Profiling the Steroid Sulfate Metabolome

Christopher Charles Jung-Hua Fitzgerald

A thesis submitted for the degree of

Doctor of Philosophy

May 2022

Research School of Chemistry

The Australian National University

Australian National University

Abstract

The study of androgenic anabolic steroids (AAS) and their metabolites form an important cornerstone of anti-doping analysis in both human and equine sports. With the advancement of ultra-high-performance liquid chromatography-tandem high resolution mass spectrometry (UHPLC-HRMS) instruments and techniques the analytical focus has shifted toward the detection of free steroid metabolites and their intact Phase II conjugates. The development of new tools to assist in this analysis of these compounds is of high priority. The work presented in this thesis aims to develop new methodologies for the detection and characterisation of AAS and their metabolites, with an emphasis on sulfate conjugates. Methodologies presented include a combination of UHPLC-HRMS, informatics and synthetic chemistry applied in a multidisciplinary way over the fields of organic synthesis, analytical chemistry, and metabolomics.

Chapter One presents a summary of the literature detailing the metabolism of AAS, and the past and current methods of detection in anti-doping laboratories. Adjacent techniques from different fields that are relevant or have potential application in anti-doping analysis are also highlighted. This includes aspects such as metabolomics, different high-resolution mass spectrometers, non-targeted acquisition methods, associated data processing software, and intelligence based anti-doping approaches.

Chapter Two presents a new data processing tool for the annotation of sulfated metabolites that utilises UHPLC-HRMS/MS instrumentation with a non-targeted data dependent acquisition (DDA) method. This new tool was applied in two separate scenarios; firstly, to identify potential steroid sulfate biomarkers of testosterone propionate (TP) doping, and secondly, to qualitatively assess the selectivity of various sulfatase enzymes. Key to this method is the use of an automated *k*-means clustering algorithm to identify putative sulfated metabolites. This technique is unique as it also retains all other information in each metabolome assessed, giving the potential for a full metabolomic analysis. In the second part of the chapter, the new analytical tool is extended to identify potential steroid sulfate metabolites of TP doping in the urinary metabolome of three horses. From this a

new semi-targeted steroidomic workflow is established, that highlights several new biomarkers that are likely related to doping.

Chapter Three presents the development of a new type of stable isotope label for sulfate conjugate metabolites. The method is unique as it places the stable isotope label on the sulfate conjugate group. It is also readily accessible following a one-pot synthesis procedure and simple SPE purification on small scales (5-10 mg). Its utility is demonstrated by the synthesis of a range of selectively labelled mono- and bis-conjugated compounds. The bis(sulfate) compounds synthesised are further analysed using MS/MS, pseudo-MS/MS/MS and energy-resolved collision induced dissociation (CID) mass spectrometry experiments, that use the labelled conjugates as probes to elucidate complex fragmentation pathways.

Chapter Four presents research that directly investigates the identity of a new steroid biomarker with energy resolved CID analysis. This chapter represents a confluence of the work from Chapter Two and Three. Specifically, the new biomarker identified was discovered in Chapter Two and the methodology for its elucidation was developed in Chapter Three. It also further develops the energy-resolved CID analysis as a comprehensive analytical technique that is applicable to both *in vitro* and *in vivo* samples, at biologically relevant concentrations.

Chapter Five presents research that extends the untargeted metabolomic techniques developed in this thesis to the analysis of a designer steroid, hemapolin. A targeted analysis of a new class of sulfonate metabolites is established using UHPLC-HRMS/MS DDA analysis. These metabolites can now be incorporated into anti-doping screening and confirmation procedures for future detection of hemapolin misuse. Following this, a semi-targeted metabolomic analysis of the urinary metabolome was also performed to putative identify further long-term metabolites of hemapolin administration.

Declaration

I declare that this thesis and the work presented herein comprises only of my own original work performed between March 2018 and May 2022 towards the completion of the Doctor of Philosophy degree. It contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of the authors knowledge this thesis does not contain material that has been previously published or written by another person, except where due reference and credit is made in the text. This thesis is less than 100,000 words in length exclusive of figure, tables, references, and appendices.

Christopher Charles Jung-Hua Fitzgerald,

May 2022.

Acknowledgements

This thesis would not have been possible without the love and support of so many people. The gratitude and the privilege of time, circumstance, and curiosity I have had that has enabled me to undergo this research to the best of my ability will always shape the person I am. I would not have been able to achieve so much without the support, wisdom, and guidance of so many people.

First and foremost, my supervisor Associate Professor Mal McLeod, who has constantly encouraged me to take my research into new directions, was always ready for an impromptu meeting to hear any new ideas I had, and whose critical thinking and great scientific and chemical expertise never ceases to amaze me. Thank you for all the time editing drafts and guiding me in the right direction over the last four years of research.

Next, to my co-supervisor Dr Adam Cawley, whose friendliness, fellow nerdiness and enthusiasm for all things to do with analytical chemistry was infectious! Thank you for taking me in during my multiple research trips to Sydney, you, and the rest of the team at ARFL made it feel like a second home and I always felt welcome.

The McLeod group who has become like family over the years, I will always be grateful for the friendly faces and the general banter offered, even during the hardest times in my research.

To my Family, Friends, and Jess, all of you have been tremendous and I thank you for letting me, constantly and enthusiastically, talk ‘your ears off’ about my research. Thank you for being there over the years.

Lastly, to the Research School of Chemistry technical staff and joint mass spectrometry facility staff. All of you, and your expertise, has allowed me to grow professionally and made me a better scientist.

Publications and other works

Publications arising from research performed toward the completion of this degree:

Fitzgerald, C. C. J. et al. Profiling Urinary Sulfate Metabolites With Mass Spectrometry. *Front. Mol. Biosci*. **9**, 829511 (2022). https://doi.org/10.3389/fmolb.2022.829511

Fitzgerald, C. C. J. & McLeod, M. D. Synthesis of stable isotope labelled steroid bis(sulfate) conjugates and their behaviour in collision induced dissociation experiments. *Org. Biomol. Chem.* **20**, 3311–3322 (2022).

Fitzgerald, C. C. J., Bowen, C., Elbourne, M., Cawley, A., and McLeod, M. D. Energy-Resolved Fragmentation Aiding the Structure Elucidation of Steroid Biomarkers. *Journal of the American Society for Mass Spectrometry*.**2022**. Submitted for review.

Other works that the author has been involved in, however, are not a subject of this thesis:

Pranata, A. and Fitzgerald, C. C. J. *et al.* Synthesis of steroid bisglucuronide and sulfate glucuronide reference materials: Unearthing neglected treasures of steroid metabolism. *Steroids* **143**, 25–40 (2019).

Pozo, O. J. and Fitzgerald, C. C. J. *et al.* Alternate steroid sulfation pathways targeted by LC–MS/MS analysis of disulfates: Application to prenatal diagnosis of steroid synthesis disorders. *J. Mol. Endocrinol.* **61**, M1–M12 (2018).

Table of Contents

Abstract i
Declaration iii
Acknowledgements v
Publications and other works vii
Table of Contents viii
List of Figures xi
List of Tables xv
List of Abbreviations xvii

CHAPTER ONE 1

1.1 Introduction ***3***
1.1.1 Anabolic androgenic steroids 3
1.1.2 Use of AAS for performance in human and equine sports 4
1.1.3 Biosynthesis of AAS in humans and horses 6
1.1.4 Metabolism of AAS in Humans and Horses 7
1.1.5 Instrumentation for the detection of AAS 9
1.1.6 Gas chromatography-mass spectrometry 10
1.1.7 Liquid chromatography mass spectrometry 11
1.1.8 Mass analysers 13
1.1.9 Low resolution mass analysers 14
1.1.10 High resolution mass analysers 16
1.1.11 Metabolomic profiling 19
1.1.12 Untargeted metabolomic workflows 22
1.1.13 Data acquisition in untargeted metabolomic profiling workflows 22
1.1.14 Data dependant acquisition 24
1.1.15 Data independent acquisition 25
1.1.16 Data Processing and Informatic Tools for Metabolomic Workflows 27
1.1.17 Best practices in untargeted metabolomics 29
1.1.18 Metabolomic detection of AAS in human and equine sports 30
1.1.19 Difficulties in untargeted metabolomic profiling 34
1.1.20 The Equine Biological Passport 34
Thesis Aims ***37***
1.2 References ***38***

CHAPTER TWO ***47***

2.1 Preface ***49***
2.1.1 Specific contributions 50
2.1.2 Summary 51
2.2 Profiling Urinary Sulfate Metabolites with Mass Spectrometry ***55***
2.3 Semi-targeted profiling of in vivo Phase II urinary metabolites of testosterone propionate in thoroughbred geldings ***71***
2.3.1 Introduction 71
2.4 Results and Discussion ***72***
2.4.1 Detection and ionisation of testosterone propionate metabolites in LC-MS analysis 73
2.4.2 Confirmations of reference materials in equine urine 73
2.4.3 Targeted individual kinetic excretion profiles of reference materials in equine urine 76
2.4.4 Semi-targeted kinetic excretion profile of a minor metabolite of testosterone propionate 80
2.4.5 Discussion of the targeted analysis 82
2.4.6 Semi-targeted metabolomic analysis of testosterone propionate administration samples; *a summary* 82
2.4.7 High throughput analysis of equine urinary metabolites 83
2.4.8 Data quality assessment and PCA analysis 83
2.4.9 OPLS-DA model establishment and validation 89

2.4.10 Exploratory discriminant analysis of the equine urinary metabolome with a focus on Phase II steroidal metabolites 94
2.4.11 Identifications of steroidal biomarkers 101
2.5 Conclusion ***118***
2.5.1 Future directions and commentary 118
2.6 Experimental ***120***
2.6.1 Animal Administration and ethics approval. 120
2.6.2 UHPLC-HRMS DDA analysis for equine urine samples 120
2.6.3 Parameters for CE spread and DDA MS experiments 121
2.6.4 Extraction of equine urine samples for UHPLC-HRMS DDA analysis 121
2.6.5 Untargeted profiling of equine urine and semi-targeted metabolite confirmation studies 122
2.6.6 Linearity 123
2.6.7 Recovery 123
2.6.8 Repeatability 124
2.6.9 Reproducibility 124
2.6.10 Sensitivity 125
2.6.11 Accuracy and precision 125
2.7 Synthesis of steroid diol monosulfates ***125***
2.7.1 Androsterone-3-sulfate, ammonium salt (5) 126
2.7.2 5α-Androstane-3α,17β-diol 3-sulfate, ammonium salt (6) 126
2.7.3 Etiocholanolone 3-sulfate, ammonium salt (7) 127
2.7.4 5β-Androstane-3α,17β-diol 3-sulfate, ammonium salt (8) 127
2.7.5 Dihydrotestosterone 3-sulfate, ammonium salt (9) 128
2.7.6 5α-Androstane-3β,17β-diol 17-sulfate, ammonium salt (10) 128
2.7.7 Epidihydrotestosterone 3-sulfate, ammonium salt (11) 129
2.7.8 5α-Androstane-3α,17α-diol 17-sulfate, ammonium salt (12) 129
2.7.9 17α-Acetoxy-5α-androstan-3-one (13) 130
2.7.10 17α-Acetoxy-5α-androstan-3β-ol 3-sulfate, ammonium salt (14) 130
2.7.11 5α-Androstane-3β,17α-diol 17-sulfate, ammonium salt (15) 131
2.7.12 17β-Acetoxy-5β-androstan-3-one (16) 132
2.7.13 17β-Acetoxy-5β-androstan-3α-ol 3-sulfate, ammonium salt (17) 132
2.7.14 5β-Androstane-3α,17β-diol 3-sulfate, ammonium salt (18) 133
2.7.15 17α-[(4-nitrobenzoyl)oxy]-5β-Androstan-3-one (18) 134
2.7.16 17α-Hydroxy-5β-androstan-3-one (19) 134
2.7.17 17α-Hydroxy-5β-androstan-3-one 17-sulfate, ammonium salt (20) 135
2.7.18 5β-Androstane-3α,17α-diol 17-sulfate, ammonium salt (21) 136
2.7.19 17α-Acetoxy-5β-androstan-3-one (22) 136
2.7.20 17α-Acetoxy-5β-androstan-3α-ol 3-sulfate, ammonium salt (23) 137
2.7.21 5β-Androstane-3α,17α-diol 3-sulfate, ammonium salt (24) 138
2.8 References ***139***

CHAPTER THREE

143
3.1 Preface ***145***
3.1.1 Summary 146
3.2 Synthesis of stable isotope labelled steroid bis(sulfate) conjugates and their behaviour in collision induced dissociation experiments ***147***
3.3 Future directions and commentary ***161***
3.4 References ***163***

CHAPTER FOUR

165
4.1 Preface ***167***
4.1.1 Summary 168
4.2 Energy-Resolved Fragmentation Aiding the Structure Elucidation of Steroid Biomarkers ***171***
4.3 Future directions and commentary ***179***
4.4 References ***181***

CHAPTER FIVE
183
5.1 Introduction ***185***
5.1.1 Designer steroids in sports and the world 185
5.1.2 Hemapolin: an episulfide designer steroid 185
5.1.3 LC-MS analysis of equine urinary metabolites of Hemapolin 187
5.1.4 *Aims: In vivo* metabolism of hemapolin and epistane in equine 189
5.2 Results and Discussion ***190***
5.2.1 Methodological design: semi-targeted profiling, a global and targeted analysis 190
5.2.2 Detection and ionisation of hemapolin metabolites in LC-MS analysis 191
5.2.3 Targeted individual kinetic excretion profiles in equine urine 191
5.2.4 Untargeted metabolomics to highlight potential biomarkers of hemapolin administration 198
5.2.5 Approach 1: Non-targeted metabolomics analysis of hemapolin administration samples 199
5.2.6 Data quality assessment and pre-treatment 199
5.2.7 Multivariate analysis and OPLS-DA model establishment 201
5.2.8 Identification of potential biomarkers of hemapolin administration 205
5.2.9 Biomarker annotation of significant features 216
5.2.10 Discussion of Approach 1 219
5.2.11 Approach 2: Semi-targeted steroidomic analysis of hemapolin administration samples 220
5.2.12 Steroid searcher algorithm and steroidomic OPLS-DA analysis 221
5.2.13 Steroidomic analysis of hemapolin administration 227
5.2.14 Results of OPLS-DA and biomarker annotation 236
5.2.15 Discussion of Approach 2 238
5.3 Conclusion ***240***
5.3.1 Future directions and commentary 240
5.4 Experimental ***242***
5.4.1 Animal Administration 242
5.4.2 UHPLC-HRMS DDA analysis for equine urine samples 242
5.4.3 Parameters for CE spread and DDA MS experiments 242
5.4.4 Extraction of equine urine samples for UHPLC-HRMS DDA analysis 243
5.4.5 Untargeted profiling of equine urine and semi-targeted metabolite confirmation studies 244
5.4.6 Linearity 245
5.4.7 Recovery 245
5.4.8 Repeatability and reproducibility 245
5.4.9 Sensitivity 246
5.4.10 Accuracy and precision 246
5.5 References ***247***
CHAPTER SIX
251
6.1 Conclusion ***253***
6.2 Future directions and commentary ***256***
6.3 References ***260***

List of Figures

Figure 1.1: Testosterone propionate (TP) showing steroid ring structure and IUPAC recommended ring lettering and atom numbering. The rings of the steroid are depicted in the plane of the paper. Atoms attached to the ring are denoted alpha (α) if it is below and beta (β) if it is above the plane of the paper.[1] **3**

Figure 1.2: Biosynthesis of androgens and oestrogens via the Δ^5 and Δ^4 pathways in mammalian systems. Adapted from multiple sources.[10,16,17] **7**

Figure 1.3: Phase I and Phase II metabolic pathways for testosterone. Adapted from Schänzer *et al* (1992).[19] **8**

Figure 1.4: General differences between Phase I and II metabolism in humans and equines. Adapted from Scarth *et al* (2011).[10] **9**

Figure 1.5: Types of common stationary phases used in UHPLC to separate steroids and their metabolites.[23] **12**

Figure 1.6: Instrumental schematic of electrospray ionisation and formation of a Taylor cone and liquid jet in negative mode. Image adapted from Crotti *et al* (2011).[27] **13**

Figure 1.7: Schematic of a triple quadrupole (QQQ) mass analyser equipped with a collision cell in quadrupole 2 (Q2). Adapted from Crotti *et al* (2011).[27] **14**

Figure 1.8: Schematic of a linear ion trap. Ions are held in a 'potential well' between 4 hyperbolic shaped rods divided into three sections. Adapted from Schwartz *et al* (2002).[30] **15**

Figure 1.9: Schematic of a single reflector ToF mass analyser. The red ion-trace has a higher initial velocity which penetrates further into the reflectron, ultimately giving it a longer flight path. It then arrives at the detector at the same time as the blue ion-trace which had a lower initial velocity. Image adapted from multiple sources.[26,38] **17**

Figure 1.10: Schematic of an Orbitrap. The black line is the typical spiral trajectory of an ion along the z- and r-axis of the spindle electrode. Adapted from Makarov *et al* (2015).[34] **18**

Figure 1.11: The relationship between the number of detected metabolites and the reliability of quantification in untargeted and targeted metabolomic assays. Adapted from Cajka and Fiehn (2016).[42] **20**

Figure 1.12: Overview of DDA-MS and DIA-SWATH techniques in mass spectrometry. A) In DDA-MS the '*n*' most abundant precursor ions in each MS1 survey scan are selected and fragmented to give MS/MS data. B) In DIA- SWATH, the survey scan is divided up into pre-defined mass isolation windows of either a constant or variable *m/z* width. All precursors in each isolation mass window are then fragmented to give MS/MS data following deconvolution. The was adapted from multiple sources.[84,85] **23**

Figure 1.13: Schematic of a broadband MS^E DIA method. Co-eluting molecules (blue and green) are simultaneously fragmented at high and low energies. The resulting MS/MS spectra is a complex mixture of precursor and product ion pairs of both molecules. Adapted from Plumb *et al* (2006).[95] **26**

Figure 1.14: A) Schematic description of how pooled quality control (pQC) samples are prepared. B) The clustering of pooled QC samples relative to all other samples in PCA, indicating good data quality. Adapted from Broadhurst *et al* (2018).[118] **30**

Figure 1.15: Constant ion loss (CIL) method for the untargeted detection of steroid bis(sulfate) metabolites. Adapted from McLeod *et al* (2017).[131] **33**

Figure 1.16: Schematic of the proposed forensics intelligence model for the selection of EBP horses. Omics would be used to discover novel biomarkers relevant for doping detection. Adapted from Cawley *et al* (2017),[138] and originally based on a model designed by Schumacher *et al* (2012) for EPO and blood doping.[139] **36**

Figure 2.1: Kinetic excretion profiles of TS (1) in three horses following the administration of testosterone propionate. Error bars represent standard deviation, *n* = 3. **77**

Figure 2.2: Kinetic excretion profiles of EAS (2) in three horses following the administration of testosterone propionate. Error bars represent standard deviation, *n* = 3. **77**

Figure 2.3: Kinetic excretion profiles of TSD (4) in three horses following the administration of testosterone propionate. Error bars represent standard deviation, *n* = 3. **81**

Figure 2.4: *k-means* clustering plots for horses A) H1, B) H3, and C) H5. **87**

Figure 2.5: Multivariate analysis. PCA of horses H1 (A), H3 (B) and H5 (C), with 5 groupings, pre-administration samples (PreT, blue), 1-2 days post-administration (Early_Post, pink), 3-18 days post-administration (Late_Post, green), non-doped control samples (PreT_Con, light-blue), and the pooled QC samples (pQC, purple). Ellipses indicate the 95 % CI interval of the groupings. **89**

Figure 2.6: A-C) OPLS-DA score plot showing the separation between PreT (green) and Early_Post (pink) samples. D-F) OPLS-DA score plot showing the separation between Early_Post (pink) and Late_Post (green) samples, in H1, H3 and H5, respectively. Ellipses indicate the 95 % CI interval of the groupings. **93**

Figure 2.7: S-plots of the 6 contrasts modelled by OPLS-DA. In general, the x-axis (p[1]) represents the covariance of the normalised abundance of each feature, and the y-axis (p(corr)[1]) represents the measure of correlation of each feature to the model. A-C) PreT and Early_Post contrasts, D-F) Early_Post and Late_Post contrast, for H1, H3 and H5, respectively. **95**

Figure 2.8: Kinetic excretion profiles. A) Testosterone sulfate. B) T3S, in H1, H3 and H5. Inserts show known or putative structures. Response ratios indicates LOWESS normalised area divided by the IS area (nandrolone sulfate). Error bars represent standard error, SE = $\sigma/(n)^{1/2}$, *n = 3.* * p-value < 0.01, compared to response ratio at time 0 h. **113**

Figure 2.9: Kinetic excretion profiles of metabolites found that were changed in two of three horses H1, H3 and H5. Putative assignments are as follows: A) estrone sulfate, B) T7S, C) T8S, D) pregnenolone sulfate (as enol), E) unknown. Inserts show putative structures (if available). Black triangles show excretion profiles of molecules that were not found to be a significant biomarker by the OPLS-DA analysis, however, were still present based on *m/z* and RT matching. Response ratios indicates LOWESS normalised area divided by the IS area (nandrolone sulfate). Error bars represent standard error, SE = $\sigma/(n)^{1/2}$, *n = 3.* * p-value < 0.01, compared to response ratio at time 0 h. **115**

Figure 2.10: Possible fragmentation mechanisms for the decomposition of estrone sulfate under CID-CES conditions. The product ions shown were taken from the MS/MS spectra of putatively assigned estrone sulfate biomarker discovered in Horse 5, H5. A) charge-directed decomposition. B) cycloreversion pathway. C) Other possible product ion structures. Adapted from Attygalle *et al* (2001)[62] and Cole *et al* (2012).[63] **117**

Figure 3.1: Selective incorporation of stable isotope labelled sulfate esters in steroidal systems allows for the investigation of fragmentation patterns using collision induced dissociation MS/MS experiments. Original artwork by Mr. Fitzgerald. **146**

Figure 4.1: Energy resolved CID utilised for structure determination to guide the synthesis of a new steroidal reference material. Original artwork by Mr. Fitzgerald. **169**

Figure 5.1: Proposed pathways for urinary Phase I metabolites of hemapolin in the horse. E, elimination; C, Cytochrome P450 oxidation; O, oxidation. Adapted from *Waller et al* (2020).[6] **186**

Figure 5.2: Structure of Epitiostanol and pro-drug mepitiostane.[8] **187**

Figure 5.3: Proposed metabolic pathways for the formation of sulfonate metabolites from the major designer steroid drug hemapolin and its minor component epistane. Adapted from Weththasinghe (2020).[10] 188

Figure 5.4: Synthesised sulfonate reference materials of potential metabolites of hemapolin and epistane.[10] 189

Figure 5.5: Kinetic excretion profiles of synthesised reference materials plotting concentration (ng/mL) over time (h) A) SW72; B) SW76; C) E2; and D) E4. Error bars represent standard deviation, n = 3. Note: times extend to 168 h, however, are only shown to 80 h for clarity. 197

Figure 5.6: Proposed untargeted workflows to identify biomarkers potentially associated with the administration of a dopant in negative mode. A) *Approach 1*; A wholly non-targeted approach for analysis of equine urine data. B) *Approach 2*; A semi-targeted approach that incorporates exact mass matching for Phase I metabolic transformations of madol and steroids from the steroidogenesis pathways. Note: that *Approach 2* emulates the steps taken in *Approach 1*, unless otherwise indicated. 199

Figure 5.7: The *k*-means clustering plot contrasting MA against IR of sulfated (571) and non-sulfated (3348) features in the negative mode screen, total features, n = 3919. 200

Figure 5.8: Principal Component Analysis (PCA) built from the linear combination of the total kinetic excretion profile (Pre, *n* = 3 green: Post, *n* = 30, pink) as well as a control horse (Con, *n* = 3, blue), and pooled QC samples (pQC, *n* = 12, light blue), with total samples, *n* = 48. A) PCA score plot of the negative mode screen. B) PCA score plot of the positive mode screen. Ellipses indicate the 95 % CI interval of the groupings. 202

Figure 5.9: The OPLS-DA plots generated between pre-administration and post administration samples with associated S-plots. Post administration samples, *red*. Pre-administration samples, *green*. A) OPLS-DA score plot of positive mode. B) S-plot of positive mode features (n = 2499).C) OPLS-DA score plot of negative mode D) S-plot of negative mode samples (n = 2006). Ellipses indicate the 95 % CI interval of the groupings. 204

Figure 5.10: Positive mode kinetic excretion profiles of the top 10 significant molecules identified in OPLS-DA analysis of hemapolin administration. Norm. Int. = LOWESS normalised intensity. The error bars show standard error, SE = σ/(n)1/2, n = 3. The error bars are shown, however, are obscured due to their small size. The asterisk * indicates a p-value < 0.01, relative to time 0 h. Note; all non- 0h samples were significant. 217

Figure 5.11: A) Negative mode kinetic excretion profiles of the top 10 significant molecules identified in OPLS-DA analysis of hemapolin administration. B) A putative steroid glucuronide metabolite, A10G, that matched theoretical steroid masses in Tables 5.4 and 5.5. Norm. Int. = LOWESS normalised intensity. The error bars show standard error, SE = σ/(n)1/2, n = 3. The error bars are shown, however, are obscured due to their small size. The asterisk * indicates a p-value < 0.01, relative to time 0 h. Note; all non- 0h samples were significant. 218

Figure 5.12: Principal Component Analysis (PCA) built from the linear combination of the total kinetic excretion profile (Pre, *n* = 3, green: Post, *n* = 30, pink) as well as a control horse (Con, *n* = 3, blue), and pooled QC samples (QC, *n* = 12, light blue), with total samples, *n* = 48. A) PCA score plot of the negative mode screen. Features based on aligned list of theoretical steroid biomarkers. B) PCA score plot of the positive mode screen). Ellipses indicate the 95 % CI interval of the groupings 223

Figure 5.13: The OPLS-DA plots generated between pre- and post-administration samples with associated S-plots. Features based on the aligned list of theoretical steroid biomarkers. Post administration samples, *red*. Pre-administration samples, *green*. A) OPLS-DA plot of positive mode base. B) S-plot of positive mode features (*n* = 260).C) OPLS-DA plot of negative mode. D) S-plot of negative mode samples (*n* = 106). Ellipses indicate the 95 % CI interval of the groupings. 225

Figure 5.14: Kinetic excretion profiles of 13 of the 14 significant molecules identified in OPLS-DA analysis of the positive mode screen. Norm. Int. = LOWESS normalised intensity. The error bars show standard error, SE = σ/(n)1/2, n = 3. The error bars are shown, however, are obscured due to their small size. The asterisk * indicates a p-value < 0.01, relative to time 0 h. Note; all non 0 h samples were significant. 233

Figure 5.15: Kinetic excretion profiles of the 10 significant molecules identified in OPLS-DA analysis of negative mode. Norm. Int. = LOWESS normalised intensity. The error bars show standard error, SE = σ/(n)1/2, n = 3. The error bars are shown, however, are obscured due to their small size. The asterisk * indicates a p-value < 0.01, relative to time 0 h. Note; all non 0 h samples were significant. 235

Figure 5.16: Proposed product ions of androstene sulfate, *m/z* 369.1734, RT 7.348 min under CID conditions with the CE spread (CES) function set at 57 ± 30 eV. 237

Figure 5.17: Proposed pathways for urinary Phase I & Phase II metabolites of hemapolin in the horse. E, elimination; C, Cytochrome P450 oxidation; O, oxidation, SULTs, sulfotransferase, UGT, UDP-glucuronosyltransferase. Red, blue, and purple structures indicate biomarkers found in negative, positive mode, or both screens, respectively. *O-Gluc*, represents a glucuronide moiety, $C_6H_8O_7$. Indicates that it could be from multiple Phase I metabolites, with only one example shown. New Phase I metabolites are shown in boxes. Metabolite A4 is not shown. Figure adapted from Waller *et al* (2020).[6] 239**

List of Tables

Table 1.1: List of DIA and DDA compatible open-source data processing packages. This represents a non-exhaustive search and is based on a similar search reported in a review by Fenaille *et al* (2017).[86] 28

Table 2.1: Confirmation of standard reference materials to their metabolites derived from the sulfate fraction of equine urine (H2), using UHPLC-HRMS/MS CES CID experiments. Each pair was tested against AORC and WADA criteria.[27,28] The CES ranges varied for each compound; testosterone sulfate (1), CES 60 ± 20. Epiandrosterone sulfate (2), CES 60 ± 30. 5α-Androstane-3β,17α-diol 3-sulfate (3), CES 60 ± 30. Androst-4-ene-3β,17β-diol 17-sulfate (4), CES 60 ± 20. Product ions were matched with a mass tolerance of ± 10 ppm and relative abundance was assessed by the peak area of the extracted ion chromatograms. 75

Table 2.2: Retention time, repeatability, reproducibility, recovery at mid- and high-range calibrators for different steroidal refence standards in surrogate blank equine urine. 79

Table 2.3: The retention time of TSD (4) in surrogate stripped matrix (reference material) and as the in vivo metabolite in horse urine. Accurate mass matches were done at ± 5 ppm, allowable retention time differences were calculated from the AORC criteria.[28] 81

Table 2.4: Descriptive statistics for *k-means* analysis for the detected features in the H1, H3, and H5 urine samples. 85

Table 2.5: OPLS-DA cross validation, permutation results, and S-Plot feature selection criteria. 91

Table 2.6: A list of reference *m/z* values related to endogenous steroid conjugates taken from steroidal biosynthetic pathways in mammalian systems. 97

Table 2.7: A list of reference *m/z* values related to possible Phase I transformations of Testosterone adapted from Section 2.2.[1] 99

Table 2.8: The list of matched *m/z* for molecules identified by the S-plots. The MS/MS spectra from the raw spectra are also shown. 103

Table 5.1: *Confirmation using UHPLC retention time and MS/MS criteria*. Urinary metabolites were derived from the sulfate fraction of equine urine after administration hemapolin (H) were matched against synthesised reference materials. Confirmation was performed according to the Association of Official Racing Chemists (AORC) retention time and MS/MS criteria.[9] All CID product ion spectra used for confirmation were acquired using a collision energy (CE) of 20 eV for positive mode and 30 eV for negative mode. All MS/MS confirmations are taken from the 4 h sample post administration. The relative abundance of precursor and product ions was derived from the peak area of a single extracted ion chromatogram of the three replicates for the reference material and metabolite, mass filters were set at a tolerance of ± 5 ppm. 193

Table 5.2: Retention time, repeatability, reproducibility, recovery at mid-range calibrators for different steroidal refence standards in surrogate blank equine urine 195

Table 5.3: A list of reference *m/z* related to possible Phase I transformations of hemapolin (H) and madol (M), up to four transformations were considered. 207

Table 5.4: A list of reference m/z values related to endogenous steroids taken from steroidal biosynthetic pathways in mammalian systems and related studies. 213

Table 5.5: The list of matched *m/z* for molecules identified by the S-plots. The MS/MS data from the raw spectra are also shown. 215

Table 5.6: The 13 molecules identified by the S-plots in the positive mode screen. The MS/MS spectra from the raw spectra are also shown. 229

Table 5.7: The 10 molecules identified by the S-plots in the negative mode screen. The MS/MS spectra from the raw spectra are also shown. 231

List of Abbreviations

Anabolic Androgenic Steroids (**AAS**)
Adverse Analytical Findings (**AAF**)
All-Ion Fragmentation (**AIF, MS^{All}, MS^{E}**)
Atypical Findings (**ATF**)
Athlete Biological Passport (**ABP**)
Australian Racing Forensics Laboratory (**ARFL**)
Association of Official Racing Chemist (**AORC**)
Association of Racing Commissioners International (**ARCI**)
Atmospheric Pressure Chemical Ionisation (**APCI**)
Atmospheric Pressure Ionisation (**API**)
Chemical Entities of Biological Interest (**ChEBI**)
Collision Energy (**CE**)
Collision Energy Spread (**CES**)
Collision Induced Dissociation (**CID**)
Constant Ion Loss (**CIL**)
Constant Neutral Loss (**CNL**)
Cytochrome P450 (**CYP**)
Data Dependent Acquisition (**DDA**)
Data Independent Acquisition (**DIA**)
Direct Current (**DC**)
Electron Ionisation (**EI**)
Electrospray Ionisation (**ESI**)
Endogenous Androgenic Anabolic Steroids (**EAAS**)
Equine Biological Passport (**EBP**)
Escherichia coli (***E. coli***)
Federation Equestrian Internationale (**FEI**)
Food and Drug Administration (**FDA**)
Fourier Transform Ion Cyclotron Resonance (**FT-ICR**)
Full Width at Half Maximum (**FWHM**)
Gas Chromatography (**GC**)
(Ultra) High Performance Liquid Chromatography (**(U)HPLC**)
Higher Energy C-trap Dissociation (**HCD**)
High Resolution (**HR**)
International Conference of Racing Analyst and Veterinarians (**ICRAV**)
International Federation of Horseracing Authorities (**IFHA**)
International Union of Pure and Applied Chemistry (**IUPAC**)
Ion Exchange Chromatography (**IEX**)
Intensity Ratio (**IR**)
Locally Weighted Scatterplot Smoothing (**LOWESS**)
Low Resolution (**LR**)
Liquid Chromatography (**LC**)
Liquid-Liquid Extraction (**LLE**)
Maximum Abundance (**MA**)
Mass to Charge ratio (***m/z***)
Mass Spectrometry (**MS**)
Molecule (**M**)
Multiple Reaction Monitoring (**MRM**)
Multiple Stage Mass Spectrometry (**MS^{n}**)
National Association of Testing Authorities (**NATA**)
National Measurement Institute (**NMI**)
Normalised Collision Energy (**NCE**)
Normal Phase Hydrophilic Interaction Liquid Chromatography (**HILIC**)
Nuclear Magnetic Resonance (**NMR**)
Orthogonal Projections to Latent Structures Discriminant Analysis (**OPLS-DA**)
Para-nitrophenol (**PNP**)
Partial Least Squares Regression (**PLS**)
Performance and Image Enhancing Drugs (**PIEDs**)
Principal Component Analysis (**PCA**)
Principal Component (**PC**)
Protein Precipitation (**PP**)
3'-Phosphoadenosine-5'-Phosphosulfate (**PAPS**)
Pooled Quality Control (**pQC**)
Quadrupole (**Q**)
Quality Assurance (**QA**)
Quality Control (**QC**)
Relative Abundance (**RA**)
Ribonucleic Acid (**RNA**)
Single-Ion Monitoring (**SIM**)
Selected Reaction Monitoring (**SRM**)
Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra (**SWATH-MS**)
Solid Phase Extraction (**SPE**)
Stable Isotope Label (**SIL**)
Sulfotransferases (**SULTs**)
Tandem Mass Spectrometry (**MS/MS**)
Testosterone-Epitestosterone Ratio (**T/E**)
Testosterone Propionate (**TP**)
Thin Layer Chromatography (**TLC**)
Time of Flight (**ToF**)
Trimethylsilyl (**TMS**)
Triple Quadrupole (**QQQ**)
uridine 5'-diphospho-glucuronosyltransferases (**UGTs**)
Uridine Diphosphate (**UDP**)
Ultraviolet (**UV**)
Variable Data Independent Acquisition (***v*DIA**)
Weak Anion Exchange (**WAX**)
World Anti-doping Agency (**WADA**)

CHAPTER ONE

Metabolomics, Mass Spectrometry, Steroids and Anti-Doping.

1.1 Introduction

1.1.1 Anabolic androgenic steroids

Steroids are a broad class of chemical compounds that widely occur in nature as lipids, hormones, and other natural products. They are characterised by a fused tetracyclic ring system that consists of three cyclohexane rings (A, B & C) and a cyclopentane ring (D), Figure 1.1. This structure gives the molecules conformational rigidity and chemical stability. Steroid subclasses are based on the number of carbon atoms in the backbone. Major classes include estranes (C18), androstanes (C19), pregnanes (C21), cholanes (C24) and cholestanes (C27).[1] Anabolic androgenic steroids (AAS) are steroids that exhibit androgenic and anabolic effects by binding to androgen receptors, a well-known example is testosterone, widely used as its injectable ester testosterone propionate (Figure 1.1). Androgenic effects of these molecules include masculinization, and the maintenance and development of secondary male sex characteristics. Anabolic effects promote muscle building, such as an increase in metabolism, skeletal muscle mass, strength, and endurance.[2]

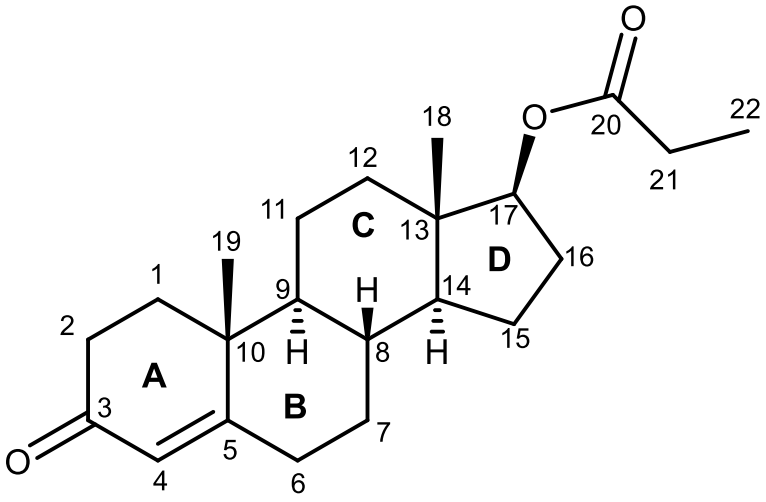


Figure 1.1: Testosterone propionate (TP) showing steroid ring structure and IUPAC recommended ring lettering and atom numbering. The rings of the steroid are depicted in the plane of the paper. Atoms attached to the ring are denoted alpha (α) if it is below and beta (β) if it is above the plane of the paper.[1]

1.1.2 Use of AAS for performance in human and equine sports

The use of ergogenic substances in sports and competition has been recorded since ancient times.[3] In the past these substances have ranged from alcohol, hallucinogenic mushrooms, herbs, bull urine, animal testicles, with little to any proven effect.[3] It wasn't until 1935 following the discovery of testosterone and its physiological effects that the nature of sports doping began to change. Over the decades following its discovery, the misuse of AAS grew in competition, to the point where they were the most widely used form of dopant in both human and equine based sports.[3] This eventuated in the prohibiting of AAS substances in human sports by 1976.[4] For equine sports, this type of anti-doping intervention came in 1974, with the first meeting of the International Federation of Horseracing Authorities (IFHA).[5] Interestingly, this mid-century period of heightened AAS usage in competition closely coincided with the 'golden age' of steroid research.[6] Out of this period came numerous and potent analogues of testosterone. This established the knowledge base that would eventually give rise to designer AAS.[6] Designer steroids have been defined as any AAS which has been prepared in such a way to evade detection by anti-doping authorities, often involving the chemical modification to the steroid core and accessed in a clandestine fashion. Designer steroids are still in wide use today and present a significant problem to anti-doping authorities, however, they are not the focus of this thesis. Various in-depth reviews can be found on the subject.[5,7,8]

As the nature of doping in sports evolved so did the efforts of anti-doping authorities trying to prevent it. This has historically played out as a 'cat and mouse' game with an anti-doping authority's rules and regulations continually adapting to new dopants.[6] Several agencies and regulatory bodies exist throughout human and equine sports to combat doping. For humans the World Anti-Doping Agency (WADA) holds prominent position and sets the standard of the human anti-doping industry.[4,8,9] Within equine sports there is a greater diversity of authorities, largely due to differences in rules and the control of drugs in different geographical regions and jurisdictions. Prominent bodies include the Association of Official Racing Chemist (AORC), Association of Racing Commissioners International (ARCI), the Federation Equestrian Internationale (FEI), and the previously mentioned IFHA.[10]

Accessibility to leading research in equine anti-doping can also be challenging, with a large amount of research published in the proceedings of the international conference of racing analyst and veterinarians (ICRAV), which is not available on sites such as Scopus, Google Scholar, and Medline.[10]

A key motivation for many of these anti-doping authorities is the widely held belief that doping is not only unfair with respect to competitors, but also presents certain dangers to human or animal health.[3] For example, long-term abuse of endogenous AAS (EAAS) has been proven in humans to lead to an increase in multiple adverse health effects, such as, increases in cardiovascular risk factors, in mood disturbances and disorders, and drug withdrawal.[2] In equine sports the risk factors are not as well studied and even less is known about the effects of AAS substance abuse.[10] What is agreed on, is that in both species there is clear evidence that long term abuse can lead to serious health and welfare consequences.[5] Further, one of the main motivators of anti-doping efforts in the protection of the racing industry.[5] However, despite the clear health risks and illegality, AAS remain in high use. As recently as 2019 WADA identified that AAS and associated anabolic agents were the most prolific type of performance enhancing substance used in sport, and accounted for 44% of adverse analytical findings (AAF) and 82% of Atypical Findings (ATF) that year.[11] A similar 2019 report from the Australian Racing Forensic Laboratory (ARFL) also identified AAS substance abuse in equine racing as a major threat within the Australian racing industry.[12]

With the wide range of performance-enhancing drugs available today, it is interesting that AAS remain so popular among sporting cheats, suggesting that they offer benefits that other classes of compounds do not. Unlike other types of compounds, such as stimulants, whose benefits are correlated with the elimination half-life of the drug, AAS have known long-lasting effects on the body, even after excretion of the drug. These effects are primarily from the increase in muscle mass and endurance that are gained and then maintained when the athlete couples AAS use with training. The increased muscle mass is retained for a short period, even after the use of the anabolic steroid is discontinued. This allows the athlete to retain the benefits of the drug while mitigating the risk of being caught during in-competition screening.[5,6] Consequently, there is a continued effort to improve the

retrospectivity and detection for AAS misuse by anti-doping authorities and researchers by the identification of new biomarkers.

1.1.3 Biosynthesis of AAS in humans and horses

An understanding of the biosynthesis and metabolism of AAS is essential in the study and detection of these compounds for anti-doping purposes. All major mammalian bio-synthetic and metabolic pathways of AAS are present in both humans and equine. Consequently, there are many reported similarities between steroid biosynthesis (steroidogenesis) and metabolism in both species.[10,13,14] Steroidogenesis is the biosynthetic process by which cholesterol is converted into all classes of endogenous steroids in the body. It is highly tissue specific and predominantly occurs in the steroidogenic tissue, such as sex organs and the adrenal cortex, Figure 1.2.[15]

Steroidogenesis is primarily driven by two broad classes of steroidogenic enzymes, hydroxysteroid dehydrogenases and cytochrome P450 (CYP) enzymes. There are several differences of steroid biosynthesis in equines and humans. A major difference exists for the CYP17a enzyme, which is responsible for catalysing both 17α-hydroxylation and 17,20-cleavage in the Δ5 and Δ4 biosynthetic pathways of androgens and oestrogens, Figure 1.2. In humans there is a preference for the 17,20-cleavage of 17α-hydroxypregnenolone which leads to the Δ5 bio-synthetic pathway predominating. This results in the steroid dehydroepiandrosterone (DHEA) as the main depot of testosterone and androgen biosynthesis in humans. Whereas, in horses the CYP17 enzyme can also cleave 17α-hydroxyprogesterone leading to both the Δ5 and Δ4 biosynthetic pathways to predominate for androgen and oestrogen biosynthesis. [10,16,17]

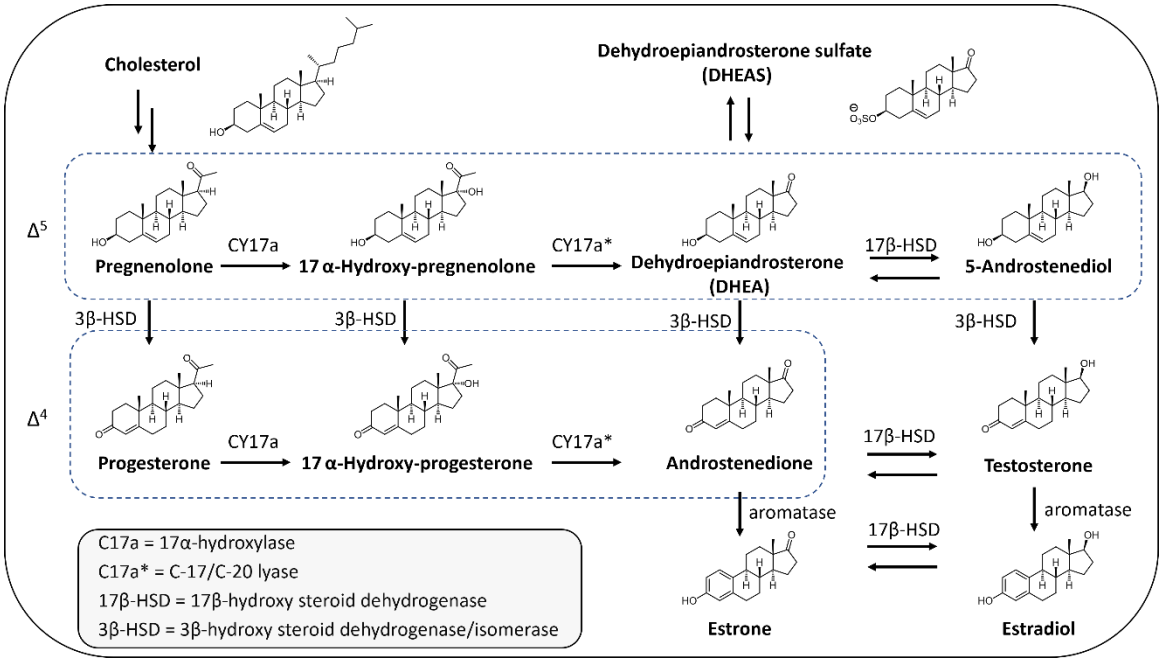


Figure 1.2: Biosynthesis of androgens and oestrogens via the Δ^5 and Δ^4 pathways in mammalian systems. Adapted from multiple sources.[10,16,17]

1.1.4 Metabolism of AAS in Humans and Horses

Almost all exogenous or endogenous small organic molecules undergo chemical modification in the body through metabolism. Metabolism is central in determining the biological properties of a compound such as biological activity, half-life, and toxicity. In the case of AAS metabolism, it is generally conserved between humans and equines.[13]

There are two types of metabolism that govern AAS that are termed Phase I and Phase II, Figure 1.3. Phase I metabolic processes are dominated by enzymatically driven modifications to the steroidal backbone, such as oxidation, reduction, and hydroxylation. This gives a variety of structurally and functionally different steroidal metabolites.[13] In Phase II metabolism, steroids and their metabolites undergo conjugation. This is dominated by the two major classes of conjugates: sulfates and glucuronides, with minor classes being methylation, acetylation, and conjugation with amino acids or glutathione. Glucuronylation is performed by uridine 5'-diphospho-glucuronosyltransferases (UGTs) and substrate uridine diphosphate (UDP) glucuronic acid, and sulfation is performed by sulfotransferases (SULTs) and substrate 3'-phosphoadenosine-5'-phosphosulfate

(PAPS).[13] The polarity gained in these reactions aid in their excretion from the body, resulting in these conjugates accounting for up to 97% of excreted steroid metabolites in the urinary metabolome of humans.[18]

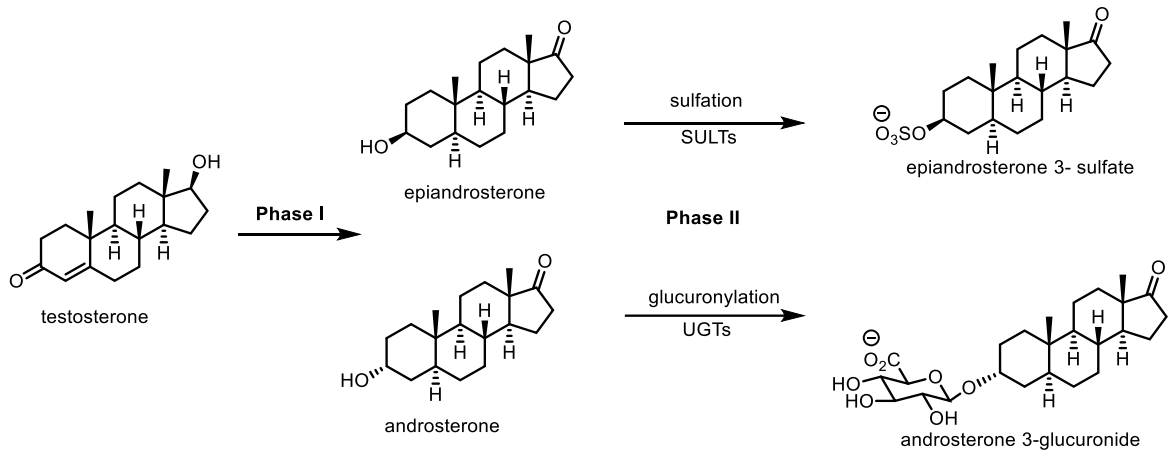


Figure 1.3: Phase I and Phase II metabolic pathways for testosterone. Adapted from Schänzer *et al* (1992).[19]

The metabolism of AAS does not vary greatly between humans and equines, however, there are several reported key differences (Figure 1.4).[10] In Phase I metabolism, hydroxylation at the 16 position of the androstane backbone occurs at a higher frequency in equines, while the 6-position hydroxylation is more commonly observed in humans. Further, it has been reported that the reduction of the 3-ketone position tends to favour 3β-hydroxylation in equines, while the 3α-epimer is favoured in humans.[10] In Phase II metabolism there is a trend for sulfation to predominate over glucuronylation in horses compared to humans. The opposite has been reported in human metabolism.[10] In the equine, sulfation has also been shown to be favoured for 17β-hydroxy groups, while glucuronylation is selective towards 17α-hydroxy groups.[14]

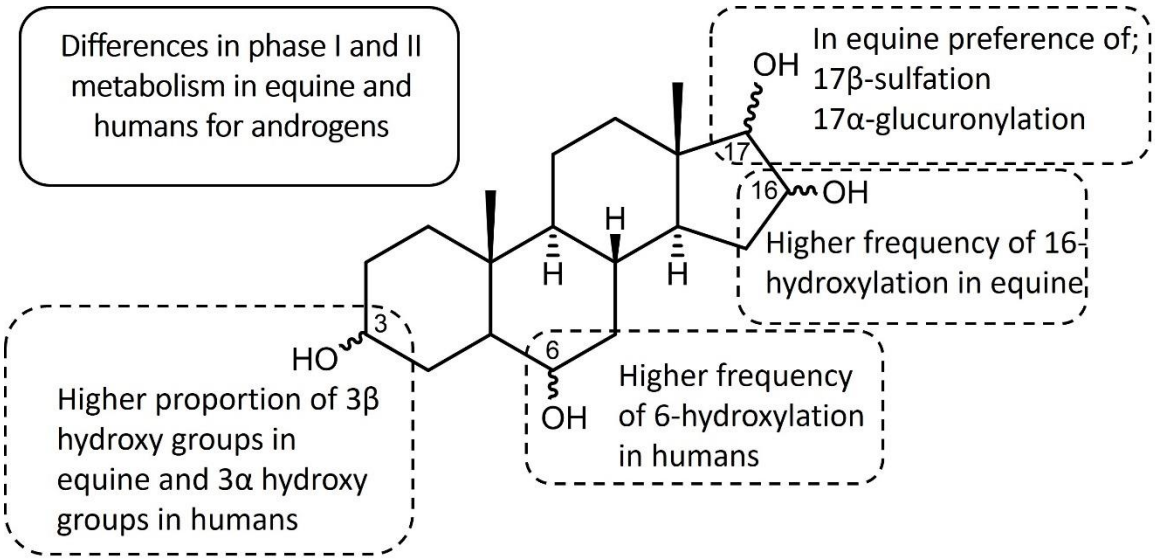


Figure 1.4: General differences between Phase I and II metabolism in humans and equines. Adapted from Scarth *et al* (2011).[10]

1.1.5 Instrumentation for the detection of AAS

The analyses used to detect AAS abuse has changed with the advancement of instrumentation over time. Historically, detection of AAS and their metabolites was first done using rudimentary thin layer chromatography tandem ultraviolet (TLC-UV) techniques as early as the 1950's. As technology progressed this was quickly surpassed by high performance liquid chromatography-UV (HPLC-UV) detection which was then followed by mass spectrometry-based techniques. In a modern context, the contemporary anti-doping laboratory relies on several techniques such as gas chromatography-mass spectrometry (GC-MS), ultra-high-performance liquid chromatography-mass spectrometry (UHPLC-MS), isotope ratio mass spectrometry (IRMS), nuclear magnetic resonance (NMR) and bio-assays.[6] The major instrumental technique used throughout this thesis is UHPLC-HRMS/MS. A brief overview of GC-MS and liquid chromatography tandem mass spectrometry (LC-MS/MS) techniques, and various mass analysers, is provided below to contextualise their importance in AAS analysis.

1.1.6 Gas chromatography-mass spectrometry

Gas chromatography mass spectrometry coupled with electron ionisation (EI) has been the method of choice in the anti-doping industry as it allows for reproducible and robust profiling of AAS.[6] This method provides excellent chromatographic resolution, high sensitivity and can be used for both qualitative and quantitative assays. An important example of the prevalence of GC-MS is that it is the main analytical technique used for the steroid module of WADA's athlete biological passport (ABP). The ABP purpose is to serve as an electronic record of individual athlete's biological variables to indirectly reveal the effects of doping behaviour. The steroid module is one component of the WADA Anti-Doping Administration and Management System (ADAMS), which is one part of an intricate anti-doping operational and intelligence framework.[20] The steroid module uses an adaptive model based on Bayesian networks for the longitudinal monitoring of athletes. Specifically, it monitors the concentrations of seven steroid biomarkers in an athlete's urine and plasma. These are testosterone, epitestosterone, androsterone, etiocholanolone, 5α-androstane-3α,17β-diol, and 5β-androstane-3α,17β-diol, the testosterone-epitestosterone ratio (T/E), together with the specific gravity of an athlete's urine sample. Any variation or unusual perturbation in these values gives a case for further investigation into potential doping behaviour within ADAMS.[21] Alongside this, GC-MS techniques were recently identified in a 2017 review by Schänzer and Thevis as essential in the identification of over 60 AAS and their metabolites as well as in the ongoing identification of novel AAS.[6]

Although GC-MS techniques form the backbone of many analytical laboratories they have several distinct disadvantages when analysing intact Phase II AAS metabolites.[22] These conjugate metabolites are in most cases not volatile or thermally stable enough for direct GC-MS analysis. Instead, they require a laborious sample preparation that is used to isolate, hydrolyse and derivatise the conjugate for analysis. Specifically, conjugates are firstly isolated either through liquid-liquid or solid phase extraction, followed by either, chemical, or enzymatic deconjugation to liberate the free steroid. They are then derivatised to more volatile chemical species for GC-MS analysis. Trimethylsilyl (TMS) ethers are commonly used for the derivatisation step due to their ease of preparation, chemical and

thermal stability, high volatility, and distinct ionisation pattern.[6,23] The main drawbacks in this sample preparation exist in the deconjugation and chemical derivatisation steps. A common chemical deconjugation technique is acid solvolysis, which provides a robust and general method of deconjugation. However, as it does not discriminate between different types of conjugate metabolites, such as sulfates and glucuronides, it can lead to a loss in chemical information regarding the conjugate type and site, and in some cases has been shown to lead to sample degradation.[24] Milder conditions such as enzymatic ones, can provide selective deconjugation and therefore retain information of the metabolites conjugate type. However, in anti-doping laboratories the most used enzyme for the deconjugation of glucuronylated metabolites, *Escherichia coli* (*E. coli*) β-glucuronidase, has been shown to result in incomplete hydrolysis of some substrates. Moreover, crude extracts of sulfatase enzymes are also routinely used for deconjugation; however, these contain considerable glucuronidase activity, and can lead to undesirable steroid conversions.[18] This leads to a loss of information when using GC-MS techniques for conjugates. Regardless, some recent methods have been developed for detecting sulfate conjugates using GC-MS.[23,25]

1.1.7 Liquid chromatography mass spectrometry

Liquid chromatography mass spectrometry has rapidly grown in importance in anti-doping and is widely used alongside GC-MS for the analysis of AAS. In contrast, LC-MS allows for the direct analysis of intact AAS conjugates and avoids the need for the time-consuming sample preparation steps required in GC-MS protocols, meaning less sample is often required for analysis. The sample preparation is straightforward, often needing only solid phase extraction (SPE) before undergoing chromatographic separation. Chromatographic separation in LC-MS is important as many AAS metabolites are isobaric. To achieve separation various column stationary phases can be used. Popular ones include reverse phase C18, C8, C2, phenyl-hexyl, and normal phase hydrophilic interaction liquid chromatography (HILIC) such as silica, ion exchange chromatography (IEX) and amino columns, Figure 1.5. These are commonly used with ultra-high performance liquid chromatography (UHPLC) systems to achieve better chromatographic resolution, sensitivity, specificity and to shorten analysis times.[23]

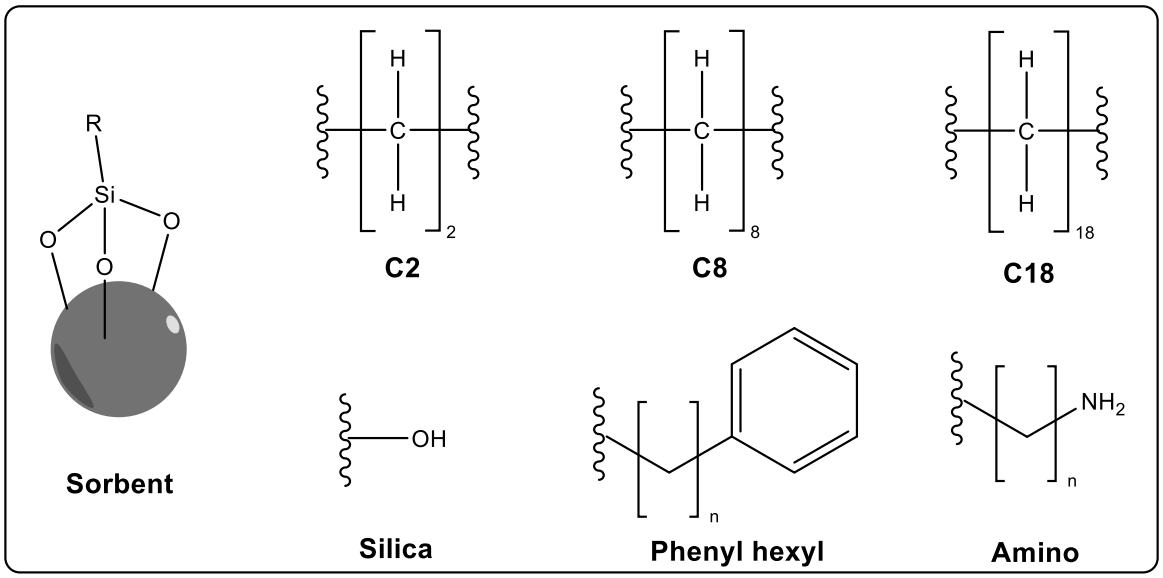


Figure 1.5: Types of common stationary phases used in UHPLC to separate steroids and their metabolites.[23]

The two main ionisation techniques used in LC-MS are electrospray ionisation (ESI) and atmospheric pressure ionisation (API), ESI is the most widely used of the two and provides efficient ionisation for polar and ionic compounds alike.[26] In ESI, ions are produced by applying a high voltage to a liquid that is passing through an emitter, typically a metallic or glass capillary. When the liquid reaches the tip of the capillary it forms a Taylor cone that emits a charged jet of liquid through its apex. This leads to the formation of highly charged droplets of liquid that are dispersed through Coulombic repulsion, forming ions, Figure 1.6.[27] This type of 'soft' ionisation is ideal for AAS conjugates and other biomolecules, that are easily destroyed during harsher ionisation techniques such as electron ionisation. In ESI AAS conjugates and free steroids alike tend to ionise as deprotonated species ($[M-H]^-$) or adducts (e.g. $[M+NH_4]^+$, $[M+H]^+$). However, ESI analysis of AAS conjugates has been reported to be susceptible to matrix effects and is often not suitable for more non-polar free AAS metabolites. The latter often lack a ketone or other polar functionality needed for efficient ionisation, such as steroid diols.[23] In these cases, harsher ionisation techniques such as atmospheric pressure chemical ionisation (APCI) tend to be used. However, this can cause increased in-source fragmentation of AAS conjugates and can lead to loss of information. Often in these cases LC-MS and GC-MS analysis will be used in a complementary manner.[23]

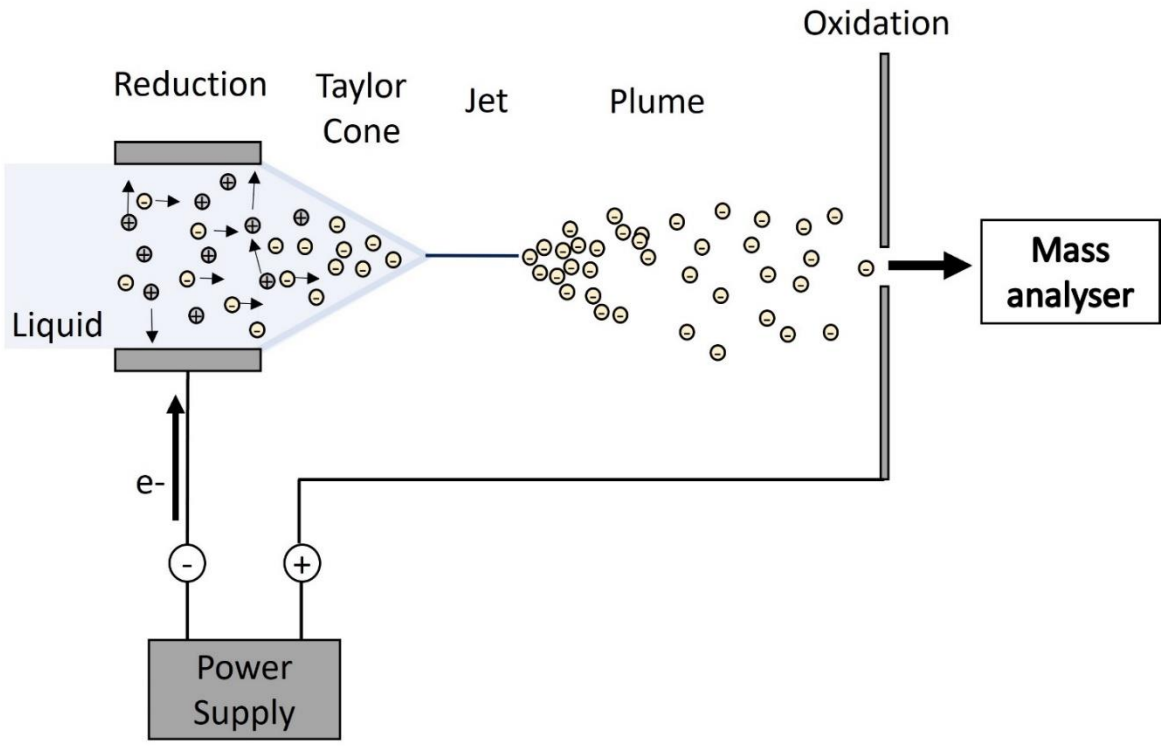


Figure 1.6: Instrumental schematic of electrospray ionisation and formation of a Taylor cone and liquid jet in negative mode. Image adapted from Crotti *et al* (2011).[27]

1.1.8 Mass analysers

Mass spectrometry plays a crucial role in the detection of AAS and the types of mass analysers available to the researcher are extensive, with each offering different advantages and disadvantages.[28] The major difference between analysers is the way by which *m/z* ratio is used to separate ions. Mass spectrometry is the principle analytical technique used throughout this thesis, and as such, several analysers are described in brief below. Specifically, the principal mass analysers used in this thesis are low resolution triple quadrupole, and for high resolution, orbitrap and quadrupole time of flight analysers.

1.1.9 Low resolution mass analysers

Quadrupole mass analysers are among the most common in MS instruments.[26] These analysers guide ion beams through an opening between four parallel rods. Fixed and alternating voltages are applied to these rods which functions to stabilise the ion beam trajectories over a given *m/z* range. Stabilised ions will pass through to the detector and be recorded, however, ions with unstable trajectories will collide with the rods and not be recorded in the final spectrum. Triple-quadrupole analysers are the preferred analyser in many anti-doping laboratories, as each quadrupole can be used in different ways, giving a variety of multistage MS (MS^n) experiments.[29] An example of this is collision induced dissociation (CID) experiments, Figure 1.7. Here, an ion can be selected in the first quadrupole, subjected to CID in the second, and then in the third, the product ion spectra analysed.[27] The high scan speeds of triple quadrupole instruments are also favoured as it allows the analyst to survey hundreds of transitions per second. In general, for quantitative analysis and routine work, low resolution instruments such as triple- or single-quadrupole mass analysers are preferred as they offer good sensitivity, selectivity, and are versatile, cost-effective, and robust to operate.[29] Further, quadrupoles are a common addition to the front of various HRMS analysers allowing for the capacity to perform MS/MS acquisitions and experiments.

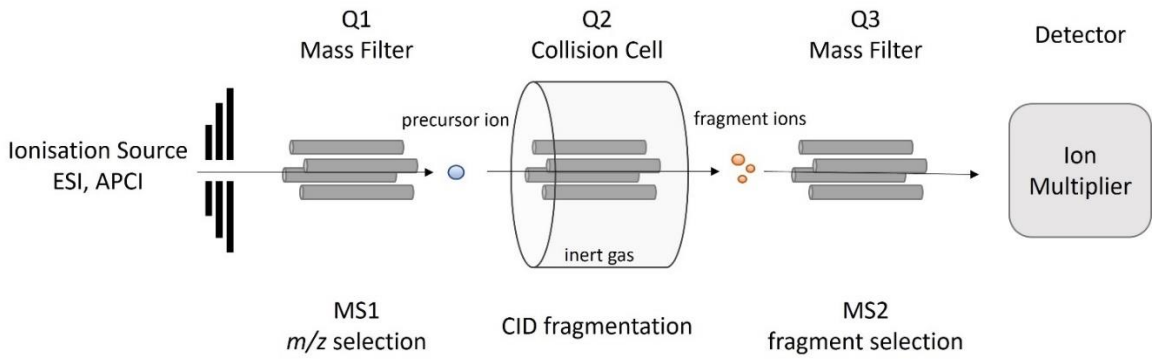


Figure 1.7: Schematic of a triple quadrupole (QQQ) mass analyser equipped with a collision cell in quadrupole 2 (Q2). Adapted from Crotti *et al* (2011).[27]

Another well-established low-resolution analyser is the quadrupole ion trap. Described as an ‘electric field test tube’, ion traps function as a device for the confinement of gaseous ions in positive or negatively charged fields.[26] Linear ion traps work by trapping ions in a potential well or ‘basin’ between 4 hyperbolic shaped rods, Figure 1.8.[30] There are three sections (front, middle and back) that have discreet DC levels that allow the containment of a packet of ions along the central axis of the middle section of the ion trap in a ‘potential well’. Spectra are measured by scanning the potential of the middle section rods, this causes ions of increasing *m/z* to be ejected from the trap through the narrow slits in the axial rods, where they are then measured by an ion detector. This has been likened to tipping the ‘basin’ of the potential well, the trapped ions then spill out of the basin in the order of ascending *m/z* value.[31] Using this approach ion traps can be used for mass selection allowing researchers to study specific ions. Ion traps are well suited for MS^n experiments as they allow many cycles of ion isolation, fragmentation, and isolation of product ions.[30] Ultimately, this allows for extensive data to be acquired and elucidation of fragmentation pathways for single molecules. One major drawback is the loss of the ion abundance after multiple fragmentation cycles. For a detailed description of ion trap mass spectrometers there are several reviews available.[31,32]

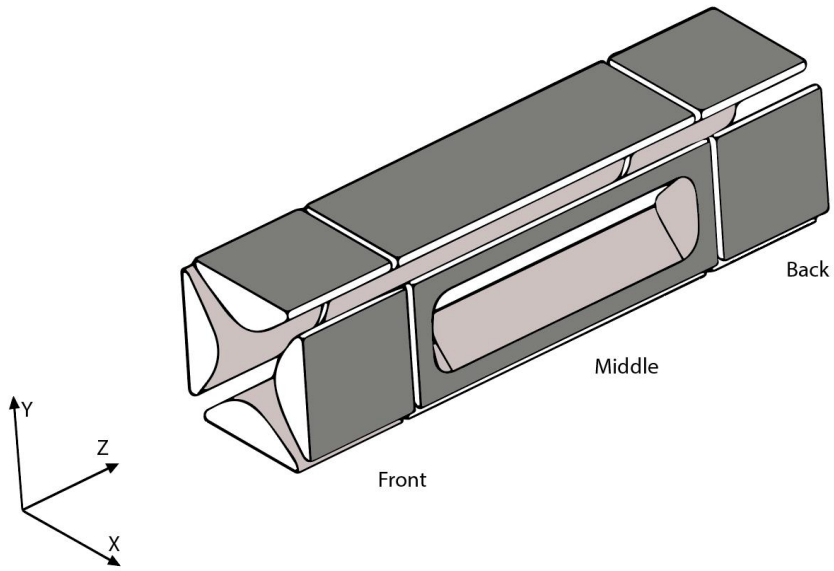


Figure 1.8: Schematic of a linear ion trap. Ions are held in a ‘potential well’ between 4 hyperbolic shaped rods divided into three sections. Adapted from Schwartz *et al* (2002).[30]

1.1.10 High resolution mass analysers

High resolution mass spectrometry (HRMS) is defined as any type of mass spectrometry where the exact mass of the molecular ions in a sample is determined, as opposed to just the nominal mass (number of protons and neutrons in a molecule).[33] Resolution is often defined in terms of the full width at half maximum (FWHM) method. Here the ion mass (*m*) is divided by the peak width (Δ*m*) at half of the peak height. If the FWHM value is greater than (>) 10,000 it is defined as high resolution.[34] ‘

There are a variety of HRMS analysers in use today, with the main ones discussed and used in this thesis including the quadrupole time-of-flight (Q-ToF) and quadrupole Orbitrap mass spectrometers.[26] Other notable analysers include Fourier transform ion cyclotron resonance (FT-ICR) which is not discussed in this thesis, however, several reviews are available on the subject.[35,36] Additionally, HRMS techniques have revolutionised detection of AAS in anti-doping analysis, providing increases in sensitivity afforded by using narrow mass windows (± 1-10 ppm).[23,37] This improves analysis as less sample is needed and allows for retrospective analysis of emerging compounds in historic samples.[6,23]

Quadrupole time-of-flight (Q-ToF) mass analysers are conceptually simple and are essentially a triple quadrupole with the last quadrupole being replaced by a ToF analyser. The ToF analyser is at the heart of the Q-ToF’s high resolution capacity. They work to differentiate ions based on their *m/z* value by measuring the time taken to traverse a flight tube. Ion packets are accelerated into a flight tube and to a detector using an electric potential. Ions of lower mass (or higher charge) will have greater velocities than higher mass ions, and the time taken to traverse the flight tube can then be translated to the *m/z* value.[26] Most modern Q-ToF analysers are now equipped with an ‘ion mirror’ (also known as a ‘reflectron’ or ‘electrostatic mirror’) to improve the overall resolution by decreasing the ion packet thickness. They consist of a series of electrodes placed at the end of the flight chamber that repel incoming ions back down the flight tube to the detector. More energic (faster) ions penetrate deeper into the ion mirror and travel a slightly greater overall distance than less energic (slower) ones. In effect, this focuses the ion package

onto the detector improving resolution, Figure 1.9. Here, all ions in an ion packet are detected simultaneously, this differs from a triple quadrupole, which individually scans through an *m/z* range. Due to their configuration, Q-ToF analysers are favoured among anti-doping analysts as they are essentially a high-resolution version of triple quadrupoles and can perform similar MS^n experiments.[26]

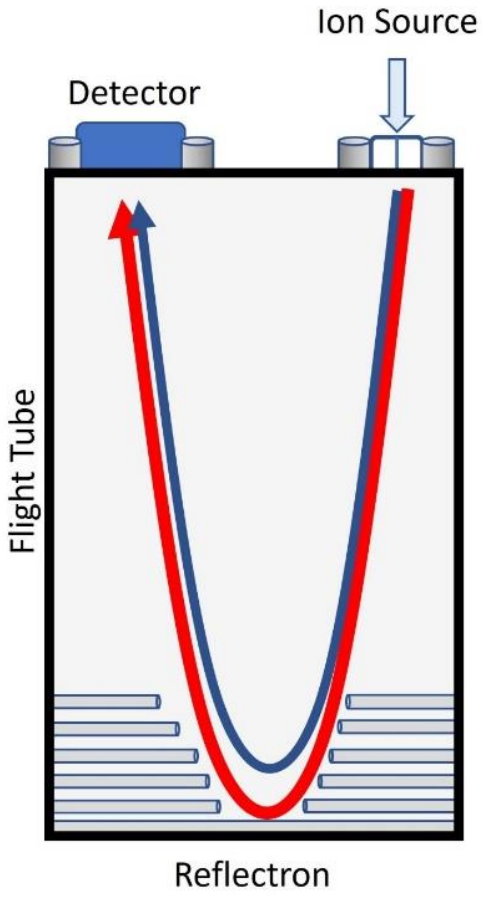


Figure 1.9: Schematic of a single reflector ToF mass analyser. The red ion-trace has a higher initial velocity which penetrates further into the reflectron, ultimately giving it a longer flight path. It then arrives at the detector at the same time as the blue ion-trace which had a lower initial velocity. Image adapted from multiple sources.[26,38]

Orbitrap analysers are one of the newest types of mass analyser, and are based on the principle of electrostatic orbital trapping of ions first proposed by Kingdon in the 1920's.[39] In these, ions are electrostatically trapped between a spindle shaped inner electrode and barrel shaped outer electrode. The outer electrode is split into two identical halves with a gap to allow for ion entry. The motion of trapped ions comprises of a complex shape with two components. Ions orbit around the central electrode while also maintaining an axial oscillation in the z-axis relative to the same electrode, Figure 1.10.[34] These are termed angular (r) and radial (z) directions, respectively. The *m/z* value can be directly related to the radial oscillation in the z-axis that is independent of the angular orbital motion. A spectrum is generated by performing a Fourier transform of the current produced from the radial oscillation frequencies and detected in the two halves of the outer electrode. One strength of the Orbitrap mass analysers is their high resolving powers (>150,000) and high mass accuracies (as low as 0.2 ppm) which are only matched by the likes of FT-ICR analysers.[34,40] However, the higher resolutions come with a trade-off, as they require longer times to acquire data and thus equate to slower scan speeds.

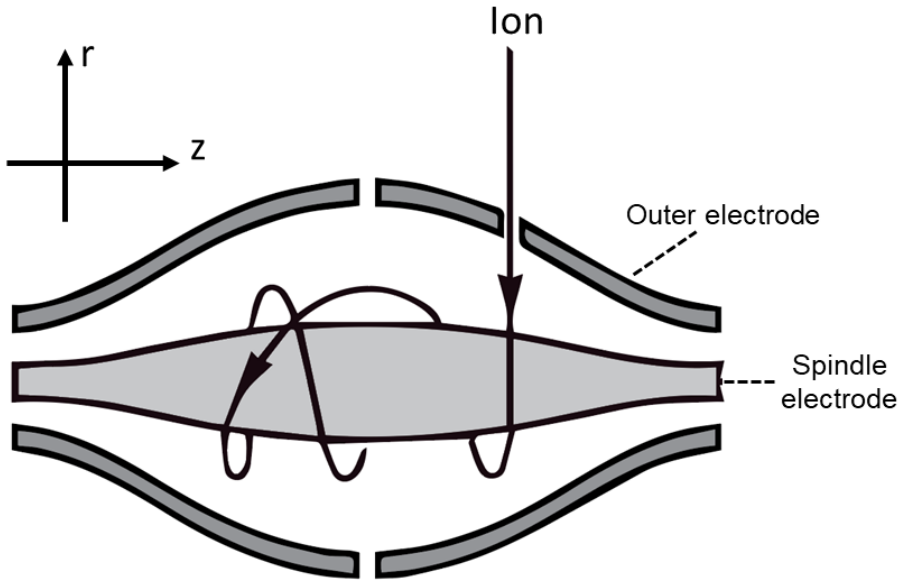


Figure 1.10: Schematic of an Orbitrap. The black line is the typical spiral trajectory of an ion along the z- and r-axis of the spindle electrode. Adapted from Makarov *et al* (2015).[34]

1.1.11 Metabolomic profiling

The field of metabolomics, first known as metabonomics, has rapidly developed over the past two decades as a complementary one to genomics (genes), proteomics (proteins) and transcriptomics (RNA transcripts), which all belong to the field of 'omics-sciences'. The central goal of metabolomic profiling is to detect as many low molecular weight molecules (50-1500 Da) in a biological system as possible.[41]

In their review of metabolomic and lipidomic based methods, Cajka and Fiehn (2016), describe the concept of a spectrum that spans between truly untargeted and targeted approaches, Figure 1.11. [42] They were able to show that most techniques exist within this spectrum, with a trade-off between the reliability for quantification of metabolites and the number of metabolites analysed. In this definition, targeted approaches are characterised by an increase in the reliability of quantification of a small number of metabolites, the opposite is true of untargeted ones, where little is known about a large number of metabolites. Ultimately, they were able to identify that most analytical approaches exist as a mixture of targeted and untargeted techniques, describing six types of possible approaches along a spectrum, Figure 1.11.[42] The detection of AAS is no different, with various techniques having been employed that would fit along this scale.[23,43,44] In this thesis the primary technique used is untargeted metabolomic profiling, however, for context brief overviews of both types of broad assays are given below.

