Identification, Characterisation and Quantification of Proteins used in Chemical Communication

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Jennifer M Unsworth

2014

Acknowledgments

Firstly I would like to thank my supervisors Professor Rob Beynon and Professor Jane Hurst for giving me the opportunity to do this PhD and for their guidance, enthusiasm and patience over the past four years. I would also like to thank Professor Elke Zimmerman and Professor Ute Radespial for allowing me to collaborate with them on the mouse lemur work.

Special thanks goes to the following: the amazing Lynn who has helped me out with pretty much everything over the past 4 years you have also been a great friend to me which I am incredibly grateful for. To Stu and Deb for their help and guidance with the QconCAT work and Lupe for answering my MUP related questions and for being so approachable and letting me chat to you about my data. Thank you also to Rachel for allowing me to collaborate with you on the harvest mouse work ☺ and to Amanda for collecting samples for me whenever I needed them and for providing musical entertainment in the lab. Many thanks to Duncan for teaching me *de novo* sequencing and the principles of anion exchange chromatography and to Philip for his technical assistance with the instruments. I apologise for all the times I woke you up from you naps with the depressing sentence – the *E. Coli* hits are down...

To my lovely friends Vicki and Yvonne – I honestly don't know what I would have done without all your support, I dread to think how much pork related produce, teacakes and wine we have gone through over the past couple of years.

A big thank you to my family – Mum, Dad, Steph and Phil for your constant love and support. I am very lucky to have you! Much love to Bubbles, Mo and Gabriella (and Nibbies) for listening to my thesis woes at all hours of the night when all you wanted to do was play on your wheels and eat seeds. Finally I would like to thank my partner Chris for your love, hugs and cups of tea and for always encouraging and reassuring me. I imagine you are just as pleased as I am that this thesis is finally written!

Abstract

Identification, characterisation and quantification of proteins used in chemical communication.

Most animals have excretory systems to remove soluble waste. In humans soluble waste is mainly excreted through the urinary system. Kidneys, urinary bladder and urethra make up this system and are responsible for the production of urine by filtration, reabsorption and secretion. Under normal circumstances urine contains water, creatinine, urea and salts. In humans, the presence of elevated levels of protein or glucose is indicative of medical conditions such as impaired kidney function and diabetes. Some animals are an exception to this. Rodents such as the house mouse (*Mus musculus*), Norway rat (*Rattus norvegicus),* bank vole (*Myodes glareosin*) and Roborovski hamster (*Phodopus roborovskii*) excrete substantial amounts of protein in their urine yet their renal function remains intact. These proteins belong to the lipocalin family and play an essential part in chemical signalling. Their size (18-19kDa) allows them to escape from being filtered out of the urine during the ultrafiltration step resulting in their excretion in the urine.

Many of these proteins share a high sequence identity and genomic data is often incomplete or absent. One aspect of this thesis looks at developing a quantification method for a set of highly homologous lipocalins in mice. Another was to characterise and identify proteins excreted in the harvest mouse (*Micromys minutus*) and mouse lemur (*Microcebus*) in the absence of genomic data and see if they are related to the lipocalin family or if they belong to a completely different group of proteins.

Using mass spectrometric techniques a method to quantify major urinary proteins (MUPS), lipocalins found in mice, was developed and implemented. A quantification concatemer (QconCAT) was designed to do this and was based on genomic data from the laboratory strain of mouse C57BL/6. MUP isoforms were successfully quantified in both male and female C57BL/6 mice. The QconCAT strategy was also used to assess MUP production during the estrous cycle in female

mice. Females express more MUP during the estrous stage with a decline in expression seen during the proestrous.

For the second part of this thesis, lipocalin expression in the harvest mouse (*Micromys minutus*) was investigated. Urine samples collected from male and female harvest mice revealed proteins approximately 18-19 kDa expressed in both sexes. The concentration of protein in urine was much lower than that observed in other rodents. Alternative areas of protein excretion were explored and revealed the same protein to be excreted in much higher concentrations from the saliva and/or paws. Again mass spectrometry was employed to identify and characterise these proteins. A preliminary discovery analysis identified proteins that shared high homology with other lipocalins including MUPS and odorant binding proteins. Intact mass analysis also confirmed the presence of three abundant proteins in both males and females. Anion exchange chromatography was used to separate the proteins for *de novo* sequence analysis which confirmed that harvest mice excrete proteins belonging to the lipocalin family.

The final section of this thesis examines characterising protein expression in the mouse lemur (*Microcebus*). Although they are classed as primates not rodents, mouse lemurs are known to respond to urinary chemosignals from their conspecifics. Urine samples were collected from two species of mouse lemur - *Microcebus murinus* and *Microcebus lehilahytsara.* As mouse lemurs have a specific breeding season samples were collected both in and out of season. Some of the male mouse lemurs from both species expressed a large amount of protein during reproductive season. No protein was observed in females. Intact mass analysis identified a protein at 9388 Da in the *M. murinus* and 9418 Da in the *M. lehilahytsara.* Unlike many members of the rodent family who excrete large quantities of lipocalins, *de novo* sequencing confirmed this protein to be a member of the Whey Acidic Protein family (WAPS). WAPS are expressed across many lineages and have a variety of functions including antibacterial and antifungal action, protease inhibition, tumour suppression and anti-inflammatory activity. No protease inhibition by the mouse lemur protein was observed and further studies will need to be established to determine the biological function of this WAP.

Table of Contents

List of Figures

List of Tables

Appendices

Supplementary data A	Extracted ion chromatograms of heavy:light peptide pairs used in the quantification of B6 male and female mice.
Supplementary data B	MS/MS spectra used to determine the harvest mouse protein sequences
Supplementary data C	MS/MS spectra used to determine the mouse lemur protein sequences

Supplementary data can be found on the disc located in the back of this thesis.

Abbreviations

2D-PAGE	Two dimensional polyacrylamide gel electrophoresis
°C	Degrees centigrade
µg	Microgram
µl	Microlitre
A	Alanine
AEX	Anion exchange chromatography
Ala	Alanine
AOB	Accessory olfactory bulb
AQUA	Absolute quantification
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
Asp N	Endoproteinase Asp N
BAPNA	N_{α}-Benzoyl-L-arginine 4-nitroanilide
BLAST	Basic Local Alignment Sequence Tool
BSA	Bovine serum albumin
C	Cysteine
$CaCl_2$	Calcium Chloride
CID	Collision induced dissociation
C-terminus	Carboxy-terminus
Cys	Cysteine
D	Aspartic acid
D	Diestrus
Da	Dalton
D-P	Diestrus-Proestrus
DTT	Dithiotreitol
E	Estrus
E	Glutamic acid
E-M	Estrus-Metestrus
E. coli	Escherichia coli
ELISA	Enzyme-linked immunosorbent assay
ESI-MS	Electrospray ionisation mass spectrometry
ESP	Exocrine gland secreting peptide
ETD	Electron transfer dissociation
F	Phenylalanine
GC-MS	Gas chromatograph mass spectrometry
GG	Grueneberg ganglion
Gln	Glutamine
Glu	Glutamic acid
Glu C	Endoproteinase Glu C
H	Histidine
HCD	Higher-energy collisional dissociation
HCl	Hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
I	Isoleucine
ICAT	Isotope-coded affinity tags
IEF	Isoelectric focusing
Ile	Isoleucine
K	Lysine

kDa	Kilodalton
kV	Kilovolts
L	Leucine
L	Litre
LC-MS	Liquid chromatography-Mass spectrometry
Leu	Leucine
Lys	Lysine
Lys C	Endoproteinase Lys C
M	Metestrus
M	Methionine
M	Molar
M-D	Metestrus-Diestrus
MALDI	Matrix assisted laser desorption ionisation
MALDI–TOF-MS	Matrix-assisted laser-desorption ionization–time of flight
MaxENT	Maximum Entropy Modeling
MES	2-(N-morpholino) ethanesulphonic acid
Met	Methionine
mg	Milligram
MHC	Major histocompatibility complex
mM	Millimolar
MOB	Main olfactory bulb
MOE	Main olfactory epithelium
MS	Mass spectromtery
MS/MS	Tandem mass spectrometry
MUP	Major urinary protein
m/z	Mass to charge ratio
N	Asparagine
Native PAGE	Native polyacrylamide gel electrophoresis
NMR	Nuclear magnetic resonance spectroscopy
N-terminus	Amino-terminus
OR	Olfactory receptor
OSN	Olfactory sensory neuron
P	Proline
P	Proestrus
P-E	Proestrus-Estrus
Phe	Phenylalanine
pI	Isoelectric point
PMF	Peptide mass fingerprinting
Pr:Cr	Protein: creatinine ratio
Pro	Proline
PSAQ	Protein standard for absolute quantification
Q	Glutamine
QconCAT	Quantification concatamer
Q-peptide	Quantification peptide (from QconCAT)
Q-ToF	Quadrupole/Orthogonal Time of Flight Mass Spectrometer
R	Arginine
rpm	Revolutions per minute
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	Stable isotope labelling by/with amino acids in cell culture
SO	Septal organ of Masera
T	Threonine
Thr	Threonine

Tris	2-amino-2-hydroxymethyl propane-1, 3-diol
Trp	Tryptophan
Tyr	Tyrosine
UPLC	Ultra performance liquid chromatography
V	Valine
V1R	Vomeronasal family 1 receptor
V2R	Vomeronasal family 2 receptor
Val	Valine
VNO	Vomeronasal organ
VSN	Vomeronasal sensory neuron
W	Tryptophan
WAP	Whey acidic protein
WFDC	Whey acidic four disulphide core protein
Y	Tyrosine

Chapter 1: Introduction

1.1 Animal communication

Animal communication can be defined as the transfer of information from one animal to another that results in a behavioural change in the receiver. Communication is often in the form of visual, auditory and olfactory cues. Visual signals frequently include gestures, facial expressions, body posture and mimicry. Herring gull chicks exhibit a begging response upon presentation of the parents' beak which signals feeding time to the chicks (Tinbergen and Perdeck, 1951; Ten Cate *et al.,* 2009). Aggressive, dominant wolves will often have high body posture and raised hackles while submissive ones lower their tails and ears and carry their bodies low (Sillero-Zubiri, 2004). Others examples of visual signalling include peacocks attracting peahens by erecting and displaying their impressive trains and chameleons change colour to reflect their physiological state and intentions to conspecifics (Stuart-Fox and Moussalli, 2008).

Auditory signals are also regularly used as a form of communication between animals. Male pacific walruses use acoustic displays to attract a female during the breeding season and to warn other competing males (Fay 1982; Stirling *et al.,* 1987). Lactating female guinea pigs respond to vocal calls from their pups (Kober *et al.,* 2007) while ultrasonic calls in red deer play a key role in sexual behaviour (Pomerantz and Clemens, 1981).

While visual and auditory cues are frequently used by most animals, olfactory signals are the primary source of communication. The advantage of olfactory signalling is information about the depositor is still able to be detected after they have left the scene, which is particularly useful for defending territories etc. Olfactory signals are deposited in the form of scent marks and provide more detailed information about the depositor.

1.2 Scent marking

Scent marking is a behaviour by which glandular secretions are deposited on the ground or onto objects in an animal's environment (Johnson, 1973). These secretions are often in the form of urine, faeces or are excreted from specialised scent glands. Most animals display scent marking behaviour and use it as a form of communication between conspecifics (Brown and MacDonald, 1985) with glandular secretions, urine and faeces being placed in noticeable places in their territories/home ranges, often in lines or along paths and boundaries (Gosling and Roberts, 2001). By placing scent marks in this manner, an individual can define and defend their territory from invading conspecifics. These scent marks contain information such as sex, species, individual identity, social status and the prescence, age and location of the marking (Brown and MacDonald, 1985; Hurst, 1993; Gosling and Roberts, 2001; Hurst *et al.,* 2001; Petrulis, 2013). Examples of marking behaviour by various mammals are outlined on Table 1.1.

Males usually mark more than females with dominant males marking more than others ensuring the markings stay fresh (Gosling, 1982). Although scent marking has advantages such as conveying information about the signaller in their absence and requiring less energy to produce than an acoustic signal, it does involve significant cost in time and risk. The reasons behind scent marking are unclear but there are several hypotheses. The first hypothesis is an individual tends to place their markings around the edge of a territory; the markings serve as fence or warning sign for conspecifics not to enter the territory. However most species studied will cross into territories despite the markings (Gosling and McKay, 1990) the exceptions being male moles and beavers who avoid marked sites (Gosling and Stone, 1990; Sun and Muller-Schwarze 1998). The second hypothesis is a trespasser will learn the scent of the signaller so if they encounter the owner of the scent they will recognise this and avoid fights they are likely to lose (Gosling and Roberts, 2001). The third hypotheses propose that animals establish boundaries with major competitors and therefore prevent costly disagreements between territory owners (Brashares and Arcese 1999; Gosling and Roberts, 2001).

Animals can counter-mark existing scents. This can be in the form of over-marking where an individual will partially or completely cover the existing scent of another individual or by adjacent marking where an individual will mark nearby the existing scent of another individual (Johnston *et al.,* 1994). Outcomes of over-marking include scent blending, individual scents that have been over-marked remain distinct or scent masking which has been observed in male Syrian hamsters who ignore earlier marks and only treat the most recent top marking as "familiar" (Johnston *et al.,* 1994). Territory owners will often verify their ownership by over-marking an intruder's scent. In mice dominant males appear not to over-mark their own scent or that of a genetically identical individual (Hurst, 1990; Nevison *et al.,* 2000).

Table 1.1 Examples of scent marking behaviours in mammals

Species	Scent marking behaviour	References
Hyaena (*Hyaena hyaena)*	Create 'scent posts' by rubbing their anal scent glands over tall grasses and shrubbery to relay information such as sex, familiarity, identity and possibly sexual status.	Drea *et al.,* 2002; Burgener *et al.,* 2009
Wolf (*Canis lupus)*	Alpha males urinate by raising one of their back legs to mark their territory. This is different to normal urination in which the animal uses a squatting technique. The alpha female will often counter mark where her mate has just urinated.	Peters and Mech, 1975; Briscoe *et al.,* 2002
Red squirrel (*Sciurus vulgaris)*	Uses secretions from oral glands for kin recognition.	Mateo, 2006
House mouse (*Mus musculus)*	Use urinary scents to provide a broad range of individual-specific information such as dominance, health and reproductive status and territorial information.	Rich and Hurst 1998; Beynon and Hurst, 2004; Hurst, 2009
Rabbit (*Oryctolagus cuniculus)*	Both male and female rabbits display "chinning" which is when the animal rubs its chin on objects or conspecifics in order to deposit secretions from the submandibular scent glands.	Mykytowycz, 1965; Arteaga *et al.,* 2008
Giant panda *(Ailuropoda melanoleuca)*	Scent mark off ground e.g. tree trunks. Use both urine and anogenital gland secretions to communicate individual identity, sex, reproductive condition, age and competitive status.	Swaisgood *et al.,* 1999; Swaisgood *et al.,* 2000; Hagey and MacDonald, 2003
Tiger (*Panthera tigris)*	Urine spraying and scraping with deposits of urine, faeces and anal gland secretions are the primary forms of marking. Other forms include clawing, cheek rubbing and vegetation flattening. Males increase the frequency of marking when females are in estrous and when marking their territory.	Smith *et al.,* 1989

Both physiological and behavioural responses have been observed in response to scent marking. In mice and rats, female puberty is accelerated in the presence of male urine and can be delayed by the presence of female urine amongst group-housed females (Drickamer, 1977; Mucignat-Caretta *et al.,* 1995; Novotny *et al.,* 1999). Urine from males will promote aggression in other males and attract females (Novotny *et al.,* 1985; Lacey *et al.,* 2007; Roberts *et al.,* 2010). In female gray short tailed opossums, estrous is only ever induced in response to a male scent mark (Harder *et al.,* 2008). Decreased testosterone levels have been observed in male gray mouse lemurs post exposure to dominant male urine (Schilling *et al.,* 1984). Male giant pandas in response to rival scent marks will significantly increase their sexual motivation and become more interested in estrous females (Bian *et al.,* 2013). Exposure to dominant male urine will suppress aggression, scent marking and production and territorial patrolling in male blackbuck antelopes (Rajagopal *et al.,* 2010).

A scent mark usually contains pheromones which are responsible for the relaying information about the signaller and cause the behavioural and physiological changes observed. Pheromones are described as 'substances which are secreted to the outside by an individual and received by a second individual of the same species, in which they release a specific reaction, e.g., a definite behaviour or a developmental process' (Karlson and Luscher, 1959). Pheromones are separated into two categories – volatile and involatile. Volatile compounds tend to be small molecules while involatile compounds usually include peptides and proteins.

1.3 Volatiles

Volatile pheromones require no extra energy investment by the signaller as they are often by –products of metabolism (Wyatt, 2009). The advantage of a volatile pheromone is that it can be detected even after the depositor has left the scene. The disadvantage is they are lost to the environment quite soon after secretion of the scent mark. Volatile pheromones have been identified in a number of mammals and have been studied extensively in rodents, mice and rats in particular. A number of volatile components identified by Gas chromatography – Mass spectrometry (GC-

MS) analysis are present in the urine of both mice and rats. 2-sec-butyl-4,5-Dihydrothiazole and 2,3-dehydro-exobrevicomin are two pheromones found in male mouse urine and promote inter-male aggression as well as puberty acceleration and estrous synchronisation in females (Novotny *et al.,* 1985; Jemiolo *et al.,* 1986; Novotny *et al.,* 1999). They also bind to involatiles (major urinary proteins) (Novotny *et al.,* 1985; Jemiolo *et al.,* 1986). Male mouse urine also contains 6-hydroxy-6-methyl-3-heptanone and (methylthio)-methanethiol which causes puberty acceleration and an attractant females respectively (Novotny *et al.,* 1999; Lin *et al.,* 2005). Female urine contains 2-heptanone and 2,5 –dimethylpyrazine which both delay puberty with 2-heptanone potentially causing a prolongation of estrous and 2,5 –dimethylpyrazine having the opposite effect and suppressing estrous in grouped females (Novotny *et al.,*1986; Ma *et al.,* 1998). Rat urine contains 2-heptanone and 4- ethylphenol both of which are attractive to females (Zhang *et al.,* 2008). 2-heptanone also serves as a fear pheromone causing anxiety and stress in rats (Sugai *et al.,* 2006; Gutierrez-Garcia *et al.,* 2007). Rat pups also emit dodecyl propionate from their preputial glands which serves as an attractant to their mothers (Brouette-Lahlou *et al.,* 1999).

Volatile pheromones have also been observed in other mammals. During estrous, the urine of female Asian elephants contain high concentrations of (Z)-7-dodecen-1-yl acetate, a sex pheromone that stimulates male sexual behaviour (Rasmussen, 1997). Female bovine urine contains 1-iodoundecane during estrous and serves as an attractant to bulls (Kumar *et al.,* 2000; Archunan and Kumar, 2013). Male black buck antelope urine has three volatile components - 3-hexanone, 6-methyl-5-hepten-2-one and 4-methyl-3-heptanone – all of which are only observed during the dominance hierarchy period by dominant males (Rajagopal *et al.,* 2010). A pheromone emitted by male goats – 4-ethyloctanal, is responsible for the activation of gonadotropin releasing hormone (GnRH) a key hormone in the regulation of estrous and reproduction, in female goats (Murata *et al.,* 2014). Three volatile pheromones have been identified in female buffaloes - 1-chlorooctane, 4-methylphenol and 9-octadecenoic acid. Isolation of the three fractions saw sexual responses such as sniffing and mounting by males in response to two of the

volatiles 4-methylphenol and 9-octadecenoic acid. No responses were observed with 1-chlorooctane (Rajanarayanan and Archunan, 2011).

1.4 Involatiles

The volatile components of a scent mark draw the receiver towards the location of the scent mark and allow them to investigate the markings further. As volatiles are metabolically produced, information such as health and reproductive status of the signaller at the time of the marking is conveyed to the recipient. Involatile pheromones have the advantage over volatile pheromones in that they are more stable and continue to be present in the scent mark for some time. Examples of involatile components are the major urinary proteins (MUPS), the major histocompatibility complex (MHC) peptides and the exocrine gland secreting peptides (ESP). These involatiles portray information such as individual identity and the receiver must make direct contact with the scent mark to collect the required information. The longevity and robustness of these pheromones means the scent mark will not be mistaken as belonging to another individual.

1.4.1 Major urinary proteins

The excretory system is responsible for removing excess and unwanted materials from an organism to prevent damage to the body and to maintain homeostasis. Most animals have excretory systems to remove soluble waste. In mammals soluble waste is mainly excreted through the urinary system (Kardasz, 2009). Kidneys, urinary bladder and urethra make up this system and are responsible for the production of urine by filtration, reabsorption and secretion.

The production of urine begins with an ultra filtration step. Filtration is one of the main functions of the kidneys and uses special filtration units known as glomeruli, which line the capillaries that make up the glomerulus (Ronco, 2007). The filtration step is aided by a blood pressure difference between two arterioles –the afferent arteriole which supplies blood to the glomerulus and the efferent arteriole in which the blood exits the glomerulus (Atherton, 2012). This blood pressure difference between the two arterioles results in small molecules such as water, sodium

chloride, urea and glucose being forced through the glomerular capillaries to form a fluid called glomerular filtrate. The capillaries have a low permeability to plasma proteins, such as albumin, so the passage of large molecular weight products are restricted (Larina *et al.,* 2013).

The majority of the glomerular filtrate is then reabsorbed back into the blood as it passes through the renal tubes. This enables the body to retain most if its nutrients (Rubenstein *et al.,* 2012). At the same time waste substances are then secreted into the tubular fluid, such as potassium ions, ammonium ions, creatinine, urea and drug metabolites (Atherton, 2006) leading to the production of urine. This not only removes excess amounts of these substances but also helps maintain a healthy blood pH (approximately 7.4) (Atherton, 2006). Urine is then excreted via the ureters, bladder and urethra.

Under normal circumstances urine contains water, creatinine, urea and salts. In humans, presence of elevated levels of protein or glucose is indicative of medical conditions such as impaired kidney function and diabetes (Bailey, 2011; Naresh *et al.,* 2013). The urine content of some rodents such as the house mouse (*Mus musculus*) and the Norway rat (*Rattus norvegicus)* has been widely studied and it is well known that these rodents excrete a substantial amount of protein in their urine yet their renal function remains intact. These proteins are known as major urinary proteins (MUPs) and play an essential part chemosignalling (Beynon and Hurst, 2003). Mice in particular excrete high concentrations of MUP (up to 20 mg/ml per day) which is a huge energy investment for each individual animal. The size (18-19kDa) of these MUP proteins allows them to escape from being filtered out of the urine during the ultrafiltration step resulting in their excretion in the urine (Neuhaus, 1986).

MUPS belong to the lipocalin family of proteins. Lipocalins are a large group of extracellular proteins. They are transport proteins that bind small hydrophobic molecules. They also have other molecular recognition properties that include binding to specific cell-surface receptors and the formation of complexes with

soluble macromolecules (Flower, 1996; Flower *et al.,* 2000). Lipocalins have highly conserved structures yet vary quite drastically on the sequence level. All lipocalins have eight β-strands which together form a cup-shaped anti parallel β-barrel which surrounds an internal ligand binding site (Flower *et al.,* 1993). The eight strands of the barrel are connected by β-hairpin loops, the first loop being slightly larger than the rest and forms a lid by folding back to close off the internal binding site (Flower *et al.,* 1993; Flower *et al.,* 2000). The other end is closed off by a short N-terminal α-helical domain (Lucke *et al.,* 1999) (Figure 1.1).

The structure of mouse MUPS has been defined by x-ray crystallography (Bocskei *et al.,* 1991; Bocskei *et al.,* 1992; Lucke *et al.,* 1999; Timm *et al.,* 2001). Similar to other lipocalins, MUPS have eight anti parallel β-strands that form a single β-sheet surrounding a ligand binding cavity. The binding cavity contains several hydrophobic residues with the highly conserved tryptophan residue (Try 19) at the centre of the cavity (Flower *et al.,* 1993)

MUP ligand binding

Mouse MUPS bind a number of volatile components in their hydrophobic cavity including the male specific volatile pheromones mentioned in section 1.3 - 2-sec-butyl-4,5-Dihydrothiazole, 2,3-dehydro-exobrevicomin and 6-hydroxy-6-methyl-3-heptanone (Bacchini *et al.,* 1992; Robertson *et al.,* 1993; Novotny *et al.,* 1999). Fractions of MUP isoforms by anion exchange chromatography has shown there is some specificity of ligand binding (Robertson *et al.,* 1993; Armstrong *et al.,* 2005). The male specific isoform known as darcin not only binds more thiazole than the other isoforms, it also binds it more tightly causing slower release of the volatile from the scent mark (Armstrong *et al.,* 2005; Roberts *et al.,* 2010). Despite functional genes being present in both sexes, darcin is only expressed in male mice (Mudge *et al.,* 2008). Behavioural studies have revealed females are most attracted to the darcin component of male urine. MUPs were separated using anion exchange chromatography and females were exposed to each individual fraction.

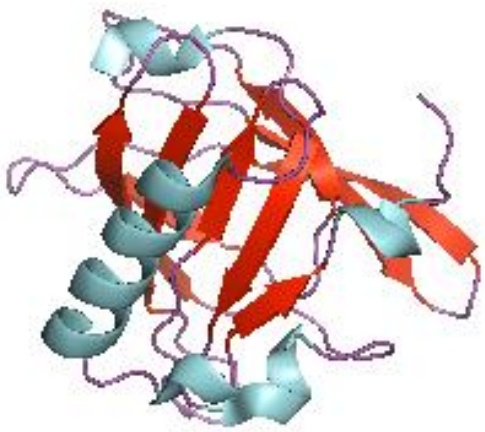


Figure 1.1 The tertiary structure of a lipocalin
A ribbon diagram demonstrating the 3D structure of aphrodisin, a lipocalin observed in female Syrian hamsters, as well as secondary domains. The β sheet forming the β barrel is highlighted in red and the α helix is illustrated in blue. The tertiary structure was generated using PyMOL molecular visualisation software (Schrodinger, Inc).

All females showed the most interest in the darcin fraction (Roberts *et al.,* 2010). To test whether this response was darcin or thiazole related a recombinant form of darcin was produced and presented to the female mice (Roberts *et al.,* 2010). Females were equally as attracted to the recombinant darcin suggesting darcin protein itself acts as a sex pheromone. As darcin is a single protein that is not polymorphic between males, it cannot provide the individual scent specific signal that females require to recognize a particular male. Recognition of an individual male is a result of a learned attraction by the females to the airborne volatiles produced by individual males. This learned attraction by females is stimulated by direct contact with darcin and results in the female learning and becoming attracted to the airborne odours of a specific individual but no to that of other males (Roberts *et al.,* 2010; Roberts *et al.,* 2012).

X-ray crystallography has been used to observe the interaction of MUPs and bound ligands (Figure 1.2) (Bocskei *et al.,* 1992; Timm *et al.,* 2001). Two pheromones 2-sec-butyl-4,5-dihydrothiazole and 6-hydroxy-6-methyl-3-heptanone have been shown to bind within hydrophobic cavity at one end of the β-barrel, formed by the side chains of Phe56, Leu58, Leu60, Ile63, Leu72, Phe 74, Met87, Val100, Tyr102, Phe108, Ala121, Leu123, Leu134, and Tyr138 (Timm *et al.,* 2001). Furthermore, the exact orientation of the ligand binding has also been recognized by hydrogen bonding between water molecules and the 2-sec-butyl-4,5-dihydrothiazole nitrogen and the ketone oxygen group in 6-hydroxy-6-methyl-3-heptanone (Timm *et al.,* 2001). It is unclear exactly how the ligands reach the binding site as the cavity is completely surrounded by side chains. A study by Zidek *et al.,* 1999 showed, using NMR relaxation techniques, the backbone flexibility of the MUP protein increases as it binds 2-sec-butyl-4,5-dihydrothiazole. Large conformational changes in the protein allow the ligand access to the binding site and also significantly stabilises the protein-pheromone complex (Zidek *et al.,* 1999). Fluorescent probe studies with various MUP isoforms proved that different amino acid compositions inside the binding pocket led to a decreased binding affinity and fluorescence yield for the probe (Darwish Marie *et al.,* 2001).

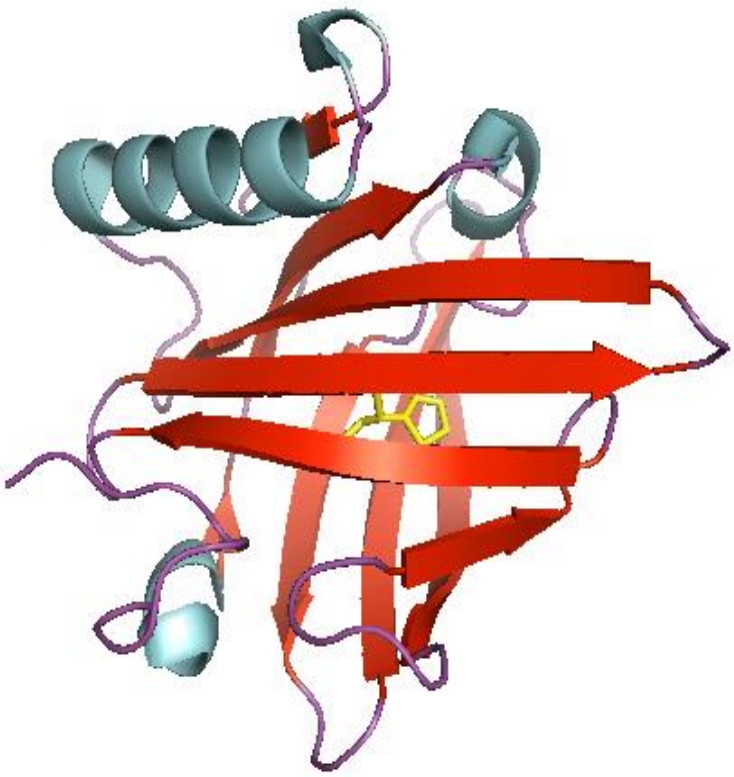


Figure 1.2 The tertiary structure of mouse MUP 1 with ligand.
A ribbon diagram demonstrating the 3D structure of mouse MUP 1 as well as secondary domains. The β sheet forming the β barrel is highlighted in red and the α helix is illustrated in blue. At the centre of the cavity is the male specific ligand (s)-2-sec-butyl -4, 5 dihydrothiazole highlighted in yellow. The tertiary structure was generated using PyMOL molecular visualisation software (Schrodinger, Inc).

One of the disadvantages of volatile ligands is they are lost to the environment after a short period of time and when the MUP ligands are not bound to MUP protein they fade away after only a few minutes (Robertson *et al.,* 2001). MUPS have the ability to delay the release of these pheromones and therefore extend the lifetime of the these chemical signals, with pheromones still being detected by conspecifics up to 24 hours later (Hurst *et al.,* 1998; Humphries *et al.,* 1999).

MUP genetics

MUPS are products of a multigene cluster located on mouse chromosome 4 (Krauter *et al.,* 1982). A gene cluster is usually defined as a set of two or more genes that encode for the same or similar products. They are created by gene duplication and divergence. A gene is accidently duplicated during cell division so its descendants have two copies of the gene which initially code for the same protein. During the course of evolution these genes diverge so the product they code for have different but related functions with genes still being adjacent to each other on the chromosome (Ohno, 1970).

Extensive sequencing of the laboratory strain of mouse C57BL/6 has enabled significant amount of information about this multigene cluster to be acquired. Mudge *et al.,* 2008 identified 19 functional MUP genes and 19 pseudogenes with further analysis by Logan *et al.,* 2008 identifying a total of 21 intact genes and 21 pseudogenes. The multigene cluster could be separated into three groups on the basis of phylogenetic analysis. Phylogenetic analysis is used to observe the evolution of a genetically related group of organisms or study the relationships between a collection of genes or proteins that are derived from a common ancestor. One group of genes within the MUP cluster consisted of pseudogenes. A second group contained functional genes with high homology to each other and a third group contained genes and pseudgenes that were more divergent and have low homology to all other MUP genes. These groups were localised within the MUP locus to two areas, referred to as central and peripheral genes by Mudge *et al* (2008). The central region is flanked at either end by the peripheral region (Figure 1.3). The central region contains 15 functional MUP genes and 16 pseudogenes.

The central genes are extremely homologous and are most likely the result of a number of gene duplications and divergence from one of the older peripheral genes. The timing of the oldest divergence event for the functional central MUP genes is approximately 1.2-2.4 Mya (Mudge *et al.,* 2008). The peripheral genes share less sequence homology and include 6 intact functional genes and 5 pseudogenes (Mudge *et al.,* 2008; Logan *et al.,* 2008). The timing for the oldest divergence for the functional MUP loci in the peripheral region is estimated to be 11.2-22.4 Mya (Mudge *et al.,* 2008).

MUP expression

MUPs are primarily synthesised in the liver and escape glomerular filtration leading to excretion in urine. MUPs account for approximately 99% of protein found in mouse urine. They are synthesised with a 19 amino acid signalling peptide that is cleaved off before entering the bloodstream (Finlayson *et al.,* 1965). Several hormones – testosterone, growth hormones, thyroxine, insulin and glucocorticoids are all thought to contribute to the control of MUP synthesis (Ruemke and Thung, 1964; Knopf *et al.,* 1983; Spiegelberg and Bishop, 1988; Johnson *et al.,* 1995).

Male laboratory mice typically excrete 10-20 mg/ml of protein per day with females excreting much less – approximately 2-10 mg/ml per day (Cheetham *et al.,* 2009). Although highly homologous, major MUP isoforms can be separated using mass spectrometry and isoelectric focussing (Robertson *et al.,* 1996, 1997; Beynon *et al.,*2002; Cheetham *et al.,* 2009; Mudge *et al.,* 2008). For laboratory strains of mice who often belong to one of only two phenotypes, these isoform profiles are virtually identical between individuals of the same sex in the same species. Wild mice profiles are more complex and unique. Both sexes of wild mice excrete up to three times more protein than laboratory strains (Beynon and Hurst, 2004). Substantial variation between unrelated individuals has been observed (Robertson *et al.,* 1997; Beynon *et al.,* 2002). Offspring inherit different MUP haplotypes from their parents leading to large variability. Wild mice will use these variations in MUP profiles rather than MHC peptides (see section 1.4.2) to avoid in-breeding (Sherborne *et al.,* 2007).

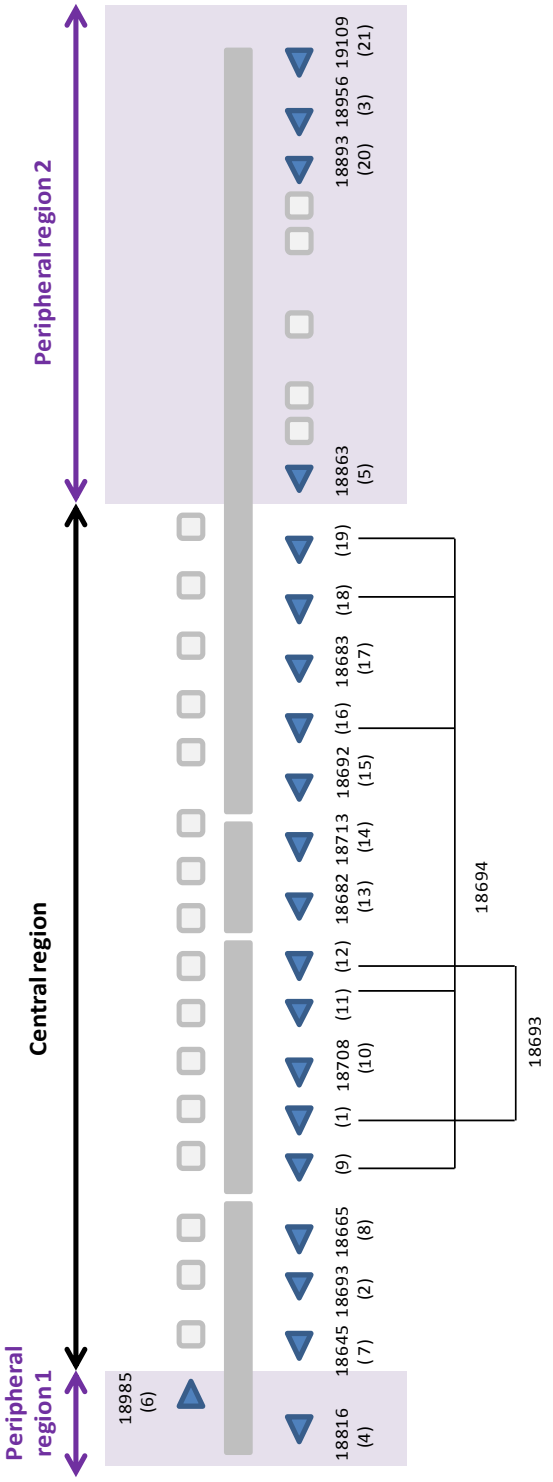


Figure 1.3 C57BL/6 mouse MUP gene cluster
Predicted MUP genes are represented by the blue triangles (1-21), pseudogenes are illustrated by the gray squares. The mature MUP masses are written underneath the blue triangles and are adjusted for the loss of the 19 amino acid signalling peptide and the presence of a disulphide bond. The MUP genes highlighted in the purple box are genes that do not cluster closely together and are termed peripheral genes. In the central region are a cluster of MUP genes that are grouped closely together after phylogenetic analysis. The MUP locus figure was adapted from Mudge *et al.*, 2008.

