacrylamide (Amresco). The final concentrations of the separation gel are as follows: 0.375 M Tris.HCl (pH 8.8), 0.1 % (w/v) SDS, 0.125 % (v/v) N,N,N',N'-tetramethylethylenediamine (TEMED) and 0.075 % (w/v) ammonium persulfate (APS, BioRad). The final solute concentrations in the stacking gel were similar to the separation gel except that 0.125 M Tris-HCl (pH 6.8) was employed. The gels were cast with a thickness of 0.75 mm and a height of 6 cm and 2 cm for separation and stacking respectively using a multiple gel caster (Amersham Biosciences). The wells were 3.5 mm wide with 15 wells in the stacking region.

The low molecular weight reference ladder (BioRad) was prepared by performing a 20 x dilution in cracking buffer (see Appendix), followed by heating at 95 °C for 5 minutes, and was then stored at -20 °C until used.

The protein samples were prepared by mixing a defined volume of protein solution (~2 µL) and making up to 10 µL with cracking buffer (see Appendix). Samples were

heated at 95 °C for 5 minutes before loading 5 µL of samples or standards into the wells. The gel electrophoresis was performed in a 1x running buffer using SE250 Mighty Small II gel tank (Hoefer) with a maximum current limited to 30 mA per gel and maximum potential difference limited to 200 V, supplied by Pharmacia LKB EP5 500/400 power supply. Electrophoresis was continued until the bromophenol blue solvent front reached the bottom of the gel (~45 minutes). The gel was then removed and washed three times in milliQ water to remove SDS before staining with BioSafe™ Coomassie blue G-250 stain (BioRad) according to manufacturer instructions.

2.6.3.2 Determination of protein concentration

The concentration of proteins was determined in two ways; by measuring UV absorbance or by Bradford assay.

2.6.3.2.1 Bradford assay

The Bradford assay is a colorimetric based assay that is based on the change of the colour of the dye Coomassie Brilliant Blue G-250 under acidic conditions upon protein binding [45]. Upon binding, the absorbance is shifted from 470 nm to 595 nm. The assay is based on the absorbance measurement at 595 nm that was measured using BioStrategy SpectraMax M2 plate reader.

The standards and samples were prepared in 0.9 g/L NaCl solution. Bovine serum albumin (BSA) was used as the standard for calibration. An initial standard of 1.4 g/L BSA (1.3 mL) was serially diluted in saline solution to produce the standard series with a range of BSA concentrations; 1.40, 1.08, 0.83, 0.64, 0.49, 0.38, 0.29, 0.22 g/L. The wells were set up by adding 100 µL water, 5 µL of protein and 250 µL of 20% Bradford reagent to give full wells with little or no meniscus. The absorbance was measured after incubating the wells for 10 minutes at room temperature. The standard curve was plotted with the absorbance at 595 nm against the BSA concentrations. The absorbance of the proteins was measured in a similar way to the standards by adding 100 µL of protein instead of the BSA standard and the concentrations of proteins were determined from the calibration plot.

2.6.3.2.2 UV absorbance

Enzyme concentrations were also determined by UV absorbance at 280 nm on a NanoDrop 1000 (Thermo Scientific) using an extinction coefficient estimated by the ProtParam program on the ExPASy Proteomic Server [46].

2.6.4 Enzyme purification

The PaS mutants and WT-PaS expressed in *E. coli* were purified for enzyme kinetics and characterization using Ni-NTA His-trap column and further purified by size-exclusion chromatography when required.

2.6.4.1 Affinity purification by Ni-NTA column

Large-scale purification:

The PaS enzyme purification was performed as described in Stevenson *et al.* (2015) with slight changes. Briefly, a starter culture was prepared by inoculating a single colony into 5 mL of Terrific Broth with 100 mg L^{-1} ampicillin (TBA) and was incubated overnight at 37 °C, 180 rpm (VWR IKA KS 4000 i control incubating shaker). The overnight starter culture was inoculated into 500 mL of TBA with 0.1 mM IPTG medium in a 2 L conical flask. The flask was incubated for ~30 hours at 30 °C or 48-50 hours at room temperature, shaking at 200 rpm. The cells were harvested by centrifuging at 8000*g*, for 30 minutes at 4 °C. The supernatant was removed and the pellet was resuspended in buffer A (35 mL of Buffer A for 5 g of wet cell pellet). This cell suspension was lysed by 40 mg mL^{-1} lysozyme (50 µL per 25 mL cell suspension) for 15 minutes at room temperature and was sonicated at 50% power and 50% pulse for 6-10 minutes by a Thermos Fisher omni sonic ruptor 400. Ammonium sulfate (0.2 g mL^{-1}) was dissolved in the cell lysate, to precipitate unwanted protein, before centrifuging at 30000*g* for 30 minutes at 4 °C. The supernatant was recovered and additional ammonium sulfate was added for a final concentration of 0.4 g mL^{-1} before overnight incubation on ice to precipitate PaS. The sample was then centrifuged at 30000*g* for 30 minutes at 4 °C. The supernatant was discarded and the pellet was resuspended in 20 mL of buffer A. This protein solution was loaded onto a 5 mL Ni-NTA His-trap column (GE Healthcare) using a peristaltic pump (8 mL min^{-1}) and

was eluted by an increasing gradient of buffer B with a flow rate of 1 mL min^{-1} (Figure 2.40A). The collected fractions were tested for sulfatase activity using *para*-nitrophenylsulfate (PNPS) and the purity of these fractions was evaluated by SDS-PAGE (Figure 2.40B). The fractions with > 90 % purity were pooled and extensively dialyzed against 50 mM Tris-HCl at pH 7.5 and stored at 4 °C.

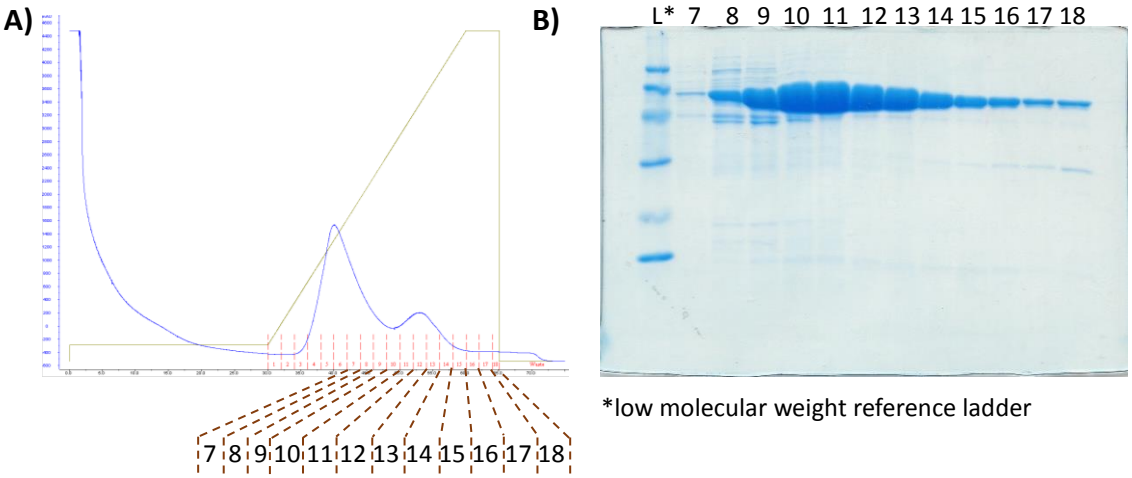


Figure 2.40: Purification of WT-PaS by Ni NTA His-trap column. A) Elution profile (blue) with the increasing buffer B concentration (green). The second peak corresponds to sulfatase activity while the first peak corresponds to non-specific protein. B) Purity determined on SDS-PAGE

Small-scale purification:

Small-scale purification was performed for the enzyme characterization purposes. The overnight starter culture (0.5 mL) was inoculated to 50 mL of TB in a 200 mL culture flask. The same steps in the large-scale purification for expression, harvesting and cell lysis were followed. The lysed cell suspension was centrifuged for 10 minutes at 4000*g*. Next, the supernatant was loaded into 1 mL Ni-NTA His-trap column (GE Healthcare) and was washed with 10 mL of buffer A (see Appendix) containing 5% imidazole. The PaS enzyme was then eluted with 10 mL of buffer B (see Appendix) using a peristaltic pump where the first 0.5 mL was discarded. The eluted protein was mixed well with 5 g of ground ammonium sulfate and left overnight at 4 °C for precipitation. This was centrifuged for 15 minutes at 30000*g* and the supernatant was discarded. The protein pellet was resuspended in 0.5-1 mL of buffer A and was transferred to a 1.5 mL Eppendorf tube and centrifuged at 15000*g* for 15 minutes to precipitate any residual matter. The supernatant was then loaded to a 5 mL Hi-trap desalting column (GE

Healthcare) that was pre-conditioned with 25 mL buffer A. The protein was eluted with buffer A using a syringe (to avoid sample tailing) where the first 1.5 mL was discarded and 1.5-3.5 mL fraction was collected and was stored at 4 °C until further use.

2.6.4.2 Gel purification

The fractions containing sulfatase activity from the large scale purification of Ni His-trap column were further purified by Superdex 200 gel filtration column. The pooled fractions were concentrated to < 2 mL (4000*g*, 4 °C) using an Amicon ultra-centrifugal filter (Milipore, 30 kDa). The concentrated protein was injected to the AKTA protein purification system (GE healthcare) and was eluted in buffer A with a flow rate of 1 ml min^{-1} (Figure 2.41A). The purity of the fractions collected were assessed by SDS PAGE (Figure 2.43B). This protocol resulted in > 30 mg of protein with crystallization grade purity.

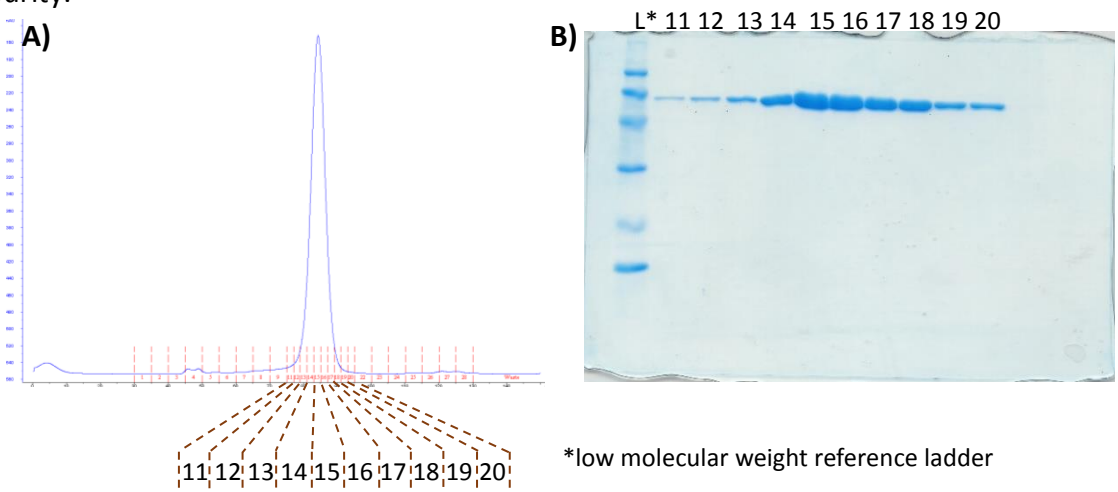


Figure 2.41: Purification of His-trap purified WT-PaS by size exclusion chromatography. A) Elution profile (blue) in buffer A. B) Purity determined on SDS-PAGE.

2.6.5 Library generation

2.6.5.1 SSM libraries (libraries 1-5)

A mega-primer PCR method reported by Sanchis *et al.* [26] was used to create libraries for the single site saturation mutagenesis where a mutagenic primer and a flanking primer were used to generate a product of 300-1100 base pairs with the required mutation in the first PCR. This amplified sequence was then used as a mega-

primer in the second PCR to anneal to the template at 60 °C with a second flanking primer in the vector, to generate the full length gene which was extracted by a gel purification. If the yield was low, then the gene was amplified in a third PCR with first and second flanking primers followed by PCR purification. This was double digested by NdeI, EcoRI restriction enzymes (see 2.6.1.4) and was ligated to NdeI, EcoRI, CIP alkyl phosphatase treated pJExpress404 ampicillin resistant vector (see 2.6.1.5).

2.6.5.2 Generating the combined mutants (PV-PaS and TV-PaS)

The P-PaS, T-PaS and V-PaS mutants were double digested by PstI and EcoRI restriction enzymes. The vector band (~4.5 kb) containing 155P/T mutation and the insert band (~1 kb) containing 330V mutation were isolated by electrophoresis (see 2.6.1.3) in an 0.8% agarose gel, purified (see 2.6.1.6) and ligated (see 2.6.1.5) to generate the double mutants.

2.6.5.3 MSM 1 library (library 6)

An assembly method reported by Gibson *et al.* [47] was used to create the MSM 1 library. Gibson assembly master mix purchased from New England Biolabs was used to anneal two fragments of overlapping DNA. The insert fragment containing the mutations and the vector fragment were generated in two separate PCRs and they were annealed at 50 °C for 15 minutes and 2 µL of this product was transformed into 50 µL electrocompetent *E. coli* DH5α (see 2.6.1.8).

2.6.5.4 MSM 2 and MSM 3 libraries (libraries 7 and 8)

Insert and vector DNA was spliced using an alternative method developed by Ruhu Qi at the Research School of Chemistry, Australian National University. Following vector amplification and purification 200 ng was added to a 20 µL reaction with an engineered E2 exonuclease with the provided buffer (engineered by Dr. Ruhu Qi, Astralian National University, not yet commercialized). DpnI was also included in the reaction to eliminate any parental DNA carried over from the PCR. The reaction mixture was incubated at 37 °C (Qualtex solidstat incubators) for 1 hr, followed by enzyme inactivation at 75 °C for 20 min. This resulted in vector DNA with about 15 bases of 5’ overhang. The insert

consisted of synthetic double-stranded oligonucleotides with 5' overhangs complementary to those on the vector. The oligonucleotides were based on a short sequence from the PaS gene and included degenerate codons at mutation points. The coding and complementary oligonucleotides were annealed by heating a solution of 50 nM of each at 72 °C for 5 minutes and then cooling to room temperature. The purified vector DNA was mixed with the synthetic insert DNA in a 1:3 molar ratio at 65 °C for 10 min before cooling to room temperature. The complementary 15 base-pair overhangs were sufficient to enable transformation of *E. coli* (see 2.6.1.8).

2.6.5.5 Shuffled library (library 9)

A similar protocol as described in section 2.6.5.1 was followed to generate the shuffling library. WT-PaS and PV-PaS were used as the templates for each half of the library to achieve K/V at 330 position in the library. A flanking primer with a tail end (see Tables A1 and A2) was used to avoid contamination from the starting template.

2.6.6 Liquid chromatography mass spectrometry (LC-MS)

The separation and detection of the steroids and their sulfate conjugates were performed on an Agilent 1260 Infinity UPLC system coupled to an Agilent 6120 quadrupole mass spectrometer employing atmospheric pressure electrospray ionization (AP-ESI). An isocratic solvent system of 65% methanol in water with 10 mM ammonium acetate was used with an Agilent Poroshell 120 C18 column (30 mm with 5 mm guard, 2.7 µm particles). Agilent 1290 infinity LC injector HTS sampler was used for screening experiments and the UPLC auto-sampler for kinetic experiments. The steroid sulfates were monitored in the negative mode as the proton loss species, $[M-H]^-$ and the free steroids were monitored in the positive mode as the proton adducts, $[M+H]^+$. Capillary voltages and fragmentation voltages were adjusted according to published settings [38].

2.6.7 Sulfatase activity screening

The screens were designed to discover variants with beneficial activity for testosterone sulfate (TS) hydrolysis. Unless otherwise stated, the buffers used for

screening and assays are as follows. 50 mM Tris-acetate was used for buffering at pH 9, 50 mM Tris-HCl for buffering at pH 7.5 and a triple buffer (TMA buffer) consisting of 50 mM Tris, 25 mM MES and 25 mM acetic acid for buffering at pH 7. The pH of buffers was measured using pH 700 by Eutech Instruments. The buffers typically contained 1 µM of nandrolone (MSM 2, 3 and shuffled libraries) as the internal standard for quantification.

2.6.7.1 Primary screen

A 96-well plate based assay was performed as the primary screen. Single colonies from a SSM library were inoculated into separate wells containing 100 µL of LBA with 0.1 mM IPTG in a round bottom 96-well plate. The sealed plates were incubated overnight at 37 °C shaking at 180 rpm. Then 40 µL of the culture from each well was transferred into a new flat bottom plate and the culture medium was removed by centrifuging (3700 rpm, 15 minutes) with 160 µL milliQ water and discarding the supernatant by inverting the plate with a gentle tap. The cell pellets were resuspended and lysed by 10 µL of 10 % BPER II (Thermo Fisher Scientific) for 10 minutes shaking at 400 rpm. Then 40 µL of the reaction mixture containing 100 µM or 20 µM TS (final concentration) buffered at pH 9 or 7 (Table A3 see Appendix), was added to the wells, and the sealed plates were incubated at 37 °C (Qualtex solidstat incubators) for 4 hours. The reactions were then quenched with 50 µL methanol and the plates were sealed with sealing film (Axygen PlateMax ®). They were mixed and centrifuged at 3700 rpm for 15 minutes using Eppendorf 5804 plate centrifuge. The supernatant was then injected into the LCMS. Data were evaluated by considering the T/TS (SSM libraries and MSM 1) or T/N (MSM 2, 3 and shuffled libraries) peak area.

2.6.7.2 Secondary and tertiary screen

The beneficial mutants from the primary or secondary screens were picked from the 96-well culture plates and streaked on LBA-agar plates before incubating at 37 °C, overnight. Isolated clones were then used for secondary or tertiary screening.

The culture growth and reactions in the SSM libraries were performed in 500 µL of LBA with 0.1 mM IPTG in 1.5 mL Eppendorf tubes. The sealed tubes were incubated overnight at 37 °C shaking at 180 rpm. Then 100 µL of the culture volume was transferred to a new 1.5 mL tube and the culture medium was removed by centrifuging

(16000*g*, 5 minutes) with 400 µL milliQ water and discarding the supernatant. The cell pellet was resuspended in 10 µL milliQ water and lysed with 10 µL of 20 % BPER II (Thermo Fisher Scientific) for 10 minutes. Then 180 µL containing 100 µM TS (final concentration) buffered at pH 9, was added to the wells, and the sealed tubes were incubated at 37 °C (Qualtex solidstat incubators) for 4 hours. The reactions were quenched with 200 µL methanol, mixed, centrifuged (16000*g*, 5 minutes) and 350 µL of the supernatant was transferred to HPLC vials for LC-MS analysis. The culture growth, reaction and LC-MS analysis of the MSM libraries were performed in 96-well plates by the same protocol as described in section 2.6.7.1. These screens for SSM and MSM libraries were executed as independent biological triplicates.

2.6.7.3 Comparison of TS and PNPS hydrolysis

The TS and PNPS hydrolysis of a small library of randomly picked ten colonies from SSM 1 and SSM 4 were tested in singlicate. A wild type colony was used in both assays as a control. Ten randomly picked colonies from each library were expressed overnight at 37 °C in 500 µL LBA with 0.1 mM IPTG in 1.5 mL Eppendorf tubes. The PNPS assay was performed by lysing 10 µL of culture with 10 µL of 20 % BPER II (Thermo Fisher Scientific) for 5 minutes followed by adding 180 µL of 0.1 M PNPS (final) and measuring the initial rates at 405 nm. The TS assay was performed as described in section 2.6.7.1.

2.6.8 Evaluation and characterization of PaS variants

The concentrations of each enzyme preparation were tested by Bradford assay with bovine serum albumin as standards (see 2.6.3.2.1).

2.6.8.1 Enzyme kinetics

Enzyme activity as a function of substrate concentration (2.5, 5, 10, 20, 40, 80, 160, 320 µM) was determined for purified proteins of WT PaS and the mutants for TS hydrolysis using LCMS. The concentration series of substrate was prepared by serial dilution of 320 µM TS in 50 mM Tris acetate at pH 9. The enzymes were prepared in the same buffer with a final concentration of 5 mg/L BSA and a final enzyme concentration of 0.4 µM for all enzymes except PVIV-, VVIV- and LEF-PaS that were at 0.08 µM. The

kinetics were monitored at 37 °C in triplicate. The substrates and the enzyme solution were centrifuged at 16000*g* for 5 minutes. Then, 360 µL of the substrate at each concentration was pipetted into separate HPLC vials and 400 µL of the enzyme solution was pipetted into a 1.5 mL tube. All were pre-incubated at 37 °C for 10 minutes in the LC-MS auto-sampler with the heater on. Next, 40 µL of the enzyme was pipetted to each vial, mixed, and the reaction rate was determined by measuring steroid concentration (T) at four time points, over which the rate was constant. Freshly prepared T standards (0.5, 1, 2, 5 µM T in 50 mM Tris acetate, pH 9) were used for quantification. The Michaelis Menten constant (K_M) and maximum velocity for the reaction (V_{max}) were calculated by fitting the data into the Michaelis Menten model using KaleidaGraph software.

2.6.8.2 Effect of pH on PaS activity

The optimum pH for the purified enzymes was assayed using WT-PaS and PVFV-PaS with two different substrates: TS at 20 µM or 300 µM or PNPS at 20 µM or 50 µM. all substrates were prepared in TMA buffer containing 50 mM Tris, 25 mM MES and 25 mM acetic acid [48, 49]. The pH of the reactions ranged from 4 to 9 for TS hydrolysis or pH 6 to 9 for PNPS hydrolysis. The TS hydrolysis was performed by mixing 10 µL of enzyme (0.4 µM final concentration) with 90 µL substrate and incubating at 37 °C for 1 hour. These reactions were quenched with 100 µL of MeOH (final concentration), centrifuged (16000*g*, 5 minutes), 150 µL of the supernatant was transferred to wells in 96-well plate and assayed by LC-MS (see 2.6.6). The PNPS hydrolysis was performed by pre-incubating the enzyme and substrate separately at 37 °C for 5 minutes, and mixing 20 µL of enzyme (10 nM final concentration) with 180 µL of substrate. After 1 minute or 3 minutes of reaction for WT-PaS and PVFV-PaS respectively, 100 µL of 1 M KOH was added to quench enzyme activity and ensure complete ionization of the *para*-nitrophenolate ion. The PNP was assayed at 405 nm using a SpectraMax M2 plate reader. The tests were performed in triplicate.

2.6.8.2.1 PaS hydrolysis in a range of buffers

The hydrolysis of WT-PaS and PVFV-PaS was tested in a range of buffers at two pHs representing each buffer. The reactions were performed at 37 °C or 50 °C. However, the same buffer preparation at 50 mM concentration was used at both temperatures

where the pH was adjusted only at 50 °C by HCl or $NaOH_{(aq)}$. The reactions were performed in 1.5 mL tubes in duplicates. First, 180 µL of the prepared buffers was pipetted to the 1.5 mL tubes. Then 10 µL of a mixture of 400 µM TS and 20 µM N was added followed by the addition of 10 µL of WT-PaS or PVFV-PaS at 0.4 µM final concentration. The reactions were incubated at 37 °C or 50 °C (Qualtex solidstat incubators) for half an hour followed by quenching the reactions with 400 µL of 75 % MeOH. The quenched reactions were centrifuged at 16000*g* for 5 minutes and transferred to HPLC vials for LC-MS analysis.