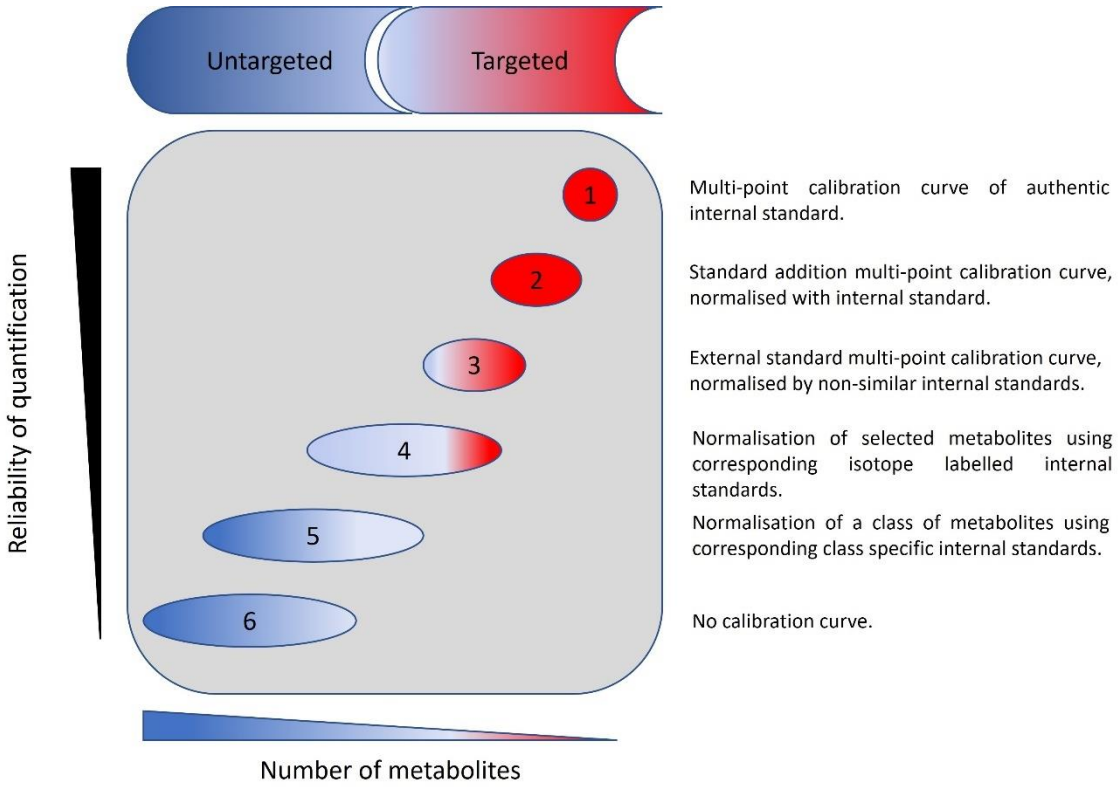


Figure 1.11: The relationship between the number of detected metabolites and the reliability of quantification in untargeted and targeted metabolomic assays. Adapted from Cajka and Fiehn (2016).[42]

Targeted metabolomics aims to quantify the presence of a pre-defined set of compounds, but due to this limited scope, it is often not considered a true omics-based approach.[45] Rather, it falls under the umbrella of direct detection or targeted methods. The use of multi-point calibration curves, certified reference materials and stable isotope labelled (SIL) internal standards gives these methodologies high confidence and accuracy in the detection and quantification of metabolites, Figure 1.11, criteria 1-3.[42,43,46] In targeted metabolomic approaches key metabolites within metabolic pathways act as representative markers that reflect changes within that pathway. Targeted approaches are limited to the assessment of a few to the low hundreds of metabolites due to factors such as the commercial availability of reference standards, the time it takes to perform these experiments in a high-throughput laboratory environment, and increased sensitivity needed for instruments in targeted mode.[47] In an anti-doping context, these targeted assays tend to focus on the longitudinal analysis of a small subset of metabolites termed a ‘metabolic fingerprint’, various examples have been presented to this end.[48,49] An

example that highlights the use of targeted AAS detection is the previously discussed ABP.[21] Here a small number of metabolites are analysed in a targeted fashion to a high degree of accuracy and associated confidence. Asides from the ABP, one early example of this type of targeted approach was a WADA funded project in 2006 called “metabonomic signature in bike athletes”.[50] This became a precursor to the steroid ABP module in ADAMS.[51] In general, the strength of targeted methods lies in the validation of one or several hypotheses, whereas, untargeted approaches attempt for the unbiased assessment of all small molecules in a sample giving the opportunity to discover novel metabolites and the development of new hypotheses. Moreover, targeted metabolomic methods are routinely and effectively used in various fields such as nutrition research,[52] plant metabolomics,[53,54] and medical diagnosis.[46,47,55–59]

Untargeted metabolomic methodologies focus on the analysis of all detectable metabolites in a sample, whether they are known or unknown, and are generally considered a true omics approach. As screening for all metabolites gives a more comprehensive picture of a systems physiological state through its metabolic signature, it can potentially reduce bias compared to targeted approaches. [60–62] The advancement of UHPLC-HRMS/MS technologies have enhanced untargeted metabolomic techniques and allowed for the rapid detection of thousands of different compounds that make up a biological sample. The metabolomic profiles obtained using these UHPLC-HRMS/MS based methods are often described by their *m/z* values, the intensity of detected ions and their retention times.[60] These methods have been used to interpret change in a biological system, such as perturbations due to disease states, drug administration or genetic engineering, and can give an unprecedented insight into biological processes down to a biomarker level. In general, this approach typically relies on the statistical (i.e. univariate or multivariate analysis) or computer science (i.e. machine learning) analysis of patterns or signatures that reflect a particular metabolic state.[63–67] Only the most thorough of studies will result in the identification of novel metabolites, as this step requires the investment of significant time and resources. Moreover, the major challenge of untargeted metabolomic profiling is often the unambiguous identification or confirmation of any significant biomarkers that are selected from a workflow.[41,68–70]

1.1.12 Untargeted metabolomic workflows

Untargeted metabolomic workflows tend to follow a general order of experimental design, consisting of sample collection, preparation, and then data acquisition steps. After these, various data processing pipelines, and statistical analysis can be employed to highlight potentially important biomarkers, this latter step is often dependent on the biological question being asked. The following section will discuss aspects of a general MS-based metabolic workflow. Various in-depth reviews exist on the subject which are outside the scope of this thesis.[43,45,71,72]

Many types of sample collection and preparation procedures have been described for metabolomic workflows and tend to be specific to type of biological matrix. In general, for untargeted metabolomic studies, sample collection and preparation should be as simple and universal as possible. It is also important that sample preparation is as consistent as possible, and accounts for multiple factors such as protein, substrate or analyte concentration, analyte polarity and stability once isolated.[60,72] Some of the most common sample preparation techniques used in MS-based metabolomics are solid phase extraction (SPE), protein precipitation (PP) and liquid-liquid extractions (LLE).[45] The former of these techniques, SPE, is used extensively throughout the sample preparation methodology in this thesis, it was chosen due to its proven generality for the extraction of AAS and their metabolites from urine.[22,73–76]

1.1.13 Data acquisition in untargeted metabolomic profiling workflows

Central to many MS-based untargeted metabolomic profiling methods are data acquisition techniques. Some common approaches are scan MS, multiple reaction monitoring (MRM) otherwise known as selected reaction monitoring (SRM), single-ion monitoring (SIM), and non-targeted acquisition. The most basic of these is scan MS. Here, all intact masses that reach the detector are measured, this can be useful in the determination of chemical formulae and the identification of known compounds. Next, the use of Multi-stage MS (MS^n) data acquisition are commonly performed with CID experiments. In an anti-doping and broader steroid-omics context this technique and other targeted acquisition techniques like MRM have been used in various applications in both low- and high-resolution instruments to

gain qualitative insight into a given set of AAS analytes. They possess significant weaknesses when it comes to the number of analytes that can be targeted.[77–83]

Lastly, non-targeted acquisition has been widely adopted for use in untargeted profiling in the broader omics (genomics, proteomics, & metabolomics) community. Two main techniques exist for this purpose, they are data dependant acquisition (DDA) and data independent acquisition (DIA). Both have the goal of collecting fragmentation information for as many detected features as feasible. A simple overview of both techniques is shown in Figure 1.12. [84] In this thesis the main non-targeted acquisition method used in DDA, however, a brief overview of both methods is given below for context.

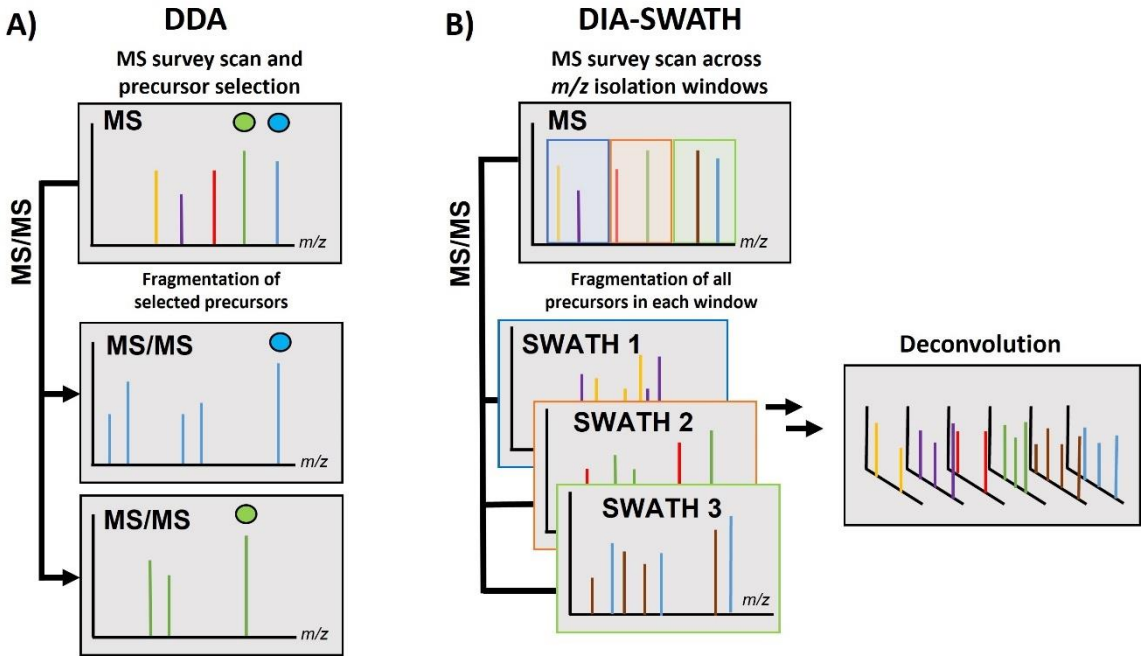


Figure 1.12: Overview of DDA-MS and DIA-SWATH techniques in mass spectrometry. **A**) In DDA-MS the '*n*' most abundant precursor ions in each MS1 survey scan are selected and fragmented to give MS/MS data. **B**) In DIA- SWATH, the survey scan is divided up into pre-defined mass isolation windows of either a constant or variable *m/z* width. All precursors in each isolation mass window are then fragmented to give MS/MS data following deconvolution. The was adapted from multiple sources.[84,85]

1.1.14 Data dependant acquisition

In DDA workflows the mass analyser will automatically switch from acquiring scan MS data to MS/MS data, only allowing a fixed number of precursor ions to be selected and analysed by MS/MS, Figure 1.12. This switching is triggered when a predefined criterion is met in the scan MS, this can be things such as, a precursor ion intensity threshold, isotope pattern, mass defect, or the presence of a diagnostic ion or characteristic neutral loss (in GC-MS settings). These precursor ions are then taken through to MS/MS and subjected to CID. When used in untargeted profiling it is common for this triggering to occur based on the abundance of precursor ions, i.e., to take the top 10-20 most abundant precursors per survey scan.[86] The power of this technique is that it directly links the precursor with its tandem MS/MS product ion spectra. This gives the researcher MS/MS information on '*n'* molecular features with a relatively simple data deconvolution process. In metabolomic studies DDA based methods are commonly used for the identification of Phase II metabolites based on diagnostic fragmentation patterns. These have included methods for the detection of sulfates, glucuronides, bile acid conjugates and acylcarnitines from a range of various bio-fluids.[87–90]

There are several weaknesses associated with the DDA method. Firstly, the presence of low amounts of 'other ions' than the precursor in the isolation window for MS/MS experiments. This presence of so called "noise signals" can hamper data interpretation and incorrectly inform spectral library matching. Several statistical and computation programs have been designed addressing these artifacts.[91,92] Another limitation is the sampling of background ions from the matrix. Due to the nature of sampling, the presence of background ions in high abundance can lead to analytes of interest going undetected. To combat this problem, it has become routine to use dynamic and static exclusion methods. Dynamic exclusion windows exclude the selected precursor ions for a set amount of time following MS/MS acquisition, to ensure that they are not selected multiple times. Static exclusion methods such as background subtraction lists are often created by running solvent or extraction blanks prior to acquisition to measure the background ions, these are then excluded from being acquired. These methods actively curate the data and increase the potential for the detection of biologically relevant

molecules.[86] Finally, DDA can also suffer from inconsistent feature detection when there is a close eluting pair or variance in abundance of precursor ions between biological samples. This is largely limited by the scan speed of the instrument, which can make retrospective analysis difficult when looking for new compounds and give an incomplete metabolomic picture. Some studies have reported a loss of up to 15% and 30% of biologically relevant metabolites when using the DDA acquisition technique between biological samples when compared to DIA methods.[86,93]

1.1.15 Data independent acquisition

In DIA, all precursor ions over a selected *m/z* range are analysed by MS/MS, allowing for the collection of scan MS and MS/MS on every detected molecule. Unlike DDA, there are many techniques that lie under the umbrella of DIA approaches. Some of these are, All-Ion Fragmentation (AIF), MS^{All}, MS^{E} and are otherwise generalised as “All-in-One” approaches used exclusively with HR instruments coupled with U(H)PLC systems. Due to the depth of this subject matter it will be covered in brief for the purposes of this thesis, however, in-depth literature reviews and commentary are available.[86,94]

In general, DIA methods are based on the isolation of predetermined *m/z* ranges from a given survey scan and the subsequent fragmentation of these (Figure 1.12).[95] This acquisition is done by variations on one of two general methods; either by 1) broadband methods, or by 2) Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra (SWATH-MS) methods. Broadband methods were first described as early as 2005 and initially evolved from predecessor techniques such as, ‘nozzle-skimmer dissociation’ and ‘shotgun CID’ methods.[94] Broadband acquisition (MS^{E} or MS^{All}) involves alternating scans in the collision cell between low and high energies, allowing the information of both precursor ions (low energy) with fragment ions and neutral losses (high energies) to be collected for each isolation window over the full *m/z* range (Figure 1.13).[94] A related technique specific to Orbitraps is that of AIF where higher energy C-trap dissociation (HCD) is performed, where all precursor ions are fragmented without pre-selection criteria. One intrinsic disadvantage of these broadband type

techniques is that they result in highly complex MS/MS spectra (mixture of MS and MS^E) and often require sophisticated software tools to deconvolute precursor-product (fragment) ion pairs.

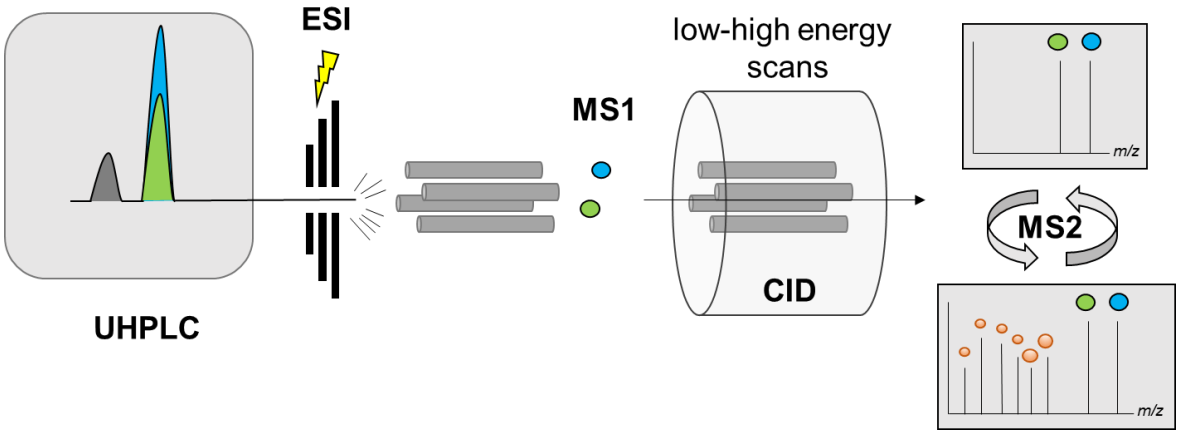


Figure 1.13: Schematic of a broadband MS^E DIA method. Co-eluting molecules (blue and green) are simultaneously fragmented at high and low energies. The resulting MS/MS (MS2) spectra is a complex mixture of precursor and product ion pairs of both molecules. Adapted from Plumb *et al* (2006).[95]

Alternatively, SWATH approaches work by reducing the number and presence of interfering ions in the MS/MS spectra and is considered a hybrid between DDA and MS^E methods (Figure 1.12, B).[84] In SWATH methods the first mass filter is stepped into discrete mass isolation windows of a constant size (typically *m/z* 25) over the whole *m/z* range of interest. The result of this is that fragmentation data is obtained over a narrow discrete range. This has the effect of improving MS/MS spectra quality and making spectra deconvolution comparatively simpler than broadband methods.[96] Moreover, SWATH methodologies allow for quantitative and full retrospective analysis of every single molecule in each sample. They have also been demonstrated to be comparable to targeted MRM techniques for the quantification of metabolites, however, they tend to have a lower sensitivity. Ultimately, unlike MS^E and DDA methods, SWATH has the capacity to allow for a combined untargeted and targeted workflow for large-scale metabolite quantification.[86,97,98]

The main drawback of SWATH techniques is the presence of chimeric spectra. Here, the presence of multiple close eluting molecules of similar masses and/or that are isobaric can cause complex MS/MS spectra, leading to a lower quality of overall data. To reduce the occurrence of these chimeric spectra a third technique has recently been reported in proteomic applications known as variable DIA (vDIA) and/or variable SWATH.[99] In these, the *m/z* range is varied depending on the abundance of molecules in a given isolation window.[99] In variable techniques, these crowded mass ranges are divided up into smaller isolation windows of variable size that are dependent on the mass difference between the analytes of interest, with mass isolation windows being as small as 1 *m/z*.[97] These allow for a higher discrimination between molecules, giving higher quality spectra. However, these type of DIA techniques are limited by the scanning speed of the mass analyser.[97]

1.1.16 Data Processing and Informatic Tools for Metabolomic Workflows

Despite these considerable advances in non-targeted MS acquisition technologies, they only make up part of an untargeted metabolomic workflow. Once collected, the correct processing of data is critical to gaining meaningful insights and analysis of a sample. Metabolomic data analysis for DIA and DDA techniques have developed rapidly over the last decade. Initially informatic tools were limited to vendor-based software that although powerful, tends to be expensive, application specific, and closed off to modification (closed-source software). These factors can hinder research more than advance it.[86] Adding to this, the more widely used open source metabolomic software tools, such as, XCMS and MZMINE, are not designed for DDA and DIA data sets.[100,101] In response to this, over the last decade a suite of various open source (free) software and code-based tools have been made available for DIA and DDA based metabolomics. A representative list of these have been made available with their uses in Table 1.1. Some more comprehensive software packages are MS-Dial, MetaboAnalyst, OpenMS, and the recently released MetaboKit.[102–104] These packages allow a full suite of data alignment, deconvolution and normalisation, database matching and basic to advanced statistical analysis for metabolomic workflows. Other software tools have more specific goals such as, evaluating purity,[91] or the improvement of data

acquisition and/or analysis.[105–110] As such, in this thesis a variety of open-source packages and programs are used for various purposes, such as data filtering, alignment and data pre-treatment. In instances were no specific program or packages could be found for the required chemo- or bio-informatic data processing, custom script was developed and applied in the coding languages R and or Python. As a result, part of this thesis has been attributed to the development of new informatic scripts to aid in the specific requirements of the research being performed. In this context and more broadly, it is as important for researchers in metabolomics, to have a rudimentary grasp on the use of chemo- and bio-informatic tools and coding languages, as it is for them to be able to perform the sample extraction and instrumental analysis required for metabolomic experiments.

Table 1.1: List of DIA and DDA compatible open-source data processing packages. This represents a non-exhaustive search and is based on a similar search reported in a review by Fenaille *et al* (2017).[86]

Package	Use	Language/Software
MS-Dial[102]	**DIA SWATH.** Deconvolution. **DDA & DIA.** Data processing suite including, alignment, and normalisation of data with some statistical analysis.	Software app
MetaboAnalyst 5.0[103]	**DIA & DDA.** Encompassing data processing and statistical suite.	R package and Webpage
msPurity[91]	**DDA.** Spectral purity score.	R package
OpenMS[111]	**DIA SWATH.** For targeted metabolomic analysis	C^{++} core platform
MetShot[105]	**DDA.** Data acquisition and processing tools.	R package
IE-Omics[106]	**DDA.** Iterative exclusion list for multiple injections.	R code
Met DIA[108]	**DIA.** Targeted extraction of metabolites from multiplexed MS/MS spectra.	R code
SWATH Tuner[107]	**SWATH DIA.** Optimisation of variable precursor isolation windows.	R code
NOFI[109]	**DIA SWATH.** A ranking algorithm to prioritise fragment ions.	R code
MetaboDIA[112]	**DIA & DDA.** Bioinformatics workflow for untargeted metabolomics analysis.	R code
LipidMatch[110]	**DIA & DDA.** Automated workflow for the identification of lipid metabolites.	R package
MetaboKit[104]	**DIA & DDA.** Untargeted metabolomics data processing and analysis.	Python

1.1.17 Best practices in untargeted metabolomics

There has been a lack of a global standardised workflow in untargeted metabolomic techniques. This is contrasted with targeted ones where guidelines are available from most leading authorities, such as the EPA, FDA, NATA, and in an anti-doping sense the AORC and WADA to name a few.[113–117] This has led to very little, to no consistency in experimental design and data validation among published semi-targeted and untargeted assays in both metabolomic and proteomic fields.[118] One major consideration of metabolomic workflows is the implementation of quality control (QC) and assurance (QA) processes.[45,118–120] Considering the rapid expansion and adoption of metabolomic techniques over the past two decades, there remains little to no community agreed-upon guidelines for QA and QC implementation for untargeted metabolomics. This inconsistency in method validation and data assurance was recently addressed by Broadhurst *et al*, in a 2018 review. In this they make a clear argument for the standardisation of untargeted assays in a clinical and broader metabolomics setting.[118] They suggest the general adoption of standardised quality assurance (QA) and quality control (QC) protocols for the promotion of good practices in both data analysis and reporting. One example given is the use of system suitability, process blank, and pooled QC samples (pQC), along with the open dissemination of data from published work. In this proposed workflow, data quality is assured by using pQC samples. This can be easily checked by principal component analysis (PCA), where a tight clustering of the pQC samples relative to the variance among biological samples, is an indicator of good data quality (Figure 1.14).[118] Further, the reporting of data analytical pipelines is also essential to ensuring replicable and robust methodologies. This would include the reporting of data pre-treatment (normalisation, transformation, scaling, filtering) as well as any analytical tool or process applied to the data, as the change of these parameters can often influence the outcome of the data interpretation.[66] These QA and QC practices serve only to increase the reliability of metabolomic analysis on any given analytical platform. Moreover, many vendors and opensource software, such as MS-Dial and MetaboAnalyst are already supporting these types of analyses and transparent practices.[102,103]

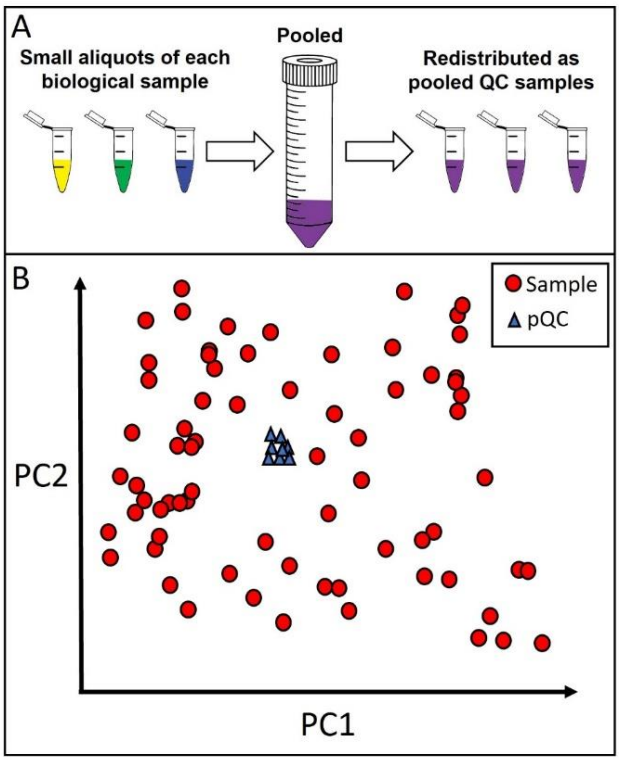


Figure 1.14: **A**) Schematic description of how pooled quality control (pQC) samples are prepared. **B**) The clustering of pooled QC samples relative to all other samples in PCA, indicating good data quality. Adapted from Broadhurst *et al* (2018).[118]

1.1.18 Metabolomic detection of AAS in human and equine sports

Metabolomic strategies for the screening of AAS doping fall under two broad categories, either employing direct or indirect detection methods. Direct methods tend to focus on the targeted detection of the steroid dopant and its immediate metabolites. Whereas, indirect methods tend to target the effect of a steroid dopant on a system related to physiological change such as perturbations in metabolic signatures.[121] Both of these approaches tend to use a mix of targeted and untargeted metabolomic methodologies; but the former tends to be sub-categorised under steroidomics due to its narrower field of analysis and are often performed as semi-targeted or targeted assays.[122,123] Both these types of approaches have strengths and limitations. Direct methods tend to be effective so long as the metabolites of interest remain present in the system. They also tend to be independent of genetic or biological factors and are usually considered as sufficient proof of malpractice. These direct methods tend to be susceptible to

degradation or clearance from the body. The narrow scope of this type of method can also make it more difficult to discover novel biomarkers. Overall, this can lead to the reduction in the length of detection windows following drug abuse. However, the analysis of intact Phase II steroid conjugates has been able to retroactively prolong these detection windows.[124–126]

Conversely, indirect strategies are ones that try to measure the metabolic signature of a steroid dopant by considering all possible metabolomic features. These techniques do not rely on specific biomarker identification and generally lead to increased detection windows for abuse based on measured perturbations to the metabolic signature. However, factors such as stress, diet, disease, and inter-individual genetic variation may also result in perturbed off-range values, leaving these methods susceptible to giving false-positive results.[120] To control for this indirect detection techniques are generally tested against populations. Further the non-specific nature of these techniques tends to make them more difficult to use in convictions, and in most cases the results must be confirmed against authentic standards.[121,126]

Recently there has been a particular focus in both human and equine anti-doping research on the use of untargeted metabolomic methodologies coupled with a supervised multivariate analysis to identify AAS biomarkers of interest. In the human anti-doping field, Boccard *et al (2011)*, has demonstrated an untargeted metabolomics work follow for the surveying and discovery of novel biomarkers following the oral administration of testosterone undecanoate.[127] Following an UHPLC-MS untargeted metabolomics acquisition of upwards of 6000 features, a pseudo-targeted steroidomics approach was taken by accurate mass alignment of raw spectral data against a list of sterol lipids from the Lipid Maps database, resulting in approximately 200 matches. As well as this untargeted analysis, they also performed a targeted assay on 10 key steroid metabolites. The combined untargeted and targeted data sets were then analysed using orthogonal projections to latent structures-discriminant analysis (OPLS-DA) using the SIMCA statistical software package. This resulted in the identification of a series of steroid glucuronide and sulfate conjugated metabolites that were significantly changed following administration.[127] A similar untargeted steroidomics study was performed

by Palermo *et al* (2017).[128] Here, they used partial least squares regression (PLS) analysis on the KNIME open-source statistical software, to annotate potential upregulated steroid conjugates following the transdermal administration of testosterone gel. Like Boccard *et al* (2011), putative annotation was done by accurate mass matching of the detected metabolic *m/z* features against a list derived from the human metabolome data base (HMDB) and ones derived from the scientific literature. Impressively, ten of these metabolites were then confirmed against authentic reference materials.[128] In both instances, untargeted metabolomic workflows were implemented and this was coupled with a pseudo-targeted steroidomics approach, using the accurate mass matching against theoretically generated lists of known steroids and metabolites. Interestingly for both applications this general approach was able to highlight significant metabolites potentially related to various steroidal biosynthetic and regulatory pathways in humans. Despite this, testosterone glucuronide was the only confirmed metabolite that was found in common between the two studies. However, there were several similar metabolites with matching accurate masses but due to differences in liquid chromatography conditions no further comparison could be made. This may indicate that introducing or developing a retention time index for the use in these types of LC-MS steroidomic studies would be of value. Other notable metabolomic applications for the detection of steroids in human systems are covered in a comprehensive review by Narduzzi *et al* (2019).[121]

Similar untargeted metabolomic analysis have also been pursued in equine sports and was comprehensively reviewed by Keen *et al* (2022).[45] In one study, Kaabia *et al* (2014), used a GC-MS analysis with a combination of OPLS-DA and statistical models to successfully highlight biomarkers that extended the detection window of the administration of the estrane AAS nandrolone, in plasma and urine matrices.[74] Following this, a similar UHPLC-HRMS/MS based study performed by Kaabia and Cloteau *et al* (2022), used selected reaction monitoring and a similar statistical method to highlight several biomarkers associated with the administration of a cocktail of testosterone esters.[129] In this case the biomarkers were not identified against reference standards and largely reflected a perturbed metabolomic state after the administration of a dopant. However, robust statistical models were able to be built around them that were indicative of testosterone ester abuse. Lastly,

Chan *et al* (2015) used OPLS-DA models to identify several steroidal biomarkers that were related to aromatase inhibitor administration.[130]

Aside from multivariate analysis the use of MS^n experiments has also been widely adopted in these steroidomic methodologies for the detection of Phase II conjugate metabolites of steroids. Here, knowledge of unimolecular fragmentation pathways characteristic for metabolite classes can be used to aid in identification. For AAS, such methods typically attempt to characterise steroids based on these common unimolecular fragmentation patterns. [6,131–133] In the case of AAS Phase II metabolites similar techniques have been used to identify compounds by their characteristic conjugate-derived fragmentation pattern.[133] As conjugated Phase II metabolites have distinct fragmentation patterns, this has been used to effectively 'target' these compounds in an untargeted assay. Various techniques have been developed for the monitoring of sulfate and glucuronide conjugate fragment ions.[18,44,131]

One study by McLeod *et al* (2017), was able to target a subclass of steroid bis(sulfate) conjugates through the monitoring the loss of hydrogen sulfate ions and the transition from doubly charged precursor to a singly charged fragment using a constant ion loss method on a triple quadrupole mass analyser, Figure 1.15. From this they were able to screen for bis(sulfate) metabolites in their sample in an untargeted manner.[131] In another example, by Pozo *et al* (2013), were able to putatively annotate steroid glucuronide conjugates based on several characteristic MS/MS transitions, in particular the neutral loss of the dehydrated glucuronide (176 Da) was highlighted as an effective transition of glucuronides.[133]

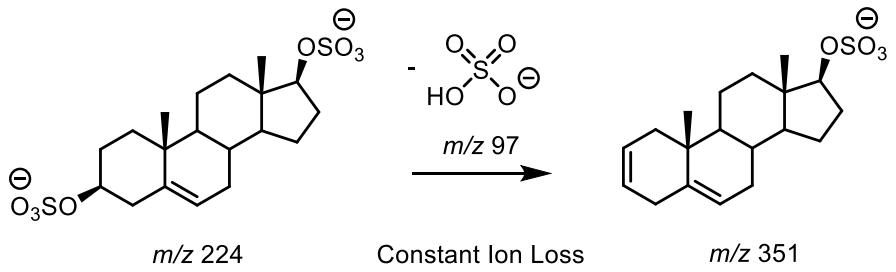


Figure 1.15: Constant ion loss (CIL) method for the untargeted detection of steroid bis(sulfate) metabolites. Adapted from McLeod *et al* (2017).[131]

The methods reviewed thus far have mainly utilised scan MS with HRMS and or triple quadrupole LRMS analysers in their untargeted metabolomic assays for anti-doping detection of AAS. Only a handful of comprehensive studies have been performed to assess the use of DIA and DDA methods for AAS detection in anti-doping fields, despite the adoption of these techniques in the wider metabolomics fields.[23,43,134]

To this end, Leung, *et al* (2021) recently described a doping control screen using vDIA for the targeted detection of prohibited substances in horse urine.[135] In this, they assessed a vDIA method on 331 basic drugs and 45 quaternary ammonium drugs, including the designer steroid 17-desmethylstanozolol. Overall, the application of DDA and DIA non-targeted acquisitions is still an emerging technique in an anti-doping context and could provide a powerful means for the detection of AAS and their metabolites.

1.1.19 Difficulties in untargeted metabolomic profiling

Although comprehensive in nature, a major hurdle common to all untargeted metabolomic profiling techniques is that they are often only qualitative or semi-quantitative in nature and can have limited ability to identify and confirm metabolites. In many cases presented in this thesis the methods tend to provide presumptive annotations of detected features, and due to these fundamental challenges, they seldom provide absolute identifications against authentic reference materials. In many cases the putative annotation can present challenges and often manifest as a lack of available data on a given metabolite. This results in low hits of a given feature to the entries in a curated database. Moreover, once a putative match is assigned, further validation steps are needed such as MS/MS fragmentation and retention time matching or confirmation against reference materials, often complicating the analysis.[43]

1.1.20 The Equine Biological Passport

Longitudinal profiling was initially implemented in the human anti-doping field through the ABP. Analogous to this, multiple versions of an equine biological passport (EBP) have been either implemented or proposed in different jurisdictions

around the world.[45,136,137] A notable effort is being made from the Gluck Equine Research Centre at the University of Kentucky, USA. In this they aim to identify specific biomarkers that can detect drug use and monitor them over time, as well as providing an omics approach with a focus on total animal welfare.[137] Another version of the EBP is also being developed by Cawley and his team at Racing NSW, Sydney, Australia, and their goals for the EBP are outlined in the perspective piece, *Intelligence-based anti-doping from an equine biological passport*.[138] In this they outline 5 essential steps in the development of an EBP. These are:

1) Approximate number of horses at any point in time that the analytical and IT infrastructure, together with human resourcing can accommodate.
2) Experimental design relating to sampling strategy – e.g., type of blood collection tubes and frequency of sample collection.
3) Requirement for ethically approved administration trials and access to research horses.
4) Statistical support for developed models in the form of specialist software and/or training.
5) Selection strategy for suitable biomarkers and method validation requirements.

This version of an EBP would be the keystone in a multifaceted forensic intelligence model for racing authorities to use as depicted in Figure 1.16. In a similar manner to WADA's ADAMs and ABP, this model would incorporate intelligence (betting trends, veterinary intelligence, surveillance), sanctioning, biomarker analysis, and routine screening in a centralised dynamic fashion. It would have the ability to provide an appropriate authority with the data needed to monitor for suspicious biological profiles that are indicative of doping. This model would also have the potential to provide a holistic view of a horse's physiology giving the added benefits of animal welfare, and act as a deterrence to would-be cheats through indirect detection of doping substances by their biological signatures.

Central to both methods are the selection of novel biomarkers to include in the EBP, with both labs identifying Omics based approaches as a powerful tool to meet this end. These would include the use of, or a combination of, genomic, proteomic

and metabolomic assays coupled with bio-informatic analysis, such as multivariate analysis, to help identify relevant biomarkers in doping detection. Overall, a move to an incorporated Omics approach for the identification of novel biomarkers would represent a broadening of the multidisciplinary anti-doping sciences into one that encompasses, forensics, small molecule chemistry, bioinformatics and metabolomics.[137,138]

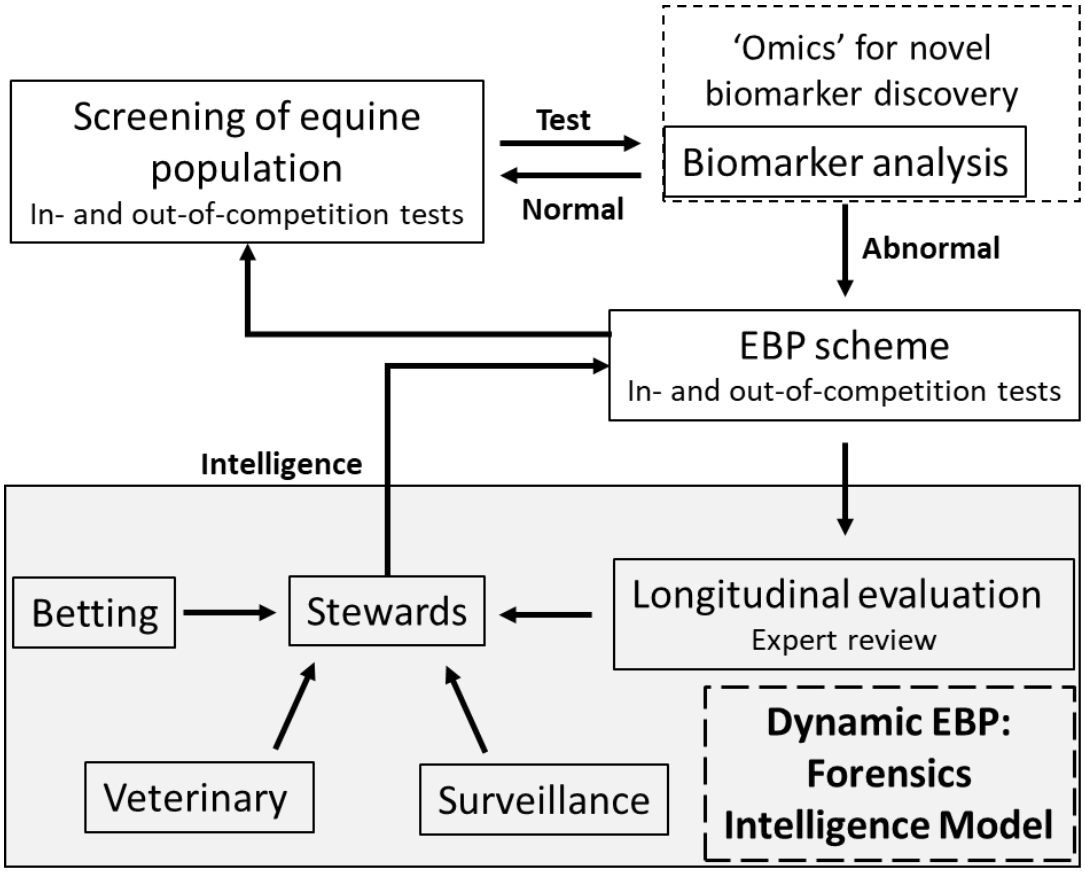


Figure 1.16: Schematic of the proposed forensics intelligence model for the selection of EBP horses. Omics would be used to discover novel biomarkers relevant for doping detection. Adapted from Cawley *et al* (2017),[138] and originally based on a model designed by Schumacher *et al* (2012) for EPO and blood doping.[139]

Thesis Aims

The study of androgenic anabolic steroids and their metabolites form an important cornerstone of anti-doping analysis in both human and equine sports. The work presented in this thesis aims to develop new methodologies for the detection and characterisation of androgenic anabolic steroids and their metabolites in the urinary matrix. Methodologies presented include a combination of UHPLC-MS, chemoinformatic and bioinformatic, and synthetic techniques applied in a multidisciplinary way over the fields of organic synthesis, analytical chemistry, and metabolomics.

Specifically, this thesis aims to undertake:

1) The development of an untargeted metabolomic profiling workflow for the detection of Phase II AAS metabolites in a urine matrix, with a focus on sulfated metabolites.
2) The application of untargeted metabolomic profiling of Phase II AAS metabolites for the identification of novel biomarkers following the administration of testosterone propionate in thoroughbred horses.
3) Synthesis of stable isotope labelled reference materials for the analysis of steroidal Phase II conjugate metabolites, including studying the fragmentation pathways of bis-conjugate metabolites.
4) Structural elucidation of metabolites using collision induced dissociation experiments.
5) The metabolomic screening of the designer steroid hemapolin following administration in a thoroughbred horse.

1.2 References

1. Moss, P. International union of pure and applied chemistry. *Analytica Chimica Acta* vol. 13 499–500 (1955).

2. Hartgens, F. & Kuipers, H. Effects of Androgenic-Anabolic Steroids in Athletes. *Sport. Med.* **34**, 513–554 (2012).

3. Conti, A. A. & Conti, A. A. Doping in Sports in Ancient and Recent Times. *Med. Secoli* **22**, 181–190 (2010).

4. Wagner, U. & Pedersen, K. M. The IOC and the doping issue-An institutional discursive approach to organizational identity construction. *Sport Manag. Rev.* **17**, 160–173 (2014).

5. Waller, C. C. & McLeod, M. D. A review of designer anabolic steroids in equine sports. *Drug Test. Anal.* **9**, 1304–1319 (2017).

6. Schanzer, W. & Thevis, M. Human Sports Drug Testing by Mass Spectrometry. *Mass Spectrom. Rev.* **36**, 16–46 (2017).

7. Alquraini, H. & Auchus, R. J. Strategies that athletes use to avoid detection of androgenic-anabolic steroid doping and sanctions. *Mol. Cell. Endocrinol.* **464**, 28–33 (2018).

8. Parr, M. K. & Schänzer, W. Detection of the misuse of steroids in doping control. *J. Steroid Biochem. Mol. Biol.* **121**, 528–537 (2010).

9. Athanasiadou, I. *et al.* Analytical progresses of the World Anti-Doping Agency Olympic laboratories: a 2016 update from London to Rio. *Bioanalysis* **8**, 2265–2279 (2016).

10. Scarth, J. P., Teale, P. & Kuuranne, T. Drug metabolism in the horse: A review. *Drug Test. Anal.* **3**, 19–53 (2011).

11. World Anti-Doping Agency (WADA). *Anti-Doping Testing Figures Report | WADA*. *WADA* https://www.wada-ama.org/en/resources/laboratories/anti-doping-testing-figures-report (2019).

12. Racing NSW. *Annual Report 2019*. *Racing NSW* https://www.racingnsw.com.au/wp-content/uploads/AR-Book-24-October-FINAL.pdf (2019).

13. Schanzer, W. Metabolism of anabolic androgenic steroids. *Clin. Chem.* **42**, 1001–1020 (1996).

14. Teale, P. & Houghton, E. Metabolism of anabolic steroids and their relevance to drug detection in horseracing. *Bioanalysis* **2**, 1085–1107 (2010).

15. Knych, H. K. D., McKemie, D. S. & Stanley, S. D. Molecular Cloning, Expression, and Initial Characterization of Members of the CYP3A Family in Horses. *Drug Metab. Dispos.* **38**, 1820–1827 (2010).

16. Schiffer, L. *et al.* Human steroid biosynthesis, metabolism and excretion are differentially reflected by serum and urine steroid metabolomes: A comprehensive review. *J. Steroid Biochem. Mol. Biol.* **194**, 105439 (2019).

17. Shackleton, C., Marcos, J., Arlt, W. & Hauffa, B. P. Prenatal diagnosis of P450 oxidoreductase deficiency (ORD): A disorder causing low pregnancy estriol, maternal and fetal virilization, and the Antley-Bixler syndrome phenotype. *Am. J. Med. Genet.* **129**, 105–112 (2004).

18. Pranata, A. *et al.* Synthesis of steroid bisglucuronide and sulfate glucuronide reference materials: Unearthing neglected treasures of steroid metabolism. *Steroids* **143**, 25–40 (2019).

19. Schanzer, W. & Donike, M. Metabolism of anabolic steroids in man: synthesis and use of reference substances for identification of anabolic steroid metabolites. *Anal. Chim. Acta*

23–48 (1992).

20. Schobinger, C., Emery, C., Schweizer-Gründisch, C. & Kuuranne, T. Support of a laboratory-hosted Athlete Biological Passport Management Unit (APMU) to the anti-doping organisations. *Rechtsmedizin* vol. 31 526–532 (2021).

21. World Anti-Doping Agency (WADA). *Athlete Biological Passport Operating Guidelines*. (2017).

22. Rodríguez-Morató, J., Pozo, Ó. J. & Marcos, J. Targeting human urinary metabolome by LC–MS/MS: a review. *Bioanalysis* **10**, 489–516 (2018).

23. Wudy, S. A., Schuler, G., Sánchez-Guijo, A. & Hartmann, M. F. The art of measuring steroids: Principles and practice of current hormonal steroid analysis. *J. Steroid Biochem. Mol. Biol.* **179**, 88–103 (2018).

24. Uduwela, D. R., Pabis, A., Stevenson, B. J., Kamerlin, S. C. L. & McLeod, M. D. Enhancing the Steroid Sulfatase Activity of the Arylsulfatase from *Pseudomonas aeruginosa*. *ACS Catal.* **8**, 8902–8914 (2018).

25. Iannone, M., Botrè, F., Martinez-Brito, D., Matteucci, R. & de la Torre, X. Development and application of analytical procedures for the GC–MS/MS analysis of the sulfates metabolites of anabolic androgenic steroids: The pivotal role of chemical hydrolysis. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **1155**, (2020).

26. El-Aneed, A., Cohen, A. & Banoub, J. Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Appl. Spectrosc. Rev.* **44**, 210–230 (2009).

27. Crotti, S., Seraglia, R. & Traldi, P. Some Thoughts on Electrospray Ionization Mechanisms: *Eur. J. Mass Spectrom.* **17**, 85–100 (2011).

28. McLuckey, S. A. & Wells, J. M. Mass analysis at the advent of the 21st century. *Chem. Rev.* **101**, 571–606 (2001).

29. Gomez, C. *et al.* Analytical strategies based on mass spectrometric techniques for the study of steroid metabolism. *TrAC - Trends Anal. Chem.* **53**, 106–116 (2014).

30. Schwartz, J. C., Senko, M. W. & Syka, J. E. P. A Two-Dimensional Quadrupole Ion Trap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* 659–669 (2002).

31. March, R. E. Quadrupole ion traps. *Mass Spectrom. Rev.* **28**, 961–989 (2009).

32. Douglas, D. J., Frank, A. J. & Mao, D. Linear ion traps in mass spectrometry. *Mass Spectrom. Rev.* **24**, 1–29 (2005).

33. Hecht, E. S., Scigelova, M., Eliuk, S. & Makarov, A. Fundamentals and Advances of Orbitrap Mass Spectrometry. *Encycl. Anal. Chem.* 1–40 (2019) doi:10.1002/9780470027318.A9309.PUB2.

34. Eliuk, S. & Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu. Rev. Anal. Chem.* **8**, 61–80 (2015).

35. Brown, S. C., Kruppa, G. & Dasseux, J. L. Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrom. Rev.* **24**, 223–231 (2005).

36. Nikolaev, E. N., Kostyukevich, Y. I. & Vladimirov, G. N. Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: Theory and simulations. *Mass Spectrom. Rev.* **35**, 219–258 (2016).

37. Gaudl, A., Kratzsch, J. & Ceglarek, U. Advancement in steroid hormone analysis by LC–MS/MS in clinical routine diagnostics – A three year recap from serum cortisol to dried blood 17α-hydroxyprogesterone. *J. Steroid Biochem. Mol. Biol.* **192**, 105389 (2019).

38. Allen, D. R. & McWhinney, B. C. Quadrupole Time-of-Flight Mass Spectrometry: A Paradigm Shift in Toxicology Screening Applications. *Clin. Biochem. Rev.* **40**, 135 (2019).

39. Kingdon, K. H. A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Phys. Rev.* **21**, 408 (1923).

40. Perry, R. H., Cooks, R. G. & Noll, R. J. Orbitrap mass spectrometry: Instrumentation, ion motion and applications. *Mass Spectrom. Rev.* **27**, 661–699 (2008).

41. Cawley, A. *et al.* The potential for complementary targeted/non-targeted screening of novel psychoactive substances in equine urine using liquid chromatography-high resolution accurate mass spectrometry. *Anal. Methods* **8**, 1789–1797 (2016).

42. Cajka, T. & Fiehn, O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal. Chem.* **88**, 524–545 (2016).

43. Reichel, C. OMICS-strategies and methods in the fight against doping. *Forensic Sci. Int.* **213**, 20–34 (2011).

44. Pozo, O. J. *et al.* Alternate steroid sulfation pathways targeted by LC–MS/MS analysis of disulfates: Application to prenatal diagnosis of steroid synthesis disorders. *J. Mol. Endocrinol.* **61**, M1–M12 (2018).

45. Keen, B., Cawley, A., Reedy, B. & Fu, S. Metabolomics in clinical and forensic toxicology, sports anti-doping, and veterinary residues. *Drug Test. Anal.* https://doi.org/10.1002/dta.3245 (2022) doi:10.1002/dta.3245.

46. Roberts, L. D., Souza, A. L., Gerszten, R. E. & Clish, C. B. Targeted metabolomics. *Curr. Protoc. Mol. Biol.* **1**, 30.2.1-30.2.24 (2012).

47. Ussher, J. R., Elmariah, S., Gerszten, R. E. & Dyck, J. R. B. The Emerging Role of Metabolomics in the Diagnosis and Prognosis of Cardiovascular Disease. *J. Am. Coll. Cardiol.* **68**, 2850–2870 (2016).

48. Ponzetto, F. *et al.* Longitudinal monitoring of endogenous steroids in human serum by UHPLC-MS/MS as a tool to detect testosterone abuse in sports. *Anal. Bioanal. Chem. 2015 4083* **408**, 705–719 (2015).

49. Nair, V. S. *et al.* Evaluation of longitudinal steroid profiling with the ADAMS adaptive model for detection of transdermal, intramuscular, and subcutaneous testosterone administration. *Drug Test. Anal.* **12**, 1419–1431 (2020).

50. Le Bouc, Y. *et al. Metabonomic signature in bike athletes: a pilot study*. *WADA* https://www.wada-ama.org/sites/default/files/resources/files/lebouc_0.pdf (2006).

51. Botrè, F., Georgakopoulos, C. & Elrayess, M. A. Metabolomics and doping analysis: promises and pitfalls. *Futur. Sci.* **12**, 719–722 (2020).

52. Scalbert, A. *et al.* Mass-spectrometry-based metabolomics: Limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* **5**, 435–458 (2009).

53. Okazaki, Y. & Saito, K. Recent advances of metabolomics in plant biotechnology. *Plant Biotechnol. Reports 2011 61* **6**, 1–15 (2011).

54. Hong, J., Yang, L., Zhang, D. & Shi, J. Plant metabolomics: An indispensable system biology tool for plant science. *Int. J. Mol. Sci.* **17**, (2016).

55. Jacob, M., Lopata, A. L., Dasouki, M. & Abdel Rahman, A. M. Metabolomics toward personalized medicine. *Mass Spectrom. Rev.* **38**, 221–238 (2019).

56. Gowda, G. A. N. *et al.* Metabolomics-based methods for early disease diagnostics. *Expert Rev. Mol. Diagn.* **8**, 617–633 (2008).

57. Anh, N. H. *et al.* Steroidomics for the prevention, assessment, and management of cancers: A systematic review and functional analysis. *Metabolites* **9**, https://doi.org/10.3390/metabo9100199 (2019).

58. Beger, R. A Review of Applications of Metabolomics in Cancer. *Metabolites* **3**, 552–574

(2013).

59. Monnerie, S. *et al.* Metabolomic and Lipidomic Signatures of Metabolic Syndrome and its Physiological Components in Adults: A Systematic Review. *Sci. Reports 2020 101* **10**, 1–13 (2020).

60. Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D. & Mclean, J. A. Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *J. Am. Soc. Mass Spectrom.* **27**, 1897–1905 (2016).

61. Nikolskiy, I., Mahieu, N. G., Chen, Y.-J., Tautenhahn, R. & Patti, G. J. An Untargeted Metabolomic Workflow to Improve Structural Characterization of Metabolites. *Anal. Chem.* **85**, 7713–7719 (2013).

62. Dettmer, K., Aronov, P. A. & Hammock, B. D. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **26**, 51–78 (2007).

63. Sen, P. *et al.* Deep learning meets metabolomics: A methodological perspective. *Brief. Bioinform.* **22**, 1531–1542 (2021).

64. Micheli, A. & Podda, M. Deep Learning in Cheminformatics. *Deep Learn. Biol. Med.* 157–195 (2022) doi:10.1142/9781800610941_0006.

65. Skinnider, M. A. *et al.* A deep generative model enables automated structure elucidation of novel psychoactive substances. *Nat. Mach. Intell.* **3**, 973–984 (2021).

66. Cook, T., Ma, Y. & Gamagedara, S. Evaluation of statistical techniques to normalize mass spectrometry-based urinary metabolomics data. *J. Pharm. Biomed. Anal.* **177**, 112854 (2020).

67. Issaq, H. J., Van, Q. N., Waybright, T. J., Muschik, G. M. & Veenstra, T. D. Analytical and statistical approaches to metabolomics research. *J. Sep. Sci.* **32**, 2183–2199 (2009).

68. Wishart, D. S. Advances in metabolite identification. *Futur. Sci.* **3**, 1769–1782 (2011).

69. Wishart, D. S. Computational strategies for metabolite identification in metabolomics. *Bioanalysis* vol. 1 1579–1596 (2009).

70. Kwak, M., Kang, K. & Wang, Y. Methods of Metabolite Identification Using MS/MS Data. *J. Comput. Inf. Syst.* **62**, 12–18 (2022).

71. Joré, C. *et al.* Liquid chromatography – high resolution mass spectrometry-based metabolomic approach for the detection of Continuous Erythropoiesis Receptor Activator effects in horse doping control. *J. Chromatogr. A* **1521**, 90–99 (2017).

72. Vinayavekhin, N. & Saghatelian, A. Untargeted metabolomics. *Current Protocols in Molecular Biology* vol. 90 1–24 (2010).

73. Waller, C. C. & McLeod, M. D. A simple method for the small scale synthesis and solid-phase extraction purification of steroid sulfates. *Steroids* **92**, 74–80 (2014).

74. Kaabia, Z. *et al.* Monitoring the endogenous steroid profile disruption in urine and blood upon nandrolone administration: An efficient and innovative strategy to screen for nandrolone abuse in entire male horses. *Drug Test. Anal.* **6**, 376–388 (2014).

75. Hintikka, L. *et al.* Liquid chromatographic–mass spectrometric analysis of glucuronide-conjugated anabolic steroid metabolites: method validation and interlaboratory comparison. *J. Mass Spectrom.* **43**, 965–973 (2008).

76. Esquivel, A., Alechaga, É., Monfort, N. & Ventura, R. Direct quantitation of endogenous steroid sulfates in human urine by liquid chromatography-electrospray tandem mass spectrometry. *Drug Test. Anal.* **10**, 1734–1743 (2018).

77. Guddat, S. *et al.* Synthesis, characterization, and detection of new oxandrolone metabolites as long-term markers in sports drug testing. *Anal. Bioanal. Chem.* **405**, 8285–8294 (2013).

78. Shah, I., Petroczi, A., Uvacsek, M., Ránky, M. & Naughton, D. P. Hair-based rapid analyses for multiple drugs in forensics and doping: application of dynamic multiple reaction monitoring with LC-MS/MS. *Chem. Cent. J. 2014 81* **8**, 1–10 (2014).

79. Guo, B., Chen, B., Liu, A., Zhu, W. & Yao, S. Liquid Chromatography-Mass Spectrometric Multiple Reaction Monitoring-based Strategies for Expanding Targeted Profiling towards Quantitative Metabolomics. *Curr. Drug Metab.* **13**, 1226–1243 (2012).

80. Lee, C. *et al.* Selective LC-MRM/SIM-MS based profiling of adrenal steroids reveals metabolic signatures of 17α-hydroxylase deficiency. *J. Steroid Biochem. Mol. Biol.* **198**, 105615 (2020).

81. Wang, Z., Zhou, X., Liu, X., Dong, Y. & Zhang, J. A novel HPLC-MRM strategy to discover unknown and long-term metabolites of stanozolol for expanding analytical possibilities in doping-control. *J. Chromatogr. B* **1040**, 250–259 (2017).

82. Hintikka, L. *et al.* Liquid chromatographic–mass spectrometric analysis of glucuronide-conjugated anabolic steroid metabolites: method validation and interlaboratory comparison. *J. Mass Spectrom.* **43**, 965–973 (2008).

83. Thevis, M. & Schänzer, W. Mass spectrometry in sports drug testing: Structure characterization and analytical assays. *Mass Spectrom. Rev.* **26**, 79–107 (2007).

84. Krasny, L. & Huang, P. H. Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. *Mol. Omi.* **17**, 29–42 (2021).

85. Guo, J. & Huan, T. Comparison of Full-Scan, Data-Dependent, and Data-Independent Acquisition Modes in Liquid Chromatography-Mass Spectrometry Based Untargeted Metabolomics. *Anal. Chem.* **92**, 8072–8080 (2020).

86. Fenaille, F., Barbier Saint-Hilaire, P., Rousseau, K. & Junot, C. Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand? *J. Chromatogr. A* **1526**, 1–12 (2017).

87. Lafaye, A. *et al.* 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 (2004).

88. Lutz, U., Lutz, R. W. & Lutz, W. K. Metabolic profiling of glucuronides in human urine by LC-MS/MS and partial least-squares discriminant analysis for classification and prediction of gender. *Anal. Chem.* **78**, 4564–4571 (2006).

89. Libert, R. *et al.* Bile acids and conjugates identified in metabolic disorders by fast atom bombardment and tandem mass spectrometry. *Clin. Chem.* **37**, 2102–2110 (1991).

90. Chace, D. H. Mass spectrometry in the clinical laboratory. *Chem. Rev.* **101**, 445–477 (2001).

91. Lawson, T. N. *et al.* msPurity: Automated Evaluation of Precursor Ion Purity for Mass Spectrometry-Based Fragmentation in Metabolomics. *Anal. Chem.* **89**, 2432–2439 (2017).

92. Hopfgartner, G., Tonoli, D. & Varesio, E. High-resolution mass spectrometry for integrated qualitative and quantitative analysis of pharmaceuticals in biological matrices. *Anal. Bioanal. Chem.* **402**, 2587–2596 (2012).

93. Benton, H. P. *et al.* Autonomous Metabolomics for Rapid Metabolite Identification in Global Profiling. *Anal. Chem.* **87**, 884–891 (2014).

94. Zhang, F., Ge, W., Ruan, G., Cai, X. & Guo, T. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **20**, 1900276 (2020).

95. Plumb, R. S. *et al.* UPLC/MSE; a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun. Mass Spectrom.* **20**, 1989–1994 (2006).

96. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, O111.016717 (2012).

97. Li, K. W., Gonzalez-Lozano, M. A., Koopmans, F. & Smit, A. B. Recent Developments in Data Independent Acquisition (DIA) Mass Spectrometry: Application of Quantitative Analysis of the Brain Proteome. *Front. Mol. Neurosci.* **13**, 248 (2020).

98. Zhu, X., Chen, Y. & Subramanian, R. Comparison of Information-Dependent Acquisition, SWATH, and MS All Techniques in Metabolite Identification Study Employing Ultrahigh-Performance Liquid Chromatography−Quadrupole Time-of-Flight Mass Spectrometry The predominant tool for identification of drug biotransfor-mation products is liquid chromatography coupled with mass. *Anal. Chem* **86**, 22 (2014).

99. Arnhard, K., Gottschall, A., Pitterl, F. & Oberacher, H. Applying 'Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra' (SWATH) for systematic toxicological analysis with liquid chromatography-high-resolution tandem mass spectrometry. *Anal. Bioanal. Chem.* **407**, 405–414 (2015).

100. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787 (2006).

101. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 1–11 (2010).

102. Tsugawa, H. *et al.* MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **12**, 523–526 (2015).

103. Pang, Z. *et al.* MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, 388–396 (2021).

104. Narayanaswamy, P. *et al.* MetaboKit: A comprehensive data extraction tool for untargeted metabolomics. *Mol. Omi.* **16**, 436–447 (2020).

105. Neumann, S., Thum, A. & Böttcher, C. Nearline acquisition and processing of liquid chromatography-tandem mass spectrometry data. *Metabolomics* **9**, 84–91 (2013).

106. Koelmel, J. P. *et al.* Expanding Lipidome Coverage Using LC-MS/MS Data-Dependent Acquisition with Automated Exclusion List Generation. *J. Am. Soc. Mass Spectrom.* **28**, 908–917 (2017).

107. Zhang, Y. *et al.* The Use of Variable Q1 Isolation Windows Improves Selectivity in LC-SWATH-MS Acquisition. *J. Proteome Res.* **14**, 4359–4371 (2015).

108. Li, H., Cai, Y., Guo, Y., Chen, F. & Zhu, Z.-J. MetDIA: Targeted Metabolite Extraction of Multiplexed MS/MS Spectra Generated by Data-Independent Acquisition. *Anal. Chem.* **88**, 8757–8764 (2016).

109. Bilbao, A. *et al.* Ranking fragment ions based on outlier detection for improved label-free quantification in data-independent acquisition LC-MS/MS. *J. Proteome Res.* **14**, 4581–4593 (2015).

110. Koelmel, J. P. *et al.* LipidMatch: An automated workflow for rule-based lipid identification using untargeted high-resolution tandem mass spectrometry data. *BMC Bioinformatics* **18**, 1–11 (2017).

111. Röst, H. L. *et al.* OpenMS: A flexible open-source software platform for mass spectrometry data analysis. *Nat. Methods* **13**, 741–748 (2016).

112. Chen, G. *et al.* Customized Consensus Spectral Library Building for Untargeted Quantitative Metabolomics Analysis with Data Independent Acquisition Mass Spectrometry and MetaboDIA Workflow. *Anal. Chem.* **89**, 4897–4906 (2017).

113. United States Environmental Protection Agency. *Appendix B to Part 136 - Definition and procedure for the determination of the method detection limit - Revision 1.11. Code of Federal Regulations*. *Code of Federal Regulations* https://www.gpo.gov/fdsys/pkg/CFR-2011-title40-vol23/pdf/CFR-2011-title40-vol23-part136-appB.pdf (2011).

114. Food and Drug Administration. *Guidance for Industry Bioanalytical Method Validation Guidance for Industry Bioanalytical Method Validation*. *Biopharmaceutics Federal register* http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htmand/orhttp://www.fda.gov/AnimalVeterinary/GuidanceComplianceEnforcement/GuidanceforIndustry/default.htm (2018).

115. NATA. *General Accreditation Guidance-Validation and verification of quantitative and qualitative test methods written authority from NATA. Validation and verification of quantitative and qualitative test methods*. https://www.nata.com.au/phocadownload/gen-accreditation-guidance/Validation-and-Verification-of-Quantitative-and-Qualitative-Test-Methods.pdf (2018).

116. Association of Offical Racing Chemist. AORC MS Criteria (modified 23 Aug 16) - Association of Official Racing Chemists. http://www.aorc-online.org/documents/aorc-ms-criteria-modified-23-aug-16/.

117. World Anti-Doping Agency (WADA). *World Anti-Doping Code 2015*. https://www.wada-ama.org/en/resources/the-code/world-anti-doping-code (2015).

118. Broadhurst, D. *et al.* Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **14**, 72 (2018).

119. Dudzik, D., Barbas-Bernardos, C., García, A. & Barbas, C. Quality assurance procedures for mass spectrometry untargeted metabolomics. a review. *J. Pharm. Biomed. Anal.* **147**, 149–173 (2018).

120. Godzien, J., Alonso-Herranz, V., Barbas, C. & Armitage, E. G. Controlling the quality of metabolomics data: new strategies to get the best out of the QC sample. *Metabolomics* **11**, 518–528 (2015).

121. Narduzzi, L., Dervilly, G., Audran, M., Le Bizec, B. & Buisson, C. A role for metabolomics in the antidoping toolbox? *Drug Test. Anal.* **12**, 677–690 (2020).

122. Eisenhofer, G., Durán, C., Chavakis, T. & Cannistraci, C. V. Steroid metabolomics: machine learning and multidimensional diagnostics for adrenal cortical tumors, hyperplasias, and related disorders. *Curr. Opin. Endocr. Metab. Res.* **8**, 40–49 (2019).

123. Olesti, E., Boccard, J., Visconti, G., González-Ruiz, V. & Rudaz, S. From a single steroid to the steroidome: Trends and analytical challenges. *J. Steroid Biochem. Mol. Biol.* **206**, 105797 (2021).

124. Marcos, J. & Pozo, O. J. Current LC–MS methods and procedures applied to the identification of new steroid metabolites. *J. Steroid Biochem. Mol. Biol.* **162**, 41–56 (2016).

125. Esquivel, A., Alechaga, É., Monfort, N. & Ventura, R. Sulfate metabolites improve retrospectivity after oral testosterone administration. *Drug Test. Anal.* **11**, 392–402 (2019).

126. Protti, M., Mandrioli, R. & Mercolini, L. Perspectives and strategies for anti-doping analysis. *Bioanalysis* **11**, 149–152 (2019).

127. Boccard, J. *et al.* A steroidomic approach for biomarkers discovery in doping control. *Forensic Sci. Int.* **213**, 85–94 (2011).

128. Palermo, A., Botrè, F., de la Torre, X. & Zamboni, N. Non-targeted LC-MS based metabolomics analysis of the urinary steroidal profile. *Anal. Chim. Acta* **964**, 112–122 (2017).

129. Cloteau, C. *et al.* From a non-targeted metabolomics approach to a targeted biomarkers strategy to highlight testosterone abuse in equine. Illustration of a methodological transfer

between platforms and laboratories. *Drug Test. Anal.* https://doi.org/10.1002/dta.3221 (2022) doi:10.1002/dta.3221.

130. Ho, G. *et al.* Targeted Metabolomics Approach To Detect the Misuse of Steroidal Aromatase Inhibitors in Equine Sports by Biomarker Profiling. (2015) doi:10.1021/acs.analchem.5b03165.

131. McLeod, M. D. *et al.* Constant Ion Loss Method for the Untargeted Detection of Bis-sulfate Metabolites. *Anal. Chem.* **89**, 1602–1609 (2017).

132. Pozo, O. J. *et al.* Steroid sulfation pathways targeted by disulfates determination. Application to prenatal diagnosis. *J. Mol. Endocrinol.* JME-17-0286 (2018) doi:10.1530/JME-17-0286.

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

134. Mullard, G. *et al.* A new strategy for MS/MS data acquisition applying multiple data dependent experiments on Orbitrap mass spectrometers in non-targeted metabolomic applications. *Metabolomics* **11**, 1068–1080 (2015).

135. Leung, D. K. K., Wong, A. S. Y., Zhou, Q. L., Wan, T. S. M. & Ho, E. N. M. Application of a non-target variable data independent workflow (vDIA) for the screening of prohibited substances in doping control testing. *Drug Test. Anal.* **13**, 1008–1033 (2021).

136. Thornton, T. . The Equine Biological Passport: Has Its Time Come? *THOROUGHBRED DAILY NEWS* https://www.thoroughbreddailynews.com/the-equine-biological-passport-has-its-time-come-shared-archive/.

137. Horohov, D. Equine Biological Passport | Gluck Equine Research Center. *University of Kentucky* http://gluck.ca.uky.edu/content/equine-biological-passport (2021).

138. Cawley, A. T. & Keledjian, J. Intelligence-based anti-doping from an equine biological passport. *Drug Test. Anal.* **9**, 1441–1447 (2017).

139. Schumacher, Y. O., Saugy, M., Pottgiesser, T. & Robinson, N. Detection of EPO doping and blood doping: the haematological module of the Athlete Biological Passport. *Drug Test. Anal.* **4**, 846–853 (2012).

CHAPTER TWO

Untargeted Profiling of Conjugated Metabolites in the Urinary Metabolome.

Publication:

Reproduced from Fitzgerald, C. C. J. et al. Profiling Urinary Sulfate Metabolites With Mass Spectrometry. *Front. Mol. Biosci*. **9**, 829511 (2022). https://doi.org/10.3389/fmolb.2022.829511.

2.1 Preface

The work presented in this chapter has partly been described in the journal, *Frontiers in Molecular Biosciences – Cellular Biochemistry*. It details a workflow for the untargeted metabolic profiling of sulfated metabolites in the urine matrix.

This publication and supporting information (100 pages containing experimental procedures, characterisation data, Python and R scripts, and further analysis relevant to the main manuscript), was primarily authored by Mr. Fitzgerald, however, it represents a highly collaborative work. The co-contributors and contributions of whom are thankfully acknowledged as:

Dr. Rikard Hedman, Dr. Dimanthi Uduwela, Ms. Bettina Paszerbovics, Dr. Adam Carroll, Dr. Teresa Neeman, Dr. Adam Cawley, Dr. Lance Brooker, and Associate Professor Malcom McLeod,

Contributions as given in the manuscript follows the Contributor Role Taxonomy (CRediT) format and are as follows:

CF: Conceptualization, Methodology, Synthesis, Script, Validation, and Writing—Original Draft. RH: Conceptualization and Methodology. DU: Conceptualization and Methodology. BP: Methodology. A, Carroll: Conceptualization and Script. TN: Methodology and Script. A, Cawley: Conceptualization and Funding. LB: Conceptualization and Methodology. MM: Conceptualization, Methodology, Funding, and Writing—Supervision, review and editing.

2.1.1 Specific contributions

The work presented in the following manuscript was initially conceived as a brief analysis in Dr. Uduwela's PhD thesis and formed a major part of Dr. Hedman's master's thesis. Mr. Fitzgerald acknowledges that the research presented in this chapter would not be possible without this initial pilot work. The work presented herein, uses the same initial animal experiments, UHPLC conditions and GC-MS experiments conceived of previously. However, this body of work primarily comprises of Mr. Fitzgerald's original research, including the synthesis of all compounds, development of sample extraction methodology, assay design and implementation and, all data-pipelines and computational analysis.

Other notable contributions, in which thanks are gratefully extended, are:

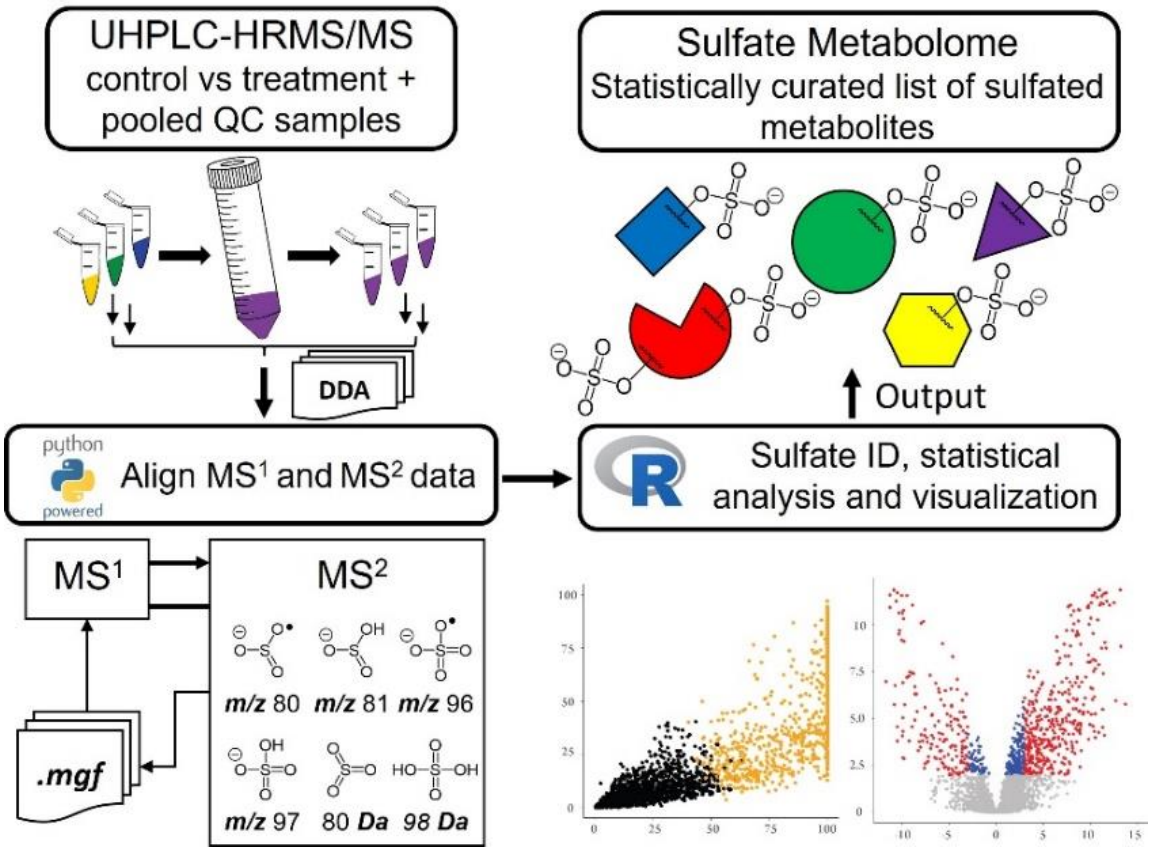
Ms. Paszerbovics who was a master's research student working with Mr. Fitzgerald. Specifically, she was instrumental in the development of the para-nitrophenol (PNP) assay and establishment of initial UHPLC conditions for the equine urine that were used in the final analysis (see supplementary information). The work she undertook was presented as a part fulfilment of her master's thesis.

Dr. Carroll who wrote the initial version of the Spectrum Analyser script in Python, that is used throughout this thesis.

Dr. Neeman who co-wrote the script for the *Limma* analysis performed in R with Mr. Fitzgerald and provided robust advice regarding statistical analysis.

Dr. Hedman who performed the initial optimisation of collision energies for steroid sulfates used in the final UHPLC-HRMS/MS analysis in both equine and human samples.

Dr. Brooker along with Dr. Hedman who performed the GC-MS assay of steroids in pooled human urine.



Scheme 2.1 Profiling Urinary Sulfate Metabolites with Mass Spectrometry. The R logo is licenced under the CC-BY-SA 4.0, https://creativecommons.org/licenses/by-sa/4.0/, with credit given to the, © 2016 The R foundation. The Python two snake logo is a trademark of the Python Software Foundation. Original artwork by Mr. Fitzgerald.

2.1.2 Summary

The study of urinary Phase II sulfate metabolites is central to understanding the role and fate of endogenous and exogenous compounds in biological systems. This has been driven in part by the developments and improvements in LC-MS technologies allowing for the direct detection of these metabolites. When considering the analysis of sulfated metabolites, although there exist robust synthetic methods for the synthesis of sulfated metabolites, especially in the case of steroid sulfates, these studies are usually performed as targeted ones, that do not consider the whole metabolome. This introduces the possibility of missing potentially important metabolic markers.

This chapter describes the development of data analytical tools which were made to specifically study sulfated metabolites in the urine matrix in an untargeted

fashion. In the first part of the chapter, presented as a manuscript, the development of a data workflow that identifies sulfated conjugates through *k-means* clustering analysis of sulfate ester derived MS/MS fragmentation intensities, is described. This was performed using ultra-high-performance liquid chromatography-high resolution tandem mass spectrometry (UHPLC-HRMS/MS) with data dependent acquisition (DDA) coupled to an automated script-based data processing pipeline and differential metabolite level analysis written by Mr. Fitzgerald with contributions from Dr. Neeman, and Dr. Carroll.

The utility of the method was highlighted in two applications. Firstly, the urinary metabolome of a thoroughbred gelding horse was examined before and after administration of the anabolic androgenic steroid (AAS) testosterone propionate. The analysis detected elevated levels of ten sulfated steroid metabolites, three of which were identified and confirmed by comparison with synthesised reference materials. This included 5α-androstane-3β,17α-diol 3-sulfate, a previously unreported equine metabolite of testosterone propionate. Secondly, the hydrolytic activity of four sulfatase enzymes on samples of pooled human urine was examined. This revealed that *Pseudomonas aeruginosa* arylsulfatases (PaS) enzymes possessed higher selectivity for the hydrolysis of sulfated metabolites than the commercially available *Helix pomatia* arylsulfatase (HpS). This was further confirmed in an adjacent GC-MS study. This method provides a rapid tool for the systematic, untargeted metabolic profiling of urinary metabolites with annotation of sulfated metabolites in a urinary matrix. This method is used throughout this thesis as a first pass exploratory approach to survey and annotate sulfated metabolites in the urinary metabolome.

The second part of the chapter describes the application of this workflow to examine the change over time of the urinary metabolomes of three horses following the administration of testosterone propionate. In this, a semi-targeted steroidomics approach is taken for the discovery of relevant biomarkers of doping. Metabolomic features are evaluated using orthogonal projections to latent structures-discriminant analysis (OPLS-DA) classification models to assess their predictive power and are performed over the kinetic excretion profiles of the three horses. Steroidal Phase II conjugated metabolites were then selected from these markers

by automatic peak selection through the matching of accurate mass to charge ratios to a list of known or derived steroidal metabolites. This highlighted a series of putatively assigned sulfated and glucuronylated steroid metabolites that could be related to testosterone propionate administration in thoroughbred horses. This method demonstrates the application of an untargeted metabolomic workflow as a first-pass method for the surveying and identification of new biomarkers of steroid administration in the equine urinary metabolome. These biomarkers can now be assessed in future studies to confirm their identify and to decide whether they will be incorporated into the routine anti-doping workflow or in longitudinal analysis.

2.2 Profiling Urinary Sulfate Metabolites with Mass Spectrometry

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 23 February 2022
doi: 10.3389/fmolb.2022.829511

Profiling Urinary Sulfate Metabolites With Mass Spectrometry

*Christopher C. J. Fitzgerald[1], Rikard Hedman[1], Dimanthi R. Uduwela[1], Bettina Paszerbovics[1], Adam J. Carroll[1], Teresa Neeman[1], Adam Cawley[2], Lance Brooker[3] and Malcolm D. McLeod[1]**

[1]Research School of Chemistry, Australian National University, Acton, ACT, Australia, [2]Australian Racing Forensic Laboratory, Racing NSW, Sydney, NSW, Australia, [3]Australian Sports Drug Testing Laboratory, National Measurement Institute, Sydney, NSW, Australia

The study of urinary phase II sulfate metabolites is central to understanding the role and fate of endogenous and exogenous compounds in biological systems. This study describes a new workflow for the untargeted metabolic profiling of sulfated metabolites in a urine matrix. Analysis was performed using ultra-high-performance liquid chromatography-high resolution tandem mass spectrometry (UHPLC-HRMS/MS) with data dependent acquisition (DDA) coupled to an automated script-based data processing pipeline and differential metabolite level analysis. Sulfates were identified through *k*-means clustering analysis of sulfate ester derived MS/MS fragmentation intensities. The utility of the method was highlighted in two applications. Firstly, the urinary metabolome of a thoroughbred horse was examined before and after administration of the anabolic androgenic steroid (AAS) testosterone propionate. The analysis detected elevated levels of ten sulfated steroid metabolites, three of which were identified and confirmed by comparison with synthesised reference materials. This included 5α-androstane-3β,17α-diol 3-sulfate, a previously unreported equine metabolite of testosterone propionate. Secondly, the hydrolytic activity of four sulfatase enzymes on pooled human urine was examined. This revealed that *Pseudomonas aeruginosa* arylsulfatases (PaS) enzymes possessed higher selectivity for the hydrolysis of sulfated metabolites than the commercially available *Helix pomatia* arylsulfatase (HpS). This novel method provides a rapid tool for the systematic, untargeted metabolic profiling of sulfated metabolites in a urinary matrix.

Keywords: sulfation, mass spectrometry, metabolomics, steroid, anti-doping, sulfatase, sulfate ester

OPEN ACCESS

Edited by:
Tarsis G. Ferreira, University of Houston, United States

Reviewed by:
Ernesto Satoshi Nakayasu, Pacific Northwest National Laboratory (DOE), United States
Fernando Ogata, Federal University of São Paulo, Brazil

****Correspondence:***
Malcolm D. McLeod
malcolm.mcleod@anu.edu.au

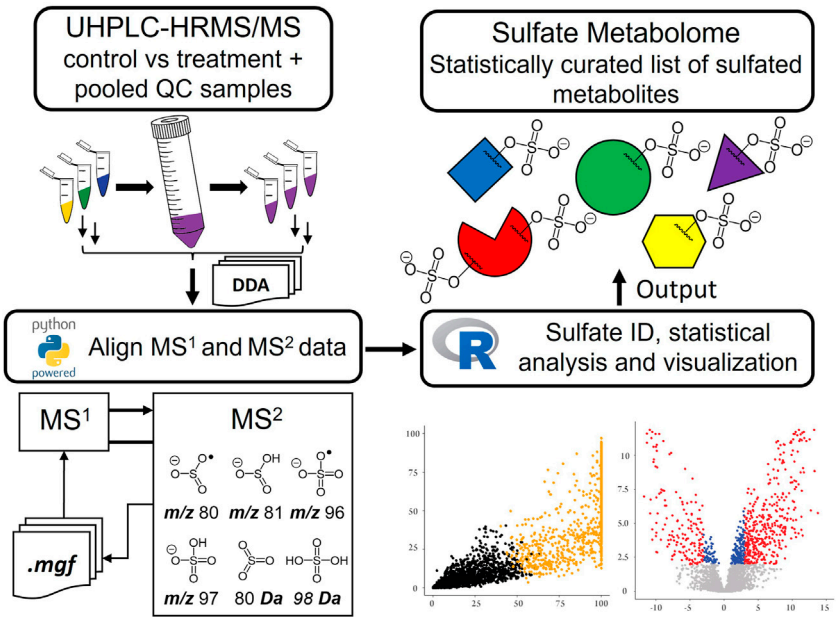
Specialty section:
This article was submitted to Cellular Biochemistry, a section of the journal Frontiers in Molecular Biosciences

Received: *05 December 2021*
Accepted: *12 January 2022*
Published: *23 February 2022*

Citation:
Fitzgerald CCJ, Hedman R, Uduwela DR, Paszerbovics B, Carroll AJ, Neeman T, Cawley A, Brooker L and McLeod MD (2022) Profiling Urinary Sulfate Metabolites With Mass Spectrometry. Front. Mol. Biosci. 9:829511. doi: 10.3389/fmolb.2022.829511

INTRODUCTION

The study of phase II metabolism is essential to understand the biochemical role and fate of endogenous and exogenous compounds and is dominated by the two major classes of conjugates: sulfates and glucuronides (Schänzer, 1996). A compelling example of the importance of phase II conjugation is provided by the field of steroid metabolism, with the conjugates accounting for up to 97% of excreted urinary metabolites (Pranata et al., 2019). Glucuronylation, performed by uridine 5′-diphospho-glucuronosyltransferases (UGTs) has traditionally been the main focus of these investigations, as the major excretory phase II pathway, and due to the ready availability of β-glucuronidase enzymes for selective deconjugation. More recently, the study of sulfate conjugates has gained prominence due to the intriguing interplay between sulfotransferase (SULT) mediated



GRAPHICAL ABSTRACT |

synthesis, sulfatase promoted hydrolysis and a range of transport phenomena (Foster and Mueller, 2018; Uduwela et al., 2018; Günal et al., 2019). The systematic study of the sulfated metabolome in a variety of contexts will be pivotal in revealing the roles of sulfation pathways in biology.

The field of metabolomics initially emerged as a complementary approach to genomics and proteomics (Fiehn et al., 2000; Lafaye et al., 2004). However, metabolic profiling has grown into its own field and serves as a powerful way to identify biomarkers in an unbiased fashion, including with relative quantification (Fiehn et al., 2000; Tolstikov and Fiehn, 2002). Metabolomic methods can be applied to the study of disease states or perturbations to homeostasis, giving unprecedented insight into biological processes (Antignac et al., 2005; Pozo et al., 2018). The study of sulfate conjugated metabolites has been of particular interest in metabolomics and found useful applications in applied sciences such as anti-doping and medical research (Pranata et al., 2019; Mueller et al., 2021). A range of tools are available, including the use of nuclear magnetic resonance (NMR), and various hyphenated chromatographic-tandem mass spectrometry (MS/MS) techniques (Lafaye et al., 2004). Early approaches to examine the sulfate metabolome employed gas chromatography-mass spectrometry (GC-MS) techniques, however, these suffered from various limitations due to the polar character of the conjugates, challenges with reliable enzymatic or chemical cleavage of sulfate esters, and loss of information about conjugation sites and levels (Gomes et al., 2009). This directed research towards liquid chromatography-mass spectrometry (LC-MS) techniques, which permit the direct detection of the intact sulfate conjugates, allowing the sulfate fraction to be specifically targeted during analysis. This type of analysis heavily relies on MS/MS techniques driven by the capacity of sulfate conjugates to readily undergo ionisation during electrospray ionisation (ESI), and display distinctive fragmentation behavior (Thevis et al., 2011; Gómez et al., 2012; Balcells et al., 2017a). Approaches based on tandem mass spectrometry include, multiple reaction monitoring (MRM), precursor ion scanning, or constant ion loss (CIL) scanning, which have all been used to selectively and directly detect sulfate conjugated metabolites (Bowers and Sanaullah, 1996; Bean and Henion, 1997; Hintikka et al., 2008; Gómez et al., 2013; Balcells et al., 2017a; McLeod et al., 2017; Schänzer and Thevis, 2017).

