EQUINE UDP-GLUCURONOSYLTRANSFERASES AND THEIR ROLE IN PHASE II METABOLISM

Natalie Dawn Pickwell, MSc, BSc.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

September 2017

If you can’t fly, then run.

If you can’t run, then walk.

If you can’t walk, then crawl,

but by all means, keep moving.

Martin Luther King Jr

Acknowledgements

Firstly, a dedication to my supervisor and mentor, Tracey Coffey. Without your encouragement, I would not have chosen the path of a PhD, and without your support and guidance I would never have finished it. You have been such a positive influence, both professionally and personally, providing an environment where I can cultivate self-confidence and scientific knowledge whilst inspiring me to try new experiences. For that you have my gratitude. Thank you.

My thanks to Adam Blanchard for being so generous with your knowledge and time, your instruction and friendship have been invaluable. My gratitude also to Ceri Staley, all the hugs, tea and sympathy helped me stay the course. To Emma Drinkall, for your encouragement, support, understanding, coffee and laughter, I thank you. It has been my privilege to work with such wonderful people.

To the friends that are truly family; Denis Smith Jr, Raphaelle Barbier-Saint-Hillaire, Russell Carter, Georgia Horth, Sammy Aboushoushah and Paul J Robinson. University brought you all into my life and though we have all grown to walk separate paths, your friendship has remained steadfast. You have cheered from the side lines, picked me up with words of encouragement and tears of laughter. Through the trials and tribulations, you have heartened me, without judgement you have given unconditional support. Though a PhD can be at times lonely, I have never been alone. My heartfelt thanks to each of you.

Finally, to my family, there are no words to fully thank you for your support over the last four years, without which I would not have made it through, it has been quite the journey. To my strong, inspirational aunt, Judith Russell, for always being in my corner, believing in me, encouraging me to keep striving forwards, for being free with your advice and time, all the phone calls really did make me feel better. Thank you. To Auntie Chris, thank you for the unconditional support and for being the voice of reason. Lastly, to my father, Keith Pickwell. You have taught me to always

work hard and to the best of my ability. Throughout the course of my education, you have been my rock, encouraging me in moments of despair, you taught me to demolish insurmountable walls one brick at a time. Your constant, unwavering support has been priceless.

To all, I give you my gratitude.

I dedicate this thesis to my Grandad, Ivan Pickwell. Your eyes shone with pride with every academic achievement, and I remember clearly the look on your face when I told you I would be returning to education to gain my doctorate. Though you are no longer here to see me complete the process, I hope that you would be proud.

Abstract

Metabolism is an essential chemical process and pathway involving multiple reactions. Oxidative or functionalisation reactions dominate phase 1 metabolism and is predominately controlled by the Cytochrome P450s (P450s). Phase 2 reactions are frequently referred to as the detoxification, or elimination, phase. Several families of enzymes are involved, and the largest of these are the Uridine diphosphate 5'-glucuronosyltransferases (UGTs). The purpose of this project was to initiate the development of an equine *in vitro* toolbox, concentrating on the UGTs. This required identification of *UGT*s in the equine genome. The first step used syntenic analysis which enabled us to utilise relative gene order conservation between species to determine whether the predicted gene encoded a member of the UGT superfamily. Further analysis of sequence relationships provided confidence that the genes under investigation were *UGTs*, but also allowed us to determine which *UGT* orthologue we were investigating. PCRs were performed to isolate the genes, and subsequent sequencing enabled the *UGTs* to be investigated for key features, including signal peptides, signature sequences, transmembrane domains, and dilysine repeats, which are characteristic of this family of membrane-bound proteins.

We isolated and characterised five putative equine *UGTs*. Subsequent analyses indicated these to be orthologous to human *UGT1A6*, *UGT2A3*, *UGT2B17*, and two *UGTs* orthologous to *UGT3A2.* Three equine UGT genes were cloned into a vector for the development of functional recombinant proteins. *UGT1A6, UGT2A3* and *UGT3A2* expression constructs were transfected into Human Embryonic Kidney 293 cells and stable cell lines generated for analysis. Four drugs were assayed to determine the functionality of the recombinant enzymes and individual substrate specificities. Whilst these studies were inconclusive, further work is required to establish function and substrate profile in order to take the first steps towards creating an *in vitro* toolbox for equine drug metabolism.

Expression of *UGT1A6* and *UGT3A2* was measured in four tissue samples from 12 horses. For both genes, expression levels in the liver were greatest whilst the brain showed negligible expression. Expression levels of both genes in the kidney and lung were similar and lower than levels detected in the liver.

Contents

List of figures and tables

List of abbreviations

ANOVA – Analysis of variance

AO – Aldehyde oxidases

aT – Annealing temperature

B2M – Beta-2-microglobulin

BLAST – Basic Local Alignment Search Tool

bp – Base pairs

BSA – bovine serum albumin

CAPSL – Calcyphosine-like

cDNA – Complementary DNA

CDS – Complete coding sequence

COMT – Cetechol O-methlytransferases

CSN1S1 Casein Alpha S1

CSN2 - Casein Beta

CYP – Cytochrome P450s

DKGD - Diacylgycerol Kinase Delta

DMSO – Dimethly sulfoxide

ER – Endoplasmic reticulum

FMOs – Flavin-containing monoxygenases

GI – Gastrointestinal tract

GSTs – Glutathione S-transferase

HEK293 – Human embryonic kidney cells 293

HPRT1 - Hypoxanthine-guanine phospho-ribosyltransferase

ILR7 - Interleukin 7 Receptor

LMBRD2 - LMBR1 Domain Containing 2

MAOs – Monoamine oxidases

MCS – Multiple cloning site

ML – Maximum likelihood

NADK2 - NAD Kinase 2, Mitochondrial

NATs – N-aceyltransferase

NCBI – National Centre for Biotechnology

PCR – Polymerase chain reaction

qPCR – Quantitative polymerase chain reaction

RANBP3L - RAN Binding Protein 3 Like

RNA – Ribonucleic acid

RPKM – Reads per Kb per million

RPL32 – 60s Ribosomal protein L32

SPP2 - Secreted Phosphoprotein 2

SKP2 - S-Phase Kinase Associated Protein 2

SPEF2 - Sperm Flagellar 2

SULTS – Sulfotransferases

TPMT – Thipurine S-methyltransferase

TRPM8 - Transient Receptor Potential Cation Channel Subfamily M Member 8

UDPGA – Uridine diphosphate glucuronice acid

UDP-Glc – Uridine diphospho-glucose

UDP-GlcNAc – Uridine diphosphate N-acetlyglucosamine

UGTS – Uridine diphosphate 5’-glucuronosyltransferase

USP40 - Ubiquitin Specific Peptidase 40

XO – Xanthine oxidases

Chapter 1: Introduction

1.1 General introduction

In 1995 the Horse Genome Project, undertaken by a consortium of 70 collaborating group, set out to map the equine genome. In 2005, mapping was superseded by sequencing the genome of Twilight, a famous thoroughbred female racehorse (Chowdhary and Raudsepp, 2008). Completed in 2006, by The Broad Institute at the Massachusetts Institute of Technology, the National Health Institute and the Equine Genome Sequencing Consortium (http://www.broadinstitute.org/mammals/horse) research has focussed primarily on traits of equine health, and identifying similarities and differences compared to the human genome (Chowdhary and Raudsepp, 2008).

There are multiple hereditary conditions in the domestic horse, *Equus caballas*, such as inflammatory and degenerative disorders, respiratory diseases, reproduction and infertility, developmental and muscular diseases (Chowdhary and Raudsepp, 2008). The Online Mendelian Inheritance in Animals (http://omia.angis.org.au/home/ - September 2017) suggests there are 128 disorders for which the horse maybe a suitable potential model for human diseases (Wade et al., 2009).

As our knowledge of the genetic basis for equine diseases evolves, new targets for therapy will be discovered. Additionally, our understanding of the action and metabolism of therapeutic drugs will expand. This knowledge is important to ensure effective treatment and will also be of substantial benefit to equestrian sports where the utilization of substances that have the potential to affect performance is regulated. Guidelines regarding the use of controlled and prohibited medications can be found in the FEI database (http://prohibitedsubstancesdatabase.feicleansport.org/).

1.2 The Liver: an overview

When investigating the metabolism of a compound, one needs to consider the organ in which this occurs. Whilst all organs in the body are equipped to metabolise compounds, the enzymes involved are more abundant in the gastrointestinal tract, lung, kidney, and liver, which is the primary site of drug metabolism (Xu et al., 2005).

The liver is fundamental to the maintenance of homeostasis of the mammalian body, cleansing the body of toxic compounds and regulating levels of endogenous and exogenous substances in the blood (Selye, 1941). Liver failure may lead to absorption of abnormal amounts of fat, problems with digestion, prolonged activity of endogenous hormones, and drugs; the consequence of these actions may ultimately be fatal (Bernal et al., 2010).

The liver is located within the rib cage situated behind the diaphragm in the horse (Abdel-Misih and Bloomston, 2010). Accounting for up to 2% of total body weight, the liver is the largest organ in the horse (Konig and Liebich., 2014), and at any one time may contain 10-15% of the total blood volume (Abdel-Misih and Bloomston, 2010). The liver acts as a buffer against volume changes. For example, it will release blood upon injury resulting in fluid loss; conversely, the liver may retain blood in its vascular system when fluid levels are increased, such as in the instance of infusions (Konig and Liebich., 2014). Unlike the majority of mammals, the domestic horse lacks a gall bladder, owing to the continuous digestive process in the horse. The liver therefore constantly performs secretory functions continually releasing bile (Abdel-Misih and Bloomston, 2010).

The liver is a highly structured and resilient organ, capable of regeneration if damaged (Michalopoulos and DeFrances, 1997), and can function adequately with as little as 25% of the tissue being healthy. It performs a diverse number of roles within the body (Best, 1934), including metabolism, detoxification of endogenous and exogenous substrates, protein synthesis, glycogen storage, decomposition of red blood cells and hormone production (Best, 1934).

Of these roles, metabolism is the mechanism responsible for regulating the levels of xenobiotics within the body. Metabolic reactions may activate or inactivate a drug, thus understanding the mechanism of drug metabolism is crucial to understanding how a drug will react, potential drug-drug interactions and drug detection.

1.2.1 Liver structure

The liver, encompassed by a fibrous capsule, is comprised of four demarcated lobes (Reece, 2009); left hepatic, right hepatic, caudate, and quadrate (Konig and Liebich., 2014). The hepatic porta and portal vein supply the liver with blood (Kune, 1969), whilst the bile duct and hepatic vessels mark the boundary between the caudate and quadrate lobes (Konig and Liebich., 2014).

Each of the four lobes is formed of multiple hexagonal lobules, which are the smallest grossly visible units of the liver. Each lobule is composed of a repeated pattern of cells formed into sheets termed laminae hepaticae (Konig and Liebich., 2014, Elias and Bengelsdorf, 1952). A branch of the portal vein receives blood which flows towards the hepatic vein via sinusoids, inside of which lie Kupffer cells; these specialised macrophage cells remove waste materials from the blood including expired erythrocytes and micro-organisms. (Reece, 2009). The sinusoids are surrounded by hepatocyte cells, which form branching plates; and between the branching plates are canaliculi which transport bile to ducts. (Reece, 2009, Elias and Bengelsdorf, 1952).

1.2.2 Hepatocytes

Hepatocytes are the functional cells of the liver, with an average life span of 5 months (Elias and Bengelsdorf, 1952). These cells contain the normal sub-cellular organelles, such as golgi apparatus and mitochondria (Figure 1.1), but are particularly abundant in endoplasmic reticulum (ER) (Campbell. N.A, 2005). It is this organelle, and to a lesser extent the nuclear membranes, to which the enzymes involved in metabolism are bound (Owens et al., 2005).

Hepatocytes produce a range of compounds including albumin, fibrinogen, enzymes, coagulation factors and hormone transporting globulins (Sjaastad et al., 2010). To perform their numerous roles the hepatocytes require access to large volumes of plasma, and to facilitate this their cell surface is covered in microvilli, maximising surface area (Figure 1.1). The hepatocytes are enclosed by rows of endothelial cells separated by small and large pores termed fenestra, which facilitate the filtration of the plasma from the blood to the hepatocytes. The epithelial cells are described as metabolically active owing to the fact they are involved in synthesis, conversions, and storage in addition to metabolism (Reece, 2009).

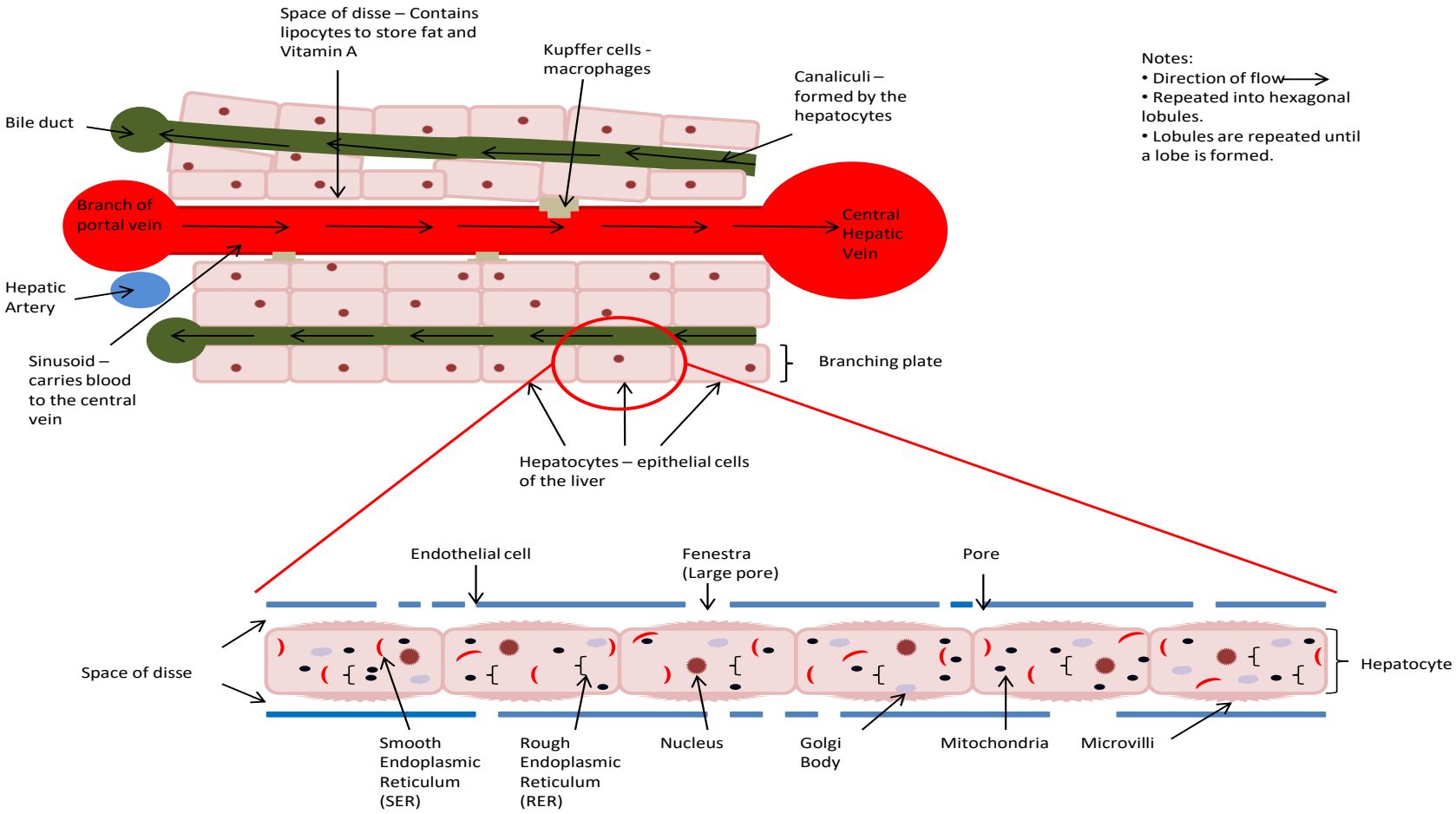


Figure 1.1: Hepatocyte structure. Branching plates of hepatocyte cells surround bile ducts forming canaliculi, these plates in turn form around the branch of the portal vein forming sinusoids, which carry blood to the central hepatic vein. Hepatocyte cells are encompassed by endothelial cells containing pores which filter the plasma from the blood.

1.3 What is metabolism?

Metabolism is the essential chemical process that occurs in the cells of all living organisms whereby compounds are produced, broken down or bio-transformed to maintain life. It is a highly controlled process occurring via a sequence of reactions termed metabolic pathways. Metabolic reactions are categorised as catabolic, breaking down molecules, or anabolic, producing molecules (Gibson and Skett, 2001).

The study of metabolism revolves around the concept of absorption, distribution, metabolism, and excretion, with each step influencing the quantity of active drug in the body (Ekins et al., 2005). How endogenous or exogenous compounds or xenobiotics are metabolised is a significant consideration, with factors such as ethnicity, gender, age, and health needing to be accounted for (Gibson and Skett, 2001). Which organ the metabolism takes place in, whether it is solely one enzyme or a sequence of enzymatic reactions. needs to be investigated. The metabolism of drugs requires the enzymatic bio-transformation of a molecule from the parent state (Gibson and Skett, 2001).

The metabolism of xenobiotics, foreign compounds such as synthetic drugs, plant or fungal derived secondary metabolites and environmental pollutants (Nebert and Russell, 2002) and endogenous compounds, naturally occurring within the body, are controlled by enzymatic reactions which require the bio-transformation of a molecule from the parent state to a functional state, altering the compound's polarity, water solubility or excretability (Meyer, 1996). Whilst the majority of these metabolic reactions result in a metabolite less reactive than the parent drug, hence detoxification, this is not exclusively the case. There are examples where the resultant bio-transformed product has increased activity which can cause a toxic effect (Meyer, 1996).

Although the liver is the main site of drug metabolism all tissues are capable of metabolic processes to some extent (Meech and Mackenzie, 1997a, Xu et al., 2005).

Metabolic pathways utilize a range of chemical reactions which in turn require a variety of enzymes to mediate these reactions (Gibson and Skett, 2001).

1.4 The two phases of metabolism

Classically, metabolism is considered to have two phases. Phase 1 is controlled predominantly by the Cytochrome P450s (CYPs) and Phase 2 is mediated by a range of enzymes (Jancova et al., 2010) of which Uridine diphosphate 5'-glucuronosyltransferases (UGTs) contribute to a significant proportion of reactions (Jancova et al., 2010).

Each phase has a specific role and function in controlling levels of potentially toxic and useful compounds within the body (Penner et al., 2012). It would be incorrect, however, to consider these phases as independent or discrete as is often portrayed in the literature. It is more appropriate to consider them to be working in tandem (Xu et al., 2005). The UGTs and CYPs can work both independently and in conjunction with each other. UGTs can catalyse a compound which has previously been modified by the CYPs (Wildt et al., 1999) and it is estimated that CYPs and UGT enzymes account for over 90% of all drug metabolism and clearance from the human body (Rowland et al., 2013).

1.4.1 An overview of phase 1 metabolism

Phase 1 metabolism is often referred to as functionalisation reactions (Gibson and Skett, 2001); a variety of reactions are performed which introduce or modify a functional group, for example hydroxyl (-OH), carboxyl (-COOH), or amine ($-NH_2$) (Penner et al., 2012).

The phase 1 reactions are mediated by numerous enzymes of which the major group are the CYP superfamily of enzymes (Gibson and Skett, 2001, Meyer, 1996). Contributions are also made by Flavin-containing monoxygenases (FMOs),

monoamine oxidases (MAOs), xanthine oxidase (XO), and aldehyde oxidases (AO), all of which perform oxidation reactions (Penner et al., 2012, Hines and McCarver, 2002, Wales and Fewson, 1994).

1.4.1.1 Cytochrome P450

The CYPs, named because of the chromophore pigment which produces a spectral peak at 450nm, is a superfamily of enzymes (Nebert and Russell, 2002). These microsomal proteins are classified into families based on the similarity of their amino acid sequences (Gibson and Skett, 2001, Nebert and Russell, 2002), gene structure, and phylogenetic criteria (Werck-Reichhart and Feyereisen, 2000). The CYPs are haemoproteins that require a co-factor of nicotinamide adenine dinucleotide phosphate (NADPH) to catalyse a vast array of oxidation reactions (Gibson and Skett, 2001). They metabolise multiple targets including xenobiotics, steroids, alkanes, and fatty acids (Bernhardt and Urlacher, 2014, Nebert and Russell, 2002).

Subfamilies are estimated to have diverged over 1 billion years ago, resulting in the diversification of sequences (McKinnon et al., 2008). Families of CYPs have been divided based on >40% amino acid sequence homology, with sub-families sharing >55% (McKinnon et al., 2008).

The CYPs have been studied extensively (Wildt et al., 1999, Nebert and Russell, 2002), with over 90,000 publications currently available (www.ncbi.nl.nih.gov June 2018). These enzymes have been found in an array of species ranging from mammals to bacteria and plants (Renault et al., 2014, Penner et al., 2012), with over 21,000 members of this superfamily currently identified (Bernhardt and Urlacher, 2014). In mammals, the CYP enzymes are membrane bound, with the majority bound to the ER and a handful existing within the mitochondria (Guengerich, 2003). Bacterial CYPs are not membrane bound proteins but are soluble and present in the cytosol (Hannemann et al., 2007). In insects and

nematodes, CYPs have been discovered in the mitochondria (Werck-Reichhart and Feyereisen, 2000).

1.4.1.2 Cytochrome P450 in equines – an overview

Human CYPs are grouped according to their sequence similarity into 18 families and 44 subfamilies. Of the 57 putatively functional human CYPs, the enzymes belonging to the CYP1, 2 and 3 families (Nebert et al., 2013) are responsible for the majority of drug metabolism (Vimercati et al., 2017). As a result, these three families have been the primary target for isolation and characterisation across numerous species (Vimercati et al., 2017, Zanette et al., 2013, Watanabe et al., 2013, Moskaleva et al., 2015). In comparison to the depth of knowledge on this major enzyme family in humans and rodents, CYPs in veterinary species, particularly the horse, remains in its infancy.

Although the first equine CYP was isolated as far back as 1993 (Komori et al., 1993), it is only within the last decade that there has been an increase in research in this area, towards the generation of an *in vitro* system to study phase 1 metabolism in the horse. Individual CYPs have been successfully isolated and cloned, and recombinant proteins have been expressed and functionally characterised (Knych and Stanley, 2008, Peters et al., 2013, DiMaio Knych et al., 2010). Information on equine CYPs remains incomplete (DiMaio Knych et al., 2010), with most coming from microsomal studies using probe substrates to infer their presence and activity (Nebbia et al., 2003).

Forty six sequences in the equine genome have been identified as *CYPs* (Orr, 2016) however, given the number of CYPs identified in other species, it is likely this will increase with further study.

1.4.1.2.1 CYP1 family

Since research began into the equine cytochromes in 1993 (Komori et al., 1993) compared to other families, little has been discovered regarding the equine CYP1 family. Analysis of the equine genome has discovered three sequences predicted to be members of the equine CYP1 family, annotated as *CYP1A2-like, CYP1A1* and *CYP1B1-like,* on chromosome 1 (Orr, 2016), which correlates with the number of CYP1 members encoded in the human genome (Shimada et al., 2017). However, in comparison to the human sequences, the equine predicted members do not appear to be full-length and their expression and function has not been experimentally confirmed.

1.4.1.2.2 CYP2 family

Equine orthologues of human *CYP2B6* were located to equine chromosome 10 identifying six potential *CYP2B6 orthologues* (Peters et al., 2013). A more recent study identified five further clans of the CYP2 family, CYP2A13, CYP2B6, CYP2C92, CYP2D50, and CYP2E1, encoded in the equine genome (Orr, 2016).

Whilst the expression of equine CYP2 members has been confirmed in the liver by western blot, gene expression of *CYP2* clans has been confirmed using conserved regions of each clan (Costas, 2006, Tyden et al., 2012). However, expression of individual *CYP*s and in depth analyses of their expression in drug metabolising tissues has only been investigated in the liver.

1.4.1.2.2.1 Recombinant protein studies

Equine CYP2B6, expressed in the V79 Chinese hamster fibroblast cell line, was found to metabolise ketamine, an anaesthetic and analgesic commonly used by veterinarians (Peters et al., 2013). With human CYP2B6 showing highest activity with ketamine this implies that CYP2B6 is the equine orthologue (Peters et al., 2013), however this has yet to be further investigated. Recombinant equine CYP2C9 was discovered to have a similar substrate profile to human CYP2C9 (DiMaio Knych

et al., 2009), but different rates of metabolism. Equine CYP2C9 catalysed diclofenac at a slower rate than human CYP2C9, but showed similar rates in the metabolism of tolbutamide and warfarin (DiMaio Knych et al., 2009).

1.4.1.2.3 CYP3 family

A comparative analysis of human *CYP3A* genes to the equine genome identified seven potential orthologues of *CYP3A* genes and one pseudogene (Schmitz et al., 2010). The length and number of exons were consistent between human CYP3A4 and the identified horse CYP3As (Schmitz et al., 2010), with *CYP3A89*, *CYP3A96* and *CYP3A97* sharing the greatest sequence similarity with human *CYP3A4* (DiMaio Knych et al., 2010).

1.4.1.2.3.1 Tissue expression of the CYP3A genes

Analysis of expression of *CYP3A* isoforms, analysed in liver and intestinal tissue, found high levels of expression of *CYP3A97*, *CYP3A89*, *CYP3A96* and *CYP3A94* in the liver, whilst *CYP3A93* and *CYP3A95* were detected at low levels in the liver, accounting for 1% and 2% respectively of total *CYP3A* expression (Tyden et al., 2012). The highest levels of *CYP3A93* and *CYP3A96* expression were in the duodenum and proximal jejunum (Tyden et al., 2012). Further to this, using conserved regions of sequence, five families of CYPs were also assessed for expression in the intestine and liver. *CYP1A*, *CYP2A*, *CYP2C*, *CYP2D* and *CYP2E* are all expressed in liver and intestine, although levels of *CYP2D* were low relative to the other four families (Tydén et al., 2014).

1.4.1.2.3.2 Protein expression studies

Previous studies had produced recombinant CYPs, CYP3A89 and CYP3A96. CYP3A89 was found to be a truncated protein, suggesting that equine CYP3A89 is a pseudogene or the result of a mutation or post-translational modification; further work is required to confirm this (DiMaio Knych et al., 2010). Equine CYP3A96 was

discovered to be functionally active with testosterone and nifedipine as the substrates (DiMaio Knych et al., 2010).

1.4.1.2.4 Further Research

Research into equine phase 1 metabolism of remains in its infancy, and whilst progress has been made, further investigation is required to fully elucidate the CYPs involved, their expression profile, and substrate profiles.

1.4.1.3 Non-CYP mediated reactions

FMOs, XOs, AOs and MAOs also contribute to phase 1 reactions (Gibson and Skett, 2001). In humans, FMOs are highly expressed in kidney, lung, small intestine, liver and brain, while XOs have a wide tissue expression profile which also includes the heart, adrenals, spleen, and Kupffer cells (Binda et al., 2002, Chen et al., 2011b). AO expression is much more tissue specific, shown to be present in brain, lung, kidney, and liver (Strolin Benedetti et al., 2006).

As with CYPs, FMOs, MAOs, AOs, and XOs have been found across species including plants, fungi, and the prokaryote kingdom (Chen et al., 2011b, Wales and Fewson, 1994, Binda et al., 2002). All of these enzymes perform oxidative reactions against numerous substrates (Gibson and Skett, 2001).

This is not a comprehensive list or overview of the enzymes involved in phase 1 metabolism but provides a brief insight into the number and diversity of enzymes involved in each category. Owing to the huge number of enzymes involved in metabolism, the implications for drug metabolism and disease is only beginning to be appreciated.

1.4.2 An overview of the minor enzymes of phase 2 metabolism

Phase 2 metabolism is often referred to as the detoxification or elimination step (Xu et al., 2005). As with phase 1, this is also under the regulation of multiple enzymes

which have been shown to have different tissue and development expression patterns (Xu et al., 2005), some of which require co-factors to function (Gibson and Skett, 2001). It is suggested that each of the enzymes has a basal level of expression, which may be further induced as a result of xenobiotic exposure (Xu et al., 2005).

UGTs account for approximately a third of the phase 2 reactions, with the sulfotransferases (SULTs) being the second most active group of enzymes in humans (Jancova et al., 2010). N-acetyltransferases (also commonly termed acetyl CoA dependent N-acetyltransferases, arylamine N-acetyltrasnferases or NATs), Glutathione S-transferase (GSTs), methyltransferases and catechol O-methyl transferases are responsible for the remainder of the phase 2 reactions (Jancova et al., 2010, Gibson and Skett, 2001). Less is known about these reactions as they are comparatively less frequent than the CYP reactions (Jancova et al., 2010).

1.4.2.1 N-acetyltransferases

NATs, a group of cytosolic enzymes found in the Kupffer cells (Jancova et al., 2010, Gibson and Skett, 2001), catalyse the acetyltransferase from acetylcoenzyme A to a substrate that may be an amine or hydrazine compound (Butcher et al., 2002).

NATs have been discovered in multiple species from humans, rodents, cats, rabbits, dogs, and zebrafish to the bacterium *Salmonella typhimurium* (Butcher et al., 2002, Jancova et al., 2010, Sim et al., 2008).

In humans they have been divided into two sub-families (Butcher et al., 2002) *NAT1* and *NAT2* (Penner et al., 2012). *NAT1* has a broad tissue expression profile, whilst *NAT2* is specific to the gut and liver (Penner et al., 2012). There is a development specific profile, with *NAT1* detected very early in development, at the four cell stage (Sim et al., 2008). All mammals have been found to encode polymorphic loci, with polymorphisms in human *NAT1* shown to be associated with the development of breast cancers (Sim et al., 2008, Grant et al., 1997). *NATs* in the prokaryotes have

also been proven to be polymorphic, with mutations attributed to the slowed growth of *Mycobacterium tuberculosis* (Sim et al., 2008).

1.4.2.2 Sulfotransferases

SULTs are responsible for the conjugation of the sulfonyl moiety from the co-factor 3'-phospho-adenosine-5'-phosphosulfate (PAPS) to a substrate (Glatt et al., 2000). Primarily SULTs conjugate phenols, but they can also metabolise alcohols, amines, and thiols (Gibson and Skett, 2001), both endogenous and exogenous (Jancova et al., 2010). Sulfonation increases a compounds solubility, which facilitates excretion from the body due to the ability of the compounds to penetrate cell membranes (Glatt and Meinl, 2004). Detoxification is not the only function that SULTs have as they are known to produce products which have toxifying effects on the body. Conjugates of benzylic and allylic alcohols are known to cause toxic effects by being converted into highly reactive pro-carcinogens binding to DNA (Glatt and Meinl, 2004, Jancova et al., 2010).

SULTs are cytosolic enzymes (Penner et al., 2012, Glatt et al., 2000, Nowell and Falany, 2006), that have a wide tissue expression profile including, liver, lung, kidney, gastrointestinal tract, and platelets (Nowell and Falany, 2006, Penner et al., 2012). This a superfamily of enzyme is divided into two classes. One class is membrane bound to the golgi apparatus (Penner et al., 2012, Glatt and Meinl, 2004), whilst the second is the soluble cytosolic form that metabolises a wide substrate range (Glatt and Meinl, 2004). The second class is divided into sub-families, based on sequence similarities (Glatt et al., 2000, Jancova et al., 2010).

In comparison to the CYPs, SULTS have not been extensively studied although they have been found in mammals ranging from humans to rodents to extending beyond mammals to the piscine, zebrafish (Nowell and Falany, 2006, Yasuda et al., 2006). Studies are beginning to identify SULTs in prokaryotes with several discovered in mycobacteria (Mougous et al., 2002).

Polymorphisms have been identified in *SULT*s; for example several have been discovered in *SULT1A* and associated with platelet enzymatic activity (Nowell and Falany, 2006). Two single nucleotide polymorphisms (SNPs) in *SULT2A1*, *G187C* and *G781A* have been associated with the development of prostate cancer (Nowell and Falany, 2006).

1.4.2.3 Glutathione S-transferases

GSTs are a superfamily of enzymes, encoded on several chromosomes, which catalyse a number of reactions (Strange et al., 2001). Most of the proteins are predominately found in the cytosol, however some are membrane bound (Jancova et al., 2010). GSTs are involved in the catalysis of endogenous prostaglandins and steroids but also metabolise an array of xenobiotics, detoxifying epoxides, ketones, and aromatic compounds, amongst others (Jancova et al., 2010, Gibson and Skett, 2001).

They have been found in a range of species. In plants they detoxify herbicides as well as auxins and cytokinins (Edwards et al., 2000), and they have been found in bacteria with specific substrate profiles (Vuilleumier and Pagni, 2002). In mammals they are found to be expressed in a range of tissues, with levels specific to each family, and shown to have a more diverse substrate profile (Thomson et al., 2004).

1.4.2.4 Methyltransferases

The major methyltransferases are Thiopurine S-methyltransferase (TPMT) and Catechol O-methyltransfersaes (COMT). TPMTs are cytosolic enzymes shown to be highly expressed in the liver and kidney, which are responsible for catalysing drugs via the addition of a methyl group, preferentially metabolising thiopurine drugs (Penner et al., 2012). Thiopurine drugs are used to treat a range of diseases and disorders, particularly notable are their roles in metabolising anti-cancer and immunosuppressant drugs (Jancova et al., 2010). A reduction in the activity of these enzymes results in the accumulation of thiopurine nucleotides causing haemopoiesis; this consequence of failing to produce red blood cells can be fatal (Jancova et al., 2010).

COMTs are found in both the cytosol and attached to membranes, and include neurotransmitters among its substrates (Penner et al., 2012). The membrane bound form of this group of enzymes is found to be expressed in the brain at the highest levels, with the cytoplasmic version found at more significant levels in the peripheral tissues (Jancova et al., 2010).

This overview of the minor phase 2 enzymes describes their involvement in conjugation reactions (Iyanagi, 2007), and whilst less well studied and understood compared to phase 1 enzymes, they are no less important. Equally as varied and present across both the eukaryote and prokaryote kingdoms - these enzymes contribute to the overall processing of xenobiotics. Mutations in these genes are not only disease-causing but also have functional implications. A large proportion of phase 2 reactions are mediated by a major class of enzymes, the UGTs (Iyanagi, 2007) which are the focus of this study.

1.5 Uridine 5'-diphosphate glucuronosyltransferases

Uridine 5'-diphosphate-glucuronosyltransferases (also referred to in the literature as Glucuronosyltransferases, Uridine Diphosphate, UDP-glucuronyly-transferase, UDP-glucuronosyltransferases or UGTs) are a diverse sub-family of enzymes belonging to a large superfamily of Glycosyltransferases. UGTs are estimated to account for a third of total phase 2 reactions (Guillemette, 2003, Jancova et al., 2010).

UGTs mediate the transfer of a polar moiety from a donor sugar to a less polar molecule, termed aglyclones (de Wildt et al., 1999), with the products of this reaction being hydrophilic glucuronides (Rowland et al., 2013). It is this change in polarity which facilitates the removal of the conjugate from the circulatory system via urine or bile (Radominska-Pandya et al., 2005b, Soars et al., 2001, Meech and Mackenzie, 1997a). There are a range of compounds which can act as a donor sugar including Uridine diphosphate glucuronic acid (UDPGA), Uridine diphosphate N-

acetylglucosamine (UDP-GLcNAc) and Uridine disphospho-glucose (UDP-GLc) (Owens et al., 2005, MacKenzie et al., 2011, Mackenzie et al., 2008). These donor sugars are mostly utilised by mammals, while invertebrates, plants, and microorganisms preferentially use UDPGLc as the donor sugar (MacKenzie et al., 2011).

The UGTs have a wide substrate profile, glucuronidating a diverse range of naturally occurring and artificial substrates, including alcohols, phenols, hydroxylamines, carboxylic acids, amines, sulphonamides, and thiols (Gibson and Skett, 2001, Rowland et al., 2013). The substrate profile includes compounds naturally occurring within the body (endogenous) such as hormones, bile, and bilirubin (Meech and Mackenzie, 1997a), and external compounds that are taken into the body (exogenous) including fat soluble vitamins, carcinogens, environmental pollutants, and drugs (xenobiotics) (Shelby et al., 2003).

The purpose of glucuronidation is to increase the polarity of a compound in order to enable excretion from the circulatory system. However, there are circumstances whereby, instead of inactivation, bio-activation occurs leading to the production of a compound which is potentially harmful (Stingl et al., 2014) . One example involves morphine which can be glucuronidated into two forms: morphine-3-glucuronide and morphine-6-glucuronide. Morphine-6-glucuronide has a 600 times more potent analgesic effect than unconjugated morphine (Shelby et al., 2003, Guillemette, 2003).

1.5.1 The UGT families

The UGTs are a large and diverse superfamily of enzymes that in vertebrates are divided into five families based upon sequence homology, termed UGT1, UGT2, UGT3, UGT5 and UGT8 (Gong et al., 2001, Owens et al., 2005, Meech and Mackenzie, 1997a, Penner et al., 2012). In mammals, four families have been identified, each containing multiple members, with a wide overlapping substrate

profile (Owens et al., 2005, Meech and Mackenzie, 1997a, Meech and Mackenzie, 2010), with the *UGT1s* being the largest and most structurally complex (Radominska-Pandya et al., 2005c). The UGT2 family is divided into two subfamilies denoted UGT2A and UGT2B, with genes within each family sharing >70% sequence homology (Meech and Mackenzie, 1997a, Meech et al., 2012a). The UGT3 and UGT8 families are the smallest, with the UGT3s encompassing two members and UGT8 a single enzyme (Meech et al., 2012b, Meech et al., 2015). The UGT8 is a unique member of this superfamily in that its role is not bio-transformative, but primarily biosynthetic. It is involved in the production of brain sphingolipids (Meech et al., 2015) and as such is not within the scope of this project.

The UGT5 family appears to belong exclusively to the piscines with 17 members of the family identified in zebrafish (Huang and Wu, 2010). UGTs have also been identified in species beyond the vertebrates, with 42 UGTs sequentially identified in the silkworm (Huang et al., 2008) as well as other species of Lepidoptera (Ahn et al., 2012). UGTs have also been identified in plants, but here they appear not to be anchored to the membrane of the ER, but present as cytosolic enzymes (Bock, 2016, Caputi et al., 2012). As with all other metabolising enzymes orthologues have also been isolated in the prokaryote kingdom (Schmid et al., 2016).

1.5.1.1 UGT1 locus structure

Of the four sub-families, *UGT1* is the largest and has the most structurally complex locus (Radominska-Pandya et al., 2005a). The human genome possesses 13 *UGT1* genes over a 200kb locus on chromosome 2, including four pseudogenes; *UGT1A13p*, *UGT1A12p*, *UGT1A11p* and *UGT1A2p* (de Wildt et al., 1999, Owens et al., 2005, Radominska-Pandya et al., 2005a, Stingl et al., 2014). Mouse *UGTs* are encoded over a region of 190kb on chromosome 1 and consist of 14 isoforms, of which five (*UGT1A14p*, *UGT1A8p*, *UGT1A6p*, *UGT1A4p* and *UGT1A3p*) are pseudogenes (Zhang et al., 2004). The rat locus, present on chromosome 9, is

smaller than both the human and mouse loci at 110kb. The rat genome encodes 11 members of the *UGT1* family which includes three pseudogenes, *UGT1A10p*, *UGT1A5p* and *UGT1A3p* (Zhang et al., 2004, Owens et al., 2005, Mackenzie et al., 2005). The zebrafish genome encodes both a *UGT1A* and *UGT1B* sub-family, each containing seven members (Huang and Wu, 2010) with a single member, *UGT1B6p*, determined to be a pseudogene (Wang et al., 2014). Of the UGT1 families investigated thus far, the presence of UGT1A and UGT1B appear unique to fish, as flounder and plaice have also been shown to express both sub-families (Leaver et al., 2007).

Figure 1.2 displays the complex locus in humans; each *UGT1* transcript encodes a protein composed of five exons, a unique exon 1 and four shared exons, 2-5 (Ohno and Nakajin, 2009, de Wildt et al., 1999). Upstream from each individual exon 1, sequencing has an identified a TATA box, the promoter element from which transcription is initiated (Meech et al., 2012a, Mackenzie et al., 2005, Owens et al., 2005, Ritter et al., 1992). Once translated exon 1 generates the amino-termini which encodes substrate specificity (Mackenzie et al., 2005, Guillemette, 2003, Rowland et al., 2013). The four shared exons are identical between all *UGT1* isozymes and form the carboxyl-terminal of the protein, which binds the enzyme to the ER membrane and facilitates interaction with the co-factor and UDPGA (Owens et al., 2005, Meech and Mackenzie, 1997a). Initiation of transcription of any given exon 1 will result in transcription of all exon 1s that follow as well as the shared exons (Meech and Mackenzie, 1997a). This creates multiple isoforms from this locus, each with a unique exon 1 but identical exons 2-5 (Owens et al., 2005). Alternative splicing joins the first exon to the shared exons creating the mature protein. The locus structure is conserved across species with humans, mice, and rats sharing this arrangement (Mackenzie et al., 2005), as well as chickens, frogs, and zebrafish, suggesting likely conservation across multiple animals (Meech et al., 2012a, Huang and Wu, 2010).

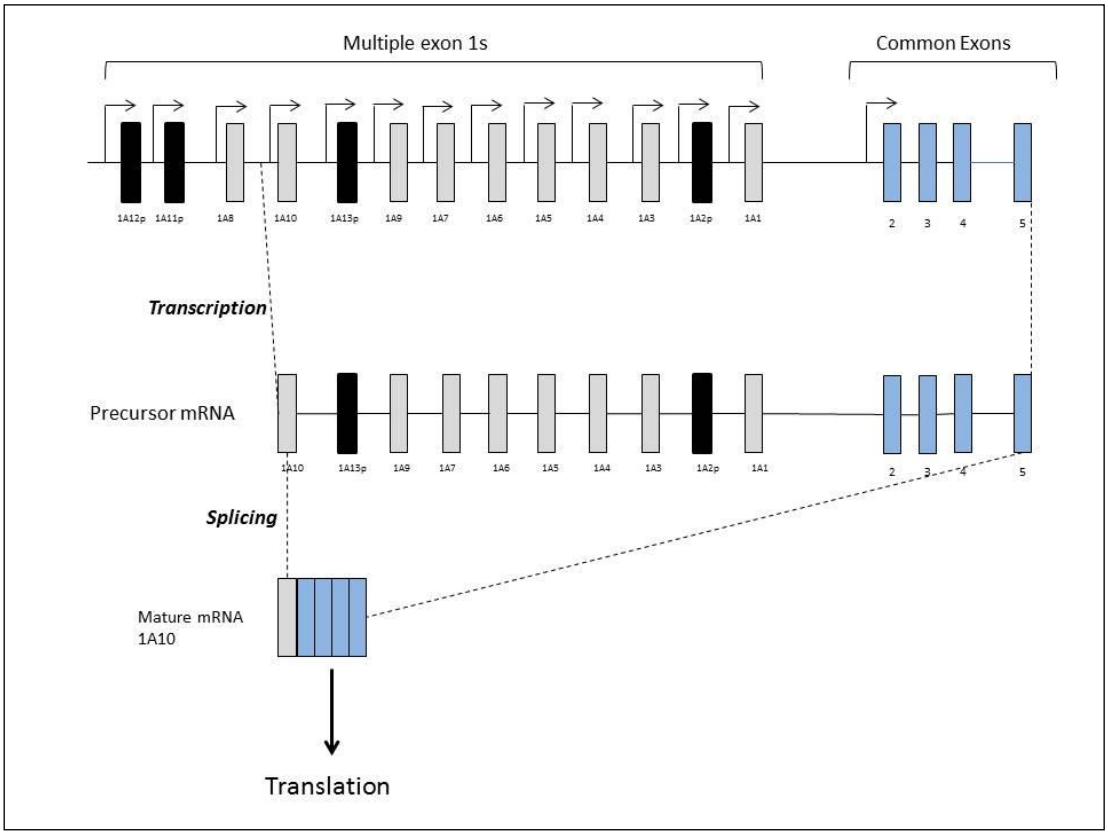


Figure 1.2: The Human *UGT1* locus and subsequent synthesis of an isoform. This figure uses the synthesis of *UGT1A10* to demonstrate how individual isoform are transcribed and translated to a functional protein. The four common exons (light blue) are located upstream of the variable exon 1s (light grey), including pseudogenes (black). The arrows indicate the beginning of the TATA sequence, this promoter element is the point from which transcription starts. When transcription begins at the promoter for *UGT1A10*, all subsequent exons are also transcribed in the precursor mRNA. The mature protein is created from splicing the first exon of the precursor mRNA to the last four exons, this sequence then proceeds to be translated into the functional protein.

1.5.1.2 The UGT2 locus

The UGT2 family differs from the other three mammalian UGT families (Figure 1.3), as it is made of two subfamilies denoted UGT2A and UGT2B (Owens et al., 2005). Both are composed of six exons, with members of each respective subfamily sharing >70% sequence homology (Meech et al., 2012a). There are three members of the UGT2A subfamily, with *UGT2A1* and *UGT2A2* encoded through the sharing of exons 2-6 and variable splicing of exon 1, which creates each individual isoform (Figure 1.3) (Owens et al., 2005, Court et al., 2008). *UGT2A3* is encoded separately in the locus, upstream to *UGT2A1* and *UGT2A2* (Court et al., 2008), amongst the *UGT2B* genes, a structure which is maintained in mice (Owens et al., 2005). The several members of the UGT2B subfamily are each encoded for individually within the genome (Meech and Mackenzie, 1997a, Mackenzie et al., 2005).

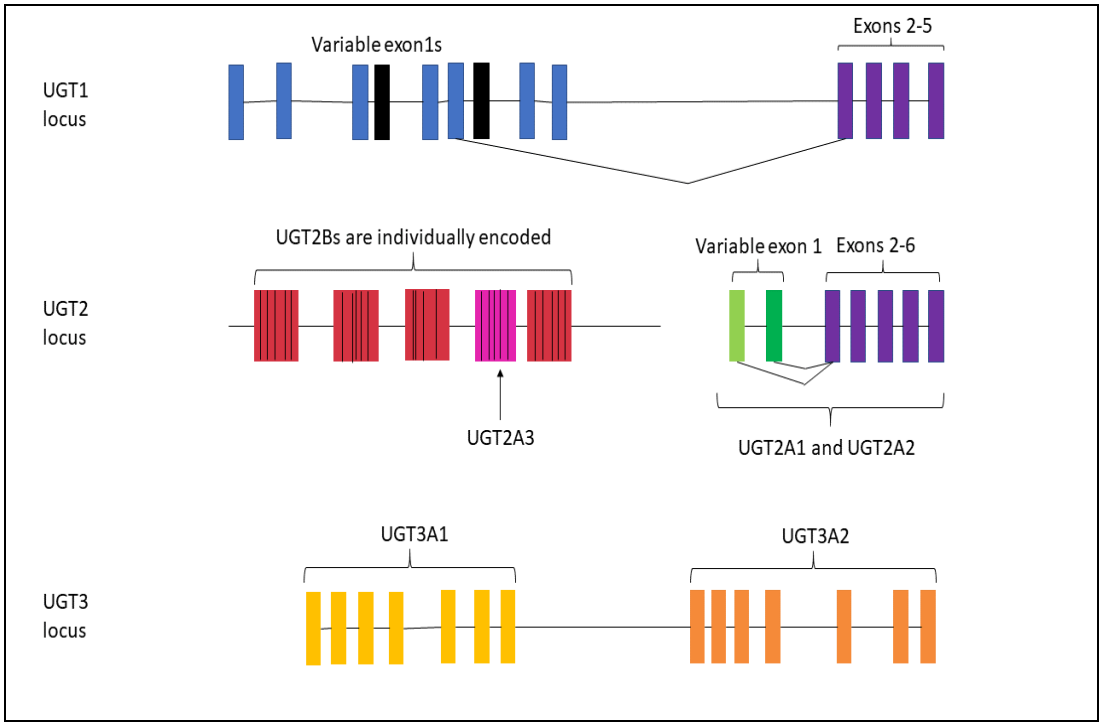


Figure 1.3: Schematic of the human *UGT* loci. The schematic displays the differences in locus structure between each of the *UGT1, UGT2 and UGT3* families. *UGT1* is a complex locus, with individual isoforms sharing exons 2-5 (highlighted in purple) whilst there are multiple exons 1s (blue) some of which are pseudogenes (black; these are joined by variable splicing to form the individual isoforms. The *UGT2* family is comprised of two sub-families, UGT2A1 and UGT2A3 which are encoded through shared exons 2-6 (purple) and individual exon 1s (green). The protein, UGT2A3, is encoded separately (pink) to UGT2A1 and UGT2A3, with *UGT2A3* situated amongst the *UGT2Bs.* Each member of the UGT2B sub-family (red) is encoded separately in the genome. Each of the *UGT3* sequences are encoded by seven exons, separately in a tandem repeat on the chromosome, *UGT3A1* yellow, *UGT3A2* pale orange.

1.5.1.3 The UGT3 locus structure

The UGT3s family, which is much smaller than either the UGT1 or UGT2 families, has been found to contain two members in humans and mice, and a single member in rats (Meech et al., 2012a). Each human and murine isoform, termed *UGT3A1* and *UGT3A2*, is encoded separately on the genome as a direct repeat (Meech et al., 2012b) and is composed of seven exons (Meech et al., 2012a, Mackenzie et al., 2008). The rat genome encodes a single *UGT3A* member, comprised of seven exons, termed *UGT3A2* (Mackenzie et al., 2005, Meech and Mackenzie, 2010).

1.5.2 Tissue profile of the UGTS

In mammals, whilst the liver is responsible for the majority of glucuronidation reactions (Radominska-Pandya et al., 2005a), UGT expression is not solely confined to this organ. UGT expression has been demonstrated in prostate, uterus, breast, placental tissue, kidney, and brain, with almost all tissues thought to express some UGTs (Guillemette, 2003, Meech and Mackenzie, 1997a, Jancova et al., 2010).

UGTs are numerous with 31 genes identified in humans (Ohno and Nakajin, 2009), and orthologues isolated and studied in mice and rats (Fay et al., 2015). As the primary site of drug metabolism, the majority of the UGTs are expressed in the liver. However there are specific isoforms that gene expression studies have shown to be expressed extra-hepatically, such as *UGT1A7* in humans (Bock, 2003). UGTs have been isolated from a range of mammals, and orthologues have also been discovered in insects, worms, yeast, bacteria, and plants (Bock, 2003). In plants a diverse range of UGTs have been found which have been discovered to be cytosolic enzymes (Caputi et al., 2012, Bock, 2016). Investigations in insects have identified over 310 UGTs from nine different species (Ahn et al., 2012).

1.5.3 Clinical impact of mutations in the UGTs

Intrinsic factors, such as age and health status, will impact on the ability to metabolise a substrate, and there are also external factors such as drug-drug interactions which affect the ability to metabolise a compound (Meech and Mackenzie, 1997a, Stingl et al., 2014). Factors at the genetic level can also alter the rate of metabolism. *UGT*s are highly polymorphic genes and, dependent on the type of mutation (insertion, deletion, substitution, truncation or recombination), reaction rates of the resulting enzyme can alter which in turn will impact on how quickly the body can clear itself of unwanted chemical compounds (Stingl et al., 2014). A mutation may not only alter how the UGT responds to and process its substrate but it can also have clinical implications. Polymorphisms in the *UGT1* family of humans have been suggested to be associated with a decreased risk of the development of colorectal cancer and also to an increased risk in liver, colorectal, and otolaryngeal cancers (Radominska-Pandya et al., 2005a).

1.6 Drug disposition in companion animals

Knowledge on the fate of a drug, its disposition (absorption, distribution, metabolism, and excretion) and pharmacokinetics is crucial in pharmaceutical research and development. In addition, an understanding of how drugs are metabolised is also essential to the prescribing veterinarian in the treatment of companion animals. The CYPs and UGTs account for 90% of hepatic clearance, whilst in most instances this detoxifies the system, in rare instances the drug may be converted to a more reactive metabolite. Understanding how animals respond to a compound will inform on the range of species to which a drug can be administered (Rowland et al., 2013).

The main challenge for veterinarians is not the selection of a drug but a rational dosing regimen which can be dependent on an animal's anatomy, biochemistry, physiology, and behaviour as well as on the nature and causes of the condition

requiring treatment, resulting in inter- and intra-species differences in drug response (Toutain et al., 2010).

1.6.1 Drug disposition in cats

Cats are known to respond differently to certain drugs when compared to other companion animal species (Court, 2013b). Pharmacokinetic studies indicate that drugs such as acetaminophen (paracetamol), carprofen, and acetylsalicylic acid (aspirin) are cleared more slowly in cats than in dogs and humans. Cats have been shown to lack major UGT enzymes, including UGT1A6 and UGT1A9, that glucuronidate acetaminophen (Court and Greenblatt, 2000). Deficient glucuronidation may also explain slower carprofen clearance, although there is no direct evidence for this. Poor aspirin clearance is thought to be mainly a consequence of slower glycine conjugation (Court, 2013b).

Cats encode few UGT2B enzymes relative to dogs, humans, and rodents which implies they glucuronidate a narrower range of substances (Kondo et al., 2017a). One example where this is important involves the conjugation of morphine and estradiol, which is mediated by UGT2B7 in humans. The use of these as probe substrates in cats confirmed that low levels of estradio-17-glucuronide and morphine-3-glucuronide were formed (van Beusekom et al., 2014), suggesting that the cat lacks a orthologue for UGT2B7 (van Beusekom et al., 2014). There may be implications for the lack of UGT2B7, for example the use of morphine as an analgesic in cats may be inappropriate owing to their inability to glucuronidate and excrete the metabolites, leading to morphine toxicity.

1.6.2 Drug metabolism in dogs

Dogs are a popular companion animal and can be used as animal models for human disease as they can develop over 360 diseases analogous to those seen in humans (Shearin and Ostrander, 2010). For drugs to be metabolised they need to be

transported across cell membranes. If the proteins involved in this process are faulty, this can affect the ability to process compounds. For example, a truncation of ATP Binding Cassette Subfamily B Member 1 (*ABCB1*), resulting from a 4bp deletion which creates a premature stop codon, has increased the sensitivity of dogs to ivermectin, a drug used to treat parasites (Deshpande et al., 2016). *ABCB1* encodes a transmembrane P-glycoprotein which transports molecules to the inside of cells (Deshpande et al., 2016), and this truncation results in a non-functional protein preventing small molecules from translocating across cellular membranes, which in turn prevents them from engaging with the metabolising enzymes. This can result in compounds reaching toxic levels, which in the case of ivermectin, can lead to death in dogs (Deshpande et al., 2016).

There are genetic differences in dog breeds which have been shown to alter the effectiveness and response to a drug (Fleischer et al., 2008). CYP2B11 is an enzyme responsible for the hydroxylation of the anaesthetic agent propofol. A comparison of CYP2B11 activity in beagles versus greyhounds found that it was higher in beagles, resulting in greyhounds taking longer to recover from anaesthesia using this agent (Fleischer et al., 2008, Hay Kraus et al., 2000). The microsomal content of CYP2B11 was looked at in male and female mixed breed dogs, and found males contain CYP2B11 at greater levels than females (Hay Kraus et al., 2000). In dogs, rats, and humans the protease inhibitor indinavir, used to treat human HIV, is metabolised by CYP3A4. A study using microsomes found female beagles metabolised indinavir two-fold quicker than males (Mugford and Kedderis, 1998). These studies suggest that there are both species and gender differences in the metabolism of drugs.

Studies examining the activity of other drug metabolising enzymes found labradors to have higher activity of TPMTs than other breeds making them more efficient at metabolising thiopurine based drugs (Fleischer et al., 2008). Interestingly, all dogs have been found to lack cytosolic NAT enzymes leaving them unable to catalyse arylamine and hydrazine compounds. This includes *p*-aminobenzoic acid, which has

multiple uses, including sunscreen and for the treatment of skin conditions, and the ingestion of this compound can cause vomiting and diarrhoea. The antibacterial agent sulfamethazine can cause loss of appetite, vomiting, diarrhoea, and, in extreme cases. liver damage making this an unsuitable for use in dogs (Trepanier et al., 1997).

1.6.2.1 Polymorphisms in canine metabolising enzymes

Breed specific mutations have been found in canine *CYP*s which affect their ability to metabolise the analgesic phenacetin and tacrine, a drug used to treat Alzheimer's disease. A single nucleotide polymorphism in *CYP1A2* results in a premature stop codon which causes an absence of functional CYP1A2 in the liver (Court, 2013a). An absence of this enzyme is prevalent in 17% of the Japanese beagle population, whilst the allele frequency in Irish wolfhounds is 42% (Court, 2013a). *In vitro* studies using CYP1A2-deficient liver microsomes found phenacetin and tacrine to be more slowly metabolised than in wildtype livers (Court, 2013a). In the case of phenacetin, use of this drug in CYP1A2-deficient dogs could result in toxicity because their livers take longer to clear the compound (Court, 2013a).

1.6.2.2 Comparison of expression of canine and human UGTs

Comparative studies looking at the tissue profile of drug metabolising enzymes in the dog have found that *UGT1A6* is highly expressed in the liver while *UGT1A2*, *UGT1A9,* and *UGT1A11* are expressed in the intestines but levels in the liver are either very low or absent (Heikkinen et al., 2015). This is different to that shown in human studies with *UGT1A7* highly expressed extra-hepatically alongside *UGT1A8* and *UGT1A10* (Izukawa et al., 2009, Ohno and Nakajin, 2009, Heikkinen et al., 2015).

1.6.3 Drug disposition in ferrets

Paracetamol has also been shown to be poorly metabolised in ferrets (Court, 2001). Whereas the lack of expression of *UGT1A6* in cats is due to deleterious mutations in *UGT1A6*, similar mutations in ferret *UGT1A6* have not been found (Court, 2001). Although this study did not look directly at gene expression, it is possible mutations exist in the regulatory components which may cause reduced gene expression (Court, 2001).

1.6.4 Drug disposition: considerations beyond the level of species

These findings highlight the need to not only investigate drug metabolism in individual species, but to also consider the effect of breed, age, and sex. Researching the expression profiles of *UGT*s and identifying polymorphisms may improve the welfare of animals, as this will change our understanding of drug metabolism and enable the veterinarian to choose the most appropriate medication and dosage regimen.

1.7 UGTs in equines

It has been estimated that 6.5% of total hospital admissions are the result of adverse drug reactions (Patel et al., 2007). Increasing our knowledge of the role of UGTs in metabolism will benefit pharmaceutical research, facilitating the development of drugs that will cause minimal adverse responses (Teale and Houghton, 2010, Scarth et al., 2011). This is extendable to the veterinary field, in the treatment of domesticated and farm animals.

Understanding the mechanism of drug metabolism in animals will enable more effective treatments to be developed with minimal side effects. Horses are both socially and economically valuable animals (Scarth et al., 2011). In some countries they are an increasingly valuable food source, with almost 1 million tonnes of horse

meat produced in 2014; globally, the main producers are China, with an annual production of ~200,000 tonnes in 2014 (http://www.fao.org/faostat/en/#search/horse). With a growing market in horse meat understanding the impact of using therapeutic drugs, given the potential for them to enter the food chain through contaminated meat, needs to be thoroughly investigated (Scarth et al., 2011).

It is beneficial to understand the mechanism of drug metabolism for investigative processes where it is essential to determine whether drugs, illegal or otherwise, have been taken, such as in the cases of overdoses, in the racing/sporting industry or equestrian events, where doping tests are mandatory (Scarth et al., 2011). Anabolic steroids are commonly abused for performance enhancement and aesthetic reasons to improve the physical appearance of the horse for breeding selection purposes (Anielski, 2008). Increasing our understanding of steroid glucuronidation will enable anti-doping regulators to more readily detect cases of abuse as they are extensively metabolised (Teale and Houghton, 2010). Steroids are also a popular 'designer' drug, synthetically modified to keep one step ahead of the regulations; being able to investigate and profile unknown substrates may also lead to more successful prosecutions for drug abuse (Teale and Houghton, 2010). Horses may not always be abused with modified endogenous substrates. The use of recombinant human erythropoietin (EPO) in horses has been reported. In addition to the illegality of this in the horse racing industry, it also poses a welfare issue. Horses treated with EPO can die suddenly as a result of increased blood viscosity or as a result of anaemia owing to the horse's immune system recognising an 'alien' compound and producing antibodies which not only attack the human EPO but also endogenous EPO (Lönnberg et al., 2012).

Current knowledge of drug metabolism in equines lags behind other species, such as humans and dogs. Historically studies have predominantly involved large, expensive *in vivo* experiments, where blood and urine were sampled and analysed (Scarth et al., 2011). With advances in science and a desire to incorporate the 3Rs

(reduction, refinement, and replacement) into research, alternative methods have been sought. A popular choice is the use of microsomal preparations, however the development of more specific *in vitro* tools is slowly becoming available for animals, including the dog and cynomolgus monkeys (Soars et al., 2001, Troberg et al., 2015, Hanioka et al., 2006). *In vitro* tools have yet to progress significantly in the horse, yet access to such tools would enable an improved understanding of the metabolism of regulated and illegal drugs. It would follow the ethos of the 3Rs and enable the profiling of current and novel drugs, increasing the likelihood of detection of post-metabolic products (Scarth et al., 2011).

1.8 Aims and Objectives

This project aims to isolate and characterise UGT enzymes from the domestic horse, by comparison to UGTs in other species, including humans, rats, and mice, and to determine which drugs are metabolised by this group of proteins with a view to improving equine medication and doping control within the horse racing and equestrian industries.

Chapter 2: Materials and Methods

2.1 Animal tissue for analysis

Biopsies of liver, kidney, brain, and lung tissue were obtained from five post-slaughter animals at an abattoir (F. Drury & Sons Ltd, Swindon UK), in accordance with the Welfare of Animals at the Time of Killing (UK) Regulations, 2015. Brain biopsies were taken using forceps and scalpel, with other tissues sampled using a 3mm punch biopsy (Fisher Scientific, UK). Separately, liver samples were collected from seven additional animals.

Biopsies were immediately placed into RNA-later and/or Allprotect (Qiagen, Manchester UK) for transportation. Information on age and gender of the animals, where available, was collected at the time of sampling. Samples were stored for a minimum of 24hr at 4°C before transferred for long-term storage at -80°C.

2.2 Molecular biology techniques

2.2.1 Primer design for polymerase chain reaction

Equine *UGT* sequences retrieved from the Ensembl database (http://www.ensembl.org/index.html) were used. Primers were designed to isolate the complete coding sequence (CDS) using Primer3 (http://bioinof.ut.ee/primer3-0.4.0/) and the Oligo-analyzer 3.1 tool (Integrated DNA technologies - http://eu.idtdna.com/calc/analyzer), and ordered from Sigma-Aldrich, UK (see Chapter 3, section 3.3.2, Table 3.3).

2.2.2 Total RNA extraction

Equine tissues were homogenised using MACs M Tubes and a MACs tissue homogenizer (Miltenyi Biotec, UK). 30mg of frozen tissues (liver, brain, kidney and lung) were placed in 1ml of Qiazol lysis reagent (Qiagen, UK), in gentleMACS™ M tube (Miltenyi Biotec), the tube sealed and placed on the gentleMACS™ Dissociator, RNA_02 program (Miltenyi Biotec). Tubes were briefly centrifuged and incubated at room temperature (RT) for 5min. The

lysate was transferred to a 1.5mL tube and 12 Units (U) of proteinase K (600AU/mL - Qiagen) solution was added and incubated at 56^{o}C for 1hr. Total RNA was extracted from each lysate using the RNeasy Mini Tissue Kit (Qiagen) as per the manufacturer's instructions.

2.2.3 DNase treatment and quantification of total RNA

Following the isolation of total RNA, contaminating genomic DNA was removed using DNA-*free* (Ambion by Life Technologies, UK), following the manufacturer's instructions. Total RNA was quantified using the Qubit RNA HS Assay (Life Technologies, Invitrogen, UK) as described by the manufacturer. The purity of the total RNA was measured using the NanoDrop ND-8000 Spectrophotometer (Thermo Scientific, UK). All total RNA (30µl) samples were stored at -80^{o}C.

2.2.4 cDNA synthesis

Total RNA (~500ng) was reverse transcribed using the Superscript III Reverse Transcriptase (Life Technologies, Invitrogen), 5X first-strand buffer and using Oligo $dT_{(18)}$ in a 20µl reaction, incubated 50^{o}C, as per the manufacturer's instructions (Life Technologies, Invitrogen). The resulting cDNA was stored at -20^{O}C.

2.2.5 Polymerase chain reaction (PCR)

2.2.5.1 Gradient PCR

The optimum annealing temperature was determined for each pair of primers using a gradient PCR (GeneTouch Thermal Cycler, BioER, China). Gradients PCRs were performed using Phusion High-Fidelity DNA Polymerase (New England Biolabs, UK). Each tube contained 2µl (25ng) of cDNA template, 1.25µl of each primer (10µM), 5µl of 5X Phusion Buffer HF, 0.5µl of 10mM dNTPs, 0.25µl of Phusion polymerase and 17.25µl of PCR grade water (Sigma-

Aldrich, UK). Thermal cycler protocol: initial denaturation 98°C for 30sec, then 35 cycles of 10sec denaturation at 98°C, 20sec annealing at variable temperature (over a 12°C range), 45sec extension at 72°C, a final extension step of 10min at 72°C and cooled to 4°C.

2.2.5.2 Standard PCR

PCRs to isolate the coding sequence (CDS) of genes were performed using Phusion High-Fidelity DNA Polymerase (New England Biolabs) as follows: 10µl 5X Phusion HF buffer, 1µl 10mM dNTPs, 2.5µl of forward and reverse primers (10µM), 0.5µl Phusion High Fidelity Polymerase, 2µl (25ng) of template cDNA and 31.25µl of water. Positive and negative controls were also included, with cycling conditions as described above (see section *2.2.5.1*).

2.2.5.3 Touchdown PCR

If multiple bands were present, a touchdown PCR was performed on the G-Storm thermal cycler (G-Storm, UK) as detailed in section *2.2.5.2*. The first six cycles on the thermal cycler programme were performed using annealing temperatures 8°C higher than the final annealing temperature (aT) before completing the program with a constant annealing temperature as established from the gradient PCRs for the remaining 35 cycles.

2.2.6 Agarose gel electrophoresis

5µl of PCR product was mixed with 1µl of purple 6X gel loading dye (New England Biolabs) prior to electrophoresis on a 1% agarose gel, made from 0.4mg of Ultra Pure™ agarose (Invitrogen) and 40ml of 1X TBE (Tris-Borate-EDTA, Fisher Scientific), placed in a horizontal electrophoresis chamber, 7 x 7cm, (Biorad, UK) 1hr at 70volts (V). Gels were post-stained for 30min in a gently agitating solution of 45mL water, 5mL of 1M NaCl (Fisher Scientific) and

15µL of Gel Red nucleic acid stain (Biotium INC, USA). PCR products were visualised using the ImageQuant 300 Imager, version 1.0.3 (GE Healthcare, USA).

2.2.7 PCR purification and quantification

PCR products were purified via one of two methods:

1) The QIAquick PCR purification kit (Qiagen) was used for PCRs which generated a single amplicon.

2) The Gel extraction kit (Qiagen) was used to excise the desired band from an agarose gel when multiple bands were produced in the reaction.

Both kits were used as described in the manufacturer's protocols. Purified products were quantified using the Qubit dsDNA BR Assay Kit (Invitrogen, Life Technologies) as per the manufacturer's protocol, and stored at -20^{o}C.

2.2.8 cDNA synthesis for Quantitative PCR

Superscript III First-strand synthesis supermix for qRT-PCR (Life Technologies, Invitrogen) was used to convert 100ng of total RNA to cDNA following the manufacturer's protocol and using the oligo $dT_{(18)}$ primer (Life Technologies, Invitrogen).

2.2.9 Quantitative PCR

2.2.9.1 Primer and probe design

Primers and probes (Sigma-Aldrich), were designed using Primer Express Software for Real-Time PCR, Version 3.0 software (Applied Biosystems). Primers were designed to cross an intron-exon boundary, and probes were

labelled with 5' FAM reporter dye and 3' TAMRA quencher (Chapter 4, section 4.3.1, Table 4.2).

2.2.9.2 Optimisation of quantitative PCR assays

Optimisation of primer concentrations was initially performed. Nine different conditions (Table 2.1) were tested in triplicate and probe concentrations were held constant at 5pmol/µl. Primer and probes were tested by quantitative PCR using the ABI Prism 7500 FAST Sequence Detection System and FAST Universal Master Mix (Applied Biosystems, California USA). The cycling conditions were as follows: initial denaturation of 20sec at 95^{o}C, then 40 cycles of 15sec denaturation at 95^{o}C and 1min annealing at 60^{o}C.

The primer concentrations that yielded the maximum ΔRn value (magnitude of specific signal generated under the above PCR conditions) and lowest Cycle threshold (Ct) were used for subsequent for quantification assays.

Once optimal primer concentrations were established, a titre was performed to determine the efficiency of the primer concentration combination. The cloned gene was used as the template. Serial dilutions of the plasmid preparation were performed to create solutions with the following copy numbers of plasmid: 300,000, 30,000, 3,000, 300, and 30 copies.

Forward primer final concentration (nM)	**Reverse primer final concentration (nM)**		
	50	**300**	**900**
50	50/50	50/300	50/900
300	300/50	300/300	300/900
900	900/50	900/300	900/900

Table 2.1: Reverse and forward primer concentrations. This is the combinations of forward and reverse primer concentrations tested in triplicate to determine optimal concentrations for quantification assays.

2.2.9.3 Quantitative PCR assays

Quantitative PCRs were performed using the cDNA template diluted to 2.5ng/µl. Results were normalised against the reference genes beta-2-microglobulin (*B2M*), 60S ribosomal protein L32 (*RPL32*) and *β-actin*. All reactions were performed in triplicate alongside a no template control.

2.2.10 Statistical analysis of quantitative PCR data

Raw data generated from the quantitative PCR was processed using MS Excel 2013. The mRNA levels were analysed as fold change relative to the reference genes. A One-Way Analysis of Variance (ANOVA) was performed, followed by a Tukey Test using GraphPad Prism V7.01 (http://www.graphpad.com). Reference genes were analysed for stability using Ref Finder (http://leonxie.esy.es/RefFinder/).

2.2.11 A-tailing of purified PCR products

A-tailing of blunt ended PCR products was performed by using Klenow Fragment (New England Biolabs) as per the manufacturer's protocol, containing 500ng of template. Products were purified and quantified as described in section *2.2.7*.

2.2.12 Cloning of PCR products into TOPO2.1 vector

A-tailed PCR products were cloned into the pCR™2.1-TOPO® vector (TOPO TA Cloning Kit, Invitrogen, Life Technologies). 1µl of pCR™2.1-TOPO® vector, 3µl

of purified PCR product, 1µl of salt solution (1.2M NaCl, 0.06M $MgCl_2$), and 1µl of water (both provided in the TOPO TA Cloning Kit) were mixed by gently flicking the tube and left to incubate for 30min at RT. The remainder of the reaction was performed as described by the manufacturer, using Transform One Shot® TOP10 competent cells (Life Technologies, Invitrogen). Ampicillin (50µg/ml – Sigma-Aldrich) was used as the selectable marker on Nutrient Agar (OXOID, UK) plates and incubated overnight at 37°C. Blue/white screening was performed using 40mg/ml of X-gal (VWR International, USA). Successful colonies were grown in 5ml of Nutrient Broth (OXOID) containing 50µg/ml of ampicillin overnight, at 37°C and shaking at 200rpm.

2.2.13 Plasmid DNA extraction

Plasmids were extracted from the overnight cultures using the QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's protocol. The resulting plasmid DNA was stored at -20°C.

2.2.14 PCRs to confirm the presence of cloned gene

PCRs were performed on the plasmid DNA, using the T7 promoter - TAATACGACTCACTATAGGG and M13R – CAGGAAACAGCTATGAC primers to confirm the insert was present (see appendix A for vector map). PCR products were cleaned up as described previously (see section *2.2.7*).

2.2.15 Sequencing using BigDye Terminator V3.1

Sequencing reactions were performed using 5X big dye buffer and the Big Dye Terminator V3.1 sequencing kit (Applied Biosystems, California USA) following the manufacturer's protocol with one amendment: in a 10µl reaction 2µl of Betaine (Sigma-Aldrich) was added to create a final concentration of 1M (See Chapter 3, section 3.3.4, Table 3.7 for table of sequencing primers).

2.2.16 Purification of sequencing reactions

The sequencing reactions were transferred from 0.2ml PCR tubes to 1.5ml centrifuge tubes and reactions cleaned up by adding the following: 15µl of DEPC-treated water (Life Technologies, Invitrogen), 2µl of 125mM Ethylenediaminetetraacetic acid, (EDTA-Fisher Scientific), 2µl of 3M sodium acetate (pH 5.2, Ambion by Life Technologies, Invitrogen) and 50µl of 100% ethanol. Tubes were left to incubate at RT for 15mins and then centrifuged at 13,000rpm for 30min at 4^{o}C (pre-cooled centrifuge). The supernatant was removed and 70µl of 70% ethanol was added to each tube and centrifuged at 13,000rpm for 15min at 4^{o}C.

The ethanol solution was removed and the lid of the tube left open for 20min at 37^{o}C to air dry the pellet. These were sent for sequencing on a 3730xl DNA Analyser at the Zoology Department at Oxford University.

2.2.17 Analysis of sequence data

Sequence data was analysed using Sequencher version 4.1.4 (Gene Codes Corporation). Using the tools available in Sequencher, a contig was generated, and a consensus sequence exported for downstream investigations.

2.2.18 Gene identification using BLAST and syntenic evaluation

Gene sequence data was used to perform nucleotide and translated BLAST searches via the NCBI database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to confirm the identity of the cloned gene. Syntenic comparisons were investigated using Ensembl (http://www.ensembl.org/index.html).

2.2.19 Double restriction digest of pCR™2.1-TOPO® UGT clones for sub-cloning

The consensus sequence obtained for each gene was placed into NEBcutter V2.0 (http://nc2.neb.com/NEBcutter2/) to find restriction enzymes to excise the whole gene from the TOPO2.1 pCR™2.1 vector which did not digest the insert, and to also facilitate directional sub-cloning into the expression vector pcDNA™3.1 (for vector map see appendix B, Invitrogen by Life Technologies). This identified two suitable enzymes per gene (Table 2.2). A double digest was performed to excise each *UGT* from the vector. The mammalian expression vector pcDNA™3.1 was also digested with the same enzymes as those used to excise the gene from pCR™2.1.

Each digest contained 250ng of cloned gene plasmid or 500ng of the expression vector. The reaction also contained 0.25µl of each enzyme (Promega, USA), 0.2µl bovine serum albumin (BSA), 2µl digest buffer, with the volume made up to 20µl with sterile water. Reactions were incubated at 37°C for 2hrs and heat inactivated at 70°C for 15mins. Digests were visualized on a 0.7% 7 x 7cm agarose gel in a horizontal electrophoresis chamber (Biorad) to verify linearization. Bands of the correct size (Table 2.2) were excised via the gel extraction method (see section *2.2.7*) and quantified as described in *2.2.7*.

Gene (Ensembl ID)	**Enzyme 1**	**Enzyme 2**	**Buffer**	**Product Size (bp)**
ENSECAG00000023519	*BamHI*	*XbaI*	B	1723
ENSECAG00000014362	*XhoI*	*BamHI*	C	1744

ENSECAG00000010396	*BamHI*	*XhoI*	C	1651

Table 2.2: List of enzymes used to excise genes from pCR™2.1 for sub-cloning into pcDNA™3.1. Each of the genes was excised from pCR™2.1 via double digest, with the pcDNA™3.1 vector digested with the same enzymes to create directional sub-cloning. Optimal buffer for the double digest was selected, and digests checked on an agarose gel, with products (bp) gel excised, purified and quantified**.**

2.2.20 Ligation of genes into pcDNA3.1 expression vector

For ligation of UGT genes into pcDNA™3.1 a 3:1 insert to vector ratio was used.

$$\frac{\text{Vector (ng)} \times \text{Size of insert (kb)}}{\text{Size of vector (kb)}} \times \text{Molar ratio of } \frac{\text{Insert}}{\text{Vector}} = \text{ng of Insert}$$

Calculated quantities of insert and vector were mixed with 1µl of Ligase 10X buffer (Promega), 1µl of T4 DNA ligase (Promega) and made up to a total volume of 20µl prior to incubation for 18hrs at 15°C.