MUP and lipocalin expression in non-rodents

MUPS have been studied extensively in mice who use MUPS as their primary source of communication. Phylogenetic analysis by Logan *et al.,* (2008) found the last common ancestor of rat and mouse had either a single or small number of MUPs which enabled them to determine the extent of MUP gene expansions across non-rodent lineages. Of the sequenced genomes available, they were able to identify genes in different species that evolved from a common ancestral gene by speciation (orthologues) and contiguous genomic sequence spanning the interval between the genes in nine additional placental mammals. Pigs, dogs, bush babies, macaques, chimpanzees and orang-utans all have one functional MUP gene. Humans have a single MUP pseudogene containing a mutation that causes mis-splicing, rendering it dysfunctional (Table 1.2, Figure 1.4).

Interestingly, two of the nine genomes did reveal further examples of lineage specific expansions (Figure 1.3). Three MUP paralogues were identified in the horse with the product of one of these previously isolated from dander and sublingual salivary glands (Gregoire *et al.,* 1996). Identified as a major horse allergen, this protein has been used to identify further expression in the submaxillary glands and liver. The gray mouse lemur was also found to have at least two MUP gene paralogues and one possible pseudogene.

Protein expression arising from these MUP genes has also been observed in the pig. Expression of a salivary lipocalin that binds sex pheromones in the submaxillary gland of male pigs has been observed (Marchese *et al.,* 1998; Loebel *et al.,* 2000). Dogs also express two lipocalins in their tongue epithelial tissue and paratoid gland that are also potent allergens to humans (Konieczny *et al.,* 1997; Saarelainen *et al.,* 2004). Cats express a number of allergen proteins one of which FEL D 4 is a lipocalin secreted from the submandibular salivary gland (Smith *et al.,* 2004). Interestingly, this lipocalin is detected through the vomeronasal organ (VNO) of mice and caused defensive behaviours. Also native odour stimuli from other species that did not contain MUP/lipocalins caused no response in mice (Papes *et al.,* 2010).

Table 1.2 A list of functional and pseudo MUP genes in 11 different species. Table was adapted from *Logan et al.,* 2008

Species	Chromosome	Functional genes	Pseudogenes	Total
Mice (*Mus musculus*)	4	21	21	42
Rat *(Rattus rattus)*	5	9	13	22
Pig (*Sus scrofa)*	1	1	0	1
Dog *(Canis lupus familiaris)*	11	1	0	1
Bush baby (*Otolemur agyisymbanus)*	Unassigned	1	0	1
Macaque (*Macaca sylvanus)*	15	1	0	1
Chimpanzee (*Pan troglodytes)*	9	1	0	1
Horse (*Equus ferus caballus)*	25	3	0	3
Mouse lemur (*Microcebus*)	Unassigned	2	1	3
Orang-utans (*Pongo borneo)*	9	1	0	1
Human (*Homo sapiens*)	9	0	1	1

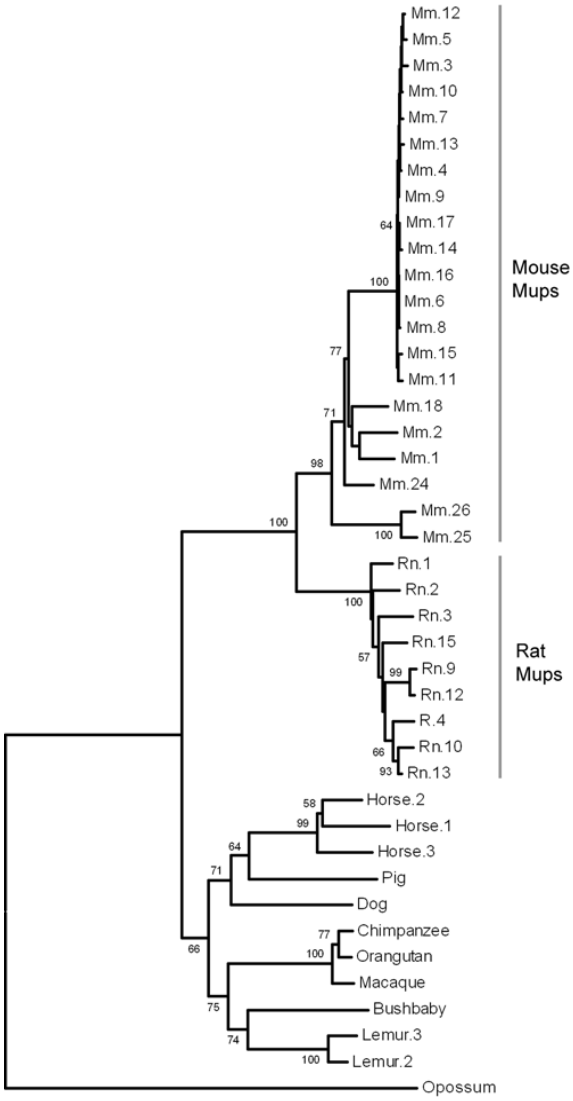


Figure 1.4 Phylogeny of MUP coding sequences in mammals
A rooted phylogenetic tree illustrating the MUP coding sequences in mammals, using a MUP-like cDNA previously described in opossums (Chamero *et al.,* 2007). The expected cDNA sequences generated from open reading frames and aligned. The repeatability was tested by bootstrapping using 1000 replicates and a random seed. Interior branches with bootstrap support 50% are shown. This diagram was taken from Logan *et al.,* 2008

1.4.2 MHC peptides

The major histocompatability complex is a large multigene area containing a number of closely linked highly polymorphic genes that play a crucial role in immunological self and non-self recognition (Klein and Figueroa 1986; Janeway, 1993). The main function of MHC proteins is to transport peptides from within a cell to the cell surface where they are presented to T-cells, which will ignore healthy cells and destroy cells containing foreign protein. Each protein binds to a specific peptide producing a set of uniquely bound peptide-MHC complexes for each individual. These complexes are then discarded from the cell during cell turnover and released into bodily fluids such as blood, saliva and urine (Singh *et al.,* 1997).

In addition to their role in immunity, the MHC is thought to participate in mate selection for many mammals through olfactory cues. Females are thought to choose a mate with a dissimilar MHC type to their own to avoid inbreeding and to improve resistance to infection (Penn and Potts, 1998; Jordan and Bruford, 1998). However, there is limited research into this heterozygote advantage of disease resistance with one study suggesting no immunological advantages (Ilmonen *et al.,* 2007).

MHC-dependant mate choice has been observed in primates such as humans and mouse lemurs. Women were exposed to odours from MHC-dissimilar and MHC-similar males and mostly preferred odours from MHC-dissimilar males. Also the MHC-dissimilar odours often reminded woman of previous partners (Gosling *et al.,* 2008). In a study with gray mouse lemurs, post-copulatory mate-choice has been observed with fathers being more MHC-dissimilar to mothers (Schwensow *et al.,* 2008). MHC-dependant mate choice has also been observed in non-mammals such as fish and birds (Von Schantz *et al.,* 1996; Von Schantz *et al.,* 1997; Olsen *et al.,* 1998; Freeman-Gallant *et al.,* 2003).

In mice, there are reservations over the role of MHC peptides as a signal of individuality as native MHC peptides have never been observed in urine. Mice have receptors for MHC peptides in their VNO and MOE. Synthetic peptides have been shown to cause pregnancy block (Leinders-Zufall *et al.,* 2004; Thompson *et al.,*

2007) and conflicting data surrounding response of VNO sensory neurons to MHC peptides has been published (Chamero *et al.,* 2007; He *et al.,* 2008; Nodari *et al.,* 2008).

1.4.3 Exocrine-gland secreting peptides (ESPs)

More recently ESPS have been observed in rodents such as mice and rats. These peptides are not secreted in urine but are found in extraorbital lachrymal gland, Harderian gland and/or submaxillary gland of with responses to these peptides observed in the vomeronasal organ but not the main olfactory epithelium (Kimoto *et al.,* 2007). They are encoded by a multigene family on chromosome 17 of the mouse and chromosome 9 of the rat and encode proteins of various lengths ranging from 5-15 kDa (Kimoto *et al.,* 2007).

There are 24 functional ESP genes in mice with expression of the various individual ESPs varying between strains. Expression also varies between sexes, a male-specific ESP has been identified in the lachrymal glands of a number of strains. When females make close nasal contact with the facial area or bedding of adult males, stimulation of vomeronasal sensory organs is observed (Kimoto *et al.,* 2005).

1.5 Pheromone detection

In most mammals, pheromones are detected using a dual olfactory system (Figure 1.5). This olfactory system consists of the main olfactory system (MOS) and the accessory olfactory system (AOS). Mammals use either one or both of these system to detect chemosensory clues present in scent marks. The main olfactory epithelium (MOE) is responsible for the conscious perception of odours while the accessory olfactory system is responsible for the detection of pheromones that elicit various behavioural and physiological responses between conspecifics.

1.5.1 Main olfactory system

The MOE is located at the posterior end of the nasal cavity and is mostly made up of olfactory sensory neurons (OSNs). These OSNs send their axons into the main olfactory bulb (MOB) which in turn sends out nerve fibres to the olfactory cortex

before proceeding to higher sensory centres. The OSN contain olfactory receptors which are heptahelical G-protein-coupled receptors (GPCR) that share a significant homology in vertebrates (Rouquier and Giorgi, 2007). The amount of receptors varies between mammals depending on how much olfactory system is required for survival. For instance, humans contain less functional olfactory receptor genes than most other mammals. A rise in pseudogenes from old world monkeys to new world monkeys suggests primates may have lost part of their olfactory ability over time (Rouquier *et al.,* 2000). In contrast to this mice and rats have over 1300 olfactory receptors that bind a broad range of odorants with different affinities (Zhang and Firestein, 2002).

The MOS is not normally associated with pheromone detection; it is usually responsible for detection and differentiation of complex chemical signals that are present in both the physical and social surroundings of individuals. However the individual sensory roles for the two olfactory systems are still unclear. Pheromone detection by the MOS has been reported in female boars. Male boars secrete a volatile steroid androstenone in their saliva that induces lordosis in females (Dorries *et al.,* 1995; Dorries *et al.,* 1997). If the female AOS is blocked off the pheromone continues its effect inducing the female mating stance, suggesting this volatile is detected in the MOS (Dorries *et al.,* 1995; Dorries *et al.,* 1997). Also preovulatory LH surge and ovulation in ewes after exposure to ram odours is thought to involve the MOS. The ewes still experience a surge in LH in response to the rams after blocking off the AOS (Cohen-Tannoudji *et al.,* 1989; Delgadillo *et al.,* 2009). (methylthio) methanethiol (MTMT) in male mouse urine is also detected by the MOS (Lin *et al.,* 2005). The rabbit mammary pheromone 2-methylbut-2-enal present in the milk of the mother encourages nipple-searching behaviour in pups. Removal of the AOS has no affect on the pups' nipple-searching efforts but removal of the MOE eliminates the behaviour completely (Hudson and Distel, 1986).

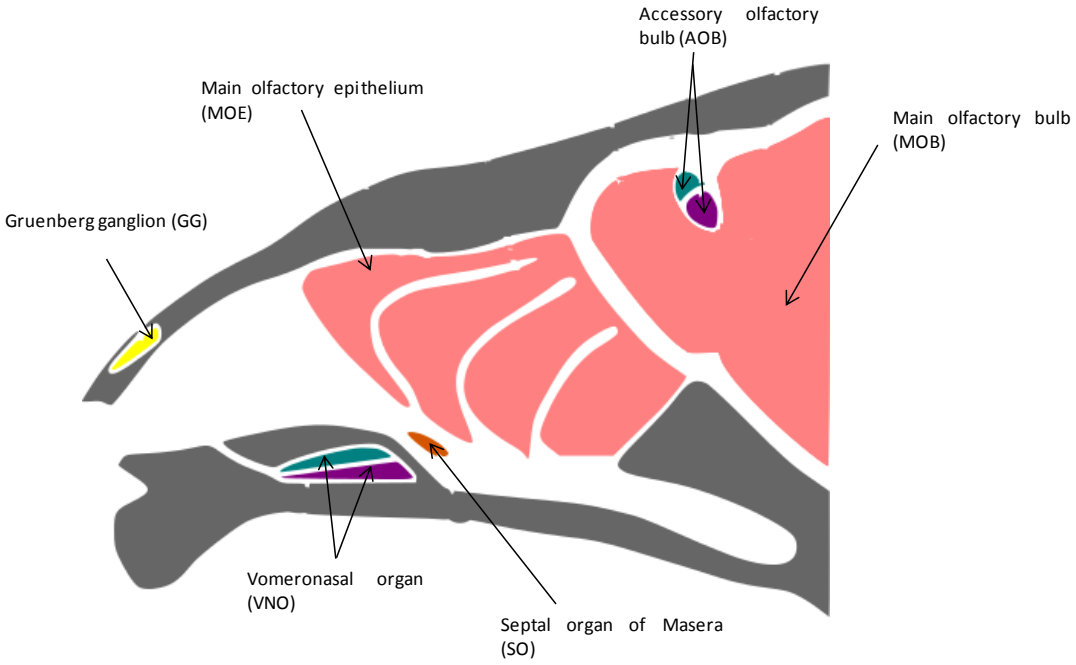


Figure 1.5 Anatomical representation of the mammalian olfactory system.
The location of the various chemosensory subsystems in the mammalian nose. A rodent was used in this example. Adapted from Brennan and Zufall , 2006.

1.5.2 Accessory olfactory system

The AOS is responsible for the detection of the majority of pheromones. A vomeronasal organ (VNO) is based in the vomer between the nose and the mouth and is responsible for detecting stimuli. Like the MOE, the VNO contains sensory receptors whose axons project into the accessory olfactory bulb (AOB). The axons that leave the AOB project into parts of the brain that stimulate aggression and mating behaviour.

The VNO contains two types of sensory receptor – VR1 and VR2 receptors. VR1 receptors detect small volatile molecules and VR2 receptors perceive involatile pheromones such as peptides and proteins (Dulac and Axel, 1995; Matsunami and Buck, 1997). Both receptors belong to two distinct super families of seven trans membrane G-protein coupled receptors. They have different molecular structures and are expressed in different locations in the VNO. VR1 receptors are linked to the G protein Gαi2 and are located in the apical region of the VNO. VR2 receptors are liked to a G protein Gαo and are based in the basal compartment of the VNO (Dulac and Torello, 2003; Mombaerts, 2004). They have a longer N terminal which is thought to be involved in pheromone binding. The V1R receptors transmit projections into the rostral part and the V2Rs into the caudal part of the AOB (Zufall and Leinders-Zufall, 2007).

Identification of VR1 genes has been made easier because of their relatively simple gene structure. At present a complete VR1 gene repertoire has been identified in human, chimpanzee, mouse, rat, dog, cow and opossum with the number of intact genes varying between species (Rodriguez and Mombaerts, 2002; Rodriguez *et al.,* 2002; Grus and Zhang, 2004; Zhang *et al.,* 2004; Young *et al.,* 2005; Grus *et al.,* 2005). However little is known about VR2 receptors and until recently these receptors had only been described in rodents and marsupials. The first functional VR2 receptor genes in a primate, the gray mouse lemur, were observed in a study by Holenbrink *et al.,* 2012.

Many mammals exhibit the Flehmen response to transfer information to their VNO. This involves the animal curling back its upper lip and exposing its front teeth. The animal then inhales over the scent and remains in that position for a few minutes to allow air to transfer from the scent mark to the VNO. In cattle blocking of the VNO significantly reduces inter-individual aggression between males (Klemm *et al.,* 1984). Removing the VNO in male mouse lemurs reduces aggression between males and reduces sexual behaviours (Aujard, 1997). Ewes could not distinguish their own lambs from lambs belonging to others after their VNO was purposely blocked (Booth and Katz, 2000) however conflicting evidence was published by Levy *et al.,* 1995. VNO- dependant pheromone responses in rodents have been studied in more detail and include:

- The Lee-Boot effect – the grouping of female mice in one area causes suppression or a modification of estrous (Van der Lee and Boot, 1955)
- The Vandenbergh effect – the onset of puberty in young female mice is accelerated by non-volatile molecules in adult male urine (Vandenbergh, 1969)
- The Bruce-Lee effect – the presence of a male (or his urine) from a different strain to her mate can prevent egg implantation in females that have recently mated (Bruce, 1960)
- The Whitten effect – synchronised estrous in a group of females in response to urinary cues from a male conspecific (Whitten, 1958)

1.6 Discovery, identification and quantification of pheromones

Biochemical analysis of volatile and non-volatile pheromones requires two very different analytical approaches. Volatile ligands have been the subject of in-depth analysis for a number of years with detection and identification methods for these pheromones well established. The complexity of secretions left by various mammals may complicate isolating individual volatiles and defining roles for each volatile in mammalian behaviour. However, significant progress in volatile pheromone isolation and detection has been made, primarily by GC –MS, and has allowed a greater insight into the biological role of volatile pheromones. As pheromone production is linked to hormonal control, monitoring volatile profiles

for animals in different behavioural situations/endocrine status will enable the identification of possible pheromones by any changes in the GC-MS volatile profiles. These potential pheromones can be analysed further to clarify the chemical structure and allow the design of a biological assay to help confirm the presumed pheromones biological and behavioural role (Novotny, 2003).

Recent advances in proteomics have allowed comprehensive analysis of non-volatile components of scent marks, proteins in particular. The term proteomics was devised in 1995 (Wilkins *et al.,* 1996) and can be defined as the study of the structure and function of proteins. The varying complexities of proteomic methodologies mean behavioural labs can often get a complete insight into the scent complexity and protein components within a scent mark. Significant developments in liquid chromatography – mass spectrometry (LC-MS) has resulted in full identification and characterisation of proteins present in a scent mark in the absence of genomic data. Also, post identification and characterisation of these proteins, quantification methods have been established allowing an assessment of the regulation of the proteins expressed by an animal.

1.6.1 Discovering the complexity of a scent mark

The complexity of a scent mark will determine the analytical approach to be taken. The most commonly used technique to asses complexity is sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Proteins are separated according to their molecular weight followed by visualisation using stain. SDS is an anionic surfactant that binds to polypeptide chains resulting in denaturisation and a negative charge on the proteins. During the electrophoresis step, the proteins are then separated by molecular weight. However, protein shape and folding can also influence where a protein will migrate to on the gel. Darcin is a good example of this as this MUP protein retains some of its shape after treatment with SDS. This results in darcin migrating further down the gel than one would expect.

Complexity is assessed by visual inspection of the gel. A large number of bands would signify the scent mark being of high complexity and the intensity of the band

gives an approximation of the relative abundance of each protein. SDS-PAGE could also be used to efficiently compare protein expression in animals in response to a social status and also compare different individuals and sexes.

Other gel methods that are slightly less simple than SDS-PAGE include isoelectric focussing (Towbin *et al.,* 2001; Friedman *et al.,* 2009) and native gel electrophoresis (Wittig and Schagger, 2008). A two-dimensional separation (2D-PAGE) can allow further exploration of the complexity of a scent mark. Proteins are first resolved by their charge then in a second dimension their molecular weight. They can resolve thousands of proteins and are particularly useful for identifying polymorphisms e.g. MUPs. While 2D gels are provide a good visualization of complexity they do require more protein than SDS-PAGE and can be quite challenging to prepare.

1.6.2 Identification of proteins in a scent mark

Following the assessment of complexity by gel electrophoresis, individual protein bands from the gel are digested with protease such as trypsin to cleave the protein into smaller fragments, referred to as peptides. Peptide masses are obtained using mass spectrometry to produce a peptide mass fingerprint which can be compared to other fingerprints in a database of known fingerprints (Perkins *et al.,* 1999). The database search engine will allocate a score to the peptide mass fingerprint – the higher the score the more likely it is the protein match is true.

If no protein matches are made from the peptide mass fingerprinting analysis, further information about each peptide in the digest can be collected using tandem mass spectrometry (MS/MS). Tandem mass spectrometry involves two stages of mass spectrometry. During the first stage the masses of the peptides are measured and are often referred to as precursor ions, during the second stage the peptides are isolated and fragmented using an inert gas such as helium in a process referred to as collisional induced dissociation (CID) (Figure 1.6). During CID, precursor ions collide with helium gas molecules which lead to activation or excitement of the peptide backbone. The kinetic energy from these collisions is converted into vibrational energy in the peptide ion, which the peptide ion then releases through fragmentation reactions (Bertsch *et al.,* 2009). The fragments of the peptide are termed b and y ions depending on where the charge has been retained. If it is retained on the N-terminus the ion will be labelled a b ion and if the charge is on the C-terminus the ion will be termed a y ion (Figure 1.7). Other common ions found in a CID MS/MS spectrum include a and x ions which are as a result of a C=O loss from b and y ions. A loss of ammonia and water from b and y ions may also be observed in the spectrum.

If working with a species whose genome is known then the fragmentation patterns are searched against all patterns of peptides that can be generated by the proteome of that organism (Cottrell, 2011). The most commonly used protein identification programme is called Mascot (www.matrixscience.com). Mascot uses statistical methods to assess the validity of a match. The strength of a peptide match is based primarily on the concurrence of masses – the precursor mass and MS/MS fragment ion masses that are present in the spectra, coinciding with the predicted masses of peptides and fragment masses calculated on the basis of the sequence from a peptide present in a protein database (Perkins et al., 1999). The strength of the statistical score can be adversely affected by the presence of unassigned peaks, which are mostly likely to be baseline noise, and by the number of peptides in the database which have the same precursor mass within a user defined search tolerance. The majority of search programmes use the precursor mass first to select a subset of fragments from the database that have the correct

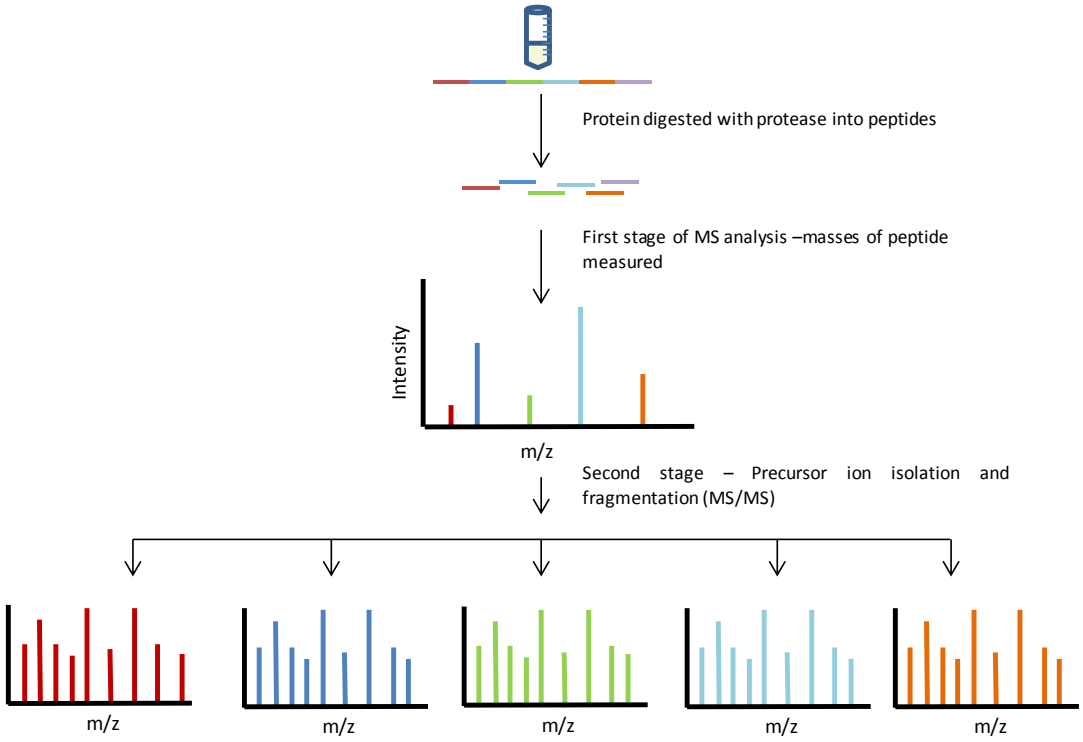


Figure 1.6 LC-MS/MS workflow
Proteins are digested overnight by incubation with enzymes to break up proteins into peptides (represented by the coloured lines). The digested proteins are then analysed by mass spectrometry. During the first stage of MS the masses of the precursor ions are measured**.** During the second stage of MS the precursor ions are isolated and fragmented with an inert gas to produce fragmentation spectra (bottom row of graphs). The distance between each ion in the fragmentation spectra corresponds to the mass of an amino acid.

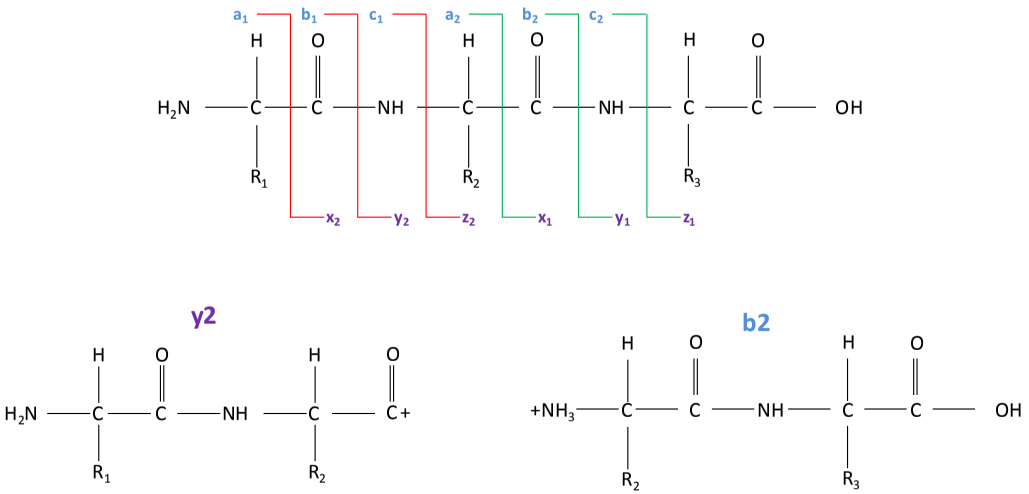


Figure 1.7 Peptide fragmentation chemistry

Top figure. The red and green lines represent areas of the peptide backbone where fragmentation can take place. If collision induced dissociation (CID) is used then fragmentation will result in b and y ions. An alternative method of fragmentation – electron transfer dissociation (ETD) will produce c and z ions.

Bottom figure. An example of a b and y ion produced by CID fragmentation. The charge is retained on the C terminus for the y ion and the N terminus for the b ion.

mass and then the algorithm compares the MS/MS peak masses to a set of predicted fragment masses, generated in silico from the sequence of each peptide in the subset. Other factors that affect scoring include choice of database, whether taxonomy has been selected and user defined mass tolerances.

Chemical communication proteins are often proteins with a high rate of evolution and under selective pressure so obtaining significant matches through a database search is unlikely. There is often incomplete or no genome data when it comes to identifying chemical signalling proteins and identification becomes more complicated. The MS/MS fragmentation spectra will require manual interpretation, referred to as *de novo* sequencing. The mass difference between fragment ions in MS/MS spectra is used to determine the amino acid sequence – each amino acid residue has its own unique mass. To interpret the sequence of tryptic peptide, it is common to start by looking for the y1 ion, which will be 147 Da if the sequence terminates with a lysine or 175 Da if the sequence ends with an arginine residue (Ma and Johnson, 2011). This provides a good starting point for interpretation. The distance is measured between the y1 ion and the next y ion in the series and an amino acid is assigned. The distance between the y2 and y3 ion is then measured and the third amino acid residue is assigned. This continues until the end of the spectrum, signified by the peptide precursor mass. The sequence is then interpreted in the opposite direction - left to right - using the b ion series to provide confidence in the sequence deduced from the y ion series (Ma and Johnson, 2011).

Unlike Mascot which uses the masses (both precursor and fragment) to obtain an identification, the short manual interpreted peptide sequences are then searched against protein databases for similar proteins using a BLAST tool (Basic local alignment search tool). This can provide information on what class the protein might belong too (Altschul *et al.,* 1990). If several peptides prove to be similar to a certain protein or certain family of proteins then this protein will be used as a model to attempt to construct the unknown protein sequence. The unknown protein is also digested with different proteases that have different specificities to

generate overlapping sequence data and therefore adding confidence to the newly generated sequence.

1.6.3 Quantification of proteins in a scent mark

Once a chemical signalling protein has been identified and characterised the next stage would be to quantify them. Proteins involved in scent communication are thought to have their expression up and down regulated depending on a social situation, season, sex and maturation. There are a number of methodologies available to quantify proteins. Non-mass spectrometry based methods include Bradford assay, quantitative western blotting, enzyme-linked immunosorbent assay (ELISA) and enzyme assays. While these techniques are still commonly used, there are occasions were a more complex approach is required for example quantification of the highly polymorphic MUP proteins (see Chapter 3). Recently developed methods have introduced protein quantification by mass spectrometry. The advantages of mass spectrometry include high sensitivity, speed of analysis and the large amount of information generated in one experiment.

There are a variety of mass analysers available each suited to certain types of analysis. All differ in terms of mass accuracy and resolution, sensitivity and selectivity. Mass accuracy indicates the accuracy of the mass to charge ratio (m/z) provided by the mass analyser. It is often expressed in parts per million (ppm) and is defined as the difference between the theoretical m/z and the measured m/z. Mass accuracy is largely linked to resolution of the instrument. Low resolution instruments have poor mass accuracy. Resolution is the ability of a mass analyser to produce two distinct signals for two ions with a small m/z difference. As the precision obtained on the mass of the analysed sample depends on the determination of the centroid of the peak, if the instrument cannot resolve two similar masses then the calculated m/z will be inaccurate leading to a large ppm error.

An Orbitrap is a mass analyser that benefits from both high resolution and mass accuracy. The Orbitrap operates by radially trapping ions about a central spindle

electrode. An outer barrel electrode is coaxial with an inner spindle electrode and m/z values are calculated from the frequency of ion oscillations, along the axis of the electric field, undergone by the orbitally trapped ions. Ion frequencies are determined by acquisition of time-domain image current transients, with subsequent Fourier transforms being used to obtain mass spectra (Hu *et al.,* 2005). A second type of mass analyser that is popular for quantitative studies is the triple quadrupole mass spectrometer. In contrast to the Orbitrap, triple quadrupoles have a much lower resolution but are more sensitive and have a larger linear dynamic range. Two quadrupole mass spectrometers (Q1 and Q3) are positioned in tandem with a non mass resolving, radio frequency only, quadrupole between them to act as a collision cell for fragmentation. Q1 and Q3 act as mass filters and ions are separated based on the stability of their trajectories in the oscillating electric fields that are applied to the quadrupoles. In a typical multiple reaction monitoring experiment (MRM, most commonly used for quantification) Q1 is set to select a certain precursor mass and Q3 is set to select the fragments of this precursor mass. This selectivity reduces the number of background and matrix ions which improves the signal to noise ratio allowing for much lower limits of detection. A time of flight mass analyser (TOF) is another routinely used mass analyser. TOF analysers use an electric field to accelerate ions through the same potential and then measure the time they take to reach the detector (Aebersold and Mann, 2003). If the ions all have the same charge then their kinetic energies will be identical and their velocities will depend only on their m/z with lighter ions reaching the detector first and the heavier ones taking longer. TOF instruments have high resolution and mass accuracy (not as high as Orbitrap), a large dynamic range, good linearity and very fast acquisition times. Sensitivity, particularly on newer generation TOFs is impressive but greater sensitivity is still observed using triple quadrupoles.

There are a variety of methods to quantify proteins using mass spectrometry. These methods can be divided into relative and absolute quantification. Relative quantification uses a comparison between two or more samples to assess changes in the levels of proteins in response to the alteration of the function of a biological system, for example healthy versus disease states. Absolute quantification uses a

labelled internal standard and is useful for the comparison of protein abundances in a single sample.

1.6.3.1 Relative quantification

Label free quantification

Label free quantification involves comparing the abundances of proteins in multiple samples without the use of stable isotopes (Chelius and Bondarenko, 2002). Label free quantification may be based on precursor ion intensity, such as peptide peak areas or peak heights, or on spectral counting. Spectral counting simply counts the total number of fragmentation (MS/MS) spectra produced by peptides belonging to a certain protein (Lundgen *et al.,* 2010). A label free approach requires excellent resolution and mass accuracy for reproducible identification and quantification. The instrument must be able to resolve co-eluting isobaric species and reduce quantification interferences which are especially important for samples of high complexity or high dynamic range. Reproducible chromatography is also imperative for efficient separation from co-eluting species that would lead to inaccurate quantification data. The mass spectrometers of choice include and Orbitrap, Fourier transform ion cyclotron (FTICR) or new generation TOFs. Label free quantification can be applied to a variety of applications including complex biomarker discovery, systems biology studies and isolated proteins and protein complexes. Proteins are extracted from samples and digested with a protease such as trypsin. The peptides are then analysed by LC-MS and identified using accurate mass precursor information and the fragmentation data. Quantification occurs at the MS level by comparing chromatographic peak areas for precursor ions between individual raw data files (Wong and Cagney, 2010).

Stable isotope labelling by amino acids in cell culture (SILAC)

SILAC based quantification is a widely used technique that uses non radioactive labelling to identify differences in protein abundances between samples (Ong *et al.,* 2002). Two populations of cells are cultured. One population contains all essential amino acids. The second population contains all but one essential amino acid for

example arginine. Arginine is replaced by a labelled version – [$^{13}C_6$] arginine. Both cell populations are combined, proteins are extracted and then digested with a protease. The resulting peptides are analysed by mass spectrometry with a +6 Da mass shift observed in peptides containing heavy arginine (Ong and Mann, 2006). The ratio of peak intensities for heavy/light peptide pairs reflects the abundance ratio for the two proteins. Quantification is performed at the MS level by comparing the intensities of the heavy and light precursor ions with protein identification based on accurate mass and fragmentation data (Ong and Mann, 2007). As with the label free approach, an Orbitrap is normally the choice of mass spectrometer for SILAC due to its high resolution and mass accuracy.

Isotope-coded affinity tags (ICATs)

ICAT uses chemical labelling reagents that consist of an affinity tag (biotin), a linker that can incorporate stable isotopes and a reactive group with specificity for thiol groups. The method was originally developed to reduce sample complexity and identify low abundance proteins and peptides in complex samples as only cysteine residues are tagged and labelled peptides are affinity purified (Gygi *et al.,* 1999). The ICAT reagent exists in both heavy (traditionally deuterium 8) and light forms. Two protein mixtures that symbolise two different cell states are treated with ICAT reagent – one with heavy and one with light. They are then combined and digested, normally with trypsin. The digested material is then subjected to an affinity chromatography step to isolate the peptides labelled with ICAT reagent (Gygi *et al.,* 1999). These peptides are then analysed by mass spectrometry and identification of peptides is completed using fragmentation data and quantification is achieved on the MS level by measuring the ratio of the signal intensity between the heavy and light peptide pairs. Again an Orbitrap is a common choice of mass spectrometer.

Isobaric labelling

Isobaric labelling is another mass spectrometry based strategy used in quantitative proteomics. Peptides or proteins are labelled with various chemical groups that are isobaric (the same mass). These isobaric tags contain reporter, balance, and reactive regions. Each sample is digested and labelled individually. All samples are

then mixed in equal ratios and analysed by mass spectrometry. As the tags have the same total molecular weight they are indistinguishable in the first MS scan. It is only during the fragmentation stage that the reporter regions disassociate to produce ion signals which reflect quantitative information about relative amounts of peptide in the sample (Thompson *et al.,* 2003; Ross *et al.,* 2004). Peptide identification and quantification is derived from the MS/MS spectrum. A database search is typically performed on the fragmentation data to identify the labelled peptides (and consequently the proteins) while the reporter ion is used to relatively quantify the peptides. Instrumentation such as Orbitraps and Q-TOFs are the normally the analysers of choice although Orbitraps are favoured as they have the option of higher-energy collisional dissociation (HCD) fragmentation. HCD does not suffer from the low mass cut-off like the standard CID and therefore is useful for isobaric tag based quantification as reporter ions can be observed (Kocher *et al.,* 2010).