An early example of LC-MS based profiling of sulfate conjugates was in human and rat urine. Here sulfated molecules were identified by monitoring the neutral loss of 80 Da (SO_3), precursors of *m/z* 80 ($^{\bullet}SO_3^-$) and of *m/z* 97 (HSO_4^-), using ESI triple-quadrupole MS/MS (Lafaye et al., 2004). These methods were used to identify tens of sulfate metabolites, some of which were then confirmed through synthesis of the corresponding reference materials (Lafaye et al., 2004). A similar study used ultra-high-performance liquid chromatography (UHPLC)/negative ion matrix-assisted laser desorption ionization tandem time-of-flight high resolution mass spectrometry (MALDI-TOF/TOF-MS) to identify 1,129 potential sulfate candidates from the urine of pregnant women with detection based on the neutral loss of 80 Da and/or the formation of the *m/z* 97 product ion. Database matches were investigated for these candidates but none were confirmed against reference materials (Yao et al., 2016). More recently a method for the detection of sulfated metabolites in gut-microbiota using UHPLC-MS in an untargeted fashion was reported (Ballet et al., 2018). In this study sulfated metabolites were identified by comparing

features that underwent change after treatment with a purified preparation of the commercially available sulfatase enzyme *Helix pomatia* arylsulfatase (HpS). Structural validation was then achieved through MS/MS fragment identification, by comparison with reference materials or through data base matching. The method was able to confirm the structures of 36 from 206 putative sulfated metabolites identified (Ballet et al., 2018). Although collectively these approaches provide useful tools to study sulfate metabolites, they also have some limitations. In general, these approaches monitor only a selection of the known sulfate ester fragmentation modes and as a result may miss important classes of conjugate or not report information that could aid in characterizing different classes of conjugate. Furthermore, methods such as enzyme hydrolysis, that assess conversions of sulfate esters to their non-sulfated counterparts, only report on species that change and do not necessarily report on those that do not change. Moreover, they may also miss those product species with low ionisation efficiencies. Herein, we report the development of an untargeted metabolomics approach to comprehensively profile the sulfate metabolome in a urinary matrix. It employs ultra-high performance liquid chromatography-tandem high resolution mass spectrometry (UHPLC-HRMS/MS) with data dependent acquisition (DDA) and a novel data analysis pipeline to systematically identify sulfated metabolites by studying fragmentation behavior. The suitability of the method was evaluated in two applications: firstly, as a screening tool to identify potential steroid markers in equine urine post doping with testosterone propionate; and secondly, to monitor the performance of sulfatase enzymes for the hydrolysis of sulfate esters in pooled human urine. This method increases the utility of untargeted metabolomics for sulfate biomarker discovery in a urinary matrix.

MATERIALS AND METHODS

Animal Administration

Animal administration was approved by both Charles Sturt University (Wagga Wagga, NSW, Australia) and Racing NSW Animal Care and Ethics Committees. Testoprop® (testosterone propionate, 250 mg) was administered by intramuscular injection in the neck, opposite to the sampling jugular catheter to a thoroughbred gelding (635 kg, 14-year-old). Urine samples were collected at −72, −48, −24, 0, 2, 4, 6, 8, 12, 24, 36, 48, 72 h and then daily to 28 days after administration. Samples were stored at −20°C at the Racing NSW or the Australian National University (ANU) in sterile Falcon tubes (polypropylene, centrifuge tubes, 50 ml).

Equine Urine Sample Preparation

The method has been reported previously and adapted according to our work (Waller et al., 2016). To an aliquot of urine (1.1 ml), phosphate buffer (0.55 ml, 100 mM, pH 7.4) was added and the samples were centrifuged (1,100 x *g*, 5 min) to pellet solids. Each supernatant was fortified with a mixture of analytical internal standards nandrolone sulfate (**S1**), cholanediol bis(sulfate) (**S2**), epiandrosterone ($^{18}O_3$)-sulfate (**S3**) and 5α-androstane-3β,17β-diol 3,($^{18}O_3$)17-bis(sulfate) (**S4**) (0.150 ml, final concentration equivalent to 300 ng/ml original urine volume (1.1 ml) per standard). At this stage, the supernatants were split equally into biological samples and pooled quality control (QC) samples (0.818 ml, equivalent to 0.5 ml of original urine volume). The latter were pooled and then redistributed as pooled QC aliquots (0.818 ml). Both supernatant and pooled QC samples (0.818 ml) were then loaded onto a Waters Oasis™ WAX SPE cartridge (3 cc), that was pre-conditioned with methanol (2 ml) and water (2 ml). Samples were washed with NaOH (2 ml, 0.1 M), phosphate buffer (2 ml, 100 mM, pH 7.4), and Milli-Q water (2 ml) before elution with a mixture of ethyl acetate: methanol: diethyl amine (25:25:1 v/v/v, 3 ml) into clean 10 ml glass tubes. Samples were then evaporated to dryness under a reduced pressure at 40°C and stored at -20°C. The dried samples were re-dissolved with acetonitrile: water (20% v/v, 50 µL), filtered using 0.2 µm spin filters and stored at 5°C for analysis.

Enzyme Hydrolysis in Human Urine Sample Preparation

The human urine samples used in this study were collected with approval by the ANU Human Research Ethics Committee, 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; they were all healthy and reported not using steroids within 1 month of supplying a sample. Urine was taken from six subjects (three females and three males ranging between 20–50 years old) and pooled, into phthalate free plastic containers (Nalgene ® bottles, style 2110) and stored at −20°C. The following procedure was based on an established method (Stevenson et al., 2015). Aliquots of pooled human urine (2.1 ml) were pipetted into 15 ml falcon tubes. The samples were either adjusted to a pH of 7.5 ± 0.2 (PaS enzymes and control) or to 4.0 ± 0.2 (HpS enzyme) with the addition of Tris buffer (1.5 ml, 0.2 M) or acetate buffer (1.5 ml, 0.2 M) with mixing, respectively. To their respective tubes, preparations of purified WT-PaS, PVFV-PaS, LEF-PaS, or crude HpS (0.4 ml, PaS preparations 2.0 mg/ml, HpS 5.9 mg/ml) and for the control sample enzyme storage buffer (0.4 ml, 0.1 mM Tris-HCl, pH 7.5 in 50% v/v glycerol) were added. The quantities of enzyme were normalized according to their rates of *para*-nitrophenyl sulfate hydrolysis (**Supplementary Table S7**) (Stevenson et al., 2015). Hydrolysis reactions were performed in triplicate for a total of 15 samples and controls by overnight incubation at 37°C. Samples were centrifuged (1,100 x *g*, 5 min) and the supernatants were split equally (1.9048 ml each, equivalent to 1 ml of original urine volume) into biological and pooled QC samples. All pooled QC samples were pooled and redistributed into 1.9048 ml pooled QC aliquots. Each supernatant and pooled QC were fortified with a mixture of analytical internal standards nandrolone sulfate (**S1**), cholanediol bis(sulfate) (**S2**), epiandrosterone ($^{18}O_3$)-sulfate (**S3**) and 5α-androstane-3β,17β-diol 3,($^{18}O_3$)17-bis(sulfate) (**S4**) (0.300 ml, final concentration equivalent to 300 ng/ml urine volume (1 ml) per standard). The treatments, control and pooled QC samples were then subjected to SPE as above, with tris-HCl buffer (2 ml, 0.2 M) being used in place of phosphate

TABLE 1 | Masses used to identify sulfate-derived fragments and neutral lossess (Attygalle et al., 2001; Yi et al., 2006; Farrell et al., 2011; Balcells et al., 2017b; McLeod et al., 2017; Esquivel et al., 2018).

Ion (*m/z*)	Nominal	Accurate
$^{\bullet}SO_3^-$	80	79.9573
HSO_3^-	81	80.9652
$^{\bullet}SO_4^-$	96	95.9523
HSO_4^-	97	96.9601
Neutral loss (Da)		
SO_3	80	79.9568
H_2SO_4	98	97.9674

buffer in the wash phase. Samples were then evaporated to dryness under reduced pressure at 40°C, and the dried samples were stored at −20°C until analysis. For analysis samples were reconstituted in MeOH: water (20% v/v, 100 µL).

UHPLC-HRMS/MS Analysis

The study was carried out using a Q-Exactive Plus Orbitrap mass spectrometer equipped with a heated electrospray ionisation source (HESI-II) interfaced to an Ulti-Mate 3000 system for chromatographic separation (all from Thermo Fisher, Scoresby, Australia). For equine urine samples: the column used was an Acquity UPLC CSH phenyl-hexyl column (2.1 × 100 mm, i.d., 1.7 µm) fixed to an Acquity UPLC CSH phenyl-hexyl VanGuard Pre-Column (2.1 × 5 mm i.d., 1.7 µm). The UHPLC separation was performed at flow rate 0.4 ml/min, using gradient mixing of two mobile phase components. Solution A: 20 mM ammonium formate in water; and solution B: 20 mM ammonium formate in acetonitrile: water (90% v/v). The gradient was 0–0.5 min (20% B), 0.5–15 min (20–58% B), 15–20.5 min (58–100% B), 20.5–21.5 min (100–20% B), 21.5–30.0 min (20% B). The injection volume was 5 µL and the column oven temperature was 40°C. For human urine samples: the column used was a polar end capped Thermo Accucore aQ C18 column (2.1 × 100 mm, 2.6 µm). The UHPLC separation was performed at a flow rate 0.4 ml/min, using gradient mixing of two mobile phase components; solution A: 5 mM ammonium formate in water; and solution B: 5 mM ammonium formate in methanol: water (99% v/v). The gradient was 0–20 min (1–100% B), 20–25 min (100% B), 25–26 min (100–1% B) and 26–35 min (1% B). The injection volume was 5 µL and the column oven temperature was 40°C. For HRMS analysis: the spray voltage was 2.50 kV, capillary temperature 250°C, S-lens RF level 50, and auxiliary heater temperature 350°C. Mass calibration was performed in negative mode: using propionic acid (*m/z* 73.0295), isobutyric acid (*m/z* 87.0452), heptanoic acid (*m/z* 129.0921), in addition to Pierce ESI negative ion calibration solution. Scan spectrum acquisition with a resolution of 70,000 (Full Width at Half Maximum; FWHM) and scan range of *m/z* 200 to 2000 was used in negative mode. The automatic gain control (AGC) was set to 3×10^6. Data dependent acquisition (DDA) MS/MS spectra (*m/z* 50–500) were collected, between 1 and 20 min, with a resolution of 17,500 (FWHM) on the top 10 precursors in each scan window. The intensity threshold (minimum intensity to initiate a DDA scan) was 8.0×10^4. The apex trigger was set to 2–6 s, and dynamic exclusion was used to exclude already selected ions for the following 3 s. A static exclusion list was also used to exclude precursor ions from DDA that did not emanate from the urine samples but were instead derived from solvents and other method components that were observed in extraction blank samples. This list consisted of the 1,000 most abundant ions detected from two incubated water extraction blank samples, which were taken through the whole sample treatment. For experiments with pooled QC samples, the injection sequence started by running two blanks followed by running at least 10 system suitability samples and two pooled QC samples, followed by the main sequence. For both applications: biological samples were performed in triplicate and analysed in a random order, with pooled QC samples dispersed evenly throughout. In targeted parallel reaction monitoring (PRM) MS/MS experiments for metabolite confirmation, a list of single *m/z* was selected and fragmented over multiple NCE's, for both reference material and biological samples.

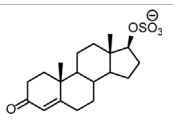
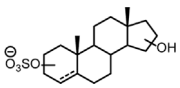
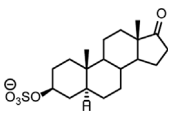
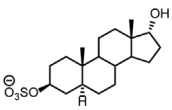
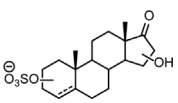
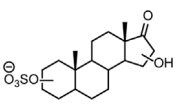
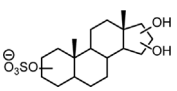
Data Workflow Pipeline and Processing

Data alignment, LOWESS normalization of total ion chromatograms (TIC), and Principal Component Analysis (PCA) analysis was performed in MS-DIAL by applying tolerances to both retention time (RT) (±3 s) and mass accuracy ($\Delta m/z$ ± 5 ppm), alongside the recommended settings for DDA-HRMS systems (Tsugawa et al., 2015). Further data analysis was performed either in python or R (**Supplementary Section S3**). A Python script was used to extract and append MS/MS data for each aligned feature that contained a sulfate-derived fragment. The *k*-means clustering function in R was used to sort metabolic features (Hartigan and Wong, 1979; Team, 2017). Two clusters were chosen as they represented the data appropriately (**Supplementary Figures S3, S4**) (Kaufman and Rousseeu, 2009). Clustering was performed based on the normalized data of eight unique parameters all derived from the MS/MS spectra. These included the six characteristic sulfate-derived fragments (**Table 1**). (Attygalle et al., 2001; Yi et al., 2006; Farrell et al., 2011; Balcells et al., 2017b; McLeod et al., 2017; Esquivel et al., 2018) The remaining two parameters were Intensity Ratio (IR): the total sum of sulfate-derived fragments divided by the sum of all fragments, and Maximum Abundance (MA): the normalized relative abundance (%) of the largest sulfate-derived fragment. Data were clustered into non-sulfate and sulfate groupings. High throughput differential metabolite level analysis and data visualization was performed in R, using the *limma* and *ggplot2* packages, respectively (Ginestet, 2011; Ritchie et al., 2015; Team, 2017). The output gives a final curated list of metabolic features that were sorted into sulfate and non-sulfate metabolites, with associated statistics [e.g., adjusted *p*-value, $\log_2$ Fold Change (FC)], UHPLC-HRMS/MS data (RT and *m/z*), and product ion data for any sulfate-derived fragments detected by MS/MS (see **Supplementary Table S14** for an example).

Synthesis of Anabolic Androgenic Steroid Sulfate Metabolite Reference Materials

Briefly, the synthesis of sulfated reference materials was performed as follows and was adapted from previous work

TABLE 2 | Structures of steroid sulfate metabolites and associated high throughput differential level analysis data from the untargeted profiling (log_2 fold change and adjusted *p*-value). Identified structures were confirmed against synthesized reference materials according to AORC retention time and MS/MS criteria, see **Supplementary Table S6**.

m/z	RT (min)	Log_2FC	Adjusted *p*-value	Structure	Identity
367.1583	8.77	10	2E-07		testosterone sulfate (**1**)
369.1739	7.69	9	6E-06		unidentified
369.1739	10.16	6	5E-06		epiandrosterone sulfate (**2**)
371.1895	10.53	8	2E-05		5α-androstane-3β-17α-diol-3-sulfate (**3**)
383.1537	6.37	7	1E-03		unidentified
385.1690	6.74	4	7E-08		unidentified
387.1845	7.59	10	2E-05		unidentified
387.1846	7.39	8	8E-03		unidentified
387.1846	4.56	9	5E-05		unidentified
387.1849	4.48	9	2E-05		unidentified

(Waller and McLeod, 2014): SO_3•py (30 mg, 188 mmol) was added to a solution of free steroid (5 mg) in dimethylformamide (0.5 ml) the resulting reaction mixture was capped and stirred at room temperature for 3 h. The reaction was then quenched with water (10 ml) and loaded onto a pre-conditioned Waters Oasis C18 SPE cartridge (3 cc). The reaction mixture was washed with aqueous ammonia solution (2 ml, 5% v/v) followed by water (2 ml). Steroid sulfates were then eluted in aqueous ammonia methanol solution (3 ml, 5% v/v). The methanolic ammonia fraction was then concentrated in vacuo to yield the desired steroid sulfate as an ammonium salt.

RESULTS AND DISCUSSION

Application 1: The effects of Testosterone Propionate Doping on the Urinary Sulfate Metabolome in the Equine

Anabolic androgenic steroids (AAS) and their metabolites are routinely detected in anti-doping urinary analysis (Schänzer and Thevis, 2017). Intact phase II AAS metabolites such as sulfate esters play an important role in anti-doping analysis as long-term biomarkers for both exogenous and endogenous steroids (Schänzer et al., 2013; Piper et al., 2016; Kiousi et al., 2021). Further, sulfation of metabolites is reported to predominate in horses when compared to humans (Scarth et al., 2011). In this application we use the untargeted profiling workflow to assess change in the sulfate urinary metabolome after the intramuscular administration of the AAS testosterone propionate in a thoroughbred gelding (horse 1). This study included a comparison of pre (−24 h) and post (+12 h) administration urine samples, performed in triplicate. These samples were also compared against samples from a non-drug treated gelding (horse 2) at three time points, to assess natural variability over time. Samples were extracted using weak anion exchange (WAX) solid phase extraction (SPE) to isolate the sulfated fraction from the urine.

Data Acquisition, Batch Analysis and Alignment

Metabolomic data was acquired using UHPLC-HRMS/MS with data dependent acquisition. The normalized collision energy (NCE) for MS/MS analysis was optimized by studying fragmentation patterns of 20 steroid monosulfate and bis(sulfate) reference materials (**Supplementary Table S1**). An NCE of 60 eV was chosen for the DDA experiments as all sulfate-derived fragments could be observed with high relative abundance (**Table 1**). Following data acquisition, detected features were aligned using suitable tolerances (RT ± 3 s and

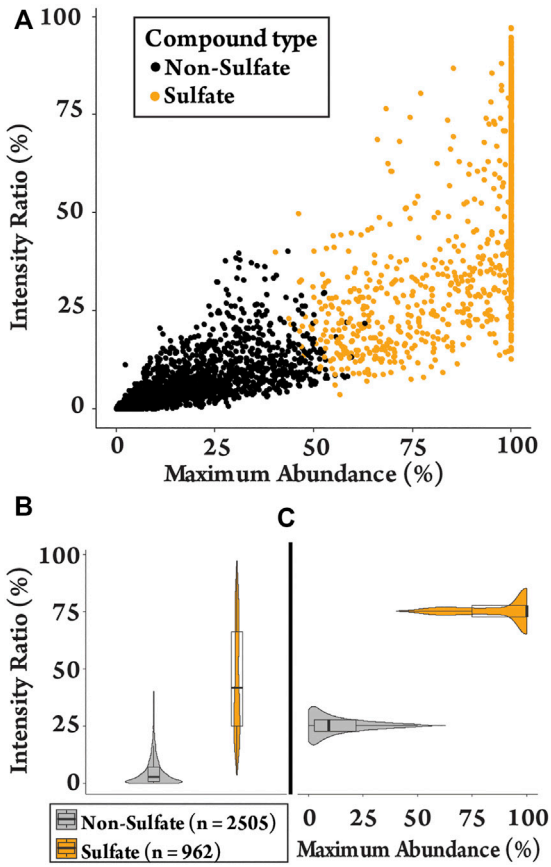


FIGURE 1 | Plots of the detected features (n = 3467) in equine urine after administration of testosterone propionate. **(A)** Clustering plot contrasting MA against IR of sulfated and non-sulfated features. **(B)** The spread of IR values for sulfated and non-sulfated features. **(C)** The spread of MA values for sulfated and non-sulfated features.

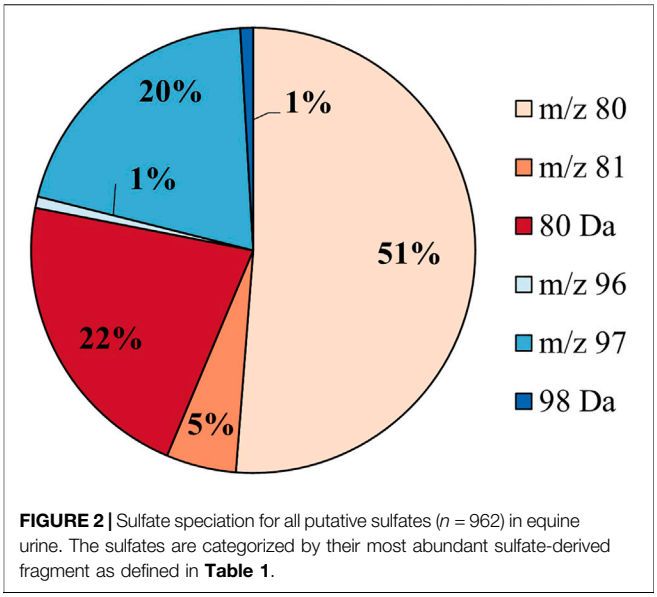


FIGURE 2 | Sulfate speciation for all putative sulfates (n = 962) in equine urine. The sulfates are categorized by their most abundant sulfate-derived fragment as defined in **Table 1**.

$\Delta m/z$ ± 5 ppm) in MS-DIAL. Normalization was performed by the use of locally weighted scatterplot smoothing (LOWESS) program (Sangster et al., 2006; Godzien et al., 2015; Broadhurst et al., 2018; Considine et al., 2018; Dudzik et al., 2018). This gave an initial list of 6,230 total detected features. Following this, features without MS/MS spectra, an S/N ratio below three and peaks that appeared in less than two samples, were excluded giving a final curated list of 3467 features. Data was assessed through PCA with good quality data indicated by the tight clustering of replicate samples and pooled QC samples (**Supplementary Figure S1**) (Broadhurst et al., 2018).

To allow identification of sulfated metabolites, MS/MS spectra of LC/MS features from pooled QC samples were exported to Mascot Generic Format files (.MGF) and analyzed with a custom Python script (**Supplementary Section S7**) designed to search for accurate mass evidence of known sulfate-associated ions and neutral losses from the precursor ion. During this process two new parameters termed intensity ratio (IR) and maximum abundance (MA) were calculated for each detected feature. These parameters were simple ratios generated from the sulfate-derived fragment ions observed in MS/MS spectra. The IR was the ratio of sulfate-derived fragment ions as a percentage of all product ions in the MS/MS spectrum. The MA was the relative abundance of the sulfate reporter ion with the highest relative abundance. As sulfate metabolites typically display MS/MS spectra dominated by characteristic product ions or neutral losses these ratios were used to help identify sulfate metabolites. These data on sulfate-derived fragments were mapped onto the alignment results from MS-DIAL on the basis of retention time and accurate precursor m/z (closest retention time match within tight m/z and RT error tolerance limits) using a second custom Python script. (**Supplementary Section S7**).

Identification of Sulfated Metabolites

k-means clustering was used to differentiate between sulfated and non-sulfated features in the samples in an unbiased manner. This aims to partition data into groups based on the minimization of the sum of squares to their assigned cluster center. Grouping was based on the six sulfate-derived fragments (**Table 1**) together with the ratios MA and IR, and were grouped to the nearest center (mean) across the eight parameters, clustered into two groups (**Supplementary Figure S3**) (Hartigan and Wong, 1979). This separated the 3467 features into 2,505 non-sulfates and 962 sulfates (28%). The clustering of features is visualised with a plot of MA vs. IR (**Figure 1A**), with a full summary of clustering provided in the supplementary information (**Supplementary Table S2**). As expected, sulfated features tended to have a larger MA value (i.e., a sulfate-derived fragment is a major peak) and a larger IR value (i.e., sulfate-derived fragments make up a relatively large proportion of total product ions). The distributions of these ratios are visualised as violin plots (**Figures 1B,C**). They show that the IR of sulfated molecules were spread with a median IR of 42% (**Figure 1B**, mean = 46%, range = 4–97%) contrasted with the non-sulfated features with a median IR of 3% (mean = 5%, range = 0–40%). The median MA of 99%

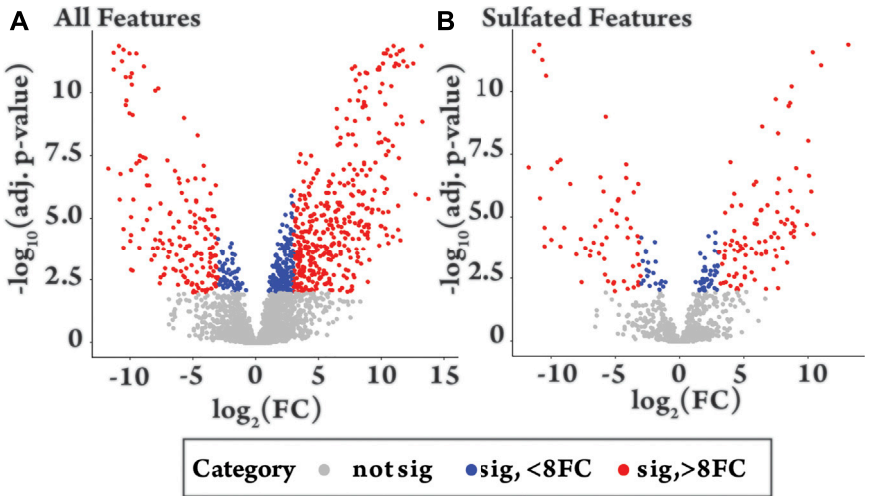


FIGURE 3 | Volcano plots of $-\log_{10}$ of the adjusted (adj.) *p*-value against $\log_2$ fold change (FC) comparing equine urine samples post-administration of testosterone propionate (+12 h) to pre-administration (−24 h). Colour groupings. Grey = not significant (adjusted *p*-value > 0.01), Blue = significant (*p*-value < 0.01) with fold change 0–8, Red = significant (adjusted *p*-value < 0.01) with a fold change >8. **(A)** "All features", shows the differential metabolite level analysis over all detected features in the plus 12 h sample [*n* = 3467, grey *n* = 2662 (77%), blue *n* = 235 (7%), red *n* = 570 (16%)]. **(B)** "Sulfated Features", shows differential metabolite level analysis over all sulfated features in the plus 12 h sample [*n* = 962, grey *n* = 747 (77%), blue *n* = 61 (6%), red *n* = 154 (16%)].

(**Figure 1C**, mean = 88%, range = 40–100%) also contrasted to the non-sulfated features with a median MA of 9% (mean = 5%, range = 0–63%). This clustering of sulfated and non-sulfated features was later used to prioritizatise sulfated metabolites for further investigation.

Speciation of Sulfates

The putative sulfated metabolites can be differentiated further based on the observed major sulfate-derived transition (**Table1**). In the equine urine samples (total sulfates, *n* = 962), the dominant sulfate-derived transition was the $^{\bullet}SO_3^-$ ion (*m/z* 80, *n* = 493, 51%), with the HSO_4^- ion (*m/z* 97, *n* = 194, 20%), and the neutral loss of SO_3 (80 Da, *n* = 209, 22%) also prominent. Other sulfate-derived fragments including the HSO_3^- (*m/z* 81, *n* = 49, 5%), $^{\bullet}SO_4^-$ (*m/z* 96, *n* = 8, 1%) ions, and H_2SO_4 neutral loss (98 Da, *n* = 9, 1%) were observed at a lower frequency (**Figure 2**). The sulfate-derived ions $^{\bullet}SO_3^-$, HSO_3^- and neutral loss SO_3 are typically associated with phenolic and other unsaturated sulfate metabolites, while the ion HSO_4^- is associated with saturated sulfates (McLeod et al., 2017). The majority of sulfate metabolites (78%) showed two or more sulfate derived fragments on applying a relative abundance threshold of 5% (data not shown).

High Throughput Differential Metabolite Level Analysis

To identify features that change significantly following drug administration, a high throughput differential analysis, commonly used to assess gene expression, was applied using the packages *Limma* and *Glimma* in R, from Bioconductor (**Supplementary Section S7**) (Ritchie et al., 2015; Su et al., 2017; Team, 2017). The results of the analysis comparing the pre- (−24 h) and post drug-administration samples (+12 h) are shown as volcano plots for all features (**Figure 3A**) and sulfated features (**Figure 3B**), with the results summarized in **Supplementary Table S3**. Similar patterns were observed for all features and sulfated features in terms of significant change (red and blue, **Figure 3**) and positive and negative fold change. Of the 962 sulfated features, 215 had significantly changed (adjusted *p*-value < 0.01) (red and blue, **Figure 3B**) representing 6% of the total detected features in equine urine after doping. Of these, a larger proportion were upregulated, specifically 136 were upregulated and 79 were downregulated, as indicated by the direction of their fold change. The analysis also showed relatively large inter-horse variation in detected features, which is likely due to individual differences in metabolism between the two horses (**Supplementary Figure S8**). Comparisons also revealed limited intra-horse variation over time (horse 2, **Supplementary Figure S9**), but a relatively large intra-horse variation pre- and post-drug administration (**Supplementary Figure S8**), a pattern of change that could be indicative of increased metabolic activity following administration of the AAS testosterone propionate.

Discovery of a Novel Steroid Sulfate Metabolite in Equine Urine

The 215 LC/MS features found to display significant changes in relative signal intensity following testosterone propionate administration could represent direct metabolites of testosterone propionate or other metabolites modulated indirectly in response to testosterone propionate administration. However, this study specifically sought to identify sulfated metabolites derived from the direct metabolism of testosterone propionate. A list of possible

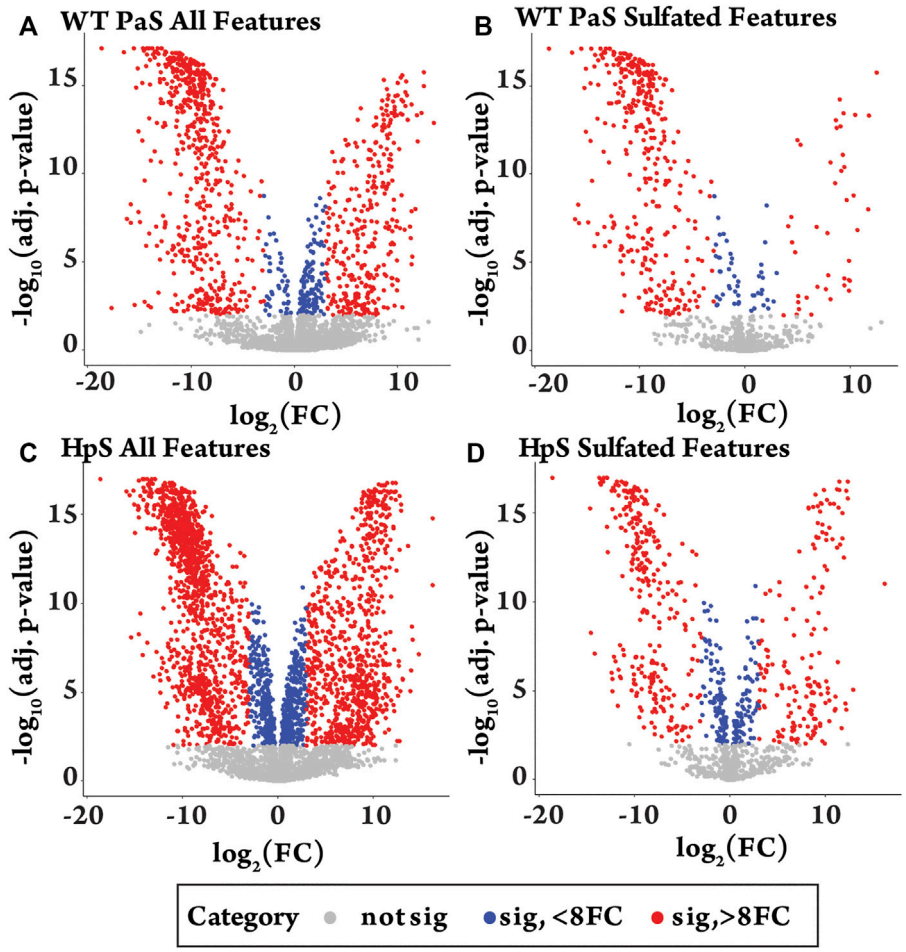


FIGURE 4 | Differential metabolite level analysis for enzyme hydrolysis of pooled human urine. **(A)** All detected features (*n* = 5774) in the WT-PaS treated urine relative to a control sample. **(B)** Sulfate features (*n* = 1430) in WT-PaS treated sample. **(C)** All detected features (*n* = 5774) in the HpS treated urine relative to a control sample. **(D)** Sulfate features (*n* = 1430) in HpS treated sample.

metabolites was generated by applying combinations of up to three phase I metabolic transformations, including hydroxylation, bond oxidation and reduction to testosterone, followed by sulfation (**Supplementary Table S4**). Steroid sulfates in the 215 molecules matching the accurate mass (*m/z* ± 5 ppm) against the generated list of predicted metabolites were identified, using R. Where possible these putative metabolites were confirmed against synthetically derived reference materials by comparison of UHPLC retention time and MS/MS behaviour (Aru et al., 2020).

From this search 30 possible steroid sulfates were identified with theoretical accurate mass matching proposed metabolic transformations (**Supplementary Table S4**). Of these molecules, 10 were found to be elevated in the range of 16–1024-fold after testosterone propionate administration (+24 h) and by comparison to the control (horse 2), **Supplementary Table S5**. The identities of testosterone sulfate (**1**) (*m/z* 367.1583), epiandrosterone sulfate (**2**) (*m/z* 369.1739), and 5α-androstane-3β,17α-diol 3-sulfate (**3**) (*m/z* 371.1895) were established after comparison to synthesised reference materials (**Table 2**, for synthesis of reference materials see **Supplementary Section S4**). Confirmation was performed according to criteria set out by the Association of Official Racing Chemists (AORC) for UHPLC retention times and MS/MS transitions (**Supplementary Table S6**) (Association of Offical Racing Chemist, 2020). Scan MS data showed the expected isotope signatures. Of the three confirmed structures, the first two are known biomarkers of testosterone metabolism in equine and human systems (Piper et al., 2017; Esquivel et al., 2019). The final metabolite 5α-androstane-3β,17α-diol 3-sulfate (**3**) has not been described in literature or in online data bases such as ChEBI (Hastings et al., 2016). The upregulated nature of these steroid sulfates also support the general idea of perturbation caused by doping with testosterone propionate (Pozo et al., 2010; Scarth et al., 2011; Vimercati et al., 2017). Future directions for this work could aim at identifying the remaining seven putative steroid sulfate metabolites against reference materials. Longitudinal or population studies would also be required to assess the

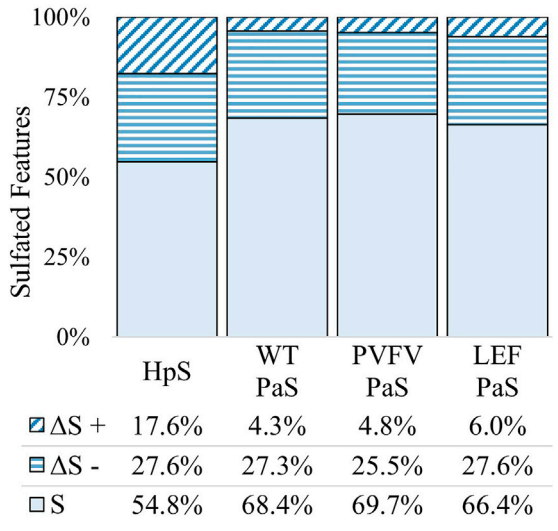


FIGURE 5 | Total change in sulfated features for each sulfatase enzyme treatment of pooled human urine. Unchanged sulfated features are represented by "S" where the adjusted *p*-value > 0.01. Changed sulfated features are represented by "ΔS", with an adjusted *p*-value < 0.01. The direction of fold change is indicated by "–" or "+" signs. Values reported as a proportion (%) of total sulfated features (n = 1430).

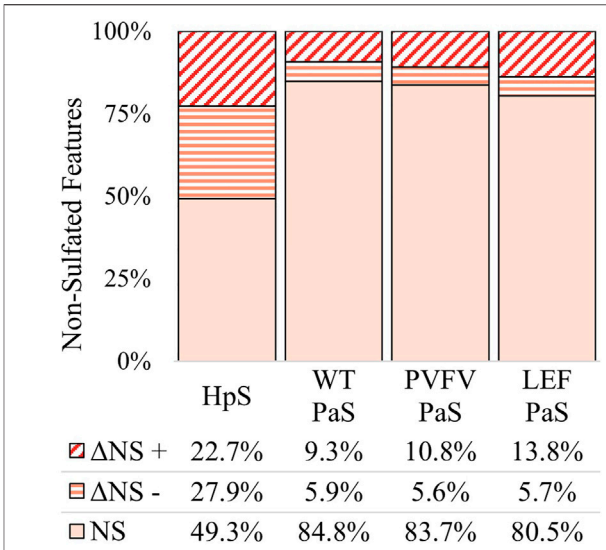


FIGURE 6 | Total change in non-sulfated features for each sulfatase enzyme treatment of pooled human urine. Unchanged non-sulfated features are represented by NS where the adjusted *p*-value > 0.01. Changed non-sulfated features are represented by ΔNS, where adjusted *p*-value < 0.01. The direction of fold change is indicated by "–" or "+" signs. Values reported as a proportion (%) of total non-sulfated features (n > 4344).

importance of these markers in doping with testosterone propionate.

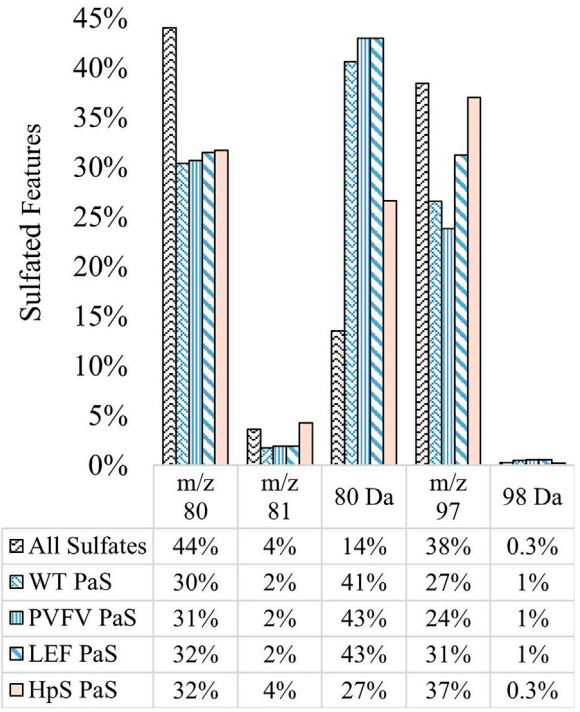
Application 2: Profiling in Sulfatase Treated Human Urine

Hydrolysis is frequently used prior to the analysis of conjugated steroids. In GC-MS analysis, sulfate conjugates are typically hydrolysed to the free steroid and derivatised in analytical workflows, to improve thermal stability and volatility. Hydrolysis is also commonly used in both GC-MS and LC-MS approaches to aid confirmation of metabolites against the more readily available unconjugated steroid reference materials. For sulfated conjugates, hydrolysis is often performed using the commercially available *Helix pomatia* aryl sulfatase (HpS). However, without extensive purification (Ballet et al., 2018), these crude enzyme preparations are known to contain additional enzyme activities such as glucuronidase, oxidase, and reductase, making it unsuitable for many applications (Gomes et al., 2009). Alternatively, chemical solvolysis can be used as a means of deconjugation, however, this can lead to analyte degradation and increased matrix interference (Gomes et al., 2009). Recently, a new recombinantly expressed and purified arylsulfatase from *Pseudomonas aeruginosa* has been investigated for the selective hydrolysis of sulfatase esters (Stevenson et al., 2015). Directed evolution was employed to improve the catalytic efficiency of testosterone sulfate hydrolysis, with improvements in substrate scope and thermostability relative to the wild-type (WT-PaS) enzyme also observed (Uduwela et al., 2018). This application sought to compare the performance of a commercially available HpS preparation to three PaS preparations for hydrolysis of urinary sulfates. Two improved mutants PVFV-PaS and LEF-PaS along with WT-PaS were selected for this evaluation, representing different points along the evolutionary pathway (Uduwela et al., 2018).

Profiling the Hydrolytic Activity of Sulfatases in Pooled Human Urine

The workflow described above was used to assess the change in the sulfate metabolome in pooled human urine after incubation with each enzyme preparation. For the hydrolysis study, aliquots of pooled human urine from six healthy people (three females and three males ranging between 20–50 years old) were treated with each enzyme preparation or a control in triplicate and incubated overnight. Enzyme activity was normalized for the hydrolysis of the common *p*-nitrophenyl sulfate prior to the experiment using recommended pH ranges (**Supplementary Table S7**) (Stevenson et al., 2015). Following this, samples were extracted and then subjected to the described workflow.

Implementing the analytical workflow resulted in 5,774 features with associated MS/MS data from an initial 15,608 detected features in the pooled human urine. The PCA analysis showed tight grouping of all PaS enzyme mutants, distinct from the control, HpS and pooled QC samples (**Supplementary Figure S2**) (Broadhurst et al., 2018). *k-means* clustering resulted in a total of 1,430 putative sulfates (25%) being identified from the 5,774 features (**Supplementary Figure S5**). In this, sulfate-derived fragment speciation was dominated by the ions *m/z* 80 (n = 630, 44%) and *m/z* 97 (n = 550, 38%), with other minor species also observed including neutral loss 80 Da (n = 194, 14%), *m/z* 81 (n = 52, 4%), neutral loss 98 Da (n = 4, 0.3%) and *m/z* 96 (n = 8, 1%)



	m/z 80	m/z 81	80 Da	m/z 97	98 Da
All Sulfates	44%	4%	14%	38%	0.3%
WT PaS	30%	2%	41%	27%	1%
PVFV PaS	31%	2%	43%	24%	1%
LEF PaS	32%	2%	43%	31%	1%
HpS PaS	32%	4%	27%	37%	0.3%

FIGURE 7 | Speciation (%) of hydrolysed sulfates for each sulfatase enzyme treatment of pooled human urine. Hydrolysed sulfates (significant negative fold-change) for each enzyme treatment (WT, n = 391, PVFV, n = 365, LEF, n = 365, HpS, n = 394) are grouped according to their most abundant sulfate derived transition. This is contrasted with the speciation of all sulfates detected in pooled human urine (n = 1430). Note: neutral loss of 96 Da was not detected in pooled human urine.

FIGURE 8 | Heat map displaying the $\log_2$(FC) of the concentration (ng/mL) of 33 free steroids detected in GC-MS when treated with either WT PaS, PVFV PaS, LEF PaS or *E. coli* β-glucuronidase. The top 6 listed free steroids make up part of the steroid module in the athletes biological passport, all listed steroids are endogenously found in human urine. Adjusted p-values of $p < 0.05$ is denoted by "*", and of $p < 0.01$ is denoted by "**". Differential metabolite level analysis and the heat map was generated using R.

(**Supplementary Figure S6**). In contrast to the equine urine samples this showed (**Supplementary Figure S10**) a higher proportion of saturated sulfates (m/z 97) than unsaturated sulfates (m/z 80, m/z 81 and 80 Da). The majority of sulfate metabolites (64%) showed two or more sulfate derived fragments on applying a relative abundance threshold of 5% (data not shown).

The differential metabolite level analysis revealed that of the four enzyme treatments, the crude HpS preparation had distinct activity from the three PaS treatments (**Figures 4–6**). HpS possessed the largest number of significantly changed molecules in both the non-sulfated and sulfated fraction of the urinary metabolome. In the sulfated fraction (**Figure 5**), HpS had a total of 45% (646/1,430) of total sulfates significantly changed, which contrasts with the three PaS treatments displaying changes ranging from 30 to 34% (WT, 452/1430, PVFV, 433/1430, LEF, 481/1430). In the non-sulfated fraction (**Figure 6**), HpS resulted in 50.6% (2,202/4344) of molecules undergoing significant change compared to the PaS treatments, 15–20% (WT, 662/4344, PVFV, 710/4344, LEF, 848/4344).

In terms of enzyme hydrolysis, both HpS and the PaS enzymes showed a similar proportion (25.5–27.6%, **Figure 5**) of the sulfated features with a significant negative fold change, consistent with sulfatase enzyme hydrolysis. However, in the case of HpS treatment, this was accompanied by a greater proportion of sulfated features with a positive fold change (**Figure 5**) as well as a greater proportion of non-sulfated features undergoing both positive and negative fold change (**Figure 6**), which was inconsistent with simple sulfate ester hydrolysis and suggested significant levels of alternative enzyme activity. For the PaS enzyme treatments 82–87% (WT, 391/452, PVFV, 365/433, LEF, 365/481) of significantly changed sulfated features underwent a negative fold change consistent with enzyme hydrolysis compared to only 63% (394/646) for HpS treated samples (**Figure 6** and **Supplementary Table S3**). Taken together, these observations show similar levels of sulfate hydrolysis for the four enzyme treatments but also clearly show higher selectivity for sulfate hydrolysis by the recombinantly expressed and purified PaS enzymes relative to the HpS crude enzyme preparation.

The substrate scope for each sulfatase treatment was assessed by comparing the sulfate-derived fragment speciation for hydrolysed sulfates. **Figure 7** shows the fragment speciation for all hydrolysed sulfates in comparison to the speciation of all detected sulfates (n = 1,430) in pooled human urine. Both enzyme classes displayed a preference for hydrolysis of sulfates, 27–43%, characterised by the neutral loss of SO_3 (80 Da), when compared to the total proportion of sulfate species 14%. This may indicate a preference for the hydrolysis of electron deficient unsaturated sulfate esters, such as phenolic sulfates, through a process of bond homolysis (to give $^{\bullet}SO_3^-$) followed by electron transfer to give SO_3 and the corresponding oxy-anion fragment. There were also distinct differences between the two classes of enzyme treatments. Specifically, the PaS treatments hydrolysed a smaller proportion of saturated sulfates compared to the HpS treatment, as indicated by the proportion of the m/z 97 ion. Within the PaS enzyme classes, there was also a small observed increase in the hydrolysis of saturated sulfates (m/z 97) from the WT-PaS to the LEF-PaS mutant. This trend aligned with the aims of the previous directed evolution study, which sought to improve the catalytic efficiency of the PaS enzyme for the hydrolysis of the saturated alkyl sulfate ester testosterone sulfate (Uduwela et al., 2018).

Evidence for Glucuronidase Activity in Crude HpS Extract

The results clearly show the HpS enzyme had lower selectivity for the hydrolysis of sulfated metabolites, as a large proportion, 55% (1,214/2,202), of non-sulfated metabolites underwent significant change, **Figure 6**. A large part of this non-specific activity is attributable to the crude nature of the HpS extract that contains a range of enzyme activities including glucuronidase, oxidase and reductase activities (Stevenson et al., 2015; Ballet et al., 2018; Uduwela et al., 2018).

To investigate the possible glucuronidase activity of the crude HpS extract, a semi-targeted search was adapted in a retrospective fashion on the data acquired for application 2. The search was performed against a list of putative steroid glucuronides either derived from known steroids or from metabolic transformations of testosterone glucuronide, in a similar approach to that adopted in application 1 (**Supplementary Table S8**). Each of the 1,214 significantly changed non-sulfated features was matched in MS and MS/MS using accurate mass (±5 ppm) and assigned as glucuronides by matching to characteristic MS/MS transitions (Fabregat et al., 2013). These transitions included neutral loss of 194 Da (loss of glucuronic acid) and 176 Da [loss of glucuronic acid–H_2O (gluc)], and the fragment ions m/z 175 [(gluc-H)$^-$], m/z 157 [(gluc-H-H_2O)$^-$], m/z 113 [(gluc-H-H_2O-CO_2)$^-$], m/z 85 [(Gluc-H-H_2O-CO_2-CO)$^-$] and m/z 75 [($HOCH_2CO_2$)$^-$]. Unlike, Fabregat et al. (2013), matching was done at high resolution, allowing for accurate masses to be used in both MS and MS/MS dimensions (**Supplementary Table S9**).

From this search 96/1,214 molecules were identified as putative steroid glucuronides and had at least two characteristic MS/MS transitions. Of these, 90/96 underwent significant hydrolysis in the HpS treated sample, while none of these 96 molecules underwent hydrolysis in the PaS treatments (**Supplementary Tables S8, S10**). The hydrolysis of glucuronide metabolites by HpS enzyme was not unexpected, with the product information sheet indicating at least 30 units of β-glucuronidase for every unit of sulfatase activity. Due to this, it has routinely been used in drug metabolism studies in both medical and anti-doping fields due to this broad substrate scope (Houghton and Maynard, 2010; Garg et al., 2018). A four-step purification of the crude HpS extract to generate higher selectivity for sulfate ester hydrolysis has recently been described, and the resulting preparation used to screen for unsaturated sulfate esters in human urine (Ballet et al., 2018). However, this purified HpS preparation is not commercially available. The PaS variants evaluated in this study show high levels of selectivity as expected for recombinantly expressed and purified enzymes and provide a convenient alternative for the study of the sulfated metabolome.

Substrate Selectivity of PaS

To demonstrate the selectivity of the PaS enzymes a targeted GC-MS analysis of free steroids was performed in treated urine (**Figure 8**, **Supplementary Table S11**). In this experiment we measured the concentrations of 33 free steroids in pooled human urine after treatment with either the PaS enzymes or with *Escherichia coli* (*E.coli*) β-glucuronidase (**Supplementary Section S3**). This β-glucuronidase was specifically chosen as that mandated for glucuronide deconjugation by the steroid module of the World Anti-Doping Agency athlete biological passport (World Anti-Doping Agency WADA, 2017). The concentration of 14 of these free steroids significantly increased after treatment with the PaS enzymes indicating some level of sulfate conjugation. This included testosterone sulfate that showed expected improvements in hydrolysis in moving from WT-PaS to PVFV-PaS and LEF-PaS mutants. Also, for these 14, the concentration of eight steroids, clustered in blue (**Figure 8**), showed a greater increase in concentration following PaS hydrolysis than observed for *E. coli* β-glucuronidase suggesting higher levels of sulfate conjugation. High levels of sulfate conjugation have also recently been reported for a range of circulating vitamin D metabolites (Jenkinson et al., 2021). These results were also compared against those from the metabolic profiling. Using an accurate mass search (m/z ± 5ppm) potential matches were found to each of these eight steroid sulfates in the UHPLC-HRMS/MS data set (**Supplementary Table S12**). Although the majority of these remain unconfirmed, epiandrosterone sulfate, was matched against an isotope labelled internal standard [epiandrosterone ($^{18}O_3$)-sulfate (**S3**)] used in the initial metabolic profiling. Epiandrosterone sulfate was found to undergo full hydrolysis for each enzyme treatment (**Supplementary Table S13**).

In this application we have demonstrated the selectivity of the PaS enzymes towards sulfated metabolites when compared to the commercially available crude HpS extract. As recombinantly expressed and purified enzymes, PaS is recommended as a preferred enzyme for studies of the sulfated metabolome. Overall, the results of this study demonstrate the usefulness of

untargeted metabolic profiling methods to monitor minute differences in the sulfate metabolome in urine. Its strength lies in the deconvolution of sulfated from non-sulfated metabolites by monitoring sulfate-derived fragment ions.

CONCLUSION

There are several clear pathways forward for this type of untargeted metabolic profiling. Due to its simplicity and relative ease of use it could be employed as a screen to look for new metabolites in applications such as in disease diagnosis or anti-doping. It is also conceivable that similar filters and scripts could be applied to related urinary metabolites such as glucuronides and phosphates, and di-anionic metabolites (McLeod et al., 2017). Limitations of this approach lie with the acquisition speeds of the MS instrumentation and the DDA method. The DDA method suffers from only sampling higher abundance ions per duty cycle, which can lead to missed detection or the misassignment of a molecule.

Overall, we have presented a novel workflow for the untargeted profiling of sulfated metabolites in urine matrices that combines UHPLC-HRMS/MS instrumentation and a new data processing pipeline. This provided a rapid tool for the qualitative assessment of the sulfate metabolome in equine and human urine. In equine urine 215 of 962 putative sulfate metabolites were found to significantly change after testosterone propionate administration in a single horse. Of these, 10 upregulated features were predicted to be steroid sulfate metabolites based on accurate mass searches. The identity of three steroid sulfates were confirmed as testosterone sulfate (**1**), epiandrosterone sulfate (**2**), and a new metabolite, 5α-androstane-3β-17α-diol-3-sulfate (**3**), according to AORC retention time and MS/MS criteria (Association of Offical Racing Chemist, 2020). The profiling method was also used to examine sulfatase activity in pooled human urine. The new workflow identified 1,430 putative sulfated metabolite features in pooled urine from a total of 5,774 features. Qualitatively, it was observed that the three PaS enzymes selectively hydrolysed sulfate esters and may be preferred in applications targeting the sulfate metabolome. Alternative β-glucuronidase activity associated with the HpS enzyme was also demonstrated. The use of our profiling-based approach could be of value in the identification and monitoring of endogenous and exogenous sulfated metabolites in urine.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ANU Human Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by the Charles Sturt University (Wagga Wagga, NSW, Australia) Animal Care and Ethics Committee and the Racing NSW Animal Care and Ethics Committee.

AUTHOR CONTRIBUTIONS

CF: Conceptualization, Methodology, Script, Validation, and Writing—Original Draft. RH: Conceptualization and Methodology. DU: Conceptualization and Methodology. BP: Methodology. AC: Conceptualization and Script. TN: Methodology and Script. AC: Conceptualization and Funding. LB: Conceptualization and Methodology. MM: Conceptualization, Methodology, Funding, and Writing—Supervision, review and editing CF

FUNDING

We thank Thermo Fischer Scientific for supporting the Charles Hocart Award to DU and the Australian Research Council Linkage Project scheme (LP180100421) for financial support.

ACKNOWLEDGMENTS

Bradley Stevenson for guidance of PaS synthesis and related assays. Alex Chen for initial methodological development of UHPLC and HRMS/MS systems.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.829511/full#supplementary-material

REFERENCES

Antignac, J. P., Brosseaud, A., Gaudin-Hirret, I., André, F., and Bizec, B. Le. (2005). Analytical Strategies for the Direct Mass Spectrometric Analysis of Steroid and Corticosteroid Phase II Metabolites. *Steroids* 70, 205–216. doi:10.1016/j.steroids.2004.11.009

Aru, V., Motawie, M. S., Khakimov, B., Sørensen, K. M., Møller, B. L., and Engelsen, S. B. (2020). First-principles Identification of C-Methyl-Scyllo-Inositol (Mytilitol) - A New Species-specific Metabolite Indicator of Geographic Origin for marine Bivalve Molluscs (Mytilus and Ruditapes spp.). *Food Chem.* 328, 126959. doi:10.1016/j.foodchem.2020.126959

Association of Offical Racing Chemist (2020). AORC MS Criteria. Association of Official Racing Chemists. Available at: http://www.aorc-online.org/documents/aorc-ms-criteria-modified-23-aug-16/.

Attygalle, A. B., García-Rubio, S., Ta, J., and Meinwald, J. (2001). Collisionally-induced Dissociation Mass Spectra of Organic Sulfate Anions. *J. Chem. Soc. Perkin Trans. 2* 24, 498–506. doi:10.1039/b009019k

Balcells, G., Gómez, C., Garrostas, L., Pozo, Ó. J., and Ventura, R. (2017). Sulfate Metabolites as Alternative Markers for the Detection of 4-chlorometandienone Misuse in Doping Control. *Drug Test. Anal.* 9 (7), 983–993. doi:10.1002/dta.2101

Balcells, G., Matabosch, X., and Ventura, R. (2017). Detection of Stanozolol O- and N- Sulfate Metabolites and Their Evaluation as Additional Markers in Doping Control. *Drug Test. Anal.* 9 (7), 1001–1010. doi:10.1002/dta.2107

Ballet, C., Correia, M. S. P., Conway, L. P., Locher, T. L., Lehmann, L. C., Garg, N., et al. (2018). New Enzymatic and Mass Spectrometric Methodology for the Selective Investigation of Gut Microbiota-Derived Metabolites. *Chem. Sci.* 9 (29), 6233–6239. doi:10.1039/c8sc01502c

Bean, K. a., and Henion, J. D. (1997). *J. Chromatogr. B* 690 (1–2), 65–75. doi:10.1016/s0378-4347(96)00403-3

Bowers, L. D., and Sanaullah, fnm. (1996). Direct Measurement of Steroid Sulfate and Glucuronide Conjugates with High-Performance Liquid Chromatography-Mass Spectrometry. *J. Chromatogr. B: Biomed. Sci. Appl.* 687 (1), 61–68. doi:10.1016/s0378-4347(96)00232-0

Broadhurst, D., Goodacre, R., Reinke, S. N., Kuligowski, J., Wilson, I. D., Lewis, M. R., et al. (2018). Guidelines and Considerations for the Use of System Suitability and Quality Control Samples in Mass Spectrometry Assays Applied in Untargeted Clinical Metabolomic Studies. *Metabolomics* 14 (6), 72. doi:10.1007/s11306-018-1367-3

Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., et al. (2016). ChEBI in 2016: Improved Services and an Expanding Collection of Metabolites. *Nucleic Acids Research* 44 (Database issue), D1214–D1219. doi:10.1093/nar/gkv1031

Considine, E. C., Thomas, G., Boulesteix, A. L., Khashan, A. S., and Kenny, L. C. (2018). *Metabolomics* 14 (1). doi:10.1007/s11306-017-1299-3

Dudzik, D., Barbas-Bernardos, C., García, A., and Barbas, C. (2018). Quality Assurance Procedures for Mass Spectrometry Untargeted Metabolomics. A Review. *J. Pharm. Biomed. Anal.* 147, 149–173. doi:10.1016/j.jpba.2017.07.044

Esquivel, A., Alechaga, É., Monfort, N., and Ventura, R. (2018). *Drug Test. Anal.* 10 (11–12), 1734–1743. doi:10.1002/dta.2413

Esquivel, A., Alechaga, É., Monfort, N., and Ventura, R. (2019). Sulfate Metabolites Improve Retrospectivity after Oral Testosterone Administration. *Drug Test. Anal.* 11 (3), 392–402. doi:10.1002/dta.2529

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 (10), 5005–5014. doi:10.1021/ac4001749

Farrell, T., Poquet, L., Dionisi, F., Barron, D., and Williamson, G. (2011). Characterization of Hydroxycinnamic Acid Glucuronide and Sulfate Conjugates by HPLC-DAD-MS2: Enhancing Chromatographic Quantification and Application in Caco-2 Cell Metabolism. *J. Pharm. Biomed. Anal.* 55 (5), 1245–1254. doi:10.1016/j.jpba.2011.03.023

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R. N., and Willmitzer, L. (2000). Metabolite Profiling for Plant Functional Genomics. *Nat. Biotechnol.* 18 (11), 1157–1161. doi:10.1038/81137

Foster, P. A., and Mueller, J. W. (2018). SULFATION PATHWAYS: Insights into Steroid Sulfation and Desulfation Pathways. *J. Mol. Endocrinol.* 61 (2), T271–T283. doi:10.1530/jme-18-0086

Garg, N., Hansson, A., Knych, H. K., Thevis, M., Hedeland, U., Globisch, D., et al. (2018). Structural Elucidation of Major Selective Androgen Receptor Modulator (SARM) Metabolites for Doping Control. *Org. Biomol. Chem.* 16 (5), 698–702. doi:10.1039/c7ob03030d

Ginestet, C. (2011). ggplot2: Elegant Graphics for Data Analysis. *Stat. Soc. Ser. A. (Statistics Soc.* 174 (1), 245–246. doi:10.1111/j.1467-985x.2010.00676_9.x

Godzien, J., Alonso-Herranz, V., Barbas, C., Armitage, E. G., and Armitage, G. (2015). Controlling the Quality of Metabolomics Data: New Strategies to Get the Best Out of the QC Sample. *Metabolomics* 11, 518–528. doi:10.1007/s11306-014-0712-4

Gomes, R. L., Meredith, W., Snape, C. E., and Sephton, M. A. (2009). Analysis of Conjugated Steroid Androgens: Deconjugation, Derivatisation and Associated Issues. *J. Pharm. Biomed. Anal.* 49, 1133–1140. doi:10.1016/j.jpba.2009.01.027

Gómez, C., Pozo, O. J., Geyer, H., Marcos, J., Thevis, M., Schänzer, W., et al. (2012). *Steroid Biochem. Mol. Biol.* 132 (3–5), 239–246. doi:10.1016/j.jsbmb.2012.05.010

Gómez, C., Pozo, O. J., Marcos, J., Segura, J., and Ventura, R. (2013). Alternative Long-Term Markers for the Detection of Methyltestosterone Misuse. *Steroids* 78 (1), 44–52. doi:10.1016/j.steroids.2012.10.008

Günal, S., Hardman, R., Kopriva, S., and Mueller, J. W. (2019). Sulfation Pathways from Red to green. *J. Biol. Chem.* 294 (33), 12293–12312. doi:10.1074/jbc.rev119.007422

Hartigan, J. A., and Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Appl. Stat.* 28 (1), 100. doi:10.2307/2346830

Hintikka, L., Kuuranne, T., Leinonen, A., Thevis, M., Schänzer, W., Halket, J., et al. (2008). Liquid Chromatographic-Mass Spectrometric Analysis of Glucuronide-conjugated Anabolic Steroid Metabolites: Method Validation and Interlaboratory Comparison. *J. Mass. Spectrom.* 43 (7), 965–973. doi:10.1002/jms.1434

Houghton, E., and Maynard, S. (2010). Some Aspects of Doping and Medication Control in Equine Sports. *Handb. Exp. Pharmacol.* 195, 369–409. doi:10.1007/978-3-540-79088-4_17

Jenkinson, C., Desai, R., McLeod, M. D., Mueller, J. W., Hewison, M., and Handelsman, D. J. (2021). Circulating Conjugated and Unconjugated Vitamin D Metabolite Measurements by Liquid Chromatography Mass Spectrometry. *J. Clin. Endocrinol. Metab.* 107 (2), 435–449. doi:10.1210/clinem/dgab708

Kaufman, L., and Rousseeu, P. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*, Vol. 344. John Wiley & Sons.

Kiousi, P., Fragkaki, A. G., Kioukia-Fougia, N., and Angelis, Y. S. (2021). Liquid Chromatography-Mass Spectrometry Behavior of Girard's Reagent T Derivatives of Oxosteroid Intact Phase II Metabolites for Doping Control Purposes. *Drug Test. Anal.* 13, 1822–1834. doi:10.1002/dta.3056

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 (6), 655–664. doi:10.1002/jms.635

McLeod, M. D., Waller, C. C., Esquivel, A., Balcells, G., Ventura, R., Segura, J., et al. (2017). Constant Ion Loss Method for the Untargeted Detection of Bis-Sulfate Metabolites. *Anal. Chem.* 89 (3), 1602–1609. doi:10.1021/acs.analchem.6b03671

Mueller, J. W., Vogg, N., Lightning, T. A., Weigand, I., Ronchi, C. L., Foster, P. A., et al. (2021). Steroid Sulfation in Adrenal Tumors. *J. Clin. Endocrinol. Metab.* 106 (12), 3385–3397. doi:10.1210/clinem/dgab18

Piper, T., Putz, M., Schänzer, W., Pop, V., McLeod, M. D., Uduwela, D. R., et al. (2017). *Drug Test. Anal.* 9 (11–12), 1695–1703. doi:10.1002/dta.2291

Piper, T., Schänzer, W., and Thevis, M. (2016). Revisiting the Metabolism of 19-nortestosterone Using Isotope Ratio and High Resolution/high Accuracy Mass Spectrometry. *J. Steroid Biochem. Mol. Biol.* 162, 80–91. doi:10.1016/j.jsbmb.2015.12.013

Pozo, O. J., Marcos, J., Khymenets, O., Pranata, A., Fitzgerald, C. C., McLeod, M. D., et al. (2018). *J. Mol. Endocrinol.* 61 (2). doi:10.1530/jme-17-0286

Pozo, O. J., Marcos, J., Ventura, R., Fabregat, A., and Segura, J. (2010). Testosterone Metabolism Revisited: Discovery of New Metabolites. *Anal. Bioanal. Chem.* 398 (4), 1759–1770. doi:10.1007/s00216-010-4082-0

Pranata, A., Fitzgerald, C. C., Khymenets, O., Westley, E., Anderson, N. J., Ma, P., et al. (2019). *Steroids*, 143.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Sangster, T., Major, H., Plumb, R., Wilson, A. J., and Wilson, I. D. (2006). A Pragmatic and Readily Implemented Quality Control Strategy for HPLC-MS and GC-MS-based Metabonomic Analysis. *Analyst* 131 (10), 1075. doi:10.1039/b604498k

Scarth, J. P., Teale, P., and Kuuranne, T. (2011). Drug Metabolism in the Horse: a Review. *Drug Test. Anal.* 3 (1), 19–53. doi:10.1002/dta.174

Schänzer, W., Guddat, S., Thomas, A., Opfermann, G., Geyer, H., and Thevis, M. (2013). *Drug Test. Anal.* 5 (11–12), 810–818. doi:10.1002/dta.1516

Schänzer, W. (1996). Metabolism of Anabolic Androgenic Steroids. *Clin. Chem.* 42 (7), 1001–1020. doi:10.1093/clinchem/42.7.1001

Schänzer, W., and Thevis, M. (2017). Human Sports Drug Testing by Mass Spectrometry. *Mass. Spec. Rev.* 36, 16–46. doi:10.1002/mas.21479

Stevenson, B. J., Waller, C. C., Ma, P., Li, K., Cawley, A. T., Ollis, D. L., et al. (2015). Pseudomonas Aeruginosaarylsulfatase: a Purified Enzyme for the Mild Hydrolysis of Steroid Sulfates. *Drug Test. Anal.* 7 (10), 903–911. doi:10.1002/dta.1782

Su, S., Law, C. W., Ah-Cann, C., Asselin-Labat, M.-L., Blewitt, M. E., and Ritchie, M. E. (2017). Glimma: Interactive Graphics for Gene Expression Analysis. *Bioinformatics* 33 (13), 2050–2052. doi:10.1093/bioinformatics/btx094

Team, R. C. R. (2017). *Foundation for Statistical Computing*. Vienna, Austria.

Thevis, M., Thomas, A., and Schänzer, W. (2011). Current Role of LC-MS(/MS) in Doping Control. *Anal. Bioanal. Chem.* 401 (2), 405–420. doi:10.1007/s00216-011-4859-9

Tolstikov, V. V., and Fiehn, O. (2002). Analysis of Highly Polar Compounds of Plant Origin: Combination of Hydrophilic Interaction Chromatography and Electrospray Ion Trap Mass Spectrometry. *Anal. Biochem.* 301 (2), 298–307. doi:10.1006/abio.2001.5513

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., et al. (2015). MS-DIAL: Data-independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. *Nat. Methods* 12 (6), 523–526. doi:10.1038/nmeth.3393

Uduwela, D. R., Pabis, A., Stevenson, B. J., Kamerlin, S. C. L., and McLeod, M. D. (2018). Enhancing the Steroid Sulfatase Activity of the Arylsulfatase from *Pseudomonas aeruginosa*. *ACS Catal.* 8 (9), 8902–8914. doi:10.1021/acscatal.8b02905

Vimercati, S., Büchi, M., Zielinski, J., Peduto, N., and Mevissen, M. (2017). Testosterone Metabolism of Equine Single CYPs of the 3A Subfamily Compared to the Human CYP3A4. *Toxicol. Vitro* 41, 83–91. doi:10.1016/j.tiv.2017.02.017

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. Pharm. Biomed. Anal.* 124, 198–206. doi:10.1016/j.jpba.2016.02.031

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. doi:10.1016/j.steroids.2014.09.006

World Anti-Doping Agency Wada (2017). *Athlete Biological Passport Operating Guidelines*.

Yao, Y., Wang, P., Shao, G., Del Toro, L. V. A., Codero, J., and Giese, R. W. (2016). Nontargeted Analysis of the Urine Nonpolar Sulfateome: a Pathway to the Nonpolar Xenobiotic Exposome. *Rapid Commun. Mass. Spectrom.* 30 (21), 2341–2350. doi:10.1002/rcm.7726

Yi, L., Dratter, J., Wang, C., Tunge, J. A., and Desaire, H. (2006). Identification of Sulfation Sites of Metabolites and Prediction of the Compounds' Biological Effects. *Anal. Bioanal. Chem.* 386 (3), 666–674. doi:10.1007/s00216-006-0495-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

In addition to the published work presented in Section 2.2 of this chapter, the following section presents related unpublished work.

2.3 Semi-targeted profiling of *in vivo* Phase II urinary metabolites of testosterone propionate in thoroughbred geldings

2.3.1 Introduction

As explored in the first part of this chapter, the use of growth promoters such as anabolic androgenic steroids (AAS) to gain advantage in competition remains a predominant method of doping in horse racing.[1] This is not only unethical but also incurs potential harm to the welfare of the animal. Thus, to protect both the integrity of the sport and the welfare of animals, the use of AAS as performance enhancing drugs is banned by the International Federation of Horseracing Authorities (IFHA).[2] Doping methods are constantly evolving with some recent examples including the use of endogenous steroids,[3] hormones,[4] pro-hormones,[5,6] designer drugs,[7] low-dosage drug-cocktails,[8] and gene-doping.[9,10] In accordance with this, the field of anti-doping has grown into a rich area of research focused on the development of new omics methodologies for the detection of AAS misuse. Some examples of these emerging methodologies include, transcriptomics, proteomics, metabolomics, and steroidomics approaches.[4,11–19] What differentiates these omics approaches is their focus on measuring the effect of a dopant on a given biological system, which differs from conventional analytical approaches focused on the direct detection of a biologically active compound(s) and metabolites.[13] Within anti-doping surveillance, there has been a significant interest in the analysis of intact Phase II conjugates of dopants, of which the main limiting factor has been the access to reference materials.[1,19–21] Further, intact Phase II AAS metabolites such as sulfate esters play an important role in anti-doping analysis as long-term biomarkers for both exogenous and endogenous steroids.[22–24] Of particular interest, sulfation of metabolites is reported to predominate in horses when compared to humans.[21]

The present study was inspired by work done by Cloteau *et al* (2021),[12] Kaabia *et al (2014)*,[3] Palermo *et al* (2017),[25] and Boccard *et al* (2011),[26] in which OPLS-DA models were used to perform untargeted steroidomic analysis of the urinary metabolome.

Herein, the kinetic excretion profiles of the urinary Phase II metabolome of three gelding horses that were administered intramuscularly with testosterone propionate (Testoprop®, 250 mg), were analysed using a semi-targeted metabolomics approach. The quantification of three sulfated metabolites discovered in the first part of this chapter was also performed, with two full kinetic excretion profiles investigated. This study builds on the techniques and methodologies established in the first part of this chapter and incorporates them into a multivariate workflow for the identification of steroidal Phase II biomarkers that could potentially be related to doping.

2.4 Results and Discussion

Urine samples were collected -72, -48, -24, 0 (pre-administration), 2, 4, 6, 8, 12, 24, 36, 48, 72 hours and then daily to 28 days after the administration of testosterone propionate from five gelding thoroughbreds (named **H1**-**H5**). However, due to sample availability those examined in this study span from −48, −24, 0 (pre-administration), 4, 12, 36, 72, 168, 192, 240, 288, 336 and 432 h after administration for three separate horses (**H1**, **H3**, & **H5**). The control samples used to account for inter-horse variability were pre-administration samples gathered from adjacent horses used in the study (i.e., for **H1**, control samples were used from **H3** and **H5**). Samples from a fourth horse (**H2**), the same horse from Section 2.2 of this chapter, were also examined at the 0, 12, and 36 h time points for metabolite confirmation, however, due to low urine sample availability, no excretion profile could be investigated.

A semi-targeted approach was adopted for the analysis of the samples. This had two main components; *Firstly*, a targeted approach focused on the quantification of two metabolites, testosterone sulfate (**TS, 1**) and epiandrosterone sulfate (**EAS, 2**) was performed. This initial analysis was aimed at their confirmation,

quantification, and then the collection of individual kinetic excretion profiles up to 432 h (18 days) following administration of testosterone propionate for three separate horses (**H1**, **H3**, & **H5**). *Next*, three metabolites were investigated, androst-4-ene-3β,17β-diol 17-sulfate (**TSD, 4**), 5α-androstane-3β,17α-diol, 3-sulfate (**ABAS, 3**) and dehydroepiandrosterone sulfate (**DHEAS**). The metabolites **TSD** and **ABAS** were confirmed in a fourth horse (**H2**) of the same administration study, however, they were not fully quantified due to the relatively small abundances in the urine samples. The concentration of **TSD** (**4**) was estimated using the calibration curve of **ABAS** (**3**) to semi-quantify the kinetic excretion profile. This was done by assuming an equal response factor, as **ABAS** (**3**) is also a steroid diol monosulfate. It should be noted that **DHEAS** was not detected in any horse in this administration study, and thus was not addressed in further analysis.

Secondly, the overall design of the experiment allowed for the untargeted profiling of the Phase II conjugated urinary steroidal metabolome of three horses over the 18-day time span of the study. This substrate generality was achieved through a WAX SPE extraction with minimal washing steps allowing for the examination of conjugated steroids. All biological samples were taken in triplicates for statistical analysis. Finally, data quality was checked with the use of pooled QC samples that were distributed uniformly throughout the analysis as done in Section 2.2 of this chapter.[1] This gave the necessary platform to allow for robust high throughput differential metabolite level analysis followed by multivariate analysis.

2.4.1 Detection and ionisation of testosterone propionate metabolites in LC-MS analysis

Detectability of all reference materials were assessed on the UHPLC-HRMS system. All steroid sulfate reference materials were detectable in the negative polarity as their $[M\text{-}NH_4]^-$ ions.

2.4.2 Confirmations of reference materials in equine urine

Confirmation of metabolites was first performed for the synthesised reference materials in the urine of **H2** of this study. This was done as the urine of **H2** was examined in Section 2.2 of this chapter. The following compounds, **TS** (**1**), **EAS**

(**2**), **ABAS** (**3**), and **TSD** (**4**), were confirmed using AORC and WADA criteria, Table 2.1. The confirmation is the subject of an application note done in collaboration with Shimadzu scientific, where it is covered in detail. Briefly, a collision energy spread (CES) function was utilised for the confirmation of the described steroid compounds. The CES function linearly ramps the collision energy through a set range. It was observed that CES allowed for a wider range of detectable product ions. This differs from the single CE conditions that were used in the first part of this chapter, where only one or two ions were present for steroid sulfates. For this reason, we used CES MS/MS in the data dependent acquisition throughout this study. Further, the synthesis and elucidation of **TSD** (**4**) is the subject of Chapter Three of this thesis. It should be noted that for the three horses (**H1**, **H3**, & **H5**) examined in this study, no matches to **ABAS** (**3**) were found. Matches based on retention time and exact mass were found for **TSD** (**4**).

Table 2.1: Confirmation of standard reference materials to their metabolites derived from the sulfate fraction of equine urine (**H2**), using UHPLC-HRMS/MS CES CID experiments. Each pair was tested against AORC and WADA criteria.[27,28] The CES ranges varied for each compound; testosterone sulfate (**1**), CES 60 ± 20. Epiandrosterone sulfate (**2**), CES 60 ± 30. 5α-Androstane-3β,17α-diol 3-sulfate (**3**), CES 60 ± 30. Androst-4-ene-3β,17β-diol 17-sulfate (**4**), CES 60 ± 20. Product ions were matched with a mass tolerance of ± 10 ppm and relative abundance was assessed by the peak area of the extracted ion chromatograms.

Reference material				**Metabolite**			
CES experiments	**Retention time (min) (%RSD, *n* =18) [theoretical *m/z*]**	**Product ion (*m/z*)**	**Relative abundance (%)**	**Retention time [tolerance] (min)**	**Relative abundance (%)**	**AORC[28] [%Tolerance]**	**WADA[27] [%Tolerance]**
Testosterone sulfate (1)	10.10 (0.12%) [367.1580]	*367.1563	27	10.08 [10.05-10.15]	32	[7-47]	[22-32]
		96.9588	100		100	[60-100]	[90-100]
		79.9567	8		3	[0-28]	[3-13]
		165.0217	0.4		1	[0-20]	[0-5]
		177.0206	4		7	[0-24]	[0-9]
		285.1831	0.4		1	[0-20]	[0-5]
		337.1089	1		2	[0-21]	[0-6]
		351.1259	6		11	[0-26]	[1-11]
Epiandrosterone sulfate (2)	10.99 (0.10%) [369.1740]	*369.1721	29	10.97 [10.94-11.04]	33	[9-49]	[23-35]
		96.9588	100		100	[60-100]	[90-100]
		79.9567	4		3	[0-24]	[0-8]
		259.0979	0.4		0.2	[0-20]	[0-5]
5α-Androstane-3β,17α-diol 3-sulfate (3)	12.32 (0.09%) [371.1898]	*371.1887	18	12.31 [12.27-12.37]	13	[0-38]	[13-23]
		96.9592	100		100	[60-100]	[90-100]
		79.9563	8		4	[0-28]	[3-13]
Androst-4-ene-3β,17β-diol 17-sulfate (4)	9.49 (0.15%) [369.1740]	*369.1721	13	9.47 [9.44-9.54]	8	[0-33]	[8-18]
		96.9588	100		100	[60-100]	[90-100]
		79.9567	9		7	[0-29]	[4-14]
		335.1288	4		4	[0-24]	[0-9]
		337.1466	46		40	[26-66]	[37-55]
		351.1598	1		4	[0-21]	[0-6]

* indicates precursor ions.

2.4.3 Targeted individual kinetic excretion profiles of reference materials in equine urine

Preliminary results for **TS** (**1**) and **EAS** (**2**) were assessed for reproducibility, repeatability, recovery, linearity, and sensitivity (i.e., limits of detection (LOD) and limits of quantification (LOQ = 3.3xLOD)), Table 2.2. These terms are defined in the experimental section. Calibration curves were performed by means of spiking 6 levels (0, 1, 2, 5, 10, 50 ng/mL) in stripped blank matrix and were generated by plotting the EIC peak area ratios of reference materials (**TS** and **EAS**) over internal standard nandrolone sulfate (negative mode) (IS concentration = 10 ng/mL). Linearity performance of the method was checked for each compound and used to define the *in vivo* ranges of the compounds in urine between a range of 0-20 ng/mL for three horses (**H1**, **H3**, **H5**). Following this, the response factors were used to obtain the kinetic excretion profiles in three horses (**H1**, **H3**, **H5**) over the time of the administration study from -48 hours to 432 hours, pre- to post-administration, Figures 2.1 & 2.2. In general, the profiles for **TS** and **EAS** were consistent for the three horses over the 18-day time span of the study. Specifically, **TS** was above the LOQ for approximately 1 week (168 h) and above detectable limits for up to 12 days (288 h). A similar profile was also observed for **EAS** but at lower concentrations, however, it was also above detectable limits for up to 12 days (288h). The profile of **H3** differed slightly from that of the other two horses and showed a smaller secondary spike in both TS and EAS levels at 288 h. This extended the detectable limit of both metabolites to two weeks (336 h) for **H3**.

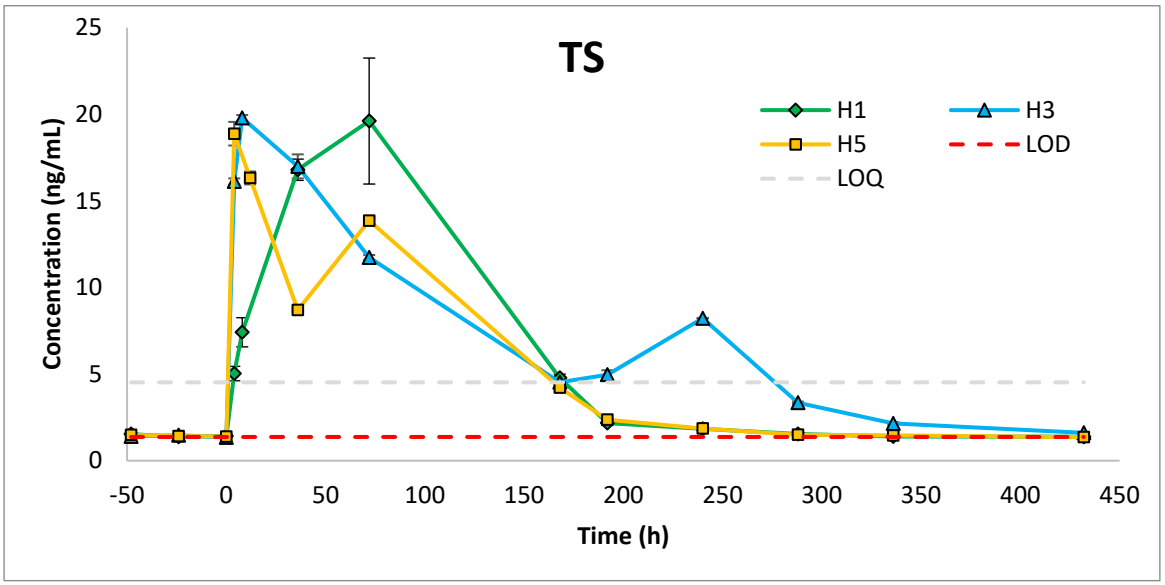


Figure 2.1: Kinetic excretion profiles of **TS** (**1**) in three horses following the administration of testosterone propionate. Error bars represent standard deviation, *n* = 3.