Chemically competent *E.coli* (TOP10) were transformed with the ligated products, as described in the TOPO TA Cloning Kit (Invitrogen, Life Technologies), and plated onto Nutrient Agar (OXOID) containing 50µg/ml ampicillin (Sigma-Aldrich) overnight at 37°C. Colonies were randomly selected and grown overnight in Nutrient Broth (OXOID) containing ampicillin.

2.2.21 Endofree plasmid DNA extraction from E.coli

The plasmid DNA was extracted from the overnight cultures using the Qiaprep Spin Miniprep Kit (Qiagen) and PCRs were performed, using T7 promoter and BGH primers, to confirm the presence of an insert as described (sees section *2.2.5.2*). PCR products were purified (see section *2.2.7*) and sequenced, using T7 promoter and BGH, (see sections *2.2.15 and 2.2.16*) to confirm the orientation of the gene in the expression vector. Once a successfully cloned product was identified a fresh starter culture was grown in 5ml of Nutrient Broth (OXOID) containing 50µg/µl ampicillin (Sigma-Aldrich) for 8hrs at 37^{o}C, with shaking at 300rpm. 100µl of the starter culture was used to inoculate 100ml of Nutrient Broth containing 50µg/ml ampicillin. The inoculated broth was incubated at 37^{o}C for 16hrs at 300rpm. The plasmid DNA was extracted from the culture using the Qiagen Plasmid Maxi Kit (Qiagen) as per the manufacturer's instructions.

2.2.22 Linearization of pcDNA3.1^{TM} plasmid

To create a stably transfected cell line, the plasmid DNA containing the sub-cloned *UGT* gene required digestion to a linear state. A restriction enzyme was chosen which would only digest the pCR^{TM}2.1 vector containing the gene within the multiple cloning site. The enzyme *SacI* (Promega) was utilized to digest pCR^{TM}2.1/*UGT* clones. 50µl digestion reactions were set up and incubated as per the manufacturer's protocols. Digests were visualised on a 0.7% agarose gel alongside undigested plasmid DNA. Digests were cleaned up using the PCR Purification kit as directed by the manufacturer (see section *2.2.7*).

2.3 Cell culture techniques

2.3.1 HEK 293 cell culture

Human Embryonic Kidney 293 cells (HEK 293 – Sigma-Aldrich) cells were chosen as the expression system for the isolated *UGTs*.

The adherent cells were cultured in complete media in Nunc® cell culture T75 flasks (Thermofisher) at 37°C and 5% CO_2 until 80% confluent. Complete media contained Dulbecco's Modified Eagle medium (DMEM, Gibco, UK), 10% foetal calf serum (Gibco), 1% Penicillin-streptomycin (Gibco), and 2mM L-Glutamine (Gibco) warmed to 37°C for 30mins prior to use.

2.3.1.1 Media changes

Media was aspirated and cells washed with 2ml PBS/EDTA (Phosphate Buffered Saline 'PBS' – Gibco and 2mM EDTA – Fisher Scientific). The PBS/EDTA was aspirated and 10ml of fresh complete media added. The flasks were immediately placed back into the 37°C, 5% CO_2 incubator for 24hrs before visually checking confluency.

2.3.1.2 Passaging cells

Cells were passaged when ~80% confluent. Media was aspirated and cells washed with 2ml of PBS/EDTA. The PBS/EDTA was aspirated, and 2ml of 0.5% Trypsin (Trypsin EDTA – Gibco) was added to the cells and the flask was placed in the incubator at 37°C for 2mins. Cells were visually checked to ensure full detachment from the flask, and 12ml of complete media was mixed with the trypsinized cells. 1ml of trypsinized cells was added to a new flask containing 10ml of complete media. The flask was placed back in the incubator, 37°C, 5% CO_2 and monitored after 24hrs.

2.3.1.3 Counting HEK293 cells

Media aspirated and cells were washed with 2ml of PBS/EDTA, which was then fully removed. Cells were incubated at 37°C, 5% CO_2 with 2ml of trypsin for 2mins. 10ml of pre-warmed completed media was added to the flask and carefully mixed with the trypsinized cells. The total volume was transferred to a 50ml tube (Falcon, Fisher scientific) and centrifuged at 125g for 10mins. The media was aspirated and the pelleted cells were re-suspended in 5ml of complete media. 10µl of cells were mixed with 10µl of Trypan blue stain 0.4% (Gibco) and counted using a haemocytometer.

2.3.2 Geneticin study

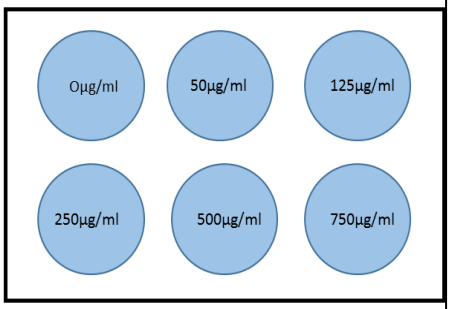
6-well plates (Nunc cell culture – Thermofisher) were seeded with $4x10^3$ HEK293 cells in 2ml of complete media. Plates were incubated at 37°C, 5% CO_2 for 48hrs to allow cells to adhere and begin to proliferate. After the 48hrs cells were provided with a media change (see section *2.3.1.1*) with 2ml of fresh, pre-warmed complete media added. Plates were numbered with the date the cells required to be counted on. The quantity of Geneticin (Sigma-Aldrich) was then calculated (Figure 2.3) and the correct amount was added to each individual well to create the final concentrations as follows: 0, 50, 125, 250, 500, 750, 1000, 2000, 5000µg/ml. The equation and an example of the calculation are shown in figure 2.3.

Plates were incubated at 37°C, 5% CO_2 for 48hrs with Geneticin, after which one plate was taken to perform a cell count (see section *2.3.1.3*). The remaining plates were provided with a media change (section *2.3.1.1*) including fresh Geneticin, and returned to the incubator, with conditions as described, for a further 48hrs. This was repeated until all counts had been performed. The data was then analysed using MS Excel to generate a kill curve

and determine the optimal concentration of Geneticin to create a stable transfected cell line.

C1 V1 = C2 V2

C1 = Stock concentration of Geneticin = 50mg/ml (50,000µg/ml)
V1 = **unknown**
C2 = Concentration in well
V2 = Volume of media in well (2ml)



C1 V1 = C2 V2

50,000µg/ml x V1 = 750µg/ml x 2ml

$$V1 = \frac{750\cancel{\mu g/ml} \times 2ml}{50{,}000\cancel{\mu g/ml}}$$

V1 = 0.03ml
V1 = 30µl

Figure 2.3: Example calculation of the amount of Geneticin to achieve the correct concentration per well. The amount of Geneticin was calculated from a stock of 50mg/ml to generate final concentrations in µg/ml. A template of the plates is provided, with the higher concentrations of 1000, 2000 and 5000µg/ml set up in a separate plate.

2.3.3 Transfection of HEK293 cells

In 6-well plates (Nunc cell culture), 3 wells were seeded with the following number of cells in 2ml of pre-warmed complete media; 6.25 x 10^5, 3.25 x10^5, and 6.25 x 10^4. All plates were incubated overnight at 37^{o}C, 5% CO_2.

Prior to transfection all media was aspirated from the cells and replaced with 500µl of fresh, pre-warmed complete media, and placed back in the incubator whilst the transfection solution was prepared, for approximately 40mins.

The transfection solution contained 100µl of Opti-MEM reduced serum media (Gibco) mixed with 0.5ug of template, either a positive control (pcDNA3.1^{TM}/*CAT*) or pcDNA3.1^{TM} (*UGT1A6*, *UGT2A3* or *UGT3A2*) and a negative control, with water replacing the template. To each tube, 3.75µl of Lipofectamine LTX (Thermo Fisher Scientific) and 2.5µl of Plus reagent (Thermo Fisher Scientific) was added and mixed by gentle pipetting. The transfection solution was left to incubate at RT for 30mins.

Upon completion of the incubation, 100µl of the transfection solution was added to each of the seeded wells, and gently mixed with the complete media. Plates were placed into an incubator for 24hrs at 37^{o}C, 5% CO_2. After the 24hrs incubation, the transfection solution was aspirated and 2ml of pre-warmed complete media was added, plates placed back into the incubator under the aforementioned conditions and left for 48hrs.

2.3.4 Selection of stably transfected cells

After 48hrs of incubation, complete media was aspirated and cells washed with 1ml of PBS/EDTA. 2ml of fresh, pre-warmed complete media was added to each well, and then 500µg/ml of Geneticin was added and gently mixed. Cells were provided with media changes (containing Geneticin) every 48hrs. Once cells were 80% confluent, they were passaged (see section *2.3.1.2*) and

placed into in Nunc® cell culture T25 flasks (Thermo Fisher Scientific), containing 5ml of pre-warmed complete media and incubated for 48hrs at 37°C, 5% CO_2. Cells grown in the T25 flasks until 80% confluent and then passaged (see section *2.3.1.2*) into T75 flasks containing 10ml of pre-warmed complete media and incubated in the same conditions as above for 48hrs. Once cells were established 500µg/ml of Geneticin was used to maintain the transfected cell line until several flasks were >80% confluent.

2.3.5 Total protein extraction and quantification from transfected HEK293 cells

Once flasks were >80% confluent, the media was fully aspirated and cells washed with 3ml of PBS/EDTA. PBS/EDTA was aspirated, and cells were detached by incubating in 5ml of PBS (Gibco) in the fridge for approximately 5mins. Cell suspensions were transferred to 15ml tubes (Falcon, Fisher Scientific) and centrifuged for 4min at 800g. The PBS was aspirated carefully to leave a pellet at the bottom of the tube, and cells were re-suspended in 5ml of PBS prior to centrifugation for 5min at 2000g. The PBS was aspirated and the cell pellet was re-suspended in 1ml of PBS and placed on ice. The cell suspension was sonicated using an Ultrasonic processor (VC 505 Sonics Vibra-cell™), whilst on ice, to disrupt the cell membranes (5 x 5sec bursts). Lysates were transferred to 1.5ml microcentrifuge tubes and centrifuged at 3,000rpm for 10mins and the resulting protein extract transferred to fresh 1.5ml tubes. The total protein extract was quantified using the Qubit™ Protein Assay Kit (Invitrogen - Thermo fisher) and stored at -20°C.

2.3.6 Detection of recombinant protein by western blots

2.3.6.1 Antibodies and positive controls

Western blot analysis was performed using antibodies designed to detect the equine recombinant UGT protein using antibodies; anti-UGT1A6 (SAB2102641), anti-UGT2A3 (SAB1407972) and anti-UGT3A1 (SAB1408324) (Sigma-Aldrich). Each western blot included the human recombinant UGT protein with a GST tag, UGT1A6 (ABNOH054578-Q01-25), UGT2A3 (ABNOH079799-P01-25) and UGT3A1 (ABNOH133688-P01-25) recombinant proteins (VWR International) as positive controls.

2.3.6.2 Western blot analysis

In 1.5ml microcentrifuge tubes, 20µl of total protein lysate or 20µg of respective positive control were mixed with 10µl of Laemmli 2x concentrate (Sigma-Aldrich) and heated to 100^{o}C for 10mins to denature the proteins.

The samples were loaded onto a 10% SDS-PAGE gel (Bio-Rad) in Tris-Glycine-SDS Buffer (Bio-Rad, UK) alongside 5µl of Precision plus protein ladder (Bio-Rad), and 200V applied to the gel for 40mins. A section of nitrocellulose membrane (0.45µm pore size, Sigma-Aldrich) was soaked in a solution of 1X Tris buffered saline (Bio-Rad) containing 500µl of Tween (Bio-Rad). Post-electrophoresis, a semi-dry transfer was performed; blotting paper, nitrocellulose membrane, gel, and then blotting paper were layered, and 20V was applied to the Trans-Blot SD semi-dry transfer cell (Biorad) for 1hr.

Confirmation of the transfer was performed by staining the nitrocellulose membrane with Ponceau S solution (Sigma-Aldrich) with gentle agitation at 10rpm for 5mins to visualise the ladder and bands. The membrane was carefully removed to visually confirm the transfer. The Ponceau S solution was removed prior to incubation with the antibodies. This was achieved by washing the nitrocellulose membrane three times with the Tris-Tween solution for 5min with gentle agitation at 70rpm.

The nitrocellulose membrane was then incubated at RT, with gentle agitation at 36rpm for 1hr, in a 5% Bovine Serum Albumin (BSA) (Semi skimmed, powdered milk, VWR International) blocking solution. The primary antibody (anti-UGT for relevant protein) was diluted 1:1000 in 3ml of 5% BSA to create a working solution in a 50ml tube. A test for reference protein was also performed using the primary antibody to Histone H3 (D1H2) (Cell signalling Technology, UK). The nitrocellulose membrane was placed into the 50ml tubes with the primary antibody (Sigma-Aldrich) and placed on a roller for 1hr at RT. The membrane was removed and washed with the Tris-Tween three times for 5mins each, with rocking at 70rpm.

The secondary antibody (either Goat anti-Rabbit secondary antibody, HRP or Goat anti-Mouse IgG secondary antibody, HRP, Thermo Fisher Scientific) was diluted 1:5000 in a 3% milk solution (Semi skimmed, powdered milk, VWR International) made with Tris-Tween solution in a 50ml tube. The nitrocellulose membrane was carefully removed from the Tris-Tween wash and placed into the 50ml tube with the secondary antibody (Sigma-Aldrich), and then placed on a roller for 1hr at 20rpm. Following incubation the nitrocellulose membrane was washed again with Tris-Tween solution three times for 5mins each, with rocking at 70rpm.

After the final wash step, the nitrocellulose membrane was gently blotted and placed on cling film. 1ml of chemiluminescent western blot detection substrate (ECL – Biorad) and 1ml of Buffer (Biorad) were mixed and carefully pipetted over the membrane and left to incubate for 5mins. The membrane was visualised by X-ray film developed using the SRX-101A developer (Konica Minolta, UK).

Chapter 3: Identifying and isolating novel equine UDP-glucuronosyltransferases

3.1 Introduction

3.1.1 Features characteristic of the UGTs

UGTs possess features characteristic of proteins which require translocating to and anchoring in membranes. At the beginning of the polypeptide is a signal peptide which is responsible for directing the protein to the ER membrane. In UGTs this is encoded in the first 22 amino acids (Ahn et al., 2012). Following cleavage, the remainder of the N-terminus is responsible for determining substrate specificity of the mature protein (Guillemette, 2003, Rowland et al., 2013).

The C-terminus has three important features: the transmembrane domain, dilysine motif, and a signature sequence. The transmembrane domain acts to anchor the enzyme to the ER membrane and is composed of 16 hydrophobic amino acids (Mackenzie et al., 2008, Ouzzine et al., 1999). Within the last five amino acids of the polypeptide is a dilysine motif, **K**X**K**XX, with the lysines situated at either position -3, -4, and -5 from the C-termini end (Andersson et al., 1999, Meech and Mackenzie, 1997a). The third key feature within the C-terminus, and located upstream of the transmembrane domain, is a signature sequence, which is 44 amino acids in length and required for binding the UDP component of the donor sugar (Meech et al., 2012a, Bock, 2016).

3.1.2 Creating an UGT in vitro toolbox

Research into the two phases of metabolism in the horse is lagging behind similar research in humans, mice, rats, and even canines. To date there are no recombinant equine UGT tools available for *in vitro* investigations. The little research that has been conducted has used microsomes and homogenised liver to assess the metabolites formed in phase I and Phase II reactions (Wong et al., 2016).

Work characterising the horse transcriptome of immunologically active tissues identified the liver enriched for phase 1 and phase 2 enzymes (Moreton et al., 2014). Here we report the isolation and characterisation of equine *UGT*s from liver tissue.

3.2 Methods and Materials

3.2.1 Sequence analyses

The horse is an economically valuable animal, particularly in sports such as equestrian eventing and horse racing. Additional to their sporting value, they offer a better model for human disorders than murine for certain diseases, yet relatively speaking they are not a well genetically characterised animal. A paper published in 2014 described the transcriptome of immunologically active tissues from the genome of the thoroughbred mare Twilight, EquCab2 – GCA000002305.1 (Moreton et al., 2014). A list of predicted equine UGTs was identified from this publication and provided as a starting point for this research by Professor R. Emes (Table 3.1 - unpublished data). The 12 genes were named by their Ensembl identification number, and information on gene expression was quantified and normalised by calculating the reads per kb per million (RPKM) values per tissue, lymphocyte, jejunum, kidney, liver, lymph node, and spleen, with 10 of the 12 genes more highly expressed in the liver compared to the other six tissues analysed. Two genes, ENSECAG00000020628 and ENSECAG00000023519, displayed particularly high expression, RPKM >60, denoted in red, relative to the other tissues analysed (Table 3.1).

The nucleotide and amino acid sequences of the predicted equine *UGTs* were retrieved from the Ensembl database http://www.ensembl.org/index.html, and analysed using the NCBI BLAST tool http://www.ncbi.nlm.nih.gov/ (Altschul et al., 1990) to confirm the predicted sequence was a *UGT*.

Orthologues of full length *UGTs* in 21 species, including mammals, insects and plants, were identified from the NCBI database (www.ncbi.nih.gov/protein/?item=) and 91 amino acid fasta files retrieved (see appendix C). Geneious (Biomatters Ltd - http://www.geneious.com/) was used to produce an alignment of the sequences and infer their relationships

using the Maximum Likelihood (ML) method with the percentage of replicate trees in the taxa clustered in the bootstrap test (500 replicates). Trees were annotated using iTOL Interative Tree Of Life (https://itol.embl.de/).

The *UGT1* and *UGT3* sub-families were analysed in greater depth, with a phylogenetic comparision of each sub-family. The *UGT1* tree included 72 amino acid sequences from 23 species, using orange and barley as the outgroup (see appendix D). The phylogenetic tree for the *UGT3* sequences involved 15 species and 20 amino acids sequences, with orange used as the outgroup (see appendix E). Many of the sequences used in both trees are predicted and have not been experimentally proven.

		RPKM (0.d.p)						
Ensembl Gene ID	Predicted UGT	Lymphocyte (a)	Jejunum (c)	Kidney (c)	Liver (c)	Lymph node (c)	Lymphocyte (b)	Spleen (c)
ENSECAG00000008247	ENSECAG00000008247|ENSECAT00000009198	0	8	0	15	0	0	0
ENSECAG00000008900	equCab2_uc003jjz.2	0	0	3	23	0	0	0
ENSECAG00000010396	equCab2_uc003jjy.2, ENSECAG00000010396|ENSECAT00000010690	0	2	1	46	0	0	0
ENSECAG00000010718	ENSECAG00000010718|ENSECAT00000011007	0	0	0	33	0	0	0
ENSECAG00000014362	equCab2_uc010ihs.3, ENSECAG00000014362|ENSECAT00000016051	0	8	0	8	0	0	0
ENSECAG00000017275	ENSECAG00000017275|ENSECAT00000018212, equCab2_uc003heg.4	0	5	0	1	0	0	0
ENSECAG00000017801	ENSECAG00000017801|ENSECAT00000018809	0	2	0	0	0	0	0
ENSECAG00000018165	equCab2_uc021xov.1	0	0	0	18	0	0	0
ENSECAG00000019112	ENSECAG00000019112|ENSECAT00000020254	0	3	1	2	0	0	0
ENSECAG00000020628	equCab2_uc011clo.2, ENSECAG00000020628|ENSECAT00000022670, equCab2_uc003heh.3	0	0	0	74	0	0	0
ENSECAG00000023519	ENSECAG00000023519|ENSECAT00000025670	0	39	60	9	0	0	0
ENSECAG00000024269	ENSECAG00000024269|ENSECAT00000026123	0	2	0	1	0	0	0

(a) "Twilight", healthy Thoroughbred (b) healthy castrated male welsh mountain pony (c) aged gelding euthanized for arthritis.

Table 3.1: Ensembl gene identification of sequences predicted to be *UGT*s within the equine genome (genome assembly EquCab2: GCA_000002305.1). Gene expression in the lymphocyte, jejunum, kidney, liver, lymph node and spleen were quantified and normalized by calculating the reads per Kb per million (RPKM), values shown per tissue. Data provided by Professor R. Emes, unpublished. Numbers in red indicate the tissues with high expression levels (≥60 RPKM).

3.3 Results

RNA was extracted from the liver of a single horse for the downstream application of isolating selected UGTs.

3.3.1 Optimisation of total RNA extractions

Use of the RNeasy Mini Kit as described by the manufacturer (see *section 2.2.2*) resulted in a low yield of total RNA. The extraction method required optimisation, and several modifications were attempted to increase the quality and quantity of total RNA extracted.

During the first homogenisation process, fragments of liver remained visible, and it was this step that was targeted for optimisation prior to use of the RNeasy Mini kit. Two alternative optimisations of the initial protocol were tested, with mechanical disruption of the tissue through bead-beating in RLT-β-Mercaptoethanol retained as the initial step. One method added a subsequent incubation step with Proteinase K for 1 hour, while the second added an additional homogenisation step through a QIAshredder prior to incubation with Proteinase K for 1 hour. Both methods increased the yield of total RNA, 771.2ng/µl (23.163µg in 30µl) and 551ng/µl (16.53µg in 30µl) respectively. For the extraction method involving the proteinase K without a second homogenisation step, although the yield was good the quality of the total RNA was not as good as that for the sample that had two homogenisation steps: 260/280 value of 1.67 compared to 1.94 (summarized in Table 3.2).

A further comparison of the double homogenisation method was made with a separate RNA extraction protocol using MACs tubes to homogenise the tissue and Qiazol as the lysate buffer, with one sample methodology including an additional proteinase K step for 1 hour. Although the method using MACs tube and Qiazol homogenisation produced RNA of very similar quality to that obtained in the previous method, the yield was considerably lower at 70ng/µl (2.1µg total). The use of MACs tubes, Qiazol and the additional proteinase K significantly improved the

yield of total RNA extracted - 3120ng/µl (93.6µg in 30µl). The quality also improved, with a 260/280 ratio of 1.99.

A

Optimisation	Quantity Total ng	Quality (260/280 ratio)
No amendments to manufacturers protocol	471	2.38
Proteinase K incubation	23,136	1.67
QiaShredder + Proteinase K incubation	16,530	1.94
MACs tube and Qiazol homogenisation	2100	2.03
MACs tube, Qiazol and Proteinase K incubation	93,600	1.99

B

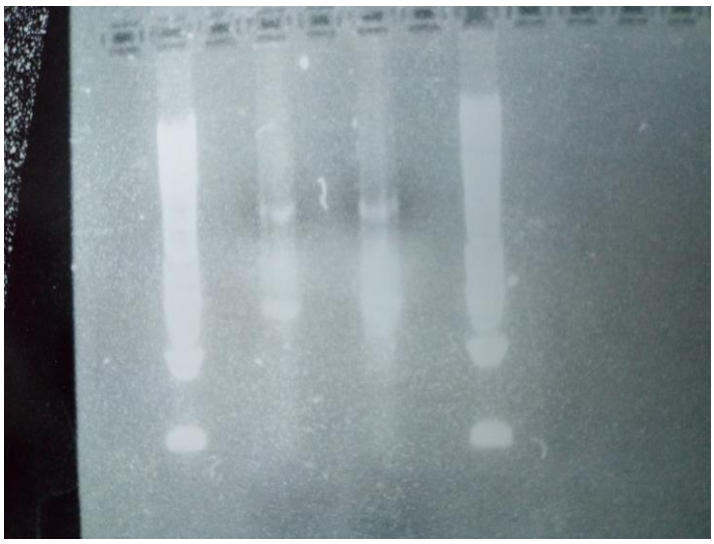


Table 3.2: Summary of the optimisation steps and the quantity and quality of total RNA extracted. A) Following the manufacturer's protocol for the RNeasy Mini Kit without amendment resulted in low yield and poor total RNA quality. Introducing additional homogenisation steps increased the quantity of total RNA produced and the quality. The use of MACs tube, Qiazol and proteinase K for the homogenisation yielded the highest quantity of total RNA, and the best quality 260/280 ratio, 1.99. B) Image of RNA on a 0.7% agarose gel, post stained.

3.3.2 Optimisation of PCR primers – determination of annealing temperatures

Primers were designed to isolate the entire coding region, from start to stop codon, of each *UGT* enzyme from the list provided by Dr Emes (*See section 3.2.1, Table 3.1*). Primer pairs for each of the 12 genes (Table 3.3) were tested in a gradient PCR to empirically determine the range of annealing temperatures over which the primer pairs produced a PCR product of approximately the correct size. Of the 12 genes, five primer pairs produced a product; Ensembl ID *ENSECAG00000008900* (predicted *UGT3A1*), *ENSECAG00000010396* (predicted *UGT3A1*), *ENSECAG00000014362* (predicted *UGT2A1*), *ENSECAG00000020628* (predicted *UGT2B31*) and *ENSECAG00000023519* (predicted *UGT1A6*). The remaining primer pairs did not yield any products.

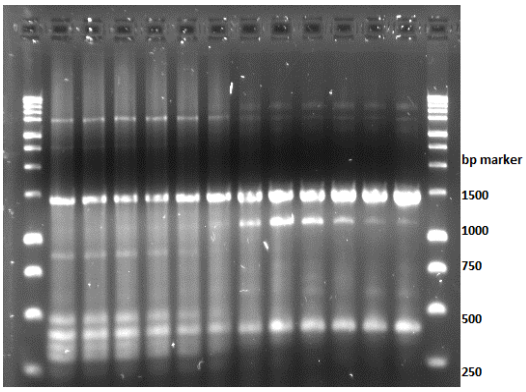
Using gradient PCR, the temperature range for producing a band of the correct size following PCR ranged from 48^{o}C to 70^{o}C (Figure 3.4). Gene *ENSECAG00000014362* displayed a relatively narrow annealing temperature range with 7^{o}C variability, 52-59^{o}C. PCR using primers to *ENSECAG00000008900*, *ENSECAG00000020268* and *ENSECAG00000023519* all produced bands of approximately the correct size over the full 12-degree range tested and *ENSECAG00000010396* over a 15-degree range. Table 3.5 summarises the temperature range over which the primers were empirically shown to produce a PCR product and the annealing temperature at which the final gene product was isolated.

Touchdown PCR was attempted to reduce the number of secondary products, as seen with genes *ENSECAG00000014362* and *ENSECAG00000008900*, with a final annealing temperature set at 54^{o}C, and while this reduced the number of secondary bands it did not fully eliminate them (Figure 3.4).

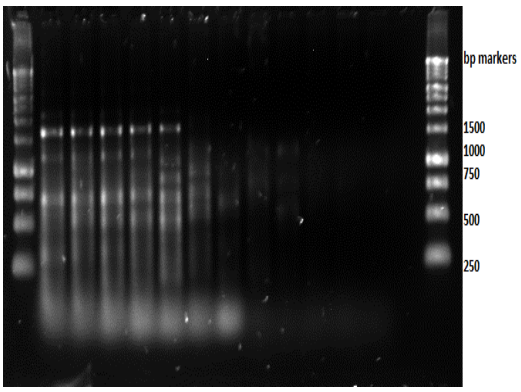
Ensembl gene ID	Predicted UGT	Primer	Primer sequence 5' - 3'	Amplicon Size (bp)
ENSECAG00000023519	1A6	1	F: ATGGCTCCTGCAATGTTCGA	1629
		2	R: AAGTTTGGTTCACTTCCCAC	
ENSECAG00000026028	2B31-like	3	F: GCATTCCACCAAGATGTCTCT	1607
		4	R: CGGGTATAGCTACTCCCT	
ENSECAG00000008247	2C1-like	10	F: ATGAAGACTGCAAAAGGC	1590
		11	R: CTACTCTCTCTTTTTCTTC	
ENSECAG00000008900	3A1-like	12	F: ATGGGGAGCCTGCGGGC	1521
		13	R: CCCACGGAGAGCACCAG	
ENSECAG00000013096	3A1-like	14	F: ATGATGAGGCCACGGG	1572
		15	R: TCAGGCCTTCTTCAGTTC	
ENSECAG00000014362	2A1	16	F: ATGGCGTCTGAGAAATG	1665
		17	R: TTTATTAAGGTCATTGTGGGC	
ENSECAG00000017275	2B31-like	18	F: ATGTCTCTGAAATGG	1587
		19	R: CTACTCCCTTTTTTCC	
ENSECAG00000017801	2C1-like	20	F: ATGAAGACTGTGGAAGG	1590
		21	R: CTACATTCTCTTTTTCTTC	
ENSECAG00000018165	2B31-like	22	F: ATGTCTCTGGAATGGATTTCAC	1587
		23	R: CTACTCCTTTTTTCCTTC	
ENSECAG00000019112	2C1-like	24	F: ATGAAGACTGTACAAGG	1590
		25	R: TTATTCTCTCTTTTTCTTC	
ENSECAG00000024269	8	26	F: ATGAAGTCTTACACTGCG	2134
		27	R: TCTGCATTCAGTTTTGAGC	

Table 3.3: Table of PCR primers utilized to isolate the full coding sequence of the Uridine diphosphate glucuronosytransferases (*UGTs).* The Ensembl identification of the predicted equine *UGTs* are listed along with the predicted isoform name. Each primer pair was designed to amplify the complete coding sequence. The expected base pair (bp) amplicon size is shown.

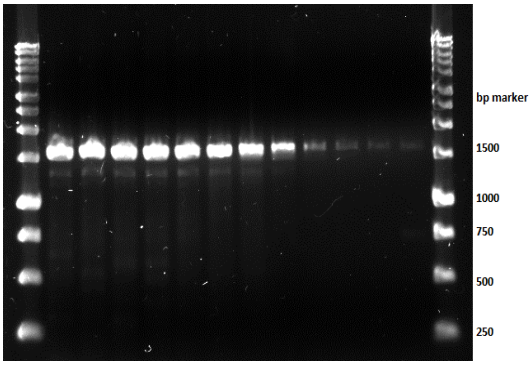
1. *ENSECAG00000008900*
predicted *UGT3A1*



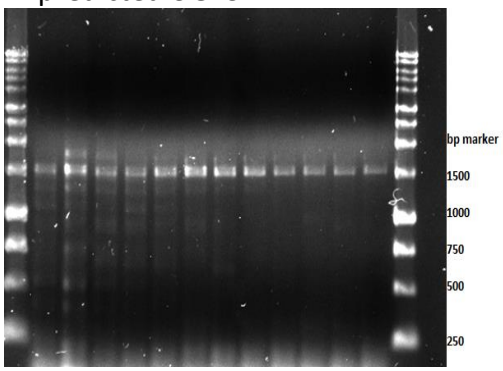
4) *ENSECAG00000014362*
predicted *UGT2A1*



2. *ENSECAG00000020628*
predicted *UGT2B31*



5) *ENSECAG00000010396*
predicted *UGT3A1*



3. ENSECAG000000203519
predicted *UGT1A6*

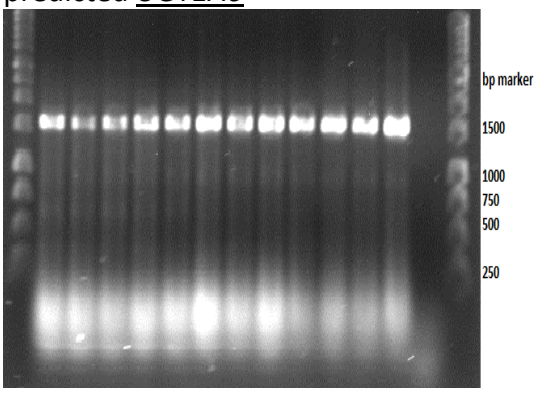


Figure 3.4: Gel images of gradient PCRs. Each PCR was performed over a 12°C range, with the calculated hypothetical annealing temperature in the centre of this range. 1) Gene *ENSECAG00000008900* yielded a product over a temperature range of 48-60°C. 2) Gene *ENSECAG00000020628*, annealing temperature range of 58-70°C. 3) Gene *ENSECAG00000023519* produced an amplicon between 55-67°C. 4) Gene *ENSECAG00000014362*, 52-59.2°C. 5) Gene *ENSECAG00000010396* produced an amplicon 45-60°C. 1kb ladder from Promega.

Gene identifier	Product Size	Gradient temperature range °C	Selected annealing temperature °C
8900	1450bp	48 – 60	54 (T)
10396	1572bp	45 – 60	58
14362	1665bp	52 – 59	54 (T)
20268	1607bp	58 – 70	58
23519	1629bp	55 – 67	65

Table 3.5: Table displaying the genes which yielded a product over a gradient PCR. Table provides the Ensembl ID number, the length of the coding sequence, range of temperatures over which a product of the correct size were produced and the annealing temperature selected for downstream PCRs. (T) = those primer pairs where touchdown PCR was used for further optimisation.

3.3.3 TOPO TA cloning

Cloning into the pCR™2.1 vector was an intermediary step to facilitate sequencing the full length gene and downstream sub-cloning into an expression vector.

Multiple optimisation steps were required to improve the transformation efficiency with an increase in the incubation step from 10mins to 30mins, with gentle mixing throughout, proving to be optimal.

Initial PCRs performed to excise the entire multiple cloning site within the UGT-pCR™2.1 clones using primers to the vector pCR™2.1, M13 forward and M13 reverse (see appendix A for vector map) produced two bands following visualisation of the PCRs on the agarose gels. Alignment of the primer sequences to all the cloned genes to the M13 Forward primer found there to be sufficient similarity to create a secondary PCR product. Therefore, an alternative primer, T7 promoter, was assessed and used alongside M13 reverse, to amplify the cloned genes (and regions of the multiple cloning site) from the TOPO 2.1 pCR™2.1 vector.

3.3.4 Optimisation of sequencing reactions

The first attempts at sequencing the *UGT* genes from the vector, using primers M13 reverse and T7 promoter and primers designed within the gene, 400bp apart (Table 3.6), were unsuccessful. The result was either too much background noise or no signal.

Initial attempts to optimise sequencing involved altering the conditions of the thermal cycler. Firstly, a reaction was set up with an initial incubation step of 96°C for 10mins, increased from the original protocol of 1min. A second reaction increased the initial temperature from 96°C to 98°C, incubation time remained at 1min. The third reaction involved an incubation step of 98°C for 10min, for this reaction only the volume was doubled to 20µl (summarised in Table 3.7). None of these modifications improved the quality of the sequence data generated.

Ensembl Gene ID	Primer	5' - 3'
ENSECA0000020628	F1	GAAAAGCTCTGTAAAGATGC
	R1	GCTACTTTACAGAGCTTTTC
	F2	TCACTGCAAACCTACCAAACC
	R2	GGTTTGGTAGGTTTGCACTGA
	F3	CATTGTTCACATGAAGGC
	R3	GCCTTCATGTGAACAATG
	F4	TGTTGATCCCAACAAACCC
	R4	CTGGAAGACGGTGTTTGGCTCC
ENSECAG0000023519	F1	GTAAAATACCTAGAAGAG
	R1	CTCTTCTAGGTATTTTAC
	F2	ATGGTTTTTATTGGTGGGATC
	R2	GATCCCACCAATAAAAACCAAT
	F3	ATTCTGATGATTTAGCAAATG
	R3	CATTTGCTAAATCATCAGAAG
	R4	TCTAGGAACATAGGAAGGAGC
ENSECAG0000008900	F1	AGTTTTGACTTCTGTCCTT
	R1	AAGGACAGAAGTCAAAACT
	F2	CTCAAGGAGATGAACACT
	R2	AGTGTTCATCTCCTTGAG
ENSECAG0000014362	F1	TGTAGTGGTTATAGACCCTG
	R1	CAGGGTGTATAACCACTACA
	F2	TTTCCTCGTCCATACTTACC
	R2	GGTAAGTATGGACGAGGAAA
	F3	GATTTGCTCAATGCCTTGAG
	R3	CTCAAGGCATTGAGCAAATC
ENSECAG0000010396	F1	GGATTTGACTTGTGTTCTC
	R1	GAGAACACAAGTCAAATCC
	F2	TTCAAGGAGTGAACAGG
	R2	CCTGTTCACTCCTTGAA

Table 3.6: A list of primers designed for the sequencing of each of the isolated *UGT* genes. Primers were utilised to sequence along the entire coding sequence (see appendix E1-5 for gene information showing primer locations).

Optimisation	1	2	3
Temperature (°C)	96	98	98
Time (min)	10	1	10
Volume (µl)	10	10	20

Table 3.7: Summary of the initial optimisations. Initial optimisations involved altering the cycling conditions. Changes involved increasing the denaturation temperature, the length of time of the denaturation step and the volume of the sequencing reaction.

Failure of cycling conditions to improve sequence data led to altering the chemical conditions. Four reactions, containing different reagents were set up using the same 40ng template to compare their effect (Kieleczawa, 2006), whilst using the standard cycling conditions;

1. Dimethyl sulfoxide (DMSO) – final concentration 5%
2. Glycerol – final concentration 5%
3. DMSO + Glycerol – final concentration 5% each
4. Betaine – final concentration 1M

A comparison of the sequence data showed the addition of 1M of Betaine to the sequencing reaction improved the quality of the data returned.

Sequence data was obtained for each of the five genes. The data for *ENSECAG00000008900*, *ENSECAG00000020628* and *ENSECAG00000014362* is discussed, alongside investigations into the identity of the *UGT* isolated (*see sections 3.3.6, 3.3.7 and 3.3.8*). Results for *ENSECAG00000023519* and ENSECAG00000010396 are reported in sections 3.3.9 and 3.3.10 respectively.

3.3.5 Sequence analysis of the predicted UGT genes

The list of predicted *UGT* sequences provided by Dr R. Emes did not provide information on the identity of the orthologue. Sequence analysis of the *UGT* amino acid sequences from 21 species, representing the UGT1, UGT2, and UGT3 proteins was performed to show their relationships, with plants, orange (ACS87992.1) and barley (ADC92549.1), as the outgroup.

Figure 3.8 displays the ML tree, with the largest group comprising *UGT1* sequences (indicated by the orange band around the outside of the tree). Within the *UGT1* clade there is clear clustering of certain isoforms, for example eight out of nine *UGT1A6* sequences cluster together and the eight *UGT1A1* sequences cluster, suggesting that these isozymes have retained sequence conservation; the singular piscine representative, Zebrafish, has a unique *UGT1B* subfamily (yellow band). The *UGT2* family is represented by 32 amino acid sequences, which are divided into two clades (Figure 3.8); *UGT2As*, indicated by a dark blue band, and the *UGT2Bs* represented by the lighter blue band. A small cluster of *UGT2C-like* sequences is present in the clade, composed of predicted equine UGT sequences.

The smallest clade on the tree comprises the *UGT3* amino acid sequences (indicated by the green band) with 14 sequences included in the analysis. Within this clade primate sequences cluster together but are divided clearly into *UGT3A1* and *UGT3A2*. Representing the ungulates is a single cow sequence, *UGT3A1,* and two predicted equine sequences, both termed *UGT3A1-like*.

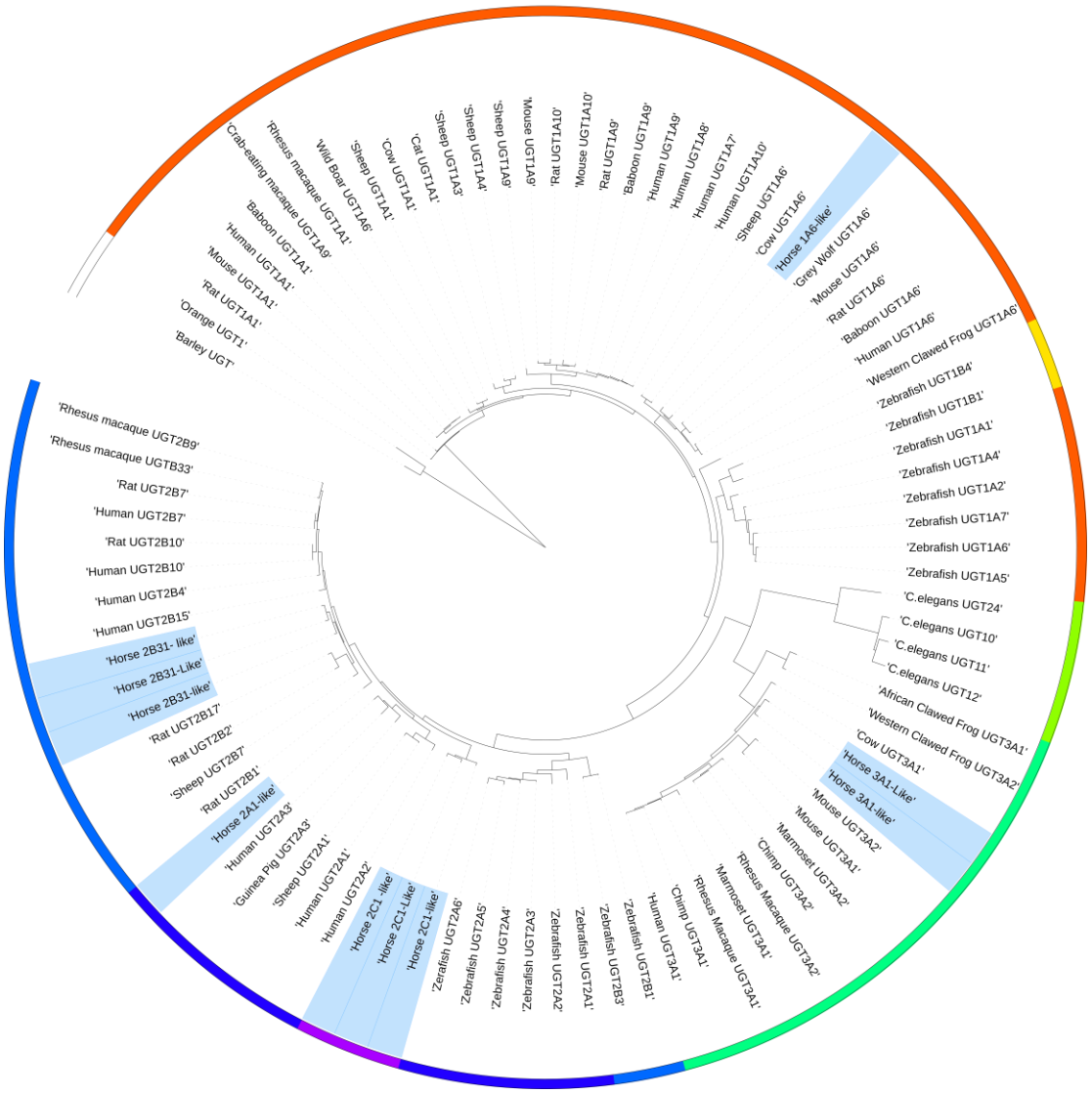


Figure 3.8: A Maximum Likelihood (ML) consensus tree. 91 Amino acid sequences, confirmed and predicted, from 21 species were retrieved from the NCBI database and analysed using maximum likelihood (ML), bootstrapped 100 times, Geneious (Biomatter Ltd). The horse sequence names (highlighted pale blue) are predicted and have not been experimentally proven. The *UGT1* family is indicated by the orange band around the tree, The UGT2s are displayed with a blue band, dark blue for the UGT2A sub-family and light blue for the UGT2B sub-family, the purple band indicates the predicted equine 2C1-like sequences. UGT3 sequences are emphasized by the green band. Two plant sequences, barley (ADC92549.1) and orange (ACS87992.1) have been used as outgroup. Full list of accession numbers for each sequence present in appendix C.

3.3.6 Analysis of ENSECAG00000008900

The list of predicted sequences provided by Dr R. Emes highlighted *ENSECAG00000008900* as a gene encoding a UGT enzyme, however it did not provide information on the identity of the orthologue. Both the nucleotide and amino acid sequences were retrieved from the Ensembl database and were analysed phylogenetically to determine which group of *UGT* sequences *ENSECAG00000008900* clustered with (Figure 3.8). Analysis found it clustered with *UGT3A1* and *UGT3A2* sequences from the human, rat, mouse, and cow genomes.

An enlarged view of the *UGT3* clade (Figure 3.9) highlights the separate clusters of the primate, ungulates, and rodent sequences. The primate sequences divide into two subfamilies, *UGT3A1* (bright green band) and *UGT3A2* (pink band). The rodent sequences are separate from that of the ungulates, which include the two predicted equine sequences, and the cow *UGT3A1* sequence (dark blue band). The predicted horse and cow *UGT3A1* amino acid sequences share over 78% of sequence identity at the amino acid level. A *UGT3* sequence from orange (XP_006469356.1) was utilised as outgroup.

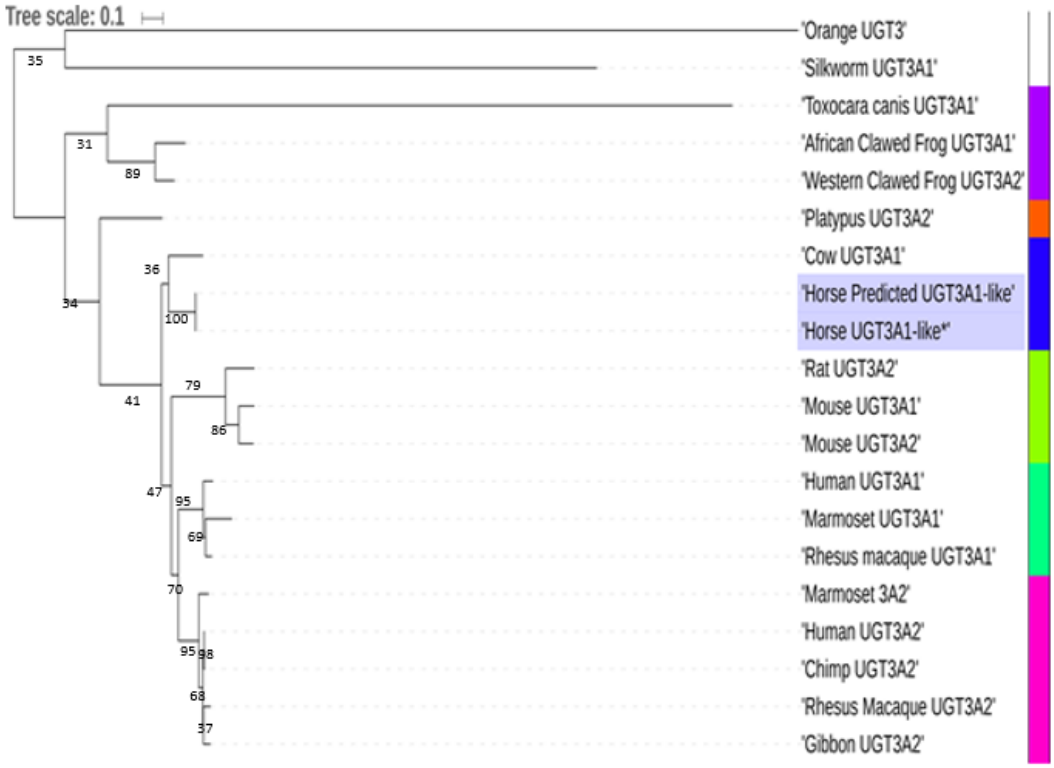


Figure 3.9: A Maximum Likelihood (ML) consensus tree of *UGT3-like* sequences. *UGT3* amino acid sequences were retrieved from the NCBI and analysed against the predicted equine *UGT3* sequences, using ML and bootstrapped 100 times. The equine genome has two predicted *UGT3A1*-like sequences, *UGT3A1-like** has not been investigated in this chapter. With the following exceptions; human, mouse and frog, all sequences taken from the NCBI are predicted. Primate sequences divide into *UGT3A1* (pink) and *UGT3A2* (bright green). Equine sequences cluster with the cow *UGT3A1*. The Orange *UGT3* (XP_006469356.1) sequence was used as outgroup.

There were insufficient sequences in the tree to establish whether *ENSECAG00000008900* is an orthologue of *UGT3A1* or *UGT3A2*. An alignment of *UGT3A1* and *UGT3A2* from human and mice, and *UGT3A1* from cow and rat (rat was a predicted sequence) showed the equine predicted *UGT3* sequence to be 16 amino acids shorter at the C-terminus. In terms of sequence homology, the equine sequence was 78% homologous to both bovine and murine *UGT3A1.* When compared to the human and mouse sequences, both equine *UGT3A1-like* sequences share a greater homology with the *UGT3A2* polypeptide, 74.75% and 65.61% for human and mice *UGT3A2* compared to 72.39% and 64.82% respectively for *UGT3A1*.

3.3.6.1 Syntenic investigation

BLAST analysis of sequence *ENSECAG00000008900* found it matched with a gene annotated in NCBI as equine *UGT3A1-like*. This gene also matched to other equus *UGT3A1-like* genes from the donkey and Przewalski's horse respectively.

A syntenic investigation found *ENSECAG00000008900*, which is encoded on horse chromosome 21, location 29,204,050-29,222,886, corresponds to a region on human chromosome 5, location 5:35,951,010-36,001,028 (Figure 3.10).

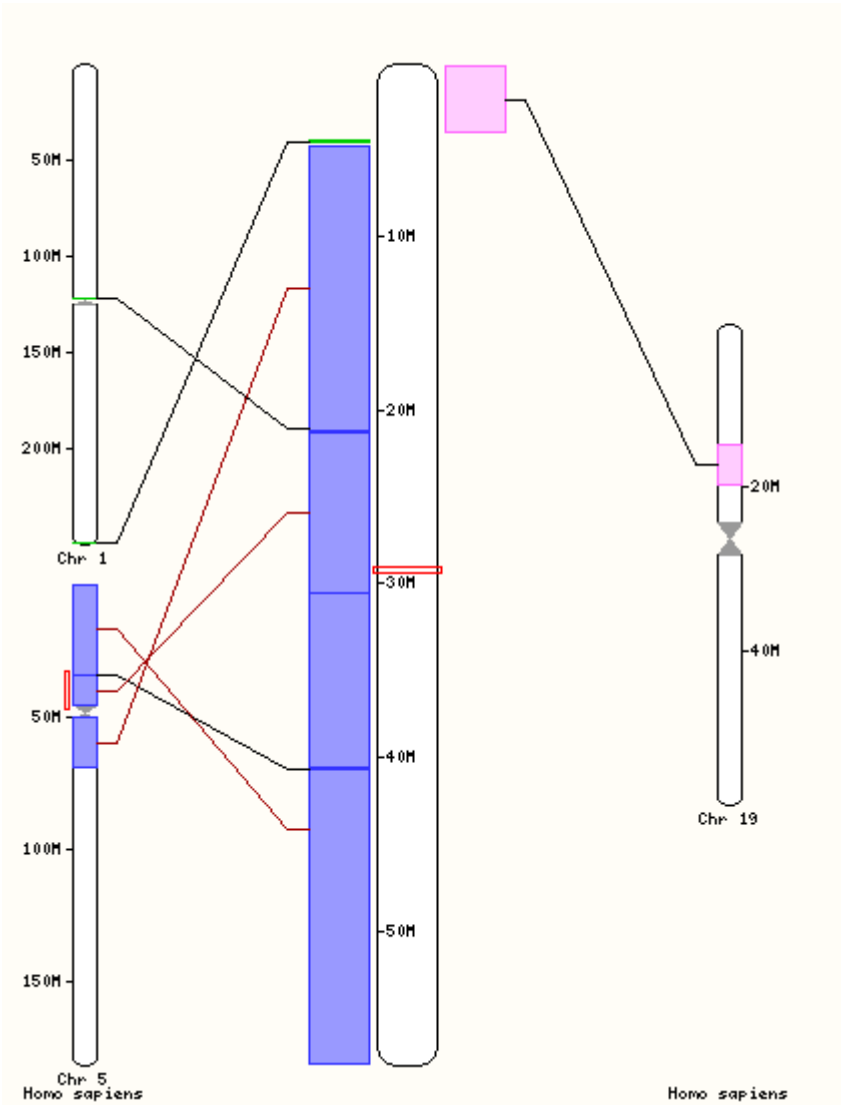


Figure 3.10: Syntenic comparison of Equine *ENSECAG00000008900* (UGT3A1-like) and the human genome. The diagram (taken from Ensembl – release 89) shows blocks of equine chromosome 21 conserved in human chromosomes 1, 5 and 19. The area of interest (demarcated by a red box) on the equine chromosome encodes *ENSECAG00000008900*, this and the surrounding region is present as a conserved block on human chromosome 5.

Looking at the specific region encoding *ENSECAG00000008900*, it showed two adjacent genes named with Ensembl identities which, when compared to the list provided by Dr Emes, were also predicted *UGT* sequences. The corresponding region on human chromosome 5 contained two *UGT3* sequences, *UGT3A1* and *UGT3A2*, encoded adjacently (Figure 3.11).

Several genes in the region around the *UGT3* locus are maintained between species. Downstream of *ENSECAG00000008900* in the equine genome three genes are encoded, Calcyphosine like (*CAPSL*), Interleukin 7 Receptor (*IL7R*), and Sperm Flagellar 2 (*SPEF2*). These are also located downstream of the *UGT3s* locus on the human genome. Upstream of the human and equine loci, LMBR1 Domain Containing 2 (*LMBRD2*), S-Phase Kinase Associated Protein 2 (*SKP2*), NAD Kinase 2, Mitochondrial (*NADK2*) and RAN Binding Protein 3 Like (*RANBP3L*) are conserved (Figure 3.11).

As humans and rats each encode two isoforms of *UGT3* the other genes in the region surrounding *ENSECAG00000008900* in the equine genome were checked to determine whether there may be a second UGT encoded within this region. Two additional UGT genes were identified with Ensembl identifiers *ENSECAG00000010396* and *ENSECAG00000010718*.

Analysis of the size of each of these genes found *ENSECAG00000010718* to be much shorter at 642 nucleotides (214 amino acids) while *ENSECAG00000010396* was 1572 nucleotides (523 amino acids) in length. BLAST analysis of both genes found both *ENSECAG00000010396* and *ENSECAG00000010718* were annotated on NCBI as *UGT3A1-like*.

A nucleotide sequence alignment of *ENSECAG00000008900*, *ENSECAG00000010396* and *ENSECAG00000010718* was performed (Figure 3.12), and showed a size difference between the three sequences. The difference in size between *ENSECAG00000008900* and *ENSECAG00000010396* is accounted for in the C-termini, with *ENSECAG00000008900* being 51 nucleotides (17 amino acids) shorter. *ENSECAG00000010718* is 642bp, 879bp shorter than *ENSECAG00000008900*,

aligning with both *ENSECAG00000008900* and *ENSECAG00000010396* 201 nucleotides (67 amino acids) from their respective start codons (highlighted green), with sequence homology matching across exons 3 and 4 of their N-termini. Annotation of *ENSECAG00000010718* revealed no start codon (ATG) though it does encode a stop codon (TAA).

Figure 3.13 displays the results of an alignment between *ENSECAG00000008900* and *ENSECAG00000010396*. These two predicted UGT3A1-like genes share 89.48% sequence homology at the nucleotide level, with the majority of the differences within the first half of the sequence. At the protein level these two genes share 83% of their amino acids.

A protein alignment with the human orthologues shows Equine ENSECAG00000008900 to be like both human UGT3A1 and UGT3A2, with 72.3% and 74.7% of the amino acids conserved between sequences respectively (Figure 3.14).

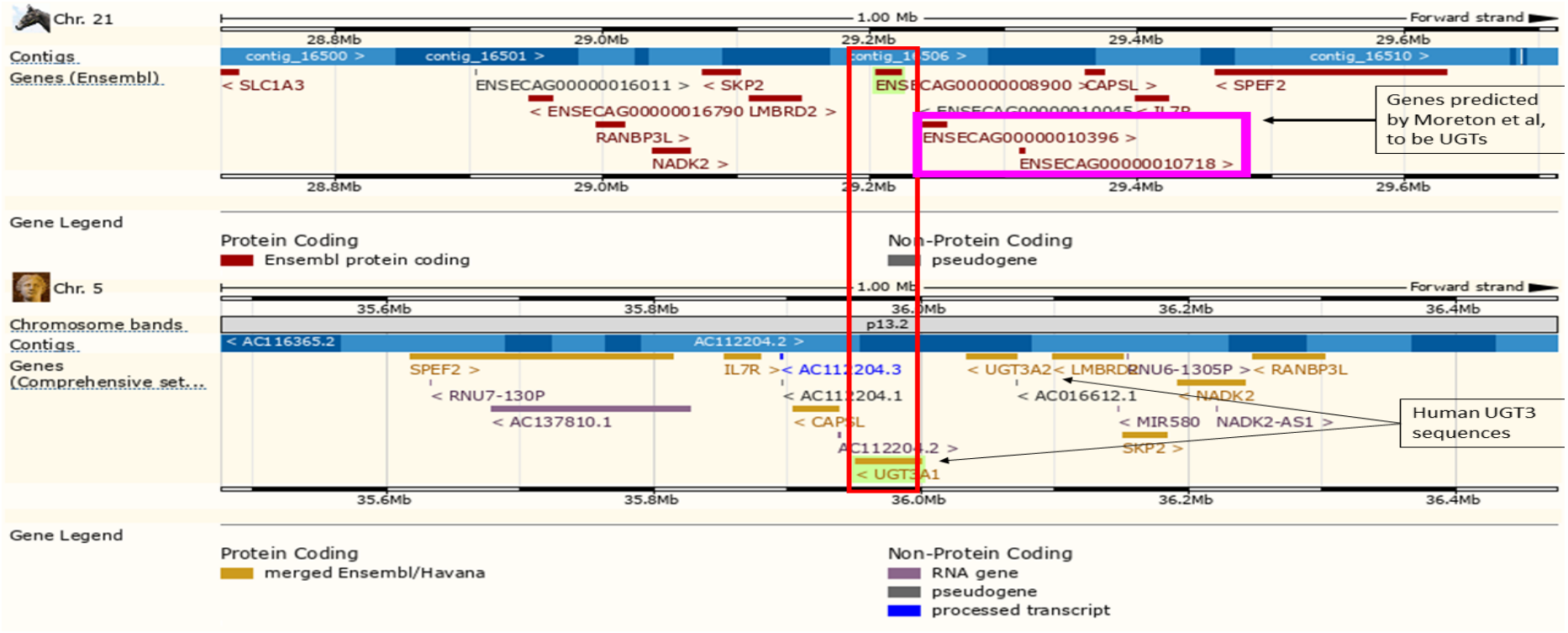
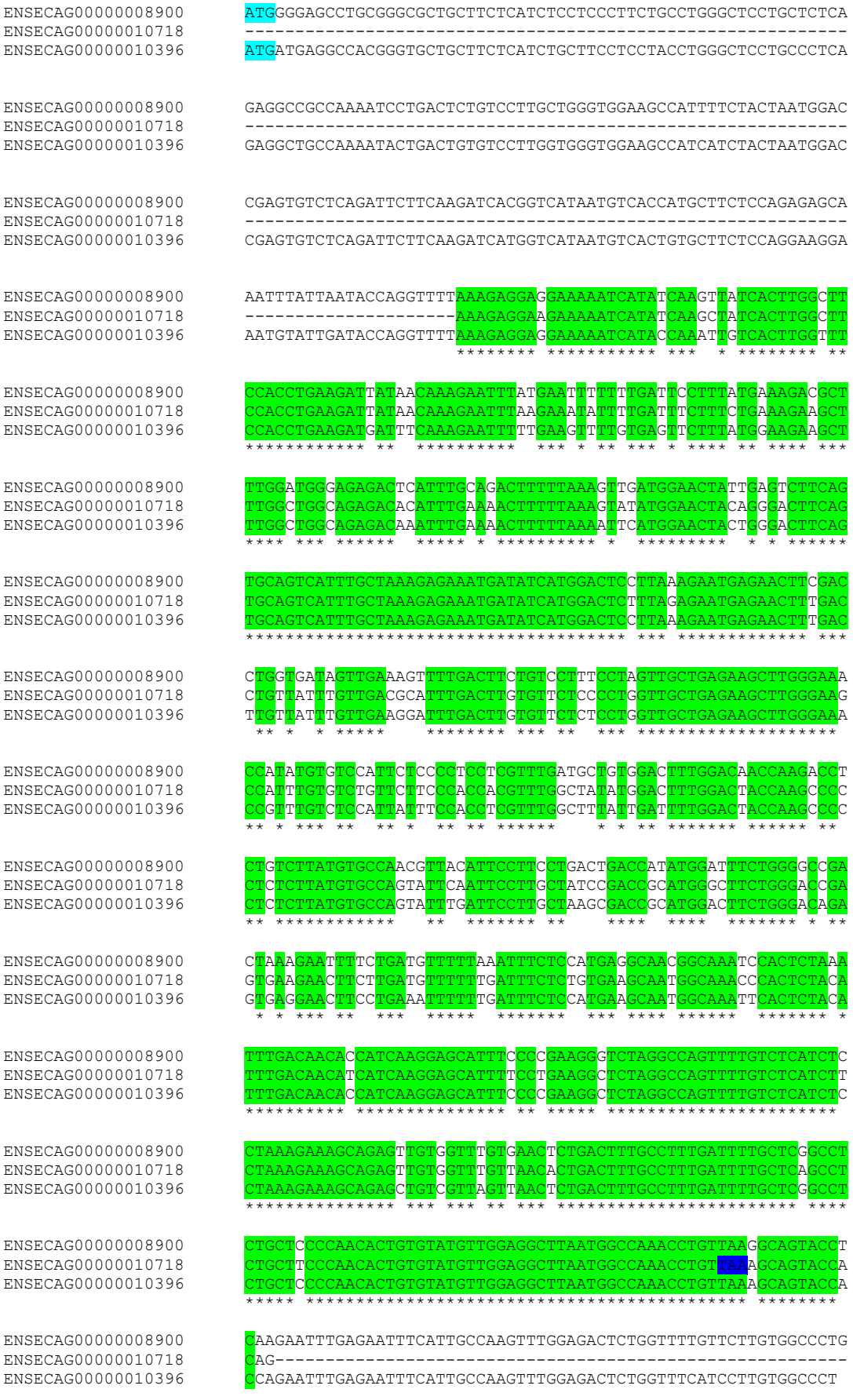


Figure 3.11: Region comparison between ENSECAG00000008900 and human *UGT3* loci. Comparing the equine gene *ENSECAG00000008900* to the human genome, found the gene to correspond to a locus on human chromosome 5 which encodes *UGT3A1*. The human genome encodes two *UGT3* members, *UGT3A1* and *UGT3A2*. The equine locus also contains two sequences predicted by Moreton et al, to be *UGTs*; *ENSECAG00000010396* and *ENSECAG00000010718*. There are several conserved genes between the equine and human loci, *SPEF2*, *IL7R* and *CAPSL* are conserved downstream of the *UGT3* locus and *SKP2*, *NADK2*, *LMBRD2* and *RANBP3L* are maintained upstream of the *UGT3* locus.


ENSECAG00000008900
ENSECAG00000010718
ENSECAG00000010396

```
ENSECAG00000008900      GGCTCCATGATCAGTGGCTCTTCATCCCAAGAATTTCTCAAGGAGATGAACACTGCCTTT
ENSECAG00000010718      ------------------------------------------------------------
ENSECAG00000010396      GGCTCTGTGGTGAACATCTTTCAGTCCCAGTATGTTTTCAAGGAGATGAACAGGGCCTTT

ENSECAG00000008900      GCTCATCTCCCTCAAGGGGTCTTATGGAGGTGTAAGCCTTCTCATTGGCCCAAAGACATC
ENSECAG00000010718      ------------------------------------------------------------
ENSECAG00000010396      GCTCATCTACCTCAAGGGGTGATATGGAAGTGTAATCCTTCTCATTGGCCTGAAGACATC

ENSECAG00000008900      AAATTAGCAGCAAATGTGAAAATTGTGGACTGGCTTCCTCAGAGTGACCTCTTGGCTCAC
ENSECAG00000010718      ------------------------------------------------------------
ENSECAG00000010396      AAATTGGCAGAAAATGTGAAAATTGTGGACTGGCTTCCTCAGAGTGACCTCCTGGCTCAC

ENSECAG00000008900      CCTCACATCCGTCTCTTTGTCACCCATGGTGGGATAAATAGCATCATGGAGGCCATCCAA
ENSECAG00000010718      ------------------------------------------------------------
ENSECAG00000010396      CCTCGCATCCGTCTCTTTGTCACCCATGGTGGGATAAATAGCATCATGGAAGCCATCCAA

ENSECAG00000008900      CATGGCGTTCCCATGGTGGGGATTCCCGTCTTTGGAGACCAGCCTGAAAACCTGTTCCGA
ENSECAG00000010718      ------------------------------------------------------------
ENSECAG00000010396      CATGGTGTGCCCATGGTGGGGATTCCCTTCTTTGGTGACCAGCCTGAAAATCTGTTCCGG

ENSECAG00000008900      GTAGAAGCCAAAAACTTTGGTGTCTCTATCCAGTTAAAGCAGATCAAGGCTGAGACACTG
ENSECAG00000010718      ------------------------------------------------------------
ENSECAG00000010396      GTAGAAGCCAAAAACTTTGGTGTCTCTATCCAGGTAAAGCAGATCAAGGCTGAGACACTG

ENSECAG00000008900      TCTCTGAAGATGAAGCAAGTCATAGAAGACAAGAGGTACAAATCTGCAGCCGTGGCCGCC
ENSECAG00000010718      ------------------------------------------------------------
ENSECAG00000010396      GCTCTGAAGATGAAGCAAGTCATAGAAGACAAGAGGTACAAATCTGCAGCCGTGGCCGCC

ENSECAG00000008900      AGCATCATCAGACGCTCCCACCCCCTCACTCCTGCCCAGCGGCTGGTGGGCTGGACCAAC
ENSECAG00000010718      ------------------------------------------------------------
ENSECAG00000010396      AGCATCATCAGGCGCTCCCACCCCCTGACTCCTGCCCAGCGGCTGGTGGGCTGGACCAAC

ENSECAG00000008900      CACATCCTGCAGACAGGGGGTGCAGCGCACCTCAAGCCCCATGCCTTCCAGCAGCCATGG
ENSECAG00000010718      ------------------------------------------------------------
ENSECAG00000010396      CACATCCTGCAGACAGGGGGTGCAGCGCACCTCAAGCCCCACGCCTTCCAACAGCCATGG

ENSECAG00000008900      TATGAACAGTACCTGCTCGATGTCTTCTTGTTCCTGCTGGTGCTCACCGTGGGCACCATG
ENSECAG00000010718      ------------------------------------------------------------
ENSECAG00000010396      TATGAACAGTACCTGCTCGATGTCTTCTTGTTCCTGCTGGTGCTCACCGTGGGCACCATG

ENSECAG00000008900      TGGCTCTGTGGGAAGCTGCTG----------------------------------------
ENSECAG00000010718      ------------------------------------------------------------
ENSECAG00000010396      TGGCTCTGTGGGAAGCTGCTGGGCATGGTGGCCAGGTGGCTGTGTGGGGCCAGGAAGCTG

ENSECAG00000008900      ------------
ENSECAG00000010718      ------------
ENSECAG00000010396      AAGAAGGCCTGA
```

Figure 3.12. Sequence alignment of the three-predicted equine UGT3 sequences. All three sequences are predicted to be *UGT3* sequences, all termed *UGT3A1-like* on NCBI. Sequences retrieved from the Ensembl database. The alignment shows *ENSECAG00000010718* to be the shortest of the three sequences, aligning to the central portion of the other two sequences. Homologous bases between all three sequences are highlighted green. *ENSECAG00000008900* is shorter than *ENSECAG00000010396* by 52 nucleotides. Start codons (light blue) are only present on two sequences, stop codons (dark blue) are present in all three sequences at different positions.


ENSECAG00000008900
ENSECAG00000010396
ATGGGGAGCCTGCGGGCGCTGCTTCTCATCTCCTCCCTTCTGCCTGGGCTCCTGCTCTCA
ATGATGAGGCCACGGGTGCTGCTTCTCATCTGCTTCCTCCTACCTGGGCTCCTGCCCTCA
GAGGCCGCCAAAATCCTGACTCTGTCCTTGCTGGGTGGAAGCCATTTTCTACTAATGGAC
GAGGCTGCCAAAATACTGACTGTGTCCTTGGTGGGTGGAAGCCATCATCTACTAATGGAC
CGAGTGTCTCAGATTCTTCAAGATCACGGTCATAATGTCACCATGCTTCTCCAGAGAGCA
CGAGTGTCTCAGATTCTTCAAGATCATGGTCATAATGTCACTGTGCTTCTCCAGGAAGGA
AATTTATTAATACCAGGTTTTAAAGAGGAGGAAAAATCATATCAAGTTATCACTTGGCTT
AATGTATTGATACCAGGTTTTAAAGAGGAGGAAAAATCATACCAAATTGTCACTTGGTTT
CCACCTGAAGATTATAACAAAGAATTTATGAATTTTTTTGATTCCTTTATGAAAGACGCT
CCACCTGAAGATGATTTCAAAGAATTTTTGAAGTTTTGTGAGTTCTTTATGGAAGAAGCT
TTGGATGGGAGAGACTCATTTGCAGACTTTTTAAAGTTGATGGAACTATTGAGTCTTCAG
TTGGCTGGCAGAGACAAATTTGAAAACTTTTTAAAATTCATGGAACTACTGGGACTTCAG
TGCAGTCATTTGCTAAAGAGAAATGATATCATGGACTCCTTAAAGAATGAGAACTTCGAC
TGCAGTCATTTGCTAAAGAGAAATGATATCATGGACTCCTTAAAGAATGAGAACTTTGAC
CTGGTGATAGTTGAAAGTTTTGACTTCTGTCCTTTCCTAGTTGCTGAGAAGCTTGGGAAA
TTGTTATTTGTTGAAGGATTTGACTTGTGTTCTCTCCTGGTTGCTGAGAAGCTTGGGAAA
CCATATGTGTCCATTCTCCCCTCCTCGTTTGATGCTGTGGACTTTGGACAACCAAGACCT
CCGTTTGTCTCCATTATTTCCACCTCGTTTGGCTTTATTGATTTTGGACTACCAAGCCCC
CTGTCTTATGTGCCAACGTTACATTCCTTCCTGACTGACCATATGGATTTCTGGGGCCGA
CTCTCTTATGTGCCAGTATTTGATTCCTTGCTAAGCGACCGCATGGACTTCTGGGACAGA
CTAAAGAATTTTCTGATGTTTTTAAATTTCTCCATGAGGCAACGGCAAATCCACTCTAAA
GTGAGGAACTTCCTGAAATTTTTTGATTTCTCCATGAAGCAATGGCAAATTCACTCTACA
TTTGACAACACCATCAAGGAGCATTTCCCCGAAGGGTCTAGGCCAGTTTTGTCTCATCTC
TTTGACAACACCATCAAGGAGCATTTCCCCGAAGGCTCTAGGCCAGTTTTGTCTCATCTC
CTAAAGAAAGCAGAGTTGTGGTTTGTGAACTCTGACTTTGCCTTTGATTTTGCTCGGCCT
CTAAAGAAAGCAGAGCTGTCGTTAGTTAACTCTGACTTTGCCTTTGATTTTGCTCGGCCT
CTGCTCCCCAACACTGTGTATGTTGGAGGCTTAATGGCCAAACCTGTTAAGGCAGTACCT
CTGCTCCCCAACACTGTGTATGTTGGAGGCTTAATGGCCAAACCTGTTAAAGCAGTACCA
CAAGAATTTGAGAATTTCATTGCCAAGTTTGGAGACTCTGGTTTTGTTCTTGTGGCCCTG
CCAGAATTTGAGAATTTCATTGCCAAGTTTGGAGACTCTGGTTTCATCCTTGTGGCCCTG
GGCTCCATGATCAGTGGCTCTTCATCCCAAGAATTTCTCAAGGAGATGAACACTGCCTTT
GGCTCTGTGGTGAACATCTTTCAGTCCCAGTATGTTTTCAAGGAGATGAACAGGGCCTTT
GCTCATCTCCCTCAAGGGGTCTTATGGAGGTGTAAGCCTTCTCATTGGCCCAAAGACATC
GCTCATCTACCTCAAGGGGTGATATGGAAGTGTAATCCTTCTCATTGGCCTGAAGACATC
AAATTAGCAGCAAATGTGAAAATTGTGGACTGGCTTCCTCAGAGTGACCTCTTGGCTCAC
AAATTGGCAGAAAATGTGAAAATTGTGGACTGGCTTCCTCAGAGTGACCTCCTGGCTCAC
CCTCACATCCGTCTCTTTGTCACCCATGGTGGGATAAATAGCATCATGGAGGCCATCCAA
CCTCGCATCCGTCTCTTTGTCACCCATGGTGGGATAAATAGCATCATGGAAGCCATCCAA
CATGGCGTTCCCATGGTGGGGATTCCCGTCTTTGGAGACCAGCCTGAAAACCTGTTCCGA
CATGGTGTGCCCATGGTGGGGATTCCCTTCTTTGGTGACCAGCCTGAAAATCTGTTCCGG

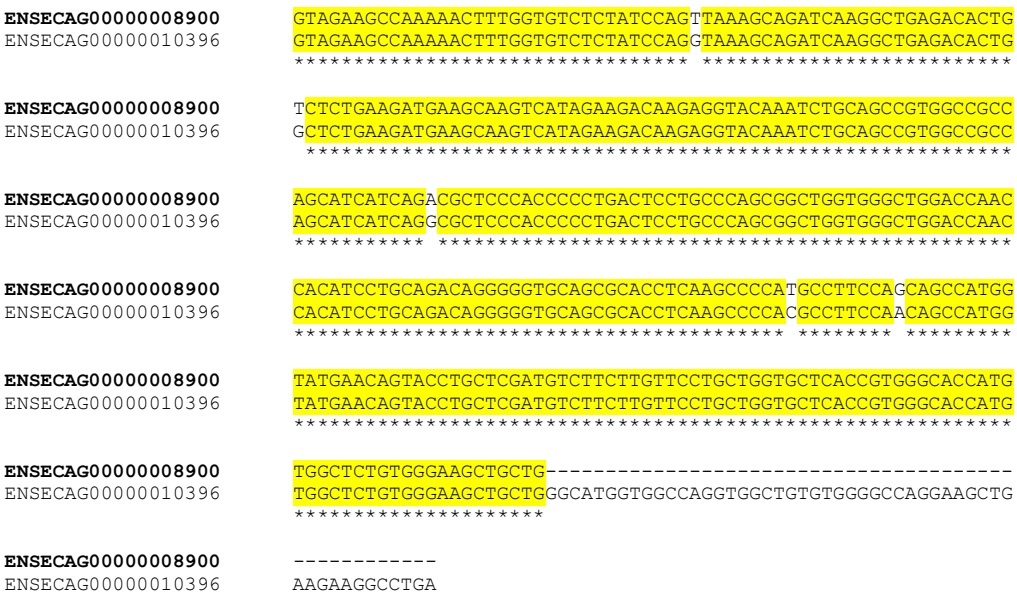


Figure 3.13: Alignment of two predicted equine UGT3A1-like sequences. The two genes, *ENSECAG00000008900* and *ENSECAG00000010396* were aligned at the nucleotide level to assess the homology between the two predicted sequences. *ENSECAG00000010396* is 51 nucleotides longer at the C-termini. *ENSECAG00000008900* and *ENSECAG00000010396* share 89% sequence homology, shared nucleotides highlighted yellow.