1.6.3.2 Absolute quantification

Absolute quantification peptides (AQUA)

An AQUA peptide is a stable isotope labelled, chemically synthesised peptide designed for the absolute quantification of proteins (Kirkpatrick *et al.,* 2005). The development and implementation of an AQUA based strategy usually begins with the amino acid sequence of the protein to be quantified being examined and a tryptic peptide chosen to be chemically synthesised and used for quantification. An AQUA peptide is then produced with the exact amino acid sequence as its counterpart in the native protein except one residue is substituted for a labelled version resulting in a mass difference between the two peptides. This allows the mass spectrometer to distinguish between the native and synthetic peptide (Kirkpatrick *et al.,* 2005). Both the AQUA peptide and the native peptide should share the same physiochemical properties for example they should ionise the same in the mass spectrometer source and should chromatograph from the LC column in the same way. This will result in a more accurate quantification.

A triple quadrupole mass spectrometer is usually the mass analyser of choice for absolute quantification. The synthetic and native peptides are analysed by LC-MS/MS and appropriate fragment ions are chosen to be incorporated into an MRM experiment for quantification (Kirkpatrick *et al.,* 2005). The targeted precursor mass is selected in Q1, fragmented in Q2 and the chosen fragments are selected in Q3. This targeted MS analysis using MRM enhances the lower detection limit for peptides by up to 100 fold (as compared to full scan ms/ms analysis) by allowing rapid and continuous monitoring of the specific ions of interest. To detect an analyte-AQUA peptide pair, two alternating MRM experiments are done during a single LC-MRM analysis. Care must be taken when setting up the MRM method to ensure enough data points (15-20) are collected across the chromatographic peak for reproducible quantification data. The number of data points can be manipulated by the scan time i.e. the time the mass spectrometer spends looking at the mass of interest. An alternative approach to absolute quantification is using an instrument with high accuracy (TOF or Orbitrap), extracting the exact m/z values out of the chromatogram for both the labelled and analyte peptides and comparing the peak intensity or area between the two peptides.

The AQUA peptide is then either added to the protein sample prior to proteolysis or added just before LC-MS analysis. During the LC-MS analysis, both versions of the peptide fragment identically in the collision cell of the mass spectrometer so the mass spectrum will contain the same fragments with some of them shifted in mass due to the isotope label. Protein quantification is determined by the ratio between specifically monitored fragment ions for the AQUA peptide and analyte (Kettenbach *et al.,* 2011).

Protein standards for absolute quantification (PSAQ)

PSAQ standards are full length stable isotope labelled proteins used for absolute quantification (Brun et al., 2007). These standards are produced using cell free systems or a bacterial expression system. A PSAQ standard is produced for each protein to be quantified and is added to the sample mixture at the beginning of the experimental process. Advantages of this method over other quantification

strategies are firstly, the digestion efficiency should be the same for both the analyte and labelled peptide producing more accurate quantification data (Brun et al., 2009). AQUA peptides are produced as ready made labelled tryptic peptides so complete digestion of the analyte is imperative for accurate quantification. A similar situation is observed when using the QconCAT method (see next section). Although the QconCAT is a labelled protein, the peptides are not in the same position as the corresponding peptides in the native protein. Secondly, PSAQ standards are compatible with pre-fractionation which has enabled them to be successfully used to quantify proteins in complex biological matrices were pre fractionation is necessary. These standards also have the added advantage of switching to an alternative reporter peptide if the peptide originally chosen for quantification is no longer suitable for example if interferences such as ionisation competition prevent detection of the peptide in a complex matrix (Brun et al., 2009; Lebert *et al.,* 2011).

Similarly to the AQUA method, quantification is routinely done using a triple quadrupole with an MRM method (Brun et al., 2007; Brun et al., 2009). An MRM method is advantageous for sensitivity, selectivity and specificity. Although selectivity can often be lost in complex matrices, this is improved by using labelled proteins and peptides as they co-elute with the target peptide. The improved sensitivity allows even the low abundant proteins to be quantified in difficult matrices. Alternatively quantification is also possible on the MS level by measuring differences between peak intensity or area between the labelled and native peptides.

Quantification concatemers (QconCATS)

A QconCAT is a stable isotope labelled protein that comprises of peptides from multiple proteins to be quantified (Pratt *et al.,* 2006). The QconCAT approach is relatively low cost compared to PSAQ standards and AQUA peptides as multiple proteins (up to 100) can be quantified using a single QconCAT protein. Peptides are chosen to represent each protein to be quantified. Peptides should be unique to the protein to be quantified, be suitable for LC-MS analysis i.e. ionise and

chromatograph well, have amino acids present that will be required for labelling and residues that could potentially cause problems with quantification such as methionine, which can oxidise and cause a mass shift, should be avoided (Eyers *et al.,* 2008). These rules for choosing peptides are discussed further in chapter 3. Once peptides have been chosen for incorporation into the QconCAT, their sequences are assembled *in silico* and a gene is constructed which encodes the assembled Q peptides using codons that yield maximum expression in *E. coli.* A His-tag motif for purification is also added onto the c-terminus. The gene is then synthesised and subcloned into an expression plasmid. Once an expression plasmid encoding the QconCAT protein has been produced, the plasmid is introduced into an appropriate *E. coli.* expression strain. A single transformant is then grown in medium containing amino acids with certain residues added to the culture in a labelled form, for example $^{13}C_6$ arginine and $^{13}C_6$ lysine for tryptic digests. Expression is induced using Isopropyl β-D-1-thiogalactopyranoside (IPTG) and analysed by SDS-PAGE to monitor expression (Pratt *et al.,* 2006; Rivers *et al.,* 2007). The QconCAT is then purified by affinity chromatography and the end product can be confirmed by either SDS-PAGE followed by an in gel tryptic digest and MALDI – TOF analysis or electrospray – mass spectrometry (ESI-MS) to obtain an accurate mass of the protein. Once the purified product is confirmed as being the QconCAT, it must be quantified to enable quantification of the analyte proteins. QconCAT protein concentration is normally measured using a protein assay (Pratt *et al.,* 2006; Rivers *et al.,* 2007).

A known concentration of QconCAT protein is then added to the mixture of proteins to be quantified. This mixture is then digested using the appropriate protease, which depends on what amino acids have been labelled, and analysed by LC-MS (Rivers *et al.,* 2007). Amounts of each protein are calculated either on the fragment ion level by using a triple quadrupole and MRM method or the MS level by comparing the intensity of the precursor masses using extracted ion chromatograms on a high mass accuracy and resolution instrument such as a TOF or Orbitrap. Again, as seen with AQUA and PSAQ strategies, the MRM approach is more popular due to the ability to detect low abundant proteins and selectivity and

specificity to provide confidence in the quantification by being able to select the exact precursor mass and the corresponding fragment ions.

1.7 Aims and objectives

The overall aim of this thesis was to investigate the protein components present in scent marks using advanced proteomic methodologies. The first aspect of this thesis aimed to develop a method for the quantification of MUPS in a laboratory strain of mouse C57BL/6. Mouse MUPS have been widely studied on a behavioural level and a considerable amount of genome data for this strain of mouse has been collected. Electrospray (ESI) mass spectrometry has identified 5 major isoforms in males and 3 isoforms in female mice in this strain of mouse. Developing a quantification method would either confirm the presence or absence of other minor MUP isoforms whose functional genes had been identified from genomic data. A QconCAT strategy was implemented to quantify all MUPs expressed by male and female B6 mice and differences between sexes were examined. The QconCAT quantification method was then used to assess MUP production in female mice during the estrous cycle.

The second part of this thesis examined protein secretion in the harvest mouse (*MIcromys minutus*). There is no genomic data available for this species so it is unknown if they have any genes related MUPs or other lipocalins observed in other rodent species. Lipocalins have been observed in the Syrian hamster (Sibger *et al.,* 1986), bank vole (Stopkova *et al.,* 2010), Roborovski hamster (Turton *et al.,* 2010) in addition to the well established MUPS in mice and rats. Proteins from the harvest mice were characterised using mass spectrometric techniques and potential behavioural aspects were also investigated.

The final section of this thesis investigated urinary proteins secreted by the mouse lemur (*Microcebus*). Mouse lemurs are the world's smallest primate and have been observed responding to urinary cues from conspecifics. They have two functional MUP genes and also functional VR2 receptors, the first in primates, have recently been discovered (Holenbrink *et al.,* 2012). Urinary proteins excreted by two species

of male mouse lemur - *Microcebus murinus* and *Microcebus lehilahytsara* were identified and fully characterised using advanced mass spectrometric techniques.

Chapter 2: Materials and methods

2.1 Sample collection

C57BL/6 laboratory mice

Urine was collected from male and female C57BL/6 laboratory mice by gentle bladder massage. The urine samples were collected between 9am-10am by technical staff at the University of Liverpool, Leahurst campus. A single sample of urine was supplied from each mouse. The mice, illustrated in figure 2.1, were housed in a temperature (20 °C) and humidity controlled environment with a 12 hour light cycle (12 hours light/12 hours' darkness). Males were housed individually; females were housed in groups of 2-3 per cage. Inbred stocks were supplied by Harlan laboratories (Bicester, Oxon, UK).

Harvest mice

Harvest mouse urine was collected using a recovery method. The rodents, illustrated in figure 2.1, were individually placed on a mesh wire grid over a glass dish with another over the top to confine the animal - they were then left for approximately 1 hour with regular checks for urine. The urine samples were collected between 9am-11am by technical staff at the University of Liverpool, Leahurst campus. A single sample of urine was supplied from each rodent. The animals were housed in a temperature (20 °C) and humidity controlled environment with a 12 hour light cycle (12 hours light/12 hours' darkness). Harvest mice were bred in an outdoor enclosure based at the University of Liverpool, Leahurst campus, UK.

The saliva from the harvest mice was collected by swabbing the inside of the cheek with a Pasteur pipette and then transferred to an eppendorf tube (0.5 ml). The body and paw washes were collected by swabbing the animal with cotton buds soaked in water (50 µl). The buds were then removed and placed in an eppendorf tube (1.5 ml) before centrifugation for five minutes at 2000 rpm.

The glass rod washes were collected in a similar manner. Rods were placed in harvest mice cages and left for two weeks. After two weeks the rods were removed and washed with cotton buds soaked in water (150 µl) prior to centrifugation at 2000 rpm for five minutes. Samples were collected by technical staff at the University of Liverpool, Leahurst campus. A single sample of each wash and saliva was supplied from each rodent.

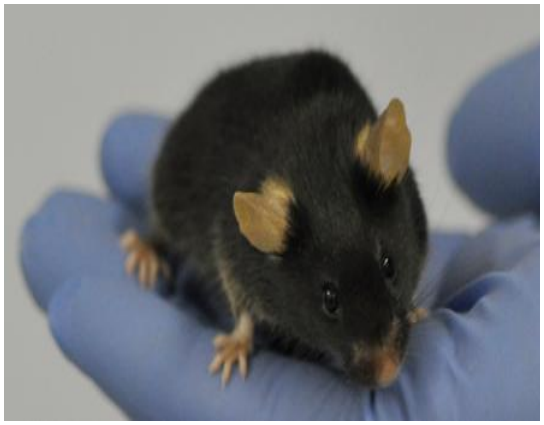
Mouse lemurs

Mouse lemur urine was collected using gentle bladder massage. The urine samples were collected early morning by technical staff at the University of Hanover, Germany. A single sample of urine was supplied from each mouse lemur for each season (breeding and non-breeding). The mouse lemurs, illustrated in figure 2.1, were housed in a temperature (23 °C) and humidity controlled environment. The animals were kept under a fluctuating, reversed light cycle with a 14-h light period and a 10-h dark period (reproduction period) or vice versa (10 h light, 14 h dark; resting period). The mouse lemurs are bred and housed in captivity in a breeding colony at the Institute of Zoology, University of Veterinary Medicine Hannover. The mouse lemurs are kept in groups of three-four animals of the same sex.

Gray mouse lemur
(*M. murinus*)

Goodmans mouse lemur
(*M. lehilahytsara)*

Harvest mouse
(M. minutus)



C57BL/6 laboratory mouse

Figure 2.1 Animals systems used to obtain samples for analysis.
Top pictures: The two species of mouse lemur used to collect urine samples from during the breeding and non breeding seasons. Photographs obtained from the Institute of Zoology, University of Veterinary Medicine Hannover.
Bottom pictures: A harvest mouse in an outdoor enclosure and a C57BL/6 laboratory mouse. Photographs supplied by technical staff at the University of Liverpool, Leahurst campus, UK.

2.2 Protein assay

Total protein concentration was measured using a Coomassie Plus protein assay kit (Pierce, Rockford, USA). A stock solution of Bovine Serum Albumin (BSA, 1 mg/ml) was prepared and diluted down appropriately to produce a standard curve (0-50 µg/ml). Samples were diluted down with purified water to make sure they were within the linear range of the assay. Absorbance readings were measured at 620 nm using a plate reader (Thermo Scientific™ Multiskan™).

2.3 Creatinine assay

Creatinine levels were measured using a creatinine assay kit (Sigma, UK). The creatinine standard curve ranged from 0-30 µg/ml. Samples were diluted down with purified water to make sure they were within the linear range of the assay. Absorbance readings were measured at 570 nm using a plate reader (Thermo Scientific™ Multiskan™).

2.4 SDS-PAGE

SDS-PAGE was performed as described by Laemmli (1970). Samples were mixed 1:1 with reducing sample buffer and heated at 95 °C. Samples were loaded onto a 15 % SDS-PAGE gel and run at a constant voltage of 200 V until the dye front reached the bottom of the gel. Protein bands were visualised with Coomassie Brilliant blue stain (Sigma) overnight and destained the following morning with a mixture of purified water (80%), acetic acid (10%) and methanol (10%).

2.5 Protein digestion

In-gel digestion

Pieces of gel were removed from the protein bands identified by SDS-PAGE. The gel pieces were destained (50:50 ACN:25 mM NH_4HCO_3) for 15 minutes at 37°C. This process was repeated until the gel pieces were fully destained. The gel plugs were then reduced in dithiothreitol (DTT, 10 mM) to reduce the disulfide bonds between the cysteine residues inside the protein. This reduction process was carried out at 60 °C for 1 hour. The DTT was discarded and Iodoacetamide (25 µl, 55 mM) was

added to the gel pieces to prevent the re-formation of the disulfide bonds between the cysteine residues by covalently binding to the thiol group of the cysteine residue. This alkylation step was carried out in the dark at room temperature for 45 minutes. The samples were then dehydrated in acetonitrile (ACN) for 15 minutes at 37 °C. Protease – trypsin, endoproteinase LysC or endoproteinase GluC (10 µl, 10 ng/ml) was added to each of the gel pieces at a 10:1 substrate:enzyme ratio and the samples were incubated for 16 hours. The digestion reaction was stopped with the addition of formic acid (1% v/v).

In-solution digestion of MUPS

Urine was diluted in 25 mM NH_4HCO_3 to produce a final concentration of 10 µg/µl of protein. This solution was incubated with *Rapi*Gest™ SF Surfactant (0.1% w/v final concentration, Waters) at 80 °C for 10 minutes. The samples were then reduced with DTT (3 mM final concentration) at 60 °C for 10 minutes followed by alkylation with iodoacetamide (9 mM final concentration) in the dark at room temperature. The protease, either trypsin (0.2 µg/µl diluted in 25 mM NH_4HCO_3), endoproteinase LysC (0.1 µg/µl diluted in 25mM Tris HCl pH 8.5) or endoproteinase GluC (0.2 µg/µl diluted in ddH_2O) was added to the digests at an substrate:enzyme ratio of 50:1 and left to incubate for 16 hours. Following incubation, a small proportion of the digested material was removed to run on an SDS-PAGE gel to check for complete digestion. The rest of the digest was treated with TFA (to a final concentration of 0.5% v/v) and incubated at 37 °C for 45 minutes to remove the *Rapi*Gest™ SF Surfactant prior to LC-MS analysis. The samples were then centrifuged at 10,000 rpm for 15 minutes and the supernatant transferred to a fresh 0.5 ml Eppendorf.

In-solution digestion of harvest mouse and mouse lemur samples

Samples were diluted in 25 mM NH_4HCO_3 to produce a final concentration of 10 µg/µl of protein. This solution was incubated with *Rapi*Gest™ SF Surfactant (0.05% w/v final concentration, Waters) at 80 °C for 10 minutes. The samples were then reduced with DTT (3 mM final concentration) at 60 °C for 10 minutes followed by alkylation with iodoacetamide (9 mM final concentration) in the dark at room

temperature. The protease, either trypsin (0.2 µg/µl diluted in 25 mM NH_4HCO_3) or endoproteinase LysC (0.1 µg/µl diluted in 25mM Tris HCl pH 8.5), was added to the digests at an substrate:enzyme ratio of 50:1 and left to incubate for 16 hours. Following incubation, a small proportion of the digested material was removed to run on an SDS-PAGE gel to check for complete digestion. The rest of the digest was treated with TFA (to a final concentration of 0.5% v/v) and incubated at 37 °C for 45 minutes to remove the *Rapi*Gest™ SF Surfactant prior to LC-MS analysis. The samples were then centrifuged at 10,000 rpm for 15 minutes and the supernatant transferred to a fresh 0.5 ml Eppendorf.

In solution digestion of glass rod anion exchange fractions

Strataclean beads (20 µl, Agilent technologies, UK) were added to fractions produced from anion exchange chromatography. The samples were vortexed for two minutes before centrifugation at 5000 rpm for two minutes. The supernatant was discarded and the beads were washed by vortexing with purified water (500 µl). The samples were again centrifuged at 5000 rpm and the supernatant discarded. This process was repeated once more. The beads were then suspended in 25 mM NH_4HCO_3 (50 µl) and incubated with *Rapi*Gest™ SF Surfactant (0.05% w/v final concentration, Waters) at 80 °C for 10 minutes. The samples were then reduced with DTT (3 mM final concentration) at 60 °C for 10 minutes followed by alkylation with iodoacetamide (9 mM final concentration) in the dark at room temperature. The protease, either trypsin (5 µl, 0.2 µg/µl diluted in 25 mM NH_4HCO_3), endoproteinase LysC (5 µl, 0.1 µg/µl diluted in 25mM Tris HCl pH 8.5) or endoproteinase GluC (0.2 µg/µl diluted in ddH_2O), was added to the digests and left to incubate for 16 hours. All stages of the digestion process were carried out using a shaking mixer (1000 rpm) to keep the beads suspended and ensure efficient digestion. Following incubation the supernatant was removed and a small proportion was removed to run on an SDS-PAGE gel to check for complete digestion. The rest of the digested material was treated with TFA (to a final concentration of 0.5% v/v) and incubated at 37 °C for 45 minutes to remove the *Rapi*Gest™ SF Surfactant prior to LC-MS analysis. The samples were then

centrifuged at 10,000 rpm for 15 minutes and the supernatant transferred to a fresh 0.5 ml Eppendorf.

2.6 Peptide mass fingerprinting (PMF)

The peptide mixtures from the in-gel digestion were analysed by MALDI-TOF-MS (matrix-assisted laser-desorption ionization–time of flight), (Bruker ultrafleXtreme™). The mass spectrometer was operated in reflectron mode with positive ion detection. Samples were mixed with MALDI matrix (α-cyano-4-hydroxycinnamic acid in 50% ACN / 0.2%TFA) in a 1:5 ratio and spotted onto a target plate before being left to air dry. The laser frequency was set to 1000Hz with the laser energy optimised to 27% of the maximum with 500 shots per spectrum. Spectra were gathered between 800-4000 m/z. The mass spectrometer was externally calibrated with a mixture of Des-Arg bradykinin (904.47 Da), angiotensin I (1296.69 Da), neurotensin (1672.92 Da), ACTH 1-17 fragment (2093.09 Da) ACTH (corticotrophin, 2465.2 Da) and ACTH 7- 38 fragment (3657.93 Da). The concentration of each was 6 pmol/µl, apart from ACTH 7- 38 which was 9 pmol/µl. All standards were sourced from Sigma.

2.7 Electrospray – mass spectrometry (ESI-MS) of intact protein

Samples were diluted in formic acid (0.1% in purified water) to produce a protein concentration of approximately 5 pmol/µl. The samples were injected onto a C4 desalting trap (Waters MassPREP ™ Micro desalting column, 2.1 x 5 mm, 20 µm particle size, 1000 Å pore size) (Waters, Manchester, UK) that was fitted onto a nano ACQUITY Ultra Performance liquid chromatography® (UPLC®) system. The chromatography system was coupled to a SYNAPT™ G1 QTof mass spectrometer fitted with an electrospray source (Waters, Manchester, UK). Protein was eluted using a mixture of solvents A and B. Solvent A was HPLC grade water with 0.1% (v/v) formic acid, and solvent B was HPLC grade acetonitrile with 0.1% v/v) formic acid. Separations were performed by applying a linear gradient of 5% to 95% solvent B over 10 min at a flow rate of 40 µl/min followed by an equilibration step (5 min at 5 % solvent B).

Data was collected between 500 – 3500 m/z. The mass spectrometer was externally calibrated with horse heart myoglobin (1 pmol/µl, Sigma). Data was processed using maximum entropy software (MAX ENT 1, Mass Lynx version 4.1, Waters). Data sets were processed at 0.5 Da/channel over a mass range of 18200 – 19200 Da (for MUPS), 16000-18000 Da (harvest mouse) and 8500-10000 Da (Mouse lemur).

2.8 Tandem mass spectrometry

MUP quantification data

LC-MS analysis was carried out using a nano UPLC® system coupled to a SYNAPT™ G2 QTof mass spectrometer fitted with a nanospray source (Waters, Manchester, UK). Peptides (500 fmol) were loaded onto a C_{18} trapping column (180 µm × 20 mm) (Waters, Manchester, UK) at 5 µl/min in 99% formic acid diluted in purified water (0.1%) and 1% formic acid diluted in ACN (0.1%) for five minutes. Peptides were then separated using an ACQUITY UPLC® BEH column C18 analytical column (75µm x 150mm, 1.7µm) over a one hour gradient using a mixture of solvents A and B. Solvent A was HPLC grade water with 0.1% (v/v) formic acid, and solvent B was HPLC grade acetonitrile with 0.1% (v/v) formic acid. Separations were performed by applying a linear gradient of 3% to 85% solvent B over 35 min at 300 nL/min followed by an equilibration step (15 min at 3 % solvent B).

The mass spectrometer was operated positive ion mode using an MSe method. Data was acquired between 300-3000 m/z. The mass spectrometer detectors were calibrated with Leucine-enkephelin (50 pmol/µl) (Waters, Manchester, UK). Glu-fibrinopeptide (5 pmol/µl) (Waters, Manchester, UK) was used for the mass calibration. The mass spectrometer conditions were as follows: capillary voltage, 3 kV; cone voltage, 45 V; source temperature, 80 °C; desolvation temperature, 150 °C; cone gas flow, 50 L/hr; desolvation gas flow, 500 L/hr.

Protein quantification was achieved on the MS level using extracted ion chromatograms. An m/z for each analyte was recovered/extracted from the entire

dataset for the chromatographic run. The mass tolerance for the extraction, which varies depending on which mass analyser is used, was set to 0.2 Da for extracting m/z values for peptides to be used in MUP quantification. Quantification was performed by comparing the extracted ion chromatogram peak intensity of the endogenous and the labelled forms of the proteotypic peptide.

Harvest mouse and mouse lemur de novo sequence data

Samples were analysed using a Ultimate 3000 nano system (Dionex/Thermo Fisher Scientific, Hemel Hempstead, UK) coupled to a QExactive mass spectrometer (Thermo Fisher Scientific, Hemel Hempstead, UK). Peptides (500 fmol) were loaded onto a trap column (Acclaim PepMap 100, 2cm x 75 µm inner diameter, C_{18}, 3 µm, 100Å) at 5 µl/min with an aqueous solution containing 0.1% (v/v) TFA and 2% (v/v) acetonitrile. After 3 min, the trap column was set in-line with an analytical column (Easy-Spray PepMap® RSLC 15cm x 75 µm inner diameter, C_{18}, 2 µm, 100Å) (Dionex). Peptides were eluted by using an appropriate mixture of solvents A and B. Solvent A was HPLC grade water with 0.1% (v/v) formic acid, and solvent B was HPLC grade acetonitrile 80% (v/v) with 0.1% (v/v) formic acid. Separations were performed by applying a linear gradient of 3.8% to 50% solvent B over 35 min at 300nL/min followed by a washing step (5 min at 99% solvent B) and an equilibration step (15 min at 3.8% solvent B).

The mass spectrometer was operated in data dependent positive (ESI+) mode to automatically switch between full scan MS and MS/MS acquisition. Survey full scan MS spectra (300-2000 m/z) were acquired in the Orbitrap with 70,000 resolution (200 m/z) after accumulation of ions to $1x10^6$ target value based on predictive automatic gain control (AGC) values from the previous full scan. Dynamic exclusion was set to 20s. The 10 most intense multiply charged ions ($z \geq 2$) were sequentially isolated and fragmented in the octopole collision cell by higher energy collisional dissociation (HCD) with a fixed injection time of 120ms and 35,000 resolution.

The mass spectrometer was calibrated using a ready to use positive ion calibration solution from the instrument manufacturer (Thermo Fisher Scientific, Hemel Hempstead, UK). The solution contains a mixture of caffeine, MRFA, Ultramark

1621, and n-butylamine in an acetonitrile:methanol:water solution containing acetic acid (1% v/v). The mass spectrometer conditions were as follows: spray voltage, 1.9kV, no sheath or auxillary gas flow; heated capillary temperature, 250°C; normalised HCD collision energy 30%. The MS/MS ion selection threshold was set to 1 x 10^4 counts and a 2 *m/z* isolation width was set.

2.9 *De novo* sequencing analysis

De novo sequencing analysis of the harvest mouse and mouse lemur proteins was assisted by PEAKS®6 software for proteomics (Bioinformatics Solutions Inc, Canada). Precursor and fragment ion error tolerances were set to 10 ppm and 0.01Da respectively. Post translational modifications, carbamidomethylation (fixed modification) and oxidation of methionine (variable modification) residues were also included in the processing set-up. Fragmentation type was set to higher-energy C-trap dissociation (HCD). The average local confidence score – a score assigned by PEAKS which reflects the likelihood of a peptide sequence being correct – was set to a 55% cut off as recommended by PEAKS.

2.10 Database searching

Raw data was imported into Peaks 6 software for proteomics (Bioinformatics Solutions Inc, Canada) and searched against a custom made lipocalin database. The parameters were set to accept 1 missed cleavage, a fixed modification of carbamidomethyl cysteine and a variable medication to include methionine oxidation. Precursor and fragment ion error tolerances were set to 10 ppm and 0.01Da respectively. Fragmentation was set to HCD.

2.11 Anion exchange chromatography

Anion exchange chromatography was performed using an Äkta chromatography system (GE Healthcare, Bucks., UK). Samples (approximately 1 mg of protein) were manually injected (100 µl) onto a UNO Q (1 ml) anion exchange column that had been previously equilibrated in 10 mM Hepes pH 8.0. Bound protein was eluted

using a linear salt gradient of 0-0.5 M NaCl. The eluent from the column was monitored at 214 nm. Fractions were collected (1ml) over a 60 minute gradient. Mass and purity of detected proteins was confirmed by ESI-MS.

2.12 QconCAT design

The MUP QconCAT was designed by Dr. S Armstrong and Dr. D Simpson, University of Liverpool, Protein Function Group. Theoretical digests (using endoproteinase LysC) of MUP sequences taken from genomic sequencing data by Mudge *et al.,* 2008 were used to select MUP peptides for inclusion in the QconCAT. Peptides were chosen based on uniqueness in the first instance. For those MUPs with no unique peptides, a subtraction method using shared MUP peptides was deployed (see Chapter 3). Once the Q peptides were chosen, the QconCAT gene was constructed and synthesised by PolyQuant GmbH, Germany using the method described in Pratt *et al.,* 2006.

2.13 Bacterial Transformation

The transformation process was carried out Mrs L McLean, University of Liverpool, Protein Function Group. Transformation is the transmissible modification of the properties of a competent bacterium by DNA from another bacterial strain. The MUP QconCAT gene was cloned into apET21a plasmid vector and transformed into *E. coli* BL21 (DE3) cells. A tube of BL21(DE3) competent *E. Coli* cells were thawed on ice for 10 minutes. The cells were gently mixed and 50 µl transferred to a separate tube and kept on ice. Plasmid DNA (5 µl) was added to the cell mixture and the contents mixed gently. The mixture was placed on ice for 10 minutes before being heat shocked in a water bath set at 42 °C for 10 seconds. The sample was then placed on ice for a further 5 minutes. A super optimal broth (SOC) solution (950 µl) supplied by Promega, UK, was added to the sample. This SOC solution contained 2% w/v tryptone, 0.5% w/v Yeast extract, NaCl (10mM), KCl (2.5mM), $MgCl_2$ (anhydrous 10mM) and deionised water. The mixture was then left to incubate (37 °C) on a mixer at 250rpm for 60 minutes. Cells were then mixed by inversion and a 10-fold dilution was performed in SOC. LB agar plates were heated at 37 °C for 10 minutes prior to the diluted transformation mixture (50 µl) being

added to the plates. The plates were incubated at 37 °C overnight. The following morning a glycerol stock of the plasmid was produced to allow for long term storage. A single colony from an LB plate was added to a culture of LB medium (5 ml) containing ampicillin (100 µg/ml). The culture was placed in an incubator at 37 °C for 5-6 hours with vigorous shaking at 300 rpm. The bacterial cells were then harvested by centrifugation 4000 rpm for 15 minutes at 4 °C. The centrifuged bacteria was then added to a sterilized 60% glycerol solution (1:1 by volume e.g. 1 ml of glycerol solution to 1 ml of bacteria). The glycerol bacterial stock was aliquoted in 100 µl aliquots and stored at -80 °C prior to protein expression.

2.14 Expression and purification of the MUP QconCAT

The glycerol stock was defrosted and streaked, using a loop and sterile technique, onto LB agar plates containing ampiciliin (50 mg/ml). The plates were incubated overnight at 37 °C. The following morning an individual colony was incubated at 37 °C in LB broth (10 ml) and ampicillin (10 µl, 50 mg/ml). After 6 hours, 100 µl of the LB culture was added to minimal medium containing disodium phosphate (0.24 M), potassium phosphate (0.11 M), sodium chloride (11 mM), ammonium chloride (93 mM), magnesium sulphate (1 M), calcium chloride (0.1 M), glucose (20%, 1 g in 5 ml), thiamine (0.5 % w/v) and deionised water. The culture was then incubated overnight at 37 °C with continuous vigorous shaking at 300 rpm. The following morning 6 mls of the overnight culture was added to 200 ml of minimal medium containing the solutions described above plus a full set of unlabelled amino acids (10 mg/ml of hydrophilic amino acids and 20 mg/ml of hydrophobic amino acids) and [$^{13}C_6$] lysine and/or [$^{13}C_6$] arginine (100 mg/L) as the only source of these amino acids. The culture was incubated at 37 °C with continuous vigorous shaking at 300 rpm. The OD (600nm) was taken every hour until it reached an absorbance reading of 0.6. QconCAT protein expression was then induced with isopropyl-D-thiogalactopyranoside (IPTG) and the cells were harvested by centrifugation at 3500 rpm at 4 °C for 15 min. Inclusion bodies containing the QconCAT protein were recovered by breaking cells using BugBuster Protein Extraction Reagent (Novagen, Nottingham, UK). Inclusion bodies were resuspended in 80 mM phosphate buffer, 6 M guanidinium chloride, 2 M NaCl, 40 mM imidazole, pH 7.4. From this solution,

QconCAT proteins were purified by affinity chromatography using a nickel-based resin (HisTrap kit, GE Healthcare, Bucks., UK). Following sample loading, HisTrap columns were washed with 80 mM phosphate buffer, pH 7.4, prior to elution of the sample with the same buffer containing a higher concentration of imidazole (80 mM phosphate, 2 M NaCl, 2 M imidazole, 6 M guanidinium chloride, pH 7.4) during which phase fractions (1 ml) were collected. The purified QconCAT was desalted by three rounds of dialysis against 100 volumes of 100 mM ammonium bicarbonate, pH 8.5, for 3 h using fresh buffer each time.

2.15 Expression and purification of labelled darcin

The darcin glycerol stock was defrosted and streaked, using a loop and sterile technique, onto LB agar plates containing ampiciliin (50 mg/ml). The plates were incubated overnight at 37 °C. The following morning an individual colony was incubated at 37 °C in LB broth (10 ml) and ampicillin (10 µl, 50 mg/ml). After 6 hours, 100 µl of the LB culture was added to minimal medium containing disodium phosphate (0.24 M), potassium phosphate (0.11 M), sodium chloride (11 mM), ammonium chloride (93 mM), magnesium sulphate (1 M), calcium chloride (0.1 M), glucose (20%, 1 g in 5 ml), thiamine (0.5 % w/v) and deionised water. The culture was then incubated overnight at 37 °C with continuous vigorous shaking at 300 rpm. The following morning 6 mls of the overnight culture was added to 200 ml of minimal medium containing the solutions described above plus a full set of unlabelled amino acids (10 mg/ml of hydrophilic amino acids and 20 mg/ml of hydrophobic amino acids) and [$^{13}C_6$] lysine and [$^{13}C_6$] arginine (100 mg/L) as the only source of these amino acids. The OD (600 nm) was taken every hour until it reached an absorbance reading of 0.6. Darcin protein expression was then induced with isopropyl-D-thiogalactopyranoside (IPTG) and the cells were harvested by centrifugation at 3500rpm at 4 °C for 15 min. Labelled darcin protein was then purified by affinity chromatography using a nickel-based resin (HisTrap kit, GE Healthcare, Bucks., UK). Following sample loading, HisTrap columns were washed with 80 mM phosphate buffer, pH 7.4, prior to elution of the sample with the same buffer containing a higher concentration of imidazole (80 mM phosphate, 2 M NaCl,

2 M imidazole, pH 7.4) during which phase fractions (1 ml) were collected. The purified darcin was desalted by three rounds of dialysis against 100 volumes of 100 mM ammonium bicarbonate, pH 8.5, for 3 h using fresh buffer each time.

Chapter 3: Quantification of mouse major urinary proteins

3.1 Introduction

Mice use olfactory chemosignals present in their urine as their main source of communication. These signals can provoke a variety of behavioural and physiological responses including the onset of puberty (Drickamer 1986, Caretta-Mucignat *et al.,* 1995), mate choice (Hurst 1990, Thom *et al.,* 2008) and aggression between males (Novotny *et al.,* 1985, Caretta-Mucignat *et al.,* 2004). Mouse MUP proteins contain the primary source of information for conspecifics and have been the subject of in-depth behavioural experiments (Cheetham *et al.,* 2007; Ramm *et al.,* 2008; Roberts *et al.,* 2012). Following the identification of MUPs and their roles in chemical signalling, the next logical step was to monitor the regulation in expression of these proteins through the development of a quantification method. Development of a quantification method will mean not only can changes in overall MUP expression be observed, but increases and decreases in individual MUP proteins in selected social situations will be also possible allowing a greater insight into intra-species communication.

3.2 Aims and objectives

This chapter will focus on the development of a method to absolutely quantify MUPS. The objectives of the study were:

- To design a QconCAT for the quantification of MUPS expressed in an inbred laboratory strain of mouse C57BL/6 (B6)
- To devise a method for complete proteolysis of the native protein,
- To quantify MUPS in both male and female B6 mice and compare expression between sexes.
- Use the QconCAT to examine female MUP production during the estrous cycle.

3.3 Results and discussion

3.3.1 Design of a QconCAT for the quantification of MUPS

A QconCAT strategy was designed and implemented for the quantification of MUP isoforms in B6 laboratory mice. The B6 strain was an ideal choice as the MUP locus in these mice has been subjected to in depth gene analysis (Mudge *et al.,* 2008; Logan *et al.,* 2008). Although the majority of the MUP cluster has been sequenced, there are still gaps suggesting there may be further uncharacterised MUP variants expressed in urine. The QconCAT was designed based on the Mudge paper as this paper was released earlier than the Logan paper. The B6 genome was sequenced using bacterial artificial chromosomes (BACS), engineered DNA molecules used to clone DNA sequences in bacterial cells. Segments of an organisms DNA is inserted into BACs. The BACs, with the inserted DNA, are then taken up by bacterial cells which grow and divide, amplifying the BAC DNA which can be then isolated and used in sequencing DNA. The sequenced parts are then rearranged *in silico* resulting in the genomic sequence of the organism. Genomic sequencing by Mudge *et al.,* 2008 identified 19 predicted genes and 18 loci that are pseudogenes. There were three gap regions identified within the tiling path indicating that the full complement of MUP loci is not yet represented. Liver transcription for 14 of the genes was confirmed, peptides for these proteins were included in the QconCAT design. A further two peptides were chosen for incorporation into the QconCAT. These peptides were from two proteins from a second strain of mouse BALB/C which would allow potential quantification of urinary proteins in this strain of mouse at a later date. This chapter will just focus on the B6 strain of mouse.