2.6.8.3 Thermostability of PaS mutants

PaS solutions were subjected to heat treatment in covered PCR strips using a PCR thermocycler (Bio-Rad C1000^{TM} or Applied Biosystems Veriti). For analysis of the apparent melt temperature (T_m^{app}), PaS aliquots were incubated at defined temperatures for 5 min. The half-life ($T_{1/2}$) was measured at 50 °C for defined times. Samples were kept on ice before and for 10 minutes after heat treatment, and untreated controls were kept on ice throughout. The PaS concentration for heat treatment was defined so that a ten-fold dilution would allow an accurate measurement of activity: 800 nM for PVFV-PaS and 200 nM for all others. The activity of controls or samples (diluted 10-fold) was assayed at 25 °C with 100 µM PNPS in 50 mM Tris acetate buffer at pH 9, or 50 mM Tris HCl buffer at pH 7.5 (enzymes diluted 10-fold in the reaction). The initial rate of PNP formation was measured at 405 nm over a period of 5 min using a BioStrategy SpectraMax M2 plate reader. Residual activity of treated samples was determined relative to controls and analysed by non-linear regression with KaleidaGraph 4.5. Observations for T_m^{app} were fitted to a modified Hill equation [49] whereas those for $T_{1/2}$ were to an exponential decay model.

2.6.8.4 Testing the substrate scope

Initial rates for hydrolysis of TS, EAS, ECS and ETS were tested with the purified proteins in triplicates. The tests were carried out in sealed 96-well plates. Two reaction conditions at pH 7.5 and 37 °C were used: i) α-configured steroid sulfates at 100 µM with 4 µM enzyme incubated for 4 hours; or ii) β-configured steroid sulfates at 20 µM with 0.08 µM enzyme incubated for 1 hour. Reaction were started by mixing 10 µL of enzyme

with 90 µL of substrate. Reactions were quenched with 100 µL of MeOH, centrifuged, and the amount of free steroid produced was measured by LCMS using the peak area ratio of the free steroid to internal standard (N). The results were quantified against the standards of free steroids prepared in a similar manner.

2.6.8.5 Levels of post-translational modification of PaS mutants

The levels of post-translational modification of cysteine to formylglycine was tested by the Ellman's assay as described by Reiner *et al.* [50]. All reagents and calibration standards required were freshly prepared. A 2 mM solution of Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid) or DTNB) was prepared in 50 mM sodium acetate. The reactions were conducted in 0.1 M Tris acetate at pH 8 and carried out in the wells of a 96-well flat bottom plate with a 200 µL total volume.

2.6.8.5.1 Determination of the molar extinction coefficient of TNB^{2-}

A standard series was prepared by using the reduced form of *l*-glutathione (GSH). The GSH stock solution for the standard series was prepared by diluting a 0.1 M $GSH_{(aq)}$ solution 200 times to make a 500 µM in water. A standard series of 5, 7.5, 10, 15, 20, 30, 40 µM (calculated as the final concentrations in the reactions) GSH was prepared by diluting the 500 µM stock solution to produce 30 µM and 40 µM standards followed by a serial dilution. The reactions for the standard series were performed by adding guanidium chloride (8 M, 70 µL), Tris acetate buffer (1 M, 20 µL), GSH standard (100 µL) and $DTNB^{2-}$ (2 mM, 10 µL). The GSH blank was prepared by adding 100 µL of water instead of a GSH standard. The absorbance at 412 nm for the release of TNB^{2-} was measured for 15 minutes at room temperature. The maximum absorbance was plotted against the GSH concentration to obtain the calibration plot and the molar extinction coefficient was determined by subtracting the blank and applying the parameters to Beer-Lambert law; $A=\varepsilon*c*l$ where, A = absorbance, ε = molar extinction coefficient, c = concentration of the standards, l = path length. The reactions were performed in triplicate.

2.6.8.5.2 Determination of remaining cysteine of each PaS mutant and WT-PaS

The enzymes were diluted to 100 µM concentration in water and the remaining amount of cysteine (thiol concentration) was tested according to the same protocol as for the standards. The reactions included the enzyme (100 µL), Tris acetate buffer (1 M, 20 µL), guanidium chloride (8 M, 70 µL) and $DTNB^{2-}$ (2 mM, 10 µL). A protein blank was prepared similar to the GSH blank by adding 100 µL of water instead of protein. The thiol concentration was determined by applying the maximum absorbance observed to the equation of Beer-Lambert law. Each absorbance measurement was subtracted by the respective blank samples. The absorbance was measured in triplicate for each enzyme.

2.6.9 Investigation of E74 residue

2.6.9.1 Site directed mutagenesis of E74 residue

Site directed mutagenesis was performed to create thee variants of PaS; E74Q, E74D and E74A. This included a single PCR amplification of the gene with a forward mutagenic primer (P27-f, P28-f and P29-f respectively for 74Q, 74D and 74A) and a reverse primer (P24-r) located at the end of the gene. The PCR product was purified from a gel extraction followed by double digestion with NcoI and EcoRI restriction enzymes. The purified digested product was ligated to NcoI, EcoRI, CIP alkyl phosphatase treated pJExpress404 ampicillin resistant vector (see 2.6.1.5). The ligated product was purified (see 2.6.1.6) followed by bacterial transformation (see 2.6.1.8).

2.6.9.2 PNPS kinetics

2.6.9.2.1 Determination of the molar extinction coefficient of para-nitrophenolate (PNP^-)

The standard series was prepared by serial dilution of an 80 µM PNP standard prepared from a freshly prepared 50 mM stock PNP solution in 50 mM Tris acetate at pH 9, 37 °C. Then 200 µL of each standard was pipetted into wells of a 96-well flat bottom plate and the end point absorbance at 405 nm was read using a BioStrategy SpectraMax M2 plate reader. The standard curve was plotted with absorbance against PNP concentration and the molar extinction coefficient was calculated by the dividing the slope of the curve by the path length that was calculated by substituting the parameters

of the Beer-Lambert law.

2.6.9.2.2 Kinetics

The enzyme kinetics with PNPS was performed for the PaS E74 mutants and WT-PaS. Enzyme activity as a function of substrate concentration (2.5, 5, 10, 20, 40, 80, 160, 320 µM) was determined for purified proteins of WT PaS and the E74 mutants using BioStrategy SpectraMax M2 plate reader. The concentration series of substrate was prepared by serial dilution of 320 µM PNPS in 50 mM Tris acetate at pH 9. The enzymes too were prepared in the same buffer with a final enzyme concentration of 2 nM for WT-PaS and 200 nM for the E74 variants. The kinetics were monitored at 37 °C in triplicate. The substrate was pipetted to the wells of a 96-well flat bottom plate with 180 µL from each substrate concentration with the lids on and incubated in the plate compartment of the plate reader with the incubator turned on at 37 °C for 5 minutes. The enzymes (1 mL) at 10 times greater concentration were incubated in a 37 °C water bath for 5 minutes in sealed 1.5 mL Eppendorf tubes. After incubation, 20 µL of the enzyme was pipetted into the wells of the substrate by multi-channel pipetting and the plate was shaken for 5 seconds by the shaker option on the plate reader prior to reading absorbance. The initial rates for PNP^- release was measured from 50-300 seconds at 37 °C and the rate with related to concentration terms was calculated by applying the absorbance measurement to Beer-Lambert law by substituting the pre-determined extinction coefficient. The Michaelis constant (K_M) and maximum velocity for the reaction (V_{max}) were calculated by fitting the data into the Michaelis Menten model using KaleidaGraph software.

2.6.10 Optimizing PaS expression in *E. coli*

Different growth conditions were tested to optimize the expression of PaS. A single colony was inoculated in 0.5 mL of LBA or TBA and incubated overnight at 37 °C. The overnight starter culture was used to inoculate 50 mL of LBA or TBA with or without IPTG (100 µM) for induction in 125 mL culture flasks. The cultures were incubated at 30 °C or 37 °C shaking at 180 rpm (VWR IKA KS 4000 i control incubating shaker). The cultures without added IPTG were incubated until the OD reached 0.8 at 595 nm

followed by addition of 100 mM IPTG to give a final concentration of 100 µM. The PNPS activity for all cultures was tested every hour between 24 and 32 hours by pipetting out 10 µL of culture into a wells of a 96-well flat bottom plate followed by lysis with 10 µL of 20 % BPER II (Thermo Fisher Scientific) for 5 minutes and measuring the initial rate of PNPS hydrolysis.

1. Culture medium – LBA (Lauria Bertani- 100 mg L^{-1} ampicillin) and TBA (Terrific Broth - 100 mg L^{-1} ampicillin) were used for comparison
2. Expression temperature – was varied between 30 °C and 37 °C under shaking conditions
3. Induction time – was varied on which time IPTG induction was carried out
    a. Early induction – the overnight starter culture was added to the medium with IPTG.
    b. Late induction – the overnight starter culture was added to the medium without IPTG and then it was later induced with IPTG when the OD reached 0.8 at 595 nm.

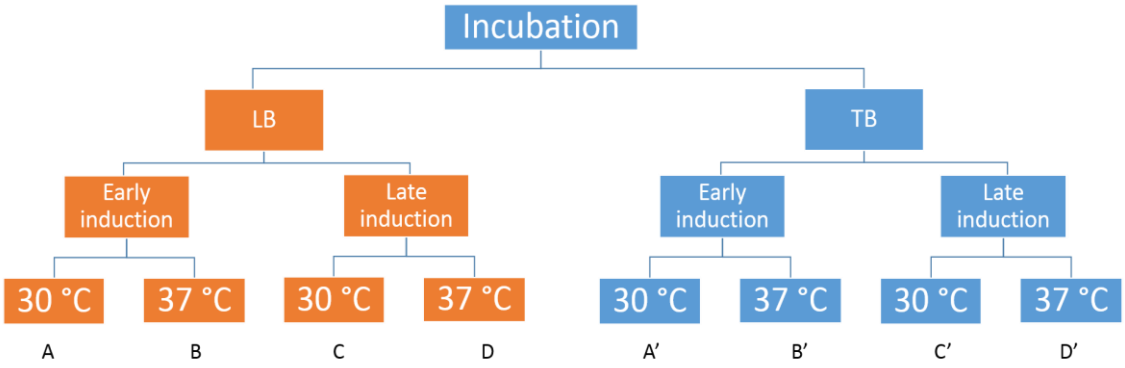


Figure 2.42: Different incubation conditions tested to optimize PaS expression

The PNPS activity over an incubation time of 24-32 hours was measured and the maximum activity was plotted as in Figure 2.41. These results suggest that TBA medium, containing more nutrients than LBA, resulted in ~2 times better expression than in LBA. Additionally, longer incubation at lower temperatures yielded better results. Expressing the protein at 30 °C by inoculating the starter culture straight to IPTG induction in TBA resulted in the best expression for PaS. Therefore, a similar condition was employed for the expression of PaS for purification purposes.

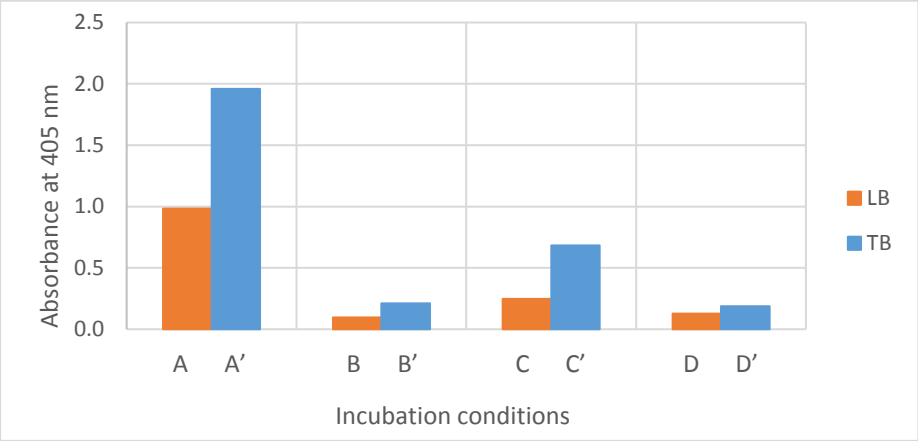


Figure 2.43: PNPS hydrolysis activity in crude cell lysate for cultures under different conditions over an incubation time of 24-32 hours

2.6.11 Minimal media selection

2.6.11.1 Preparation of MMTS plates

The MMTS was prepared according to the recipe developed in the McLeod laboratory adapted from Sambrook *et al.* [51] that uses the trace element recipe described by EMBL (obtained online in 2012). The MMTS medium was poured into petri dishes to achieve similar depth among the dishes in a sterile lamina-flow cabinet. The medium was allowed to solidify by drying for 10 minutes and were used immediately.

2.6.11.2 Preparation of library cultures

An *E. coli* transformation and recovery was performed that was followed by incubation as described in 2.6.1.10. Next, the cells were centrifuged at 6000*g* for 1 minute and 400 µL of the supernatant was removed. The cell pellet was resuspended in the remaining media and was plated on a LBA-agar plate and was incubated overnight at 37 °C (Qualtex solidstat incubators). The cells were recovered by agitating about 10 sterile glass beads and 2 mL of sterile water so that the beads displace the colonies over the entire surface. Then 1 mL of the resuspended cells were transferred to a sterile 1.5 mL tube that was centrifuged at 6000*g* for 1 minute. The supernatant was removed and the cells were resuspended in 1 mL of no sulfur storage media and 100 µL aliquots were prepared in 1.5 mL tubes and stored at -80 °C. One aliquot was thawed and diluted 100-

1000 times in water and 50 µL was plated on LBA-agar to estimate the dilution needed to obtain approximately 300 colonies per plate.

2.6.11.3 Selection on MMTS

An aliquot was thawed and 10 g L^{-1} thiamine was added to give a final concentration of 0.4 mg mL^{-1} in the diluted culture (DH5α is deficient in thiamine biosynthesis). Then 50 µL was plated on MMTS plates and the culture was uniformly distributed by agitating 10-15 sterile glass beads until the culture was absorbed. The glass beads were removed by collecting the beads to the lid by inversion of the plate. The plates were incubated at 30 °C until the colonies were visible and were transferred to 37 °C for faster growth.

2.7 References

[1] Christakopoulos, P., and Topakas, E. (2012) Editorial Note: Advances in enzymology and enzyme engineering, *Comput. Struct. Biotechnol. Journal 2*, *DOI: 10.5936/csbj.201209001*.

[2] Landon, M., Evans, W. H., and Smith, E. L. (1968) Subtilisin Carlsberg III. Isolation and amino acid composition of chymotryptic peptides, *J. Biol. Chem. 243*, 2165-2171.

[3] Aiba, H., Nishiya, Y., Azuma, M., Yokooji, Y., Atomi, H., and Imanaka, T. (2015) Characterization of a thermostable glucose dehydrogenase with strict substrate specificity from a hyperthermophilic archaeon *Thermoproteus* sp. GDH-1, *Biosci. Biotech. Bioch. 79*, 1094-1102.

[4] Fujimura, S., Hirai, K., and Shibata, Y. (2006) Dipeptidyl peptidase with strict substrate specificity of an anaerobic periodontopathogen *Porphyromonas gingivalis*, *Fems Microbiol. Lett. 209*, 127-131.

[5] Barrozo, A., Duarte, F., Bauer, P., Carvalho, A. T., and Kamerlin, S. C. (2015) Cooperative electrostatic interactions drive functional evolution in the alkaline phosphatase superfamily, *J. Am. Chem. Soc. 137*, 9061-9076.

[6] Lutz, S. (2010) Beyond directed evolution-semi-rational protein engineering and design, *Curr. Opin. Biotech. 21*, 734-743.

[7] Farinas, E. T., Bulter, T., and Arnold, F. H. (2001) Directed enzyme evolution, *Curr. Opin. Biotech. 12*, 545-551.

[8] Molina-Espeja, P., Vina-Gonzalez, J., Gomez-Fernandez, B. J., Martin-Diaz, J., Garcia-Ruiz, E., and Alcalde, M. (2016) Beyond the outer limits of nature by directed evolution, *Biotechnol. Adv. 34*, 754-767.

[9] Renata, H., Wang, Z. J., and Arnold, F. H. (2015) Expanding the enzyme universe: accessing non-natural reactions by mechanism-guided directed evolution, *Angew. Chem. Int. Edit. 54*, 3351-3367.

[10] Packer, M. S., and Liu, D. R. (2015) Methods for the directed evolution of proteins, *Nat. Rev. Genet. 16*, 379-394.

[11] Cobb, R. E., Sun, N., and Zhao, H. M. (2013) Directed evolution as a powerful synthetic biology tool, *Methods 60*, 81-90.
[12] Reetz, M. T., Kahakeaw, D., and Lohmer, R. (2008) Addressing the numbers problem in directed evolution, *Chembiochem 9*, 1797-1804.
[13] Hibbert, E. G., and Dalby, P. A. (2005) Directed evolution strategies for improved enzymatic performance, *Microb. Cell Fact. 4,* 29-34.
[14] Zhang, Y. H. P., Himmel, M. E., and Mielenz, J. R. (2006) Outlook for cellulase improvement: screening and selection strategies, *Biotechnol. Adv. 24*, 452-481.
[15] Chica, R. A., Doucet, N., and Pelletier, J. N. (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design, *Curr. Opin. Biotech. 16*, 378-384.
[16] Turner, N. J. (2009) Directed evolution drives the next generation of biocatalysts, *Nat. Chem. Biol. 5*, 568-574.
[17] Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D. X., and Vogel, A. (2005) Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test, *Angew. Chem. Int. Edit. 44*, 4192-4196.
[18] Li, K. (2013) MSc. thesis, The Australian National University.
[19] Taylor, S. V., Kast, P., and Hilvert, D. (2001) Investigating and engineering enzymes by genetic selection, *Angew. Chem. Int. Edit. 40*, 3310-3335.
[20] Patrick, W. M., Firth, A. E., and Blackburn, J. M. (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries, *Protein. Eng. 16*, 451-457.
[21] Reetz, M. T., and Wu, S. (2008) Greatly reduced amino acid alphabets in directed evolution: making the right choice for saturation mutagenesis at homologous enzyme positions, *ChemComm 43*, 5499-5501.
[22] Reetz, M. T., Kahakeaw, D., and Sanchis, J. (2009) Shedding light on the efficacy of laboratory evolution based on iterative saturation mutagenesis, *Mol. BioSyst. 5*, 155-122.
[23] Walter, K. U., Vamvaca, K., and Hilvert, D. (2005) An active enzyme constructed from a 9-amino acid alphabet, *J. Biol. Chem. 280*, 37742–37746.
[24] Boltes, I., Czapinska, H., Kahnert, A., von Bulow, R., Dierks, T., Schmidt, B., von Figura, K., Kertesz, M. A., and Uson, I. (2001) 1.3 angstrom structure of arylsulfatase from *Pseudomonas aeruginosa* establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family, *Structure 9*, 483-491.
[25] Zheng, L., Baumann, U., and Reymond, J. L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol, *Nucleic Acids Res. 32*, e115.
[26] Sanchis, J., Fernandez, L., Carballeira, J., Drone, J., Gumulya, Y., Hobenreich, H., Kahakeaw, D., Kille, S., Lohmer, R., Peyralans, J., Podtetenieff, J., Prasad, S., Soni, P., Taglieber, A., Wu, S., Zilly, F., and Reetz, M. (2008) Improved PCR method for the creation of saturation mutagenesis libraries in directed evolution: application to difficult-to-amplify templates, *Appl. Microbiol. Biot. 81*, 387-397.
[27] Pavelka, A., Chovancova, E., and Damborsky, J. (2009) HotSpot wizard: a web server for identification of hot spots in protein engineering, *Nucleic Acids Res. 37*, W376-W383.
[28] Kintses, B., Hein, C., Mohamed, M. F., Fischlechner, M., Courtois, F., Leine, C., and Hollfelder, F. (2012) Picoliter cell lysate assays in microfluidic droplet compartments for directed enzyme evolution, *Chem. Biol. 19*, 1001-1009.
[29] Firth, A. E., and Patrick, W. M. (2005) Statistics of protein library construction, *Bioinformatics 21*, 3314-3315.
[30] Firth, A. E., and Patrick, W. M. (2008) GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries, *Chem. Biol. 36*, W281-W285.

[31] Welch, M., Govindarajan, S., Ness, J. E., Villalobos, A., Gurney, A., Minshull, J., and Gustafsson, C. (2009) Design parameters to control synthetic gene expression in *Escherichia coli*, *Plos One 4*.e7002.
[32] Hanson, S. R., Best, M. D., and Wong, C. H. (2004) Sulfatases: structure, mechanism, biological activity, inhibition, and synthetic utility, *Angew. Chem. Int. Edit. 43*, 5736-5763.
[33] Bojarova, P., Denehy, E., Walker, I., Loft, K., De Souza, D. P., Woo, L. W. L., Potter, B. V. L., McConville, M. J., and Williams, S. J. (2008) Direct evidence for ArO-S bond cleavage upon inactivation of *Pseudomonas aeruginosa* arylsulfamates by aryl sulfatase, *Chembiochem 9*, 613-623.
[34] Waters, M. L. (2002) Aromatic interactions in model systems, *Curr. Opin. Chem. Biol. 6*, 736-741.
[35] Acharya, P., Rajakumara, E., Sankaranarayanan, R., and Rao, N. M. (2004) Structural basis of selection and thermostability of laboratory evolved *Bacillus subtilis* lipase, *J. Mol. Biol. 341*, 1271-1281.
[36] Kumar, S., Tsai, C. J., and Nussinov, R. (2000) Factors enhancing protein thermostability, *Protein Eng. 13*, 179-191.
[37] Kumar, S., Ma, B. Y., Tsai, C. J., and Nussinov, R. (2000) Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers, *Proteins 38*, 368-383.
[38] Stevenson, B. J., Waller, C. C., Ma, P., Li, K., Cawley, A. T., Ollis, D. L., and McLeod, M. D. (2015) *Pseudomonas aeruginosa* arylsulfatase: a purified enzyme for the mild hydrolysis of steroid sulfates, *Drug Test Anal*. *7,* 903-911.
[39] You, L., and Arnold, F. H. (1996) Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide, *Protein Eng. 9*, 77-83.
[40] Riddles, P. W., Blakeley, R. L., and Zerner, B. (1979) Ellmans reagent - 5,5'-dithiobis(2-nitrobenzoic acid) - re-examination, *Anal. Biochem. 94*, 75-81.
[41] Kraut, J. (1977) Serine proteases - structure and mechanism of catalysis, *Annu. Rev. Biochem. 46*, 331-358.
[42] Waller, C. C., and McLeod, M. D. (2014) A simple method for the small scale synthesis and solid-phase extraction purification of steroid sulfates, *Steroids 92*, 74-80.
[43] IDT OligoAnalyzer 3.1, http://sg.idtdna.com/calc/analyzer.
[44] Laemmli, U. K. (1970) Cleavage of structural proteins during assembly of head of bacteriophage-T4, *Nature 227*, 680-&.
[45] Compton, S. J., and Jones, C. G. (1985) Mechanism of dye response and interference in the Bradford protein assay, *Anal. Biochem. 151*, 369-374.
[46] ExPASy ProtParam tool, http://au.expasy.org/tools/protparam.html.
[47] Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases, *Nat. Methods 6*, 343-U341.
[48] Liu, M., Sergienko, E. A., Guo, F. S., Wang, J., Tittmann, K., Hubner, G., Furey, W., and Jordan, F. (2001) Catalytic acid-base groups in yeast pyruvate decarboxylase I. Site-directed mutagenesis and steady-state kinetic studies on the enzyme with the D28A, H114F, H115F, and E477Q substitutions, *Biochemistry-US 40*, 7355-7368.
[49] Stevenson, B. J., Liu, J. W., and Ollis, D. L. (2008) Directed evolution of yeast pyruvate decarboxylase I for attenuated regulation and increased stability, *Biochemistry-US 47*, 3013-3025.
[50] Riener, C. K., Kada, G., and Gruber, H. J. (2002) Quick measurement of protein sulfhydryls with Ellman's reagent and with 4,4'-dithiodipyridine, *Anal. Bioanal. Chem. 373*, 266-276.
[51] Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) In *Molecular Cloning: A laboratory manual* 2nd Edition ed., Cold Spring Harbor Laboratory Press.