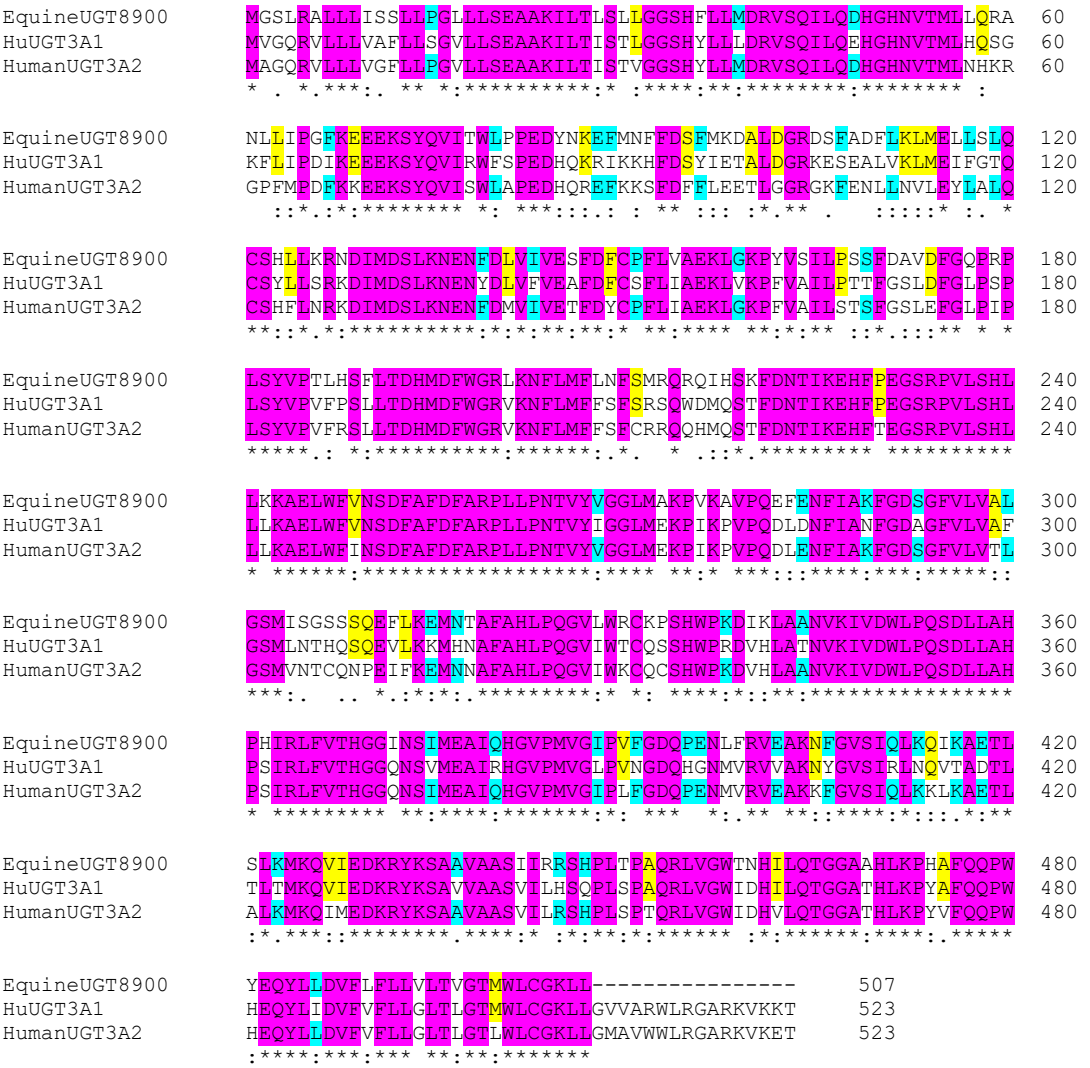


Figure 3.14: Alignment of Equine ENSECAG00000008900 and the human orthologues. The amino acid sequence of ENSECAG00000008900 was aligned against the human orthologues, UGT3A1 and UGT3A2. 64% of the amino acids are conserved between all three sequences (highlighted pink), 72.3% of the amino acids are conserved between the equine sequence and human UGT3A1 (highlighted yellow), and 74.7% of the amino acids are conserved between the equine sequence and human UGT3A2 (turquoise).

3.3.6.2 Annotating the sequence ENSECAG00000008900 for features characteristic of the UGT3s

A sequence of 1452bp was isolated and its translated sequence (484 amino acids) annotated for features characteristic to the *UGT* sequences (figure 3.15). The annotated sequence shows the signal peptide, the first 22 amino acids, as indicated by the blue box, while the green box highlights the signature sequence. This sequence lacks two features characteristic of the UGT enzymes; there is no transmembrane domain or dilysine motif at the end of the sequence, both of which anchor the UGT to the ER membrane.

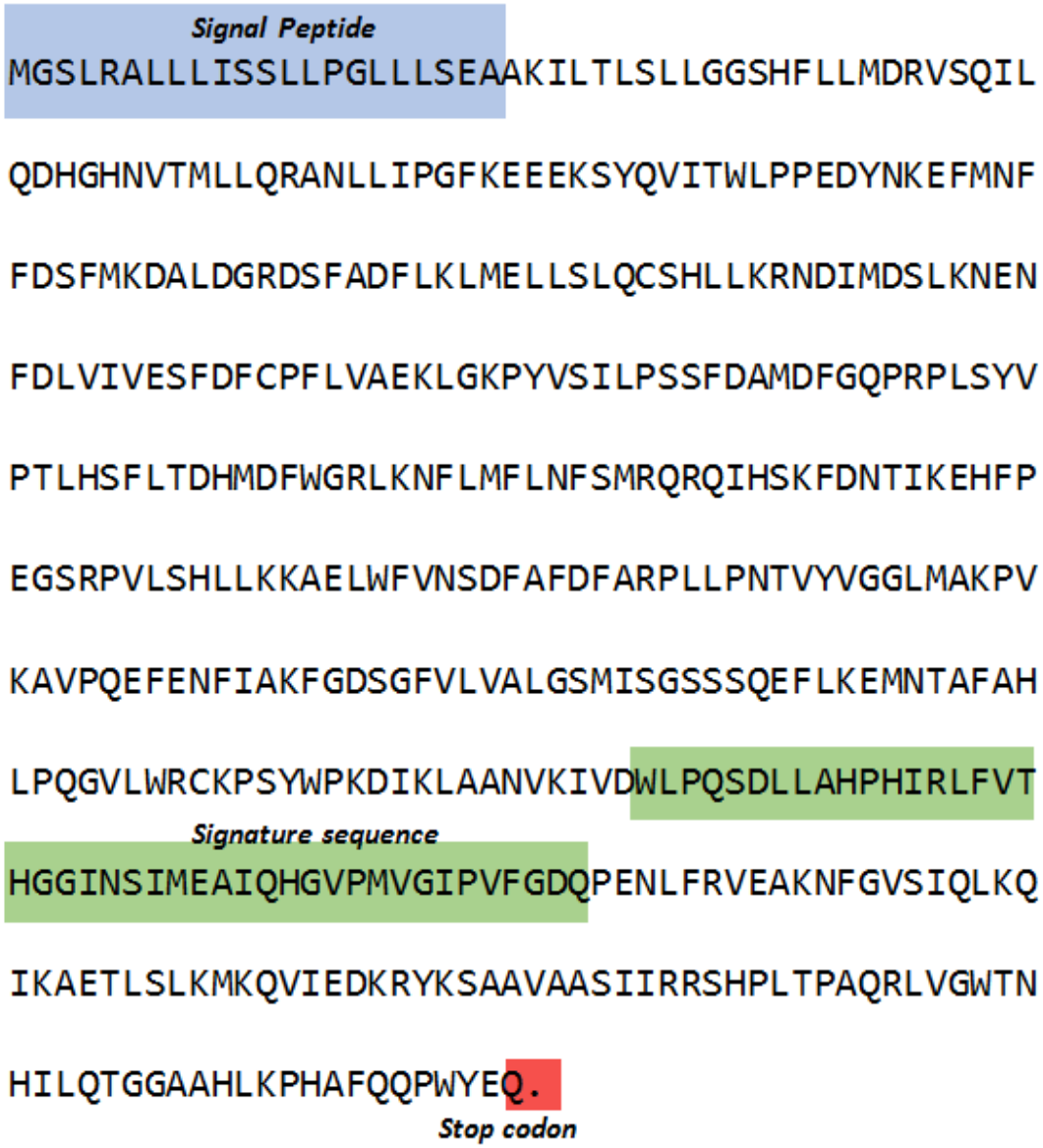


Figure 3.15. Annotation of the translated *UGT3A1-like* sequence. The translated sequence of the isolated *UGT* has been annotated for features characteristic of the UGT enzymes. The motif in the blue box is the predicted signal peptide, which direct the mature protein to the ER membrane. The green box indicates the signature sequence, which is responsible for binding the co-factor. Two important features are missing, the transmembrane domain and dilysine motif, both of which are required to anchoring the mature protein to the ER. The stop codon is represented by the full-stop, in red.

3.3.7 Determining the UGT identity of ENSECAG00000020628

ENSECAG00000020628 was a *UGT* identified in the transcriptome of immunological tissue and listed in the table provided by Dr R.Emes. BLAST analysis of this sequence revealed it was named as *UGT2B31-like*, with a high level of similarity to *UGT2B31-like* from the donkey and Przewalski's horse. The nucleotide and amino acid sequences were retrieved from the Ensembl database and analysed phylogenetically against a range of *UGT* sequences covering the *UGT1*, *UGT2* and *UGT3* families (See section 3.3.5, Figure 3.8). This sequence was found to cluster on the tree with *UGT2B* sequences from humans, rat, sheep, and rhesus macaque, none of which have the nomenclature of *UGT2B31*.

3.3.7.1 Syntenic comparisons of horse UGT2B-like regions with the human genome

A syntenic comparison of *ENSECAG00000020628* with the human genome found the equine gene, which is located on chromosome 3, location 66,060,357-66,070,907, corresponds to location 68,537,184-68,568,527 on chromosome 4 of the human genome.

Looking at both regions in detail shows that the predicted equine *UGT2B31-like* corresponds to human *UGT2B17*. This region on the human genome contains several *UGT2s*, including *UGT2B15, UGT2B10, UGT2A3* (Figure 3.16 – highlighted by green boxes) and a pseudogene *UGT2B29P* (indicated by a purple box). There is a lack of annotated genes within the equine *UGT* region which makes it difficult to establish how large the conserved region is and what genes have been conserved here. A single gene, YTH Domain Containing 1 (*YTHDC1*), is present in the equine region downstream of the predicted *UGT2B31-like* sequence, which is also present downstream of human *UGT2B17*.

Given the lack of annotation of this region of the equine genome and the number of UGTs encoded in the corresponding human region, this region could potentially encode equine UGT genes. A comparison of the Ensembl named genes against the list provided by Dr. Emes found two more genes in this region that the equine transcriptome work had predicted to be *UGTs* (Moreton et al., 2014). These are

labelled *ENSECAG000000107801*, *ENSECAG00000019112*, *ENSECAG00000017275* and *ENSECAG00000018165* (Figure 3.16 – highlighted by bright pink boxes). A BLAST analysis of each sequence in the NBCI database found ENSECAG00000017275 and *ENSECAG00000018165* are both predicted to be *UGT2B31-like*, whilst *ENSECAG000000107801* and *ENSECAG00000019112* are both predicted as *UGT2C1-like* enzymes.

ENSECAG00000020628 was aligned against the human *UGT2B17*, *UGT2B10*, *UGT2B15* and *UGT2A3* amino acid sequences to determine which gene it shared the greatest homology with. *UGT2B10* had the highest level of homology at 83.2%, however this sequence was considerably shorter than all other sequences at 280 amino acids, whereas the equine sequence and *UGT2B17* and *UGT2B15* are all 526 amino acids, *UGT2A3* is 527 residues. *ENSECAG00000020628* was 80% homologous to human *UGT2B17* and showed least homology with human *UGT2A3* (63.5%). Figure 3.17 displays the conserved amino acid residues between equine *ENSECAG00000020628* and human *UGT2B17*.

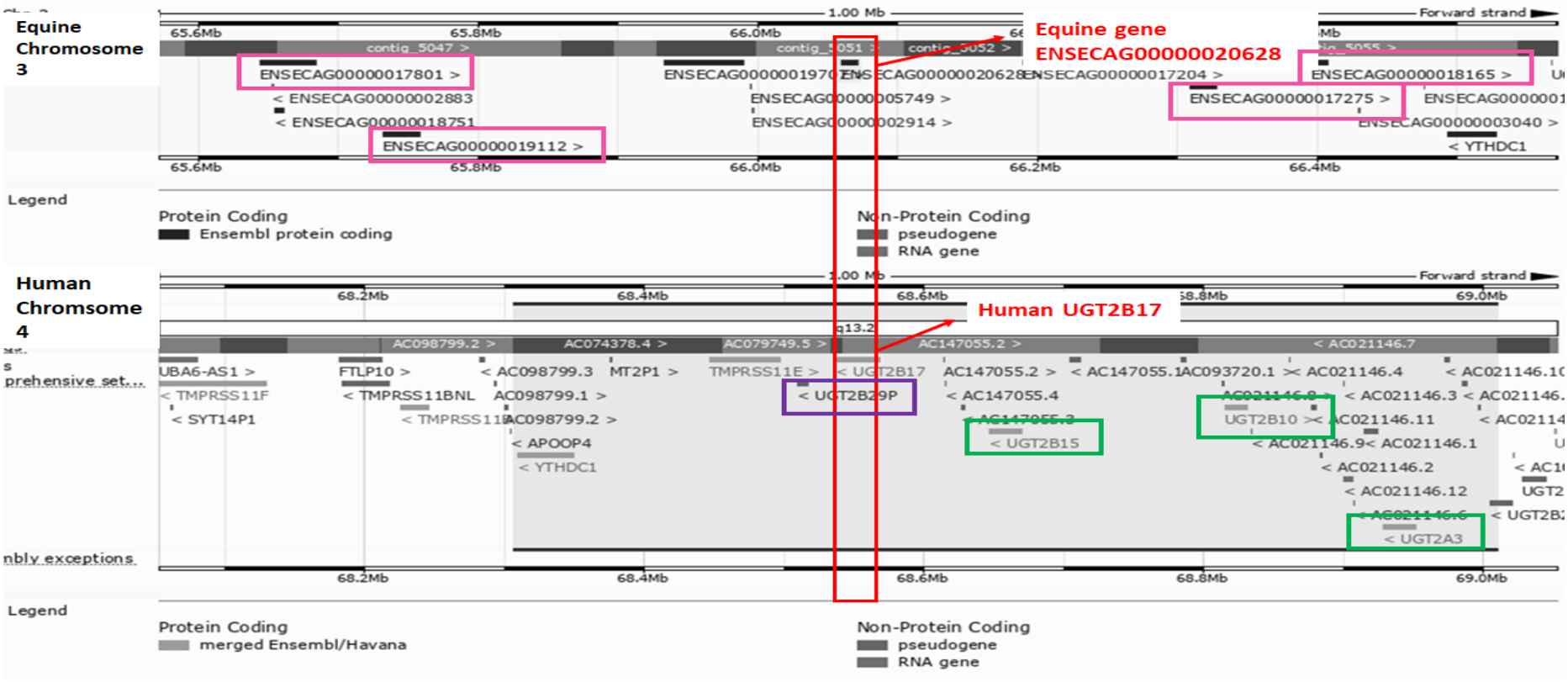
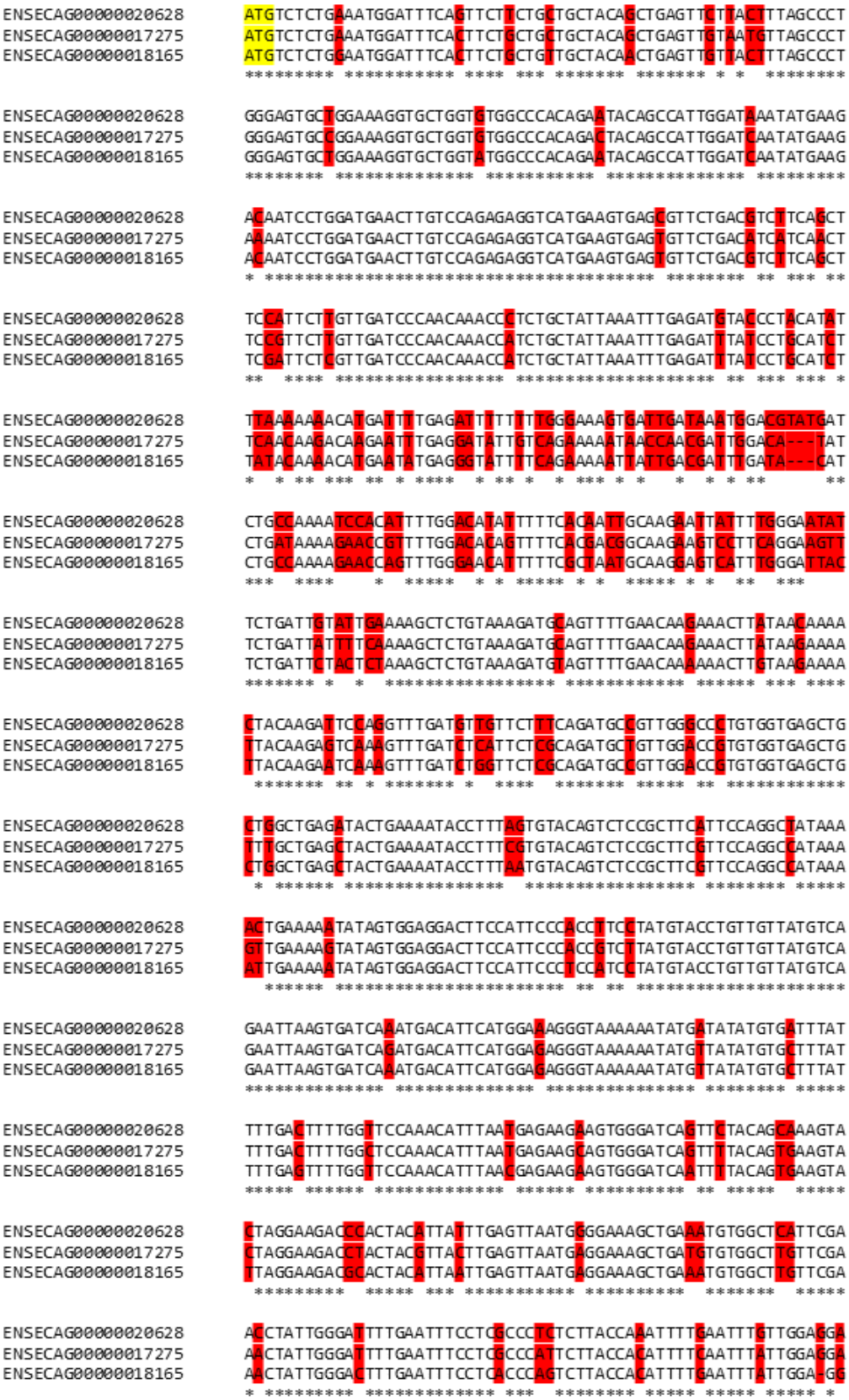


Figure 3.16: Region comparison of the predicted equine *UGT2B31-like* and human chromosome 4. The equine sequence (*ENSECAG00000020628*) correlates to human *UGT2B17*. Several human *UGTs* are found to be encoded nearby, *UGT2B15*, *UGT2B10* and *UGT2A3* (green boxes) and a pseudogene, *UGT2B29P* (purple box). The genes on the equine chromosome were compared against the list provided by Dr Emes to find that four additional UGTs are predicted on the equine chromosome (pink boxes).

Figure 3.17: Alignment of equine *ENSECAG00000020628* and human *2B17*. The alignment of the amino acids of *ENSECAG00000020628* and the human orthologue of *UGT2B17* displays the conserved residues (highlighted pink), 80% of the amino acids are conserved between the two sequences.

3.3.7.2 Sequence similarities of the three predicted UGT2B31-like enzymes

The three predicted equine *UGT2B31-like* sequences were aligned to determine the homology between their nucleotide sequences (Figure 3.18). The most variation between the three sequences is in the first half of the sequence, particularly between position 240 and 360bp. Overall there is >89% homology between all three sequences.


ENSECAG00000020628
ENSECAG00000017275
ENSECAG00000018165

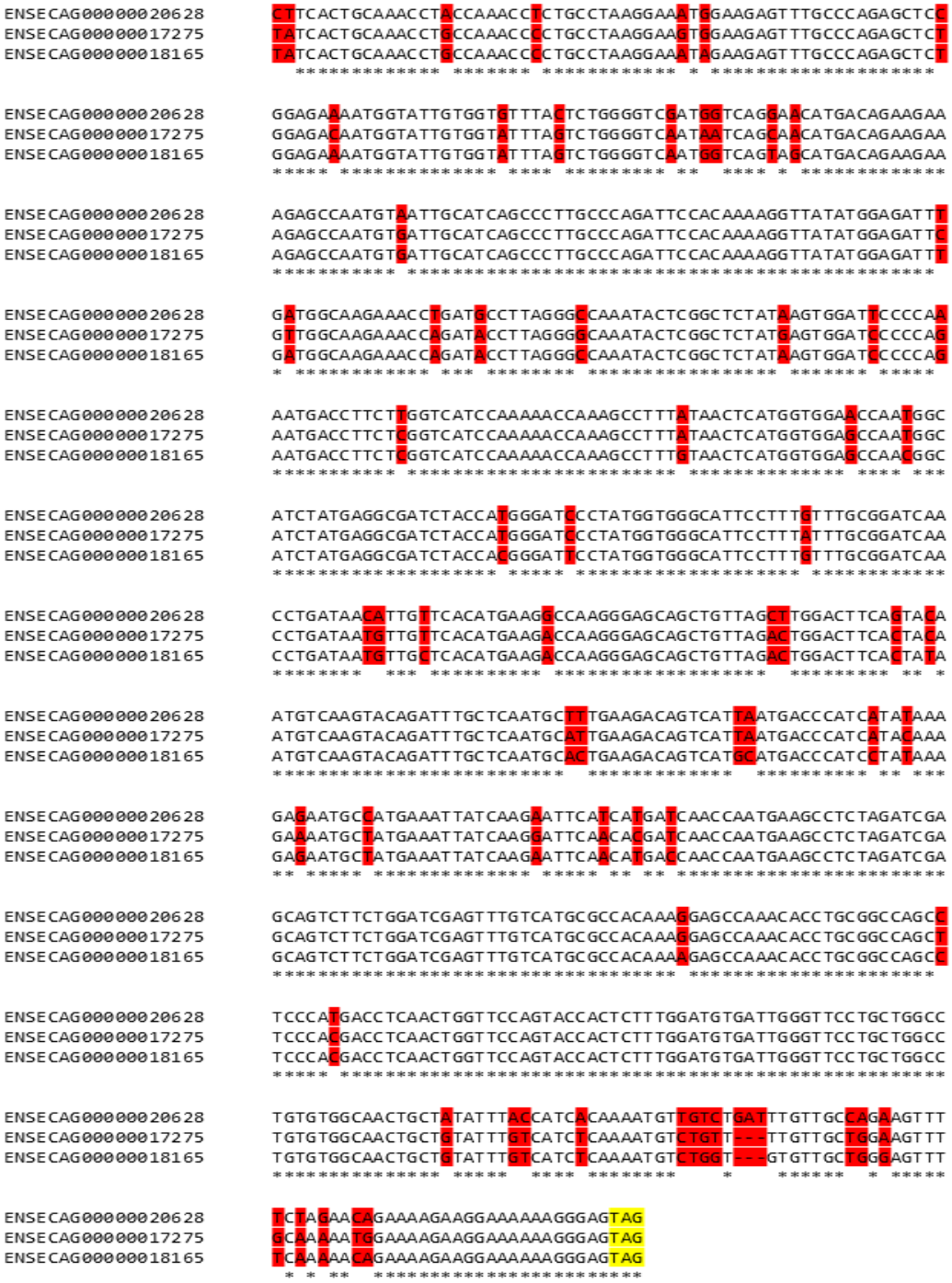


Figure 3.18: Alignment of the three-predicted equine UGT2B31-like genes. The three genes, *ENSECAG00000020628*, *ENSECAG00000017275* and *ENSECAG00000018165* were aligned at the nucleotide level to assess the homology between the three predicted equine *UGTS*. Most of the difference exists within the first half of the sequences (highlighted red), with a large number of difference between positions 240-360. A large proportion of the size difference for ENSECAG00000020628 is in the C-termini, which is 14 nucleotides longer than the two other sequences. *ENSECAG00000020628* shares 89% and 90.2% homology with *ENSECAG00000017275* and *ENSECAG00000018165* respectively. Start and stop codons are highlighted in yellow.

3.3.7.3 Annotating the sequence ENSECAG00000020628 for features characteristic to the UGTs

The isolated CDS sequence was translated and annotated for features characteristic to the UGTs. Figure 3.19 displays the annotated *UGT2B31-like* polypeptide. The first 22 residues, highlighted green, encode the signal peptide at the beginning of the N-terminus. In the C-terminus is a signature sequence spanning 44 amino acids (highlighted bright blue). This is followed by the transmembrane domain (pale blue) and the dilysine repeat (highlighted yellow, lysine residues in bold). There is also a glycosylation signal of three amino acids ('NMT' highlighted pink) in the sequence.

Investigation of the gene structure (genome construct EquCab 2.0) on the NBCI database implied that this gene has three variants (Figure 3.20). Variant 1 correlates with the approximate size of the majority of *UGT*s, at 531 amino acids. Variant 2 is a shorter polypeptide of 447 amino acids due to a partial deletion of exon 1. Variant 3 is a little larger than variant 1; it contains an insertion at exon 2 and complete deletion of exon 3. Sequencing alone could not provide conclusive results as to which variant had been isolated. The annotated translation (Figure 3.19) is 530 amino acids in length, suggesting it could be variant 1.

A PCR was designed to determine conclusively whether variant 1 had been isolated, but to also determine whether variants 2 and 3 are co-expressed in the liver. To achieve this, exon specific primers were designed. Forward primers were designed to cross the exon 1 and 2 boundary. As this boundary was different for each variant, owing to deletions and insertions, each primer was designed to a unique exon boundary. The reverse primer was designed to exon 4. The difference in size of the PCR product should enable us to differentiate between expression of the three variants. The semi-quantitative PCR confirmed that all three variants were expressed (data not shown).

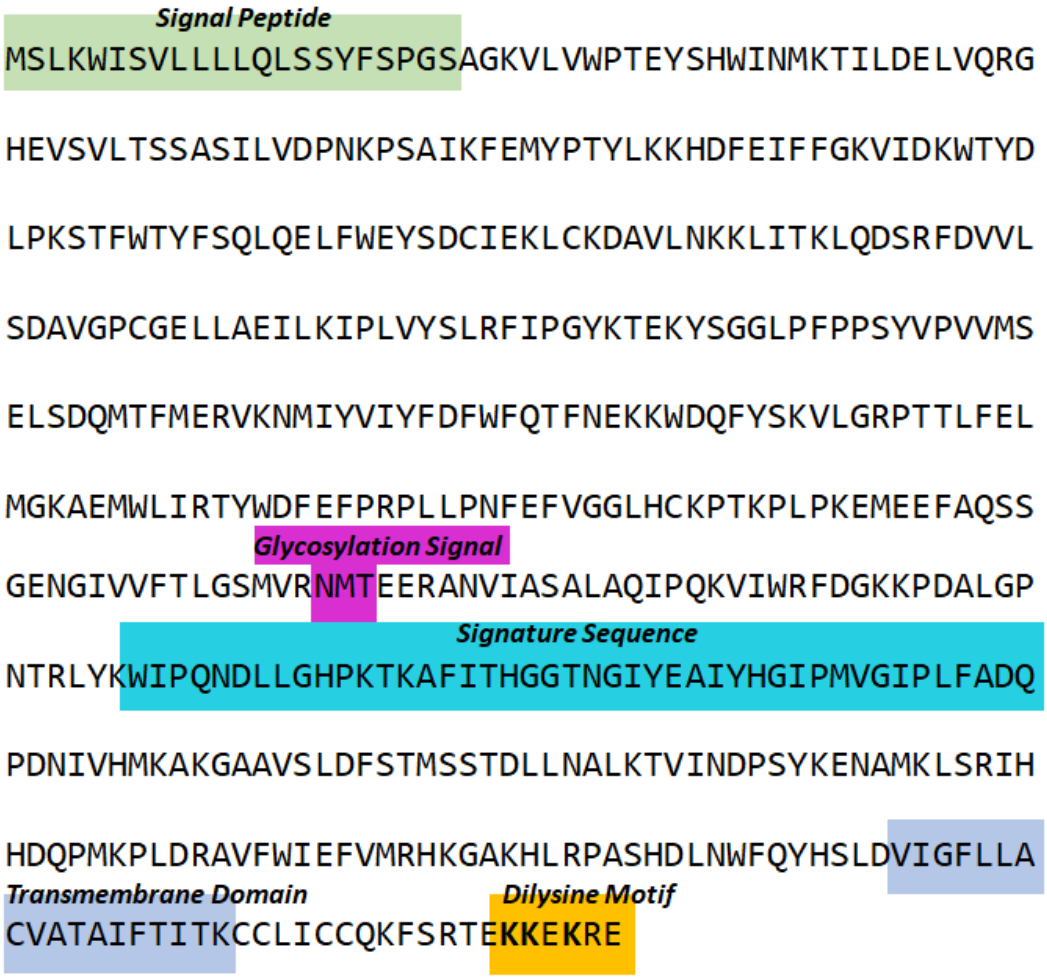


Figure 3.19: The translated sequence of the *UGT2B31-like* gene, annotated for features characteristic to the *UGTs*. The first 22 amino acids, highlighted green, highlight the signal peptide. The signature sequence, bright blue, is a 44 amino acid motif which binds the donor sugar. The transmembrane domain, pale blue, anchors the protein to the ER membrane, the yellow box indicated the dilysine motif, with lysine residues in bold typeface.

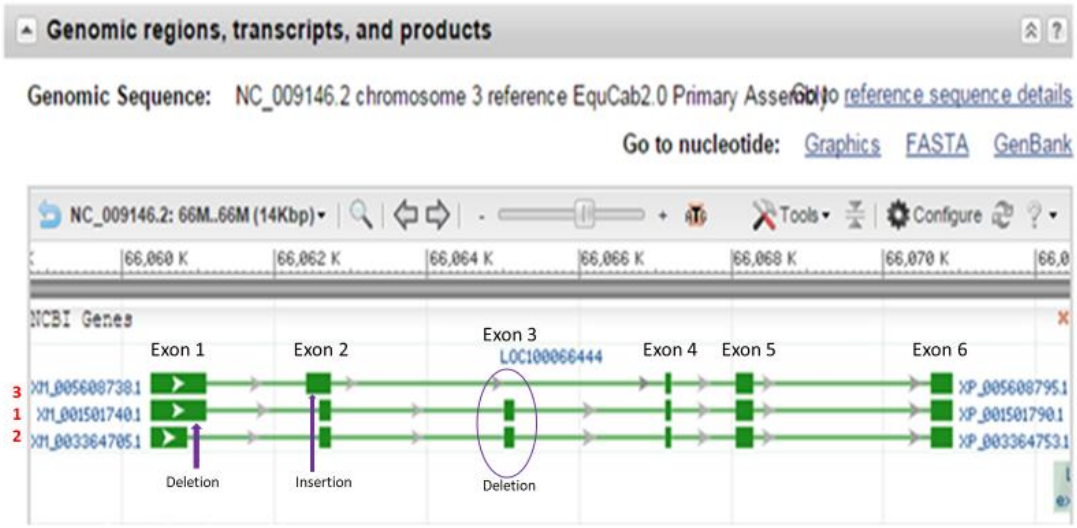


Figure 3.20: Three structural variants of *ENSECAG00000020628*. The NCBI database suggest that there are three variants of *UGT2B31-like*. Variant 1 is 531 amino acids, which is a length complementary to other investigated UGTs. Variant 2 is shorter at 447 amino acids owing to a partial deletion of exon 1. Variant 3 is the largest at 545 residues, due to an insertion in exon 2 and a complete deletion of exon 3.

3.3.8 Evidence for ENSECAG00000014362 as a novel equine UGT

The Ensembl database annotates *ENSECAG00000014362* as *UGT2A3*. However a BLAST analysis of this sequence in the NCBI database identifies this gene as *UGT2A1*, transcript variant 2.

3.3.8.1 Syntenic comparison

The Ensembl database shows the *ENSECAG00000014362* gene to be encoded on equine chromosome 3, location 65,351,083-65,371,301. A syntenic comparison to the human genome correlates this to chromosome 4 in humans, location, 68,928,463-68,951,791.

There is a lack of gene annotation for this region of the equine genome with three named genes present; Casein Beta (*CSN2*), Casein Alpha S1 (*CSN1S1*), and Sulfotransferase Family 1E Member 1 (*SULT1E1*). The equivalent region on the human chromosome encodes multiple UGT genes, *UGT2B10*, *UGT2B15* and *UGT2B17* all encoded downstream of *UGT2A3*, while *UGT2B7*, *UGT2B11* and *UGT2B28* are encoded upstream. The equine gene correlates to the human *UGT2A3* gene in the Ensembl database, but a BLAST analysis in the NCBI database (Figure 3.21), identified it as *UGT2A1*

Figure 3.21 displays the numerous *UGTs* encoded in this region of the human genome. Further analysis of the equine region surrounding *ENSECAG00000014362* was made and revealed two additional genes predicted to be equine UGTs, *ENSECAG00000008247* and *ENSECAG00000017801* (pink boxes - these have not been isolated within this study). Upstream of *ENSECAG00000017801* (*See section 3.3.7.1, figure 3.16*) are several more predicted *UGTs* (reported in section 3.3.7.1) this corresponds to the multiple genes seen in the human *UGT2* locus.

Both *ENSECAG00000008247* and *ENSECAG00000017801* were analysed via BLAST analyses in the NCBI database, and both shown to be annotated as *UGT2C-like* genes. The three genes, *ENSECAG00000014362*, *ENSECAG00000008247* and

ENSECAG00000017801 were aligned to assess homology. *ENSECAG00000014362* shares less than 70% homology with two aligned sequences, whereas *ENSECAG00000008247* and *ENSECAG00000017801* share 90% of their nucleotide sequence.

To determine which human UGT2 had the greatest level of homology with equine *ENSECAG00000014362* comparisons were made between the amino acid sequences of several *UGT2*s from humans, mice, and rats. They shared the greatest level of homology with human *UGT2A1* (74.57%), 74% homologous to murine *UGT2A2* and *74%* homologous to rat *UGT2A1*, 73.72%. As the equine sequence correlates to a human orthologue in an area that encodes numerous human *UGT2B* enzymes, a comparison of homology was also made with the *UGT2B* enzymes located in the region. The human *UGT2Bs* all showed above 70% sequence homology to *ENSECAG00000014362*,with the greatest similarity to human *UGT2B17* (80%). Murine *UGTB2s* showed 64-75% sequence similarity, with the highest level of homology to murine *UGT2B1*; rat *UGTs* displayed similar levels of homology as the mouse, ranging from 64-76%.

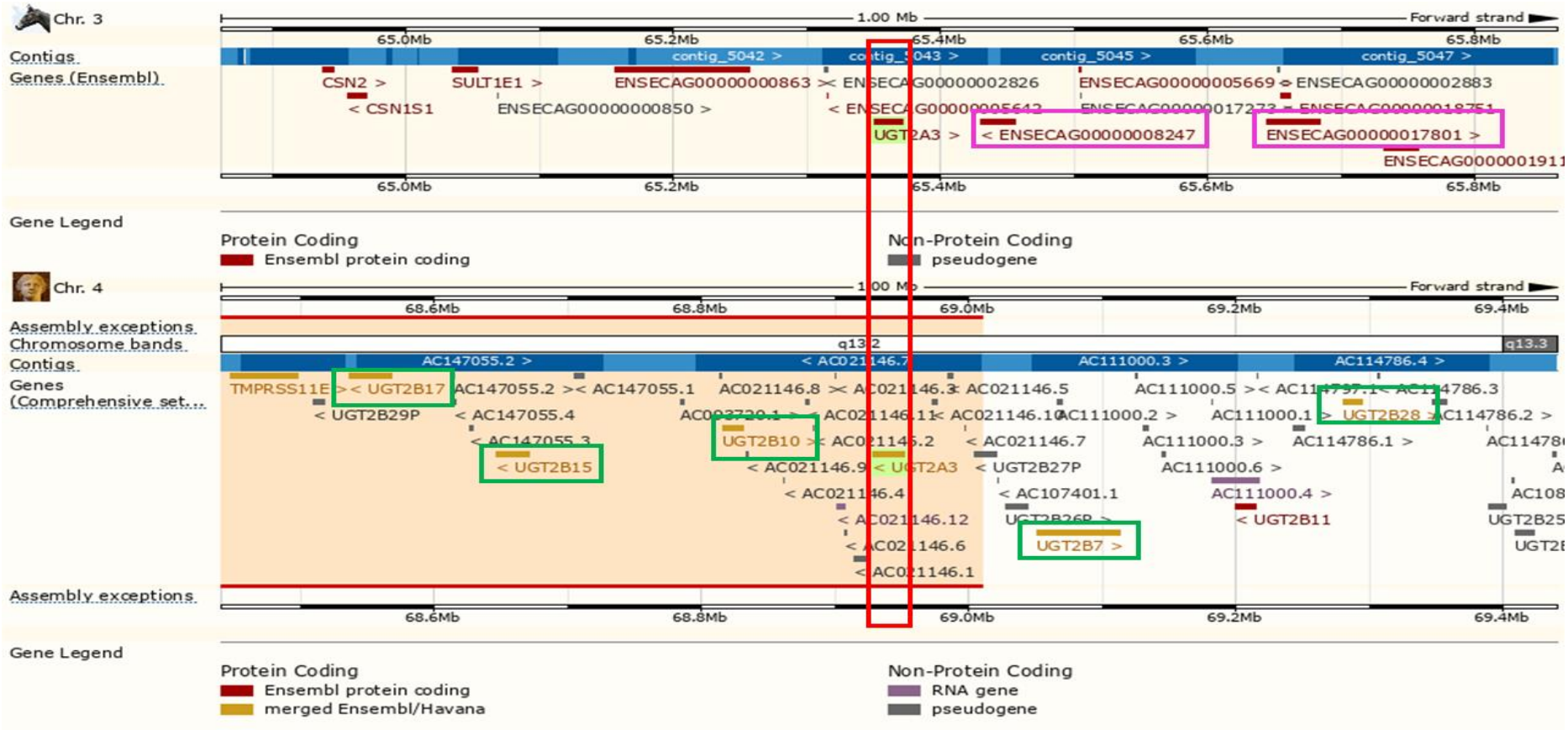


Figure 3.21: Region comparison of equine *ENSECAG00000014326* and human chromosome 4. The equine sequence (*ENSECAG00000014362*) Correlates to human *UGT2A3*. Several human *UGTs* are found to be encoded nearby, *UGT2B15*, *UGT2B10* and *UGT2B17* (green boxes). The genes on the equine chromosome were compared against the list provided by Dr Emes to find that two additional *UGTs* are predicted on the equine chromosome (pink boxes).

3.3.8.2 Gene structure of ENSECAG00000014362

A product of 1665bp was isolated and sequenced, and the translated amino acid sequence (555 amino acids) annotated for features that are characteristic to the UGT superfamily of enzymes.

Figure 3.22 displays the annotated polypeptide, with features highlighted. The signal peptide is encoded in the first 22 amino acids (green box), while the signature sequence, a motif involved in binding the donor sugar, is shown in a blue box. Towards the end of the sequence are two of the key sequences, the transmembrane sequence (highlighted yellow) and the dilysine motif (pink). The sequence has two asterisks (highlighted red) which indicated the location of the stop codons, with two stop codons identified in this polypeptide.

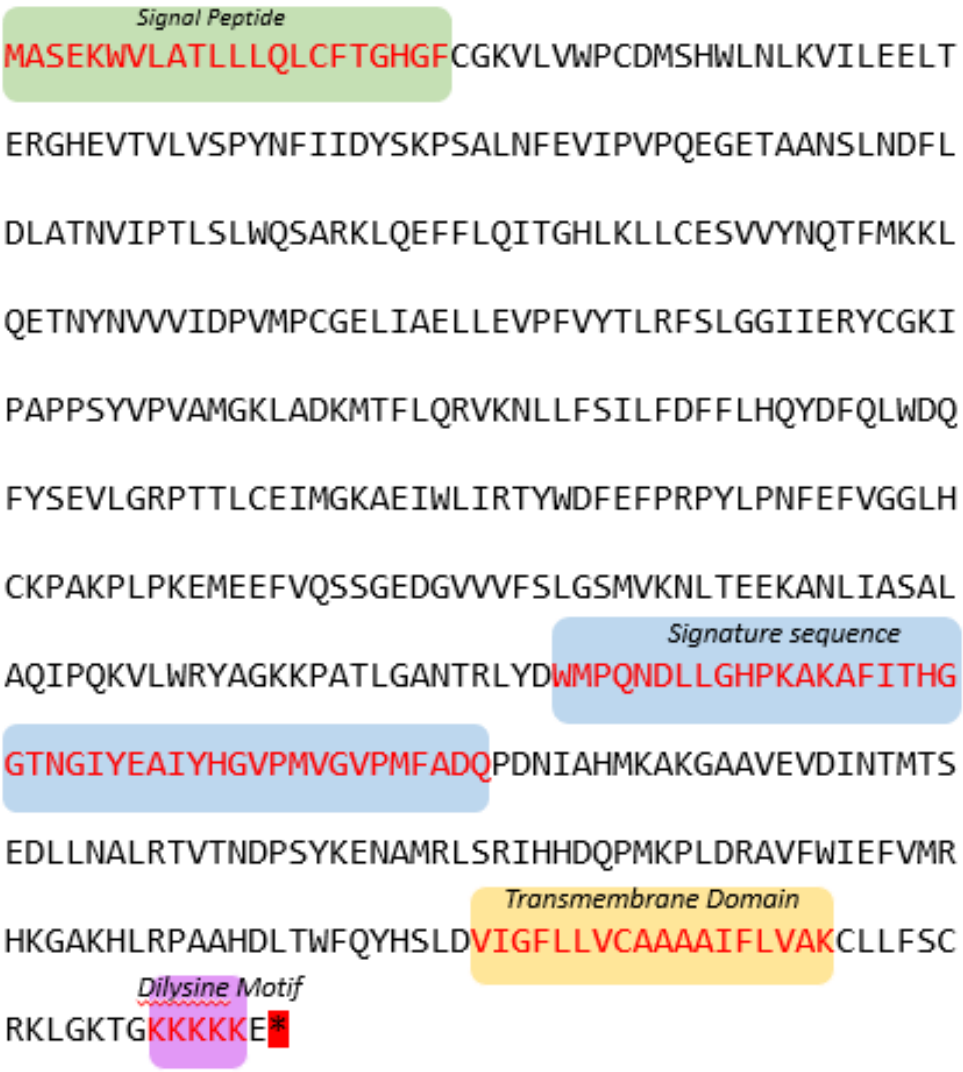


Figure 3.22: Annotated translation of the isolated *ENSECAG00000014362*. The amino acid sequence was annotated for characteristic features of the *UGTs*. This encompasses the signal peptide (pale green), signature sequence (pale blue), and transmembrane domain (yellow) and dilysine motif (pink). The asterisk (highlighted red) denote stop codon.

3.3.9 Identifying equine UGT1A6

3.3.9.1 Phylogenetic analysis of equine UGT1A6

The ML tree displays the phylogenetic analysis of 72 *UGT* sequences from the *UGT1* sub-family, representing 23 species, with barley (ADC92549.1) and orange (ACS87992.1) as outgroups (Figure 3.23, see appendix D for list of accession numbers). The predicted equine *UGT1A6* sequence was retrieved from Ensembl, *ENSECAT00000025758.1.*

The tree displays four distinct clusters and the separate outgroup containing barley and orange. The four clusters are broadly separated into *UGT1A6s* from multiple species, the Zebrafish *UGT* sequences, *UGT1A7-12* sequences and *UGT1A1-5* sequences.

The Zebrafish represent the singular piscine member included in this analysis, with the tree displaying ten Zebrafish *UGT* sequences (Figure 3.23, denoted by blue lines). The Zebrafish *UGTs* cluster in a separate clade, with the closest related species included in the analysis being the western clawed frog *UGT1A6*. Within the Zebrafish clade the sequences divide into two sub-groups, *UGT1A* and *UGT1B*.

In the three remaining clades, the rodent sequences always cluster together while the primates cluster separately. Ungulate *UGTs* are under-represented in the databases; however those that could be included in the analysis sit separate to both the rodent and primate sequences.

Within the cluster containing the *UGT1A1-5* sequences, the *UGT1A1* sequences cluster together on a separate branch. Within this clade is an anomaly, the Wild Boar *UGT1A6* (Figure 3.23, highlighted yellow) sits within this clade and appears most closely related to the cow and sheep *UGT1A1* sequences and not *UGT1A6*. Cat

UGT1A1 deviates from this group of isozymes appearing more closely related to the *UGT1A2-5* sequences. The second branch in this clade shows a clear division between the rodents, rat, and mouse, and the human and other primate sequences. This division of clustering by animal is repeated in a separate clade containing the *UGT1A7-12* sequences.

There is clear clustering of the *UGT1A6* isozymes (Figure 3.23, highlighted light purple). Within this cluster there is a clear division of the rodent sequences, the primate sequences and other mammals including the equine sequence (dark purple) grouping with the *UGT1A6* in the grey wolf, cow, and sheep. The robustness of the ML tree, confirmed through bootstrap analysis (100 times), supported the confidence of the relationship.

3.3.9.2 Comparison of the sequence identities of UGT1A6

A comparison of the *UGT1A6* sequences between human, cow, mouse, rat, sheep and grey wolf (Figure 3.24) showed that the horse and human sequences share the highest level of similarity at the amino acid level (82.9%). The horse sequence is also similar to that of the grey wolf (82.2%), and least similar to the *UGT1A6* from the rat and mouse (79.2% and 78.2% respectively).

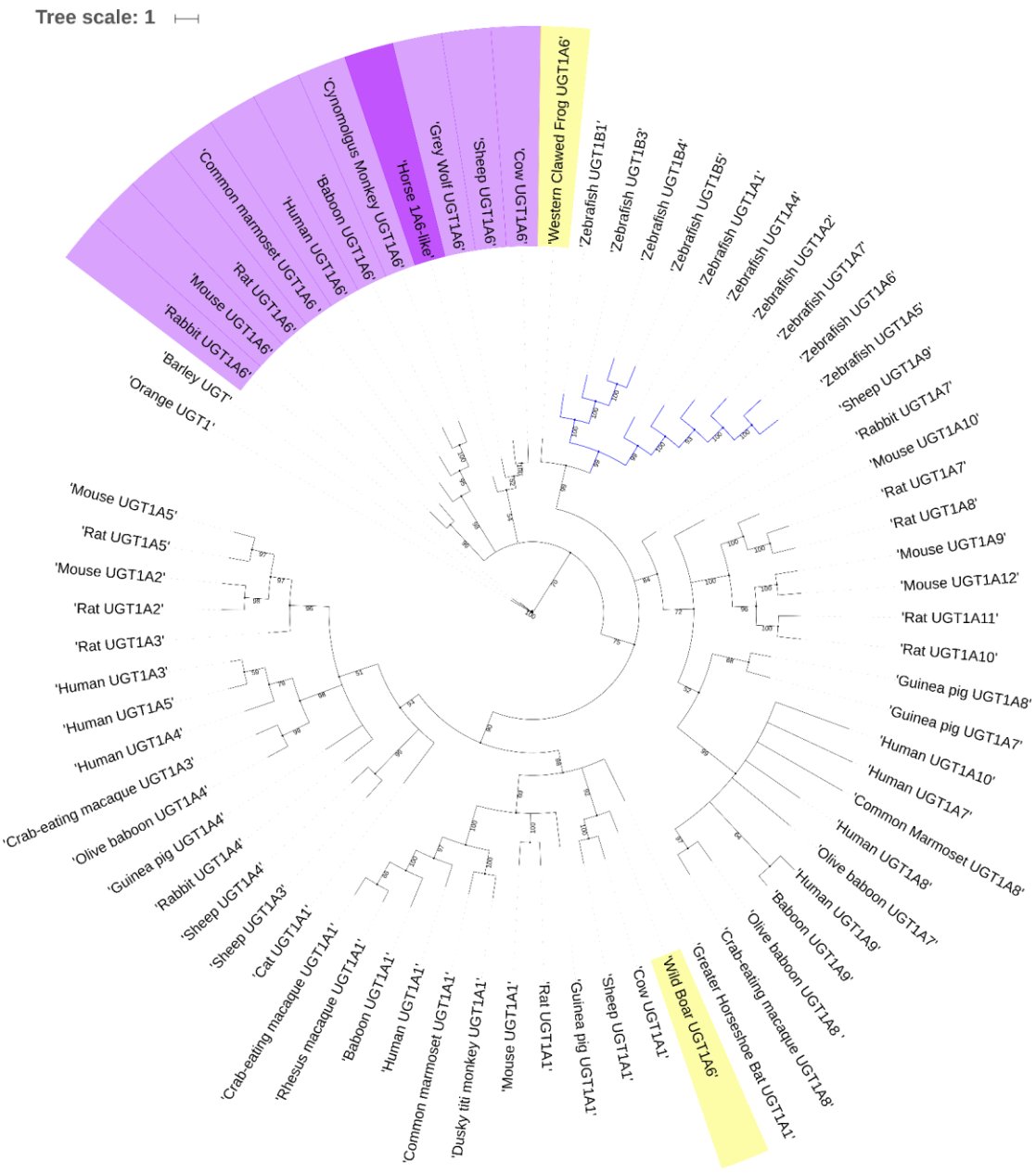


Figure 3.23. A Maximum likelihood (ML) consensus tree of *UGT1* sequences. 72 *UGT1* sequences from 21 mammals and 2 plants were retrieved from the NCBI database, these included both confirmed and predicted UGT sequences (see appendix B for full list of accession numbers). Sequences analysed using ML and bootstrapped 100 times, to infer relationships using Geneious (Biomatters Ltd). Sequence names highlighted in light purple represent the clustering of the *UGT1A6* sequences, the dark purple highlights the predicted equine sequence. There are two *UGT1A6* sequences (highlighted pale yellow) which fail to cluster with the majority of the *UGT1A6* isozymes; the Wild Boar *UGT1A6* which is more closely associated with the *UGT1A1s* and the Western Clawed Frog *UGT1A6* which sits adjacent to the Zebrafish *UGT* sequences. The equine sequence (dark purple) clusters with the *UGT1A6*s, close to grey wolf, sheep, and cow sequences.

	Horse 1A6-like	Cow 1A6	Sheep 1A6	Grey Wolf 1A6	Human 1A6	Mouse 1A6	Rat 1A6
Horse 1A6-like	100.0	79.8	80.2	82.2	82.9	78.2	79.2
Cow 1A6	79.8	100.0	94.9	80.5	79.4	76.7	77.3
Sheep 1A6	80.2	94.9	100.0	80.2	79.4	76.0	76.4
Grey Wolf 1A6	82.2	80.5	80.2	100.0	80.8	78.4	79.1
Human 1A6	82.9	79.4	79.4	80.8	100.0	79.5	79.9
Mouse 1A6	78.2	76.7	76.0	78.4	79.5	100.0	93.4
Rat 1A6	79.2	77.3	76.4	79.1	79.9	93.4	100.0

Figure 3.24: Percentage sequence identities between *UGT1A6*. The percentage of amino acid residues conserved between the horse *UGT1A6* sequence and the equivalent from the rat (AAL67853.1), mouse (NP_659545.2), human (NP_001063.2), cow (NP_777187.1), sheep (NP_001192075.1) and grey wolf (NP_001003078.1) *UGT1A6* sequences. On a colour scale of red to pink indicating the fewest conserved amino acids between sequences and red indicates the highest number of amino acids conserved between sequences. The greatest levels of similarity between the sequences included in the analysis are between the cow and sheep *UGT1A6* sequences (94.9%). The equine predicted *UGT1A6-like* sequence displays the highest percentage of conserved amino acid residues with the human *UGT1A6* sequence, 82.9%.

3.3.9.3 Syntenic comparison – a cross species comparison to infer enzyme identity

The predicted equine orthologue of human *UGT1A6* is encoded in the horse genome on chromosome 6, location 20,439,984 – 20,543,149. The *UGT1A6* locus and the sequence flanking it on the chromosome are conserved in humans on chromosome 2 (Figure 3.25).

In humans the *UGT1* locus is complex, with the variable slicing and sharing of exons creating the multiple isozymes. As equine *UGT1A6* is the first member to be isolated and due to a lack of annotation it is difficult to infer whether the equine locus is also complex. However an analysis of the region showed a level of conservation between humans and equines (Figure 3.26). Several genes locations flanking the *UGT1* locus in the human genome appear to have been conserved, including Diacylgycerol Kinase Delta (*DGKD*), Ubiquitin Specific Peptidase 40 (*USP40*), which are both located upstream of the *UGT1A6* gene and Secreted Phosphoprotein 2 (*SPP2*) and Transient Receptor Potential Cation Channel Subfamily M Member 8 (*TRPM8*) which are located downstream.

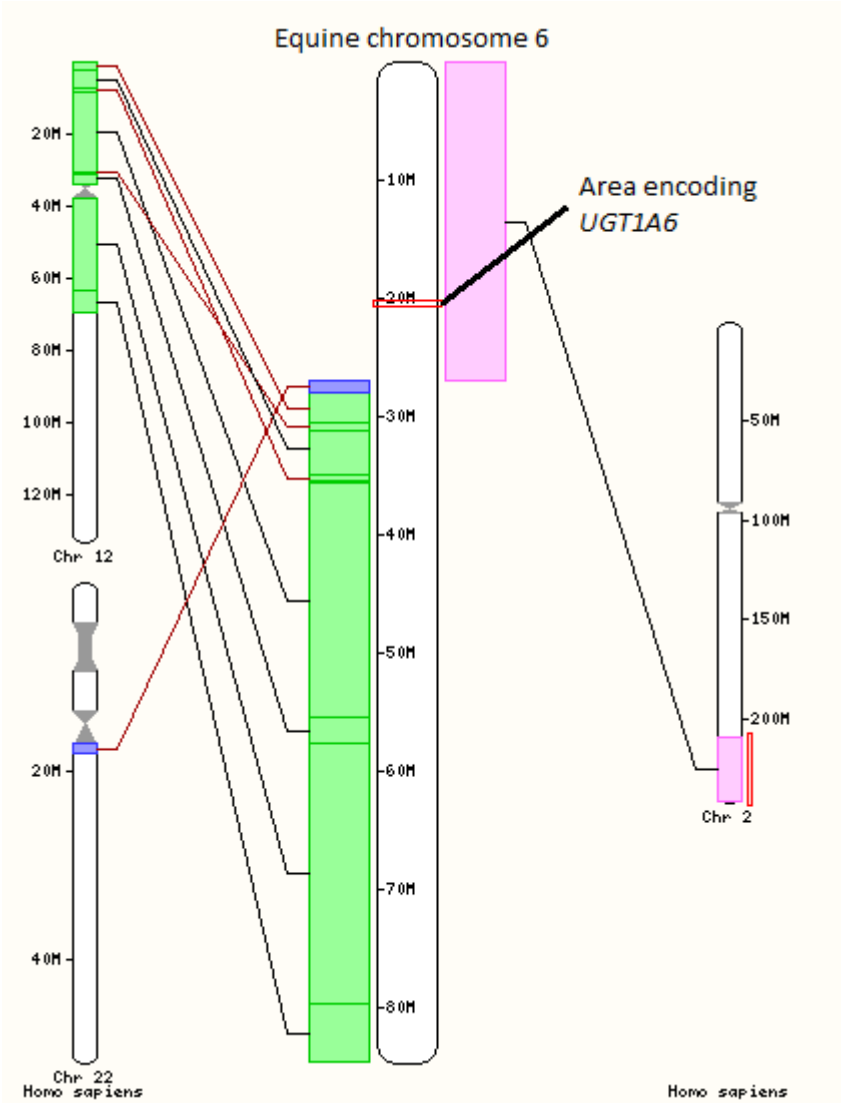


Figure 3.25: Syntenic comparison of equine chromosome 6 with the human genome. The diagram (taken from Ensembl -release 89) shows blocks of equine chromosome 6 conserved in human chromosomes 2, 12 and 22. The area of interest (demarcated by a red box) on the equine chromosome encodes *UGT1A6*, this and the surrounding region is present as a conserved block on human chromosome 2.

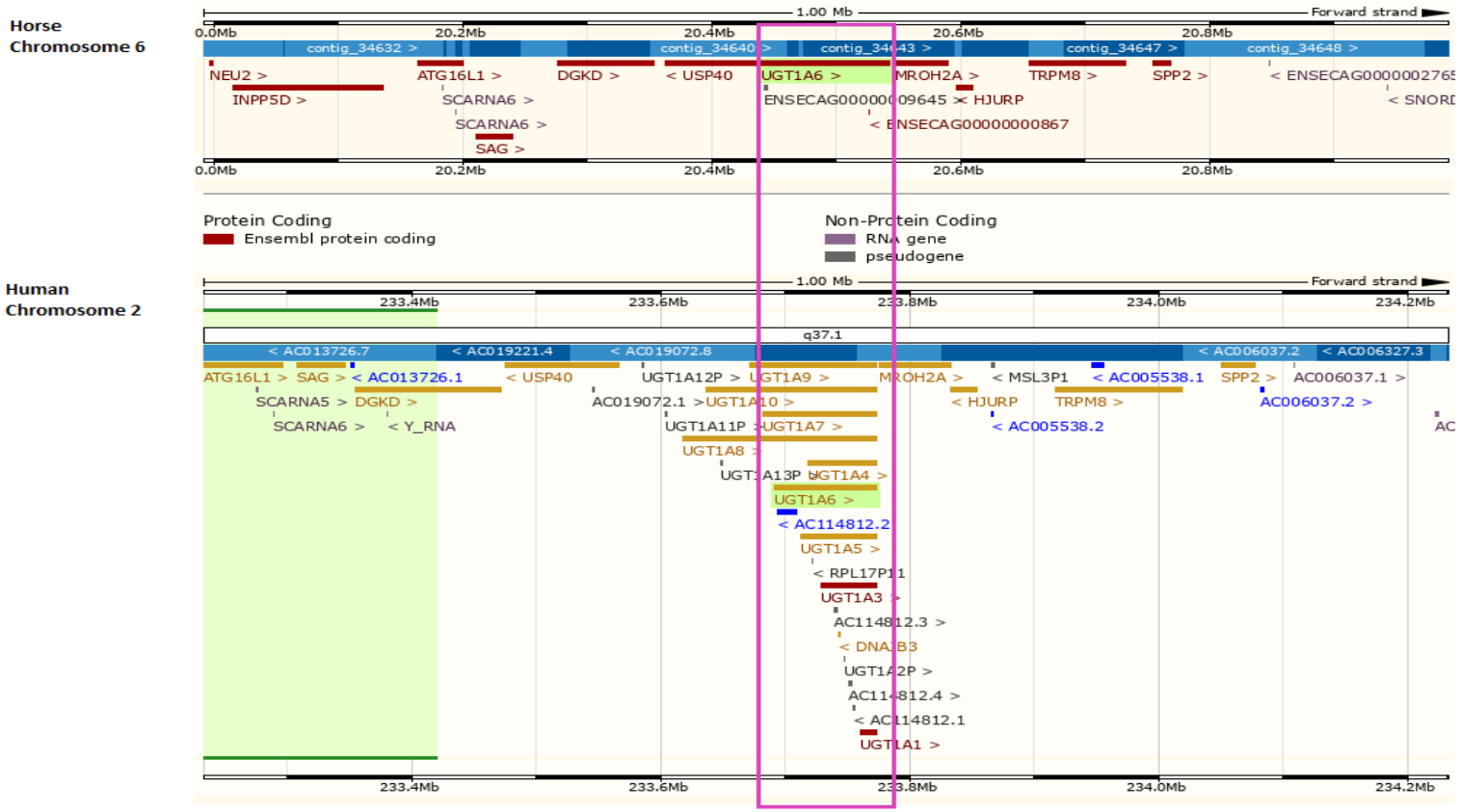


Figure 3.26: Genomic comparison between *UGT1A6* sequences of equine and human. The region of the equine *UGT1A6* sequence correlates to the *UGT1A6* sequence region in humans. Multiple genes in the region appear to have been conserved, including; Discylglycerol Kinase Delta (*DGKD*), Ubiquitin Specific Peptidase 40 (*USP40*) upstream Secreted Phosphoprotein 2 (*SPP2*) and Transient Receptor Potential Cation Channel Subfamily M member 8 (*TRPM8*) both located downstream of the *UGT1A6* gene.

3.3.9.4 Gene structure

A sequence of 1629bp for equine *UGT1A6*, encoding a protein of 533 amino acids, was successfully isolated and annotated for features characteristic of *UGT* sequences (Figure 3.27). *UGTs* are comprised of two domains: the N-terminus, implicated in substrate specificity, and the C-terminus which anchors the protein to the ER and binds the donor sugar (Meech and Mackenzie, 1997b).

A comparison of equine *UGT1A6* to orthologues from the rat, mouse, sheep, cow, dog, and human showed that 62.85% of amino acids are conserved between all sequences, with the greatest conservation in the C-termini and the N-termini showing the highest level of variability (Figure 3.27).

At the N-terminus, the first 22 amino acids of the equine *UGT1A6* encode the signal peptide, which directs the mature protein to the ER. A comparison of the equine and human signal peptides shows 17 residues (77%) are conserved (Figure 3.27 – highlighted light grey). Further comparison of signal peptide sequences from rat, mouse, sheep, and cow *UGT1A6* showed the rat and mouse have 86% similarity. The rat shares 63% and mouse 59% of the 22 amino acids with humans. The dog sequence shows the greatest difference, with the shortest signature sequence of 18 amino acids, with only 50% conserved with humans.

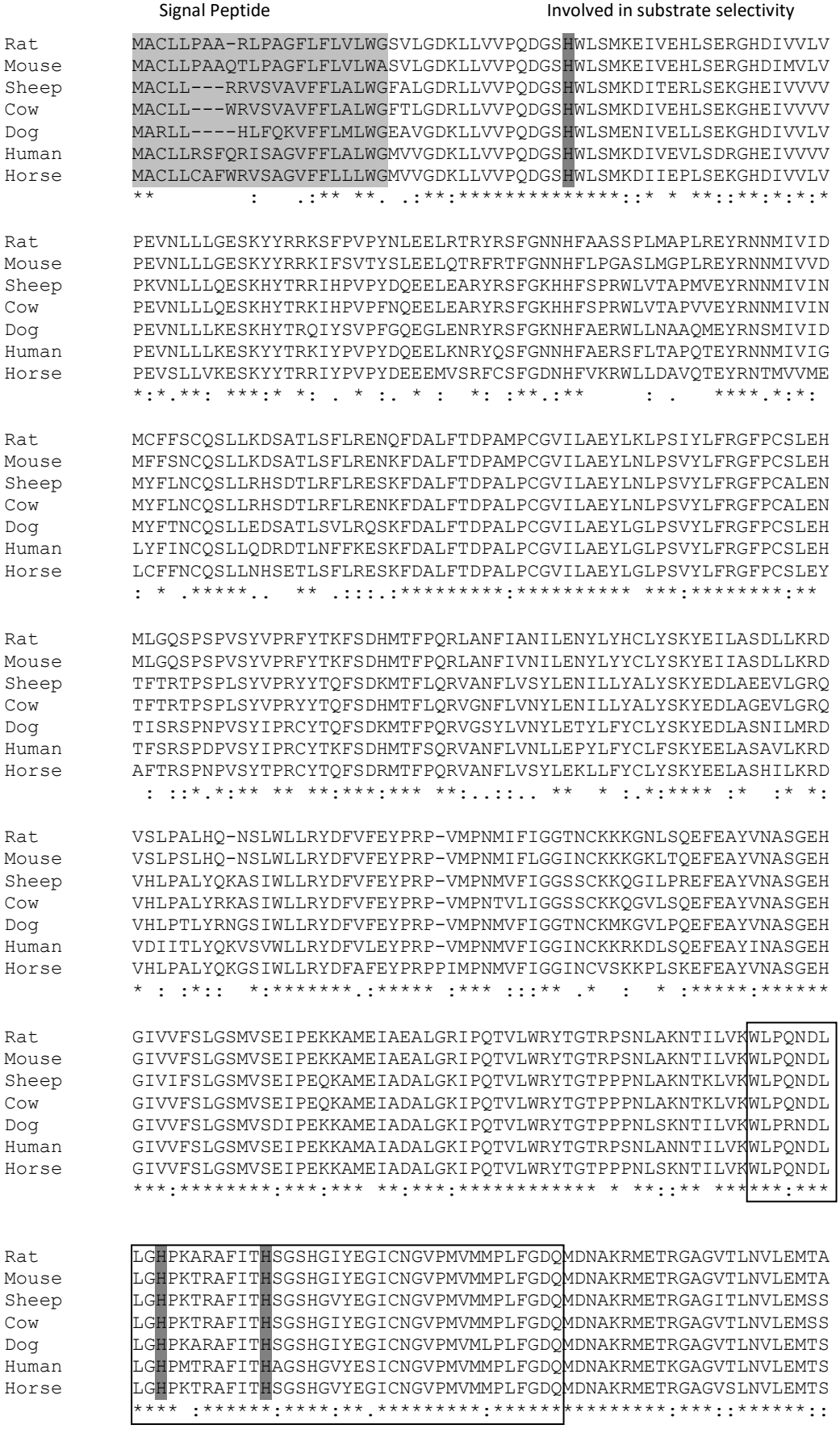
A histidine at position 38 (Figure, 3.27, dark grey), implicated in substrate selectivity in humans, is present in the equine *UGT1A6* sequence and is also conserved in the mouse, rat, cow, sheep, and dog *UGT1A6.*

Within the C-terminus there is a signature sequence, transmembrane domain and a dilysine repeat, all of which show a high level of conservation between the sequences analysed (Figure 3.27). Within the signature sequences analysed (boxed area – Figure 3.27), 84% of the all amino acids are conserved. A comparison between

the human and equine *UGT1A6* signature sequences shows this level of conservation to be 93%.

The transmembrane domain is a region comprised of 16 amino acids, of which 70% of the residues are conserved between the seven sequences aligned (Figure 3.27). Downstream of the transmembrane domain, in the final five residues of the sequence, is a dilysine repeat (Figure 3.27, highlighted in bold typeface with a black background), which is present in all seven *UGT1A6* sequences.

In addition to the histidine at position 38, human studies have identified three additional histidine residues to be of importance in the role of glycosylation and donor sugar specificity (Figure 3.27, highlighted dark grey). These three histidine amino acids are all present within the C-termini of the sequences. Two of the histidines are situated closely together at positions 361 and 370, with the third immediately upstream of the transmembrane domain at position 485. All three histidine residues are conserved in equine *UGT1A6* and in all the *UGT1A6* sequences included in the alignment.


Signal Peptide
Involved in substrate selectivity
Rat MACLLPAA-RLPAGFLFLVLWGSVLGDKLLVVPQDGSHWLSMKEIVEHLSERGHDIVVLV
Mouse MACLLPAAQTLPAGFLFLVLWASVLGDKLLVVPQDGSHWLSMKEIVEHLSERGHDIMVLV
Sheep MACLL---RRVSVAVFFLALWGFALGDRLLVVPQDGSHWLSMKDITERLSEKGHEIVVVV
Cow MACLL---WRVSVAVFFLALWGFTLGDRLLVVPQDGSHWLSMKDIVEHLSEKGHEIVVVV
Dog MARLL----HLFQKVFFLMLWGEAVGDKLLVVPQDGSHWLSMENIVELLSEKGHDIVVLV
Human MACLLRSFQRISAGVFFLALWGMVVGDKLLVVPQDGSHWLSMKDIVEVLSDRGHEIVVVV
Horse MACLLCAFWRVSAGVFFLLLWGMVVGDKLLVVPQDGSHWLSMKDIIEPLSEKGHDIVVLV
Rat PEVNLLLGESKYYRRKSFPVPYNLEELRTRYRSFGNNHFAASSPLMAPLREYRNNMIVID
Mouse PEVNLLLGESKYYRRKIFSVTYSLEELQTRFRTFGNNHFLPGASLMGPLREYRNNMIVVD
Sheep PKVNLLLQESKHYTRRIHPVPYDQEELEARYRSFGKHHFSPRWLVTAPMVEYRNNMIVIN
Cow PEVNLLLQESKHYTRKIHPVPFNQEELEARYRSFGKHHFSPRWLVTAPVVEYRNNMIVIN
Dog PEVNLLLKESKHYTRQIYSVPFGQEGLENRYRSFGKNHFAERWLLNAAQMEYRNSMIVID
Human PEVNLLLKESKYYTRKIYPVPYDQEELKNRYQSFGNNHFAERSFLTAPQTEYRNNMIVIG
Horse PEVSLLVKESKYYTRRIYPVPYDEEEMVSRFCSFGDNHFVKRWLLDAVQTEYRNTMVVME
Rat MCFFSCQSLLKDSATLSFLRENQFDALFTDPAMPCGVILAEYLKLPSIYLFRGFPCSLEH
Mouse MFFSNCQSLLKDSATLSFLRENKFDALFTDPAMPCGVILAEYLNLPSVYLFRGFPCSLEH
Sheep MYFLNCQSLLRHSDTLRFLRESKFDALFTDPALPCGVILAEYLNLPSVYLFRGFPCALEN
Cow MYFLNCQSLLRHSDTLRFLRENKFDALFTDPALPCGVILAEYLNLPSVYLFRGFPCALEN
Dog MYFTNCQSLLEDSATLSVLRQSKFDALFTDPALPCGVILAEYLGLPSVYLFRGFPCSLEH
Human LYFINCQSLLQDRDTLNFFKESKFDALFTDPALPCGVILAEYLGLPSVYLFRGFPCSLEH
Horse LCFFNCQSLLNHSETLSFLRESKFDALFTDPALPCGVILAEYLGLPSVYLFRGFPCSLEY
Rat MLGQSPSPVSYVPRFYTKFSDHMTFPQRLANFIANILENYLYHCLYSKYEILASDLLKRD
Mouse MLGQSPSPVSYVPRFYTKFSDHMTFPQRLANFIVNILENYLYYCLYSKYEIIASDLLKRD
Sheep TFTRTPSPLSYVPRYYTQFSDKMTFLQRVANFLVSYLENILLYALYSKYEDLAEEVLGRQ
Cow TFTRTPSPLSYVPRYYTQFSDHMTFLQRVGNFLVNYLENILLYALYSKYEDLAGEVLGRQ
Dog TISRSPNPVSYIPRCYTQFSDKMTFPQRVGSYLVNYLETYLFYCLYSKYEDLASNILMRD
Human TFSRSPDPVSYIPRCYTKFSDHMTFSQRVANFLVNLLEPYLFYCLFSKYEELASAVLKRD
Horse AFTRSPNPVSYTPRCYTQFSDRMTFPQRVANFLVSYLEKLLFYCLYSKYEELASHILKRD
Rat VSLPALHQ-NSLWLLRYDFVFEYPRP-VMPNMIFIGGTNCKKKGNLSQEFEAYVNASGEH
Mouse VSLPSLHQ-NSLWLLRYDFVFEYPRP-VMPNMIFLGGINCKKKGKLTQEFEAYVNASGEH
Sheep VHLPALYQKASIWLLRYDFVFEYPRP-VMPNMVFIGGSSCKKQGILPREFEAYVNASGEH
Cow VHLPALYRKASIWLLRYDFVFEYPRP-VMPNTVLIGGSSCKKQGVLSQEFEAYVNASGEH
Dog VHLPTLYRNGSIWLLRYDFVFEYPRP-VMPNMVFIGGTNCKMKGVLPQEFEAYVNASGEH
Human VDIITLYQKVSVWLLRYDFVLEYPRP-VMPNMVFIGGINCKKRKDLSQEFEAYINASGEH
Horse VHLPALYQKGSIWLLRYDFAFEYPRPPIMPNMVFIGGINCVSKKPLSKEFEAYVNASGEH
Rat GIVVFSLGSMVSEIPEKKAMEIAEALGRIPQTVLWRYTGTRPSNLAKNTILVKWLPQNDL
Mouse GIVVFSLGSMVSEIPEKKAMEIAEALGRIPQTVLWRYTGTRPSNLAKNTILVKWLPQNDL
Sheep GIVIFSLGSMVSEIPEQKAMEIADALGKIPQTVLWRYTGTPPPNLAKNTKLVKWLPQNDL
Cow GIVVFSLGSMVSEIPEQKAMEIADALGKIPQTVLWRYTGTPPPNLAKNTKLVKWLPQNDL
Dog GIVVFSLGSMVSDIPEKKAMEIADALGKIPQTVLWRYTGTPPPNLSKNTILVKWLPRNDL
Human GIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVKWLPQNDL
Horse GIVVFSLGSMVSEIPEKKAMEIADALGKIPQTVLWRYTGTPPPNLSKNTILVKWLPQNDL
Rat LGHPKARAFITHSGSHGIYEGICNGVPMVMMPLFGDQMDNAKRMETRGAGVTLNVLEMTA
Mouse LGHPKTRAFITHSGSHGIYEGICNGVPMVMMPLFGDQMDNAKRMETRGAGVTLNVLEMTA
Sheep LGHPKTRAFITHSGSHGVYEGICNGVPMVMMPLFGDQMDNAKRMETRGAGITLNVLEMSS
Cow LGHPKTRAFITHSGSHGIYEGICNGVPMVMMPLFGDQMDNAKRMETRGAGVTLNVLEMSS
Dog LGHPKARAFITHSGSHGIYEGICNGVPMVMLPLFGDQMDNAKRMETRGAGVTLNVLEMTS
Human LGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTLNVLEMTS
Horse LGHPKTRAFITHSGSHGVYEGICNGVPMVMMPLFGDQMDNAKRMETRGAGVSLNVLEMTS

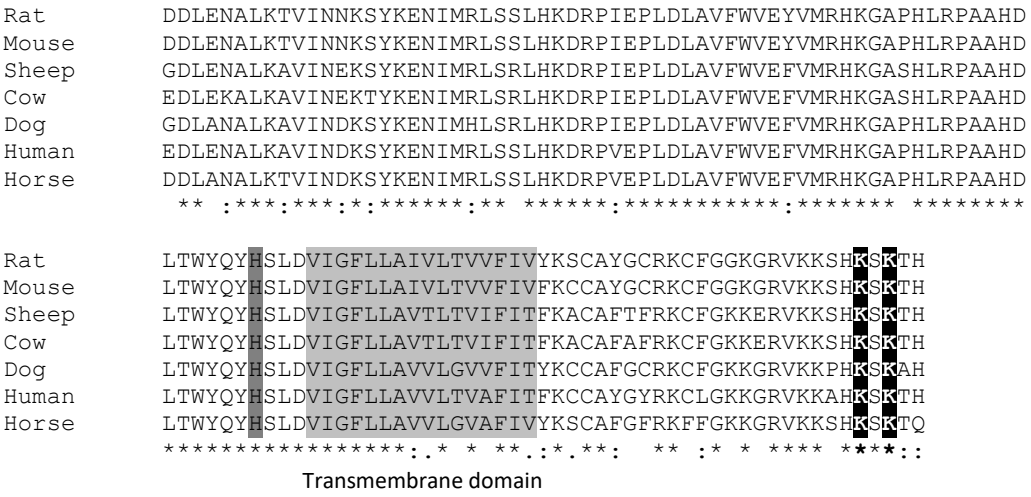


Figure 3.27: Amino acid alignment of *UGT1A6* sequences. *UGT1A6* sequences were retrieved from the NCBI database for rat (AAL67853.1), mouse (NP_659545.2), sheep (NP_001192075.1), cow (NP_777187.1), dog (NP_001003078.1), human (NP_001063.2) and the predicted equine *UGT1A6* sequence. These were aligned using Clustal Omega, * indicated residues conserved between all species. The first 22 amino acids (light grey) indicate the signal peptide region. The signature sequence, indicated by a boxed area, is involved in binding the sugar donor, the transmembrane domain is highlighted light grey and implicated in anchoring the enzyme to the ER, the dilysine (KXK) repeat present in the last five residues of the sequence is highlighted in bold typeface with a black background. Four histidines (H) are involved in substrate selection and glycosylation (highlighted dark grey).

3.3.10 Identification of equine UGT3A1

3.3.10.1 Cross species comparison of the UGT3 locus

A comparison of the equine region encoding the predicted *UGT3A1-like* gene, chromosome 21, 29,238,823-29,257,168, and the human genome using Ensembl (release 89) found the equivalent genetic loci was on human chromosome 5, with the orthologues at position 35,951,010-36,071,358. Figure 3.28 displays this syntenic relationship, with the region on the central chromosome (equine chromosome 21) encoding our gene of interest denoted by the red box; this co-localises to the region on human chromosome 5. Blocks of equine chromosome 21 also co-localise to areas on chromosomes 1 and 19 of humans.

A more detailed look at the locus structure found several of the genes surrounding the *UGT3* genes have been conserved between humans and equines (Figure 3.29). Downstream of the *UGTs*, Calcyphosine like (*CAPSL*), Interleukin 7 Receptor (*IL7R*), and Sperm flagellar 2 (*SPEF2*) are maintained across species. Upstream LMBR1 Domain containing 2 (*LMBRD2*), S-Phase kinase associated protein 2 (*SKP2*), NAD kinase 2, mitochondiral (*NADK2*) and RAN binding protein 3 like (*RANBP3L*) are also conserved. This detailed syntenic comparison was also made with the mouse genome, which encodes two *UGT3* genes on chromosome 15. Analysis of the murine genome up and downstream of the *UGT3* locus found identical gene conservation in these flanking regions (Figure 3.29).