ESI –MS has been previously been employed to map MUP variants by virtue of their molecular mass and using this information, compare the MUP urinary phenotype between inbred strains, wild populations, gender, and individuals (Evershed *et al.,* 1993; Robertson *et al.,* 1996; Beynon *et al.,* 2002; Robertson *et al.,* 2007, Dr S Armstrong, thesis). However, only the mass of the MUP isoform can be reported with any real confidence, and even then minor MUP masses may be obscured by more dominant MUP species. In the B6 lab strain, five major MUP isoforms were

identified in males and three major isoforms in females using ESI-MS analysis of intact urine. B6 male urine was separated by anion exchange and the separate fractions collected and subjected to ESI-MS. As the dominant isoforms were in separate fractions, further minor MUP variants could be detected including MUPs 3, 13, 17 and 14 (Dr S Armstrong, thesis). By designing a QconCAT for the quantification of MUPs, it was anticipated that quantification data for every MUP could be collected. Quantifying on the peptide level eliminates some of the issues observed with the ESI-MS data such as the major isoforms dominating the signal and resolving the MUP isoforms which have very similar masses.

The design of a QconCAT for MUP quantification was constrained by the high sequence homology between individual MUPS (Figure 3.1). Ideally a unique peptide would be chosen to represent each protein to be quantified in the QconCAT. Finding unique tryptic fragments for all MUPS was difficult. Choosing an alternative protease such as endoproteinase LysC (LysC) that created larger peptides upon proteolysis increased the number of unique fragments available for quantification (Figure 3.2).

When selecting peptides to be used for quantification, a number of factors should be considered. Firstly is the peptide in an area of the protein where complete proteolytic digestion will be consistently achieved? For example if trypsin is the choice of protease, cleavage sites near negatively charged amino acids will prove more difficult to cleave. Trypsin contains an aspartate residue inside its binding pocket that attracts basic residues such as arginine (Arg) and lysine (Lys) to form salt bridges with the aspartate, an essential part of the binding process (Hedstrom, 2002). If there are negatively charged residues near the cleavage site in the protein, these can form salt bridges with nearby basic residues disrupting the recognition process leading to missed or partial cleavages (Siepen *et al.,* 2007). Therefore optimisation of the digestion method is imperative. Other factors to consider when deciding on peptides for quantification include peptide suitability for the type of analysis to be used e.g. LC-MS (Eyers *et al.,* 2011) and potential for a post-translational modification to occur such as deamidation or oxidation of methionine. A post translational modification will alter the mass of the peptide which would

cause problems for quantification. If using an MRM method with a precise mass set in the MS method, the peptide would go undetected because it would have a different mass. If quantifying on the MS level using extracted ion chromatograms the mass shift will again result in a reduced or no signal in the extracted ion chromatogram window. The degree of modification may also be different in the analyte compared to the Q peptide. For example the storage conditions of the Q peptides may promote a certain PTM like deamidation which would lead to inaccurate quantification data.

For MUPS many of these "rules" for choosing peptides cannot be applied because the number of unique peptides is very restrictive even when using a protease such as LysC. Using the predicted amino acid sequences of mature MUPs (sourced from Mudge *et al.,* 2008), LysC peptides were chosen for inclusion in the QconCAT. The choice of peptides to be used was limited as there were few unique peptides to choose from. In some MUPS there were no unique peptides at all making the strategy for quantification quite complicated.

Sixteen peptides were chosen to be incorporated into the MUP QconCAT (Figure 3.3). Calculating amounts of those MUPS with no unique peptide involved using a subtraction method. For example MUP 17 has no unique peptide so the amount will be calculated using peptide 4 in figure 3.3 and subtracting off MUP 13 which shares peptide 4 but has also has a unique peptide 6 and can therefore be quantified. This in turn can then be used to calculate MUP 7 by subtracting amounts of MUP 13 and 17 away from peptide 7 which they both share with MUP 7. MUPs 1 and 12 can then be quantified by subtracting the amount calculated for MUP 7 away from peptide 12 which they both share. Finally MUPs 9, 11, 16, 18 and 19 can be calculated by subtracting MUP 2 which has a unique peptide and MUP 1 and 12 away from peptide 13 which they all share. Alternatively, MUPs 9, 11, 16, 18 and 19 could be calculated using either peptide 1 or 5 and subtract away all MUPS that share those peptides. The strategy for quantification is outlined in Figure 3.4.

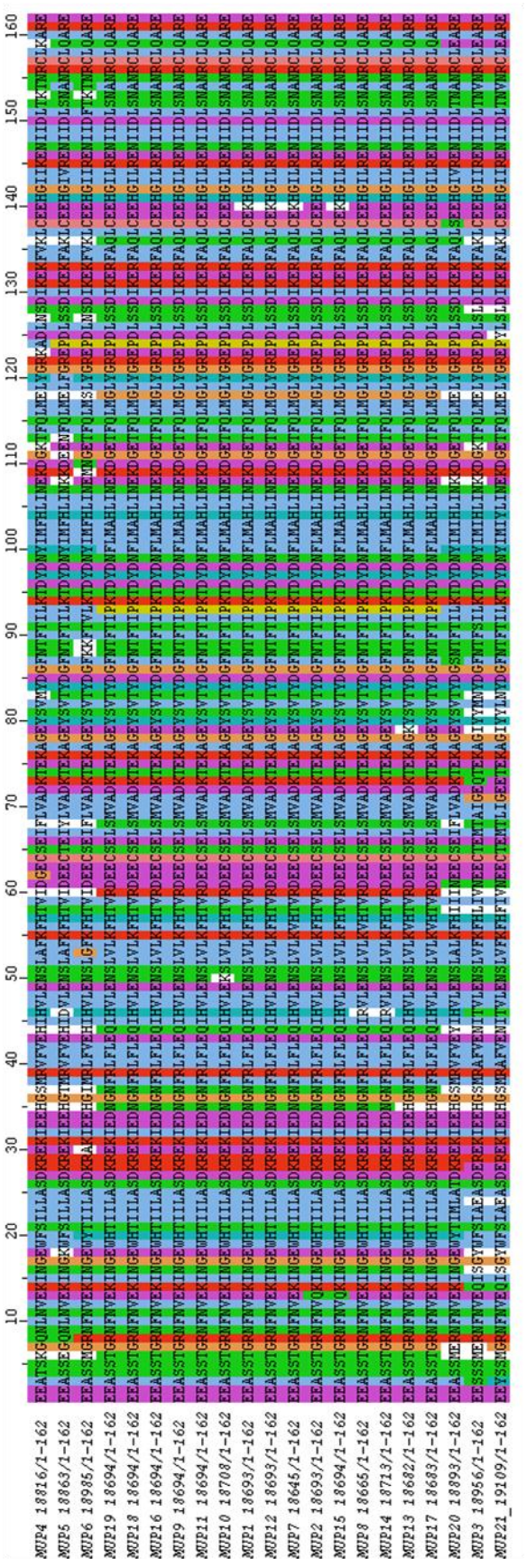


Fig.3.1 Alignment of MUP amino acid sequences based on the C57BL/6 genome.
Sequences were aligned using the ClustalW2 alignment tool for multiple sequences. Jalview bioinformatics software was used to colour code the sequences to reflect the high sequence homology between MUP variants. Each sequence was adjusted for the loss of the 19aa signal peptide. Each mass was adjusted for the formation of a disulphide bond (-2 Da).

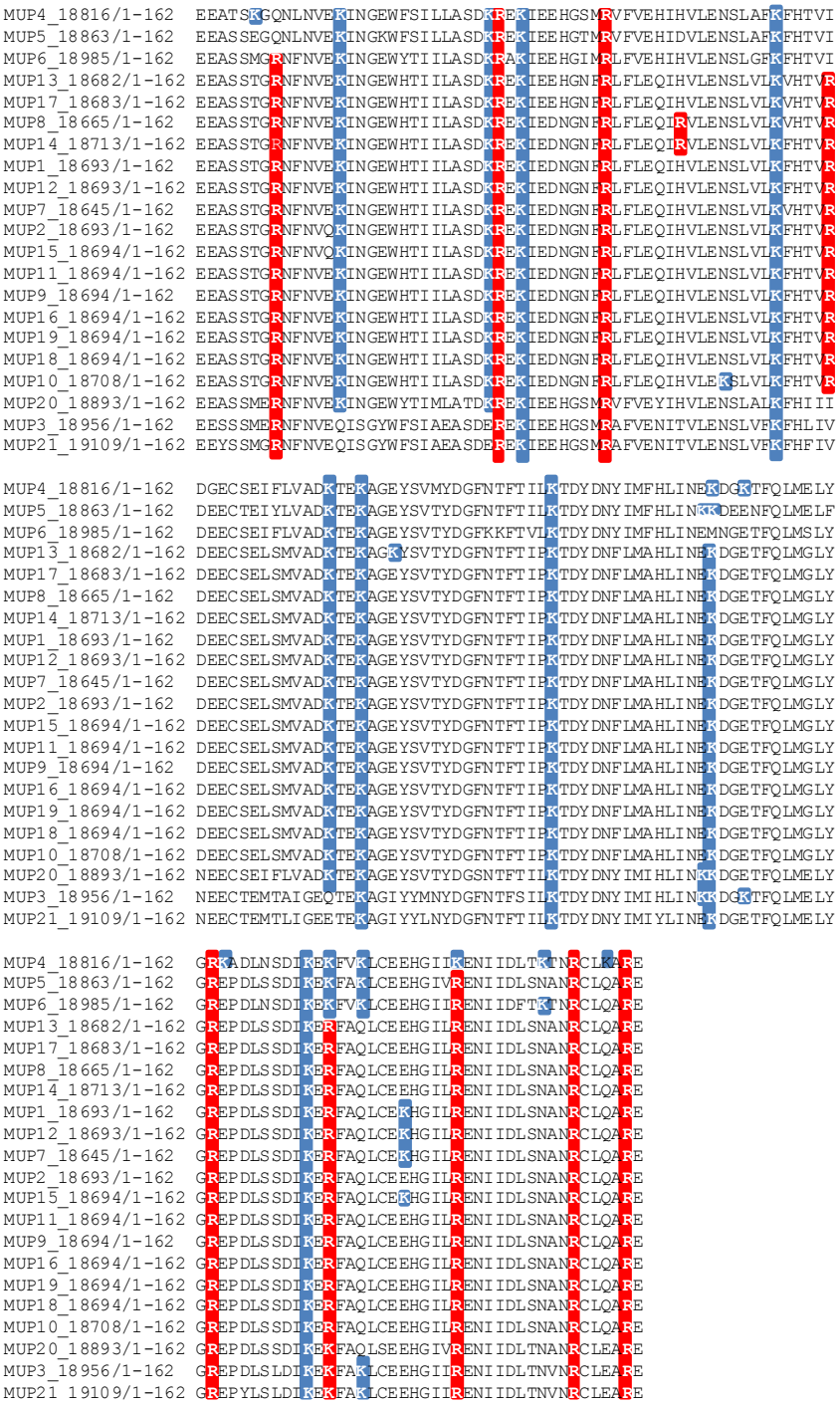
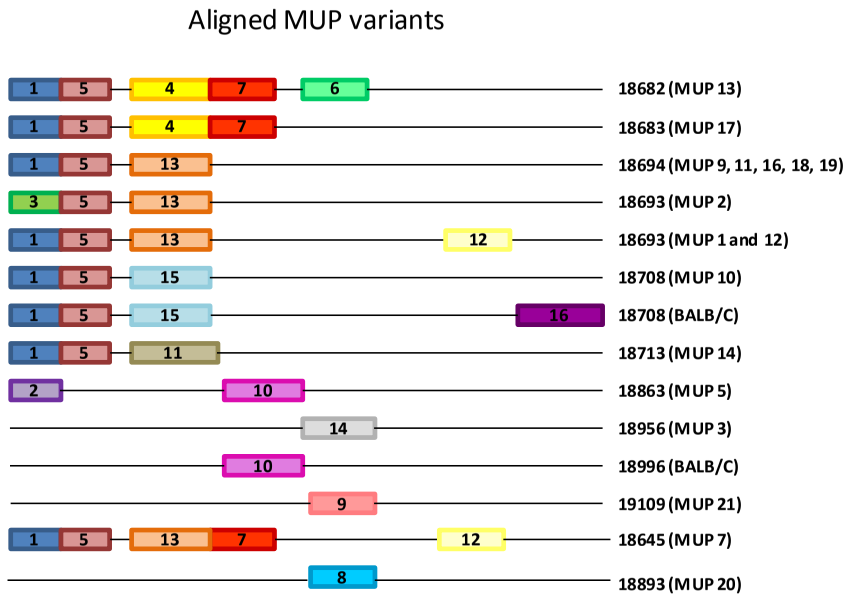


Fig 3.2 Proteolytic maps of MUP amino acid sequences.
Sequences were aligned using the ClustalW2 alignment tool for multiple sequences. The cleavage sites for LysC and trypsin are highlighted blue (lysine) and red (arginine).

The MUP QconCAT was expressed in *E. coli* (as described in the methods) with a single label [$^{13}C_6$] lysine. Samples were analysed by SDS-PAGE to ensure expression had taken place post IPTG induction (Figure 3.4). The QconCAT was then purified and aliquots of the wash and elution steps analysed by SDS-PAGE (Figure 3.5). To check that the correct product had been produced with complete labelling, an in-gel digest of the purified product followed by MALDI-TOF analysis was completed (Figure 3.6).



LysC fragment	Peptide mass[M+H⁺]	Sequence
1	1567.72	EEASSTGRNFNVEK
2	1533.69	EEAASSEGQNLNVEK
3	1566.74	EEASSTGRNFNVQK
4	2877.55	IEEHGNFRLFLEQIHVLENSLVLK
5	1596.82	INGEWHTIILASDK
6	1752.84	YSVTYDGFNTFTIPK
7	2104.93	VHTVRDEECSELSMVADK
8	1965.93	AGEYSVTYDGSNTFTILK
9	2113.05	AGIYYLNYDGFNTFTILK
10	2182.03	FHTVIDEECTEIYLVADK
11	2859.56	IEDNGNFRLFLEQIRVLENSLVLK
12	1180.54	ERFAQLCEK
13	2840.51	IEDNGNFRLFLEQIHVLENSLVLK
14	2116.99	AGIYYMNYDGFNTFSILK
15	2314.2	IEDNGNFRLFLEQIHVLEK
16	3123.48	LCEEHGILRENIILSNANRCLQARE

Fig. 3.3 LysC Peptides selected for incorporation into the MUP QconCAT and their relative location in the MUP protein.
Peptides chosen to be included in the QconCAT are represented by the coloured boxes. The sequences on the left were used as a model to assemble the QconCAT gene. The masses of those peptides containing a cysteine residue have been corrected for alkylation with iodoacetamde (mass increase of 58 Da).

[MUP 13] = ratio between unique MUP peptide 6 and the corresponding QconCAT peptide.

[MUP 17] = ratio between MUP peptide 4 and the corresponding QconCAT peptide - [MUP 13}

[MUP 7] = ratio between MUP peptide 7 and the corresponding QconCAT peptide – [MUP 13 + MUP 17]

[MUP 1+12] = ratio between MUP peptide 12 and the corresponding QconCAT peptide – [MUP 7]

[MUP 2] = ratio between unique MUP peptide 3 and the corresponding QconCAT peptide.

[MUP 14] = ratio between unique MUP peptide 11 and the corresponding QconCAT peptide.

[MUP 10] = ratio between unique MUP peptide 15 and the corresponding QconCAT peptide.

[MUP 5] = ratio between unique MUP peptide 2 and the corresponding QconCAT peptide.

[MUP 3] = ratio between unique MUP peptide 14 and the corresponding QconCAT peptide.

[MUP 21] = ratio between unique MUP peptide 9 and the corresponding QconCAT peptide.

[MUP 20] = ratio between unique MUP peptide 8 and the corresponding QconCAT peptide.

[MUP 9, 11, 16, 18, 19] = ratio between peptide 13 and the corresponding QconCAT peptide – [MUP 2 + MUP 1 and 12 + MUP 7]

Alternative calculations for MUPs 9, 11, 16, 18 and 19

[MUP 9, 11, 16, 18, 19] = ratio between peptide 1 and the corresponding QconCAT peptide – [MUP 13 + MUP 17 + MUP 1and 12 + MUP 10 + MUP 14 + MUP 7]

[MUP 9, 11, 16, 18, 19] = ratio between peptide 5 and the corresponding QconCAT peptide – [MUP 13 + MUP 17 + MUP 2 + MUP 1and 12 + MUP 10 + MUP 14 + MUP 7]

Fig. 3.4 Strategy for the quantification of MUPs.
Each individual MUP will be calculated using either a unique peptide or by subtracting away amounts calculated for other MUPs that have a shared peptide. (See figure 3.3)

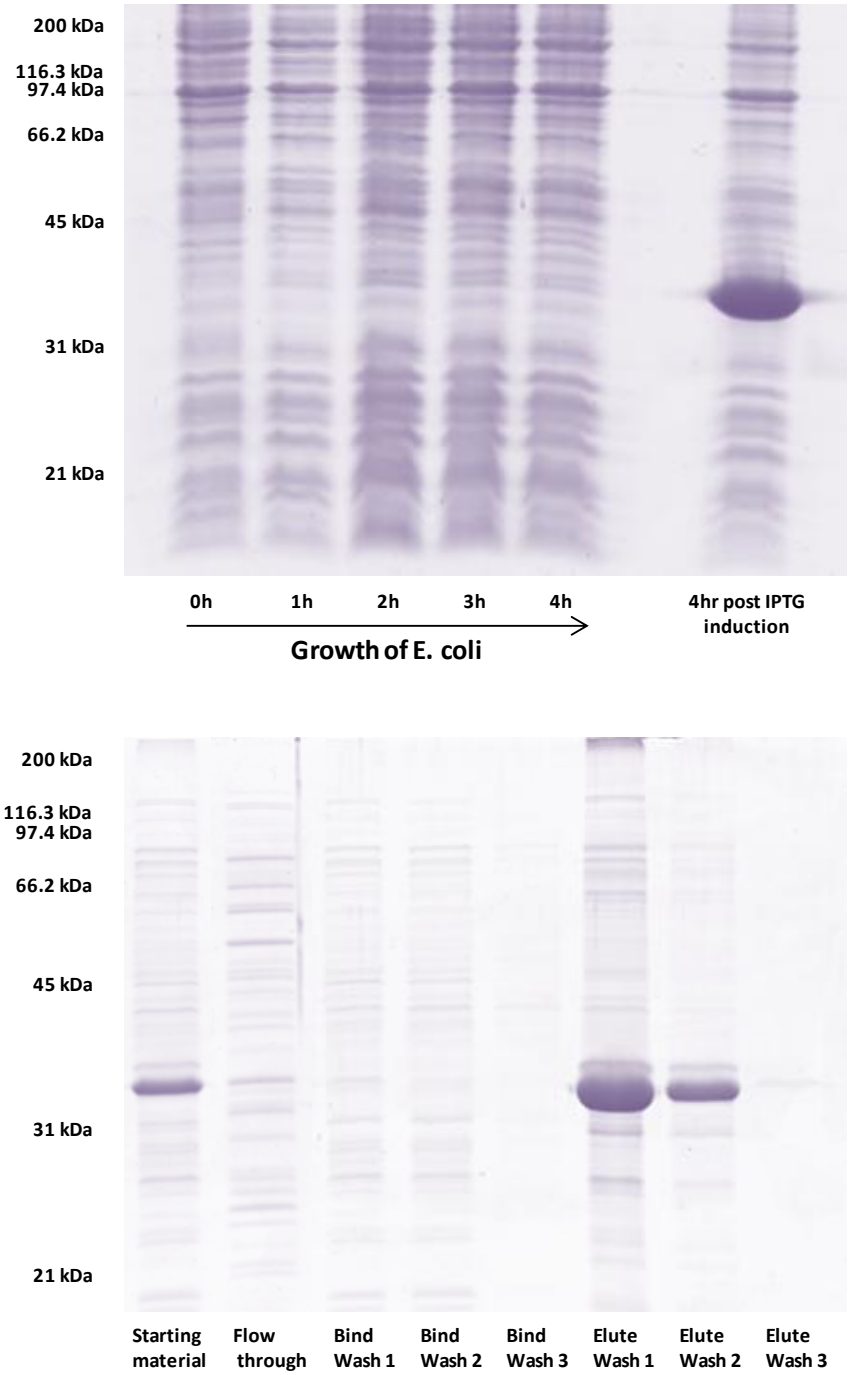


Fig.3.5 Expression and purification of the MUP QconCAT.
Top gel. The MUP QconCAT was expressed in *E. coli and* labelled with $^{13}C_6$ Lysine. The OD (600nm) of *E. coli* was taken every hour until it reached an absorbance reading of 0.6. IPTG was then added to the culture to induce E. coli to synthesise the protein. Bottom gel. A MUP QconCAT cell pellet was then purified by solubilising the inclusion bodies in NaCl (2 M), sodium phosphate (80 mM, pH 7.4), GnCl (6 M) and imidazole (40 mM). The solubilised inclusion body was then filtered and passed through a 1 ml HisTrap column. The purified protein was eluted in NaCl (2 M), sodium phosphate (80 mM, pH 7.4), GnCl (6 M) and imidazole (2 M). Elute wash 1 and 2 were combined and dialysed overnight in NH_4CO_3 (100 mM , pH 8.5). Samples were run on a 15% SDS gel and stained with coomassie blue.

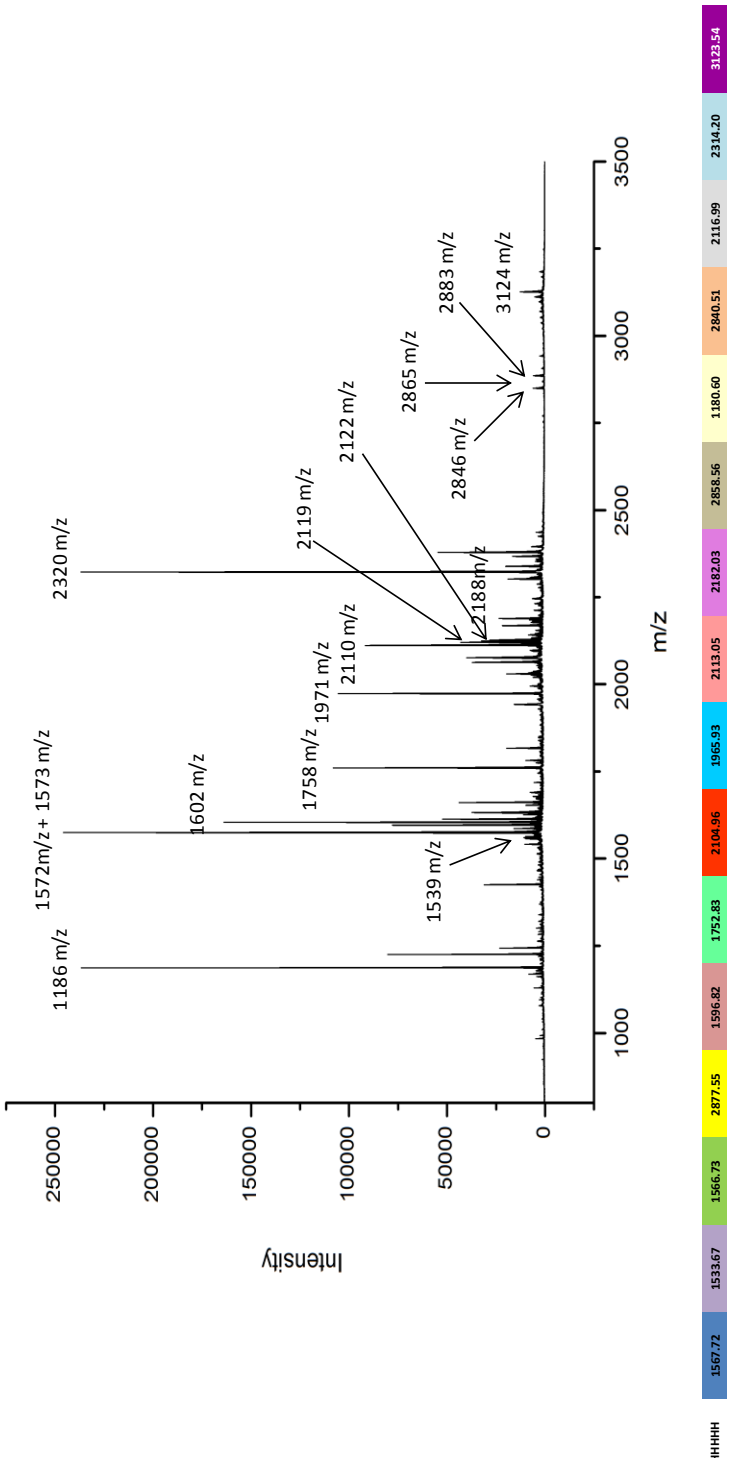


Fig.3.6 MALDI-TOF analysis of purified QconCAT.
A small piece of gel was removed from the SDS-PAGE analysis of the purified QconCAT and reduced and alkylated with DTT and iodoacetamide before being incubated overnight at 37 °C with LysC. The peptides were recovered the following day and mixed 1:1 with α-Cyano-4-hydroxycinnamic acid dissolved in 50% ACN, 0.1% TFA. The mixture (1 µl) was spotted onto a MALDI target plate and left to dry at room temperature before being analysed by MALDI-TOF. The coloured boxes underneath the peptide mass fingerprint represent LysC fragments of the QconCAT, 100% coverage of the protein was achieved.

3.3.2 Optimising proteolysis of the native protein

Complete digestion of the analyte and concatenated standard is essential for absolute quantification. Complete digestion of the MUP QconCAT should be easily achieved as it lacks secondary structure (Figure 3.7) (Pratt *et al.,* 2006). The MUP proteins present in the urine are more challenging to digest because they have an extensive beta sheet conformation (Flower *et al.,* 1993; Flower, 1996), making them difficult to digest. Using a standard digest protocol of reduction and alkylation followed by enzyme proteolysis is not very effective against MUPS (Figure 3.7). A comparison of undigested and digested MUP by SDS-PAGE shows very little proteolysis with a minor shift in molecular weight for the digested material (Dr S Armstrong thesis, Wu *et al.,* 1999). This is thought to be due to the digested MUPS missing the N and C termini that have been cleaved off by the protease. The rest of the protein is intact and resistant to further proteolysis (Wu *et al.,* 1999).

Previous attempts at digesting MUPS have included the use of denaturing reagents such as urea (8 M) combined with increased enzyme: protein ratios (Dr S Armstrong, thesis; Dr D Simpson personal communication). While this method has had some success, it can be time consuming and many denaturing reagents are not compatible with LC-MS analysis. An alternative more LC-MS compatible reagent *Rapi*Gest™ SF Surfactant provides a feasible alternative to the harsh denaturing buffers. *Rapi*Gest™ makes proteins more soluble and therefore more susceptible to enzymatic cleavage. Unlike denaturing reagents it does not suppress protease activity or modify substrates. After overnight incubation with protease, digests are acidified with TFA to break down *Rapi*Gest™ into by-products that do not interfere with LC-MS analysis (Yu YQ *et al.,* 2003; 2004).

Initially, MUPS (100 µg) were digested using the recommended protocol for *Rapi*Gest™. *Rapi*Gest™ was added to the sample to be digested (total concentration 0.05%) and the sample heated at 80 °C for 10 minutes. This was followed by reduction with DTT (3 mM final concentration) and alkylation with iodoacetamide (9 mM final concentration). The digest was then incubated overnight with trypsin (50:1 substrate: protease). Trypsin was used in the optimisation experiments

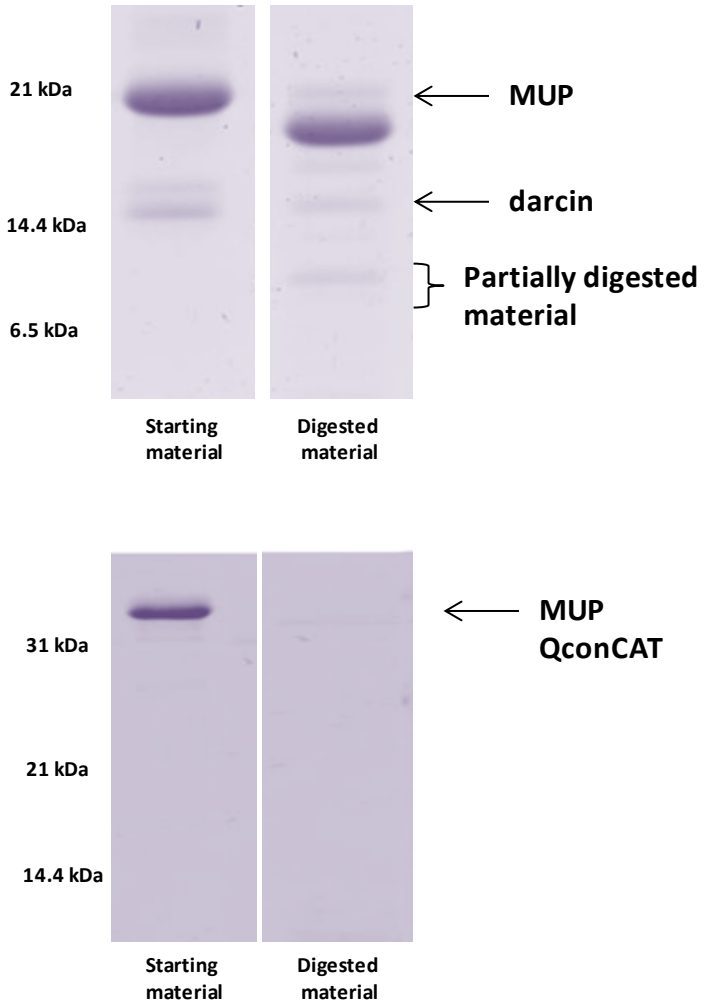


Fig.3.7 Digestion of MUPs and QconCAT using a standard in-solution digest protocol.
Top gel. MUPS (100 µg) from C57/BL6 male mouse urine were reduced with DTT (3 mM final concentration) and alkylated with iodoacetamide (9 mM final concentration). Trypsin was added (50:1 substrate:enzyme) and the digest incubated overnight at 37 °C. An aliquot of the starting material (100 µg MUP diluted into 25 mM NH_4CO_3) was reserved to compare to the digested material. The same amount of starting and digested material was loaded onto the gel so a direct comparison could be made. Samples were run on a 15% SDS gel and stained with coomassie blue stain. Bottom gel. QconCAT(100 µg) was reduced with DTT (3 mM final concentration) and alkylated with iodoacetamide (9 mM final concentration). Trypsin was added (50:1 substrate:enzyme) and the digest incubated overnight at 37 °C. An aliquot of the starting material (100 µg QconCAT diluted into 25 mM NH_4CO_3) was reserved to compare to the digested material. The same amount of starting and digested material was loaded onto the gel so a direct comparison could be made. Samples were run on a 15% SDS gel and stained with coomassie blue

because it was more readily available and more cost effective. An aliquot of the digest was removed the following morning before removal of the *Rapi*Gest™ and resolved by SDS-PAGE along with the starting material to assess the degree of digestion. The majority of the protein appeared to be digested with just a faint band seen in the digested material (Figure 3.8). The digest was repeated again using increased concentrations of *Rapi*Gest™ 0.05%, 0.1%, 0.2%, and 0.4%. Slightly more protein appeared to be digested using the 0.1% *Rapi*Gest™ solution but not much improvement was seen increasing the concentration after that (Figure 3.9). Using 0.1% *Rapi*Gest™ as the new standard concentration, a time course experiment was done to see if this could further improve digestion (Figure 3.10). Six digests were prepared and the *Rapi*Gest™ solution (0.1 %) was added to the samples, one digest was heated for the standard 10 minutes, one digest heated for 20 minutes etc up to 60 minutes. Again the samples were resolved by SD-PAGE and level of digestion compared. There did not appear to be a significant improvement after ten minutes so the protocol was not amended.

The reason for incomplete digestion was unclear. It was either due to the trypsin activity reducing over time, which could be solved using an enzyme top up step, or the MUPS were still forming inhibitory products making them resistant to complete proteolysis despite the introduction of *Rapi*Gest™ to the protocol. A set of three digests were prepared in duplicate using the *Rapi*Gest™ (0.1%) protocol. One digest contained just MUP protein, one digest contained just bovine albumin and one digest contained both MUP and albumin. Following overnight incubation, an additional amount trypsin was added to just one replicate of each. The samples were left to incubate for a further 6 hours.

Aliquots of both replicates were compared by SDS-PAGE. In the MUP only digest there was no improvement in proteolysis, the addition of extra enzyme had no effect. The albumin only digest went on to be completely digested after the enzyme top up. In the co-digest there was no improvement seen in MUP digestion but once again the albumin was completely digested after the top up.

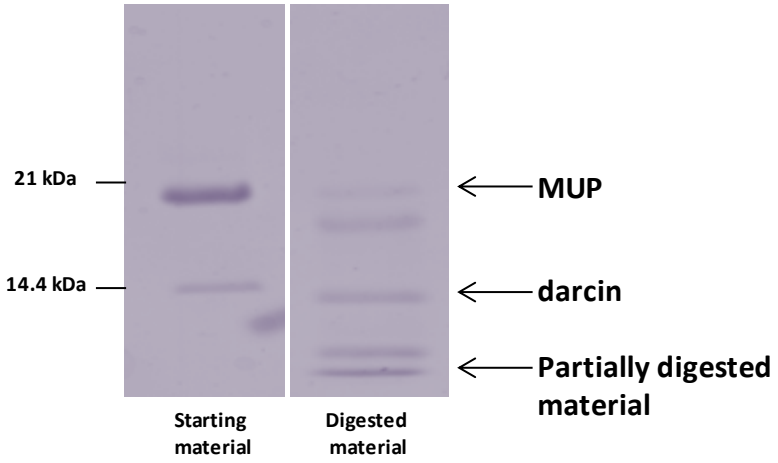


Fig.3.8 Digestion of MUPS using a standard in-solution digest protocol plus *Rapi*Gest™.
MUPS (100 µg) from C57/BL6 male mouse urine were digested using a 0.05% solution of *Rapi*Gest™ followed by a reduction with DTT (3 mM final concentration) and alkylation with iodoacetamide (9 mM final concentration) step. Trypsin was added (50:1 substrate:enzyme) and the digest incubated overnight at 37 °C. An aliquot of the starting material (100 µg MUP diluted into 25 mM NH_4CO_3) was reserved to compare to the digested material. The same amount of starting and digested material was loaded onto the gel so a direct comparison could be made. Samples were run on a 15% SDS gel and stained with coomassie blue stain.

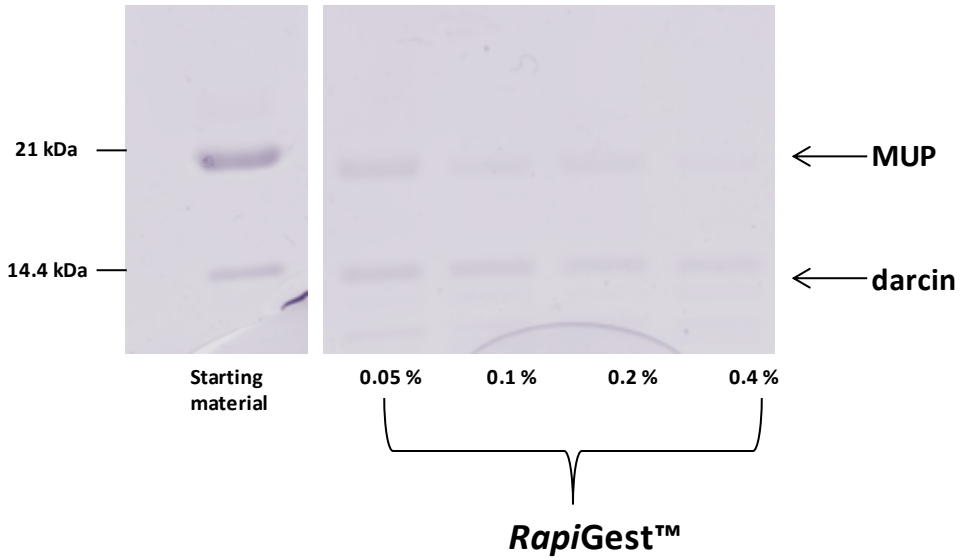


Fig.3.9 Digestion of MUPs using increasing concentrations of *Rapi*Gest™.
Four individual MUP digests (100 µg) were prepared. A different concentration of *Rapi*Gest™ was used in each sample – 0.05%, 0.1%, 0.2% and 0.4% followed by reduction with DTT (3 mM final concentration) and alkylation with iodoacetamide (9 mM final concentration). Trypsin was added (50:1 substrate:enzyme) and the digests incubated overnight at 37 °C. The same amount of starting and digested material was loaded onto the gel so a direct comparison could be made. Samples were run on a 15% SDS gel and stained with coomassie blue stain.