Chapter 3: Application of PaS enzyme for urine samples

3.1 Introduction

With new developments in the analytical techniques such as mass spectrometry and chromatography, untargeted or metabolomic approaches are attracting attention in the anti-doping arena and becoming a promising strategy for the identification of new markers or metabolites. These methods have resulted in identification of new steroid sulfate metabolites [1-3] that extend detection windows for exogenous steroids or serve as markers for exogenous administration of endogenous steroids. With the growing importance of steroid sulfate markers in anti-doping screens, the demonstration that WT-PaS or its engineered variants can hydrolyse sulfate esters in real-world samples would be useful.

The monitoring of the illicit use of steroids in the anti-doping laboratories is typically conducted by testing urinary steroids [4], as urine can be acquired in large volumes. However, components within the urine matrix can interfere with enzyme activity and provide an additional barrier for the introduction of a new enzyme [5-7] aimed at improving sample processing. The PaS enzyme engineered in defined buffers will be of greater utility if it has a similar performance in a complex matrix like urine. This chapter aims to investigate the practical application of PaS enzyme in urine including an effort using an untargeted method to detect a broad range of sulfated urinary metabolites and determine whether these are PaS substrates.

The studies in this chapter include the application of enzyme hydrolysis of steroid sulfates either spiked into urine, present in urine as metabolites of endogenous steroids or exogenous administration. The enzyme hydrolysis was performed by PaS mutants produced in the course of enzyme engineering alongside with WT-PaS for comparison. This chapter consists of three studies; 3.2.1: quantification of d_3-testosterone sulfate (d_3TS) in a spiked urine sample, 3.2.2: hydrolysis of endogenous dehydroepiandrosterone sulfate (DHEAS) in human urine and 3.2.3: untargeted

screening of urinary sulfate metabolites and their enzyme hydrolysis in both humans and horses (section 3.2.3).

3.2 Results and Discussion

3.2.1 Quantification of d_3-TS hydrolysis in a spiked urine sample

In this thesis, protein engineering of PaS has been performed to improve TS hydrolysis. For this reason, the PaS variant with the highest V_{max} for TS hydrolysis, LEF-PaS, was tested in comparison with wild-type PaS. These tests were performed by using a female urine sample. The reaction conditions employed were analogous to those used for glucuronide hydrolysis with β-glucuronidase in a typical anti-doping screen. The tests were performed in duplicate and d_3-TS (Figure 3.1) was used as the substrate instead of TS to avoid any endogenous TS complicating the results. Urine was spiked with 100, 200, or 300 ng/mL of d_3-TS (61.6, 123.3, 185.0 ng mL^{-1} d_3-T equivalent respectively) and incubated overnight at pH 7.5 and 37 °C with 100 µg/ mL or 10 µg/mL of either version of PaS as described in section 3.4.2.

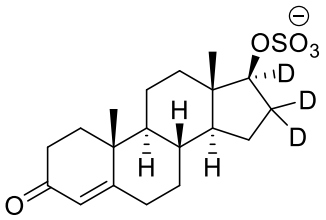


Figure 3.1: Structure of d_3-testosterone sulfate

First, the hydrolysis was performed with 100 µg/mL enzyme concentration (Figure 3.2A) and LEF-PaS showed complete conversion at all concentrations while WT conversion was lower (35-60%). Within this experiment, the hydrolysis at 100 ng/mL d_3-TS by WT-PaS is exceptional with a large error observed. Since the observed extent of hydrolysis between WT-PaS and LEF-PaS were not as dramatically discernible as expected, the enzyme concentration was reduced by one order of a magnitude allowing greater differentiation between the rates after overnight incubation (Figure 3.2B).

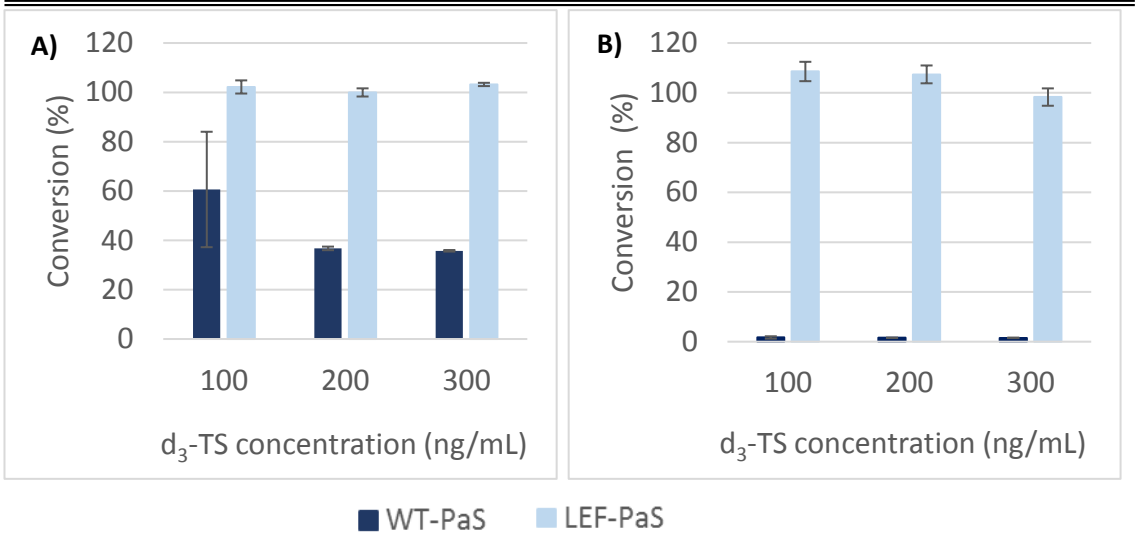


Figure 3.2: The percentage of d_3TS hydrolysed by WT-PaS and LEF-PaS at A) 100 µg/mL B) 10 µg/mL enzyme concentration at pH 7.5. The error bars represent the standard error of the duplicates.

These results show that 10 µg/mL of LEF-PaS can completely hydrolyse the d_3-TS at all substrate concentrations whereas WT-PaS conversion was estimated as < 2% at all concentrations (Figure 3.2B), that was below the lower limit of quantification (Table 3.13). The lowest enzyme concentration at which LEF-PaS can completely hydrolyse d_3-TS under these conditions was not determined and could be lower than 10 µg/mL. In addition, the LEF-PaS mutant has proven to show good hydrolytic activity compared to WT-Pas in the presence of a urine matrix despite the potential for inhibitory substances.

3.2.2 Testing endogenous DHEAS hydrolysis in human urine

Dehydroepiandrosterone (DHEA) is the most abundant circulating steroid in humans [8] and is a metabolic intermediate on the biosynthetic pathway to testosterone. It is reversibly converted to its sulfate conjugate (DHEAS) that exists in a greater concentration than the free steroid [9, 10]. The amount of urinary DHEAS hydrolysis was tested with WT-PaS in comparison with PVFV-PaS as described in section 3.4.3. The PVFV-PaS mutant was selected because it has the broadest substrate range (section 2.3.3). The DHEAS hydrolysis was performed with 100 µg/mL enzyme concentration and an overnight reaction at 37 °C at two different pHs (Figure 3.3).

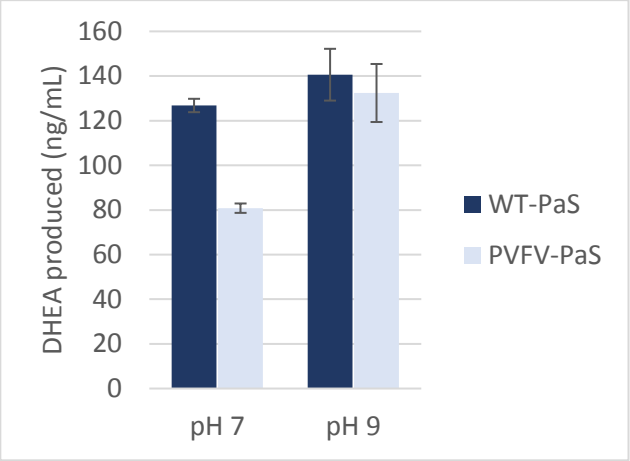


Figure 3.3: The amount of DHEA detected after enzyme hydrolysis at pH 7 and pH 9

The amount of free DHEA in a urine sample was analysed with no enzyme treatment and was estimated as less than 1 ng/mL and so considered negligible. Hence, the amount of DHEA detected from the enzyme treated samples is the DHEA produced from the hydrolysis of endogenous DHEAS. The hydrolysis at pH 9 compared to pH 7 was greater with both versions of PaS. At pH 9, both WT-PaS and PVFV-PaS showed levels of hydrolysis that were identical within error. However, the absolute values of DHEA produced by PVFV-PaS at pH 7 were masked by an additional peak with the same mass as DHEA that had a retention time difference of < 7% that eluted after DHEA resulting in a lower integrated peak area. The amount hydrolysed by WT-PaS at pH 7 is slightly lower than that at pH 9 but is identical within error. Slower hydrolysis at neutral pH could possibly reflect lower stability. The DHEA produced from hydrolysis is within the range of quantification (Table 3.13). However, quantification of the amount of DHEAS present was not carried out to determine the percentage hydrolysed. A better discrimination of the amounts hydrolysed between enzymes could have observed if a shorter reaction time and a lower enzyme concentration was used.

3.2.3 Untargeted screening of urinary metabolites

The main aim of this section of work was to more broadly study the hydrolysis patterns of WT-PaS in comparison with PVFV-PaS. To achieve this, an untargeted detection method for urinary steroid sulfates was developed and employed to samples

pre- and post-enzyme hydrolysis. In addition, another aim of the study was to provide some structural information about the steroid sulfates observed. Furthermore, the equine urine study aims to identify metabolites of testosterone propionate administration by comparison of pre- and post-administration urine samples that could lead to the identification of potential markers.

This study was performed as a part of a prize (Charles Hocart Award, ACT Mass Spectrometry Symposium, 2015) awarded to the candidate. The sample analysis on a Q-Exactive plus LC-MS/MS (see section 3.4.5.2) was conducted by Mr Alex Chen of Thermo Fisher, Scoresby, Victoria. The methods were developed through consultation between the two parties.

The Q-Exactive plus employs an orbitrap mass analyser that offers high resolution accurate mass (HRAM) performance. This can analyse a mass range of *m/z* 50-6000 and can be used for untargeted metabolomics. However, the mass range analysed in this study was limited to *m/z* 200-800. This study included the analysis on full scan mode (at 70000 resolution, FWHM) that includes continuous acquisition and has the potential of detecting any compound if it is ionisable. Also data dependent tandem mass spectrometry (ddMS/MS) acquisition was employed (at 17500 resolution, FWHM) which selects prominent precursor ions [11] (top 10 per duty cycle in this study) to be fragmented and analysed by MS/MS. Therefore, this approach detects only the analytes selected for ddMS/MS that are later identified by applying the required filters. This technique is limited by the duty cycle rate as a faster duty cycle enables more acquisitions [11]. This ddMS/MS scan (17500, FWHM) and the high-resolution full MS scan (70000, FWHM) were acquired in parallel. This study targeted the identification of steroid sulfate metabolites although a range of putative sulfate metabolites were observed that did not appear to be steroidal in nature (Figure 3.4).

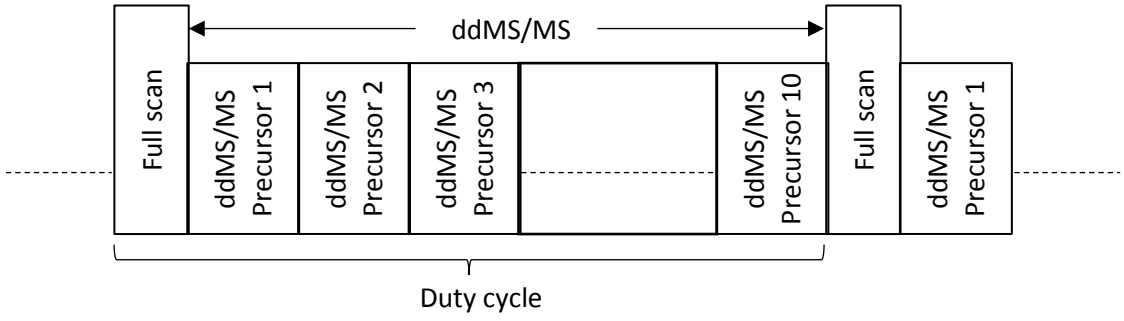


Figure 3.4: A schematic representation of a duty cycle.

The study encompassed four urine samples; (1) human – female urine (female); (2) human – male urine (male); (3) horse – 0 hr blank urine (horse 0 hr); (4) horse – 12 hr post administration of testosterone propionate (horse 12 hr). Each sample had three different treatments; i) control sample with no enzyme treatment; ii) treatment with WT-PaS; iii) treatment with PVFV-PaS. The enzyme hydrolysis was performed overnight at 37 °C, pH 7.5 in 75 mM (final concentration) Tris-HCl buffer (see section 3.4.5.1). The metabolites were extracted by solid phase extraction (SPE) with a weak anion exchange (WAX) cartridge and eluted with a mixture of ethyl acetate : methanol : diethylamine (25:25:1, by volume). All samples contained nandrolone and nandrolone sulfate as internal standards and were treated identically. Due to limited availability of the urine samples and analysis time, this study was performed as technical duplicates.

3.2.3.1 Data analysis of ddMS/MS spectra and assignment of putative structures

The data analysis included five major steps: 1) Analysis of ddMS/MS data with sulfate filters; 2) Integration of the peak areas for all sulfate metabolite precursor ions; 3) Determination of the possible molecular formulae for selected precursor ions; 4) Attempted correlation of observed metabolites with the possible structures of previously reported compounds; 5) Integration of the calculated positive mode ion assuming hydrolysis to a sulfate free metabolite.

First, filters for sulfates within a 10 ppm mass tolerance considering the loss of *m/z* = 97 or 80 corresponding to HSO_4^- and $^{\cdot}SO_3^-$ respectively [12] were applied on the ddMS/MS data. All the precursor ions corresponding to ≥ 20% of the relative abundance in the sulfate filters applied (Figure 3.5) were selected for further evaluation. As an example, from approximately 1050 data dependent acquisitions in the female control sample, 15 precursor ions were selected for data analysis.

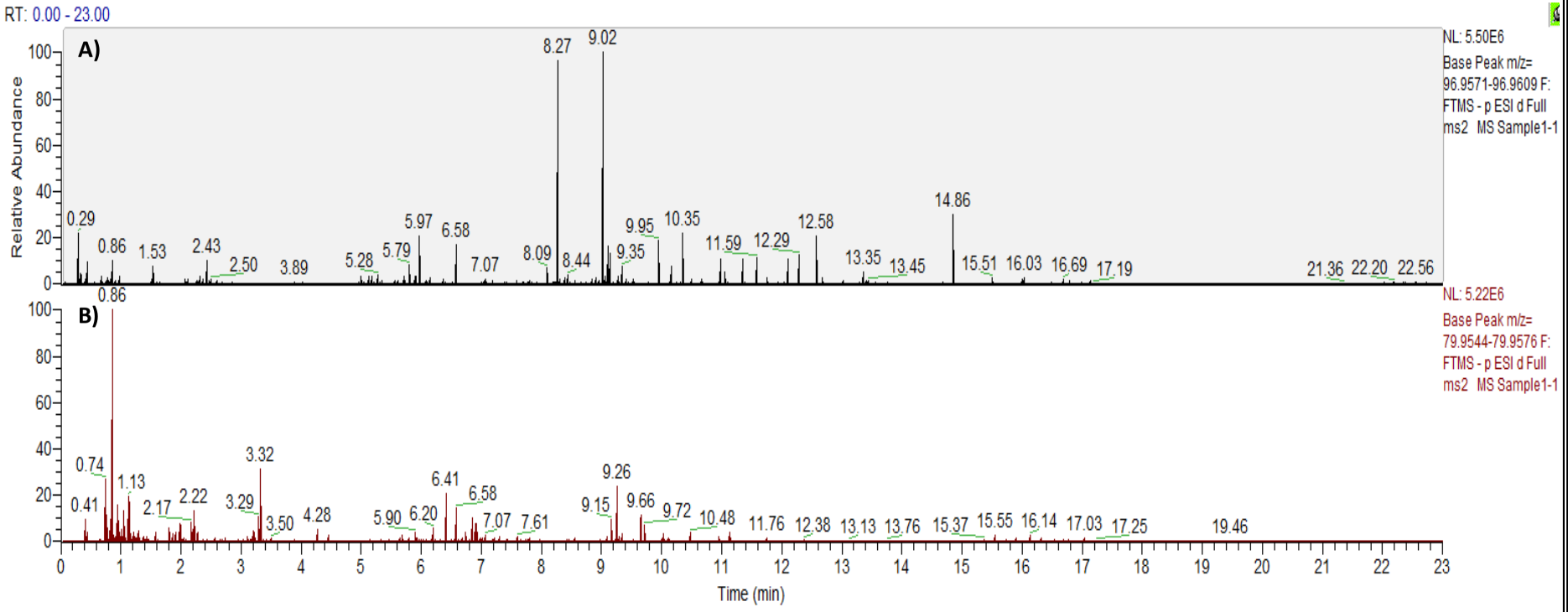


Figure 3.5: Extracted ion chromatograms for filters targeting A) HSO_4^- (*m/z* = 97) and B) $^{\cdot}SO_3^-$ (*m/z* = 80) applied to ddMS/MS data derived from the female-control urine sample. Chromatographic peaks with a relative abundance ≥ 20% of the most abundant peak (excluding the internal standard) were selected for further investigation.

By applying sulfate filters to all samples analysed in duplicate, a range of putative sulfate metabolites were identified (Table 3.1). Similar profiles were observed for duplicates.

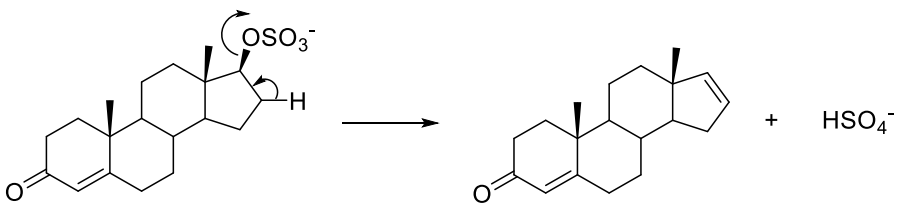
Table 3.1: The number of peaks selected by applying the sulfate filters

Sample no.	Sample name	Number of precursor ions selected from the sulfate filters*
(1) i)	female _control	15
(1) ii)	female_ WT	19
(1) iii)	female_ PVFV	8
(2) i)	male _control	6
(2) ii)	male_ WT	12
(2) iii)	male_ PVFV	6
(3) i)	horse 0 hr _control	11
(3) ii)	horse 0 hr _ WT	16
(3) iii)	horse 0 hr _ PVFV	14
(4) i)	horse 12 hr _control	10
(4) ii)	horse 12 hr _ WT	14
(4) iii)	horse 12 hr _ PVFV	12

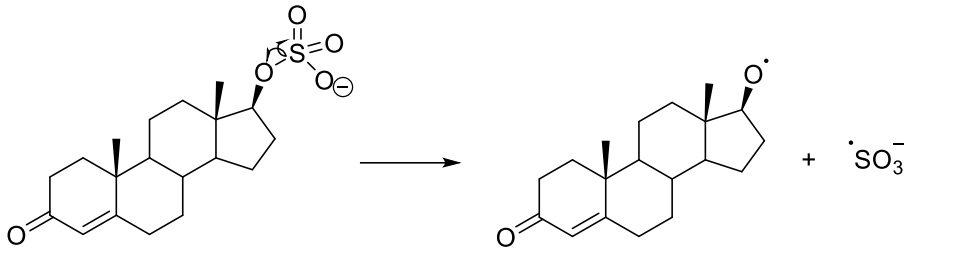
* Number of precursor ions selected with a relative abundance ≥ 20% of the most abundant peak

Then the fragmentation patterns and the product ions for each of the precursor ions were analysed by inspection of the ddMS/MS scans. In the majority instances, this allowed the precursor ions to be identified, and in some instances some assignment of metabolite structure. The typical fragmentation patterns expected for the mono-sulfates and bis-sulfates based on previous studies [12, 13] are exemplified in Figure 3.6. The previous studies report that mono-sulfates and bis-sulfates lose HSO_4^-, $^{\cdot}SO_3^-$ or SO_3 of *m/z* 97, *m/z* 80 and 80 Da respectively.