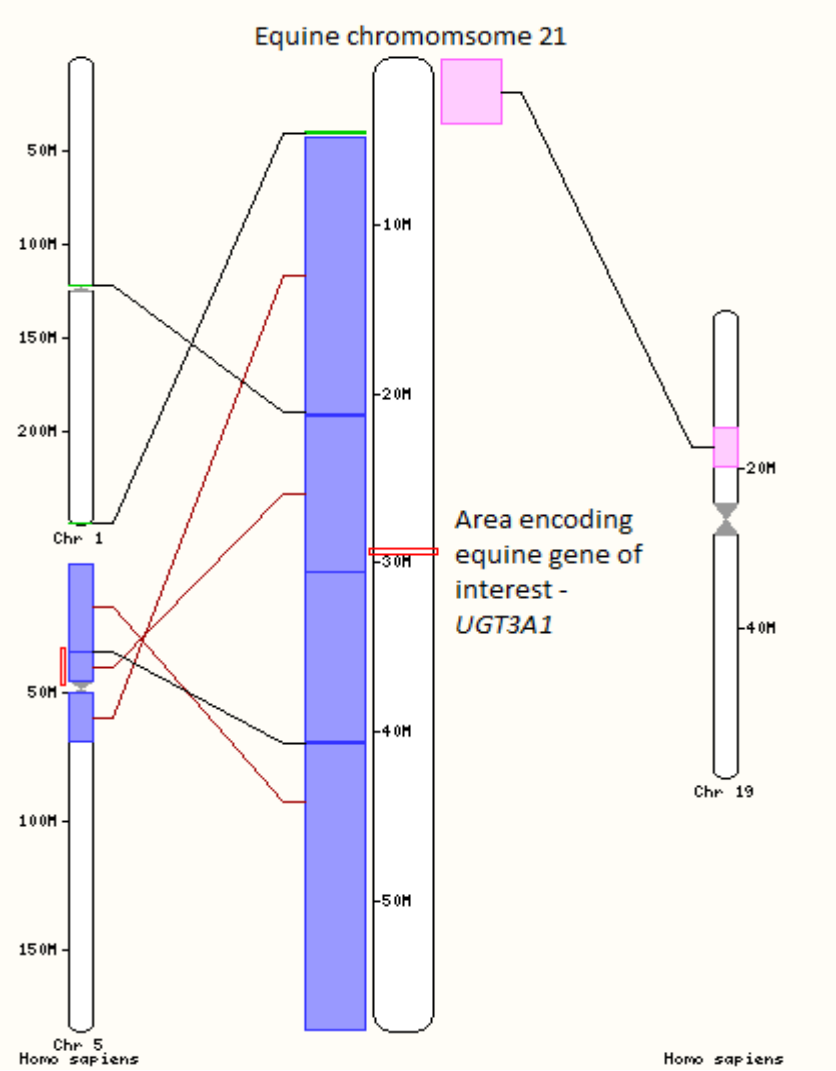


Figure 3.28: Syntenic comparison of equine chromosome 21 and the human genome. The figure (taken from Ensembl, release 89) shows blocks of equine chromosome 21 conserved on human chromosomes 1, 5 and 19. The area on the equine chromosome denoted by the red box co-localises to a region on human chromosome 5, which encodes the human *UGT3A1* and *UGT3A2* genes, surrounding genes are also conserved.

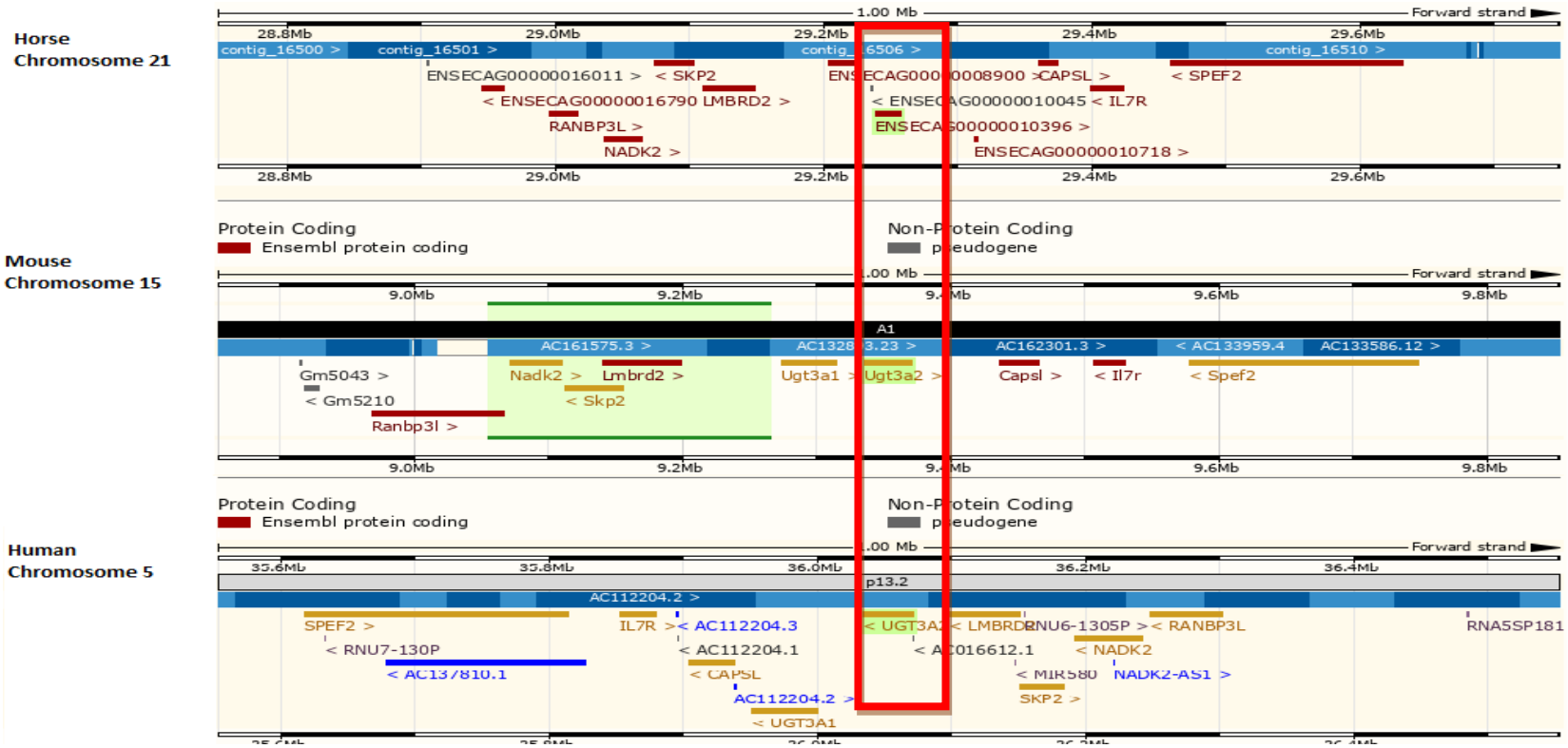


Figure 3.29: Region comparison of *UGT3A1-like* sequences between horse, mouse and human. The predicted horse *UGT3A* sequences are denoted as ENSECAG00000008900 and ENSECAG00000010396. The regions between animals show multiple genes in the region have been conserved throughout evolution, including Calcyphosine like (*CAPSL*), Interluekin 7 receptor (*ILR7*) and Sperm flagellar 2 (*SPEF2*) downstream of the *UGTs* and LMBR1 domain containing 2 (*LMBRD2*), S-Phase kinase associated protein 2 (*SKP2*), NAD kinase 1, mitochondrial (*NADK2*) and RNA binding protein 3 like (*RANBP3L*) upstream.

The *UGT3* genes are composed of seven exons, with exons 1-4 encoding the amino (N) termini of the protein and exons 5-7 encoding the carboxyl (C) termini, conserved to the configuration of the human and mice *UGT3* genes (Figure 3.30). In both humans and mice, *UGT3A2* is located upstream of *UGT3A1*. This appears to be reversed in equines with *UGT3A2-like* located downstream of *UGT3A1* (Figure 3.30). The equine UGT3A locus is 40% smaller than that of the human or mouse locus.

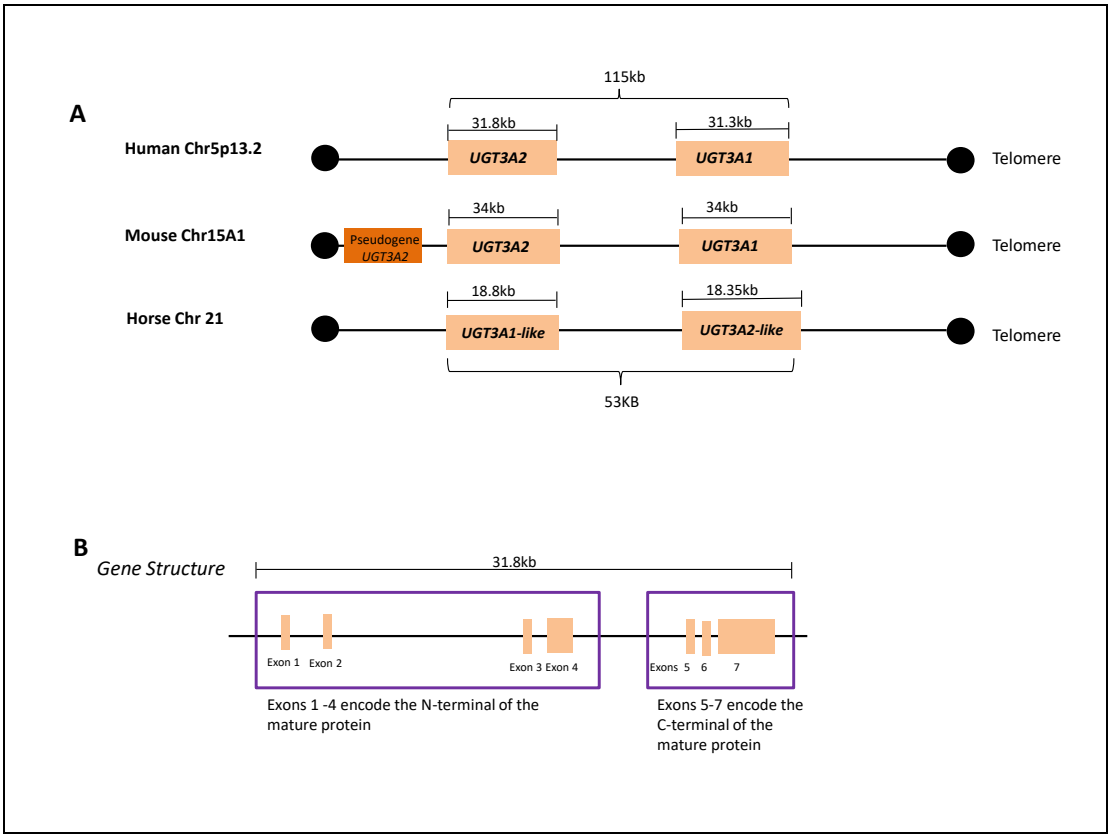


Figure 3.30: **Comparative schematic of the *UGT3* loci in humans, mice, and horse**. Both mice and humans encode two isoforms of the *UGT3s* on chromosome 15A1 and 5p13.2 respectively (A), with *UGT3A2* located upstream of *UGT3A1*. This appears to be reversed in the equine genome. The *UGT3* genes are encoded by 7 exons (B), exons 1-4 comprise the N-termini, exons 5-7 the C-termini, and are located on equine chromosome 21. The human locus is 115kb, with each *UGT* encompassing 32kb, the equine locus is 53kb, with each gene encompassing 18kb.

The equine sequences were assessed for homology against the murine and human *UGT3A1* and *UGT3A2* sequences (Figure 3.31). The heat map displays the results, using a green to red scale with sequences with the highest percentage of shared residues in green. Both equine sequences share a greater sequence identity with *UGT3A2* sequences of humans and mice. The equine sequence identified in Ensembl as ENSECAG00000010396 shares 74.4% sequence similarity to human *UGT3A2* and 63.5% with mouse *UGT3A2*, whereas it displays 71.9% and 62.2% shared residues with human and mouse *UGT3A1* sequences respectively. The equine gene identified as ENSECAG00000008900 was 74.8% identical to human *UGT3A2* and 65.4% to murine *UGT3A2*, while a comparison to human and mouse *UGT3A1* sequences showed 72.4% and 64.6% identity respectively.

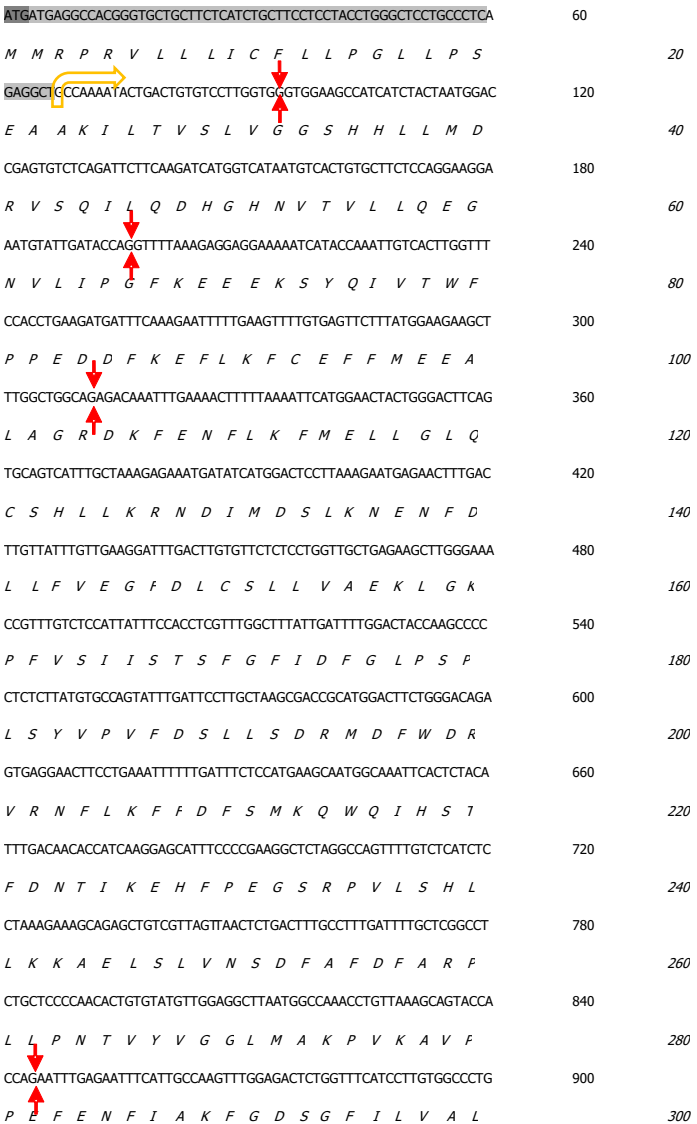
	Mouse *UGT3A1*	Mouse *UGT3A2*	Human *UGT3A1*	Human *UGT3A2*	Horse *UGT3A1-like* (10396)	Horse *UGT3A1-like* (8900)
Mouse *UGT3A1*	100	87.2	64.3	65.6	62.2	64.6
Mouse *UGT3A2*	87.2	100	64.1	65.5	63.5	65.4
Human *UGT3A1*	64.3	64.1	100	78.4	71.9	72.4
Human *UGT3A2*	65.6	65.5	78.4	100	74.4	74.8
Horse *UGT3A1-like* (10396)	62.2	63.5	71.9	74.4	100	83.2
Horse *UGT3A1-like* (8900)	64.6	65.4	72.4	74.8	83.2	100

Figure 3.31: Heat map of percentage conserved residues between *UGT3* sequences. Sequences were pairwise aligned and the percentage of conserved residues was calculated. The spectrum of green to red highlights the high to lower level of homology. Horse *UGT3A1-like* (10396 – retrieved from Ensembl *ENSECAG00000010396*) and horse *UGT3A1-like* (8900 – retrieved from Ensembl *ENSECAG00000008900*) were most similar to human *UGT3A2* (NP_777574.2). Mouse *UGT3A1* (NP_997099.2) and *UGT3A2* (NP_659094.1) and human *UGT3A1* (NP_689617.3) sequences were retrieved from the NCBI.

3.3.10.2 Gene structure

We isolated a 1572bp coding sequence for equine *UGT3A1-like*, encoding a protein of 523 amino acids (Figure 3.32) with a predicted molecular weight of 53kDa.

UGT enzymes can be conceptually divided into two domains. The N-terminus determines the substrate specificity and the C-terminus contains the signature sequence, transmembrane domain, and dilysine motif. UGTs are localised to the ER, spanning the membrane with a type I topology, such that the major part of the protein is located within the lumen. Constructed from seven exons the complete sequence contains a signal peptide, which directs the mature protein to the ER, comprised of the first 22 amino acids (Figure 3.32). Equine *UGT3A1-like* also contains a signature sequence from amino acid position 350 to 396 (greyed box, Figure 3.32) which is responsible for determining the sugar donor specificity, and a putative transmembrane domain and dilysine motif. The signal peptide, signature sequence, transmembrane domain (highlighted by a green box) and dilysine motif (underlined bold typeface) are features characteristic of UGT enzymes (MacKenzie et al., 2011).



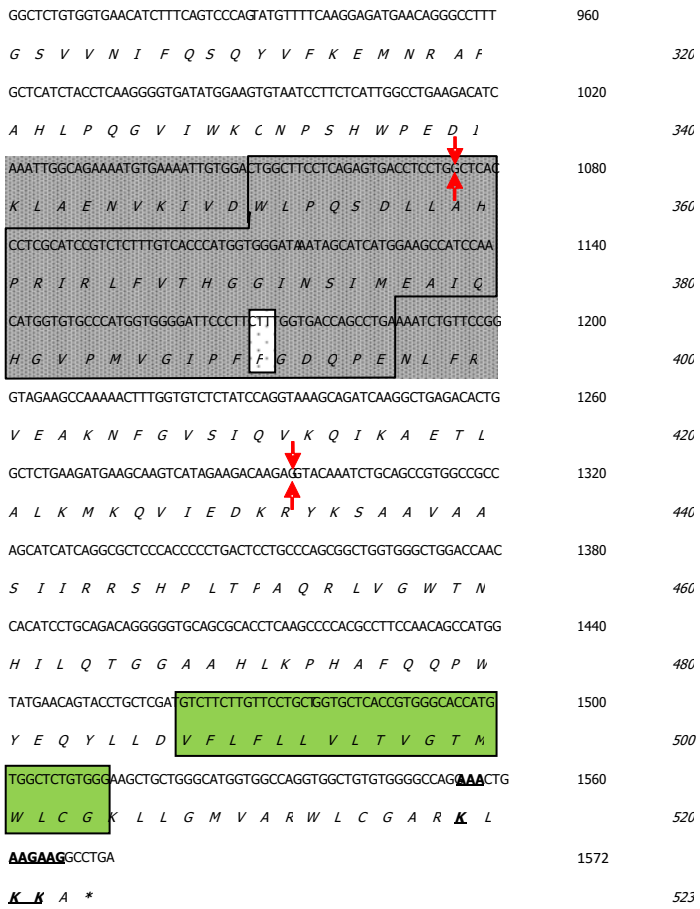


Figure 3.32Nucleotide and protein sequence of equine*UGT3A1*-like. Nucleotides 1-66 denote the signal peptide; the yellow arrow marks the start of the mature protein. Red arrows demarcate exon boundaries. Residues 350 -396 (dark grey box) contains the signature sequence, residue 391 (pale box) determines donor sugar specificity. The putative transmembrane sequence is shown as a green box, with the dilysine motif designated in underlined bold typeface

3.4 Discussion

Glucuronidation is a major pathway in phase 2 metabolism (Jancova et al., 2010, Gibson and Skett, 2001), which involves the conjugation of glucuronic acid to a hydrophobic substrate in order to change the polarity to a more hydrophilic state (Meech and Mackenzie, 2010), enabling the body to excrete the compound via bile and urine (de Wildt et al., 1999, Guillemette, 2003). Whilst the pharmaceutical industry is making progress understanding drug metabolism in people, advances in veterinary animals lag behind. Developing *in vitro* tools to understand drug metabolism in horses will be beneficial to improving health and welfare, as well as detection of illegal drug use in sports.

3.4.1 Why optimise RNA extractions?

The majority of UGTs are predominately expressed in the liver (Radominska-Pandya et al., 2005a), making it the optimal tissue to extract the RNA for the isolation and characterisation of this diverse superfamily of enzymes. Liver tissue is fibrous, with type III collagen reticular fibres, which in turn provide scaffolding to the liver cells. The fibrous nature of liver tissues presents challenges with regards to isolating nucleic acids from the cells. Optimal homogenisation of the tissue was essential for separation and lysing of cells.

The Qiagen RNeasy mini handbook stated that the yield from liver of total RNA should be 40-60µg (for mouse/rat tissue). Initial extractions for equine liver resulted in low quantities and poor quality total RNA.

Optimisation sought to produce a protocol which improved yield and quality of total RNA isolated with the focus of improving homogenisation of liver tissue, as maximising cell lysis should increase the quantity of total RNA available for isolation.

A comparison of several techniques found a two-step protocol was optimal for RNA extraction. The initial homogenisation step was performed using MACs tubes and Qiazol lysis reagent. Step two was to incubate the homogenate with proteinase K

for 1 hour, with the final step being the extraction total RNA using the RNeasy mini kit. This three-step approach increased total RNA isolation by 198-fold and yielded of RNA suitable quality, making this the optimal method for extraction of total RNA from equine liver, suitable for downstream applications.

3.4.2 Why do primer annealing temperatures require testing?

The annealing temperature (aT) of the primers was calculated using Oligo Analyzer, which uses an algorithm to determine the aT based on factors that remain constant; these are: target DNA, Oligo conc 0.25µM, Mg^{2+} 0mM and Na^{+} conc 50mM. These conditions are not always reflective of the chemical environment in which PCR is performed given the broad array of polymerases available for use. This work utilised Phusion High Fidelity Polymerase, so optimisation of the aT specific for use with this polymerase was required. This project required the use of a high fidelity enzyme which is important to ensure accurate amplification for PCR and downstream sequencing reactions (Bryksin and Matsumura, 2010).

The gradient PCRs were performed over a 12-degree range with the Oligo Analyser calculated aT in the centre of the range. PCRs for five of the 12 genes successfully produced a band of the correct size, with *ENSECAG00000008900* and *ENSECAG00000014362* also producing several secondary bands. For both these genes a touchdown PCR was performed which reduced the number of secondary bands but did not eliminate them.

3.4.3 Optimisation of cloning

Initial attempts at cloning were unsuccessful, resulting in alteration of the ligation steps to improve cloning efficiency. Tripling the ligation time and increasing the quantity of insert used in the reaction proved to be sufficient optimisation steps, facilitating downstream applications of sequencing and sub-cloning.

3.4.4 Optimisation of sequencing reactions

The initial attempts at sequencing failed to produce any interpretable sequence data. There are several possible reasons for this including the GC content, the quantity, and size of di- and tri-nucleotide repeats. and whether the sequence contains hairpin structures (Kieleczawa, 2006).

Initial research suggested that it was possible to optimise the thermal cycling parameters of the sequencing reaction to improve the data quality. However, modification to the cycling conditions failed to improve the quality.

We investigated the GC content of each of the genes to determine whether this might impact on sequencing. *ENSECAG00000008900*, *ENSECAG00000010396* and *ENSECAG00000023519* all had a 48% GC content, while *ENSECAG00000014362* and *ENSECAG00000020268* had 44% and 42% GC respectively.

An evaluation of the GC content of the horse genome found the mean GC content to be 48%, similar to that of other placental animals, including humans (~46%), cow (~49.9%), rat and mouse (both ~51%) (Romiguier et al., 2010). It is generally considered that genes with GC content >60% may lead to problems with sequencing (Kieleczawa, 2006). However, with none of the predicted *UGT* genes had a GC content > 48%, and with no extensive repeats identified, these were discounted as contributory factors.

There was the potential for the formation of hairpins to be creating the issues observed. Although increasing the temperature of the initial incubation step is suggested to be a sufficient step for most templates, studies have found the addition of reagents such as DMSO and betaine to be more effective (Kieleczawa, 2006). Contacting our sequencing service provider for advice, they also suggested 5% glycerol and a mix of DMSO + glycerol, at 5% each. Whilst the addition of DMSO, glycerol or a mix of the two did not result in an improvement in the quality of the

sequence data generated, the addition betaine (1M) resulted in good sequence data.

3.4.5 ENSECAG00000008900 – a member of the UGT3 superfamily

In mammals, there are three families of UGTs, UGT1, UGT2, and UGT3. The *UGT1s* are described as having a complex locus with unique exon 1s and shared exons 2-5 (Owens et al., 2005), with the *UGT2As* sharing this structure. The *UGT2Bs* are each encoded separately while the *UGT3s* contain only two members, encoded adjacently in the genome (Jancova et al., 2010).

3.4.5.1 Syntenic comparison of the UGT3 locus

Throughout evolution, with the diversification of species, blocks of genetic information have been maintained although translocation, and recombination events have resulted in these blocks moving to different chromosomes. Syntenic investigations are used to look at specific blocks of genetic information in order to establish orthologues between species (Nomiyama et al., 2013). The investigated sequence, *ENSECAG00000008900*, is predicted to be the equine orthologue of human UGT3A1, termed equine *UGT3A1-like*. Syntenic and sequence analyses provided additional inference that the isolated sequence is a member of the UGT3 superfamily.

The region of equine chromosome 21 encoding *ENSECAG00000008900* was found to correspond to a region on chromosome 5 of the human genome where there are two *UGT3s*, termed *UGT3A1* and *UGT3A2*. Several genes in the region of the human *UGT3s* were found to also be present in the region of *ENSECAG00000008900* on the equine genome. *RANBP3L*, *NADK2*, *SKP2,* and *LMBRD2* are all located upstream of human *UGT3A1*, and also present upstream of the equine gene. Gene locations were also found to be maintained downstream of the investigated equine gene and the human *UGT3* locus. Using Ensembl, the equivalent loci from mouse and rats were also investigated to see whether these genes were maintained in additional species. In both rodents this level of gene location conservation was maintained.

This simple syntenic comparison of genomes demonstrates that the locus structure is maintained between species, and provides confidence that the gene *ENSECAG00000008900* is a member of the *UGT3* family. With mice and humans having two members of the UGT3 family encoded in the genome, a comparison of against the unannotated genes present in the region around *ENSECAG00000008900* (Figure 3.11) was carried out highlighting two additional genes that are predicted to be *UGTs*. This included *ENSECAG00000010396* which forms part of this overall study and has been reported in depth in section 3.3.10. Briefly, *ENSECAG00000010396* showed a high level (89%) of sequence homology to *ENSECAG00000008900*. Interestingly this region in the equine genome also included a third predicted *UGT*, *ENSECAG00000010718*, which has been identified by Moreton et al, (2014) as a potential *UGT*, whereas the human and mouse genomes only include two *UGT3* family members.

ENSECAG00000010718 has not been isolated within this study. However a brief investigation of this gene found that over the length of *ENSECAG00000010718* it shared 85% homology with *ENSECAG00000008900* over exons 3 and 4. It is a much shorter sequence of 642bp compared to 1452bp for *ENSECAG00000008900*, with no start codon evident. The *UGTs* are known to encode pseudogenes (Gong et al., 2001) and truncated versions. For example in patients with Crigler-Najjar syndrome, the truncated versions of UGT1A1 have still been found to co-localise to the ER membrane (Suzuki et al., 2014). Without further study it is not possible to say whether this gene is a pseudogene, splice variant or truncated UGT.

The phylogenetic analysis of *ENSECAG00000008900* against *UGTs* from all three families and a range of species, found it clustered with *UGT3* sequences, most closely associating with cow *UGT3A1* (See section 3.3.6, Figure 3.9). This clustering combined with the syntenic analysis provided additional confidence that this sequence is a member of the UGT3 family but does not confirm which particular enzyme this gene encodes.

3.4.6 Analysing ENSECAG00000020628 - isolated novel member of the UGT2B sub-family

ENSECAG00000020628 was predicted by Moreton et al (2014) to be a transcriptionally active UGT enzyme. This gene was successfully isolated from liver tissue and further investigated in order to establish gene identity.

3.4.6.1 Syntenic analysis of ENSECAG00000020628

The genomic region containing the equine gene *ENSECAG00000020628* was compared against the human genome to determine if the locus had remained conserved. *ENSECAG00000020628* was found to co-localise to a region of the human genome containing multiple UGT genes on chromosome 4. Five UGT genes were encoded in the human genome, all belonging to the UGT2 family of enzymes. Of the five enzymes present, one was a member of the UGT2A sub-family, and four were members of the UGT2B sub-family, of which one is a pseudogene. In both the equine and human genomes there was a lack of gene annotation present which makes it difficult to infer how conserved this locus is. A single annotated gene from the equine locus, *YHDC1*, is present also in the human locus, downstream of *ENSECAG00000020628* and downstream of human *UGT2B17*.

Given the numerous *UGT2*s present in the human locus it is difficult to determine from synteny alone which orthologue has been isolated. In subsequent phylogenetic analyses reported in section 3.3.5 (figure 3.8), *ENSECAG00000020628* was found to cluster with *UGT2B* sequences, suggesting it may be a member of the UGT2B sub-family rather than the UGT2As. Assessing the homology of *ENSECAG00000020628* against the human *UGT2Bs* found greatest sequence similarity to human *UGT2B17*. An attempt to compare *ENSECAG00000020628* against *UGT2Bs* of mice failed to successfully identify a murine orthologue. This is due to issues with nomenclature of *UGT2B17* which also had the alias of *UGT2B5*. The rat genome encodes a *UGT2B17*, and this was the initial sequence used to assess for homology. However, this gene has also been given the synonym of

UGT2B1, *UGT2B10, UGT2B15* and *UGTB34*. Whilst *ENSECAG00000020628* shows 65% homology to rat '*UGT2B17*', we have no confidence in which rat UGT2B isozyme this gene encodes. Therefore, we can only use synteny and homology with human *UGT2* sequences to infer the identity of *ENSECAG00000020628*, and these suggest that we have isolated the orthologue of human *UGT2B17*.

3.4.7 ENSECAG00000014362, an orthologue of the human UGT2A sub-family

ENSECAG00000014362 is identified in the Ensembl database as *UGT2A3*, whereas in the NCBI database this gene is given the alias *UGT2A1*. The computational analyses used to identify this gene appear to suggest it is a member of the UGT2A sub-family. Additional syntenic, phylogenetic, and genetic investigations were performed to clarify the identity of this gene.

3.4.7.1 What can synteny tell us about ENSECAG00000014362?

There is a lack of descriptive annotation for the region of the equine genome around *ENSECAG00000014362* making it difficult to ascertain how well maintained this locus is across species. *ENSECAG00000014362* co-localises to *UGT2A3* in the human genome. Three additional annotated genes are present, namely: *CSN2*, *CSN1S1,* and *SULT1*, which are all upstream of *ENSECAG00000014362* in the equine genome and are also located upstream of human *UGT2A3*. *SULT1* is a fellow conjugating enzyme belonging to the superfamily termed sulfotransferases (Glatt et al., 2000). This group of enzymes contributes to phase 2 metabolism and are only superseded in their contributions to phase 2 by the UGTs (Jancova et al., 2010, Glatt and Meinl, 2004).

Analysis of *ENSECAG00000014362* found it clustered with *UGT2A3* sequences from humans and guinea pigs (information in subsequent chapters). To be confident that the gene was a *UGT2A* and not *UGT2B,* alignments were performed with all the relevant human *UGT* sequences to assess homology. All of the *UGT2B* sequences shared less than 65% homology, with human *UGT2A1* and *UGT2A2* both sharing 65% identity with *ENSECAG00000014362*. Human *UGT2A3* displayed the highest of homology with *ENSECAG00000014362*, with 75% similarity at the amino acid level. Homology with the *UGT2A* sequences from rat and mouse was also assessed, with the murine genome encoding three *UGT2A* isozymes *UGT2A1*, *UGT2A2,* and *UGT2A3*, but only two reported in the rat genome, *UGT2A1* and *UGT2A3*. *ENSECAG00000014362* shared 74% identity with mouse *UGT2A3*, but <67% with mouse *UGT2A1* and *UGT2A2*. Comparison with the rat isozymes found

ENSECAG00000014362 most homologous to rat *UGT2A3*, with 74% identity at the amino acid level. From these initial investigations, it appears the isolated equine gene, *ENSECAG00000014362,* is an orthologue of *UGT2A3*.

3.4.8 Equine ENSECAG00000025319, an orthologue of human UGT1A6

3.4.8.1 What do syntenic investigations tell us?

Throughout evolution blocks of genomic information are maintained across species, often translocating to different chromosomes. Genes are often maintained within blocks, and the conserved synteny between species can help to identify orthologues (Nomiyama et al., 2013). The investigated equine sequence (*ENSECAG00000023519*) was predicted to be an orthologue of human *UGT1A6* (Moreton et al., 2014). Investigations into the synteny between the chromosomal regions encoding this gene in horses and humans provided confidence that the equine sequence may encode a UGT1 enzyme.

The region of equine chromosome 6 encoding our gene of interest was found to correspond to a region on chromosome 2 of humans. The locations of several genes in the region are conserved between horse and humans, including *DGKD* and *USP40* which are both found downstream of *UGT1A6*, and *SPP2* and *TRMP8* positioned upstream. It is difficult to make syntenic comparisons between this region of the equine genome and the rat genome to confirm locus structure as nomenclature of the rat UGT1 family is not clear. The only rat UGT annotated in the Ensembl database identifies this as *UGT1A5*, however the information provided by Ensembl states that the NCBI database has annotated this same gene as *UGT1A6*. Further investigation on the NCBI database highlights the need for both clarity and confidence in the naming of genes as it also states that *UGT1A6* may also be known as *UGT1A7*. Analysis of the genomic region around the location of '*UGT1A5*' of rats in Ensembl shows multiple transcripts of this gene. I hypothesis that each of these transcripts is a different *UGT1*, and this is supported by information available on the NBCI database. A broader look at the chromosome 9 region 95,161,157 - 95,302,822 on the rat genome where '*UGT1A5*' is located shows other genes have been maintained and conserved, with *USP40* and *DGKD* downstream and *TRPM8* and *SPP2* upstream of this *UGT*, as seen in both the human and equine genomes. In addition, the

'*UGT1A5'* clusters with *UGT1A6* sequences from several animals, suggesting *UGT1A6* may be the correct nomenclature for this gene. The mouse genome is more comprehensively annotated for the multiple isozymes of the *UGT1* family, encoded on chromosome 1 region 88,134,809 - 88,218,997, and displays this conserved complex locus and the corresponding locations of *DGKD*, *USP40*, *TRPM8* and *SPP2*.

This conservation of the locus structure and flanking genes between species supports the identified location on equine chromosome 6 for the *UGT1* locus. Currently the equine genome is annotated for a single member of the *UGT1* family, predicted to be *UGT1A6*. While the gene position suggests that this is the orthologue of human *UGT1A6*, the high level of sequence similarity between the *UGT1* isozymes as found in human, rats, and mice (Meech and Mackenzie, 1997a) means synteny alone cannot be relied upon to confirm which equine isozyme is encoded here, and further characterisation is needed.

3.4.8.2 amino acid sequence homology

A ML tree was produced to assess the relationship of the predicted equine *UGT1A6* to *UGT* sequences from 23 species. Homology between the most closely associated *UGT* sequences to equine *UGT1A6* was investigated further.

There was distinct clustering within the phylogenetic tree, with the predicted equine sequence clustering with the *UGT1A6* sequences from primates, rodents, cow, sheep and grey wolf (section 3.3.9.1, figure 3.23), independently of the other *UGT1* sequences. This information suggests that the predicted sequence is likely to be a genuine *UGT1* sequence. The clustering of this equine gene with characterised *UGT1A6* sequences over other *UGT1* isoforms supports the hypothesis that this predicted *UGT* is the equine orthologue of *UGT1A6*.

A comparison of the sequence identities between rat, mouse, human, cow, sheep, and grey wolf *UGT1A6* and the equine sequence showed a high level of conservation, ranging from 78.2-82.9%. The highest level of homology was between the human and grey wolf *UGT1A6* sequences (82%). The sequence similarity between the equine *UGT1A6* with other *UGT1*s is ≤61%. Collectively this suggests that the predicted gene sequence is a *UGT1* family member and most similar to *UGT1A6* than any other *UGT1* isozyme.

3.4.8.3 Size differences in the UGT1A6 sequence

The *UGT1A6* sequences were found to vary in length from 528 to 533 amino acids. The putative equine sequence was the longest, at 533 amino acids, with the human *UGT1A6* sequence a single amino acid shorter, and the dog the shortest at 528 amino acids. This variation is predominantly a result of differences within the signal peptide, which is a 22-amino acid motif in the horse, human, and mouse compared to 21 amino acids in the rat, 19 in the sheep and cow, and 17 in the dog.

3.4.8.4 Substrate specificity of UGT1A6 orthologues

A sequence comparison between the *UGT1A6* amino acid sequences from six animals (rat, mouse, sheep, cow, dog, and human) and the horse found the majority of the sequence variation in the N-terminus. A direct comparison of the N-terminus of the human and equine *UGT1A6* sequences found that 25% of the amino acids differed. In humans UGT1A6 has been established as a phenol-conjugating isozyme with substrates including 1-Naphthol, Paracetamol, and Serotonin (Burchell et al., 2005, Krishnaswamy et al., 2003). Human UGT1A6 also glucuronidates resveratrol (Uchihashi et al., 2012), albeit at a reduced rate, compared to UGT1A8 and UGT1A10 which are the primary enzymes involved in resveratrol glucuronidation (Dellinger et al., 2014). Glucuronidated resveratrol has also been shown to be a substrate of mouse UGT1A6 (Uchihashi et al., 2012). Given the N-terminus of the protein is responsible for substrate selectivity (Rowland et al., 2013, Mackenzie et al., 2005),

this level of sequence homology suggests that each UGT1A6, including equine UGT1A6, may have a broadly similar substrate profile, with an affinity for phenolic compounds, but perhaps altered capacity to glucuronidate specific substrates.

3.4.8.5 Importance of the histidine residues

Studies of the 16 human *UGTs* from the *UGT1* and *UGT2* families have suggested there are four key histidine residues within UGT1A enzymes that have been implicated in glycosylation and substrate specificity (Ouzzine et al., 2000, Kerdpin et al., 2009). A histidine at position 38 is situated in exon 1 and may be involved in substrate selectivity, with mutation of this amino acid to proline altering substrate selection (Kerdpin et al., 2009). Histidine 485, in the C-terminus, has been proven by mutational studies to be important in the structure and function of the protein as abolition of histidine at this position removed enzyme function (Ouzzine et al., 2000). Histidine 361 is postulated to be involved in binding of the donor sugar, and evidence suggests histidine 370 has a catalytic role (Ouzzine et al., 2000). The three histidine residues at amino acids 361, 370, and 485 have been found to be important glycosylation sites in human UGTs (Fujiwara et al., 2009). While the importance of these histidines have not been elucidated in other species, all four are conserved in the *UGT1A6* sequences analysed (Figure 3.27) suggesting they may serve the same role in horse and other animals UGTs as they do for the human UGTs, although mutational and functional studies are required to confirm this hypothesis.

3.4.9 UGT3A1, a novel member of the UGT3 family

3.4.9.1 Syntenic comparison of the UGT3 locus and gene structure

The annotated equine genome encodes two predicted *UGT3A1-like* genes situated on chromosome 21, with Ensembl identifiers *ENSECAG00000008900* and *ENSECAG00000010396*.

Syntenic comparison of the equine chromosomal region containing the *UGT3A* locus with the human genome confirmed that this genetic locus is conserved throughout evolution, with the equine region to human chromosome 5. Using Ensembl, a comparison of the *UGT3A* loci with characterised animals such as the rat and mouse has found the *UGT3A* locus and surrounding genes have been maintained across species. Investigations into the locus of other mammals, including chimp, macaque, and cow also found the locus structure to be conserved (Meech and Mackenzie, 2010). For the locus in each animal, *LMBRD2* sits adjacent to *UGT3A2*, with *NADK2*, *SKP2* and *RANBP3L* all encoded upstream. *CAPSL* sits adjacently to *UGT3A1*, with *ILR7* and *SPEF2* encoded further downstream.

This conserved clustering of genes and the structure of the *UGT3A1* and *UGT3A2* genes are indicative of a duplication event (Meech and Mackenzie, 2010). Additional evidence from the human *UGT3s* of a duplication event comes from the promoter and 5' untranslated region sequences and these elements upstream of *UGT3A2* which share 70% homology with a segment upstream of *UGT3A1* and a region at the 3' end of *UGT3A2* (Meech and Mackenzie, 2010). As the *UGT3* family is small, it has not been subjected to the same level of investigation as the *UGT1* and *UGT2* families. The *UGT1* locus is structurally complex (Ohno and Nakajin, 2009). However many members of the *UGT2* family are individually encoded within the genome (Owens et al., 2005), and it is possible that members of the *UGT2* family have arisen from duplications events. Research into genome-wide duplication events investigated segmental duplications, regions larger than 1kb with >90% homology, in multiple species. Segmental duplications are involved in genomic rearrangements and recombination events (Feng et al., 2017). Genome-wide investigations in cattle, sheep, horse, and pigs found segmental duplication regions were enriched in locations related to xenobiotic metabolism. In particular these species were enriched for segmental duplications in the *UGT2* regions (Feng et al., 2017). This evidence suggests that copies of *UGT2*s may have arisen because of duplication and rearrangement events. This study included genes based on copy numbers, and as the *UGT3* family is very small and contains few copies, it was probably excluded from

this analysis. However, it is plausible that segmental duplications are present within the *UGT3* region and contributed to the duplication of this gene - a hypothesis supported from the high level of homology seen between the promoter and 5' UTR regions (Meech and Mackenzie, 2010).

3.4.9.2 What does the sequence analysis of UGT3A1 tell us?

Both of the predicted '*UGT3A1-like*' amino acid sequences, *ENSECAG00000008900* and *ENSECAG00000010396*, were analysed for homology with *UGT* sequences from 21 species via a ML tree. Both the predicted '*UGT3A1-like*' sequences were shown to cluster with UGT3 sequences from cows, humans, and mice, in a separate clade to that of the UGT1 and UGT2 genes. The sequence analysis, in addition to the information acquired through syntenic investigations, provides confidence that both sequences are *UGT*s and a member of the *UGT3* family and orthologous to one of the human *UGT3As*. What the tree does not indicate is which predicted '*UGT3A1-like'* correlates to which orthologue, *UGT3A1* or *UGT3A2*.

For this, additional information is required from assessing sequence homology, expression, and functional data. In this chapter the sequence isolated, ENSECAG00000010396, shall be the focus of homologous and expression comparisons.

3.4.9.3 Assessing sequence homology, what does this tell us?

As the syntenic and sequence analyses did not provide sufficient evidence to confirm the exact identity of the equine *UGT3s*, in particular *ENSECAG00000010396*, a series of pairwise alignments was performed against human and mouse *UGT3A1* and *UGT3A2* amino acid sequences. Unfortunately this failed to clarify which isozyme was isolated. Both *ENSECAG00000008900* and *ENSECAG00000010396* shared greatest homology with the *UGT3A2* sequence from humans and mice.

The phylogenetic tree of the *UGT3*s (Figure 3.9) displays clustering of the UGT3 isozymes in the primates, which is not seen in other animals. Human, chimp, Rhesus macaque, and gibbon *UGT3A2* isozymes cluster separately from their *UGT3A1* sequences. This clustering supports investigations performed by Meech and Mackenzie (2010), which also found primate isozymes clustering, but this was not replicated in the rat, mouse, horse, or cow (Meech and Mackenzie, 2010). Assessment of the homology of cow, dog, horse, mouse, and rat *UGT3* sequences found their *UGT3A1* and *UGT3A2* orthologues were more homologous to human *UGT3A2*. The same was true when the comparison was made to chimp *UGT3s* (Meech and Mackenzie, 2010). This means, on the basis of sequence homology alone, it is difficult to determine whether *ENSECAG00000010396* is an equine orthologue of *UGT3A1* or *UGT3A2*. Meech and Mackenzie (2010) put forth the suggestion that it may be more appropriate to name the isozymes based on their locations relative to the *LMBRD2* and *CAPSL* genes as these are maintained across species.

3.4.10 Assessing isolated sequences for features characteristic to the UGTs

The UGT polypeptides, as with all proteins, are comprised of two halves: the N-terminus, responsible for determining substrate specificity (Guillemette, 2003), and the C-terminus which selects for the donor sugar and anchors the enzyme to the ER membrane (Ouzzine et al., 1999). The *UGTs* encode certain features characteristic of membrane bound proteins: a signal peptide which directs the polypeptide to the ER membrane (Ahn et al., 2012), a transmembrane domain, and a dilysine motif which anchor the enzyme to the ER membrane (Mackenzie et al., 2008, Ouzzine et al., 1999, Andersson et al., 1999). Additionally in the C-terminus is a signature sequence which determines the donor sugar utilised in the conjugation reactions (Meech et al., 2012b).

3.4.10.1 The signal peptide

Proteins are transcribed in ribosomes in the cytosol. However to be directed to a specific organelle within a cell they require a motif to direct their transport. For proteins directed to the ER membrane, this is called a signal peptide (Walter and Johnson, 1994). Research into insect *UGTs* discovered these initial 22 amino acids are responsible for directing the mature protein to the ER (Ahn et al., 2012), and this has subsequently been proven in humans (Meech et al., 2012b).

Signal peptides can be of variable lengths, on average 16 to 30 amino acid residues, and composed of different amino acid, yet they maintain a constant tripartite structure (Kapp; et al., 2009, Martoglio, 2003). This tripartite structure is constructed from three regions termed n-, h- and c- (Figure 3.33) (Martoglio, 2003). The n- region is variable in length and is the amino terminus of the polypeptide, diverse with regards to amino acids and length (Martoglio, 2003), The equine n-region is composed of only two amino acids, methionine and alanine, which are hydrophobic amino acid, and this is true of all the UGTs represented in figure 3.34. The h- region is also variable, but typically composed of 5-15 hydrophobic amino acids, the majority of which are hydrophobic (Walter and Johnson, 1994, Kapp; et al., 2009). interacts with a signal recognition particle (SRP) present in the cytosol (Figure 3.35) (Kapp; et al., 2009). This interaction facilitates the targeting of the polypeptide-ribosome complex to the ER where the SRP interacts with the SRP receptor on the ER membrane. The docked complex enables the amino acid chain to interact with the ER translocation channel. Once this linkage is established the synthesis of the polypeptide chain continues. As protein synthesis continues the protein is translocated to the ER membrane (Kapp; et al., 2009, Seppen et al., 1996). The c-region encodes a cleavage signal where signal peptidase removes the signal peptide once the polypeptide is anchored to the ER membrane, figure 3.35 (Kapp; et al., 2009).

The tripartite structure of the signal peptide, with the n-, h-, and c-regions, appears to be characteristic of the signal peptide, however there is a great diversity in size and sequence of the signal peptides (Martoglio, 2003). There is diversity in the length of the signal peptides of rodents, sheep, cow, dog, humans, and horse. The first five amino acids (Figure 3.27) are conserved in the seven sequences, which accounts for the n-region and the first few residues of the h-region. It is the h-region where the diversity between sequences exists. These sequence differences may have no effect, or they may impede the ability of the enzymes to translocate to the ER membrane. The exact tripartite structure of the equine sequence remains to be experimentally confirmed. Mutational studies, including truncations, will confirm the importance of individual residues and the importance of the signal peptide.

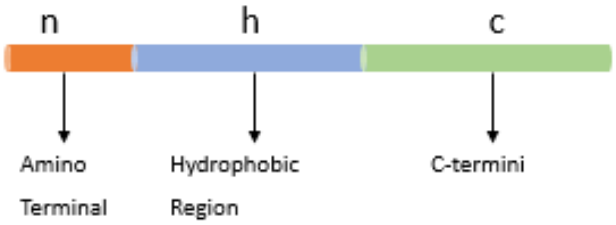


Figure 3.33: Schematic of the signal peptide. The signal peptide can be of variable length, comprised of a tripartite structure. The n-region encodes the amino terminus, the h- region is the hydrophobic core and the c-region encodes a cleavage site. Once the polypeptide is anchored into the ER membrane, enzymes cleave the signal peptide and the mature protein is formed.

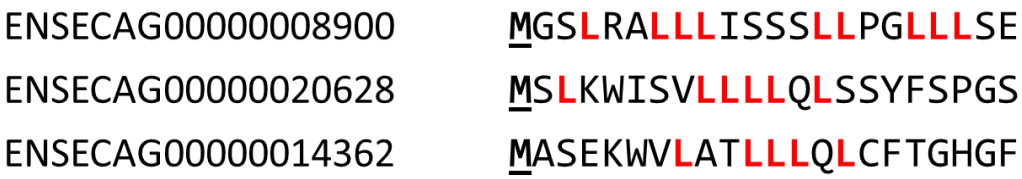


Figure 3.34: Alignment of equine signal peptides from equine *ENSECAG00000008900*, *ENSECAG00000020628* and *ENSECAG00000014362*. The 22 amino acid signal peptides of the three equine genes were aligned to assess homology. Methionine (M – bold underlined) is the only residue conserved in all three sequences. *ENSECAG000000089000* is enriched for leucine (L – highlighted red) residues. *ENSECAG00000020628* and *ENSECAG00000014362*, also members of the UGT2 family, display differing amino acid compositions for their signal peptides.

A comparison of the signal peptides of the three of the equine genes isolated, *ENSECAG00000008900*, *ENSECAG00000020628* and *ENSECAG00000014362,* show sequences of similar length but different amino acid composition. The only amino acid that is conserved in all of the equine signal peptides is the first amino acid (Figure 3.34), indicating the start of the polypeptide, methionine. *ENSECAG00000008900* encodes a leucine-rich signal peptide, with nine of the 22 residues a leucine, which is non-polar. The signal peptides from *ENSECAG00000020628* and *ENSECAG00000014362* include less than six leucines, demonstrating that even within one species, the signal peptide can vary in composition.

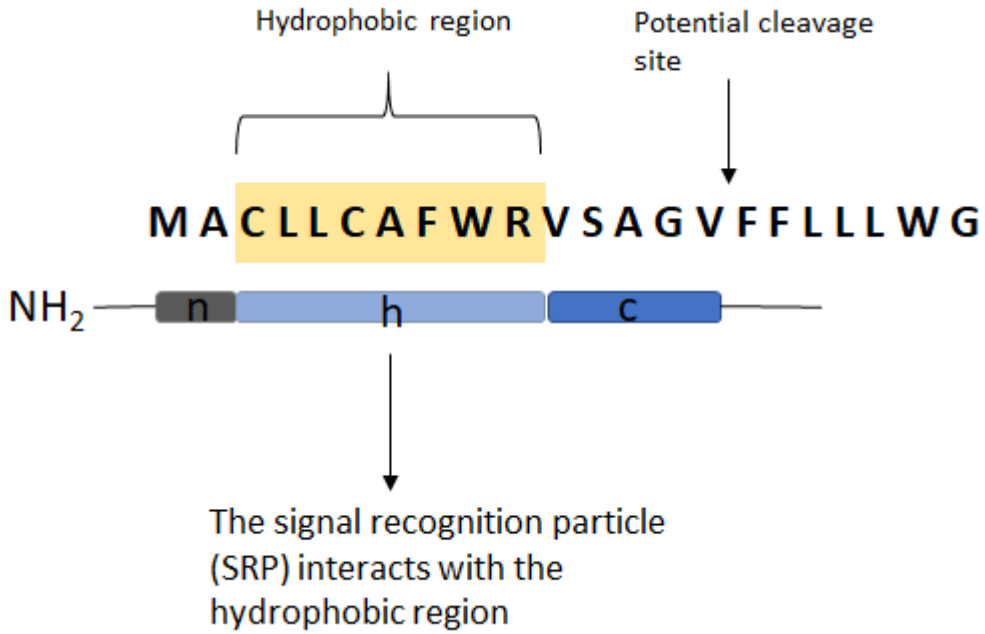


Figure 3.35: Equine signal peptide. The 22-amino acid signal peptide has a tripartite structure, n-region at the beginning of the sequence, an h-region of variable length which is hydrophobic, and the c-region and cleavage site. The hydrophobic region interactions with a signal recognition particle (SRP) present in the cytosol, the polypeptide/SRP/ribosome complex translocate to the ER membrane where the SRP interacts with the SRP receptor on the ER membrane. The polypeptide interacts with the ER translocation channel and as synthesis continues the UGT is anchored to the ER membrane. The SRP is cleaved by the signal peptidase.

Looking at the *UGT1* sequences from humans shows how variable the signal peptide can be within one group of enzymes from one species. The longest putative signal peptide is 30 amino acids (human UGT1A3, UGT1A4 and UGT1A5) while the shortest signal peptide of 24 amino acids is seen in UGT2B15 and UGT2B4 (Kerdpin et al., 2009). As demonstrated in the equine sequences above, the human *UGT1*s also have no overall consensus sequence for the signal peptides they encode, proving not only to be variable in length but sequence as well (Kerdpin et al., 2009).

Both of the human *UGT3* sequences encode a signal peptide of 22 amino acids (MacKenzie et al., 2011). A consensus sequence of these two human *UGT3* signal peptides shows the 22-amino acid region to be leucine rich and 86% homologous with three variable residues between the two sequences (Figure 3.36, green 'x'). A comparison against the signal peptide of *ENSECAG00000008900* shows the equine signal peptide is highly hydrophobic and also leucine rich. In the hydrophobic core are three serine (S) residues (Figure 3.36, highlighted yellow), the human UGT2B sequences contain a single serine in their hydrophobic region (Kerdpin et al., 2009), serine is a polar amino acid. Of the 22 amino acids, 11 (50%) are conserved in the signal peptides between the human and equine *UGT3* sequences. This shows that although there is diversity within a species in the signal peptide, there can also be homology in the signal peptide between species.

Like the human (Mackenzie et al., 2008) and mouse UGT1A6 signal peptide, the homologous region in the horse is also 22 amino acids long. Between the human and equine sequences 16 of the 22 amino acids are conserved, suggesting that the role of the signal peptide in directing the mature protein to the ER is also a requirement of equine *UGT1A6*.

Figure 3.36: Comparison of signal peptides. The signal peptide of equine *ENSECAG00000008900* and a consensus sequence of human *UGT3A1* and *UGT3A2* highlights the leucine rich nature of the sequences. The equine sequence contains three polar serine (S) residues in the centre of the hydrophobic region.

The signal peptide is cleaved once the polypeptide chain is incorporated and anchored to the ER membrane by signal peptidase (Kapp; et al., 2009, Walter and Johnson, 1994), which then degrades to leave the mature protein. The remainder of the N-terminus encodes the substrate specificity of the isozyme (Guillemette, 2003).

3.4.10.1.1 Comparing signal peptides across species

The annotated equine sequence contained a signal peptide encoded by the first 66 nucleotides; the signal peptide directs the mature protein to the ER membrane (Mackenzie et al., 2008, Ahn et al., 2012). A comparison of the signal peptide of the human *UGT3*s found the 22 amino acid sequence to be conserved, with only 3 amino acids different between *UGT3A1* and *UGT3A2* (MacKenzie et al., 2011). A comparison of both equine sequences, *ENSECAG00000008900* and *ENSECAG00000010396*, at the nucleotide level found 81% similarity between both predicted equine *UGT3s* signal peptides. A comparison of the translated sequence found 15 of the 22 amino acids conserved between the two predicted equine *UGT3A* sequences, a much lower level of conservation than is seen between the human isozymes (Figure 3.37).

Figure 3.37: Intraspecies comparison of the signal peptide. The signal peptide of *UGT3A1* and *UGT3A2* of humans was compared for homology, the sequences share 86% of amino acids. A comparison of the two predicted equine sequences, *ENSECAG00000008900* and *ENSECAG00000010396* (abbreviated to 8900 and 10396) shows the homology in the equine sequences is lower, 68%.

In the human sequences, the valine to alanine change and alanine to glutamine does not alter the charge in sequence as all amino acids are non-polar, however the change from Serine (S) to Proline (P) alters the amino acid from polar to non-polar. In the equine sequences three of the amino acid differences result in a change of polarity in the signal peptide; Serine (S) to Arginine (R) is a change of a polar residue to a positively charged residue. Serine to Cytosine (C) and Serine to Phenylalanine (F) both result in a change of polar to non-polar amino acids.

Figure 3.38 exemplifies the differences between human and equine *UGT3* signal peptides, and the additional comparison of equine *UGT1A6* and ENSECAG00000010396, which shows just how diverse these two signal peptide sequences can be within the *UGTs* within one species (Figure 3.38b). Figure 3.38b displays the diversity between *UGT* sequences across species, with *UGTs* represented from humans (NP_689617.3), silkworm (*bombyx mori* – NP_001135960.1) and *Arabiopsis thaliana* (ANM58449.1). An additional comparison was also made with an enzyme from phase 1 metabolism, human CYP45-2D6 (AIA09571.1), and a non-drug metabolising enzyme, human cyclophilin B (M60857), which is a protein involved in the binding of the immunosuppressive drug cyclosporin A (Price et al., 1991). This schematic displays the different amino acids and lengths of regions within the signal peptides. The signal peptides have been

described as typically containing a positively charged n-region (Kapp; et al., 2009), but equine *UGT1A6* and human *UGT3A1* contain non-polar residues in the n-region. Equine *ENSECAG00000010396* contains two positively charged amino acids. The hydrophobic region varies in length mostly consisting of non-polar residues (denoted by ':'). Interestingly the signal peptide of the *UGT* from *Arabidopsis thaliana* has charged amino acids in what should be the 'hydrophobic region'. It is the hydrophobic region which interacts with the SRP to translocate the nascent polypeptide chain to the ER membrane (Kapp; et al., 2009, Seppen et al., 1996, Martoglio, 2003). The c-region appears to be short in the *UGTs* and, as seen in the n- and h-regions, is constructed from a variety of polar and non-polar amino acids. This region encodes a signal peptidase cleavage site which, once the polypeptide chain is at the ER membrane and anchored, is cleaved to produce the mature protein (Martoglio, 2003). This exemplifies the diversity of the UGTs and their signal peptides, and how little we know about how they work.

The N-terminus is responsible for determining the substrate specificity of the UGT. Human UGT3A1 and UGT3A2 have been shown to metabolise different substrates, with UGT3A1 metabolising ursodeoxycholic acid (Mackenzie et al., 2008) and UGT3A2 1-naphthol and genistein (MacKenzie et al., 2011). One would hypothesise that as they metabolise different substrates, elements of the amino acid sequence of the N-termini would be different between the two enzymes and this was found to hold true, with the region between positions 60 and 120 in the N-termini to be only 50% homologous (Meech and Mackenzie, 2010). A comparison of the same regions of *ENSECAG00000008900* and *ENSECAG00000010396* found that it to show the greatest amount of variation in the N-termini; homology in this region was 68%. Whilst mutational and truncation studies need to be performed to confirm this region as the substrate binding domain, it may prove to be significant in substrate specificity.

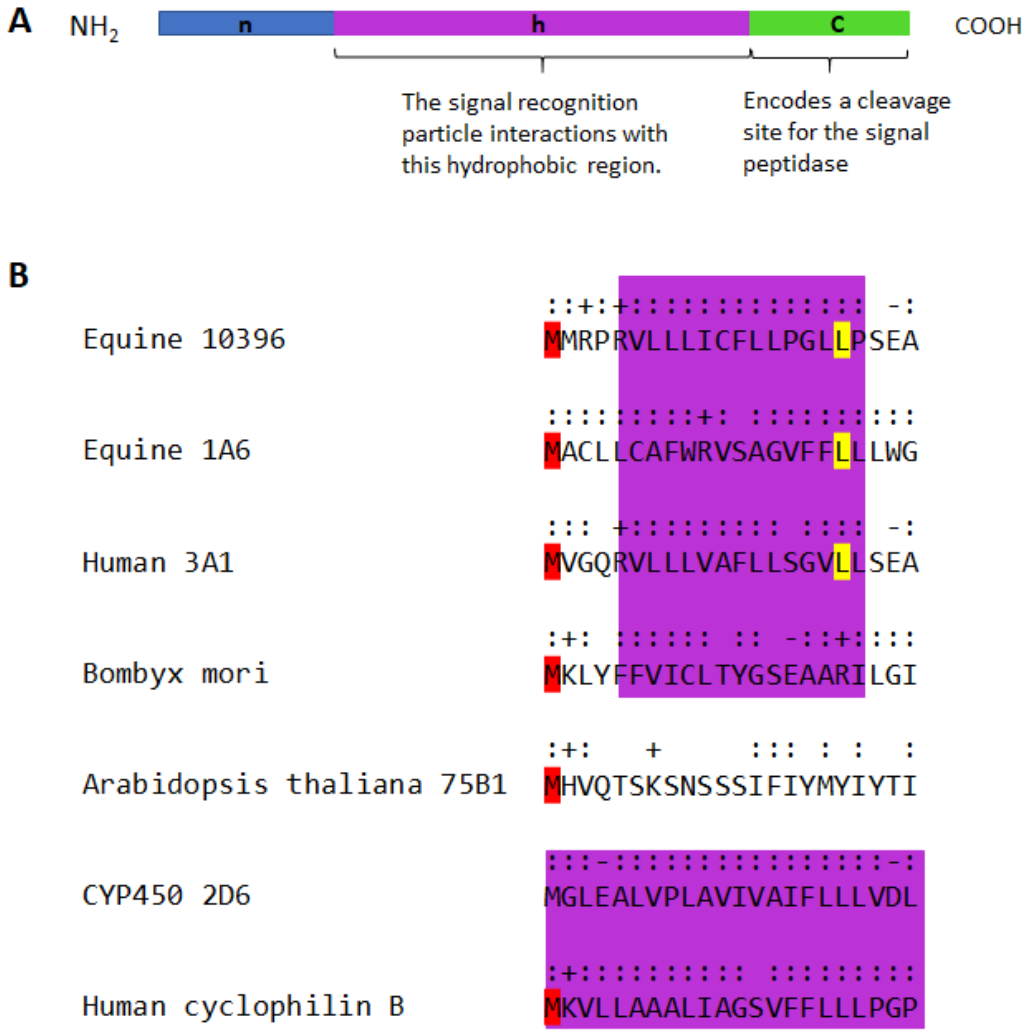


Figure 3.38: Schematic and comparisons of multiple signal peptides. A) Displays the tripartite construct of the signal peptides three regions, regions n-, h-, and c-. The n- region at the beginning of the sequence, the h-region is of variable length, it is this hydrophobic region which interacts with a signal recognition particle (SRP) which ultimately translocate the nascent polypeptide chain to the ER membrane and the c-region which encodes the cleavage site for the signal peptidase. B) displays the signal peptides from multiple species including equine predicted *UGT3* (*ENSECAG00000010396* – abbreviated to 10396), equine *UGT1A6* (sequence from chapter 4), human *UGT3A1* (NP_689617.3 – abbreviated to 3A1), *Bombyx mori* (Silk worm - NP_001135960.1) and *Arabiposis thaliana* (ANM58449.1) including a human phase I metabolising enzyme Cytochrome P450-2D6 (AIA09571.1) and enzyme that is neither a Cytochrome nor a *UGT*, human cyclophilin B (M60857). These sequences show the diversity in the signal peptides and the range of charged amino acids, '+' indicates positively charged, '-` indicated negatively charged, ':' indicated non-polar and where there is no symbol above a letter, the amino acid is considered polar.

3.4.10.1.2 Implications of mutational studies on the signal peptide

Mutational studies within this hydrophobic domain of a human bilirubin metabolising UGT were found to inhibit the translocation of the protein to the ER membrane, with a specific mutation at position 15 resulting in the development of Crigler Najjar type II (Seppen et al., 1996). Another study found the mutation at position 18, Cysteine – Arginine, in the hydrophobic core of the signal peptide of the human preproparathyroid hormone resulted in the development of autosomal dominant familial isolated hypoparathyroidism (Datta et al., 2007). A separate study, involving truncation of the signal peptide of human UGT1A6 found that removal of the N-terminus did not impede the ability of the isozyme to anchor to the ER membrane. The translocation of the synthesized polypeptide to the ER membrane was not impaired until the N-terminus region spanning amino acids 140-240, was truncated (Ouzzine et al., 1999), suggesting that there is a secondary and internal signal motif which can direct the protein.

3.4.10.2 The signature sequence

The UGTs across bacteria, insects, plants, and mammals have been found to include a signature sequence which has been implicated in the binding of the donor sugar (Ahn et al., 2012, Ross et al., 2001, Ahn et al., 2014, Huang et al., 2008). Silkworms have been found to encode 42 *UGTs* (Huang et al., 2008), and plants have been shown to encode a diverse range of *UGTs* (Ross et al., 2001) which would appear to suggest that the signature sequence is a crucial feature of the *UGT*s. The signature sequence, which is 44 amino acids long, is present in the C-terminus of the polypeptide (Ouzzine et al., 1999).

3.4.10.2.1 Signature sequence of ENSECAG00000020628

From the syntenic, phylogenetic and homology analyses, *ENSECAG00000020628* was found to be a member of the *UGT2B* family, showing greatest similarity to human

UGT2B17. To determine if *ENSECAG00000020628* might utilise the same donor sugar as human UGT2Bs, the signature sequences were aligned.

The alignment of *ENSECAG00000020628* against human *UGT2B4*, *UGT2B7*, *UGT2B11*, *UGT2B15*, *UGT2B17* and *UGT2B28* found the signature sequences were highly conserved with only 6 amino acids differing between them. Homology of 86% in the signature sequence (Figure 3.39 – conserved residues indicated by an asterisk) would support the suggestion that the equine UGT encoded by *ENSECAG00000020628* would have a preference to utilise the same donor sugar as the human UGT2Bs, namely UDPGA (Owens et al., 2005).

```
Human UGT2B4     WIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVGVPLFADQ
Human UGT2B7     WIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVGIPLFADQ
Human UGT2B11    WIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVGIPLFFDQ
Human UGT2B28    WIPQNDLLGLPKTRAFITHGGANGIYEAIYHGIPMVGIPLFWDQ
EMBL20628        WIPQNDLLGHPKTKAFITHGGTNGIYEAIYHGIPMVGIPLFADQ
Human UGT2B15    WLPQNDLLGHPKTKAFITHGGTNGIYEAIYHGIPMVGIPLFADQ
Human UGT2B17    WLPQNDLLGHPKTKAFITHGGTNGIYEAIYHGIPMVGIPLFADQ
                 *:******* ***:*******:***************:*** **
```

Figure 3.39: Alignment of signature sequences. Human *UGT2B* signature sequences were aligned against the equine *ENSECAG00000020628* signature sequence (Abbreivated to EMBL20628 – highlighted yellow). 86% of the amino acids were conserved between all of the sequences, indicated by an asterisk.

3.4.10.2.2 Assessing homology of the signature sequence of *ENSECAG00000014362*

The syntenic investigations inferred that *ENSECAG00000014362* was an orthologue of human *UGT2A3*, supported by results from the phylogenetic and homology analyses. The greatest level of sequence homology was with *UGT2A3* from human, mouse, and rat (74%), with homology to *UGT2A1* and *UGT2A2* sequences from these species below 65%.

Despite this, an alignment of all *UGT2A* signature sequences shows a high level of homology has been maintained between all sequences. 73% of the amino acids have been conserved between all nine sequences included in the analysis (Figure 3.40). This indicates that it is highly likely that all *UGT2* sequences in mice, rats, humans and also the horse use the same donor sugar, UDPGA (Owens et al., 2005).

```
Human UGT2A1   WIPQNDLLGHPKTKAFITHGGTNGIYEAIYHGVPMVGVPMFADQPDNIAHMKAKGA
Mouse UGT2A1   WIPQNDLLGHPKTRAFITHGGTNGIYEAIYHGIPMVGVPMFADQPDNIAHMKAKGA
Rat UGT2A1     WIPQNDLLGHPKTRAFITHGGTNGIYEAIYHGIPMVGVPMFADQPDNIAHMKAKGA
Human UGT2A2   WIPQNDLLGHPKTKAFITHGGTNGIYEAIYHGVPMVGVPMFADQPDNIAHMKAKGA
Mouse UGT2A2   WIPQNDLLGHPKTRAFITHGGTNGIYEAIYHGIPMVGVPMFADQPDNIAHMKAKGA
Mouse UGT2A3   WIPQNDLLGHPKTKAFITHGGTNGIYEAIYHGVPMVGVPMLGDQPHNIAHMEAKGA
Rat UGT2A3     WIPQNDLLGHPKTRAFITHGGTNGIYEAIYHGVPMVGIPMFGDQPYNIAHMEAKGA
Human UGT2A3   WIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGA
Equine 14362   WMPQNDLLGHPKAKAFITHGGTNGIYEAIYHGVPMVGVPMFADQPDNIAHMKAKGA
               *:**********::******* **********:****:*::.**  *****:****
```

Figure 3.40: Alignment of UGT2A sub-family signature sequence. Signature sequences from humans, mice, rat and horse have been aligned to assess homology. Humans and mice encode three UGT2A members, rat encode two, UGT2A1 and UGT2A3. Over the 44 amino acids, 73% are conserved in all ten sequences.

3.4.10.2.3 Signature sequence of *UGT3*s

Across mammals the signature sequences of the *UGT3*s appear to be highly conserved. In equines, the signature sequence is present at amino acids 351-396. Meech et al (2012) compared the signature sequences from *UGT3A1* and *UGT3A2* of primates and several non-primates, and found within a species the signature sequence is highly conserved (Meech et al., 2012b). A comparison of the signature sequence of *UGT3A1* and *UGT3A2* within primates consistently found homology to be >82%. Comparison of the mouse and dog signature sequences showed homology was >90% (Meech et al., 2012b). A comparison of 33 signature sequences determined the level of conservation over the 44 amino acid structure to be 52% (Meech et al., 2012b).

In humans UGT3A1 and UGT3A2 utilise different donor sugars, with UGT3A1 using UDPGlcNAc whilst UGT3A2 uses UDPGlc (Mackenzie et al., 2008, Meech et al., 2012b). Mutational studies determined the amino acid at position 391 to determine the preference of donor sugar for each isozyme (Mackenzie et al., 2008), with *UGT3A1* possessing asparagine and *UGT3A2* possessing phenylalanine at this position (MacKenzie et al., 2011). Meech et al (2012) looked at the signature sequence of the *UGT3*s in several primates, including baboon, marmoset, and gorilla, and found this distinction to be present in all animals. Comparing the signature sequence of *ENSECAG00000008900* against the signature sequence of human *UGT3A1* and *UGT3A2* found homology between the three motifs to be 82% (Figure 3.41). The residue that confers sugar specificity (highlighted red) in the equine sequence is phenylalanine, which would suggest that this sequence is an orthologue of human *UGT3A2*.

```
ENSECAG00000008900      WLPQSDLLAHPHIRLFVTHGGINSIMEAIQHGVPMVGIPVFGDQ
Human UGT3A1            WLPQSDLLAHPSIRLFVTHGGQNSVMETIRHGVPMVGLPVNGDQ
Human  UGT3A2           WLPQSDLLAHPSIRLFVTHGGQNSIMEAIQHGVPMVGIPLFGDQ
```

Figure 3.41: Alignment of signature sequences. The signature sequence for human *UGT3A1* and *UGT3A2* were aligned against the isolated equine sequence *ENSECAG00000008900*. The three sequences show 82% homology, with 8 amino acids different between them (in bold green). Position 391 has been found in humans to confer sugar specificity (highlighted red). The equine sequence encodes phenylalanine at position 391, suggesting it utilises UDPGlc as the donor sugar and is an orthologue of human *UGT3A2*.

A comparison of the signature sequence across *UGT3A1* and *UGT3A2* enzymes from six mammals, plus both of the equine sequences found the region to possess 32 amino acids (70%) conserved between all sequences (Figure 3.42). Within the signature sequence a single residue, position 391, has been found in humans to determine donor sugar preference (Meech et al., 2012b). Human *UGT3A1* typically encodes asparagine at position 391 and with functional studies showing that this resulted in this enzyme preferentially using N-acetyl-glucosamine. A comparison of *UGT3A1* sequences from chimps, gorilla, baboon, and marmoset also found position 391 to encode asparagine (MacKenzie et al., 2011, Meech et al., 2012b). However, functional assessment of human UGT3A2 in the presence of N-acetyl-glucosamine resulted in no activity. Sequence analysis of human *UGT3A2* identified phenylalanine at position 391 and the resulting enzyme was shown to be functionally active in the presence of UDP-glucose (MacKenzie et al., 2011, Meech and Mackenzie, 2010). This suggests that it is possible to discriminate between *UGT3A1* and *UGT3A2* based on the amino acid present at position 391. This certainly appears to be true for primates, comparisons of the signature sequences from human, chimp, marmoset and macaque *UGT3A1s* showed all encode asparagine at position 391 whilst the *UGT3A2* sequences encode phenylalanine (Figure 3.42) (Meech et al., 2012b). However, analysis of bovine and murine *UGT3A1* and *UGT3A2* sequences throws this theory into doubt. If the theory of discriminating between enzymes based on the

amino acid at position 391 were to hold true then cow *UGT3A1* and mouse *UGT3A1* would both encode asparagine. However they both have phenylalanine at this position. Interestingly both murine *UGT3A1* and *UGT3A2* both have phenylalanine at position 391. A look at both the equine predicted *UGT3A1-like* sequences found they encode phenylalanine in each sequence, suggesting the importance of residue 391 to confer sugar specificity may be limited solely to primates.

```
Mouse      3A1          WLPQIDLLAHPSIRLFVTHGGMNSVMEAVHHGVPMVGIPFFGDQPE
Mouse      3A2          WLPQTDLLAHPSIRLFVTHGGMNSVMEAVHHGVPMVGIPFFFDQPE
Cow        3A1          WLPQNDLLGHPRIRLFVSHGGMNSIMEAIQHGVPMVGIPLFGDQHE
Human      3A1          WLPQSDLLAHPSIRLFVTHGGQNSVMEAIRHGVPMVGLPVNGDQHG
Chimp      3A1          WLPQSDLLAHPSIRLFVTHGGQNSVMEAIRHGVPMVGLPVNGDQHG
R.Macaque  3A1          WLPQSDLLAHPSIRLFVTHGGQNSVMEAIRHGVPMVGLPVNGDQHG
Marmoset   3A1          WLPQSDLLAHPSIRLFVTHGGQNSIMEAIRHGVPMVGLPVNGDQHG
Human      3A2          WLPQSDLLAHPSIRLFVTHGGQNSIMEAIQHGVPMVGIPLFGDQPE
Chimp      3A2          WLPQSDLLAHPSIRLFVTHGGQNSIMEAIQHGVPMVGIPLFGDQPE
R.Macaque  3A2          WLPQSDLLAHPSIRLFVTHGGQNSIMEAIQHGVPMVGIPLFGDQPE
Marmoset   3A2          WLPQSDLLAHPSIRLFVTHGGQNSIMEAIQHGVPMVGIPVFGDQPE
Equine 3A1-like         WLPQSDLLAHPRIRLFVTHGGINSIMEAIQHGVPMVGIPFFGDQPE
Equine 3A2-like         WLPQSDLLAHPHIRLFVTHGGINSIMEAIQHGVPMVGIPVFGDQPE
                        **** ***.** *****:*** **:***::*******:*.  **
```

Figure 3.42: Alignment of the amino acid signature sequences. The signature sequences from seven species were aligned, 70% of residues are conserved between all sequences. Residue 391, highlighted in grey, confers the sugar donor specificity of the human enzymes. *UGT3A1* sequences from the primates; Human (NP_689617.3), Chimp (XP_526949.3), Rhesus Macaque (XP_001093373.1) and Marmoset (XP_002763568.2), the latter three are predicted sequences, all encoding for asparagine. The *UGT3A2* protein sequence from Humans (NP_777574.2), Chimp (XP_003310798.1), Rhesus Macaque (XP_014995351.1) and Marmoset (XP_002745106.1) encode phenylalanine. Both members of the *UGT3* family in mice encode for phenylalanine at position 391, both sequences in the horse also encode phenylalanine as does the predicted bovine *UGT3A1*.

3.4.10.2.4 Investigating the signature sequence of UGT1A6

The signature sequence of the seven mammalian UGT1A6 sequences analysed was highly conserved (84%). Figure 3.43 compares the signature sequences from *UGT* representatives from plants, silkworm, bacteria, worms, fish, and mammals to that from equine *UGT1A6*. The alignment shows six conserved amino acids to be conserved (highlighted in yellow) between all thirteen sequences, which represent multiple UGT families. In addition, ten residues (highlighted green) have been found to be conserved in ten or more of the species. This may potentially imply that these 16 amino acids are the critical residues which interact in the binding of the donor sugar. The red letters indicate the residues which are conserved between the mammalian sequences and, where present, if those residues are present in the non-mammalian sequences. The signature sequence homology between all the mammalian *UGT* sequences analysed is 96%.