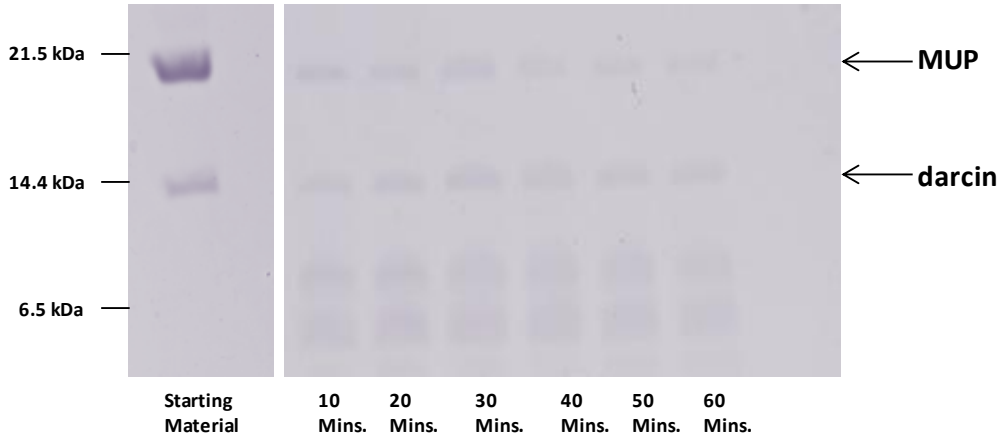


Fig.3.10 *Rapi*Gest™ time course experiment.
A MUP digest (100 µg) was prepared and mixed with *Rapi*Gest™ (0.1% total concentration) and incubated at 80 °C. An aliquot was removed from the incubator every 10 minutes. After the *Rapi*Gest™ incubation step, each aliquot was then reduced with DTT (3 mM final concentration) and alkylated with iodoacetamide (9 mM final concentration). Trypsin was added (50:1 substrate:enzyme) and the digests incubated overnight at 37 °C . Starting material was again retained for comparison purposes. The same amount of starting and digested material was loaded onto the gel so a direct comparison could be made. Samples were run on a 15% SDS gel and stained with coomassie blue stain.

A fourth sample was also prepared containing just MUP. This sample was digested with trypsin overnight with an additional amount of albumin added the following morning. The sample was left to incubate for a further 6 hours and an aliquot taken and resolved on SDS-PAGE. There was some digestion of albumin indicating that trypsin was still active. This concludes that MUPS are still forming inhibitory products making them resistant to further proteolysis by trypsin (Figure 3.11 and 3.12).

The digestion protocol used 100 µg of protein. It was anticipated that reducing the amount of substrate to be digested will in turn reduce the amount of inhibitory products formed. Three digests were prepared – one that contained 100 µg, one that contained 50 µg and one that contained 10 µg of MUP protein. All three were digested using the *Rapi*Gest™ protocol and incubated overnight with trypsin. Aliquots were removed the following day and analysed by SDS-PAGE (Figure 3.13). The same amount of protein was loaded onto the gel so a direct comparison could be made between all three digests. There looked to be complete digestion in the 10 µg digest.

As the QconCAT was labelled with [$^{13}C_6$] lysine, the samples for quantification would have to be digested using LysC. MUP (10 µg) was digested with LysC to ensure that the method optimised using trypsin was appropriate for digestion with LysC. Starting material and digested material was resolved by SDS-PAGE (Figure 3.14). No bands were identified in the digested material indicating full digestion had been achieved.

3.3.3 Co-digestion and LC-MS analysis of analyte and QconCAT

Following optimisation of the digest, both analyte and QconCAT were mixed in a 1:1 (by protein ratio) and digested using the protocol listed in the methods section. The samples were analysed by LC-MS and each heavy: light pair was examined individually. Many of the heavy: light pairs were not detected.

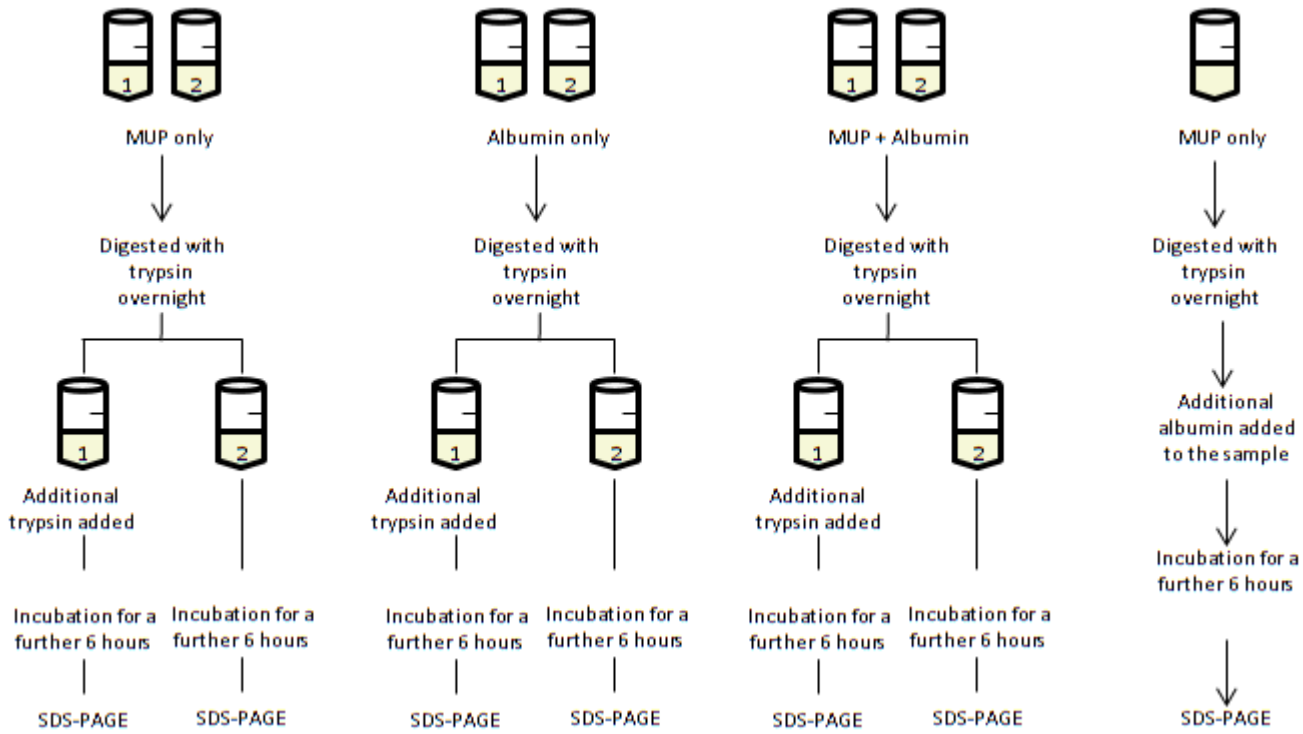


Fig. 3.11 Workflow to determine the cause of incomplete digestion of MUPS.
A set of three digests were prepared in duplicate using the *Rapi*Gest™ (0.1%) protocol. The first set of digests contained just MUP protein, the second set contained just bovine albumin and the third set contained both MUP and albumin. Following overnight incubation with trypsin, an additional amount trypsin was added to just one replicate of each. All samples were left to incubate for a further 6 hours. A fourth sample was also prepared containing just MUP. This sample was digested with trypsin overnight with an additional amount of albumin added the following morning. This sample was also left to incubate for a further 6 hours. An aliquot of each samples was taken for SDS-PAGE analysis (Figure 3.12) to determine if incomplete digestion was due to trypsin loosing activity or if MUPS were still resistant to proteolysis despite the introduction of *Rapi*Gest™ to the protocol

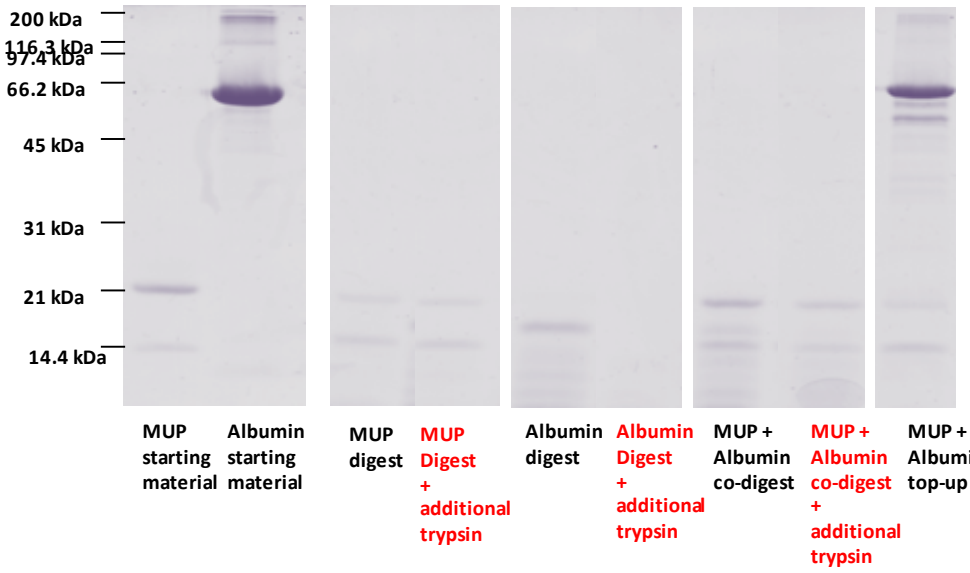


Fig.3.12 SDS-PAGE analysis to identify the cause of incomplete digestion of MUPS.
A set of three digests were prepared in duplicate using the *Rapi*Gest™ (0.1%) protocol. The first set of digests contained just MUP protein, the second set contained just bovine albumin and the third set contained both MUP and albumin. Following overnight incubation with trypsin, an additional amount trypsin was added to just one replicate of each. All samples were left to incubate for a further 6 hours. A fourth sample was also prepared containing just MUP. This sample was digested with trypsin overnight with an additional amount of albumin added the following morning. This sample was also left to incubate for a further 6 hours. An aliquot of each samples was taken for SDS-PAGE analysis to determine if incomplete digestion was due to trypsin loosing activity or if MUPS were still managing to inhibit the protease despite the introduction of *Rapi*Gest™ to the protocol. Starting material was again retained for comparison purposes. The same amount of starting and digested material was loaded onto the gel so a direct comparison could be made. Samples were run on a 15% SDS gel and stained with coomassie blue stain.

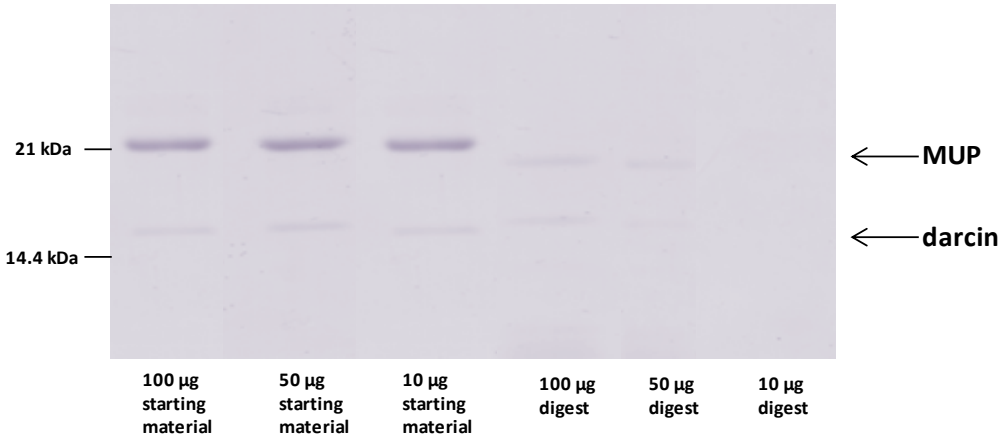


Fig.3.13 Observing the degree of digestion using decreasing amounts of MUP in the starting material .

Three digests were prepared containing different amounts of MUPS – 100 µg, 50 µg and 10 µg. All were incubated with 0.1% *Rapi*Gest™ at 80 °C before being reduced and alkylated with DTT (3 mM) and iodoacetamide (9 mM) respectively. All were digested overnight with trypsin (substrate:enzyme 50:1) at 37 °C. Degree of digestion was then compared by SDS-PAGE. Starting material was again retained for comparison purposes. The same amount of starting and digested material was loaded onto the gel so a direct comparison could be made. Samples were run on a 15% SDS gel and stained with coomassie blue stain.

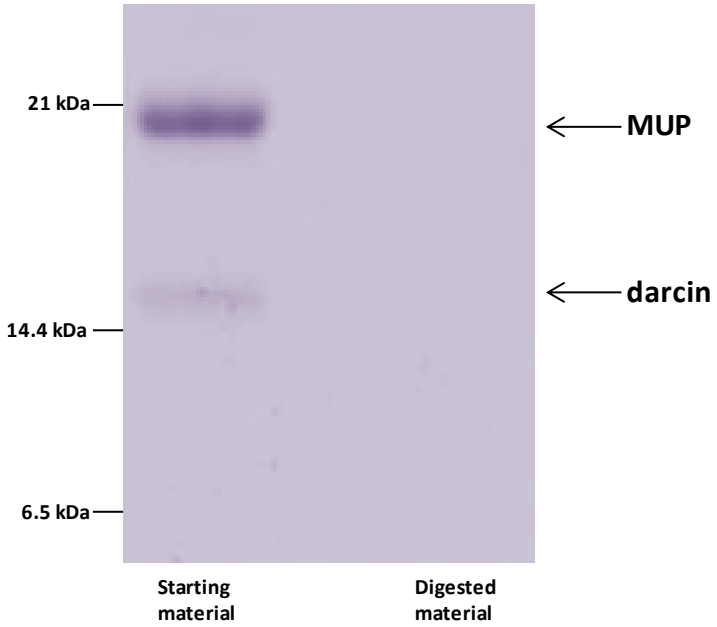


Fig.3.14 Proteolysis of MUPS with LysC using the optimised digest method.
MUP (10 µg) was incubated with 0.1% *Rapi*Gest™ at 80 °C before being reduced and alkylated with DTT (3 mM) and iodoacetamide (9 mM) respectively before overnight digestion with LysC (substrate:enzyme 50:1) at 37 °C. Degree of digestion was then compared by SDS-PAGE. Starting material was again retained for comparison purposes. The same amount of starting and digested material was loaded onto the gel so a direct comparison could be made. Samples were run on a 15% SDS gel and stained with coomassie blue stain.

It was unclear why these peptide pairs were not observed. It was possible that the digest had not been successful and there were missed cleavages that would have been too large to see by LC-MS or the peptides not observed were not suitable for analysis by LC-MS. As all the Q peptides had previously been identified using MALDI-TOF during the purification step, the digest was also re-analysed using MALDI-TOF. All the Q peptides could now be seen using MALDI-TOF analysis which indicated that maybe some peptides were unsuitable for the type of chromatography being used. Inspection of the Q peptide sequences indicated a high proportion of hydrophobic residues. The stationary phase of an LC column consists of hydrophobic alkyl side chains that interact with the analyte. These carbon chains can vary in length C4, C8 and C18, C18 being the most hydrophobic and C4 the least hydrophobic. C4 columns are used to analyse large molecules and proteins, the idea being that they will have more hydrophobic sites and will therefore only require a shorter side chain on the stationary phase to interact with. Peptides are routinely separated using C18 stationary phase because they are smaller and therefore have less hydrophobic sites and are more easily detained by the longer hydrophobic carbon side chains.

A C18 reverse phase column had previously been used to analyse the MUP digests. As many of the Q peptides were hydrophobic a C4 column was used as an alternative to the C18 and the results compared (Figure 3.15). The data was processed using maximum entropy software (MaxEnt 3, MassLynx 4.1, Waters) to deconvolute the spectra to make a visual comparison and identification of peptides less complicated. The software takes multiply charged spectra and deconvolutes it into singly charged spectra. This is particularly useful for analysing the MUP Q peptides as there are masses that are very similar and only differ by 1 Da and therefore have overlapping isotopic patterns. A significant improvement was seen in the number of peptides detected using the C4 column. Unfortunately the more hydrophilic peptides that were detected using the C18 column had not been detained on the C4 column and it is most likely they were lost during the trapping step of the LC method.

Investigating other column options was not feasible as there were some peptides that also didn't appear to ionise very well – peptide 4 (2883 m/z), peptide 11 (2865 m/z) and peptide 13 (2846 m/z) – giving a low signal to noise ratio making quantification difficult. During replicate runs of the digests, there were occasions when these three peptides were not detected at all. This would affect quantification considerably as many of the MUPS are quantified using the subtraction method explained previously so the absence of these three peptides would have meant no quantification data for a number of MUPS.

Another approach would have to be taken to quantify MUPS. The current QconCAT (Figure 3.3) was theoretically digested with trypsin. The smaller tryptic fragments should ionise better and also be more compatible with a C18 column. The peptide sequences were examined to see how many MUPS now shared the new tryptic fragments and how this would affect the quantification. Fortunately this did not have a large impact on the strategy for quantification. The MUPS that had unique peptides still had a unique fragment (Figure 3.16). Peptide 7 was the only peptide that when digested with trypsin gave two fragments that were shared with many more MUPS disrupting the strategy for quantification considerably. Using a combination of two proteases LysC and trypsin would eliminate this problem.

There were also three MUPS, 20, 21 and 3, who had Q peptides that did not have an internal arginine residue. These were also peptides that were not compatible with C18 chromatography. As a plasmid was available for MUP 20, more commonly known as darcin, a labelled version of darcin was made and used as a standard. New peptides were chosen for these three MUPS. Again this was based on a subtraction method. MUP 20 (darcin) contained all 3 peptides so quantification was possible. The amount of MUP 20 was to be calculated using peptide 8. MUPS 3 and 21 could then be quantified by subtracting their unique peptides away from the value calculated for darcin. The revised strategy for quantification is outlined in Figures 3.17 and 3.18.

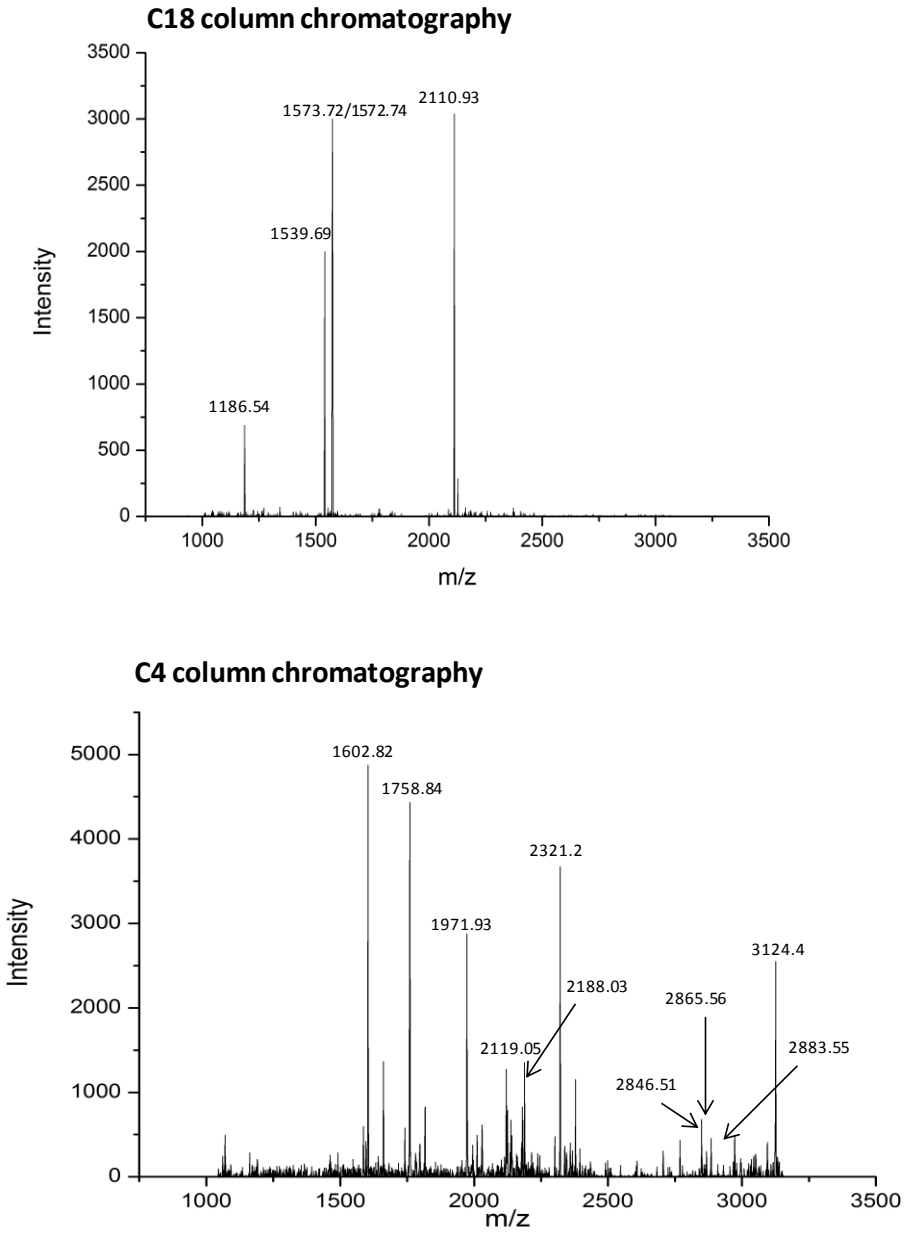


Fig.3.15 Comparison between C18 and C4 column chromatography.
MUP QconCAT was digested using the protocol described in section 3.2 and analysed by LC-MS. One chromatography system was set up with a C18 column (top graph) and one system set up with a C4 column (bottom graph) and the digest analysed on both systems. Only the least hydrophobic peptides have been identified using the C18 column. These peptides were not seen when C4 chromatography was used, they have most likely been lost at the trapping stage prior to the start of the gradient. The more hydrophobic peptides are absent from the C18 analysis but were observed when analysed using C4 chromatography. Spectra were deconvoluted using maximum entropy software (MassLynx 4.1, Waters).

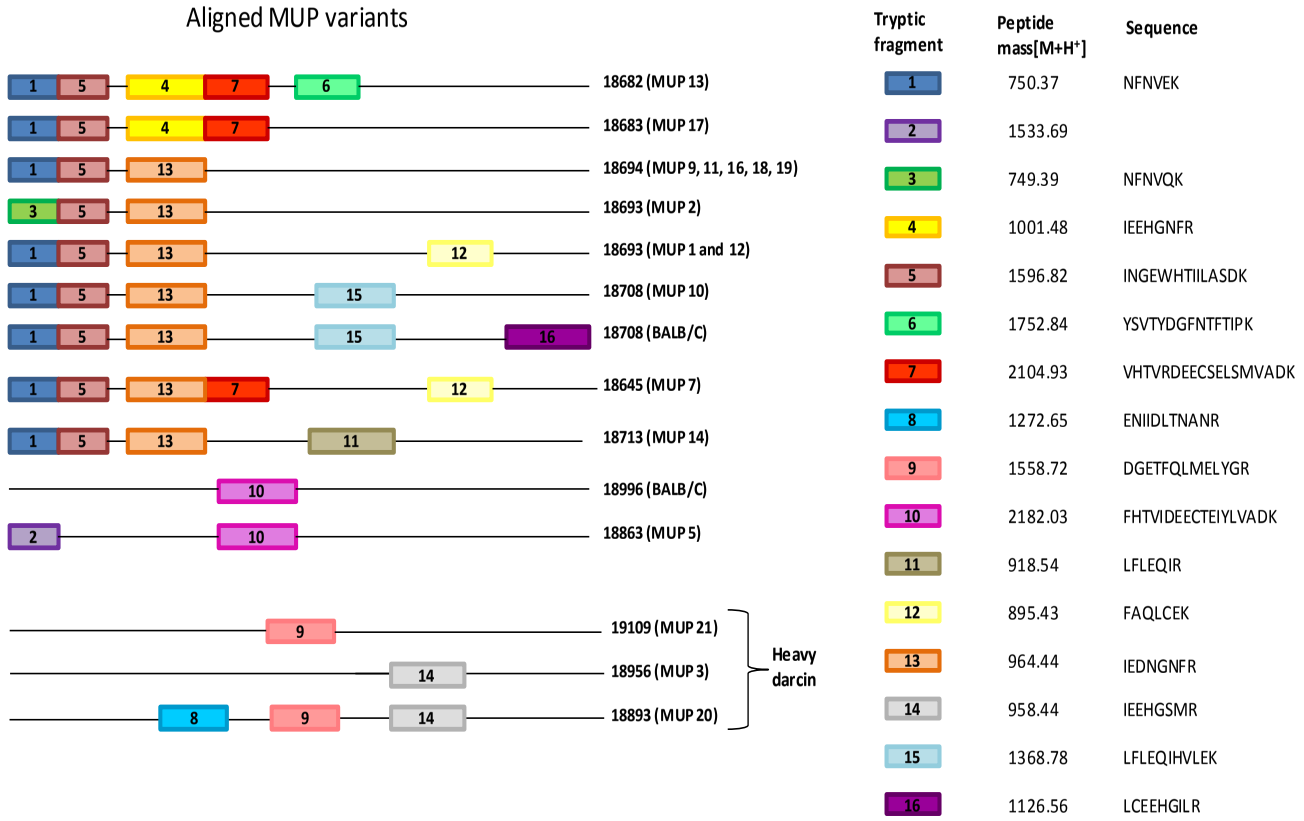


Fig. 3.16 Tryptic fragments selected for incorporation into the MUP QconCAT and their relative location in the MUP protein.
Tryptic fragments chosen based on the original LysC peptides chosen for the QconCAT are represented by the coloured blocks. The peptide masses on the right hand side have again been corrected for alkylation with iodoacetamide. MUPS 21, 3 and 20 had completely new peptides chosen as there was no internal arginine cleavage site. These MUPS will be quantified using heavy labelled darcin as a standard.

[MUP 13] = ratio between unique MUP peptide 6 and the corresponding QconCAT peptide.

[MUP 17] = ratio between MUP peptide 4 and the corresponding QconCAT peptide - [MUP 13}

[MUP 7] = ratio between MUP peptide 7 and the corresponding QconCAT peptide – [MUP 13 + MUP 17]

[MUP 1+12] = ratio between MUP peptide 12 and the corresponding QconCAT peptide – [MUP 7]

[MUP 2] = ratio between unique MUP peptide 3 and the corresponding QconCAT peptide.

[MUP 14]= ratio between unique MUP peptide 11 and the corresponding QconCAT peptide.

[MUP 10] = ratio between unique MUP peptide 15 and the corresponding QconCAT peptide.

[MUP 5] = ratio between unique MUP peptide 2 and the corresponding QconCAT peptide.

[MUP 9, 11, 16, 18, 19] = ratio between MUP peptide 13 and the corresponding QconCAT peptide – [MUP 2 + MUP 1 and 12 + MUP 7 + MUP 10+ MUP 14]

Alternative calculations for MUPs 9, 11, 16, 18 and 19

[MUP 9, 11, 16, 18, 19] = ratio between MUP peptide 1 and the corresponding QconCAT peptide – [MUP 13 + MUP 17 + MUP 1and 12 + MUP 10 + MUP 14 + MUP 7]

[MUP 9, 11, 16, 18, 19] = ratio between MUP peptide 5 and the corresponding QconCAT peptide – [MUP 13 + MUP 17 + MUP 2 + MUP 1and 12 + MUP 10 + MUP 14 + MUP 7]

[MUP 20] = ratio between unique MUP peptide 8 and the corresponding labelled darcin peptide.

[MUP 3] = ratio between MUP peptide 14 and the corresponding labelled darcin peptide – [MUP 20]

[MUP 21] = ratio between MUP peptide 9 and the corresponding labelled darcin peptide.- [MUP 20]

Fig. 3.17 Strategy for the quantification of MUPS using trypsin.
Each individual MUP will be calculated using either a unique peptide or by subtracting away amounts calculated for other MUPS that have a shared peptide. (See figure 3.16)

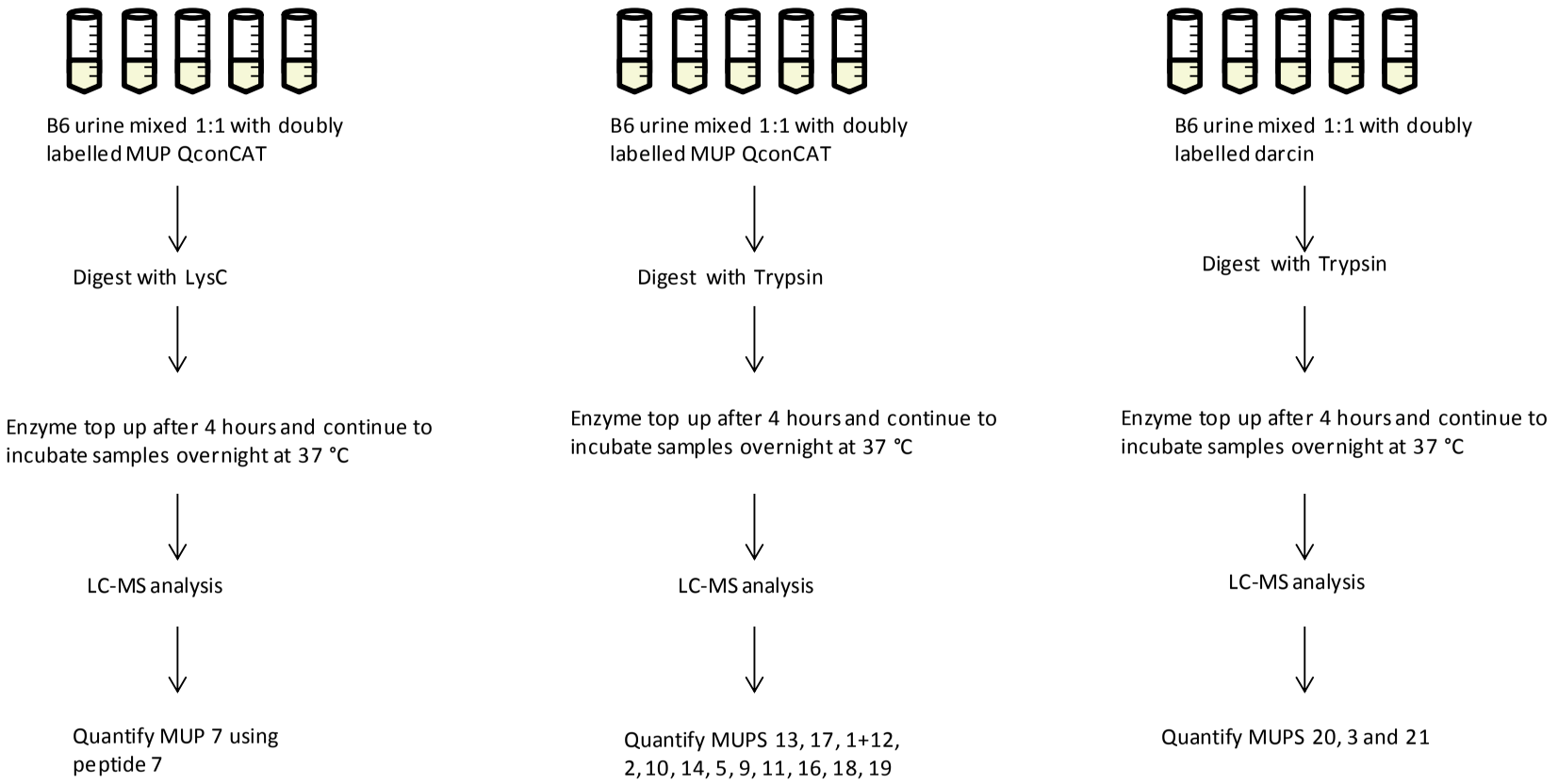


Fig. 3.18 Revised MUP quantification workflow.
Three separate digests for each mouse will be prepared, two with QconCAT and one with heavy darcin. The first set of urine and QconCAT sample will be digested with LysC to enable quantification of MUP 7 using peptide 7 which is no longer restricted to MUPS 7, 13 and 17 when digested with trypsin making quantification impossible. The second set of urine and QconCAT samples will be digested with trypsin to quantify the majority of the other MUPS 13, 17, 1+12, 2, 10, 14, 5, 9, 11, 16, 18, 19. The third set of samples are B6 urine mixed 1:1 with labelled darcin. These will be digested with trypsin to allow quantification of MUPS 2, 20 and 21.

3.3.4 Quantification of MUPS using a doubly labelled QconCAT and darcin

The MUP QconCAT was again expressed in *E. coli* and labelled with [$^{13}C_6$] $_{lysine}$ and [$^{13}C_6$] arginine. Samples were again resolved by SDS-PAGE to ensure expression had taken place post IPTG induction (Figure 3.19). The QconCAT was then purified and aliquots of the wash and elution steps analysed by SDS-PAGE (Figure 3.19). An in-gel digest and MALDI-TOF analysis were carried out to check labelling had occurred (Figure 3.20). A labelled version of darcin was also expressed and purified in the same manner as the QconCAT (Figure 3.21) followed by MALDI-TOF analysis to confirm labelling (Figure 3.22).

Prior to co-digestion with B6 urine, both the QconCAT and darcin were individually digested and analysed by LC-MS using a C18 column. The Q peptides chosen to represent MUPS 20 and 3 in the heavy darcin standard were ideal for analysis. The peptide chosen for MUP 21 unfortunately was not suitable and as there were no other options to quantify, this amount of this MUP could not be calculated. All tryptic fragments from the QconCAT were identified by LC-MS. The isotopic patterns for peptides 5 and 13 were slightly unusual. Both sequences for these peptides contain the Asn-Gly (N-G) that can result in deamidation. Deamidation is a non-enzymatic process (Robinson and Rudd, 1974) in which the side chain of asparagine is converted into aspartic or isoaspartic acid. The side chain of asparagine attacks the peptide group on the C terminal side which leads to the formation of a succinamide intermediate. This intermediate is then hydrolysed to form aspartic or isoaspartic acid resulting in a mass shift of +1 Da (Geiger and Clarke 1987; Bischoff *et al.,* 1993). Deamidation is more likely to occur if asparagine is followed by glycine in the amino acid sequence. As glycine is small with a low steric hindrance it is more open to attack by the asparagine side chain (Robinson NE *et al.,* 2001; Rivers *et al.,* 2008). The unusual isotopic pattern of peptide 5 and 13 suggests that there is a mixture of deamidated and non-deamidated forms (Figure 3.23). If both the acid and amide forms both ionised in a similar manner then the monoisotopic masses for each form could be added up and used for quantification. Peptide 5 was used in an experiment to test whether or not the amide and acid forms behaved the same when analysed by mass spectrometry.

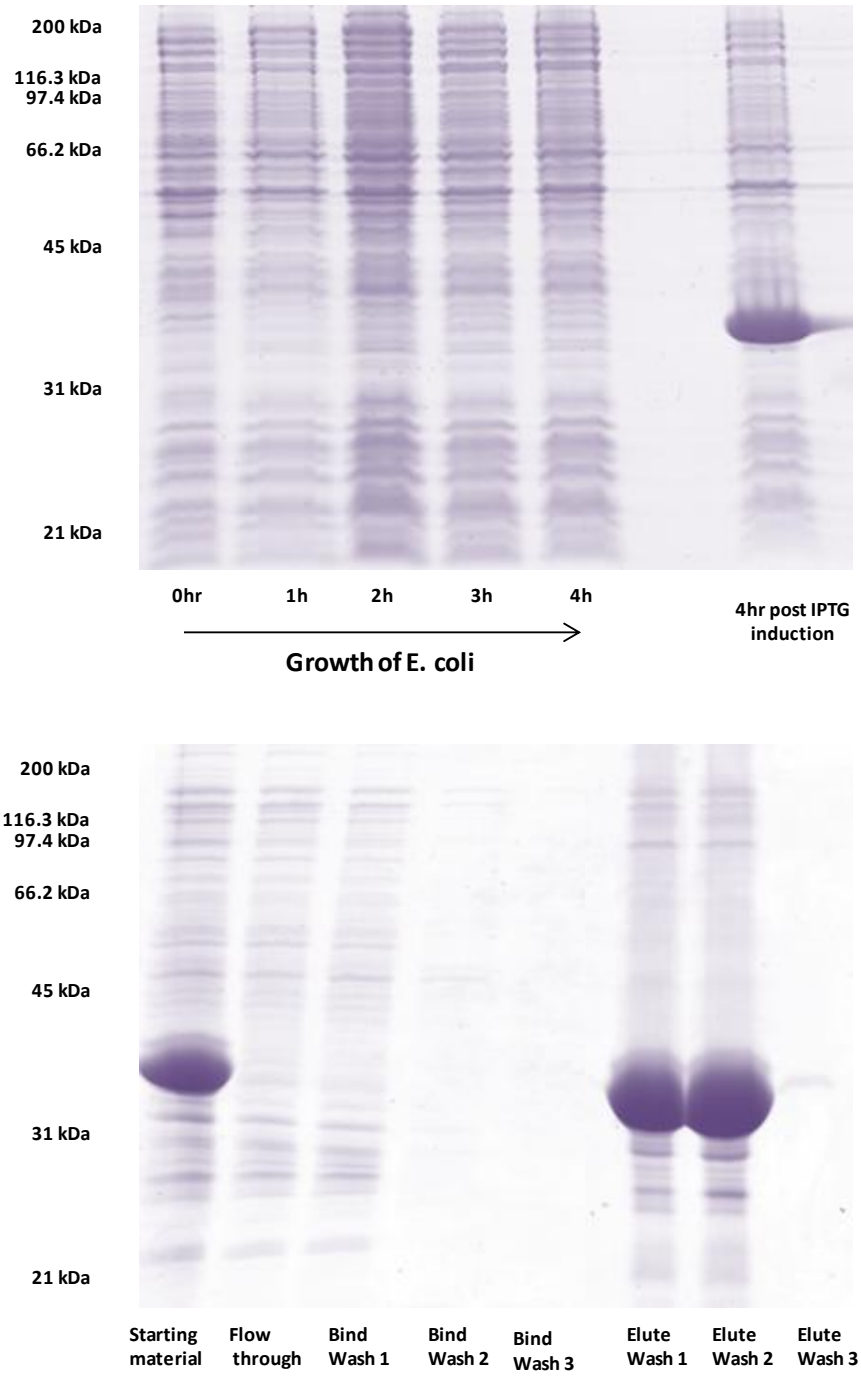


Fig.3.19 Expression and purification of a doubly labelled MUP QconCAT.
Top gel. The MUP QconCAT was expressed in *E. coli* and labelled with $^{13}C_6$ Lysine and $^{13}C_6$ arginine . The OD (600nm) of *E. coli* was taken every hour until it reached an absorbance reading of 0.6. IPTG was then added to the culture to induce *E. coli* to synthesise the protein. Bottom gel. A MUP QconCAT cell pellet was then purified by solubilising the inclusion bodies in NaCl (2 M), sodium phosphate (80 mM, pH 7.4), GnCl (6 M) and imidazole (40 mM). The solubilised inclusion body was then filtered and passed through a 1 ml HisTrap column. The purified protein was eluted in NaCl (2 M), sodium phosphate (80 mM, pH 7.4), GnCl (6 M) and imidazole (2 M). Elute wash 1 and 2 were combined and dialysed overnight in NH_4CO_3 (100 mM , pH 8.5). Samples were run on a 15% SDS gel and stained with coomassie blue.