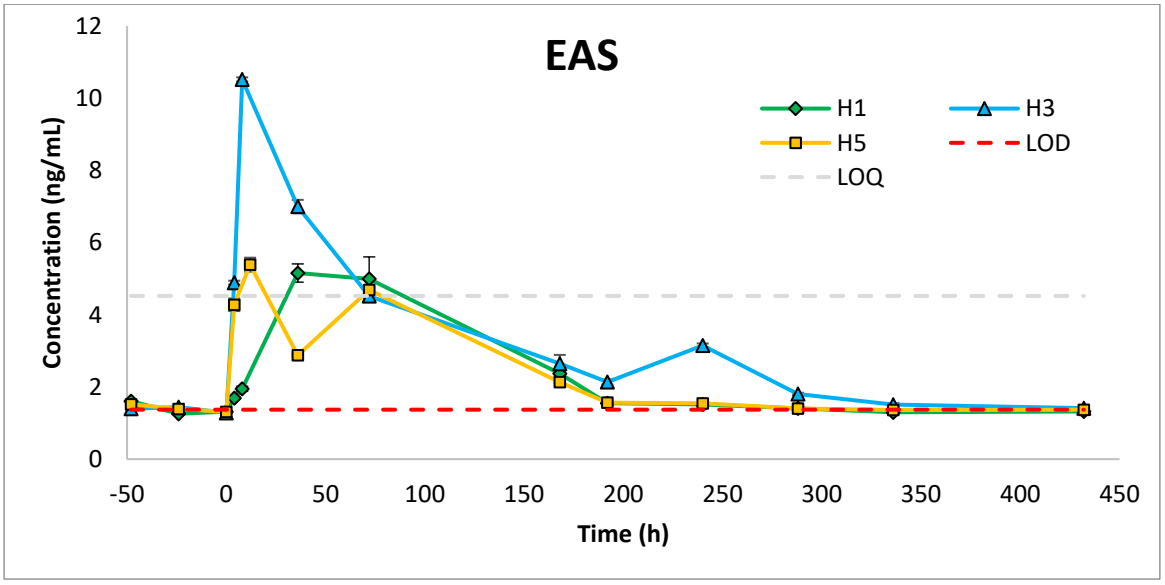


Figure 2.2: Kinetic excretion profiles of **EAS** (**2**) in three horses following the administration of testosterone propionate. Error bars represent standard deviation, *n* = 3.

Table 2.2: Retention time, repeatability, reproducibility, recovery at mid- and high-range calibrators for different steroidal refence standards in surrogate blank equine urine.

Compound	RT (min)	%RSD	Repeat a-bility (%RSD)[a] @ 10 ng/mL *n* = 4	Reproducibil -ity (%RSD)[a] @ 10 ng/mL *n* = 5	Lineari -ty (R^2)	Recove -ry (%) @ 100 ng/mL	LOD (ng/mL)	LOQ (ng/mL)
Testosterone sulfate (1)[b]	10.09[c]	0.11[c]	0.11	0.10	y = 0.23x – 0.2665 (0.99)	92	1.37	4.50
Epiandrostero ne sulfate (2)[b]	10.98[c]	0.13[c]	0.13	0.12	y = 0.19x – 0.1421 (0.99)	91	1.33	4.40
5α-Androstane-3β-17α-diol, 3-sulfate (3)	12.31[c]	0.11[c]	0.11	0.10	y = 0.14x – 0.1897 (0.99)	95	1.51	4.98

[a] Value for RSDmax given by the Horwitz equation is 32% for a concentration of 10 ng/mL.[29]

[b] Concentration range for calibration curve 0, 2, 5, 10, 50 ng/mL.

[c] n = 17.

2.4.4 Semi-targeted kinetic excretion profile of a minor metabolite of testosterone propionate

The metabolite, **TSD** (**4**), was one of ten putative steroid sulfate metabolites that were identified in part one of this chapter.[1] The elucidation and synthesis is the subject of its own study that used collision induced dissociation energy resolved experiments to aid in its confirmation, this is covered in Chapter Three of the thesis.

The kinetic excretion profile **TSD** (**4**) was qualitatively assessed in the three horses (Figure 2.3), the concentration was estimated from the response factor of **ABAS** (**3**), Table 2.2. This was justified based on both compounds being steroid diol mono-sulfates. However, it was acknowledged that **TSD** (**4**) has an un-saturated backbone, thus, a qualitative approach was taken. Further, the retention time of the *in vivo* metabolite was also matched to the synthesised reference material, Table 2.3. Following this, it was observed that **TSD** (**4**) had a similar excretion pattern to that detected for the two major metabolites **TS** (**1**) and **EAS** (**2**), suggesting that all three are from a related clearance pathway for testosterone propionate. Similarly, **TSD** (**4**) was also detectable to 12 days (288 h), however, the smaller concentrations likely indicate its status as a minor metabolite.

Table 2.3: The retention time of **TSD** (**4**) in surrogate stripped matrix (reference material) and as the in vivo metabolite in horse urine. Accurate mass matches were done at ± 5 ppm, allowable retention time differences were calculated from the AORC criteria.[28]

Matrix	Average RT [min]	Allowable difference
Reference material TSD (**4**)	9.47 [a]	9.44-9.54
H2 urine	9.49 [a]	
H1 urine	9.47 [b]	
H3 urine	9.47 [b]	
H5 urine	9.46 [b]	

[a] *n* = 18 samples.

[b] *n* = 56 samples.

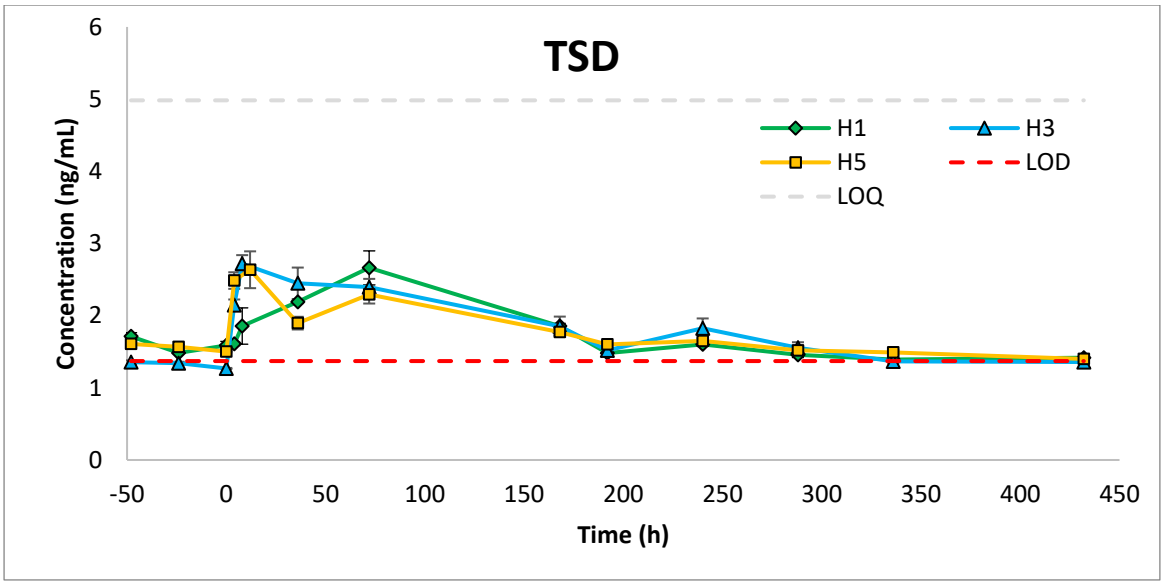


Figure 2.3: Kinetic excretion profiles of **TSD** (**4**) in three horses following the administration of testosterone propionate. Error bars represent standard deviation, *n* = 3.

2.4.5 Discussion of the targeted analysis

So far in this chapter, four metabolites have been examined in four separate horses. In horse two (**H2**), the identities of **TS** (**1**), **EAS** (**2**), **ABAS** (**3**) and **TSD** (**4**) were confirmed according to the relevant AORC criteria, Table 2.1. Moreover, **TS** and **EAS** are known sulfate metabolites of testosterone propionate administration.[1] **TS** is a direct metabolite of testosterone and **EAS** is one from epiandrosterone, a known downstream metabolite of testosterone in the biosynthetic pathway of steroids.[30,31] The metabolites were then studied in the remaining three horses, **H1**, **H3**, and **H5**. In these, only three of the four metabolites, **TS** (**1**), **EAS** (**2**), and **TSD** (**4**) were found based on accurate mass (± 5 ppm) and retention time matching. The absence of **ABAS** (**3**) is likely an example of inter-individual metabolic variability, and could be a unique metabolite of **H2**.[3,12,15] The remaining three metabolites were found to have similar patterns in their kinetic excretion profiles in **H1**, **H3** and **H5**. An excretion profile could not be obtained for **H2** due to lack of available sample. The similarities in excretion likely indicate that the three metabolites are of a related clearance pathway.

2.4.6 Semi-targeted metabolomic analysis of testosterone propionate administration samples; *a summary*

A brief exploratory discriminate analysis of each horse's (**H1**, **H3**, and **H5**) urinary profile was performed over the 18-day time span of the study. This analysis was aimed at targeting the Phase II metabolites of the 'steroidome'. To do this the urinary metabolome of each horse was analysed separately through the data processing pipeline and differential metabolite level analysis seen in Section 2.2 of this chapter. This included the MS-Dial driven alignment and LOWESS normalisation coupled with *k-means* analysis for putative sulfate conjugate identification.[1] Following this, multivariate analysis, specifically PCA and OPLS-DA, was used to highlight potential biomarkers. A pseudo-targeted approach was then adopted, where data was aligned using accurate mass against a list of putative Phase II steroidal conjugates either derived from known steroids or from the metabolic transformation of testosterone, a similar approach was adopted in both applications 1 and 2 in Section 2.2 of this chapter and later in Chapter Five.

Similar steroidomic approaches have also been reported by Boccard *et al* (2011),[26] and Palermo *et al* (2017),[25] which was previously discussed in the thesis introduction.

2.4.7 High throughput analysis of equine urinary metabolites

Data was acquired and processed using the same general workflow for each horse (**H1**, **H3** and **H5**). Acquisition of raw data was performed using the same UHPLC-HRMS/MS with data dependent acquisition as the targeted portion of this study. The collision energy was set at 57 ± 15 eV based on the high relative abundance of sulfate derived fragments observed in related work for steroid sulfates.[1,32] Following data acquisition, detected features were aligned, using suitable tolerances (RT ± 6 s and mass error (Δ) ± 5 ppm), and then normalised (LOWESS) in MS-Dial. This created an initial list of 8429, 14912, 14282 features, in **H1**, **H3** and **H5**, respectively. Features without MS/MS were removed and then each metabolome was annotated using the *k-means* analysis developed in Section 2.2 of this chapter.[1] This resulted in a total of 5971 (1170 sulfates and 4801 non-sulfates, **H1**), 9338 (1978 sulfates and 7360 non-sulfates, **H3**), and 8741 (2312 sulfates and 6429 non-sulfates, **H5**) remaining in each horse's metabolome, Table 2.4 and Figure 2.4.

2.4.8 Data quality assessment and PCA analysis

The multivariate analysis and data pre-treatment was performed in MetaboAnalyst 5.0.[33] As discussed extensively in Section 2.2 of this chapter, data quality is of major importance when performing non-targeted analysis of a metabolome since it ensures a non-biased and accurate interpretation of the outcome.[1,34] Data filtering was performed based on the comparison to the relative standard deviation (RSD) of the pooled QC samples, and based on inter-quantile range according to methods developed by *Hackstadt and Hess*. 2009.[35] Specifically, features with low repeatability were filtered out if their RSD was greater than > 25% of the RSD observed for that feature in the pooled QC samples. Also, features with near-constant values were excluded on the basis of their Inter-quartile range.[35] This resulted in final datasets of 4359 (**H1**), 6234 (**H3**), and 5910 (**H5**) that was subjected to multivariate analysis.

Table 2.4: Descriptive statistics for *k-means* analysis for the detected features in the **H1**, **H3**, and **H5** urine samples.

Horse	Annotation	*n*	Maximum Abundance (MA, %)			Intensity Ratio (IR, %)		
			Mean	Std.Dev.	Median	Mean	Std.Dev.	Median
H1	**Sulfates**	2321	95	10	100	63	25	67
	Non-Sulfates	4801	3	9	0	0	4	0
H3	**Sulfates**	1978	90	15	100	63	25	67
	Non-Sulfates	7360	3	10	0	1	4	0
H5	**Sulfates**	2312	93	13	100	54	21	54
	Non-Sulfates	6438	4	10	0	2	5	0

Std. Dev. = standard deviation.

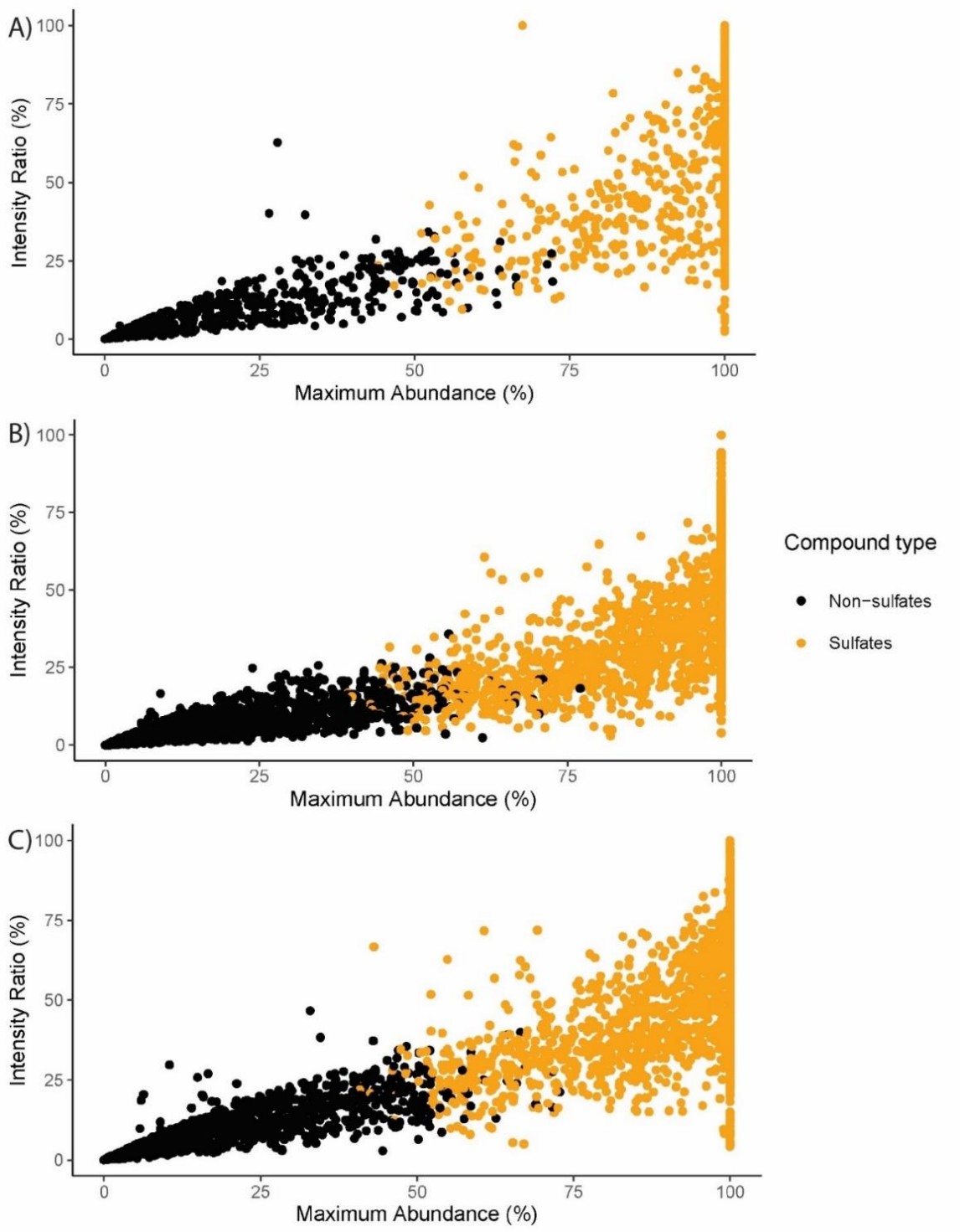


Figure 2.4: *k-means* clustering plots for horses **A) H1**, **B) H3**, and **C) H5**.

2.4.9 OPLS-DA model establishment and validation

The PCA of each data set showed tight grouping of the pQC samples compared to the variability observed across the collected samples, confirming the analytical robustness of each data set (Figure 2.5 A-C). To undergo OPLS-DA the data was grouped based on clustering observed in the non-supervised PCA.[36]

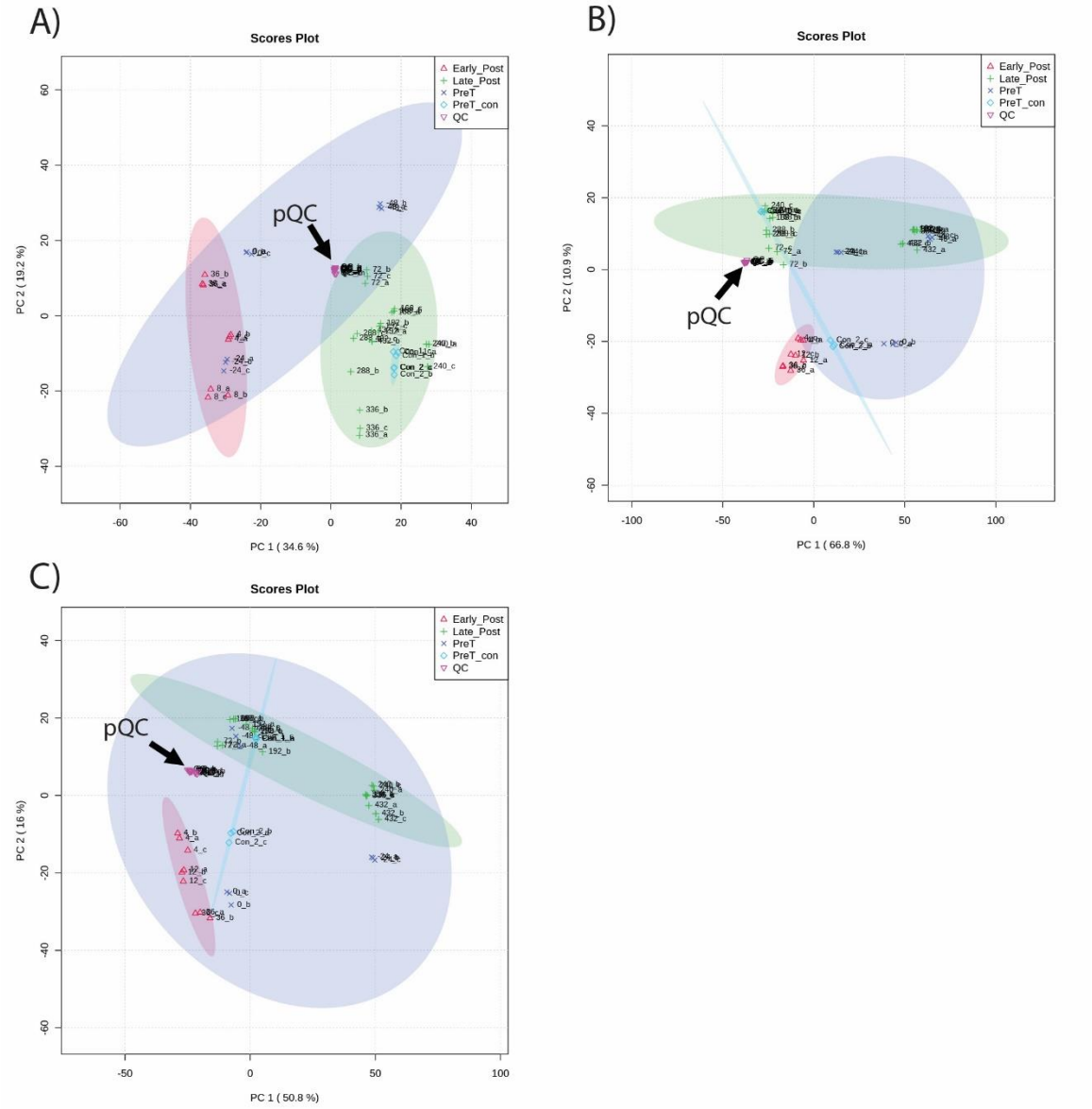


Figure 2.5: Multivariate analysis. PCA of horses **H1** (**A**), **H3** (**B**) and **H5** (**C**), with 5 groupings, pre-administration samples (PreT, blue), 1-2 days post-administration (Early_Post, pink), 3-18 days post-administration (Late_Post, green), non-doped control samples (PreT_Con, light-blue), and the pooled QC samples (pQC, purple). Ellipses indicate the 95 % CI interval of the groupings.

For all three horses, samples were assigned into three groupings, pre-administration (Pre_T, -48 h, -24 h, 0h, *n* =9), early post-administration (4h, 12 h (or 8 h in the case of **H1**), and 36h, *n* = 9), and late post-administration (72h-436 h, *n* = 21). The OPLS-DA was performed by examining the three possible permutations for each horse. Specifically, contrasts were modelled for the pre- and early post-administration, pre- and late post-administration, and early- and late post-administration.

Validation is an essential step in multivariate analysis and allows the omission of false correlations and ensures model robustness, ultimately, allowing for interpretation of the data for a specific biological question.[37] The OPLS-DA model used in this analysis was generated from the MetaboAnalyst 5.0 package,[33] which is a wrapper of the R package *rpols* from Bioconductor.[37] This package performs a cross validation model of the data which creates R2Y (R2), R2X and Q2 parameters. These explain the proportion of variance from the X matrix (or peak intensities), and Y matrix (or class/ groups), and the calculated validation, respectively. The R2 (goodness of fit) and Q2 (goodness of prediction) together are known as quality metrics and the values for each indicate the optimal number of components to use in model construction. Ideally Q2 and R2 should not be less than (>) 0.2 apart, as large variation can indicate overfitting from the use of too many components. The threshold of Q2 is variable in the literature, with an arbitrary value of Q2 being more than (>) 0.5 indicating good predictive ability of the model. However, values below this threshold have routinely been use and validated through the use of a permutation test, although in these instances caution should be used making final conclusions.[38] Lastly, it is customary to evaluate the cross validation with a permutation test. This was also performed in MetaboAnalyst 5.0. Here the same three values are calculated but are created from 1000 random permutations of the Y matrix (class/group category). It is important that the Q2, R2 and number of components used in the final model are reported to allow the reader to access the quality of a given model.[37]

In all three horses two of descriptive models were successfully validated using cross-validation. These were the pre- and early post-administration, and the, early- and late post-administration contrasts, Table 2.5, and Figure 2.6. The validation of the final contrast, pre- and late post-administration, failed in all horses. Specifically, all had Q2 values less than (<) 0.40, and the differences between R2 and Q2 were greater than (>) 0.2. Due to this, these contrasts were not examined any further to avoid superfluous conclusions from potential overfitting.

Table 2.5: OPLS-DA cross validation, permutation results, and S-Plot feature selection criteria.

Horse	**Contrast**	**Cross validation**			**Permutation (*n*=1000)**			**S-plot selection criteria**		
		R2(X)	**R2**	**Q2**	**R2**	**Q2**	**p-value**	**p(corr)[1]**	**p[1]**	**Features selected**
H1	PreT and Early_Post	0.24	0.66	0.59	0.99	0.99	<0.001	0.7	15	82
	Early_Post and Late_Post	0.29	0.855	0.788	0.94	0.93	<0.001	0.7	20	282
H3	PreT and Early_Post	0.54	0.82	0.80	0.99	0.99	<0.001	0.7	30	336
	Early_Post and Late_Post	0.19	0.62	0.60	0.99	0.99	<0.001	0.7	15	127
H5	PreT and Early_Post	0.33	0.75	0.69	0.99	0.99	<0.001	0.7	25	77
	Early_Post and Late_Post	0.32	0.76	0.74	0.99	0.99	<0.001	0.7	25	130

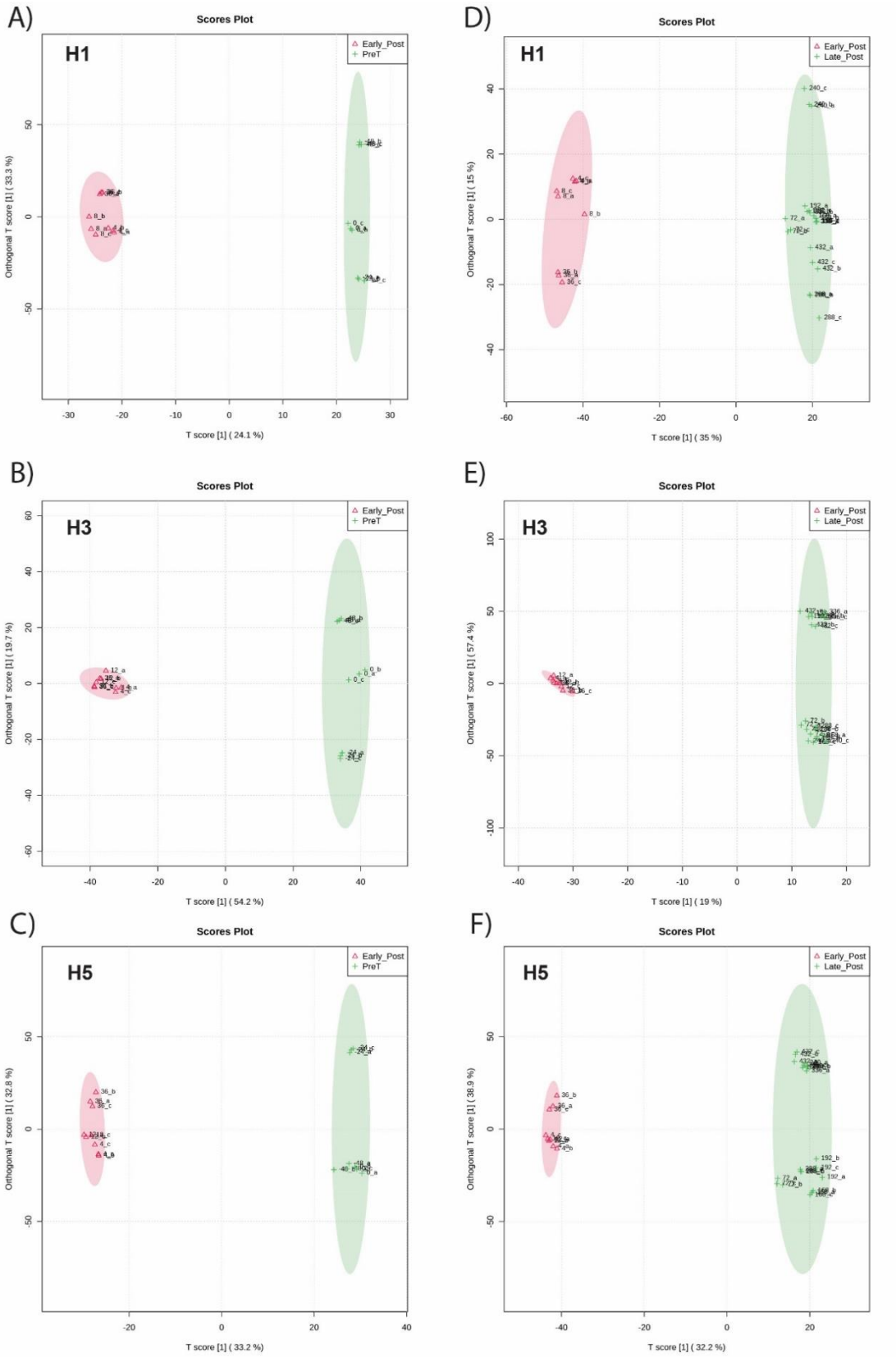


Figure 2.6: **A-C**) OPLS-DA score plot showing the separation between PreT (green) and Early_Post (pink) samples. **D-F**) OPLS-DA score plot showing the separation between Early_Post (pink) and Late_Post (green) samples, in **H1, H3** and **H5**, respectively. Ellipses indicate the 95 % CI interval of the groupings.

2.4.10 Exploratory discriminant analysis of the equine urinary metabolome with a focus on Phase II steroidal metabolites

The S-plot is a visualisation method that shows the covariance (p[1], x-axis) and modelled correlation (p(corr)[1], y-axis) to allow pinpointing of interesting features. It should be noted that the use of S-plots generally has a higher validity when there is clear class separation in the first component of the PCA, Figure 2.5 and 2.6.[3,12,36] In this analysis S-Plots were used to highlight the most predictive features in each contrast. Features were initially chosen based on being equal to or above the absolute values of arbitrary bounds for correlation (p(corr)[1]) and relative covariance (p[1]) in each plot, using the criteria outlined in Table 2.5.[36,39] The criteria were set to a broad threshold for correlation (p(corr)[1]) of 0.70, but the threshold for the relative covariance were determined on a case by case basis as deemed appropriate, Figure 2.7.[36,39]

Features within these initial bounds were then matched using their accurate mass (± 5 ppm) against the calculated masses of predicted Phase II metabolites derived from either the metabolic transformations of testosterone or the steroid biosynthetic pathway, Tables 2.6 and 2.7. Similar approaches to discern potential steroidal metabolites were used in Section 2.2 of this chapter and have been used in the literature for steroidomic analysis.[25,26,40] All matched masses were also searched (± 0.03Da) for on the MassBank and ChEBI online databases, with no matches being found.[41,42] This potentially highlights the need for a library of steroid Phase II metabolites to aid similar analysis.

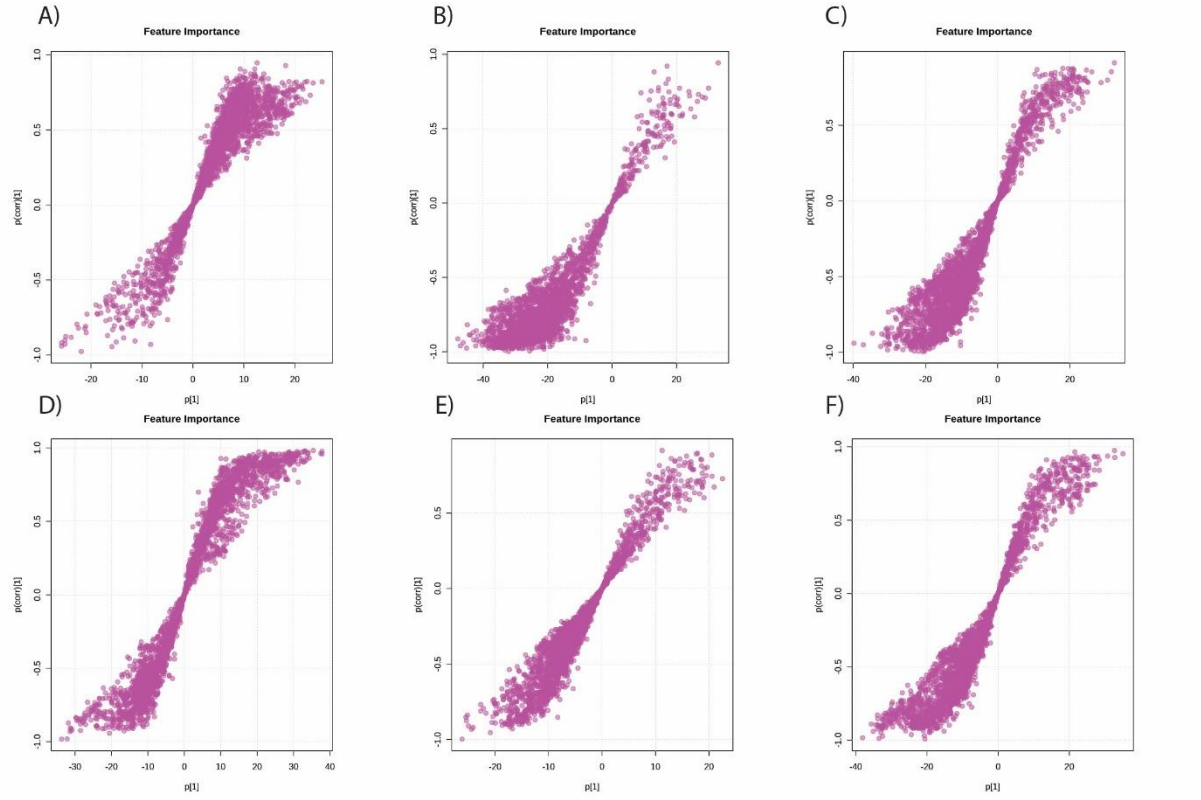


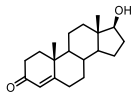
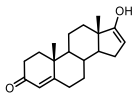
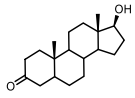
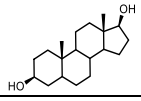
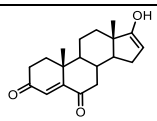
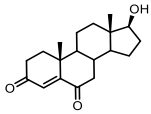
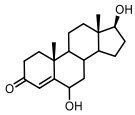
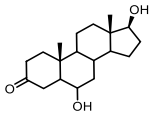
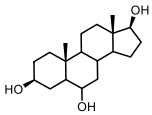
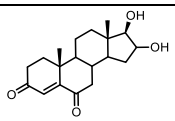
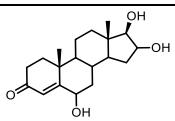
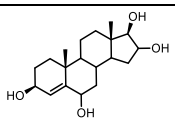
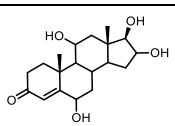
Figure 2.7: S-plots of the 6 contrasts modelled by OPLS-DA. In general, the x-axis (p[1]) represents the covariance of the normalised abundance of each feature, and the y-axis (p(corr)[1]) represents the measure of correlation of each feature to the model. **A-C**) PreT and Early_Post contrasts, **D-F**) Early_Post and Late_Post contrast, for **H1, H3** and **H5**, respectively.

Table 2.6: A list of reference *m/z* values related to endogenous steroid conjugates taken from steroidal biosynthetic pathways in mammalian systems.

Possible biomolecule	Formula of free steroid	Exact mass [M]	Sulfated $[M-H]^-$	Glucuronylated $[M-H]^-$
*Estrone	$C_{18}H_{22}O_2$	270.16198	349.111	445.1862
*Estradiol	$C_{18}H_{24}O_2$	272.17763	351.1266	447.2019
*Estriol	$C_{18}H_{24}O_3$	288.1725	367.1215	463.1968
*DHEA *Testosterone *Epitestosterone	$C_{19}H_{28}O_2$	288.20893	367.1579	463.2332
*Epiandrosterone *Androsterone *Etiocholanolone *DHT	$C_{19}H_{30}O_2$	290.22458	369.1736	465.2488
5β-Androstane-3α,17β-diol[43] *Androstane diol	$C_{19}H_{32}O_2$	292.24023	371.1892	467.2645
6β-Hydroxy-androstenedione[44] *11β-Hydroxy-androstenedione *11β-OH-androsterone *11β-OH-etiocholanolone	$C_{19}H_{26}O_3$	302.1882	381.1372	477.2125
*16α-Hydroxy-DHEA	$C_{19}H_{28}O_3$	304.20385	383.1528	479.2281
3,16-Dihydroxy-androstan-17-one[45] (no stereochemistry given)	$C_{19}H_{30}O_3$	306.2195	385.1685	481.2438
5α-Androstane-3β,16β,17β-triol[46]	$C_{19}H_{32}O_3$	308.23515	387.1841	483.2594
*Progesterone (as an enol)	$C_{21}H_{30}O_2$	314.22458	393.1736	489.2488
*Pregnenolone	$C_{21}H_{32}O_2$	316.24023	395.1892	491.2645
*3β-hydroxy-5α-pregnane-20-one	$C_{21}H_{34}O_2$	318.25588	397.2049	493.2801
*5β-Pregnane-3α, 20α-diol	$C_{21}H_{36}O_2$	320.27153	399.2205	495.2958
*17α-Hydroxy-progesterone[47]	$C_{21}H_{30}O_3$	330.2195	409.1685	505.2438
*17α-Hydroxy-pregnenolone	$C_{21}H_{32}O_3$	332.23515	411.1841	507.2594
*5α,3α-Tetrahydro deoxy-corticosterone	$C_{21}H_{34}O_3$	334.2508	413.1998	509.2751
*11-Dehydrocorticosterone	$C_{21}H_{28}O_4$	344.19876	423.1478	519.223
*11-Deoxycortisol *Corticosterone	$C_{21}H_{30}O_4$	346.21441	425.1634	521.2387
*Cortisone	$C_{21}H_{28}O_5$	360.19368	439.1427	535.2179
*Cortisol	$C_{21}H_{30}O_5$	362.20933	441.1583	537.2336
*18-Oxocortisol	$C_{21}H_{28}O_6$	376.18859	455.1376	551.2129
*18-Hydroxycortisol	$C_{21}H_{30}O_6$	378.20424	457.1532	553.2285

*biomarker found in multiple sources.[14,48–55]

Table 2.7: A list of reference *m/z* values related to possible Phase I transformations of Testosterone adapted from Section 2.2.[1]

Name	Formula of free steroid	Theoretical Free steroid Structure	Transformation	Exact mass	Sulfated	Glucuronylated
				[M]	$[M-H]^-$	$[M-H]^-$
Testosterone	$C_{19}H_{28}O_2$		No change	288.2090	367.1579	463.2332
T1	$C_{19}H_{26}O_2$		1 x Oxidation (enolation)	286.1933	365.1423	461.2175
T2	$C_{19}H_{30}O_2$		1 x Reduction	290.2246	369.1736	465.2488
T3	$C_{19}H_{32}O_2$		2 x Reduction	292.2402	371.1892	467.2645
T4	$C_{19}H_{24}O_3$		1 x Hydroxylation 2 x Oxidation (enolation)	300.1726	379.1215	475.1968
T5	$C_{19}H_{26}O_3$		1 x Hydroxylation 1 x Oxidation	302.1882	381.1372	477.2125
T6	$C_{19}H_{28}O_3$		1 x Hydroxylation	304.2038	383.1528	479.2281
T7	$C_{19}H_{30}O_3$		1 x Hydroxylation 1 x Reduction	306.2195	385.1685	481.2438
T8	$C_{19}H_{32}O_3$		1 x Hydroxylation 2 x Reduction	308.2352	387.1841	483.2594
T9	$C_{19}H_{26}O_4$		2 x Hydroxylation 1 x Oxidation	318.1831	397.1321	493.2074
T10	$C_{19}H_{28}O_4$		2 x Hydroxylation	320.1988	399.1478	495.223
T11	$C_{19}H_{30}O_4$		2 x Hydroxylation 1 x Reduction	322.2144	401.1634	497.2387
T12	$C_{19}H_{28}O_5$		3 x Hydroxylation	336.1937	415.1427	511.2180

2.4.11 Identifications of steroidal biomarkers

From the discriminate analysis 9, 7 and 16 putative steroid Phase II metabolites were matched against features for **H1**, **H3**, and **H5,** with a total 32 features, of which 23 were unique. The results, including the *k-means* annotation, MS, MS/MS spectra, the OPLS-DA model, and relative statistics are reported in Table 2.8.

Of the 32 features, 22 were assigned as sulfate metabolites, 5 were assigned as glucuronides and the remaining had inconclusive MS/MS spectra. Pleasingly, after manual assessment of the MS/MS in the majority of cases the *k-means* correctly assigned 18 of the 22 sulfated metabolites. There were only two cases where the annotation of a metabolite was deemed to be incorrect. The first one was for **estrone sulfate**, which was detected in both horses **H1** and **H5**. In both instances the MA and IR of estrone was more than one standard deviation away from the mean for non-sulfates (Table 2.4). Specifically, **estrone sulfate** had an average MA (36 in **H1**, and 35 in **H5**) and IR (17 in **H1**, and 15 in **H5**). However, the algorithm did not cluster it with sulfated metabolites as it is very different to the mean centres that were used to classify a sulfated metabolite, Table 2.4. A similar pattern was observed for the mis-annotation of the putatively assigned metabolite **T11S_B** in **H5**, a sulfated derivative of T11, Table 2.7. Again, the relatively small IR and MA were higher than that of the mean centres used to classify non-sulfated metabolites, specifically, MA was 19 and IR was 10, however, it fell outside of the classification of sulfated metabolites, Table 2.4. In each case the features are likely sulfated. This highlights the limitation of the non-supervised *k-means* algorithm, where sulfated metabolites with lower abundances are poorly classified, such as aromatic or conjugated sulfates. This could be potentially fixed in future iterations of the sulfate annotation workflow by use of a different unsupervised clustering algorithm or machine learning for pattern recognition. Analogous techniques have been used for the analysis of new psychoactive substances.[56,57]

Table 2.8: The list of matched *m/z* for molecules identified by the S-plots. The MS/MS spectra from the raw spectra are also shown.

Horse 1									
Putative assignment	***k-means***	**Contrast**	**Row Id**	**RT [min]**	***m/z***	**p [1]**	**p [corr][1]**	**MS/MS spectra: *m/z* (%RA)**	**Final MS/MS assignment**
Estrone sulfate*	**Non-Sulfate**	**PreT and Early_Post**	**2824**	**9.17**	**349.1116**	**-16**	**-0.76**	**269.1562(50), 183.080(14), 159.0825 (31), 145.0670(4), 143.0497(42)**	**Sulfate**
Testosterone sulfate*	**Sulfate**	**PreT and Early_Post**	**3265**	**10.06**	**367.1584**	**-22**	**-0.98**	**367.1570 (2), 351.1259(5), 177.0215 (6), 96.9593(100), 95.9522 (19), 79.9564 (13)**	**Sulfate**
T3S*	**Sulfate**	**PreT and Early_Post**	**3350**	**10.04**	**371.1896**	**-22**	**-0.82**	**96.9593 (100)**	**Sulfate**
T7S*	**Sulfate**	**PreT and Early_Post**	**3839**	**7.70**	**385.1690**	**-26**	**-0.92**	**96.9595 (100)**	**Sulfate**
Progesterone Sulfate* (as an enol)	**Sulfate**	**Both**	**4282**	**12.25**	**393.1741**	**29 [A]** **18 [B]**	**0.96 [A]** **0.73 [B]**	**96.9596 (100), 69.8181(8)**	**Sulfate**
T3G*	**Non-Sulfate**	**PreT and Early_Post**	**6333**	**11.30**	**467.2651**	**-24**	**-0.88**	**317.7739 (100)**	**N/A**
T7G	**Non-Sulfate**	**Early_Post and Late_Post**	**6655**	**7.96**	**481.2447**	**20**	**0.88**	**103.8167 (100)**	**N/A**
17α-Hydroxy-progesterone glucuronide*	**Non-Sulfate**	**Early_Post and Late_Post**	**7180**	**14.58**	**505.2445**	**-25**	**-0.78**	**281.2508 (100)**	**N/A**
Cortisol glucuronide	**Non-Sulfate**	**Early_Post and Late_Post**	**7660**	**12.61**	**537.2344**	**-24**	**-0.75**	**250.0486 (100)**	**N/A**

Table 2.8: continued…

Horse 3									
Putative assignment	***k-means***	**Contrast**	**Row Id**	**RT [min]**	***m/z***	**p [1]**	**p [corr][1]**	**MS/MS spectra: *m/z* (%RA)**	**Final MS/MS assignment**
Testosterone sulfate*	**Sulfate**	**PreT and Early_Post**	**5075**	**10.05**	**367.1589**	**-45**	**-0.98**	**351.1264 (6), 177.0222(6), 96.9596 (100),79.9565 (12)**	**Sulfate**
Epiandrosterone sulfate	**Sulfate**	**PreT and Early_Post**	**5140**	**10.95**	**369.1742**	**-37**	**-0.93**	**140.7885 (8), 96.9598 (100), 95.0494 (16),79.9567 (16)**	**Sulfate**
T3S*	**Sulfate**	**PreT and Early_Post**	**5215**	**10.02**	**371.1898**	**-40**	**-0.93**	**96.9600 (100)**	**Sulfate**
T8S*	**Sulfate**	**PreT and Early_Post**	**6081**	**9.98**	**387.1848**	**-46**	**-0.94**	**387.34061 (7), 96.9589 (100)**	**Sulfate**
Corticosterone sulfate	**Sulfate**	**PreT and Early_Post**	**8332**	**7.91**	**425.1650**	**-35**	**-0.93**	**397.0386 (6), 315.196 (21), 186.102 (16), 149.0231 (12), 114.0305 (6), 110.9747 (23), 96.9592 (100), 79.9588 (25), 79.0996(12),**	**Sulfate**
T3G*	**Sulfate**	**PreT and Early_Post**	**10060**	**11.28**	**467.2650**	**-37**	**-0.93**	**449.26721 (20), 307.4281 (7), 192.06461 (7), 96.9588 (100), 79.9586 (12)**	**Sulfate**
T5G	**Non-Sulfate**	**Early_Post and Late_Post**	**10521**	**11.23**	**477.2118**	**-15**	**-0.88**	**477.2179 (1), 306.14679 (21), 301.18231 (100), 257.155 (48), 119.7264 (22),**	**Glucuronide**

Table 2.8: continued…

Horse 5									
Putative assignment	***k-means***	**Contrast**	**Row Id**	**RT [min]**	***m/z***	**p [1]**	**p [corr][1]**	**MS/MS spectra: *m/z* (%RA)**	**Final MS/MS assignment**
Estrone sulfate*	**Non-Sulfate**	**PreT and Early_Post**	**4150**	**9.15**	**349.1110**	**-26**	**-0.98**	**269.1537 (32), 267.1403 (1), 183.0805 (10), 171.0808 (3), 169.0649 (2), 159.0811 (15), 145.065 (100), 143.0496 (33), 79.9564 (1)**	**Sulfate**
T1S	**Sulfate**	**PreT and Early_Post**	**4637**	**9.60**	**365.1422**	**-27**	**-0.88**	**349.1105 (6), 213.1272 (3), 183.9797 (2), 173.0237 (2), 96.9591 (1), 79.9564 (17)**	**Sulfate**
Testosterone sulfate*	**Sulfate**	**PreT and Early_Post**	**4696**	**10.04**	**367.1586**	**-30**	**-0.98**	**351.1264 (9), 177.0214(7), 96.9593 (100), 79.9565 (12)**	**Sulfate**
T3S*	**Sulfate**	**PreT and Early_Post**	**4820**	**10.01**	**371.1892**	**-30**	**-0.95**	**371.18829 (6), 106.9804 (3), 96.9594 (100), 79.9569 (17)**	**Sulfate**
T7S_B	**Sulfate**	**PreT and Early_Post**	**5434**	**8.88**	**385.1687**	**-30**	**-0.94**	**259.0997 (3),151.0400 (5), 96.9593 (100),79.9567 (9)**	**Sulfate**
T7S*	**Sulfate**	**PreT and Early_Post**	**5435**	**7.69**	**385.1690**	**-27**	**-0.88**	**287.027 (7), 271.1675 (7), 271.0661 (7), 162.0675 (6), 141.4577 (3), 135.0483 (7), 124.7723 (3), 96.9600 (100), 79.9568 (33)**	**Sulfate**
T8S*	**Sulfate**	**PreT and Early_Post**	**5564**	**9.97**	**387.1840**	**-40**	**-0.94**	**387.18491 (3), 96.9595 (100), 79.957 (10)**	**Sulfate**
Progesterone Sulfate* (as an enol)	**Sulfate**	**Early_Post and Late_Post**	**6105**	**12.24**	**393.1737**	**28**	**0.91**	**96.9592 (100), 79.9560 (16)**	**Sulfate**
T11S_A	**Sulfate**	**PreT and Early_Post**	**6544**	**8.69**	**401.1632**	**-30**	**-0.90**	**329.1424 (57), 301.1457(30), 287.1281 (40),271.101 (100), 245.1555 (41),109.0306 (18), 97.1361 (9), 95.9507 (15), 79.957 (28)**	**Sulfate**

Table 2.8: continued…

T11S_B	**Non-Sulfate**	**PreT and Early_Post**	**6545**	**8.77**	**401.1633**	**-31**	**-0.84**	**329.1521 (13), 301.1582 (21), 271.1031 (100), 219.072(22), 145.9005 (9), 118.767 (9), 79.9569 (19), 68.1482 (9)**	**Sulfate**
T11S_C	**Sulfate**	**PreT and Early_Post**	**6549**	**8.70**	**401.1649**	**-35**	**-0.87**	**329.1428 (100), 287.1298(73), 271.09949 (57), 219.0693 (26), 142.2317 (9), 129.0704 (21), 96.9584 (40), 95.9507 (31)**	**Sulfate**
Pregnenolone glucuronide	**Non-Sulfate**	**Early_Post and Late_Post**	**10098**	**11.09**	**491.2654**	**-32**	**-0.91**	**315.1958 (100)**	**Glucuronide**
17a-Hydroxy-pregnenolone glucuronide	**Non-Sulfate**	**Early_Post and Late_Post**	**10937**	**14.03**	**507.2596**	**-25**	**-0.90**	**317.5416 (19), 299.2003 (100)**	**N/A**
5α,3α-Tetrahydro-deoxy-corticosterone glucuronide	**Non-Sulfate**	**Early_Post and Late_Post**	**11054**	**14.64**	**509.2735**	**-26**	**-0.85**	**380.9835(17), 349.3172 (10), 333.2407 (100), 303.234 (73), 302.3379 (16), 299.2033 (14), 220.1702 (10), 133.8114 (11), 122.4146 (11), 87.0311 (13)**	**Glucuronide**
	Non-Sulfate	**Early_Post and Late_Post**	**11057**	**14.03**	**509.2757**	**-27**	**-0.91**	**333.2428 (100), 303.2319 (63), 85.0279 (2)**	**Glucuronide**
	Non-Sulfate	**Early_Post and Late_Post**	**11058**	**14.65**	**509.2757**	**-28**	**-0.89**	**333.2430 (100), 303.2315 (63), 139.18851 (12), 115.002 (25), 57.0265 (34)**	**Glucuronide**

[A] indicates data for PreT and Early_Post contrast. [B] indicates data for Early_Post and Late_Post contrast. * found in at least two horses. Row ID represents MS-Dial alignment identifier. RT [min] and *m/z* represent averaged values from alignment. For MS/MS spectra only major ions are listed with their relative abundances (%). **S** indicates a sulfated metabolite. **G** indicates a glucuronylated metabolite.

Next, 11 features across the three horses matched the exact mass of theoretical steroidal glucuronides, Table 2.8. Upon manual examination of the MS/MS spectra, 5 were putatively assigned as glucuronides by using accurate mass (± 5 ppm) and matching to characteristic MS/MS transitions. These transitions included neutral loss of 194 Da (loss of glucuronic acid) and 176 Da [loss of glucuronic acid–H_2O (gluc)], and the fragment ions *m/z* 175 [(gluc-H)-], *m/z* 157 [(gluc-H-H_2O)$^-$], *m/z* 113 [(gluc-H-H_2O-CO_2)$^-$], *m/z* 85 [(Gluc-H-H_2O-CO_2-CO)$^-$] and *m/z* 75 [($HOCH_2CO_2$)$^-$], the same approach was used in Section 2.2 of this chapter. Of the remaining 6 features, 5 had inconclusive information in the MS/MS to give an assignment, and 1 was classified as a sulfate.

Interestingly, several compounds were found in common to Boccard *et al.* (2011),[26] although several caveats exist for this comparison. Their study was performed in humans using a different testosterone derivative (testosterone undecanoate), that was orally administered, and the raw spectra was acquired using a different UHPLC-QToF analyser, and acquisition was performed using all ions fragmentation (MS^E) in MS/MS with a ramped collision energy. Despite these clear differences it is interesting to examine the crossover of significant biomarkers that were found in both studies. Specifically, the only biomarker confirmed against authentic standards in both studies was testosterone sulfate. This is not surprising as testosterone sulfate is a known urinary metabolite of testosterone.[26] Other similarities include the detection of several un-identified steroid sulfates with the same mass in both studies, of note in their study were, *m/z* 369.1752, 383.1539, 467.2649, 477.2073, 481.2446, that matched putatively identified molecules in our own study to a mass accuracy of ± 10 ppm, Table 2.8.

The differences between the two studies are also quite apparent. Our study takes advantage of the non-targeted MS/MS spectra by annotating sulfates in a semi-automated fashion, as well as the manual annotation of glucuronides. This gives the researcher more information on the unknown biomarkers, giving a better indication of their potential identity. Further, differences are notable in the proportion of putative sulfates and glucuronide metabolites detected. In our study, most putative metabolites were sulfated (22/32 approximately 69%), however, Boccard *et al.* (2011), found mainly glucuronylated metabolites (15/20, 75%). This

is likely a reflection of equine Phase II metabolism to be more selective towards sulfation over glucuronylation.[21] Once again, this assessment is only suggested as most of the biomarkers have not been confirmed against reference materials in our study.

Further, among the 32 features, only 2 were found to be significant by the OPLS-DA analysis in all horses, and 5 features were found in at least two horses, Table 2.8. To reduce any generalisations and the effect of inter-individual variability on the final outcomes of this study, only the 7 features that were found to be of significance in at least two horses were further analysed through the reconstruction of their kinetic excretion profiles, Figures 2.8 and 2.9. For the 5 molecules that were only identified in two of the three horses, the kinetic excretion profiles were reconstructed for the third horse (if possible) and are marked in Figure 2.9 with a triangle. As the OPLS-DA will only select the most predictive biomarker in each individual horse separately. It should be noted that many of the profiles marked with a triangle also follow similar patterns as the other two and may be more useful to give a general assessment of the biomarker in each horse. This difference in biomarker selection by the OPLS-DA models is expected, and likely highlights the inter-individual difference in metabolism and biological diversity that is possible between members of the same species. This is a common limitation in many single animal metabolomic and bioanalytical experiments.[3,12,58,59] One potential way around this would be to analyse all samples in a single experiment, however, due to time limitations this was not possible.

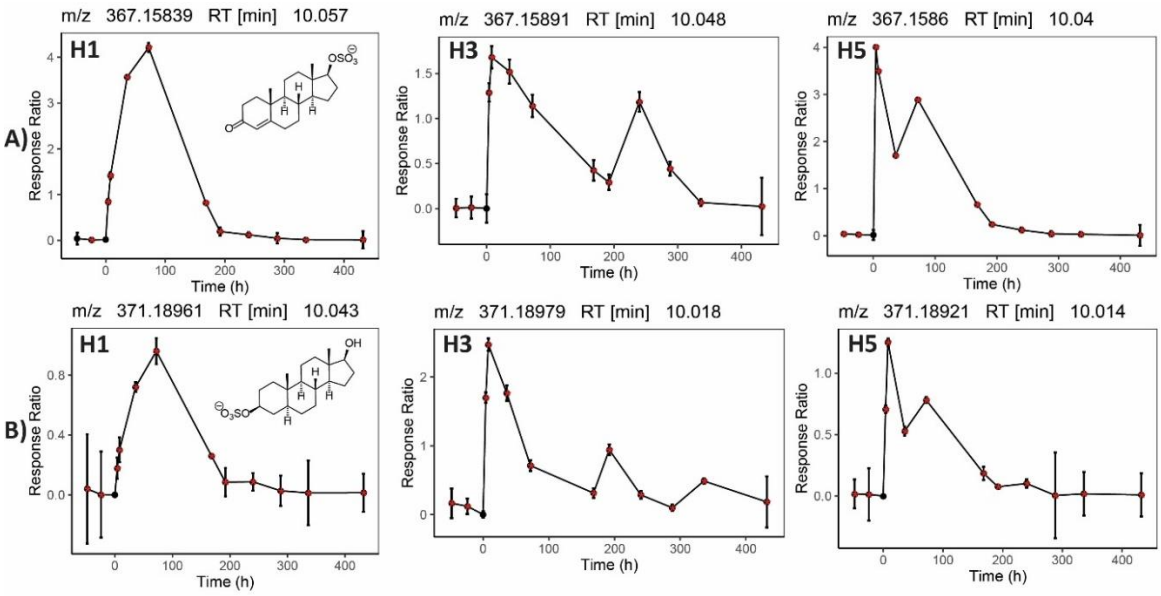


Figure 2.8: Kinetic excretion profiles. **A) Testosterone sulfate**. **B) T3S**, in **H1**, **H3** and **H5**. Inserts show known or putative structures. Response ratios indicates LOWESS normalised area divided by the IS area (nandrolone sulfate). Error bars represent standard error, SE = $\sigma/(n)^{1/2}$, $n = 3$. * p-value < 0.01, compared to response ratio at time 0 h.

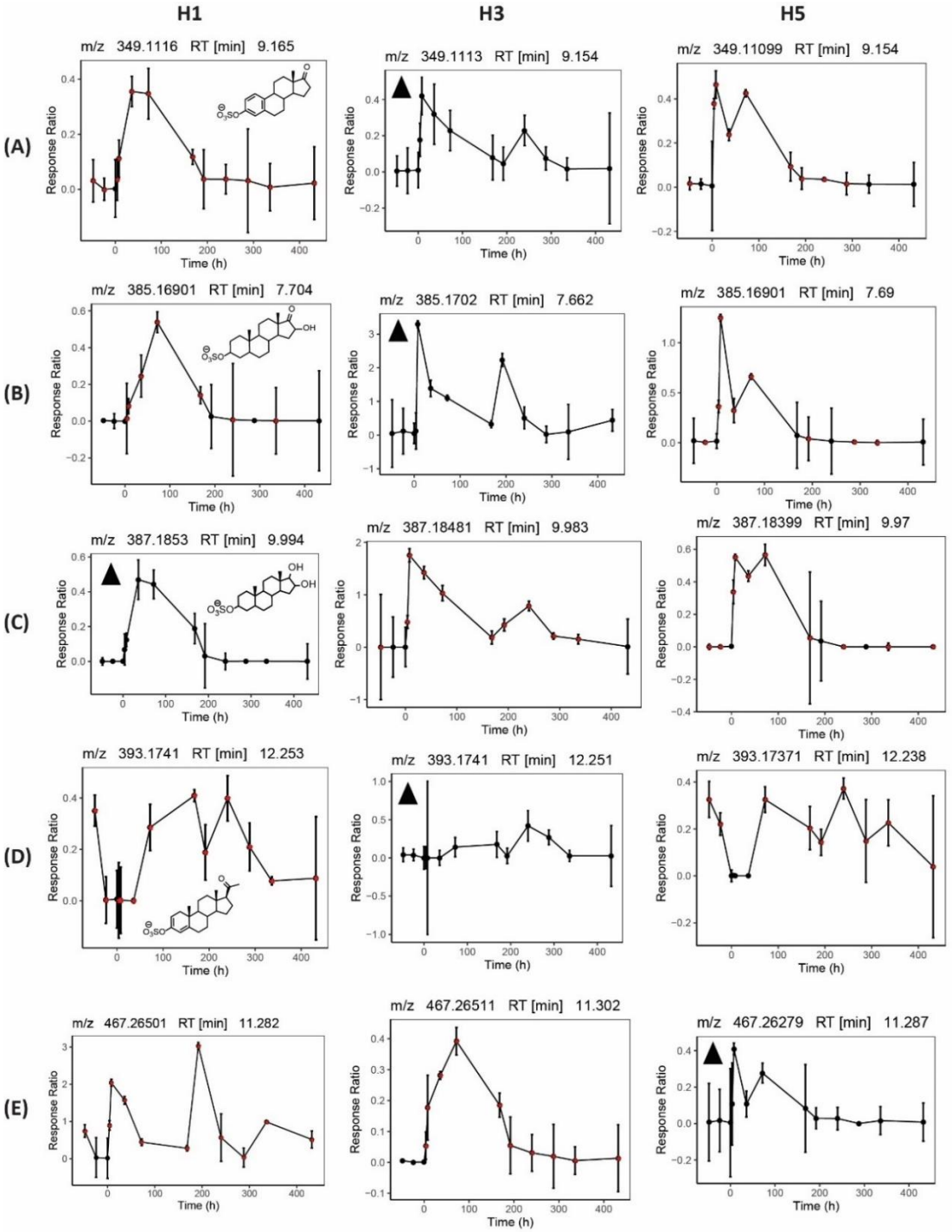


Figure 2.9: Kinetic excretion profiles of metabolites found that were changed in two of three horses **H1**, **H3** and **H5**. Putative assignments are as follows: **A) estrone sulfate**, **B) T7S**, **C) T8S**, **D) pregnenolone sulfate (as enol)**, **E) unknown**. Inserts show putative structures (if available). Black triangles show excretion profiles of molecules that were not found to be a significant biomarker by the OPLS-DA analysis, however, were still present based on *m/z* and RT matching. Response ratios indicates LOWESS normalised area divided by the IS area (nandrolone sulfate). Error bars represent standard error, SE = $\sigma/(n)^{1/2}$, $n = 3$. * p-value < 0.01, compared to response ratio at time 0 h.

Lastly, the 7 biomarkers were all identified in the OPLS-DA model as the most predictive upregulated features for the pre-administration to early post-administration contrasts. This suggested that Phase II steroidal metabolites are large drivers of differences in the urinary metabolome early post administration within the first 36h following administration of testosterone propionate. A similar pattern was found by Boccard *et al*. (2011) in the human urinary metabolome of orally administered testosterone undecanoate. Here, they found that the urinary steroidomic profile was at its most perturbed 12 hours following administration for measured steroidal Phase II conjugate markers. Although, in our study the biomarkers have in many cases not been confirmed against reference materials, they could still theoretically be used in the future as potential qualitative indicators of testosterone doping, to prompt a researcher to perform further analysis of a given sample.

The one exception of these general trends was the putatively assigned **progesterone sulfate (as enol)**, which was downregulated going from the pre- to the early post-administration samples (observed in **H1**), followed by being upregulated going to the late post administration samples (observed in **H1** and **H5**), Figure 2.9.

Of the two features found in all horses, one had already been identified in the first part of this study as **testosterone sulfate (1)**. This was found to be in the 10 most up-regulated features in each horse, for the pre-administration to early post-administration contrast. The second feature putatively assigned as a steroid mono-sulfate diol, **T3S**, was also found to be highly predictive, additionally it followed a similar excretion pattern to **TS**, as seen for **TSD** (**4**) at the start of this chapter. However, confirmation against an authentic reference material would have to be performed to assess this.

Interestingly, the two biomarkers assigned as putative downstream Phase I metabolites of testosterone sulfate, **T7S** and **T8S**, also possessed similar kinetic excretion profiles to **TS**, in their respective horse, Figure 2.9. This was also the case for the putatively assigned **estrone sulfate** that was found to be significant in **H1** and **H5**, which is a known downstream metabolite of testosterone.[21,60,61]

The assignment of **estrone sulfate** can be strengthened based on characteristic product ions obtained in its MS/MS spectra, Table 2.8 and Figure 2.10.[32,62,63] Specifically, the charge driven neutral loss of sulfur trioxide, 80 Da, giving the phenolic anion, *m/z* 269, and the phenolate ions, *m/z* 145 and 143, that are products of cycloreversion of the C-ring, and *m/z* 171 and 169, that are products of further charge driven decomposition. These 5 transitions are well-known and diagnostic of both estrone sulfate and related sulfated phenolic-estrane metabolites.[62,63] Interestingly, the product ions, *m/z* 183, 171 and 169, have only been observed in the literature as products of low-energy CID.[63] However, they are observed in this study, despite the use of high energy collision energy spread (CES, 57 eV ± 15 eV) acquisition. This indicates that the use of CES may give the ability to simultaneously access low and high energy product ions for these types of metabolites in a single acqusition.

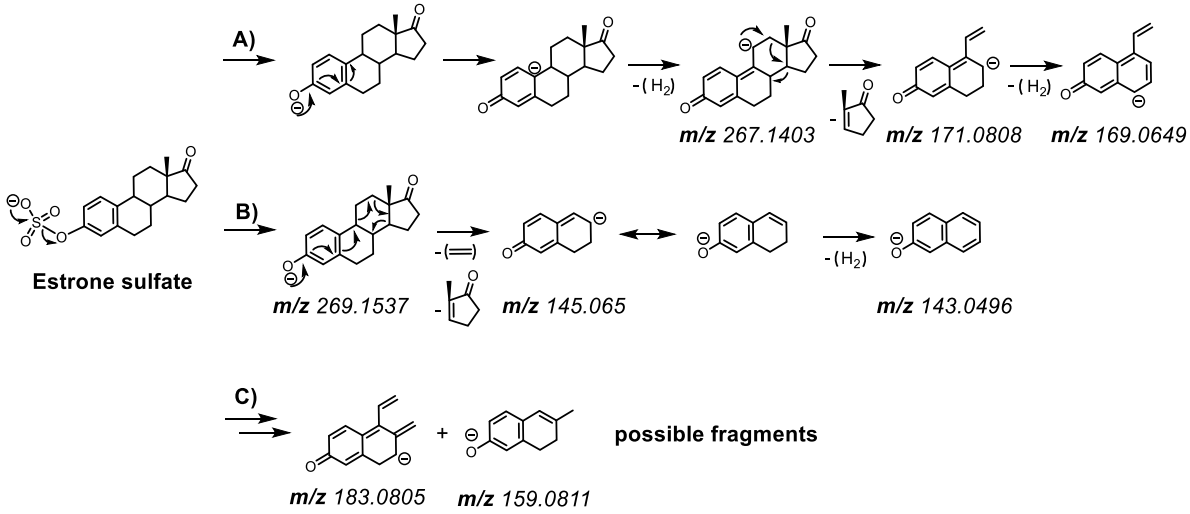


Figure 2.10: Possible fragmentation mechanisms for the decomposition of estrone sulfate under CID-CES conditions. The product ions shown were taken from the MS/MS spectra of putatively assigned estrone sulfate biomarker discovered in Horse 5, **H5**. A) charge-directed decomposition. B) cycloreversion pathway. C) Other possible product ion structures. Adapted from Attygalle *et al* (2001)[62] and Cole *et al* (2012).[63]

Although these putatively assigned steroidal biomarkers give promising leads in explaining the metabolism of testosterone propionate in the three horses studied, confirmation would have to be performed against an authentic reference material to ascertain further biological relevance. This represents a large body of work outside the scope of this thesis and would be the subject of future work.

2.5 Conclusion

The use of endogenous AAS remains a predominant method of mal-practice in horse racing. This study examined the *in vivo* metabolism of the synthetic ester derivative of testosterone, testosterone propionate, following intra-muscular administration in three thoroughbred geldings. The use of a semi-targeted metabolomics UHPLC-HRMS/MS approach allowed for the quantification of direct sulfate metabolites against authentic samples as well as a metabolite level analysis of the data for qualitative biomarker discovery. Two sulfated metabolites, **testosterone sulfate (1)** and **epiandrosterone sulfate (2)**, were quantified and their kinetic excretion profiles were established over an 18-day period post administration. Following this, the excretion profile of a third metabolite, **TSD** (**4**), was also qualitatively examined.

The importance of alternative methods for anti-doping testing analysis is becoming more relevant with the establishment of intelligence based anti-doping. The use of metabolomic approaches gives a suitable complementary analysis to already established targeted methods.[64] In this study, the use of exploratory multivariate discriminate analysis of the three horse's urinary metabolome highlighted that Phase II steroidal metabolites are likely drivers of differences in early *in vivo* metabolism post drug administration. Moreover, the utilisation of MS/MS data allowed for the putative annotation of possible conjugate groups in a semi-automated fashion. Ultimately, this semi targeted approach highlighted seven putative steroidal metabolites of testosterone propionate administration that were significantly upregulated in more than one horse.

2.5.1 Future directions and commentary

Various avenues of future research can be taken from this work. Firstly, in the multivariate analysis, several key steroidal metabolites were tentatively identified. Confirmation against authentic synthetic reference materials would allow for the use of these biomarkers for doping control purposes. Of these, the commercially available (100 mg is approximately AUD $220 from Merck, March 2022)[65], **estrone sulfate** remains a strong candidate based on the observation of characteristic

fragment ions observed in the MS/MS spectra. Moreover, the metabolite **T3S**, is also another strong candidate for identification, as a library of similar compounds were synthesised as the subject of Section 2.2 of this chapter (synthesis described below in the experimental section). However, due to time constraints the necessary confirmation analysis could not be undertaken in this thesis.

Next, the datasets developed and analysed in this semi-targeted study did not consider biomarkers that were not putative steroidal metabolites. Future analysis of the datasets could be retroactively performed to search for other types of biomarkers that may report on doping status. This would be an important analysis as it could find biomarkers that increase the time-frame for doping detection of testosterone propionate. Similar to analyses performed in studies of related steroids.[12] A good starting point would be the human metabolome data base, although this was initially not considered due to the equine subjects being used.[66] Ultimately, due to time constraints no such analysis was carried out.

Lastly, in both parts 1 and 2 of this chapter, and later in Chapter Five of this thesis, the automated annotation of conjugated metabolites using MS/MS data and *k-means* clustering was successfully employed for sulfated metabolites. This type of annotation approach could feasibly be extended to any conjugate with characteristic fragmentation in MS/MS. A strong candidate for the extension of this work would be glucuronide metabolites. Unlike sulfated metabolites, these have characteristic fragmentation in positive as well as the negative mode, increasing the potential applications of this type of analysis. Moreover, the use of data independent acquisition (DIA) could feasibly allow for full conjugate annotation of a given metabolome, however, certain caveats exist such as spectra deconvolution, chimeric spectra, greatly enlarged data-file size and more computationally expensive data processing.[67–69]

2.6 Experimental

2.6.1 Animal Administration and ethics approval.

Please see Section 2.2 of this chapter for details of the animal administration study and ethics approval.

2.6.2 UHPLC-HRMS DDA analysis for equine urine samples

The instrumentation used in this second study was different from the first study in Chapter 2.1. Here, the screen and metabolite confirmations used the following conditions. Analysis was carried out on a 9030 HR-Q-ToF mass spectrometer equipped with a standard ESI source interfaced to a Nexera LC-40D X3 pump system for chromatographic separation (all from Shimadzu). The column used was a Shim-pack Velox C18 column (1.8 μm, 2.1 x 100 mm, Shimadzu). The UHPLC separation was performed at flow rate 0.4 mL/min, using gradient mixing of two phases: Solution A; 20 mM ammonium formate in 100% water; and solution B; 20 mM ammonium formate in 100% methanol. The gradient was 0-0.5 min (10% B), 0.5-20.5 min (10-100% B), 20.5-23.0 min (100% B), 23.0-23.01 min (100-10% B), 23.01-27.5 min (10% B). The injection volume was 5 µL and column oven temperature was 40 ºC.

The use of the Q-ToF gave a greater scanning speed in DDA, which was roughly doubled with respect to the orbitrap used in Chapter 2.1. However, the Q-ToF had a lower resolution than the orbitrap (70000). For example, the average resolution of 30000 (FWMH) of the Q-ToF cannot resolve testosterone sulfate (*m/z* 367.1585) from testosterone phosphate (*m/z* 367.1680), the minimum resolution needed is approximately 38650 using the doublet method ($R = m_2/(m_2-m_1)$), however, would be able to differentiate their product ions of hydrogen sulfate (*m/z* 96.9601) and dihydrogen phosphate (*m/z* 96.9700), where the minimum resolution needed is approximately 9795.[70] This is sufficient for the *k-means* to differentiate between these molecules.

2.6.3 Parameters for CE spread and DDA MS experiments

For HRMS analysis mass calibration was performed in negative mode: using sodium iodide (NaI) clusters. Acquisition was performed at a resolution of approximately 30,000 for both scan (*m/z* 100 to 750) and for DDA (*m/z* 50-750). HRMS/DDA was collected between 1 and 23.5 minutes for the top 20 precursors in each survey scan. The event time was set to 0.1 ms (10 Hz) for the scan and 0.035 ms (28.6 Hz) for subsequent DDA scans, with a total loop time (duty cycle) of 0.800 seconds over a total of 21 events with dynamic mass exclusion windows of 3s. Collision induced dissociated experiments were performed using the CE spread (CES) function at CE 60 ± 15 eV in negative mode acquisition, with argon used as the collision gas. An exclusion list for precursor ions that did not emanate from the urine samples but were instead observed in extraction blank samples was employed. For experiments with QC samples, the injection sequence started by running two blanks followed by running at least five system suitability samples followed by two pooled QC samples, followed by the main sequence. Biological samples were analysed in a random order, with pooled QC samples dispersed evenly throughout. In targeted parallel reaction monitoring (PRM) MS/MS experiments for metabolite confirmation, a list of single *m/z* was selected and fragmented over multiple collision energies (CE), for both the reference material and biological samples. For data analysis, extracted ion chromatograms from these PRM experiments were then integrated using a mass filter of ± 5 ppm.

2.6.4 Extraction of equine urine samples for UHPLC-HRMS DDA analysis

The extraction of equine urine samples were based on an established method.[1] To an aliquot of urine (0.6 mL), phosphate buffer (0.3 mL, 100 mM, pH 7 at rt) was added and the samples were centrifuged (3000 x *g*, 5 min) to pellet solids. Supernatant (0.750 mL) was fortified with a mixture of analytical internal standards nandrolone sulfate, cholanediol bis(sulfate), and epiandrosterone ($^{18}O_3$)-sulfate, (**S1**, **S2** and **S3**, respectively, in Section 2.2 of this chapter), the final concentration was equivalent to 100 ng/mL original urine volume (0.5 mL) per standard. Biological pooled quality control (pQC) was made up of pooled aliquots (0.65 mL) of each biological sample which were redistributed and prepared as above.

Extraction was performed using Oasis WAX SPE cartridges (3 cc), that were pre-conditioned with methanol (2 mL) and water (2 mL). Supernatants were than loaded and washed with NaOH (2 mL, 0.1 M), phosphate buffer (2 mL, 100 mM, pH 5.4), and Milli-Q water (2 mL) before elution with a mixture of ethyl acetate: methanol: diethyl amine (25:25:1 v/v/v, 3 mL) into clean 10 mL glass tubes. All samples were then evaporated to dryness under a reduced pressure at 40 °C and stored at -20 °C until analysis. For analysis, the dried samples were re-dissolved in MeOH and filtered using 0.2 µm spin filters, then dried down under reduced pressure, and reconstituted with filtered MeOH: water (100 µL, 10 % v/v) and stored at 5 °C until analysis. This gave samples that were approximately 5 times the concentration of the original urine sample (0.5 mL).

2.6.5 Untargeted profiling of equine urine and semi-targeted metabolite confirmation studies

Untargeted profiling analysis for all biological samples were run in triplicate in a random order, with pQC samples dispersed evenly throughout. Data alignment and normalisation of total ion chromatograms (TIC) was done in MS Dial ver. 4.80 using tight tolerances in both retention time (RT) (+/- 6 s) and mass accuracy (*m/z* +/- 5 ppm), alongside the recommended settings for HRMS systems.[68] Normalisation was performed using the pQC samples via LOWESS function. The data quality and resulting features were assessed through PCA analysis in R. For negative mode screen MS/MS data was processed to identify potential sulfates in python and R.[1] Following this data was then exported in Excel. Data was then pre-treated (filtered, scaled and transformed) and then multivariate analysis (PCA and OPLS-DA) was performed in MetaboAnalyst 5.0.[33] Briefly, data filtering was performed based on comparison to the relative standard deviation (RSD, $x < 25\%$) of the pooled QC samples, and near zero values were filtered based on inter-quantile ranges. This filtering by variance has been shown to decrease the false discovery rate and were performed using the MetaboAnalyst 5.0 default program settings for data pre-treatment and done according to methods developed by *Hackstadt and Hess*. 2009.[35] The data is scaled via pareto scaling and then $\log_{10}$ transformed. This curated dataset is then subjected to PCA and OPLS-DA modelling. Following the OPLS-DA, the features identified in the S-Plot data are exported to R and

matched based on accurate mass to charge matching (± 5ppm). This was done using a steroid searcher algorithm that was written in R as a custom script and matched the *m/z* of detected peaks to the list of theoretical steroidal biomarkers within ± 5 ppm, for the list see Tables 2.7 and 2.8. The steroid searcher algorithm is available in the supplementary information of the manuscript presented in Chapter 2.2. These matched features are then processed using the Limma package from *bio-conductor* for univariate statistics and ggplot2 for kinetic excretion profiles.[71]

Metabolite confirmation studies were performed using PRM experiments and based on retention time, accurate mass and MS/MS match to an in-house synthetic library of standards. Metabolites were matched to reference standards according to AORC criteria.[28]

2.6.6 Linearity

Calibration curves were performed by means of spiking 6 levels (0, 1, 2, 5, 10, 50 ng/mL) in stripped blank matrix and were generated by plotting the EIC peak area ratios of reference materials (testosterone sulfate (**1**), epiandrosterone sulfate (**2**) and 5α-androstane-3β,17α-diol 3-sulfate (**3**)) over internal standard nandrolone sulfate (negative mode) (IS concentration = 10 ng/mL). Linearity of the method was checked for each compound and used to define the *in vivo* ranges of the compounds in urine between a range of 0-50 ng/mL. All linearity equations are given in Table 2.2. The coefficient of determination (r^2) ranged between 0.9942-0.9949 indicating good linearity for each measured compound in the study. All calibration curves were taken as technical triplicates, with two repetitions occurring before and after the sample batch, and one occurring on a different day to the initial run.

2.6.7 Recovery

Extraction recovery was performed for each measured reference material at a concentration of 100 ng/mL. Two separate samples were made up, pre and post addition samples. The general preparation is as follows; an aliquot of blank urine (0.6 mL), was added to phosphate buffer (0.3 mL, 100 mM, pH 7 at rt), and the

mixture was centrifuged (3000 x *g*, 5 min) to pellet solids. The supernatant (0.750 mL) was then treated to create pre- or post-addition samples. Pre-addition: the supernatant (0.750 mL) was fortified with a mixture of 100 ng/mL of reference material, the sample was then subjected to WAX SPE and reconstituted as above. Post-addition: the supernatant (0.750 mL) underwent WAX SPE and then fortified with the same mixture of reference materials. In both cases following extraction, eluents were fortified with 100 ng/ mL of internal standard (nandrolone sulfate) before reconstitution. The extraction recovery was calculated as the area ratios (areas: reference material/IS) of the post-addition over that of the pre-addition sample.

2.6.8 Repeatability

Repeatability was determined through the calculation of the relative standard deviation (RSD) of 4 urine samples spiked with 10 ng/mL of QC calibration points. Maximum values for repeatability were defined by the Horwitz equation as 32% at a concentration of 10 ng/mL.[29] The Horwitz equation is the normalised performance parameter indicating the acceptability of methods of analysis, usually with respect to inter-run or laboratory reproducibility (precision) but can also be used for repeatability (intra-run) evaluation. Repeatability is considered acceptable if the calculated RSD does not exceed 2/3 of the maximum RSD (RSD_{max}) predicted by the Horwitz equation. If this is criteria is met the analytical method is able to provide reputable (and or reproducible) measurements.[3,29,72]

2.6.9 Reproducibility

Reproducibility was determined through the calculation of the relative standard deviation (RSD) of 5 urine samples spiked with 10 ng/mL of QC calibration points on different days and different runs. Maximum values for reproducibility were defined by the Horwitz equation as 32% at a concentration of 10 ng/mL.[29]

2.6.10 Sensitivity

Limits of detection (LOD = 3 σ/b) [73] were estimated according to the calibration curves of individual analyte samples as response ratios, where σ is the standard deviation of the regression, b is the slope of the calibration curve.

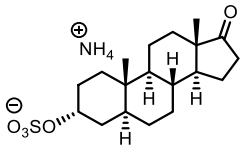
2.6.11 Accuracy and precision

Good accuracy was obtained for epiandrosterone sulfate (**2**) over the 10 ng/mL QC samples (n = 4) with a percentage relative error (%RE) ranging from -6% to -8%. However, poor accuracy was obtained over the 10 ng/mL QC samples for testosterone sulfate (**1**) with %RE ranging from -20% to -30%. Good precision was obtained for both quantified molecules across the QC samples with a percentage coefficient of variation (%CV) of 6.5% for testosterone sulfate (**1**) and 7.6% for epiandrosterone sulfate (**2**). It should be noted that the poor accuracy of testosterone sulfate (**1**) may have to be rectified in the future as this experiment could not be performed again due to time limitations.

2.7 Synthesis of steroid diol monosulfates

In addition to the published work presented in Chapter 2.1, the following section presents related unpublished work detailing the synthesis of the library of steroid diol monosulfate compounds used to identify 5α-androstane-3β,17α-diol 3-sulfate (**3**) in **H2**. The general procedures for synthesis and list of instrumentation, materials, reagents, and chemicals used for synthesis can be found in the supporting information of Section 2.2 of this chapter.

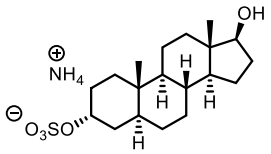
2.7.1 Androsterone-3-sulfate, ammonium salt (5)



Androsterone (5.00 mg, 17.2 µmol) was reacted according to general procedure S.4.3, using unlabelled sulfuric acid. This gave the title compound as a colourless solid (> 98% conversion).

^{1}H NMR (400 MHz, CD_3OD) δ 4.62 (m, 1H,C3-H), 2.44 (m, 1H, C16-H), 2.12 – 1.91 (m, 3H), 1.82 (m, 1H), 1.80 – 1.43 (m, 9H), 1.43 – 1.17 (m, 6H), 1.07 (m, 1H), 0.87 (s, 3H, C18-H_3), 0.86 (s, 3H, C19-H_3), 0.82 (m, 1H); LRMS (-ESI): *m/z* 369.3 (100%, $[C_{19}H_{29}O_5S]^-$); HRMS (-ESI) *m/z* calcd. for $[C_{19}H_{29}O_5S]^-$ ($[M\text{-}NH_4]^-$) 369.1738, found 369.1730. Spectroscopic data and spectra was found to match the literature for the compound.[74]

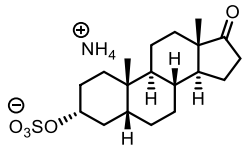
2.7.2 5α-Androstane-3α,17β-diol 3-sulfate, ammonium salt (6)



Androsterone 3-sulfate, ammonium salt **5** (5.00 mg, 13.6 µmol) was reacted according to general procedure S.4.4. This gave the title compound as a colourless solid (> 98% conversion).