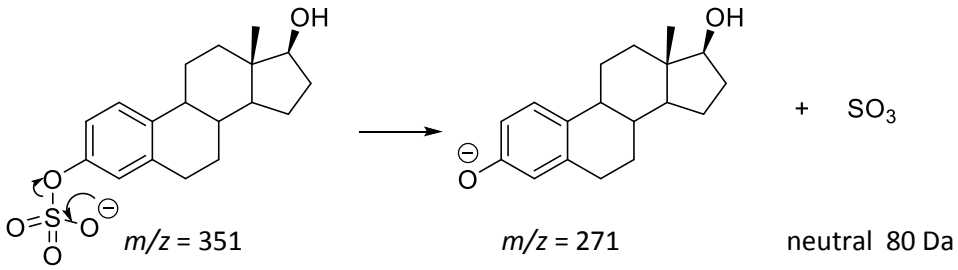
<u>Mono-sulfate ion loss fragmentation for testosterone sulfate</u>



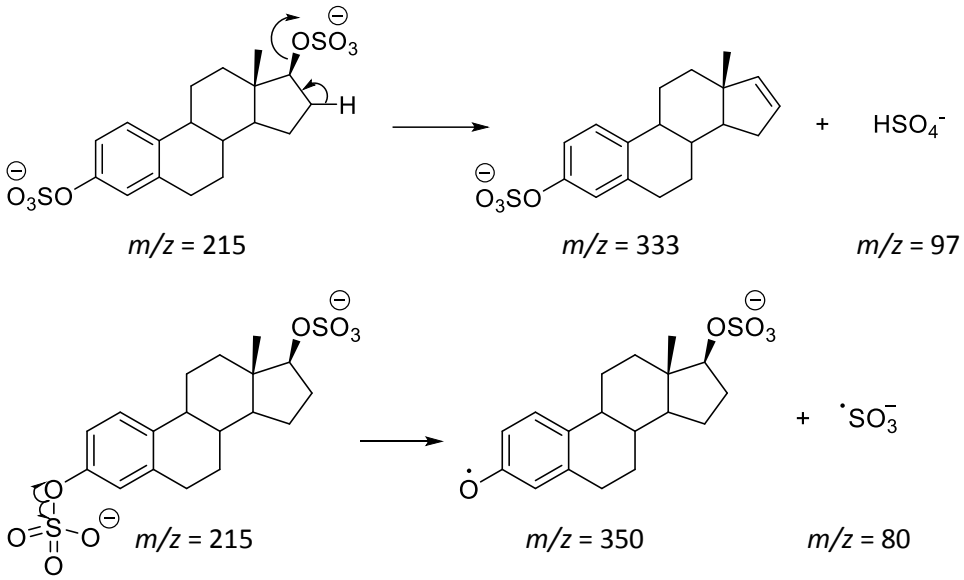
m/z = 367 neutral 270 Da *m/z* = 97



Mono-sulfate neutral loss of SO_3



Bis-sulfate ion loss fragmentation for estradiol bis-sulfate



Bis-sulfate neutral loss of SO_3 for estradiol bis-sulfate

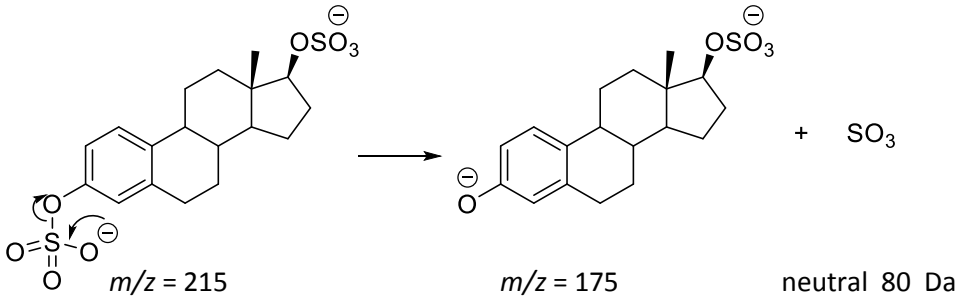


Figure 3.6: Typical fragmentation patterns observed for bis-sulfates and mono-sulfates

In the ddMS/MS data, fragmentation patterns of the mono-sulfates predominantly consisted of a single product ion: the loss of a sulfate ion; or two product ions: loss of a $^{\cdot}SO_3^-$ ion and the ion derived from of the neutral loss of SO_3. Despite the scans targeting only the ion loss of the sulfate moieties, a neutral loss of SO_3 was occasionally observed in the product ion spectra due to ion loss of $^{\cdot}SO_3^-$ in addition to the neutral loss of SO_3. This provided a second ion with which to confirm the identity of the precursor ion. Due to this, metabolites that only lost SO_3 (neutral loss) would not be observed by this method. Similarly, a product ion derived from the loss of 95 Da was observed in mono-sulfates that could be due to the consecutive neutral losses of SO_3 (80 Da) and $^{\cdot}CH_3$ (15 Da). Again, such cases would only be observed when accompanied by an ion loss of the sulfate derived ion at *m/z* 80 or *m/z* 97 leading to the mono-sulfate being observed as a precursor ion (Figure 3.7).

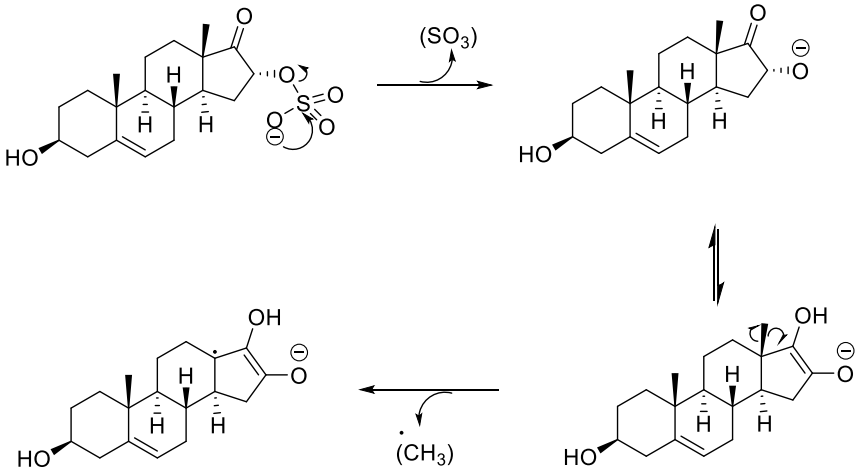


Figure 3.7: The proposed pathway for the loss of 95 Da in a mono-sulfate.

Furthermore, a product ion of *m/z* 81 was often observed that corresponds to an ion loss of a HSO_3^- [12] which provides potential for an additional filter.

The identification of bis-sulfates was achieved with much less ambiguity due to the additional product ions typically afforded by fragmentation including observation of a mono-sulfate with a greater *m/z* value than that of the precursor ion. The fragmentation pattern of bis-sulfates therefore typically consisted of two product ions: the loss of a sulfate ion and the ion corresponding to the resulting mono-sulfate.

In addition to a loss of *m/z* = 97 or 80, an apparent loss of *m/z* = 95 was infrequently observed in the bis-sulfates that could be due to a loss of $^{\cdot}SO_3^-$ followed by a loss of $^{\cdot}CH_3$ [12] that could occur in a reaction scheme as shown in Figure 3.8.

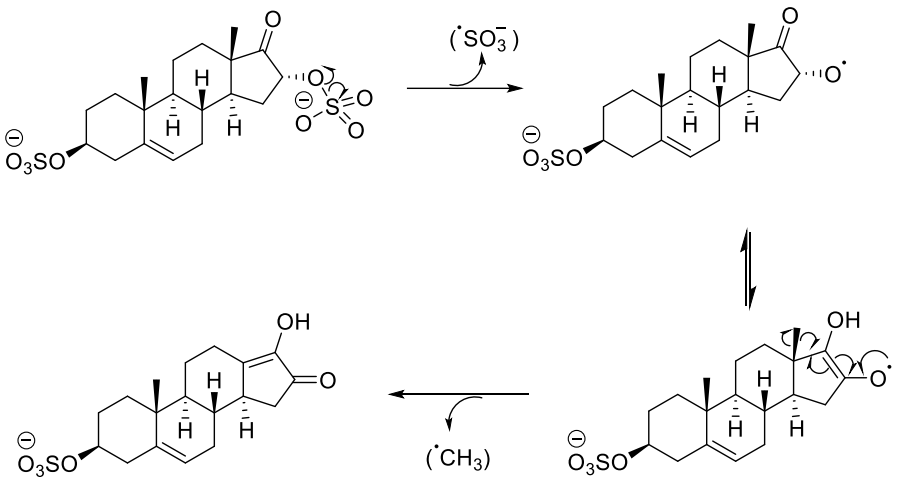
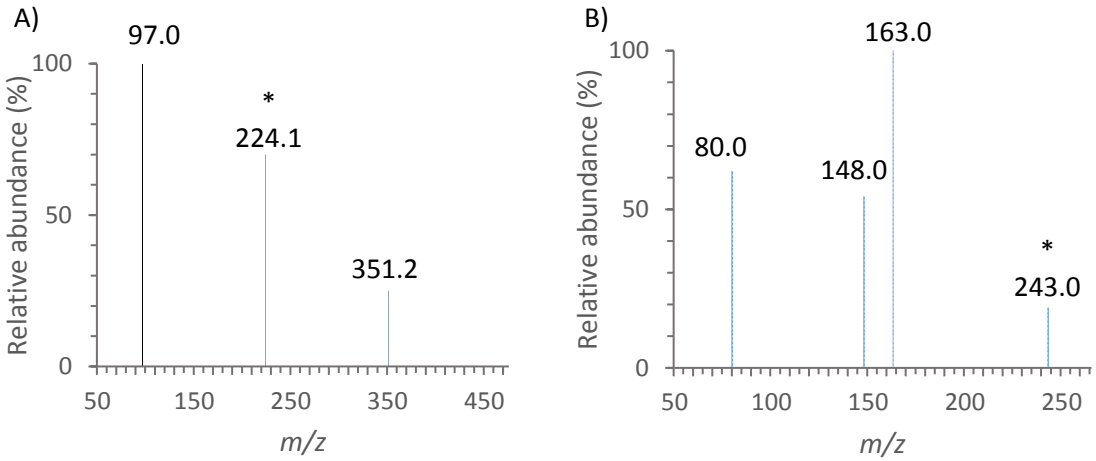


Figure 3.8: The proposed pathway for the loss of *m/z* = 95 in a bis-sulfate. Figure adapted from reference article [12].

Figure 3.9 shows two exemplary product ion spectra. The MS behaviour of the precursor ion (*m/z* = 224) in Figure 3.9A corresponds to that of a bis-sulfate. The loss of a HSO_4^- ion (*m/z* = 97) from the precursor ion resulted in a product ion of *m/z* = 351. Whereas the MS behaviour of the precursor ion (*m/z* = 243) in Figure 3.9B corresponds to a mono-sulfate that shows product ions at *m/z* = 163 that corresponds to the neutral loss of SO_3 and *m/z* = 148 that corresponds to the succeeding loss of and $^{\cdot}CH_3$ (15 Da). A complemented ion loss of $^{\cdot}SO_3^-$ (*m/z* = 80) is also observed indicating why this mono-sulfate was observed in the filtered ddMS/MS data.



The *m/z* values are shown to one decimal point
and the fragments with < 5% relative abundance are not shown for clarity.

Figure 3.9: Exemplary product ion spectra observed in negative mode ddMS/MS corresponding to A) bis-sulfate loss of *m/z* 97 and B) mono-sulfate loss of 95 Da and *m/z* 80. The precursor ions are marked (*).

This study targeted steroidal sulfate analytes. Potential analytes range from estrone (270 Da) to cholesterol (386 Da) leading to notional mono-sulfates ($[M-H]^-$) from *m/z* = 349 to *m/z* = 465 and bis-sulfates ($[M-2H]^{2-}$) from *m/z* = 215 to *m/z* = 272 were considered. The free steroids ($[M+H]^+$) range from *m/z* = 271 to *m/z* = 388. However, the scans employed in the analysis range from *m/z* = 200 to *m/z* = 800 in both polarities. This allowed other classes of sulfate metabolites to be detected by this approach, which may or may not fall within the range of steroids.

3.2.3.1.1 Analysis of the female urine samples

1) Application of sulfate filters to ddMS/MS scans

The scan for the sulfate ion fragments in the female samples with ≥ 20% relative abundance of the most abundant peak in each sample resulted in metabolites that were only observed either in one sample or as common to more than one sample. In total, 33 distinct metabolites were observed with limited overlap of six metabolites between the different samples (Figure 3.10).

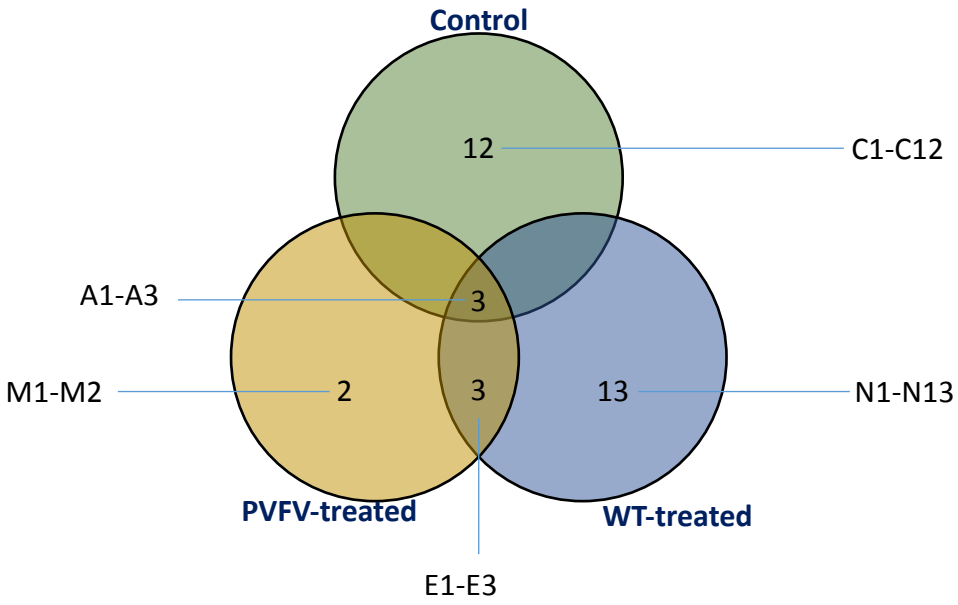


Figure 3.10: Venn diagram representing the number of precursor ions observed for each treatment of female urine.

These metabolites were classified as mono-sulfates (mono) or bis-sulfates (bis) according to the fragmentation patterns observed. They were also classified as steroidal or non-steroidal (NS) based on the steroidal mass range (section 3.2.3.1) and to

observation of even ions when considering mono-sulfates within the mass range (Table 3.2).

Table 3.2: The precursor ions selected by applying the sulfate filters, the retention times, the product ions observed and the prediction as a bis-sulfate or a mono-sulfate.

	Precursor ion	Ret. time (min)	Product ions					Bis /mono sulfate	Non-steroid
			sulfate		Other ions				
			97	80					
All samples									
A1	224 (40)	8.91	x		351 (12)			bis	
A2	383 (20)	9.89	x					mono	
A3	265 (80)	14.71	x					mono	ns
WT and PVFV treated samples									
E1	222 (100)	0.88		x	178 (12) 107 (10)	109 (12) 148 (8)	91 (10) 81 (5)	mono	ns
E2	224 (25)	1.21	x	x	124 (15)	351 (8)	312 (5)	bis	
E3	248 (100)	1.84		x	124 (12)	107 (10)	81 (5)	mono	ns
Control only									
C1	203 (45)	0.74		x	123 (100)	108 (15)		mono	ns
C2	212 (80)	0.86		x	132 (20)	81 (100)		mono	ns
C3	243 (19)	3.32		x	163 (100)	148 (54)		mono	ns
C4	232 (80)	5.97	x		367 (15)			bis	
C5	370 (10)	6.41		x	290 (100)	291 (25)		mono	ns
C6	231 (35)	6.58	x	x	367 (15) 381 (5)	365 (5)	81 (60)	bis	
C7	224 (70)	8.27	x		351 (25)			bis	
C8	354 (10)	9.26		x	274 (100)			mono	ns
C9	352 (5)	9.66		x	272 (100)			mono	ns
C10	238 (30)	10.35	x		379 (8)			bis	
C11	367 (25)	11.59	x					mono	
C12	601 (100)	12.58	x					mono	ns
WT treated sample only									
N1	229 (100)	0.69		x	107 (30) 121 (6)	147 (25) 81 (20)	150 (12)	mono	ns
N2	309 (5)	0.75	x		129 (8)			mono	ns
N3	201 (5)	1.01	x		100 (5)	59 (8)	142 (5)	mono	ns
N4	232 (100)	2.33	x	x	367 (15)	59 (20)	215 (8)	bis	
N5	305 (10)	2.36	x					mono	ns
N6	232 (50)	7.38	x	x	215 (50)	155 (40)	59 (100)	mono	ns

					173 (20) 129 (8)	234 (30) 81 (42)	233 (20)		
N7	385 (50)	8.95	x		386 (10)			mono	
N8	383 (50)	9.07	x		384 (10)			mono	
N9	385 (30)	9.12	x		386 (10)			mono	
N10	383 (20)	10	x		384 (5)			mono	
N11	369 (100)	12.14	x		370 (15)			mono	
N12	369 (100)	12.3	x		370 (25)			mono	
N13	277 (100)	15.36		x				mono	ns
PVFV treated sample only									
M1	357 (15)	4.26		x	243 (5)	67 (20)		mono	
M2	246 (35)	8.16	x		395 (8)			bis	

The relative ion abundance observed in the MS/MS spectrum is shown within brackets.
x denotes identified by filters for *m/z* 80 and/or *m/z* 97.
mono: mono-sulfates; bis: bis-sulfates; NS: non-steroids
Blue highlights: Putative steroids.
Grey highlights: Even mass mono-sulfates within the steroidal mass range.

From the 33 metabolites filtered in the three female samples, 17 metabolites (52%) could possibly be steroids (highlighted in blue). The remaining 16 metabolites are considered non-steroidal (NS) either because they do not fall within the steroidal range or due to the mono-sulfates containing an even-mass ion. In many instances, the sulfate fragment was observed as the base peak.

Eight bis-sulfates were identified (A1, E2, C4, C6, C7, C10, N4 and M2). The loss of a HSO_4^- (*m/z* 97) was common to all the bis sulfates and the resulting mono-sulfate was observed as a minor product ion ranging from 5% - 25% relative intensity. The ion loss of $^{\cdot}SO_3^-$ (*m/z* 80) was observed by three precursor ions (E2, C6 and N4), however, the resultant mono-sulfate fragment was observed with < 1% relative abundance and was absent for N4. This indicates that the loss of *m/z* 97 is favoured with the bis-sulfates when compared to the loss of *m/z* 80. For C6 there was a prominent loss of *m/z* 80 and 15 Da to result in *m/z* 367 ion.

The loss of *m/z* 97 and/or *m/z* 80 was observed in the mono-sulfates. Typically, the loss of *m/z* 97 did not result in other fragment ions. However, six mono-sulfates (C1, C2, C3, C5, C8 and C9) with the loss of *m/z* 80 contained two product ions: one corresponding to the loss of 80 Da and the second at *m/z* 80. This could be due to differences in fragmentation resulting in a mono-anionic product from a neutral loss of SO_3 (80 Da) accompanied by the ion loss of *m/z* 80 as explained previously. Two mono-

sulfates (C1 and C3) fragmented to give sequential loss of 80 Da SO_3 and 15 Da $^{\cdot}CH_3$ (Figure 3.6). In addition, an ion loss of HSO_3^- was observed at *m/z* 81 in seven product ion spectra (E1, E3, C2, C6, N1, N2 (< 2%) and N6) both in bis-sulfates and mono-sulfates indicating the applicability of using *m/z* 81 as an additional filter for sulfate derived ions.

2) Peak integration of the precursor ions

Peak integration was performed on the control and enzyme treated samples in the full scan data, for all the precursor ions selected from any sample regardless of their steroidal nature, and normalized with respect to the internal standard in each sample and presented as a fraction of the respective peak area of the control sample (Table 3.3).

Table 3.3: The integrated peak areas for each precursor ion normalized to the internal standard and presented as a fraction of the respective peak area of the control sample.

	m/z	Ret. time (min)	WT/control	PVFV/control
A1	224	8.91	0.173	0.766
A2	383	9.89	0.591	0.664
A3	265	14.71	0.550	1.520
E1	222	0.88	1.519	2.016
E2	224	1.21	0.823	1.054
E3	248	1.84	1.072	1.176
C1	203	0.74	0.000	0.000
C2	212	0.86	0.000	0.001
C3	243	3.32	0.000	0.000
C4	232	5.97	0.135	0.558
C5	370	6.41	0.000	0.000
C6	231	6.58	0.000	0.000
C7	224	8.27	0.000	0.000
C8	354	9.15	0.000	0.000
C9	352	9.66	0.000	0.000
C10	238	10.35	0.000	0.000
C11	367	11.59	0.044	0.040
C12	601	12.58	0.000	0.000
N1	229	0.69	N/A	N/A
N2	309	0.75	3.674	4.880
N3	201	1.01	1.380	1.572
N4	232	2.33	0.615	0.885
N5	305	2.36	0.705	0.960
N6	232	7.38	0.291	0.205
N7	385	8.95	0.514	0.630
N8	383	9.07	0.239	0.286

N9	385	9.12	0.341	0.573
N10	383	10	0.898	1.119
N11	369	12.14	0.740	0.787
N12	369	12.3	0.723	0.783
N13	277	15.36	0.082	1.262
M1	357	4.26	0.000	0.348
M2	246	8.16	0.000	0.295

Blue and grey highlights: Selected for molecular formula determination.
N/A: Not applicable as the peak was absent in the control sample.

Despite some ions not being selected as common to all samples from the ddMS/MS scan due to the criteria set (≥ 20% relative abundance), in most instances they were found in lower abundance in all samples. The absence of some signals when applying the sulfate mass filters for *m/z* 97 and *m/z* 80 may also reflect the data-dependent acquisition that only selects the most prominent ions observed in the full-scan data for subsequent MS/MS or product ion scan. Importantly, many sulfates selected in the control sample were absent or diminished considerably in the enzyme treated samples indicating significant hydrolysis of those sulfates (C1-C12). From the 33 metabolites, the WT-PaS and PVFV-PaS treated samples exhibit significant changes of < 33% or > 300% compared to the control samples in 20 and 16 metabolites respectively. From those, twelve metabolites (A1, C4-C11, N8, M1 and M2) fall within the mass range of steroids, and nine metabolites (E1, C1, C2, C3, C12, N1, N2, N6 and N13) are non-steroidal. The two metabolites that the enzyme treated samples show a change of > 300% than the control sample (N1 and N2) could be mono sulfates resulting from partial hydrolysis of bis-sulfates. However, all metabolites that fall within the mass range of steroids (Table 3.3 highlighted in blue and grey) regardless of the amounts hydrolysed are of interest and were selected for molecular formula determination.

3) Molecular formulae determination

As the next step, the possible molecular formulae were determined on the 20 metabolites highlighted from Table 3.3. This included all the putative steroids and the three non-steroids within the steroidal range. The elements C, H, O, S, N and P were considered for this determination. Despite the fact that steroid sulfates are unlikely to contain nitrogen and phosphorus, these elements were included in the determination as some metabolites investigated were non-steroidal and it was possible that the masses

that potentially matched steroids were in fact non-steroidal in nature. The molecular formulae were determined as described in section 3.4.5.3. The search encompassed a criterion to contain at least one S and four O atoms in mono-sulfates and two S and eight O atoms in the putative bis-sulfates. Provided that the selected metabolites are sulfates, this criterion agrees well with the mono-sulfates. However, with the bis-sulfates it was assumed that the two anionic charges are derived from two sulfate groups. The possible molecular formulae within a mass tolerance of 3 ppm were compared with that of the product ions and the only formulae that were comparable are reported (Table 3.4).

Table 3.4: The possibilities of molecular formula for selected precursor ions.