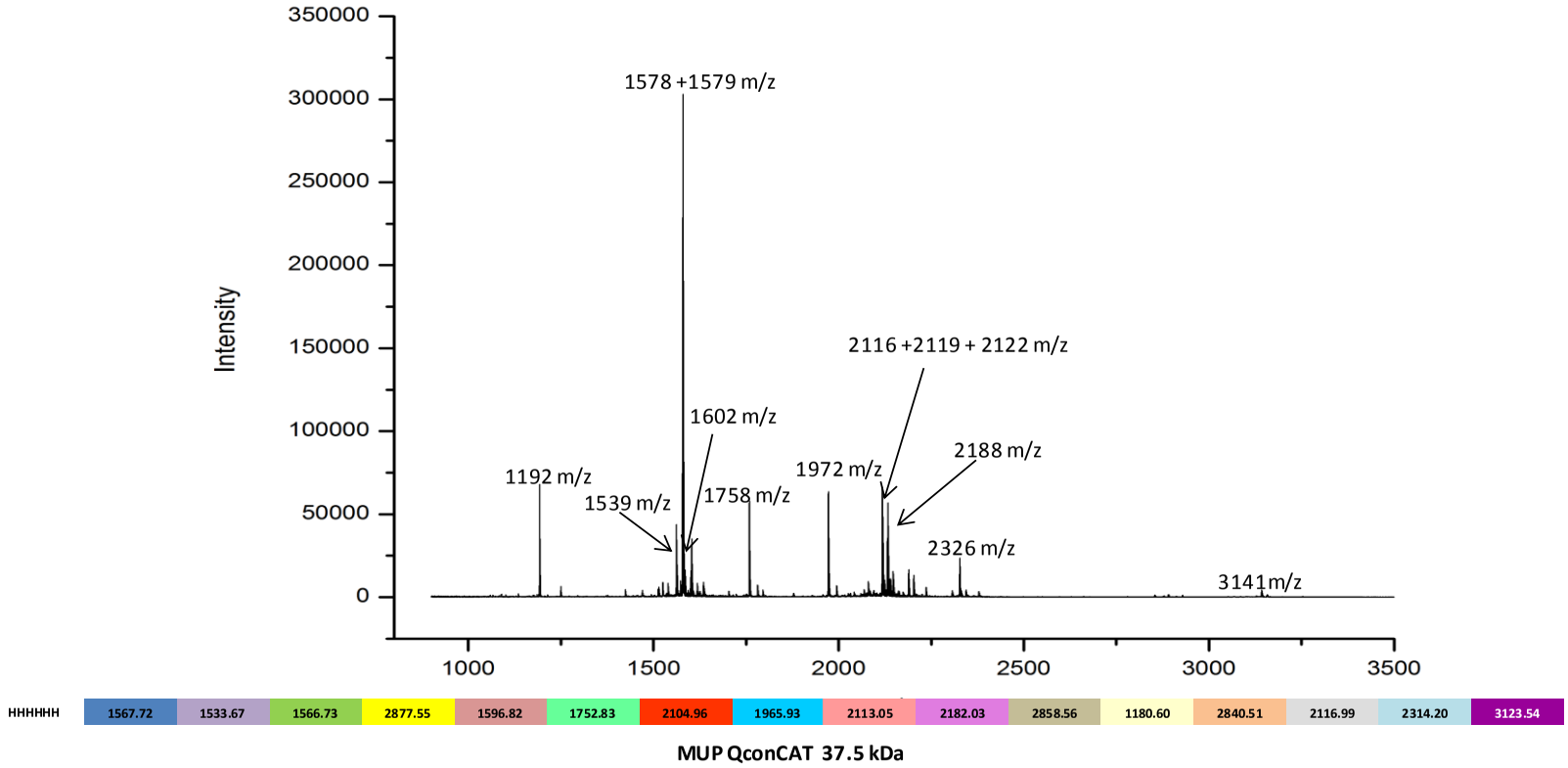
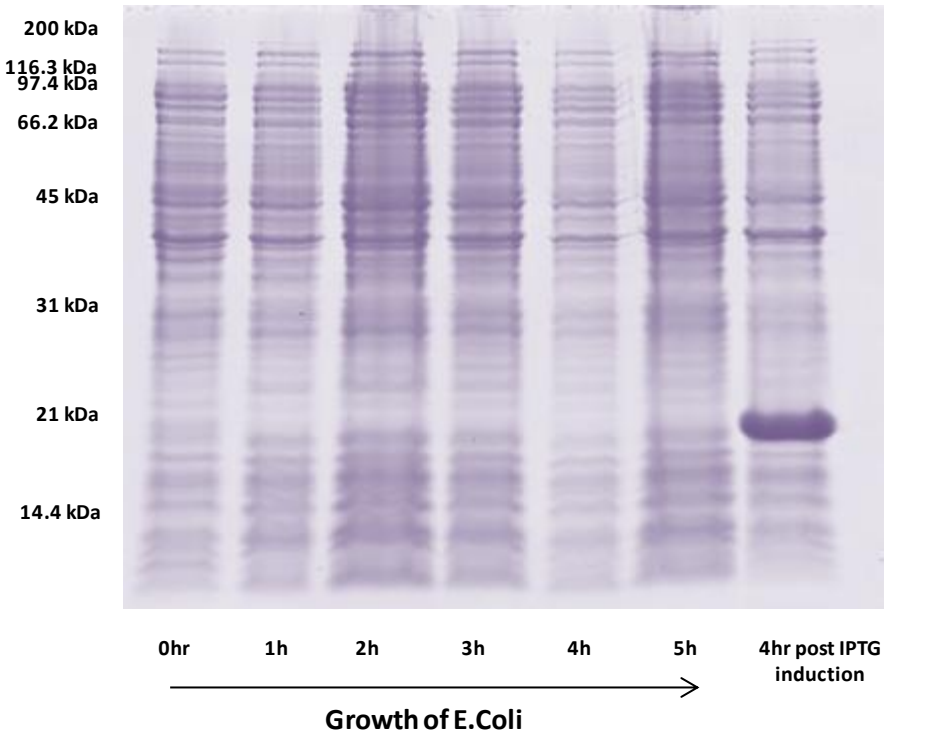


Fig.3.20 MALDI-TOF analysis of purified doubly labelled QconCAT. A small piece of gel was removed from the SDS-PAGE analysis of the purified QconCAT and reduced and alkylated with DTT and iodoacetamide before being incubated overnight at 37 °C with LysC. LysC was chosen over trypsin as some masses of the tryptic fragments were too small to identify by MALDI-TOF. The peptides were recovered the following day and mixed 1:1 with α-Cyano-4-hydroxycinnamic acid dissolved in 50% ACN, 0.1% TFA. The mixture (1 µl) was spotted onto a MALDI target plate and left to dry at room temperature before being analysed by MALDI-TOF. The coloured boxes underneath the peptide mass fingerprint represent LysC fragments of the QconCAT, 100% coverage of the protein was achieved.



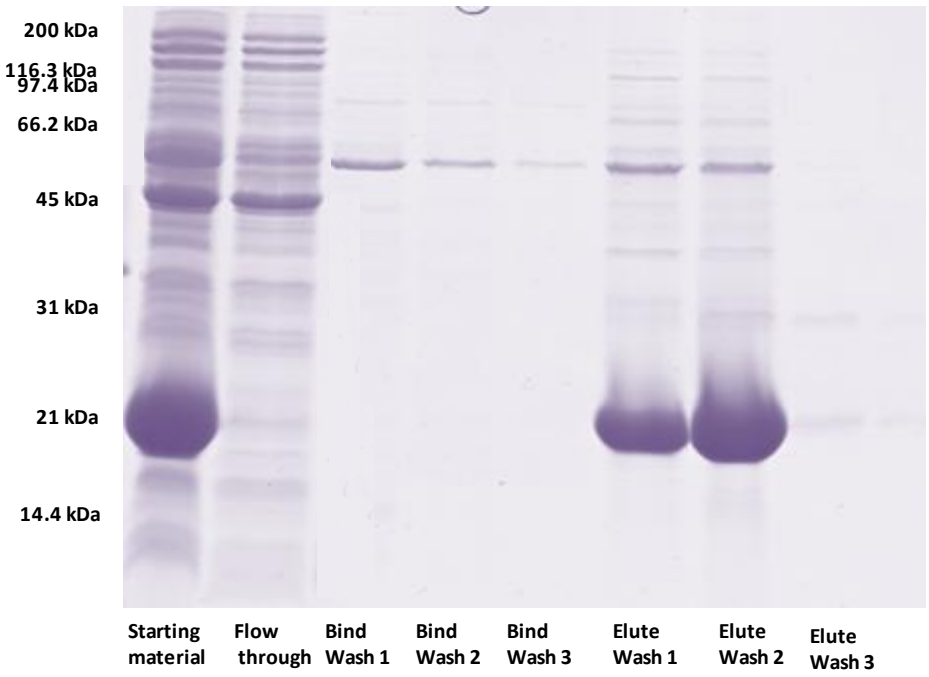


Fig.3.21 Expression and purification of a doubly labelled darcin standard.
Top gel. The labelled darcin was expressed in *E. coli* and labelled with $^{13}C_6$ Lysine and $^{13}C_6$ arginine . The OD (600nm) of *E. coli* was taken every hour until it reached an absorbance reading of 0.6. IPTG was then added to the culture to induce *E. coli* to synthesise the protein. Bottom gel. Unlike the QconCAT which forms inclusion bodies, labelled darcin is found in the soluble fraction therefore GnCl is not required. The soluble fraction was filtered and passed through a 1 ml HisTrap column. The purified protein was eluted in NaCl (2 M), sodium phosphate (80 mM, pH 7.4) and imidazole (2 M). Elute wash 1 and 2 were combined and dialysed overnight in NH_4CO_3 (100 mM , pH 8.5). Samples were run on a 15% SDS gel and stained with coomassie blue.

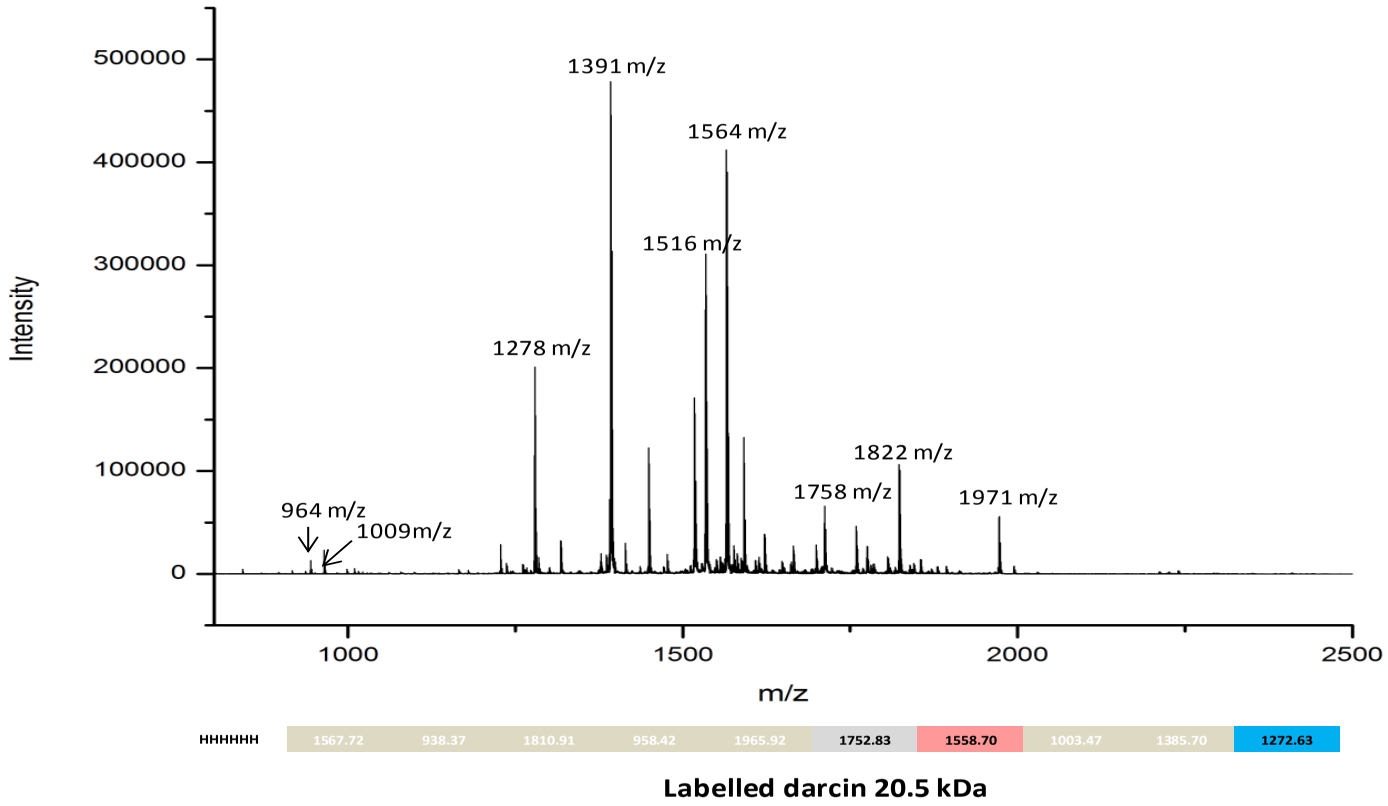


Fig.3.22 MALDI-TOF analysis of purified labelled darcin.

A small piece of gel was removed from the SDS-PAGE analysis of the purified labelled darcin and reduced and alkylated with DTT and iodoacetamide before being incubated overnight at 37 °C with trypsin. The peptides were recovered the following day and mixed 1:1 with α-Cyano-4-hydroxycinnamic acid dissolved in 50% ACN, 0.1% TFA. The mixture (1 µl) was spotted onto a MALDI target plate and left to dry at room temperature before being analysed by MALDI-TOF. The coloured boxes underneath the peptide mass fingerprint represent tryptic fragments of the labelled darcin. The beige boxes with white text represent peptides not used in quantification. The three peptides that are to be used are coloured coded as per figure 3.12, 100% coverage of the protein was achieved.

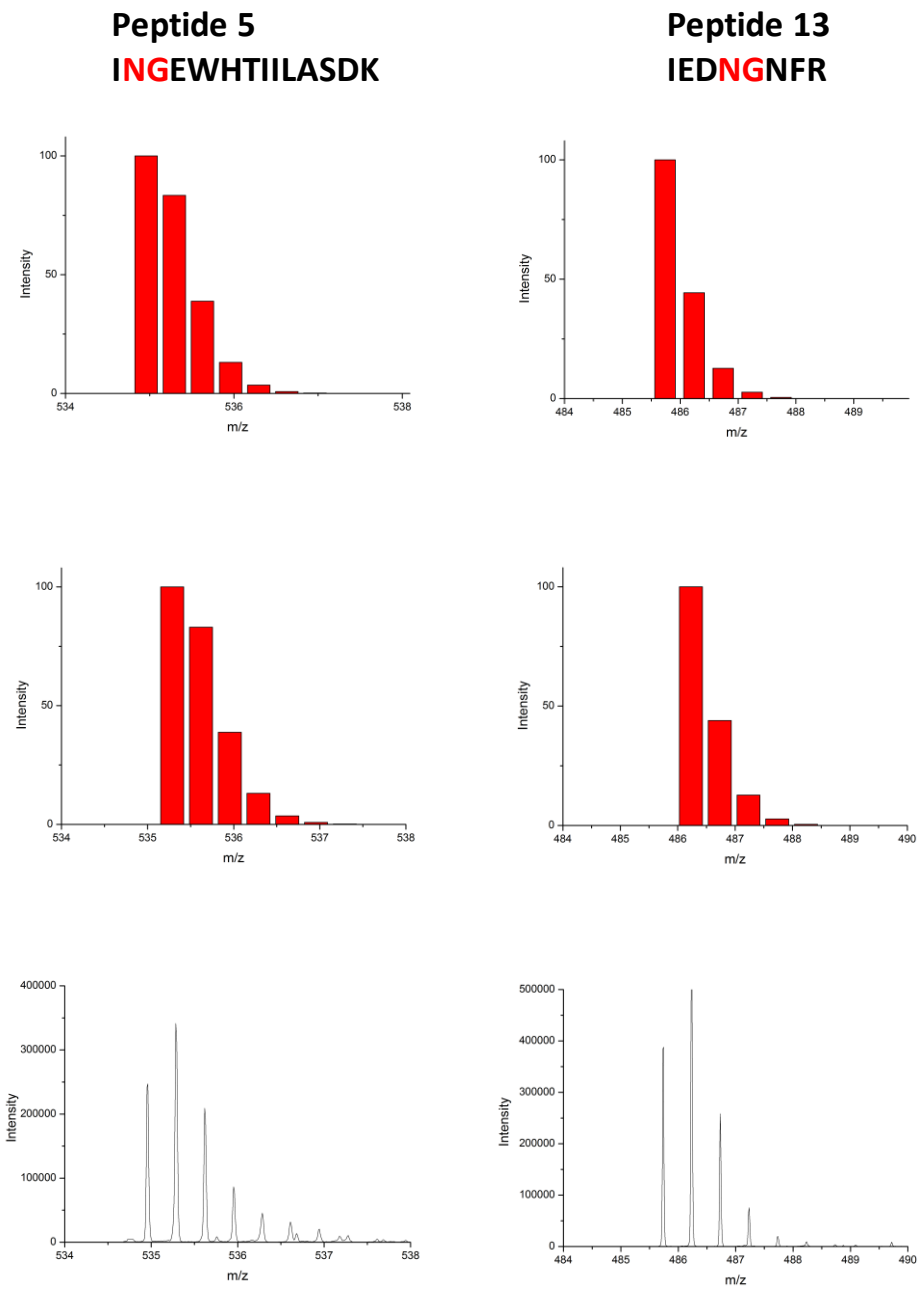


Fig.3.23 Isotopic profiles of peptides 5 and 13.
Top graphs. Predicted isotopic patterns for peptides 5 and 13 was done using MS isotope (Protein prospector tool, University of California). Middle graphs. Predicted isotopic patterns for deamidation at N-G for peptides 5 and 13 was also done using the MS isotope tool. Bottom graphs. Actual isotopic patterns obtained from LC-MS analysis showing a mixture of both deamidated and non-deamidated forms.

The QconCAT was stored in 100 mM ammonium bicarbonate at pH 8.5 to prevent precipitation of the protein and allow long term storage at 4 °C. A deamidation reaction is more likely to occur at a higher pH and an increase in temperature. This more alkaline pH coupled with the first stage of the in-solution digest protocol – heating the sample at 80°C – was most likely responsible for the deamidation of peptide 5 and 13. A fresh pellet of QconCAT was purified and dialysed into 50mM ammonium bicarbonate at pH 7. Aliquots (5 µg/ml) were then taken and diluted into 50mM ammonium bicarbonate all at different pH values – pH 7, 8, 9 and 10. These four aliquots were then digested using the in solution protocol stated in the methods. They were then analysed by LC-MS and the ion of interest – peptide 5 – was extracted. The isotopic pattern of Q peptide 5 digested at pH 7 suggested very little if any deamidation had occurred. The sample at pH 8 showed some degree of deamidation. Extensive deamidation was observed in the samples at pH 9 and 10 (Figure 3.24). To assess whether the ionisation was affected by the deamidation reaction the sum of the intensities across the isotopic envelope were compared for each pH (Figure 3.25). The sum of the intensities was plotted against degree of deamidation and remained constant across the pH range. This suggests that the conversion of amide to acid during the deamidation reaction does not affect ionisation. Therefore adding up the intensities of both forms should be acceptable when it comes to quantification.

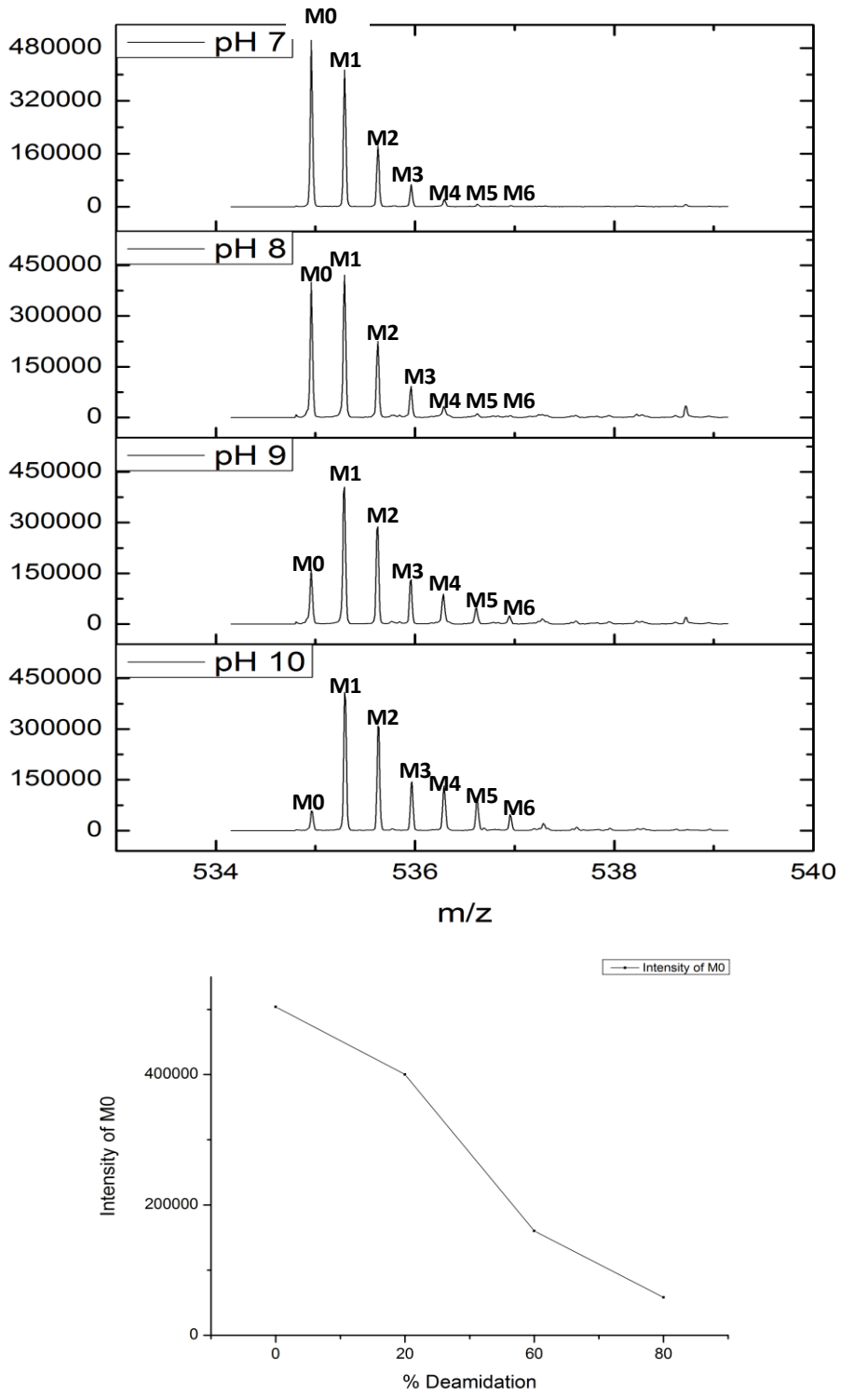


Fig.3.24 The conversion of amide to acid with increasing pH.
Top graphs. Traces from LC-MS analysis. The intensity of the monoisotopic mass (M0) decreases as the pH increases as a result of the conversion of asparagine to aspartic acid. Bottom graph. The intensity of M0 plotted against percentage of deamidation. The percentage of deamidation at each pH was calculated using the MS solver programme in Microsoft excel.

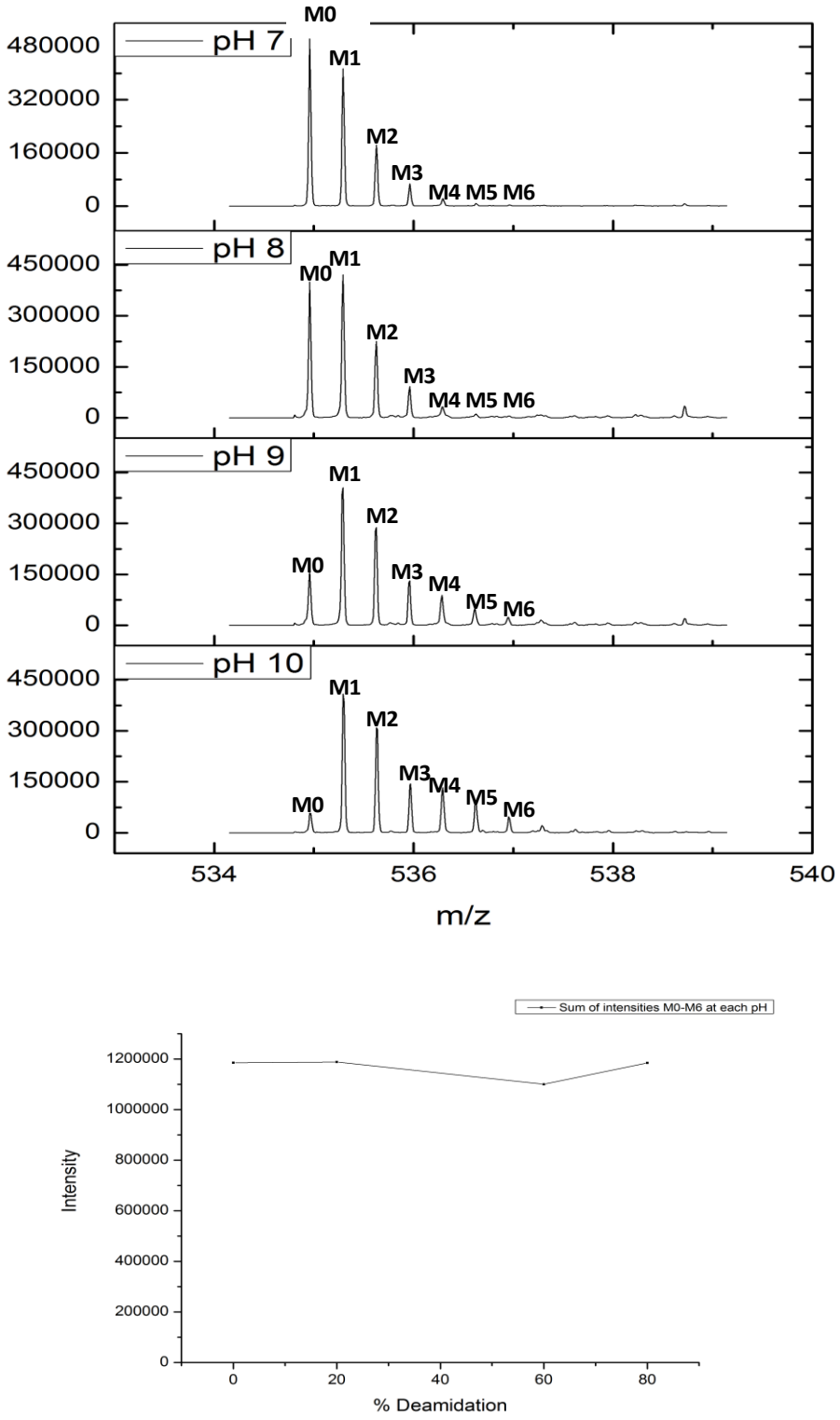


Fig.3.25 The effect of deamidation on ionisation.
Top graphs. Traces from LC-MS analysis. The intensity of the monoisotopic mass (M0) decreases as the pH increases as a result of the conversion of asparagine to aspartic acid. Bottom graph. The sum of intensities (M0-M6) was added up for each sample and plotted against the percentage of deamidation previously worked out using MS solver in Microsoft excel.

Five B6 male and five B6 female urine samples were individually co-digested with QconCAT in a 1:1 protein ratio. Urine samples from the same five males and five females were also co-digested with heavy darcin in a 1:1 protein ratio. All samples were analysed by LC-MS and heavy: light peptide pairs extracted for quantification (supplementary data A). The amount of MUP 20 was quantified using peptide 8 followed by quantification of MUP 3 by subtracting peptide 14 away from 8. Using the QconCAT LysC digest amounts of MUPS 13 and 17 were calculated first followed by MUP 7. Peptide 6 was used to calculate MUP 13 which was then subtracted from peptide 4 to calculate MUP 17. MUP 7 was then calculated by subtracting the amounts of MUP 13 and 17 away from peptide 7 which all three share. MUPS 2, 10, 14 and 5 were then quantified using their unique fragments produced in the tryptic digest. MUPS 1 and 12 were calculated by subtracting the amount of MUP 7 away from peptide 12. The last MUPS to be quantified were 9, 11, 16, 18 and 19 which relied on using either peptide 1, 5 or 13 and subtracting away the amounts of the other MUPS that share those peptides (Figure 3.17). Using peptides 5 and 13 for quantification resulted in negative numbers even when adding up both the amide and acid forms produced during the deamidation reaction. This could be due to the deamidation reaction stopping at the intermediate stage for both analyte and QconCAT but at different rates leading to an inaccurate ratio between the two for quantification. If this was the case then a signal for this intermediate would be detected at -17Da. No intermediate was observed for either peptide 5 or 13. Another reason for this may be down to the reproducibility of digestion. Even though a digestion method was optimised for MUPS (section 3.3.2) this may not be reproducible each time. Both peptides 5 and 3 are next to each other in the analyte but not the QconCAT (Figure 3.26). In the analyte the cleavage site and the surrounding residues are as follows D- K-R-E-K. This is the most challenging part of the MUP sequence to digest. As mentioned previously, the two acidic residues will make it difficult for trypsin to cleave successfully. There was also the added complication of two basic residues being adjacent to each other. There may well be partial missed cleavages around this site which would result in a lower signal for the

MUP QconCAT

EEASSTGRNF NVEKEEASSE GQNLNVEKEE ASSTGRNFNV QKIEEHGNFR LFLEQIHVLE
NSLVLKINGE WHTIILASDK YSVTYDGFNT FTIPKVHTVR DEECSELSMV ADKAGEYSVT
YDGSNTFTIL KAGIYYLNYD GFNTFTILKF HTVIDEECTE IYLVADKIED NGNFRLFLEQ
IRVLENSLVL KERFAQLCEK IEDNGNFRLF LEQIHVLENS LVLKAGIYYM NYDGFNTFSI
LKIEDNGNFR LFLEQIHVLE KLCEEHGILR ENIILSNANR CLQARE

18694

EEASSTGRNF NVEKINGEWH TIILASDKRE KIEDNGNFRL FLEQIHVLEN SLVLKFHTVR
DEECSELSMV ADKTEKAGEY SVTYDGFNTF TIPKTDYDNF LMAHLINEKD GETFQLMGLY
GREPDLSSDI KERFAQLCEE HGILRENIID LSNANRCLQA RE

Fig.3.26 Sequence comparison between MUP QconCAT and 18694.
The sequences for peptides 5 (pink) and 13 (orange) are not next to each other in the QconCAT but are in the native protein. In the native protein the sequence for peptide 5 ends in D-K followed by R-E-K (highlighted in red) which may make it more difficult for trypsin to cleave causing missed cleavages and inaccurate quantification.

analyte and the negative numbers calculated for quantification. They may not be easily visualised on a 1D gel, something that is routinely done post digest before the addition of TFA. No partial missed cleavages were found when searching manually or using software (PLGS) but some of the partial missed cleavages would be quite large and may not ionise or chromatograph well making detection difficult.

Using peptide 1 produced more reasonable data (Figure 3.27). The major isoforms expressed in males are MUP 7, MUP 10, MUP 20, MUP 1, 2 and 12 and MUPS 9, 11, 16, 18 and 19. There were also low levels of minor isoforms detected – MUP 13, 17, 3 5 and 14 some of which correspond to the ESI-MS analysis on fractionated urine (Dr S Armstrong, thesis). As the mice are genetically identical, less variation between individuals would have been expected. Males are housed individually to prevent fighting and the environment that all five males were kept in at the time of sample collection was the same. It is possible that the degree of digestion of the native protein varied between each sample but this is unlikely as the QconCAT data is in agreement with SDS-PAGE, protein assay and ESI-MS (see pages 96-102). There was a slight difference in age between the five males sampled and fully matured mice have an increased MUP expression compared to juveniles (unpublished data). Also, even though the mice are caged separately, their cages are placed next to each other and therefore they will be aware of the presence of other males due to their highly efficient olfactory system. This may influence their MUP expression particularly if a dominant male is caged nearby. The females showed slightly less variation between individuals although two females did express larger quantities of MUP 10 than the other three. Females were also examined to make sure they not in the estrus stage of their estrous cycle, however this is just a visual check and the stage of the cycle is estimated, it could be that 3 of the females were in not in estrus but the other two females were just entering the estrus stage so their protein expression started to increase (see section 3.3.5). Like the male B6 mice, females express MUP 2, MUPS 9, 11, 16, 18 and 19and MUP 10. There was no evidence of MUP 20or MUP 7expression in females which agrees with published data (Armstrong *et al.,* 2005; Roberts *et al.,* 2010). A comparison of total MUP abundance between males and females (Figure 3.28) confirms that males express a

higher concentration of MUPs compared to females (Cheetham *et al.,* 2009). A summary of which Q peptides were successfully used to quantify MUPs are outlined in table 1.1.

Table 3.1 A summary of peptides used for the quantification of MUPs in male and female B6 lab mice. Peptides 10 and 16 were not included as these peptides are used to quantify MUPs present in a another strain of lab mouse BALB/C.