The signature sequence confers the specificity of the donor sugar (Meech et al., 2012b), and this high level of homology between the mammalian *UGT1A6* and equine *UGT1A6* sequences makes it probable that they utilise the same donor sugar.

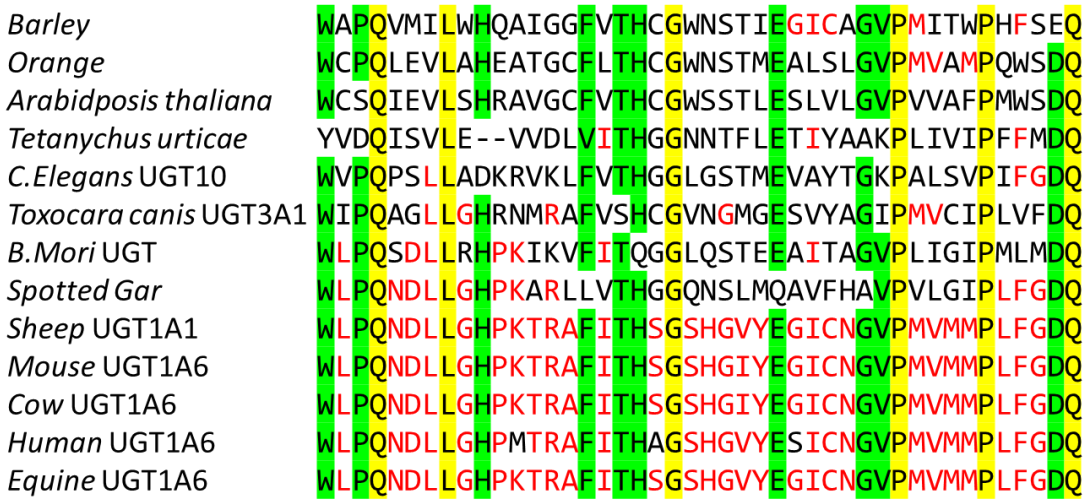


Figure 3.43: Alignment of signature sequences. The signature sequences of barley (ADC92549.1), orange (XP_006469356.1) and *Arabidposis thaliana* (ANM58449.1) represent *UGTs* from the plant kingdom. Tetanchus urticae (AHX56839) is a bacterial *UGT*, C.elegans (NP_504313.2) and Toxocara canis (KHN88569.1) represent small worms. The insects are represented by Bombyx mori – the silkworm (NP_001243978.1) and one fish is present, the spotted Gar (XP_006627360.1). Five mammals include sheep (NP_001192076.1), mouse (NP_659545.2), cow (NP_777187.1), human (NP_001063.2) and horse (ENSECAG00000023519) *UGTs*. The amino acids highlighted yellow are conserved across all sequences, and those highlighted green are conserved in ten or more sequences. Letters which are red show the conserved amino acids between the mammalian sequences.

3.4.10.3 Transmembrane domains

Features characteristic of membrane bound proteins are the presence of a transmembrane domain and a dilysine motif (for details on dilysine motif *see section 3.4.10.4*), which are important for binding the UGT to the ER membrane (Meech et al., 1996). They are composed of hydrophobic amino acids, and this lack of polarity facilitates their anchoring to the membrane (Mackenzie et al., 2008).

In *UGT*s the transmembrane domain comprised of 16 amino acids in the C-termini which is non-polar, and this lack of polarity facilitates interaction with the ER membrane to anchor the enzyme (Mackenzie et al., 2008, Ouzzine et al., 1999). A study investigating transmembrane domains in eukaryotes, including both vertebrates and fungi, found organelle specific properties (Sharpe et al., 2010). They vary in length and residue composition and structurally form an α-helix. Organelle membranes vary in protein and lipid content as well as thickness. For example the plasma membrane is thicker than the ER membrane, and therefore a transmembrane domain for a protein incorporated in to the plasma membrane is longer than for one anchored to the ER membrane (Sharpe et al., 2010).

3.4.10.3.1 *ENSECAG00000008900* – a gene absent of the transmembrane domain.

Analysis of *ENSECAG00000008900* revealed the absence of the transmembrane domain. At 484 amino acids, this polypeptide is shorter than isolated and functionally characterised human and murine *UGT3s*. It is difficult to state without further research whether *ENSECAG00000008900* encodes a fully functional membrane bound UGT. The presence of a stop codon is suggestive of this polypeptide being a truncated version of a *UGT.* Whilst studies have established that truncation of the UGT from the C-terminus can reduce enzymatic activity they have still been found anchored to the membrane (Meech et al., 1996), and it is

plausible that if this is confirmed as a truncated protein, it may still have some function.

Studies using tags are required to see if this enzyme can be embedded into the ER membrane and studies are required to determine if this UGT will be functional. Initial evidence suggests this to be an orthologue of human *UGT3A2*, with a preference for the use of UDPGlc as the donor sugar. Functional studies will establish whether this equine enzyme uses the same donor sugar, and mutational studies are also needed to determine whether residue 391 is singularly responsible for sugar specificity.

3.4.10.3.2 Transmembrane domains of the equine UGT2 members

The regions encoding the transmembrane domains of *ENSECAG00000020628* and *ENSECAG00000014362* show conservation in the first six residues, VIGFLL, all of which are hydrophobic amino acids. A comparison of *UGT*s from humans, rats, mice, and zebrafish found this short six amino acid motif conserved across the *UGT1*s and *UGT2*s of rats, mice, and humans and in the *UGT1A*s of zebrafish. The remainder of the transmembrane domain is composed of mostly hydrophobic amino acids but is more variable in composition. *ENSECAG00000014362* contains several polar threonine residues in the latter part of the transmembrane domain, which may suggest parts of this domain is not in contact with the ER membrane.

3.4.10.3.3 Importance of the transmembrane domain in UGT1A6

Further comparative analysis also confirmed the presence of the putative transmembrane domain conserved across the UGT1A6 sequences of seven mammals (Figure 3.27), which is key to anchoring the mature protein to the ER membrane (Meech et al., 1996).

The first six amino acids appear conserved in all the mammals. A consensus sequence for this region between our sequences, VIGFLLA(V/I)(V/T)L(/T)V(X)FI(V/T), highlights that the residues, with the exception of threonine, are non-polar. This lack of polarity enables the protein to anchor to the membrane via this domain (Mackenzie et al., 2008, Ouzzine et al., 1999).

3.4.10.3.4 The transmembrane domain of equine UGT3A1

A comparison of the *UGT3* transmembrane domains from multiple species found the length of this region to be constant across species with a high level of conservation producing a consensus of VFLLGLTLGT(L/V/M)WLCGK. Within the 16 amino acids there are two threonine residues and one variable position, encoding either leucine, valine, or methionine. Of those sequences investigated, the transmembrane domains of amphibians, the spotted gar, *Toxocara canis,* and European rabbit were sequentially different. The transmembrane domain of ENSECAG00000010396 shared 10 of the 16 amino acids with the consensus sequence formed of the other mammalian transmembrane domains. (Figure 3.44).

Analysis of the transmembrane domains from the non-vertebrates found this region from the moth and the corn earworm is the same length, 16 amino acids, but their composition is valine, leucine, and alanine rich (Ahn et al., 2012). Of the 16 UGTs analysed, valine, alanine, and leucine accounted for 57% of all amino acids. This hypothetically suggests that in addition to their being organelle specific properties of the transmembrane domains (Sharpe et al., 2010), there may be species specific amino acids within this key region of *UGT*s.

Figure 3.44: Transmembrane domain sequence. The consensus sequence for the *UGT3* transmembrane domain, 'L' is a variable amino acid, encoded by either a leucine, valine or methionine. Equine *ENSECAG00000010396* shares 62.5% identity with the consensus sequence.

3.4.10.3.5 Implications for the mutation or absence of the transmembrane domain

In a study that found truncation of the signal peptide made little difference to the ability of the enzyme to insert into the ER membrane (Ouzzine et al., 1999), removal of the c-terminus including the transmembrane domain was found not to prohibit the enzyme from anchoring to the ER membrane (Ouzzine et al., 1999). What this study did not establish is the functional capabilities of the enzyme due to the truncations. A separate study analysed the effect of truncation of the transmembrane domain on enzyme activity and found decreasing activity with increasing size of the truncation (Meech et al., 1996). They also found that appending six or more amino acids to the C-termini reduced enzymatic activity.

Studies investigating the effect of mutations in the cystic fibrosis transmembrane conductance regulator transmembrane domain found that mutations can have a severe phenotypic response (Choi et al., 2005). An insertion of a proline residue affected the ability of the protein to insert into the membrane and had a direct effect on protein conformation and function contributing to the development of cystic fibrosis (Choi et al., 2005). Therefore insertions, deletions, and truncations can markedly affect the transmembrane domain and its function.

3.4.10.4 The dilysine motif

The dilysine motif (**K**X**K**XX) is an important feature of membrane bound proteins, with two lysine residues at positions -3, -4, or -5 from the end of the C-termini (Meech and Mackenzie, 1997a, Levesque et al., 2001, Andersson et al., 1999), and is an additional signal by which the UGT is targeted for membrane retention.

Analysis of this region of *ENSECAG00000020628* found the motif to be **K**E**K**RE. Comparison of the *UGT2B* sequences from rats, mice, and humans showed that two lysines at positions -3 and -5 was the most common form of the motif, with two exceptions in humans, with *UGT2B15* and *UGT2B17* containing three lysines at positions -3, -4, and -5. The motif for *ENSECAG00000014362* was **KKKK**E. Analysis of this motif in *UGT2A* sequences from rats, mice, and humans found this was conserved, with four lysines present in all the sequences.

For the equine sequence *ENSECAG00000010396* the motif is **K**L**KK**A. The lysine at positions -5 and -3 appear to be the more important; a sequence alignment of all the *UGT3* sequences used to create the phylogenetic tree (Figure 3.9) found a lysine present in each sequence at position -5 and -3. Conversely to what was stated by Meech and Mackenzie (1997a), a lysine was found at position -2 in many of the sequences. The lysine at position -2 is more variable and it is not present in either *UGT3A1* or *UGT3A2* sequences for the mouse, nor is it present in the rat or frog sequences. An interesting division was seen in the primates with the *UGT3A1* sequences encoding lysine at -2, but the *UGT3A2* sequences having glutamic acid.

A comparative investigation of the sequences used to create Figure 3.8, which included sequences from the *UGT1*, *UGT2* and *UGT3* families, showed that all mammalian *UGTs* contained a dilysine motif, however sequences from barley and orange did not. No lysine residues were present in the last five amino acids and no homology existed between the last five residues of the barley and orange sequences. An investigation of 88 *UGT* sequences from *Arabidopsis thaliana* found none of these contained a dilsyine motif. It is hypothesised that plant *UGTs* are

cytosolic and therefore do not require the retention signal (Ross et al., 2001). This is supported by investigations of 91 additional *UGT* sequences from *A.thaliana*, which found that a transmembrane motif could not be identified (Li et al., 2001).

Chapter 4: Identification, characterisation, and expression of the equine orthologues of human UDP-Glucuronosyltransferase 1A6 and 3A1

4.1 Introduction

4.1.1 Gene expression profile

4.1.1.1 Tissue expression profile of the UGT1s

The expression of *UGTs* has been extensively studied in humans, with the majority of *UGTs* found to be expressed in the liver (Guillemette, 2003, Fay et al., 2015, Moreton et al., 2014). The primary site of glucuronidation is the liver and this metabolically active organ has high levels of *UGT1A1*, *UGT1A3*, *UGT1A4*, *UGT1A6,* and *UGT1A9* (Finel et al., 2005). However, *UGTs* are also expressed in other tissues, including the brain, uterus, kidney, gastro-intestinal tract (GI), lung, adrenals, and skin (Radominska-Pandya et al., 2005a, Guillemette, 2003). Some *UGTs* are exclusively extra-hepatically expressed. In the human *UGT1* family these include *UGT1A5, 1A7, 1A8,* and *1A10* (Rowland et al., 2013), with *UGT1A8* and *1A10* highly expressed in the GI tract (Stingl et al., 2014). Conversely in rats *UGT1A5* expression is highest in the liver whereas *UGT1A2* and *1A7* are primarily detected extra-hepatically (Shelby et al., 2003); this profile is also seen in mice with the addition that *UGT1A1* is also highly expressed hepatically. In mice *UGT1A6* is highly expressed in multiple tissues, particularly liver and intestines (Buckley and Klaassen, 2007).

4.1.1.2 Tissue expression profile of the UGT3s

Comparative to the *UGT1s*, little is known regarding the expression profile of human *UGT3As. UGT3A2* has been detected at extremely low levels in intestines and liver (Meech et al., 2012b) and shown to be absent in the heart, lung, and brain (Mackenzie et al., 2008), but highly expressed in the testes and thymus (MacKenzie et al., 2011). In mice expression of *UGT3A1* and *UGT3A2* is highest in the kidney, with low levels in the liver; no expression has been reported in the GI tract, lung, heart, or brain.

4.1.2 Gender and age-specific expression of UGTs

Differences in *UGT1A1*, *UGT1A6* and *UGT2B1* expression between male and female rats were investigated in liver and two regions of ocular tissue: lens and extra-lenticular (Nakamura et al., 2005). Expression levels for the three genes were similar in liver tissues, but much higher in the female ocular tissues than the male ocular tissue, with *UGT1A6* expression the highest of the three *UGTs* (Nakamura et al., 2005). A separate study performed in rats looked at the expression of *UGT1A6* in the brain and the olfactory epithelium and bulb from animals from 1 day to 24 months old (Leclerc et al., 2002). Expression levels were variable with age in the olfactory epithelium and brain, with the olfactory bulb displaying a clear increase in expression with age, with a steady increase from 1 day to 12 months and then a sharp increase by 24 months (Leclerc et al., 2002). The same study also investigated the expression of *UGT2A1*, and found expression was absent in the brain, with the olfactory epithelium and bulb both showing increasing levels until 3 months of age, and expression levels decreasing by 12 months (Leclerc et al., 2002).

A study investigating age-specific expression in humans ranging in age from 0-25 years (Neumann et al., 2016) showed that expression of several *UGT1s* (*UGT1A1, UGT1A3, UGT1A4, UGT1A5, UGT1A6*) and two members of the *UGT2* family (*UGT2B7* and *UGT2B15*) increased with age (Neumann et al., 2016). A separate, study looking at 16 paediatric patients (0-24 months) and 12 adults (25-75 years) found that expression of *UGT1A9* and *UGT2B4* increased with age (Vysko et al., 2013).

A comparative gender study performed in mice investigated the expression of eight *UGTs* across 12 tissues. *UGT1A1* was most highly expressed in female livers, as was *UGT1A5*, however *UGT1A2* was not detected in any tissues of either gender except for female kidney samples. *UGT1A10* was also expressed at significantly higher levels in female kidney samples. In the other tissues and isoforms sampled expression levels were similar between male and females (Buckley and Klaassen, 2007).

A more limited study in humans investigated the gender difference of a singular isoforms, *UGT2B17,* in 103 liver samples, and found males had expression levels four times higher than females (Gallagher et al., 2010). These studies indicate there may be disparities in expression between genders and with increasing age and while the underlying causalities behind this remain to be fully elucidated, it is entirely plausible that such differences will be seen in other species.

4.1.3 Clinical impact of genetic polymorphisms

The *UGTs* are highly polymorphic enzymes (Stingl et al., 2014). Owing to the complex nature of the *UGT1s* locus a polymorphism present in exon 1 would affect a specific isozyme, whereas mutations in exons 2-5 may potentially affect all UGT1 enzymes synthesised (de Wildt et al., 1999). An example of a clinical manifestation of mutations in the UGT1 family is Crigler Najjar Syndrome types I and II, both of which impede the ability to metabolise bilirubin (Jancova et al., 2010, de Wildt et al., 1999, Meech and Mackenzie, 1997a). Major risk factors for the development of various cancers are hormones whilst dietary carcinogens have been linked to the ontogenesis of colorectal cancers. Both dietary components and hormones are metabolised to an excretable form by UGTs (Hu et al., 2016). Multiple polymorphisms in the promoter and enhancer regions as well as the within the coding region of the gene itself have been implicated in the development of cancers. For example, changes in the TATA box in terms of the number of TA repeats has been shown to result in decreased promoter activity for UGT1A1 and has been suggested to affect the susceptibility of breast, endometrial, and colorectal cancers (Hu et al., 2016). Human UGT1A6 in particular has been linked to breast and lung cancer (Hu et al., 2016). Genetic polymorphisms may lead to reduced enzyme functionality or conversely to bio-activation, as has been seen with morphine 6-O glucuronides which have increased analgesic effect (Stingl et al., 2014, de Wildt et al., 1999, Ritter, 2000). Disease, ethnicity, age, and environmental influences can

also affect the metabolic capabilities of the UGTs (Jancova et al., 2010, Meech and Mackenzie, 1997a), thus altering the body's response to xenobiotics.

The use of hepatic metabolic *in vitro* tools for the prediction of human pharmacokinetics has become a stalwart in the discovery of new drugs. With research into UGTs predominately occurring in man and rodents, research of veterinary animals is slowly catching up to improve disease, health, and welfare of animals. *In vitro* tools are slowly becoming available for animals such as the dog and cynomolgus monkeys (Soars et al., 2001, Troberg et al., 2015, Hanioka et al., 2006), with little progress made in applying these tools to the horse. Access to such tools would improve disease, health, and welfare through a mechanistic understanding of how foreign chemicals are metabolised and enable improved understanding of the metabolism of regulated and illegal drugs. In this study, the expression levels of equine orthologues of human *UGT1A6* and *UGT3A1* have been investigated in four tissues as well as a pilot study investigating the effect of age on expression. Additionally, the genes have been examined for the presence of polymorphisms.

4.2 Methods and Materials

4.2.1 Summary of horse details

Table 4.1 summarises the age, gender and tissues sampled for each of the horses used for the quantification of *UGT1A6* and *UGT3A1*. For five animals, brain, liver, lung, and kidney tissue was acquired. Liver tissue was collected from an additional seven horses. Where possible, the sex of the animal was noted.

Animal	Age	Sex	Tissue
1	1	-	Liver
2	5	-	Liver
3	5	-	Liver
4	6	Mare	Brain, Liver, Lung, Kidney
5	6	Gelding	Brain, Liver, Lung, Kidney
6	7	Gelding	Liver
7	11	Mare	Brain, Liver, Lung, Kidney
8	13	Mare	Liver
9	17	-	Liver
10	19	-	Liver
11	19	Mare	Brain, Liver, Lung, Kidney
12	23	Mare	Brain, Liver, Lung, Kidney

Table 4.1: Summary of age, sex and tissue. The age and tissues sampled from 12 animals, and where possible the sex was also recorded. For five animals, tissue was collected from brain, liver, lung, and kidney. A biopsy of liver was collected from an additional seven horses.

4.2.2 Sequencing to identify polymorphisms

From each of the 12 horses used to generate the gene expression data the *UGT1A6* and *UGT3A1* genes were isolated by PCR (see sections 2.2.5, 2.2.6, and 2.2.7) and cloned (as described in sections 2.2.12 and 2.2.13). A single clone from each animal was PCR'd and prepared for sequencing (see sections 2.2.14, 2.2.15, and 2.2.16) sent to be sequenced at the Zoology Department at Oxford University. Sequences were aligned to the reference sequence to determine the presence of any novel polymorphisms (as described in section 2.2.17).

4.3 Results

4.3.1 Selection of reference genes

Reference genes are required to account for variability in template concentrations in quantitative PCR analyses. Based on published works looking at expression in equine skin samples (Bogaert et al., 2006), four reference genes were selected for testing in each of the tissues to be analysed for *UGT1A6* expression.

Hypoxanthine-guanine phosphor-ribosyltransferase (*HPRT1*), Beta-2-microglobulin (*B2M*), 60S Ribosomal protein L32 (*RPL32*) and *β-actin* sequences were retrieved from the NCBI database, and primers and probes were designed, table 4.2.

Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')	Probe (5'-3')
β-actin	CGGGCGACGACGCTC	CATGGGCCAGAAGGACTCA	CAGGGCGTGATGGT
B2M	CCCTGATAGTTAAGTGGGATCGA	CTTGAAGATTCCTCATTTGGATTTG	CTCTAACCAGCATCATGAA
HPRT1	ATGACCAGTCAACAGGGGACA	TTGCTTTCCTTGGTCAAGGAG	GTGATTGGTGGAGATGA
RPL32	TGTGCAACAAATCGTACTGTGC	CATCGTGGAGAAAGCAGCC	GAGATTGCTCACAACGTCTCCTCCAAGAAC
ENSECAG00000023519	CTCCTTGGATAGTGGTTTTTTGCTCACAC	CTCCTTGGATAGTGGTTTTTTGCT	CCACAATTCCATGTTCTCCAGAAGCATTGA
ENSECAG00000010396	GGACTTCAGTGCAGTCATTTGC	GGACTTCAGTGCAGTCATTTGC	AAGCTTTGGCTGGCAGAGACAAATTTGA

Table 4.2: Primers and probes for quantitative PCR. Primers and probes for use in quantification of transcripts for the four reference genes, *β-actin* (NM_001081838.1), Beta-2-microglobulin (*B2M* – NM_001082502.3), Hypoxanthine phosphoribosyltransferase 1 (*HPRT1* - AY372182.1) and 60S Ribosomal protein (*RPL32* – XM_001501497.4) and the two equine genes ENSECAG00000023519 (predicted *UGT1A6*) and ENSECAG00000010396 (predicted *UGT3A1-like*). One of the primers per gene was designed to sit across an intron-exon boundary and probes were labelled with 5'-FAM reporter dye and 3'-TAMRA quencher.

The results showed *HPRT1* to be the most variable in expression across the tissues. The cycle threshold value (Ct value) of *HPRT1* ranged from 22 to 31, with a mean Ct for liver, kidney, and brain of 25, however for the lung the mean Ct was 27 suggesting *HPRT1* was expressed at a lower level in the lung than the other three tissues (Figure 4.3). *β-actin* and *B2M* showed a similar expression profile across the four tissues, with the average Ct value for both reference genes ranging from 21.6 to 23.1, although *β-actin* displays a greater expression range in brain tissue with Ct values ranging from 19.3 to 28. *RPL32* resulted in the most consistent expression across all tissues with the least amount of variability. In each of the tissues the range of Ct values remained within the 20.5 to 23.3 range, with the mean Ct value across the four tissues ranging from 21.3 to 22. Based on these results, *β-actin, B2M* and *RPL32* displayed the most similar and consistent expression levels and were selected as reference genes for quantitative PCR of *UGT1A6* expression.

To confirm our results the data was analysed using Ref Finder four algorithms, GeNorm, NormFinder, Delta C_T, and BestKeeper. A comprehensive analysis then assessed the results of the four algorithms and ranked the reference genes according to their stability (Table 4.4 and Figure 4.5). Across each of the tissues, for each of the algorithms, *HPRT1* was found to be the least stably expressed, with *RPL32* consistently the most stably expressed of the four genes.

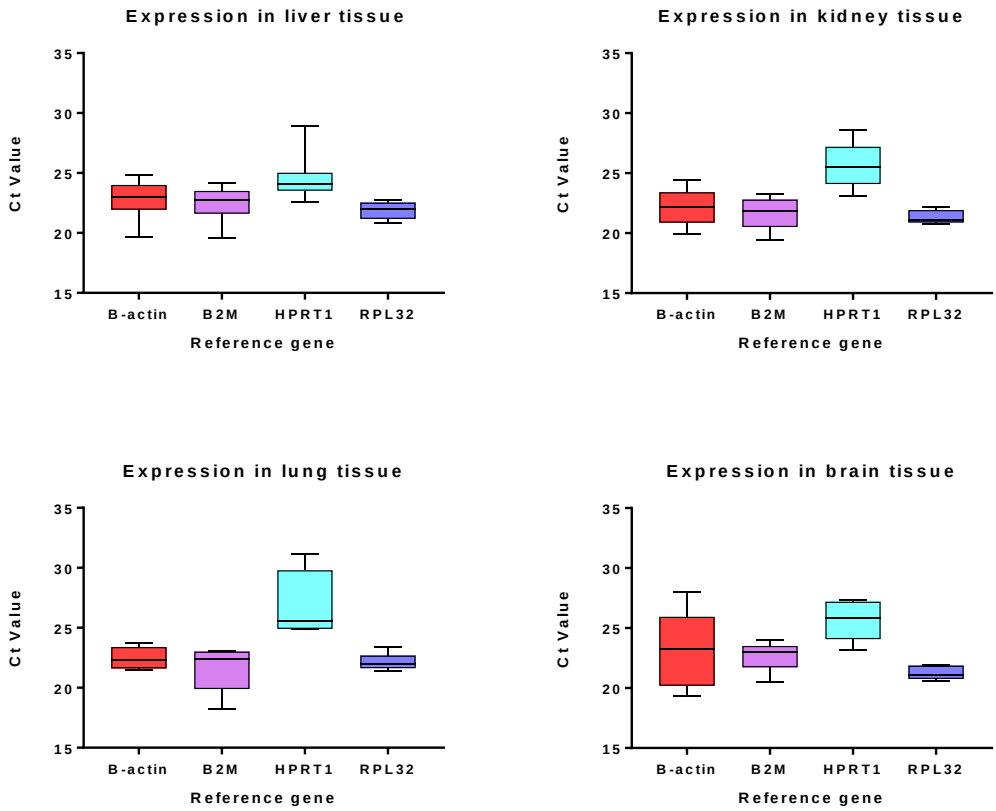


Figure 4.3: Box and whisker plot showing cycle threshold (C_T) variation for each of the reference genes. Reference genes, *β-actin*, Beta-2-microglobulin (*B2M*), Hypoxanthine-guanine phosphor-ribosyltransferase (*HPRT1*) and Ribsomal Protein L32 (*RPL32*), were tested for in each tissue; liver (n=12), lung, brain and kidney (n=5). The median and quartiles were calculated. Whiskers show minimum and maximum C_T values. The box and whisker plots show *HPRT1* to have the highest C_T values in all the tissues and also the greatest variability in expression. *B2M*, *RPL32* and *β-actin* all dislayed similar levels of expression and mean C_T values across the four tissues, making them the optimal choice of reference genes for normalising gene expression data.

BestKeeper		**Delta CT**		**geNorm**		**NormFinder**		**Comprehensive Analysis**	
Gene	Stability value	Gene	Stability value	Gene	Stability value	Gene	Stability value	Gene	Stability value
B-actin	1.35	*B-actin*	2.31	*B-actin*	1.99	*B-actin*	1.73	*B-actin*	2.71
B2M	1.16	*B2M*	2.28	*B2M*	1.38	*B2M*	1.83	*B2M*	1.86
RPL32	0.56	*RPL32*	1.81	*RPL32*	1.38	*RPL32*	0.69	*RPL32*	1.00
HPRT1	1.56	*HPRT1*	4.41	*HPRT1*	2.20	*HPRT1*	1.96	*HPRT1*	4.00

Table 4.4: Gene expression stability analysis using RefFinder software. Four reference genes were analysed in liver, lung, brain and kidney for expression stability and suitability for normalising data (n=27). RefFinder software ranked gene expression stability, comparing four computational algorithms and producing a comprehensive analysis. *HPRT1* was the least stable gene in each analysis.

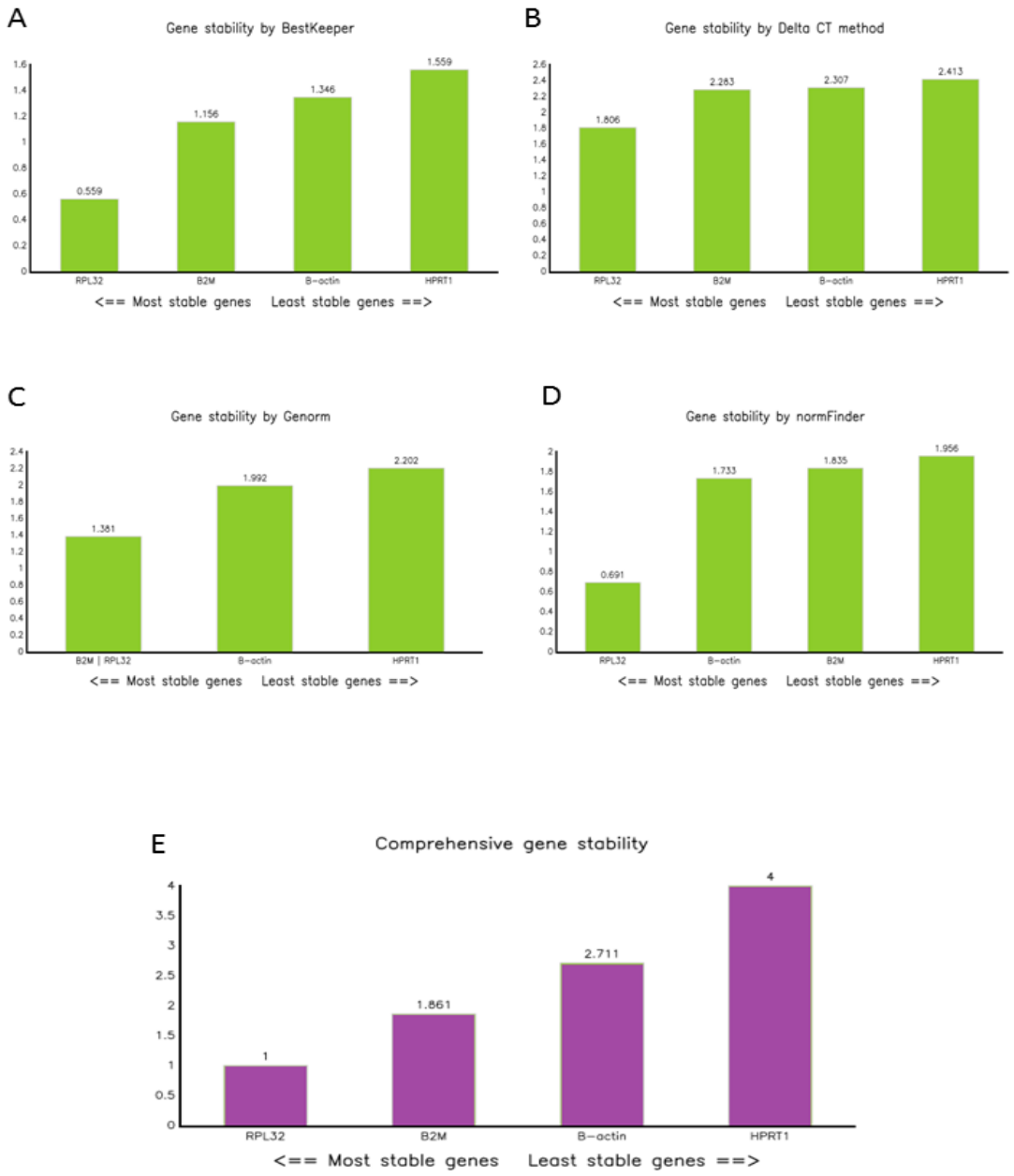


Figure 4.5: Gene Expression stability analysis using RefFinder software. The expression of *B-actin, B2M, RPL32 and HPRT1* were analysed using RefFinder to determine the reference genes which were stably expressed across liver, lung, kidney, and brain. The graphs show the four different algorithms used in the analysis. Graph F displays the comprehensive analysis, which incorporates all the results of all the algorithms and ranks genes in order of stability. *HPRT1* was found to be the least stable gene.

4.3.2 Optimisation of reference gene primer concentrations for qPCR

The primer concentratiosn for amplification of the reference genes were optimised by testing at nine different concentrations ranging from 50nM to 900nM (see section *2.2.9.2*, Table 2.1). A fixed probe concentration of 100nM was used in each reaction. The primer pair which yielded the best amplification (greatest fluorescence) with the lowest C_T value was selected for use in subsequent testing of the four tissues for transcript expression (Table 4.6 and Figure 4.7).

	Forward Primer	**Reverse Primer**
B2M	900nM	900nM
B-actin	50nM	900nM
RPL3	900nM	300nM

Table 4.6: Concentration of forward and reverse primers of reference genes. Forward and reverse primer concentrations for each of the reference genes was empirically established.

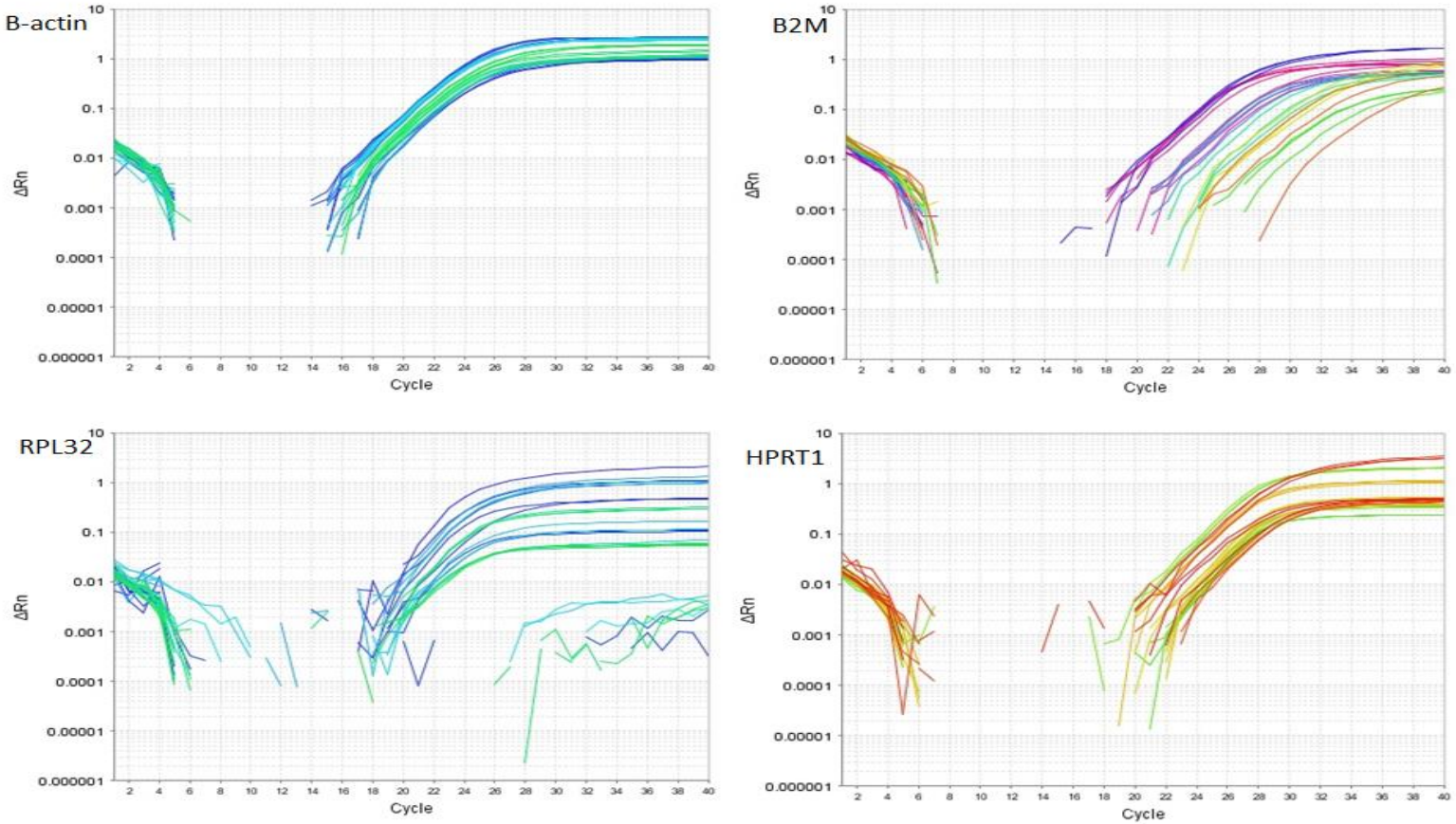


Figure 4.7: Amplification plots of reference genes. Primer titres were performed to determine the optimal concentrations for forward and reverse primers for the qPCR assays. Amplification plots show all combinations of concentrations used in the titre for *β-actin*, *B2M*, *RPL32* and *HPRT1*.

4.3.3 Optimisation of UGT1A6 primers for qPCR

A primer titration was also performed for the qPCR primers for equine *UGT1A6*, to determine the optimal concentration for each primer. With a fixed probe concentration of 100nM, the primer concentrations which yielded the greatest fluorescence (best amplification) and the lowest C_T value were investigated. The optimal forward and reverse primer concentrations were 50nM and 900nM respectively which yielded a C_T of 26.3.

4.3.4 Efficiency of the UGT3A1 qPCR primers

Development of a qPCR assay for equine *UGT3A1* started with the determination of the efficiency of the primers (Figure 4.8). Testing a serial dilution of plasmid containing the cloned gene yielded a R^2 value of 0.9997.

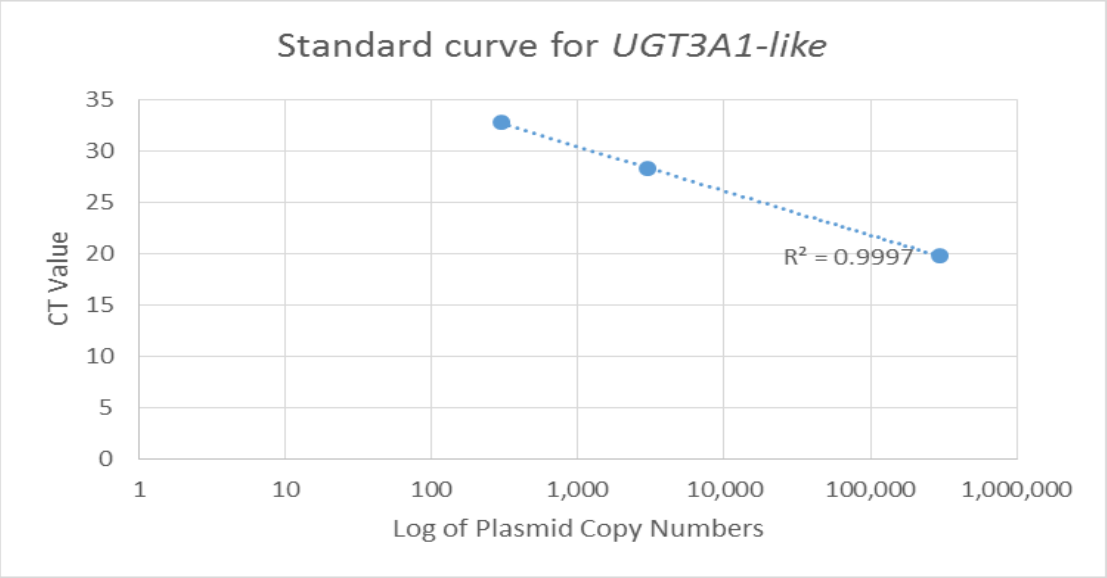


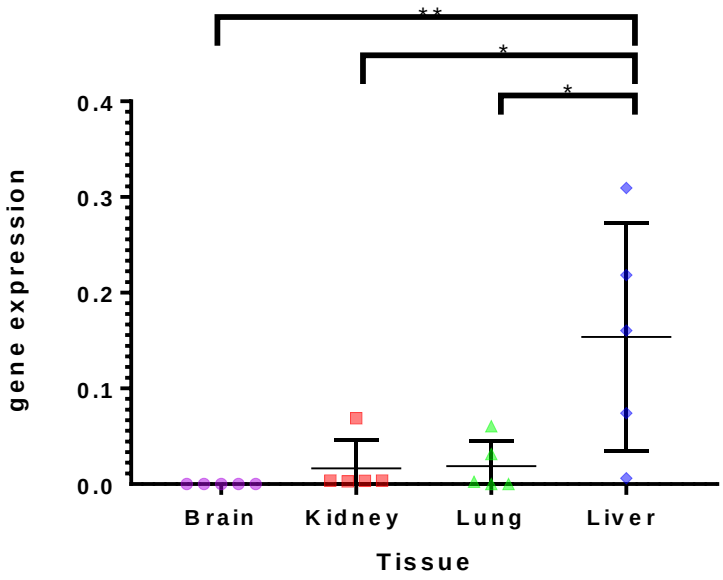
Figure 4.8: Standard curve of *UGT3A1*. Standard curve for *UGT3A1-like* was produced from a serial dilution of cloned gene against cycle threshold (Ct) values. This shows a reverse linear correlation with coefficient of 0.9997. Each point represent a mean of measurements performed in triplicate.

4.3.5 Expression of UGT1A6 in equine liver, brain, kidney and lung

Total mRNA from brain, kidney, lung, and liver tissue (n=5) were analysed for expression of equine *UGT1A6* transcripts (Figure 4.9). Expression levels of *UGT1A6* in the brain tissue were negligible in all five samples. Relative gene expression levels in the kidney and lung were similar across both tissues. *UGT1A6* expression levels were highest in the liver and showed most intra-tissue variation.

The differences in relative expression between kidney and liver and lung and liver are statistically significant ($p<0.05$). The differential expression between brain and liver was highly significant ($p<0.01$).

A



B

Animal	Brain	Kidney	Lung	Liver
1	0.00002	0.00389	0.00002	0.21842
2	0.00000	0.00380	0.00004	0.00602
3	0.00009	0.06883	0.03143	0.16041
4	0.00000	0.00272	0.00257	0.07431
5	0.00000	0.00334	0.06046	0.30942

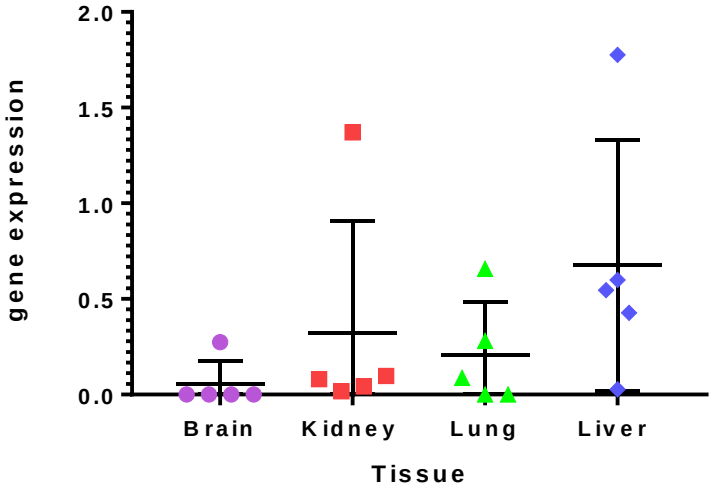
Figure 4.9: Comparison of gene expression of *UGT1A6* in equine tissues. Figure 4.9a displays a graphical representation of the quantitative real-time PCR was performed using reverse transcribed total mRNA from brain, kidney, lung, and liver tissues (n=5). Table 4.9b summarizes the gene expression values for each animal. Expression levels were normalized against three reference genes; Beta-2-Microglobulin (*B2M* – NM_001082502.3), *β-actin* (NM_001081838.1) and 60S Ribosomal protein L32 (*RPL32* – XM_001501497.4). Results; mean ± SD. Significant * (P<0.05), highly significant ** (P<0.01). Data analysed by One Way Analysis of Variance (ANOVA) using Graphpad Prism 7.01. *UGT1A6* levels in the brain were negligible in the five animals. Levels of expression in the kidney and lung were low, with little within tissue variability of expression. The liver samples displayed both the highest and most variable levels of expression.

4.3.6 UGT3A1-like expression in equine liver, brain, kidney and lung

Liver, brain, kidney, and lung tissue (n=5) were analysed for expression of equine *UGT3A1-like* (Figure 4.10), with data normalized against three reference genes, *B2M, RPL32,* and *β-actin.* Relative expression levels were extremely low in brain tissue, with expression in three of the animals below the level of detection of the assay, and in one animal the relative expression levels were so low as to be considered negligible.

Analysis of *UGT3A1-like* expression in the kidney identified a single horse (horse #12) with high transcript levels. Analysis of the kidney from the remaining four animals found expression levels of *UGT3A1-like* at least 15 times lower than this horse. In the lung tissues, expression levels were more variable with three animals yielding minimal expression. Expression was higher in the liver compared to brain, kidney, and lung, although there was considerable variation in expression levels within tissues. Whilst one animal yielded minimal expression of *UGT3A1-like* in the liver, three animals displayed similar levels, which were 25 times that of the animal with the least expression. A single animal yielded relatively high expression levels in the liver - 88 times the expression level detected in the animal with the least expression. The differences in expression of *UGT3A1-like* between the four tissues were not statistically significant ($p>0.05$).

A



B

Animal	Brain	Kidney	Lung	Liver
1	0.00147	0.09786	0.00042	1.77581
2	0.00000	0.42550	0.00057	0.02724
3	0.27342	1.37073	0.65832	0.59879
4	0.00000	0.08110	0.08781	0.42732
5	0.00000	0.01731	0.28121	0.54606

Figure 4.10: **Comparison of gene expression of *UGT3A1-like* in equine tissues.** Figure 4.10a graphically displays the quantitative real-time PCR was performed using reverse transcribed mRNA from brain, kidney, lung, and liver tissues (n=5). Table 4.10b summarises the gene expression values in all tissues for each animal. Expression levels were normalized against three reference genes; Beta-2-Microglobulin (*B2M* - NM_001082502.3), *β-actin* (NM_001081838.1) and 60S Ribosomal protein L32 (*RPL32* - XM_001501497.4). Results, mean ± SD, were not significant (P>0.05), data analysed by ANOVA using GraphPad Prism 7.01. Expression in the brain was minimal, four animals displayed minimal expression in the kidney, with a single animal yielding relatively high levels of UGT3A1-like. Three animals show minimal expression in the lung. Liver tissue Liver showed the greatest variability in expressions, with the mean expression highest of all the tissues tested.

4.3.11 Identification of single nucleotide polymorphisms in UGT1A6

The coding sequence for the equine *UGT1A6* gene was isolated and bidirectionally sequenced from the liver of nine horses. The analysis of sequence data identified four polymorphisms. At position 124 the nucleotide was either homozygous guanine (20%) or homozygous adenine (80%), which would alter the amino acid from aspartic acid to asparagine. A homozygous silent mutation was present in two animals at nucleotide position 748 with the presence of either thymine or cytosine. A polymorphism at nucleotide 1235 was homozygous altering the codon from CAT to CGT, resulting in either a histidine or arginine at amino acid 411; this was present in a third of the horses sampled. The fourth polymorphism identified was at position 1559 altering the amino acid from proline (CGG) or glutamine (CAG) in 44% of the animals.

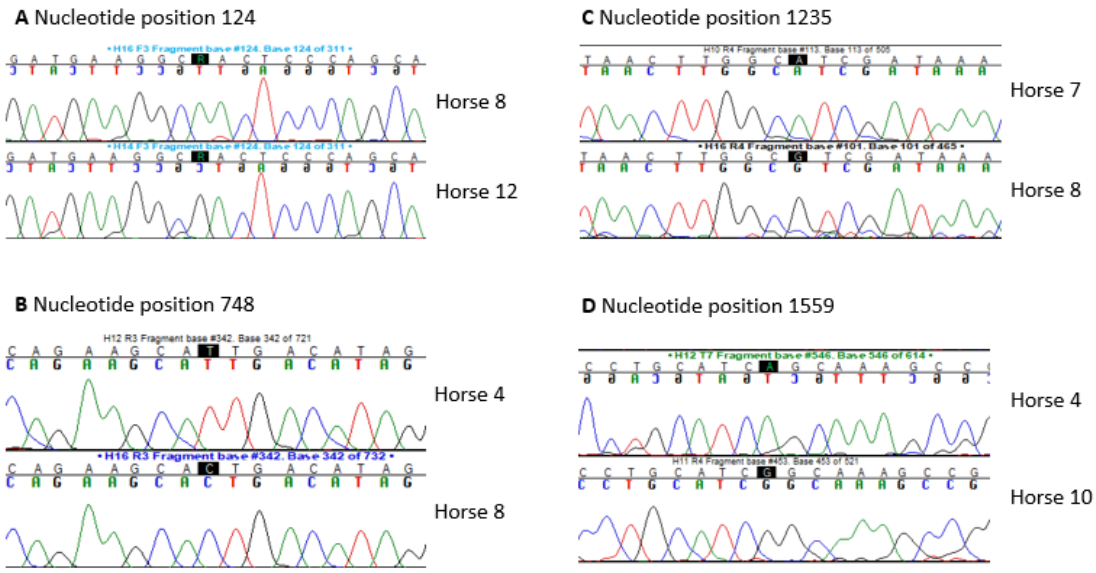


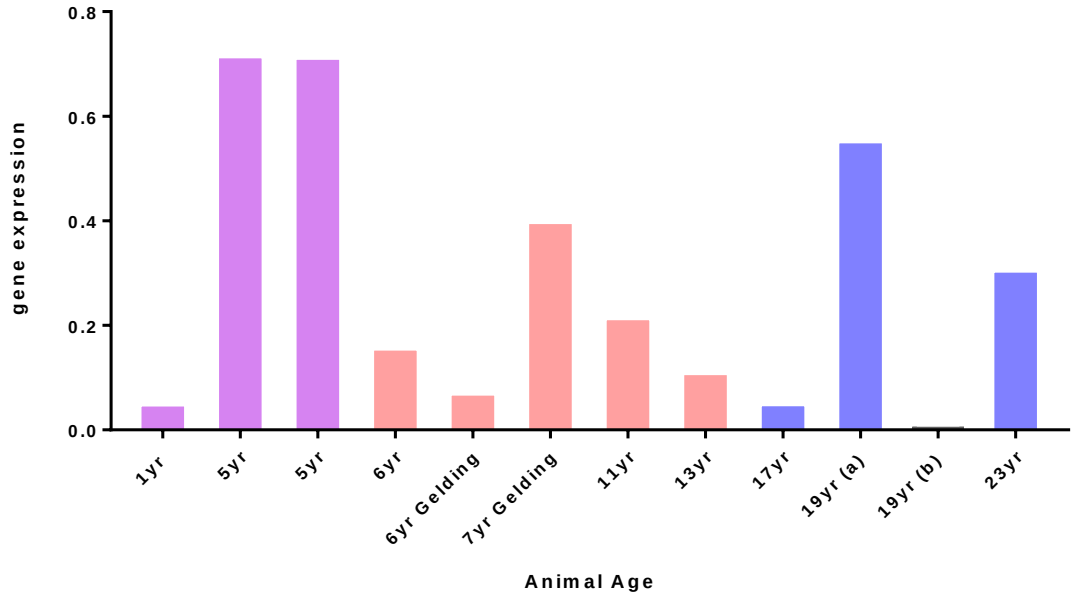
Figure 4.15: Representative sequence traces of SNPs in equine *UGT1A6*. Sequence traces representing the different nucleotides encoded in the transcripts of equine UGT1A6 at positions 124, 748, 1235, 1559. The nucleotide of each SNP is highlighted with a black background.

4.3.7 Age specific expression levels of UGT1A6

A preliminary investigation looking into the effect of age on the expression of *UGT1A6* isolated from 12 animals ranging in age from 1-23yrs yielded no clear trend (Figure 4.11), with only three ages present in duplicate; 5yr, 6yr, and 19yr old animals. The highest levels of *UGT1A6* expression were detected in the two 5yr old animals, with the lowest in one of the 19yr old animals.

Expression in the 1yr old was low, with expression seeming to increase with age as both 5yr old animals showed expression 14-fold greater than that of the 1yr old. Expression of *UGT1A6* was lower in both 6yr olds, falling by over 77%, with the mare showing higher expression levels than the gelding. *UGT1A6* expression increases in the 7yr old, however lower levels were detected in the animals aged 11, 13, and 17yr, with the 17yr old showing expression levels 8-fold lower than the 7yr old. The two 19yr old animals displayed very different levels of expression, with animal 19(a) showing an increase in expression of 91% compared to the 17yr old, and animal 19(b) showing negligible expression of *UGT1A6*. Expression of *UGT1A6* in the 23yr old animal was 50% less than that of animal 19(a).

A



B

1yr	5yr	5yr	6yr	6yr Gelding	7yr Gelding
0.05369	0.71943	0.71653	0.16041	0.07431	0.40240

11yr	13yr	17yr	19yr (a)	19yr (b)	23yr
0.21842	0.11368	0.05392	0.55671	0.00602	0.30942

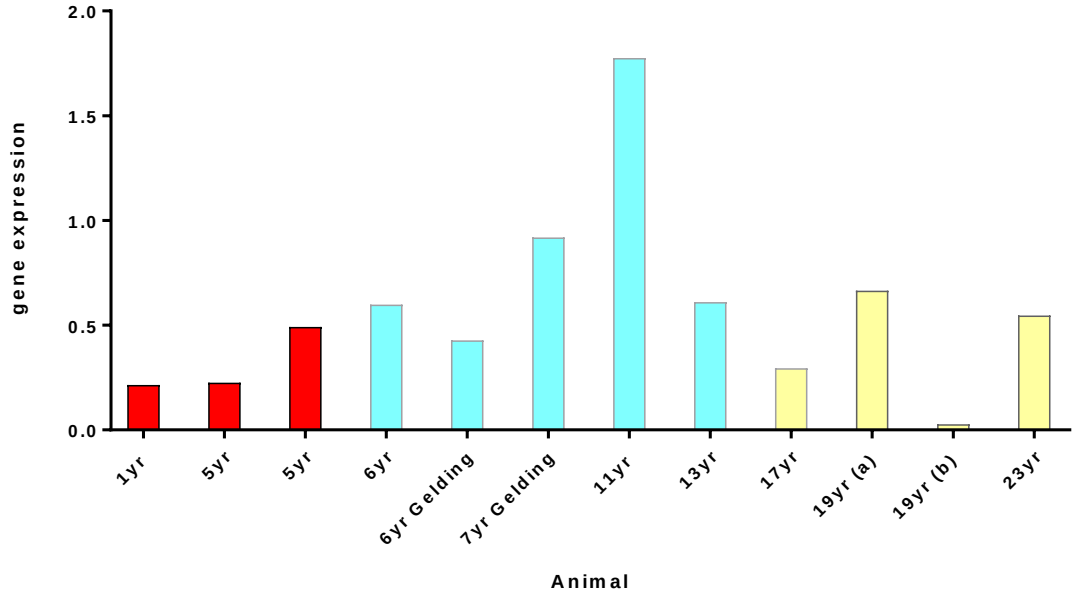
Figure 4.11: Comparison of gene expression of *UGT1A6* in 12 equine liver samples. Figure 4.11a graphically displays the quantitative real-time PCR was performed using reverse transcribed total mRNA from 12 liver samples, from horses ranging in age from 1yr to 23yrs (n=1 except 5yrs, n=2, 6yrs n=2 and 19yrs n=2). Table 4.11b summarises the gene expression for each horse liver analysed. Expression levels were normalized against three reference genes; Beta-2-Microglobulin (*B2M* – NM_001082502.3), *β-actin* (NM_001081838.1) and 60S Ribosomal protein L32 (*RPL32* – XM_001501497.4). Results display relative expression levels using Prism 7.01.

4.3.8 Age specific expression of UGT3A1

A preliminary study of the expression levels in 12 liver samples was performed to investigate *UGT3A1-like* expression in animals of different ages, ranging from 1 to 23 years (figure 4.12), n=1 except for 5yrs, 6yrs, and 19yrs (n=2)

Quantitative data was normalised against three reference genes *B2M, RPL32,* and *β-actin.* The data showed a trend of increasing expression up to the age of 11yrs, and then a subsequent decrease with age. The two 5yr animals displayed different levels of expression of *UGT3A1-like*, with the second animal expressing double the number of transcripts. Of the two 6yr old animals analysed, a mare and a gelding, the mare displayed higher expression levels than that of the gelding. The two 19yr old animals displayed large differences in expression levels, with one expressing *UGT3A1-like* at levels 33 times greater.

A



B

1yr	5yr	5yr	6yr	6yr Gelding	7yr Gelding
0.21428	0.22553	0.49169	0.59879	0.42732	0.91950

11yr	13yr	17yr	19yr (a)	19yr (b)	23yr
1.77581	0.61084	0.29409	0.66493	0.02724	0.54606

Figure 4.12: Comparison of gene expression of *UGT3A1-like* in 12 animals ranging in age from 1-23yrs. Liver samples were analysed from 12 animals ranging in age from 1yr to 23 yrs, with all ages n=1 except 5yrs (n=2), 6yrs (n=2) and 19yrs (n=2). Figure 4012a graphically displays the expression of *UGT3A1-like* was normalised against three reference genes, *B2M*, *RPL32* and *β-actin*. *UGT3A1-like* expression was found to increase up to 11yrs before decreasing with age. Table 4.12b summarises the relative gene expression from the liver of each horse.

4.3.9 Demographic analysis of age specific expression of UGT1A6

To determine if there was a demographic effect of age on expression the animals were grouped into three age ranges: young (1-5yrs, n=3), medium (6-15yrs, n=5), and older animals (16+yrs, n=4). Within each age range the expression levels are variable (Figure 4.13). Using the mean gene expression from each group suggested that the young animals have the highest levels of expression. The levels of *UGT1A6* decreased in the medium age group of horses, with the older animals showing a mean increase in transcript levels for *UGT1A6* (Figure 4.13). These differences were not statistically significant.

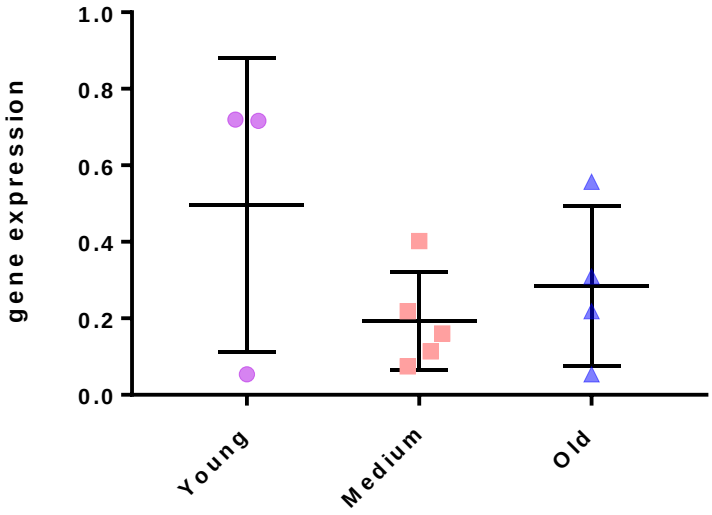


Figure 4.13: Comparison of gene expression of *UGT1A6* in animals grouped by age. The results of the quantitative real-time PCR were divided into age groups; young (1-5yrs, n=3), medium (6-15yrs, n=5) and old (16yrs+, n=4). Results; mean ±SD. Data analysed by One Way Analysis of Variance (ANOVA) using Graphpad Prism 7.01, results were not statistically significant ($p>0.05$). The young animals had a higher mean gene expression, with the medium age groups of animals showing the lowest level of expression. Expression then rose slightly in the older animals.

4.3.10 Demographic analysis of age specific expression of UGT3A1

Due to the low number of animals sampled, they were grouped into three age groups to establish whether the general demographic pattern was statistically significant. The age grouping selected was young (1-5yrs, n=3), medium (6-15yrs, n=5), and older animals (16yrs+, n=4). The mean gene expression of *UGT3A1* in the livers from the young animals is the lowest of the three groups. For the medium age group the mean relative expression of *UGT3A1-like* was 2.7 times higher than the mean expression of *UGT3A1-like* of the young group. The average gene expression of the older animals dropped by 56% relative to the mean gene expression of *UGT3A1-like* in the medium age group. The observed differences in expression of *UGT3A1-like* between the different age groups were not statistically significant (Figure 4.14).

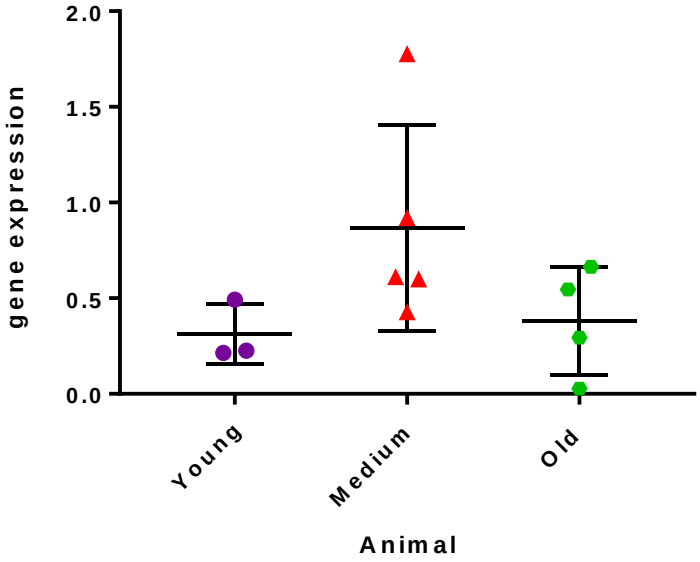


Figure 4.14: Comparison of gene expression of *UGT3A1-like* in animals grouped by age. Animals were grouped into three categories based on age; young (1-5yrs, n=3), medium (6-15yrs, n=5) and old (16yrs+, n=4). The medium aged animals displayed the greatest range of expression and the highest average expression level of *UGT3A1-like*. A One-Way ANOVA was performed using Graphpad Prism 7.01. The differences in expression levels between the three groups were not statistically significant, $p>0.05$.

4.3.12 Identification of single nucleotide polymorphisms in UGT3A1

The full gene sequence was obtained from liver biopsies of nine horses. The nucleotide sequences were interrogated for genetic variation with one SNP identified in our data. At position 193 the nucleotide sequence was either a cytosine or guanine, with 56% of sequences homozygous G/G, 33% homozygous C/C and a single animal heterozygous (G/C) at this position. This created a nucleotide change in the first base of the codon, changing CCA to GCA, altering the amino acid from proline to alanine, both of which are non-polar (Figure 4.16).

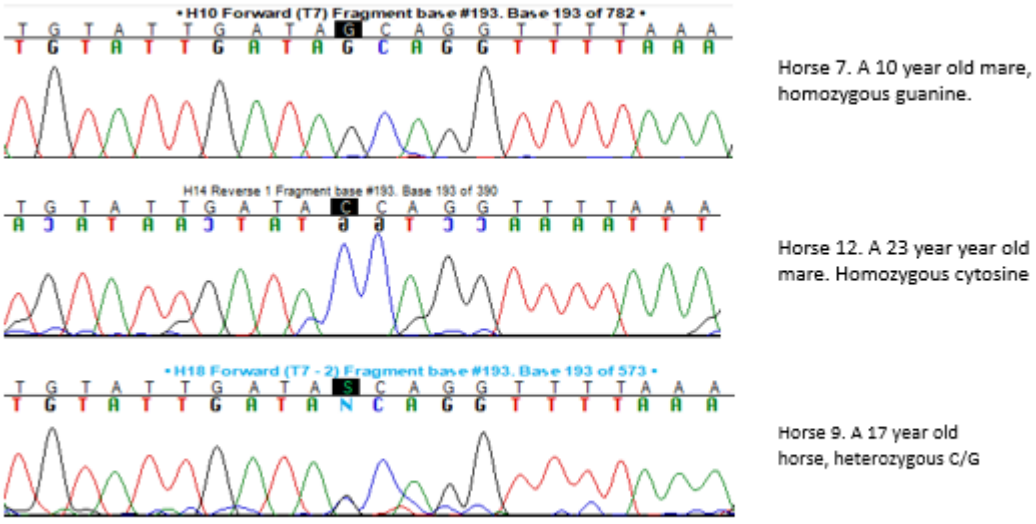


Figure 4.16. Sequencing traces from three horses. The figure displays a single sequence trace from three horses. Horse 7 was identified as homozygous guanine at position 193. Sequencing of a clone of *UGT3A1* from horse 12 found the position to be homozygous cytosine. Horse 9 was identified as heterozygous C/G at position 193. Position 193 is highlighted by the black background behind the called nucleotide.

4.4 Discussion

One of the major pathways of phase 2 metabolism involves glucuronide formation, which is mediated by the UGT family of enzymes (Jancova et al., 2010). This method of biotransformation provides a pathway to eliminating exogenous and endogenous compounds (Gibson and Skett, 2001, Tukey and Strassburg, 2000). Mediated by the UGT family of enzymes this presents a method of biotransformation and elimination of potentially toxic xenobiotic and endogenous compounds (Izukawa et al., 2009, Tukey and Strassburg, 2000, Gibson and Skett, 2001). UGTs are bound to the ER and catalyse the addition of a polar glucuronide moiety from a donor sugar to a substrate facilitating elimination of a compound from the circulatory system (Meech and Mackenzie, 2010). Developing our understanding of metabolism in horses will contribute to improving health and welfare and better disease treatment. In this chapter, we report on the isolation and characterisation of an equine *UGT*, the orthologue of human *UGT1A6* and *UGT3A1*.

4.4.1 Determination of suitable reference genes and optimisation of primer concentrations

Quantitative real-time PCR is the most sensitive method for the detection of mRNA transcripts (Bustin et al., 2009, Kozera and Rapacz, 2013, Guénin et al., 2009). Establishing suitable internal controls is necessary to account for variability in template concentrations, cDNA synthesis, and pipetting errors. It is important to identify reference genes, often referred to as 'housekeeping' genes, which display stable expression across the tissues to be analysed. Using multiple reference genes minimises variation between biological replicates and between tissues (Kozera and Rapacz, 2013, Bustin et al., 2009). Typically, 'housekeeping' genes are selected as they are endogenously expressed, however stable expression across tissues needed to be empirically determined as they can vary among tissues or cells and may change under certain circumstances. Studies with *GAPDH*, for example, have proven expression can be affected by physiological variables, such as inflammation, oxidative stress, and hypoxia (Cummings et al., 2014), and/or expressed at very low

levels in some tissues such as human umbilical vein endothelial cells (Chen et al., 2013). Thus the selection of reference genes is critical for gene expression studies.

Four reference genes were selected based on previously published work (Bogaert et al., 2006); Beta-2-Microglobulin (*B2M*), Ribosomal Protein L32 (*RPL32*), *β-actin* and hypoxanthine phosphoribosyltransferase 1 (*HPRT1*). As these genes have only been established as suitable for equine skin, we validated their suitability for gene expression studies in brain, liver, lung, and kidney. For our tissue selection, analysis by comparing C_T values found *HPRT1* to have variable expression across the tissues and a higher cycle threshold than *B2M, RPL32,* and *β-actin*. It was disregarded from further investigations. The cycle threshold values for *B2M*, *RPL32* and *β-actin* were similar across the tissues with stable expression profiles, and therefore were selected for normalization of qPCR data. Algorithmic assessment of the reference genes found *HPRT1* to be the least stably expressed across the tissues, whereas *RPL32* was consistently the most stable in expression of the four genes. Using this information, we excluded *HRPT1* from downstream assays, and *RPL32*, *B2M* and *β-actin* were used to normalise expression data. It is important to assess the stability of reference genes, due to the influence from underlying conditions, such as those mentioned above. It is also important when the breed of the animals sampled are unknown. A study looking at the expression levels of nine reference genes in two breeds of horse, thoroughbred and the Jeju pony, found expression of *18S Ribosomal RNA*, *Ubiquitin B* and *Succinate dehydrogenase complex subunit A* different between the two breeds, making them unsuitable for the normalization of expression data (Ahn et al., 2011).

4.4.2 Expression profile of UGT1A6 in brain, kidney, lung and liver

Gene expression levels of equine *UGT1A6* were investigated in brain, kidney, lung, and liver tissue. All animals sampled were processed for consumption and would therefore have undergone the minimum withdrawal period from drugs prior to transport to the abattoir. Any differences seen in expression levels would therefore not have resulted from reflect prior medication.

UGT1A6 expression was greatest in the liver which correlates to the organ where the majority of drug metabolism takes place (Izukawa et al., 2009). However, this contradicts the data provided by Dr. Emes which indicated the kidney expressed higher levels of *UGT1A6* than the liver. Expression values were not assessed for lung and brain tissue in this previous study.

We found similar levels of *UGT1A6* expression in the kidney and lung tissues which were lower than expression levels in the liver reflecting the expression profile of *UGT1A6* found in human liver, kidney, and lung tissue (Münzel et al., 1996), with higher expression in human liver tissue and the least amount of expression detected in the lung. Analysis of brain tissue resulted in minimal levels of expression being detected in only two samples. Analysis of tissue specific *UGT1A6* expression in mice (Buckley and Klaassen, 2007) partially supports the expression profile seen in the equine samples. In mice, *UGT1A6* was found to be most highly expressed in the liver and minimal expression in the brain with both kidney and lung expressing *UGT1A6* at lower levels than in the liver and higher than the brain tissue (Buckley and Klaassen, 2007). However, in mice, lung tissues expressed *UGT1A6* at double the level seen in the kidney tissue, which is different to that observed in the horse (Buckley and Klaassen, 2007). Such expression profiling has also been performed in rats, which also found *UGT1A6* to be minimally expressed in the brain, however this study showed kidney expressed the greatest quantity of *UGT1A6*, with liver and lung displaying similar lower levels of *UGT1A6* (Shelby et al., 2003) This suggests UGT1A6 contributes to metabolism in the liver but the involvement of UGT1A6 in

metabolising compounds in the brain is negligible, implying there is little physiological need for *UGT1A6* to be expressed in the brain.

4.4.3 Expression profile of predicted UGT3A1-like in liver, lung, kidney and brain

Expression of equine *ENSECAG00000010396* was investigated in liver, lung, kidney, and brain, and expression levels were found to be the highest in the liver. This is unsurprising as this tissue is the location of the majority of metabolic reactions. Lung and kidney tissue displayed similar levels of gene expression whilst expression was detected in only two of the five samples of brain tissue. These results in part support the information provided by Dr Emes, which indicated expression in the liver to be high, while the RPKM for kidney of one horse indicating expression to be low; this previous study did not report RPKM values for lung or brain tissue. The extremely low levels of detection in the brain suggest the requirement for this gene in the brain is low.

The expression profile of *ENSECAG00000010396*, predicted *UGT3A1-like,* is different to that reported in the mouse where expression of both *UGT3A1* and *UGT3A2* is low in the liver and highly expressed in kidney tissue (Buckley and Klaassen, 2007). Expression was not detected in several other tissues, including lung and brain (Buckley and Klaassen, 2007); this expression profile in mice does not match what was observed in this study. *UGT3A1* expression has not been investigated in human tissues, however *UGT3A2* has been analysed and shown to be expressed in the kidney, testes, and thymus but absent from the liver (MacKenzie et al., 2011). Based on this information we can propose that the equine *ENSECAG00000010396* displays an expression pattern more consistent with that of mouse *UGT3A1* and *UGT3A2*, in that expression is present in liver and kidney but low in lung and brain. Additionally, *UGT3A2* expression in humans is absent in the liver and we have detected expression in this organ. This would support the hypothesis that we have isolated and quantified the equine orthologue of *UGT3A1*.

4.4.4 Preliminary investigations into age specific expression of UGT1A6

The preliminary investigations into expression levels of UGT1A6 and whether this differs with age showed no significant results. The data (Figure 4.11) suggested that in the first 5 years of life there is an increase in expression of *UGT1A6* with levels then decreasing before increasing again in the 7yr animals. There is a second decrease in expression prior to levels rising again in the older animals. Due to a low number of replicates it is not possible to determine the accuracy and significance of this apparently bimodal pattern of expression. To gather a better understanding of the effect of age, analysis of gene expression in more animals is required.

Grouping the animals in age ranges showed that expression is highly variable. With such low number of replicates it is difficult to determine whether the pattern seen is genuine and reflective of the populations. Expression and regulation of the *UGT*s is not completely clear and understood. In humans *UGT1A6* and *UGT2B17* have been shown to be impacted by age (Neumann et al., 2016), but how this impact is driven remains unclear. The expression of *UGTs* is partially impacted by hormone signalling (Neumann et al., 2016) and dietary flavonoids have been found to induce expression of *UGTs*, notably *UGT1A1* (Moon et al., 2006). With dietary and/or hormone components potentially responsible for the expression pattern observed in the horse, a larger cohort is required to elucidate the expression pattern and attribute potential causalities for these differences.

4.4.5 Preliminary age specific investigations into the expression of UGT3A1

The preliminary investigation into age-specific expression has suggested there may be a correlation between expression and age. The trend in the data suggested an increased requirement for the expression of *UGT3A1* up to 11yrs. Whilst the underlying cause for age specific expression is unknown, factors such as hormone levels and diet have been implicated in affecting *UGT* expression (Moon et al., 2006, Neumann et al., 2016). It may be that dietary requirements or hormones produced in mid-aged animals are having some effect on *UGT3A1* expression, but the precise

cause and relationship, if any, between age and expression remains to be elucidated. Age effect on *UGT3A1* expression has not been studied in any mammal, and due to the low number of replicates the information should be interpreted cautiously. A larger data set would provide robust data with which to investigate age-specific *UGT* expression.

4.4.6 Sex specific expression

There were two 5yr, 6yr, and 19yr old animals included in the analyses, of which there is only information on sex for the 6yr old animals. A difference in the expression levels of *UGT1A6* can be seen between the mare and gelding, with the mare displaying double the expression levels seen in the gelding. Gender differences in the expression of *UGT1A6* have been detected in rat ocular tissue (Nakamura et al., 2005, Buckley and Klaassen, 2007, Leclerc et al., 2002), which showed that females expressed a greater quantity of *UGT1A6* in the lens and extra-lenticular tissue of the eyes, whilst a study involving a wider range of tissues showed males to express significantly higher levels of *UGT1A6* in the lung. Further investigations involving a larger number of males and females are required to determine whether UGT expression in the horse is sex dependent.

The sexual dimorphism in expression of metabolic enzymes can result in gender specific metabolism and pharmacokinetics, with over 1000 genes identified in rat and mice that have implications in sex specific metabolism (Waxman and Holloway, 2009). A sex difference may result in the potential to respond to drugs differently; the rate at which drugs become bioavailable or the rate at which they are cleared from the body may be different (Waxman and Holloway, 2009). With regards to race horses, the administration and withdrawal periods are the same for either gender, however it may be that females can metabolise a compound much faster than males, or vice versa, in which case the withdrawal time to eliminate the drug completely from the body could be shorter. Conversely if a compound is metabolised quickly by one gender, it may mean illegal doping is more difficult to detect. As the metabolism

and pharmacokinetics of drugs is elucidated, it is plausible that the racing industry may need to consider the possibility of introducing sex specific regulations to drug administration.

4.4.7 Breed specific UGT expression

It is possible that there are differences in gene expression between breeds and breed specific SNPs may play a role. Studies in cattle have found breed-specific *UGT1A6* expression (Giantin et al., 2008). Three breeds of cattle were investigated for expression and activity, with *UGT1A6* expression in Charolais found to be 1.5-fold higher than in Piedmontese and 8-fold higher than in Blonde d'Aquitaine animals, although enzyme activity was only 1.5- and 2-fold higher (Giantin et al., 2008). Expression of *UGT1A1* was also investigated and shown to have the opposite pattern, with Blonde d'Aquitaine animals expressing *UGT1A1* at higher levels than that observed in Piedmontese and Charolais animals (Giantin et al., 2008). Differences in expression may influence the bioavailability and efficacy of xenobiotics and lead to differences between breeds in terms of clearance. In several countries, where the horse is a food source, breed differences in the metabolism of compounds may mean that pollutants and xenobiotics are not fully eliminated prior to entering the food chain (Giantin et al., 2008, Sallovitz et al., 2002). It is possible that breed differences may affect the response of horses to drugs, legal and illicit, which may have both positive and/or negative effects on their health and well-being. It is possible one breed may metabolise a steroidal compound quickly and as such this illegal drug may not be detected in doping tests. The same may also be true of dogs used in racing, one breed may metabolise an illicit compound quicker than another and thus avoid detection from anti-doping tests.

Information on breed specific SNPs may ultimately lead to tighter breed specific anti-doping regulations for horse and canine racing. For veterinary species in general it may potentially lead to breed specific treatments and doses of drugs. A larger study is required to identify breed specific polymorphisms, and this would need to be

followed by functional assays to determine the impact of the SNP on enzyme kinetics and substrate specificity.

4.4.8 Single nucleotide polymorphisms (SNPs) in UGTs

4.4.8.1 Identification and potential implication for SNPS in equine UGT1A6

From the comparison of *UGT1A6* sequence data from nine horses four polymorphisms were identified. Three of these SNPs conferred amino acid changes that may impact on enzyme function. Mutations at position 124 and 158 are in the amino part of the protein and have the potential to impact on substrate selectivity, whereas mutations 1235 and 1559 are in the C-terminal and could potentially affect the ability to bind to the ER or accept the donor sugar (Figure 4.17).