	Precursor ion	Product ions	
A1	224.0614 (-2)	351.1637 (-1)	
i	C15 H31 O8 N P S2	C15 H30 O4 N P S	
ii	C19 H28 O8 S2	C19 H27 O4 S	
A2	383.1534 (-1)		
i	C12 H27 O4 N6 S2		
ii	C15 H30 O6 N P S		
iii	C19 H27 O6 S		
E2	224.0433 (-2)		
i	C14 H27 O9 N P S2	C14 H26 O5 N P S	
C4	232.0590 (-2)	367.1587 (-1)	
i	C15 H31 O9 N P S2	C15 H30 O5 N P S	
ii	C19 H28 O9 S2	C19 H27 O5 S	
C5	370.0966 (-1)	290.1398 (-1)	
i	C6 H24 O6 N6 P2 S	C6 H24 O3 N6 P2	
ii	C7 H18 O4 N10 S2	C7 H18 O N10 S	
iii	C9 H20 O5 N7 S2	C9 H20 O2 N7 S	
iv	C12 H23 O7 N2 P S	C12 H23 O4 N2 P	
v	C16 H20 O7 N S	C16 H20 O4 N	
C6	231.0511 (-2)	367.1218 (-1)	365.1419 (-1)
i	C17 H24 O8 N3 S2	C16 H21 O5 N3 S	C17 H23 O4 N3 S
ii	C19 H26 O9 S2	C18 H23 O6 S	C19 H25 O5 S
C7	224.0614 (-2)	351.1637 (-1)	
i	C15 H31 O8 N P S2	C15 H30 O4 N P S	
ii	C19 H28 O8 S2	C19 H27 O4 S	
C8	354.1011 (-1)	274.1452 (-1)	
i	C12 H23 O6 N2 P S	C12 H23 O3 N2 P	
ii	C16 H20 O6 N S	C16 H20 O3 N	
C9	352.0883 (-1)	272.1295 (-1)	
i	C5 H21 O4 N8 P S2	C5 H21 O N8 P S	

ii	C8 H24 O6 N3 P2 S	C8 H24 O3 N3 P2	
iii	C9 H18 O4 N7 S2	C9 H18 O N7 S	
iv	C10 H24 O9 S2	C10 H24 O6 S	
C10	238.0773 (-2)	379.1949 (-1)	
i	C17 H35 O8 N P S2	C17 H34 O4 N P S	
ii	C21 H32 O8 S2	C21 H31 O4 S	
C11	367.1588 (-1)		
i	C13 H28 O4 N4 P S		
ii	C15 H30 O5 N P S		
iii	C17 H25 O4 N3 S		
iv	C19 H27 O5 S		
N4	232.0590 (-2)	367.1586 (-1)	
i	C15 H31 O9 N P S2	C15 H30 O5 N P S	
ii	C19 H28 O9 S2	C19 H27 O5 S	
N7	385.1696 (-1)		
i	C12 H29 O4 N6 S2		
ii	C15 H32 O6 N P S		
iii	C19 H29 O6 S		
N8	383.1538 (-1)		
i	C12 H27 O4 N6 S2		
ii	C15 H30 O6 N P S		
iii	C19 H27 O6 S		
N9	385.1694 (-1)		
i	C12 H29 O4 N6 S2		
ii	C15 H32 O6 N P S		
iii	C19 H29 O6 S		
N10	383.1539 (-1)		
i	C12 H27 O4 N6 S2		
ii	C15 H30 O6 N P S		
iii	C19 H27 O6 S		
N11	369.1745 (-1)		
i	C15 H32 O5 N P S		
ii	C19 H29 O5 S		
N12	369.1744 (-1)		
i	C15 H32 O5 N P S		
ii	C19 H29 O5 S		
M1	357.1016 (-1)		
i	C7 H19 O4 N9 S2		
ii	C10 H22 O6 N4 P S		
iii	C14 H19 O6 N3 S		
iv	C16 H21 O7 S		
M2	246.0749 (-2)	395.1898 (-1)	
i	C17 H35 O9 N P S2	C17 H34 O5 N P S	
ii	C21 H32 O9 S2	C21 H31 O5 S	

Blue highlights: Steroidal molecular formula

Most of the ions resulted in more than one possibility for the molecular formulae and some were determined as non-steroids. As expected, the even mass mono-sulfates (C5, C8, and C9) did not result in steroidal molecular formulae. From the 17 metabolites that were assumed as steroids from their *m/z* values (Table 3.2), only 15 metabolites resulted in a possible steroidal molecular formula with a number of carbon atoms ≥ 18 (number of carbon atoms in estrone is 18) and are highlighted in blue. The E2 and M1 metabolites did not result in a steroidal molecular formula. Interestingly all these 15 metabolites contain a loss of *m/z* 97 out of which only two metabolites exhibited a loss of *m/z* 80 (C6 and N4). Some metabolites possessing the same *m/z* value with different retention times (A1 and C7; A2, N8 and N10; C4 and N4; N7 and N9; N11 and N12) will be considered together for the structure determination. Only the metabolites with a steroidal formula will be considered for structure determination.

4) Steroid structure determination

The metabolites with a steroidal formula (A1, A2, C4, C6, C7, C10, C11, N4, N7-N12, and M2) were selected for structure determination. The structures were elucidated from the derived molecular formulae and the fragmentation patterns for the ions that could possibly possess a steroidal structure. The structure search was based on the reported structures on SciFinder molecular formula search tool. The structures for all derived molecular formulae (steroidal or non-steroidal) in these 15 metabolites were investigated. All the structures that resulted from this tool were filtered based on the following criteria: 1) should contain at least one $-OSO_3^-$ group; 2) no heterocyclic steroids (as endogenous steroids with a heterocyclic structure are not reported); 3) no heavy-atom isotopes. Interestingly the non-steroidal formulae for the metabolites (e.g.: $(C17\ H35\ O9\ N\ P\ S2)^{2-}$ of M2) did not result in any structures and the steroidal molecular formulae resulted only in steroidal structures but no non-steroidal structures. The results of the structure search are shown in Table A6 (see Appendix). From these 15 putative steroid sulfates, eight were significantly hydrolysed by either enzyme or both (Table 3.3) out of which six are bis-sulfates and three are mono-sulfates.

However, this structural search has limitations as some of the published structures have inadequate evidence, as they are not compared with reference standards. Given many possibilities that exist for the structures of a single ion, they could be confirmed

by comparison with synthesized standards. However, a confirmation step was not carried due to limited time and the off-site analysis conducted as part of this study.

5) Integration of the positive mode derivatives

The expected masses of the complete hydrolysed products in the positive mode were calculated to the fourth decimal point by subtracting the mass of the lost sulfate group(s) from the sulfate metabolite precursor ion. The calculated *m/z* for $[M+H]^+$ were then filtered from the positive polarity data in the full scan mode, for the 15 metabolites having a steroidal molecular formula (Table 3.5).

Table 3.5: Peak integration of corresponding positive derivatives normalized to the internal standard and determination of molecular formulae.

negative derivative			positive derivative: mass tolerance=5 ppm					
	m/z*	molecular formula	calculated m/z	ret time (min)	average area x 10^{-3}			molecular formula#
					control	WT	PVFV	
A1 C7	224.0614 (-2)	C15 H31 O8 N P S2 C19 H28 O8 S2	291.2319	11.9	18.388	16.763	16.677	C15 H34 O2 N P C19 H31 O2
N11 N12	369.1744 (-1)	C15 H32 O5 N P S C19 H29 O5 S						
C4 N4	232.059 (-2)	C15 H31 O9 N P S2 C19 H28 O9 S2	307.2268	8.24 12.25	0.561 0.000	0.774 0.068	4.744 0.077	C15 H34 O3 N P C19 H31 O3
N7 N9	385.1696 (-1)	C15 H32 O6 N P S C19 H29 O6 S C12 H29 O4 N6 S2		17.71	1.526	1.238	1.174	
C6	231.0511 (-2)	C17 H24 O8 N3 S2 C19 H26 O9 S2	305.2111	8.93 9.3 10.19	2.085 0.000 6.065	0.958 2.090 1.354	2.296 0.728 1.612	C17 H27 O2 N3 C19 H29 O3
A2 N8 N10	383.1538 (-1)	C12 H27 O4 N6 S2 C15 H30 O6 N P S C19 H27 O6 S		12.65 13.38 13.6 16.66	0.028 0.218 2.053 0.904	2.648 4.668 3.552 1.705	3.861 3.671 3.770 1.571	
C10	238.0773 (-2)	C17 H35 O8 N P S2 C21 H32 O8 S2	319.2632	-	N/D	N/D	N/D	Not determined
C11	367.1588 (-1)	C13 H28 O4 N4 P S C17 H25 O4 N3 S C19 H27 O5 S C15 H30 O5 N P S	289.2162	8.24	1.536	3.501	3.914	C19 H29 O2
M2	246.0749 (-2)	C17 H35 O9 N P S2 C21 H32 O9 S2	335.2581	12.96 16.72	1.232 0.080	1.302 1.746	1.262 2.256	C17 H38 O3 N P C21 H35 O3

N/D: Not detected.
*The charge is specified in brackets.
#Derived from observed precursors with 3 ppm mass tolerance.
Grey highlights: enzyme treated samples > 3 times than control.

In some cases, the mono-sulfates and bis-sulfates give rise to the same completely hydrolysed product and so are shown together in Table 3.5. This suggests that these mono-sulfates could be the partially hydrolysed products of the respective bis-sulfates. Despite the partially hydrolysed products of bis-sulfates C10 and M2 that result in mono-sulfates at *m/z* 397 and *m/z* 413 respectively being observed when the sulfate filters were applied in the ddMS/MS (Table 3.2), they were not reported as the relative abundance was < 2%.

The presence of a single prominent peak was observed for the hydrolysed derivatives of A1/C7/N11/N12 and C11 metabolites whereas multiple peaks for the same *m/z* value were observed in some instances (C4/N4/N7 and N9, C6/A2/N8 and N10, and M2). No peaks in the positive mode were detected for the calculated *m/z* value for the hydrolysed derivative of C10.

Integration of the peak areas in the positive mode resulted in three ions with peaks that integrate to more than 3 times the peak area in the enzyme treated sample over the control sample (highlighted in grey). However, the analysis of the positive mode derivatives is challenging for a few reasons; there lies some uncertainty regarding the functionality of the free steroids. Those without functionality such as a carbonyl group to be protonated would not be detected in the positive mode. Furthermore, the extent of hydrolysis for the bis-sulfates is unknown as whether they result in completely hydrolysed free steroids or partially hydrolysed mono-sulfated products.

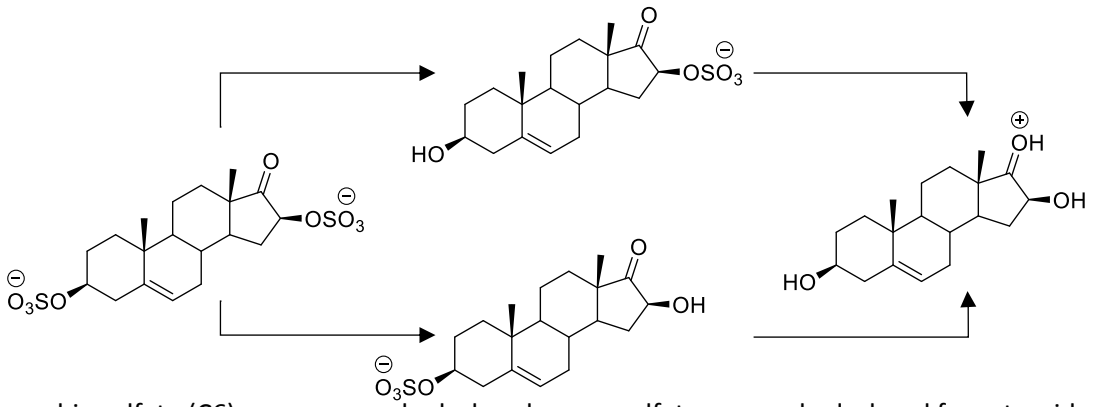
The putative molecular formulae of the positive mode derivatives were determined in a similar way to step 3, and only the formulae that match with the formulae derived from negative mode data are reported. From these 15 putative steroid sulfates, eight showed significant hydrolysis in the negative mode data (A1, C4, C6, C7, C10, C11, N8, M1 and M2) by both WT-PaS and PVFV-PaS except for A1 and C4 that showed significant hydrolysis only by WT-PaS. The amount of hydrolysis shown in the negative mode data do not correlate well with that of the positive mode data for A1, C7 and C10 metabolites, which showed significant hydrolysis in the negative mode data but not in the positive mode data. However, the other metabolites contained peaks that show significant hydrolysis in both negative and positive mode data.

Analysis of the C6 metabolite

By proceeding through all the steps explained above, the C6 metabolite serves as suitable case for further discussion. The C6 metabolite was identified as a bis-sulfate (*m/z* 231) and the absence of this metabolite in the enzyme treated samples indicates hydrolysis by the enzymes (Table 3.3). The bis-sulfate suggests the possible presence of two mono-sulfates by partial hydrolysis of either sulfate group. The enzyme hydrolysed mono-sulfate product of C6 is calculated as *m/z* 383. This value corresponds to that of the A2, N8 or N10 metabolites that were all observed to some extent in the controls. The enzyme treated samples containing reduced amounts of the three mono-sulfate metabolites compared to the control sample (except for the PVFV treated N10 metabolite) and no additional mono-sulfate peaks at *m/z* 383 in the enzyme treated samples suggests extensive hydrolysis of the bis-sulfate to the free steroid rather than the predominance of partial hydrolysis (Table 3.3).

The scan in the positive mode for the calculated mass (*m/z* 305) of the complete hydrolysed product resulted in seven peaks out of which three resulted in peak areas > 10 times compared to the control (highlighted in grey). This observation is consistent with the number of peaks expected from the complete hydrolysis of the three mono-sulfates (two of which could be derived from the bis-sulfate) and/or one bis-sulfate metabolite that should give a number between two to four hydrolysed products.

The C6 metabolite identified as a bis-sulfate shows a fragmentation pattern with product ions for the loss of *m/z* 97 (corresponding to the loss of HSO_4^-) and *m/z* 95 (corresponding to the loss of $^{\cdot}SO_3^-$ followed by the loss of $^{\cdot}CH_3$). The respective fragmentations could occur as depicted in Figure 3.7. Previous studies of bis-sulfates suggest the loss of *m/z* 95 is associated with a sulfate adjacent to ketone or alkene [12]. Taken these points into consideration with the structure search results (Table A6, see Appendix), the following structure or a stereoisomer (Figure 3.11) can be proposed for the C6 metabolite. Confirmation of this proposed structure would require access to reference materials that would allow comparisons of LC retention times and MS behaviour.



bis-sulfate (C6)
m/z 231

hydrolysed mono-sulfate
m/z 383

hydrolysed free steroid
m/z 305

Figure 3.11: The proposed structure for the C6 metabolite and the enzyme hydrolysed products. Candidates for the mono-sulfates and the free steroid were observed in the negative and positive mode data respectively.

3.2.3.1.2 Analysis of the male urine samples

The number of metabolites selected from the application of sulfate filters was comparatively lower for all the male samples than for the female samples (Table 3.1). Most of the metabolites observed in the male samples were present in the female samples; control sample: C2, C7, A1, C10, C11 and M2; WT treated sample: N1, E2, A2, N11, A3 and N13; PVFV treated sample: E2, M1, M2, A1 and A2. In contrast, E1, E3, C1, C3-C6, C8, C9, C12, N2-N10 and N12 metabolites were not identified in the male samples with ≥ 20% relative abundance.

The male samples identified five bis-sulfates, which were also observed in the female samples (A1, E2, C7, C11 and M2). The control, WT treated and PVFV treated samples contained zero, six and one new sulfate metabolites respectively, that were not observed in the female samples (Table 3.6).

Table 3.6: The additional precursor ions observed in the male samples, the retention times, the product ions observed and the prediction as a bis-sulfate or a mono-sulfate

	Precursor ion	Ret. time (min)	Product ions				Bis/mono sulfate	Non-steroid
			sulfate		other ions			
			97	80				
WT treated sample only								
N14	309 (100)	15.18	x				mono	ns
N15	277 (100)	15.56		x			mono	ns
N16	353 (100)	15.59	x				mono	
N17	291 (100)	15.75		x			mono	ns
N18	291 (100)	16.47		x			mono	ns
N19	305 (100)	16.6		x			mono	ns
PVFV treated sample only								
M3	214 (100)	0.42		x	178 (55)	134 (35)	mono	ns

The relative ion abundance observed in the MS/MS spectrum is shown within brackets.
x denotes identified by filters for *m/z* 80 and/or *m/z* 97.
mono: mono-sulfates; bis: bis-sulfates; ns: non-steroids
Blue highlights: Putative steroids;

The analysis of the male urine samples contained some uncertainty due to the incompatibility of the retention times between samples possibly due to slight changes in column temperature. The WADA chromatographic criterion specify the difference in retention time (ΔRT) between the chromatographic peaks of the same analyte should not exceed 1% or ± 0.1 minutes. Complying this criterion was challenging when selecting the peaks for integration in the male samples. The corresponding analytes between samples were selected to the best ability by considering similar peak patterns and the shifts in the NS internal standard.

Table 3.7: The integrated peak areas for each precursor ion normalized to the internal standard and presented as a fraction of the respective peak area of the control sample

	m/z	Ret. time (min)	WT/control	PVFV/control
A1	224	8.88	0.106	1.075
A2	383	9.77	0.294	0.524
A3	265	14.38	1.796	0.820
E2	224	1.21	1.389	1.602
C2	212	0.80	0.000	0.000
C7	224	8.09	0.000	0.000
C10	238	10.13	0.000	0.000
C11	367	11.53	0.016	0.020
N1	229	0.40	12.858	1.496
N11	369	11.16	0.171	0.019
N13	277	15.11	12.472	1.050
N14	309	15.18	6.378	0.425
N15	277	15.56	15.849	1.029
N16	353	15.59	5.354	0.433
N17	291	15.75	22.564	1.179
N18	291	16.47	18.599	0.864
N19	305	16.45	18.973	1.434
M1	357	4.09	0.000	0.734
M2	246	7.85	0.000	0.418
M3	214	0.42	0.827	2.026

Blue highlights: Putative steroids

Comparison of the peak areas (Table 3.3 and Table 3.7) of the metabolites common to the female samples reveals that some metabolites (C2, C7, C10 and C11) are absent or reduced in the enzyme treated samples indicating considerable hydrolysis. Knowing the presence of DHEAS in high abundance [14] and observing DHEAS in human urine in a study performed to target sulfate metabolites in human urine [13], the C11 metabolite present in both female and male samples could be DHEAS. In contrast, A1, A2, A3, E2, N1, N11, N13 M1 and M2 metabolites show different levels of hydrolysis compared to the female samples. The N14-N19 metabolites show an interesting pattern of integrated peak areas as these metabolites in the WT treated sample are 5-23 times greater than that of the control sample, whereas the PVFV treated sample contains comparable amounts of these metabolites to the control sample. From the 20 metabolites identified, the WT-PaS and PVFV-PaS treated samples exhibit significant changes of < 33% or > 300% compared to the control samples in 17 and 5 metabolites respectively.

The N16 metabolite is the only additional metabolite observed in the male samples that falls within the steroidal mass range and was selected for molecular formula determination. However, it resulted in a single formula, $(C16\ H33\ O6\ S)^{1-}$ that is non-steroidal and will not be discussed further.

3.2.3.1.3 Analysis of the horse urine samples

This portion of the study focussed on identifying metabolites of testosterone propionate (TP) (Figure 3.12). This was based on comparing thoroughbred gelding urine samples collected at 0 and 12 hours post-administration of intramuscularly injected TP (see 3.1.1). In a recent study [15] the administration of TP to thoroughbred gelding was performed to test the T/ET ratio, however, studies on TP administration for untargeted steroidomics was not found on literature. Further, the enzyme hydrolysis of the sulfate metabolites was tested by performing the same treatment as in the human urine samples. In this text, the pre-administration and post-administration samples are referred to as 0 hr and 12 hr samples respectively.

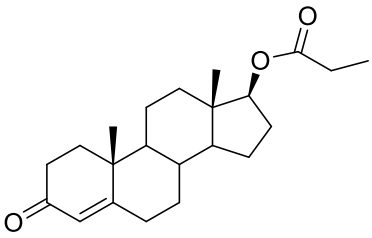


Figure 3.12: Structure of testosterone propionate

In the filtered ddMS/MS data, the peaks with ≥ 20% relative abundance compared to the most abundant peak of three samples each (one control and two enzyme treated) of 0 hr and 12 hr post-administration urine were selected. The analysis of these samples resulted in a number of sulfate metabolites (Table 3.1) out of which only a single metabolite (P27) in the 12 hr sample was identified as a bis-sulfate. The peak areas of all these metabolites (40 unique metabolites) were integrated in all samples irrespective of which sample the metabolite was selected from. The peak integrated results of the controls of the 0 hr and 12 hr samples are reported in Table 3.8.

Table 3.8: The integrated peak areas in the 0 hr and 12 hr control samples for each precursor ion resulting from the sulfate filters in the ddMS/MS normalized to the internal standard.

	Ret. time (min)	m/z	Average area		
			0hr	12 hr	12hr/0hr
P1	0.29	253	0.076	0.133	1.75
P2	0.31	217	0.133	0.194	1.46
P3	0.4	210	0.010	0.014	1.36
P4	0.63	243	0.059	0.162	2.75
P5	0.69	212	11.153	10.377	0.93
P6	0.78	212	33.862	60.489	1.79
P7	0.84	215	0.168	0.961	5.72
P8	0.85	230	0.415	0.715	1.72
P9	0.95	212	1.835	1.915	1.04
P10	0.96	273	0.052	0.013	0.24
P11	0.99	201	0.011	0.016	1.49
P12	1.22	226	2.979	3.417	1.15
P13	1.27	230	4.361	5.172	1.19
P14	1.32	210	0.012	0.010	0.87
P15	1.64	254	0.155	0.832	5.36
P16	1.73	247	0.133	0.022	0.16
P17	1.83	254	0.603	0.419	0.69
P18	1.99	240	0.354	0.447	1.26
P19	2.51	226	32.721	39.028	1.19
P20	2.86	301	0.202	0.632	3.13
P21	2.88	301	2.727	8.317	3.05
P22	4.53	209	0.173	1.120	6.46
P23	4.77	279	0.579	0.777	1.34
P24	5.53	270	0.487	0.122	0.25
P25	5.79	291	0.135	0.297	2.19
P26	6.06	284	0.157	0.041	0.26
P27	6.4	232	0.000	0.203	N/A
P28	6.53	233	0.476	0.132	0.28
P29	6.72	287	0.012	0.057	4.63
P30	6.83	289	0.430	2.381	5.54
P31	6.84	209	0.335	0.061	0.18
P32	6.95	289	0.430	0.456	1.06
P33	8.01	319	0.102	0.092	0.90
P34	8.84	344	0.129	1.192	9.28
P35	9.04	395	0.045	0.035	0.77
P36	10.15	314	0.027	0.067	2.48
P37	11.47	369	0.001	1.423	1998.33
P38	14.11	265	0.631	0.724	1.15
P39	15.26	353	0.031	0.034	1.10
P40	15.66	293	0.263	0.299	1.14

Blue highlights: Putative steroids
Grey highlight: 12 hr area >3 times than 0 hr area
N/A: not applicable

Within the defined criteria in the selection of metabolites, only three metabolites were found common to horse and human samples. The P6, P11 and P38 all non-steroidal metabolites were present in the human samples as C2, N3 and A3 respectively. The integrated peak areas of these metabolites in the controls of horse and human samples are presented in Table 3.13.

Table 3.13: The integrated peak areas of the three common metabolites in horse and human control samples.

m/z	Horse control samples			Human control samples		
		Average area			Average area	
		0 hr	12 hr		female	male
212	P6	33.862	60.489	C2	6.509	15.740
201	P11	0.011	0.016	N3	0.009	0.000
265	P38	0.631	0.724	A3	1.268	0.770

The variation between the human and horse samples are not very significant for P11 and P38 metabolites. However, the P6 metabolite varies significantly in each sample with the horse samples containing 2-10 times greater amounts than that in the human samples. The enzyme hydrolysis of this metabolite reduced the amount of sulfate by more than 500 fold compared to the control sample in both horse (not shown) and human samples.

From the 40 metabolites, ten metabolites had an average area that was three times higher in the 12 hr sample than in the 0 hr sample (grey highlights). Interestingly, six metabolites were observed to decrease > 3 times in the 12 hr sample compared to the 0 hr sample. Only four metabolites fell within the mass range of steroids (blue highlights: P27, P35, P37 and P39). The details of the MS/MS data of these metabolites are reported (Table 3.9).

Table 3.9: The product ions of the metabolites observed with > 5 times in the 12 hr post-administration sample than the 0 hr sample and/or metabolites within the steroidal mass range.