Peptide	Sequence	Used in Quantification?
1	NFNVEK	Yes – used to quantify MUPs 9, 11, 16, 18 and 19
2	EEAASSEGQNLNVEK	Yes – used to quantify MUP 5
3	NFNVQK	Yes – used to quantify MUP 2
4	IEEHGNFR	Yes – used to quantify MUP 17
5	INGEWHTIILASDK	No – possible incomplete digestion of native protein (Figure 3.26)
6	YSVTYDGFNTFTIPK	Yes – used to quantify MUP 13
7	VHTVRDEECSELSMVADK	Yes - used to quantify MUP 7
8	ENIIDLTNANR	Yes – used to quantify MUP 20
9	DGETFQLMELYGR	No – does not ionise well in mass spectrometer
11	LFLEQIR	Yes – used to quantify MUP 14
12	FAQLCEK	Yes – used to quantify MUPs 1 and 12
13	IEDNGNFR	No – possible incomplete digestion of native protein (Figure 3.26)
14	IEEHGSMR	Yes – used to quantify MUP 3
15	LFLEQIHVLEK	Yes – used to quantify MUP 10

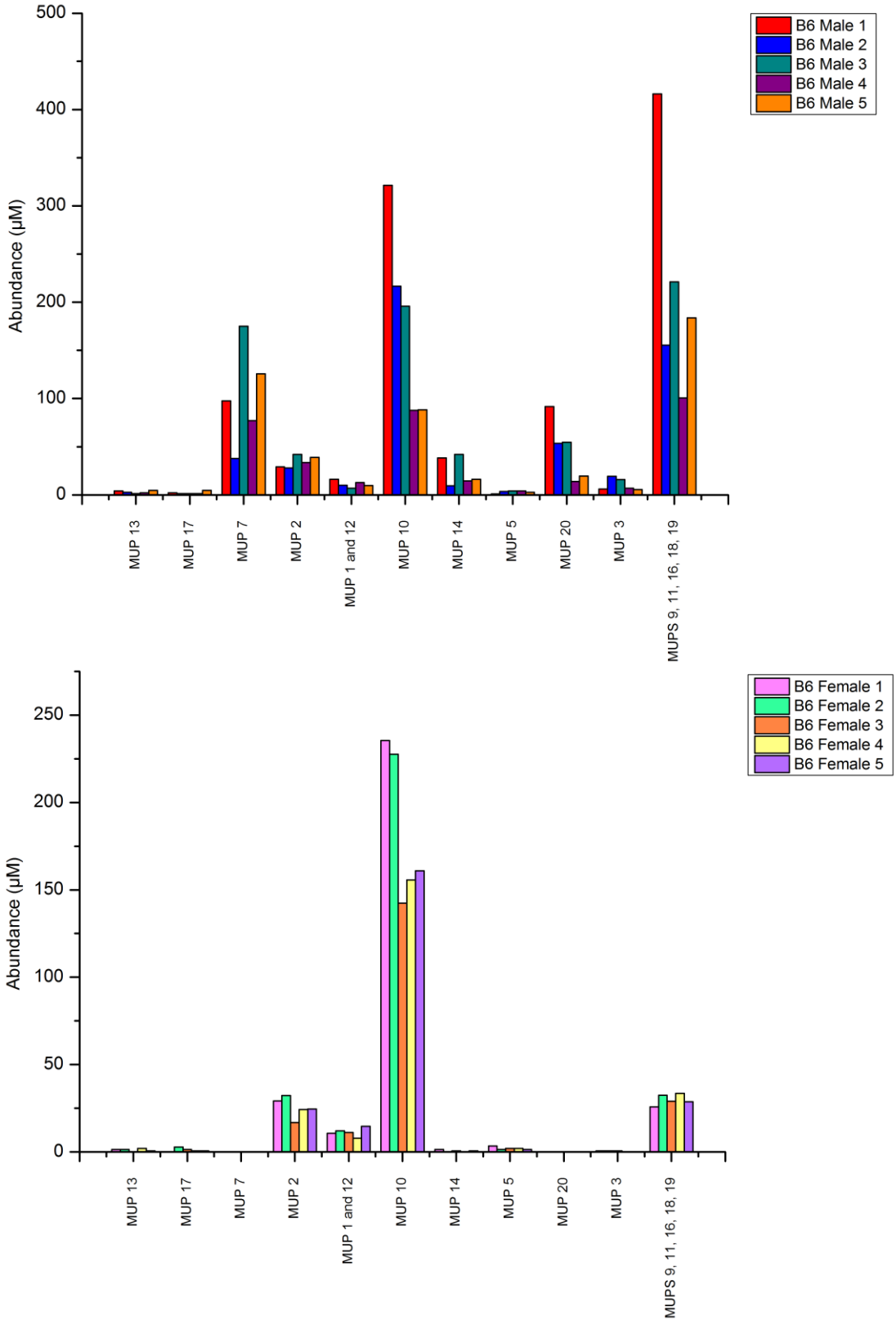


Fig.3.27 Quantification of individual MUP variants expressed in male and female C57BL/6 mice.
Individual MUP isoforms were quantified in five B6 male (top graph) and five B6 female (bottom graph) mice using the MUP QconCAT.

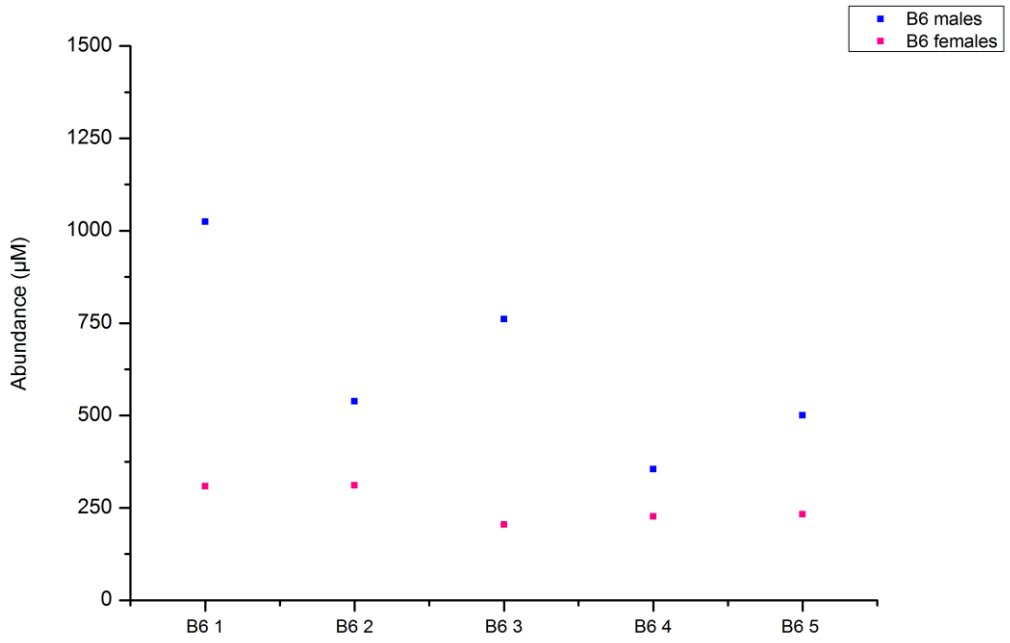


Fig.3.28 A comparison between C57BL/6 male and female MUP expression.
Individual MUP isoforms were quantified in five B6 male and five B6 female mice using the MUP QconCAT. The amount of MUP was summed up for each mouse and plotted on the same scale to observe differences between sexes.

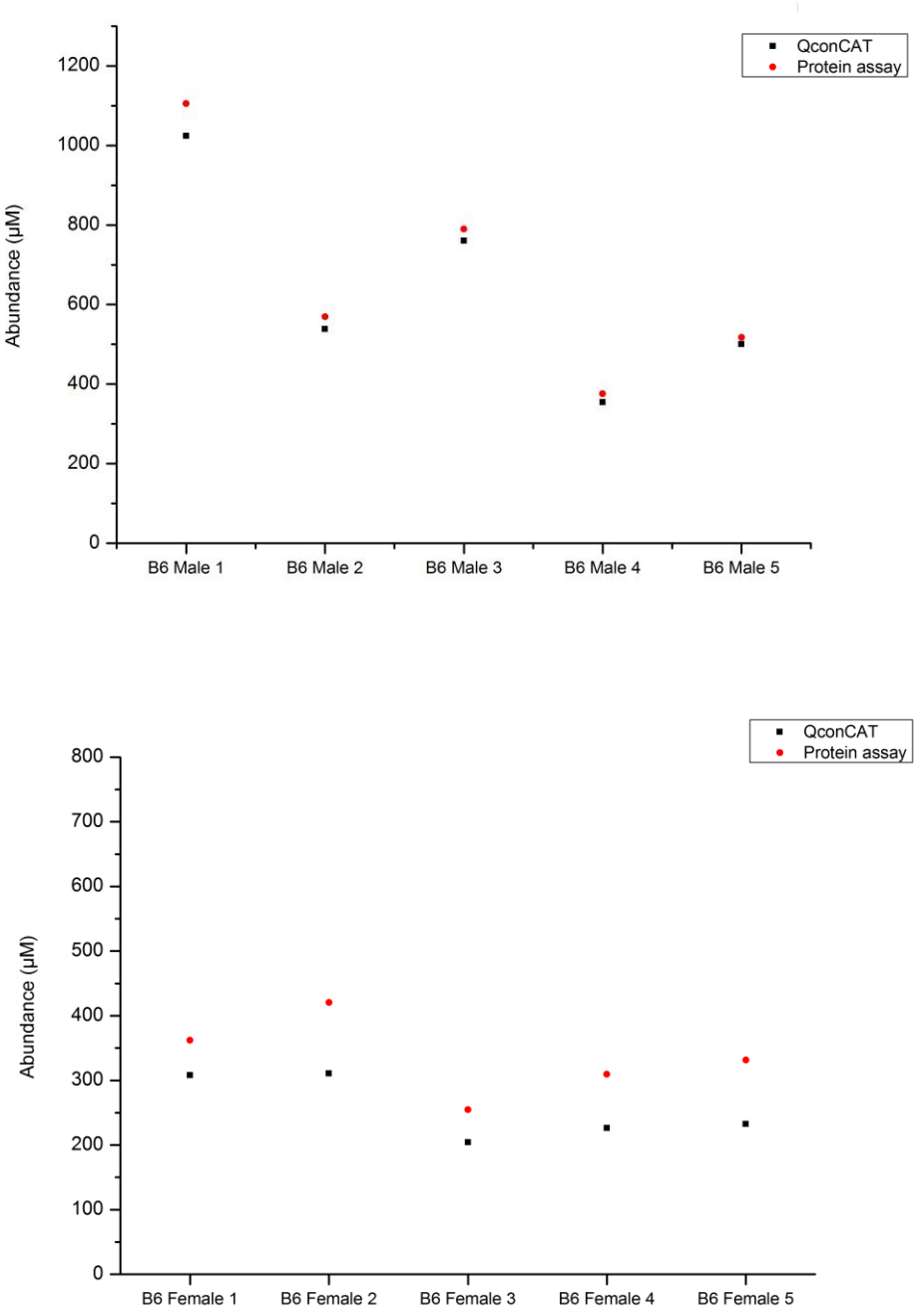


Fig.3.29 Comparison of QconCAT quantification and protein assay.
The amount of each MUP calculated using the QconCAT in each individual mouse was summed to get a "total MUP" amount. The protein content (99% MUP) of each mouse urine sample was established by a Coomassie dye binding assay.

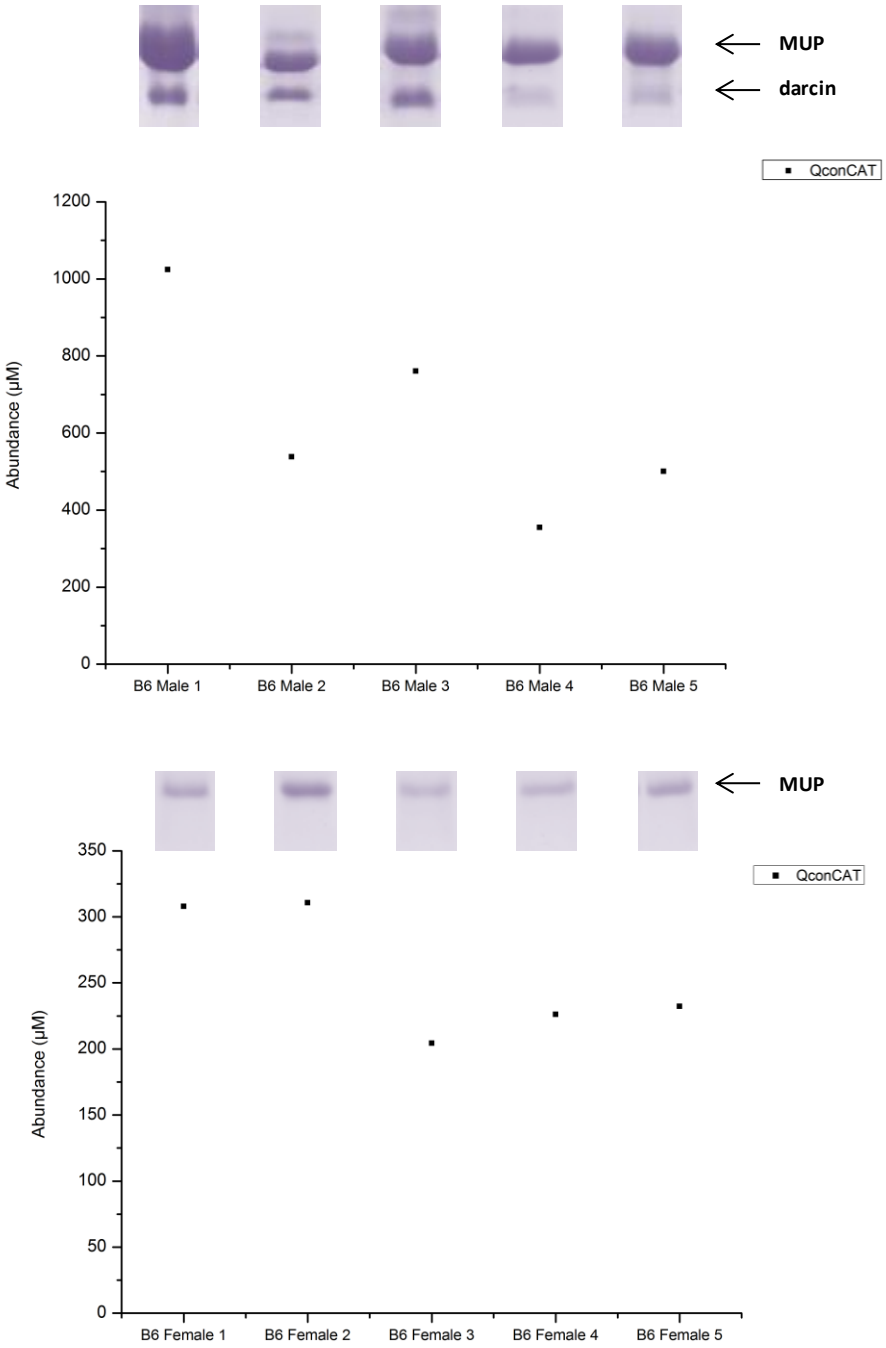


Fig.3.30 Comparison of QconCAT total MUP quantification and SDS-PAGE.
Urine from each of the five males and females was mixed 1:1 with sample buffer and analysed by SDS-PAGE. Samples were run on a 15% SDS gel and stained with coomassie blue. Total amount of MUP was calculated for each mouse using the QconCAT data (graphs).

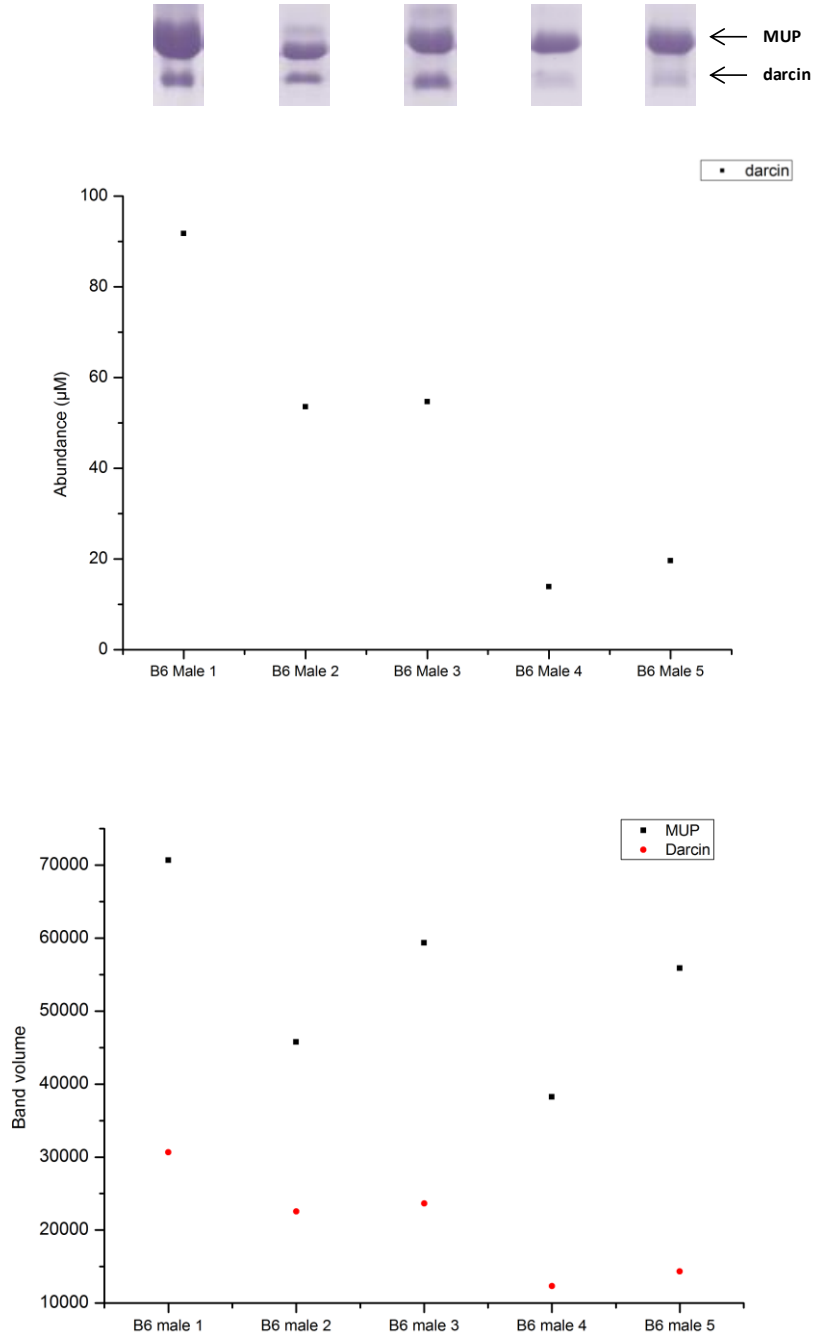


Fig.3.31 Comparison of QconCAT darcin quantification and SDS-PAGE with densitometry analysis.
Top graph: Urine from each of the five males was mixed 1:1 with sample buffer and analysed by SDS-PAGE. Samples were run on a 15% SDS gel and stained with coomassie blue.
Bottom graph: Densitometry on SDS-PAGE analysis was performed using Total Lab™ software and the relative volumes of the main MUP band and the fast migrating MUP 20 band were measured. No albumin was present on the gel so no densitometry analysis was possible. Rolling ball background subtraction (radius 300) was performed on the gel image. Band detection had a minimum slope of 0 , noise reduction was at 5% maximum peak and band edge detection was automatic.

The QconCAT quantification was compared with data obtained from a protein assay, SDS-PAGE and intact mass analysis. The protein assay measures total protein in the urine so the amounts for each individual MUP was added up to produce a total and then amounts compared to the protein assay. The results correlated well for both sexes with a slightly increased concentration calculated when using the protein assay (Figure 3.29). This was most likely caused by the presence of low level amounts of other protein e.g. albumin. As the protein content of mouse urine is 99% MUP, SDS-PAGE analysis can also provide an estimate on how much MUP protein is present in a urine sample (Figure 3.30). Again this correlated well with the QconCAT data, particularly darcin which according to the QconCAT data, B6M 4 and 5 expressed much lower levels than B6M1-3 (Figure 3.31). Densitometry analysis also confirmed this (Figure 3.31).

As MUP isoforms are so similar in sequence and structure it is likely that the peak intensities acquired from intact mass analysis (Figures 3.32 and 3.33) do reflect the relative amounts of MUP isoforms present (Dr S Armstrong, thesis). The QconCAT and intact mass compare adequately illustrating that the B6 males express larger quantities of MUP 10, MUPs 1, 2 and 12, MUPs 9, 11, 16, 18 and 19, MUP 7and MUP 20. Decreased levels of darcin are also observed in the intact mass spectrum of B6M4 and 5. The intact mass data also confirmed the variation in MUP isoform expression observed between individual males. Due to the resolution of the mass spectrometer, it was difficult to identify MUPs 1, 2 and 12 on the intact mass spectra as only 1 Da separates this mass from MUPs 9, 11, 16, 18 and 19. It is most likely that a peak for MUP 1, 2 and 12 is underneath the peak observed for MUPs 9, 11, 16, 18 and 19. Also due to the high abundance of MUP 10 and the resolution of the mass spectrometer, it was difficult to identify MUP 14 particularly in B6M 1 and 3 which have levels similar to MUPs 1, 2 and 12. The intact mass spectra acquired for each individual female also reflected the amounts quantified for each of the major isoforms using the QconCAT method. The most abundant isoform was confirmed to MUP 10 and again no

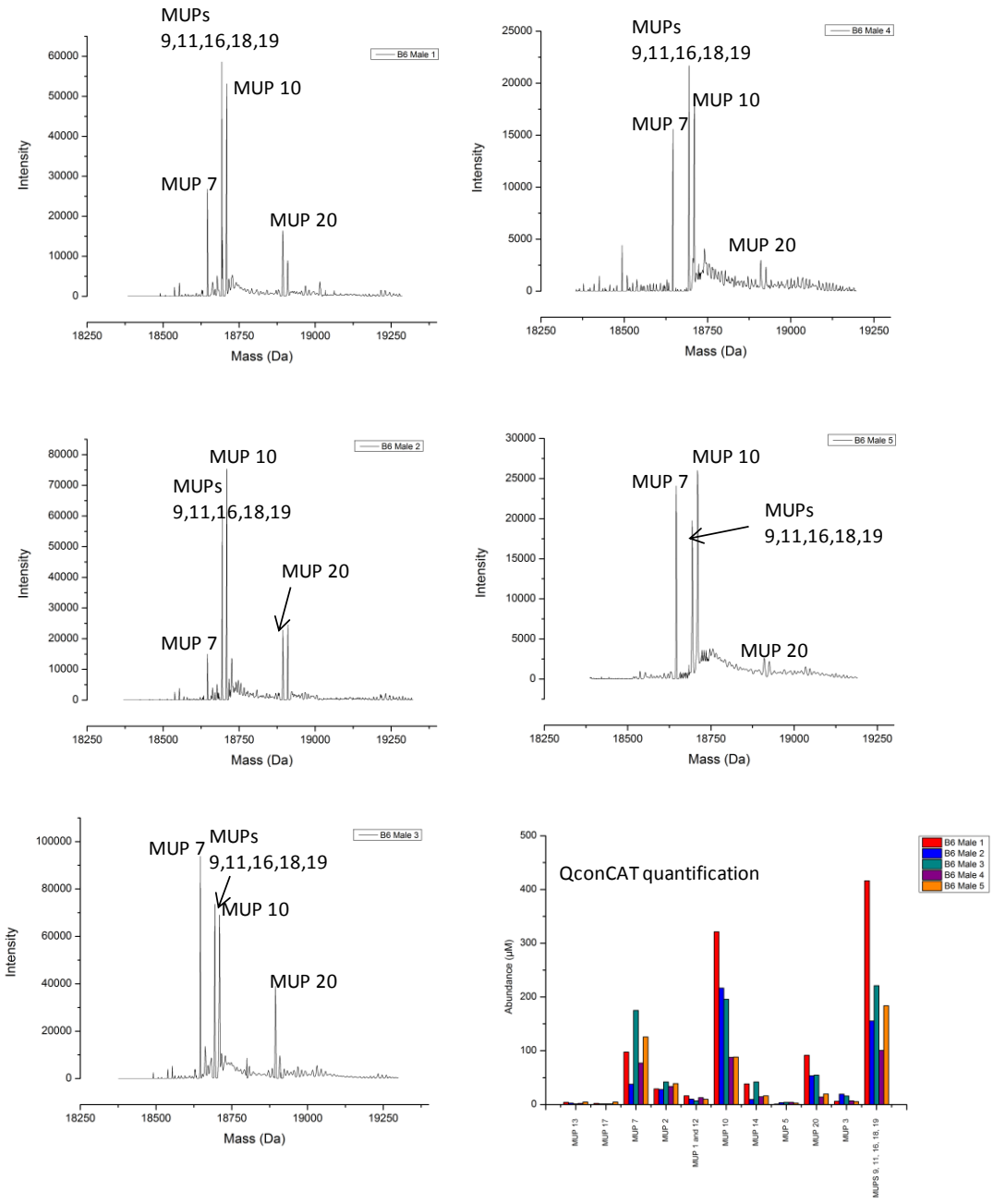


Fig.3.32 Comparison of C57BL/6 male QconCAT quantification and ESI-MS analysis. Urine from the five B6 males was diluted into formic acid (0.1 %) to produce a final concentration of 5 pmol/µl. The samples were then injected onto a C4 desalting trap and masses of MUPS present were determined by ESI-MS. The envelope of multiply charged protein ions was deconvoluted using MAXENT 1 software to yield the true mass composition of the sample (MaxEnt 1, MassLynx 4.1, Waters). The peak directly after MUP 20 is an oxidation of a methionine residue resulting in a mass increase of 16 Da.

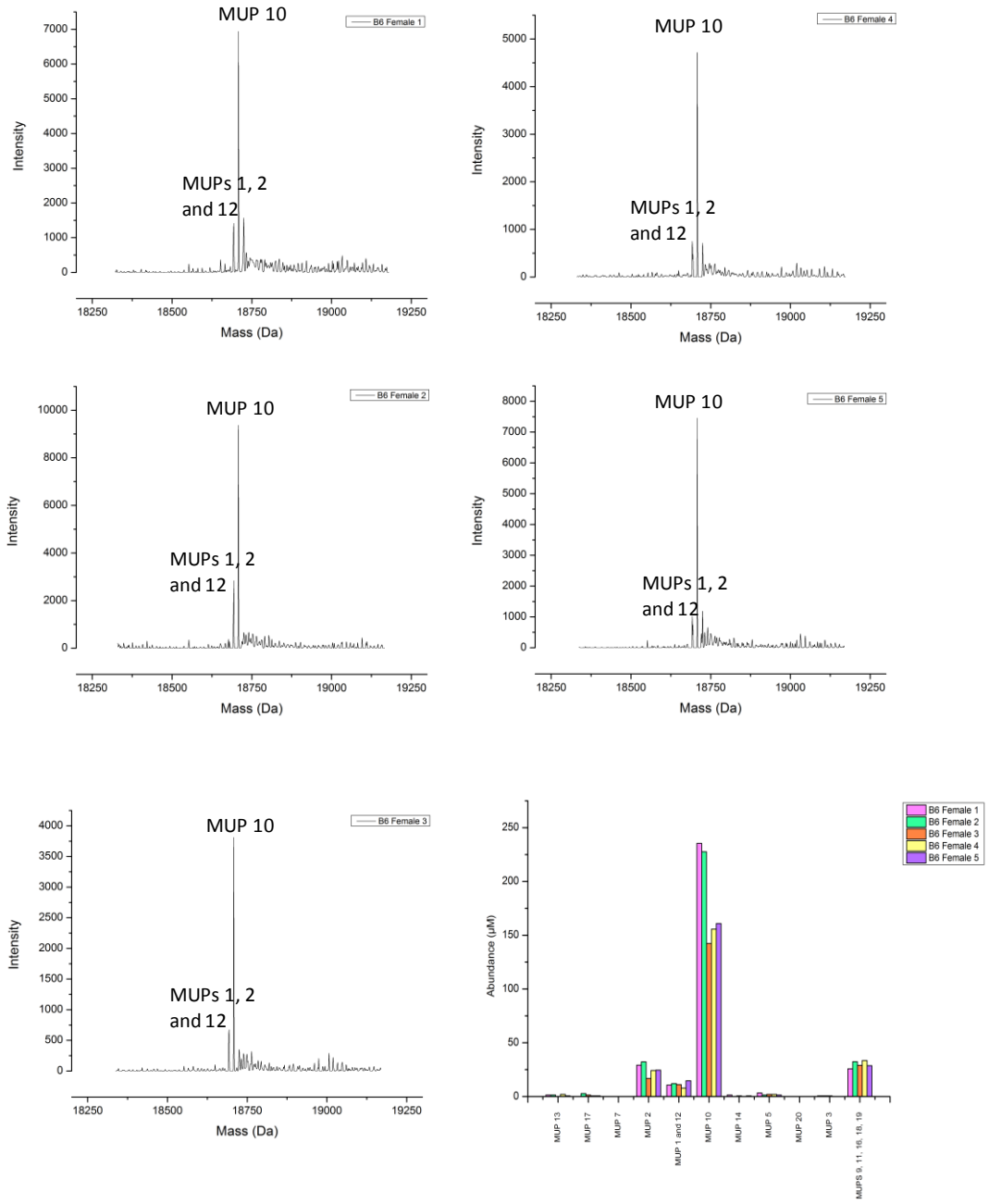


Fig.3.33 Comparison of C57BL/6 female QconCAT quantification and ESI-MS analysis.

Urine from the five B6 females was diluted into formic acid (0.1 %) to produce a final concentration of 5 pmol/µl. The samples were then injected onto a C4 desalting trap and masses of MUPS present were determined by ESI-MS. The envelope of multiply charged protein ions was deconvoluted using MAXENT 1 software to yield the true mass composition of the sample (MaxEnt 1, MassLynx 4.1, Waters).

evidence of expression of the male specific isoforms MUP 7and MUP 20. Again it was difficult to distinguish between MUPs 1, 2 and 12 and MUPs 9, 11, 16, 18 and 19. After the mass spectrum was deconvoluted, the software identified the mass to be MUPs 1, 2 and 12 in the females, probably because these MUPs were more abundant than MUPs 9, 11, 16, 18 and 19 which were expressed in much higher levels in the males hence the peak in male intact mass data being identified as MUPs 9, 11, 16, 18 and 19. A summary of MUP variants quantified in both male and female B6 mice is outlined in table 3.2

Table 3.2 A summary of MUP proteins quantified in male and female B6 lab mice

	MUPs present in B6 males	MUPs present in B6 females
Major isoforms	MUP 7	MUP 2
	MUP 2	MUP 1 and 12
	MUP 1 and 12	MUPs 9,11,16,18,19
	MUP 20	MUP 10
	MUPs 9,11,16,18,19	
	MUP 10	
Minor isoforms	MUP 14	MUP 13
	MUP 13	MUP 17
	MUP 17	MUP 5
	MUP 3	
	MUP 5	

3.3.5 Investigating MUP production during the estrous cycle

In addition to behavioural and physiological responses such as inter-male aggression, mate choice and puberty, MUPs are thought to play a role in the oestrous cycle. Stopka *et al.,* 2007 have presented a study in which they found both sexes up regulate their MUP production during social contact and that females advertise their reproductive status by modifying their MUP production during the estrous cycle (Stopka *et al.,* 2007).

The mouse estrous cycle lasts approximately 4-5 days. There are 4 stages in the cycle - proestrus, estrous, metestrus and diestrus. Each stage can be defined on cell types present from a vaginal swab (Marcondes *et al.,* 2002). During the proestrus stage the female is not yet sexually receptive, a swab will show the presence of

mostly epithelial cells and the endometrium will start to grow as the level of oestrogen rises. This is followed by an increase in gonadotropic hormones causing ovulation which leads into the estrus phase. This can be defined by clusters of squamous epithelial cells that are uneven in shape and contain no nucleus. The female is sexually receptive at this stage. The level of oestrogen starts to decrease and the corpus luteum begins to form and as a consequence progesterone levels start to rise. This is the metestrus stage which is characterised by the presence of leucocytes and nucleated epithelial cells. This stage is closely followed by the diestrus phase which is verified by presence of mainly leukocytes (Parkening *et al.,* 1982, Walmer *et al.,* 1992, Caligioni 2009).

The aim of this study was to use the MUP QconCAT to not only look at increases and decreases of MUP expression during estrus but to identify any significant changes of specific MUP isoforms. Six adult females were used in this study and urine samples were collected at each stage. All samples were mixed 1:1 protein ratio and digested using the protocol described in the methods before being analysed by LC-MS.

The amount of each MUP present in each sample was quantified using the same workflow previously used to quantify the five B6 males and five B6 females. As it was sometimes difficult to define the exact stage of the cycle some samples were labelled as being in between phases for example diestrus-proestrus (D-P). Overall there was an increase in MUP production during the estrus (E) phase with the lowest amount of MUP expressed during the proestrus stage (P) (Figure 3.34 and 3.35).

B6 female A (B6FA) had samples taken at the following stages – D-P, estrus-metestrus (E-M), metestrus (M) and diestrus (D). MUP expression peaked in the E-M sample and was at its lowest in the D-P sample. The M and D samples showed similar expression, with slightly more amounts in the M sample. B6 female B (B6FB) had samples taken at D-P, proestrus-estrus (P-E), E-M and metestrus-diestrus (M-D). The P-E sample had most MUP expression followed closely by the E-M sample.

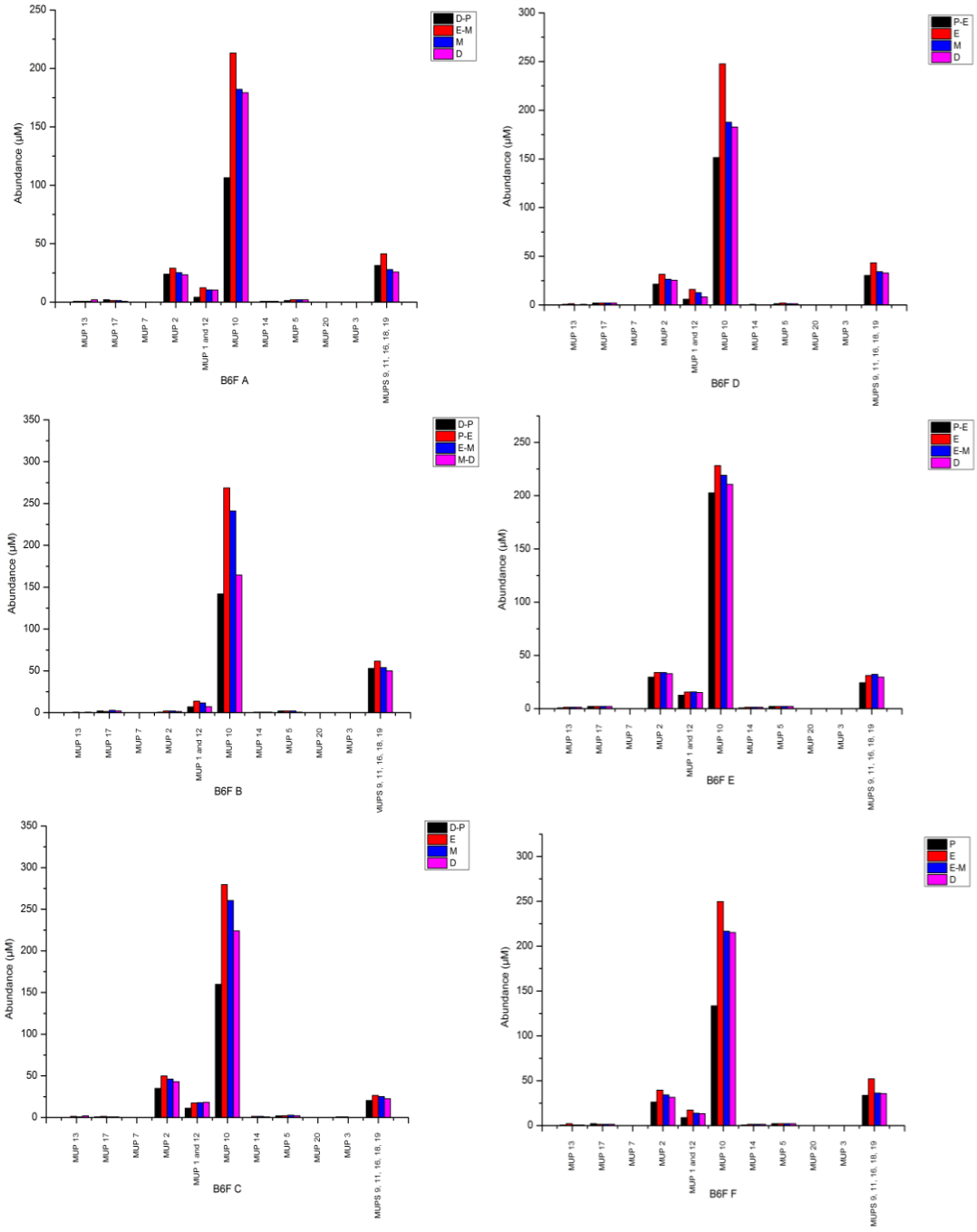


Fig.3.34 Quantification of individual MUP isoforms expressed during the mouse estrous cycle. Six B6 females (labelled A-F) had samples taken at four different stages in their estrous cycle. The samples were mixed 1:1 with QconCAT and digested using the protocol described in section 3.2.3 and analysed by LC-MS. Amounts of individual MUP variants calculated at each stage in the estrous cycle.