^{1}H NMR (400 MHz, CD_3OD): δ 4.59 (m, 1H, C3-H), 3.57 (t, J 8.7, 1H, C17-H), 2.00 – 1.92 (m, 2H), 1.83 (m, 1H), 1.76 – 1.54 (m, 5H), 1.52 – 1.17 (m, 10H), 1.09 – 0.90 (m, 3H), 0.84 (s, 3H, C18-H_3), 0.75 (m, 1H), 0.72 (s, 3H, C19-H_3); ^{13}C NMR (101 MHz, CD_3OD) δ 82.5 (C17), 79.7, 55.9, 52.3, 46.4, 44.1, 38.3, 38.1, 36.9, 36.6, 36.4, 32.8, 30.7, 29.8, 24.3, 21.9, 12.7 (C18), 11.7 (C19), one peak obscured or overlapping; LRMS (-ESI): *m/z* 371.3 (100%, $[C_{19}H_{31}O_5S]^-$); HRMS (-ESI) *m/z* calcd. for $[C_{19}H_{31}O_5S]^-$ ($[M\text{-}NH_4]^-$) 371.1898, found 371.1889.

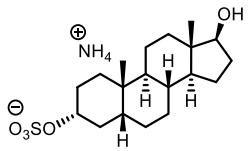
2.7.3 Etiocholanolone 3-sulfate, ammonium salt (7)



Etiocholanolone (5.00 mg, 17.2 µmol) was reacted according to general procedure S.4.3, using unlabelled sulfuric acid. This gave the title compound as a colourless solid (> 98% conversion).

^{1}H NMR (400 MHz, CD_3OD) δ 4.29 (m, 1H, C3-H), 2.43 (dd, J = 19.2, 8.6 Hz, 1H), 2.14 – 2.02 (m, 1H, C16-H), 2.02 – 1.81 (m, 5H), 1.77 (ddt, J = 11.6, 5.4, 2.5 Hz, 2H), 1.71 – 1.20 (m, 12H), 1.06 (m, 1H), 0.99 (s, 3H, C18-H_3), 0.87 (s, 3H,C19-H_3); LRMS (-ESI): *m/z* 369.3 (100%, $[C_{19}H_{29}O_5S]^-$; HRMS (-ESI) *m/z* calcd. for $[C_{19}H_{29}O_5S]^-$ ($[M\text{-}NH_4]^-$), 369.1737, found 369.1730. Spectroscopic data and spectra was found to match the literature for the compound.[74]

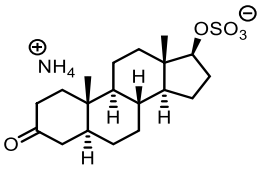
2.7.4 5β-Androstane-3α,17β-diol 3-sulfate, ammonium salt (8)



Etiocholanolone 3-sulfate, ammonium salt **7** (5.00 mg, 13.6 µmol) was reacted according to general procedure S.4.4. This gave the title compound as a colourless solid (> 98% conversion).

^{1}H NMR (700 MHz, CD_3OD) δ 4.28 (m, 1H), 3.59 (t, J=8.6, 1H), 2.04 – 1.80 (m, 6H), 1.76 (m, 1H), 1.59 (m, 1H), 1.52 – 1.39 (m, 7H), 1.36 – 1.19 (m, 3H), 1.15 – 0.98 (m, 3H), 0.96 (s, 3H), 0.72 (s, 3H); ^{13}C NMR (176 MHz, CD_3OD) δ 82.5, 80.4, 52.3, 44.2, 43.7, 42.1, 38.2, 37.3, 36.5, 35.7, 34.6, 30.7, 28.9, 28.1, 27.2, 24.3, 23.8, 21.5, 11.7; LRMS (-ESI): *m/z* 371.2 (100%, $[C_{19}H_{31}O_5S]^-$); HRMS (-ESI) *m/z* calcd. for $[C_{19}H_{31}O_5S]^-$ ($[M\text{-}NH_4]^-$), 371.18977, found 371.18892.

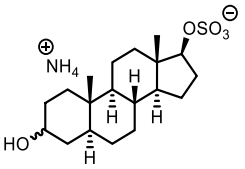
2.7.5 Dihydrotestosterone 3-sulfate, ammonium salt (9)



Dihydrotestosterone (5.00 mg, 17.2 µmol) was reacted according to general procedure S.4.3, using unlabelled sulfuric acid. This gave the title compound as a colourless solid (> 98% conversion).

^{1}H NMR (400 MHz, CD_3OD) δ 4.22 (t, J = 8.5 Hz, 1H), 2.49 (m, 1H), 2.35 (m, 1H), 2.28 – 1.92 (m, 5H), 1.80 – 1.12 (m, 13H), 1.07 (s, 3H), 1.07 (m, 1H), 0.83 (s, 3H), 0.77 (m, 1H); LRMS (-ESI): *m/z* 369.3 (100%, $[C_{19}H_{29}O_5S]^-$; HRMS (-ESI) *m/z* calcd. for $[C_{19}H_{29}O_5S]^-$ ($[M\text{-}NH_4]^-$) 369.17412, found 369.17353. Spectroscopic data and spectra was found to match the literature for the compound.[74]

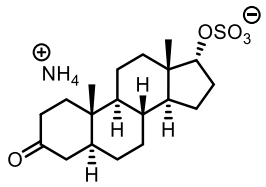
2.7.6 5α-Androstane-3β,17β-diol 17-sulfate, ammonium salt (10)



Dihydrotestosterone 3-sulfate, ammonium salt **9** (5.00 mg, 13.6 µmol) was reacted according to general procedure S.4.4. This gave the title compound as a colourless solid (> 98% conversion as a mixture of 3α-OH : 3β-OH in a 1:11 ratio). Data reported for major epimer where appropriate.

^{1}H NMR (600 MHz, CD_3OD) δ 4.21 (m, 1H), 3.51 (m, 1H), 2.15 (m, 1H), 1.94 (m, 1H), 1.78 – 1.67 (m, 4H), 1.65 – 1.49 (m, 4H), 1.48 – 1.09 (m, 9H), 1.06 – 0.88 (m, 2H), 0.84 (s, 3H), 0.79 (s, 3H), 0.68 (m, 1H); ^{13}C NMR (151 MHz, CD_3OD) δ 141.7, 123.0, 88.1, 79.8, 52.1, 51.7, 43.8, 40.4, 38.5, 37.8, 37.8, 33.2, 32.6, 30.0, 29.2, 24.4, 21.7, 19.8, 12.0; LRMS (-ESI): *m/z* 371.2 (100%, $[C_{19}H_{31}O_5S]^-$); HRMS (-ESI) *m/z* calcd. for $[C_{19}H_{31}O_5S]^-$ ($[M\text{-}NH_4]^-$), 371.18977, found 371.18908.

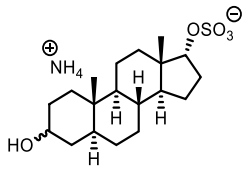
2.7.7 Epidihydrotestosterone 3-sulfate, ammonium salt (11)



Epidihydrotestosterone (5.00 mg, 17.2 µmol) was reacted according to general procedure S.4.3, using unlabelled sulfuric acid. This gave the title compound as a colourless solid (> 98% conversion).

^{1}H NMR (400 MHz, CD_3OD) δ 4.33 (d, J = 5.7 Hz, 1H), 2.49 (m, 1H), 2.37 (m, 1H), 2.28 – 1.90 (m, 5H), 1.81 – 1.63 (m, 4H), 1.61 – 1.31 (m, 8H), 1.22 (m, 1H), 1.07 (s, 3H), 1.01 (m, 1H), 0.82 (m, 1H), 0.77 (s, 3H); LRMS (-ESI): *m/z* 369.2 (100%, $[C_{19}H_{29}O_5S]^-$; HRMS (-ESI) *m/z* calcd. for $[C_{19}H_{29}O_5S]^-$ ($[M\text{-}NH_4]^-$) 369.17412, found 369.17353. Spectroscopic data and spectra was found to match the literature for the compound.[74]

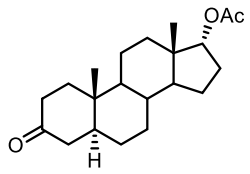
2.7.8 5α-Androstane-3α,17α-diol 17-sulfate, ammonium salt (12)



Epidihydrotestosterone 3-sulfate, ammonium salt **11** (5.00 mg, 13.6 µmol) was reacted according to general procedure S.4.4. This gave the title compound as a colourless solid (>98 % conversion, with a ratio of 1:11 of the 3α-OH : 3β-OH epimers). The data is reported for the major epimer where appropriate.

^{1}H NMR (600 MHz, CD_3OD) δ 4.31 (d, *J* = 5.8 Hz, 1H), 3.54 (m, 1H), 2.14 (m, 1H), 1.96 (m, 1H), 1.79 – 1.49 (m, 9H), 1.46 – 1.09 (m, 7H), 1.09 – 0.93 (m, 3H), 0.84 (s, 3H), 0.74 (s, 3H), 0.69 (m, 1H); LRMS (-ESI): *m/z* 371.2 (100%, $[C_{19}H_{31}O_5S]^-$); HRMS (-ESI) *m/z* calcd. for $[C_{19}H_{31}O_5S]^-$ ($[M\text{-}NH_4]^-$), 371.18977, found 371.18902. Spectroscopic data and spectra was found to match the literature for the compound.[20]

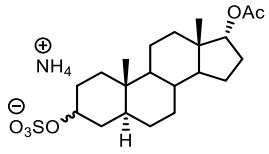
2.7.9 17α-Acetoxy-5α-androstan-3-one (13)



Acetic anhydride (0.5 mL, 5.29 mmol) was added dropwise to a stirring solution of epidihydrotestosterone (20.00 mg, 68.8 µmol) and DMAP (2 mg, 17.2 µmol) in pyridine (0.5 mL), and left at room temperature under a nitrogen atmosphere for 5 hours. The reaction was then extracted with water (1.5 mL), followed by ether (7.5 mL), and the organic phase washed with HCl (3 mL, 1M) followed by saturated sodium bicarbonate solution (3 mL). The organic layer was then dried with anhydrous magnesium sulfate and concentrated under reduced pressure. The crude was subjected to flash chromatography, which yielded the title compound as a colourless solid (20.2 mg, 60.8 µmol, 88 % yield).

^{1}H NMR (600 MHz, CD_3OD); δ 4.81 (d, *J* = 6.4 Hz, 1H), 2.47 – 2.25 (m, 3H), 2.19 (m, 1H), 2.11 (m, 1H), 2.03 (s, 3H), 2.01 (m, 1H), 1.81 – 1.71 (m, 2H), 1.68 – 1.10 (m, 13H), 1.01 (s, 3H), 0.96 (m, 1H), 0.76 (s, 3H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 212.14, 170.85, 82.00, 53.65, 50.01, 46.84, 44.90, 44.83, 38.75, 38.32, 35.89, 35.74, 32.13, 31.93, 30.14, 29.03, 24.81, 21.45, 21.00, 16.81, 11.64; LRMS (+ESI): *m/z* 687.4 (100 %, $[(C_{21}H_{32}O_3)_2Na]^+$); HRMS (+ESI) *m/z* calcd. for $[C_{21}H_{32}O_3Na]^+$ ($[M+Na]^+$), 355.2249, found 355.2236. Spectroscopic data and spectra was found to match the literature for the compound.[20]

2.7.10 17α-Acetoxy-5α-androstan-3β-ol 3-sulfate, ammonium salt (14)

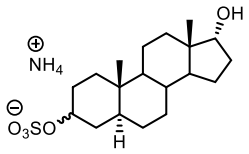


17α-Acetoxy-5α-androstan-3-one (**13**) (20 mg, 60.2 µmol) was reacted according to general procedure S.4.4. Following this the crude 17α-acetoxy-5α-Androstan-3β-ol (12.8mg, 38.3 µmol, 64 % isolated yield) was reacted according to general procedure S.4.3, using un-labelled sulfuric acid. This gave the title compound as a

colourless solid (14.4 mg, 34.8 µmol, 58 % isolated yield with a ratio of 1:7 of the 3α-OH : 3β-OH epimers). Data reported for major epimer.

^{1}H NMR (600 MHz, MeOD); δ 4.76 (d, J = 6.3 Hz, 1H), 4.27 (m, 1H), 2.18 (m, 1H), 1.85 – 1.72 (m, 4H), 1.67 – 1.28 (m, 13H), 1.27 – 1.14 (m, 2H), 1.09 – 0.96 (m, 3H), 0.86 (s, 3H), 0.82 (m, 1H), 0.77 (s, 3H), 0.71 (m, 1H); ^{13}C NMR (151 MHz, MeOD) δ 172.66, 83.50, 79.63, 55.55, 51.34, 46.26, 46.01, 38.29, 37.04, 36.55, 36.33, 33.57, 33.10, 30.82, 29.79, 29.77, 25.60, 21.74, 21.11, 17.09, 12.65; LRMS (-ESI): *m/z* 413.1 (100%, $[C_{21}H_{33}O_6S]^-$); HRMS (-ESI) *m/z* calcd. for $[C_{21}H_{33}O_6S]^-$ ($[M\text{-}NH_4]^-$), 413.1998 found 413.2001.

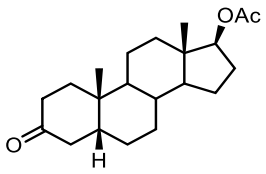
2.7.11 5α-Androstane-3β,17α-diol 17-sulfate, ammonium salt (15)



17α-Acetoxy-5α-androstan-3β-ol 3-sulfate, ammonium salt **14** (14.4 mg, 34.8 µmol) and potassium carbonate (12 mg) were suspended in methanol (2 mL), the solution was then stirred for 3 days. The methanol was then removed under reduced pressure, the crude was then dissolved in 10 mL of water and then purified according to general procedures S.4.1 and S.4.2, giving the title compound as a colourless solid (4 mg, 10.8 µmol, 31 % isolated yield, reported as ratio of 14:1 of the 3α-OH : 3β-OH epimers). The data is reported for the major epimer where appropriate.

^{1}H NMR (400 MHz, MeOD) δ 4.26 (m, 1H), 3.62 (d, J = 6.1 Hz, 1H), 2.22 – 2.08 (m, 2H), 2.09 – 1.97 (m, 2H), 1.89 – 1.70 (m, 4H), 1.71 – 1.53 (m, 2H), 1.52 – 1.28 (m, 7H), 1.27 – 1.13 (m, 2H), 1.12 – 0.97 (m, 2H), 0.88 (s, 3H), 0.72 (m, 1H), 0.68 (s, 3H); ^{13}C NMR (101 MHz, MeOD) δ 80.74, 79.69, 55.69, 49.94, 46.48, 46.30, 38.32, 37.17, 36.57, 36.35, 33.70, 32.83, 32.76, 29.87, 29.78, 25.61, 21.87, 17.59, 12.68; LRMS (-ESI): *m/z* 371.2 (100%, $[C_{19}H_{31}O_5S]^-$); HRMS (-ESI) *m/z* calcd. for $[C_{19}H_{31}O_5S]^-$ ($[M\text{-}NH_4]^-$), 371.1892 found 371.1903.

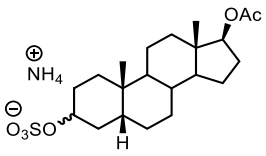
2.7.12 17β-Acetoxy-5β-androstan-3-one (16)



Acetic anhydride (0.5 mL, 5.29 mmol) was added dropwise to a stirring solution of 17β-hydroxy-5β-androstan-3-one (15 mg, 51.6 µmol) and DMAP (2 mg, 17.2 µmol) in pyridine (0.5 mL), and left at room temperature under a nitrogen atmosphere for 5 hours. The reaction was then diluted with water (1.5 mL), followed by extraction with ether (2x7.5 mL), the combine organic phases were then washed with HCl (3 mL, 1M) followed by saturated sodium bicarbonate solution (3 mL). The organic phase was then dried with anhydrous magnesium sulfate and concentrated under reduced pressure. The crude was then extracted using C18 SPE according to general procedure S.4.1, which gave the title compound as a colourless solid (16.5 mg, 49.6 µmol, 96 % isolated yield).

^{1}H NMR (600 MHz, $CDCl_3$) δ 4.62 (t, *J* = 9.4, 7.8, 1.7 Hz, 1H), 2.68 (m, 1H), 2.33 (m, 1H), 2.20 – 2.15 (m, 2H), 2.04 (s, 3H), 1.91 – 1.77 (m, 3H), 1.66 – 1.48 (m, 9H), 1.40 – 1.21 (m, 4H), 1.17 – 1.10 (m, 2H), 1.03 (s, 3H), 0.81 (s, 3H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 213.18, 171.32, 82.86, 50.94, 44.45, 42.93, 42.49, 41.02, 37.34, 37.20, 35.55, 35.16, 27.75, 26.65, 25.56, 23.67, 22.81, 21.33, 20.85, 12.28 (one peak missing or overlapping); LRMS (+ESI): *m/z* 687.4 (20 %, $[(C_{21}H_{32}O_3)_2Na]^+$); HRMS (+ESI) *m/z* calcd. for $[C_{21}H_{32}O_3Na]^+$ ($[M+Na]^+$), 355.2249, found 355.2250.

2.7.13 17β-Acetoxy-5β-androstan-3α-ol 3-sulfate, ammonium salt (17)

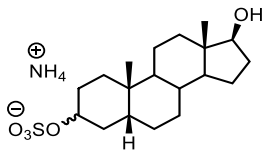


The steroid 17β-Acetoxy-5β-androstan-3-one (**16**) (4.1 mg, 12.3 µmol) was reacted according to general procedure S.2.4. Following this the crude, 17β-acetoxy-5β-androstan-3α-ol, was reacted according to general procedure S.4.3, using

unlabelled sulfuric acid. This gave the title compound as a colourless solid (4.2 mg, 10.2 µmol, 83 % isolated yield with a ratio of 11:1 of the 3α-OH : 3β-OH epimers). Data reported for major epimer where appropriate.

^{1}H NMR (600 MHz, MeOD) δ 4.49 (m, 1H), 4.17 (tt, *J* = 11.3, 4.8 Hz, 1H), 2.06 (m, 1H), 1.91 – 1.89 (m, 3H), 1.85 – 1.72 (m, 4H), 1.69 – 1.61 (m, 2H), 1.57 (m, 1H), 1.44 – 1.30 (m, 5H), 1.27 – 1.00 (m, 8H), 0.92 (m, 1H), 0.85 (s, 3H), 0.70 (s, 3H); 13C NMR (151 MHz, MeOD) δ 173.03, 84.43, 80.29, 51.98, 43.91, 43.67, 41.88, 38.31, 37.02, 36.47, 35.69, 34.57, 28.84, 28.61, 28.12, 27.12, 24.51, 23.79, 21.44, 20.95, 12.49 ; LRMS (-ESI): *m/z* 413.1 (100%, $[C_{21}H_{33}O_6S]^-$); HRMS (-ESI) *m/z* calcd. for $[C_{21}H_{33}O_6S]^-$ ($[M\text{-}NH_4]^-$), 413.1998 found .413.2024

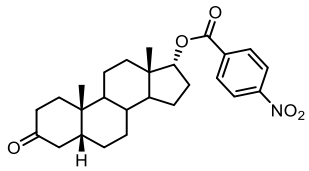
2.7.14 5β-Androstane-3α,17β-diol 3-sulfate, ammonium salt (18)



The steroid 17α-Acetoxy-5α-androstan-3α-ol 3-sulfate, ammonium salt **17** (4.2 mg, 10.2 µmol) and potassium carbonate (30 mg) were suspended in methanol (2mL), the solution was then stirred for 3 days. After which the methanol was removed under reduced pressure, the crude was then dissolved in 10 mL of water and then purified according to general procedures S.4.1 and S.4.2 giving the title compound as a colourless solid (1.2 mg, 3.2 µmol, 31 % isolated yield, reported as ratio of 6:1 of the 3α-OH : 3β-OH epimers). The data is reported for the major epimer where appropiate.

^{1}H NMR (700 MHz, MeOD) δ 4.28 (m, 1H), 3.59 (t, *J* = 8.7 Hz, 1H), 1.98 (m, 1H), 1.93 – 1.81 (m, 6H), 1.77 – 1.74 (m, 2H), 1.59 (m, 1H), 1.48 – 1.41 (m, 4H), 1.31 – 1.20 (m, 5H), 1.12 – 1.01 (m, 3H), 0.96 (s, 3H), 0.72 (s, 3H); ^{13}C NMR (101 MHz, MeOD) δ 82.50, 80.33, 52.34, 44.19, 43.72, 42.05, 38.17, 37.29, 36.49, 35.70, 34.58, 30.72, 28.85, 28.14, 27.18, 24.33, 23.84, 21.54, 11.65.; LRMS (-ESI): *m/z* 371.1 (100%, $[C_{19}H_{31}O_5S]^-$); HRMS (-ESI) *m/z* calcd. for $[C_{19}H_{31}O_5S]^-$ ($[M\text{-}NH_4]^-$), 371.1892 found 371.1897.

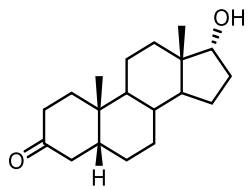
2.7.15 17α-[(4-nitrobenzoyl)oxy]-5β-Androstan-3-one (18)



Diethyl azodicarboxylate (DEAD, 44 µL, 0.25 mmol) was added dropwise to a stirring solution of 17β-Hydroxy-5β-androstan-3-one (30 mg, 0.1 mmol), 4-nitrobenzoic acid (42 mg, 0.25 mmol), triphenyl phosphine (70 mg, 0.25 mmol), in toluene (5 mL). The reaction was then heated to 80°C under an atmosphere of nitrogen and left for 3 days. The reaction was then quenched with water (15 mL) and extracted with ethyl acetate (45 mL). The organic layer was dried ($MgSO_4$) and concentrated under reduced pressure. The crude was then subject to flash chromatography, giving the title compound as a colourless solid (20.8 mg, 47.3 µmol, 47 % isolated yield).

^{1}H NMR (600 MHz, $CDCl_3$) δ 8.34 – 8.28 (m, 2H), 8.23 – 8.18 (m, 2H), 5.10 (d, J = 6.3, 0.9 Hz, 1H), 2.74 (m, 1H), 2.41 – 2.27 (m, 2H), 2.18 (m, 1H), 2.09 – 1.99 (m, 2H), 1.98 – 1.81 (m, 2H), 1.75 (m, 1H), 1.69 – 1.60 (m, 2H), 1.49 – 1.37 (m, 4H), 1.37 – 1.18 (m, 6H), 1.15 (m, 1H), 1.05 (s, 3H), 0.86 (s, 3H); ^{13}C NMR (151 MHz, $CDCl_3$) δ 213.13, 164.41, 150.66, 136.25, 130.78, 123.73, 83.94, 50.75, 45.55, 44.24, 42.48, 40.91, 37.33, 37.09, 36.01, 35.19, 32.53, 30.39, 26.72, 26.36, 24.99, 22.82, 20.75, 16.92 (two peaks overlapping); LRMS (+ESI): *m/z* 462.2 (60%, $[C_{26}H_{33}NO_5Na]^+$); HRMS (+ESI) *m/z* 462.2256 calcd. for $[C_{26}H_{33}NO_5Na]^+$ ($[M+Na]^+$), found 462.2266.

2.7.16 17α-Hydroxy-5β-androstan-3-one (19)

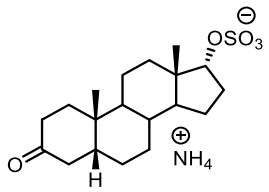


Sodium hydroxide (450 µL, 5 M) was added to a stirring solution of 17α-[(4-nitrobenzoyl)oxy]-5β-androstan-3-one **18** (5.8 mg, 13.2 µmol) in methanol (2 mL), and left at room temperature for 16 hours. The methanol was then taken off under

reduced pressure, and the crude was then dissolved in 10 mL of water and then purified according to general procedures S.4.1 and S.4.2 giving the title compound as a colourless solid (3.8 mg, 13.1 µmol, 99 % isolated yield).

^{1}H NMR (600 MHz, $CDCl_3$) δ 3.75 (m, 1H), 2.69 (m, 1H), 2.35 (m, 1H), 2.20 – 2.10 (m, 2H), 2.10 – 1.98 (m, 2H), 1.94 – 1.74 (m, 3H), 1.69 – 1.36 (m, 9H), 1.31 – 1.12 (m, 4H), 1.03 (s, 3H), 0.69 (s, 3H); ^{13}C NMR (151 MHz, $CDCl_3$) δ 213.38, 80.11, 48.97, 45.60, 44.57, 42.51, 40.83, 37.39, 37.33, 35.94, 35.18, 32.67, 31.75, 26.74, 26.31, 22.83, 20.92, 17.22 (one peak overlapping); LRMS (+ESI): *m/z* 313.2 (100%, $[C_{19}H_{30}O_2Na]^+$); HRMS (+ESI) *m/z* 313.2143 calcd. for $[C_{19}H_{30}O_2Na]^+$ ($[M+Na]^+$), found 313.2157.

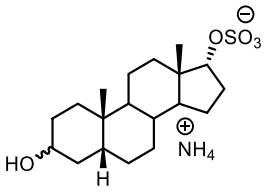
2.7.17 17α-Hydroxy-5β-androstan-3-one 17-sulfate, ammonium salt (20)



17α-Hydroxy-5β-androstan-3-one **19** (3.8 mg, 13.1 µmol) was reacted according to general procedure S.4.3, using unlabelled sulfuric acid. This gave the title compound as a colourless solid (> 98% conversion).

^{1}H NMR (600 MHz, CD_3OD) δ 4.13 (d, J = 5.8 Hz, 1H), 2.66 (t, J = 14.3 Hz, 1H), 2.27 (m, 1H), 2.03 – 1.84 (m, 3H), 1.80 – 1.69 (m, 3H), 1.67 – 1.52 (m, 2H), 1.47 – 0.95 (m, 12H), 0.84 (s, 3H), 0.57 (s, 3H); ^{13}C NMR (151 MHz, MeOD) δ 216.31, 87.98, 50.78, 46.34, 46.13, 43.27, 41.57, 38.40, 37.99, 37.22, 36.15, 32.96, 31.27, 27.77, 27.20, 25.53, 23.02, 21.79, 17.26; LRMS (-ESI): *m/z* 369.2 (100%, $[C_{19}H_{29}O_5S]^-$); HRMS (-ESI) *m/z* calcd. for $[C_{19}H_{29}O_5S]^-$ ($[M-NH_4]^-$), 369.1736 found 369.1742.

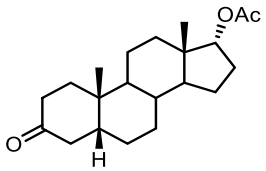
2.7.18 5β-Androstane-3α,17α-diol 17-sulfate, ammonium salt (21)



The steroid 5β-Androstane-3-one 17α-sulfate, ammonium salt **20** (5.00 mg, 13.6 µmol) was reacted according to general procedure S.4.4. This gave the title compound as a colourless solid (>98 % conversion with a ratio of 6:1 of the 3α-OH : 3β-OH epimers). The data is reported for the major epimer where appropriate.

^{1}H NMR (600 MHz, CD_3OD) δ 4.32 (d, J = 5.8 Hz, 1H), 3.51 (m, 1H), 2.01 – 1.69 (m, 6H), 1.60 – 1.26 (m, 12H), 1.26 – 1.14 (m, 3H), 1.02 (m, 1H), 0.96 (s, 3H), 0.74 (s, 3H); ^{13}C NMR (176 MHz, CD_3OD) δ 88.05, 72.55, 51.01, 46.32, 43.68, 41.73, 37.53, 37.25, 36.68, 35.80, 33.06, 31.30, 30.68, 28.34, 28.01, 25.59, 23.95, 21.41, 17.29; LRMS (-ESI): *m/z* 371.2 (100%, $[C_{19}H_{31}O_5S]^-$); HRMS (-ESI) *m/z* calcd. for $[C_{19}H_{31}O_5S]^-$ ($[M\text{-}NH_4]^-$), 371.1892, found 371.1894.

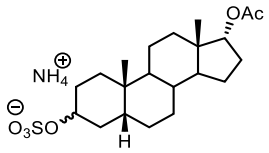
2.7.19 17α-Acetoxy-5β-androstan-3-one (22)



Acetic anhydride (0.5 mL, 5.29 mmol) was added dropwise to a stirring solution of 17α-hydroxy-5β-androstan-3-one **19** (5 mg, 17.2 µmol) and DMAP (2 mg, 17.2 µmol) in pyridine (0.5 mL), and left at room temperature under a nitrogen atmosphere for 5 hours. The reaction was then extracted with water (1.5 mL), followed by ether (7.5 mL), and the organic phase washed with HCl (3 mL, 1M) followed by saturated sodium bicarbonate solution (3 mL). The combined organic phases were then dried with anhydrous magnesium sulfate and concentrated under reduced pressure. The crude was then extracted using C18 SPE according to general procedure S.4.1, which gave the title compound as a colourless solid (5 mg, 49.6 µmol, 96 % isolated yield).

^{1}H NMR (400 MHz, $CDCl_3$) δ 4.82 (d, J = 6.2 Hz, 1H), 2.71 (m, 1H), 2.44 – 2.29 (m, 2H), 2.26 – 2.14 (m, 2H), 2.05 (s, 3H), 1.96 – 1.72 (m, 4H), 1.64 – 1.34 (m, 9H), 1.35 – 1.28 (m, 2H), 1.03 (s, 3H), 0.95 (m, 1H), 0.88 (m, 1H), 0.76 (s, 3H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 213.39, 170.86, 82.04, 50.24, 44.99, 44.39, 42.49, 40.74, 37.37, 37.20, 35.89, 35.14, 32.13, 30.23, 26.69, 26.26, 24.78, 22.80, 21.48, 20.77, 16.82; LRMS (+ESI): *m/z* 355.2 (100 %, $[C_{21}H_{32}O_3Na]^+$); HRMS (+ESI) *m/z* calcd. for $[C_{21}H_{32}O_3Na]^+$ ($[M+Na]^+$), 355.2249, found 355.2247.

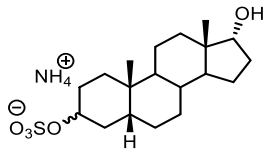
2.7.20 17α-Acetoxy-5β-androstan-3α-ol 3-sulfate, ammonium salt (23)



The steroid 17α-acetoxy-5β-androstan-3-one **22** (5 mg, 15.0 µmol) was reacted according to general procedure S.4.4. Following this the crude, 17βacetoxy-5β-androstan-3β-ol, was reacted according to general procedure S.4.3. using unlabelled sulfuric acid. This gave the title compound as a colourless solid (5 mg, 12.1 µmol, 81 % isolated yield with a ratio of 5:1 of the 3α-OH : 3β-OH epimers). Data is reported for major epimer where appropriate.

^{1}H NMR (600 MHz, MeOD) δ 4.76 (d, 1H), 4.27 (m, 1H), 2.20 (m, 1H), 1.95 – 1.84 (m, 4H), 1.81 – 1.74 (m, 3H), 1.59 – 1.41 (m, 10H), 1.36 – 1.27 (m, 3H), 1.27 – 1.19 (m, 3H), 1.05 (m, 1H), 0.96 (s, 3H), 0.76 (s, 3H); ^{13}C NMR (151 MHz, MeOD) δ 172.77, 83.58, 80.30, 51.36, 46.06, 43.63, 41.71, 37.39, 36.48, 35.68, 34.57, 33.23, 30.92, 28.86, 28.16, 27.91, 25.64, 23.78, 21.36, 21.16, 17.07; LRMS (-ESI): *m/z* 413.1 (100%, $[C_{21}H_{33}O_6S]^-$); HRMS (-ESI) *m/z* calcd. for $[C_{21}H_{33}O_6S]^-$ ($[M-NH_4]^-$), 413.1998 found 413.2031.

2.7.21 5β-Androstane-3α,17α-diol 3-sulfate, ammonium salt (24)



The steroid 17α-acetoxy-5β-androstan-3β-ol 3-sulfate, ammonium salt **23** (5.00 mg, 13.6 μmol) was reacted according to general procedure S.4.4. This gave the title compound as a colourless solid (>98 % conversion with a ratio of 5:1 of the 3α-OH : 3β-OH epimers). The data is reported for the major epimer where appropriate.

^{1}H NMR (700 MHz, MeOD) δ 4.27 (m, 1H), 3.63 (d, J = 6.1 Hz, 1H), 2.14 (m, 1H), 1.96 – 1.84 (m, 4H), 1.79 – 1.70 (m, 2H), 1.61 (m, 1H), 1.59 – 1.38 (m, 7H), 1.36 – 1.27 (m, 3H), 1.25 – 1.10 (m, 2H), 1.05 (m, 1H), 0.96 (s, 3H), 0.75 (m, 1H), 0.66 (s, 3H); ^{13}C NMR (176 MHz, MeOD) δ 80.84, 80.40, 49.97, 46.56, 43.72, 41.82, 37.53, 36.53, 35.71, 34.60, 32.99, 32.86, 28.89, 28.25, 28.04, 25.63, 23.86, 21.50, 17.58; LRMS (-ESI): *m/z* 371.1 (100%, $[C_{19}H_{31}O_5S]^-$); HRMS (-ESI) *m/z* calcd. for $[C_{19}H_{31}O_5S]^-$ ($[M\text{-}NH_4]^-$), 371.1892 found 371.1899.

2.8 References

1. Fitzgerald, C. C. J. *et al.* Profiling Urinary Sulfate Metabolites With Mass Spectrometry. *Front. Mol. Biosci.* **9**, 829511 (2022).

2. International Federation of Horse Racing Authorities. International agreement on race horse welfare. www.IFHAonline.org.

3. Kaabia, Z. *et al.* Monitoring the endogenous steroid profile disruption in urine and blood upon nandrolone administration: An efficient and innovative strategy to screen for nandrolone abuse in entire male horses. *Drug Test. Anal.* **6**, 376–388 (2014).

4. Joré, C. *et al.* Liquid chromatography – high resolution mass spectrometry-based metabolomic approach for the detection of Continuous Erythropoiesis Receptor Activator effects in horse doping control. *J. Chromatogr. A* **1521**, 90–99 (2017).

5. Viljanto, M. *et al.* Monitoring dehydroepiandrosterone (DHEA) in the urine of Thoroughbred geldings for doping control purposes. *Drug Test. Anal.* **10**, 1518–1527 (2018).

6. Viljanto, M. *et al.* 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.* **9**, 1328–1336 (2017).

7. Waller, C. C. & McLeod, M. D. A review of designer anabolic steroids in equine sports. *Drug Test. Anal.* vol. 9 1304–1319 (2017).

8. Higgins, A. J. From ancient Greece to modern Athens: 3000 years of doping in competition horses. *J. Vet. Pharmacol. Ther.* **29**, 4–8 (2006).

9. Tozaki, T. & Hamilton, N. A. Control of gene doping in human and horse sports. *Gene Ther.* https://doi.org/10.1038/s41434-021-00267-5 (2021) doi:10.1038/s41434-021-00267-5.

10. Wilkin, T., Baoutina, A. & Hamilton, N. Equine performance genes and the future of doping in horseracing. *Drug Test. Anal.* **9**, 1456–1471 (2017).

11. Nebbia, C. *et al.* Novel strategies for tracing the exposure of meat cattle to illegal growth-promoters. *Vet. J.* **189**, 34–42 (2011).

12. Cloteau, C. *et al.* From a non-targeted metabolomics approach to a targeted biomarkers strategy to highlight testosterone abuse in equine. Illustration of a methodological transfer between platforms and laboratories. *Drug Test. Anal.* https://doi.org/10.1002/dta.3221 (2022) doi:10.1002/dta.3221.

13. Wudy, S. A., Schuler, G., Sánchez-Guijo, A. & Hartmann, M. F. The art of measuring steroids: Principles and practice of current hormonal steroid analysis. *J. Steroid Biochem. Mol. Biol.* **179**, 88–103 (2018).

14. Kirchmeier, A., van Herwaarden, A. E., van der Kolk, J. H., Sauer, F. J. & Gerber, V. Plasma steroid profiles before and after ACTH stimulation test in healthy horses. *Domest. Anim. Endocrinol.* **72**, 106419 (2020).

15. Reichel, C. OMICS-strategies and methods in the fight against doping. *Forensic Sci. Int.* **213**, 20–34 (2011).

16. Knych, H. K. D., McKemie, D. S. & Stanley, S. D. Molecular Cloning, Expression, and Initial Characterization of Members of the CYP3A Family in Horses. *Drug Metab. Dispos.* **38**, 1820–1827 (2010).

17. Yuan, M., Breitkopf, S. B. & Asara, J. M. Serial-omics characterization of equine urine. *PLoS One* **12**, e0186258 (2017).

18. Fragkaki, A. G., Kioukia-Fougia, N., Kiousi, P., Kioussi, M. & Tsivou, M. Challenges in detecting substances for equine anti-doping. *Drug Test. Anal.* **9**, 1291–1303 (2017).

19. Taylor, P., Scarth, J. P. & Hillyer, L. L. Use of in vitro technologies to study phase II conjugation in equine sports drug surveillance. *Bioanalysis* **2**, 1971–1988 (2010).

20. Pranata, A. *et al.* Synthesis of steroid bisglucuronide and sulfate glucuronide reference materials: Unearthing neglected treasures of steroid metabolism. *Steroids* **143**, 25–40 (2019).

21. Scarth, J. P., Teale, P. & Kuuranne, T. Drug metabolism in the horse: A review. *Drug Test. Anal.* **3**, 19–53 (2011).

22. Waller, C. C., Cawley, A. T., Suann, C. J., Ma, P. & McLeod, M. D. In vivo and in vitro metabolism of the designer anabolic steroid furazadrol in thoroughbred racehorses. *J. Pharm. Biomed. Anal.* **124**, 198–206 (2016).

23. Kiousi, P., Fragkaki, A. G., Kioukia-Fougia, N. & Angelis, Y. S. Liquid chromatography–mass spectrometry behavior of Girard's reagent T derivatives of oxosteroid intact phase II metabolites for doping control purposes. *Drug Test. Anal.* **13**, 1822–1834 (2021).

24. Piper, T., Schänzer, W. & Thevis, M. Revisiting the metabolism of 19-nortestosterone using isotope ratio and high resolution/high accuracy mass spectrometry. *J. Steroid Biochem. Mol. Biol.* **162**, 80–91 (2016).

25. Palermo, A., Botrè, F., de la Torre, X. & Zamboni, N. Non-targeted LC-MS based metabolomics analysis of the urinary steroidal profile. *Anal. Chim. Acta* **964**, 112–122 (2017).

26. Boccard, J. *et al.* A steroidomic approach for biomarkers discovery in doping control. *Forensic Sci. Int.* **213**, 85–94 (2011).

27. World Anti-Doping Agency. *WADA Technical Document – TD2021IDCR - Minimum Criteria For Chromatographic-Mass Spectrometric Confirmation Of The Identity Of Analytes For Doping Control Purposes*. https://www.wada-ama.org/sites/default/files/resources/files/td2021idcr_final_eng_0.pdf (2020).

28. Association of Offical Racing Chemist. AORC MS Criteria (modified 23 Aug 16) - Association of Official Racing Chemists. http://www.aorc-online.org/documents/aorc-ms-criteria-modified-23-aug-16/.

29. Horwitz, W. & Albert, R. The Horwitz ratio (HorRat): A useful index of method performance with respect to precision. *J. AOAC Int.* **89**, 1095–1109 (2006).

30. Kihel, L. El. Oxidative metabolism of dehydroepiandrosterone (DHEA) and biologically active oxygenated metabolites of DHEA and epiandrosterone (EpiA) - Recent reports. *Steroids* **77**, 10–26 (2012).

31. Hennebert, O. *et al.* Epimerase activity of the human 11β-hydroxysteroid dehydrogenase type 1 on 7-hydroxylated C19-steroids. *J. Steroid Biochem. Mol. Biol.* **114**, 57–63 (2009).

32. Fitzgerald, C. C. J. & McLeod, M. D. Synthesis of stable isotope labelled steroid bis(sulfate) conjugates and their behaviour in collision induced dissociation experiments. *Org. Biomol. Chem.* **20**, 3311–3322 (2022).

33. Pang, Z. *et al.* MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, 388–396 (2021).

34. Broadhurst, D. *et al.* Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **14**, 72 (2018).

35. Hackstadt, A. J. & Hess, A. M. Filtering for increased power for microarray data analysis. *BMC Bioinformatics* **10**, 1–12 (2009).

36. Worley, B. & Powers, R. PCA as a Practical Indicator of OPLS-DA Model Reliability. *Curr. Metabolomics* **4**, 97–103 (2016).

37. Thévenot, E. A., Roux, A., Xu, Y., Ezan, E. & Junot, C. Analysis of the Human Adult Urinary

Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. *J. Proteome Res.* **14**, 3322–3335 (2015).

38. Triba, M. N. *et al.* PLS/OPLS models in metabolomics: The impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol. Biosyst.* **11**, 13–19 (2015).

39. Szymańska, E., Saccenti, E., Smilde, A. K. & Westerhuis, J. A. Double-check: Validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **8**, 3–16 (2012).

40. Olesti, E., Boccard, J., Visconti, G., González-Ruiz, V. & Rudaz, S. From a single steroid to the steroidome: Trends and analytical challenges. *J. Steroid Biochem. Mol. Biol.* **206**, 105797 (2021).

41. Chemical Entities of Biological Interest (ChEBI). https://www.ebi.ac.uk/chebi/init.do.

42. Kind, T. *et al.* Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom. Rev.* **37**, 513–532 (2018).

43. Dehennin, L., Ferry, M., Lafarge, P., Pérès, G. & Lafarge, J. P. 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 (1998).

44. Hisahiko, S. 19-Hydroxyandrostenedione and 6β-hydroxyandrostenedione: New steroids regulated by the renin-angiotensin system in man. *J. Steroid Biochem.* **20**, 383–385 (1984).

45. Dumasia, M. C. & Houghton, E. Studies related to the metabolism of anabolic steroids in the horse: The identification of some 16-oxygenated metabolites of testosterone and a study of the phase II metabolism. *Xenobiotica* **11**, 323–331 (1981).

46. Hungerford, N. L., McKinney, A. R., Stenhouse, A. M. & McLeod, M. D. Selective manipulation of steroid hydroxyl groups with boronate esters: efficient access to antigenic C-3 linked steroid–protein conjugates and steroid sulfate standards for drug detection. *Org. Biomol. Chem.* **4**, 3951–3959 (2006).

47. McKinney, A. R., Suann, C. J. & Stenhouse, A. M. Detection of 17α-hydroxyprogesterone caproate in equine plasma by gas chromatography/tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **20**, 1855–1858 (2006).

48. Chavatte, P., Holtan, D., Ousey, J. C. & Rossdale, P. D. Biosynthesis and possible biological roles of progestagens during equine pregnancy and in the newborn foal. *Equine Vet. J. Suppl.* **29**, 89–95 (1997).

49. Ghayee, H. K. & Auchus, R. J. Basic concepts and recent developments in human steroid hormone biosynthesis. *Rev. Endocr. Metab. Disord.* **8**, 289–300 (2007).

50. Short, R. V. delta5 3beta-Hydroxy steroids in the follicular fluid of the mare. *J. Endocrinol.* **23**, 277–283 (1961).

51. Cameron, E. H. & Grant, J. K. Biochemistry of histologically defined zones in the adrenal cortex: cortisol synthesis in the horse. *J. Endocrinol.* **37**, 413–420 (1967).

52. El-Sheikh Ali, H. *et al.* Steroid synthesis and metabolism in the equine placenta during placentitis. *Reproduction* **159**, 289–302 (2020).

53. Mostaghel, E. A. Beyond T and DHT - Novel steroid derivatives capable of wild type androgen receptor activation. *Int. J. Biol. Sci.* **10**, 602–613 (2014).

54. Schiffer, L. *et al.* Human steroid biosynthesis, metabolism and excretion are differentially reflected by serum and urine steroid metabolomes: A comprehensive review. *J. Steroid Biochem. Mol. Biol.* **194**, 105439 (2019).

55. Storbeck, K.-H. *et al.* Steroid Metabolome Analysis in Disorders of Adrenal Steroid Biosynthesis and Metabolism. *Endocr. Rev.* **40**, 1605–1625 (2019).

56. Pasin, D., Cawley, A., Bidny, S. & Fu, S. Current applications of high-resolution mass spectrometry for the analysis of new psychoactive substances: a critical review. *Anal. Bioanal. Chem.* **409**, 5821–5836 (2017).

57. Skinnider, M. A. *et al.* A deep generative model enables automated structure elucidation of novel psychoactive substances. *Nat. Mach. Intell.* **3**, 973–984 (2021).

58. Kaufmann, A. High-resolution mass spectrometry for bioanalytical applications: Is this the new gold standard? *J. Mass Spectrom.* **55**, e4533 (2020).

59. Dervilly-Pinel, G., Chereau, S., Cesbron, N., Monteau, F. & Le Bizec, B. LC-HRMS based metabolomics screening model to detect various β-agonists treatments in bovines. *Metabolomics* **11**, 403–411 (2015).

60. Mueller, J. W., Gilligan, L. C., Idkowiak, J., Arlt, W. & Foster, P. A. The regulation of steroid action by sulfation and desulfation. *Endocr. Rev.* **36**, 526–563 (2015).

61. Lievertz, R. W. Pharmacology and pharmacokinetics of estrogens. *Am. J. Obstet. Gynecol.* **156**, 1289–1293 (1987).

62. Attygalle, A. B., García-Rubio, S., Ta, J. & Meinwald, J. Collisionally-induced dissociation mass spectra of organic sulfate anions. *J. Chem. Soc. Perkin Trans. 2* **4**, 498–506 (2001).

63. Rannulu, N. S. & Cole, R. B. Novel fragmentation pathways of anionic adducts of steroids formed by electrospray anion attachment involving regioselective attachment, regiospecific decompositions, charge-induced pathways, and ion-dipole complex intermediates. *J. Am. Soc. Mass Spectrom.* **23**, 1558–1568 (2012).

64. Cawley, A. *et al.* The potential for complementary targeted/non-targeted screening of novel psychoactive substances in equine urine using liquid chromatography-high resolution accurate mass spectrometry. *Anal. Methods* **8**, 1789–1797 (2016).

65. Catalogue, M. Estrone 3-sulfate =98 1240-04-6. https://www.sigmaaldrich.com/AU/en/product/sigma/e9145.

66. Wishart, D. S. *et al.* HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **37**, 603–610 (2008).

67. Benedetto, A., Pezzolato, M., Biasibetti, E. & Bozzetta, E. Omics applications in the fight against abuse of anabolic substances in cattle: challenges, perspectives and opportunities. *Curr. Opin. Food Sci.* **40**, 112–120 (2021).

68. Tsugawa, H. *et al.* MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **12**, 523–526 (2015).

69. Cajka, T. & Fiehn, O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal. Chem.* **88**, 524–545 (2016).

70. Beynon, J. H. Recommendations for symbolism and nomen clature for mass spectroscopy: International union of pure and applied chemistry physical chemistry division commission on Molecular Structure and Spectroscopy. *Pure Appl. Chem.* **50**, 65–74 (1978).

71. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, 47–48 (2015).

72. Rivera, C. & Rodriguez, R. Horwitz equation as quality benchmark in ISO/IEC 17025 testing laboratory. in *IIE Annual Conference and Expo 2010 Proceedings* (2010).

73. Food and Drug Administration. *Guidance for Industry Bioanalytical Method Validation Guidance for Industry Bioanalytical Method Validation*. *Biopharmaceutics Federal register* http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htmand/orhttp://www.fda.gov/AnimalVeterinary/GuidanceComplianceEnforcement/GuidanceforIndustry/default.htm (2018).

74. Waller, C. C. & McLeod, M. D. A simple method for the small scale synthesis and solid-phase extraction purification of steroid sulfates. *Steroids* **92**, 74–80 (2014).

CHAPTER THREE

Stable Isotope Labelled Steroid Bis(sulfate) Conjugates

Publication:

Reproduced from Fitzgerald, C. C. J. & McLeod, M. D. Synthesis of stable isotope labelled steroid bis(sulfate) conjugates and their behaviour in collision induced dissociation experiments. *Org. Biomol. Chem.* **20**, 3311–3322 (2022). With permission from the Royal Society of Chemistry, CCC request ID 600076072.

3.1 Preface

The following manuscript has been published in the journal of Organic and Biomolecular Chemistry. This work describes the developments of new isotope labelling tools to aid the analysis of sulfate-conjugated metabolites. This publication and supporting information (145 pages containing relevant experimental procedures, analysis, characterisation data and copies of NMR and MS spectra) was authored by Mr. Christopher Fitzgerald and Associate Professor Malcolm McLeod. All experimental work, data analysis, and preparation of the original draft and supporting information was done by Mr. Fitzgerald. Associate Professor McLeod initially conceived and provided funding for the project and assisted in the editing and revision of the manuscript. A portion of this work detailing: the synthesis of deuterium labelled estrone sulfates, synthesis of the SIL-labelled mono sulfate library of molecules, and the bis(sulfates), 3β,16α-dihydroxy-5α-androstan-17-one 3,16[$^{18}O_3$]-bis(sulfate), estradiol [$^{18}O_3$]-bis(sulfate), 5α-androstane-3β,17β[$^{18}O_3$]-diol bis(sulfate), and 5α-androstane-3α,17β[$^{18}O_3$]-diol bis(sulfate). As well as the initial splitting tree analysis of these bis(sulfates) and rudimentary collision induced dissociation experiments were performed as part of Mr Fitzgerald's Honours thesis titled "Development and Application of Stably Labelled Steroid Derivatives as Internal Standards and Mass Spectrometry Probes" (ANU, 2017). All bis(sulfate) compounds were re-synthesised and re-analysed, as well as the development of the CID-energy resolved analysis methodology in the publication, were all revised and improved during Mr. Fitzgerald's PhD studies. All other work presented in the following manuscript was performed during Mr. Fitzgerald's PhD candidature.

3.1.1 Summary

The study of diverse sulfate conjugates is essential to reveal the complex interplay of synthesis, hydrolysis and transport inherent in sulfation pathways. Recent research has focused on the direct detection of intact Phase II steroid conjugate metabolites using liquid chromatography-tandem mass spectrometry as it provides more information on the metabolite conjugation site and type. This paved the way for the substantiation of a relatively understudied class of doubly conjugated metabolites, examples of which include steroid bis(sulfate), steroid bisglucuronide, and steroid sulfate glucuronide conjugates. To further research into the bisconjugated metabolites, access to stable isotope labelled internal standards is an essential steppingstone. The research presented in the following manuscript and chapter introduces a new type of stable isotope labelling for sulfate conjugates. The method is unique as it introduces the stable isotope on the sulfate conjugate. The stable isotope labelled conjugates are readily accessed in a facile one-pot synthesis, which allows selective incorporation at free hydroxyl groups. These labelled conjugates have already been used in published work as internal standards in medical diagnostic assays. In this manuscript, the synthetic methodology for the labelling process is detailed and selectively incorporated into bisconjugated steroids for the first time and used to study MS fragmentation pathways in unprecedented detail.

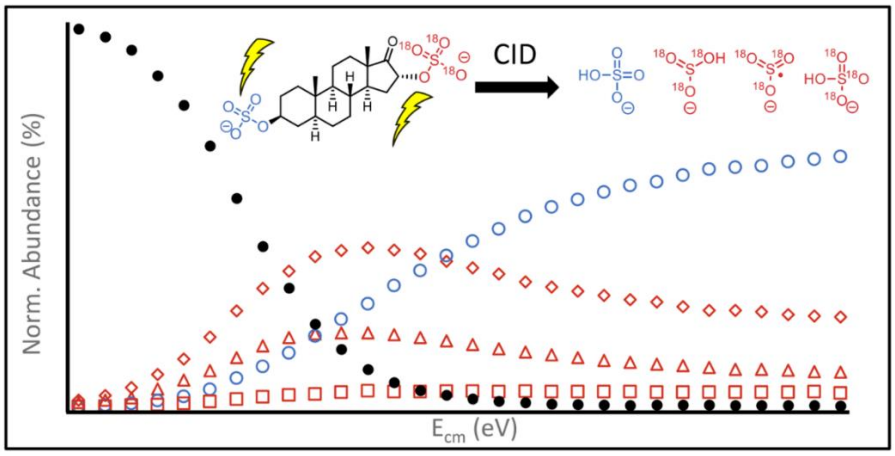


Figure 3.1: Selective incorporation of stable isotope labelled sulfate esters in steroidal systems allows for the investigation of fragmentation patterns using collision induced dissociation MS/MS experiments. Original artwork by Mr. Fitzgerald.

3.2 Synthesis of stable isotope labelled steroid bis(sulfate) conjugates and their behaviour in collision induced dissociation experiments

Organic & Biomolecular Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Org. Biomol. Chem.*, 2022, **20**, 3311

Synthesis of stable isotope labelled steroid bis (sulfate) conjugates and their behaviour in collision induced dissociation experiments†

Christopher C. J. Fitzgerald and Malcolm D. McLeod *

Received 23rd February 2022,
Accepted 25th March 2022

DOI: 10.1039/d2ob00375a

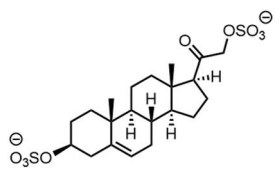
rsc.li/obc

Steroid bis(sulfate) metabolites derived from the two-fold sulfation of unconjugated precursors represent an important yet understudied portion of the steroid profile. The investigation of these compounds in fields such as medicine or anti-doping science relies on mass spectrometry (MS) as the principal tool to identify and quantify biomarkers of interest and depends in turn on access to steroid reference materials and their stable isotope labelled (SIL) derivatives. A new [^{18}O] stable isotope label for sulfate metabolites is reported, which allows for the selective, late-stage and 'one-pot' synthesis of a variety of SIL–steroid conjugates suitable as MS probes and internal standards. The method is applied to more comprehensively study the MS behaviour of steroid bis(sulfate) compounds through collision-induced dissociation (CID) experiments.

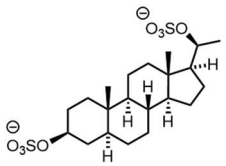
Introduction

The chemistry and biochemistry of the steroid metabolome, including steroid hormones, steroidal drugs and their metabolites have been a rich area of study in medicine and anti-doping science.[1–6] Phase II metabolites, including steroid glucuronides and sulfates, make up a large proportion of the urinary steroid metabolome, with these conjugates accounting for up to 97% of excreted steroid metabolites.[7–9] Therefore, they serve as important markers for upstream metabolic processes.[1,2,9] This is particularly relevant for sulfated metabolites, as steroid sulfates, including dehydroepiandrosterone sulfate (DHEAS) and estrone sulfate can serve as an endogenous depot for circulating steroid hormones in the body.[2,4–6]

Research has increasingly focused on the direct detection of intact phase II steroid metabolites using liquid chromatography-tandem mass spectrometry (LC-MS/MS) techniques, as it circumvents the time consuming deconjugation and derivatisation steps typically associated with gas chromatography-mass spectrometry (GC-MS) analysis.[10] Moreover, direct analysis provides information about conjugation patterns and levels that may provide additional insights into metabolic pathways.[11,12,13,14,16] Intact phase II metabolites are easily observed using LC-MS/MS technologies due to the ready ionisation of the conjugates under electrospray ionisation (ESI) conditions. In the case of sulfate conjugates, collision-induced dissociation (CID) shows distinctive sulfate-derived fragments in the product ion spectra. These are typically observed as the dominant transitions and include the product ions hydrogen sulfate (HSO_4^-, *m*/*z* 97), sulfate radical anion ($SO_4^{\bullet-}$, *m*/*z* 96), hydrogen sulfite (HSO_3^-, *m*/*z* 81), sulfur trioxide radical anion ($^{\bullet}SO_3^-$, *m*/*z* 80), or neutral losses of either sulfur trioxide (SO_3, 80 Da) and sulfuric acid (H_2SO_4, 98 Da).[15–23] This unique fragmentation behaviour in CID has been applied in studies characterising steroid monosulfates.[19,21,22,24–26] In contrast, the closely related steroid bis(sulfate) conjugates (Fig. 1), also



3β,21-dihydroxypregn-5-en-20-one bis(sulfate)



5α-pregnane-3β,20*S*-diol bis(sulfate)

Fig. 1 Examples of steroid bis(sulfate) metabolite biomarkers for STSD 3β,21-dihydroxypregn-5-en-20-one bis(sulfate), and PORD 5α-pregnane-3β,20*S*-diol bis(sulfate).[16,21]

Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia. E-mail: malcolm.mcleod@anu.edu.au

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2ob00375a

known as di-sulfates, have been largely understudied as a class of phase II metabolites in the broader steroid-omics literature, since their discovery in the 1960s in urine and bile.[27–30] This can be attributed to their status as minor metabolites in the overall steroid profile, but also a lack of reference materials and selective MS methods for their detection. Notable contributions have been made in the fields of endocrinology and medical diagnosis.[17,30–36]

Recently, robust synthetic methods to access bisconjugated reference materials have been described, including the synthesis of steroid bis(sulfate), bisglucuronide and mixed sulfate–glucuronide conjugates.[17,21,37] The availability of synthetically pure reference materials have enabled the study of these metabolite classes in greater detail for the first time, reigniting interest in the area.[32,33] For bis(sulfate) metabolites, the synthesis of 23 reference materials was described and used in the development of an LC-MS/MS constant ion loss (CIL) scan method for their detection in a urine matrix.[21] These methodologies were applied in a subsequent study, in which maternal urinary bis(sulfate) metabolites were identified as novel biomarkers in the prenatal diagnosis of inborn errors of steroid biosynthesis associated with Steroid Sulfatase Deficiency (STSD), Smith-Lemli-Opitz Syndrome (SLOS), and cytochrome P450 Oxido-Reductase Deficiency (PORD), (Fig. 1).[16]

The fragmentation behaviour of steroid bis(sulfate) compounds is significantly more complicated than that of their monosulfate counterparts, with product ion spectra showing characteristic features, such as the presence of monosulfate fragments containing the steroid backbone. However, in many cases no distinction can be drawn between which of the two sulfate groups present in the bis(sulfate) conjugate had undergone fragmentation. In other instances fragmentation modes were predicted for compounds such as estradiol bis(sulfate), but experimental verification for this behaviour could not be obtained.[19,21] Herein, a new method for the selective incorporation of a stable isotope label into sulfate esters is reported. From this, a range of selectively labelled steroid bis(sulfate) compounds were prepared for use as internal standards[16] or MS probes. The selectively labelled compounds have been applied to examine MS fragmentation behaviour of steroid bis (sulfate) conjugates under CID conditions. The application of energy resolved CID experiments to examine the dissociation threshold energies of sulfate fragmentation for steroid bis (sulfate) compounds is also described.

Results

Synthesis of stable isotope labelled (SIL) steroid sulfates

Sulfate conjugation reactions were achieved by adapting a synthetic method first described for the preparation of bisphenol bis(sulfate) compounds.[38] This used a one-pot mixture of sulfuric acid, acetic anhydride, and the hydroxylated substrate in pyridine. High conversions could be achieved in the synthesis of steroid sulfate compounds by first stirring the sulfuric acid and acetic anhydride in pyridine followed by the later addition of the steroid alcohol. This allowed for the *in situ* formation of sulfur trioxide–pyridine complex.

By using enriched concentrated sulfuric acid ($H_2S[^{18}O_4]$, 95 atom% in $H_2[^{18}O]$, 95 atom%), selective incorporation of an SIL–sulfate conjugate group could be achieved. This method was tested against a range of different steroidal and other compounds with varying stereochemistries and degrees of unsaturation at the hydroxyl conjugation site, resulting in a series of SIL (**1–16**, Fig. 2) and non-labelled sulfate conjugates (**17–23**, Fig. S1†). Purification was achieved through established solid phase extraction (SPE) methods developed in earlier studies, affording in the majority of cases the SIL–steroid sulfate as the ammonium salt on a 5–10 mg scale.[17,37] High conversions (>98%) were observed for label incorporation across all saturated compounds (Fig. 2), as assessed by a SPE method (Section S3†) involving elution of a combined fraction containing both free and sulfated steroids, followed by ^{1}H NMR integration of a suitable signal (typically C3-H or C17-H) from both the starting steroid and steroid sulfate product. Lower conversions were found among aromatic compounds, such as, estrone sulfate (**6**, 70%), karapinchamine A sulfate[39] (**7**, 96%) and *N*-Boc-tyrosine methyl ester sulfate[40] (**8**, 89%). This may be explained by reduced nucleophilicity of phenolic hydroxyl groups.[37] The cost of enriched sulfuric acid for a preparation of a SIL steroid sulfate on 5–10 mg scale was approximately USD$25.

Selectively labelled bis(sulfated) compounds (**9–16**) were accessed through a stepwise strategy, similar to approaches used in earlier work to make selectively labelled bisglucuronide and sulfate–glucuronide compounds.[17] Examples of this strategy show late and early incorporation of the SIL sulfate group, respectively (Scheme 1). Here, the two compounds 5α-androstane-3β,17β-diol 3,17[$^{18}O_3$]-bis(sulfate) (**11**) and 5α-androstane-3β,17β-diol 3[$^{18}O_3$]-sulfate 17-glucuronide (**24**) were synthesised from the common steroid epiandrosterone. Sulfation of epiandrosterone afforded epiandrosterone sulfate (**22**) which was then reduced to the steroid diol monosulfate (**25**). Late-stage introduction of the labelled sulfate then afforded selectively labelled steroid bis(sulfate) (**11**, Table S1†). Alternatively, sulfation with labelled sulfuric acid afforded epiandrosterone 3[$^{18}O_3$]-sulfate (**1**, Table S1†) which was then reduced to the SIL–steroid diol monosulfate (**26**). Glucuronylation then afforded the final steroid sulfate glucuronide (**24**).[17,41] This selective labelling approach could also be used to introduce the SIL labelled conjugate in structurally more complex compounds, such as the 3β,16α-dihydroxy-5α-androstan-17-one 3,16[$^{18}O_3$]-bis(sulfate) (**9**), in the final step of the in the synthetic pathway (Scheme S1†).

The ^{1}H NMR spectra were identical for the SIL labelled compounds to their unlabelled counterparts as the ^{18}O label is not spin active (I = 0). Instead, label incorporation was assessed using LC-MS on a sample set of SIL-conjugates. Incorporation of 72–85% tri labelled ($^{18}O_3$), 15–28% di labelled ($^{18}O_2$), <1% mono-labelled ($^{18}O_1$), no unlabelled material was observed in all steroid monosulfate and bis(sulfate) com-

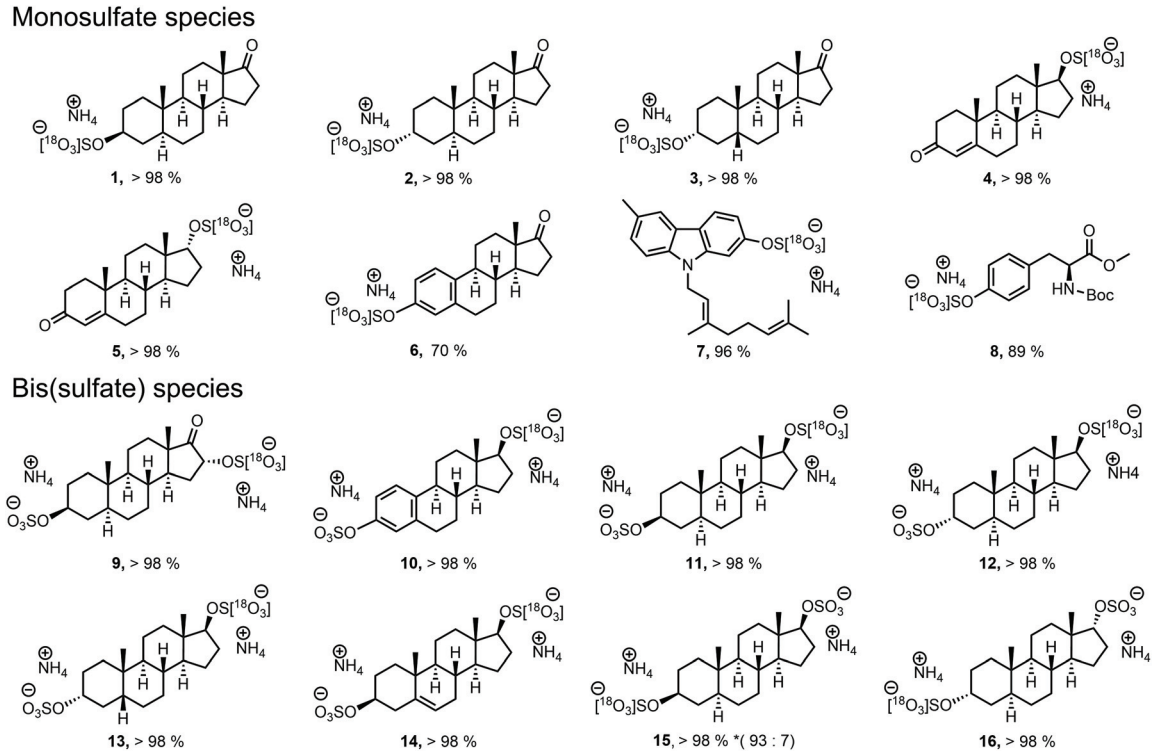


Fig. 2 Substrate scope of SIL incorporation. Percentage conversion (%) for the introduction of label as determined by ^{1}H NMR spectroscopy. [a] Reduction of dihydrotestosterone 17-sulfate to prepare compound (**15**) provided a 93 : 7 ratio of 3β : 3α diastereomers.

pounds tested (Table S1†). This aligned with the theoretical extent of labelling expected for 95 atom% sulfuric acid used in this study, with 85.7% tri-labelled ($^{18}O_3$), 13.5% di-labelled ($^{18}O_2$), 0.7% mono-labelled ($^{18}O_1$) and 0.0% of unlabelled compound.

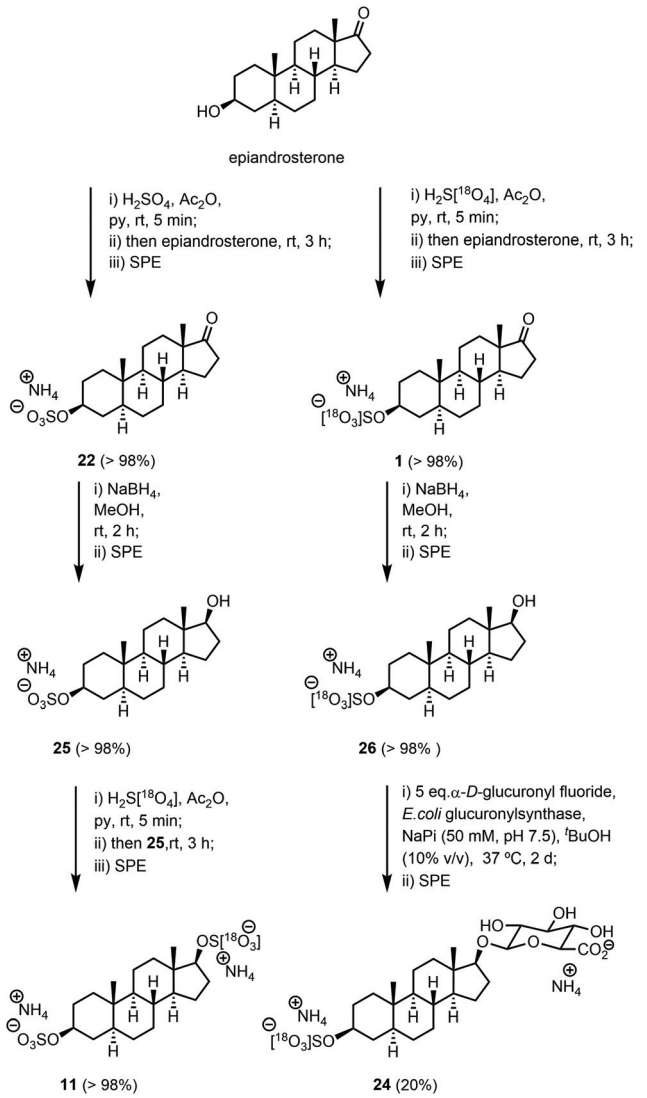
MS analysis of unlabelled and labelled steroid sulfates

Ionisation and fragmentation of steroid monosulfates. In the scan MS with a cone voltage of 26 V, SIL steroid monosulfates (**1–6**) preferentially ionised to the mono-anion ($[M - NH_4]^-$) under negative ESI conditions. Under single energy CID experiments the major product ion observed across species (**1–5**) was the corresponding SIL hydrogen sulfate (HSO$[^{18}O_3]^-$, *m*/*z* 103). The aromatic SIL–estrone sulfate (**6**) showed two product ions corresponding to the SIL sulfur trioxide radical anion ($^{\bullet}S[^{18}O_3]^-$, *m*/*z* 80), and estrone phenolate anion ($[M - S[^{18}O_3] - NH_4]^-$, *m*/*z* 269). This ionisation and fragmentation was consistent with the behaviour of unlabelled steroid monosulfates in previous studies.[17,21,37]

Ionisation and fragmentation of steroid bis(sulfate) compounds. Under the same conditions as the SIL–steroid monosulfates the SIL–steroid bis(sulfate) conjugates (**9–16**) preferentially ionised to the di-anion ($[M - 2NH_4]^{2-}$). Once again this behaviour was consistent with previous studies of unlabelled bis(sulfate) compounds.[19,21] However, under single energy CID experiments the selective placement of the SIL–sulfate conjugate provided unambiguous information on the fragmentation at each end of the steroid bis(sulfate) to generate distinguishable product ions, which in the majority of cases (compounds **11–16**) included two mono-anions ($[M - 2NH_4 - HSO_4]^-$ and $[M - 2NH_4 - HSO[^{18}O_3]]^-$) and two hydrogen sulfate ions (HSO$[^{18}O_3]^-$ and HSO_4^-). Only two of the SIL–bis(sulfate) compounds showed different fragmentation behaviour. These were the 3β,16α-dihydroxy-5α-androstan-17-one 3,16[$^{18}O_3$]-bis(sulfate) (**9**) and estradiol 3,17[$^{18}O_3$]-bis(sulfate) (**10**), which had eight and five distinct product ions, respectively. This delineation of product ions has not previously been observable with steroid bis(sulfate) compounds due to the presence of two indistinguishable sulfate esters in unlabelled compounds.[19,21]

Pseudo-MS^3 experiments of bis(sulfate) compounds. To further examine the fragmentation behaviour, four SIL–bis(sulfate) conjugates (**9–12**) were examined using "pseudo-MS/MS/MS" (pseudo-MS^3) experiments. In these experiments the cone voltage was increased to promote in-source fragmentation of the di-anions during ESI. The in-source fragments corresponding to the mono-anion product ions (such as $[M - 2NH_4 - HSO_4]^-$ and $[M - 2NH_4 - HSO[^{18}O_3]]^-$) identified above were selected in the first quadrupole (Q1). This was followed by CID (Q2) giving further fragmentation. Finally, the product ion spectra were acquired by scanning the third quadrupole (Q3). This allowed a further tracing of fragmentation pathways, and the construction of CID fragmentation trees for each of four steroid bis(sulfate) compounds.

The aliphatic bis(sulfate) compounds **11** and **12** were chosen as they represented the general behaviour observed for the compounds **11–16** in CID experiments. The splitting trees



Scheme 1 Synthesis of the SIL–steroid bis-conjugates (**11** and **24**) from epiandrosterone. Both pathways demonstrate the stepwise synthesis that allowed for the selective incorporation of the SIL label at an early (**24**) or late (**11**) stage. Percentage conversion (% conv.) was determined by ^{1}H NMR spectroscopy.

for compounds **11** and **12** showed two branches of fragmentation over two levels corresponding to the sequential loss of both sulfate units (Fig. 3 & S2†). The first level of MS/MS fragmentation produced the hydrogen sulfate ions (HSO_4^-, *m/z* 97 and $HSO[^{18}O_3]$, *m/z* 103) from a *syn*-elimination process in similar relative abundance, together with the corresponding monosulfate ions ($[M - 2NH_4 - HSO_4]^-$, *m/z* 359 and $[M - 2NH_4 - HSO[^{18}O_3]]^-$, *m/z* 353 respectively). The second level of fragmentation showed two fragmentation pathways. The major involving formation of hydrogen sulfate (HSO_4^-, *m/z* 97 or $HSO[^{18}O_3]$, *m/z* 103), and the minor involving formation of sulfur trioxide radical anion ($^{\bullet}SO_3^-$, *m/z* 80 or $^{\bullet}S[^{18}O_3]^-$, *m/z* 86), from the fragmentation of hydrogen sulfate at higher collision energies.

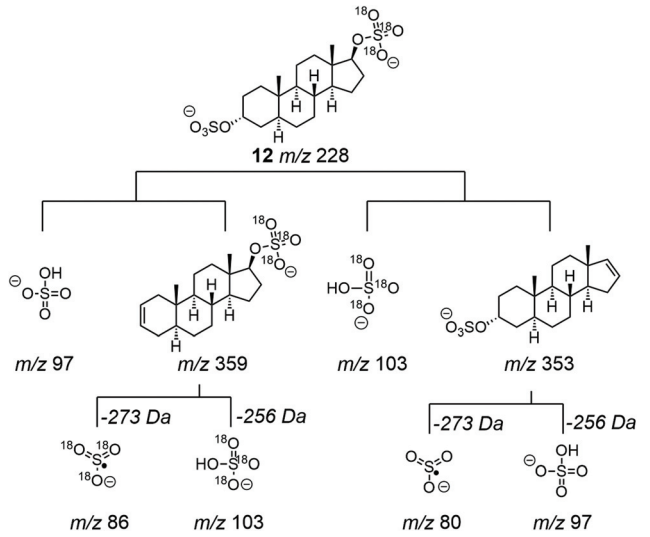


Fig. 3 Proposed CID fragmentation tree for 5α-androstane-3α,17β-diol 3,17$[^{18}O_3]$-bis(sulfate) (**12**) based on MS/MS and pseudo-MS^3 experiments.

The 3β,16α-dihydroxy-5α-androstan-17-one 3,16$[^{18}O_3]$-bis (sulfate) (**9**) displayed more complex fragmentation, with four pathways and eight distinct product ions on the first, and ten product ions on the second level (Fig. 4). Specifically, the first level consisted of steroid monosulfate ions together with their corresponding sulfate-derived fragments. These were sulfur trioxide radical anion ($^{\bullet}S[^{18}O_3]^-$, *m/z* 86), hydrogen sulfite ($HS[^{18}O_3]^-$, *m/z* 87) and hydrogen sulfate ($HSO[^{18}O_3]^-$, *m/z* 103), originating from the D-ring SIL sulfate, and hydrogen sulfate (HSO_4^-, *m/z* 97) from the A-ring sulfate. The second level of fragmentation consisted of product ions derived from the steroid monosulfate fragments generated in the first level, as seen for the fragmentation of compounds **11** and **12**.

For compound **9**, the formation of hydrogen sulfate ($HSO[^{18}O_3]^-$, *m/z* 103) originating from the D-ring 16-sulfate has not been noted previously for these 16-hydroxyandrostan-17-one 16-sulfate species, as it was previously indistinguishable from A-ring fragmentation processes in unlabelled derivatives.[21] From these observations a feasible fragmentation pathway is shown in Fig. 5, which accounts for the three sulfate-derived product ions originating from the 16-sulfate group. Moreover, these observations confirm the proposed mechanism put forward in a previous study, however, suggest that enolisation likely proceeds the formation of the radical anion, *m/z* 384, followed by hydrogen atom abstraction (Fig. 5), or loss of a methyl radical to give the enone, *m/z* 369 (Fig. 4).[21] No further fragmentation could be performed on the fragment *m/z* 369 due to low ion abundance. This complex fragmentation resulting in the 'D-ring' product ions may serve as a unique identifier for 16-hydroxyandrostan-17-one 16-sulfate or related species.

Estradiol bis(sulfate) (**10**) showed three fragmentation pathways. This consisted of two product ion pairs, the sulfur trioxide radical anion ($^{\bullet}SO_3^-$, *m/z* 80) and SIL–hydrogen sulfate

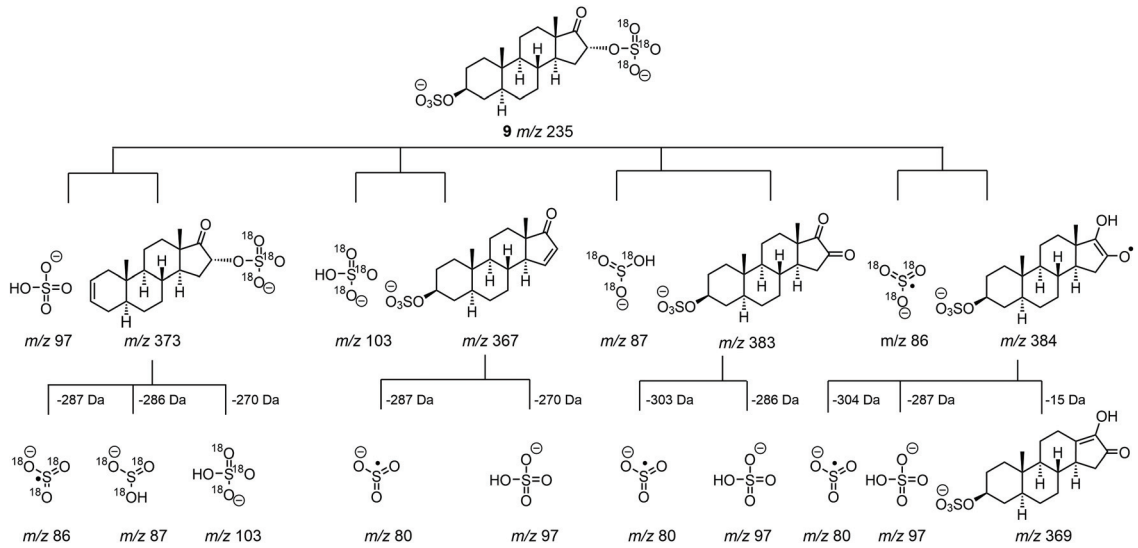


Fig. 4 The CID fragmentation tree for 3β,16α-dihydroxy-5α-androstan-17-one 3,16[$^{18}O_3$]-bis(sulfate) (**9**), based on MS/MS and pseudo-MS^3 experiments.

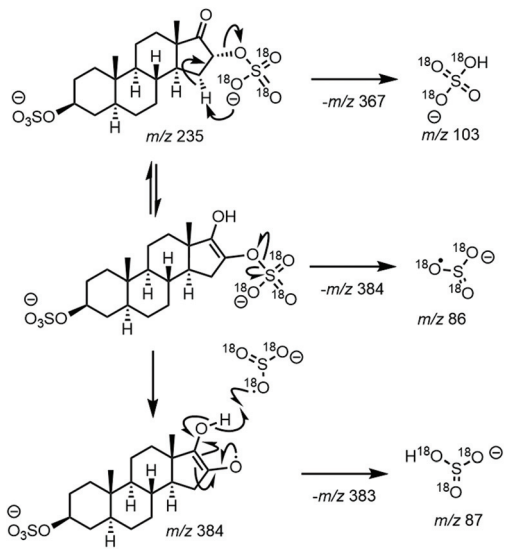


Fig. 5 Competing enol/keto systems leading to the formation of three sulfate-derived product ions from compound (**9**). Formation of *m*/*z* 103 through *syn*-elimination, *m*/*z* 86 from O–S bond homolysis, and *m*/*z* 87 from a subsequent hydrogen atom abstraction.

($HSO[^{18}O_3]^-$, *m*/*z* 103) with their corresponding monosulfate ions ($[M - 2NH_4 - {}^{\bullet}SO_3]^-$, *m*/*z* 356 and $[M - 2NH_4 - HSO[^{18}O_3]]^-$, *m*/*z* 333 respectively), and lastly the di-ion ($[M - 2NH_4 - SO_3]^{2-}$, *m*/*z* 178) derived from neutral loss of sulfur trioxide (Fig. 6). All these primary fragmentations have been noted previously in CID studies of estradiol bis(sulfate).[19] The second level of fragmentation was considerably more complex in nature than previously noted, containing a total of 10 product ions. Interestingly, there were several cases where different fragmentation pathways led to the same fragment ions, including the phenolate anions, *m*/*z* 145 and *m*/*z* 253.

For compound **10**, the results support previous mechanistic proposals that the first two branches ($^{\bullet}SO_3^-$, *m*/*z* 80 and $[M - 2NH_4 - SO_3]^{2-}$, *m*/*z* 178), arise through the homolytic cleavage of the S–O bond of the A ring sulfate ester, with the latter being formed by subsequent electron transfer.[19] Moreover, the phenolate ion, *m*/*z* 145, was observed from either the primary product ions, *m*/*z* 178 or *m*/*z* 333, likely due to a charge-directed cycloreversion. Previously this was also only thought to have originated from the di-anion ($[M - 2NH_4 - SO_3]^{2-}$, *m*/*z* 178). However, these results suggest that it also arises from the primary product mono-anion ($[M - 2NH_4 - HSO[^{18}O_3]]^-$, *m*/*z* 333) but not the phenolic radical ($[M - 2NH_4 - {}^{\bullet}SO_3]^-$, *m*/*z* 356).[19] At higher energies the phenolate anion, *m*/*z* 145, was also observed arising from phenolate anion *m*/*z* 253, derived from monosulfate ion *m*/*z* 333.