	Precursor	Ret. time (min)	Product ions					Bis/mono sulfate
			97	80	other ions			
P7	215 (20)	0.84		x	178 (100)	135 (8)		mono-sulfate
P15	254 (3)	1.64	x	x	174 (100)			mono-sulfate
P22	209 (55)	4.53	x					mono-sulfate
P27	232 (40)	6.4	x	x	233 (40)	369 (20)	155 (35)	bis-sulfate
					59 (80)	383 (16)	81 (70)	
					367 (5)			
P30	289 (40)	6.83	x					mono-sulfate
P34	344 (100)	8.84		x	124 (60)	59 (70)	67 (50)	mono-sulfate
P35	395 (12)	9.04	x		257 (20)			mono-sulfate
P37	369 (100)	11.47	x		370 (25)			mono-sulfate
P39	353 (90)	15.26	x					mono-sulfate

The relative ion abundance observed in the MS/MS spectrum is shown within brackets.

As shown in Table 3.9, the P27 metabolite exhibits the typical fragmentation pattern of a bis-sulfate showing fragmented product ions derived from the losses of *m/z* 97 (*m/z* 367), *m/z* 81 (*m/z* 383) and *m/z* 80 + 15 Da (*m/z* 369). From the four metabolites that fell within the range of steroids, P35 and P39 did not show significant changes in pre- and post- administration samples (Table 3.8) and will not be discussed. Of the non-steroids that show a significant increase (P7, P15, P22, P30 and P34) in the 12 hr sample the increase was below 10-fold when compared to the 0 hr sample. However, the putative steroidal P27 and P37 metabolites dramatically increased in the 12 hr post-administration samples > 1000 times compared to the 0 hr sample and therefore could be potential metabolites that result from TP administration.

The peak integration of P27 and P37 metabolites in the 12 hr enzyme treated samples did not result in significant changes (> 3 fold) relative to the control (Table 3.10). However, the amount of P37 metabolite in the enzyme treated samples indicates approximately half of this sulfate metabolite is hydrolysed.

Table 3.10: The integrated peak areas of the 12 hr samples of P27 and P37 normalized to the internal standard and presented as a fraction of the respective peak area of the control sample.

	m/z	Ret. time (min)	WT/control	PVFV/control
P27	232	6.4	1.038	1.514
P37	369	11.47	0.551	0.527

The positive mode scan of the hydrolysed free steroid products of P27 and P37 (hereafter named as X27 and X37 respectively) at *m/z* 307 and 291 respectively resulted in a single peak at each mass that could be integrated (Table 3.11) and a range of smaller peaks that were poorly resolved.

Table 3.11: Peak integration of the corresponding hydrolysed free steroid products of P27 and P37 metabolites in the 12 hr samples normalized to the internal standard and presented as a fraction of the respective peak area of the control sample.

negative derivative			positive derivative: mass tolerance=5 ppm					
	*m/z**	molecular formula		calculated m/z	Ret. time (min)	average area normalized to control		molecular formula[#]
						WT/ control	PVFV/ control	
P27	232.0590 (-2)	C13 H29 O8 N4 P S2 C14 H30 O8 N3 S3 C15 H31 O9 N P S2 C17 H26 O8 N3 S2 C19 H28 O9 S2	X27	307.2268	15.14	1.18	1.30	- - C15 H34 O3 N P - C19 H31 O3
P37	369.1743 (-1)	C13 H30 O4 N4 P S C14 H31 O4 N3 S2 C15 H32 O5 N P S C17 H27 O4 N3 S C19 H29 O5 S	X37	291.2319	17.22	1.03	1.15	- - C15 H34 O2 N P - C19 H31 O2

*The charge is specified in brackets.

[#]Derived from observed precursors with 3 ppm mass tolerance.

The integration of the X27 and X37 peak areas in the positive mode resulted in lower values in the control compared to the enzyme treated samples. However, observing the respective peaks in the control sample indicates the presence of the free steroid in the initial sample in considerable amounts. When both metabolites are considered, the PVFV mutant shows greater hydrolysis by 10-12% than WT (Table 3.11).

The peak integration of the positive mode derivatives X27 and X37 metabolites in the 0 hr control sample (Table 3.12) indicates the presence of these products prior to

administration. A similar amount of X27 derivative was observed in the control samples pre- and post-administration. This indicates the X27 is an endogenous metabolite that does not significantly change concentration upon TP administration. However, the concentration is slightly enhanced upon enzyme hydrolysis. The X37 metabolite on the other hand increases in concentration significantly upon TP administration and further increases upon enzyme hydrolysis. The molecular formulae derived for the positive derivatives further reduces the number of possibilities available for the negative derivatives. However, it should be noted that the positive mode peaks assigned for these metabolites are putative and may or may not be the accurate derivative of the respective metabolites.

Table 3.12: Peak integration of corresponding positive derivatives of P27 (X27) and P37 (X37) metabolites in the 0 hr and 12 hr control samples.

	average area * 10^{-3}	
	0 hr	12 hr
X27	16.754	16.167
X37	1.093	6.190

A steroidal structure search (on SciFinder molecular formula search tool) of the P27 and P37 metabolites was performed in a way similar to that of the female metabolites (section 3.2.3.1.1) and the structures for all derived molecular formulae are reported in Table A6 (see Appendix). Considering the fragmentation pattern of the P27 bis-sulfate (with the product ions derived from the ion losses at *m/z* 97, *m/z* 81 and *m/z* 95) which is similar to that of the C6 metabolite, the following structure or a stereoisomer can be proposed.

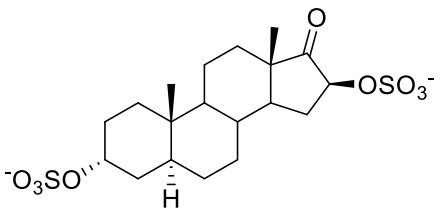


Figure 3.13: The proposed structure for the P27metabolite.

The P37 metabolite is consistent with AS, EAS, ECS or an isomer. However, AS, or ECS are unlikely due to considerable hydrolysis of P37 by both versions of PaS (Table

3.10). Other structures are also found in SciFinder and are reported in Table A6 (see Appendix). Confirmation of structures by comparison with synthesized standards could further establish these metabolites (P27 and P37) to potential markers of TP administration. Furthermore, analysing urine samples for longer time-points to establish the detection window for P27 and P37 metabolites, and also performing a population study to test if similar patterns are observed broadly or in some individuals only, would be important work to be conducted in future.

3.3 Conclusions and Future directions

The initial studies on d_3-TS and DHEAS hydrolysis of PaS enzyme has proven that PaS enzyme and its variants show good hydrolytic activity in the presence of a urine matrix despite the potential for inhibitory substances present such as phosphate. The LEF-PaS mutant at 10 µg mL^{-1} can completely hydrolyse 100, 200 and 300 ng mL^{-1} d_3-TS in an overnight reaction at 37 °C, pH 7.5 analogous to the reaction conditions used by β-glucuronidase enzyme in a typical anti-doping screen. Whereas, d_3-TS hydrolysis by WT-PaS under the same conditions was < 2% at all substrate concentrations. When DHEAS hydrolysis using 100 µg mL^{-1} of enzymes of either WT-PaS or PVFV-PaS, was performed in an overnight reaction at 37 °C, at pH 7 and pH 9, the amount of DHEA produced from hydrolysis was similar for both versions of PaS at pH 9. However, the production of DHEA at pH 7 was slightly lower than that at pH 9, which could be due to lower enzyme stability at pH 7.

The untargeted screening of the urinary sulfate metabolites of the human urine samples discovered mono-sulfates as well as bis-sulfates. The treatment of PaS enzyme on the female urine samples resulted in a decrease in some metabolites compared to the untreated sample. Ten metabolites in the female samples were lowered at least a 2-fold in both the enzyme treated samples. Despite the two versions of PaS showing considerable changes with respect to the control sample, the discrepancy between WT and PVFV treated samples was not significant except for a range of metabolites observed in the male samples (Table 3.7: N14-N19). This could be due to the high enzyme concentration (0.2 mg/mL) used in the reaction and the low concentrations of steroids present in the urine samples. When TS is concerned, WT-PaS has a slightly greater

affinity than PVFV-PaS at pH 9, however, the kinetic parameters closer to neutral pH were not determined. The comparison of female and male samples resulted in 13 common metabolites, 20 and 7 unique metabolites in the female and male samples respectively. From these, WT-PaS and PVFV-PaS treated samples exhibit significant changes of < 33% or > 300% compared to the control samples in 29 and 17 metabolites respectively in total out of which 12 and 7 are putative steroidal metabolites.

The analysis of the 0 hr and 12 hr TP post-administration equine urine samples identified no bis-sulfates in the 0 hr sample and one bis-sulfate in the 12 hr sample. Interestingly, this study discovered two putative steroid sulfates, P27 and P37, in the 12 hr sample with > 1000 fold increase over the 0 hr sample. The presence of the presumed hydrolysed products of these metabolites, X27 and X37 in the 0 hr control sample indicates the endogenous nature of the positive mode derivatives or the free urinary steroids. The level of X27 does not increase in the 12 hr control sample, but that of X37 significantly increases in the 12 hr sample compared to the 0 hr sample. Further, X27 and X37 are slightly increased upon enzyme treatment. Therefore, P27, P37 and X37 could be used as potential markers to detect TP administration. This highlights the importance of sulfates as markers of drug administration as described in section 1.4 (Chapter 1). The elevated level of X27 and X37 metabolites upon enzyme hydrolysis shows the sulfatase could be of use in sample processing (e.g.: for GC-MS).

Due to the limited resources, this study was performed as technical duplicates. Performing this study as biological replicates is one way the study could be extended and improved. Observing an ion loss of *m/z* 81 corresponding to HSO_3^- indicates that it could be used as an additional filter when filtering the precursor ions in the ddMS/MS that would make the selection more comprehensive. In addition, the neutral loss of SO_3 (80 Da) has not been investigated by this method. Further, the criterion followed when selecting the precursor ions in ddMS/MS of having ≥ 20% relative abundance would mask some potential markers. If faster techniques were available for data analysis, more metabolites would have been identified and more potential markers would have been discovered by considering a lower percentage of relative abundance for the selection.

The future direction to continue this study is to confirm the elucidated structures of the identified metabolites with synthesized standards by comparison of retention times under the same conditions

3.4 Experimental

Unless otherwise stated the chemicals were purchased from Sigma Aldrich. Dehydroepiandrosterone (DHEA) was purchased from BDH (Poole, UK) and DHEAS was synthesized according to a published method [16]. The d_3-T and d_3-TS (NEt_3 salt) were purchased from the National Measurement Institute (NSW, Australia). MilliQ water was used in all aqueous solutions and HPLC grade methanol was used for all LC-MS related work. The LC-MS system specifications are detailed in section 2.6.4 (see Chapter 2).

3.4.1 Acquisition of urine samples

A 14-year-old 635 kg thoroughbred gelding was recruited by Charles Sturt University (CSU) School of Animal and Veterinary Sciences for this study. Testoprop® (testosterone propionate, 250 mg) was administered by intramuscular injection in the neck on the opposite side to the sampling jugular catheter. Urine samples were collected at -72, -48, -24, 0, 2, 4, 6, 8, 12, 24, 36, 48 and 72 hours, then daily to 28 days post-administration. Urine samples were stored frozen at -20 °C before and after transportation to the Australian Racing Forensic Laboratory (ARFL) or Australian National University (ANU) until analysed. This study was approved by Charles Sturt University's Animal Care and Ethics Committee and was in accordance with the Australian Code of Practice for the Care and Use of Experimental Animals for Scientific Purposes (Approval number 12/083). In addition, the research proposal and approval was reviewed and supported by the Racing NSW Animal Care and Ethics Committee. The horse tolerated the drug well with no adverse effects reported.

One male and one female were recruited for this study. Urine samples were collected within two days of analysis from a healthy male aged 23 years and a healthy female aged 31 years. The collected urine samples were stored at 4 °C until analysed. The study was approved by the ANU Human Research Ethics Committee and was in accordance with the 2007 National Statement on Ethical Conduct in Human Research (Approval number 2013/654). Volunteers gave written informed consent prior to participation.

3.4.2 Quantification of d_3-TS hydrolysis in a spiked urine sample

All reactions were performed in 1.5 mL of 0.2 M Tris-HCl buffer (pH 7.5 at 37 °C, 1 µM N as the internal standard) with a 2.1 mL urine component and 0.4 mL of enzyme or enzyme storage buffer (0.1 mM Tris HCl, pH 7.5 in 50% glycerol) in 10 mL glass tubes. A series of d_3-T calibrators at 0, 1, 5, 10, 50, 100, 200, and 400 ng mL^{-1} in human urine was prepared by serial dilution of 200 µg mL^{-1} d_3-T in methanol. A 2.1 mL aliquot of each calibrator was added to 1.5 mL Tris-HCl buffer and 0.4 mL of enzyme storage buffer. A d_3-TS spiked urine sample was prepared from a 200 µg mL^{-1} d_3-TS in methanol to give final d_3-TS concentrations of 100, 200 or 300 ng mL^{-1} (61.6, 123.3, 185.0 ng mL^{-1} d_3-T equivalent respectively). A 2.1 mL aliquot of blank/d_3-TS spiked urine in 1.5 mL Tris-HCl buffer was reacted with 0.4 mL of 1 mg mL^{-1} or 0.1 mg mL^{-1} WT-PaS or LEF-PaS. Control reactions with either blank or spiked urine were performed with enzyme storage buffer. Following overnight incubation at 37 °C, 4 mL of methyl *tert*-butyl ether was added to each tube and an end-over-end rotation was performed for 30 minutes. Next, these tubes were centrifuged for 15 minutes at 4000 rpm and the top layer was decanted into new glass tubes. These samples were evaporated to dryness with a flow of N_2 at 60 °C. The dried samples were then reconstituted with 200 µL of 60% MeOH and were transferred into HPLC vials with inserts. Test samples and calibrators (10 µL) were injected into UPLC-MS system and signals for d_3-T and N were detected in the positive mode as the proton adduct, $[M+H]^+$ at *m/z* 292.4 and *m/z* 275.2 respectively.

3.4.3 Testing endogenous DHEAS hydrolysis in human urine

All reactions were performed in 1.5 mL of 0.2 M tris-acetate buffer (pH 7 or pH 9 at 37 °C, 1 µM nandrolone as the internal standard) with a 2.1 mL urine component and 0.4 mL of enzyme or enzyme storage buffer (0.1 mM tris HCl, pH 7.5 in 50% glycerol) in 10 mL glass tubes. A series of DHEA calibrators of 0, 1, 5, 10, 50, 100, 500, 1000, 1500 ng mL^{-1} was prepared in human urine by serial dilution of 150 µg mL^{-1} DHEA in methanol. Then 2.1 mL of each calibrator was added to 1.5 mL tris-acetate buffer and 0.4 mL of enzyme storage buffer. A 2.1 mL aliquot of blank urine was added to 1.5 mL tris-acetate buffer and was reacted with 0.4 mL of 1 mg mL^{-1} WT-PaS or PVFV-PaS. Control reactions

at both pHs were performed with enzyme storage buffer instead of enzyme. Following overnight incubation at 37 °C, 4 mL of methyl *tert*-butyl ether was added to each tube and an end-over-end rotation was performed for 30 minutes. Next, these tubes were centrifuged for 15 minutes at 4000 rpm and the top layer was decanted into new glass tubes. These samples were evaporated to dryness with a flow of N_2 at 60 °C. The dried samples were then reconstituted with 200 µL of 60% MeOH and were transferred into HPLC vials with inserts. Finally, 20 µL of the test samples and calibrators were injected into LCMS system and signals for DHEA and nandrolone were detected in the positive mode as the proton adduct, $[M+H]^+$ at *m/z* 289.2 and *m/z* 275.2 respectively.

3.4.4 Calibration plots of the standards

The LINEST function in Microsoft Excel was used to determine the detection limit of the tests carried out in this chapter. This function uses the least squares method to calculate a straight line that best fits the data (of the calibrators). A new set of *y* values were assigned by using the calculated slope and the intercept from the LINEST function. The lowest calibrator where the deviation of the observed and the calculated values are less than 20% was considered the lower limit of quantification (LLOQ).

Table 3.13: The range and limit of quantification of each experiment.

Section	Experiment	Range (ng mL^{-1})	LLOQ (ng mL^{-1})
3.2.1	Quantification of d3-TS hydrolysis in a spiked urine sample with 100 µg mL^{-1} enzyme	10-400	10
3.2.1	Quantification of d3-TS hydrolysis in a spiked urine sample with 10 µg mL^{-1} enzyme	5-200	5
3.2.2	Testing endogenous DHEAS in human urine at pH 7	10-500	10
3.2.2	Testing endogenous DHEAS in human urine at pH 9	10-500	10

3.4.5 Untargeted screening

3.4.5.1 Sample treatment

Four urine samples were subjected to untargeted screening; the two human urine samples, and two equine urine sample at 0 and 12 hours post-administration of

testosterone propionate (TP). Three reactions were performed for each sample in singlicates; a control reaction with the enzyme storage buffer, a WT-PaS (2 mg mL^{-1}) reaction, and a PVFV-PaS (2 mg mL^{-1}) reaction. All reactions were performed in 1.5 mL of 0.2 M Tris-HCl buffer (pH 7.5 at 37 °C, 1 µM nandrolone as the internal standard for positive mode detection) with 2.1 mL urine sample and 0.4 mL of enzyme storage buffer (0.1 mM Tris HCl, pH 7.5 in 50% glycerol for the control reaction) or enzyme (WT-PaS or PVFV-PaS) in 10 mL glass tubes. All samples were incubated overnight at 37 °C. Following incubation, the samples were centrifuged at 4000 rpm for 15 minutes (Rotofix 32 Hettich centrifuge). Next, solid phase extraction (SPE) was performed on each sample using Waters Oasis WAX SPE cartridges by a protocol described by Waller *et al.*[17]. Briefly, the cartridges were preconditioned with MeOH (1 mL) followed by water (2 mL). The supernatants of the centrifuged samples were loaded to the preconditioned cartridges and washed with 0.1 M NaOH (2 mL), 0.2 M Tris-HCl buffer (2 mL) and water (2 mL). The metabolites and the hydrolysed products were eluted into new 10 mL glass tubes with a mixture of ethyl acetate : methanol : diethylamine (25:25:1 v/v/v, 2 mL). Next 1.5 µL of 1 mM nandrolone sulfate (NS) was added to each tube as an internal standard for the negative mode detection. All samples were evaporated to dryness with a flow of N_2 at 60 °C. The dried samples were stored at -20 °C until analysis. The samples were reconstituted in 200 µL of 60% MeOH before the analysis.

3.4.5.2 Analytical method

Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis was performed on ThermoFisher QExactive Plus ™ (Thermo Fisher, Scoresby, Victoria) Orbitrap mass spectrometer equipped with a Thermo Accucore C18 AQ column (2.1 mm x 100 mm x 2.6 µm). The HPLC method consisted of a flow rate of 0.4 mL/min with a gradient elution of the following mobile phases; A: 10 mM ammonium formate in 100% water, B: 10 mM ammonium formate in 99% methanol:water, gradient: 0-3 min A:B (99:1, v/v), 3-15 min A:B (40:60, v/v), 15-18 A:B (0:100, v/v), 18-20 (0:100, v/v), 20-20.1 min A:B (99:1, v/v) and 2.9 min equilibration time. The column oven temperature was set to 40 °C, and the injection volume was 1 µL. The samples were analysed as two injections each, in both polarities in full scan mode at a resolution of 70000 and the top 10 precursor ions per duty cycle on ddMS/MS at a resolution of 17500 and a normalized

collision energy of 35%. The scan range targeted was *m/z* 200-800 in full scan and 50-800 in ddMS/MS.

3.4.5.3 Data analysis

Data analysis was performed using XCalibur 4.0 software that involved several key steps; applying filters to identify metabolites corresponding to a loss of a sulfate moiety on ddMS/MS data; integration of the areas under the peaks of the selected analytes in full scan mode; identifying the possibilities of molecular formulae; determination of the possible structures from the reported compounds and identifying and integrating the corresponding hydrolysis products in positive mode.

The previously reported characteristic mass spectrometric behaviour (Figure 3.6) of the sulfated metabolites [12] was considered in the data dependent acquisition. A precursor ion scan was performed in ddMS/MS negative mode by considering the common ion loss of *m/z* 97 corresponding to HSO_4^- and *m/z* 80 corresponding to $^{\cdot}SO_3^-$ with a mass tolerance of 10 ppm. The product ion spectra of all precursor ions corresponding to ≥ 20% of the relative abundance in the sulfate filters applied were analysed. The *m/z* value of the precursor ion and the other ions with more than 5% relative abundance present in the product ion spectra were recorded with their relative abundance. The data analysis targeted the identification of mono-sulfated and bis-sulfated steroid metabolites. Therefore, based on the fragmentation patterns and the previously reported data, the precursor ions were predicted as mono-sulfates or bis-sulfates. These metabolites were also classified according to the calculated mass range for steroids; the metabolites out of the steroidal mass range and the mono-sulfates within the mass range but with an even mass ion as non-steroids.

Next, the amount of steroid sulfate detected was evaluated by the integration of the chromatographic peaks of the selected precursor ions with a mass tolerance of 5 ppm in the full scan mode. The integrated peak areas were normalized with respect to the nandrolone sulfate internal standard ($[M-H]^-$) in the corresponding run. The averages of the normalized peak areas of the two technical replicates of the enzyme treated samples were compared with the corresponding average of the control sample. All the metabolites that fell within the calculated steroidal mass range were selected for further investigation. The possibilities for the molecular formulae of the selected

analytes were determined by including C, H, O, S, N and P as the expected elements composing the metabolites with the appropriate charge and constraining the search to a mass tolerance of 3 ppm. The search encompassed limits for the mono-sulfates to contain a minimum of one S and four O atoms and for the bis-sulfates to contain two S and eight O atoms.

Based on the calculated masses of the completely hydrolysed products (Figure 3.11), the positive derivatives were determined. These masses were scanned in the positive polarity of full scan data with a mass tolerance of 5 ppm for the metabolites that resulted in a steroidal molecular formula with the minimum number of C atoms ≥ 18. The probable peaks were integrated and were normalized with respect to the nandrolone internal standard ($[M+H]^+$). They were compared with the respective control sample for evaluation of the enzyme hydrolytic activity. The presence of the correlated molecular formulae of the hydrolysed products were determined by a similar approach to the formulae determination of the analytes in the negative mode. Finally, the structures were proposed from the derived molecular formulae and the fragmentation patterns based on the reported structure on SciFinder molecular formula search tool. All the structures that resulted from this tool were filtered based on the following criteria: 1) should contain at least one $-OSO_3^-$ group; 2) no heterocyclic steroids (as endogenous steroids with a heterocyclic structure are not reported); 3) no heavy-atom isotopes.

3.5 References

[1] Fabregat, A., Pozo, O. J., Marcos, J., Segura, J., and Ventura, R. (2013) Use of LC-MS/MS for the Open detection of steroid metabolites conjugated with glucuronic acid, *Anal. Chem. 85*, 5005-5014.

[2] Dervilly-Pinel, G., Weigel, S., Lommen, A., Chereau, S., Rambaud, L., Essers, M., Antignac, J. P., Nielen, M. W., and Le Bizec, B. (2011) Assessment of two complementary liquid chromatography coupled to high resolution mass spectrometry metabolomics strategies for the screening of anabolic steroid treatment in calves, *Anal. Chim. Acta 700*, 144-154.