The transversion, in 20% of the animals, of guanine to cytosine at position 124 within the N-termini alters the amino acid from aspartic acid to asparagine - a change from acidic to neutral polarity. This mutation occurs in the first exon of *UGT1A6* which is implicated in substrate specificity. Similar studies in humans suggest that mutations in the first exon can result in a reduction in the ability to metabolise aspirin and non-steroidal anti-inflammatory drugs (NSAIDs) (Chan et al., 2005). If *UGT1A6* is proven to metabolise equine NSAIDs, it is possible that the mutation at position 124 could impair the ability of the enzyme to recognise and metabolise the NSAID. The SNP at position 158, present in two animals, is a silent mutation, cytosine to thymine, and is not anticipated to have any functional impact.

In the C-terminus the polymorphism at position 1235 (3/9 horses) is non-synonymous creating a change from histidine to arginine (amino acid 411). This alteration occurs immediately downstream of the signature sequence which is a key motif in the binding of the donor sugar encoded from position 366-394 (Meech et al., 2012b). Whilst several residues, including amino acids 373, 382, 374, and 376, within the signature sequence have been implicated in the binding of UDPGA owing

to their ability to form hydrogen bonds with the donor sugar, amino acid 411 has not been shown to play a role (Nair et al., 2015). As such, there is no evidence to suggest that the polymorphism at position 1235 would have functional implications, however functional studies are required to confirm the significance of this change.

The SNP at position 1559 is situated after the putative transmembrane domain resulting in the alteration of proline to glutamine. The transmembrane domain is an important motif in membrane bound proteins, with investigations into the truncation of the human UGT2B1 transmembrane domain resulting in abolished enzymatic activity (Meech et al., 1996). Looking beyond the UGTs, transmembrane domains are an important component for multiple proteins including fibroblast growth factor receptor 3 (*FGFR3*). A mutation in the transmembrane domain of *FGFR3* at position 1138 has been linked to Achondroplasia or dwarfism (Shiang et al.), demonstrating that there are phenotypic implications for the presence of mutations within transmembrane domains. As the mutation at position 1559 is not within the transmembrane domain, there may be no functional implications in terms of retention to the ER membrane. However, the result of this mutation is a change from a non-polar to a polar amino acid. In addition to the change in charge, proline is a cyclic imino acid whereas glutamine is not, but instead having a side chain, which is the polar moiety. This may potentially cause a structural change in the protein and have implications on the potential interactions between the protein and the ER membrane; this remains to be functionally elucidated.

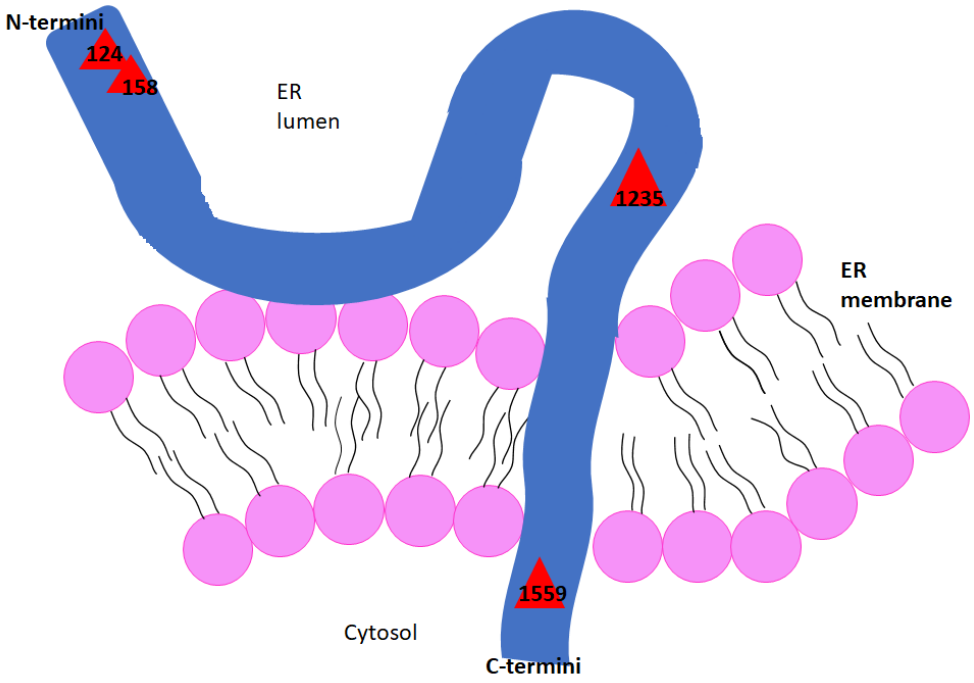


Figure 4.17: A schematic of mutations on the membrane bound UGT. The schematic displays the position of the mutations in UGT1A6 in relation to the membrane bound structure of the enzyme.

4.4.8.2 Potential implications for SNPs identified in equine UGT3A1

Sequence data from nine horses identified a single SNP which alters the amino acid at position 65, in exon 1, from proline to alanine. Whilst this change does not alter the charge, as both proline and alanine are hydrophobic amino acids, it may result in a conformational change as proline is a cyclic amino acid and alanine is not, instead it has a short hydrophobic side chain. The cyclic structure of proline means this amino acid provides conformational rigidity to the protein and puts a 'kink' in the protein at -65^{o}; this change of proline to alanine removes this angle from the polypeptide (Szabados and Savouré, 2010). The functional importance of this SNP remains to be elucidated, however as this falls within the signal peptide it may have implications for the binding of the protein to the ER. The horse SNP database (http://snugenome2.snu.ac.kr/HSDB/search.php) was interrogated to determine if this SNP, or others within this gene, had been identified previously. Three SNPs are reported within this sequence, none of which correlate to position 65 in the sequence.

The first reported mutation, in exon 4, is synonymous. However the second and third SNPs yield a change in amino acid. In exon 1, the polymorphism changes the nucleotide from C to T, altering the codon from CCT – TCT which results in the change of amino acid from the non-polar cyclic proline to a non-polar leucine which is not cyclic, but instead has a hydrophobic side chain. The third SNP is in exon 2, altering the codon from CAG to GAG; this causes a change from polar glutamine to glutamic acid, which is negatively charged. In the *UGT3*s, exons 1-4 comprise the N-termini which is the substrate binding domain (Meech et al., 2012b). The functional and structural implications of these changes in amino acids has not been investigated but it is possible that they may alter the substrate specificity of the enzyme. The change from a cyclic proline to the hydrophobic side chain amino acid of leucine may potentially create an alteration in the structure of the protein and the change in charge between glutamine and glutamic acid may affect internal or external interactions. Functional studies are required to establish the implications of these reported SNPs and the SNP detected in this research.

4.4.9 Why is the identification and characterisation of SNPs important?

Investigations into the *UGTs* in humans have determined this group of enzymes to be highly polymorphic (de Wildt et al., 1999). Over 63 mutations are linked to diseases which impair the ability of the human body to process bilirubin (Jancova et al., 2010, Guillemette, 2003). A SNP in exon 1 of *UGT1A1* alters the amino acid from an arginine to a glycine resulting in the development of Gilbert's syndrome (Jancova et al., 2010). Gilbert's syndrome causes mild, unconjugated hyper-bilirubinemia, which reduces the efficient excretion of bilirubin by approximately 30% in 10% of the population (Jancova et al., 2010, Bosma et al., 1995).

In human *UGT1A6*, three SNPs have been found to alter protein activity with pharmacological implications such as the reduced capacity to metabolise phenolic compounds (Ciotti et al., 1997). Information available on equine SNPs at the Broad Institute (https://www.broadinstitute.org/horse/horse-single-nucleotide-polymorphisms) indicates that 948,609 SNPS have been identified in the genome, of which 37,443 exist on chromosome 6. Using the Ensembl database and the horse SNP database (HSDB - http://snugenome2.snu.ac.kr/HSDB/index.php), 22 SNPs in equine *UGT1A6* in the database were identified, of which eight are synonymous and not thought to have any impact on enzyme function. Of the 14 non-synonymous SNPs, only four have been validated; two are in intronic regions and the other two are situated upstream of the gene. Functional implications of the remaining SNPs need to be validated, but it shows that equine *UGTs*, just as human *UGTs*, are polymorphic. None of the polymorphisms found in this study correlate with SNPs in the horse SNP database suggesting novel polymorphisms were discovered.

Population-specific SNPs with functional implications have been discovered in human *UGT1A9* and *UGT1A7* (Villeneuve et al., 2003). A substitution in *UGT1A9* of methionine to threonine at position 33 resulted in a decrease in the ability to metabolise 7-ethyl-10-hy-droxycamptothecin, but this mutation was only found in

Caucasians. In addition two SNPs in *UGT1A7* that are only found in African-American populations result in to decreased metabolism of 7-ethyl-10-hy-droxycamptothecin (Villeneuve et al., 2003). Regarding the *UGT* family specifically, the NCBI database has SNPs reported for human *UGT3A1*, but the functional importance of these remains to be established. A single SNP, resulting in the alteration of an amino acid (T316G) in human *UGT3A1,* was found to abolish catalytic activity (Meech and Mackenzie, 2010). There is no evidence in the NCBI SNP database for SNPs in *UGT3*s from other species; this may be due in part to this being such a small family and relatively 'unimportant' compared to the range of substrates that the UGT1s and UGT2s metabolise.

Due to a lack of information on breed and the number of horses sampled, it cannot be ascertained as to whether any of these SNPs are breed specific. Studies in humans have found particular polymorphisms to have higher prevalence in certain populations. For example a mutation in *UGT1A1*, denoted *UGT1A1*28*, associated with Gilberts Syndrome, was found in 8% of Egyptians, 16% of Europeans and 23% of African Americans, which means African Americans are more likely to develop the disease (Ehmer et al., 2012). Although mammals have not been well phenotyped for breed specific mutations, studies in dogs have found drug responses vary between breeds across a range of metabolizing enzymes (Fleischer et al., 2008). Therefore, it is likely that SNPs may be identified in the equine *UGTs* that are breed specific. This is important as it may result in breed specific doses of drugs being prescribed and may improve predictions for drug products in terms of extrapolating breed specific toxicology as well as altering detection limits with regards to illegal dosing. The implications of the identified polymorphisms in equine *UGT1A6* require additional studies to determine whether they impact on protein function.

Chapter 5: Developing a functional *in vitro* system

5.1 Introduction

As a major participant of phase 2 conjugations, UGTs control the glucuronidation reactions (Jancova et al., 2010). Bound to the ER and present in the cellular cytosol (Owens et al., 2005), the active site of the enzymes faces into the lumen. Substrates must translocate from the cytosol to the lumen, either by diffusion or via transporters to be metabolised (Meech and Mackenzie, 1997a, Bock, 2003). While the liver is the primary site of drug metabolism, almost all tissues have been found to have some drug metabolising capabilities (Jancova et al., 2010, Izukawa et al., 2009).

5.1.1 What are the donor sugars for the UGTs?

For the UGTs to be capable of mediating the conjugation of glucuronic acid to a target substrate, a co-factor is required that can donate its sugar group (Mackenzie et al., 2005). For the UGT1 and UGT2 sub-families this is UDP-glucuronic acid (Jancova et al., 2010, Gibson and Skett, 2001). The presence of the carboxyl (-COOH) and multiple hydroxyl groups (-OH) make this molecule polar (Figure 5.1). In the case of the two members of the human UGT3A sub-family, UGT3A1 and UGT3A2, they have been proven to use alternative donor sugars; uridine diphosphate N-acetylglucosamine (UDP-GLcNAc) and uridine diphosphate glucose (UDP-GLc) respectively (Figure 5.1) (Meech et al., 2012b, Mackenzie et al., 2008).

Figure 5.1: Chemical structures of the UGT co-factors

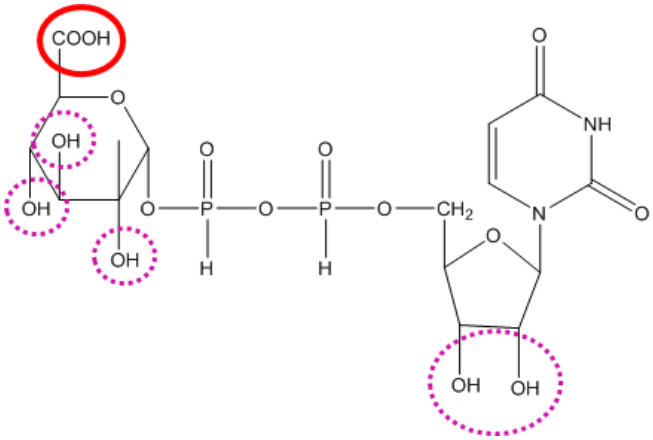


Figure 5.1a: UDP-α-D-glucuronic Acid

Figure 5.1a: Chemical structure of UDP-α-D-glucuronic acid (UDPGA). In humans UDPGA is the co-factor for Uridine- 5'-diphospho-glucuronosyltransferase (UGT) sub-family 1 and 2 enzymes. This donor sugar is highly polar owing to the carboxyl group (circled in red) and the multiple hydroxyl groups (encircled by pink, dashed lines).

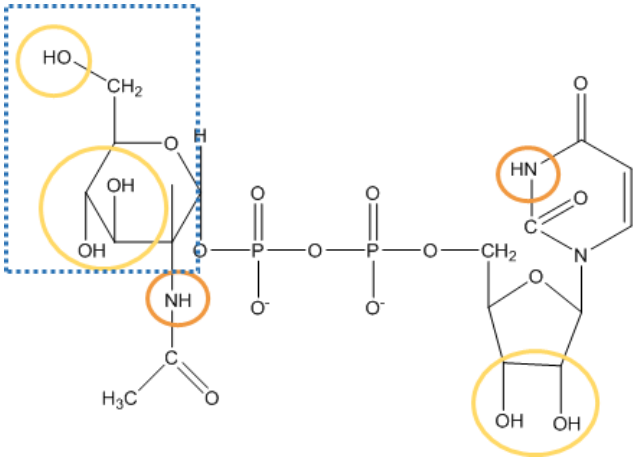


Figure 5.1b UDP-N-acetylglucosamine

Figure 5.1b: Chemical structure of UDP-N-acetylglucosamine (UDP-LGcNAc). UGT3A1 in humans uniquely utilizes UDP-GLcNAc as the donor sugar. Polarity is due to the presence of the hydroxyl groups (circles yellow) and the amine groups (circled orange).

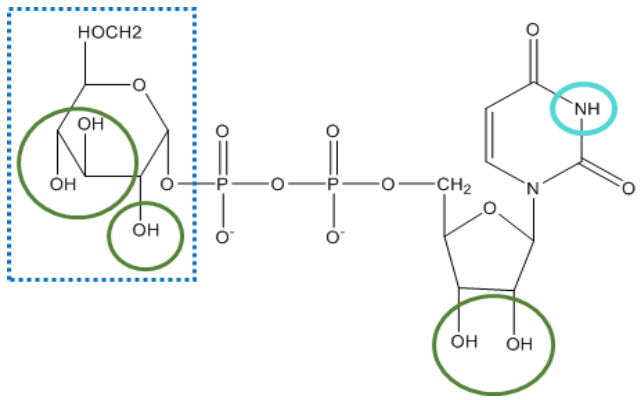


Figure 5.1c: UDP-diphosphate-glucose

Figure 5.1c: Chemical structure of UDP-diphosphate glucose (UDP-GLc). Human UGT3A2 uses UDP-GLc as the donor sugar for conjugation reactions. It is similar in structure to UDP-GLcNAc, with polarity created by the hydroxyl groups (circled green) and the amine groups (circled turquoise).

The structural differences between UDP-GLc and UDP-GLcNAc are highlighted by the blue box, on figures 5.1b and 5.1c.

Glucuronidation has the potential to occur if the substrate contains the appropriate functional groups, these include hydroxyl (-OH), carboxyl (-COOH), amines ($-NH_2$), and sulfhydryl (-SH) (Tukey and Strassburg, 2000, Gibson and Skett, 2001). Depending on which functional group the UGT reacts with determines the linkage created during conjugation resulting in three types of glucuronide formations: O-glucuronide, N-glucuronide and S-glucuronide (Gibson and Skett, 2001). The O-glucuronides are formed from the conjugation of UDP-glucuronic acid with hydroxyls and carboxyl groups, found on compounds such as phenols, alcohols, and carboxylic acids. N-glucuronides are the results of conjugations with amines and sulphonamides, and S-glucuronides are the result of reactions with thiols (Gibson and Skett, 2001). Irrespective of the type of linkages that occurs, the addition of glucuronic acid from the UDPGA donor sugar to the target substrate results in the alteration of substrate polarity from hydrophobic to hydrophilic (Guillemette, 2003, de Wildt et al., 1999, Gibson and Skett, 2001), which enables excretion into bile or urine. This reaction is exemplified in figure 5.2; oxazepam is a weakly polar substrate with a polar surface area of 61.7 Angstroms squared (Å), whereas UDP-glucuronic acid (UDPGA) is highly polar with a value of 309 Å. UGT facilitates the conjugation of UPDGA to oxazepam, which in turn increases the polar surface of the conjugate to 158 Å. This change in polarity enables the systemic removal of oxazepam.

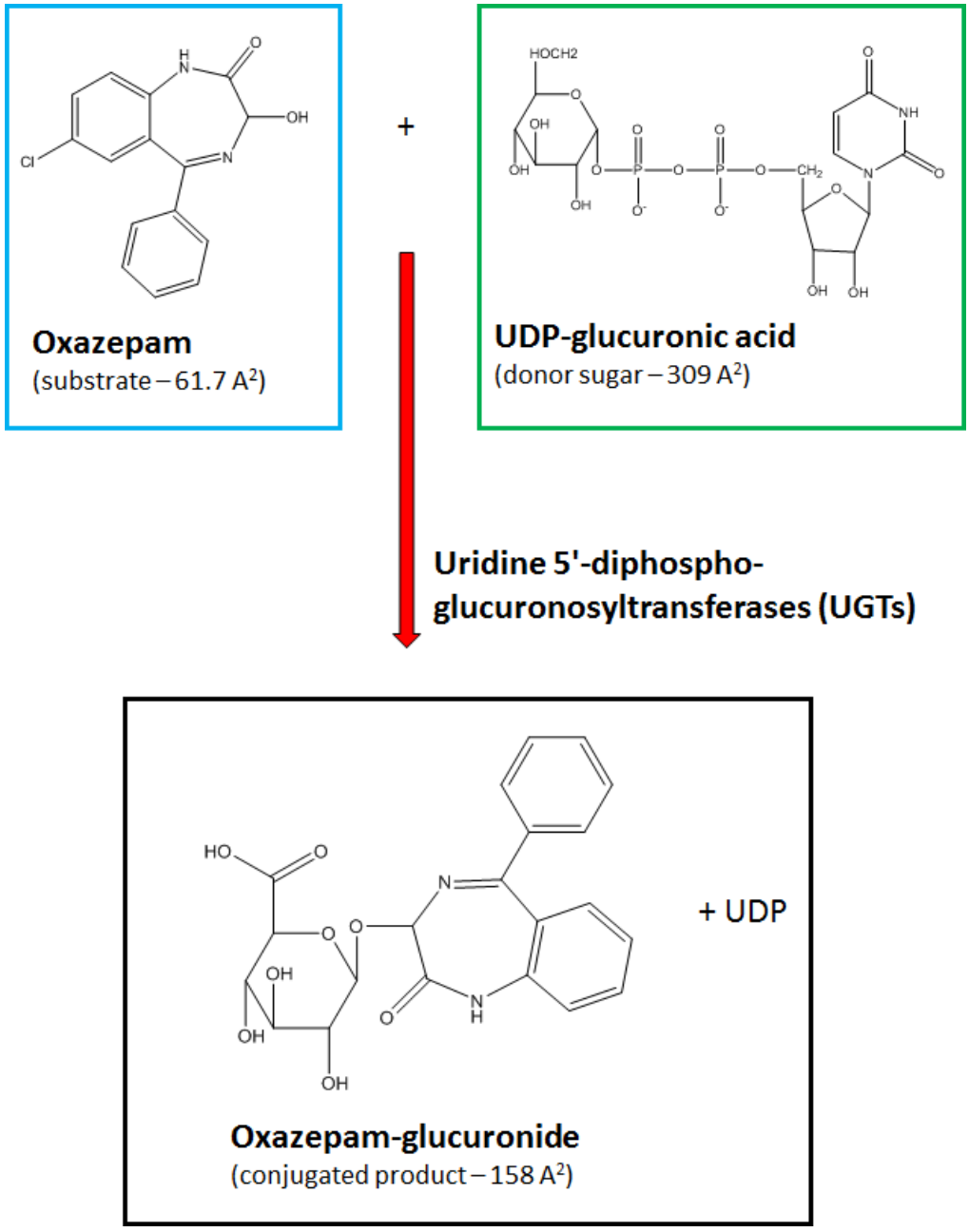


Figure 5.2: The conjugation of oxazepam to UDP-glucuronic acid. Oxazepam is a weakly polar molecule with a polar surface area of 61.7 Angstroms (Å), UDP-glucuronic acid (UDPGA) is a highly polar molecule with a value of 309 Å. Uridine 5'-diphospho-glucuronosyltransferases (UGTs) facilitate the addition of UDPGA to oxazepam, which in turn increased the surface polarity of the molecule to 158 Å. This change in polarity means the conjugate can be removed from the body.

As a major pathway of phase 2 metabolism, the UGTs have a diverse substrate range (Bock, 2003), including endogenous and exogenous targets such as bilirubin, hormones, fat soluble vitamins, dietary products, carcinogens and environmental pollutants as well as drugs (Krishnaswamy et al., 2005, Jancova et al., 2010). A single UGT isozyme often recognises multiple substrates. For example, in humans UGT1A1 recognises bilirubin, estradiol, thyroxin endobiotics as well as paracetamol and irinotecan (used in cancer treatment). Human UGT2B15 can metabolise paracetamol in addition to oxazepam and testosterone (Bock, 2010).

5.1.2 How do in vitro systems benefit the pharmaceutical industry?

In developing a new drug the pharmaceutical industry needs to understand the adsorption, distribution, metabolism, excretion, and toxicity (abbreviated to ADMET) parameters of the compound (Ekins et al., 2005). Dependent on how well a drug fulfils the ADMET parameters can determine whether its development is progressed or terminated (Zhang et al., 2012). *In vitro* systems can provide substantial knowledge on the metabolic stability, protein binding, and drug-drug interactions of compounds being developed (Bowes et al., 2012, Zhang et al., 2012).

Development of widespread *in vitro* systems is a cost effective mechanism by which a vast number of potential compounds can be screened for drug-like properties (Zhang et al., 2012) with a limited quantity of the test drug, which is often in short supply at the early stages of development. It enables the identification of non-target interactions and drug-drug interactions (Bowes et al., 2012). This leads not only to decreased costs but also efficiency savings as *in vitro* systems have limited confounding factors and compounds with very few or no non-target activity require less time on *in vivo* safety investigations, resulting in quicker development (Zhang et al., 2012, Bowes et al., 2012). Perhaps significantly, it side-steps the issues of side effects seen in non-target model organisms (Bowes et al., 2012, Zhang et al., 2012). The use of an equine *in vitro* system will provide more accurate estimates of clinical outcomes in equines. An additional benefit of *in vitro* systems is the application to

the 3Rs – replacement, reduction, and refinement. There are ethical considerations to take into account when using animals in science, as such effort is placed into using the smallest number of animals possible. *In vitro* systems not only reduce the cost but additionally reduce and even replace the number of animals required in a study (Graham and Prescott, 2015).

5.1.3 Applications of in vitro profiling to the horse racing industry

Horses are a popular companion animal, used in competitive sports such as polo and eventing, and are a source of meat in many countries. Therefore, knowledge of the mechanism of drug metabolism will be of benefit to the prescribing veterinarian and enable the pharmaceutical companies to test and predict the impact of drug-drug interactions and understand the kinetic profile of a drug.

Horse racing is a multi-million-pound industry; the British Horse Racing Authority (BHA) estimates the value to the economy of the United Kingdom (UK) to be £3.45 billion annually (http://www.britishhorseracing.com/bha/what-we-do/industry-leadership/). In the UK, the BHA is responsible for the regulations regarding the administrations of drugs to racing horses, with strict regulations as to what drugs can and cannot be administered, withdrawal times and testing (full details available on the BHA website, http://rules.britishhorseracing.com/Home).

The development of a recombinant UGT *in vitro* system will be beneficial to the pharmaceutical industry, enable mass testing and profiling of drugs currently used in equine health and welfare, but will also enable testing of a wide variety of compounds not used in the treatment of equines which may have as yet unknown benefits to horses. With regards to anti-doping, profiling the glucuronidation of multiple drugs will provide substantial information on the half-lives of compounds, determining any breed differences in metabolism, what effect drug-drug interactions have and the possibility to investigate un-identified compounds from drug screens.

5.2 Aim

To determine the functionality of the recombinant UGTs isolated within this research and to establish their phenotype through substrate profiling. Of the five equine UGTs, three were successfully cloned into the pcDNA3.1 vector. These were used for the subsequent investigations.

5.3 Method and materials

5.3.1 Selection of UGTs for further study

Of the five isolated and sequence *UGT* genes three were selected for further study. One member from each sub-family was selected. *ENSECAG0000023519* (*UGT1A6*) was the only sequence isolated from the UGT1 sub-family. ENSECAG00000014362 (*UGT2A3*) was selected to represent the UGT2 sub-family, whilst for the UGT3 sub-family, *ENSECAG00000010396* (*UGT3A1*) was chosen as this appears to be a full length isoform, whereas ENSECAG00000008900 may encode a premature stop codon.

5.3.2 Sub-cloning UGT sequences into the expression vector

Three of the five isolated and sequenced *UGT* genes were selected to be taken forward to the expression stage. The *UGT* genes required digesting out of the pCR™2.1 vector and ligating in to the expression vector pcDNA™3.1. The multiple cloning sites (MCS) of both pCR™2.1 and pcDNA™3.1 were compared to compile a list of enzymes that digested both vectors. Each isolated UGT gene was analysed in NEBcutter (http://nc2.neb.com/NEBcutter2/) to correlate restriction enzymes that cut within the MCS with ones that did not digest the gene of interest (Table 5.3). Double digests were performed to facilitate directional sub-cloning. Products digested from TOPO2.1 pCR™2.1 vector were visually checked for linearity and size on a gel prior to excision.

Gene	Enzyme 1	Enzyme 2
23519	*BamHI*	*Xba I*
14362	*Xho I*	*BamHI*
10396	*BamHI*	*Xho I*

Table 5.3: Table of enzymes used to sub-clone the full-length genes. Full-length genes cloned into pCR™2.1 were excised from the vector using two enzymes to enable directional sub-cloning into the pcDNA™3.1 expression vector. 23519 – *ENSECAG00000023519*, 14362 – *ENSECAG00000014362* and 10396 – *ENSECAG00000010396*.

Once orientation was confirmed, and sufficient quantities of the sub-cloned UGT had been obtained, the expression vector needed to be linearized ready for transfection in to the HEK293 cells. The pcDNA™3.1/UGT was digested using an enzyme which would only digest within the MCS. Linearization was checked by electrophoresis on an agarose gel.

5.3.3 Drug metabolism studies

5.3.3.1 In vitro drug metabolism studies

Three incubations were set up for each recombinant protein (rUGT) extracts, containing the following concentrations of total protein: 0.1mg/ml, 0.5mg/ml and 1.0mg/ml. For two assays, human UGT1A6 supersomes ™ (Corning® - Scientific laboratory supplies, UK) and equine microsomes (provided by Khalid Shibany) were also utilised, with concentrations as above.

Cluster tubes (Corning® – Scientific laboratory supplies, UK) were placed into a heated Bioshake iQ (QInstruments, Germany) set to 37°C. To each tube the following reagents were added: 3µl of Magnesium chloride (5mM – VWR International), alamethicin between 1.5-6µl (calculated as 50ug per mg of protein – Sigma-Aldrich) and protein to create one of the final concentrations mentioned

above. PBS (Gibco) was added to create total volume of 980µl (see Table 5.4 for clarity on reagents volumes). Tubes were sealed with cluster caps (Corning – Scientific Laboratory Supplies) and the block set to pre-incubate at 37°C, 500rpm for 30mins to mix and bring all reagents to equilibrium. 5 mins prior to the end of the pre-incubation, 20µM (2µl) of the test drug (Table 5.5) was added.

	Final Total Protein content		
	0.1mg/ml	**0.5mg/ml**	**1.0mg/ml**
Alamethicin	1.5	3	6
$MgCl_2$	3	3	3
Drug	2	2	2
UDPGA	20	20	20
Protein	X	X	X
PBS	X	X	X
Total Volume	**1000µl**	**1000µl**	**1000µl**

Table 5.4: Volumes of reagents in recombinant protein (rUGT) incubates. The amount of PBS in the reaction was variable depending on the amount of protein.

After the pre-incubation was complete, 10ul (5mM) of uridine 5'-diphosphoglucuronic acid tri-sodium salt (abbreviated to UDPGA - Sigma- Aldrich) was added and the heating block was set to shake at 500rpm, 37^{o}C for 60mins. At 0min and 60min time points, 50µl of the reaction was removed and added to 100µl of ice cold methanol (Sigma-Aldrich) in 1.5ml microcentrifuge tubes to quench the reaction.

Tubes were centrifuged at max speed for 2mins. 100µl of the supernatant was transferred to Snap ring polypropylene vials (Supelco, USA) and sealed using Snap cap with PTFE/Red rubber natural caps (Supleco) ready for analysis on the mass spectrophotometer.

Drug	Classification
17alpha-trenbolone	Steroid
Oxazepam	Benzodiapezine
Ketoprofen	Non-steroidal Anti-inflammatory
Morphine	Opioid

Table 5.5: List of drugs. The four drugs tested in the *in vitro* assays were 17alpha-trenbolone, oxazepam, ketoprofen and morphine; their respective medical classifications are given. All drugs were supplied by LGC Ltd.

5.3.3.2 Drug analysis

Analysis was performed at LGC Ltd (Medication and Doping Control Division) in Newmarket, Cambridgeshire. An injection volume of 10µl was used for the LC-MS/MS analysis. Samples were processed through the Orbitrap Discovery MS LTQ LX (Thermo-scientific) containing an Atlantis T3 µM column (2.1 x 100mm – Waters, UK) with the Accela auto-sample and pump in positive ion mode.

The analysis used a flow rate of 400µl/min, isocratic method. Solvent A contained 0.1% Acetic Acid containing 300ng/ml Uracil. Solvent B contained 0.1% Acetic Acid in Acetonitrile with 300ng/ml Uracil (reagents provided by LGC Ltd).

Each drug tested was analysed alongside a standard of pure drug in methanol, and where possible deuterated and glucuronidated forms were also utilised as standards.

5.3.3.3 Analysis of glucuronidation rate

Mass spectra was analysed using Xcalibur™ software (Thermo Scientific). Profiles were created based on the exact mass of the drug and drug-conjugate to identify the presence of the glucuronidated form. The area under the peak was measured and used to calculate whether conjugation had taken place. The area mass values were placed into a MS Excel spreadsheet and the intrinsic clearance was calculated, using the following formula:

A = LN(area mass of peak at 0 min)

B = LN(area mass of peak at 60 min)

SLOPE(A:B,0:60) * 1000 = ul/min

For each incubate account for protein content so final value for Clint = ul/min/mg of total protein

5.3.3.4 A rationale for use of intrinsic clearance

If the product (metabolite) of an enzyme reaction is monitored for the purposes of determining the intrinsic clearance, then the concentration of the metabolite (in molar units) is required and therefore a standard curve is necessary. This is because the rate of production of the metabolite initially follows zero order kinetics giving a linear increase of metabolite concentration with time. However, no metabolites were observed at the limit of detection and therefore this methodology cannot be used.

Alternatively, the intrinsic clearance can be determined by monitoring the disappearance of substrate (in this case the drug). Disappearance of the substrate will follow a first order exponential decay and if the data for disappearance is natural log transformed and plotted against time the depletion rate constant can be calculated from the slope. The initial rate of disappearance can be simply calculated

(as the initial drug concentration (in molar) in the incubation is known from the outset) by multiplying the depletion rate constant by the initial starting concentration. There is no need to use a standard curve and the units of intrinsic clearance can be expressed in uL/min/mg protein rather than % remaining/min/mg protein. This is a well-established method and is one of the main advantages of using drug disappearance over metabolite formation monitoring (Nath and Atkins, 2006).

5.4 Results

5.4.1 Geneticin – Kill curve

The kill curve was performed using Geneticin antibiotic over a concentration range of 0-5000 μg/ml. Over the course of 17 days samples were collected every 48hrs and cell counts performed to determine the number of living and non-viable cells. Figures 5.6a and 5.6b display the results of the cell counts as the percentage of viable cells.

The number of cells increased with 80% of cells viable at day 5 (Figure 5.6a). A decrease in viable cells was then observed over the course of days 7 to 17, although total cell death was never reached. The 50 μg/ml concentration showed an initial increase in the number of cells surviving the addition of Geneticin during the first five days. At day 5, cell viability peaked at 87% before showing a steady decline from days 7 to 17. By the end of the study at day 17, cell viability was reduced to 28%. Incubations containing 125 μg/ml Geneticin displayed a decrease in viable cells from 100% to 42% over the first five days, then the number of viable cells plateaued at approximately 33% for the remainder of the test period. Treatment with 250 μg/ml of Geneticin resulted in no live cells after nine days of incubation, while both 500 μg/ml and 750 μg/ml of Geneticin yielded no viable cells between days 5 and 7 of the study period.

Three assays were performed with Geneticin in excess, as positive controls, with concentrations of 1000 μg/ml, 2000 μg/ml, and 5000 μg/ml (Figure 5.6b). Both 2000 μg/ml and 5000 μg/ml concentrations resulted in complete cell death within three days, and treatment with 1000 μg/ml resulted in no viable cells by day five.

Based on these results a concentration of 500 μg/ml of Geneticin was selected to derive and maintain the transfected cell lines.

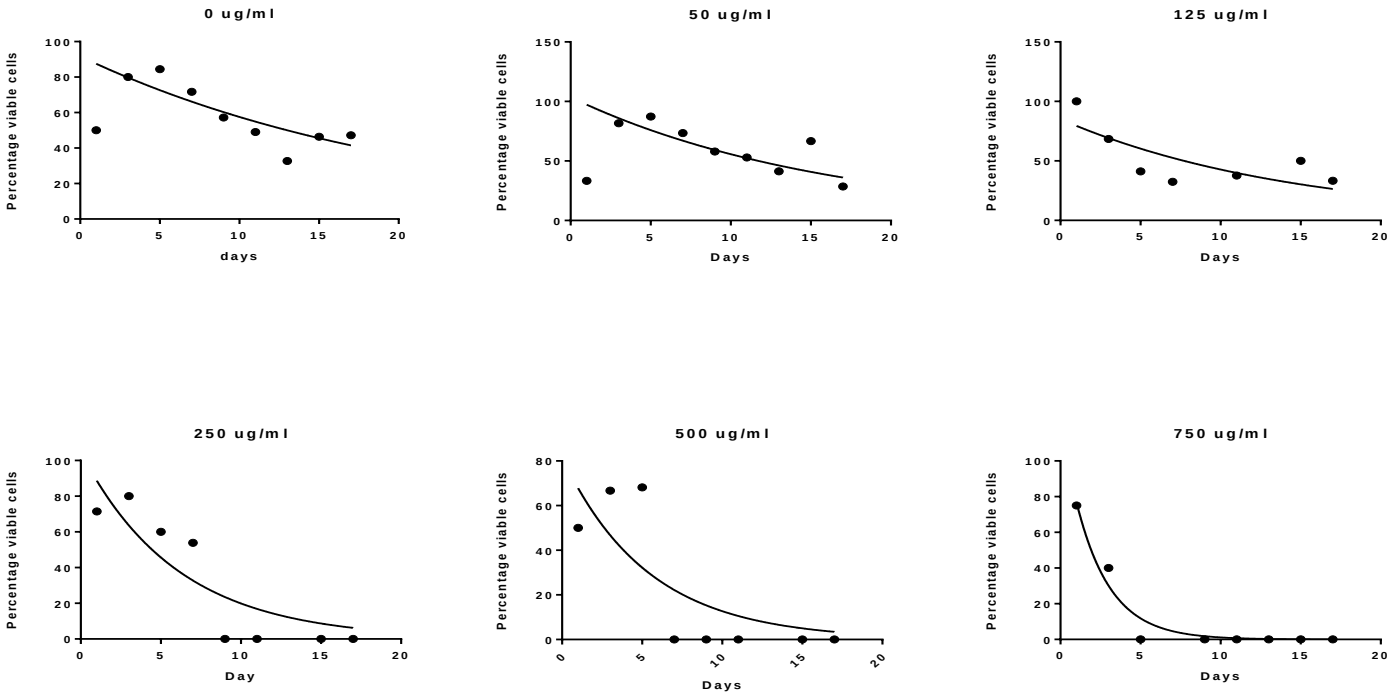


Figure 5.6a: Geneticin kill curves. HEK293 cells were incubated with Geneticin (G418) for a period of 17 days over a concentration range of 0-5000 µg/ml. This figure displays the percentage cell viability for concentrations 0 µg/ml to 750 µg/ml. 0 µg/ml is the negative control, cell viability was seen to decrease although total cell death was never reached. 50-125 µg/ml incubates showed a steady decrease in viable cells over the course of 17 days. 250-500 µg/ml displayed an initial increase in the number of viable cells for the first 3-5 days prior to displaying complete cell death. 750 µg/ml incubates resulted in complete cell death by day 5.

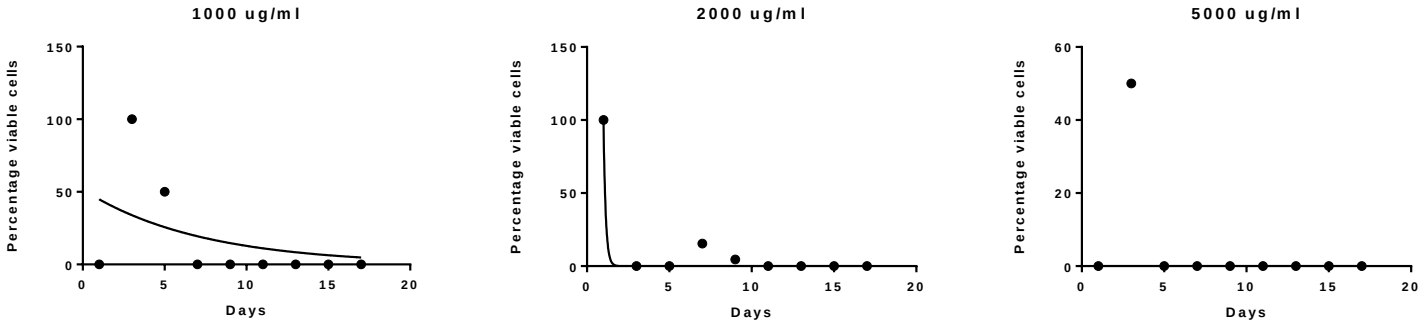


Figure 5.6b: Geneticin kill curves. Cells were incubated with Geneticin for 17 days. Concentrations of 2000 µg/ml and 5000 µg/ml were positive controls to ensure the cells responded to the presence of the antibiotic. The 1000 µg/ml incubate displayed complete cell death by day seven, whilst the highest concentrations, 2000 and 5000 µg/ml both resulted complete cell death by day five.

5.4.2. Western blots

5.4.2.1 Western blot to detect the presence of recombinant equine UGT1A6

A western blot was performed, repeated twice by Dr James, to detect the presence of recombinant equine UGT1A6. The polyclonal antibody used in this assay was human anti-UGT1A6 produced in rabbits which according to the manufacturer had been tested for cross-reactivity in several species including humans and equines. A positive control, human recombinant UGT1A6 with a glutathione S-transferase (GST) tag, was included, and this target protein was detected; this protein was slightly larger (78kDa) than the expected size for human UGT1A6 (60.7kDa) due to the GST tag. Analysis of the cell lysate from the equine UGT1A6 transfected HEK293 cells detected two proteins (Figure 5.7). The first, ~ 50kDa, is the more intense of the two and correlated to the expected size of the equine recombinant UGT1A6 protein. The second protein detected was estimated to be 40kDa in weight.

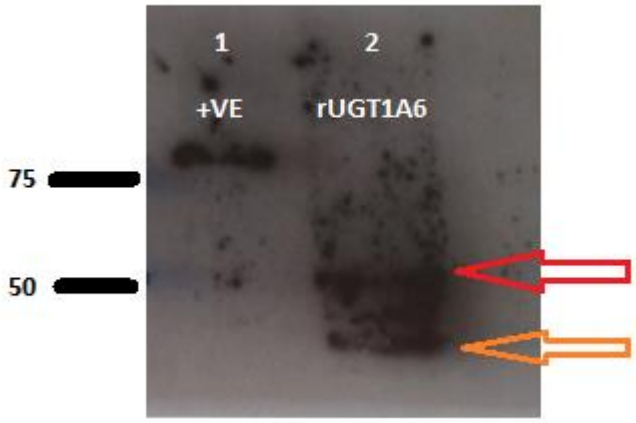


Figure 5.7: Western blot of recombinant UGT1A6. Lane 1 (+ve) contains the positive control, recombinant human UGT1A6 with a glutathione S-transferase tag, with a protein detected at 78kDA. Lane 2 contains the HEK293–equine UGT1A6 cell lysate (from HEK293 cells transfected with equine UGT1A6). The top band (red arrow) is 50kDa, the predicted size expected of the recombinant equine UGT1A6 protein there is a second, lower molecular weight band, 40kDa.

5.4.2.2 Western blot to detect the presence of recombinant equine UGT2A3

A polyclonal antibody produced in mouse against human UGT2A3, anti-UGT2A3, was used to test for the presence of recombinant equine UGT2A3. Purified recombinant human UGT2A3 with a GST tag was used as the positive control, and a negative control of cell lysates from untransfected HEK293 cells was also included in the analysis. Dr James twice repeated the western blot.

The negative control produced multiple bands with non-specific binding detecting two proteins both close to the 50kDA marker (Figure 5.8). The positive control sample of recombinant human UGT2A3 protein failed to detect a band. The transfected HEK293 cell lysate with equine rUGT2A3 produced two bands, both >100kDa.

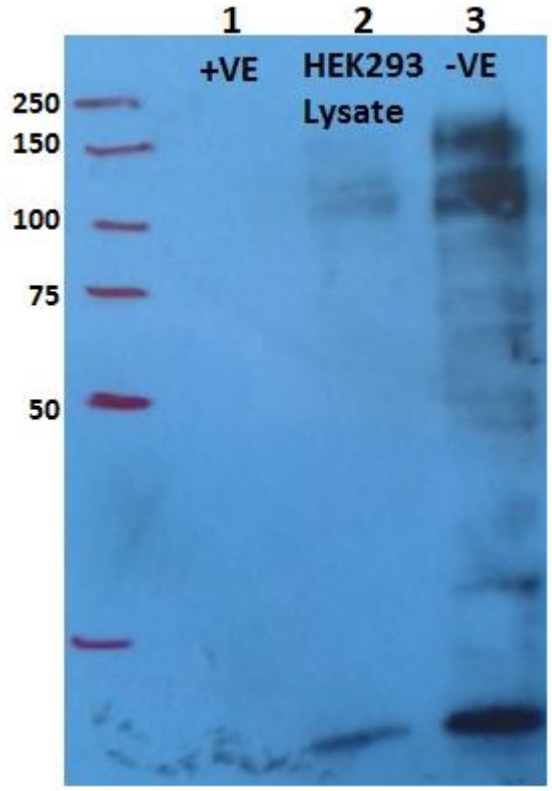


Figure 5.8: Western blot to detect recombinant equine UGT2A3. Lane 1 (+ve) contains the positive control, recombinant human UGT2A3 with a glutathione S-transferase (GST) tag, which was not detected in this assay. Lane 2 contains HEK293-equine UGT2A3 lysate (HEK293 cells transfected with equine UGT2A3), which produced bands of the incorrect size for the recombinant protein, >100kDA. Lane 3 (-ve) contains un-transfected HEK293 cell lysate as the negative control, which shows non-specific binding of the antibody used.

5.4.2.3 Western blot to determine the presence of recombinant equine UGT3A1

Western blots, twice repeated by Dr James, to detect the presence of recombinant equine UGT3A1 were performed using a mouse polyclonal anti-UGT3A1 raised against the human UGT3A1 protein. A positive control, human recombinant UGT3A1 with a GST tag failed to produce a band (Figure 5.9). A negative control of untransfected HEK203 cell lysate produced multiple bands. The transfected HEK293 cell lysate containing equine rUGT3A1 produced multiple bands, none of which are approximately 50kDA.

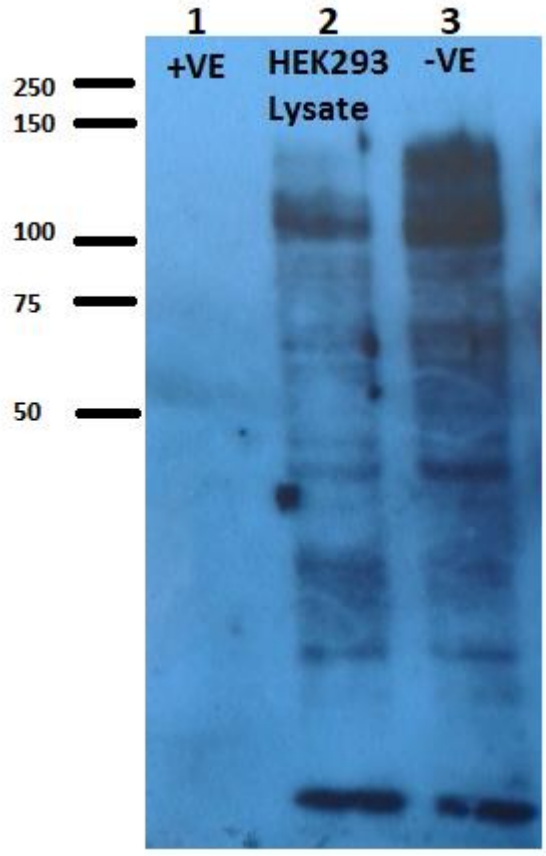


Figure 5.9: Western blot to detect the presence of recombinant equine UGT3A1. Lane 1 (+ve) is the positive control, recombinant human UGT3A1-GST tag, this failed to produce a band of 78kDA. Lane 2 contains cell lysate from HEK293-equine UGT3A1 (HEK293 transfected with equine UGT3A1), this sample produced several non-specific bands. Lane 3 is the negative control, containing un-transfected HEK293 cell lysate, which displays multiple non-specific bands.

5.4.3 Functional metabolism assays

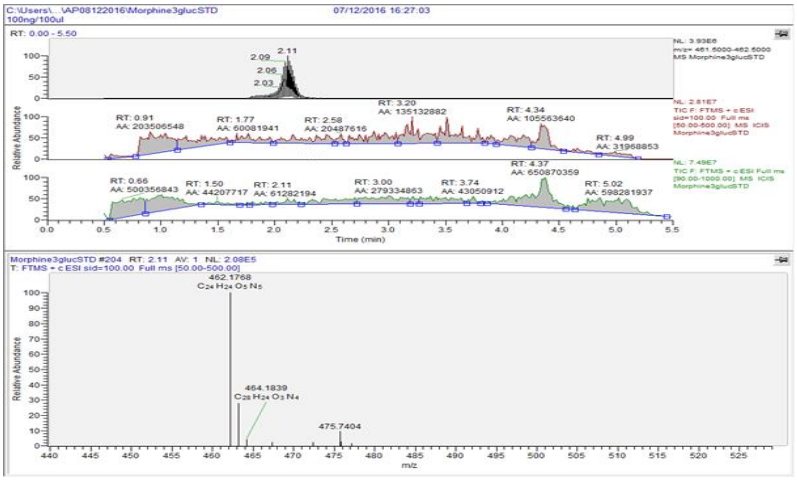
A list of drugs known to be metabolised in the horse was compiled. From this, drugs were selected that were of interest to the industrial partners, as they can potentially be used to enhance racing performance. Those selected for metabolism assays were of interest and/or deemed mostly likely to produce a positive result.

5.4.3.1 Testing the functionality of the recombinant UGTs using morphine as a substrate

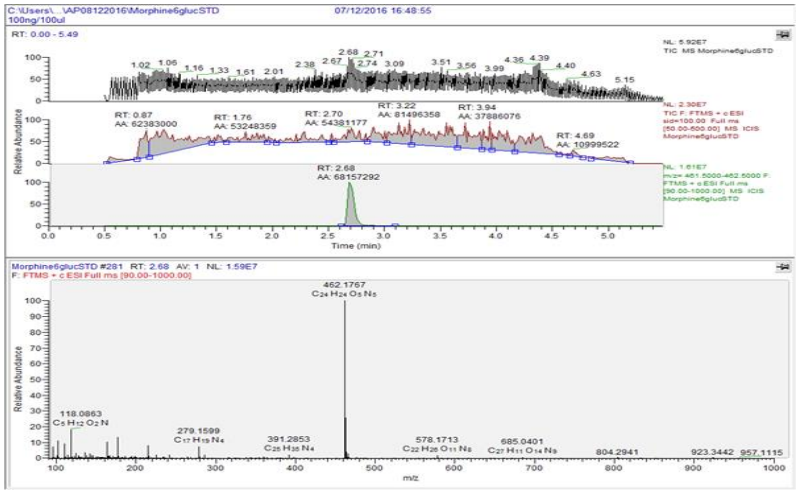
The Orbitrap mass spectrometer detects chemical compounds based on their exact mass, to an accuracy of four decimal places. The exact masses of morphine and morphine-glucuronide is 286.1445 g/mol and 462.1768 g/mol respectively. Morphine can be glucuronidated into two forms: morphine-3-glucuronide and morphine-6-glucuronide with both forms having identical masses. The two glucuronidated forms of morphine were be identified on the orbitrap by their retention times which differ due to separation by ultra-performance liquid chromatography (UPLC). Pure standards were analysed on the UPLC/Orbitrap system to determine the retention times of morphine and its glucuronidated forms in the absence of confounding factors. The retention time for the parent drug of pure morphine was 2.38 min, whereas morphine-3-glucuronide had a shorter retention time of 2.11 min and morphine-6-glucuronide a longer retention time of 2.68 min (Figure 5.10).

For each equine recombinant protein three incubations were performed containing different quantities of total protein. At two-time points, 0 min and 60 min, samples were taken and analysed on the UPLC/Orbitrap system. Using the exact mass and retention times, the incubations were analysed for the presence of the conjugated form of morphine and for decreasing quantities of the parent compound. The area mass for chromatogram peaks of correct mass and retention times were used to determine the intrinsic clearance (Cl_{int}) value for each recombinant UGT (rUGT) and

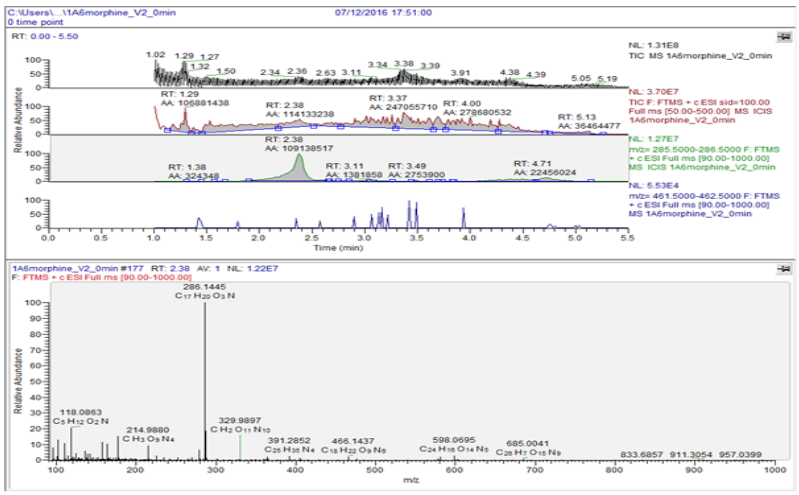
for each protein incubate (Table 5.11). Intrinsic clearance values (Cl_{int}) were calculated in units of µl/min/mg of total protein, based on the disappearance of the parent drug.



A) Morphine-3-Glucuronide standard.



B) Morphine-6-Glucuronide Standard



C) 0.5mg/ml UGT1A6 incubated with Morphine

Figure 5.10: LCMS traces. LCMS traces of (A) morphine-3-glucuronide and (B) morphine-6-glucuronide standards and the results of (C) 0.5mg/ml of UGT1A6 protein incubated with morphine.

Protein	UGT1A6	UGT2A3	UGT3A1
0.1mg/ml	10.002	-22.348	17.082
0.5mg/ml	-2.946	0.395	-0.489
1.0mg/ml	-3.874	6.7549	-1.795

Table 5.11: Cl_{int} values per mg/ml incubate per recombinant protein. Samples were taken at two time points, 0 and 60 min, using the area mass from the chromatogram peaks. The intrinsic clearance **(**Cl_{int}) values were calculated for each incubate and normalized for protein content. Each Cl_{int} = µl/min/mg of total protein and are based on the disappearance of the parent drug. Any Cl_{int} negative values indicate that no metabolism has taken place, these have been represented as zero on the graph.

The intrinsic clearance values for the rUGT1A6 incubates are below zero for the 0.5 mg/ml and 1.0 mg/ml incubations, but a Cl_{int} value of 10.00 is obtained for the incubation with the lowest total protein concentration, 0.1mg/ml (Table 5.11). The Cl_{int} values obtained for the rUGT2A3 incubations were below 0 (-22.3) for the lowest protein content incubate (0.1mg/ml), and 0.395 and 6.7549 for the 0.5 mg/ml and 1.0 mg/ml incubations, respectively. The rUGT3A1 incubates showed a decrease in Cl_{int} values with increasing protein content, with the 0.1 mg/ml yielding the highest intrinsic clearance value, 17.08, and the 1.0 mg/ml displaying the lowest value, -1.79 (Figure 5.12).

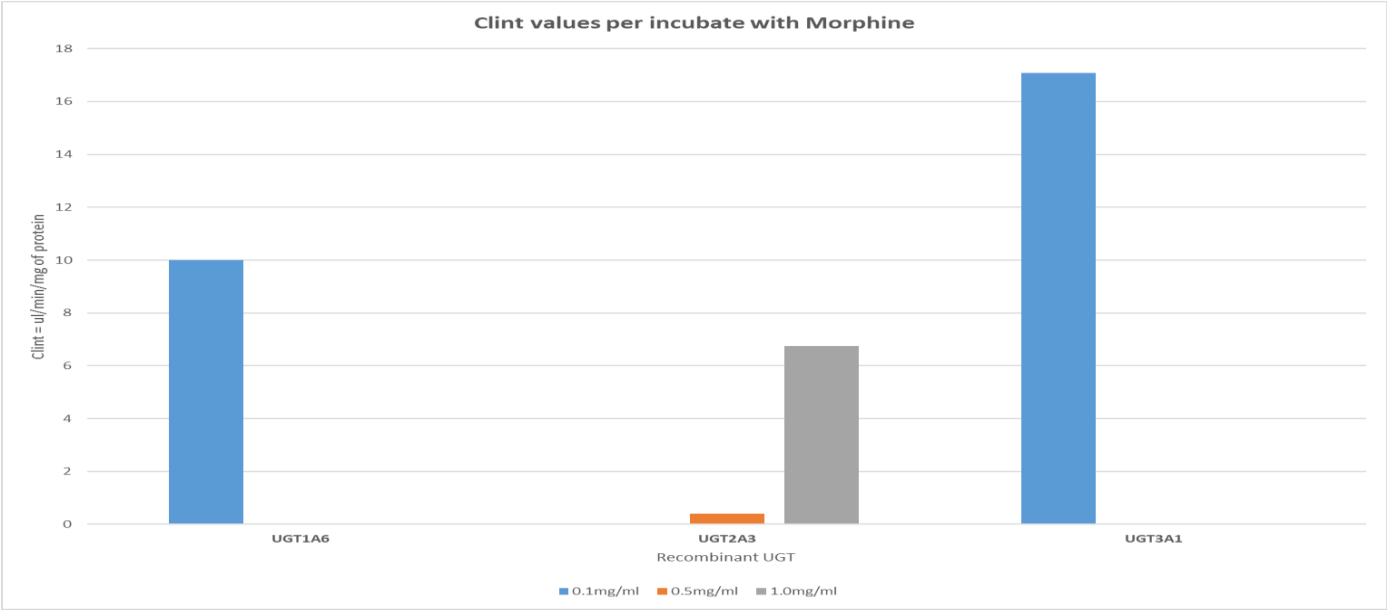
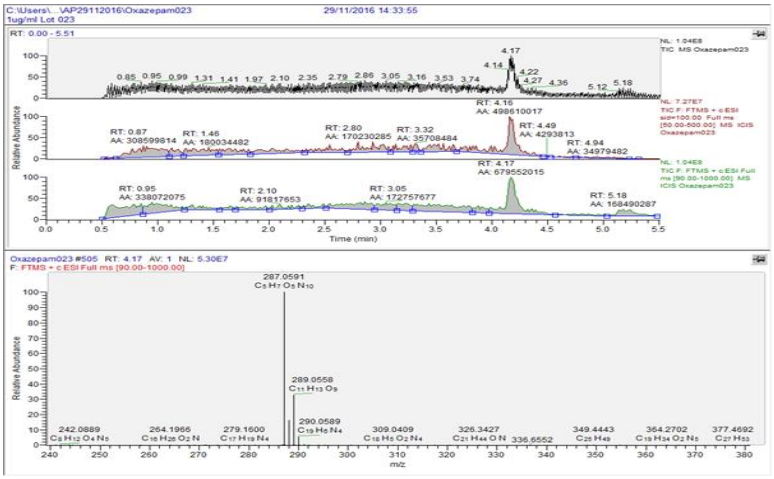


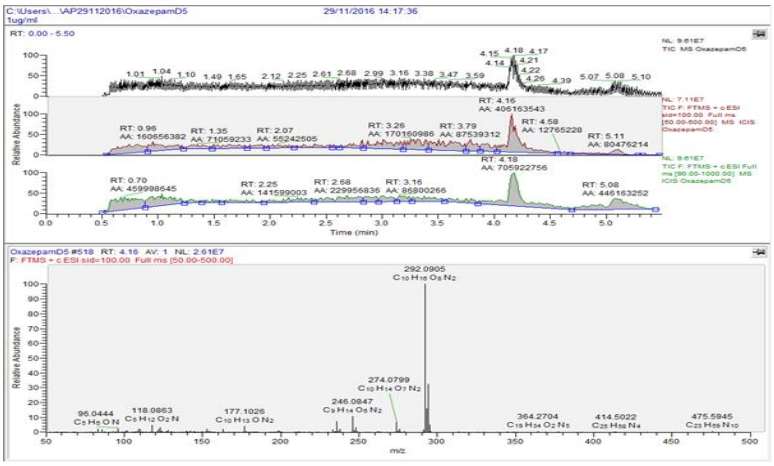
Figure 5.12: Graphs displaying the intrinsic clearance (Cl_{int}) values per recombinant protein over three different concentrations of total protein. A constant quantity of morphine was placed into each incubation, samples were taken at two time points, 0 min and 60 min. The area mass from the chromatograms at the relative retention time was used to calculate the Cl_{int} value. A value below zero indicated no metabolism of morphine had taken place, a value above zero suggests the parent compound may have been conjugated to the glucuronide. Three concentrations of total protein were tested per recombinant UGT, 0.1 mg/ml, 0.5 mg/ml and 1.0 mg/ml. For rUGT1A6, a positive Cl_{int} value was calculated for the 0.1 mg/ml incubation only. For rUGT2A3, a negative intrinsic clearance value was calculated for the lowest of the three protein concentrations, with a small positive value detected for the 0.5 mg/ml incubate and a value just above 6 for the 1.0 mg/ml incubate. rUGT3A1 displayed a relatively large Cl_{int} of 17 for the 0.1 mg/ml sample, and negative values for the higher two protein concentrations.

5.4.3.2 Testing the functionality of the recombinant UGTs using oxazepam as a substrate

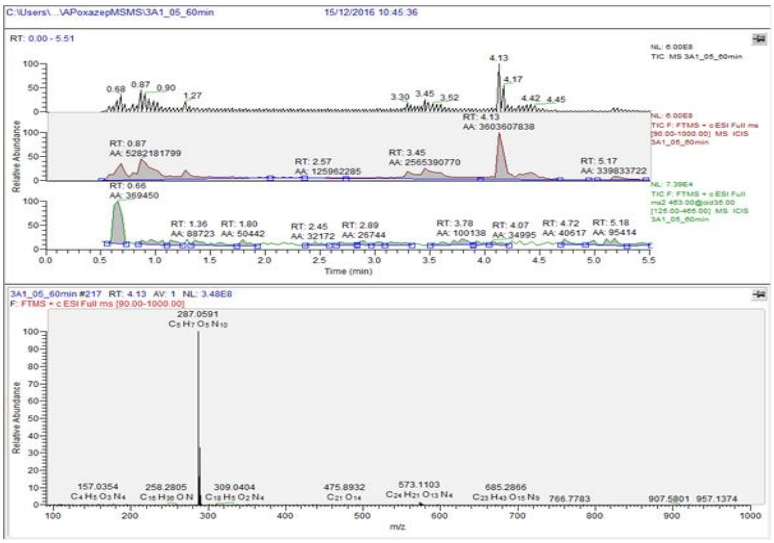
Pure standards of oxazepam, deuterated oxazepam D5 (deuterium has replaced five hydrogen atoms) and oxazepam-glucuronide were analysed on the UPLC/mass spectrometer system, and their retention times determined. Oxazepam has an exact mass of 286.71 g/mol and retention time of 4.17min, whilst the deuterated form has the same retention time but a larger exact mass of 291.74 g/mol. The conjugated form, oxazepam-glucuronide has an exact mass of 462.84 g/mol, but a shorter retention time of 3.91min (see Figure 5.13).



A) Oxazepam Standard



B) Deuterated Oxazepam (D5) standard



C) O.5mg/ml UGT3A1 incubated for 60mins with Oxazepam

Figure 5.13: LCMS traces for Oxazepam metabolism studies. Displayed are the LCMS traces for pure Oxazepam (A), used as a standard, Deuterated Oxazepam d5 (B) as a second standard and the results of 0.5mg/ml of UGT3A1 incubated for 60minutes with Oxazepam, which shows the presence of the parent Oxazepam and no glucuronide.

Three incubations per recombinant protein, UGT1A6, UGT2A3 and UGT3A1, were set up with three different total protein concentrations; 0.1 mg/ml, 0.5 mg/ml and 1.0 mg/ml. Samples were taken for analysis at two time points, 0 min and 60. The areas for these masses were obtained from the chromatograms produced and Cl_{int} values calculated for each incubation (Table 5.14).

Protein	**UGT1A6**	**UGT2A3**	**UGT3A1**
0.1mg/ml	-19.998	-24.190	-15.985
0.5mg/ml	4.791	1.106	-1.422
1.0mg/ml	0.077	-1.377	-2.546

Table 5.14: Cl_{int} values per incubate per recombinant protein. Cl_{int} values were calculated for each incubate using two time points, 0 min and 60 min, and normalized for protein content. Each Cl_{int} = µl/min/mg of total protein. Any Cl_{int} values below zero indicate that no metabolism has taken place. Cl_{int} values were based on disappearance of parent oxazepam only.

Six of the nine incubates gave negative values, suggesting no metabolism has occurred and that oxazepam has not been conjugated to the glucuronidated form. Positive Cl_{int} values were obtained for the UGT1A6 0.5 mg/ml and 1.0 mg/ml and UGT2A3 0.5 mg/ml incubates (Table 5.14 and Figure 5.15), however while the chromatograms showed that there was a diminishing quantity of oxazepam, no glucuronide metabolite was detected.

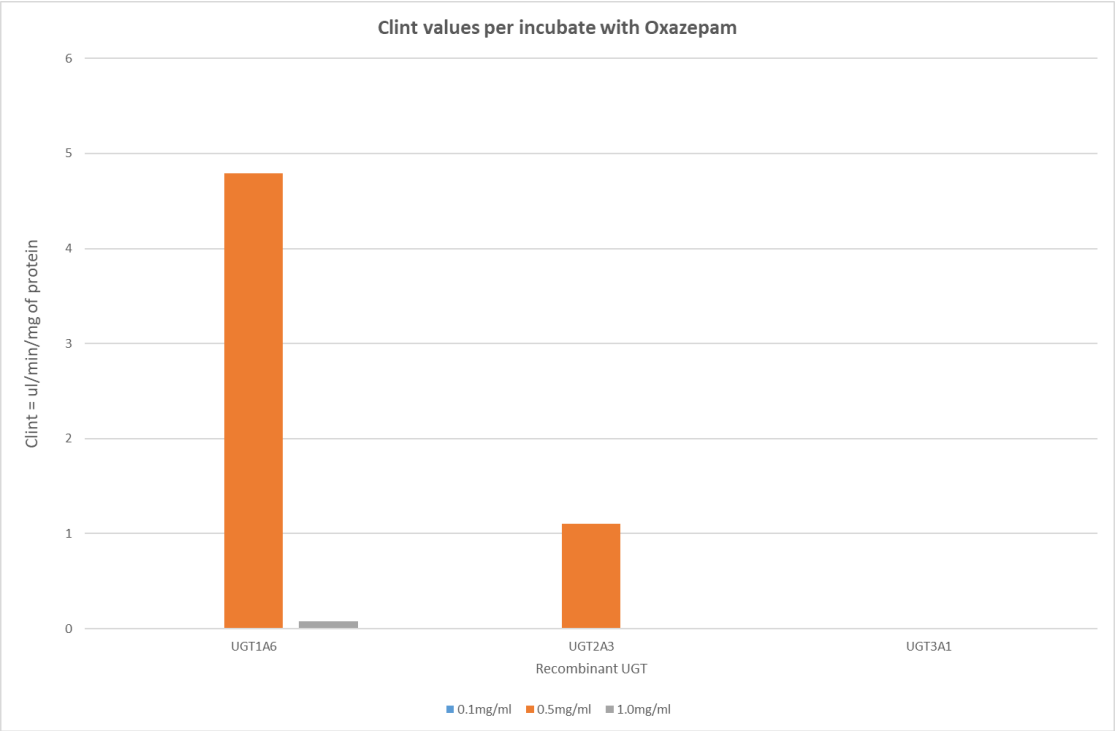
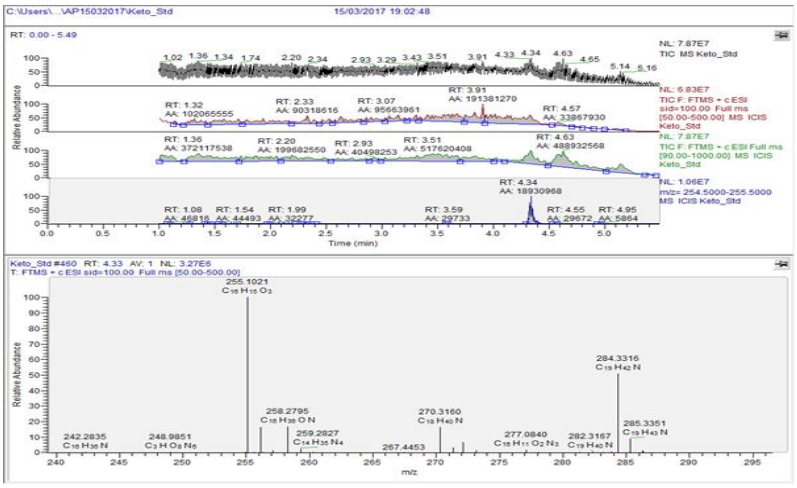


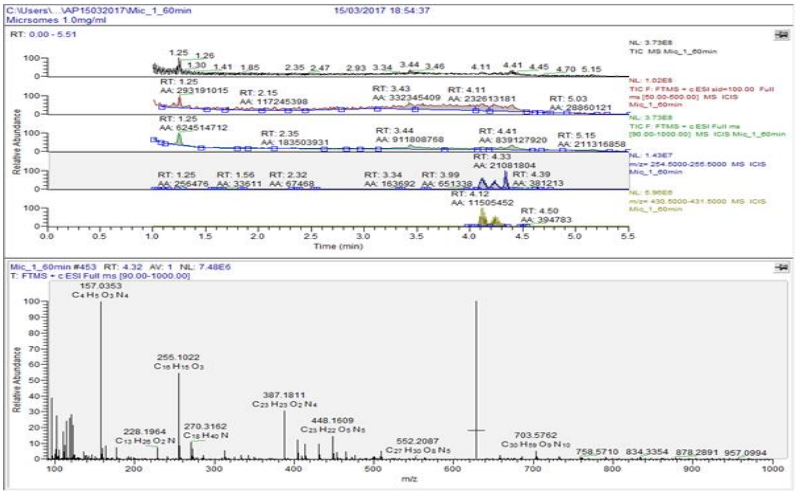
Figure 5.15: Graph displaying the intrinsic clearance (Cl_{int}) values per recombinant over three different protein concentrations. Samples were taken at two time points, 0 min and 60 min. The area mass from the chromatograms at the relative retention time was used to calculate the Cl_{int} value. A value below zero indicated no metabolism of oxazepam to the conjugated form, a value above zero is indicative of glucuronidation. Of the nine reactions, only two incubations produced a positive Cl_{int}; the UGT1A6 - 0.5 mg/ml and UGT2A3 - 0.5 mg/ml incubates. The remaining incubations all produced negative intrinsic clearance values.

5.4.3.3 Testing the functionality of the recombinant UGTs using ketoprofen as a substrate

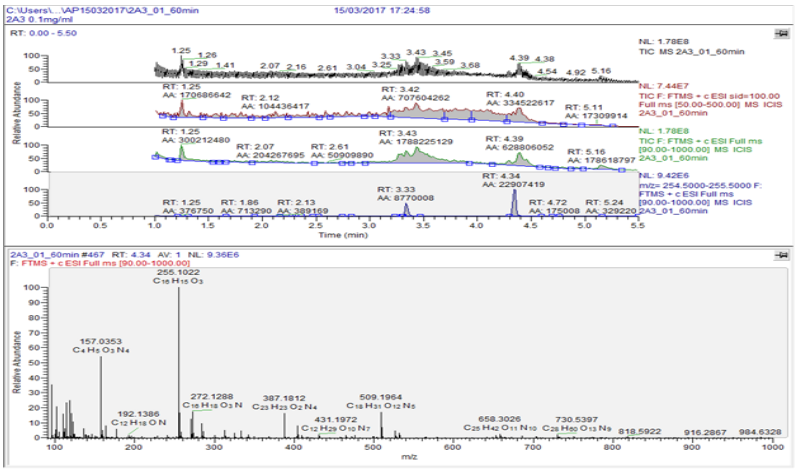
The initial analysis of a ketoprofen standard on the UPLC/Orbitrap system, gave an exact mass of 254.0937 g/mol and a retention time of 4.34 min (see Figure 5.16). No standard of Ketoprofen-glucuronide was available for analysis to determine retention time. The theoretical exact mass of ketoprofen-glucuronide is 430.1258 g/mol. Additional controls using human recombinant UGT1A6 and equine microsomes were used to supply confidence to the results obtained from the equine recombinant UGT studies.



A) Ketoprofen standard



B) 1.0mg/ml Microsomes incubated for 60minutes with ketoprofen



C) 0.1mg/ml of rUGT2A3 incubated for 60minutes with ketoprofen

Figure 5.16: Representative LCMS traces for metabolism studies using ketoprofen as substrate. LCMS traces for pure ketoprofen (A), 1.0mg/ml of microsomal protein (B) incubated for 60mins with ketoprofen and 0.1mg/ml recombinant equine UGT2A3 incubated for 60mins with ketoprofen.

Protein	UGT1A6	UGT2A3	UGT3A1	Human UGT1A6 (Corning)	Microsomes
0.1mg/ml	64.709	106.113	109.848	-8.134	0.308
0.5mg/ml	-3.441	16.334	2.563	-7.031	9.160
1.0mg/ml	9.729	8.132	2.082	-15.559	5.947

Table 5.17: Cl_{int} values per incubate per recombinant protein. Cl_{int} values were calculated for each incubate using two time points, 0 min and 60 min, and normalised for protein content. Each Cl_{int} = µl/min/mg of total protein. Any Cl_{int} values below zero indicate that no metabolism has taken place, this is based on the disappearance of the parent drug, ketoprofen.

For each of the recombinant equine UGTs, the 0.1 mg/ml incubates gave a Cl_{int} value above zero, suggesting metabolism of Ketoprofen to have occurred (Table 5.17). UGT1A6 produced a negative Cl_{int} for the 0.5 mg/ml incubate, and a positive Cl_{int} value for the 1.0 mg/ml and 0.1 mg/ml incubations. The 1.0 mg/ml Cl_{int} value is 6.6 times lower than the Cl_{int} for the 0.1 mg/ml incubation. UGT2A3 shows a high intrinsic clearance for the 0.1 mg/ml incubations, Cl_{int} = 106 µL/min/mg protein, but the values were reduced for the incubations with the higher concentrations of total protein content. UGT2A3 0.5 mg/ml produced a Cl_{int} of 16.33 µl/min/mg protein and the 1.0 mg/ml incubation produced a Cl_{int} of 8.13 µl/min/mg protein; this is a reduction of 6.49- and 13-fold respectively (Figure 5.18). UGT3A1 produced a very high Cl_{int} of 109 µl/min/mg of protein, over 43 times greater than the calculated Cl_{int} for the 0.5 µg/ml and 1.0 µg/ml incubates. In all three incubations for human recombinant UGT1A6, a negative intrinsic clearance value was produced. The equine microsomes produced negligible metabolism for the 0.1 mg/ml microsomal protein incubation while the 0.5 mg/ml and 1.0 mg/ml protein incubates indicated low levels of metabolism, with Cl_{int} values of 9.1 and 5.9 µl/min/mg protein respectively.