Although *m*/*z* 253 is not a primary product ion observed in MS/MS experiments, generation by in-source fragmentation and pseudo-MS^3 identified it as a likely intermediate on a charge-directed cyclo-reversion process (Fig. S3†).[19] Fragmentation of the phenolic radical ($[M - 2NH_4 - {}^{\bullet}SO_3]^-$, *m*/*z* 356) led to formation of the D-ring fragment, *m*/*z* 183 in a previously unappreciated and distinct radical mediated fragmentation process.[19]

As demonstrated the SIL sulfate group has enabled detailed mapping of fragmentation pathways. Uncovering this complex fragmentation behaviour may prove useful in identifying novel bis(sulfate) biomarkers through CID analysis.

Energy resolved fragmentation profiles of steroid sulfates

Energy resolved CID experiments were conducted for the seven non-SIL steroid monosulfate **17–23** (Fig. S1†), and the eight bis (sulfate) **9–16** (Fig. 2) compounds to examine if this group of largely similar compounds could be differentiated by this type of analysis.[42–44]

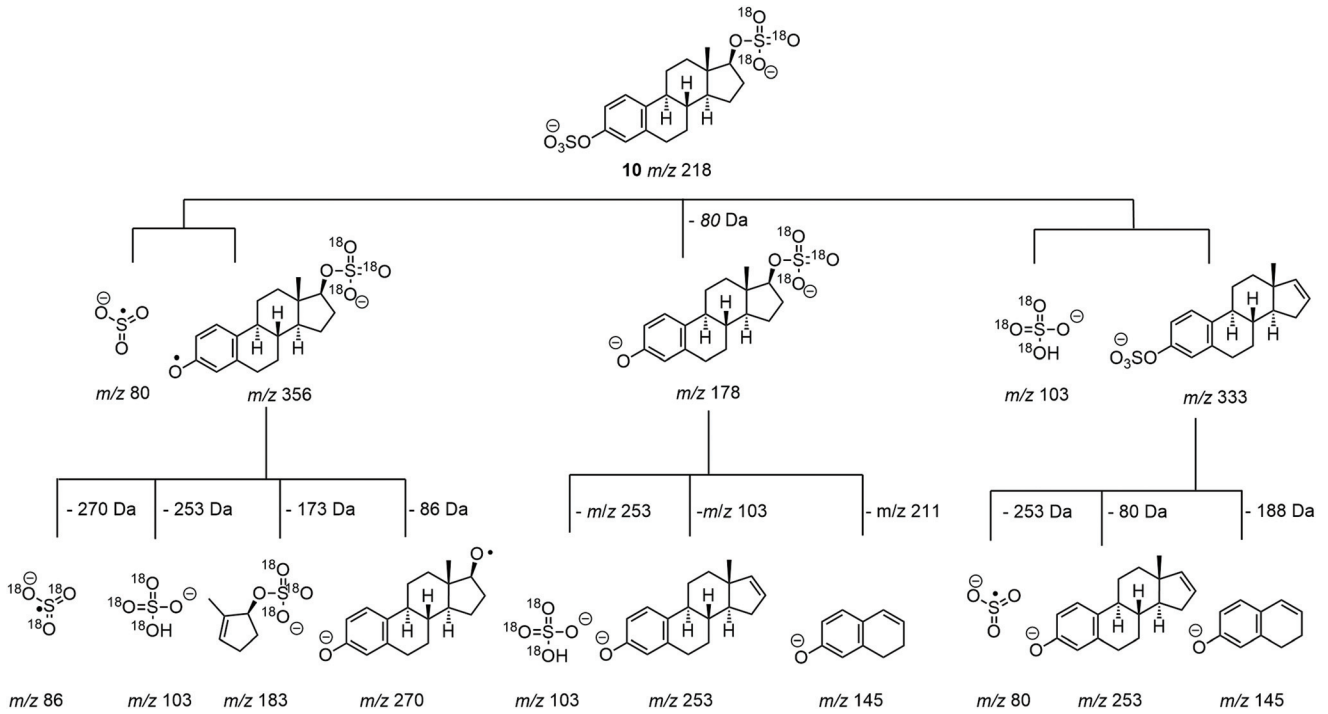


Fig. 6 The CID fragmentation tree for estradiol 3,17[$^{18}O_3$]-bis(sulfate) (**10**) based on MS/MS and pseudo-MS^3 experiments.

Steroid monosulfates. The energy resolved profiles of seven unlabelled steroid monosulfates **17–23** (Fig. S1†) were obtained and fitted to eqn (1) by least squares methods to calculate the dissociation threshold energy ($E_{5\%}$). An example of this analysis is given for androsterone sulfate (**19**) in the ESI† (Fig. S5). The results showed a clear relationship between the stereochemistry and adjacent unsaturation of the sulfate ester and its $E_{5\%}$ (Fig. 7A & Table S2†). The aromatic estrone sulfate (**17**) possessed significantly lower threshold energy reflecting the ability of this ion to undergo homolytic cleavage to generate radical fragments, including $^{\bullet}SO_3^-$. This was then followed by a sequence of increasing energies for the alicyclic ring systems indicating an increase in energy associated with the *syn* elimination reaction to give hydrogen sulfate (HSO_4^-, *m*/*z* 97). Equatorially oriented sulfates at the 3-position and β-orientated sulfates at the 17-position were found to have higher energies then their respective axial or α-orientated counterparts. An example of this is the β-configured DHT 17-sulfate (**23**) that showed higher threshold energy than the α-configured *epi*-DHT 17-sulfate (**20**). The structural dependence of this energy-resolved CID analysis could feasibly be used to differentiate the sulfation sites and stereochemistries of steroid monosulfates based on fragmentation patterns and threshold energy. This could be useful in analytical settings where only limited amounts of compound are present in biological samples, and that rely on MS methods to characterise the species present.

In contrast to previous CID studies,[19] the energy resolved profile of estrone sulfate (**21**) revealed a third fragmentation pathway, with hydrogen sulfite (HSO_3^-, *m*/*z* 81) also observed at moderate collision energies (9–48 eV, Fig. S4†), in addition to the well-established, phenolate anion (*m*/*z* 269) and sulfur trioxide radical anion ($^{\bullet}SO_3^-$, *m*/*z* 80).[19,22,45,46] The SIL–estrone sulfate (**6**) also showed the formation of hydrogen sulfite as the corresponding isotopically enriched counterpart (HS[$^{18}O_3$]$^-$, *m*/*z* 87, Fig. S5†). Two deuterated derivatives of estrone sulfate, **27** and **28** (Fig. 8A), were synthesised to probe the origin of the hydrogen atom in this fragmentation process.[47,48] These compounds were selected due to the placement of the deuterium relative to the sulfate ester; one having them directly adjacent on the aromatic ring (**27**), and the other at the benzylic positions (**28**) some distance from the sulfate ester. Under CID conditions, compound **27** fragmented to hydrogen sulfite (HSO_3^-, *m*/*z* 81), ruling out the possibility of hydrogen atom abstraction *ortho* to the sulfate ester (Fig. S6†). In the case of compound **28**, deuterium sulfite (DSO_3^-, *m*/*z* 82), was observed as the product ion (Fig. S7†). This indicated that the sulfur trioxide radical anion ($^{\bullet}SO_3^-$, *m*/*z* 80) abstracted a benzylic hydrogen atom and implicates the formation of an ion/neutral complex at moderate collision energies that allows for the sulfur trioxide radical anion to migrate with respect to the steroid radical fragment prior to hydrogen atom abstraction. Similar ion neutral migrations accompanied by proton abstraction have been observed under photoionization conditions in the gas phase for the fragmentation of 3,20-diamino-pregnane cations.[49] It should be noted that the experiments cannot discern the precise location of hydrogen atom abstraction, this being possible at either the C6 or C9 labelled positions. The fragmentation tree for estrone 3-sulfate (**17**) shows up to three fragmentation pathways (Fig. 8B), depending on collision energy. Interestingly, the closely

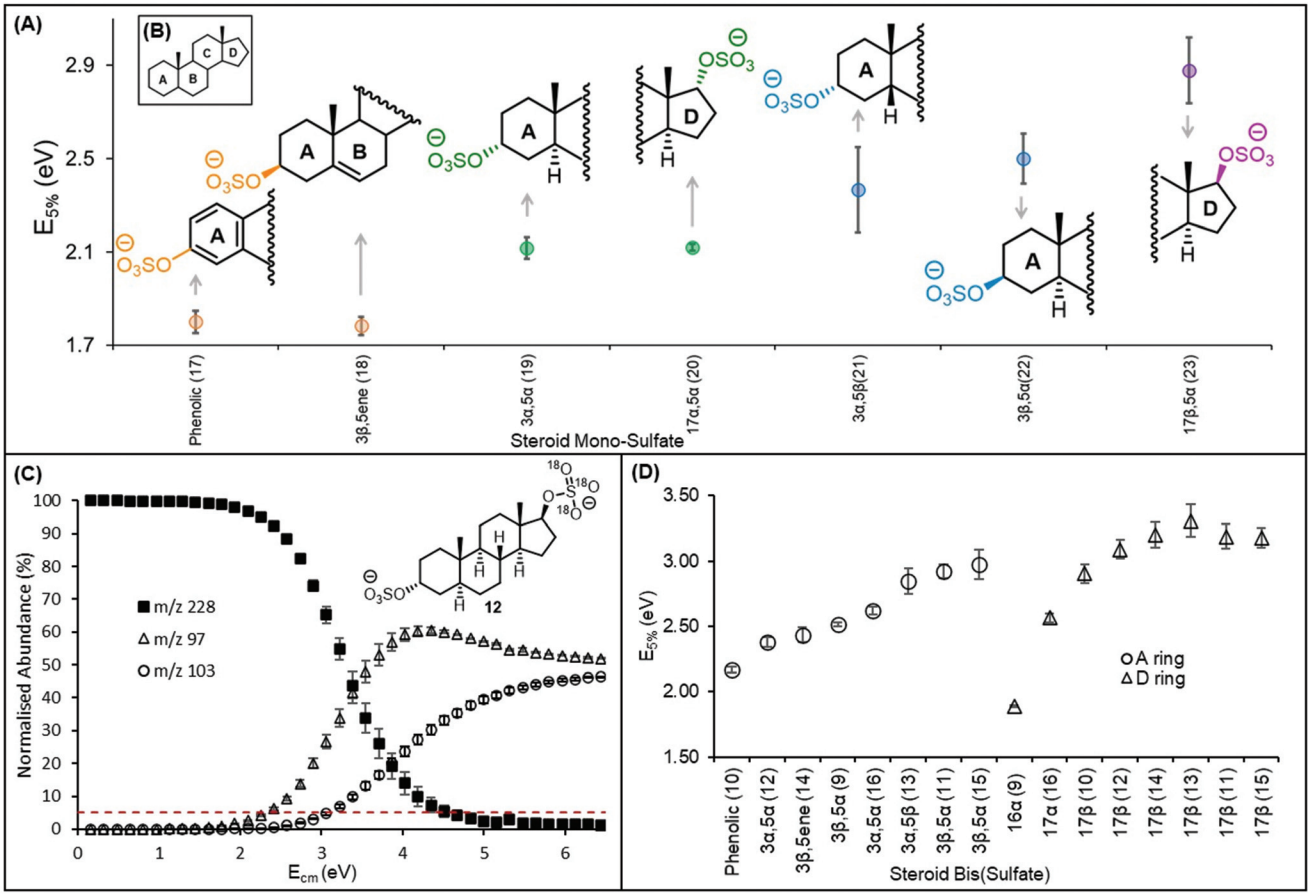


Fig. 7 Energy resolved CID experiments. (A) Dissociation threshold energies ($E_{5\%}$, eV) derived from the energy profiles of steroid monosulfates (**17–23**) with associated sulfate ester bond stereochemistries. (B) General structure of an androstane (C19) steroid backbone consisting of a fused A, B, C & D rings. (C) Example of an energy resolved profile for a steroid bis(sulfate). Normalised abundance of product ions arising from the *syn* elimination of hydrogen sulfate from 5α-androstane-3α,17β-diol 3,17[$^{18}O_3$]-bis(sulfate), ammonium salt (**12**) as a function of applied collision energy in the centre of mass frame (E_{cm}). The $E_{5\%}$ is calculated as the energy where the modelled product ion curves equal 5% normalised abundance for each product ion, indicated by the red line. (D) The mean $E_{5\%}$ (eV) for the fragmentation of steroid bis(sulfate) molecules (**9–16**) under energy resolved CID conditions. Error bars represent standard error of the mean (±SEM, n = 3) for each measured energy (eV).

related estradiol bis(sulfate) (**10**) showed different fragmentation behaviour (Fig. S8†) to its monosulfate counterpart estrone sulfate (**6**) (Fig. S5†). Estradiol bis(sulfate) (**10**) showed major loss of sulfur trioxide radical anion ($^{\bullet}SO_3^-$, m/z 80) with minor neutral loss of sulfur trioxide (SO_3, 80 Da), but did not show the formation of hydrogen sulfite (HSO_3^-, m/z 81). This would be expected given the di-anionic nature of the precursor and the ion–ion repulsion that results between fragment ions.

Steroid bis(sulfate) conjugates. The energy resolved profiles of the eight SIL–steroid bis(sulfate) compounds **9–16** (Fig. 2), showed substantially more complex behaviour than their steroid monosulfate counterparts. This necessitated additional measures during the data analysis of energy resolved experiments. Firstly, only the major sulfate-derived fragment was analysed where product ion pairs were observed, as a lower relative abundance of the corresponding monosulfate anion was typically observed. In addition, all secondary fragments were summed back into their major sulfate pathway, as is standard practice in systems with multiple levels of fragmentation.[42] Lastly, at higher energies, a decrease in the normalised abundance was observed for some product ions in some bis(sulfate) species. This is clearly seen for compound **12** and the hydrogen sulfate (HSO_4^-, m/z 97) product ion (Fig. 7C) and is expected due to further fragmentation of the corresponding monosulfate fragment ion ($[M - 2NH_4 - HSO_4]^-$, m/z 359) at higher energies (Fig. 3 and Fig. S10†). To address this, the analysis of bis(sulfate) compounds was constrained to medium fragmentation energies, around the maximum normalised abundance of each product ion curve, to obtain a better fit of the data to eqn (1) (Fig. S11†). A better fit (Fig. 7C) for the steroid bis(sulfate) (**12**) up to moderate energies (E_{cm} = 4.7 eV) was indicated by a reduction in Chi squared values for, m/z 97, χ^2 = 189 to 7, and m/z 103, χ^2 = 15 to 1, respectively. It should be noted that the absolute ion abundances also reduce as collision energy is increased, also supporting this approach (Fig. S12†). Therefore, the selective introduction of SIL sulfate esters allows for the analysis of trends in fragmentation energies for the bis(sulfate) compounds, but the fitted data may

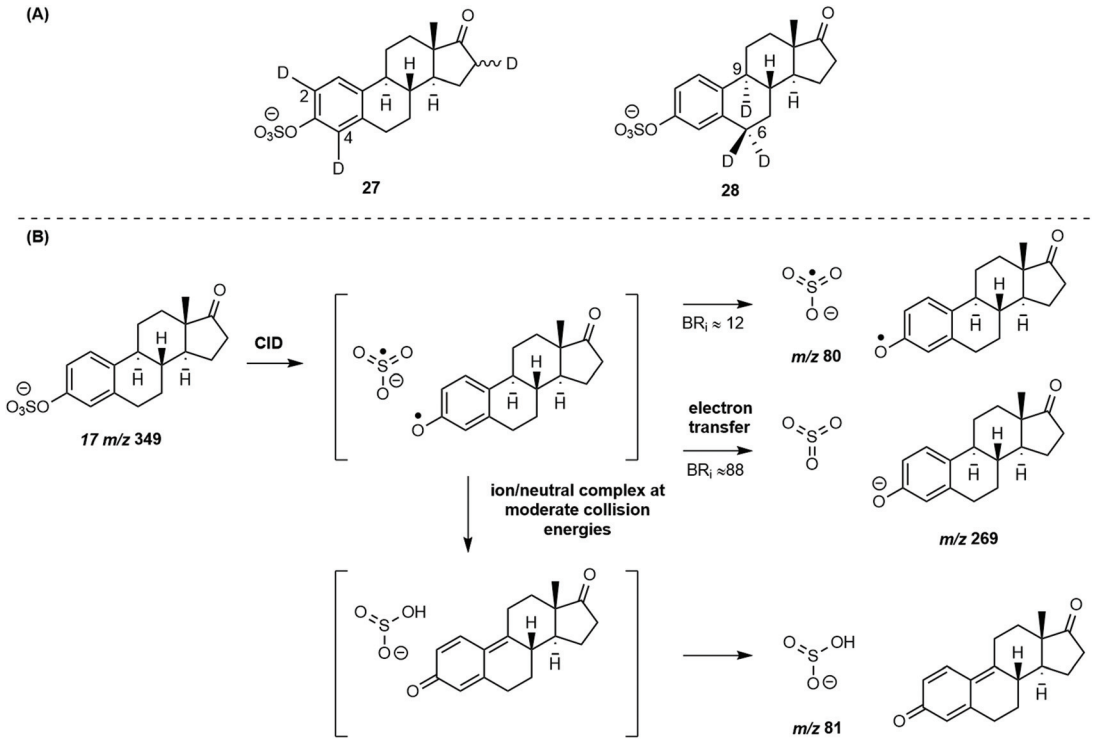


Fig. 8 Evidence for an ion/neutral complex of estrone sulfate (**17**). (A) Structures of selectively deuterated estrone 3-sulfate at the *ortho*-positions (**27**) and benzylic positions (**28**). (B) Proposed fragmentation pathway for estrone 3-sulfate (**17**). The branching ratios (BR$_i$) of both the electron transfer to give *m*/*z* 269, and homolysis to give *m*/*z* 80 are 88 and 12, respectively.

not provide accurate estimates for branching ratios or the higher energy fragmentation pathways. Importantly, no significant difference was observed between the threshold energy of labelled and unlabelled sulfate esters (Tables S3 & S4,† compounds **11** & **15**).

The results of this analysis showed that, like the steroid monosulfates, general trends in fragmentation threshold energy for each steroid bis(sulfate) could be explained based on sulfate ester stereochemistry and the degree of adjacent unsaturation (Fig. 7D). Unsaturated sulfates **9**, **10** and **14** displayed the lowest threshold energies. For saturated sulfates, those with an axial orientation (3α,5α, **12**, **16**) were observed to have lower fragmentation energies than those with an equatorial orientation (3α,5β, **13**; 3β,5α, **11**, **15**). The homoallylic sulfate androst-5-ene-3β,17β-diol 3β,17β[$^{18}O_3$]-bis(sulfate) (3β,5-ene, **14**), with an equatorial 3-sulfate, provided one exception to this trend, but an elimination pathway to generate a conjugated diene product could explained this deviation. The 16α-hydroxyandrostane-17-one (**9**) was also found to have a lower-than-expected fragmentation energy for an A-ring equatorial sulfate, with this likely the effect of the multiple lower energy D-ring fragmentation processes that are available for this molecule (Fig. 4). Finally, within the 5-membered D-ring, 17α oriented sulfates were observed to have lower fragmentation energies than those with a 17β orientation, in the same manner as steroid monosulfates.

The analysis of energy resolved CID profiles of di-anionic steroid bis(sulfate) conjugates could be achieved due to the selective incorporation of a stable isotope label and revealed a clear relationship between the adjacent unsaturation and stereochemistry of the sulfate ester bond and its fragmentation threshold energy. The combination of steroid bis(sulfate) fragmentation pathways and energies provides information on sulfate ester environment and stereochemistry that could be applied in the direct mass spectrometric characterisation of these compounds or closely related derivatives.

Discussion

The synthetic methodology developed herein provides for stable isotope label incorporation for sulfate esters (R-OS[$^{18}O_3$]$^-$). The results show that this label can be effectively introduced into a range of different steroidal and non-steroidal compounds, including selective and late-stage incorporation. The SIL-conjugates produced by this method show high isotopic enrichment, resulting in the incorporation of 72–85% tri labelled ($^{18}O_3$) sulfate, and importantly no non-labelled compound, making them suitable for a range of analytical applications. Recently, two SIL-labelled steroids from this study, 5α-androstane-3β,17β-diol 3,17[$^{18}O_3$]-bis(sulfate) (**11**) and 5α-androstane-3α,17β-diol 3,17[$^{18}O_3$]-bis(sulfate) (**12**), were

employed as SIL-internal standards in analytical methods targeting the prenatal diagnosis of steroid synthesis disorders through the direct LC-MS/MS detection of steroid bis(sulfate) biomarkers in maternal urine.[16]

This new methodology for SIL-incorporation provides several advantages over conventional deuterium (^{2}H) and ^{13}C incorporation.[50–52] One advance lies in the ready access to SIL-labelled standards after a simple one-pot synthesis in a selective and late-stage manner. This is projected to result in significantly shorter synthetic sequences than conventional approaches that require incorporation of the stable isotope onto the steroid backbone prior to conjugation.[50–52] The use of the ^{18}O isotope leads to high isotopic enrichment and mass difference ($\Delta m/z$ +6) to these compounds, giving them substantial m/z separation for MS applications. Stable isotope labels are known to have similar physicochemical properties to their unlabelled counterparts. Specifically, no significant change in chromatographic retention times was observed when compared to unlabelled material, a problem commonly seen using deuterium labelled reference materials.[16,53] Finally, like their ^{13}C labelled counterparts, label incorporation on the sulfate group provides high chemical stability for saturated derivatives including **1–5**, **8**, **11–16**. However, some care is warranted for the preparation of sulfates derived from tertiary of allylic alcohols as these may exhibit lower stability due to hydrolysis of elimination.[21,54–56] Phenolic sulfates may also display lower hydrolytic stability than their saturated counterparts.[57] One limitation of this approach is that labelled derivatives are not suitable for GC-MS methods that involve hydrolysis or elimination of the sulfate ester as part of the analytical process as the label will be lost.[17,58] Even so, this new SIL sulfate ester can readily provide internal standards for the direct detection of steroid sulfate metabolites in LC-MS applications. One avenue for future work could be to generate both deuterium and sulfate labelled compounds to study the interplay of sulfate synthesis, hydrolysis, and transport inherent in sulfation pathways.

This work also shows that the selective placement of an SIL–sulfate can be used to study the different fragmentation behaviours of monosulfate and bis(sulfate) ions using two different CID approaches. Single energy CID experiments showed similar fragmentation among monosulfates to earlier work, with the bis(sulfate) compounds showing more complex behaviour. The combination of selectively SIL–steroid bis(sulfate) compounds and pseudo-MS^3 experiments allowed for the detailed examination of these fragmentation behaviours for the first time. Of interest were estradiol and 3β,16α-dihydroxy-5α-androstan-17-one bis(sulfate) compounds, which have both been noted in earlier studies to possess highly complex fragmentation behaviour.[19,21] The fragmentation of 3β,16α-dihydroxy-5α-androstan-17-one 3,16[$^{18}O_3$]-bis(sulfate) (**9**) showed four main primary product ion pairs, with the observations largely in agreement with an earlier study. However, labelling revealed the existence of a pathway generating hydrogen sulfate (HSO[$^{18}O_3$]$^-$, m/z 103) from the D-ring 16-sulfate.[21] The fragmentations of estradiol bis(sulfate) (**10**) showed the existence of distinct charge- and radical-directed cyclo-reversion products occurring in multiple fragmentation pathways. While these observations are clearly of theoretical interest in the study of gas-phase ion chemistry, the results may also help to identify unknown steroid bis(sulfate) biomarkers that frequently rely solely on the use mass spectrometric methods.

The application of energy resolved CID experiments to steroid monosulfate and bis(sulfate) conjugates is also reported, with distinct relationships observed between the sulfate ester bond position and stereochemistry, and the fragmentation threshold energy ($E_{5\%}$). This approach could distinguish between different isomeric compounds and could feasibly be used for partial structure elucidation in biomarker discovery. The development of a SIL–sulfate conjugate was integral to these studies of fragmentation behaviour and energy for the bisconjugated compounds. Until recently, steroid bis(sulfate) conjugates have been an understudied portion of the steroid profile. The development of an efficient synthetic method to access SIL-labelled sulfates, coupled with the recent progress in MS analytical approaches,[17,21] will serve to advance the study of these intriguing steroid conjugates.

Conclusions

A new method for synthesis of labelled sulfate esters (R-OS[$^{18}O_3$]) is reported, suitable for the preparation of SIL internal standards and MS probes. The SIL sulfate was successfully incorporated into a range of phenol and alcohol derivatives. A library of eight steroid monosulfate, eight steroid bis(sulfate) and one steroid sulfate glucuronide conjugates were prepared, demonstrating the selectively and utility of this approach. The SIL bisconjugates enabled a comprehensive investigation of fragmentation behaviour as a function of the applied energy in CID experiments. This provided for the detailed analysis of fragmentation pathways using splitting trees, revealing the origins of a range of sulfate-derived fragments, including, hydrogen sulfate (HSO_4^-) and sulfur trioxide radical anion ($^{\bullet}SO_3^-$) and sulfur trioxide (SO_3). These fragmentation experiments also revealed the presence of hydrogen sulfite (HSO_3^-) derived from the formation of an ion/neutral complex on fragmentation of estrone sulfate at moderate energies. The labelling approach also exposed clear differences in fragmentation threshold energy ($E_{5\%}$) dependent on the stereochemistry and position of the sulfate ester on the steroid backbone. The information provided by this labelling approach opens avenues for the purely MS/CID-based structure determination of sulfated compounds.

Experimental

Materials and instruments associated with chemical synthesis and characterisation are reported in detail in the ESI† (S6).

General procedures for synthesis of SIL-conjugates

General procedure 1 for small scale steroid sulfation reaction. This was used to conjugate either sulfate or SIL–sulfate to a steroid. The procedure was adapted from the literature.[38] Concentrated sulfuric acid (10.0 μL, 173 μmol), either as an enriched ^{18}O-labelled sulfuric acid (SIL) or non-enriched sulfuric acid, and freshly distilled acetic anhydride (20.0 μL, 212 μmol) were added to a stirring solution of distilled pyridine (200 μL, 2.48 mmol). The resulting solution was then stirred at room temperature for 5 min. The steroid (5.0–10 mg) in distilled pyridine (200 μL, 2.48 mmol) was added, the reaction vessel was then capped and stirred at room temperature for 3 h. The reaction was quenched with water (10 mL) and subjected to SPE according to general procedures 2 and 3.

General procedure 2 for the small-scale steroid conjugate purification by WAX SPE. Weak anion exchange (WAX) solid phase extraction (SPE) was used to separate a steroid conjugate from any unreacted steroid following a reaction. The procedure was adapted from the literature.[17,37,41] A WAX SPE cartridge (6 cc, Waters Oasis®) was preconditioned with methanol (5 mL) followed by water (15 mL), under positive pressure of nitrogen. The conjugate/free steroid mixture (1 mg mL^{-1}, water) was then loaded onto the cartridge, which was then washed with formic acid in water (2% v/v, 15 mL), water (15 mL) and then methanol (15 mL) to elute any free steroid. Finally, the steroid conjugate was eluted using aqueous ammonia solution in methanol (5% v/v, 15 mL). The eluted fraction was concentrated *in vacuo* to yield the steroid sulfate as the corresponding ammonium salt.

General procedure 3 for small scale de-salting of steroid conjugate by C18 SPE. This C18 SPE was used to separate steroid conjugates from excess ammonium salts following a reaction or after extraction with WAX SPE. The procedure was adapted from the literature.[17,37,41] A C18 SPE cartridge (3 cc) was preconditioned with methanol (3 mL) followed by water (3 mL), under positive pressure of nitrogen. The reaction mixture (1 mg mL^{-1}, water) was then loaded onto the cartridge, which was then washed with aqueous ammonia solution (2% v/v, 2 mL), and then water (3 mL). The sulfate conjugate was then eluted with methanol (9 mL) and concentrated *in vacuo* to yield the steroid sulfate as the corresponding ammonium salt.

Epiandrosterone 3[$^{18}O_3$]-sulfate, ammonium salt (1). Epiandrosterone (5.00 mg, 17.2 μmol) was reacted according to general procedure 1 with enriched sulfuric acid (SIL). This gave the title compound as a colourless solid (>98% conversion). ^{1}H NMR (400 MHz, CD_3OD): δ 4.25 (m, 1H, C3-H), 2.43 (dd, J 19.1, 8.8, 1H, C16-H), 2.13–1.90 (m, 3H), 1.86–1.16 (m, 15H), 1.10–0.96 (m, 2H), 0.88 (s, 3H, C18-H_3), 0.87 (s, 3H, C19-H_3), 0.76 (m, 1H); ^{13}C NMR (101 MHz, CD_3OD) δ 224.06 (C17), 79.53 (C3), 55.78, 52.70, 46.28, 38.16, 36.71, 36.65, 36.36, 36.30, 32.80, 32.04, 29.73, 29.58, 22.72, 21.58, 14.20 (C18), 12.60 (C19), one carbon overlapping or obscured; LRMS (−ESI): *m*/*z* 375.3 (100%, $[C_{19}H_{29}[^{18}O_3]O_2S]^-$); HRMS (−ESI) *m*/*z* calcd for $[C_{19}H_{29}[^{18}O_3]O_2S]^-$ ($[M - NH_4]^-$) 375.1863, found 375.1864. The NMR spectra matched the literature for the unlabelled compound.[37]

Estradiol 3,17[$^{18}O_3$]-bis(sulfate), ammonium salt (10). Estradiol 3-sulfate, ammonium salt (5.00 mg, 13.5 μmol) was reacted according to general procedure 1 with enriched sulfuric acid (SIL). This gave the title compound as a colourless solid (>98% conversion). ^{1}H NMR (400 MHz, CD_3OD): δ 7.24 (d, J 8.4, 1H, C1-H), 7.06–6.96 (m, 2H, C2-H & C4-H), 4.32 (m, 1H, C17-H), 2.90–2.81 (m, 2H), 2.35 (m, 1H), 2.29–2.18 (m, 2H), 2.09 (m, 1H), 1.91 (m, 1H), 1.83–1.68 (m, 2H), 1.53–1.20 (m, 6H), 0.85 (s, 3H, C18-H_3); ^{13}C NMR (101 MHz, CD_3OD) δ 151.7, 138.8, 138.1, 127.0, 122.4, 119.7, 88.1, 50.9, 45.5, 44.2, 40.1, 38.0, 30.6, 29.2, 28.3, 27.4, 24.1, 12.1 (C18); LRMS (−ESI): *m*/*z* 218.2 (100%, $[C_{18}H_{22}[^{18}O_3]O_5S_2]^{2-}$); HRMS (−ESI) *m*/*z* calcd for $[C_{18}H_{23}[^{18}O_3]O_5S_2]^-$ ($[M - 2NH_4 + H]^-$) 437.0969, found 437.0962. NMR spectra matched the literature for the unlabelled compound.[21]

Further details on synthesis can be found in the ESI† in Sections S6–S9.

Triple quadrupole MS/MS collision induced dissociation (CID) experiments for the analysis of steroid sulfates

Negative mode tandem mass spectrometry (MS/MS) analysis was undertaken using a Waters 2695 Alliance Separations Module coupled to a Waters triple quadrupole detector. A pure reference material stock solution (1 μg mL^{-1}, MeOH) was directly infused into the ESI source over 1 minute at a flow rate of 15 μL min^{-1}. All MS experiments employed ESI in negative scan MS (*m*/*z* 50–750), or targeted MS/MS modes, monitoring for the mono-anion ($[M - NH_4]^-$) for steroid monosulfates, and the di-anion ($[M - 2NH_4]^{2-}$) for steroid bis(sulfate) compounds. The cone voltage was set to 26 V for scan MS and targeted MS/MS experiments and varied between 50–70 V to promote in-source fragmentation in the pseudo-MS/MS/MS (pseudo-MS^3) experiments. In all collision induced dissociation (CID) experiments (MS/MS and pseudo-MS^3), ions of interest were selected in Q1, subjected to collisions with argon gas in Q2 at a pressure of approximately 3×10^{-4} mbar and gas flow of 0.10 mL min^{-1}, and detected in Q3. Source and de-solvation temperatures were 150 °C and 350 °C, respectively. For energy resolved CID experiments the laboratory frame collision energy (E_{lab}) was varied between 0–80 eV (monosulfate) or 0–40 eV (bis(sulfate), in increments of 2 and 1 eV, respectively.

Energy-resolved fragmentation

The synthetically derived and purified compounds were directly infused onto the mass spectrometer and the energy resolved plots acquired by changing the CID energy in a stepwise fashion followed by scan analysis of product ions. Collision energies were converted to the centre-of-mass frame, $E_{cm} = z(m/(M + m))E_{lab}$, where, *m*, *M*, and *z*, stand for the mass of the collision gas, the mass of the target ion, and the charge of the target ion, respectively. Where possible, the ions derived from secondary fragmentation were summed back to their primary product ion. The normalised abundance of precursor and product ions was obtained from the area of extracted ion chromatograms, with the analysis performed in triplicate.

Breakdown curves were obtained by plotting the normalised abundance (intensity) of the product ion(s) of interest, against the energy in the centre-of-mass frame (E_{cm}), followed by least-squares fitting to the sigmoidal function of the type:

$$I_i(E_{cm}) = \frac{BR_i}{1 + e^{(E_{1/2} - E_{cm})b_i}} \quad (1)$$

where; I_i is the normalised abundance of the product ion of interest, BR_i is the branching ratio of the product ion, b_i describes the rise of the sigmodal curve, and $E_{1/2}$ is the energy at which the function has reached half of its maximum value.[42,59–61] Curve fitting was performed using KaleidaGraph® 4.5 by Synergy software. The dissociation threshold energy ($E_{5\%}$) was defined as the energy (E_{cm}) required to achieve 5% fragmentation. The appearance energy (AE, *X*-axis intercept of the tangent to the curve at $E_{1/2}$) was also investigated as part of this work but was not adopted and is not described in detail, due to its dependence on higher energy regions of the breakdown curve. Details on the derivation of AE and $E_{5\%}$ can be found in the ESI† (S4).

Author contributions

C. C. J. F.: conceptualization, methodology, investigation, formal analysis, data curation, validation, visualization, & writing – original draft. M. D. M.: conceptualization, methodology, funding acquisition, & writing – supervision, review & editing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank Prof. Stephen Blanksby (Queensland University of Technology), Dr Adam Cawley (Racing NSW), Prof. Michelle Coote and Dr Adam Mater (ANU) for helpful and insightful discussions. A special thanks to Dr Hideki Onagi for technical support and advice.

Notes and references

1 W. Schanzer, *Clin. Chem.*, 1996, **42**(7), 1001–1020.
2 J. W. Mueller, L. C. Gilligan, J. Idkowiak, W. Arlt and P. A. Foster, *Endocr. Rev.*, 2015, **36**(5), 526–563.
3 I. Athanasiadou, S. Voss, E. Lyris, A. Aljaber, M. Alsayrafi and C. Georgakopoulos, *Bioanalysis*, 2016, **8**(21), 2265–2279.
4 T. T. Gibbs, S. J. Russek and D. H. Farb, *Pharmacol., Biochem. Behav.*, 2006, **84**(4), 555–567.
5 D. Belelli and J. J. Lambert, *Nat. Rev. Neurosci.*, 2005, **6**(7), 565–575.
6 E. E. Baulieu, P. Robel and M. Schumacher, *Int. Rev. Neurobiol.*, 2001, **46**, 1–32.
7 W. Schanzer and M. Donike, *Anal. Chim. Acta*, 1993, **275**(1–2), 23–48.
8 W. Schänzer and M. Thevis, *Mass Spectrom. Rev.*, 2009, **47**(12), 987–992.
9 R. L. Gomes, W. Meredith, C. E. Snape and M. A. Sephton, *J. Pharm. Biomed. Anal.*, 2009, **49**, 1133–1140.
10 C. Gomez, A. Fabregat, Ó. J. Pozo, J. Marcos, J. Segura and R. Ventura, *TrAC, Trends Anal. Chem.*, 2014, **53**, 106–116.
11 C. E. Galuska, M. F. Hartmann, A. Sánchez-Guijo, K. Bakhaus, J. Geyer, G. Schuler, K.-P. Zimmer and S. A. Wudy, *Analyst*, 2013, **138**, 3702–3801.
12 J. Lu, M. Fernández-Álvarez, S. Yang, G. He, Y. Xu and R. Aguilera, *J. Mass Spectrom.*, 2015, **50**(1), 153–159.
13 C. H. L. Shackleton and K. M. Straub, *Steroids*, 1982, **40**(1), 35–51.
14 Y. Zhao, J. M. Boyd, M. B. Sawyer and X.-F. Li, *Anal. Chim. Acta*, 2014, **806**, 172–179.
15 P. Kiousi, A. G. Fragkaki, N. Kioukia-Fougia and Y. S. Angelis, *Drug Test. Anal.*, 2021, **13**(11–12), 1822–1834.
16 O. J. Pozo, J. Marcos, O. Khymenets, A. Pranata, C. C. Fitzgerald, M. D. McLeod and C. Shackleton, *J. Mol. Endocrinol.*, 2018, **61**(2), M1–M12.
17 A. Pranata, C. C. Fitzgerald, O. Khymenets, E. Westley, N. J. Anderson, P. Ma, O. J. Pozo and M. D. McLeod, *Steroids*, 2019, **143**, 25–40.
18 A. Esquivel, É Alechaga, N. Monfort and R. Ventura, *Drug Test. Anal.*, 2018, **10**(11–12), 1734–1743.
19 A. B. Attygalle, S. García-Rubio, J. Ta and J. Meinwald, *J. Chem. Soc., Perkin Trans. 2*, 2001, 498–506.
20 T. Farrell, L. Poquet, F. Dionisi, D. Barron and G. Williamson, *J. Pharm. Biomed. Anal.*, 2011, **55**(5), 1245–1254.
21 M. D. McLeod, C. C. Waller, A. Esquivel, G. Balcells, R. Ventura, J. Segura and Ó. J. Pozo, *Anal. Chem.*, 2017, **89**(3), 1602–1609.
22 L. Yi, J. Dratter, C. Wang, J. A. Tunge and H. Desaire, *Anal. Bioanal. Chem.*, 2006, **386**(3), 666–674.
23 G. Balcells, X. Matabosch and R. Ventura, *Drug Test. Anal.*, 2017, **9**(7), 1001–1010.
24 A. Lafaye, C. Junot, B. Ramounet-Le Gall, P. Fritsch, E. Ezan and J.-C. Tabet, *J. Mass Spectrom.*, 2004, **39**(6), 655–664.
25 Y. Yao, P. Wang, G. Shao, L. V. A. Del Toro, J. Codero and R. W. Giese, *Rapid Commun. Mass Spectrom.*, 2016, **30**(21), 2341–2350.
26 H. Alquraini and R. J. Auchus, *Mol. Cell. Endocrinol.*, 2018, **464**, 28–33.
27 C. H. L. Shackleton, R. W. Kelly, P. M. Adhikary, C. J. W. Brooks, R. A. Harkness, P. J. Sykes and F. L. Mitchell, *Steroids*, 1968, **12**(6), 705–716.
28 C. H. Shackleton, J. R. Livingstone and F. L. Mitchell, *Steroids*, 1968, **11**, 299–311.
29 T. Laatikainen and R. Vihko, *Steroids*, 1969, **14**(2), 119–131.
30 J. R. Pasqualini and M.-F. Jayle, *J. Clin. Invest.*, 1962, **41**(5), 981–987.

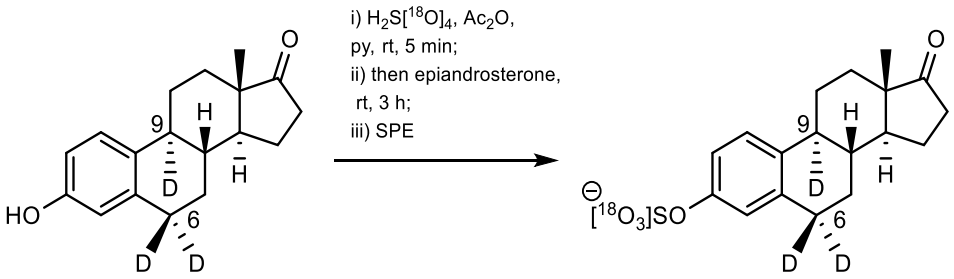
31 C. Shackleton, J. Marcos, W. Arlt and B. P. Hauffa, *Am. J. Med. Genet., Part A*, 2004, **129A**(2), 105–112.
32 L. Schiffer, L. Barnard, E. S. Baranowski, L. C. Gilligan, A. E. Taylor, W. Arlt, C. H. L. Shackleton and K. H. Storbeck, *J. Steroid Biochem. Mol. Biol.*, 2019, **194**, 105439.
33 T. A. Lightning, T. F. Gesteira and J. W. Mueller, *Mol. Cell. Endocrinol.*, 2021, **524**, 111161.
34 K. D. R. Setchell, B. Almé, M. Axelson and J. Sjövall, *J. Steroid Biochem.*, 1976, **7**(8), 615–629.
35 O. Jänne, R. Vihko, J. Sjövall and K. Sjövall, *Clin. Chim. Acta*, 1969, **23**(3), 405–412.
36 L. Kornel and Z. Saito, *J. Steroid Biochem.*, 1975, **6**(8), 1267–1284.
37 C. C. Waller and M. D. McLeod, *Steroids*, 2014, **92**, 74–80.
38 S. Bureeva, J. Andia-Pravdivy, G. Petrov, M. Igumnov, S. Romanov, E. Kolesnikova, A. Kaplun and L. Kozlov, *Bioorg. Med. Chem.*, 2005, **13**(4), 1045–1052.
39 Q. Yan, E. Gin, M. Wasinska-Kalwa, M. G. Banwell and P. D. Carr, *J. Org. Chem.*, 2017, **82**(8), 4148–4159.
40 Y. Vo, B. D. Schwartz, H. Onagi, J. S. Ward, M. G. Gardiner, M. G. Banwell, K. Nelms and L. R. Malins, *Chem. – Eur. J.*, 2021, **27**(38), 9830–9838.
41 P. Ma, N. Kanizaj, S.-A. Chan, D. L. Ollis and M. D. McLeod, *Org. Biomol. Chem.*, 2014, **12**(32), 6208–6214.
42 D. L. Marshall, G. Gryn'ova, M. L. Coote, P. J. Barker and S. J. Blanksby, *Int. J. Mass Spectrom.*, 2015, **378**, 38–47.
43 A. Toyama, H. Nakagawa, K. Matsuda, T.-A. Sato, Y. Nakamura and K. Ueda, *Anal. Chem.*, 2012, **84**(22), 9655–9662.
44 E. L. Zins, C. Pepe and D. Schröder, *J. Mass Spectrom.*, 2010, **45**(11), 1253–1260.
45 H. Zhang and J. Henion, *Anal. Chem.*, 1999, **71**(18), 3955–3964.
46 L. O. G. Weidolf, E. D. Lee and J. D. Henion, *Biomed. Environ. Mass Spectrom.*, 1988, **15**(5), 283–289.
47 P. S. Kiuru and K. Wähälä, *Tetrahedron Lett.*, 2002, **43**(18), 3411–3412.
48 T. Kurita, K. Hattori, S. Seki, T. Mizumoto, F. Aoki, Y. Yamada, K. Ikawa, T. Maegawa, Y. Monguchi and H. Sajiki, *Chem. – Eur. J.*, 2008, **14**(2), 664–673.
49 P. Longevialle, *Mass Spectrom. Rev.*, 1992, **11**(3), 157–192.
50 S. A. Wudy, *Steroids*, 1990, **55**(10), 463–471.
51 C. Bong-Chul, J. P. Mallamo, P. E. Juniewicz and C. H. L. Shackleton, *Steroids*, 1992, **57**(11), 530–536.
52 S. A. Wudy, G. Schuler, A. Sánchez-Guijo and M. F. Hartmann, *J. Steroid Biochem. Mol. Biol.*, 2018.
53 T. Berg and D. H. Strand, *J. Chromatogr. A*, 2011, **1218**(52), 9366–9374.
54 W. Schänzer, G. Opfermann and M. Donike, *Steroids*, 1992, **57**(11), 537–550.
55 H. Bi and R. Massé, *J. Steroid Biochem. Mol. Biol.*, 1992, **42**(5), 533–546.
56 H. Bi, R. Massé and G. Just, *Steroids*, 1992, **57**(7), 306–312.
57 J. L. Kice and J. M. Anderson, *J. Am. Chem. Soc.*, 1966, **88**(22), 5242–5245.
58 L. De Wilde, K. Roels, P. Van Renterghem, P. Van Eenoo and K. Deventer, *J. Chromatogr. A*, 2020, **1624**, 461231.
59 G. Bouchoux and D.-A. Buisson, *Int. J. Mass Spectrom.*, 2006, **249–250**, 412–419.
60 D. Schröder, M. Engeser, H. Schwarz, E. C. E. Rosenthal, J. Döbler and J. Sauer, *Inorg. Chem.*, 2006, **45**(16), 6235–6245.
61 D. Schröder, M. Engeser, M. Brönstrup, C. Daniel, J. Spandl and H. Hartl, *Int. J. Mass Spectrom.*, 2003, **228**(2–3), 743–757.

3.3 Future directions and commentary

There are several avenues for the new type of SIL-conjugates and energy-resolved fragmentation experiments developed in this chapter. The latter of these, as mentioned in the manuscript, has the capacity to offer a wholly MS/CID based approach for the structure determination, including compound constitution and conjugate configuration, of steroid sulfate biomarkers *in vivo.* This in fact is the main idea presented in proceeding Chapter Four. Here we investigate the use of energy-resolved fragmentation to elucidate the stereochemistry at the conjugation site of a novel steroid mono-sulfate that was discovered in Chapter Two. The next chapter is markedly different as it transfers the energy-resolved fragmentation experiments to a UHPLC-HRMS/MS system and is performed at *in vivo* concentrations (ng/mL).

Other avenues could involve the use of the SIL-sulfate conjugates as labels for use in pharmacological and endocrinological studies. A large theme in this thesis has surrounded the discussion of conjugate metabolites as clearance products of steroid metabolism. However, this is only part of the story, as several steroid sulfates are known as active species in various physiological pathways and mechanisms, which are relevant to the fields of endocrinology and pharmacology. Two salient examples are that of DHEAS and sulfated estrogens.[1] The interplay between circulating DHEA and DHEAS is well known in human systems, where DHEAS acts a depot for other androgens such as testosterone and also as an important precursor to adrenal androgens in short and long-term stress responses. This has not been explored as thoroughly in animals such as horses where distinct physiological differences are apparent.[2] The clearance rates of unconjugated and conjugated estrogens are well known in both equine and humans. In equine systems, sulfated estrogens are known to have a lower metabolic clearance rate from the plasma than that of the free steroids. As such, conjugated estrogens, like estrone sulfate presented in this manuscript, act as circulating reservoirs from which biologically active unconjugated forms are continuously being produced for downstream physiological effects.[3,4]

In both cases, it is hypothetically possible to study the interplay between the conjugated and unconjugated form of these steroids. One way to do this would be to combine the SIL-sulfate conjugate with a standard deuterated free SIL steroid. For instance, either of the deuterated estrone sulfate presented in this chapter could be conjugated with an SIL-sulfate group instead of the unlabelled sulfate group that was synthesised, Scheme 3.1. By using such an approach, the conjugation and de-sulfation of these compounds could be studied from the same starting point, as the free steroid would still be traceable via the deuterium label, and in this way events like re-sulfation could also be identified. Of course, one of the main synthetic considerations would be the scaling of the synthetic techniques presented in this chapter to the 100 mg scale that would likely be needed for *in vivo* studies.



Scheme 3.1: The possible synthesis of a doubly SIL-labelled estrone sulfate that could be used to study the circulation in the body.

The applications for the use of these new type of SIL sulfate labels have already been demonstrated as suitable for the use of SIL-internal standards in Chapter Two and in independent work carried out by collaborators.[5] As well as this there is potential that they could be used as SIL-biomarkers to study the interplay of circulating steroids in the body.

3.4 References

1. Mueller, J. W., Gilligan, L. C., Idkowiak, J., Arlt, W. & Foster, P. A. The regulation of steroid action by sulfation and desulfation. *Endocr. Rev.* **36**, 526–563 (2015).

2. Gabai, G., Mongillo, P., Giaretta, E. & Marinelli, L. Do dehydroepiandrosterone (DHEA) and its sulfate (DHEAS) play a role in the stress response in domestic animals? *Front. Vet. Sci.* **7**, 1–15 (2020).

3. Lievertz, R. W. Pharmacology and pharmacokinetics of estrogens. *Am. J. Obstet. Gynecol.* **156**, 1289–1293 (1987).

4. Bhavnani, B. R. & Stanczyk, F. Z. Pharmacology of conjugated equine estrogens: Efficacy, safety and mechanism of action. *J. Steroid Biochem. Mol. Biol.* **142**, 16–29 (2014).

5. Pozo, O. J. *et al.* Alternate steroid sulfation pathways targeted by LC–MS/MS analysis of disulfates: Application to prenatal diagnosis of steroid synthesis disorders. *J. Mol. Endocrinol.* **61**, M1–M12 (2018).

CHAPTER FOUR

Identification of a New Biomarker assisted by Energy-Resolved Collision Induced Dissociation Analysis

Publication:

4.1 Preface

The work presented in this chapter has been published in the Journal of the American Society for Mass Spectrometry. This manuscript and supporting information (60 pages containing experimental procedures, synthesis, characterisation data and copies of NMR and MS spectra) was primarily authored by Mr. Fitzgerald. All synthesis, UHPLC-HRMS/MS assays, sample preparation, data analysis, and writing of the original manuscript was performed by Mr. Fitzgerald. The manuscript also represents a highly collaborative work with multiple authors. The contributions of whom are thankfully acknowledged as;

Ms. Madysen Elbourne performed the UHPLC-MS/MS assay for the chromatographic resolution of androst-5-ene-3α,17β-diol 17-sulfate, compound (**12**).

Mr. Christopher Bowen collaborated in the design of the data processing pipeline using the Shimadzu Scientific Insight software.

Dr. Adam Cawley and Associate Professor Malcolm McLeod were the academic supervisors and provided funding, equipment and assisted in the editing of the manuscript prior to submission.

The supporting information relevant to this chapter is presented electronically alongside this thesis.

Author Contributions in CRedIT format:

CF: Conceptualization, Methodology, Script, Validation, and Writing—Original Draft. CB: Methodology. ME: Methodology: AC: Methodology and Conceptualization. MM: Conceptualization, Methodology, Funding, and Writing—Supervision, review and editing.

4.1.1 Summary

This chapter describes the use of energy-resolved collision induced dissociation (CID) experiments to assist in the structure determination of a new steroid sulfate biomarker that was detected following the administration of the anabolic androgenic steroid testosterone propionate in a thoroughbred horse in Chapter Two of this thesis. This study employs a solely mass spectrometry based method and uses the data derived from energy-resolved fragmentation of the new biomarker, together with data derived from readily available reference materials to significantly narrow the number of potential structures of the unknown metabolite. This allowed for the synthesis of a small but comprehensive group of candidate reference materials. From this selection the identity of the new biomarker was unambiguously confirmed using relevant industry analytical criteria.

The work presented in this chapter is closely related to, and builds on, the methods developed in the preceding two chapters. This is especially the case for the work presented in Chapter Three. In this chapter the mechanisms that drive the differences that are seen between different stereoisomers are likely similar to that underpinned the CID behaviour observed in Chapter Three. The techniques presented in this chapter advances the methodology in two key respects:

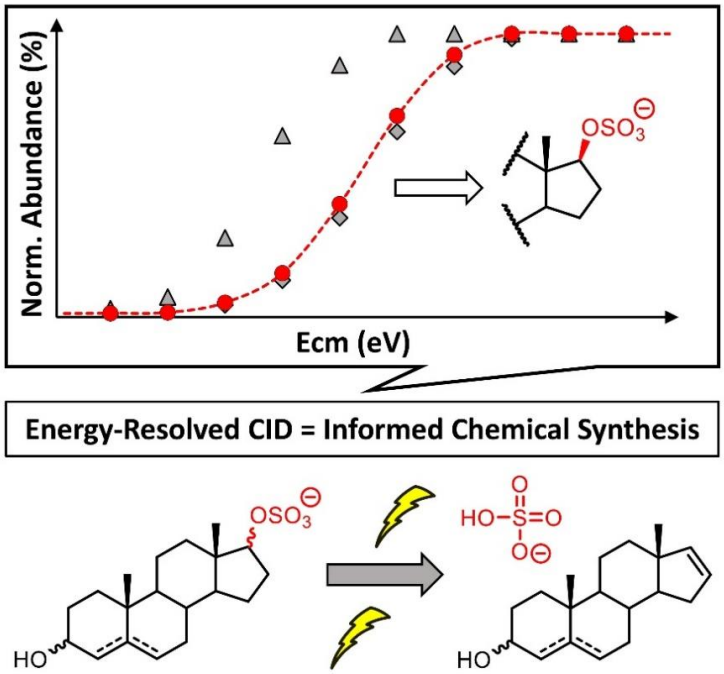


Figure 4.1: Energy resolved CID utilised for structure determination to guide the synthesis of a new steroidal reference material. Original artwork by Mr. Fitzgerald.

Firstly, the study uses UHPLC-HRMS/MS with a QToF mass analyser, to enable the energy resolved CID data to be acquired in a rapid and selective fashion, from samples derived *in vivo*. This was achieved from UHPLC chromatographic peaks derived from single injections for biomarkers present at low concentrations in a complex urinary matrix. This is in contrast to the methodology developed in Chapter Three, which infused relatively high concentrations (approximately 1 mg/mL) of pure synthesised compounds directly into the source of a low-resolution triple quadrupole mass analyser.

Secondly, this study takes the simple yet effective step of comparing the energy-resolved CID data of readily accessible reference materials to the new biomarker studied. This removes the temporal and instrumental variations that were a limiting feature of the previous work.

Lastly, this study introduces a greater efficiency regarding the use of a semi-automated data processing workflow by using a combination of the Insight (Shimadzu) and KaleidaGraph software. In Chapter Three, peak areas and intensities were manually extracted, which could take weeks to months of analysis to finalise. Whereas, in this chapter the data was obtained, analysed, and

processed over a 3-day period, making it a viable and scalable methodology for use in routine applications.

This approach has broader implications than the identification and confirmation of steroid sulfate metabolites and has the potential to streamline the discovery and identification process of a range of new biomarkers arising from metabolomic research.

4.2 Energy-Resolved Fragmentation Aiding the Structure Elucidation of Steroid Biomarkers

ACS Partner Journal

Journal of the American Society for Mass Spectrometry

pubs.acs.org/jasms

Research Article

Energy-Resolved Fragmentation Aiding the Structure Elucidation of Steroid Biomarkers

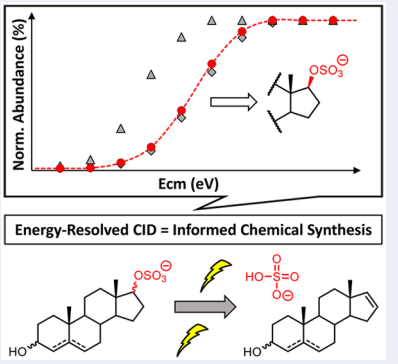
Christopher C. J. Fitzgerald, Christopher Bowen, Madysen Elbourne, Adam Cawley, and Malcolm D. McLeod*

Cite This: *J. Am. Soc. Mass Spectrom.* 2022, 33, 1276−1281

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The identification and confirmation of steroid sulfate metabolites in biological samples are essential to various fields, including anti-doping analysis and clinical sciences. Ultra-high-performance liquid chromatography with tandem mass spectrometry (UHPLC-MS/MS) is the leading method for the detection of intact steroid conjugates in biofluids, but because of the inherent complexity of biological samples and the low concentration of many targets of interest, metabolite identification based solely on mass spectrometry remains a major challenge. The confirmation of new metabolites typically depends on a comparison with synthetically derived reference materials that encompass a range of possible conjugation sites and stereochemistries. Herein, energy-resolved collision-induced dissociation (CID) is used as part of UHPLC-HRMS/MS analysis to distinguish between regio- and stereo-isomeric steroid sulfate compounds. This wholly MS-based approach was employed to guide the synthesis of reference materials to unambiguously confirm the identity of an equine steroid sulfate biomarker of testosterone propionate administration.



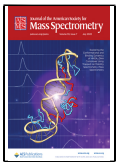
INTRODUCTION

Anabolic androgenic steroids (AAS) are among the most commonly detected group of compounds in world sports.[1] The use of AAS is proscribed by sporting bodies and included in the World Anti-Doping Agency's (WADA) "Prohibited List" and the International Federation of Horseracing Authorities' (IFHA) "International Agreement on Breeding, Racing and Wagering."[2,3] The development of new methods to detect these steroids and their metabolites form an essential and ongoing part of antidoping efforts to maintain fair competition and protect the welfare of participants.

Endogenous and exogenous AAS are mainly excreted in urine as their phase II metabolites.[4] In phase II metabolism, steroids undergo enzymatically driven conjugation, which is dominated by two major classes: sulfates and glucuronides. These conjugates account for up to 97% of excreted steroid metabolites.[5−7] In equine AAS metabolism, a larger proportion of sulfated conjugates has been reported relative to that observed in humans.[8] Intact phase II metabolites play an important role in antidoping analysis as long-term biomarkers. Several studies have demonstrated longer urinary excretion times for sulfate metabolites.[9−15] Common to these studies is the confirmation of new biomarkers using chromatographic retention time and tandem mass spectrometry (MS/MS) experiments. In the case of AAS sulfate conjugates, the spectra are usually dominated by the precursor ion and transitions to sulfate-derived fragments, including hydrogen sulfate, m/z 97, HSO_4^-.[9,16−19] This fragmentation behavior has been a key identifier in various antidoping and clinical studies characterizing steroid monosulfates.[17,20−24]

One limiting factor in identifying new steroid sulfate biomarkers is determining the position and stereochemistry of the sulfate ester. Typically, such studies rely solely on mass spectrometric methods because insufficient material is available to perform more comprehensive nuclear magnetic resonance (NMR) spectroscopic analysis or X-ray crystallography. Confirmation by MS/MS thus requires access to a wide range wide of synthetically derived regio- and stereo-isomeric reference materials that are then compared with the new biomarker using chromatographic and MS/MS analysis. This approach is dependent on highly specialized synthetic skills and is labor intensive and expensive.[18] Alternatively, hydrolysis followed by gas chromatography−mass spectrometry (GC−MS) analysis can be used to identify the unconjugated metabolite by comparison with the wide range of commercially available unconjugated reference materials, but depending on the structure, the original site of conjugation may not be readily ascertained.[16,18,25,26]

Received: March 30, 2022
Revised: May 22, 2022
Accepted: May 24, 2022
Published: June 7, 2022



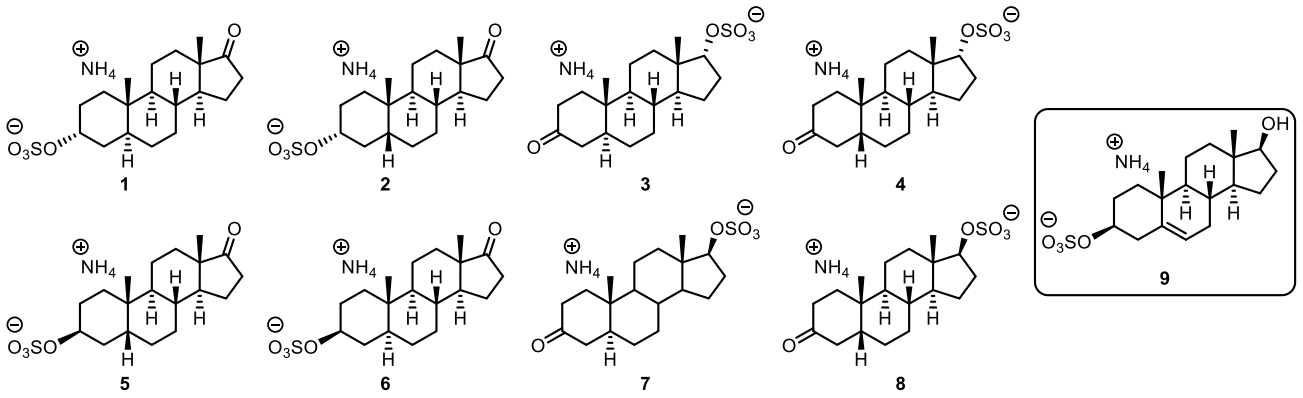


Figure 1. Synthesized library of reference materials (**1**–**9**), including hydroxyandrostanone monosulfates (**1**–**8**) and an androstenediol monosulfate (**9**).

Recently developed MS-based methodologies, such as energy-resolved collision induced dissociation (CID) and ion mobility spectrometry (IMS) experiments, can provide additional information to aid the characterization of steroid biomarkers. In IMS, the use of collision cross section (CCS) values has been demonstrated to be an effective additional molecular descriptor that allows for increased selectivity and improvements in the detection of phase I and intact phase II sulfate and glucuronide conjugates of AAS at biologically relevant concentrations in urine, although the resolution of isomers remains challenging.[27−29] The CCS approach has also been shown to be reproducible across different instrumental platforms, thereby allowing for a consolidation of steroidal CCS data in databases.[29]

Energy-resolved CID experiments, where the fragmentation of precursor ions is studied across a wide range of applied collision energies, has been shown to strengthen structural interpretation in distinguishing both isomeric small molecules and metal oxide clusters on the basis of experimentally derived appearance energies (AE).[30−34] The use of estimated threshold energies ($E_{5\%}$) has also been applied to explore competing modes of bond homolysis in alkoxyamines.[35] Energy-resolved CID has also been used to study the dissociation pathways of the Beta-2 agonist class of performance-enhancing drugs[36] and to aid the structural characterization of *N*-glycans from their fragmentation patterns.[37] Recently, these approaches have been extended to reveal characteristic differences between sets of related steroid monosulfates on the basis of estimated threshold energies ($E_{5\%}$).[38] A common drawback in this energy-resolved CID analysis is the sensitivity of the measured energies to small changes in experimental and instrumental parameters.[39] This variability suggests the need for a reference set of sulfated molecules to improve the reliability of experimentally measured threshold energies when used on different systems.[39]

This work describes the use of energy-resolved CID methods on an ultra-high-performance liquid chromatography with tandem mass spectrometry (UHPLC-MS/MS) instrument equipped with a high-speed quadrupole-time of flight (Q-ToF) mass analyzer to differentiate the sulfate ester conjugation site and stereochemistry for a set of isomeric steroid sulfate compounds. The threshold energies ($E_{5\%}$) derived from single injections were used to aid in the structure elucidation of an unknown equine urinary metabolite found at elevated levels after testosterone propionate administration to a thoroughbred horse. The information derived from this approach helped guide the synthesis of candidate reference materials, thereby leading to the rapid identification and confirmation of a novel metabolite.

METHODS

Experimental Section. The materials and instruments associated with chemical synthesis and characterization are reported in detail in the Supporting Information (Section S2).

UHPLC-HRMS/MS Detection of AAS Intact Phase II Metabolites. Energy-resolved CID experiments employed a 9030 Q-ToF mass spectrometer equipped with an ESI source interfaced to a Nexera LC-40D X3 pump system for chromatographic separation (all from Shimadzu, Rydalmere, Australia). The column used was a Shim-pack Velox C18 column (1.8 μm, 2.1 × 100 mm, Shimadzu). The UHPLC separation was performed at flow rate of 0.4 mL/min using a gradient mixing of two phases: Solution A, 20 mM ammonium formate in 100% water, and solution B, 20 mM ammonium formate in 100% methanol. The gradient was 0−0.5 min (10% B), 0.5−20.5 min (10−100% B), 20.5−23.0 min (100% B), 23.0−23.01 min (100−10% B), 23.01−27.5 min (10% B). The injection volume was 5 μL and the column oven temperature was 40 °C.

Parameters for Energy-Resolved CID MS/MS Experiments. Steroid sulfate compounds readily ionized as their monoanions [$(M\text{-}NH_4)^-$] under negative mode ESI. Scan MS (m/z 100−1000) measurements were collected between 0.5 and 20.5 min. For the duration of this acquisition, energy-resolved CID−MS/MS acquisition (m/z 50−1000) was performed for each reference material targeting the precursor ion m/z 369.1740 with a Q1 width of m/z 1. Energy-resolved CID−MS/MS spectra were acquired over a CE ramp from 10 to 65 eV in increments of 5 eV and immediately after higher energies were sampled at 75 and 80 eV. This gave a loop time (duty cycle) of 0.80 s comprising one scan MS event (0.1 s) and 14 MS/MS events (0.05 s) to give a total scan speed of approximately 19 Hz. The resolution was approximately 30 000 (fwhm) for scan MS and MS/MS experiments. The peak areas for the precursor (m/z 369.1740) and the sulfate-derived product ions (m/z 96.9596 and 79.9568) were calculated from extracted ion chromatograms with a tolerance of ±10 ppm using LabSolutions Insight software (Shimadzu).

Calculation of Estimated Threshold Energy ($E_{5\%}$). Breakdown curves were obtained by plotting the normalized abundance (peak area) of the product ion(s) of interest against

https://doi.org/10.1021/jasms.2c00092
J. Am. Soc. Mass Spectrom. 2022, 33, 1276−1281

the energy in the center-of-mass frame (E_{cm}), followed by a least-squares fitting to the sigmoidal function of the type:

$$I_i(E_{cm}) = \frac{BR_i}{1 + e^{(E_{1/2} - E_{cm})b_i}} \quad (1)$$

where I_i is the normalized abundance of the product ion of interest, BR_i is the branching ratio of the product ion, b_i describes the rise of the sigmodal curve, and $E_{1/2}$ is the energy at which the function has reached half of its maximum value.[33−35,40] Curve fitting was performed using KaleidaGraph 4.5 by Synergy Software. Where possible, the ions derived from secondary fragmentation were summed back to their primary product ion. The estimated threshold energies ($E_{5\%}$) were derived from the energy (E_{cm}) when I_i = 5, i.e. the calculated energy (E_{cm}) required to achieve 5% fragmentation, and were calculated as previously described.[35,38]

RESULTS AND DISCUSSION

A Novel Metabolite of Testosterone Propionate in the Equine. In a recent study aimed at the untargeted profiling of equine urinary sulfate metabolites after testosterone propionate administration,[41] two putative steroid sulfate metabolites with the same exact mass (m/z 369.1739) were discovered. The identity of one of these metabolites was confirmed as epiandrosterone sulfate (**6**) after matching it to synthetically derived reference materials (**1**−**8**, Figure 1). The reference materials possessed all possible regio- and stereoisomeric combinations at the C3, C5, and C17 positions along a general C_{19} hydroxyandrostanone backbone. However, no match was found for the second metabolite, which showed a significant elevation (500-fold change) 12 h after drug administration. The new biomarker was observed to have an earlier retention time (7.69 vs 10.16 min for epiandrosterone sulfate) in reverse phase chromatography, leading to the inference that the remaining metabolite possessed higher polarity than epiandrosterone sulfate. Two assumptions were made about the structure of the unknown metabolite: (1) an increase in polarity resulted from an additional hydroxy group on the steroid backbone, generally at the C3 or C17 positions,[8] and (2) the addition of a C═C double bond was required to maintain the correct accurate mass, with this type of functionally commonly occurring at the C4 or C5 positions among androstenones, c.f., testosterone and dehydroepiandrosterone (DHEA).[8] On the basis of these assumptions, 16 possible C_{19} androstenediol monosulfate isomers, stemming

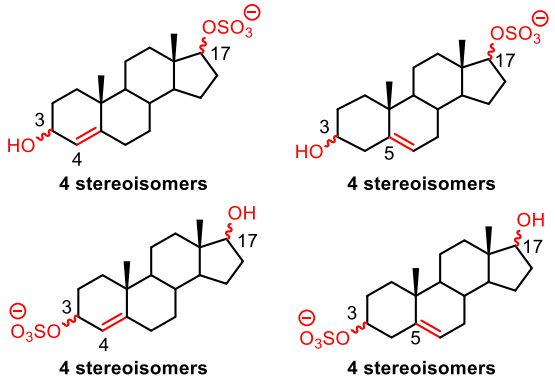


Figure 2. Possible isomers for the unknown metabolite (m/z 369.1741). There are 16 possible isomers for the androstenediol monosulfate class of molecules when the conjugation site is varied between the C3 and C17 positions and the double bond is either in the C4 or C5 positions. All 16 isomers are shown in the Supporting Information, Figure S1.

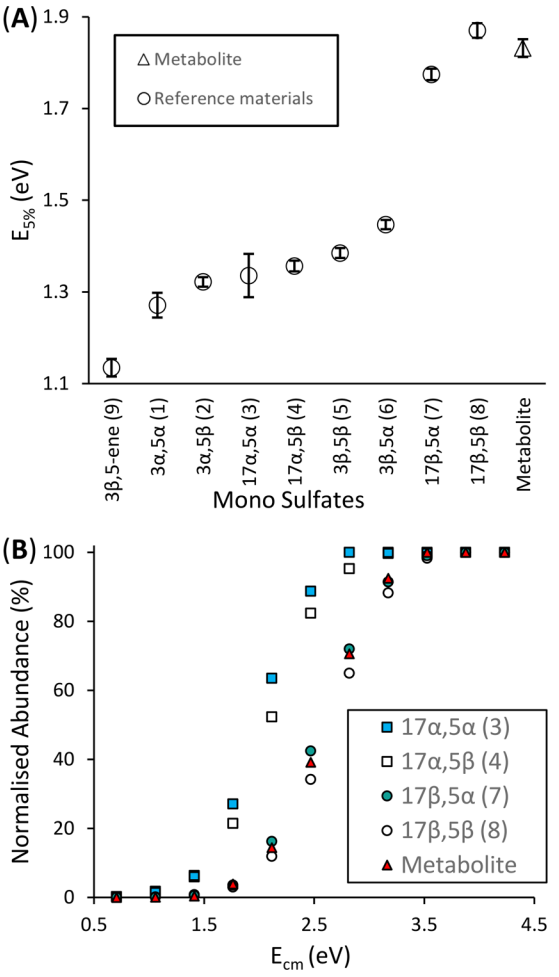


Figure 3. Evaluation of synthesized reference materials and the new biomarker. (A) Mean threshold energies ($E_{5\%}$ ± SD, n = 3) of the steroid monosulfate reference materials (**1**−**9**) and the new biomarker (Metabolite). (B) Breakdown curves for the four 17-conjugated reference materials (**3**, **4**, **7**, and **8**) and the new biomarker (Metabolite). Each reference material is identified by the stereochemistry at the conjugation site, followed by the stereochemistry of the 5-position [i.e., 17β-hydroxy-5α-androstan-3-one, 17-sulfate (**7**) becomes 17β,5α (**7**)].

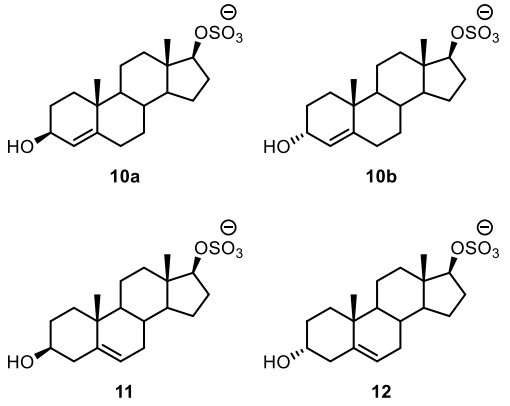


Figure 4. Four synthetically derived 17β-oriented sulfate conjugates.

https://doi.org/10.1021/jasms.2c00092
J. Am. Soc. Mass Spectrom. 2022, 33, 1276−1281

Table 1. Compound Confirmation Using UHPLC Retention Time and MS/MS Analysis

Reference material					Metabolite			
compound name	retention time (RT) (min) (% RSD, n = 12) [theoretical m/z]	product ion (m/z)	RA (%)	threshold energy ($E_{5\%}$, eV) (SD, n = 3)	RT (min) (% RSD, n = 3) [tolerance]	RA (%)	AORC[43] [% tolerance]	threshold energy ($E_{5\%}$, eV) (SD, n = 3)
androst-4-ene-3β,17β-diol, 17-sulfate (**10a**)	9.49 (0.02) [369.1740]	369.1721	38	1.82 (0.01)	9.49 (0.13) [9.45−9.55]	45	[27−50]	1.83 (0.02)
		351.1598	2			2	[0−12]	
		337.1466	39			31	[27−51]	
		96.9588	100			100	[70−100]	
		79.9567	1			2	[0−11]	

The new biomarker (Metabolite) was detected using the sulfate fraction of equine urine after administration of testosterone propionate. It was matched against the synthesized reference material (**10a**) according to the Association of Official Racing Chemists (AORC) retention time and MS/MS criteria.[43] All CID product ion spectra used for confirmation were acquired using a collision energy (CE) of 40 eV derived from the energy-resolved CID experiments. The relative abundance of product ions was derived from the peak area of extracted ion chromatograms with a mass tolerance of ±10 ppm. The threshold energy ($E_{5\%}$) for both compounds was obtained according to the Experimental Section and matched within one standard deviation (n = 3).

from four general core structures, can be considered as potential candidates for the unknown metabolite (Figure 2 and Figure S1).

Determination of Sulfate Conjugation Site and Stereochemistry Using Mass Spectrometry. The general workflow used to acquire an energy-resolved breakdown curve of a sulfate ion is now described. After UHPLC separation and negative mode ESI, the compound undergoes scan MS followed by MS/MS. In MS/MS, the collision energy is applied in a stepped fashion over 12 scans. The high scan speeds available for the Q-TOF mass analyzer allowed for the sampling of product ion data from low to high energies in small energy increments (5 eV, 10−65 eV) and at high resolution (30 000 fwhm). The energy profiles derived from the chromatographic peaks from single injections were then plotted, and the threshold energy was calculated after least-squares fitting of the data to eq 1. The effects of inter-run variability on the measured energy profiles were minimized[39] by running all reference materials and samples in triplicate in the same injection sequence.

The energy-resolved CID profile of the unknown metabolite was compared against the library of synthesized isomeric reference materials (**1**−**8**, Figure 1). These reference materials were chosen because they were readily accessible from their free steroid using previously established synthetic methodology.[42] It was hypothesized that the profiles of the eight hydroxy-androstane sulfate isomers could be used as proxies to identify the stereochemistry at the conjugation site of the new biomarker and ultimately inform the synthesis of the relevant reference material. The energy-resolved CID profile of androst-5-ene-3β,17β-diol 3-sulfate (**9**, boxed Figure 1) was also obtained, as it was readily available from a previous synthesis.[18]

The results of the energy profiling workflow are displayed in Figure 3A and show the estimated threshold energies ($E_{5\%}$) of the nine synthesized isomers (**1**−**9**) and unknown metabolite (Metabolite).

All reference materials were chromatographically resolved (Table S1). Moreover, good separation was observed between biologically relevant epimers, such as androsterone sulfate [**1**, retention time (RT) = 11.86 min], epiandrosterone sulfate (**6**, RT = 11.00 min), and etiocholanolone sulfate (**2**, RT = 11.72 min).[8,13,19,44−47] The new biomarker has a high threshold energy comparable to the 17β-oriented sulfates. The retention time of unsaturated androstenediol monosulfate (**9**) was observed to be closer to the unknown metabolite than the eight hydroxyandrostanone monosulfate compounds (**1**−**8**) because of the increased polarity resulting from the free hydroxy group. The diol monosulfate compound (**9**) was also found to have the lowest estimated threshold energy ($E_{5\%}$) because of a relatively low energy fragmentation leading to a conjugated diene product. The pattern in the $E_{5\%}$ for compounds (**1**−**8**) indicated that sulfate ester stability is largely governed by the stereochemistry at the conjugation site, while the configuration at a greater distance seems to have little effect (Figure 3A). This pattern is clearly demonstrated in the four reference materials that possess 17-conjugation (**3**,**4**,**7**, and **8**; Figure 3B). Here, despite there being large differences in the configuration of the A-ring, the 17β-oriented sulfates (**7** and **8**) show only a small difference in threshold energies. This was also observed for the 17α-oriented sulfates (**3** and **4**). Importantly, a large difference is observed between the 17α and 17β pairs in both their estimated threshold energies (Figure 3A) and energy-resolved CID profiles (Figure 3B). From this analysis we concluded that these functionally simpler hydroxyandrostanone sulfate molecules could be used as proxies for related species so long as the functionality and stereochemistry at the conjugation site are conserved. After evaluation of the new biomarker, the $E_{5\%}$ was found to aligned closely with the two 17β-sulfated reference materials (**7** and **8**, Figure 3). If we assume that the conjugation site and stereochemistry are the main contributors to the observed $E_{5\%}$ energy, we could therefore putatively assign 17β stereochemistry to the sulfate conjugation site of the new biomarker.

Confirmation of a Novel Metabolite. We used the $E_{5\%}$ energies from the energy-resolved CID as a guide to narrow the identity of the new biomarker from 16 posibilities to down to 4 isomers, specifically those that contain a 17β-oriented sulfate ester (Figure 2 and Figure S1). After the synthesis of the four isomers (Scheme S1), the new biomarker was matched to androst-4-ene-3β,17β-diol, 17-sulfate (**10a**, Figure 4). All four isomers were chromatographically resolved (Table S3). The confirmation was performed according to the Association of Official Racing Chemists (AORC) retention time and MS/MS criteria and is summarized in Table 1.[43] Further confirmation was also achieved through standard addition (spiking) of the synthesized reference materials into the urine samples because of the close elution of the **10a** and **11** isomers (Figures S2 and S3). Moreover, compounds **10a** and **11** were also analyzed using energy-resolved CID. The profile and $E_{5\%}$ of compound **10a** (1.82 ± 0.01 eV) also closely match the urinary metabolite (1.83 ± 0.02 eV) within one

https://doi.org/10.1021/jasms.2c00092
J. Am. Soc. Mass Spectrom. 2022, 33, 1276−1281

standard deviation. Feasibly, this could be used as a secondary MS-based method in the future for matching this type of steroidal metabolite (Table 1, Table S2, Figures S4 and S5). The energy profile of **10b** was not obtained because of low precursor abundance.

CONCLUSION

We report a novel application of energy-resolved CID experiments for the structure elucidation of the sulfate ester conjugate site and stereochemistry in steroid sulfate metabolites. The energy profiles could be derived from chromatographic peaks observed in single injections and were found to depend on the location and stereochemistry at the conjugation site, with more distant changes to the steroid skeleton having little effect on estimated threshold energies. The energy-resolved CID was used to guide the synthesis of reference materials for an unknown AAS sulfate urinary biomarker detected following testosterone propionate administration to a thoroughbred horse. This study shows that a library of structurally simple reference materials can be used to aid the identification of the sulfate ester conjugate site and stereochemistry for a new steroid sulfate biomarker. This energy-resolved CID−MS/MS method has wider potential to aid in the structure elucidation of new biomarkers for other metabolite classes where access to suitably diverse collections of reference materials is accessible.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jasms.2c00092.

Supplementary figures, tables, and schemes, and additional methodology; general procedures for synthesis of the steroid monosulfate library and the steroid diol monosulfate reference materials; and spectroscopic data for the synthesized compounds (PDF)

AUTHOR INFORMATION

Corresponding Author

Malcolm D. McLeod − *Research School of Chemistry, Australian National University, Canberra, Australian Capital Territory 2601, Australia;* orcid.org/0000-0002-2343-3226; Email: malcolm.mcleod@anu.edu.au

Authors

Christopher C. J. Fitzgerald − *Research School of Chemistry, Australian National University, Canberra, Australian Capital Territory 2601, Australia*

Christopher Bowen − *Mass Spectrometry Business Unit, Shimadzu Scientific Instruments (Australasia), Rydalmere, New South Wales 2116, Australia*

Madysen Elbourne − *Centre for Forensic Science, University of Technology Sydney, Broadway, New South Wales 2007, Australia;* orcid.org/0000-0002-1994-8577

Adam Cawley − *Australian Racing Forensic Laboratory, Racing NSW, Sydney, New South Wales 2000, Australia*

Complete contact information is available at: https://pubs.acs.org/10.1021/jasms.2c00092

Author Contributions

C.C.J.F.: conceptualization, methodology, script, validation, and writing—original draft. C.B.: methodology. M.E.: methodology. A.C.: methodology and conceptualization. M.D.M.: conceptualization, methodology, funding, and writing—supervision, review and editing. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

This work was presented at the ANZSMS28 conference in November 2021, Canberra, ACT, Australia.

ACKNOWLEDGMENTS

The authors thank Dr. Komba Thomas (ANU) for synthetic guidance and the provision of synthetic material **17**. The authors also offer a special thanks to Professor Stephen Blanksby (QUT) for insightful discussions on energy-resolved methodologies. The authors thank the Australian Research Council Linkage Project Scheme (LP180100421) for financial support.

REFERENCES

(1) Protti, M.; Mandrioli, R.; Mercolini, L. Perspectives and Strategies for Anti-Doping Analysis. *Bioanalysis* **2019**, *11* (3), 149−152.

(2) International Federation of Horse Racing Authorities. *International agreement on breeding, racing and wagering and appendixes.* https://www.ifhaonline.org/Default.asp?section=IABRW&area=15 (accessed on January 6, 2022).