[3] Raro, M., Ibanez, M., Gil, R., Fabregat, A., Tudela, E., Deventer, K., Ventura, R., Segura, J., Marcos, J., Kotronoulas, A., Joglar, J., Farre, M., Yang, S., Xing, Y., Van Eenoo, P., Pitarch, E., Hernandez, F., Sancho, J. V., and Pozo, O. J. (2015) Untargeted metabolomics in doping control: detection of new markers of testosterone misuse by ultrahigh performance liquid chromatography coupled to high-resolution mass spectrometry, *Anal. Chem. 87*, 8373-8380.

[4] Dehennin, L., and Matsumoto, A. M. (1993) Long-term administration of testosterone enanthate to normal men - alterations of the urinary profile of androgen metabolites potentially useful for detection of testosterone misuse in sport, *J. Steroid. Biochem. 4*, 179-189.
[5] Tudela, E., Deventer, K., Geldof, L., and Van Eenoo, P. (2015) Urinary detection of conjugated and unconjugated anabolic steroids by dilute-and-shoot liquid chromatography-high resolution mass spectrometry, *Drug Test. Anal. 7*, 95-108.
[6] Rocchitta, G., Spanu, A., Babudieri, S., Latte, G., Madeddu, G., Galleri, G., Nuvoli, S., Bagella, P., Demartis, M. I., Fiore, V., Manetti, R., and Serra, P. A. (2016) Enzyme biosensors for biomedical applications: strategies for safeguarding analytical performances in biological fluids, *Sensors-Basel 16*, e780.
[7] Bowers, L. D., and Sanaullah. (1996) Direct measurement of steroid sulfate and glucuronide conjugates with high-performance liquid chromatography mass spectrometry, *J. Chromatogr. B 687*, 61-68.
[8] Dehennin, L., Ferry, M., Lafarge, P., Peres, G., and Lafarge, J. P. (1998) Oral administration of dehydroepiandrosterone to healthy men: alteration of the urinary androgen profile and consequences for the detection of abuse in sport by gas chromatography mass spectrometry, *Steroids 63*, 80-87.
[9] Hahner, S., and Allolio, B. (2010) Dehydroepiandrosterone to enhance physical performance: myth and reality, *Endocrin. Metab. Clin. 39*, 127-139.
[10] Cawley, A. T., Hine, E. R., Trout, G. J., George, A. V., and Kazlauskas, R. (2004) Searching for new markers of endogenous steroid administration in athletes: "looking outside the metabolic box", *Forensic Sci. Int. 143*, 103-114.
[11] Kalli, A., Smith, G. T., Sweredoski, M. J., and Hess, S. (2013) Evaluation and optimization of mass spectrometric settings during data-dependent acquisition mode: focus on LTQ-orbitrap mass analyzers, *J. Proteome Res. 12*, 3071-3086.
[12] McLeod, M. D., Waller, C. C., Esquivel, A., Balcells, G., Ventura, R., Segura, J., and Pozo, O. J. (2017) A constant ion loss method for the untargeted detection of bissulfate metabolites, *Anal. Chem. 89*, 1602-1609.
[13] Lafaye, A., Junot, C., Ramounet-Le Gall, B., Fritsch, P., Ezan, E., and Tabet, J. C. (2004) Profiling of sulfoconjugates in urine by using precursor ion and neutral loss scans in tandem mass spectrometry. Application to the investigation of heavy metal toxicity in rats, *J. Mass Spectrom. 39*, 655-664.
[14] Labrie, F., Luu-The, V., Belanger, A., Lin, S. X., Simard, J., Pelletier, G., and Labrie, C. (2005) Is dehydroepiandrosterone a hormone?, *J. Endocrinol. 187*, 169-196.
[15] Viljanto, M., Scarth, J. P., Hincks, P., Hillyer, L., Cawley, A., Suann, C. J., Noble, G., Walker, C. J., Kicman, A. T., and Parkin, M. C. (2016) Application of testosterone to epitestosterone ratio to horse urine – a complementary approach to detect the administrations of testosterone and its pro-drugs in thoroughbred geldings, *Drug Test. Anal*. *DOI: 10.1002/dta.2109.*
[16] Waller, C. C., and McLeod, M. D. (2014) A simple method for the small scale synthesis and solid-phase extraction purification of steroid sulfates, *Steroids 92*, 74-80.
[17] Waller, C. C., Cawley, A. T., Suann, C. J., Ma, P., and McLeod, M. D. (2016) In vivo and in vitro metabolism of the designer anabolic steroid furazadrol in thoroughbred racehorses, *J. Pharmaceut. Biomed. 124*, 198-206.

Chapter 4 Concluding remarks and Future Directions

Exploitation of steroids by athletes is an ongoing concern in the sporting arena that demands improved doping control solutions. With the growing importance of detecting the sulfate metabolites of steroids, the significance of a suitable sulfatase enzyme was discussed. Therefore, the primary objective of this study was to engineer an enzyme for improved steroid sulfate hydrolysis. From a range of sulfatases previously reported for the hydrolysis of steroid sulfates [1], the arylsulfatase from *Pseudomonas aeruginosa* (PaS) was selected as the potential candidate. However, it required improvements on substrate scope and activity to be used for anti-doping applications. The PaS enzyme was engineered by a semi-rational approach to discover mutants with improved catalytic activity for testosterone sulfate (TS) hydrolysis and to improve the steroid sulfate substrate scope. This included mutating five residues in turn by SSM, three small groups of neighbouring residues simultaneously and finally shuffling all the beneficial mutations found in the pathway of enzyme engineering.

The substrate binding mode for PaS was unknown. With the structural data in hand, two poses for PaS binding were proposed by molecular modelling in collaboration with Dr. Thomas Balle at the Faculty of Pharmacy, University of Sydney. Therefore, as a first step, the residues around the two poses proposed for substrate binding were explored in turn by SSM. The results of these SSM libraries strongly implicated that pose II is the preferred pose for substrate binding where R155 and K330 residues are located at the opening of the pocket (Figure 4.1).

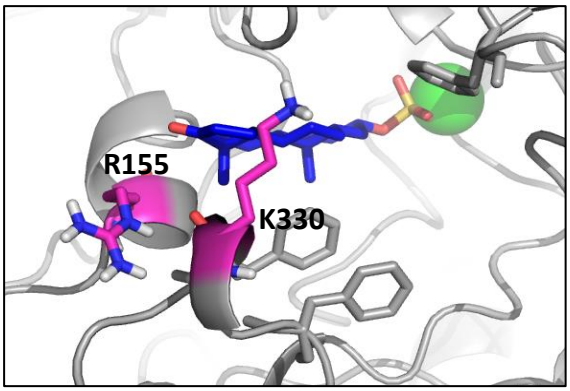


Figure 4.1: Pose II with the substrate, DHEAS, (dark blue) docked into the model derived from the crystal structure (1HDH) [2]. The R155 and K330 residues mutated by SSM are shown in pink.

Employing LC-MS technique for screening, limited the size of the MSM libraries to a few thousand. Hence the library size was restricted by reducing the amino acid alphabets [3-5] that the targeted residues were mutated to. Though this strategy resulted in significantly improved mutants it nevertheless had the disadvantage of potentially missing some positive mutants. However, as a high throughput colourimetric assay was thought unlikely to lead to improved mutants for TS hydrolysis, the coverage had to be compromised to minimise the library size.

After screening eight libraries for TS hydrolysis by cautious selection of twelve residues and one shuffled library of the beneficial mutants from the previous rounds, mutants were achieved with significant improvements in catalytic activity, substrate scope or both. Lowering the substrate concentration used for screening from MSM 2 library onwards, to mimic the conditions expected for real-world samples helped in identifying beneficial mutants with a remarkably favourable K_M that would be more suitable with the samples encountered in anti-doping screens. The LEF-PaS mutant 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. In addition to LEF-PaS mutant, PVIV-PaS and VVIV-PaS mutants also exhibited a catalytic efficiency of more than 150 times than WT-PaS and significant improvement in substrate affinity. However, PVFV-PaS identified in an earlier round shows the broadest substrate range. Importantly, the detection of significant hydrolytic activity with PVFV-PaS for the α-configured steroid sulfates derived from etiocholanolone and epitestosterone was a breakthrough. The PVFV-PaS mutant

hydrolyses ECS in a rate of two orders of magnitude higher than WT-PaS. This revealed that although the substrate scope improved at the early stages of screening, continued screening for a single substrate was detrimental for the substrate scope. The PVFV-PaS mutant now serves as the starting point for further research targeting improved mutants for the hydrolysis of α-configured steroid sulfates. Optimising the screening process for simultaneous evolution for multiple substrates that consist of α-configured or β-configured steroid sulfates, would be useful to greatly minimize the screening efforts. With improved variants for different substrates, a cocktail of enzymes could be used for steroid sulfate hydrolysis in anti-doping. The PVFV-PaS and LEF-PaS mutants also show modest improvements in thermostability.

It was also found that the optimum pH of PaS for the hydrolysis of alkyl steroid sulfates such as testosterone sulfate lies towards a slightly acidic pH. This indicates the importance of a general acid for the protonation of the leaving group of such substrates whereas aryl sulfates such as PNPS do not essentially require a general acid for the protonation of the leaving group [6] and hence the alkaline pHs would be best suitable for PNPS hydrolysis. Further, phosphate buffer was found to inactivate PaS and should not be used for PaS related experiments or application.

The importance of E74 residue in the hydrolysis of steroid sulfates was uncovered by the inactive MSM 3 library that resulted upon mutagenesis of this residue. Targeted mutagenesis of E74 confirmed the necessity of this residue. This result has led to the hypothesis that PaS requires a catalytic triad where E74 polarises H115 that then activates the FGH for the nucleophilic substitution on the sulfate ester or elimination of the sulfate group. Future study of this previously unrecognized catalytic residue could shed light on mechanism of PaS and related enzymes.

For further improvements of PaS, understanding the interactions of the catalytic residues with substrates of different orientations and steriochemistry will also be important. Performing molecular modelling or crystallization would provide useful information on this. A comprehensive modelling of PaS enzyme is currently being carried out to further characterize enzyme-substrate interactions.

Hydrolysis of steroids by PaS in real world samples indicated that the enzyme has good catalytic activity in a urine matrix and can be efficiently used in the conditions analogous to β-glucuronidase used in a typical anti-doping screen. The untargeted

screen of pre- and post-administration of testosterone propionate (TP) horse urine samples identified three metabolites that could be potential markers for TP administration. The untargeted screen would provide more important information if the predicted structures can be confirmed by comparing with synthesized reference materials. Specially confirming the identified potential markers of TP administration would be very important to be useful in anti-doping.

The PaS enzyme has already received international interest from the anti-doping community and WT-PaS and PVFV-PaS have been sent to WADA accredited laboratories in Barcelona, Cologne and Rome for testing. The outcome of these investigations are eagerly awaited.

References

[1] Stevenson, B. J., Waller, C. C., Ma, P., Li, K., Cawley, A. T., Ollis, D. L., and McLeod, M. D. (2015) *Pseudomonas aeruginosa* arylsulfatase: a purified enzyme for the mild hydrolysis of steroid sulfates, *Drug Test. Anal*. *7,* 903-911.

[2] Boltes, I., Czapinska, H., Kahnert, A., von Bulow, R., Dierks, T., Schmidt, B., von Figura, K., Kertesz, M. A., and Uson, I. (2001) 1.3 angstrom structure of arylsulfatase from *Pseudomonas aeruginosa* establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family, *Structure 9*, 483-491.

[3] Reetz, M. T., and Wu, S. (2008) Greatly reduced amino acid alphabets in directed evolution: making the right choice for saturation mutagenesis at homologous enzyme positions, *ChemComm 43*, 5499-5501.

[4] Reetz, M. T., Kahakeaw, D., and Sanchis, J. (2009) Shedding light on the efficacy of laboratory evolution based on iterative saturation mutagenesis, *Mol. BioSyst. 5*, 155-122.

[5] Walter, K. U., Vamvaca, K., and Hilvert, D. (2005) An active enzyme constructed from a 9-amino acid alphabet, *J. Biol. Chem. 280*, 37742–37746.

[6] Barrozo, A., Duarte, F., Bauer, P., Carvalho, A. T., and Kamerlin, S. C. (2015) Cooperative electrostatic interactions drive functional evolution in the alkaline phosphatase superfamily, *J. Am. Chem. Soc. 137*, 9061-9076.

Recipes and reagents

Unless otherwise specified the reagents were purchased from Sigma Alrich and solutions are made in miliQ water.

LB (Lysogeny Broth also known as Luria Burtani media)

Dissolve 5 g of yeast extract (Difco), 10 g of tryptone (Difco), 10 g of NaCl in 900 mL of water and adjust the pH to pH 7 with $NaOH_{(aq)}$ and top up to 1 L with water. Sterilize by autoclaving.

LBA (Lysogeny Broth with ampicillin)

Prepare LB as above and when cooled after autoclaving add filter sterilized ampicillin (Astral) at 100 g/L to give a final concentration of 100 mg/L.

LB-agar

Add 15 g/L of agar (Difco) to LB prepared as above. Sterilize by autoclaving.

LBA-agar

Prepare LB-agar as above and when cooled add filter sterilized ampicillin (Astral) at 100 g/L to give a final concentration of 100 mg/L.

TBA (Terrific Broth with ampicilin)

Dissolve 12 g of tryptone, 24 g of yeast extract and 4 mL of glycerol in 900 mL of water. Separately dissolve 2.31 g of KH_2PO_4 and 12.54 g of K_2HPO_4 in 100 mL of water and sterilize the two solutions separately by autoclaving. Once cooled down mix the two solutions and add filter sterilized ampicillin at 100 g/L to give a final concentration of 100 mg/L.

YENB (Yeast Extract-Nutrient Broth)

Dissolve 7.5 g/L of yeast extract and 8 g/L of nutrient broth (Difco) in water and sterilize by autoclaving.

5 X M9 salts

Dissolve 64 g of $Na_2HPO_4.7H_2O$ (or 34 g of the anhydrous salt), 15 g of KH_2PO_4, 2.5 g of NaCl and 5 g of NH_4Cl in 1 L of water and sterilize by autoclaving.

EMBL trace elements

Dissolve 5 g of EDTA, 0.83 g of $FeCl_3.6H_2O$, 84 mg of $ZnCl_2$, 13 mg of $CuCl_2.2H_2O$, 10 mg of $CoCl_2.6H_2O$, 10 mg of H_3BO_3 ,1.6 mg of $MnCl_2.6H_2O$ in 1 L of water and filter sterilize.

MMTS

Add 1.6 g of agarose (molecular biology grade) in a thoroughly rinsed autoclaved bottle and mix with 156 mL of water and bring the mixture to boil using a microwave oven to prepare the molten agarose solution. Next cool to about 60 °C and add 40 mL of 5 X M9 salts, 2 mL of 100 X EMBL trace elements, 1.27 mL of 50 % (v/v) glycerol (autoclaved sterilized), 400 µL of 1 M $MgCl_2$ & 50 mM $CaCl_2$ (prepared in 70 % ethanol), 200 µL of 100 g/L ampicillin (prepared in 70 % ethanol), 0.2 mM final concentration of TS (prepared in 70 % ethanol) 20 µL of 0.1 M IPTG (prepared in 70 % ethanol). Mix the solution gently but thoroughly by tipping end over end to avoid bubbles and to ensure uniform distribution.

No sulfur storage media

Filter sterilize 20 mL of 50 % (v/v) glycerol, 10 mL of 5 X M9 salts, 500 µL of EMBL trace elements and 100 µL of 1M $MgCl_2$ and 50 mM $CaCl_2$ dissolved in 19.4 mL of water.

Cracking buffer (include β-mercaptoethanol)

Add 120 µL of 0.1 % (w/v) bromophenol blue (Merck), 1.12 mL of 80 % (v/v) glycerol (Merck), 800 µL of 2 M Tris (pH 6.8), 2 mL of 10 % (w/v) SDS, 31 mg/mL dithiothreitol and volume up to 10 mL with water. Prior to use, β-mercaptoethanol is added to give a final amount of 5 % (v/v).

1x Running buffer (for SDS-PAGE)

Dissolve 14.4 g of glycine (Merck), 3 g Tris and 1 g of SDS (Merck) in 1 L of water.

1x SB (Sodium borate) buffer

Prepare 0.01 M NaOH (Ajax) in water at pH 8, pH adjusted by boric acid.

Buffer A

Prepare 50 mM Tris-HCl at pH 7.5, 150 mM NaCl, and 0.1 mM $CaCl_2$. The pH is adjusted by HCl.

Buffer B

Prepare 50 mM Tris-HCl at pH 7.5, 150 mM NaCl, 0.1 mM $CaCl_2$ and 200 mM imidazole. The pH is adjusted by HCl.

TMA buffer

Prepare a 2X concentrated buffer by dissolving 100 mM Tris, 50 mM MES and 50 mM acetic acid and adjusting the pH by HCl or $NaOH_{(aq)}$.

Steroid stock solutions

The 10 mM, 5 mM or 1 mM stock solutions of steroids were prepared in HPLC grade MeOH.

Steroid sulfate stock solutions

The 20 mM, 10 mM or 5 mM stock solutions of steroid sulfates were prepared in milliQ water.

Oligonucleotides

Table A1: Oligonucleotide primers used for mutagenesis

Primer No.	Primer sequence
P1-f	GAACTG**(NNS)**GGTATGCGTGCTATCCG
P2-r	CCTTTCAGAAT**(SNN)**CGGGGTAGACTC
P3-r	GAATTTCGG**(SNN)**CGCCTCCAGC
P4-r	CCGGACCGAA**(SNN)**CGGGAACGCC
P5-r	GATCCGGACC**(SNN)**TTTCGGGAAC
P6-f	CTGACCCGTGAGTGGGAGG
P7-f	GACGAGTCTACCCCGCCC**(NYS)**CTGAAAGGTACCC
P8-r	GATCCGGACC**(SRN)**CACCGGGAACGCCTC**(SRN)**CAGCGCACCCTC
P9-f	CGCTG**(NYS)**GAGGCGTTCCCGGTG**(NYS)**GGTCCGGATC
P10-r	CTTTCAG**(SRN)**GGGCGGGGTAG
P11-f	TATGTGGAGGACGAACGCTATCTGGATACT
P12-r	GGGTAGACTCGTCGTACGG
P13-f	GAGCCGCCGTACGACGAGTCTACCCCGCCC**(VTT)(NYS)(NYS)**GGT**(NYS)**CCGGCTCTG
P14-r	CCAGATAGCGTTCGTCCTCCACATACAGAGCCGG**(SRN)**ACC**(SRN)(SRN)(AAB)**GGGCG
P15-f	GGCACCGACCACCATATCGCGGGTATTGGTACC**(NYS)**GCA**(NYS)**GCCTTGAC
P16-r	CGTAACCCGGCTTACCCTCCAATTCTGGGGTCAAGGC**(SRN)**TGC**(SRN)**GGTACCAA
P17-f	CCCAGAATTGGAGGGTAAGCC
P18-r	TACCCGCGATATGGTGGTC
P19-f	GTGCGCTG**(TTS)**GAGGCGTTC
P20-f	CTACCCCG**(CST)(VTT)(STG)(RWA)**GGT**(WCC)**CCGGCTCTGTATG
P21-f	CACAATTCCACAACGGTTTC
P22-r	CGAAAGGCTCAGTCGAAAG
P23-f	CACAACGGTTTCCCTCTAGAAATAATTTTG
P24-r	CTGCCAGGCATCAAATAAAACGAAAG
P25-r	GAGAGAGAGAGAGAGAGAGAGAGAGGCCGATGGTAGTGTGGGGAC
P26-r	GAGAGAGAGAGAGAGAGAGAGAGAGG
P27-f	ATTGGTACCATGGCA**(CAA)**GCCTTGAC
P28-f	ATTGGTACCATGGCA**(GAT)**GCCTTGAC
P29-f	ATTGGTACCATGGCA**(GCA)**GCCTTGAC
P30-f	CTCGAAAATAATAAAGGGAAAATCAG
P31-r	TGGTAGTGTGGGGACTC

The mutated codons are in bold text within brackets.

Table A2: Randomized positions, methods of mutagenesis and oligomers used in each library. Library name abbreviations are specified in the main text.

Library Index	Library name	amino acid positions mutated	Mutagenesis method	Oligomers used
1	SSM 1	F328	Megaprimer method	PCR 1: P21-f / P3-r PCR 2: PCR 1 / P22-r
2	SSM 2	F463	Megaprimer method	PCR 1: P1-f / P24-r PCR 2: P23-f / PCR 1
3	SSM 3	F331	Megaprimer method	PCR 1: P6-f / P5-r PCR 2: P21-f / PCR 1 PCR 3: PCR 1 / P22-r PCR 4: P21-f / PCR 2 + PCR 3 / P22-r
4	SSM 4	K330	Megaprimer method	PCR 1: P6-f / P4-r PCR 2: P21-f / PCR 1 PCR 3: PCR 1 / P22-r PCR 4: P21-f / PCR 2 + PCR 3 / P22-r
5	SSM 5	R155	Megaprimer method	PCR 1: P21-f / P2-r PCR 2: PCR 1 / P22-r
6	MSM 1	I156/L325/F331	Gibson Assembly method	PCR 1: P7-f / P8-r PCR 2: P9-f / P10-r
7	MSM 2	I156/L157/K158/T160	RQ method	PCR 1: P11-f / P12-r Insert: P13-f / P14-r
8	MSM 3	M72/E74	RQ method	PCR 1: P17-f / P18-r Insert: P15-f / P16-r
9	Shuffled	R155/I156/L158/K158/ T160/L325/K330	Megaprimer method	PCR 1: P19-f / P25-r PCR 2: P20-f / PCR 1 / P26-r PCR 3: P23-f / PCR 2 / P26-r

Table A3: The library template and the screening conditions of each library

Library Index	Library name	Template	Screening conditions
1	SSM 1	WT-PaS	100 µM TS, pH 9, 37 °C
2	SSM 2	WT-PaS	100 µM TS, pH 9, 37 °C
3	SSM 3	WT-PaS	100 µM TS, pH 9, 37 °C
4	SSM 4	WT-PaS	100 µM TS, pH 9, 37 °C
5	SSM 5	WT-PaS	100 µM TS, pH 9, 37 °C
6	MSM 1	PV-PaS	100 µM TS, pH 9, 37 °C,
7	MSM 2	PV-PaS	20 µM TS, pH 7, 37 °C, Nandrolone-IS
8	MSM 3	PVFV-PaS	20 µM TS, pH 7, 37 °C, Nandrolone-IS
9	Shuffled	WT-PaS/PV-PaS	20 µM TS, pH 7, 37 °C, Nandrolone-IS

AtsA sequence

Amino acid sequence of native PaS with the additional hexa-histidine tag at the C-terminus The residues subjected to mutagenesis are highlighted in green.

```
         10          20          30          40          50          60
MSKRPNFLVI VADDLGFSDI GAFGGEIATP NLDALAIAGL RLTDFHTAST CSPTRSMLLT

         70          80          90         100         110         120
GTDHHIAGIG TMAEALTPEL EGKPGYEGHL NERVVALPEL LREAGYQTLM AGKWHLGLKP

        130         140         150         160         170         180
EQTPHARGFE RSFSLLPGAA NHYGFEPPYD ESTPRILKGT PALYVEDERY LDTLPEGFYS

        190         200         210         220         230         240
SDAFGDKLLQ YLKERDQSRP FFAYLPFSAP HWPLQAPREI VEKYRGRYDA GPEALRQERL

        250         260         270         280         290         300
ARLKELGLVE ADVEAHPVLA LTREWEALED EERAKSARAM EVYAAMVERM DWNIGRVVDY


        310         320         330         340         350         360
LRRQGELDNT FVLFMSDNGA EGALLEAFPK FGPDLLGFLD RHYDNSLENI GRANSYVWYG

        370         380         390         400         410         420
PRWAQAATAP SRLYKAFTTQ GGIRVPALVR YPRLSRQGAI SHAFATVMDV TPTLLDLAGV

        430         440         450         460         470         480
RHPGKRWRGR EIAEPRGRSW LGWLSGETEA AHDENTVTGW ELFGMRAIRQ GDWKAVYLPA

        490         500         510         520         530         540
PVGPATWQLY DLARDPGEIH DLADSQPGKL AELIEHWKRY VSETGVVEGA SPFLVRHHHH


HH
```