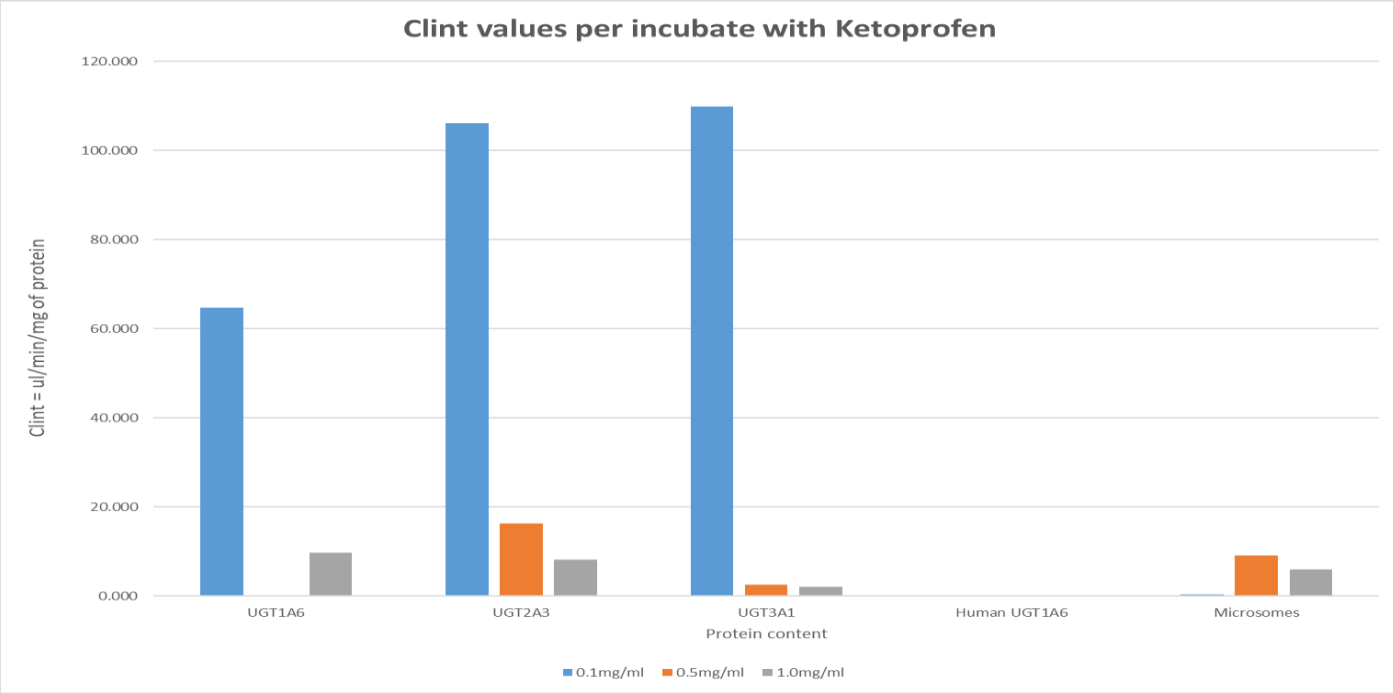
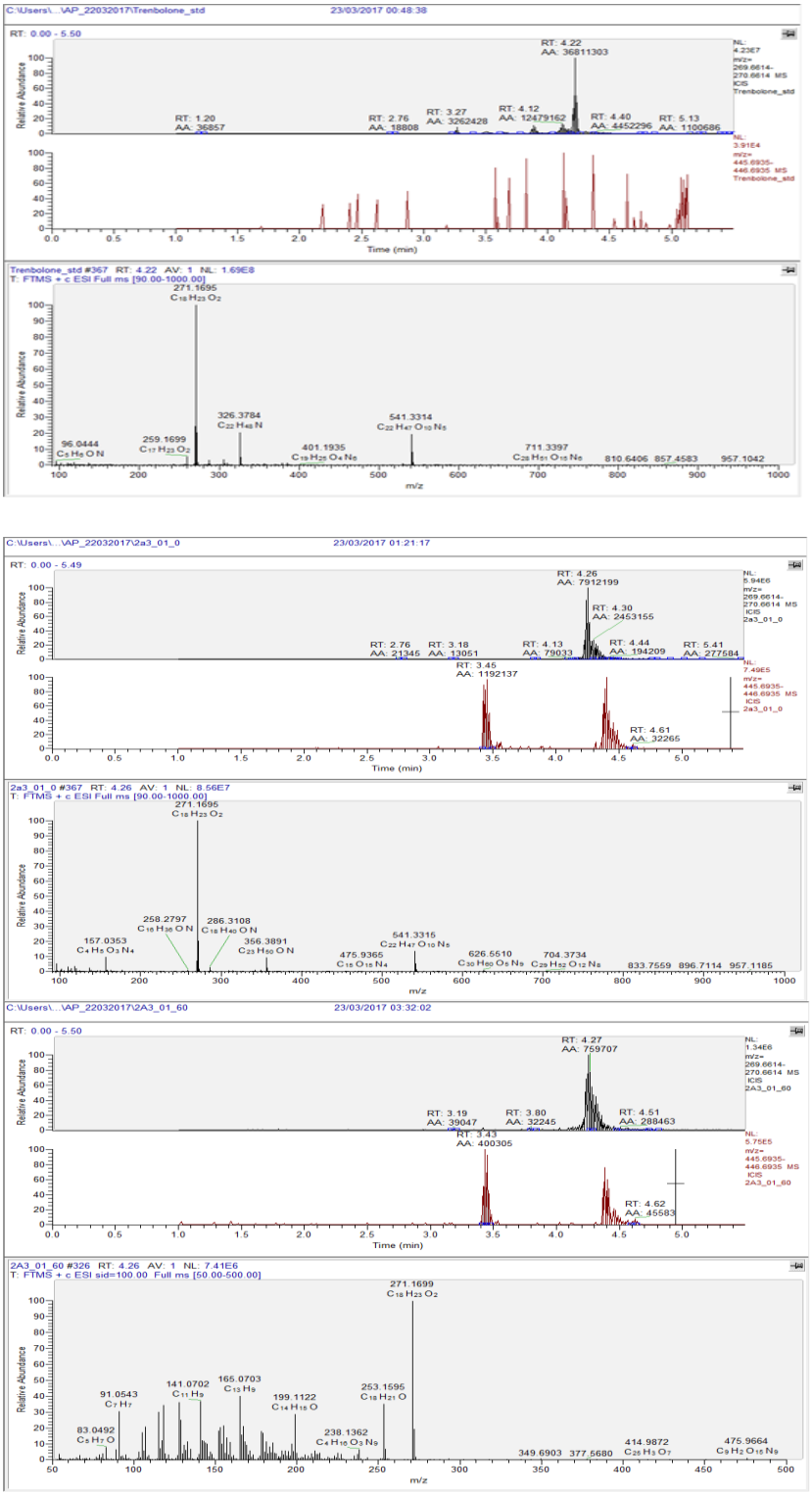


Figure 5.18: Intrinsic clearance values per incubation of total protein. Intrinsic clearance values below zero imply no glucuronidation of ketoprofen. Values above zero suggest ketoprofen has been glucuronidated. Equine UGT1A6 produced positive intrinsic clearance values for the 0.1 mg/ml and 1.0 mg/ml incubations. UGT2A3 displayed positive Cl_{int} values for all three incubations, intrinsic clearance values decreasing with an increase in total protein content. UGT3A1 displays a high level of intrinsic clearance in the 0.1 mg/ml sample and very low levels of intrinsic clearance in the 0.5 mg/ml and 1.0 mg/ml incubations. The Human UGT1A6 recombinant protein incubations all resulted in negative Cl_{int} values, the equine microsomes resulted in negligible intrinsic clearance in the 0.1 mg/ml incubation and low levels in the 0.5 mg/ml and 1.0 mg/ml incubates.

5.4.3.4 Testing the functionality of the recombinant UGTs using 17α-trenbolone as a substrate

The exact mass of 17α-trenbolone is 270.1695 g/mol, and analysis of a standard on the UPLC/Orbitrap system established the retention time for this compound to be 4.22 min (Figure 5.19). No standard of 17α-trenbolone-glucuronide was available for analysis. Additional controls using equine microsome and human UGT1A6 (Corning) were also performed.



A) Trenbolone standard

B) 0.1/ml recombinant equine UGT2A3 at 0 minutes with trenbolone

C) 0.1mg/ml of rUGT2A3 incubated for 60minutes with trenbolone

Figure 5.19: Representative traces for the analysis of trenbolone. LCMS traces for pure trenbolone (A), 0.1mg/ml of recombinant equine UGT2A3 (B) incubated for 60mins with trenbolone and 0.1mg/ml recombinant equine UGT2A3 incubated for 60mins with trenbolone.

Protein	UGT1A6	UGT2A3	UGT3A1	Human UGT1A6 (Corning)	Microsomes
0.1mg/ml	19.609	287.527	422.860	80.751	147.286
0.5mg/ml	22.795	59.114	12.397	83.862	62.996
1.0mg/ml	16.062	20.372	-31.506	13.692	28.058

Table 5.20: Clint values per incubate per recombinant protein. Clint values were calculated for each incubate using two time points, 0 min and 60 min, and normalised for protein content. Each Cl_{int} = µl/min/mg of total protein. Any Cl_{int} values below zero indicate that no metabolism has taken place, this is based on the disappearance of the parent drug, 17α-trenbolone.

With the sole exception of a single incubate, UGT3A1 1.0 mg/ml, all incubations analysed produced a positive intrinsic clearance value (Table 5.20). All three of the incubations for recombinant equine UGT1A6 resulted in small Cl_{int} values, with the 0.5 mg/ml sample producing the highest level of intrinsic clearance - Cl_{int} = 22.79 µl/min/mg protein. Equine UGT2A3 showed a decrease in clearance with an increase in the quantity of total protein, with the 0.1 mg/ml reaction producing a Cl_{int} of 287 µl/min/mg protein and the 1.0 mg/ml producing a Cl_{int} of 20.3 µl/min/mg protein (14-fold reduction in clearance). UGT3A1 displayed the highest level of intrinsic clearance of all the incubations (Figure 5.21), with a Cl_{int} of 422.89 µl/min/mg protein. The intrinsic clearance levels decreased with increasing protein concentration in the incubate, with the 1.0 mg/ml producing a negative figure, suggesting 17α-trenbolone was not conjugated in this incubation.

5.4.3.4.1 Additional controls using human recombinant UGT1A6 and equine microsomes

All incubations for the human recombinant UGT1A6 and the equine microsomes produced positive Cl_{int} values. The level of intrinsic clearance for the human UGT1A6 was similar in the 0.1 mg/ml and 0.5 mg/ml incubates, with Cl_{int} values of 80 and 83 µl/min/mg protein respectively. The 1.0 mg/ml incubation showed a 6 fold reduction in intrinsic clearance. The equine microsomes displayed decreasing intrinsic clearance with increasing protein content. The 0.1 mg/ml sample, Cl_{int} = 147 µl/min/mg protein, is 5.25 fold higher than the 1.0 mg/ml sample.

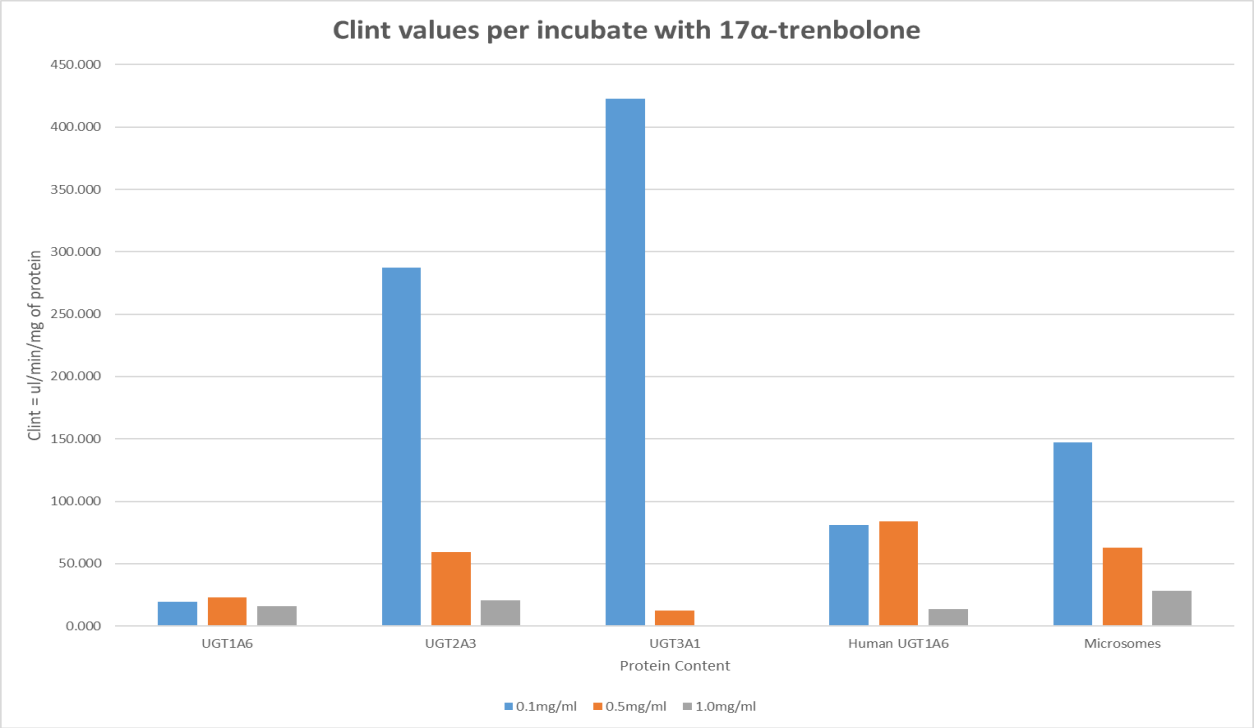


Figure 5.21: Cl_{int} values per incubation. Negative Cl_{int} values indicate no metabolism has taken place, positive Cl_{int} values indicate the substrate has been glucuronidated. A single incubation, UGT3A1 1.0 mg/ml, produced a negative value. Equine UGT1A6 shows a low intrinsic clearance, with a mean Cl_{int} of 19 µl/min/mg protein. UGT2A3 and UGT3A1 both show a decreasing level of intrinsic clearance with increasing protein content. The controls of human recombinant UGT1A6 and equine microsomes produced positive intrinsic clearance values. Human UGT1A6 resulted in similar levels in the 0.1 mg/ml and 0.5 mg/ml incubations, and a 6 fold reduction in the 1.0 mg/ml sample. The equine microsomes show decreasing intrinsic clearance with increasing protein content, with the Cl_{int} value dropping from 147 µl/min/mg protein in the 0.1 mg/ml incubate to 28 µl/min/mg protein in the 1.0 mg/ml incubate.

5.5 Discussion

UGTs are a major component of phase II metabolism, conjugating glucuronic acid to a target substrate in order to alter substrate polarity and aid elimination from the circulatory system via bile and urine (Jancova et al., 2010, Mackenzie et al., 2005, de Wildt et al., 1999, Gibson and Skett, 2001).

Knowledge of this process is advantageous to the pharmaceutical industry, as understanding the ADMET parameters of a compound is critical to drug development. *In vitro* systems are a cost effective method of screening numerous compounds to determine these parameters (Ekins et al., 2005, Zhang et al., 2012), enabling the identification of novel non-target and drug-drug interactions (Bowes et al., 2012), creating efficiency savings. *In vitro* systems also enable investigations into the effect of SNPs on metabolism. The development of *in vitro* systems also complements the concept of the 3Rs, the refinement, reduction, and replacement of animals in research (Guhad, 2005). *In vitro* analytical systems refine research of metabolism and lead to a reduction in the number of *in vivo* studies. It also enables the study of designer drugs, which have unknown toxicological effects, without any discomfort to the animal (Guhad, 2005). Such tools are available in humans and increasingly so in canines (McGinnity and Riley, 2001, Zhang et al., 2012, Soars et al., 2001), but have yet to be developed in other key species.

Current research in equines focuses on the use of microsomes which, although highly informative, requires regular collection of equine tissue. The creation of recombinant UGT *in vitro* systems will be novel and of industrial value, whilst fulfilling the requirements set out by the 3Rs (Guhad, 2005). To create a recombinant system, the UGTs need to be stably expressed and the functionality of proteins established and validated.

5.5.1 Determining the optimal concentration of Geneticin

Geneticin (G418), an aminoglycoside commonly used as a selective agent for eukaryotic cells, interferes with the 80S ribosome and protein synthesis function in eukaryotic cells and, through the incorporation of resistance genes into mammalian expression vectors, can be used to select for cells successfully transfected with the cloned gene of interest. This study used the mammalian expression vector pcDNA^{TM}3.1 (see appendix B) which expresses the bacterial aminoglycoside phosphotransferase gene which is derived from Tn5 (ThermoFisher Scientific) the Geneticin resistance gene. The first step was to determine if our chosen cell line, HEK293, was sensitive to Geneticin.

5.5.1.1 Why perform a kill curve?

The purpose of performing a kill curve was to demonstrate that the HEK293 cell line was susceptible to the antibiotic Geneticin, and to determine the optimum concentration to allow stable colonies of transfected cells to establish.

The supplier of the Geneticin (Sigma-Aldrich) suggest a concentration range of 300-500 µg/ml was suitable for animal cell lines. Published work on the cloning of UGTs using Geneticin G418 as the selective antibiotic has suggested a range of concentrations can be used. The selection of cloned UGTs in transfected Chinese hamster fibroblast cells has used concentrations ranging from 100 µg/ml to 1 mg/ml (Soars et al., 2001, Forsman et al., 2000). In studies using HEK293 cells, concentrations of Geneticin G418 ranging from 400 µg/ml to 1 mg/ml have been reported (Barbier et al., 2000, Chen et al., 2011a). A review of the published studies suggested a possible working concentration ranging from 100 µg/ml to 1 mg/ml. Therefore, it was important to establish the concentration of Geneticin required for the stable transfection of HEK293 using the expression vector of choice, pcDNA^{TM}3.1, for this particular study. The initial assay covered the full range of concentrations, 0 µg/ml (negative control) to 1000 µg/ml. Two assays were set up with Geneticin in

excess, 2000 μg/ml and 5000 μg/ml, to act as positive controls ensuring that full cell death will occur in the presence of this antibiotic.

5.5.1.2 What concentration of Geneticin was chosen and does this support previous work?

It was important to perform a negative control for this study to establish whether any of the reagents interfered with the establishment of the cell line and the proliferation of the cells, but also to determine at what rate natural cell death occurs. Cell death did occur in the negative control during the 17 day period (Figure 5.6a), but this appeared to be from overcrowding of cells and insufficient nutrients to maintain them. Positive controls, where Geneticin was in excess, are also required to ensure that the selection antibiotic is capable of causing complete cell death and that the untransfected HEK293 cells did not acquire resistance. Geneticin at concentrations of 2000 μg/ml and 5000 μg/ml showed cells were not able to establish and proliferate.

It is important when choosing the optimal concentration of Geneticin to select the concentration at which total cell death has occurred by day seven. This allows transfected cells to establish and proliferate and ensures untransfected cells are not maintained. The concentration range of 500 μg/ml to 1 mg/ml resulted in complete cell death by days five to seven. However, a visual inspection of the wells showed that live adherent cells were still present at day 5 for the 750 μg/ml and 1 mg/ml incubations. As such, it was evident that the lower concentration of Geneticin, 500 μg/ml, would be suitable to permit the successful adherence and proliferation of transfected cells whilst ensuring untransfected HEK293 would not be present five days after Geneticin administration.

The kill curve results support the information from the manufacturer that a concentration in the range of 300-500 μg/ml would be suitable for the selection and maintenance of transfected animal cells. This also corresponds to published work where the chosen expression vector utilised was pcDNA^{TM}3.1. HeLa cells transfected

with pcDNA^{TM}3.1-CD14, a human antigen, were stably maintained using 500 µg/ml (Ning and Tang, 2012). Expression of human UGT1A3 was achieved in Chinese hamster lung cells using 400 µg/ml of Geneticin (Chen et al., 2005). These studies provide confidence that the concentration of Geneticin selected for the HEK293-pcDNA^{TM}3.1 expression system was suitable.

5.5.2 Using western blots to confirm the presence of recombinant UGTs in the HEK293 expression system

5.5.2.1 Testing for the presence of recombinant equine UGT1A6

The presence of the equine recombinant UGT1A6 was tested for using an anti-UGT1A6 polyclonal antibody, raised in rabbits, against the C-termini of human UGT1A6. The manufacturers state that this antibody shows cross reactivity in sheep, canines, horse, and mouse, with publications confirming its successful use for the detection of murine UGT1A1 only (Bortolussi et al., 2014). Recombinant human UGT1A6 was used as a positive control and following western blot analysis a protein of ~78kDa was detected. There were two bands detected in the cell lysate from the equine UGT1A6-transfected HEK293 cells; one at 50kDa, which is the predicted size of the equine UGT1A6 protein and one <50kDa in size.

5.5.2.1.1 Why was there a secondary band in the transfected HEK293 lysate?

Polyclonal antibodies recognise multiple epitopes, and as such any protein with sufficient sequence similarity to the epitope recognised by the antibody could potentially be detected resulting in non-specific binding. There are a multitude of reasons for a high level of back ground noise: non-specific binding may result from the presence of too much primary antibody, or due to the wash step to remove unbound primary antibody being insufficient. Recognition of multiple epitopes means if a degraded product or truncated product is produced, the antibody would be able to recognise these and produce bands of different sizes.

5.5.2.1.2 Do the results of the western blot confirm the presence of UGT1A6?

Anti-UGT1A6 has been raised towards the C terminus of the UGT1A6 protein in humans, which is the portion of the protein shared between all the UGT1s. Figure 5.22 shows there is 100% sequence similarity between the epitope regions of the C-terminal of the human UGT1s recognised by the antibody, which means this particular antibody could recognise every human UGT1 isozyme. Figure 5.23 shows the sequence similarity between the epitope recognised by the anti-UGT1A6 antibody and the equine UGT1A6 amino acid sequence (84%). If all equine UGT1s shared this region this antibody would indiscriminately detect all equine UGT1s.

Given this information, due to the complexity of the UGT1 locus, we cannot conclusively state that what is detected is recombinant equine UGT1A6. Given the results of the western, we can say that a recombinant UGT1 may be present in the lysate. To confirm which isozyme of UGT is present, an N-termini specific antibody would need to be raised towards equine UGT1A6 and tested.

```
Epitope          ----------------------APHLRPAAHDLTWYQYHSLDVIGFLLAV
UGT1A1           DRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAV
UGT1A10          DRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAV
UGT1A4           DRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAV
UGT1A6           DRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAV
UGT1A8           DRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAV
UGT1A9           DRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAV
                                       ****************************

Epitope          VLTVAFITFKCCAYGYRKCLGK-------------
UGT1A1           VLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
UGT1A10          VLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
UGT1A4           VLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
UGT1A6           VLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
UGT1A8           VLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
UGT1A9           VLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
                 ********************
```

Figure 5.22: Alignment of human UGT1 sequences and the antibody's immunogen. Alignment of multiple UGT1 isoforms and the immunogen sequence recognised by the antibody. This displays the sequence similarity between the UGT1s C-terminus to be 100% identical, as such the antibody is unable to distinguish between the human isozymes.

```
Epitope          ----------------------------------------APHLRPAAHD
Equine UGT1A6    VINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHD
                                                         **********

Epitope          LTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGK-----------
Equine UGT1A6    LTWYQYHSLDVIGFLLAVVLGVAFIVYKSCAFGFRKFFGKKGRVKKSHKSK
                 ******************** ****.:*.**:*:** :**
```

Figure 5.23: Alignment of antibody immunogen sequence to equine *UGT1A6* sequence. This alignment shows an 84% sequence similarity between the equine amino acid sequence and the epitope recognised by the antibody.

5.5.2.2 Confirming the presence of recombinant equine UGT2A3 by western blot

The polyclonal antibody used to detect the presence of recombinant equine UGT2A3 was raised in mouse against the full-length human *UGT2A3* sequence. According to the manufacturer the antibody has not been tested for cross reactivity in any other species. Given the antibody is raised to the full-length sequence, theoretically if there was enough sequence similarity between human and equine *UGT2A3* then the polyclonal antibody should recognise the equine sequence, with potentially reduced binding affinity. A sequence alignment (Figure 5.24) between the antibody immunogen, human *UGT2A3* and the equine *UGT2A3* sequences shows 74.57% of the amino acids are conserved.

The cell lysate from HEK293 cells transfected with equine UGT2A3 did not show a band at 50kDa, but bands were observed at >100kDa. This suggests that the HEK293 cells are producing a protein with sufficient sequence similarity for the UGT2A3 antibody to bind. The negative control (untransfected cells) shows multiple bands, indicative of non-specific binding. Interestingly, the antibody failed to detect the positive control, human UGT2A3. The polyclonal nature of the antibody and that it was raised against human UGT2A3, alongside the high background in the negative, makes this failure difficult to explain. One potential reason is that the boiling step in the protocol was insufficient and that the protein did not fully denature for the antibody to be able to detect enough epitopes for a positive result. Further tests and optimisations are required to conclusively prove whether this antibody will or will not recognise equine UGT2A3. The failure of the antibody to detect equine UGT2A3 in conjunction with a failure of the positive control means we are unable to conclusively state that the antibody does not cross react with equine UGT2A3 sequences. In addition, we can cannot determine if the transfected HEK cells were expressing equine UGT2A3.

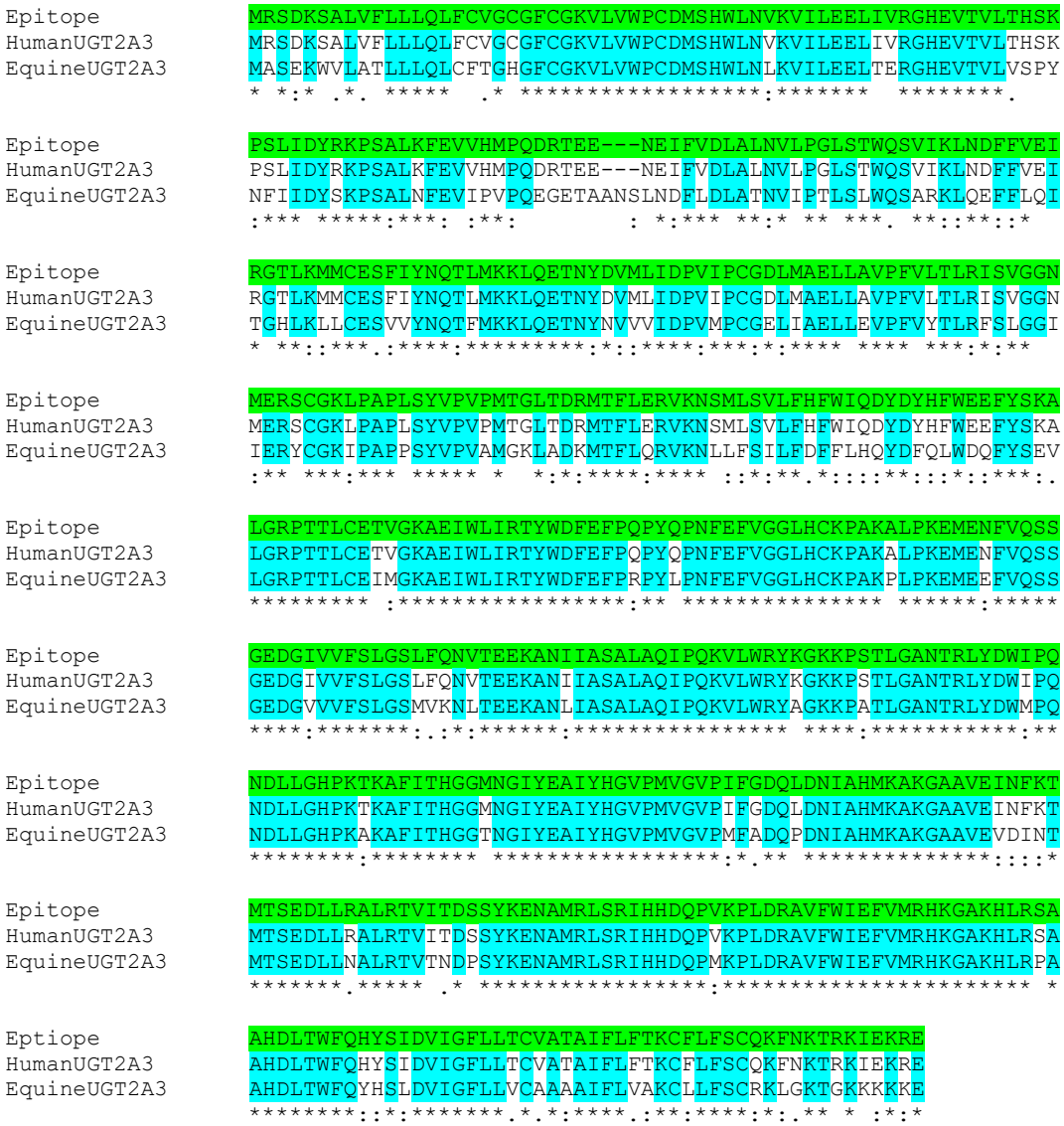


Figure 5.24: Alignment of antibody immunogen to human *UGT2A3* and equine *UGT2A3* amino acid sequences. The antibody is raised to the full length human sequence. There is a 74.57% sequence similarity between the equine *UGT2A3* sequence and the antibody and human *UGT2A3* sequence.

5.5.2.3 Confirming the expression of recombinant equine UGT3A1 by western blot

The polyclonal antibody used to detect the presence of recombinant equine UGT3A1 from transfected HEK293 cell lysate was raised in mouse against the full-length human UGT3A1 protein. According to the manufacturer the antibody has not been tested in horses, as such, recombinant human UGT3A1 was used as a positive control.

As a polyclonal antibody it is capable of recognising multiple epitopes, with sufficient sequence similarity between the target protein and the equine sequence, 72% (Figure 5.25). It should be able to detect parts of the equine sequence, with reduced binding affinity. A polyclonal antibody may also be able to recognise non-UGT proteins with similar epitopes which may in part explain the results obtained from the negative control of cell lysate from un-transfected HEK293 cells. Non-specific binding may in part be due to truncated or degraded protein or the use of too much primary antibody. The transfected cell lysate, containing the equine UGT3A1, showed non-specific binding, but no bands were present in the 50kDa region. This would suggest that no recombinant equine UGT3A1 was produced within the HEK293 expression system. However, as the positive control also failed to yield a band, there is no confidence that the negative result in the transfected lysate was due to no or low expression and may be due to ineffective binding of the polyclonal antibody.

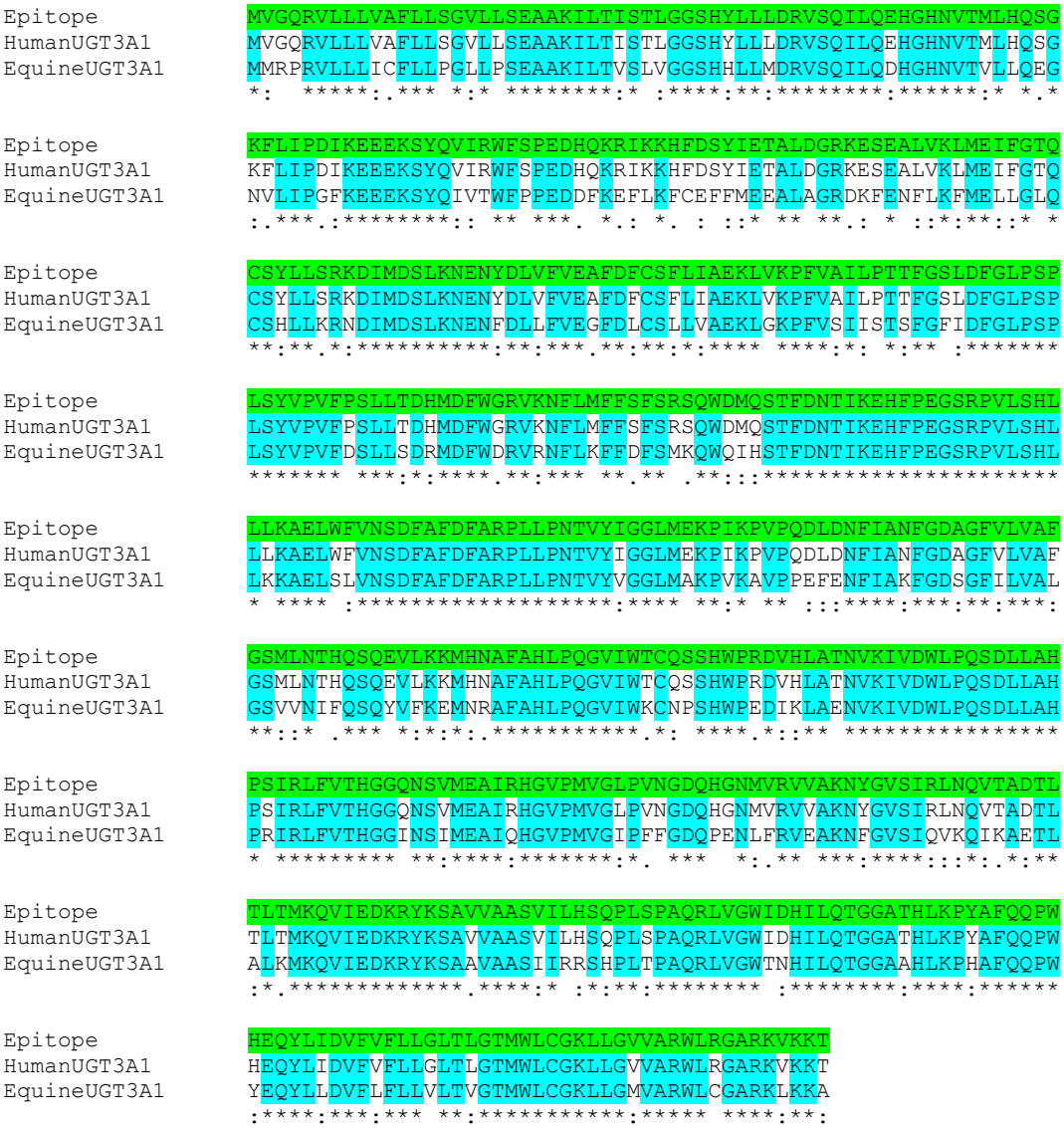


Figure 5.25: Alignment of the antibody immunogen against the human *UGT3A1* and equine *UGT3A1* amino acid sequences. Between the antibody, human *UGT3A1* sequence and the equine sequence, the shared homology was 72%.

5.5.3 Functional assays – Determining whether the recombinant enzymes are functionally active

Studying the genetic sequences of the UGT enzymes is informative regarding syntenic investigations, identifying orthologues, hypothesising functional role based on sequence and structural similarities. It is also useful to look for breed specific SNPs and to hypothesise potential implications of the SNPs and to assess the tissue specific expression profiles of distinct UGT isozymes. However, genetics has its limitations owing to the high level of sequence similarity between isozymes due to the sharing of exons; identifying the precise isozyme isolated requires functional studies. For example, in humans, paracetamol is known to be selectively conjugated by UGT1A6 (Bock et al., 1994). If the equine sequence was thought to be *UGT1A6* but also showed a high level of sequence similarity with *UGT1A10*, with the resulting protein not conjugating paracetamol (Stingl et al., 2014), then the ability of the recombinant protein to metabolise paracetamol would determine which isoform had been isolated from the equine genome. This is made more complicated by the fact that drugs can be metabolised by multiple isozymes. In this example, in humans, paracetamol is also conjugated by UGT1A1, UGT1A9 and UGT2B15 (Stingl et al., 2014).

The aim of the functional studies was to provide evidence that the recombinant equine UGTs that have been correctly synthesised and that any post-translationally modifications in the HEK293 expression system have resulted in functional proteins. In order to establish which isozyme was isolated and create a substrate profile for the recombinant UGT and produce a full kinetic profile of substrate metabolism calculating the intrinsic clearance of a drug with the rUGT, additional studies are needed. Given the time limitations and availability of equipment with the industry partners, four compounds were selected to be tested in the *in vitro* systems.

5.5.3.1 Are the recombinant equine UGTs functionally active with morphine as substrate?

Morphine is an opioid analgesic used both in humans and animals to treat acute and chronic pain (Ohno et al., 2008, Stone et al., 2003) and was accordingly selected for this study based on its wide usage. Studies in humans have identified two metabolites: morphine-3-glucuronide (M3G) and morphine-6-glucuronide (M6G) (Coffman et al., 1997).

The preferred metabolic pathway in human *in vivo* studies show 60% of clearance to be by conversion of morphine to M3G (Ohno et al., 2008). Only 5-10% is processed to M6G (Coffman et al., 1997), which has been proven to have a more potent analgesic effect (600 times) than morphine itself (Stone et al., 2003, Ohno et al., 2008, Radominska-Pandya et al., 2005a).

In vitro studies of human UGTs have found that UGT2B7 is the major enzyme to be involved in glucuronidation of morphine (Stone et al., 2003). This particular isozyme has been demonstrated to convert morphine to both M3G and M6G (Coffman et al., 1997). Several other UGTs, UGT1A1, UGT1A3, UGT1A6, UGT1A8, UGT1A9 and UGT1A10, have also been implicated in morphine metabolism (Ohno et al., 2008), however the contribution of UGT1A1, UGT1A6 and UGT1A9 is minor, based on very low affinities (Stone et al., 2003). Rats and mice are only capable of conjugating morphine to the M3G glucuronide. In rats, two UGTs have been identified as being involved in this reaction; UGT2B1 is the major enzyme for morphine metabolism, whilst UGT1A1 exhibited low glucuronidation rates (King et al., 1997).

The calculation of the intrinsic clearance (Cl_{int}), which is used to calculate the quantity of parent drug, morphine, removed per unit time and normalised for total protein content, was performed for each rUGT incubation. Of the nine incubations performed four yielded positive Cl_{int} values, suggesting metabolism of morphine was occurring in incubates; rUGT1A6 - 0.1 mg/ml, UGT2A3 - 1.0 mg/ml and for UGT3A1, both 0.1 mg/ml and 0.5 mg/ml incubates. Interestingly, a positive Cl_{int} value was seen in UGT1A6 and UGT3A1 in the incubations containing the smallest amount of

protein. Logically, more metabolism would be expected in incubates containing the larger quantity of total protein, but there was no evidence for this in the 1.0 mg/ml samples.

By looking at the chromatograms, for each sample, if metabolism were to have occurred, then chromatographic peaks would have shown an abundance of morphine at the 0 min time point and a decreased peak at the 60 min time point, whilst seeing the opposite for the glucuronidated metabolite. As morphine can be conjugated in two positions, a peak with exact mass of 462.1768 g/mol, would be expected at either 2.11 min or 2.68 min depending on the form, M3G or M6G, produced. A visual inspection of the chromatograms showed no peaks of the correct mass and retention time for the conjugate in any incubates.

The three theories pertaining to these are thus: 1) No metabolism was taking place and positive Cl_{ints} are false positives. 2) Metabolism took place at such low levels that the glucuronides were not detectable. 3) No metabolism was seen in the incubates containing 1.0 mg/ml of total protein due to some confounding factor inhibiting the metabolic reaction. It is possible that these recombinant equine UGTs have little affinity for morphine; it is not reported to be metabolised by UGT2A3 or UGT3A1 in humans and at low levels by UGT1A6 (Ohno et al., 2008). It is therefore plausible that an alternative equine UGT is the major glucuronidating enzyme for morphine.

5.5.3.2. Using oxazepam as substrate, can any glucuronide formation be detected with the recombinant equine UGTs?

Oxazepam belongs to the benzodiazepine class of drugs. Benzodiazepines are a widely prescribed drug due to their properties ranging from sedative, muscle relaxant, and anti-convulsant to anti-anxiety (Fu et al., 2010, Marland et al., 1999). When administered, oxazepam will readily and almost completely conjugate and it is also worth noting that oxazepam is a metabolite of diazepam (Scarth et al., 2011).

There are two stereo-isoforms of oxazepam, S-oxazepam and R-oxazepam. Human *in vitro* studies showed UGT2B15 to be the main contributing UGT isozyme for the glucuronidation of S-oxazepam. UGT2B7 and UGT1A9 have also been implicated as the main enzymes for the metabolism of R-oxazepam. In addition UGT1A1, UGT1A6 and UGT1A7 have also been implicated in oxazepam glucuronidation, although these show considerably reduced metabolism (Court et al., 2002).

Based on the calculation of intrinsic clearance (Cl_{int}), UGT1A6 does appear to be displaying affinity for metabolising oxazepam. The Cl_{int} value is higher in the 0.5 mg/ml total protein incubate (4.79 µl/min/mg of protein) compared to the 1.0 mg/ml protein incubate (0.07 µl/min/mg of protein), which is so low as to be considered not to be contributing to the conjugation of oxazepam. UGT2A3 also resulted in a very small Cl_{int} value in one incubate only, suggesting minor contribution of this enzyme to metabolism. All intrinsic clearance values for UGT3A1 were negative, implying no metabolism of oxazepam occurs via this isozyme, which was expected because this is not a known substrate of the human UGT3 family.

The availability of standards for oxazepam and oxazepam-glucuronide enabled these compounds to be profiled for mass and retention times on the UPLC/mass spectrometer system. This enabled the examination of the spectra and chromatograms for incubates where the Cl_{int} values suggest metabolism to be occurring and look for the oxazepam-glucuronide metabolite. None of the chromatograms showed a peak of the correct exact mass or retention time. Taking these two pieces of evidence together, it suggests that some glucuronidation of oxazepam may possibly be occurring in incubations. However, the levels of glucuronidation formation are too low for the UPLC/orbitrap system to detect or there is also the possibility that ion suppression is preventing the detection of the glucuronidated form.

5.5.3.3 Ketoprofen as a substrate to determine functionality of the recombinant proteins

Ketoprofen is a carboxylic acid compound belonging to the non-steroidal anti-inflammatory (NSAID) class of therapy drugs (Kuehl et al., 2005). NSAIDs are widely used in the treatment of pain and inflammatory conditions, such as joint problems and soft tissue damage, particularly in humans, but it is used widely in veterinary practice, including in the treatment of lameness (Kuehl et al., 2005, Terrier et al., 1999).

Studies performed in humans identified three contributing UGTs, UGT1A3, UGT1A9 and UGT2B7 (Sakaguchi et al., 2004). A look at the metabolism of ketoprofen in rats identified UGT2B1 to be responsible for the conjugation to a glucuronide (Terrier et al., 1999). With this knowledge, it was not expected that glucuronidation would be observed with the equine recombinant proteins, so in addition to analyses with the rUGTs, incubations were also performed with human rUGT1A6 and equine microsomes. It was thought that some level of metabolism would be detected in equine microsomes because they contain the full complement of expressed UGTs. A positive result for metabolism in equine microsomes would confirm that the reaction conditions were sufficient for glucuronidation to take place, but a negative in the recombinant UGT systems would indicate the recombinants either display no affinity for ketoprofen or are not functional.

The human rUGT1A6 clearly displayed no evidence of metabolism in any of incubates. The microsome intrinsic clearance values in all three incubates were very low, thus implying some conjugation of ketoprofen may be taking place. The Cl_{int} values for rUGT1A6, rUGT2A3 and rUGT3A1 showed results which are the opposite of what would be expected. High levels of metabolism occurred in the lowest, 0.1 mg/ml, protein incubates. For rUGT3A1 the Cl_{int} value was 52 times higher in the 0.1 mg/ml than the 1.0 mg/ml protein incubation. For rUGT2A3 there was 13 times more conjugation occurring in the 0.1 mg/ml versus the 1.0 mg/ml protein incubates. The

difference between the 0.1 mg/ml and 1.0 mg/ml incubates for rUGT1A6 was nine times higher in the 0.1 mg/ml incubation.

The results must be cautiously interpreted, given the high levels of metabolism in the 0.1 mg/ml protein incubates compared to the 1.0 mg/ml protein incubations. Due to no Ketoprofen-glucuronide standard being available, it was not possible to profile the exact mass and retention time. Therefore, it was not possible to interrogate the mass spectra thoroughly for the increased presence of the conjugated form. Without further study, it is not possible to conclude that this was a genuine result, indicative of glucuronidation, or whether it was an anomaly.

5.5.3.4 Is 17α-trenbolone glucuronidated by the recombinant UGTs?

17α-trenbolone (also called Epitrenbolone) belongs to the steroidal drug class. As an androgen, it can affect tissues in the body, the most desirable effect of which is protein building in skeletal muscle and bone (Teale and Houghton, 2010, Kicman, 2008); as such is regarded as a performance enhancer.

The precise human UGTs which glucuronidate 17α-trenbolone have yet to be identified, as such it was not possible to postulate which of the recombinant equine UGTs might metabolise the steroid. As the UGT2 family have been implicated in the metabolism of steroids, rUGT2A3 was investigated for its ability to metabolise 17α-trenbolone (Hum et al., 1999).

Only a single incubation produced a negative intrinsic clearance value, the rUGT3A1 1.0 mg/ml protein incubate. Analysis of rUGT2A3, rUGT3A1 and microsomes all showed parent drug to decrease with time, however higher Cl_{int} values were produced for the incubates with the least amount of total protein content; this is the reverse of what would be expected. This may suggest that there is potentially an inhibitory effect in the higher protein incubates. Both the human rUGT1A6 and the equine rUGT1A6 both show low Cl_{int} values, suggesting their contribution to the metabolism of 17α-trenbolone is minimal.

As n=1 in each instance and no authentic glucuronide standard was profiled it is not possible to say with confidence that glucuronidation was occurring. Whilst these results would appear to indicate that glucuronidation occurred, without further testing and suitable replicates, caution must be applied to their interpretation. There is also no evidence indicating that 17alpha-trenbolone is glucuronidated by human UGT1A6.

5.5.4 Study limitations of the functional assays

During the course of this study, several limiting factors were identified.

5.5.1 Successful production of recombinant equine UGT

There is no conclusive evidence that the protein sequences have been fully translated and post-translationally modified in the HEK293 cells. The western blot for rUGT1A6 suggests that a UGT1 may have been produced, however, as the antibody was generated to the C-termini there is no certainty on the exact isozyme detected, also a untransfected HEK lysate would need testing to confirm the presence/absence of endogenous UGT1A6. The western blots for rUGT2A3 and rUGT3A1 both failed to show the presence of recombinant protein, however as the control also failed, it could not be confirmed whether the problem was an issue with the antibody or as a result of a lack of protein being produced.

5.5.2 Authentic standard

Given this is the first study into equine UGTs, there is no authentic positive control as there is no fully functional, validated recombinant equine UGT available for use. Such an equine positive control, with known substrate specificity would allow us to effectively test the conditions of our assays.

5.5.3 Generating area mass values

The Xcalibur software automatically integrates peaks on the chromatogram to create an area mass value under the peak representing the mass of the compound being tested. However, phosphate buffered saline (PBS) creates a lot of background noise on the mass spectra and the peaks were not always 'clean'. As such it is highly probable that the software automatically integrated too many peaks. Owing to the amount of noise, a manual peak integration was not necessarily the better option.

5.5.4 Pilot study limitations

This was a pilot study to determine functionality of the protein with a potential substrate. A positive result from one incubate out of three per recombinant UGT is only suggestive of metabolism but does not provide confidence in the result owing to the lack of replicates, n=1.

5.5.5 Choice and number of drugs tested

Due to time constraints for the study and limitations on machine availability, only four drugs were tested. This represents a very small sample compared to the number of drugs commercially available. Continued testing of a wider number of compounds may have provided evidence to support the functionality of the rUGTs.

5.5.6 Quantity of recombinant protein in the incubates

The incubations were set up at three different total protein concentrations; 0.1 mg/ml, 0.5 mg/ml and 1.0 mg/ml. Protein concentration is a result of total protein extracted from the transfected HEK293 cell lysate. However, the quantity of recombinant UGT protein present within the total protein is unknown. It is possible that very little recombinant protein was present relative to the total protein.

An unknown issue is whether there is non-specific binding occurring with proteins from the HEK293 cell lysate. The greater the concentration of total protein the greater amount of non-specific binding that could occur, therefore you would see a decrease in Cl_{int} values with increasing protein content, as seen in these assays. A high Cl_{int} for the 0.1 mg/ml incubations compared to the 1.0 mg/ml incubations may be the result of enzyme saturation in the 1 mg/ml incubations. Enzyme saturation is theoretically possible in the recombinant system due to the expected high quantity of rUGT produced. However, it is unlikely to occur in microsomes as UGTs will be present at lower concentrations. This makes the results seen in these assays unlikely to be a result of enzyme saturation but testing a range of protein concentrations in microsomes would determine if this was occurring.

5.5.7 Glucuronide standards

A lack of authentic drug-conjugate standards in two of the assays meant that it was not possible to profile the exact mass and retention on the particular mass spectrometer used. This meant that the incubations could not be fully interrogated for the formation of glucuronides.

In this chapter we report the first steps in developing a functional *in vitro* recombinant equine UGT tool whereby phase 2 metabolites can be studied. Such *in vitro* tools will enable the assessment of drugs prescribed to horses to determine their pharmacokinetic profiles, identify drug-drug interactions, and investigate undesirable interactions. *In vitro* tools will also allow for the investigation into breed and gender differences and the impact of polymorphisms on UGT function. A recombinant *in vitro* tool box will further equine research at reduced cost and increased speed and effectiveness, with the ultimate aim of improving equine health and welfare. Additionally, it will enable anti-doping agencies to thoroughly investigate unidentified metabolites in cases of suspected drug abuse and tighten regulations were necessary.

Chapter 6: Discussion

6.1 Introduction

6.1.1 Why develop an in vitro system?

There is a great push in the pharmaceutical industry to increase our understanding of drug metabolism, to understand how a compound responds and the relation to adverse reactions (Milne et al., 2011). Factors such as age, gender, and ethnicity can all affect how a compound is metabolised (Gibson and Skett, 2001). The ultimate goal is the development of personalised medicine, with the creation of an *in vitro* system that allows the assessment of drugs to cater for the aforementioned variable factors (Lewis, 2005).

In vitro tools are becoming available for humans and animals such as dogs and cynomolgus monkeys (Soars et al., 2001, Troberg et al., 2015, Hanioka et al., 2006), but there is a dearth of basic bioscience in the horse that prevents progress in the development of such *in vitro* tools. Availability of such tools would be highly beneficial to improving equine health and welfare, profiling and detection of illegal drug use and new drug development.

6.2 Summary of results

Throughout the course of this project, five equine *UGTs* have been isolated, sequenced, assessed for homology to *UGTs* from other species and investigated for features characteristic to the *UGT* superfamily.

Using methods such as phylogenetic comparisons and synteny, which looks at the retention of neighbouring genes to relative positions on chromosomes throughout speciation, can help to infer the presence of genes on unannotated or poorly annotated genomes (Catchen et al., 2009, Stackebrandt and Goebel, 1994). Such comparative methods in conjunction with assessment of homology of these equine genes against characterised UGTs from humans, rats and mice identified the five

genes, *ENSECAG00000008900*, *ENSECAG00000010396*, *ENSECAG00000014362*, *ENSECAG00000020628* and *ENSECAG00000023519* as *UGTs*.

6.2.1 Equine ENSECAG00000023519 – orthologue of human UGT1A6

Equine *ENSECAG00000023519* was identified at the genetic level to be the orthologue of human *UGT1A6*. In humans UGT1A6 is known to metabolise a popular over the counter drug, paracetamol (Bock et al., 1994). Additionally it has been shown to make a minor contribution to the metabolism of morphine (Ohno et al., 2008, Stone et al., 2003). Side effects of morphine are common, most will cease once morphine use has been withdrawn, yet occasionally an individual may have a severe reaction (Glare et al., 2006). Understanding why this occurs will reduce incidence of adverse drug responses. In equines, as in humans, morphine is used to minimalize pain. It is a prohibited substance in horse racing as it could be abused to ensure an animal in pain races when it is not fit to do so. Morphine is produced from poppy seeds which can accidently enter the equine food chain. When poppy seeds are concentrated in food this can be metabolised by the horse to morphine, creating a false positive on race days (Kollias-Baker and Sams, 2002). It is therefore important to assess whether the equine orthologue can metabolise morphine, not only to detect use, but to increase the sensitivity of drug detection assays and refine knowledge of elimination times. Horses and other veterinary animals are all capable of adverse drug responses, further work looking at breed and gender differences may identify populations more susceptible to adverse responses.

6.2.2 Identifying members of the equine UGT2 family

ENSECAG00000020628 and *ENSECAG00000014362* were both identified as members of the *UGT2* family, which is comprised of two sub-families, *UGT2A* and *UGT2B*. *ENSECAG00000020628* was shown to be a *UGT2B*, specifically the equine

orthologue of human *UGT2B17*. *ENSECAG00000014362* was identified as a *UGT2A* member, specifically the equine orthologue of human *UGT2A3*.

The UGT2 family are an important family of enzymes for study. Members of this family are involved in the glucuronidation of steroids, UGT2B4, UGT2B7 and UGT2B17 show activity for estradiol (Kondo et al., 2017b), NSAIDS and to a lesser extent opioids (Schänzer, 1996, Kondo et al., 2017b, Coffman et al., 1997). In humans UGT2B17 displays a high affinity for the metabolism of testosterone (Sten et al., 2009), UGT2A3 is active against hyodeoxycholic acid, deoxycholic acid, chenodeoxycholic acid, and ursodeoxycholic acid, which are bile acids (Court et al., 2008). Steroids enhance muscle growth and performance and as such are prohibited for use in sports. They are a popular 'designer' drug, with small modifications made to the steroid structure which are sufficiently different as to avoid detection by accredited testing laboratories (Kazlauskas, 2010). It is a constant challenge to find and profile new designer drugs. Understanding their metabolism in humans, canines, and horse will be enhanced by development of the *in vitro* tool. As more recombinant UGT2s are produced a wider drug profile will be created, increasing the likelihood of post metabolite detection (Scarth et al., 2011, Teale and Houghton, 2010).

6.2.3 Identifying members of the UGT3 family

ENSECAG00000008900 and *ENSECAG00000010396* were both classified as members of the *UGT3* family. The evidence suggests that both *ENSECAG00000010396* and *ENSECAG00000008900* are the equine orthologues of human *UGT3A2*.

Considerably less is known about the UGT3s. They are less numerous than the UGT1s and UGT2s, with only two members identified in humans and mice and a single member in rats (Meech et al., 2012a). They are novel with regards to use of donor sugar, utilising either UDPGlcNAc or UDPGlc (Meech et al., 2012b). UGT3A2 in humans has been identified with having activity towards oestrogen and

bioflavones (MacKenzie et al., 2011) but no activity towards androgens. UGT3A1 also has activity towards oestrogens (Mackenzie et al., 2008), and given that both isozymes show activity towards oestrogens, it may be that the UGT3 enzymes are capable of metabolising a larger range of hormones, such as cortisol, this remains to be elucidated. More work remains to be done to elucidate the role of the UGT3s in humans or any species, but if they are predisposed to metabolising hormones then, as with the UGT2s, they may also be implicated in the metabolism of designer steroids.

6.2.4 Reviewing UGT nomenclature

As the number of UGTs isolated and identified from across the eukaryote and prokaryote kingdom increases, there is an argument to be made for reviewing and updating the UGT nomenclature.

The issues surrounding the nomenclature of enzyme superfamilies is not a novel concern. In 1987 the number and diversity of characterised CYPs and the plethora of names used resulted in a review and standardisation of nomenclature (McKinnon et al., 2008), which has been updated several times since.. A more descriptive nomenclature increases the accuracy of the software tools that annotate new genomes. Glycosyl hydrolase nomenclature has also been updated, originally classified on the basis of 300 sequences, a review was performed when the number identified reached 480 (Henrissat and Bairoch, 1996), with over 950 now identified.

As UGTs have been identified in yeast, bacteria, and plants, the issue of nomenclature has been previously addressed (Mackenzie et al., 1997, Mackenzie et al., 2005). A standardised nomenclature is now in use, with the root symbol 'UGT' followed by an Arabic numeral, a letter to denote subfamily and Arabic numeral to denote individual gene (Mackenzie et al., 1997). Whilst this has been accepted and newly identified UGTs are named using this standard nomenclature, there remains a requirement for improvement. During this project, genes have been identified in rats and mouse that have been given multiple aliases. For example, *UGT2B17* from

rat has multiple aliases of *UGT2B1*, *UGT2B10*, *UGT2B15,* and *UGT2B34*, likewise in the mouse *UGT2B17* is also known as *UGT2B5*. This makes it unclear as to whether a single gene has been isolated but ineffectively characterised or whether multiple genes have been identified. Reviewing the status of these genes will help to establish identities of newly isolated UGTs from across the species. Currently, predictive software tools for genome annotation will identify a *UGT*, but when it is not identified with certainty it is sequentially named chronologically (Mackenzie et al., 1997); this may in part explain why *UGT*s numbered in the 30s are being named when lower numbers do not exist.

6.3 How can this work be progressed?

6.3.1 Choice of expression vector and cell line

The work undertaken within this project encountered certain challenges. Firstly, the appropriate selection of expression vector, the vector pcDNA^{TM}3.1, was selected alongside the use of the HEK293 cell line as previous publications have successfully used these in the cloning and expression of recombinant UGTs (Radominska-Pandya et al., 2005a).

However, publications have reported that both the transient and stable transfections of HEK293 cells resulted in poor expression of the recombinant enzyme (Court et al., 2008). A separate study found the use of pcDNA^{TM}3.1 in Chinese hamster lung cells to produce sufficient quantities of recombinant protein for functional assays (Chen et al., 2005). Alternatively the baculovirus-insect cell system has been used in recombinant protein expression owing to the ability of the cells to perform most of the necessary post-translational modifications, fold proteins correctly and due to their high level of protein expression (Schneider and Seifert, 2010).

A separate method to consider is the creation of a fusion protein, which commonly uses a HIS-tag, the addition of several histidine residues added to one end of the

protein. This has been used in assays involving human UGT1s and can provide an alternative to working with 'dirty' cell lysates, as HIS-tagged proteins can be purified by metal-chelating chromatography (Kurkela et al., 2003).

In future work, it may worth considering using a different vector or cell system to optimise the production of recombinant equine UGT enzymes.

6.3.2 Difficulties in quantifying recombinant proteins

It is difficult to determine, within the HEK cell lysate, the quantity of recombinant protein present within the total protein lysate. The first step is to establish the presence of recombinant protein is by western blot. For detection of equine UGTs finding a suitable, validated antibody was problematic. Only a single UGT antibody, for UGT1A6, was available and validated for cross reactivity in equines. For UGT2A3 and UGT3A2, the antibodies available for use had only been tested in humans.

When successful, western blotting will detect a single protein in a mixture of proteins. It is a semi-quantitative method relative to a reference protein (Mahmood and Yang, 2012). The lack of antibodies for our recombinant enzymes meant we were unable to perform this test. However, there are limitations with a semi-quantitative method as it would only confirm with certainty that recombinant equine UGT was present but not absolute values.

One method by which absolute quantification is possible is via the use of mass spectrometry (Liebler and Zimmerman, 2013). Multiple-reaction monitoring mass spectrometry is a sensitive method for the precise quantification of specific peptide sequences, and this could be adapted for quantifying recombinant UGTs in cell lysates by measuring the amount of specific peptide sequence, unique to the UGT being expressed, in the total protein cell lysate (Liebler and Zimmerman, 2013). This would enable a comparison of expression systems to be made, to identify the most effective vector/cell line combination for the expression of recombinant equine UGTs.

6.3.3 Assessing functionality

Time limitations in this project meant only four drugs were used as potential substrates to assess the functionality of the recombinant UGTs. In the absence of a known substrate the functionality of the equine recombinant enzymes cannot be confirmed. Given the broad array of substrates associated with UGTs, continuing to test additional drugs may have led to a positive result.

Additionally, the positive control used, human UGT1A6, is known to metabolise serotonin. However, this was not available for testing in this project. Human UGT1A6 was chosen as a positive control as we had isolated the equine orthologue and would therefore be a suitable positive control for our assay conditions, but also confirm the substrate specificity of recombinant equine UGT1A6. Several human recombinant UGTs are commercially available, and it would be prudent in further work, to test these in order to find a positive control, preferably one with a wide substrate profile, which can be used to test the conditions and protocol of our assays.

Future work, particularly with the UGT3s needs to also focus on the donor sugar, as both human UGT3A1 and UGT3A2 use unique donor sugars (Meech et al., 2012b). Activity may not have been detected with the predicted equine orthologue of UGT3A2, as all assays were performed in the presence of UDPGA. Activity, or lack of, needs to be established with each of the donor sugars used by human UGTs, it cannot be assumed that all species and all UGTs use the same donor sugar. While each human UGT3 uses different donor sugars, in mice both UGT3A1 and UGT3A2 use UDP-Glc, with neither enzyme active in the presence of UDP-GlcNAc (Meech et al., 2012b).

6.4 The wider applications of this work

6.4.1 Applications to health and welfare of horses and the wider animal population

In humans, it is estimated that 6.5% of hospital admissions annually in the US are due to adverse drug reactions, of which 0.32% (in the region of 100,000) are fatal (Meyer, 2000). In the mid-20th century it was realised that genetics may play a part in an individual's response to a drug and so began the era of pharmacogenetics.

Advances in technology have brought science to the point where systems have been developed whereby the metabolism of a drug can be assessed prior to being delivered to the public. The use of *in vitro* tools facilitates this research, it enables the assessment of ethnicity and polymorphisms on drug metabolism as well as investigating the effect of drug-drug interactions (Meyer, 2000).

Whilst the use of *in vitro* tools is reasonably advanced in humans, it is lagging behind in our fellow mammals. Such tools are naturally extendable to the veterinary field in the treatment of farm and companion animals. Current drug metabolism studies involve large scale administration studies, where urine and blood samples are analysed, with microsomes or the use of other animals, often the zebrafish as models (Strähle et al., 2012). Health and welfare in animals is a complex and multifactorial issue (Niklason et al., 1999). *In vitro* systems will provide a more efficient method to investigate drug responses, assess for illegal doping and screen for novel and more effective medicines (Blomme et al., 2009).

6.4.2 Utility of in vitro tools to the equine sporting industry

The anti-doping agencies are fighting a constant battle to enforce the regulations set out by various regulatory bodies. In the United Kingdom, the British Horseracing Authority has a long list of regulations regarding the use of drugs in horses. There is a zero-tolerance policy in place for anabolic steroids, meaning that they cannot be given to a horse at any point in its racing lifetime

(http://www.britishhorseracing.com/regulation/anti-doping-medicationcontrol), with other drugs having to be withdrawn for a set period of time prior to racing. As such new 'designer' drugs are constantly being developed, and these are frequently modified steroids (Teale and Houghton, 2010). The development of *in vitro* tools will allow new drugs to be quickly screened to establish suitable withdrawal times in addition to profiling for detection from urine, blood, and hair. It will also enable the profiling of current drugs and increase the likelihood with which post-metabolic products are detected (Scarth et al., 2011).

In vitro tools will also provide a mechanism for the assessment of polymorphisms on the ability and rate at which a compound is metabolised. Population specific polymorphisms have been identified in humans and studies in canines have found breed specific drug responses (Ehmer et al., 2012, Fleischer et al., 2008). Therefore, the assessment of breed specific polymorphisms will be of benefit to the health and welfare of the animals, and of interest to the horse racing industry and veterinary profession. Breed specific responses may result in the metabolism of a drug in a thoroughbred being a different rate than in a Welsh pony.

6.4.3 Breed and gender specific expression and drug metabolism

As well as the impact of breed specific polymorphisms, breed specific expression can also affect rates of drug metabolism. Three breeds of cattle have been identified as expressing *UGT1A6* at different levels, which in turn resulted in differences in enzymatic activity (Giantin et al., 2008). Sexual dimorphism has been found in over 1000 genes from rats and mice (Waxman and Holloway, 2009), with studies in mice having shown gender specific expression of UGTs (Buckley and Klaassen, 2007). Alterations to enzymatic activity may alter the effectiveness of a drug, or time it takes to clear the drug from the body and as such may potentially have positive and negative effects on health and well-being, but also on withdrawal times prior to racing (Giantin et al., 2008, Sallovitz et al., 2002).

6.5 Concluding statement

There is much potential with this area of research. The creation of an *in vitro* system for UGTs is necessary to develop our understanding of phase II drug metabolism in the horse. Such tools will make the discovery and development of new drugs more cost effective and efficient. It may lead to a reduction in adverse drug responses, inform on gender and breed specific doses, potentially leading to gender and breed specific anti-doping regulations in addition to improved screening for drug abuse in the racing industry. The concept of an *in vitro* system should not be limited to the horse, whilst such tools are slowly becoming available in canines, they would be of scientific benefit to the development of health and welfare of all companion and livestock animals.

Chapter 7: References

ABDEL-MISIH, S. R. Z. & BLOOMSTON, M. 2010. Liver Anatomy. *The Surgical clinics of North America,* 90**,** 643-653.

AHN, K., BAE, J.-H., NAM, K.-H., LEE, C.-E., PARK, K.-D., LEE, H.-K., CHO, B.-W. & KIM, H.-S. 2011. Identification of reference genes for normalization of gene expression in thoroughbred and Jeju native horse(Jeju pony) tissues. *Genes & Genomics,* 33**,** 245.

AHN, S.-J., DERMAUW, W., WYBOUW, N., HECKEL, D. G. & VAN LEEUWEN, T. 2014. Bacterial origin of a diverse family of UDP-glycosyltransferase genes in the Tetranychus urticae genome. *Insect Biochemistry and Molecular Biology,* 50**,** 43-57.

AHN, S. J., VOGEL, H. & HECKEL, D. G. 2012. Comparative analysis of the UDP-glycosyltransferase multigene family in insects. *Insect Biochem Mol Biol,* 42**,** 133-47.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

ANDERSSON, H., KAPPELER, F. & HAURI, H.-P. 1999. Protein Targeting to Endoplasmic Reticulum by Dilysine Signals Involves Direct Retention in Addition to Retrieval. *Journal of Biological Chemistry,* 274**,** 15080-15084.

ANIELSKI, P. 2008. Hair analysis of anabolic steroids in connection with doping control—results from horse samples. *Journal of Mass Spectrometry,* 43**,** 1001-1008.

BARBIER, O., GIRARD, C., BRETON, R., BÉLANGER, A. & HUM, D. W. 2000. N-Glycosylation and Residue 96 Are Involved in the Functional Properties of UDP-Glucuronosyltransferase Enzymes†,‡. *Biochemistry,* 39**,** 11540-11552.

BERNAL, W., AUZINGER, G., DHAWAN, A. & WENDON, J. 2010. Acute liver failure. *The Lancet,* 376**,** 190-201.

BERNHARDT, R. & URLACHER, V. 2014. Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations. *Applied Microbiology and Biotechnology,* 98**,** 6185-6203.

BEST, C. 1934. The role of the liver in the metabolism of carbohydrate and fat. *Lancet,* 226**,** 1155-1277.

BINDA, C., MATTEVI, A. & EDMONDSON, D. E. 2002. Structure-Function Relationships in Flavoenzyme-dependent Amine Oxidations: A COMPARISON OF POLYAMINE OXIDASE AND MONOAMINE OXIDASE. *Journal of Biological Chemistry,* 277**,** 23973-23976.

BLOMME, E. A. G., YANG, Y. & WARING, J. F. 2009. Use of toxicogenomics to understand mechanisms of drug-induced hepatotoxicity during drug discovery and development. *Toxicology Letters,* 186**,** 22-31.

BOCK, K. W. 2003. Vertebrate UDP-glucuronosyltransferases: functional and evolutionary aspects. *Biochem Pharmacol,* 66**,** 691-6.

BOCK, K. W. 2010. Functions and transcriptional regulation of adult human hepatic UDP-glucuronosyl-transferases (UGTs): mechanisms responsible for interindividual variation of UGT levels. *Biochem Pharmacol,* 80**,** 771-7.

BOCK, K. W. 2016. The UDP-glycosyltransferase (UGT) superfamily expressed in humans, insects and plants: Animal□plant arms-race and co-evolution. *Biochemical Pharmacology,* 99**,** 11-17.

BOCK, K. W., SCHRENK, D., FORSTER, A., GRIESE, E.-U., MÖRIKE, K., BROCKMEIER, D. & EICHELBAUM, M. 1994. The influence of environmental and genetic factors on

CYP2D6, CYP1A2 and UDP-glucuronosyltransferases in man using sparteine, caffeine, and paracetamol as probes. *Pharmacogenetics and Genomics,* 4**,** 209-218.

BOGAERT, L., VAN POUCKE, M., DE BAERE, C., PEELMAN, L., GASTHUYS, F. & MARTENS, A. 2006. Selection of a set of reliable reference genes for quantitative real-time PCR in normal equine skin and in equine sarcoids. *BMC Biotechnology,* 6**,** 24.

BORTOLUSSI, G., ZENTILLIN, L., VANIKOVA, J., BOCKOR, L., BELLAROSA, C., MANCARELLA, A., VIANELLO, E., TIRIBELLI, C., GIACCA, M., VITEK, L. & MURO, A. F. 2014. Life-long correction of hyperbilirubinemia with a neonatal liver-specific AAV-mediated gene transfer in a lethal mouse model of Crigler-Najjar Syndrome. *Hum Gene Ther,* 25**,** 844-55.

BOSMA , P. J., CHOWDHURY , J. R., BAKKER , C., GANTLA , S., DE BOER , A., OOSTRA , B. A., LINDHOUT , D., TYTGAT , G. N. J., JANSEN , P. L. M., ELFERINK , R. P. J. O. & CHOWDHURY , N. R. 1995. The Genetic Basis of the Reduced Expression of Bilirubin UDP-Glucuronosyltransferase 1 in Gilbert's Syndrome. *New England Journal of Medicine,* 333**,** 1171-1175.

BOWES, J., BROWN, A. J., HAMON, J., JAROLIMEK, W., SRIDHAR, A., WALDRON, G. & WHITEBREAD, S. 2012. Reducing safety-related drug attrition: the use of in vitro pharmacological profiling. *Nat Rev Drug Discov,* 11**,** 909-922.

BRYKSIN, A. V. & MATSUMURA, I. 2010. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques,* 48**,** 463-465.

BUCKLEY, D. B. & KLAASSEN, C. D. 2007. Tissue- and gender-specific mRNA expression of UDP-glucuronosyltransferases (UGTs) in mice. *Drug Metab Dispos,* 35**,** 121-7.

BURCHELL, B., LOCKLEY, D. J., STAINES, A., UESAWA, Y. & COUGHTRIE, M. W. H. 2005. Substrate Specificity of Human Hepatic Udp-Glucuronosyltransferases. *Methods in Enzymology,* 400**,** 46-57.

BUSTIN, S. A., BENES, V., GARSON, J. A., HELLEMANS, J., HUGGETT, J., KUBISTA, M., MUELLER, R., NOLAN, T., PFAFFL, M. W., SHIPLEY, G. L., VANDESOMPELE, J. & WITTWER, C. T. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem,* 55**,** 611-22.

BUTCHER, N. J., BOUKOUVALA, S., SIM, E. & MINCHIN, R. F. 2002. Pharmacogenetics of the arylamine N-acetyltransferases. *Pharmacogenomics J,* 2**,** 30-42.

CAMPBELL. N.A, R. J. B. 2005. *Biology*, Pearson.

CAPUTI, L., MALNOY, M., GOREMYKIN, V., NIKIFOROVA, S. & MARTENS, S. 2012. A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *The Plant Journal,* 69**,** 1030-1042.

CATCHEN, J. M., CONERY, J. S. & POSTLETHWAIT, J. H. 2009. Automated identification of conserved synteny after whole-genome duplication. *Genome Research,* 19**,** 1497-1505.

CHEN, G., ZHAO, L., FENG, J., YOU, G., SUN, Q., LI, P., HAN, D. & ZHOU, H. 2013. Validation of Reliable Reference Genes for Real-Time PCR in Human Umbilical Vein Endothelial Cells on Substrates with Different Stiffness. *PLoS ONE,* 8**,** e67360.

CHEN, M., ZHANG, J., LI, N., QIAN, Z., ZHU, M., LI, Q., ZHENG, J., WANG, X. & SHI, G. 2011a. Promoter Hypermethylation Mediated Downregulation of FBP1 in Human Hepatocellular Carcinoma and Colon Cancer. *PLOS ONE,* 6**,** e25564.

CHEN, Y.-K., LI, X., CHEN, S.-Q. & ZENG, S. 2005. Heterologous expression of active human uridine diphosphate glucuronosyltransferase 1A3 in Chinese hamster lung cells. *World Journal of Gastroenterology : WJG,* 11**,** 118-121.

CHEN, Y., PATEL, N. A., CROMBIE, A., SCRIVENS, J. H. & MURRELL, J. C. 2011b. Bacterial flavin-containing monooxygenase is trimethylamine monooxygenase. *Proceedings of the National Academy of Sciences,* 108**,** 17791-17796.

CHOI, M. Y., PARTRIDGE, A. W., DANIELS, C., DU, K., LUKACS, G. L. & DEBER, C. M. 2005. Destabilization of the Transmembrane Domain Induces Misfolding in a Phenotypic Mutant of Cystic Fibrosis Transmembrane Conductance Regulator. *Journal of Biological Chemistry,* 280**,** 4968-4974.

CHOWDHARY, B. P. & RAUDSEPP, T. 2008. The horse genome derby: racing from map to whole genome sequence. *Chromosome Res,* 16**,** 109-27.

CIOTTI, M., MARRONE, A., POTTER, C. & OWENS, I. S. 1997. Genetic polymorphism in the human UGT1A6 (planar phenol) UDP-glucuronosyltransferase: pharmacological implications. *Pharmacogenetics,* 7**,** 485-95.

COFFMAN, B. L., RIOS, G. R., KING, C. D. & TEPHLY, T. R. 1997. Human UGT2B7 catalyzes morphine glucuronidation. *Drug Metab Dispos,* 25**,** 1-4.

COSTAS, I. 2006. Cytochrome P450 Expression in the Liver of Food-Producing Animals. *Current Drug Metabolism,* 7**,** 335-348.

COURT, M. H. 2001. Acetaminophen UDP-glucuronosyltransferase in ferrets: species and gender differences, and sequence analysis of ferret UGT1A6. *Journal of Veterinary Pharmacology and Therapeutics,* 24**,** 415-422.

COURT, M. H. 2013a. Canine cytochrome P450 (CYP) pharmacogenetics. *The Veterinary clinics of North America. Small animal practice,* 43**,** 1027-1038.

COURT, M. H. 2013b. Feline drug metabolism and disposition: pharmacokinetic evidence for species differences and molecular mechanisms. *The Veterinary clinics of North America. Small animal practice,* 43**,** 10.1016/j.cvsm.2013.05.002.

COURT, M. H., DUAN, S. X., GUILLEMETTE, C., JOURNAULT, K., KRISHNASWAMY, S., VON MOLTKE, L. L. & GREENBLATT, D. J. 2002. Stereoselective Conjugation of Oxazepam by Human UDP-Glucuronosyltransferases (UGTs): S-Oxazepam Is Glucuronidated by UGT2B15, While R-Oxazepam Is Glucuronidated by UGT2B7 and UGT1A9. *Drug Metabolism and Disposition,* 30**,** 1257-1265.

COURT, M. H. & GREENBLATT, D. J. 2000. Molecular genetic basis for deficient acetaminophen glucuronidation by cats: UGT1A6 is a pseudogene, and evidence for reduced diversity of expressed hepatic UGT1A isoforms. *Pharmacogenetics and Genomics,* 10**,** 355-369.

COURT, M. H., HAZARIKA, S., KRISHNASWAMY, S., FINEL, M. & WILLIAMS, J. A. 2008. Novel polymorphic human UDP-glucuronosyltransferase (UGT) 2A3: Cloning, functional characterization of enzyme variants, comparative tissue expression, and gene induction. *Molecular pharmacology,* 74**,** 744-754.

CUMMINGS, M., SARVESWARAN, J., HOMER-VANNIASINKAM, S., BURKE, D. & ORSI, N. M. 2014. Glyceraldehyde-3-phosphate Dehydrogenase is an Inappropriate Housekeeping Gene for Normalising Gene Expression in Sepsis. *Inflammation,* 37**,** 1889-1894.

DATTA, R., WAHEED, A., SHAH, G. N. & SLY, W. S. 2007. Signal sequence mutation in autosomal dominant form of hypoparathyroidism induces apoptosis that is corrected by a chemical chaperone. *Proceedings of the National Academy of Sciences,* 104**,** 19989-19994.

DE WILDT, S. N., KEARNS, G. L., LEEDER, J. S. & VAN DEN ANKER, J. N. 1999. Glucuronidation in humans. Pharmacogenetic and developmental aspects. *Clin Pharmacokinet,* 36**,** 439-52.

DELLINGER, R. W., GOMEZ GARCIA, A. M. & MEYSKENS, F. L. 2014. Differences in the Glucuronidation of Resveratrol and Pterostilbene: Altered Enzyme Specificity and Potential Gender Differences. *Drug metabolism and pharmacokinetics,* 29**,** 112-119.

DESHPANDE, D., HILL, K. E., MEALEY, K. L., CHAMBERS, J. P. & GIESEG, M. A. 2016. The Effect of the Canine ABCB1-1Δ Mutation on Sedation after Intravenous Administration of Acepromazine. *Journal of Veterinary Internal Medicine,* 30**,** 636-641.

DIMAIO KNYCH, H., MCKEMIE, D. & STANLEY, S. 2010. Molecular cloning, expression and initial characterization of members of the CYP3A family in horses. *Drug Metabolism and Disposition*.

DIMAIO KNYCH, H. K., DESTEFANO SHIELDS, C., BUCKPITT, A. R. & STANLEY, S. D. 2009. Equine cytochrome P450 2C92: cDNA cloning, expression and initial characterization. *Archives of Biochemistry and Biophysics,* 485**,** 49-55.

EDWARDS, R., DIXON, D. P. & WALBOT, V. 2000. Plant glutathione S-transferases: enzymes with multiple functions in sickness and in health. *Trends in Plant Science,* 5**,** 193-198.

EHMER, U., KALTHOFF, S., FAKUNDINY, B., PABST, B., FREIBERG, N., NAUMANN, R., MANNS, M. P. & STRASSBURG, C. P. 2012. Gilbert syndrome redefined: A complex genetic haplotype influences the regulation of glucuronidation. *Hepatology,* 55**,** 1912-1921.

EKINS, S., NIKOLSKY, Y. & NIKOLSKAYA, T. 2005. Techniques: application of systems biology to absorption, distribution, metabolism, excretion and toxicity. *Trends Pharmacol Sci,* 26**,** 202-9.

ELIAS, H. & BENGELSDORF, H. 1952. THE STRUCTURE OF THE LIVER OF VERTEBRATES. *Cells Tissues Organs,* 14**,** 297-337.

FAY, M. J., NGUYEN, M.-T., SNOUWAERT, J. N., DYE, R., GRANT, D. J., BODNAR, W. M. & KOLLER, B. H. 2015. Xenobiotic Metabolism in Mice Lacking the UDP-Glucuronosyltransferase 2 Family. *Drug Metabolism and Disposition*.