(3) WADA. *The Prohibited List.* World Anti-Doping Agency. https://www.wada-ama.org/en/prohibited-list (accessed on January 6, 2022).

(4) Schanzer, W. Metabolism of Anabolic Androgenic Steroids. *Clin. Chem.* **1996**, *42* (7), 1001−1020.

(5) Schanzer, W.; Donike, M. Metabolism of Anabolic Steroids in Man: Synthesis and Use of Reference Substances for Identification of Anabolic Steroid Metabolites. *Anal. Chim. Acta* **1993**, *275* (1−2), 23−48.

(6) Schänzer, W.; Thevis, M. Human Sports Drug Testing by Mass Spectrometry. *Mass Spectrom. Rev.* **2017**, *36*, 16−46.

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

(8) Scarth, J. P.; Teale, P.; Kuuranne, T. Drug Metabolism in the Horse: A Review. *Drug Test. Anal.* **2011**, *3* (1), 19−53.

(9) Kiousi, P.; Fragkaki, A. G.; Kioukia-Fougia, N.; Angelis, Y. S. Liquid Chromatography-Mass Spectrometry Behavior of Girard's Reagent T Derivatives of Oxosteroid Intact Phase II Metabolites for Doping Control Purposes. *Drug Test. Anal.* **2021**, *13* (11−12), 1822−1834.

(10) Gómez, C.; Pozo, O. J.; Marcos, J.; Segura, J.; Ventura, R. Alternative Long-Term Markers for the Detection of Methyltestosterone Misuse. *Steroids* **2013**, *78* (1), 44−52.

(11) Nair, V. S.; Doman, C. E.; Morrison, M. S.; Miller, G. D.; Husk, J.; Eenoo, P.; Crouch, A. K.; Eichner, D. Evaluation of Epiandrosterone as a Long-term Marker of Testosterone Use. *Drug Test. Anal.* **2020**, *12* (11−12), 1554−1560.

(12) Balcells, G.; Gómez, C.; Garrostas, L.; Pozo, Ó. J.; Ventura, R. Sulfate Metabolites as Alternative Markers for the Detection of 4-Chlorometandienone Misuse in Doping Control. *Drug Test. Anal.* **2017**, *9* (7), 983−993.

(13) Esquivel, A.; Alechaga, É.; Monfort, N.; Ventura, R. Sulfate Metabolites Improve Retrospectivity after Oral Testosterone Administration. *Drug Test. Anal.* **2019**, *11* (3), 392−402.

(14) Schänzer, W.; Guddat, S.; Thomas, A.; Opfermann, G.; Geyer, H.; Thevis, M. Expanding Analytical Possibilities Concerning the Detection of Stanozolol Misuse by Means of High Resolution/High Accuracy Mass Spectrometric Detection of Stanozolol Glucuronides

in Human Sports Drug Testing. *Drug Test. Anal.* **2013**, *5* (11−12), 810−818.
(15) Piper, T.; Schänzer, W.; Thevis, M. Revisiting the Metabolism of 19-Nortestosterone Using Isotope Ratio and High Resolution/High Accuracy Mass Spectrometry. *J. Steroid Biochem. Mol. Biol.* **2016**, *162*, 80−91.
(16) Pozo, O. J.; Marcos, J.; Khymenets, O.; Pranata, A.; Fitzgerald, C. C.; McLeod, M. D.; Shackleton, C. Alternate Steroid Sulfation Pathways Targeted by LC-MS/MS Analysis of Disulfates: Application to Prenatal Diagnosis of Steroid Synthesis Disorders. *J. Mol. Endocrinol.* **2018**, *61* (2), M1−M12.
(17) McLeod, M. D.; Waller, C. C.; Esquivel, A.; Balcells, G.; Ventura, R.; Segura, J.; Pozo, Ó. J. Constant Ion Loss Method for the Untargeted Detection of Bis-Sulfate Metabolites. *Anal. Chem.* **2017**, *89* (3), 1602−1609.
(18) Pranata, A.; Fitzgerald, C. C.; Khymenets, O.; Westley, E.; Anderson, N. J.; Ma, P.; Pozo, O. J.; McLeod, M. D. Synthesis of Steroid Bisglucuronide and Sulfate Glucuronide Reference Materials: Unearthing Neglected Treasures of Steroid Metabolism. *Steroids* **2019**, *143*, 25−40.
(19) Esquivel, A.; Alechaga, É.; Monfort, N.; Ventura, R. Direct Quantitation of Endogenous Steroid Sulfates in Human Urine by Liquid Chromatography-Electrospray Tandem Mass Spectrometry. *Drug Test. Anal.* **2018**, *10* (11−12), 1734−1743.
(20) Lafaye, A.; Junot, C.; Ramounet-Le Gall, B.; Fritsch, P.; Ezan, E.; Tabet, J.-C. 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.* **2004**, *39* (6), 655−664.
(21) Yao, Y.; Wang, P.; Shao, G.; Del Toro, L. V. A.; Codero, J.; Giese, R. W. Nontargeted Analysis of the Urine Nonpolar Sulfateome: A Pathway to the Nonpolar Xenobiotic Exposome. *Rapid Commun. Mass Spectrom.* **2016**, *30* (21), 2341−2350.
(22) Alquraini, H.; Auchus, R. J. Strategies That Athletes Use to Avoid Detection of Androgenic-Anabolic Steroid Doping and Sanctions. *Mol. Cell. Endocrinol.* **2018**, *464*, 28−33.
(23) Yi, L.; Dratter, J.; Wang, C.; Tunge, J. A.; Desaire, H. Identification of Sulfation Sites of Metabolites and Prediction of the Compounds' Biological Effects. *Anal. Bioanal. Chem.* **2006**, *386* (3), 666−674.
(24) Attygalle, A. B.; García-Rubio, S.; Ta, J.; Meinwald, J. Collisionally-Induced Dissociation Mass Spectra of Organic Sulfate Anions. *J. Chem. Soc. Perkin Trans. 2* **2001**, *4*, 498−506.
(25) Marcos, J.; Pozo, O. J. Current LC-MS Methods and Procedures Applied to the Identification of New Steroid Metabolites. *J. Steroid Biochem. Mol. Biol.* **2016**, *162*, 41−56.
(26) Rodríguez-Morató, J.; Pozo, Ó. J.; Marcos, J. Targeting Human Urinary Metabolome by LC-MS/MS: A Review. *Bioanalysis* **2018**, *10* (7), 489−516.
(27) Thevis, M.; Dib, J.; Thomas, A.; Höppner, S.; Lagojda, A.; Kuehne, D.; Sander, M.; Opfermann, G.; Schänzer, W. Complementing the Characterization of in Vivo Generated N-Glucuronic Acid Conjugates of Stanozolol by Collision Cross Section Computation and Analysis. *Drug Test. Anal.* **2015**, *7* (11−12), 1050−1056.
(28) Davis, D. E.; Leaptrot, K. L.; Koomen, D. C.; May, J. C.; Cavalcanti, G. D. A.; Padilha, M. C.; Pereira, H. M. G.; McLean, J. A. Multidimensional Separations of Intact Phase II Steroid Metabolites Utilizing LC-Ion Mobility-HRMS. *Anal. Chem.* **2021**, *93* (31), 10990−10998.
(29) Velosa, D. C.; Rivera, M. E.; Neal, S. P.; Olsen, S. S. H.; Burkus-Matesevac, A.; Chouinard, C. D. Toward Routine Analysis of Anabolic Androgenic Steroids in Urine Using Ion Mobility-Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (1), 54−61.
(30) Bursey, M. M.; Nystrom, J. A.; Hass, J. R. Survey of Reaction Types in Low-Energy Colliosional Activation of Protonated Methyl Alkyl Ketones. *Anal. Chim. Acta* **1984**, *159* (C), 265−274.
(31) Fetterolf, D. D.; Yost, R. A. Energy-Resolved Collision-Induced Dissociation in Tandem Mass Spectrometry. *Int. J. Mass Spectrom. Ion Phys.* **1982**, *44* (1−2), 37−50.
(32) McLuckey, S. A.; Glish, G. L.; Cooks, R. G. Kinetic Energy Effects in Mass Spectrometry/Mass Spectrometry Using a Sector/Quadrupole Tandem Instrument. *Int. J. Mass Spectrom. Ion Phys.* **1981**, *39* (2), 219−230.
(33) Schröder, D.; Engeser, M.; Brönstrup, M.; Daniel, C.; Spandl, J.; Hartl, H. Ion Chemistry of the Hexanuclear Methoxo-Oxovanadium Cluster V6O7(OCH3)12. *Int. J. Mass Spectrom.* **2003**, *228* (2−3), 743−757.
(34) Schröder, D.; Engeser, M.; Schwarz, H.; Rosenthal, E. C. E.; Döbler, J.; Sauer, J. Degradation of Ionized OV(OCH3)3 in the Gas Phase. From the Neutral Compound All the Way down to the Quasi-Terminal Fragments VO+ and VOH+. *Inorg. Chem.* **2006**, *45* (16), 6235−6245.
(35) Marshall, D. L.; Gryn'ova, G.; Coote, M. L.; Barker, P. J.; Blanksby, S. J. Experimental Evidence for Competitive N - O and O - C Bond Homolysis in Gas-Phase Alkoxyamines. *Int. J. Mass Spectrom.* **2015**, *378*, 38−47.
(36) Carlo, M. J.; York, P. M.; Patrick, A. L. Gas-Phase Dissociation Pathways of Beta-2 Agonists. *Int. J. Mass Spectrom.* **2021**, *463*, 116548.
(37) Toyama, A.; Nakagawa, H.; Matsuda, K.; Sato, T.-A.; Nakamura, Y.; Ueda, K. Quantitative Structural Characterization of Local N-Glycan Microheterogeneity in Therapeutic Antibodies by Energy-Resolved Oxonium Ion Monitoring. *Anal. Chem.* **2012**, *84* (22), 9655−9662.
(38) Fitzgerald, C. C. J.; McLeod, M. D. Synthesis of Stable Isotope Labelled Steroid Bis(Sulfate) Conjugates and Their Behaviour in Collision Induced Dissociation Experiments. *Org. Biomol. Chem.* **2022**, *20* (16), 3311−3322.
(39) Zins, E. L.; Pepe, C.; Schröder, D. Energy-Dependent Dissociation of Benzylpyridinium Ions in an Ion-Trap Mass Spectrometer. *J. Mass Spectrom.* **2010**, *45* (11), 1253−1260.
(40) Bouchoux, G.; Buisson, D.-A. Gas Phase Basicity of X(CH2)3Y(X, Y = OH, NH2) by the Thermokinetic Method. *Int. J. Mass Spectrom.* **2006**, *249−250*, 412−419.
(41) Fitzgerald, C. C. J.; Hedman, R.; Uduwela, D. R.; Paszerbovics, B.; Carroll, A. J.; Neeman, T.; Cawley, A.; Brooker, L.; McLeod, M. D. Profiling Urinary Sulfate Metabolites With Mass Spectrometry. *Front. Mol. Biosci.* **2022**, *9*, 829511.
(42) Waller, C. C.; McLeod, M. D. A Simple Method for the Small Scale Synthesis and Solid-Phase Extraction Purification of Steroid Sulfates. *Steroids* **2014**, *92*, 74−80.
(43) Association of Offical Racing Chemists. *AORC MS Criteria*. 2016.http://www.aorc-online.org/documents/aorc-ms-criteria-modified-23-aug-16/ (accessed 2020-11-05).
(44) Galuska, C. E.; Hartmann, M. F.; Sánchez-Guijo, A.; Bakhaus, K.; Geyer, J.; Schuler, G.; Zimmer, K.-P.; Wudy, S. A. Profiling Intact Steroid Sulfates and Unconjugated Steroids in Biological Fluids by Liquid Chromatography- Tandem Mass Spectrometry (LC-MS-MS). *Analyst* **2013**, *138* (138), 3792−3801.
(45) Shackleton, C. H. L.; Kletke, C.; Wudy, S.; Pratt, J. H. Dehydroepiandrosterone Sulfate Quantification in Serum Using High-Performance Liquid Chromatography/ Mass Spectrometry and a Deuterated Internal Standard: A Technique Suitable for Routine Use or as a Reference Method. *Steroids* **1990**, *55* (10), 472−478.
(46) Piper, T.; Putz, M.; Schänzer, W.; Pop, V.; McLeod, M. D.; Uduwela, D. R.; Stevenson, B. J.; Thevis, M. Epiandrosterone Sulfate Prolongs the Detectability of Testosterone, 4-Androstenedione, and Dihydrotestosterone Misuse by Means of Carbon Isotope Ratio Mass Spectrometry. *Drug Test. Anal.* **2017**, *9* (11−12), 1695−1703.
(47) Cawley, A. T.; Kazlauskas, R.; Trout, G. J.; George, A. V. Determination of Urinary Steroid Sulfate Metabolites Using Ion Paired Extraction. *J. Chromatogr. B* **2005**, *825* (1), 1−10.

https://doi.org/10.1021/jasms.2c00092
J. Am. Soc. Mass Spectrom. 2022, 33, 1276−1281

4.3 Future directions and commentary

One of the major limitations in metabolomic analysis is the identification of biomarkers against authentic synthesised reference materials. The main difficulty of this was exemplified in this chapter and the previous Chapter Two, where a single biomarker found by metabolomic analysis had 16 potential possible identities, on consideration of reasonable limitations. This is not uncommon and makes the identification of these types of biomarkers difficult and expensive. The idea of a wholly MS based approach for the elucidation of regio- or stereo-isomers of a biomolecule *in vivo* is uniquely situated, as all an analyst needs is a mass analyser with MS^n capabilities, with a high-resolution one obviously preferred. The energy resolved fragmentation developed in Chapter Three and applied in this chapter, speaks to this problem, and offers a solution. It could be further applied to other situations, some of which are described below.

It could be used to classify types of sulfate conjugates from one another *in vivo.* For example, the multivariate analysis presented in Chapter Two and later in Chapter Five, both found multiple sulfate molecules with the same accurate mass to charge ratios, e.g., four molecules with an *m/z* 387.1845 were discovered in Chapter 2.2. A reference set of compounds could be feasibly used to classify isobaric steroid sulfates from each other based on their conjugation site. A few things would have to be investigated. For instance, in the example given previously, *m/z* 387.1845, was given the general structure of an androstane triol mono-sulfate. This would likely have hydroxyl groups that do not exist on the 3 or 17 positions along the steroid backbone. To accommodate for this, the library of compounds used to classify the biomarkers would either be expanded or tested to see if they can be used to still discern between sulfates with more than two hydroxylation sites. If successful, the energy resolved fragmentation could be used as a ‘first-pass’ method in tandem with methods such as the sulfate *k*-means classifier to give a fuller idea of biomarker identify.

Investigation of other conjugate groups would also be a central avenue of future studies. The suitability of energy resolved method would have to be tested on other types of bonds, other than the sulfate ester C-O bond predominantly explored in

this chapter. A prime candidate for this would be steroid glucuronides. As discussed in Chapter Two, glucuronides make up the other major conjugated urinary metabolite class. Being able to discern between regio- or stereo-isomers of both glucuronides and sulfates by using solely MS-based methods would be a powerful approach. Although predominately speculative at this point, the energy resolve methods may not work as well on glucuronides. In their 2013 study Pozo *et al.* found that glucuronides predominately fragmented through the neutral loss of dehydrated glucuronide (176 Da).[1] This fragmentation does not involve the C-O bond at the stereocentre, as it does for the putative syn-elimination in saturated sulfate conjugates. However, several other characteristic fragmentations for glucuronides were also identified by Pozo that could give this information, such as the neutral loss of glucuronic acid (194 Da). This would have to be investigated further, however, it represents a potentially interesting application of the present work. Moreover, if energy resolved fragmentation was successfully applied to glucuronides an interesting example from work done in an earlier study already exist for investigation. Specifically, in a 2019 study by Pranata *et al.* when studying the fragmentation patterns of SIL-labelled bisglucuronides found a preference for phenolic glucuronide conjugates to undergo fragmentation over that of aliphatic ones.[2] Nonetheless, energy resolved fragmentation may be able to discern between the two types of glucuronide conjugation sites, if not the stereochemistry.

The potential of a wholly MS-based approach for partial structure elucidation remains at the heart of the energy resolved technique developed and is described in the last two chapters. It is hoped that this technique will be further developed into a viable tool to help aid in putative conjugate biomarker identification.

4.4 References

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

2. Pranata, A. *et al.* Synthesis of steroid bisglucuronide and sulfate glucuronide reference materials: Unearthing neglected treasures of steroid metabolism. *Steroids* **143**, 25–40 (2019).

CHAPTER FIVE

Semi-Targeted Metabolomic Profiling of *in vivo* Urinary Metabolites of Hemapolin

5.1 Introduction

5.1.1 Designer steroids in sports and the world

Misuse of designer anabolic steroids has become easier over recent years due to their availability as additives in “dietary” and “nutritional” supplements, accessible through a largely unregulated global online market for performance and image enhancing drugs (PIEDs).[1] Typically, these preparations are more akin to snake oil than therapeutics and regularly appear on the FDA’s Health Fraud Product Database,[2] as they can falsely promise many things such as ‘empowerment’, weight loss, mood modification, increased sexual drive, ‘glowing’ skin, enhanced cognitive function and muscle building capabilities.[1] Moreover, these dietary preparations have been found to contain steroidal compounds that have never been tested as therapeutic agents for medical or veterinary purposes.[1,3,4] Little is known about the toxicity or efficacy of these synthetic or designer steroids, with the presumption that they are biologically active based solely on their structural similarity to EAAS.[1,3] However, of the discovered anabolic compounds most have been found to be chemical derivatives of EAAS, and are likely produced in bulk in a cheap and unregulated manner.[1] Due to this they can be the subject of mal-practice in both human and equine sports, with their unusual chemical structures and substitution patterns increasing the potential to pass undetected in current anti-doping screens.[3,5] Therefore, to combat the growing threat of designer steroids, *in vivo* metabolic studies are essential intelligence for anti-doping authorities.

5.1.2 Hemapolin: an episulfide designer steroid

Recently a study carried out by the McLeod group examined the *in vivo* equine metabolism of the synthetic steroid hemapolin (**H**, 17α-methyl-2α,3α-epithio-5α-androstan-17β-ol) using GC-MS analysis.[6] Hemapolin was first reported by Okano *et al*. (2009),[7] and was discovered in the steroidal component of a “dietary” supplement, along with the related steroid epistane (**Epi,** 17α-methyl-2β,3β-epithio-5α-androstan-17β-ol). Hemapolin along with epistane and a handful of other designer steroids such as epitiostanol and mepitiostane, are uniquely substituted, as they possess a 2,3-episulfide group, a motif that occurs rarely in nature due to its general chemical

instability.[6–8] In the reported GC-MS study, Dr. Sumudu Weththasinghe and Dr. Christopher Waller, were able to establish robust synthetic methodology to access **H** and several of its urinary metabolites. These included madol (**M,** 17α-methyl-5α-androst-2-en-17β-ol) and the metabolites **T1** (17α-methyl-5α-androstane-2β,3α,17β-triol), **E2** (17β-hydroxy-17α-methyl-5α-androst-3-en-2-one), and **E4** (17β-hydroxy-17α-methyl-5α-androst-2-en-4-one), which were all detected in post administration equine urine extracts and identified according to AORC retention time and MS criteria.[9] These metabolites hint at extensive metabolic modification of the A ring in hemapolin, and from this, a metabolic pathway was proposed, Figure 5.1. Madol was also observed, however, it was unclear if it was derived from metabolism or the instability of **H** under the sample preparation (hydrolysis and derivatisation) conditions used for GC-MS. Further, the analysis also revealed several possible metabolites of **H** based on MRM transitions, including four major and two minor hydroxylated metabolites (**M1**-**M3**, **M6**) and (**M4**, **M5**), respectively, three dihydroxylated metabolites (**M7**-**M9**), and two tri-hydroxylated and reduced metabolites (**M10** and **M11**), Figure 5.1. Finally, several of the compounds including **H**, **E2**, **E4**, and **M** were found to be potent androgens in yeast cell-based an equine androgen receptor bioassay.

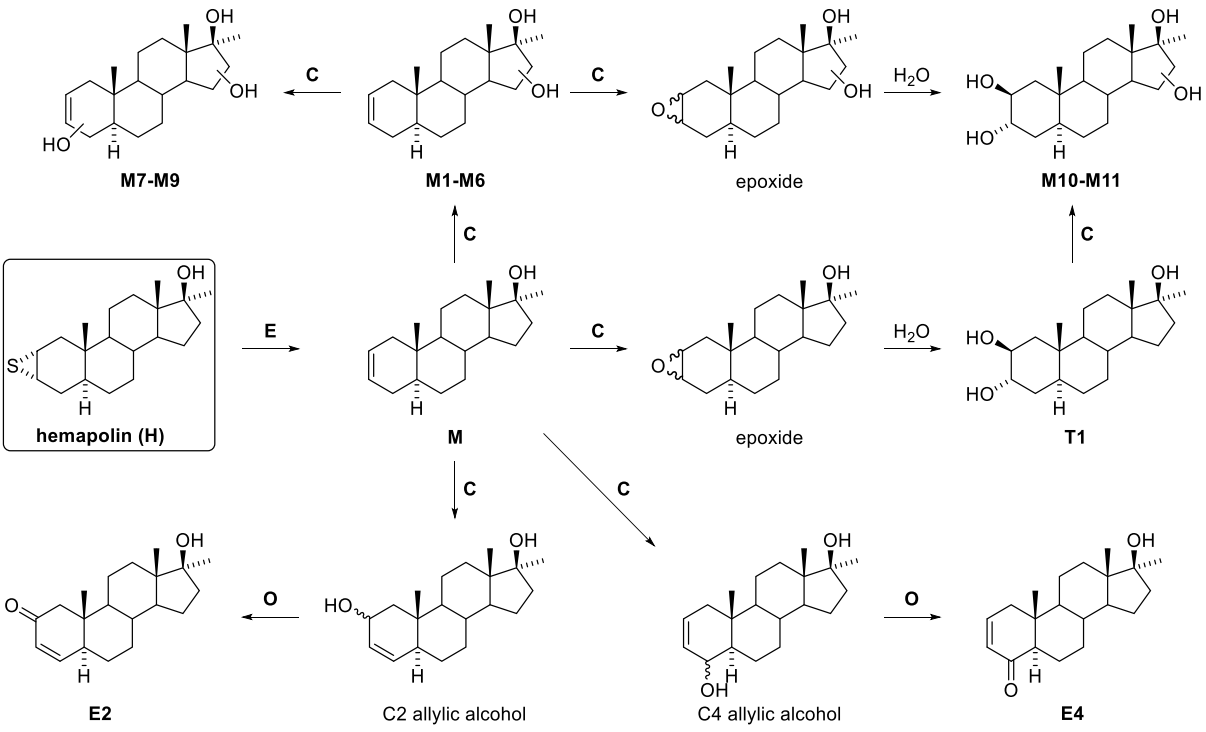


Figure 5.1: Proposed pathways for urinary Phase I metabolites of hemapolin in the horse. E, elimination; C, Cytochrome P450 oxidation; O, oxidation. Adapted from *Waller et al* (2020).[6]

5.1.3 LC-MS analysis of equine urinary metabolites of Hemapolin

In the GC-MS study, no steroid metabolites containing the episulfide or remnants of it, were detected. However, a related, yet un-published LC-MS analysis by Dr. Waller, found potential sulfur-containing metabolites based on their molecular formula. This was analogous to, and partially motivated by a study performed by Okano *et al* (2015),[8] on the metabolism of the episulfide containing designer steroid epitiostanol and its pro drug mepitiostane in human urine using LC-MS analysis, Figure 5.2. The major metabolite of epitiostanol was identified as the S-oxide derivative, which was observed to gradually convert, via dethionylation, to give 5α-androst-2-en-17β-ol.[8] Neither of these were found suitable as long term markers of epitiostanol/mepitiostane doping, and 5α-androst-2-en-17β-ol was found to be only a provisional indirect marker for doping. The reason for this was that it could only be identified in the urine of the Asian elephant, however, have not yet been observed in other systems.[8]

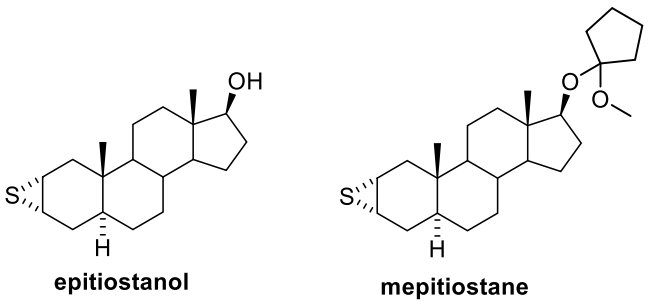


Figure 5.2: Structure of Epitiostanol and pro-drug mepitiostane.[8]

Nonetheless, upon UHPLC-HRMS analysis of the hemapolin administration samples Waller noted the presence of a putative ‘madol sulfate’ metabolite, based on the molecular formula derived from HRMS. However, this metabolite exhibited a fragmentation pattern uncharacteristic of an aliphatic sulfate ester compound, as it gave sulfur trioxide radical anion (*m/z* 80) as a major sulfate-derived fragment. To explain this, Dr. Weththasinghe,[10] proposed the presence of a minor epimer, epistane, derived from the synthesis of hemapolin. In this she hypothesised that these episulfides could metabolise via oxidation to a sulfoxide, followed by intramolecular rearrangement to sulfenic acid type metabolites than on further oxidation would then give metabolites with a sulfonate group. Four of these new metabolites were proposed

and were simply termed “sulfonates” and were proposed as an explanation for the irregular MS/MS fragmentation observed by Waller, Figure 5.3.

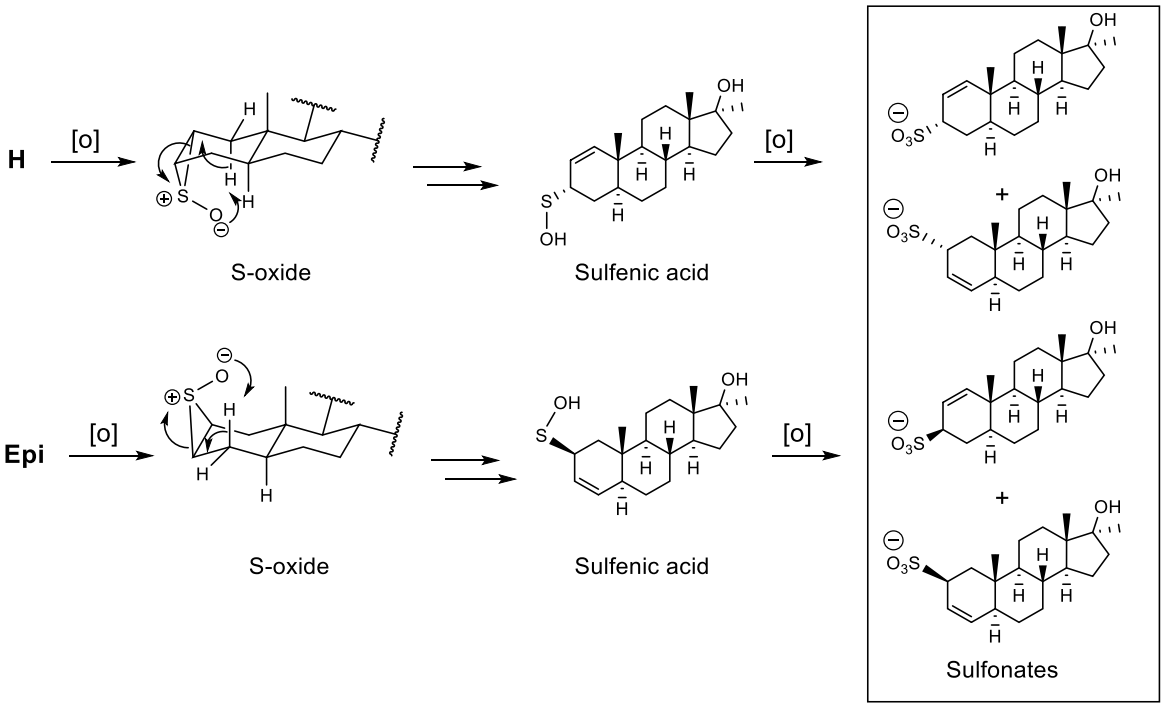
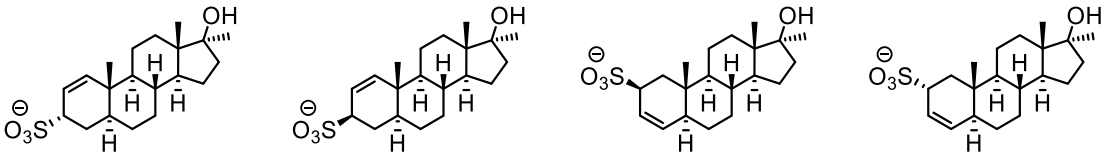


Figure 5.3: Proposed metabolic pathways for the formation of sulfonate metabolites from the major designer steroid drug hemapolin and its minor component epistane. Adapted from Weththasinghe (2020).[10]

The existence of epistane was identified following reanalysis of the original synthesised hemapolin reference material that was used in the administration study for the GC-MS analysis.[6] A small amount of epistane in a 1 in 10 ratio with hemapolin was observed. Therefore, it was hypothesised that similar metabolic processes would occur for both hemapolin and epistane leading to four isomeric sulfonates, and thus, synthesis was undertaken by Dr Weththasinghe with this in mind, Figure 5.3. Several reference materials were accessed through synthesis that could represent potential sulfonate metabolites, these are **SW56** (17β-hydroxy-17α-methyl-5α-androst-1-ene-3α-sulfonate), **SW66** (17β-hydroxy-17α-methyl-5α-androst-1-ene-3β-sulfonate), **SW72** (17β-hydroxy-17α-methyl-5α-androst-3-ene-2β-sulfonate), and **SW76** (17β-Hydroxy-17α-methyl-5α-androst-3-ene-2α-sulfonate), Figure 5.4. It should be noted that the synthesis was extensive and represented a substantial contribution by Dr. Weththasinghe, of which the details can be found in her subsequent thesis.[10] However, the synthesis description is outside the scope of this thesis.



SW56 **SW66** **SW72** **SW76**

Figure 5.4: Synthesised sulfonate reference materials of potential metabolites of hemapolin and epistane.[10]

5.1.4 *Aims: In vivo* metabolism of hemapolin and epistane in equine

The sulfonates provide excellent candidates for the unknown 'madol sulfate' first identified by Dr Waller, as they could feasibly give rise to the observed sulfur trioxide radical anion (*m/z,* 80). The remainder of this chapter will focus on the analysis of urine fractions taken post administration of hemapolin covering a range from 0h to 168h, using the same extraction, UHPLC-HRMS/MS and code-based methodologies developed throughout this thesis. There are several aims to this work. These are as follows:

1. Quantify the presence of any of the sulfonate metabolites (**SW56**, **SW66**, **SW72** and **SW76**) synthesised by Dr Weththasinghe in the post administration urine samples.

2. Quantify the presence of any Phase I metabolites synthesised by Dr Waller and used in the GC-MS analysis, using LC-MS.

3. Take kinetic profiles for the excretion of any confirmed metabolites.

4. Perform an untargeted metabolomic analysis to identify any biomarker signatures related to hemapolin administration that serve to extend the detection time.

5.2 Results and Discussion

The same single horse hemapolin administration samples used both by Dr. Waller and Dr. Weththasinghe, were used in this study.[6] This study involved the collection of urine at 0, 2, 4, 6, 8, 12, 24, 48, 72, 96, 144, and 168 hours after the administration of hemapolin. A control sample was also gathered from a related horse that was not administered with any dopant, to account for inter-horse variability.

5.2.1 Methodological design: semi-targeted profiling, a global and targeted analysis

A semi-targeted approach was adopted for the analysis of the hemapolin administration urine samples. This approach involved two main components.

Firstly, a targeted approach for all synthesised reference materials was performed, aimed at their confirmation, quantification, and then collection of individual kinetic excretion profiles over the full period of the administration study (168h). This was done for the eight synthesised reference standards, specifically: **SW56**, **SW66**, **SW72**, **SW76**, **T1**, **M**, **E2** and **E4**. Due to the nature of the reference standards, the screen was undertaken in both positive and negative modes, for the free steroids (**T1**, **M**, **E2** and **E4**) and sulfonates (**SW56**, **SW66**, **SW72** and **SW76**), respectively.

Secondly, the overall design of the experiment allowed for the untargeted profiling of the urinary steroidal metabolome. This was achieved through a general WAX SPE extraction with minimal washing steps allowing for the examination of free and conjugated steroids. All biological samples were prepared in triplicate, allowing for statistical analysis. Finally, data quality was checked with the use of pQC samples that were distributed uniformly throughout the analysis as performed in previous work.[11] This gave the necessary platform to allow for robust, high throughput differential metabolite level analysis followed by multivariate analysis to be performed on the data.

5.2.2 Detection and ionisation of hemapolin metabolites in LC-MS analysis

Detectability of all reference materials were assessed on the UHPLC-HRMS system. For the free steroids only the enone metabolites **E2** and **E4** were detected in positive mode ESI-UHPLC-MS/MS analysis as their $[M+H]^+$ ions. This could be rationalised due to the lack of ionisable functional groups on the triol (**T1**) and madol (**M**), and/or degradation over long periods of time stored at -20 °C, as fresh sample was not available. All sulfonates were detectable in the negative polarity as their $[M-NH_4]^-$ ions.

5.2.3 Targeted individual kinetic excretion profiles in equine urine

Confirmation of *in vivo* metabolites were first performed for each synthesised reference material against the 4 h post administration urine sample. For the sulfonate series **SW72** and **SW76** were detected and confirmed to be metabolites in the post administration sample, Table 5.1. The other two sulfonates **SW56** and **SW66** did not provide matches. In the positive mode analysis both **E2** and **E4** were also confirmed as metabolites, Table 5.1. All confirmed matches were performed according to AORC criteria for LC retention time and HRMS/MS analysis.[9] It should be noted that the sulfonate metabolites **SW72** and **SW76** meet the minimum required criteria for HRMS analysis with three observable transitions in MS/MS.

Table 5.1: *Confirmation using UHPLC retention time and MS/MS criteria*. Urinary metabolites were derived from the sulfate fraction of equine urine after administration hemapolin (**H**) were matched against synthesised reference materials. Confirmation was performed according to the Association of Official Racing Chemists (AORC) retention time and MS/MS criteria.[9] All CID product ion spectra used for confirmation were acquired using a collision energy (CE) of 20 eV for positive mode and 30 eV for negative mode. All MS/MS confirmations are taken from the 4 h sample post administration. The relative abundance of precursor and product ions was derived from the peak area of a single extracted ion chromatogram of the three replicates for the reference material and metabolite, mass filters were set at a tolerance of ± 5 ppm.

Reference material				**Metabolite**		
Compound name	**RT (min)** [theoretical *m/z*] ***n* =3**	**Product ion** (*m/z*)	**RA (%)**	**RT (min)** [tolerance] ***n* =3**	**RA** (%)	**AORC**[9] [%Tolerance]
E2	13.98 [303.2324]	303.2302	10	13.98 [13.93-14.03]	7	[0-20]
		285.2210	67		48	[47-87]
		267.2110	76		62	[53-99]
		245.1895	18		14	[8-28]
		227.1791	100		100	[70-100]
		187.1474	44		31	[31-57]
		159.1163	56		45	[39-73]
		145.1013	33		25	[23-43]
E4	13.56 [303.2324]	303.2311	92	13.55 [13.51-13.61]	100	[64-100]
		285.2214	41		48	[29-53]
		267.2108	100		94	[70-100]
		245.1910	13		12	[3-23]
		227.1794	64		70	[45-83]
		187.1475	45		32	[32-59]
		159.1165	74		53	[52-96]
		145.1011	63		44	[44-82]
		97.0657	4		5	[0-14]
SW72	12.47 [369.1943]	367.1943	86	12.46 [12.42-12.52]	81	[60-100]
		79.9566	100		100	[70-100]
		80.9644	1		3	[0-11]
SW76	11.48 [367.1943]	367.1943	100	11.50 [11.43-11.53]	100	[70-100]
		79.9566	25		28	[15-35]
		80.9644	3		4	[0-13]

Table 5.2: Retention time, repeatability, reproducibility, recovery at mid-range calibrators for different steroidal refence standards in surrogate blank equine urine

Compound	**RT (min)** ***n=34***	**%RSD** ***n=34***	**Repeatability (%RSD)[a] @ 10 ng/mL** ***n* = 4**	**Reproducibility (%RSD)[a] @ 10 ng/mL** ***n* = 5**	**Linearity (R^2)**	**Recovery (%) @ 100 ng/mL**	**LOD (ng/mL)**	**LOQ (ng/mL)**
E2[b]	13.97	0.1	0.07	0.07	y = 0.0043x – 0.0035 (0.99)	84	0.53	1.74
E4[b]	13.55	0.1	0.07	0.06	y = 0.0083x – 0.0068 (0.99)	91	0.59	1.94
SW72[b]	12.47	0.2	0.30	0.27	y = 0.048x – 0.04 (0.99)	107	2.17	7.17
SW76[c]	11.53	0.2	0.34	0.31[d]	y = 0.059x – 0.035 (0.97)	105	2.37	7.81

[a] Value for $\%RSD_{max}$ given by the Horwitz equation is 32% for a concentration of 10 ng/mL.[12]

[b] Concentration range for calibration curve 0, 2, 5, 10, 20, 50 ng/mL.

[c] Concentration range for calibration curve 0, 2, 5, 10, 20, 50, 100, 150, 200, 250 ng/mL.

[d] *n* = 7.

Preliminary results for each target steroid were taken for reproducibility, repeatability, recovery, linearity, and sensitivity (i.e., limits of detection (LOD) and limits of quantification (LOQ) where, LOQ = 3.3xLOD), Table 5.2. Further methodological detail can be found in the experimental section.

Following this, each target steroid's individual kinetic excretion profiles were established in the urine samples. All four confirmed metabolites followed similar profiles over the time of the experiment, Figure 5.5. These primarily consisted of the steroids being above the limits of quantification for less than 12 hours. Specifically, the **SW76** was above the LOQ from 2-4 h and both were above the LOD for **SW72** 2-6 h and **SW76** 2-8 h, Figure 5.5 A) & B). The enones were above LOQ and LOD from **E2** 2-6 h and **E4** 2-6 h, Figure 5.5 C) & D).

The profiles of the enones aligned with the findings of the GC-MS analysis, however, a considerable decrease in detection window was found for the **E4** metabolite from 72 h in GC-MS to 6 h by this LC-MS based analysis. The concentration of the enone metabolites were also considerably lower than that found in the GC-MS study, where peak excretions for **E4** were measured at 146 ng/mL. This could likely be explained by the contribution from the glucuronide conjugated fraction. This was not measured in this LC-MS study as no hydrolysis step was performed in the sample preparation. However, in the GC-MS study an acid solvolysis step was routinely performed using acetyl chloride with in sample preparation.[6] The LC-MS analysis would thus measure the contribution from the free un-conjugated fraction of the enone metabolites. Due to an absence of glucuronylated or sulfated reference materials this could not be explored further. Interestingly, the sulfonate metabolites were quickly cleared from the system, despite the large concentrations observed for **SW76** (i.e., approximately 200 ng/mL). This analysis provides the first observation of steroid sulfonate containing metabolites detected in any mammalian systems.

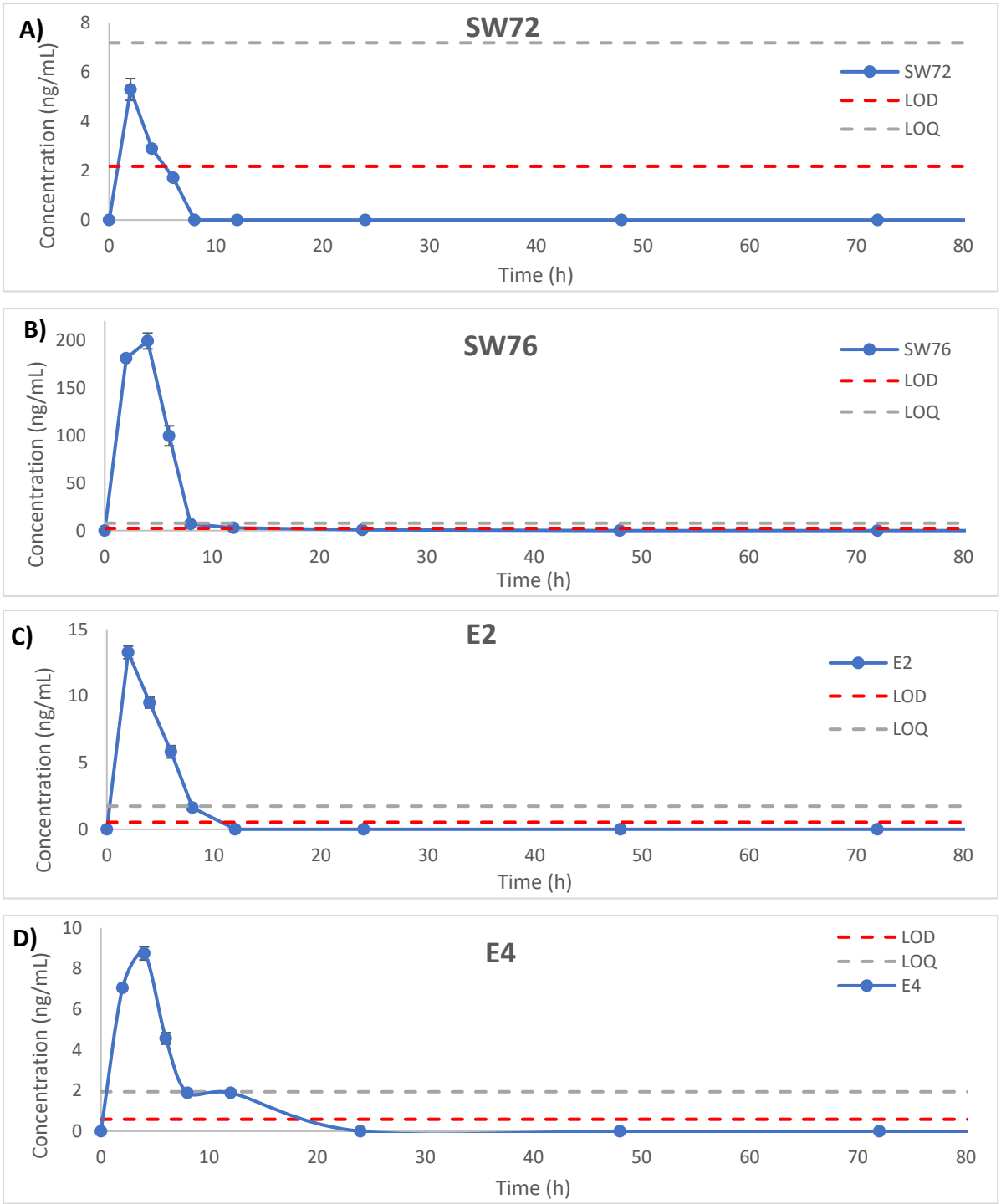


Figure 5.5: Kinetic excretion profiles of synthesised reference materials plotting concentration (ng/mL) over time (h) **A) SW72**; **B) SW76**; **C) E2**; and **D) E4**. Error bars represent standard deviation, n = 3. Note: times extend to 168 h, however, are only shown to 80 h for clarity.

5.2.4 Untargeted metabolomics to highlight potential biomarkers of hemapolin administration

Two data pipelines were designed and applied to the same UHPLC-MS/MS DDA urine sample datasets to highlight biomarkers that are potentially associated with hemapolin administration, Figure 5.6. Both data pipelines, 'approaches', have several things in common. Firstly, they both incorporate the high-throughput differential metabolite level analysis seen in Chapter Two of this thesis, including MS-Dial driven data alignment and LOWESS normalisation coupled with *k*-means analysis for putative sulfate conjugate identification for negative mode screens.[11] Secondly, multivariate analysis was also performed using the open-source software MetaboAnalyst 5.0. Here data pre-treatment was performed, including, filtering, scaling, and transformation in preparation for multivariate analysis, specifically, principal component analysis (PCA) and orthogonal projections to latent structures-discriminant analysis (OPLS-DA). The outcome of both approaches was to produce a final list of molecules that are potentially biological relevant, Figure 5.6.[13]

The approaches differ in the features that they consider as inputs in the multivariate analysis. *Approach 1*, uses all detected features as inputs into the OPLS-DA. The design of this portion of the analysis is similar to the one performed in Chapter 2.3, and was also influenced by the studies of Cloteau *et al.* (2021), Palermo *et al*. (2017) and Boccard *et al*. (2011).[14–16] Approach 2, was similar, but first isolated features based on theoretical Phase I transformations of madol, and or, from steroidogenesis, based on exact mass matching using a steroid searcher function written in R. In this way a theoretical steroid metabolome or 'steroidome' can be isolated to undergo analysis, and is similar to the steroidomics approach taken by Boccard *et al.* (2011).[16] This steroidomic subset of features is then used as the input for the OPLS-DA. Like *Approach 1* OPLS analysis is also used to monitor the perturbation arising from hemapolin administration in equine urine. A further data pipeline that was not explored was constraining all features based on accurate *m/z* matching to the theoretical list of steroids, *following* the OPLS-DA and S-Plot analysis of all features (*Approach 1*). This was not pursued as it could give superfluous results, as it may not pick the most predictive biomarkers in that given dataset.

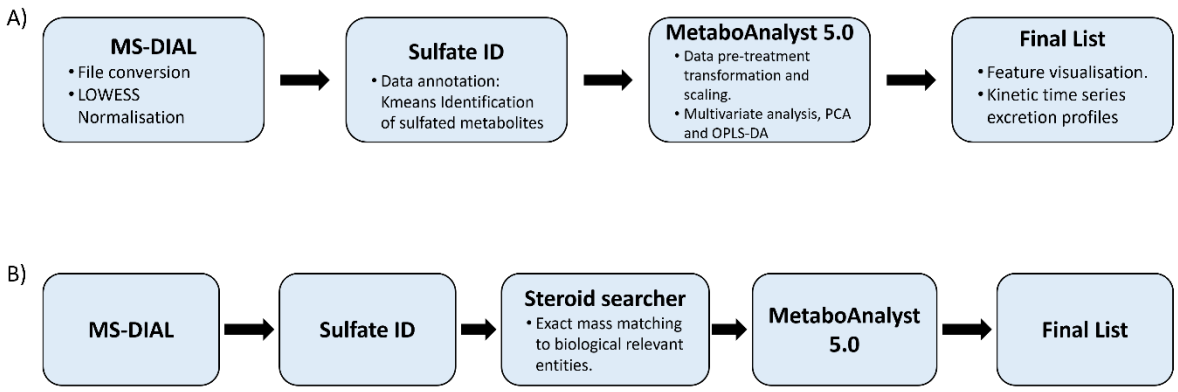


Figure 5.6: Proposed untargeted workflows to identify biomarkers potentially associated with the administration of a dopant in negative mode. **A)** *Approach 1*; A wholly non-targeted approach for analysis of equine urine data. **B)** *Approach 2*; A semi-targeted approach that incorporates exact mass matching for Phase I metabolic transformations of madol and steroids from the steroidogenesis pathways. Note: that *Approach 2* emulates the steps taken in *Approach 1*, unless otherwise indicated.

5.2.5 Approach 1: Non-targeted metabolomics analysis of hemapolin administration samples

5.2.6 Data quality assessment and pre-treatment

Data quality is of vital importance in untargeted metabolomic screens and provides the platform for an unbiased analysis of a given biological question.[17] The methodology used to secure good data quality was established in Chapter Two of this thesis and a similar approach is applied in this analysis.[11] First the data was aligned with reasonable tolerance (RT ± 3s and $\Delta m/z$ ± 5 ppm) and normalised (LOWESS) in MS-DIAL. This created a list of 21678 features with MS1 in the positive mode screen that were all retained for further analysis. In negative mode a list of 16755 features was created following alignment but only data with associated DDA MS/MS spectra were brought forward into the analysis, to identify sulfate metabolites as performed in Chapter Two.[11] After removal, this gave 3919 features in the negative screen. Following *k*-means analysis of the negative mode data, the metabolome was split into 3348 non-sulfate features and 571 putative sulfated features, Figure 5.7. Of these the descriptive statistics for the sulfates were an average MA of 87 (standard deviation =

17, median = 98) and IR of 56 (standard deviation = 23, median = 55), and for the non sulfates, an average MA of 2 (standard deviation = 7, median = 0) and IR of 1 (standard deviation = 5, median = 0) was obtained, indicating tight clustering of the two groups by the *k*-means function.[11]

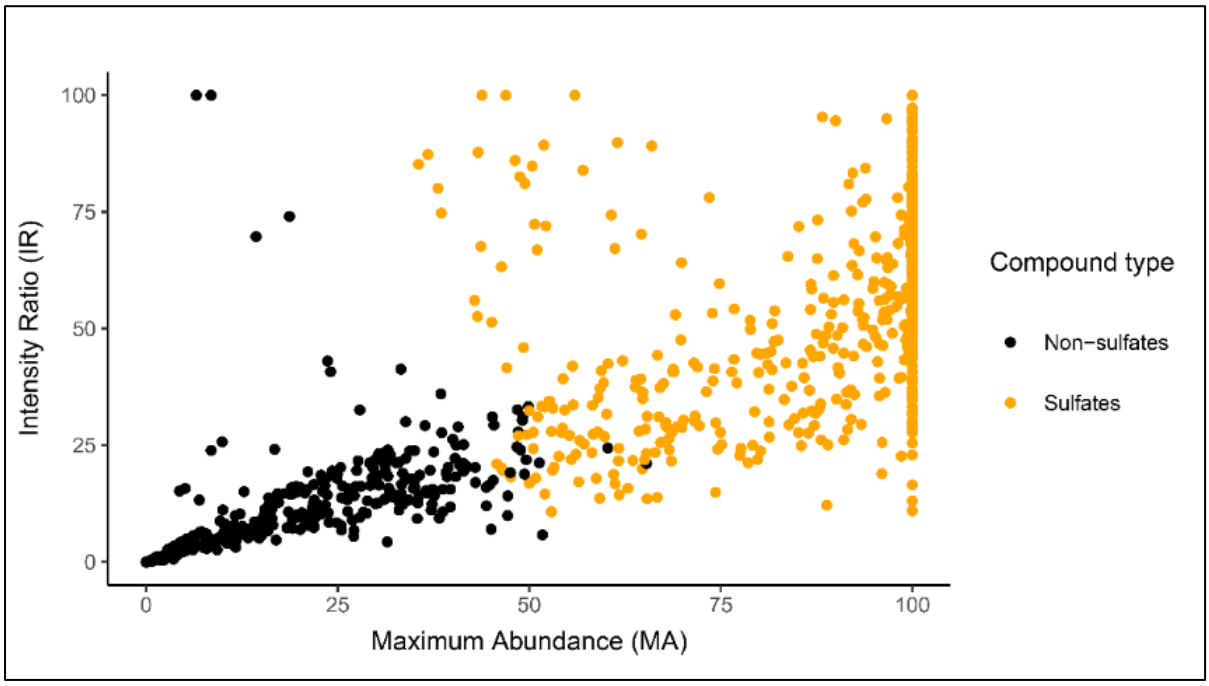


Figure 5.7: The *k*-means clustering plot contrasting MA against IR of sulfated (571) and non-sulfated (3348) features in the negative mode screen, total features, n = 3919.

Following this both the positive and negative data underwent pre-treatment consisting of data filtering, scaling, and transformation in MetaboAnalyst 5.0. Data filtering was performed based on the comparison to the relative standard deviation (RSD) of the pooled QC samples, and further filtered based on inter-quantile range as performed in Chapter 2.3.[18] This resulted in two curated datasets of similar sizes, 2499 features in positive mode and 2006 features (389 sulfates and 1617 non-sulfates) for negative mode. Following this, Pareto scaling and $\log_{10}$ transformation were applied on the datasets before undergoing multivariate analysis, as performed in Chapter 2.3.

5.2.7 Multivariate analysis and OPLS-DA model establishment

In preparation for OPLS-DA analysis the datasets were then divided into four groups based on their type and clustering observed in unsupervised PCA, Figure 5.8.[14,18,19] These were pre-administration (Pre) which consisted of the 0 h samples (*n* = 3), post-administration (Post) which consisted of time points 2-168 h (*n* = 10 x 3), the pooled QC (QC) samples (*n* = 12) and an un-doped control horse urine (Con) sample (n = 3). The groupings used were identical in the negative and positive mode screens.

The groupings also differed to the analogous study done in Chapter 2.3, as in this study, clear groupings were observed between the pre- and post-administration samples in the PCA score plots. Furthermore, good data quality was indicated by the tight clustering of the pooled QC samples relative to the total observed variability, in both positive and negative mode analysis, Figure 5.8.[11,14,17] In both cases, three principal components (PCs) could explain 65% and 70 % of variance in the positive and negative modes, respectively. Groupings were also consistent throughout both polarity modes, and the pre- and post-administration samples were distinct. These observations make the data suitable for further supervised analysis.[14]

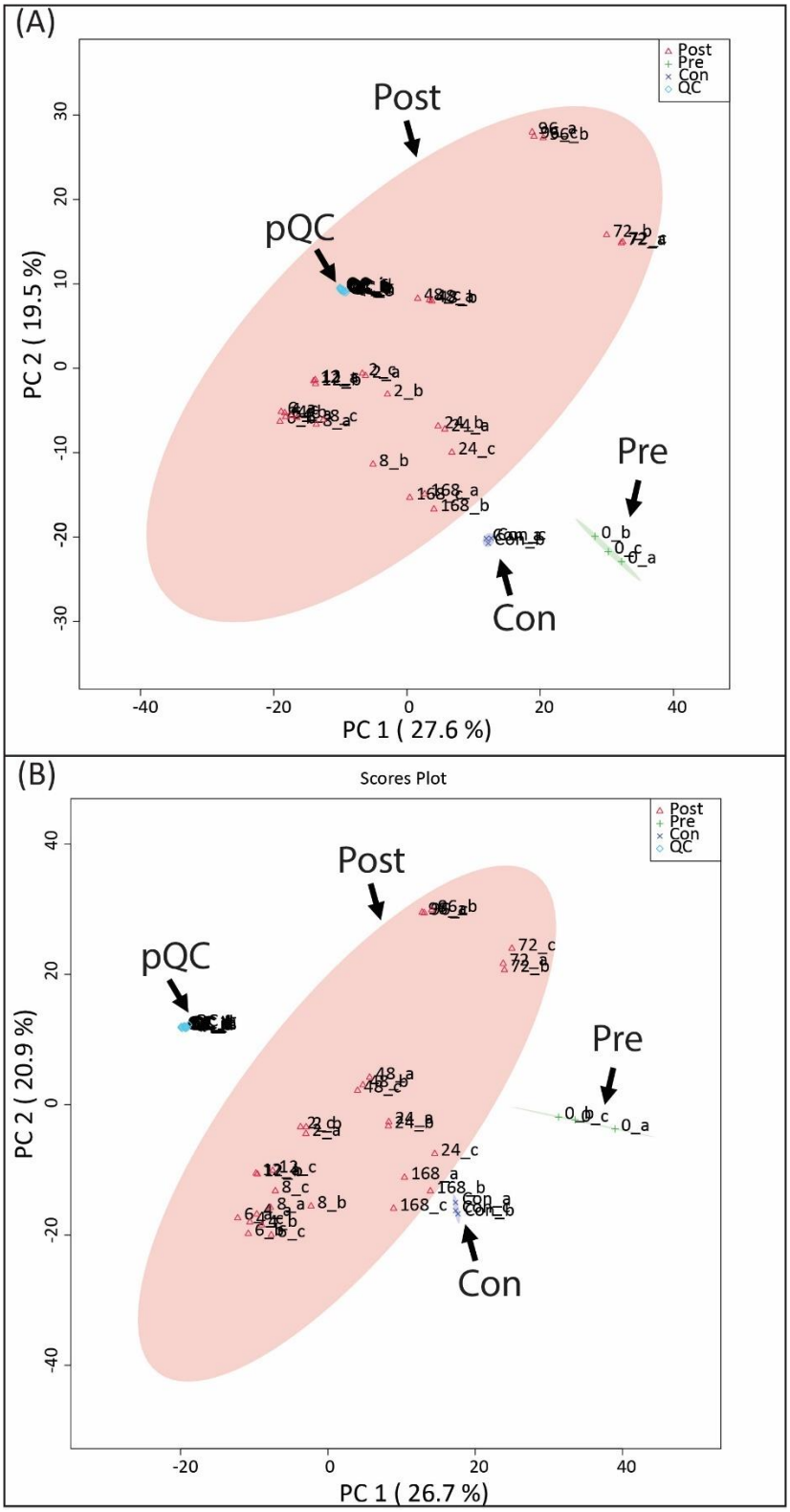


Figure 5.8: Principal Component Analysis (PCA) built from the linear combination of the total kinetic excretion profile (Pre, *n* = 3 green: Post, *n* = 30, pink) as well as a control horse (Con, *n* = 3, blue), and pooled QC samples (pQC, *n* = 12, light blue), with total samples, *n* = 48. **A)** PCA score plot of the negative mode screen. **B)** PCA score plot of the positive mode screen. Ellipses indicate the 95 % CI interval of the groupings.

Following this, supervised OPLS-DA was used to model the status, i.e., undoped and doped, of the pre (*n* = 3) and post administration (*n* = 30) groups. In these models the control horse and pQC samples were omitted. As discussed in Chapter Two, OPLS-DA is a powerful tool for dimension reduction and the identification of features that are responsible for variation between groupings.[20] Modelling the status of both groups was performed using the supervised OPLS-DA function in MetaboAnalyst 5.0, which generated the descriptive model for both positive and negative modes, Figure 5.9.[13]

Both models were validated and gave the following performances: Positive mode cross validation was done, giving R2(X) =0.21, R2(Y) = 0.86, and Q2 = 0.84, and permutation test = R2(Y) = 0.99, and Q2 = 0.95 (p-value < 0.001). Negative mode cross validation was done, giving Q2 =0.25, R2(Y) = 0.83, and Q2 = 0.81, and permutation test = R2(Y) = 0.99, and Q2 = 0.94 (p-value < 0.002, permutations = 1000).

The models allowed for evidence of distinct clusters of pre-administration and post-administration samples in both positive and negative mode data, indicated by the relative closeness in the R2(Y) and Q2 values, and the observed increase of the values in the permutation test.[20,21] Both the models clearly distinguish the chronological based groupings of urinary metabolites.

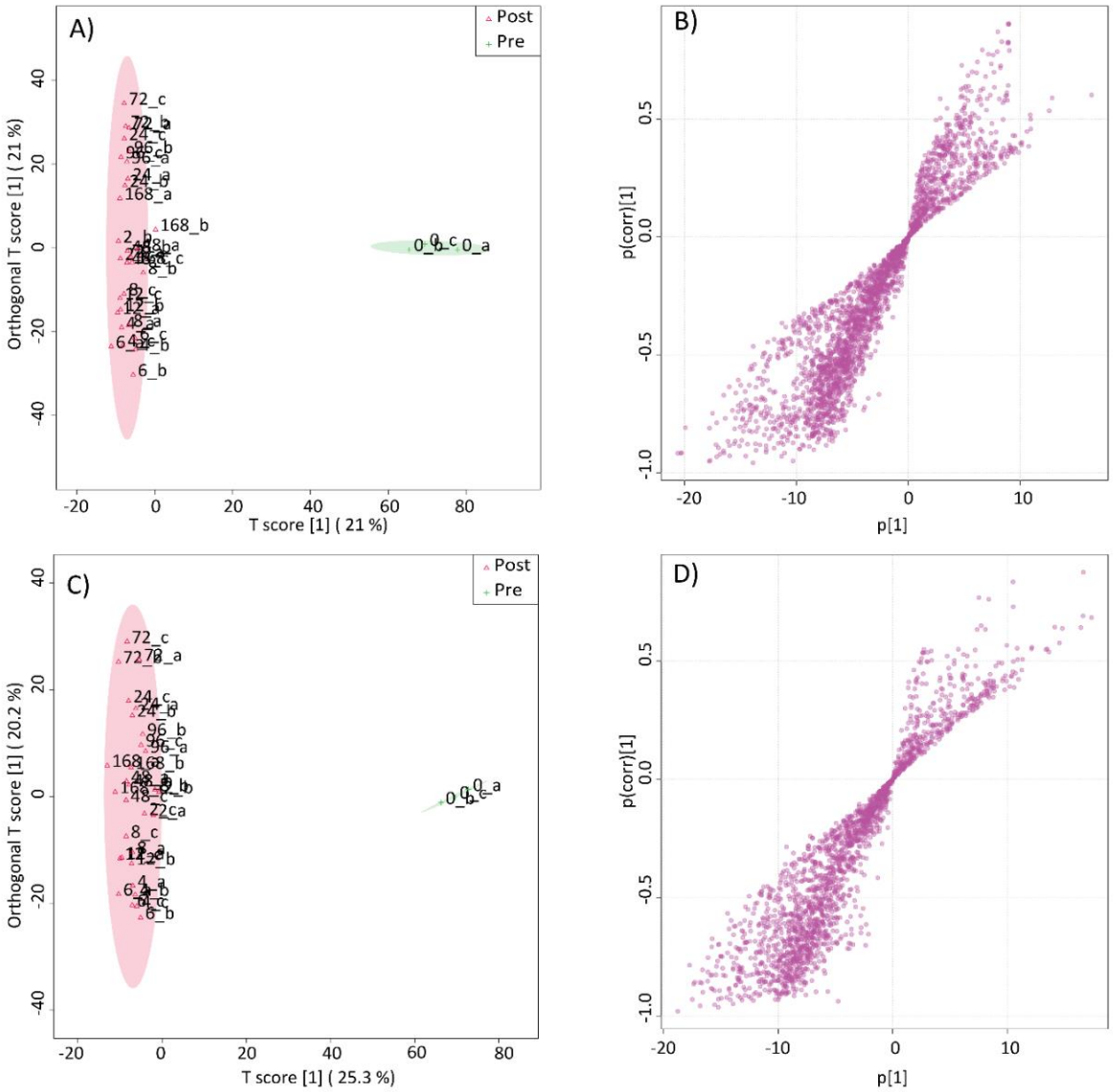


Figure 5.9: The OPLS-DA plots generated between pre-administration and post administration samples with associated S-plots. Post administration samples, *red*. Pre-administration samples, *green*. **A)** OPLS-DA score plot of positive mode. **B)** S-plot of positive mode features (n = 2499).**C)** OPLS-DA score plot of negative mode **D)** S-plot of negative mode samples (n = 2006). Ellipses indicate the 95 % CI interval of the groupings.

5.2.8 Identification of potential biomarkers of hemapolin administration

In this analysis, like in Chapter Two, S-Plots were used as the primary analytical tool to highlight the most predictive features in each OPLS-DA. In both screens the top 10 significant features were chosen for further analysis. In addition to these, all features above the absolute values for 15 for p[1] **and** above an absolute value of 0.75 for p(corr)[1], were aligned against a list of theoretical steroidal biomarkers using their accurate mass (± 5 ppm), similar to the method used in Chapter Two and by Boccard *et al*. (2011), Tables 5.3 and 5.4.

In general, the OPLS-DA model in both positive and negative screens only picked molecules that were up-regulated when compared to the pre-administration time point. Following the application of the steroid searcher algorithm to the features isolated by the S-plot thresholds, no matches were found in the positive screen. In the negative screen one steroid was matched, as **A10G**. In summary, from this selection criteria a list of 10 putative biomarkers were found in positive mode and 11 were identified in the negative mode analysis, the results are shown in Table 5.5.

Table 5.3: A list of reference *m/z* related to possible Phase I transformations of hemapolin (**H**) and madol (**M**), up to four transformations were considered.

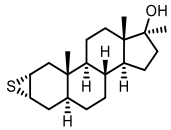
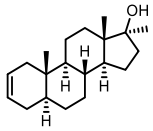
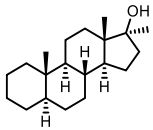
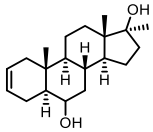
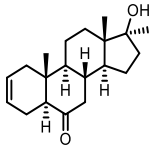
Name	Formula	Theoretical Free Structure	Transformation	Exact mass	Deprotonated mass	Protonated	Sulfated		Glucuronylated	
				[M]	$[M-H]^-$	$[M+H]^+$	$[M]^-$	$[M+2H]^+$	$[M]^-$	$[M+2H]^+$
Hemapolin	$C_{20}H_{32}OS$		None	320.21739	319.209561	321.225211	399.1664	401.181478	495.2417	497.256752
Madol	$C_{20}H_{32}O$		None	288.24532	287.23749	289.25314	367.1943	369.209407	463.2696	465.284681
A1	$C_{20}H_{34}O$		1 x Reduction	290.26097	289.25314	291.26879	369.2100	371.225057	465.2852	467.300331
A2	$C_{20}H_{32}O_2$		1x Hydroxylation	304.24023	303.232405	305.248055	383.1892	385.204322	479.2645	481.279596
A3	$C_{20}H_{30}O_2$		1 x Hydroxylation 1 x Oxidation	302.22458	301.216755	303.232405	381.1736	383.188672	477.2488	479.263946

Table 5.3: continued…

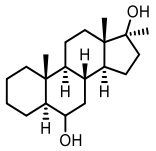
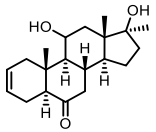
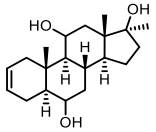
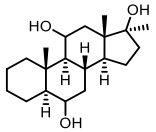
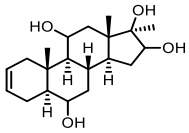
A4	$C_{20}H_{34}O_2$		1 x Hydroxylation 1 x Reduction	306.25588	305.248055	307.263705	385.2049	387.219972	481.2801	483.295246
A5	$C_{20}H_{30}O_3$		2 x Hydroxylation 1 x Oxidation	318.2195	317.21167	319.22732	397.1685	399.183587	493.2438	495.258861
A6	$C_{20}H_{32}O_3$		2 x Hydroxylation	320.23515	319.22732	321.24297	399.1841	401.199237	495.2594	497.274511
A7	$C_{20}H_{34}O_3$		2 x Hydroxylation 1 x Reduction	322.2508	321.24297	323.25862	401.1998	403.214887	497.2751	499.290161
A8	$C_{20}H_{32}O_4$		3 x Hydroxylation	336.23006	335.222235	337.237885	415.1791	417.194152	511.2543	513.269426

Table 5.3: continued…

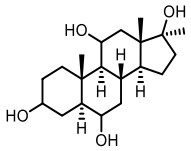
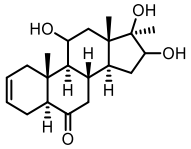
A9	$C_{20}H_{34}O_4$		3 x Hydroxylation 1 x Reduction	338.24571	337.237885	339.253535	417.1947	419.209802	513.27	515.285076
A10	$C_{20}H_{30}O_4$		3 x Hydroxylation 1 x Oxidation	334.21441	333.206585	335.222235	413.1634	415.178502	509.2387	511.253776

Table 5.4: A list of reference m/z values related to endogenous steroids taken from steroidal biosynthetic pathways in mammalian systems and related studies.

Possible biomolecule	Formula	Exact mass	Deprotonated mass	Protonated	Sulfated		Glucuronylated	
		[M]	$[M-H]^+$	$[M+H]^+$	$[M]^-$	$[M+2H]^+$	$[M]^-$	$[M+2H]^+$
*Estrone	$C_{18}H_{22}O_2$	270.16198	269.154155	271.169805	349.111	351.126072	445.1862	447.201346
*Estradiol	$C_{18}H_{24}O_2$	272.17763	271.169805	273.185455	351.1266	353.141722	447.2019	449.216996
*Estriol	$C_{18}H_{24}O_3$	288.172545	287.1647	289.1804	367.12154	369.1366	463.19681	465.21191
*DHEA *Testosterone	$C_{19}H_{28}O_2$	288.20893	287.201105	289.216755	367.1579	369.173022	463.2332	465.248296
*Androsterone	$C_{19}H_{30}O_2$	290.22458	289.216755	291.232405	369.1736	371.188672	465.2488	467.263946
*Androstane diol	$C_{19}H_{32}O_2$	292.24023	291.232405	293.248055	371.1892	373.204322	467.2645	469.279596
6β-Hydroxy-androstenedione[22]	$C_{19}H_{26}O_3$	302.1882	301.18037	303.19602	381.1372	383.152287	477.2125	479.227561
*16α-Hydroxy-DHEA	$C_{19}H_{28}O_3$	304.20385	303.19602	305.21167	383.1528	385.167937	479.2281	481.243211
3,16-Dihydroxy-androstan-17-one[23]	$C_{19}H_{30}O_3$	306.2195	305.21167	307.22732	385.1685	387.183587	481.2438	483.258861
5α-Androstane-3β,16β,17β-triol[24]	$C_{19}H_{32}O_3$	308.23515	307.22732	309.24297	387.1841	389.199237	483.2594	485.274511
*Progesterone (as an enol)	$C_{21}H_{30}O_2$	314.22458	313.216755	315.232405	393.1736	395.188672	489.2488	491.263946
*Pregnenolone	$C_{21}H_{32}O_2$	316.24023	315.232405	317.248055	395.1892	397.204322	491.2645	493.279596
*3β-hydroxy-5α-Pregnane-20-one	$C_{21}H_{34}O_2$	318.25588	317.248055	319.263705	397.2049	399.219972	493.2801	495.295246
*5β-Pregnane-3α, 20α-diol	$C_{21}H_{36}O_2$	320.27153	319.263705	321.279355	399.2205	401.235622	495.2958	497.310896
*17α-Hydroxy-progesterone[25]	$C_{21}H_{30}O_3$	330.2195	329.21167	331.22732	409.1685	411.183587	505.2438	507.258861
*17a-Hydroxy-pregnenolone	$C_{21}H_{32}O_3$	332.23515	331.22732	333.24297	411.1841	413.199237	507.2594	509.274511
*5α,3α-Tetrahydro deoxy-corticosterone	$C_{21}H_{34}O_3$	334.2508	333.24297	335.25862	413.1998	415.214887	509.2751	511.290161
*11-Dehydrocorti-costerone	$C_{21}H_{28}O_4$	344.19876	343.190935	345.206585	423.1478	425.162852	519.223	521.238126
*Corticosterone	$C_{21}H_{30}O_4$	346.21441	345.206585	347.222235	425.1634	427.178502	521.2387	523.253776
*Cortisone	$C_{21}H_{28}O_5$	360.19368	359.18585	361.2015	439.1427	441.157767	535.2179	537.233041
*Cortisol	$C_{21}H_{30}O_5$	362.20933	361.2015	363.21715	441.1583	443.173417	537.2336	539.248691
*18-Oxocortisol	$C_{21}H_{28}O_6$	376.18859	375.180765	377.196415	455.1376	457.152682	551.2129	553.227956
*18-Hydroxycortisol	$C_{21}H_{30}O_6$	378.20424	377.196415	379.212065	457.1532	459.168332	553.2285	555.243606

*biomarker found in multiple sources.[26–34]

Table 5.5: The list of matched *m/z* for molecules identified by the S-plots. The MS/MS data from the raw spectra are also shown.

Putative assignment	Row Id	RT [min]	*m/z*	p[1]	p[corr][1]	MS/MS annotation	MS/MS spectra: *m/z* (%RA)
Positive mode							
N/A	3805**	6.45	255.1231	-17	-0.93	N/A	181.0863 (57), 166.06261 (100), 153.05431 (27), 149.0598 (46), 148.052 (45), 117.0694 (17), 91.0537 (31),
N/A	13831	1.64	394.0938	-18	-0.95	N/A	394.0945 (100), 283.4260 (11)
N/A	15077**	5.85	410.1596	-17	-0.91	N/A	410.15961 (25), 410.1489 (100), 267.142 (17), 248.09689 (25)
N/A	15914**	11.73	422.2754	-16	-0.93	N/A	229.2148 (100), 211.20779 (49), 89.7614 (12)
N/A	16856**	6.45	439.1587	-18	-0.95	N/A	439.1578 (100), 364.1124 (11), 277.1069 (9), 202.0596 (4), 201.0525 (9), 176.2143 (1)
N/A	18025**	6.39	466.1715	-16	-0.92	N/A	413.1268 (38), 273.1107 (94), 163.0751 (37), 161.0955 (50), 137.0584 (55), 123.0433 (100),
N/A	20290**	14.48	549.3016	-17	-0.92	N/A	549.3016 (100)
N/A	20862**	9.55	624.0103	-20	-0.92	N/A	624.3431 (27), 624.0104 (100), 529.3099 (14), 472.2863 (13),159.0764 (12)
N/A	20865**	9.55	624.3441	-20	-0.92	N/A	624.3440 (100), 624.0091 (60), 29.3103 (33), 159.0764 (29), 112.0854 (24)
N/A	20866**	9.55	624.3468	-20	-0.92	N/A	N/A
Negative mode							
N/A	1815**	4.40	247.0279	-16	-0.89	N/A	123.0795 (100)
N/A	5230**	7.09	330.1012	-16	-0.89	N/A	250.1450 (11), 232.1330 (7), 206.1550 (100), 190.1250 (13), 177.1160 (21), 149.0840 (45), 133.0520 (7), 120.0450(28)
N/A	8570	7.96	389.0252	-17	-0.93	Sulfate	389.20252 (100), 371.20651 (6), 329.2012 (11), 309.06729 (50), 246.1873 (8), 215.85339 (4), 154.9903 (24), 151.1134 (13), 129.01981 (8), 123.0816 (11), 116.0102 (13), 113.0237 (59),99.0077 (5),95.0136 (19),85.0287 (32)
N/A	8662	8.08	389.2741	-19	-0.98	N/A	389.27411 (100)
N/A	8666	8.07	389.3004	-15	-0.96	N/A	389.30042 (100)
N/A	8750	7.91	389.6833	-16	-0.95	N/A	390.2202(100), 310.0710(35), 272.1940 (25), 200.067 (10), 132.006 (23), 129.017 (25), 116.0110 (37), 114.026 (21), 113.023 (27), 90.0299 (29), 85.0283 (29), 56.9987 (30)
N/A	9815	11.76	403.2339	-17	-0.90	N/A	96.9574 (100)
N/A	11505	5.40	433.1141	-15	-0.89	N/A	135.0470 (100), 107.5990 (36), 97.6275 (48)
N/A	14142	14.93	503.3224	-15	-0.93	N/A	503.3224 (100), 58.6504 (17)
A10G (Glucuronide)	14256**	10.96	509.2390	-16	-0.75	Glucuronide	333.2070 (100), 102.4430 (10)
N/A	15082**	14.64	573.3251	-17	-0.96	N/A	505.3399 (100)

**indicates that the peak was also detected in the control sample.

5.2.9 Biomarker annotation of significant features

The kinetic excretion profiles of the features in Table 5.5 were then reconstructed from their normalised peak height intensity and are shown in Figures 5.10 and 5.11. All features were then searched for on the Chemical Entities of Biological Interest (ChEBI) database by exact mass matching (0.003 Da). No matches were found. Next the MS/MS were examined. Most features in both modes had associated MS/MS spectra, Table 5.5. Of the negative mode molecules, *m/z* 389.02521 (RT 7.094 min), was putatively identified by the *k*-means analysis as a sulfated metabolite based on the neutral loss of sulfur trioxide, SO_3. Additionally, the steroid searcher algorithm found an accurate mass match to a putative steroid glucuronide metabolite, **A10G**. This metabolite was then manually assigned as a glucuronide based on MS/MS transitions common to glucuronides, a sifmilar approach was taken in Chapter Two.[35] Although inconclusive, as previously discussed in this thesis finding no matches to biomarkers is not an uncommon occurrence in untargeted metabolomic profiling and represents the most difficult part of this type of metabolomic assay. Despite this, similar studies have shown that it is possible to build classification models around features without clear identifications.[14]

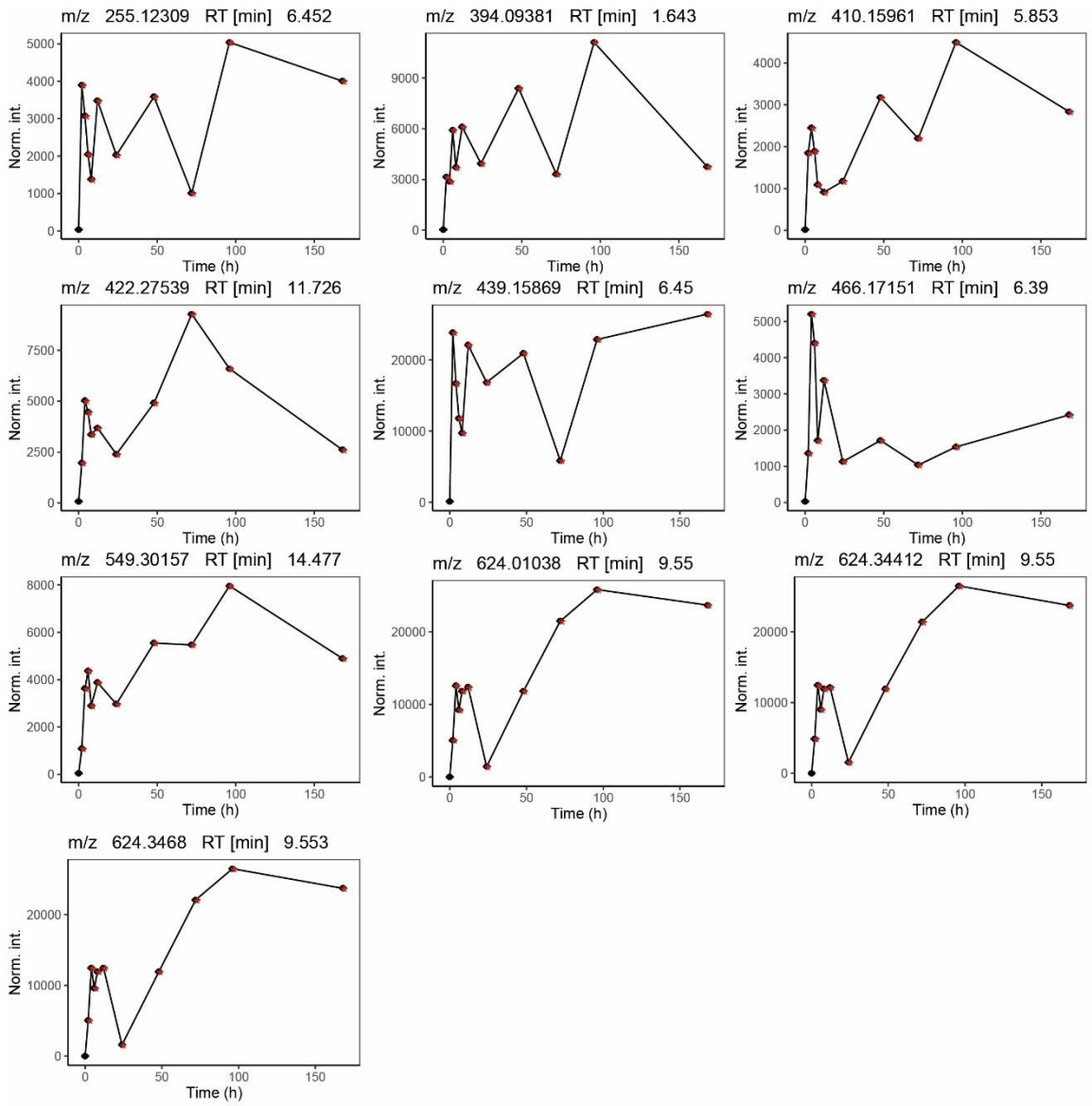


Figure 5.10: Positive mode kinetic excretion profiles of the top 10 significant molecules identified in OPLS-DA analysis of hemapolin administration. Norm. Int. = LOWESS normalised intensity. The error bars show standard error, SE = σ/(n)1/2, n = 3. The error bars are shown, however, are obscured due to their small size. The asterisk * indicates a p-value < 0.01, relative to time 0 h. Note; all non- 0h samples were significant.