PaS enzyme parameters

Important enzyme (native) parameters of WT-PaS

- Number of amino acids: 536
- Theoretical pI: 5.50
- Extinction coefficient: 102790 M^{-1} cm^{-1}

Beneficial mutations

Table A4: Molecular weight of the beneficial PaS mutants and WT-PaS

	Mutational residue							Molecular weight with the hexa-histidine tag (Da)
PaS enzyme	155	156	157	158	160	325	330	
WT-PaS	R	I	L	K	T	L	K	60768
P-PaS	P	I	L	K	T	L	K	60709
T-PaS	T	I	L	K	T	L	K	60713
V-PaS	R	I	L	K	T	L	V	60739
TV-PaS	T	I	L	K	T	L	V	60684
PV-PaS	P	I	L	K	T	L	V	60680
PVFV-PaS	P	V	L	K	T	F	V	60700
PVIV-PaS	P	I	V	I	T	L	V	60651
VVIV-PaS	R	V	V	I	T	L	V	60696
LEF-PaS	R	L	L	E	T	F	K	60803

Grey highlights: Mutations

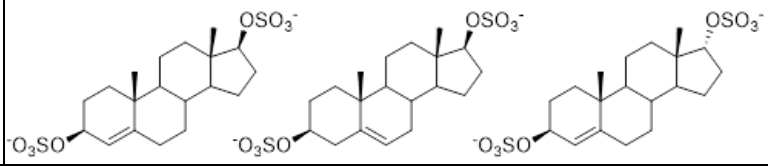
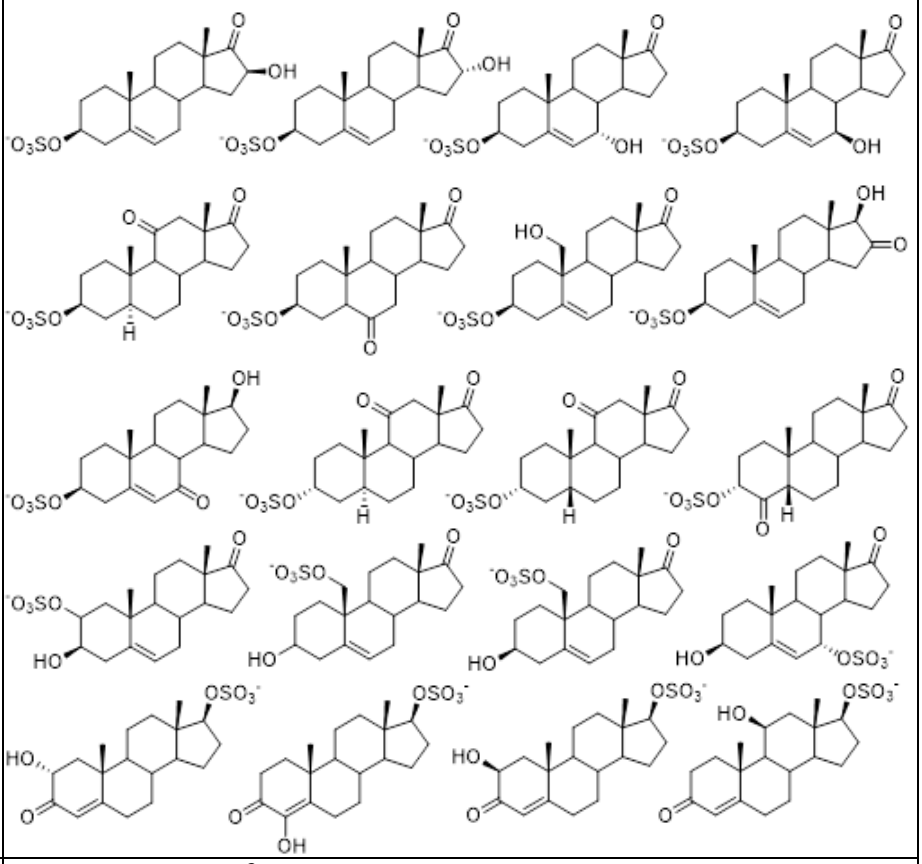
LC-MS parameters

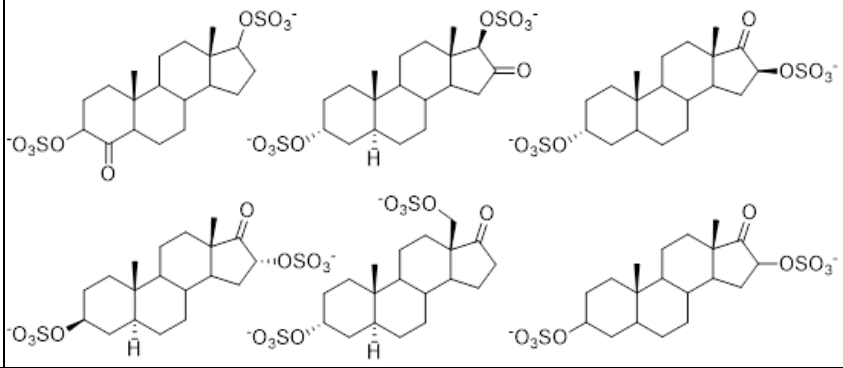
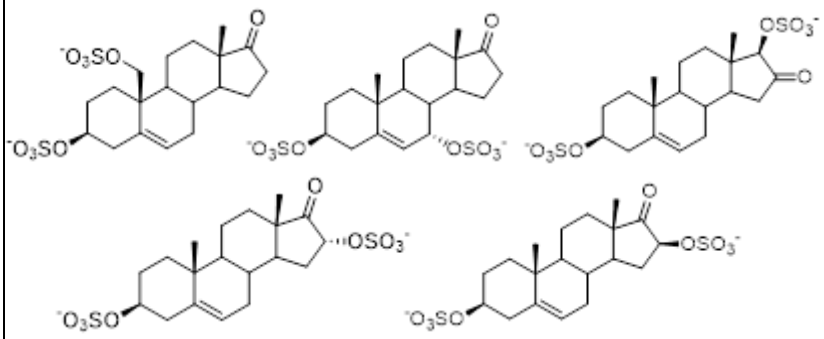
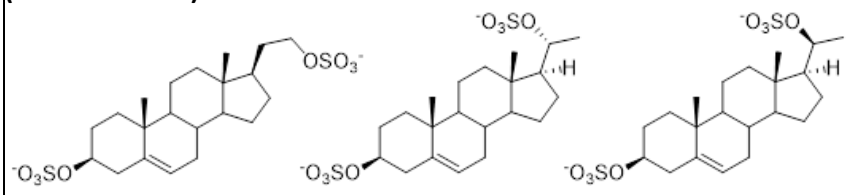
Table A5: Parameters for analyte detection by LC-MS using ESI

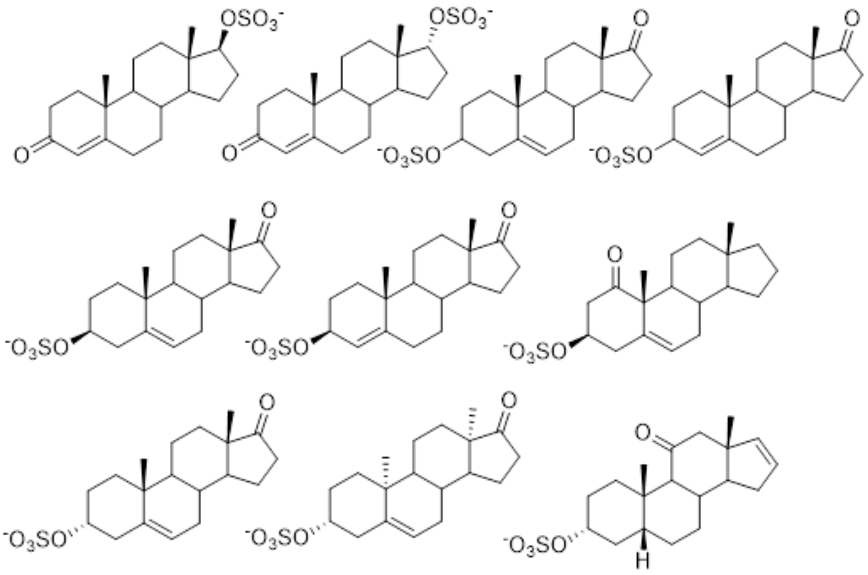
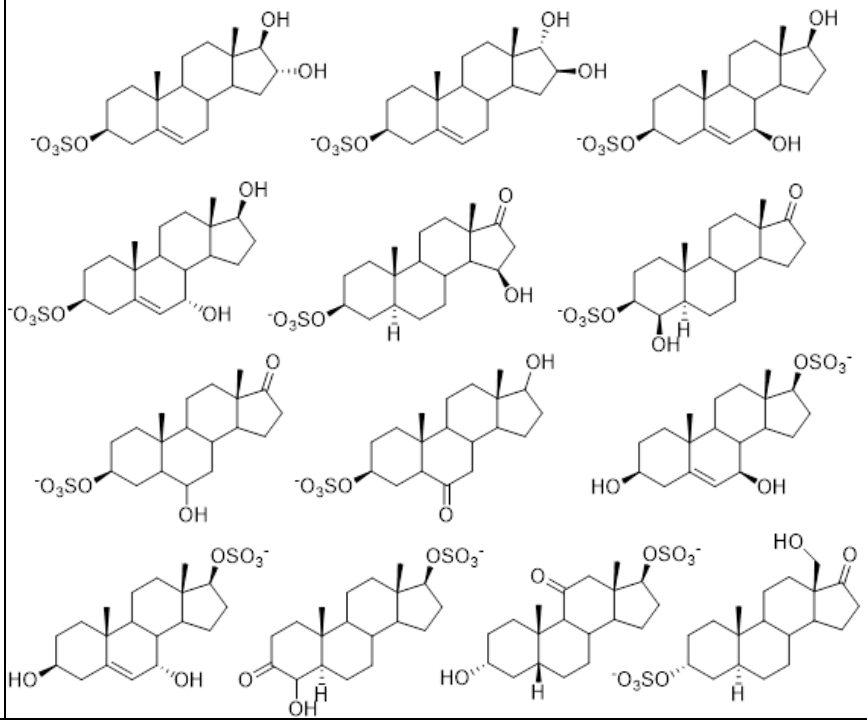
Analyte	Abbr.	Ion	*m/z*	Capillary V	Frag. V
androsterone	A	$[M+H]^+$	291.2	3000	120
androsterone 3-sulfate	AS	$[M-H]^-$	369.2	3000	200
boldenone	B	$[M+H]^+$	287.2	1500	100
boldenone 17-sulfate	BS	$[M-H]^-$	365.1	3000	200
dehydroepiandrosterone	DHEA	$[M+H]^+$	289.2	4000	100
dehydroepiandrosterone 3-sulfate	DHEAS	$[M-H]^-$	367.2	3000	200
epiandrosterone	EA	$[M+H]^+$	291.2	5000	110
epiandrosterone 3-sulfate	EAS	$[M-H]^-$	369.2	3000	200
epitestosterone	ET	$[M+H]^+$	289.2	1500	120
epitestosterone 17-sulfate	ETS	$[M-H]^-$	367.2	3000	200
estrone	E	$[M-H]^-$	269.2	4000	180
estrone 3-sulfate	ES	$[M-H]^-$	349.1	4000	180
etiocholanolone	EC	$[M+H]^+$	291.2	4000	120
etiocholanolone 3-sulfate	ECS	$[M-H]^-$	369.2	3000	200
nandrolone	N	$[M+H]^+$	275.2	1500	100
nandrolone 17-sulfate	NS	$[M-H]^-$	353.1	3000	200
testosterone	T	$[M+H]^+$	289.2	1500	120
testosterone 17-sulfate	TS	$[M-H]^-$	367.2	3000	200
d_3-testosterone	d_3-T	$[M+H]^+$	292.4	1500	120
d_3-testosterone	d_3-TS	$[M+H]^+$	370.4	3000	200

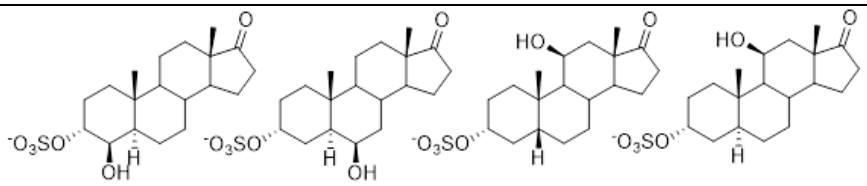
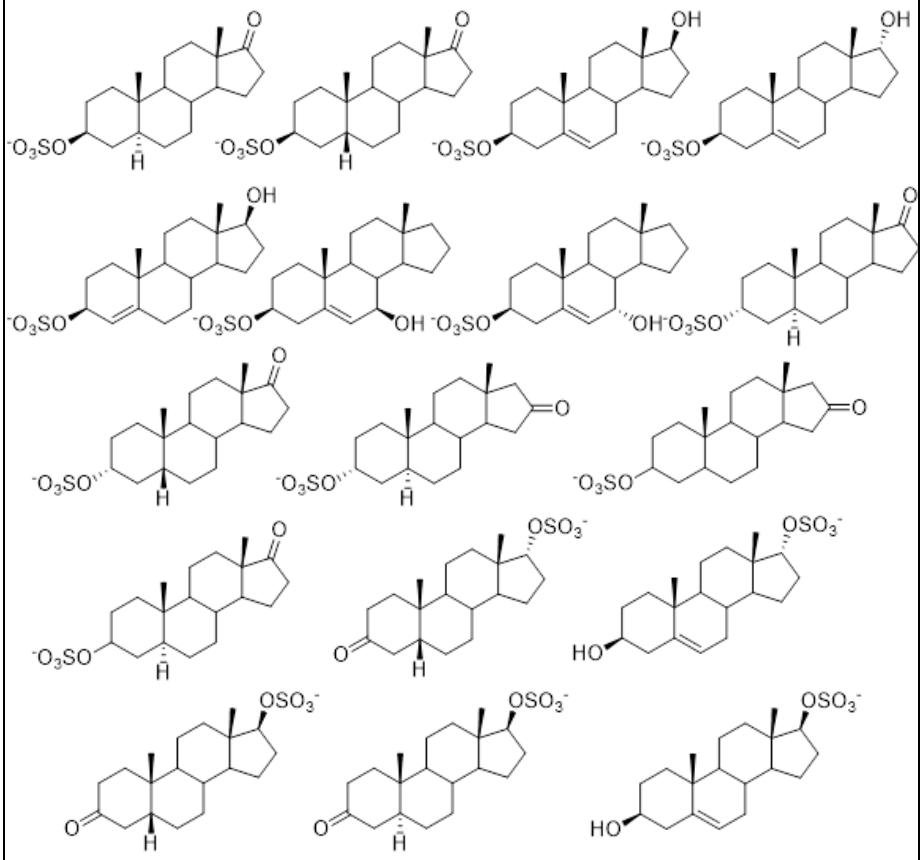
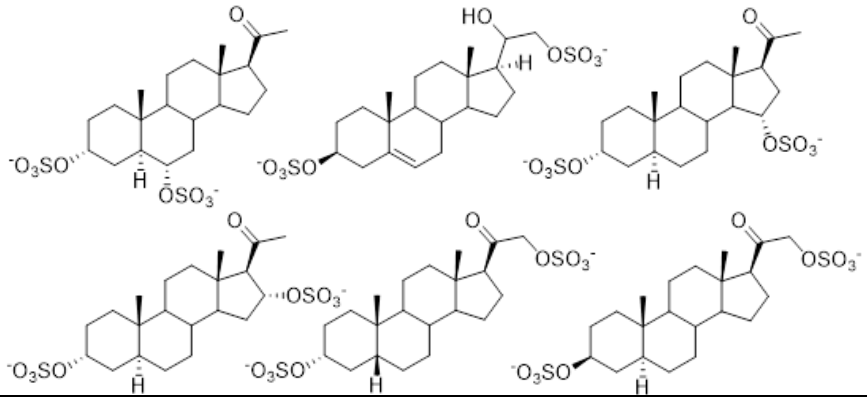
Steroidal structure search

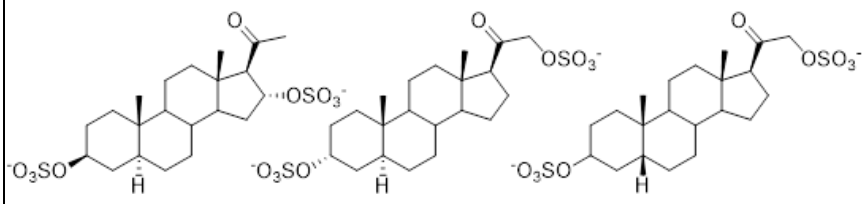
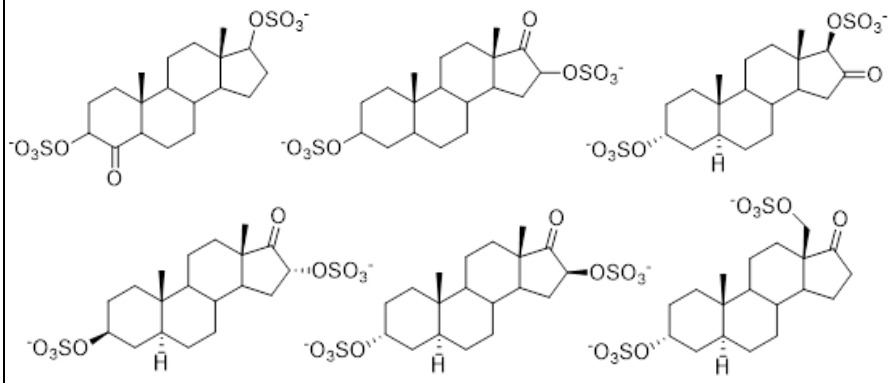
Table A6: The resulted steroidal structures for each possible chemical formula in the female and horse samples of interest

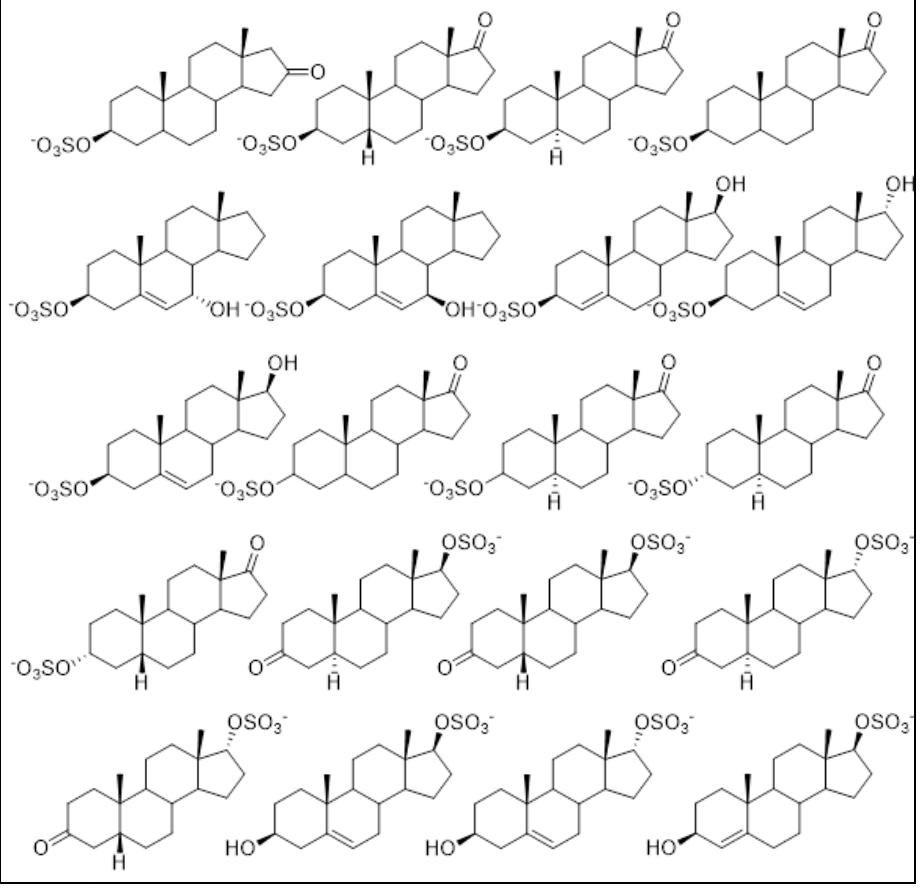
	m/z	Steroidal structure
A1 C7	224.0614 (-2)	**$(C15\ H31\ O8\ N\ P\ S2)^{2-}$** No matches **$(C19\ H28\ O8\ S2)^{2-}$** 
A2 N8 N10	383.1538 (-1)	**$(C12\ H27\ O4\ N6\ S2)^{1-}$** No matches **$(C15\ H30\ O6\ N\ P\ S)^{1-}$** No matches **$(C19\ H27\ O6\ S)^{1-}$** 
C4	232.059	**$(C15\ H31\ O9\ N\ P\ S2)^{2-}$**

N4	(-2)	No matches **(C19 H28 O9 S2) [2-]** 
C6	231.0511 (-2)	**C17 H24 O8 N3 S2** No matches **(C19 H26 O9 S2) [2-]** 
C10	238.0773 (-2)	**(C17 H35 O8 N P S2) [2-]** No matches **(C21 H32 O8 S2) [2-]** 
C11	367.1588 (-1)	**(C13 H28 O4 N4 P S) [1-]** No matches **(C15 H30 O5 N P S) [1-]** No matches **(C17 H25 O4 N3 S) [1-]** No matches

		(C19 H27 O5 S) $^{1-}$ 
N7 N9	384.1475 (-1)	**(C12 H29 O4 N6 S2)** $^{1-}$ No matches **(C15 H32 O6 N P S)** $^{1-}$ No matches **(C19 H29 O6 S)** $^{1-}$ 

		
N11 N12	369.1745 (-1)	**(C15 H32 O5 N P S)** $^{1-}$ No matches **(C19 H29 O5 S)** $^{1-}$ 
M2	246.0749 (-2)	**(C17 H35 O9 N P S2)** $^{2-}$ No matches **(C21 H32 O9 S2)** $^{2-}$ 

		
P27	232.0590 (-2)	**(C13 H29 O8 N4 P S2)** $^{2-}$ No matches **(C14 H30 O8 N3 S3)** $^{2-}$ No matches **(C15 H31 O9 N P S2)** $^{2-}$ No matches **(C17 H26 O8 N3 S2)** $^{2-}$ No matches **(C19 H28 O9 S2)** $^{2-}$ 

P37	369.1743 (-1)	**(C13 H30 O4 N4 P S)** $^{1-}$ No matches **(C14 H31 O4 N3 S2)** $^{1-}$ No matches **(C15 H32 O5 N P S)** $^{1-}$ No matches **(C17 H27 O4 N3 S)** $^{1-}$ No matches **(C19 H29 O5 S)** $^{1-}$ 

*The structures containing ^{35}S, ^{14}C, D and T isotopes are disregarded