FENG, X., JIANG, J., PADHI, A., NING, C., FU, J., WANG, A., MRODE, R. & LIU, J.-F. 2017. Characterization of genome-wide segmental duplications reveals a common genomic feature of association with immunity among domestic animals. *BMC Genomics,* 18**,** 293.

FINEL, M., LI, X., GARDNER-STEPHEN, D., BRATTON, S., MACKENZIE, P. I. & RADOMINSKA-PANDYA, A. 2005. Human UDP-glucuronosyltransferase 1A5: identification, expression, and activity. *J Pharmacol Exp Ther,* 315**,** 1143-9.

FLEISCHER, S., SHARKEY, M., MEALEY, K., OSTRANDER, E. A. & MARTINEZ, M. 2008. Pharmacogenetic and Metabolic Differences Between Dog Breeds: Their Impact on Canine Medicine and the Use of the Dog as a Preclinical Animal Model. *The AAPS Journal,* 10**,** 110-119.

FORSMAN, T., LAUTALA, P., LUNDSTROM, K., MONASTYRSKAIA, K., OUZZINE, M., BURCHELL, B., TASKINEN, J. & ULMANEN, I. 2000. Production of human UDP-glucuronosyltransferases 1A6 and 1A9 using the Semliki Forest virus expression system. *Life Sci,* 67**,** 2473-84.

FU, S., LEWIS, J., WANG, H., KEEGAN, J. & DAWSON, M. 2010. A novel reductive transformation of oxazepam to nordiazepam observed during enzymatic hydrolysis. *J Anal Toxicol,* 34**,** 243-51.

FUJIWARA, R., NAKAJIMA, M., YAMANAKA, H. & YOKOI, T. 2009. Key amino acid residues responsible for the differences in substrate specificity of human UDP-glucuronosyltransferase (UGT)1A9 and UGT1A8. *Drug Metab Dispos,* 37**,** 41-6.

GALLAGHER, C. J., BALLIET, R. M., SUN, D., CHEN, G. & LAZARUS, P. 2010. Sex Differences in UDP-Glucuronosyltransferase 2B17 Expression and Activity. *Drug Metabolism and Disposition,* 38**,** 2204-2209.

GIANTIN, M., CARLETTI, M., CAPOLONGO, F., PEGOLO, S., LOPPARELLI, R. M., GUSSON, F., NEBBIA, C., CANTIELLO, M., MARTIN, P., PINEAU, T. & DACASTO, M. 2008. Effect of breed upon cytochromes P450 and phase II enzyme expression in cattle liver. *Drug Metab Dispos,* 36**,** 885-93.

GIBSON, G. G. & SKETT, P. 2001. *Introduction to drug metabolism*, Nelson Thornes.

GLARE, P., WALSH, D. & SHEEHAN, D. 2006. The adverse effects of morphine: a prospective survey of common symptoms during repeated dosing for chronic cancer pain. *Am J Hosp Palliat Care,* 23**,** 229-35.

GLATT, H., ENGELKE, C. E., PABEL, U., TEUBNER, W., JONES, A. L., COUGHTRIE, M. W., ANDRAE, U., FALANY, C. N. & MEINL, W. 2000. Sulfotransferases: genetics and role in toxicology. *Toxicol Lett,* 112-113**,** 341-8.

GLATT, H. & MEINL, W. 2004. Pharmacogenetics of soluble sulfotransferases (SULTs). *Naunyn Schmiedebergs Arch Pharmacol,* 369**,** 55-68.

GONG, Q.-H., CHO, J. W., HUANG, T., POTTER, C., GHOLAMI, N., BASU, N. K., KUBOTA, S., CARVALHO, S., PENNINGTON, M. W., OWENS, I. S. & POPESCU, N. C. 2001. Thirteen UDPglucuronosyltransferase genes are encoded at the human UGT1 gene complex locus. *Pharmacogenetics and Genomics,* 11**,** 357-368.

GRAHAM, M. L. & PRESCOTT, M. J. 2015. The multifactorial role of the 3Rs in shifting the harm-benefit analysis in animal models of disease. *European Journal of Pharmacology,* 759**,** 19-29.

GRANT, D. M., HUGHES, N. C., JANEZIC, S. A., GOODFELLOW, G. H., CHEN, H. J., GAEDIGK, A., YU, V. L. & GREWAL, R. 1997. Human acetyltransferase polymorphisms. *Mutat Res,* 376**,** 61-70.

GUENGERICH, F. P. 2003. Cytochromes P450, drugs, and diseases. *Mol Interv,* 3**,** 194-204.

GUÉNIN, S., MAURIAT, M., PELLOUX, J., VAN WUYTSWINKEL, O., BELLINI, C. & GUTIERREZ, L. 2009. Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *Journal of Experimental Botany,* 60**,** 487-493.

GUHAD, F. 2005. Introduction to the 3Rs (Refinement, Reduction and Replacement). *Journal of the American Association for Laboratory Animal Science,* 44**,** 58-59.

GUILLEMETTE, C. 2003. Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. *Pharmacogenomics J,* 3**,** 136-58.

HANIOKA, N., TAKEDA, Y., JINNO, H., TANAKA-KAGAWA, T., NAITO, S., KOEDA, A., SHIMIZU, T., NOMURA, M. & NARIMATSU, S. 2006. Functional characterization of human and cynomolgus monkey UDP-glucuronosyltransferase 1A6 enzymes. *Chemico-Biological Interactions,* 164**,** 136-145.

HANNEMANN, F., BICHET, A., EWEN, K. M. & BERNHARDT, R. 2007. Cytochrome P450 systems—biological variations of electron transport chains. *Biochimica et Biophysica Acta (BBA) - General Subjects,* 1770**,** 330-344.

HAY KRAUS, B. L., GREENBLATT, D. J., VENKATAKRISHNAN, K. & COURT, M. H. 2000. Evidence for propofol hydroxylation by cytochrome P4502B11 in canine liver microsomes: breed and gender differences. *Xenobiotica,* 30**,** 575-588.

HEIKKINEN, A. T., FRIEDLEIN, A., MATONDO, M., HATLEY, O. J., PETSALO, A., JUVONEN, R., GALETIN, A., ROSTAMI-HODJEGAN, A., AEBERSOLD, R., LAMERZ, J., DUNKLEY, T., CUTLER, P. & PARROTT, N. 2015. Quantitative ADME proteomics - CYP and UGT enzymes in the Beagle dog liver and intestine. *Pharm Res,* 32**,** 74-90.

HENRISSAT, B. & BAIROCH, A. 1996. Updating the sequence-based classification of glycosyl hydrolases. *Biochemical Journal,* 316**,** 695-696.

HINES, R. N. & MCCARVER, D. G. 2002. The Ontogeny of Human Drug-Metabolizing Enzymes: Phase I Oxidative Enzymes. *Journal of Pharmacology and Experimental Therapeutics,* 300**,** 355-360.

HU, D. G., MACKENZIE, P. I., MCKINNON, R. A. & MEECH, R. 2016. Genetic polymorphisms of human UDP-glucuronosyltransferase (UGT) genes and cancer risk. *Drug Metab Rev,* 48**,** 47-69.

HUANG, F.-F., CHAI, C.-L., ZHANG, Z., LIU, Z.-H., DAI, F.-Y., LU, C. & XIANG, Z.-H. 2008. *The UDP-glucosyltransferase multigene family in Bombyx mori*.

HUANG, H. & WU, Q. 2010. Cloning and Comparative Analyses of the Zebrafish Ugt Repertoire Reveal Its Evolutionary Diversity. *PLOS ONE,* 5**,** e9144.

HUM, D. W., BELANGER, A., LEVESQUE, E., BARBIER, O., BEAULIEU, M., ALBERT, C., VALLEE, M., GUILLEMETTE, C., TCHERNOF, A., TURGEON, D. & DUBOIS, S. 1999. Characterization of UDP-glucuronosyltransferases active on steroid hormones. *J Steroid Biochem Mol Biol,* 69**,** 413-23.

IYANAGI, T. 2007. Molecular mechanism of phase I and phase II drug-metabolizing enzymes: implications for detoxification. *Int Rev Cytol,* 260**,** 35-112.

IZUKAWA, T., NAKAJIMA, M., FUJIWARA, R., YAMANAKA, H., FUKAMI, T., TAKAMIYA, M., AOKI, Y., IKUSHIRO, S., SAKAKI, T. & YOKOI, T. 2009. Quantitative analysis of UDP-glucuronosyltransferase (UGT) 1A and UGT2B expression levels in human livers. *Drug Metab Dispos,* 37**,** 1759-68.

JANCOVA, P., ANZENBACHER, P. & ANZENBACHEROVA, E. 2010. Phase II drug metabolizing enzymes. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub,* 154**,** 103-16.

KAPP;, K., SCHREMPF;, S., LEMBERG;, M. K. & DOBBERSTEIN, B. 2009. Post-Targeting Functions of Signal Peptides. *In:* ZIMMERMANN, R. (ed.) *Protein Transport into the Endoplasmic Reticulum.* Austin (TX): Landes Bioscience.

KAZLAUSKAS, R. 2010. Designer steroids. *Handb Exp Pharmacol***,** 155-85.

KERDPIN, O., MACKENZIE, P. I., BOWALGAHA, K., FINEL, M. & MINERS, J. O. 2009. Influence of N-Terminal Domain Histidine and Proline Residues on the Substrate Selectivities of Human UDP-Glucuronosyltransferase 1A1, 1A6, 1A9, 2B7, and 2B10. *Drug Metabolism and Disposition,* 37**,** 1948-1955.

KICMAN, A. T. 2008. Pharmacology of anabolic steroids. *British Journal of Pharmacology,* 154**,** 502-521.

KIELECZAWA, J. 2006. Fundamentals of Sequencing of Difficult Templates—An Overview. *Journal of Biomolecular Techniques : JBT,* 17**,** 207-217.

KING, C. D., RIOS, G. R., GREEN, M. D., MACKENZIE, P. I. & TEPHLY, T. R. 1997. Comparison of Stably Expressed Rat UGT1.1 and UGT2B1 in the Glucuronidation of Opioid Compounds. *Drug Metabolism and Disposition,* 25**,** 251-255.

KNYCH, H. K. D. & STANLEY, S. D. 2008. Complementary DNA cloning, functional expression and characterization of a novel cytochrome P450, CYP2D50, from equine liver. *Biochemical Pharmacology,* 76**,** 904-911.

KOLLIAS-BAKER, C. & SAMS, R. 2002. Detection of morphine in blood and urine samples from horses administered poppy seeds and morphine sulfate orally. *J Anal Toxicol,* 26**,** 81-6.

KOMORI, M., HIGAMI, A., IMAI, Y., IMAOKA, S. & FUNAE, Y. 1993. Purification and Characterization of a Form of P 450 from Horse Liver Microsomes. *The Journal of Biochemistry,* 114**,** 445-448.

KONDO, T., IKENAKA, Y., NAKAYAMA, S. M. M., KAWAI, Y. K., MIZUKAWA, H., MITANI, Y., NOMIYAMA, K., TANABE, S. & ISHIZUKA, M. 2017a. Uridine Diphosphate-Glucuronosyltransferase (UGT) 2B Subfamily Interspecies Differences in Carnivores. *Toxicological Sciences,* 158**,** 90-100.

KONDO, T., IKENAKA, Y., NAKAYAMA, S. M. M., KAWAI, Y. K., MIZUKAWA, H., MITANI, Y., NOMYAMA, K., TANABE, S. & ISHIZUKA, M. 2017b. UDP-glucuronosyltransferase (UGT) 2B subfamily interspecies differences in carnivores. *Toxicol Sci*.

KONIG & LIEBICH. 2014. *Veterinary Anatomy of Domestic Mammals: Textbook and Colour Atlas*, Schattauer.

KOZERA, B. & RAPACZ, M. 2013. Reference genes in real-time PCR. *Journal of Applied Genetics,* 54**,** 391-406.

KRISHNASWAMY, S., DUAN, S. X., VON MOLTKE, L. L., GREENBLATT, D. J. & COURT, M. H. 2003. Validation of serotonin (5-hydroxtryptamine) as an in vitro substrate probe for human UDP-glucuronosyltransferase (UGT) 1A6. *Drug Metab Dispos,* 31**,** 133-9.

KRISHNASWAMY, S., HAO, Q., AL-ROHAIMI, A., HESSE, L. M., VON MOLTKE, L. L., GREENBLATT, D. J. & COURT, M. H. 2005. UDP glucuronosyltransferase (UGT) 1A6 pharmacogenetics: I. Identification of polymorphisms in the 5'-regulatory and exon 1 regions, and association with human liver UGT1A6 gene expression and glucuronidation. *J Pharmacol Exp Ther,* 313**,** 1331-9.

KUEHL, G. E., LAMPE, J. W., POTTER, J. D. & BIGLER, J. 2005. GLUCURONIDATION OF NONSTEROIDAL ANTI-INFLAMMATORY DRUGS: IDENTIFYING THE ENZYMES RESPONSIBLE IN HUMAN LIVER MICROSOMES. *Drug Metabolism and Disposition,* 33**,** 1027-1035.

KUNE, G. 1969. The Anatomical Basis of Liver Surgery. *Australian and New Zealand Journal of Surgery,* 39**,** 117-126.

KURKELA, M., GARCíA-HORSMAN, J. A., LUUKKANEN, L., MÖRSKY, S., TASKINEN, J., BAUMANN, M., KOSTIAINEN, R., HIRVONEN, J. & FINEL, M. 2003. Expression and Characterization of Recombinant Human UDP-glucuronosyltransferases (UGTs): UGT1A9 IS MORE RESISTANT TO DETERGENT INHIBITION THAN THE OTHER UGTs AND WAS PURIFIED AS AN ACTIVE DIMERIC ENZYME. *Journal of Biological Chemistry,* 278**,** 3536-3544.

LEAVER, M. J., WRIGHT, J., HODGSON, P., BOUKOUVALA, E. & GEORGE, S. G. 2007. Piscine UDP-glucuronosyltransferase 1B. *Aquat Toxicol,* 84**,** 356-65.

LECLERC, S., HEYDEL, J. M., AMOSSE, V., GRADINARU, D., CATTARELLI, M., ARTUR, Y., GOUDONNET, H., MAGDALOU, J., NETTER, P., PELCZAR, H. & MINN, A. 2002. Glucuronidation of odorant molecules in the rat olfactory system: activity, expression and age-linked modifications of UDP-glucuronosyltransferase isoforms, UGT1A6 and UGT2A1, and relation to mitral cell activity. *Brain Res Mol Brain Res,* 107**,** 201-13.

LEVESQUE, E., TURGEON, D., CARRIER, J. S., MONTMINY, V., BEAULIEU, M. & BELANGER, A. 2001. Isolation and characterization of the UGT2B28 cDNA encoding a novel human steroid conjugating UDP-glucuronosyltransferase. *Biochemistry,* 40**,** 3869-81.

LEWIS, L. D. 2005. Personalized drug therapy; the genome, the chip and the physician. *British Journal of Clinical Pharmacology,* 60**,** 1-4.

LI, Y., BALDAUF, S., LIM, E.-K. & BOWLES, D. J. 2001. Phylogenetic Analysis of the UDP-glycosyltransferase Multigene Family of Arabidopsis thaliana. *Journal of Biological Chemistry,* 276**,** 4338-4343.

LIEBLER, D. C. & ZIMMERMAN, L. J. 2013. Targeted Quantitation of Proteins by Mass Spectrometry. *Biochemistry,* 52**,** 3797-3806.

LÖNNBERG, M., BONDESSON, U., CORMANT, F., GARCIA, P., BONNAIRE, Y., CARLSSON, J., POPOT, M.-A., ROLLBORN, N., RÅSBO, K. & BAILLY-CHOURIBERRY, L. 2012. Detection of recombinant human EPO administered to horses using MAIIA lateral flow isoform test. *Analytical and Bioanalytical Chemistry,* 403**,** 1619-1628.

MACKENZIE, P. I., BOCK, K. W., BURCHELL, B., GUILLEMETTE, C., IKUSHIRO, S., IYANAGI, T., MINERS, J. O., OWENS, I. S. & NEBERT, D. W. 2005. Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenet Genomics,* 15**,** 677-85.

MACKENZIE, P. I., OWENS, I. S., BURCHELL, B., BOCK, K. W., BAIROCH, A., BELANGER, A., GIGLEUX, S. F., GREEN, M., HUM, D. W., IYANAGI, T., LANCET, D., LOUISOT, P., MAGDALOU, J., ROY CHOWDHURY, J., RITTER, J. K., TEPHLY, T. R., SCHACHTER, H., TEPHLY, T., TIPTON, K. F. & NEBERT, D. W. 1997. The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics and Genomics,* 7**,** 255-269.

MACKENZIE, P. I., ROGERS, A., ELLIOT, D. J., CHAU, N., HULIN, J.-A., MINERS, J. O. & MEECH, R. 2011. The Novel UDP Glycosyltransferase 3A2: Cloning, Catalytic Properties, and Tissue Distribution. *Molecular Pharmacology,* 79**,** 472-478.

MACKENZIE, P. I., ROGERS, A., TRELOAR, J., JORGENSEN, B. R., MINERS, J. O. & MEECH, R. 2008. Identification of UDP glycosyltransferase 3A1 as a UDP N-acetylglucosaminyltransferase. *J Biol Chem,* 283**,** 36205-10.

MAHMOOD, T. & YANG, P.-C. 2012. Western Blot: Technique, Theory, and Trouble Shooting. *North American Journal of Medical Sciences,* 4**,** 429-434.

MARLAND, A., SARKAR, P. & LEAVITT, R. 1999. The urinary elimination profiles of diazepam and its metabolites, nordiazepam, temazepam, and oxazepam, in the equine after a 10-mg intramuscular dose. *J Anal Toxicol,* 23**,** 29-34.

MARTOGLIO, B. 2003. Intramembrane proteolysis and post-targeting functions of signal peptides. *Biochem Soc Trans,* 31**,** 1243-7.

MCGINNITY, D. F. & RILEY, R. J. 2001. Predicting drug pharmacokinetics in humans from in vitro metabolism studies. *Biochem Soc Trans,* 29**,** 135-9.

MCKINNON, R. A., SORICH, M. J. & WARD, M. B. 2008. Cytochrome P450 Part 1: Multiplicity and Function. *Journal of Pharmacy Practice and Research,* 38**,** 55-57.

MEECH, R. & MACKENZIE, P. I. 1997a. Structure and function of uridine diphosphate glucuronosyltransferases. *Clin Exp Pharmacol Physiol,* 24**,** 907-15.

MEECH, R. & MACKENZIE, P. I. 1997b. UDP-Glucuronosyltransferase, the Role of the Amino Terminus in Dimerization. *Journal of Biological Chemistry,* 272**,** 26913-26917.

MEECH, R. & MACKENZIE, P. I. 2010. UGT3A: novel UDP-glycosyltransferases of the UGT superfamily. *Drug Metab Rev,* 42**,** 45-54.

MEECH, R., MINERS, J. O., LEWIS, B. C. & MACKENZIE, P. I. 2012a. The glycosidation of xenobiotics and endogenous compounds: Versatility and redundancy in the UDP glycosyltransferase superfamily. *Pharmacology & Therapeutics,* 134**,** 200-218.

MEECH, R., MUBAROKAH, N., SHIVASAMI, A., ROGERS, A., NAIR, P. C., HU, D. G., MCKINNON, R. A. & MACKENZIE, P. I. 2015. A novel function for UDP glycosyltransferase 8: galactosidation of bile acids. *Mol Pharmacol,* 87**,** 442-50.

MEECH, R., ROGERS, A., ZHUANG, L., LEWIS, B. C., MINERS, J. O. & MACKENZIE, P. I. 2012b. Identification of Residues That Confer Sugar Selectivity to UDP-Glycosyltransferase 3A (UGT3A) Enzymes. *Journal of Biological Chemistry,* 287**,** 24122-24130.

MEECH, R., YOGALINGAM, G. & MACKENZIE, P. I. 1996. Mutational analysis of the carboxy-terminal region of UDP-glucuronosyltransferase 2B1. *DNA Cell Biol,* 15**,** 489-94.

MEYER, U. 1996. Overview of enzymes of drug metabolism. *Journal of Pharmacokinetics and Biopharmaceutics,* 24**,** 449-459.

MEYER, U. A. 2000. Pharmacogenetics and adverse drug reactions. *The Lancet,* 356**,** 1667-1671.

MICHALOPOULOS, G. K. & DEFRANCES, M. C. 1997. Liver Regeneration. *Science,* 276**,** 60-66.

MILNE, A. M., BURCHELL, B. & COUGHTRIE, M. W. H. 2011. A Novel Method for the Immunoquantification of UDP-Glucuronosyltransferases in Human Tissue. *Drug Metabolism and Disposition,* 39**,** 2258-2263.

MOON, Y. J., WANG, X. & MORRIS, M. E. 2006. Dietary flavonoids: Effects on xenobiotic and carcinogen metabolism. *Toxicology in Vitro,* 20**,** 187-210.

MORETON, J., MALLA, S., ABOOBAKER, A. A., TARLINTON, R. E. & EMES, R. D. 2014. Characterisation of the horse transcriptome from immunologically active tissues. *PeerJ,* 2**,** e382.

MOSKALEVA, N., MOYSA, A., NOVIKOVA, S., TIKHONOVA, O., ZGODA, V. & ARCHAKOV, A. 2015. Spaceflight Effects on Cytochrome P450 Content in Mouse Liver. *PLOS ONE,* 10**,** e0142374.

MOUGOUS, J. D., GREEN, R. E., WILLIAMS, S. J., BRENNER, S. E. & BERTOZZI, C. R. 2002. Sulfotransferases and sulfatases in mycobacteria. *Chem Biol,* 9**,** 767-76.

MUGFORD, C. A. & KEDDERIS, G. L. 1998. Sex-Dependent Metabolism of Xenobiotics. *Drug Metabolism Reviews,* 30**,** 441-498.

NAIR, P. C., MEECH, R., MACKENZIE, P. I., MCKINNON, R. A. & MINERS, J. O. 2015. Insights into the UDP-sugar selectivities of human UDP-glycosyltransferases (UGT): a molecular modeling perspective. *Drug Metabolism Reviews,* 47**,** 335-345.

NAKAMURA, K., FUJIKI, T. & TAMURA, H.-O. 2005. Age, gender and region-specific differences in drug metabolising enzymes in rat ocular tissues. *Experimental Eye Research,* 81**,** 710-715.

NATH, A. & ATKINS, W. M. 2006. A Theoretical Validation of the Substrate Depletion Approach to Determining Kinetic Parameters. *Drug Metabolism and Disposition,* 34**,** 1433-1435.

NEBBIA, C., DACASTO, M., ROSSETTO GIACCHERINO, A., GIULIANO ALBO, A. & CARLETTI, M. 2003. Comparative Expression of Liver Cytochrome P450-dependent Monooxygenases in the Horse and in other Agricultural and Laboratory Species. *The Veterinary Journal,* 165**,** 53-64.

NEBERT, D. W. & RUSSELL, D. W. 2002. Clinical importance of the cytochromes P450. *Lancet,* 360**,** 1155-62.

NEBERT, D. W., WIKVALL, K. & MILLER, W. L. 2013. Human cytochromes P450 in health and disease. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 368**,** 20120431.

NEUMANN, E., MEHBOOB, H., RAMÍREZ, J., MIRKOV, S., ZHANG, M. & LIU, W. 2016. Age-Dependent Hepatic UDP-Glucuronosyltransferase Gene Expression and Activity in Children. *Frontiers in Pharmacology,* 7**,** 437.

NIKLASON, L., GAO, J., ABBOTT, W., HIRSCHI, K., HOUSER, S., MARINI, R. & LANGER, R. 1999. Functional arteries grown in vitro. *Science,* 284**,** 489-493.

NING, B.-T. & TANG, Y.-M. 2012. Establishment of the cell line, HeLa-CD14, transfected with the human CD14 gene. *Oncology Letters,* 3**,** 871-874.

NOMIYAMA, H., OSADA, N. & YOSHIE, O. 2013. Systematic classification of vertebrate chemokines based on conserved synteny and evolutionary history. *Genes to Cells,* 18**,** 1-16.

NOWELL, S. & FALANY, C. N. 2006. Pharmacogenetics of human cytosolic sulfotransferases. *Oncogene,* 25**,** 1673-8.

OHNO, S., KAWANA, K. & NAKAJIN, S. 2008. Contribution of UDP-Glucuronosyltransferase 1A1 and 1A8 to Morphine-6-Glucuronidation and Its Kinetic Properties. *Drug Metabolism and Disposition,* 36**,** 688-694.

OHNO, S. & NAKAJIN, S. 2009. Determination of mRNA expression of human UDP-glucuronosyltransferases and application for localization in various human tissues by real-time reverse transcriptase-polymerase chain reaction. *Drug Metab Dispos,* 37**,** 32-40.

ORR, C. 2016. *Characterisation of equine cytochrome P450s.* University of Nottingham.

OUZZINE, M., ANTONIO, L., BURCHELL, B., NETTER, P., FOURNEL-GIGLEUX, S. & MAGDALOU, J. 2000. Importance of Histidine Residues for the Function of the Human Liver UDP-Glucuronosyltransferase UGT1A6: Evidence for the Catalytic Role of Histidine 370. *Molecular Pharmacology,* 58**,** 1609-1615.

OUZZINE, M., MAGDALOU, J., BURCHELL, B. & FOURNEL-GIGLEUX, S. 1999. An Internal Signal Sequence Mediates the Targeting and Retention of the Human UDP-Glucuronosyltransferase 1A6 to the Endoplasmic Reticulum. *Journal of Biological Chemistry,* 274**,** 31401-31409.

OWENS, I. S., BASU, N. K. & BANERJEE, R. 2005. UDP-glucuronosyltransferases: gene structures of UGT1 and UGT2 families. *Methods Enzymol,* 400**,** 1-22.

PATEL, H., BELL, D., MOLOKHIA, M., SRISHANMUGANATHAN, J., PATEL, M., CAR, J. & MAJEED, A. 2007. Trends in hospital admissions for adverse drug reactions in England: analysis of national hospital episode statistics 1998–2005. *BMC Clinical Pharmacology,* 7**,** 9.

PENNER, N., WOODWARD, C. & PRAKASH, C. 2012. Drug-metabolizing enzymes and biotransformation reactions. *ADME-enabling technologies in drug design and development*.

PETERS, L. M., DEMMEL, S., PUSCH, G., BUTERS, J. T. M., THORMANN, W., ZIELINSKI, J., LEEB, T., MEVISSEN, M. & SCHMITZ, A. 2013. Equine cytochrome P450 2B6 — Genomic identification, expression and functional characterization with ketamine. *Toxicology and Applied Pharmacology,* 266**,** 101-108.

PRICE, E. R., ZYDOWSKY, L. D., JIN, M. J., BAKER, C. H., MCKEON, F. D. & WALSH, C. T. 1991. Human cyclophilin B: a second cyclophilin gene encodes a peptidyl-prolyl isomerase with a signal sequence. *Proceedings of the National Academy of Sciences,* 88**,** 1903-1907.

RADOMINSKA-PANDYA, A., BRATTON, S. & LITTLE, J. M. 2005a. A historical overview of the heterologous expression of mammalian UDP-glucuronosyltransferase isoforms over the past twenty years. *Curr Drug Metab,* 6**,** 141-60.

RADOMINSKA-PANDYA, A., OUZZINE, M., FOURNEL-GIGLEUX, S. & MAGDALOU, J. 2005b. Structure of UDP-glucuronosyltransferases in membranes. *In:* SIES, H. & PACKER, L. (eds.) *Phase Ii Conjugation Enzymes and Transport Systems.* San Diego: Elsevier Academic Press Inc.

RADOMINSKA-PANDYA, A., OUZZINE, M., FOURNEL-GIGLEUX, S. & MAGDALOU, J. 2005c. Structure of UDP-glucuronosyltransferases in membranes. *Methods Enzymol,* 400**,** 116-47.

REECE, W. O. 2009. *Fucntional Anatomy and Physiology of Domestic Animals*, Wiley-Blackwell.

RENAULT, H., BASSARD, J.-E., HAMBERGER, B. & WERCK-REICHHART, D. 2014. Cytochrome P450-mediated metabolic engineering: current progress and future challenges. *Current Opinion in Plant Biology,* 19**,** 27-34.

RITTER, J. K. 2000. Roles of glucuronidation and UDP-glucuronosyltransferases in xenobiotic bioactivation reactions. *Chemico-Biological Interactions,* 129**,** 171-193.

RITTER, J. K., CHEN, F., SHEEN, Y. Y., TRAN, H. M., KIMURA, S., YEATMAN, M. T. & OWENS, I. S. 1992. A novel complex locus UGT1 encodes human bilirubin, phenol, and other UDP-glucuronosyltransferase isozymes with identical carboxyl termini. *Journal of Biological Chemistry,* 267**,** 3257-3261.

ROMIGUIER, J., RANWEZ, V., DOUZERY, E. J. P. & GALTIER, N. 2010. Contrasting GC-content dynamics across 33 mammalian genomes: Relationship with life-history traits and chromosome sizes. *Genome Research,* 20**,** 1001-1009.

ROSS, J., LI, Y., LIM, E. & BOWLES, D. J. 2001. Higher plant glycosyltransferases. *Genome Biol,* 2**,** REVIEWS3004.

ROWLAND, A., MINERS, J. O. & MACKENZIE, P. I. 2013. The UDP-glucuronosyltransferases: their role in drug metabolism and detoxification. *Int J Biochem Cell Biol,* 45**,** 1121-32.

SAKAGUCHI, K., GREEN, M., STOCK, N., REGER, T. S., ZUNIC, J. & KING, C. 2004. Glucuronidation of carboxylic acid containing compounds by UDP-glucuronosyltransferase isoforms. *Archives of Biochemistry and Biophysics,* 424**,** 219-225.

SALLOVITZ, J., LIFSCHITZ, A., IMPERIALE, F., PIS, A., VIRKEL, G. & LANUSSE, C. 2002. Breed Differences on the Plasma Availability of Moxidectin Administered Pour-on to Calves. *The Veterinary Journal,* 164**,** 47-53.

SCARTH, J. P., TEALE, P. & KUURANNE, T. 2011. Drug metabolism in the horse: a review. *Drug Testing and Analysis,* 3**,** 19-53.

SCHÄNZER, W. 1996. Metabolism of anabolic androgenic steroids. *Clinical Chemistry,* 42**,** 1001-1020.

SCHMID, J., HEIDER, D., WENDEL, N. J., SPERL, N. & SIEBER, V. 2016. Bacterial Glycosyltransferases: Challenges and Opportunities of a Highly Diverse Enzyme Class Toward Tailoring Natural Products. *Frontiers in Microbiology,* 7**,** 182.

SCHMITZ, A., DEMMEL, S., PETERS, L. M., LEEB, T., MEVISSEN, M. & HAASE, B. 2010. Comparative human–horse sequence analysis of the CYP3A subfamily gene cluster. *Animal Genetics,* 41**,** 72-79.

SCHNEIDER, E. H. & SEIFERT, R. 2010. Sf9 cells: A versatile model system to investigate the pharmacological properties of G protein-coupled receptors. *Pharmacology & Therapeutics,* 128**,** 387-418.

SELYE, H. 1941. ON THE RÔLE OF THE LIVER IN THE DETOXIFICATION OF STEROID HORMONES AND ARTIFICIAL ESTROGENS. *Journal of Pharmacology and Experimental Therapeutics,* 71**,** 236-238.

SEPPEN, J., STEENKEN, E., LINDHOUT, D., BOSMA, P. J. & OUDE ELFERINK, R. P. J. 1996. A mutation which disrupts the hydrophobic core of the signal peptide of bilirubin UDP-glucuronosyltransferase, an endoplasmic reticulum membrane protein, causes Crigler-Najjar type IIs. *FEBS Letters,* 390**,** 294-298.

SHARPE, H. J., STEVENS, T. J. & MUNRO, S. 2010. A Comprehensive Comparison of Transmembrane Domains Reveals Organelle-Specific Properties. *Cell,* 142**,** 158-169.

SHEARIN, A. L. & OSTRANDER, E. A. 2010. Leading the way: canine models of genomics and disease. *Disease Models & Mechanisms,* 3**,** 27-34.

SHELBY, M. K., CHERRINGTON, N. J., VANSELL, N. R. & KLAASSEN, C. D. 2003. Tissue mRNA expression of the rat UDP-glucuronosyltransferase gene family. *Drug Metab Dispos,* 31**,** 326-33.

SHIANG, R., THOMPSON, L. M., ZHU, Y.-Z., CHURCH, D. M., FIELDER, T. J., BOCIAN, M., WINOKUR, S. T. & WASMUTH, J. J. Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia. *Cell,* 78**,** 335-342.

SHIMADA, T., MURAYAMA, N., KAKIMOTO, K., TAKENAKA, S., LIM, Y. R., YEOM, S., KIM, D., YAMAZAKI, H., GUENGERICH, F. P. & KOMORI, M. 2017. Oxidation of 1-chloropyrene by human CYP1 family and CYP2A subfamily cytochrome P450 enzymes: catalytic roles of two CYP1B1 and five CYP2A13 allelic variants. *Xenobiotica***,** 1-11.

SIM, E., LACK, N., WANG, C. J., LONG, H., WESTWOOD, I., FULLAM, E. & KAWAMURA, A. 2008. Arylamine N-acetyltransferases: structural and functional implications of polymorphisms. *Toxicology,* 254**,** 170-83.

SOARS, M. G., SMITH, D. J., RILEY, R. J. & BURCHELL, B. 2001. Cloning and Characterization of a Canine UDP-Glucuronosyltransferase. *Archives of Biochemistry and Biophysics,* 391**,** 218-224.

STACKEBRANDT, E. & GOEBEL, B., M. 1994. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic and Evolutionary Microbiology,* 44**,** 846-849.

STEN, T., BICHLMAIER, I., KUURANNE, T., LEINONEN, A., YLI-KAUHALUOMA, J. & FINEL, M. 2009. UDP-Glucuronosyltransferases (UGTs) 2B7 and UGT2B17 Display Converse Specificity in Testosterone and Epitestosterone Glucuronidation, whereas UGT2A1 Conjugates Both Androgens Similarly. *Drug Metabolism and Disposition,* 37**,** 417-423.

STINGL, J. C., BARTELS, H., VIVIANI, R., LEHMANN, M. L. & BROCKMOLLER, J. 2014. Relevance of UDP-glucuronosyltransferase polymorphisms for drug dosing: A quantitative systematic review. *Pharmacol Ther,* 141**,** 92-116.

STONE, A. N., MACKENZIE, P. I., GALETIN, A., HOUSTON, J. B. & MINERS, J. O. 2003. ISOFORM SELECTIVITY AND KINETICS OF MORPHINE 3- AND 6-GLUCURONIDATION BY HUMAN UDP-GLUCURONOSYLTRANSFERASES: EVIDENCE FOR ATYPICAL GLUCURONIDATION KINETICS BY UGT2B7. *Drug Metabolism and Disposition,* 31**,** 1086-1089.

STRÄHLE, U., SCHOLZ, S., GEISLER, R., GREINER, P., HOLLERT, H., RASTEGAR, S., SCHUMACHER, A., SELDERSLAGHS, I., WEISS, C., WITTERS, H. & BRAUNBECK, T. 2012. Zebrafish embryos as an alternative to animal experiments—A commentary on the definition of the onset of protected life stages in animal welfare regulations. *Reproductive Toxicology,* 33**,** 128-132.

STRANGE, R. C., SPITERI, M. A., RAMACHANDRAN, S. & FRYER, A. A. 2001. Glutathione-S-transferase family of enzymes. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 482**,** 21-26.

STROLIN BENEDETTI, M., WHOMSLEY, R. & BALTES, E. 2006. Involvement of enzymes other than CYPs in the oxidative metabolism of xenobiotics. *Expert Opin Drug Metab Toxicol,* 2**,** 895-921.

SUZUKI, M., HIRATA, M., TAKAGI, M., WATANABE, T., IGUCHI, T., KOIWAI, K., MAEZAWA, S. & KOIWAI, O. 2014. Truncated UDP-glucuronosyltransferase (UGT) from a Crigler–

Najjar syndrome type II patient colocalizes with intact UGT in the endoplasmic reticulum. *Journal of Human Genetics,* 59**,** 158-162.

SZABADOS, L. & SAVOURÉ, A. 2010. Proline: a multifunctional amino acid. *Trends in Plant Science,* 15**,** 89-97.

TEALE, P. & HOUGHTON, E. 2010. Metabolism of anabolic steroids and their relevance to drug detection in horseracing. *Bioanalysis,* 2**,** 1085-107.

TERRIER, N., BENOIT, E., SENAY, C., LAPICQUE, F., RADOMINSKA-PANDYA, A., MAGDALOU, J. & FOURNEL-GIGLEUX, S. 1999. Human and Rat Liver UDP-Glucuronosyltransferases Are Targets of Ketoprofen Acylglucuronide. *Molecular Pharmacology,* 56**,** 226-234.

THOMSON, R. E., BIGLEY, A. L., FOSTER, J. R., JOWSEY, I. R., ELCOMBE, C. R., ORTON, T. C. & HAYES, J. D. 2004. Tissue-specific Expression and Subcellular Distribution of Murine Glutathione S-transferase Class Kappa. *Journal of Histochemistry & Cytochemistry,* 52**,** 653-662.

TOUTAIN, P.-L., FERRAN, A. & BOUSQUET-MÉLOU, A. 2010. Species Differences in Pharmacokinetics and Pharmacodynamics. *In:* CUNNINGHAM, F., ELLIOTT, J. & LEES, P. (eds.) *Comparative and Veterinary Pharmacology.* Berlin, Heidelberg: Springer Berlin Heidelberg.

TREPANIER, L. A., RAY, K., WINAND, N. J., SPIELBERG, S. P. & CRIBB, A. E. 1997. Cytosolic arylamine n-acetyltransferase (NAT) deficiency in the dog and other canids due to an absence of NAT genes. *Biochemical Pharmacology,* 54**,** 73-80.

TROBERG, J., JÄRVINEN, E., MUNIZ, M., SNEITZ, N., MOSORIN, J., HAGSTRÖM, M. & FINEL, M. 2015. Dog UDP-Glucuronosyltransferase Enzymes of Subfamily 1A: Cloning, Expression, and Activity. *Drug Metabolism and Disposition,* 43**,** 107-118.

TUKEY, R. H. & STRASSBURG, C. P. 2000. Human UDP-Glucuronosyltransferases: Metabolism, Expression, and Disease. *Annual Review of Pharmacology and Toxicology,* 40**,** 581-616.

TYDEN, E., LÖFGREN, M., PEGOLO, S., CAPOLONGO, F., TJÄLVE, H. & LARSSON, P. 2012. Differential gene expression of CYP3A isoforms in equine liver and intestines. *Journal of Veterinary Pharmacology and Therapeutics,* 35**,** 588-595.

TYDÉN, E., TJÄLVE, H. & LARSSON, P. 2014. Gene and protein expression and cellular localisation of cytochrome P450 enzymes of the 1A, 2A, 2C, 2D and 2E subfamilies in equine intestine and liver. *Acta Veterinaria Scandinavica,* 56**,** 69.

UCHIHASHI, S., NISHIKAWA, M., SAKAKI, T. & IKUSHIRO, S.-I. 2012. The critical role of amino acid residue at position 117 of mouse UDP-glucuronosyltransfererase 1a6a and 1a6b in resveratrol glucuronidation. *The Journal of Biochemistry,* 152**,** 331-340.

VAN BEUSEKOM, C. D., FINK-GREMMELS, J. & SCHRICKX, J. A. 2014. Comparing the glucuronidation capacity of the feline liver with substrate-specific glucuronidation in dogs. *Journal of Veterinary Pharmacology and Therapeutics,* 37**,** 18-24.

VILLENEUVE, L., GIRARD, H., FORTIER, L.-C., GAGNÉ, J.-F. & GUILLEMETTE, C. 2003. Novel Functional Polymorphisms in the UGT1A7 and UGT1A9 Glucuronidating Enzymes in Caucasian and African-American Subjects and Their Impact on the Metabolism of 7-Ethyl-10-hydroxycamptothecin and Flavopiridol Anticancer Drugs. *Journal of Pharmacology and Experimental Therapeutics,* 307**,** 117-128.

VIMERCATI, S., BÜCHI, M., ZIELINSKI, J., PEDUTO, N. & MEVISSEN, M. 2017. Testosterone metabolism of equine single CYPs of the 3A subfamily compared to the human CYP3A4. *Toxicology in Vitro,* 41**,** 83-91.

VUILLEUMIER, S. & PAGNI, M. 2002. The elusive roles of bacterial glutathione S-transferases: new lessons from genomes. *Applied Microbiology and Biotechnology,* 58**,** 138-146.

VYSKO, #X10D, ILOV, #XE1, , E., SZOT, #XE1, KOV, #XE1, , B., SK, #XE1, LOV, #XE1, , L., #XE1, RT, #XED, KOV, #XE1, , H., HLAV, #XE1, #X10D, OV, #XE1, , J., BOU, #X161, OV, #XE1 & , I. 2013. Age-Related Changes in Hepatic Activity and Expression of Detoxification Enzymes in Male Rats. *BioMed Research International,* 2013**,** 10.

WADE, C. M., GIULOTTO, E., SIGURDSSON, S., ZOLI, M., GNERRE, S., IMSLAND, F., LEAR, T. L., ADELSON, D. L., BAILEY, E., BELLONE, R. R., BLÖCKER, H., DISTL, O., EDGAR, R. C., GARBER, M., LEEB, T., MAUCELI, E., MACLEOD, J. N., PENEDO, M. C. T., RAISON, J. M., SHARPE, T., VOGEL, J., ANDERSSON, L., ANTCZAK, D. F., BIAGI, T., BINNS, M. M., CHOWDHARY, B. P., COLEMAN, S. J., DELLA VALLE, G., FRYC, S., GUÉRIN, G., HASEGAWA, T., HILL, E. W., JURKA, J., KIIALAINEN, A., LINDGREN, G., LIU, J., MAGNANI, E., MICKELSON, J. R., MURRAY, J., NERGADZE, S. G., ONOFRIO, R., PEDRONI, S., PIRAS, M. F., RAUDSEPP, T., ROCCHI, M., RØED, K. H., RYDER, O. A., SEARLE, S., SKOW, L., SWINBURNE, J. E., SYVÄNEN, A. C., TOZAKI, T., VALBERG, S. J., VAUDIN, M., WHITE, J. R., ZODY, M. C., PLATFORM, B. I. G. S., TEAM, B. I. W. G. A., LANDER, E. S. & LINDBLAD-TOH, K. 2009. Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse. *Science,* 326**,** 865-867.

WALES, M. R. & FEWSON, C. A. 1994. NADP-dependent alcohol dehydrogenases in bacteria and yeast: purification and partial characterization of the enzymes from Acinetobacter calcoaceticus and Saccharomyces cerevisiae. *Microbiology,* 140 (Pt 1)**,** 173-83.

WALTER, P. & JOHNSON, A. E. 1994. Signal Sequence Recognition and Protein Targeting to the Endoplasmic Reticulum Membrane. *Annual Review of Cell Biology,* 10**,** 87-119.

WANG, Y., HUANG, H. & WU, Q. 2014. Characterization of the Zebrafish Ugt Repertoire Reveals a New Class of Drug-metabolizing UDP Glucuronosyltransferases. *Molecular Pharmacology*.

WATANABE, K. P., KAWAI, Y. K., IKENAKA, Y., KAWATA, M., IKUSHIRO, S.-I., SAKAKI, T. & ISHIZUKA, M. 2013. Avian Cytochrome P450 (CYP) 1-3 Family Genes: Isoforms, Evolutionary Relationships, and mRNA Expression in Chicken Liver. *PLOS ONE,* 8**,** e75689.

WAXMAN, D. J. & HOLLOWAY, M. G. 2009. Sex Differences in the Expression of Hepatic Drug Metabolizing Enzymes. *Molecular Pharmacology,* 76**,** 215-228.

WERCK-REICHHART, D. & FEYEREISEN, R. 2000. Cytochromes P450: a success story. *Genome Biology,* 1**,** reviews3003.1-reviews3003.9.

WILDT, S., KEARNS, G., LEEDER, J. S. & ANKER, J. 1999. Glucuronidation in Humans. *Clinical Pharmacokinetics,* 36**,** 439-452.

WONG, J. K. Y., CHAN, G. H. M., LEUNG, D. K. K., TANG, F. P. W. & WAN, T. S. M. 2016. Generation of phase II in vitro metabolites using homogenized horse liver. *Drug Testing and Analysis,* 8**,** 241-247.

XU, C., LI, C.-T. & KONG, A.-N. 2005. Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Archives of Pharmacal Research,* 28**,** 249-268.

YASUDA, S., LIU, M.-Y., YANG, Y.-S., SNOW, R., TAKAHASHI, S. & LIU, M.-C. 2006. Identification of novel hydroxysteroid-sulfating cytosolic SULTs, SULT2 ST2 and SULT2 ST3, from zebrafish: Cloning, expression, characterization, and developmental expression. *Archives of Biochemistry and Biophysics,* 455**,** 1-9.

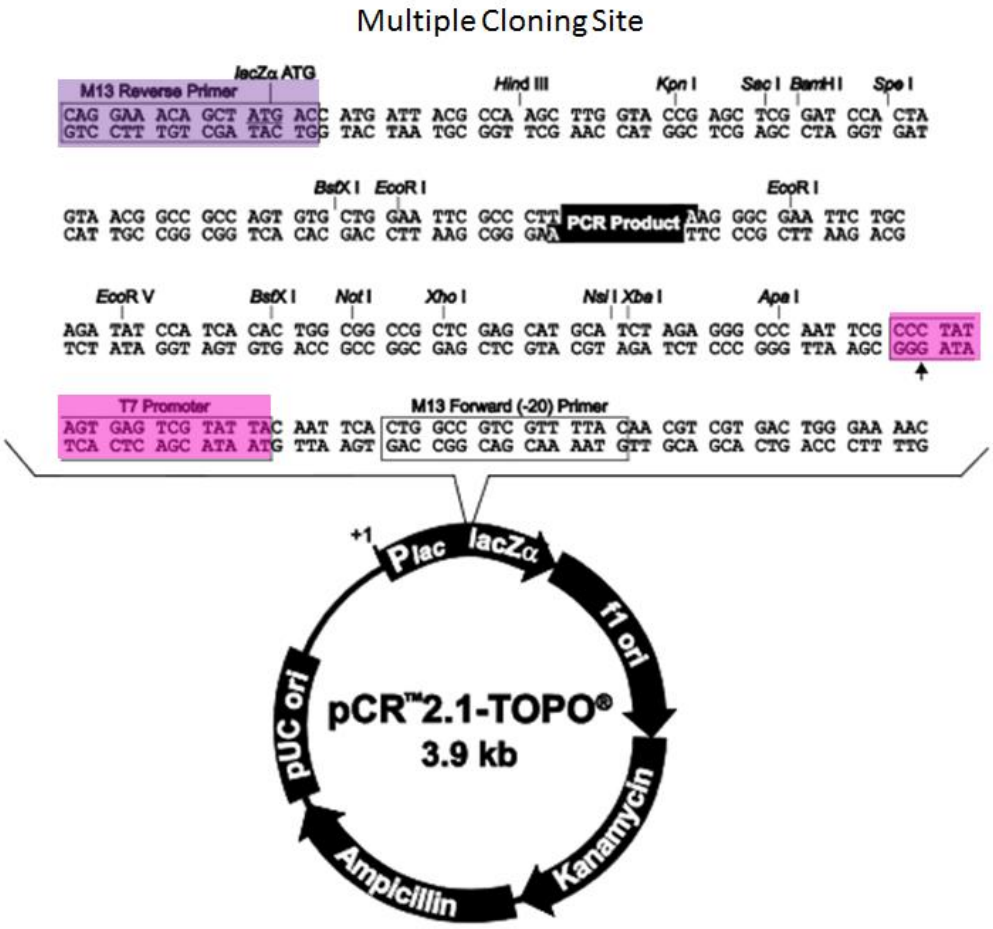
ZANETTE, J., JENNY, M. J., GOLDSTONE, J. V., PARENTE, T., WOODIN, B. R., BAINY, A. C. D. & STEGEMAN, J. J. 2013. Identification and Expression of Multiple CYP1-like and CYP3-like Genes in the Bivalve Mollusk Mytilus edulis. *Aquatic toxicology (Amsterdam, Netherlands),* 0**,** 10.1016/j.aquatox.2012.11.017.

ZHANG, D., LUO, G., DING, X. & LU, C. 2012. Preclinical experimental models of drug metabolism and disposition in drug discovery and development. *Acta Pharmaceutica Sinica B,* 2**,** 549-561.

ZHANG, T., HAWS, P. & WU, Q. 2004. Multiple variable first exons: a mechanism for cell- and tissue-specific gene regulation. *Genome Res,* 14**,** 79-89.

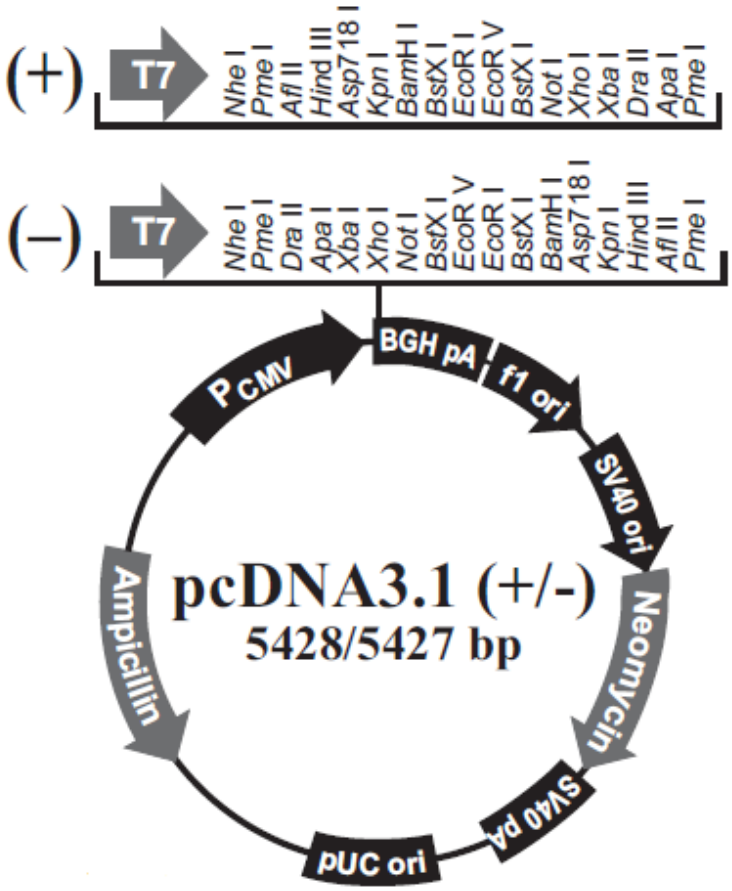
Appendix

Appendix A



Appendix A: Vector map for pCR™2.1-TOPO. The vector map, with the multiple cloning site displayed. The forward primer, T7 promoter highlighted pink, and the reverse primer, M13 Reverse highlighted purple. pCR™2.1 was used to amplify the genes and to ensure each cloned gene was fully sequenced.

Appendix B



Appendix B: Vector map for pcDNA™3.1. The vector map for pcDNA™3.1, the multiple cloning site for plus and minus vector is shown. Genes were cloned in to the appropriate vector depending on the use of the restriction enzymes relative to the start codon, ATG. The vector pcDNA™3.1 was selected to be the expression vector in the mammalian HEK293 cell system.

Appendix C

Animal/UGT	Accession
Barley UGT	ADC92549.1
Orange UGT	ACS87992.1
Baboon UGT1A1	NP_001106091.1
Baboon UGT1A6	NP_001106092.1
Baboon UGT1A9	NP_001106093.1
Cat UGT1A1	XP_006935740.1
Cow UGT1A6	NP_777187.1
Cow UGT1A1	NP_001099106.1
Cow UGT3A1	NP_001069555.1
Crab-eating macaque UGT1A9	NP_001270367.1
Chimp UGT3A1	XP_003310800.1
Chimp UGT3A2	XP_517805.2
C.elegans UGT10	NP_504313.2
C.elegans UGT11	NP_504311.1
C.elegans UGT12	NP_504309.1
C.elegans UGT24	NP_500931.3
Grey Wolf UGT1A6	NP_001003078.1
Guinea Pig UGT2A3	NP_001166497.1
Human UGT1A1	NP_000454.1
Human UGT1A3	AAR95639.1
Human UGT1A4	AAG30422.1
Human UGT1A5	AAG30421.1
Human UGT1A6	NP_001063.2
Human UGT1A7	AAB81536.1
Human UGT1A8	NP_061949.3
Human UGT1A9	NP_066307.1
Human UGT1A10	NP_061948.1
Human UGT2A1	NP_006789.3
Human UGT2A2	NP_001288162.1
Human UGT2A3	NP_079019.3
Human UGT2B4	NP_066962.2
Human UGT2B7	NP_001065.2
Human UGT2B10	NP_001066.1
Human UGT2B15	NP_001067.2
Human UGT3A1	NP_689617.3
Horse UGT1A6-like	ENSECAG00000023519
Horse UGT2B31-like (1)	ENSECAG00000018165
Horse UGT2B31-like (2)	ENSECAG00000017275
Horse UGT2B31-like (3)	ENSECAG00000020628
Horse UGT2C1-like (1)	ENSECAG00000008247
Horse UGT2C1-like (2)	ENSECAG00000019112
Horse UGT2C1-like (3)	ENSECAG00000017801
Horse UGT3A1-like (1)	ENSECAG00000008900
Horse UGT3A1-like (2)	ENSECAG00000010396
Mouse UGT1A1	AAP48593.1
Mouse UGT1A6	NP_659545.2
Mouse UGT1A9	NP_964006.2

Mouse UGT1A10	NP_964004.1
Mouse UGT3A1	NP_997099.2
Mouse UGT3A2	NP_659094.1
Marmoset UGT3A1	XM_002763522.3
Marmoset UGT3A2	XM_002745060.3
Rat UGT1A1	NP_036815.1
Rat UGT1A6	AAL67853.1
Rat UGT1A9	NP_787040.2
Rat UGT1A10	NP_958828.1
Rat UGT2B1	NP_775417.1
Rat UGT2B2	NP_113721.4
Rat UGT2B7	NP_775445.1
Rat UGT2B10	NP_001178605.1
Rat UGT2B17	NP_695226.2
Rhesus macaque UGT1A1	NP_001028041.1
Rhesus macaque UGT2B9	NP_001028199.1
Rhesus macaque UGT2B33	NP_001028002.1
Rhesus macaque UGT3A1	XM_001093373.3
Rhesus macaque UGT3A2	XM_015139865.1
Sheep UGT1A1	NP_001192076.1
Sheep UGT1A3	NP_001192077.1
Sheep UGT1A4	NP_001192078.1
Sheep UGT1A6	NP_001192075.1
Sheep UGT1A9	ADZ11102.1
Sheep UGT2B7	NP_001192080.1
Zebrafish UGT1A1	NP_001032505.2
Zebrafish UGT1A2	NP_001166241.2
Zebrafish UGT1A4	NP_001170815.2
Zebrafish UGT1A5	NP_001170811.2
Zebrafish UGT1A6	NP_001170810.2
Zebrafish UGT1A7	NP_001170805.1
Zebrafish UGT1B1	NP_001170917.1
Zebrafish UGT1B4	NP_001166239.2
Zebrafish UGT2A1	NP_001177979.1
Zebrafish UGT2A2	NP_001018306.2
Zebrafish UGT2A3	NP_001170814.2
Zebrafish UGT2A4	NP_001170804.2
Zebrafish UGT2A5	NP_001070111.2
Zebrafish UGT2A6	NP_001138283.1
Zebrafish UGT2B1	NP_001170809.1
Zebrafish UGT2B3	NP_001170812.1
African Clawed frog UGT3A1	NP_001088053.1
Western Clawed frog UGT1A6	NP_001107366.1
Western clawed frog UGT3A2	NP_001005027.3

Appendix C: List of accession numbers. 91 sequence files were retrieved from the NCBI and Ensembl. Using Genious (Biomatters Ltd), sequences were aligned and their relationships inferred using maximum likelihood method, bootstrapped, 500 replicates.

Appendix D

Animal/UGT	Accession Number
Orange	ACS87992.1
Barley	ADC92549.1
Baboon UGT1A1	NP_001106091.1
Baboon UGT1A6	NP_001106092.1
Baboon UGT1A9	NP_001106093.1
Cat UGT1A1	XP_006935740.1
Cow UGT1A1	NP_001099106.1
Cow UGT1A6	NP_777187.1
Crab-eating macaque UGT1A1	NP_001270367.1 1
Crab-eating macaque UGT1A3	NP_001336958.1
Crab-eating macaque UGT1A8	NP_001336952.1
Cynomolgus monkey UGT1A6	NP_001336957.1
Common marmoset UGT1A1	ABY79101.1
Common marmoset UGT1A6	JAB08411.1
Common marmoset UGT1A8	ABY79098.1
Dusky titi monkey UGT1A1	ACA57873.1
Greater Horseshoe Bat UGT1A1	ACC62110.1
Grey Wolf UGT1A6	NP_001003078.1
Guinea pig UGT1A1	ALO62045.1
Guinea pig UGT1A4	ALO62047.1
Guinea pig UGT1A7	ALO62048.1
Guinea pig UGT1A8	ALO62046.1
Human UGT1A1	NP_000454.1
Human UGT1A3	AAR95639.1
Human UGT1A4	AAG30422.1
Human UGT1A5	AAG30421.1
Human UGT1A6	NP_001063.2
Human UGT1A7	AAB81536.1
Human UGT1A8	NP_061949.3
Human UGT1A9	NP_066307.1
Human UGT1A10	NP_061948.1
Horse UGT1A6-like	ENSECAG00000023519
Mouse UGT1A1	AAP48593.1
Mouse UGT1A2	AAI45970.1
Mouse UGT1A5	NP_964005.2
Mouse UGT1A6	NP_659545.2
Mouse UGT1A9	NP_964006.2
Mouse UGT1A10	NP_964004.1
Mouse UGT1A12	AAP48599.1
Olive baboon UGT1A4	NP_001106089.1
Olive Baboon UGT1A7	NP_001106095.1

Olive Baboon UGT1A8	NP_001106088.1
Rabbit UGT1A4	NP_001082791.1
Rabbit UGT1A6	NP_001082788.1
Rabbit UGT1A7	AAB65795.1
Rat UGT1A1	NP_036815.1
Rat UGT1A2	AAR95631.1
Rat UGT1A3	NP_958827.1
Rat UGT1A5	NP_001034638.1
Rat UGT1A6	AAL67853.1
Rat UGT1A7	AAB18360.1
Rat UGT1A8	AAR95635.1
Rat UGT1A10	NP_958828.1
Rat UGT1A11	AAR95630.1
Rhesus macaque UGT1A1	NP_001028041.1
Sheep UGT1A1	NP_001192076.1
Sheep UGT1A4	NP_001192078.1
Sheep UGT1A3	NP_001192077.1
Sheep UGT1A6	NP_001192075.1
Sheep UGT1A9	ADZ11102.1
Western clawed frog UGT1A6	NP_001107366.1
Wild Boar UGT1A6	NP_001265679.1
Zebrafish UGT1A1	NP_001032505.2
Zebrafish UGT1A2	NP_001166241.2
Zebrafish UGT1A4	NP_001170815.2
Zebrafish UGT1A5	NP_001170811.2
Zebrafish UGT1A6	NP_001170810.2
Zebrafish UGT1A7	NP_001170805.1
Zebrafish UT1B1	NP_001170917.1
Zebrafish UGT1B4	NP_001166239.2
Zebrafish UGT1B3	NP_001170817.1
Zebrafish UGT1B5	NP_001170813.1

Appendix D: List of accession numbers. Accession numbers for 71 UGT1 sequences from characterised and predicted UGTs. Files retrieved from the NCBI database and using Geneious (Biomatters Ltd) sequences were aligned and their relationships inferred using maximum likelihood method, bootstrapped, 500 replicates.

Appendix E

Animal/UGT	Accession Number
African Clawed Frog UGT3A1	NP_001088053.1
Chimp UGT3A2	NP_001088053.1
Cow UGT3A1	NP_001069555.1
Gibbon UGT3A2	XP_003274394.1
Horse Predicted UGT3A1-like	ENSECAG00000010396
Horse UGT3A1-like*	ENSECAG00000008900
Human UGT3A1	NP_689617.3
Human UGT3A2	NP_777574.2
Marmoset UGT3A1	XP_002763568.2
Marmoset UGT3A2	XP_002745106.1
Mouse UGT3A1	NP_997099.2
Mouse UGT3A2	NP_659094.1
Platypus UGT3A2	XP_001517181.4
Orange UGT3	XP_006469356.1
Rat UGT3A2	XP_008759055.1
Rhesus macaque UGT3A1	XP_001093373.1
Rhesus macaque UGT3A2	EHH26448.1
Silkworm UGT3A1	NP_001161187.1
Toxocara canis UGT3A1	KHN88569.1
Western clawed frog UGT3A2	NP_001005027.3

Appendix E: List of accession numbers. List of accession numbers from characterised and predicted UGTs, retrieved from the NCBI and Ensembl databases. Sequences were aligned and relationships inferred using maximum likelihood, bootstrapped, 500 times, using Geneious software (Biomatters Ltd).

Appendix F1

Appendix F1: Primers used for sequencing. Diagram displays the location in the gene of forward and reverse primers used to sequence ENSECAG00000020628. Primers were designed 400-450pb apart, to create sequence overlap to enable a contig to be constructed. Start and stop codons (highlighted red underlined) were sequenced from the vector into the gene

Appendix F2

ATGGCTCCTGCAATGTTGACCGGCTCCCTTCCTCTATGTGTGTGTCTCCTGCTGACACCCGGCTTTGC

TGACGCAGGCTGGCTGCTGGTGGTACCCATGGATGGGAGCCACTGGTTTACCATGCATTCGGTTGTGG

AGAAACTCATCCACAGAGGGCATGAGGTGGTCATAGTCATGCCAGAGGTGAGTTGGCACATAGAGAAA

TCACTCAATTTTACGGTAAAGACATATTCTACGTTTTACACTCTGGAGGAGCTGGATCCTCAGTTCAA

CATTTTCTCTGAGGCTCACTGGAAAGGTCAGGAACAAAGTTTACTTTCTACGTTGCTGACTTCATCTG

GTGATAGTTTTATTGAACACTTTTACTCACATTGTAGGAGTCTGTTTAATGACGCCAAGTTAGTAAAA

Forward 1

TACCTAGAAGAGAATTCTTTTGATGCGGTCTTTCTGGATCCTTTTGATATGTGTGGCTTCATTGTAGC

Reverse 1

CAAATATTTTTCCCTCCCATCTGTGGTCTTCACCAAGGTAGTAATTTGCCACCATCTTGAAGAGGGTA

CGCAGTGTCCCAGTGCTCCTTCCTATGTTCCTAGATTTCTCTCAGGGTTCCCGGACACCTTGACTTTC

Reverse 4

AGGGAGAGAGTGCGGAAACATATCTTCTACTTTGAGGAATATTTATTTTGCCGCTACTTCATCAAAAA

TGTTTTAGAATTTGCTTCTGAGATTTTCCAAAAGACGGTCACAGAATATGATCTCTTAAGCCATACGT

CAATTTGGTTGTTACGAACTGACTTTGTGTTTGACTATCCCAAACCTGTGATGCCTAACGTGATCTTC

Forward 2

ATCGGCGGTATCAACTGCCATCAGGGAAAGCCACTGACAAAGCCCATCATGCCCAATATGGTTTTTAT

Reverse 2

TGGTGGGATCAACTGTGTGAGCAAAAAACCACTATCCAAGGAGTTTGAAGCCTATGTCAATGCTTCTG

GAGAACATGGAATTGTGGTTTTCTCTTTGGGCTCCATGGTCTCAGAGATTCCGGAGAAGAAAGCGATG

GAAATTGCTGATGCTTTGGGAAAAATACCTCAGACAGTCCTGTGGCGGTACACTGGAACTCCACCACC

AAATCTTTCGAAGAACACAATACTCGTCAAGTGGCTGCCCCAAAATGATCTGCTTGGTCACCCGAAGA

Forward 3

CTCGTGCCTTTATTACACATTCTGGCTCCCATGGTGTATATGAAGGAATCTGCAATGGCGTTCCAATG

Reverse 3

GTCATGATGCCCTTGTTTGGTGATCAGATGGACAATGCAAAGCGCATGGAGACCCGGGGAGCTGGAGT

GTCCTTGAACGTCCTGGAAATGACTTCTGATGATTTAGCAAATGCCCTAAAAACTGTCATCAATGACA

AAAGCTATAAGGAAAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCCGTGGAGCCACTGGAC

CTGGCCGTGTTCTGGGTGGAGTTCGTGATGAGGCACAAGGGGGCCCCGCACCTGCGCCCTGCAGCCCA

CGACCTCACGTGGTACCAGTACCACTCTTTGGACGTGATCGGCTTCCTCCTGGCCGTCGTGCTGGGAG

TCGCCTTCATCGTCTATAAATCTTGTGCCTTCGGCTTCCGGAAGTTCTTTGGGAAAAAAGGGCGAGTT

AAGAAGTCCCACAAATCCAAGACACAGTGAGAAGTGGGTGGGAAGTGA

Appendix F2: Positions of primers for sequencing. Primers used to sequence ENSECAG00000023519 are highlighted red in a blue box. The ends of the genes were sequenced in from the vector to capture the start and stop codons (red underlined).

Appendix F3

ATGGGGAGCCTGCGGGCGCTGCTTCTCATCTCCTCCCTTCTGCCTGGGCTCCTGCTCTCAGAGGCCG

CCAAAATCCTGACTCTGTCCTTGCTGGGTGGAAGCCATTTTCTACTAATGGACCGAGTGTCTCAGAT

TCTTCAAGATCACGGTCATAATGTCACCATGCTTCTCCAGAGAGCAAATTTATTAATACCAGGTTTT

AAAGAGGAGGAAAAATCATATCAAGTTATCACTTGGCTTCCACCTGAAGATTATAACAAAGAATTTA

TGAATTTTTTTGATTCCTTTATGAAAGACGCTTTGGATGGGAGAGACTCATTTGCAGACTTTTTAAA

GTTGATGGAACTATTGAGTCTTCAGTGCAGTCATTTGCTAAAGAGAAATGATATCATGGACTCCTTA

AAGAATGAGAACTTCGACCTGGTGATAGTTGAAAGTTTTGACTTCTGTCCTTTCCTAGTTGCTGAGA

AGCTTGGGAAACCATATGTGTCCATTCTCCCCTCCTCGTTTGATGCTGTGGACTTTGGACAACCAAG

ACCTCTGTCTTATGTGCCAACGTTACATTCCTTCCTGACTGACCATATGGATTTCTGGGGCCGACTA

AAGAATTTTCTGATGTTTTTAAATTTCTCCATGAGGCAACGGCAAATCCACTCTAAATTTGACAACA

CCATCAAGGAGCATTTCCCCGAAGGGTCTAGGCCAGTTTTGTCTCATCTCCTAAAGAAAGCAGAGTT

GTGGTTTGTGAACTCTGACTTTGCCTTTGATTTTGCTCGGCCTCTGCTCCCCAACACTGTGTATGTT

GGAGGCTTAATGGCCAAACCTGTTAAGGCAGTACCTCAAGAATTTGAGAATTTCATTGCCAAGTTTG

GAGACTCTGGTTTTGTTCTTGTGGCCCTGGGCTCCATGATCAGTGGCTCTTCATCCCAAGAATTTCT

CAAGGAGATGAACACTGCCTTTGCTCATCTCCCTCAAGGGGTCTTATGGAGGTGTAAGCCTTCTCAT

TGGCCCAAAGACATCAAATTAGCAGCAAATGTGAAAATTGTGGACTGGCTTCCTCAGAGTGACCTCT

TGGCTCACCCTCACATCCGTCTCTTTGTCACCCATGGTGGGATAAATAGCATCATGGAGGCCATCCA

ACATGGCGTTCCCATGGTGGGGATTCCCGTCTTTGGAGACCAGCCTGAAAACCTGTTCCGAGTAGAA

GCCAAAAACTTTGGTGTCTCTATCCAGTTAAAGCAGATCAAGGCTGAGACACTGTCTCTGAAGATGA

AGCAAGTCATAGAAGACAAGAGGTACAAATCTGCAGCCGTGGCCGCCAGCATCATCAGACGCTCCCA

CCCCCTGACTCCTGCCCAGCGGCTGGTGGGCTGGACCAACCACATCCTGCAGACAGGGGGTGCAGCG

CACCTCAAGCCCCATGCCTTCCAGCAGCCATGGTATGA

Forward 1
Reverse 1
Forward 2
Reverse 2

Appendix F3: Primer locations to sequence ENSECAG00000008900. Forward and reverse primers were used to sequence ENSECAG000000008900, start and stop codons captured by sequencing from the vector.

Appendix F4

ATGGCGTCTGAGAAATGGGTTTTGGCAACTCTGCTGCTGCAGCTCTGCTTCACTGGCCATGGATTCT

GTGGGAAGGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTCAATCTAAAGGTTATTCTGGAGGA

ACTTACGGAAAGGGGCCACGAGGTGACTGTGTTGGTTTCTCCATATAATTTCATCATTGACTACAGC

AAGCCTTCCGCACTGAATTTCGAGGTGATTCCTGTGCCACAGGAGGGAGAGACTGCTGCAAATTCAC

TAAATGACTTTTTAGACTTGGCTACCAATGTCATACCAACATTGTCACTCTGGCAGTCTGCAAGAAA

ACTGCAAGAGTTCTTTCTTCAAATTACTGGACATTTAAAACTTCTGTGTGAGAGTGTAGTCTACAAC

Forward 1 →

CAGACGTTCATGAAGAAACTCCAGGAAACCAACTACAATGTAGTGGTTATAGACCCTGTGATGCCCT

← Reverse 1

GTGGAGAGCTGATTGCTGAGTTGCTGGAAGTCCCTTTTGTGTACACGCTAAGGTTCTCTCTGGGTGG

TATTATTGAGAGATACTGTGGGAAAATTCCAGCTCCACCTTCCTACGTGCCTGTTGCCATGGGAAAA

CTAGCAGACAAAATGACCTTTCTGCAAAGGGTAAAAAATTTATTGTTTTCTATTTTATTTGACTTTT

TCCTCCACCAATATGACTTTCAGCTTTGGGACCAGTTTTACAGTGAAGTATTAGGAAGACCCACTAC

ATTATGTGAGATTATGGGGAAAGCAGAAATTTGGCTAATACGGACATATTGGGATTTTGAATTTCCT

Forward 2 →

CGTCCATACTTACCTAATTTTGAGTTTGTAGGAGGATTGCATTGTAAACCTGCCAAACCGTTACCTA

Reverse 2 ←

AGGAAATGGAAGAATTTGTCCAAAGTTCAGGTGAAGATGGTGTTGTGGTGTTTTCTCTGGGGTCAAT

GGTTAAAAATCTCACAGAAGAAAAAGCCAATCTCATTGCCTCAGCCCTCGCCCAGATTCCACAGAAG

GTTTTATGGAGGTACGCAGGAAAGAAACCAGCCACATTAGGAGCCAATACTCGGCTCTATGACTGGA

TGCCACAGAATGATCTTCTTGGTCATCCCAAAGCAAAAGCTTTTATCACTCATGGTGGAACCAATGG

TATCTATGAAGCTATCTATCATGGGGTCCCTATGGTGGGAGTTCCTATGTTTGCTGATCAGCCTGAT

AACATTGCTCACATGAAGGCCAAAGGAGCAGCTGTGGAGGTGGACATAAACACAATGACAAGTGAAG

Forward 3 →

ATTTGCTCAATGCCTTGAGAACAGTAACTAATGATCCTTCTTATAAAGAGAATGCTATGAGATTATC

Reverse 3 ←

AAGAATTCACCATGATCAGCCAATGAAGCCTCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATG

CGCCACAAAGGAGCCAAACACCTGCGGCCAGCCGCCCATGACCTCACTTGGTTCCAGTACCACTCTT

TGGATGTGATTGGGTTCCTGCTGGTCTGTGCAGCAGCTGCTATATTCCTGGTCGCAAAATGTCTTTT

GTTTTCTTGTCGAAAATTAGGTAAAACAGGAAAGAAGAAAAAGAAGGAATAGATCTTTCTAAGTTTT

GTAAAGGCCTGAAAGGGCAATCCTGTTCATTCTAGCCACAATGACCTTAATAA

Appendix F4: Primer locations for sequencing ENSECAG00000014362. Forward and reverse primers used to sequence the equine gene, start and stop codons (red, underlined) sequenced from the vector in.

Appendix F5

ATGATGAGGCCACGGGTGCTGCTTCTCATCTGCTTCCTCCTACCTGGGCTCCTGCCCTCAGAGGCTGC

CAAAATACTGACTGTGTCCTTGGTGGGTGGAAGCCATCATCTACTAATGGACCGAGTGTCTCAGATTC

TTCAAGATCATGGTCATAATGTCACTGTGCTTCTCCAGGAAGGAAATGTATTGATACCAGGTTTTAAA

GAGGAGGAAAAATCATACCAAATTGTCACTTGGTTTCCACCTGAAGATGATTTCAAAGAATTTTTGAA

GTTTTGTGAGTTCTTTATGGAAGAAGCTTTGGCTGGCAGAGACAAATTTGAAAACTTTTTAAAATTCA

TGGAACTACTGGGACTTCAGTGCAGTCATTTGCTAAAGAGAAATGATATCATGGACTCCTTAAAGAAT

GAGAACTTTGACTTGTTATTTGTTGAAGGATTTGACTTGTGTTCTCTCCTGGTTGCTGAGAAGCTTGG

GAAACCGTTTGTCTCCATTATTTCCACCTCGTTTGGCTTTATTGATTTTGGACTACCAAGCCCCCTCT

CTTATGTGCCAGTATTTGATTCCTTGCTAAGCGACCGCATGGACTTCTGGGACAGAGTGAGGAACTTC

CTGAAATTTTTTGATTTCTCCATGAAGCAATGGCAAATTCACTCTACATTTGACAACACCATCAAGGA

GCATTTCCCCGAAGGCTCTAGGCCAGTTTTGTCTCATCTCCTAAAGAAAGCAGAGCTGTCGTTAGTTA

ACTCTGACTTTGCCTTTGATTTTGCTCGGCCTCTGCTCCCCAACACTGTGTATGTTGGAGGCTTAATG

GCCAAACCTGTTAAAGCAGTACCACCAGAATTTGAGAATTTCATTGCCAAGTTTGGAGACTCTGGTTT

CATCCTTGTGGCCCTGGGCTCTGTGGTGAACATCTTTCAGTCCCAGTATGTTTTCAAGGAGATGAACA

GGGCCTTTGCTCATCTACCTCAAGGGGTGATATGGAAGTGTAATCCTTCTCATTGGCCTGAAGACATC

AAATTGGCAGAAAATGTGAAAATTGTGGACTGGCTTCCTCAGAGTGACCTCCTGGCTCACCCTCGCAT

CCGTCTCTTTGTCACCCATGGTGGGATAAATAGCATCATGGAAGCCATCCAACATGGTGTGCCCATGG

TGGGGATTCCCTTCTTTGGTGACCAGCCTGAAAATCTGTTCCGGGTAGAAGCCAAAAACTTTGGTGTC

TCTATCCAGGTAAAGCAGATCAAGGCTGAGACACTGGCTCTGAAGATGAAGCAAGTCATAGAAGACAA

GAGGTACAAATCTGCAGCCGTGGCCGCCAGCATCATCAGGCGCTCCCACCCCCTGACTCCTGCCCAGC

GGCTGGTGGGCTGGACCAACCACATCCTGCAGACAGGGGGTGCAGCGCACCTCAAGCCCCACGCCTTC

CAACAGCCATGGTATGAACAGTACCTGCTCGATGTCTTCTTGTTCCTGCTGGTGCTCACCGTGGGCAC

CATGTGGCTCTGTGGGAAGCTGCTGGGCATGGTGGCCAGGTGGCTGTGTGGGGCCAGGAAGCTGAAGA

AGGCCTGA

Forward 1

Reverse 1

Forward 2

Reverse 2

Appendix F5: Primers for sequencing. Forward and reverse primers used to sequence ENSECAG00000010396, start and stop codons were captures by sequencing from the vector.