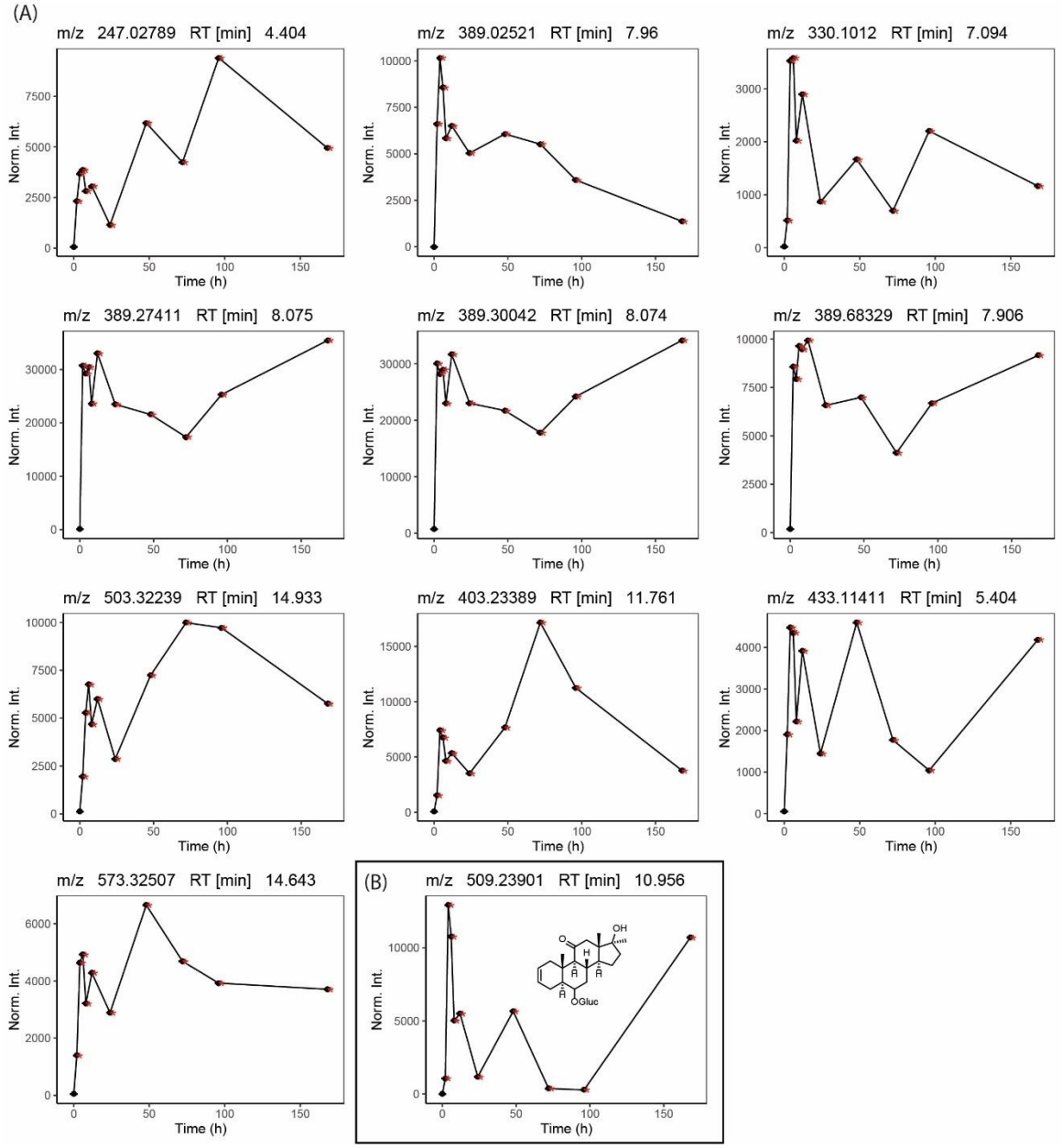


Figure 5.11: **A)** Negative mode kinetic excretion profiles of the top 10 significant molecules identified in OPLS-DA analysis of hemapolin administration. **B)** A putative steroid glucuronide metabolite, **A10G**, that matched theoretical steroid masses in Tables 5.4 and 5.5. Norm. Int. = LOWESS normalised intensity. The error bars show standard error, SE = σ/(n)1/2, n = 3. The error bars are shown, however, are obscured due to their small size. The asterisk * indicates a p-value < 0.01, relative to time 0 h. Note; all non- 0h samples were significant.

5.2.10 Discussion of Approach 1

The goal of this untargeted analysis was to find potential markers of hemapolin doping in equine. One of the major challenges in LC-MS based metabolomics is metabolite annotation. This largely stems from the complexity of the metabolome, the lack of metabolites that have been structurally identified or are available. Therefore, in this study, despite being the cause of large variation as identified by the OPLS-DA model, the highlighted molecules have not been structurally identified or confirmed and do not equate to being biologically relevant. Proper identification or population studies would be needed to validate them. However, the type of analysis used in *Approach 1*, does have benefits, as it gives a small subset of molecules to focus on, relative to the thousands of initial features that were detected in the urinary metabolome.

In this approach putative identifications were made to a sulfated metabolite and a putative steroid glucuronide metabolite (**A10G**) based on accurate mass matching as well as identification of characteristic MS/MS transitions. Despite these two putative identifications no other matches were found or inferences made about the identity of the highlighted molecules. From this point in the analysis there are several directions forward in biomarker identification. One being the exact mass alignment to searchable online databases (e.g. ChEBI[36]) that was already trialled. Other analyses include, the comparison to mass spectral libraries (e.g. Manchester Metabolomics Database[37], NIST, MassBank, MoNA, LipidBlast, Wiley MSforID, METLIN, and the Human Metabolome data base).[38] These approaches may provide useful guidance to the researcher and help them identify molecule classes to further guide the preparation of suitable reference materials. However, they are still largely reliant on statistically based scoring algorithms.[38] Another approach would be the use of accurate mass and isotope patterns for the establishment of a molecular formular, which may also suggest the class or type of molecule. This next data annotation step would represent a considerable effort and is outside the scope of this current workflow and thesis due to time constraints. However, it would be a beneficial avenue to pursue in the future.

Lastly, a classification approach could be evaluated based on the OPLS-DA performed in this study. Classification models are an alternative method to traditional direct metabolite screening, and have been successfully used to identify testosterone ester and nandrolone doping in equine.[14,19] The benefit and limitation of these models is that they do not rely on knowing the absolute identification of biomarkers. Instead, they build a statistical model around these biomarkers of interest, and use this resulting metabolic signature to detect for abnormalities in samples.[14,19]

5.2.11 Approach 2: Semi-targeted steroidomic analysis of hemapolin administration samples

As explored in *Approach 1*, untargeted metabolomics can provide a comprehensive overview of the metabolome, however, its limitations lie in failing to identify the highlighted biomarkers of interest. Pathway-based approaches have been developed to circumvent this problem. In their recent paper, Xu *et al*. (2019), outlined a comprehensive method for this type of approach.[39] In this method they use the MRM transitions of known metabolites from online data bases to inform their untargeted analysis. This provided an indication of the partial structure or class of the metabolites in the untargeted analysis, which then allowed the researcher to focus on potentially relevant molecules from data acquired in an untargeted fashion.[39] Analogous approaches were described in Chapter Two, based on the work of Boccard *et al.* (2011) and Palermo *et al.* (2017), that have been previously explained.[15,16] *Approach 2* is like these approaches, as it isolates features based on theoretical Phase I transformations of madol (**M**), and or, from steroidogenesis pathways, based on exact mass matching. However, our approach differs as it utilises MS/MS data for putative annotation of features when possible, giving further information about the unknown features.

As hemapolin is a designer steroid with a 17-methyl group (C20), its metabolites would be distinct from endogenous androgens (C19). Therefore, a list of potential metabolites was made to carry out the pseudo-targeted analysis. This list was already utilised in *Approach 1* and is shown in Tables 5.3 & 5.4. It consisted of two main components. Firstly, a list of possible metabolites was generated by applying

combinations of up to four Phase I metabolic transformations to madol (**M**), including hydroxylation, bond oxidation and reduction, followed the Phase II transformations sulfation or glucuronylation. This approach would capture any additional sulfonate metabolites due to the isobaric nature of madol sulfate and the sulfonate metabolites. A similar method was used in Chapter Two of this thesis and the original GC-MS analysis, Table 5.3.[6,11] Secondly, possible metabolites were generated from the steroidogenesis or related pathways found in literature, and their theoretical masses matched in their free, glucuronylated or sulfated forms, Table 5.4. The rationale for this second list of molecules was based on the known androgenic activity of hemapolin in horses and the potential of this steroid to perturb the metabolic pathways associated with endogenous steroids.[6]

5.2.12 Steroid searcher algorithm and steroidomic OPLS-DA analysis

The two datasets were first processed in an identical fashion to *Approach 1*, resulting in a curated list of 21678 features with MS1 in the positive mode screen, and 3919 features (3348 non-sulfate features and 571 sulfated features) in the negative mode screen. The features were then aligned against the list of molecules given in Tables 5.3 and 5.4, in their respective polarities and relevant ionisation states. The alignment was done by an algorithm written in R, which matched masses within a mass tolerance of ± 5 ppm, identical to the process performed in Chapter Two. This resulted in a filtered dataset with matches to 260/21687 and 106/3919 putative steroidal markers in positive and negative mode, respectively. In negative mode 30 of the 106 molecules were putatively identified as sulfates according to their MS/MS spectra by the *k*-means algorithm. Of these 29 were matched to the theoretical masses of sulfates from Tables 5.4 and 5.5. The last putative sulfated feature coincidently matched the mass for deprotonated cortisol, so was discounted as a steroidal metabolite, however, its information was still retained. It should be noted that matches were also found to the four quantified and validated molecules examined in the first part of this chapter, **SW72**; **SW76**; **E2**; and **E4**. Moreover, matches were found across positive and negative mode, and to most putative metabolites and/or their Phase II metabolites identified in the initial GC-MS analysis.[6] Following the steroid filtering step, both datasets were analysed as in approach 1, with only slight differences. Filtering in data-pre-

treatment was not performed. This was justified due to the relatively small size of the data sets (<500 features), and that they had effectively been filtered by alignment to putative steroids.[13,16,40] The data was then Pareto scaled, $\log_{10}$ transformed and subjected to multivariate analysis. The PCA of both datasets showed clear groupings of the pre- and post-administration samples. In both cases, with three PCs explaining 70% and 74% of the variance in positive and negative mode screens, respectively, Figure 5.12. It should be noted that this is slightly more on both accounts than *Approach 1*. These observations made the data relevant for further supervised analysis using OPLS-DA modelling.[14]

As done in *Approach 1*, modelling the status of both groups was performed using the OPLS-DA function in MetaboAnalyst 5.0, which generated the descriptive model for both positive and negative modes (Figure 5.13).[13] Both models were validated and gave the following performances: Positive mode cross validation was done, giving R2(X) =0.44, R2(Y) = 0.81, and Q2 = 0.79, and permutation test; R2(Y) = 0.99, and Q2 = 0.92 (p-value < 0.001). Negative mode cross validation was done, giving Q2(X) =0.28, R2(Y) = 0.78, and Q2 = 0.75, and permutation test; R2(Y) = 0.99, and Q2 = 0.96 (p-value < 0.001, permutations = 1000). These models provided evidence of distinct clusters of pre- and post-administration samples in both ionisation modes, as indicated by the relative closeness in the R2(Y) and Q2 values, and observed increase in the permutation test.[20,21] Both the models clearly illustrated the chronological based groupings of from the urinary steroidomic group of metabolites.

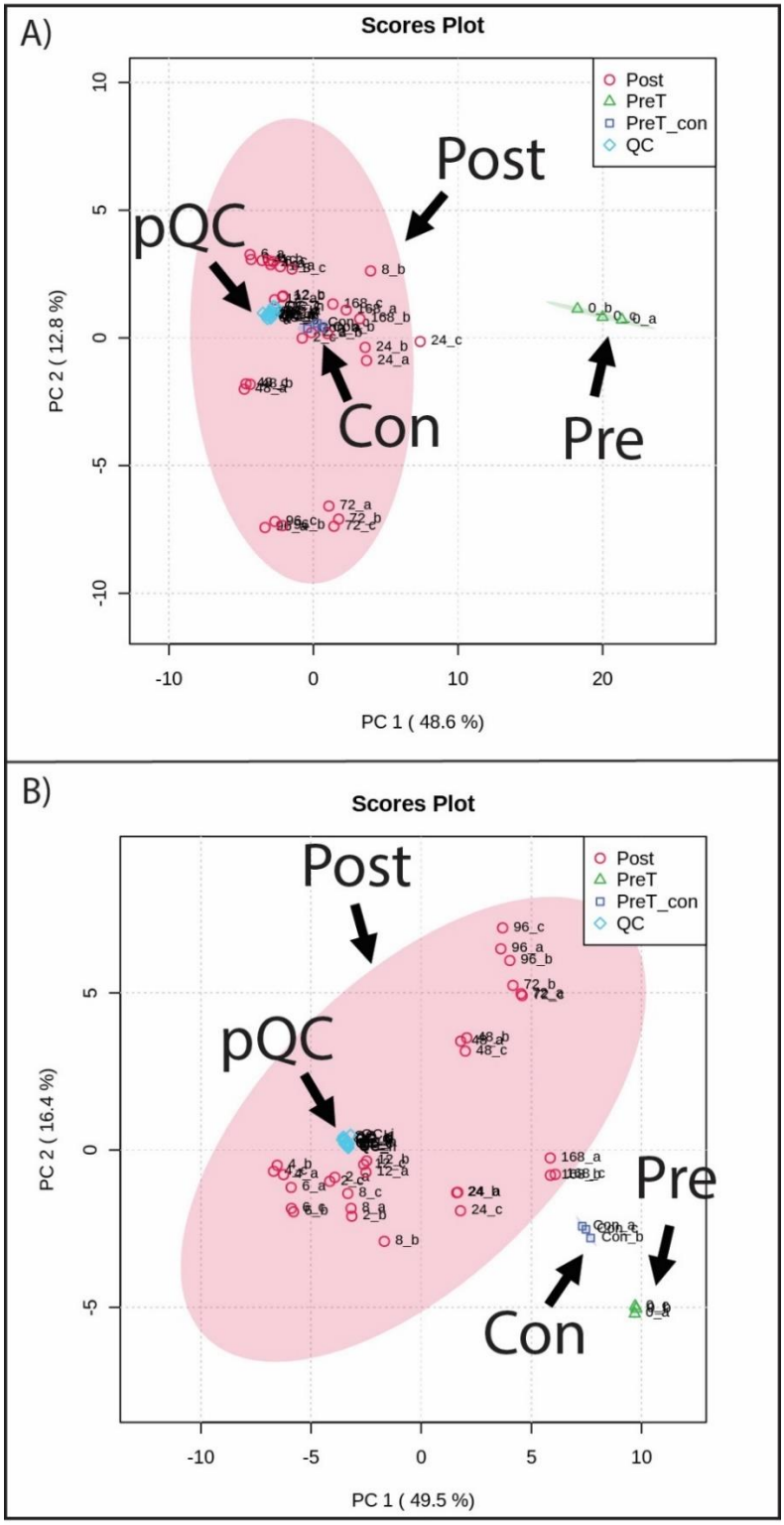


Figure 5.12: Principal Component Analysis (PCA) built from the linear combination of the total kinetic excretion profile (Pre, *n* = 3, green: Post, *n* = 30, pink) as well as a control horse (Con, *n* = 3, blue), and pooled QC samples (QC, *n* = 12, light blue), with total samples, *n* = 48. **A)** PCA score plot of the negative mode screen. Features based on aligned list of theoretical steroid biomarkers. **B)** PCA score plot of the positive mode screen). Ellipses indicate the 95 % CI interval of the groupings

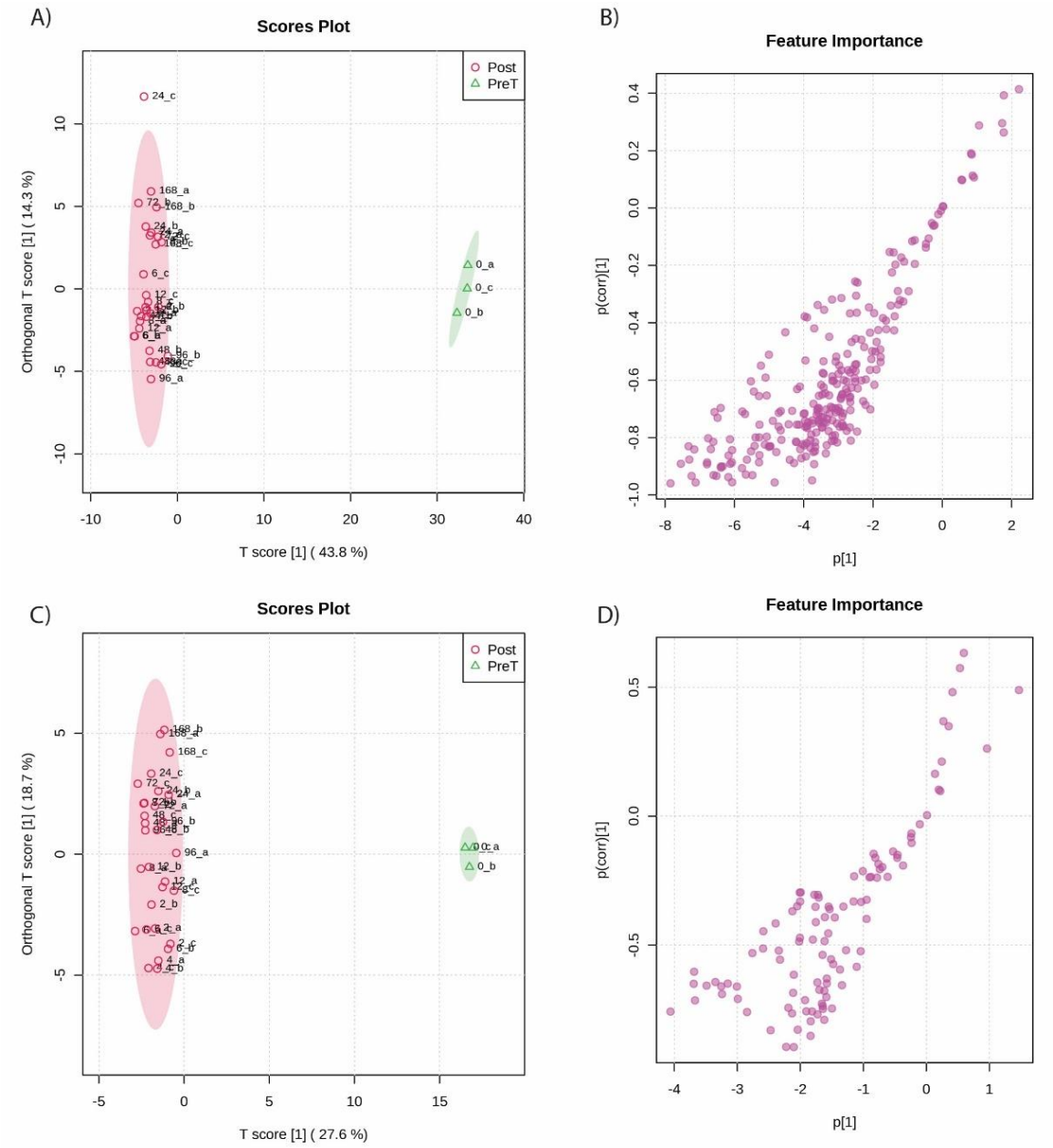


Figure 5.13: The OPLS-DA plots generated between pre- and post-administration samples with associated S-plots. Features based on the aligned list of theoretical steroid biomarkers. Post administration samples, *red*. Pre-administration samples, *green*. **A)** OPLS-DA plot of positive mode base. **B)** S-plot of positive mode features (*n* = 260).**C)** OPLS-DA plot of negative mode. **D)** S-plot of negative mode samples (*n* = 106). Ellipses indicate the 95 % CI interval of the groupings.

5.2.13 Steroidomic analysis of hemapolin administration

The most predictive features were identified and chosen using the associated S-Plots, Figure 5.13. Features were chosen based on the following; an absolute value of 6.5 (positive) and 3 (negative) for p[1] **and** above an absolute value of 0.7 (positive) and 0.5 (negative) for p(corr)[1]. These thresholds were set on a case by case basis as deemed reasonable.[20,21] This allowed the selection of the most predictive features in both data sets, 14 in positive and 10 in negative mode (5 sulfates, 2 glucuronides and 3 non-specified), Tables 5.6 and 5.7.

The main advantage of this semi-targeted steroidomic approach is that it only considers features that already have a putative identification to a biomarker. However, this may not represent the actual identity of these biomarkers. A further manual curation step was performed on the features that were putatively identified as hemapolin metabolites from Table 5.3. These samples were also checked against the control samples. If they were detected in the control sample they were not considered in further analysis, as they were unlikely to be direct hemapolin metabolites that would be detected in an un-doped control horse's urine due to the 17-methyl group present. This resulted in the omission of one feature that was putatively identified as madol from the positive screen and none from the negative screen. The kinetic excretion profiles of the remaining features were then reconstructed and are shown in Figures 5.14 and 5.15.

Table 5.6: The 13 molecules identified by the S-plots in the positive mode screen. The MS/MS spectra from the raw spectra are also shown.

Positive mode						
Theoretical assignment	**Row ID**	**RT [min]**	***m/z***	**p[1]**	**p[corr][1]**	**MS/MS spectra (%RA)**
A1	**6018**	**12.74**	**291.2682**	**-7.3**	**-0.83**	**N/A**
A4	**7104**	**10.56**	**307.2630**	**-6.8**	**-0.89**	**N/A**
A4	**7111**	**13.31**	**307.2638**	**-6.7**	**-0.84**	**N/A**
Hydroxyandrostenone sulfate (e.g. DHEAS, Testosterone S, Epitestosterone S)	**11755****	**7.35**	**369.1734**	**-7.2**	**-0.93**	**369.1728 (100), 253.1937 (41), 133.1010 (47), 121.1008 (27), 61.0094 (37)**
18-Hydroxycortisol	**12624**	**12.71**	**379.2110**	**-6.8**	**-0.8**	**N/A**
***11-Deoxycortisol sulfate *Corticosterone sulfate**	**16144**	**17.23**	**427.1787**	**-7.1**	**-0.96**	**395.7424**
Estradiol glucuronide	**17321****	**10.53**	**449.2152**	**-7.2**	**-0.84**	**N/A**
***Androstane diol glucuronide**	**18181**	**12.60**	**469.2808**	**-7.3**	**-0.88**	**N/A**
6β-Hydroxy androstenedione glucuronide	**18528****	**10.81**	**479.2274**	**-7.5**	**-0.89**	**N/A**
A3G	**18536**	**9.48**	**479.2645**	**-6.6**	**-0.81**	**N/A**
A2G	**18595**	**11.91**	**481.2805**	**-7.8**	**-0.96**	**N/A**
A7G	**19209**	**12.14**	**499.2913**	**-6.8**	**-0.89**	**N/A**
***11-Deoxycortisol glucuronide *Corticosterone glucuronide**	**19815**	**9.26**	**523.2547**	**-6.5**	**-0.93**	**N/A**

G indicates glucuronylation to base structure.

S indicates sulfation to base structure.

Table 5.7: The 10 molecules identified by the S-plots in the negative mode screen. The MS/MS spectra from the raw spectra are also shown.

Negative mode							
Theoretical assignment	**Row ID**	**RT [min]**	***m/z***	**p[1]**	**p[corr][1]**	**MS/MS Annotation**	**MS/MS spectra (%RA)**
A6S	**9524**	**8.40**	**399.1843**	**-3.3**	**-0.64**	**Sulfate**	**399.1860 (55), 96.9598 (100)**
A8S	**10535**	**5.74**	**415.1792**	**-3.5**	**-0.66**	**N/A**	**415.1771 (100), 244.3514 (57)**
A9S	**10684**	**6.22**	**417.1946**	**-3.0**	**-0.66**	**Sulfate**	**417.1955(22), 337.03519(46), 309.03769(4), 281.37479(2), 281.04559(21), 265.04919(7), 246.8703(1), 237.0556(6), 209.0584(2), 192.00571(31), 183.0459(2), 164.01089(100), 163.00281(6), 161.02361(20), 147.0071(2), 136.0155(12), 135.0081(4), 133.0282(7), 119.0143(5), 117.0337(2), 108.0213(10), 96.9596(13),**
A9S	**10686**	**5.55**	**417.1948**	**-4.1**	**-0.76**	**Sulfate**	**417.1948 (100), 96.9598 (66)**
A9S	**10688**	**4.61**	**417.1951**	**-3.69**	**-0.65**	**Sulfate**	**417.19501, 96.9598 (61)**
A9S	**10689**	**6.49**	**417.1952**	**-3.3**	**-0.66**	**Sulfate**	**417.1954 (100), 258.2700 (6), 96.5900 (53)**
Estriol Glucuronide	**12795**	**13.68**	**463.1950**	**-3.2**	**-0.69**	**N/A**	**354.0123 (100)**
A7G	**13975**	**9.22**	**497.2756**	**-3.7**	**-0.60**	**Glucuronide**	**497.2761(100), 319.2258 (6), 157.0152 (8), 113.0235 (15), 85.0264 (3)**
A7G	**13976**	**7.21**	**497.2757**	**-3.2**	**-0.65**	**N/A**	**497.27371 (100), 232.2410 (38)**
A10G	**14256**	**10.96**	**509.2390**	**-3.7**	**-0.71**	**Glucuronide**	**333.2070 (100), 102.4430 (10)**

G indicates glucuronylation to base structure.

S indicates sulfation to base structure.

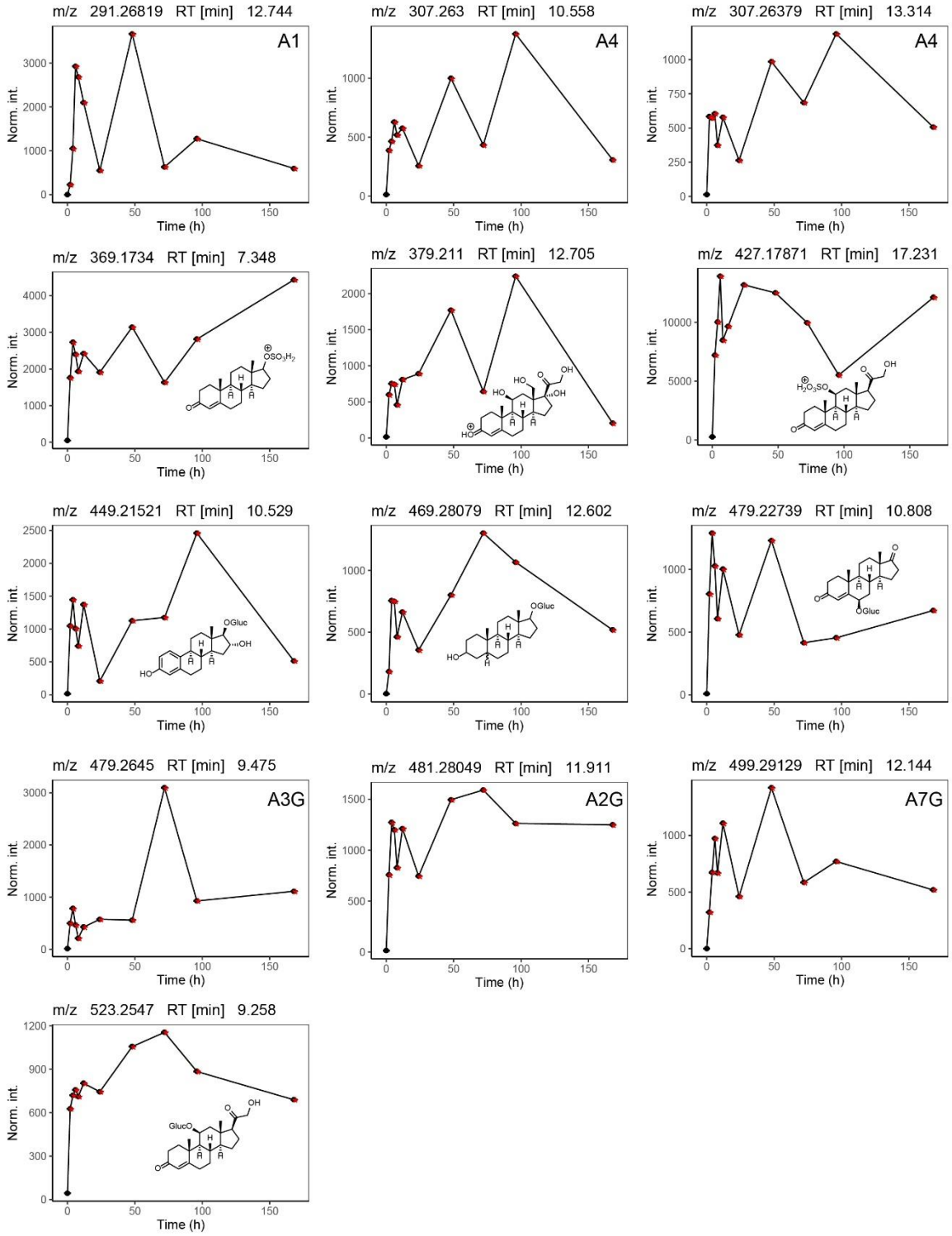


Figure 5.14: Kinetic excretion profiles of 13 of the 14 significant molecules identified in OPLS-DA analysis of the positive mode screen. Norm. Int. = LOWESS normalised intensity. The error bars show standard error, SE = σ/(n)1/2, n = 3. The error bars are shown, however, are obscured due to their small size. The asterisk * indicates a p-value < 0.01, relative to time 0 h. Note; all non 0 h samples were significant.

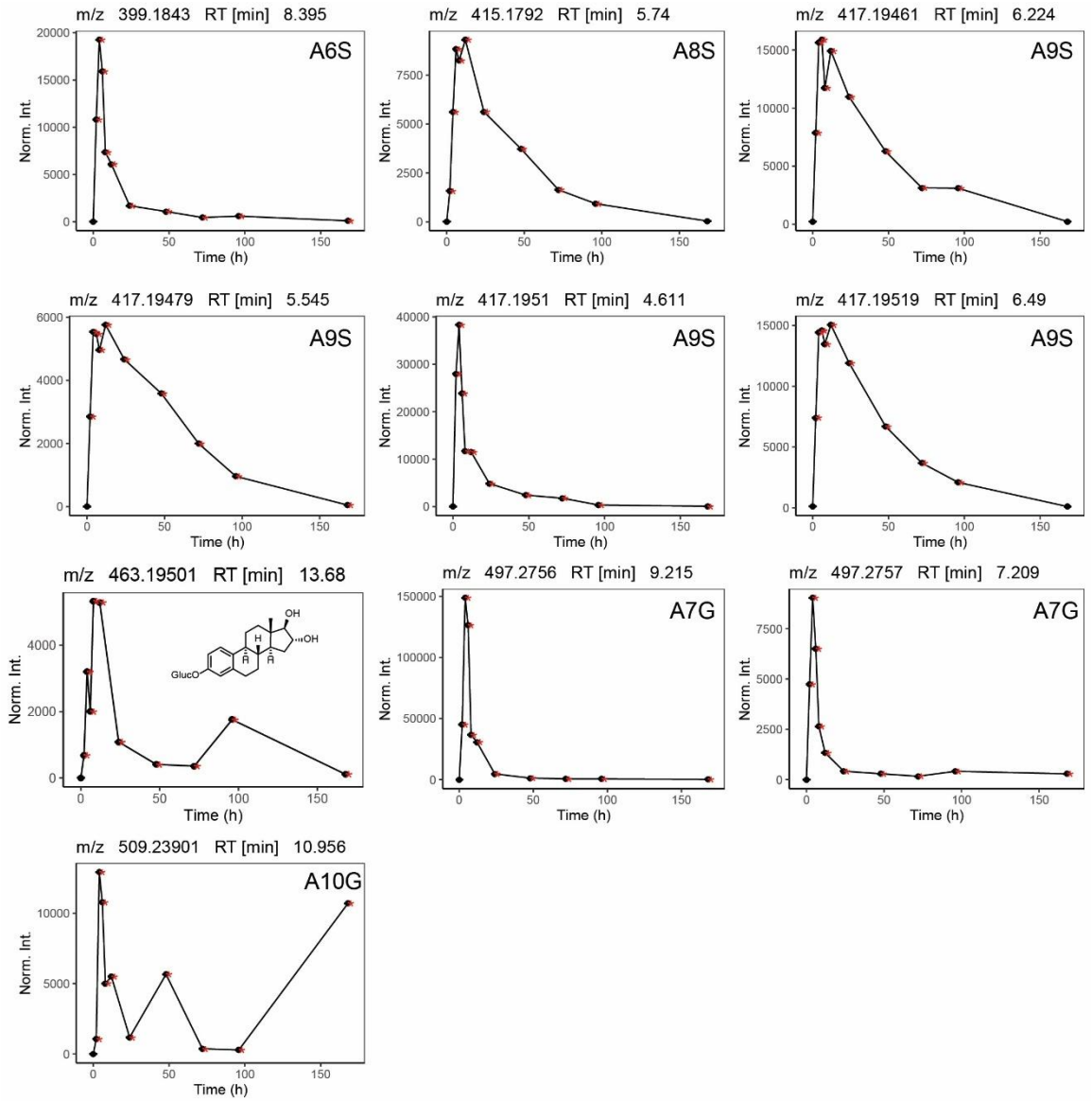


Figure 5.15: Kinetic excretion profiles of the 10 significant molecules identified in OPLS-DA analysis of negative mode. Norm. Int. = LOWESS normalised intensity. The error bars show standard error, SE = σ/(n)1/2, n = 3. The error bars are shown, however, are obscured due to their small size. The asterisk * indicates a p-value < 0.01, relative to time 0 h. Note; all non 0 h samples were significant.

5.2.14 Results of OPLS-DA and biomarker annotation

In the positive mode screen, 6 of the final 13 features were annotated as putative metabolites of madol (*c.f.* hemapolin) and the remaining features were annotated as putative metabolites from the steroid bio-synthetic pathway or associated pathways, Figure 5.14, Table 5.6.

Of the putative hemapolin metabolites, 3 of the 6 features were assigned theoretical masses equivalent to the Phase II glucuronylated metabolites of **T1**, **E4** (or **E2**), and **C4** (or **C2**, or **M1**-**M6**), specifically, **A7G**, **A3G**, and **A2G**, respectively. Of the remaining four putative metabolites, one aligned with the masses for reduced free madol, **A1**, and the remaining three aligned with reduced and hydroxylated madol, **A4**. However, based on the removal of madol, these latter free steroids, **A1** and **A4**, would have to be scrutinised further.

Of the 7 molecules assigned to the steroid biosynthetic pathway, features were matched to the theoretical exact masses of, a sulfated hydroxyandrostenone (e.g., DHEAS, TS, or ETS), an androstanediol glucuronide, 18-hydroxycortisol, estradiol glucuronide, corticosterone sulfate, corticosterone glucuronide, and 6β-hydroxy-androstenedione glucuronide. Interestingly, although highly speculative at this stage of the analysis, many of these markers represent known downstream markers of stress response systems and/or are end products of adrenal steroidogenesis in human systems. [33,41–43] However, less evidence is available of possible effector pathways in horses.

Finally, of the 13 features detected in positive mode, only one possessed comprehensive, MS/MS spectra, *m/z* 369.1734, RT 7.348 min, from the DDA data. This was putatively assigned as an hydroxyandrostenone sulfate, such as DHEAS, TS, EpiTS. Based on the product ions this could be feasible, (*m/z* 369.1728 (100), *m/z* 253.1937 (41), 133.1010 (47), *m/z* 121.1008 (27), *m/z* 61.0094 (37)). With *m/z* 253.1937 matching the exact mass for the carbo-cation of the dehydrated and de-sulfated four ring androstane core, *m/z* 133.1010 matching the C and D ring structure, and *m/z* 121.1008 matching the C and D ring structure without the C19 methyl, Figure 5.16. Despite this, a corresponding peak was not detected in the

negative mode analysis, and therefore, further confirmatory analysis would have to be performed against an authentic reference material.

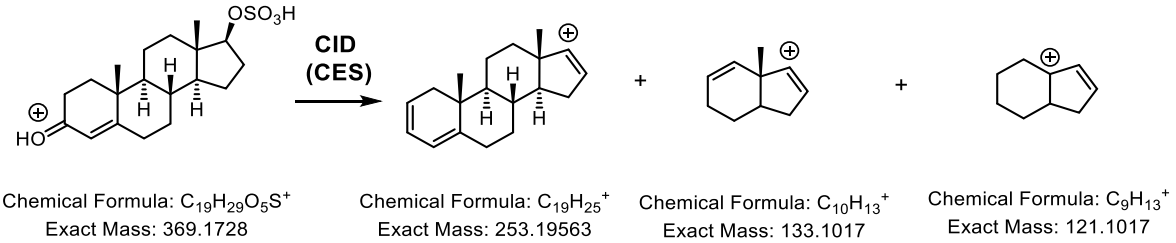


Figure 5.16: Proposed product ions of androstene sulfate, *m/z* 369.1734, RT 7.348 min under CID conditions with the CE spread (CES) function set at 57 ± 30 eV.

In negative mode, 9 of 10 features were identified as putative metabolites of hemapolin and the remaining was putatively identified as **estriol glucuronide**, Figure 5.15, Table 5.7. All features were matched to Phase II conjugates with 6 sulfates and 4 glucuronides putatively assigned. Of the sulfates, 5 were annotated as sulfated by the *k*-means algorithm, and the final molecule had inconclusive MS/MS spectra, Table 5.7. Of the 4 glucuronylated metabolites, 2 were manually annotated as glucuronides based on the observed characteristic MS/MS transitions, as performed in Chapter Two. The remaining 2 had inconclusive MS/MS spectra, which is not surprising considering that this method was not optimised for glucuronides for the negative mode MS/MS.

Many of the features detected in the negative mode screen matched potential Phase II conjugates of the metabolites proposed in the initial GC-MS analysis. The biomarkers, **A6S**, **A7G**, **A9S**, represent either glucuronylated or sulfated conjugates of **C4** (also **C2**, **M1-M6**), **M7-M9**, **T1**, and **M10-M11**, respectively, Figure 5.1. Interestingly, **A7G** was found in the positive and negative mode at similar retention times of 12.144 min & 12.037 min. Moreover, the biomarker **A10G** was also identified in both *Approach 1* and *Approach 2* as a feature of interest by the multivariate analysis. Lastly, the remaining feature was the sulfated metabolite of triply hydroxylated madol, **A8S**. It should be noted that the accurate mass of a hydroxylated and sulfated madol metabolite is isobaric to that of a sulfonated steroid metabolite. These have not been considered as structures but would have to be in future studies for biomarker identification.

5.2.15 Discussion of Approach 2

As in the GC-MS study, 15 of the 23 putatively assigned hemapolin metabolites detected in *Approach 2* provide evidence of extensive metabolic modification. From this an extended metabolic pathway can be drawn with the putative Phase II metabolites added in, Figure 5.17. All these markers identified from the S-plots were upregulated in respect to the pre-administration samples, and potentially highlight the main excretion pathways for hemapolin in equine metabolism, with many of the markers showing increased elevation even at the 1-week time point (168 hours). Although these only represent putative identifications, the increased time frame offered by these biomarkers is substantially longer than the <12 hours found for the **E2**, **E4**, and **SW72** and **SW76** biomarkers. Moreover, they give the researcher a focused group of Phase I & II biomarkers to target when synthesising potential long-term metabolites. This is particularly important for Phase II metabolites as they represent well over half (19/23) of all metabolites of interest. From the putatively assigned hemapolin metabolites, ones to easily pursue in future would be **A7G**, **A3G**, **A2G** and **A1**, as they may represent the conjugated or transformed versions of known hemapolin metabolites that have been synthesised. Further, knowing the absolute identity of **A7G**, **A3G**, and **A2G** would be of particular interest, as their excretion profiles are still highly perturbed up to 168h post administration of hemapolin.

Although the MS/MS analysis gives a promising start towards identifying the biomarkers, the CID conditions were not optimised for glucuronide or free steroid analysis. With improved conditions there is the potential for assignment of the features based on MS/MS transitions. Future improvement could also be gained from the use of data independent acquisition (DIA) or increasing the number of scans in the DDA. This last point is especially salient in the case of the positive mode screen and to a lesser extent for negative mode. Here only two features of the 20 initially selected had MS/MS information. This low precursor selection was likely due to the required feature intensity needed to trigger the DDA being set an order of magnitude too high, something that can readily be fixed in the future.

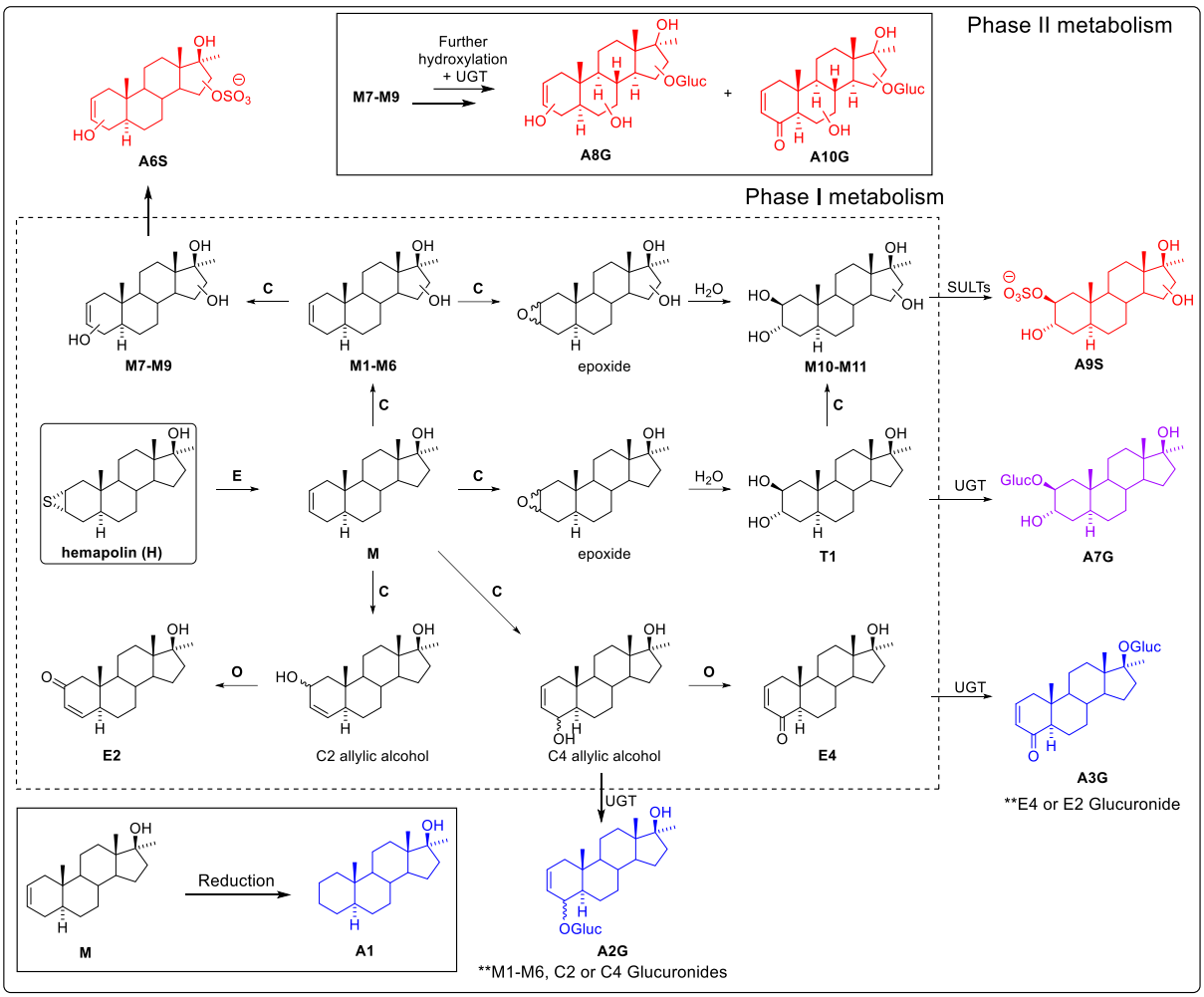


Figure 5.17: Proposed pathways for urinary Phase I & Phase II metabolites of hemapolin in the horse. E, elimination; C, Cytochrome P450 oxidation; O, oxidation, SULTs, sulfotransferase, UGT, UDP-glucuronosyltransferase. Red, blue, and purple structures indicate biomarkers found in negative, positive mode, or both screens, respectively. *O-Gluc*, represents a glucuronide moiety, $C_6H_8O_7$.** Indicates that it could be from multiple Phase I metabolites, with only one example shown. New Phase I metabolites are shown in boxes. Metabolite **A4** is not shown. Figure adapted from Waller *et al* (2020).[6]

5.3 Conclusion

Designer steroids such as hemapolin (H), if left unchecked, have the potential to harm the integrity of sports. This study building on previous ones before it, has added to the overall understanding surrounding the *in vivo* metabolism of hemapolin in the horse.[6] A semi-targeted UHPLC-HRMS/MS approach provided confirmation for a novel Phase I metabolic pathway which leads to the formation of steroid sulfonate metabolites, by quantifying and profiling the metabolites **SW72** and **SW76**. The excretion profiles of the enones, **E2** and **E4**, were also taken, and aligned with the findings from the GC-MS study.[6] The detection time of hemapolin based on these four metabolites was established to be less than 12 hours using LC-MS analysis. These targeted profiles have offered information on key metabolites of hemapolin metabolism using UHPLC-HRMS/MS analysis for the first time, information that can be incorporated into routine anti-doping screening and confirmation profiles.

Next, multivariate analysis was applied to the datasets to identify potential biomarkers affected by hemapolin administration. The non-targeted, *Approach 1*, found 21 possible biomarkers over both positive and negative mode screens, however, only two molecules were putatively annotated as a sulfated and a glucuronylated metabolite. The semi-targeted, *Approach 2*, utilised a steroidomics approach resulting in the putative assignment of 23 steroidal biomarkers, some of which were elevated for more than one week after administration. Of these, 15 were putatively assigned as metabolites of hemapolin and the remaining 8 were conjectured to be hormones related to an elevated stress response. These two multivariate analyses represent a qualitative exploratory examination of the urinary metabolome following hemapolin administration. Further studies would have to be performed to validate the biomarkers and the methodology, before routine application.

5.3.1 Future directions and commentary

Several avenues exist to finalise the work that stemmed from the untargeted metabolomic profiling of the equine urine. Both untargeted approaches were able to identify several upregulated biomarkers that have the potential to increase the detection window of hemapolin administration. In both approaches, identification

through synthetically derived reference materials would be the gold-standard in elucidating these bio-markers further. Being wholly non-targeted in nature, *Approach 1*, allowed the identification of molecules that were the greatest predictors of differences between pre- and post-administration samples. In most cases these biomarkers were elevated well past the detection window of the study. However, as the putative identity of the biomarkers could not be assigned, in such an approach, validation of these signals would have to be done, such as on population studies, or additional *in vivo* studies, to be considered as a tool for doping control.[14]

The semi-targeted *Approach 2* was designed to reduce the number of potential candidates by their accurate *m/z*. Compared to *Approach 1*, this examined molecules with smaller magnitudes of change and significance. Despite this, it resulted in similar overall groupings of the data in the PCA and OPLS-DA. This approach is powerful as it is more likely to find molecules of biological significance, although it sacrifices information about the overall metabolome and may give unreliable matches, as seen from the omission of the putative madol metabolite when compared to the control horse sample. Therefore, *Approach 2* finds its strength in narrowing possible reference materials to a small but significant group of metabolites. Moreover, a longitudinal follow-up of these biomarkers once identified could further reveal their discriminating power with respect to a reference population in the context of routine anti-doping control.

Improvements to both approaches could be made by increasing the use of control samples, this would give an increase to the biological variability of metabolites, as done in Chapter Two. Future improvements could also be made to the DDA parameters. As expected in the DDA approach precursor selection of MS/MS was good for features with high magnitudes. However, features with lower abundances were poorly selected for, especially in the case of *Approach 2*. This could be rectified by lowering the intensity threshold, increasing the number of scans of the DDA or using DIA analysis, as discussed in Chapter Two. Aside from this, a rapid PRM method could also be devised to check the MS/MS spectra of the features identified in *Approach 2* that did not have MS/MS information.

5.4 Experimental

5.4.1 Animal Administration

The following was taken from the previous GC-MS study.[6] The study (RP75) was approved by the Racing NSW Animal Care and Ethics Committee (ARA 62). Animal administration was conducted on a single horse on ethical grounds, to limit exposure to an agent that was not approved for veterinary use, and as it was expected that the metabolites produced from this synthetically derived steroid would be distinct from endogenous metabolites such that the determination of threshold concentrations would not be required. A sample of synthetic **H** (200 mg) was administered orally as a suspension in water by way of a nasogastric tube to a thoroughbred gelding (21 years old,580 kg), with urine samples collected at 0, 1, 2, 4, 6, 8, 12, 24, 48, 72,96, 120, 144, and 168 h post-administration by conditioned spontaneous voiding. The samples were stored at -20°C until required for analysis. Food and water were freely available before and during the drug administration trial.

5.4.2 UHPLC-HRMS DDA analysis for equine urine samples

The hemapolin screen and metabolite confirmations used the following conditions. Analysis was carried out on a 9030 HR-Q-ToF mass spectrometer equipped with a standard ESI source interfaced to a Nexera LC-40D X3 pump system for chromatographic separation (all from Shimadzu). The column used was a Shim-pack Velox C18 column (1.8 μm, 2.1 x 100 mm, Shimadzu). The UHPLC separation was performed at flow rate 0.4 mL/min, using gradient mixing of two phases: Solution A; 20 mM ammonium formate in 100% water; and solution B; 20 mM ammonium formate in 100% methanol. The gradient was 0-0.5 min (10% B), 0.5-20.5 min (10-100% B), 20.5-23.0 min (100% B), 23.0-23.01 min (100-10% B), 23.01-27.5 min (10% B). The injection volume was 5 µL and column oven temperature was 40 ºC.

5.4.3 Parameters for CE spread and DDA MS experiments

The HRMS analysis largely followed the settings used in Chapter Two, however, as positive mode was also used in this study there were slight differences. For HRMS analysis, mass calibration was performed in negative and positive mode: using sodium

iodide (NaI) clusters. Acquisition was performed at a resolution of approximately 30,000 for both scan (*m/z* 100 to 750) and for DDA (*m/z* 50-750). HRMS/DDA was collected between 1 and 23.5 minutes for the top 20 precursors in each survey scan. The event time was set to 0.1 ms (10 Hz) for the scan and 0.035 ms (28.6 Hz) for subsequent DDA scans, with a total loop time (duty cycle) of 0.800 seconds over a total of 21 events with dynamic mass exclusion window of 9 seconds. Collision induced dissociated experiments were performed using the CE spread (CES) function at CE 27 ± 15 eV in positive mode and CE 57 ± 30 eV in negative mode acquisition, with argon used as the collision gas. An exclusion list for precursor ions that did not emanate from the urine samples but were instead observed in extraction blank samples was used. For experiments with QC samples, the injection sequence started by running two blanks followed by running at least five system suitability samples followed by two pooled QC samples, followed by the main sequence. Biological samples were analysed in a random order, with pooled QC samples dispersed evenly throughout. In targeted parallel reaction monitoring (PRM) MS/MS experiments for metabolite confirmation, a list of single *m/z* (mass accuracy ± 5 ppm) was selected for and fragmented over multiple collision energies (CE), for both the reference material and biological samples.

5.4.4 Extraction of equine urine samples for UHPLC-HRMS DDA analysis

The extraction of equine urine samples was based on an established method.[11] To an aliquot of urine (0.6 mL), phosphate buffer (0.3 mL, 100 mM, pH 7 at rt) was added and the samples were centrifuged (3000 x *g*, 5 min) to pellet solids. Supernatant (0.750 mL) was fortified with a mixture of analytical internal standards nandrolone sulfate, d_3-testosterone, d_3-testosterone sulfate, d_3-testosterone glucuronide (final concentration equivalent to 100 ng /mL original urine volume (0.5 mL) per standard). The SPE was performed using Oasis WAX cartridges (3cc), that were pre-conditioned with methanol (2 mL) and water (2 mL). Sample supernatant was than loaded and washed with NaOH (2 mL, 0.1 M), phosphate buffer (2 mL, 100 mM, pH 5.4), and Milli-Q water (2 mL) before elution with a mixture of ethyl acetate: methanol: diethyl amine (25:25:1 v/v/v, 3 mL) into clean 10 mL glass tubes. All samples were then evaporated to dryness under a reduced pressure at 40 °C and stored at -20 °C until analysis. For analysis, the dried samples were re-dissolved in MeOH and filtered using 0.2 µm spin filters,

then dried down under reduced pressure, and reconstituted with filtered MeOH: water (100 µL, 10 % v/v) and stored at 5°C until analysis. This gave samples that were approximately 5 times the concentration of the original urine sample (0.5 mL).

Pooled quality control (pQC) samples were made up of pooled aliquots (0.65 mL) of each biological sample which were re-aliquoted and prepared as above. These were then evenly distributed throughout the sample batch during acquisition.

5.4.5 Untargeted profiling of equine urine and semi-targeted metabolite confirmation studies

Untargeted profiling analysis for all biological samples were run in triplicate with pQC samples dispersed evenly throughout the sequence and analysed in a randomised order. Data alignment and normalisation of total ion chromatograms (TIC) was performed using MS Dial ver. 4.80 with tight tolerances for both retention time (RT) (+/- 6 s) and mass accuracy (*m/z* +/- 5 ppm), alongside the recommended settings for HRMS systems.[44] Normalisation was performed using the pQC samples via the LOWESS function. The data quality and resulting features were assessed through PCA analysis in R. For negative mode screen MS/MS data was processed to identify potential sulfates in python and R.[11] Following this data was then exported to Excel. Data was then pre-treated (filtered, scaled and transformed) before multivariate analysis (PCA and OPLS-DA) was performed in MetaboAnalyst 5.0.[13] Briefly, data filtering, if undertaken, was based on the comparison to the relative standard deviation (RSD, $x < 25\%$) of the pooled QC samples, and based on inter-quantile range as performed in Chapter Two.[18] Curated data from OPLS-DA analysis was then processed in R using the Limma package from *bio-conductor* for univariate statistics.[45] OPLS-DA model validation was done in an identical manner to that described in Chapter Two. The steroid searcher algorithm was written in R as a custom script and matched the *m/z* of detected peaks to the list of theoretical steroidal biomarkers within ± 5 ppm, Tables 5.3 and 5.4. Metabolite confirmation studies were performed using PRM experiments and based on retention time, scan MS accurate mass and MS/MS match to an inhouse synthetic library of standards. Metabolites were matched to reference standards according to AORC retention time and MS/MS criteria.[9]

5.4.6 Linearity

Calibration curves were performed by means of spiking 6 levels (0, 1, 2, 5, 10, 50 ng/mL) in stripped blank matrix and were generated by plotting the EIC peak area ratios of reference materials (**E2**, **E4**, & **SW72**) over internal standard d3-testosterone (positive mode) and nandrolone sulfate (negative mode) (IS concentration = 10 ng/mL). Note due to its large concentration, 10 levels (0, 1, 2, 5, 10, 50, 100, 150, 200, 250 ng/mL) were taken for **SW76**. Linearity performance of the method was checked for each compound and used to define the in vivo ranges of the compounds in urine between a range of 0-200 ng/mL. All linearity equations are given in Table 5.1. The coefficient of determination (r^2) ranged between 0.97-0.99 indicating good linearity for each measured compound in the study. All calibration curves were taken as technical triplicates, with two repetitions occurring before and after the sample batch, and one occurring on a different day to the initial run.

5.4.7 Recovery

Extraction recovery was performed for each measured reference material (**E2**, **E4**, **SW72** & **SW76**) at a concentration of 100 ng/mL. Two separate samples were made up, pre and post addition samples. The general preparation is as follows; to an aliquot of blank urine (0.6 mL), phosphate buffer (0.3 mL, 100 mM, pH 7 at rt) was added and the samples were centrifuged (3000 x *g*, 5 min) to pellet solids. Supernatant (0.750 mL) was then treated to create pre or post addition samples. Pre-addition: the supernatant (0.750 mL) was fortified with a mixture of 100 ng/mL of reference material, the sample was then subjected to WAX SPE and reconstituted as above. Post-addition: the supernatant (0.750 mL) underwent WAX SPE and then fortified with the same mixture of reference materials. In both cases following extraction, eluents were fortified with 100 ng/ mL of internal standard (nandrolone sulfate) before reconstitution. The extraction recovery was calculated as the area ratios (areas: reference material/IS) of the post-addition over that of the pre-addition sample.

5.4.8 Repeatability and reproducibility

Repeatability and reproducibility were calculated in the same manner as was done in Chapter Two.

5.4.9 Sensitivity

Limits of detection (LOD = 3 σ/b)[46] were estimated according to the calibration curves using the concentration ranges of 0, 1, 2, 5, 10, 50 ng/mL, of individual analyte samples as response ratios, where σ is the standard deviation of the regression, b is the slope of the calibration curve.

5.4.10 Accuracy and precision

In negative mode reasonable accuracy (%RE) and precision (%CV) was obtained at approximately the LOQ for the 10 ng/mL samples (QC and calibrators, n = 5 (**SW72**), n = 7 (**SW76**)). These were, for **SW72** %RE ranged between -14% and 4% with a %CV of 7.2%, and for **SW76** %RE ranged between -16% and 3% with a %CV of 12.2%. In positive mode reasonable accuracy (%RE) and precision (%CV) was also obtained over the 10 ng/mL samples (QC and calibrators, n = 5). These were, for **E2** %RE ranged between -16% and -4% with a %CV of 5.1, and for **E4** %RE ranged between -16% and -1% with a %CV of 6.6%. It should be noted that the -16% %RE in both cases emanated from the same QC sample injection and is likely an outlier. Removing this would result in an %RE range of -13% to 3% for **E4** and -11% to 5% for all samples above the LOQ.

5.5 References

1. Abbate, V. *et al.* Anabolic steroids detected in bodybuilding dietary supplements - A significant risk to public health. *Drug Test. Anal.* **7**, 609–618 (2015).

2. Health Fraud Product Database | FDA. https://www.fda.gov/consumers/health-fraud-scams/health-fraud-product-database.

3. Waller, C. C. & McLeod, M. D. A review of designer anabolic steroids in equine sports. *Drug Test. Anal.* **9**, 1304–1319 (2017).

4. De Cock, K. J. S. *et al.* Detection and determination of anabolic steroids in nutritional supplements. *J. Pharm. Biomed. Anal.* **25**, 843–852 (2001).

5. Alquraini, H. & Auchus, R. J. Strategies that athletes use to avoid detection of androgenic-anabolic steroid doping and sanctions. *Mol. Cell. Endocrinol.* **464**, 28–33 (2018).

6. Waller, C. C. *et al.* In vivo metabolism of the designer anabolic steroid hemapolin in the thoroughbred horse. *Drug Test. Anal.* **12**, 752–762 (2020).

7. Okano, M., Sato, M., Ikekita, A. & Kageyama, S. Analysis of non-ketoic steroids 17α-methylepithiostanol and desoxymethyltestosterone in dietary supplements. *Drug Test. Anal.* **1**, 518–525 (2009).

8. Okano, M., Sato, M., Kojima, A. & Kageyama, S. Determination of mepitiostane metabolites in human urine by liquid chromatography/tandem mass spectrometry for sports drug testing. *J. Pharm. Biomed. Anal.* **115**, 236–244 (2015).

9. Association of Offical Racing Chemist. AORC MS Criteria (modified 23 Aug 16) - Association of Official Racing Chemists. http://www.aorc-online.org/documents/aorc-ms-criteria-modified-23-aug-16/.

10. Weththasinghe, S. A. Synthesis and in vitro metabolism studies of selected steroids for anti-doping analysis. (Australian National University, 2020).

11. Fitzgerald, C. C. J. *et al.* Profiling Urinary Sulfate Metabolites With Mass Spectrometry. *Front. Mol. Biosci.* **9**, 829511 (2022).

12. Horwitz, W. & Albert, R. The Horwitz ratio (HorRat): A useful index of method performance with respect to precision. *J. AOAC Int.* **89**, 1095–1109 (2006).

13. Pang, Z. *et al.* MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, 388–396 (2021).

14. Cloteau, C. *et al.* From a non-targeted metabolomics approach to a targeted biomarkers strategy to highlight testosterone abuse in equine. Illustration of a methodological transfer between platforms and laboratories. *Drug Test. Anal.* https://doi.org/10.1002/dta.3221 (2022) doi:10.1002/dta.3221.

15. Palermo, A., Botrè, F., de la Torre, X. & Zamboni, N. Non-targeted LC-MS based metabolomics analysis of the urinary steroidal profile. *Anal. Chim. Acta* **964**, 112–122 (2017).

16. Boccard, J. *et al.* A steroidomic approach for biomarkers discovery in doping control. *Forensic Sci. Int.* **213**, 85–94 (2011).

17. Broadhurst, D. *et al.* Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **14**, 72 (2018).

18. Hackstadt, A. J. & Hess, A. M. Filtering for increased power for microarray data analysis. *BMC Bioinformatics* **10**, 1–12 (2009).

19. Kaabia, Z. *et al.* Monitoring the endogenous steroid profile disruption in urine and blood upon nandrolone administration: An efficient and innovative strategy to screen for nandrolone abuse

in entire male horses. *Drug Test. Anal.* **6**, 376–388 (2014).

20. Worley, B. & Powers, R. PCA as a Practical Indicator of OPLS-DA Model Reliability. *Curr. Metabolomics* **4**, 97–103 (2016).

21. Szymańska, E., Saccenti, E., Smilde, A. K. & Westerhuis, J. A. Double-check: Validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **8**, 3–16 (2012).

22. Hisahiko, S. 19-Hydroxyandrostenedione and 6β-hydroxyandrostenedione: New steroids regulated by the renin-angiotensin system in man. *J. Steroid Biochem.* **20**, 383–385 (1984).

23. Dumasia, M. C. & Houghton, E. Studies related to the metabolism of anabolic steroids in the horse: The identification of some 16-oxygenated metabolites of testosterone and a study of the phase II metabolism. *Xenobiotica* **11**, 323–331 (1981).

24. Hungerford, N. L., McKinney, A. R., Stenhouse, A. M. & McLeod, M. D. Selective manipulation of steroid hydroxyl groups with boronate esters: efficient access to antigenic C-3 linked steroid–protein conjugates and steroid sulfate standards for drug detection. *Org. Biomol. Chem.* **4**, 3951–3959 (2006).

25. McKinney, A. R., Suann, C. J. & Stenhouse, A. M. Detection of 17α-hydroxyprogesterone caproate in equine plasma by gas chromatography/tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **20**, 1855–1858 (2006).

26. Chavatte, P., Holtan, D., Ousey, J. C. & Rossdale, P. D. Biosynthesis and possible biological roles of progestagens during equine pregnancy and in the newborn foal. *Equine Vet. J. Suppl.* **29**, 89–95 (1997).

27. Ghayee, H. K. & Auchus, R. J. Basic concepts and recent developments in human steroid hormone biosynthesis. *Rev. Endocr. Metab. Disord.* **8**, 289–300 (2007).

28. Short, R. V. delta5 3beta-Hydroxy steroids in the follicular fluid of the mare. *J. Endocrinol.* **23**, 277–283 (1961).

29. Cameron, E. H. & Grant, J. K. Biochemistry of histologically defined zones in the adrenal cortex: cortisol synthesis in the horse. *J. Endocrinol.* **37**, 413–420 (1967).

30. El-Sheikh Ali, H. *et al.* Steroid synthesis and metabolism in the equine placenta during placentitis. *Reproduction* **159**, 289–302 (2020).

31. Kirchmeier, A., van Herwaarden, A. E., van der Kolk, J. H., Sauer, F. J. & Gerber, V. Plasma steroid profiles before and after ACTH stimulation test in healthy horses. *Domest. Anim. Endocrinol.* **72**, 106419 (2020).

32. Mostaghel, E. A. Beyond T and DHT - Novel steroid derivatives capable of wild type androgen receptor activation. *Int. J. Biol. Sci.* **10**, 602–613 (2014).

33. Schiffer, L. *et al.* Human steroid biosynthesis, metabolism and excretion are differentially reflected by serum and urine steroid metabolomes: A comprehensive review. *J. Steroid Biochem. Mol. Biol.* **194**, 105439 (2019).

34. Storbeck, K.-H. *et al.* Steroid Metabolome Analysis in Disorders of Adrenal Steroid Biosynthesis and Metabolism. *Endocr. Rev.* **40**, 1605–1625 (2019).

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

36. Chemical Entities of Biological Interest (ChEBI). https://www.ebi.ac.uk/chebi/init.do.

37. Brown, M. *et al.* Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst* **134**, 1322–1332 (2009).

38. Kind, T. *et al.* Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom. Rev.* **37**, 513–532 (2018).

39. Xu, J. *et al.* Development of a metabolic pathway-based pseudo-targeted metabolomics method using liquid chromatography coupled with mass spectrometry. *Talanta* **192**, 160–168 (2019).

40. Chong, J. & Xia, J. MetaboAnalystR: An R package for flexible and reproducible analysis of metabolomics data. *Bioinformatics* **34**, 4313–4314 (2018).

41. Zuidema, L. J., Khan-Dawood, F., Dawood, M. Y. & Work, B. A. Hormones and cervical ripening: Dehydroepiandrosterone sulfate, estradiol, estriol, and progesterone. *Am. J. Obstet. Gynecol.* **155**, 1252–1254 (1986).

42. Lievertz, R. W. Pharmacology and pharmacokinetics of estrogens. *Am. J. Obstet. Gynecol.* **156**, 1289–1293 (1987).

43. Lenders, J. W., Williams, T. A., Reincke, M. & Gomez-Sanchez, C. E. 18-Oxocortisol and 18-hydroxycortisol: Is there clinical utility of these steroids? *Eur. J. Endocrinol.* **178**, R1–R9 (2018).

44. Tsugawa, H. *et al.* MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **12**, 523–526 (2015).

45. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, 47–48 (2015).

46. Food and Drug Administration. *Guidance for Industry Bioanalytical Method Validation Guidance for Industry Bioanalytical Method Validation*. *Biopharmaceutics Federal register* http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htmand/orhttp://www.fda.gov/AnimalVeterinary/GuidanceComplianceEnforcement/GuidanceforIndustry/default.htm (2018).

CHAPTER SIX

Conclusions and Commentary on Further Studies

6.1 Conclusion

Steroids are a broad and complex class of chemical compounds that widely occur in nature as lipids, hormones, and other natural products. The development of new techniques to detect and study steroids and their Phase II conjugate metabolites are important in many fields. Alongside this, the use of metabolomic techniques has become an increasingly relevant and powerful tool to investigate biological questions. This thesis aimed to further the understanding of anabolic androgenic steroid metabolism in humans and equines. A central theme in this thesis has been around the detection of Phase II conjugated metabolites, specifically, steroid sulfates, and more broadly, the sulfated metabolites in a urinary matrix. One of the main ways this has been achieved is through the development of new software tools and UHPLC-HRMS/MS techniques. These have been approached in a cross disciplinary manner, integrating metabolomics, analytical and synthetic chemistry. Contextually, the methodologies in this thesis have largely been applied as a tool in anti-doping analysis. As introduced in Chapter One, the science of sports drug testing is an all-encompassing and fascinating field and represents a relentless and ongoing effort to combat doping in sport. It is the hope of the author that the methodologies and contributions developed in this thesis, have not only added to the richness of the anti-doping field, but has also contributed to related fields such as clinical diagnosis in medicine.

Chapter Two introduces new techniques for the untargeted profiling of conjugated metabolites in urine. The methodology developed uses a new informatic workflow to identify sulfated metabolites from all known sulfate-derived MS/MS fragments appearing in a metabolomic dataset that is acquired in an untargeted fashion using DDA. Features of importance are highlighted using a combination of univariate statistics and the sulfate conjugate annotation. This methodology is powerful as it allows for the annotation of putative sulfated metabolites in a metabolome without the loss of information of overall features. The versatility allowed the workflow to be applied in various ways. Firstly, it was used to investigate the *in vivo* metabolites of testosterone propionate in a thoroughbred gelding. Here it considerably decreased the investigative work required by the analyst, by narrowing down the metabolome to a list of candidate metabolites, resulting in the discovery of ten putative biomarkers, with the

eventual confirmation of four of these against authentic synthesised reference materials. Secondly, the workflow was used in a qualitative manner, to investigate the selectivity of different sulfatase enzymes on the human urinary metabolome, allowing for the assessment of the performance of these enzymes over thousands of candidate metabolites, in a single assay. Lastly, the workflow was extended to a semi-targeted analysis, surveying the change of steroidal metabolites in the urinary metabolome following the administration of testosterone propionate to three thoroughbred geldings. This last application identified several putative steroidal markers related to doping among the three horses. The annotation and rapid narrowing down of a metabolome provided a clear way forward for future identifications of these metabolites and offers a more general examination of temporal changes to the overall urinary metabolome at the metabolite level following drug administration.

In Chapter Three a new method for the synthesis of stable isotope labelled (SIL) sulfate esters, was reported. The simple one-pot synthesis of these stable isotopes allowed for their incorporation in a simple and selective way to free hydroxyl groups along the steroid backbone. The SIL incorporation was also found to be suitable for these molecules to be used as SIL-internal standards, as seen in practice in Chapter Two. The selective placement of these labels enabled the analysis of steroid bis(sulfate) fragmentation pathways in unprecedented detail for the first time. This revealed the origin of a range of different sulfate derived fragments, including, hydrogen sulfate, hydrogen sulfite, sulfur trioxide radical anion, and sulfur trioxide, in these species. Further, energy resolved collision induced dissociation analysis of both monosulfate and bis(sulfate) compounds was performed for the first time. This found a distinct relationship between the energy needed to fragment a sulfate ester bond and its orientation on the steroid backbone. This provided the blueprint for the development of methodology for a purely MS/CID based structure determination of sulfated compounds.

Identification of new biomarkers discovered by metabolomic analysis remains a significant hurdle, not only in the field of anti-doping but also in the fields of medical diagnosis and metabolomics. Usually, identification remains as the final and crucial step for obtaining mechanistic insights into cellular processes and confirming any hypothesis or conjectures. The complexity of the metabolome combined with the vast

number of unknown metabolites that have not yet been structurally identified, makes this one of the most difficult obstacles to an analyst. Alongside this, a combination of variables exacerbates this problem. Firstly, the biomarkers in question usually exist at low ng/mL concentrations making spectroscopic techniques such as NMR and X-ray crystallography unattainable. Secondly, the lack of authentic reference materials and synthetic techniques to access said materials. Although, there exist a growing number of online databases, these usually only offer presumptive matches based on MS or MS/MS data.

Chapter Four presents work that brings together findings and methodology developed in Chapters Two and Three. Specifically, a wholly UHPLC-HRMS/MS based approach to distinguish between regio- and stereo-isomeric steroid sulfates is presented. This method takes the energy resolved fragmentation CID analysis developed in Chapter Three and applies it to the analysis of a novel *in vivo* steroid sulfate biomarker discovered in Chapter Two. The methodological jump between chapters is significant. In Chapter Three, energy resolved analysis was performed by direct infusion of high concentration (at approximately 1 mg/mL) samples straight into the MS source. This is contrasted with the methodology used in Chapter Four, that chromatographically resolved compounds using UHPLC, at *in vivo* concentrations (ng/mL) of the biomarker. Further, the energy resolved analysis was compared against a library of related but simpler steroid reference materials that ultimately allowed for the determination of the location and stereochemistry of the conjugation site of the new biomarker. This led to the unambiguous identification of the new biomarker through an informed synthetic route. This energy-resolved CID MS/MS method has wider potential to aid in the structure elucidation of new biomarkers *in vivo* at low concentrations for other conjugate metabolite classes where access to suitable diverse collections of reference materials is accessible.

Chapter Five uses the techniques developed throughout the previous chapters and details a study of the *in vivo* metabolism of the designer steroid hemapolin, using UHPLC-HRMS/MS metabolomic analysis. A semi-targeted approach was adopted like the one presented in Chapter Two. Firstly, the analysis confirmed the presence of two new metabolites **SW72** and **SW76** in negative mode. These metabolites are the first recorded cases of a new type of Phase I metabolite, called sulfonates. Along with

these sulfonate metabolites, the enone metabolites, **E2** and **E4** were also confirmed. These metabolites can now be incorporated into routine anti-doping screening and confirmation protocols for future detection. The untargeted portion of the study also identified putative mid- to long-term Phase II metabolites of hemapolin. However, further confirmation work will have to be performed to gain further insight into their possible usefulness as biomarkers.

In summary, the work presented in this thesis contributes to several different yet important areas of anti-doping research. It addresses the major issue of merging targeted and untargeted UHPLC-HRMS/MS analytical techniques to gain insight into intact Phase II urinary metabolites, such as steroid sulfates, in an anti-doping context. Through the development of a new workflow that annotates sulfate conjugates in an untargeted fashion the capability exists to elucidate the perturbation of the urinary metabolome at the level of individual metabolites. This has allowed for a quick and accessible methodology to qualitatively survey the urinary metabolome for the identification of new and meaningful sulfated biomarkers. Further, the development of new synthetic methodologies for the incorporation of stable isotope labels has provided a suitable synthetic method to readily access compounds for use as SIL-internal standards. This also allowed the comprehensive examination of the fragmentation of sulfated metabolites, leading to the feasibility of wholly MS-based elucidation techniques. This thesis has detailed how using complementary non-targeted and targeted techniques in an interdisciplinary manner led to an overall increase in successful identifications of several new biomarkers related to doping, as shown with hemapolin.

6.2 Future directions and commentary

The methodologies described in this thesis require further development and refinement but have shown promise in realising the overall goal of the untargeted profiling and identification of Phase II AAS urinary metabolites. Various avenues remain for each study presented in this thesis; the following commentary is to offer insight towards further directions of the research.

Steroid profiling forms a cornerstone of both the anti-doping and clinical chemistry fields, with each field having a history of contributing to one another.[1] In the human field this has led to a transition towards personalised athlete steroid profiles as seen in WADA's ABP.[2,3] Although effective, metabolomic analysis still represents an important complementary analytical tool to the targeted approaches offered in the modules that support the ABP. In particular, the rise of steroidomics in clinical research represents a promising avenue for the adaptation to anti-doping analysis.[4,5] However, even these approaches tend to ignore Phase II conjugate metabolites of steroids. The research method presented in Chapter Two and Chapter Five, was demonstrated to be a feasible tool for the qualitative assessment of Phase II steroid sulfate metabolites in the urinary metabolome for both endogenous and designer AAS. Another way this could be used is to study steroid metabolic modulators, such as adrenocorticotropic hormones (e.g. *Synacthen*).[6] In these studies the methodology developed in this thesis could be used to look for perturbations in the urinary metabolome in an untargeted metabolomic or steroidomic focused analysis.

Another strong avenue of future work would be the development of the MS/MS annotation tool to encompass more than one type of conjugate. A strong candidate is the glucuronide conjugates. However, other examples could include amino acids, glutathione, and phosphate metabolites. The aptness for the glucuronide moieties to be annotated in the described manner has already been shown in the studies presented in Chapter Two and Five. In these chapters, glucuronylated metabolites were putatively identified through a combination of retrospective analysis and manual assessment of their characteristic MS/MS spectral data. Albeit, this latter technique was initially conceived using low resolution mass spectrometry by Fabregat and Pozo *et al. (2013)*.[7] The automation of this process for glucuronides, as has been done for sulfates, could feasibly be achieved by altering a handful of lines of code within the analytical data pipeline developed in this thesis.

This direction would provide rich datasets and potentially have the capacity to pre-annotate 97% of conjugate Phase II metabolites in the steroid urinary metabolome, while simultaneously maintaining all data (MS and MS/MS) on features that do not fall in this category, that being glucuronylated and/or sulfated metabolites.[8] Moreover, glucuronides have the added benefit of being detected in both positive and negative

mode polarities. If pursued, this would be a very powerful tool as it would give the analyst an annotated metabolome based on the observed MS attributes of a feature, this is different from the a priori approaches used to identify features using online databases.

Along with the promise, there are several foreseeable difficulties with pursing this type of 'all-encompassing' MS/MS based conjugate annotation technique. The largest would be the optimisation of MS/MS collision energies for both glucuronide and sulfate conjugate metabolites. It has been shown that the fragmentation energies of the two species are vastly different, with glucuronides usually undergoing full dissociation at lower energies compared to sulfates.[8] This could potentially be abridged with the collision energy spread function, as used in Chapters Two and Five. Further, a dual polarity mode could be used, where glucuronides could be detected in positive mode and sulfates in negative. Another possible restraint would be achieving full acquisition of both glucuronides and sulfates, this could be feasibly done by the adoption of DIA, however, more work must be done to realise this. Lastly, the capacity for the general profiling of Phase II steroid conjugates could be greatly improved with the establishment of a repository of MS and MS/MS data. Such data would have to be acquired in a standardised way using synthesised reference materials, however, could easily be added to already existing online data bases such as MassBank or ChEBI.[9,10]

In the clinical, forensic and anti-doping fields, the use of internal standards for the confirmation of a biomarker's absolute identify is considered the industry gold standard and non-negotiable.[4,8,11–14] Stable isotope labelled internal standards of Phase II conjugates are still largely provided by conjugation of a deuterated steroidal biomarker.[15,16] This process is expensive and often can result in chromatographic differences of the standard to the metabolite.[17] In Chapter Three a solution to this problem was presented in the form of a simple and readily accessible 'single-pot' incorporation of an SIL sulfate group. Moreover, in other related work in his honours work the author developed methodology for an analogues technique in glucuronide conjugates by the incorporation of ^{13}C-labelled glucuronide moieties onto a steroid backbone, that was not discussed in this thesis.[8] These two techniques are both selective and economical, however, they may need considerable advancement to allow for scalability. This type of conjugate SIL label could feasibly be combined with

deuterated steroids to study the relationship between sulfation and de-sulfation in the body. Further, the selective nature of these labelling techniques also allowed for the in-depth analysis of the relatively understudied steroid bis(sulfate) metabolites. Initial pilot CID analyses have been carried out for selectively labelled bisglucuronides, however, not to the same extend as were for sulfates in this thesis.[8] The study of the fragmentation patterns of SIL-bisglucuronides would also provide an interesting opportunity for future studies derived from this work.

Finally, in Chapter Four we reported a novel application for energy-resolved CID experiments for the structure elucidation of sulfate ester conjugate site and stereochemistry in a new steroid sulfate biomarker. This methodology demonstrated that a library of structurally simple reference materials can be used as a proxy to ascertain the conjugation site of an unknown biomarker. This technique has wider potential to provide a wholly MS-based ‘first-pass’ for the identification of unknown biomarkers. Moreover, it could be extended to other classes of metabolites. For instance, the conjugation site of steroid glucuronides could feasibly be analysed using this methodology. One caveat being that this methodology may have limitations for biomolecules with multiple fragmentation pathways as seen in the bis(sulfate) example in Chapter Three, and largely depending on how separable the different fragment ions are from one another. Ultimately, it is hoped that this technique will offer valuable information on new biomarkers and offer a viable path forward in the discovery and identification process surrounding them.

6.3 References

1. Cawley, A. T. & Keledjian, J. Intelligence-based anti-doping from an equine biological passport. *Drug Test. Anal.* **9**, 1441–1447 (2017).

2. PE Sottas, M. S. C. S. Endogenous steroid profiling in the athlete biological passport. *Endocrinol Metab Clin* **39**, 59–73 (2010).

3. Schobinger, C., Emery, C., Schweizer-Gründisch, C. & Kuuranne, T. Support of a laboratory-hosted Athlete Biological Passport Management Unit (APMU) to the anti-doping organisations. *Rechtsmedizin* vol. 31 526–532 (2021).

4. Anh, N. H. *et al.* Steroidomics for the prevention, assessment, and management of cancers: A systematic review and functional analysis. *Metabolites* **9**, https://doi.org/10.3390/metabo9100199 (2019).

5. Eisenhofer, G., Durán, C., Chavakis, T. & Cannistraci, C. V. Steroid metabolomics: machine learning and multidimensional diagnostics for adrenal cortical tumors, hyperplasias, and related disorders. *Curr. Opin. Endocr. Metab. Res.* **8**, 40–49 (2019).

6. Fragkaki, A. G., Kioukia-Fougia, N., Kiousi, P., Kioussi, M. & Tsivou, M. Challenges in detecting substances for equine anti-doping. *Drug Test. Anal.* **9**, 1291–1303 (2017).

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

8. Pranata, A. *et al.* Synthesis of steroid bisglucuronide and sulfate glucuronide reference materials: Unearthing neglected treasures of steroid metabolism. *Steroids* **143**, 25–40 (2019).

9. Chemical Entities of Biological Interest (ChEBI). https://www.ebi.ac.uk/chebi/init.do.

10. Horai, H. *et al.* MassBank: A public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **45**, 703–714 (2010).

11. Wudy, S. A. Synthetic procedures for the preparation of deuterium-labeled analogs of naturally occurring steroids. *Steroids* **55**, 463–471 (1990).

12. De Wilde, L., Roels, K., Van Renterghem, P., Van Eenoo, P. & Deventer, K. Steroid profiling in urine of intact glucuronidated and sulfated steroids using liquid chromatography-mass spectrometry. *J. Chromatogr. A* **1624**, 461231 (2020).

13. Stokvis, E., Rosing, H. & Beijnen, J. H. Stable isotopically labeled internal standards in quantitative bioanalysis using liquid chromatography/mass spectrometry: necessity or not? *Rapid Commun. Mass Spectrom.* **19**, 401–407 (2005).

14. Cajka, T. & Fiehn, O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal. Chem.* **88**, 524–545 (2016).

15. Bong-Chul, C., Mallamo, J. P., Juniewicz, P. E. & Shackleton, C. H. L. Synthesis of deuterium-labeled 5α-androstane-3α,17β-diol and its 17β-glucuronide. *Steroids* **57**, 530–536 (1992).

16. Wudy, S. A., Schuler, G., Sánchez-Guijo, A. & Hartmann, M. F. The art of measuring steroids: Principles and practice of current hormonal steroid analysis. *J. Steroid Biochem. Mol. Biol.* **179**, 88–103 (2018).

17. Berg, T. & Strand, D. H. 13C labelled internal standards-A solution to minimize ion suppression effects in liquid chromatography-tandem mass spectrometry analyses of drugs in biological samples? *J. Chromatogr. A* **1218**, 9366–9374 (2011).