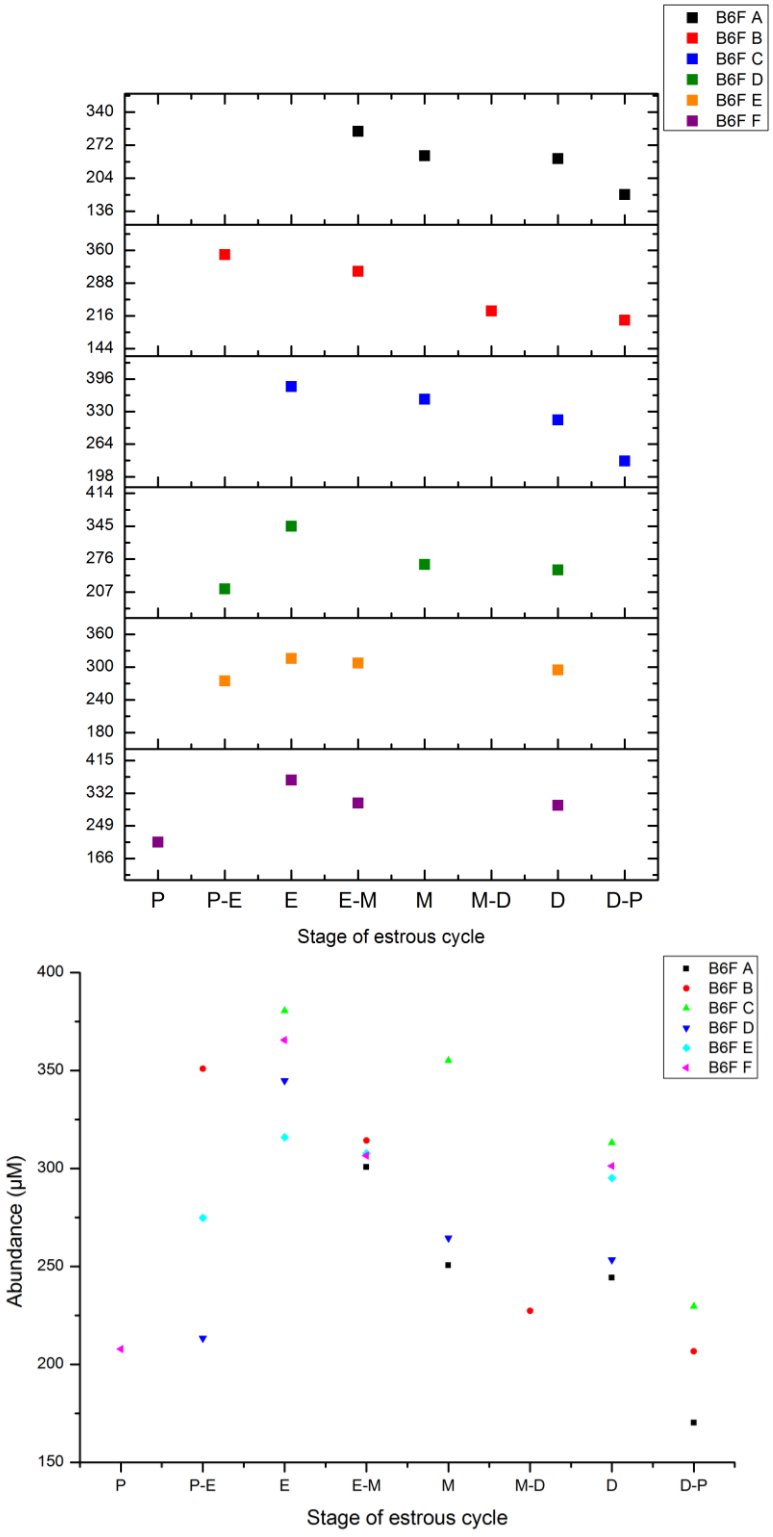


Fig.3.35 Comparison of MUP expression in individual females during the estrous cycle.
Six B6 females (labelled A-F) had samples taken at four different stages in their estrous cycle. The samples were mixed 1:1 with QconCAT and digested using the protocol described in section 3.2.3 and analysed by LC-MS. Total abundance was calculated at each stage of the cycle for each individual mouse and plotted against the stage of the estrous cycle. The bottom graph illustrates the total abundances at each stage of the estrous cycle for all six females.

Similar to B6 female A, the D-P sample had the lowest amount of MUP expression. The third female, B6 female C (B6FC), had samples taken at D-P, E, M and D. Similar to B6 female A and B, MUP expression was at its lowest in the D-P sample. An increase in MUP concentration was observed in the E sample. As seen in the first sample B6FA, the M and D samples were very similar in protein concentration with a slight raise in the M sample. B6 female D (B6FD) had samples collected at P-E, E, M and D. MUP production was at its highest in the E sample which coincides with B6FC. The lowest amount of MUP expression was seen in the P-E sample. B6FB also had a P-E sample taken which had the highest amount of MUP in. It is possible the B6FB mouse was more into the estrus phase in the P-E sample which is why this sample had most protein in. B6FD could have been more into the proestrus phase in their P-E sample hence the low amount of protein expression. The M and D samples again had similar concentrations of MUP proteins. There was less of a difference seen between stages for B6 female E (B6FE). Samples were taken at P-E, E, E-M and D. The sample that contained most MUP protein was the E sample followed closely by the E-M sample. Even though the P-E showed the least amount of MUP expression, the amount calculated was not that much lower than the E and E-M samples. This could be because the mouse was more into the estrus stage when the P-E and E-M samples were taken. Finally B6 female F (B6FF) had samples collected at P, E, E-M and D. As with the other samples, there was an increase in MUP expression in the estrus sample and the lowest amount of protein was expressed in the proestrus sample. E-M and D were once again very similar in quantity with a slightly more protein calculated in the E-M sample.

3.4 Conclusions

Despite the high sequence similarity between MUP variants, a QconCAT strategy was designed and implemented for quantification for these homologous protein isoforms. Many of the rules for QconCAT design were unable to be applied, particularly when it came to finding a unique peptide for each protein to be quantified. This was not possible for all MUP variants and required a logical

subtraction method. Surrogate peptides should be unique to the protein being quantified and easily detectable by LC-MS. The choice of peptides suitable for detection by mass spectrometry is crucial for the sensitivity of the assay (Picotti and Aebersold, 2012). Potential Q peptides should ideally be analysed by LC-MS prior to being chosen for quantification. Alternatively, information may have been collected about these peptides and stored in online suppositories such as Peptide Atlas and Global Proteome Machine Database which contain information on how often the peptide has been observed by mass spectrometry. If the peptide does not ionise efficiently then detection by mass spectrometry unachievable. Also peptides that are too hydrophilic will not bind the stationary phase of the LC column and if they are too hydrophobic then it is unlikely to be eluted. These peptides should also be avoided (Eyers *et al.,* 2011; Picotti and Aebersold, 2012).

The native peptides should also be in a region that undergoes complete proteolysis. Although trypsin is a highly selective and efficient protease, peptides with two neighbouring basic residues and the presence of acidic residues surrounding the cleavage site should be avoided as these situations often lead to missed cleavages. Surrogate peptides should not be subjected to any post translational modifications such as deamidation as these will alter the mass of the peptide and lead to inaccurate quantification data. If using an MRM method then the peptide would not be detected at all due to the mass shift (Lange *et al.,* 2008; Picotti and Aebersold, 2012).

Various mass spectrometer parameters should also be optimised to ensure the best sensitivity is achieved for each peptide. Factors that influence signal intensity are precursor charge state, ion source parameters, and if using an MRM method, fragment ion type and collision energy. Using the most dominant charge state is essential for sensitive detection of the peptide. Charge state detection is influenced by solvents, background/noise and flow rates. Optimising ion source parameters such as voltages, source temperatures and gas flows will influence the detectability of peptides. If quantifying using fragments then fragments with a mass close to that of the precursor ion should generally be avoided as these transitions often have a higher signal to noise ratio. For highest selectivity, fragment ions with an m/z value

above that of the precursor should be considered as singly charged background ions can not result in fragments with higher m/z values than the precursor. The collision energy should be optimised. If it is too low insufficient fragmentation will take place and if set too high there will be losses due to secondary fragmentation events (Lange *et al.,* 2008; Picotti *et al.,* 2010; Picotti and Aebersold, 2012).

Following the rules for Q peptide selection was not possible due to the high sequence similarity between the MUP isoforms. It was not possible to find unique peptides for MUPs 9,11,16,18 and 19 and the original [$^{13}C_6$] Lys labelled QconCAT contained peptides that ionised poorly and were too hydrophobic for the stationary phase in the column. Although a [$^{13}C_6$] Lys and [$^{13}C_6$] Arg labelled QconCAT followed by a tryptic digest eliminated these issues, peptides 5 and 13 could still not be used for quantification due to the incomplete proteolysis of the native peptides (Figure 3.26). These peptides also showed some degree of deamidation.

Deamidation is a PTM that can occur in peptides containing an Asn residue. The reaction is more likely to occur when the Asn residue is followed by a Gly residue. As glycine is a small, flexible amino acid with a low steric hindrance, the peptide group is more vulnerable and open to attack by the Asn side chain (Figure 3.36). The side chain of Asn attacks the peptide group of Gly residue which results in the formation of a succinamide intermediate. A hydrolysis reaction results in the formation of aspartic or isoaspartic acid and peptide mass shift of +1 Da. As discussed in the results section, deamidation reactions are accelerated by increasing the pH and exposure to high temperatures. The QconCAT was stored in ammonium bicarbonate at pH 10 to prevent it coming out of solution. This storage combined with the first section of the digest protocol – heating the samples to 80 °C – resulted in extensive deamidation of peptides 5 and 13 which both contain Asn-Gly. The isotopic patterns of these peptides indicated a degree of deamidation had occurred in both the Q peptides and native protein but at different rates due the storage of the QconcAT solution. Unfortunately the storage issues surrounding the QconCAT were not known

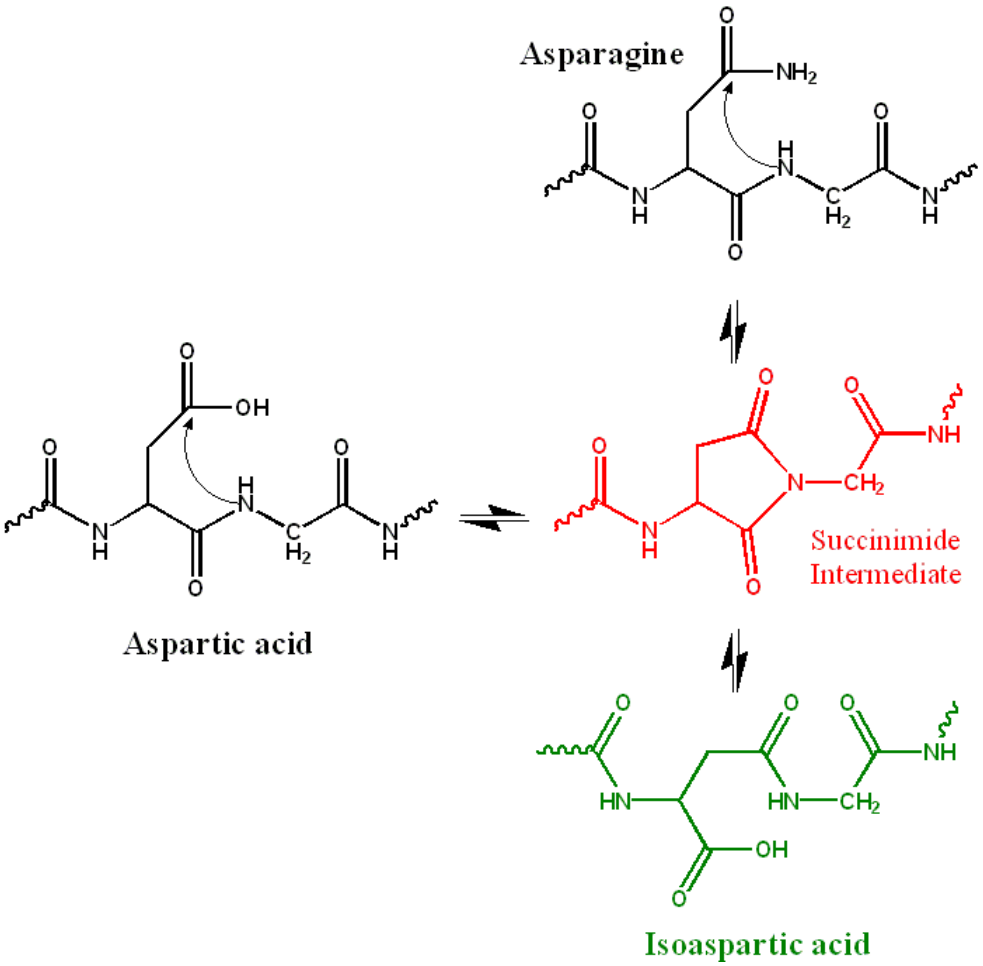


Figure 3.36. Deamidation reaction between asparagine and glycine residues

Deamidation reaction of Asn-Gly (top right) to Asp-Gly (at left) or iso(Asp)-Gly (in green at bottom right). The succinamide intermediate is represented in red. "Deamidation Asn Gly". Licensed under Creative Commons Attribution-Share Alike 3.0 via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Deamidation_Asn_Gly.png#mediaviewer/File:Deamidation_Asn_Gly.png

prior to the selection of these peptides for inclusion in the QconCAT. Also there were limited options for choosing other peptides due to the homologous nature of the MUPs. However, both the acid and amide forms ionised in a similar fashion (section 3.3.4) and therefore adding the addition of the intensities of the monoisotopic masses of both the acid and amide forms was acceptable.

Deamidation was not the primary problem with peptides 5 and 13; the inaccurate quantification was most likely caused by incomplete digestion in the native protein. Alternative approaches such as filter aided sample preparation (FASP) may improve the digestion of MUP proteins. A FASP method combines the advantages of both in gel and in solution digestion processes. A FASP protocol begins with solubilisation of the proteins in detergent such as SDS followed by reduction with DTT and disassociation of detergent micelles and protein detergent complexes with urea (8 M). The DTT, detergent and low molecular weight are removed by ultracentrifugation. Iodoacetamide is then added to the sample to prevent the reformation of disulphide bonds, the iodoacetamide is then removed by ultracentrifugation. Urea is used to wash away any excess detergent. The proteins are then digested with protease and incubated. The sample is then filtrated again and peptides collected with high molecular weight substances retained on the filter (Wisniewski et al., 2009). Another approach would be to incorporate flanking regions in the QconCAT. Flanking regions are sequences from the proteins that surround the tryptic fragments (Q-peptides) to create an identical amino acid composition around the cleavage site. This should lead to the same rate of proteolysis in both the native and QconCAT protein. The presence of the flanking regions eliminates concerns surrounding dibasic cleavage site (seen between peptide 5 and 13 in the native protein) and acidic residues at P2' position (Kito *et al.,* 2007; Nanavati *et al.,* 2008; Chen *et al.,* 2012).

Sample analysis was carried out using a TOF mass analyser and quantification was done on the MS level by extracting the exact masses of the analyte and QconCAT precursor ions. Traditionally TOF instruments were used for qualitative applications rather than quantitative experiments due to the limited dynamic range of these mass analysers. However, newer generation instruments are now equipped with

technology that allows both qualitative and quantitative analysis. The Waters G2 Synapt has QuanTof technology which incorporates novel detector electronics and hardware features that enable modern TOF mass spectrometers to generate spectra at speeds which enable narrow UPLC peaks to be accurately profiled without compromising mass resolution. At the same time QuanTof provides a proportionate response across a wide range of signal intensities, regardless of spectral complexity, so that accurate quantitative results can be obtained even in crude sample extracts. QuanTof technology contains an analogue-to-digital converter (ADC) that records the intensity of detector response over time. This enables the very fast signals produced by the detector to be correctly represented and arrival time and intensity to be calculated accurately. This allows TOF spectra to be recorded with high mass resolution, high mass accuracy and high dynamic range at very fast data acquisition rates. The Waters G2 Synapt can display up to four orders of linear magnitude. Linearity is important for accurate quantification and most quantitative experiments use standard curves to extrapolate values for unknown samples. A standard curve provides a visual representation of the dynamic range of measurement and the limit of linearity – the point at which the relationship between response and concentration are no longer linear which would result in inaccurate quantification data.

The most commonly used mass spectrometry based technique for absolute quantification is MRM methodology on a triple quadrupole mass analyser. Although the newer generation TOF analysers have improved dynamic range, triple quadrupoles have even better with instrument vendors now offering up to six orders of dynamic range. The non scanning mode of operation of triple quadrupole MRM results in increased sensitivity by up to two orders of magnitude compared to full scan techniques used by other mass analysers. It also produces the wider linear dynamic ranges which is useful for the detection of low abundant proteins in highly complex sample mixtures (Lange *et al.,* 2008). The wider dynamic range and improved sensitivity of the MRM technique would be advantageous for MUP quantification because of the large difference in intensities between the major isoforms and minor MUP variants. For instance, a lot of the major isoforms

contribute to the signal of one shared peptide making the response calculated for that peptide close to the saturation limit of the instrument. The improved sensitivity could possibly improve the quantification of the lower abundant MUP variants.

The objectives of this chapter were to develop a method for the absolute quantification of MUPs in male and female B6 lab mice and identify the differences between sexes. Using a QconCAT strategy, MUPs were successfully quantified in five male and five female B6 lab mice. The major MUP isoforms present in males were MUPs 9,11,16,18 and 19, MUPs 1, 2 and 12, MUP 10 and two male specific isoforms MUP 7 and MUP 20. The dominant variants present in females were MUPs 9,11,16,18 and 19, MUPs 1, 2 and 12 and MUP 10. These findings were in agreement with previous research (Armstrong *et al.,* 2005). Using the newly developed method, MUP production in the estrous cycle was also assessed. Although it was difficult to identify the exact stage of the cycle, there was a reoccurring pattern in MUP expression throughout the estrous cycle in the 6 females tested. MUP expression peaked during the estrus stage and declined during the proestrus stage. The biological significance of this is most likely advertisement of reproductive status by the females. The results agree with observations made by Stopka *et al* who identified female mice do vary their MUP production during the estrous cycle with females up-regulating MUP expression at the beginning of estrus (Stopka *et al.,* 2007). Overall the objectives were achieved but there were some limitations. Five animas of each sex were samples and only a single sample was supplied for each. Ideally, analysis of at least 3 urine samples per mouse, taken on different days, would provide a more accurate assessment of the concentration of MUPs present in each animal and better assess the reproducibility of the method. This is particularly important with this sample set as differences in MUP expression was observed between animals of the same sex, which is unusual as they are genetically identical mice. It would also be beneficial to monitor inter and intravariability of the assay by analysing the samples multiple times on the LC-MS system.

Rodents such as mice and rats are generally deemed as 'pests' and their behaviours can often have a detrimental effect to humans in particular. In developing countries, it is estimated that rats and mice are responsible for 25% of infectious disease cases. These diseases are often fatal due to the limited amount of health care available. An ongoing long term project is currently investigating pest control strategies in developing countries by trying to manipulate rodent behaviour. It is anticipated that this work will contribute towards this project. Knowing what MUPs are up and down regulated in social situations will provide a greater insight into behaviours displayed by rodents. This work will also complement studies taking place on animal welfare. Animal welfare research primarily focuses on the welfare of animal in captivity for example laboratory rodents. Aggression is common between laboratory rodents and current research is centred towards what triggers this aggression and is there a certain MUP protein that is responsible for it.

Chapter 4: Protein Secretion in the harvest mouse (*Micromys minutus*)

4.1 Introduction

Proteins used in rodent scent communication have been widely studied in mice. These proteins, termed MUPs, belong to the lipocalin super family of proteins. Other members of the rodent family have also been found to excrete lipocalins and use them as a form of chemical communication (Table 4.1). The majority of these lipocalins share very limited sequence homology with mouse MUPs except for the highly conserved G-X-W residues, were X is any amino acid. The exception to this are rats whose scent communication proteins were previously termed α-2-globulins but have now been renamed rat MUPs (rMUPs) as they share some homology with the well established mouse MUPs.

Table 4.1 Examples of lipocalin expression in rodents.

Species	Sex	Protein	Excretion	References
Bank Vole (*Myodes glareolus*)	Male and Females	Odorant binding proteins	Urine, saliva	Stopkova *et al.,* 2010
Bank vole (*Myodes glareolus*)	Males	Glareosin	Urine	Dr M Turton, thesis
Syrian hamster (*Mesocricetus auratus)*	Females	Aphrodisin	Vaginal secretion	Singer *et al.,* 1986
Roborovski hamster *(Phodopus roborovskii)*	Males and Females	Roborovskin	Urine	Turton *et al.,* 2010
House mouse (*Musculus domesticus*)	Males and Females	Major urinary proteins (MUPS)	Urine, saliva	Finlayson and Baumann, 1957; Finlayson *et al.,* 1965
Rat (*Rattus rattus*)	Males	Rat major urinary proteins (rMUPS)	Urine	Roy and Neuhaus, 1966; Cavaggioni and Mucignat-Carretta, 2000; Hurst *et al.,* 2007

These lipocalins are found in various bodily fluids and the concentrations excreted are similar to the mouse MUPs observed in laboratory mice (approximately 4-10 mg/ml). The complexity of these proteins is much lower with only 1-4 proteins

detected observed in each species, although genomic data is either incomplete or non existent for these rodents.

The most widely studied lipocalin is aphrodisin, a female specific lipocalin secreted by Syrian hamsters (Singer *et al.,* 1986; Henzel *et al.*, 1988). This protein, found in vaginal secretions, facilitates the mounting behaviour of males via activation of a specialized sensory structure named the vomeronasal organ, which activates the accessory olfactory bulb (Clancy *et al.*, 1984; Kroner *et al.*, 1996; Jang *et al.*, 2001). Five major pheromones specifically bound onto natural aphrodisin have been identified as 1-hexadecanol (44.7%), 1-octadecanol (19.5%), Z-9-octadecen-1-ol (18.2%), E-9-octadecen-1-ol (15.4%) and hexadecanoic acid (2.2%) (Briand *et al.,* 2004). Interestingly, aphrodisin has been observed in vaginal discharges before females reach fertility, suggesting another unknown function for this lipocalin (Mägert *et al.*, 1999).

Roborovskin, a protein discovered in the urine of roborovski hamster, has not been the focus of behavioural studies but observations made on the protein chemistry level show that there is only a single protein present and that unlike MUPs and other lipocalins no sexual dimorphism is observed and both males and females secrete the similar concentrations in urine (Turton *et al.,* 2010). Bank voles secrete three odorant binding proteins primarily in their urine but also secrete the same proteins in their saliva at lower concentrations (Stopkova *et al.,* 2010). Male bank voles express an additional sex specific lipocalin protein named glareosin (Dr M Turton, thesis) but the behavioural significance of this is yet to be investigated.

There has been little investigation into olfactory communication and protein expression in the harvest mouse. The harvest mouse, *Micromys minutus*, is a small rodent approximately 2.5-3inches long that is native to Europe and Asia. Weighing approximately 6-8g, they live in fields of cereal crops or among tall grasses (Harris and Trout, 1991). They have a highly prehensile tail that is used for climbing and balance, with slightly broader feet that are used for secure gripping on to grasses leaving the front paws free to collect food (Ishwaka and Mori, 1999). They build spherical nests that are often found suspended above ground and nest sharing,

particularly during the winter, has been observed (Ishiwaka *et al.,* 2010). Wild harvest mice numbers can temporarily decline over winter, this is rectified over the spring and summer months when frequent breeding occurs.

The limited behavioral data available suggests scents influence female mate choice. Their specialised adaptations to exploit patches of seeds in tall grasses leads to local high density populations where animals defend small individual territories. Unlike most other rodents, females can display highly aggressive behaviour similar to males. The aim of this chapter was to identify and characterise the protein component in male and female harvest mice. The following objectives were set:

- Examine areas of protein secretion in male and female harvest mice
- Characterise the primary structure of these proteins using mass spectrometric techniques.
- Establish the extent of structural heterogeneity and sexual dimorphism in protein expression.
- Investigate possible roles of these proteins in scent communication.

4.2 Results and discussion

4.2.1 Examination of the urine content of the harvest mouse

The harvest mice originated from Chester Zoo (Upton-by-Chester, UK) and were housed in an outdoor enclosure of 250 square metres. Male and female harvest mice were humanely captured from this outdoor enclosure and transferred to individual cages indoor. Urine was collected from each rodent using the recovery technique described in the methods section (Chapter 2, section 2.1) and analysed (5 µl) by 1D SDS-PAGE. In both sexes two abundant protein bands were observed around 16-18 kDa (Figure 4.1) which is consistent with the mass of other lipocalins. In contrast to other rodent species, the concentration of these proteins was quite low. Unfortunately as the harvest mice only excrete between 5-10 µl of urine at most, it was difficult to obtain protein concentration as there inadequate urine for a protein assay.

4.2.2 Peptide mass fingerprinting of urine samples

To establish any differences between sexes, pieces of gel from each protein band identified by SDS-PAGE were digested following the protocol listed in Methods section 2.5. Following overnight incubation with trypsin, the digested material was analysed by MALDI-TOF. The peptide masses produced were compared to a protein database (SwissProt) and statistically analysed to see if there were any matches. No significant matches were identified in either sex.

A comparison between protein bands highlighted differences in peptide masses with few masses that were common in both proteins (Table 4.2). These proteins may therefore be from the same protein family but with some sequence variation between them. Although differences were observed between protein bands (Figure 4.2), no differences were observed between males and females (Figure 4.3 and 4.4).

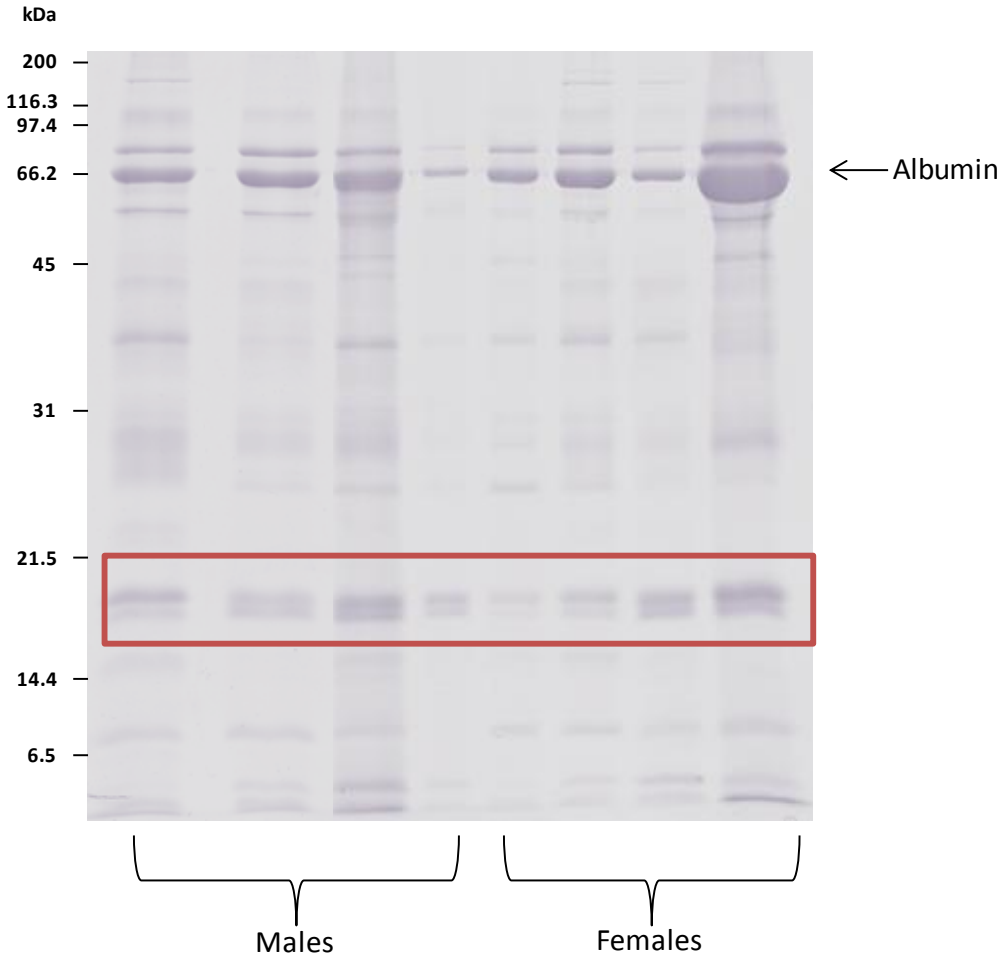


Figure 4.1 SDS-PAGE analysis of male and female harvest mouse urine.
Urine (5 µl) from both male and female harvest mice was mixed 1:1 with sample buffer and resolved by SDS-PAGE. Samples were run on a 15% SDS gel. The gel was stained with coomassie blue stain. The band at 66 kDa was later confirmed as albumin by PMF analysis. The potential lipocalins (approximately 16-18 kDa) are highlighted by the red box.

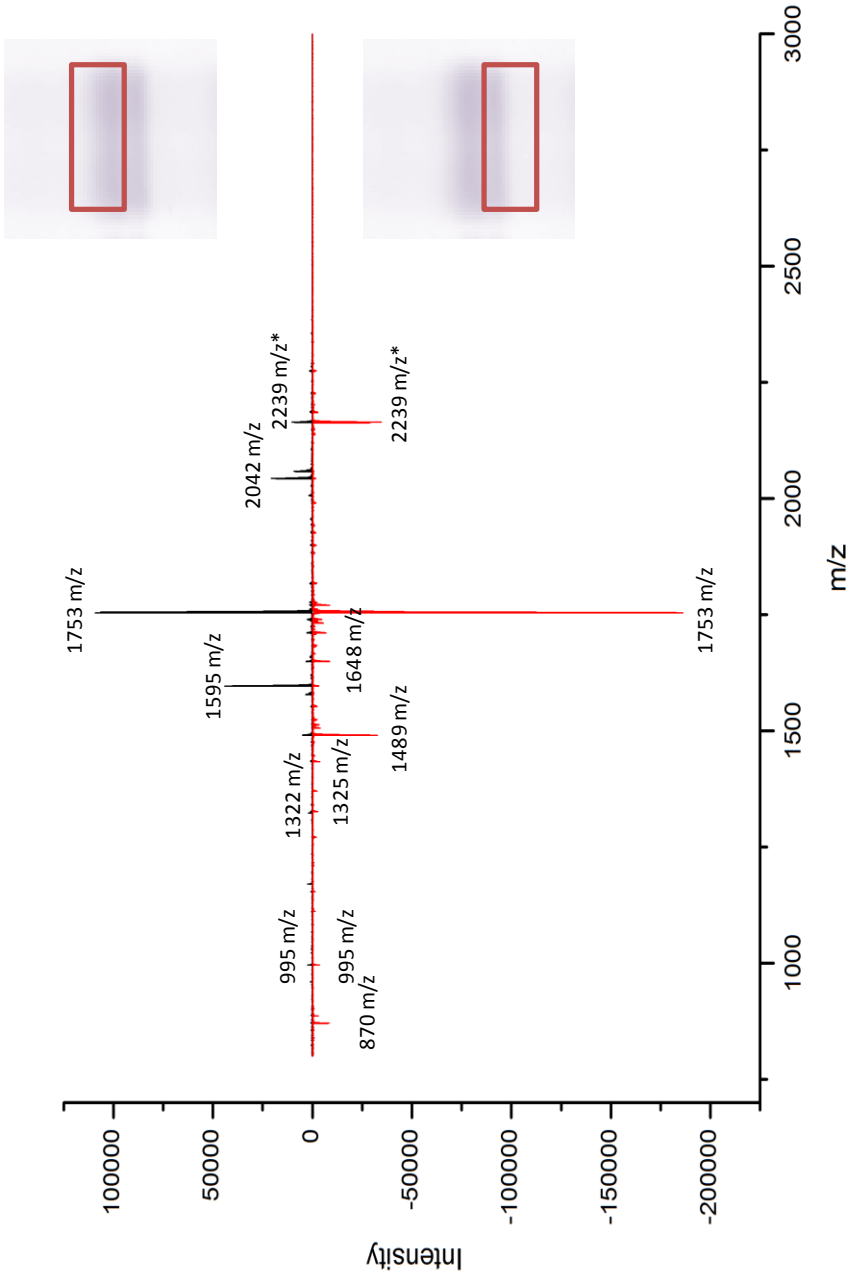


Figure 4.2 Peptide mass fingerprint comparison of the two protein bands identified by SDS-PAGE.
Small pieces of gel were extracted from the protein bands of interest from the SDS-PAGE analysis of *M. minutus* male and female urine samples. These pieces of gel were destained in 50:50 ACN:NH_4CO_3 before being reduced and alkylated in DTT (10 mM) and iodoacetamide (60 mM) respectively. Following overnight incubation at 37 °C with trypsin, the peptides were collected and mixed 1:1 with α-Cyano-4-hydroxycinnamic acid dissolved in 50% ACN, 0.1% TFA. The mixture (1 µl) was spotted onto a target plate and left to dry at room temperature before being analysed by MALDI-TOF. *trypsin autolysis peak.

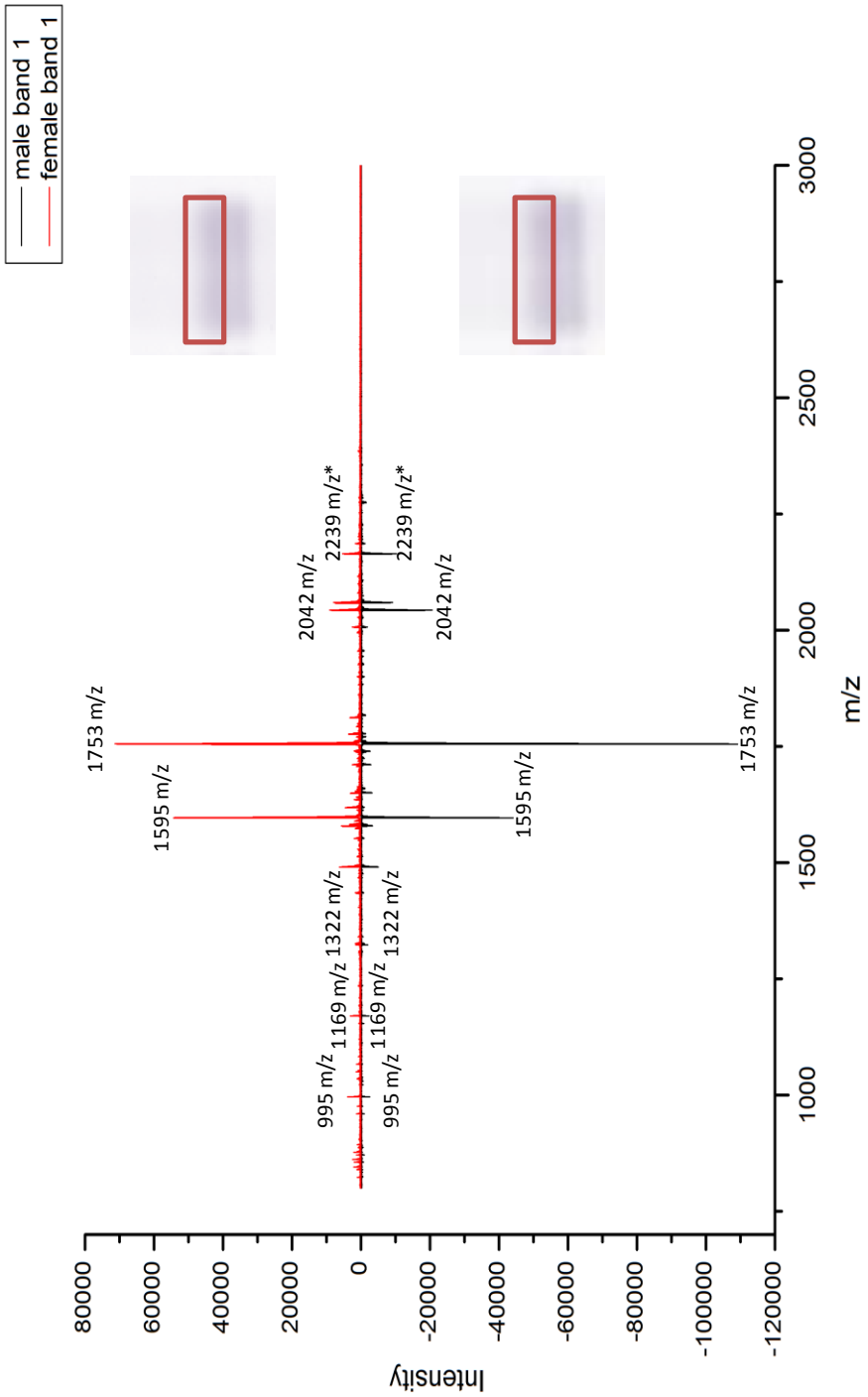


Figure 4.3 Peptide mass fingerprint comparison between male and female *M. minutus* – protein band 1
Small pieces of gel were extracted from the protein bands of interest from the SDS-PAGE analysis of *M. minutus* male and female urine samples. These pieces of gel were destained in 50:50 ACN:NH_4CO_3 before being reduced and alkylated in DTT (10 mM) and iodoacetamide (60 mM) respectively. Following overnight incubation at 37 °C with trypsin, the peptides were collected and mixed 1:1 with α-Cyano-4-hydroxycinnamic acid dissolved in 50% ACN, 0.1% TFA. The mixture (1 µl) was spotted onto a target plate and left to dry at room temperature before being analysed by MALDI-TOF. *trypsin autolysis peak.

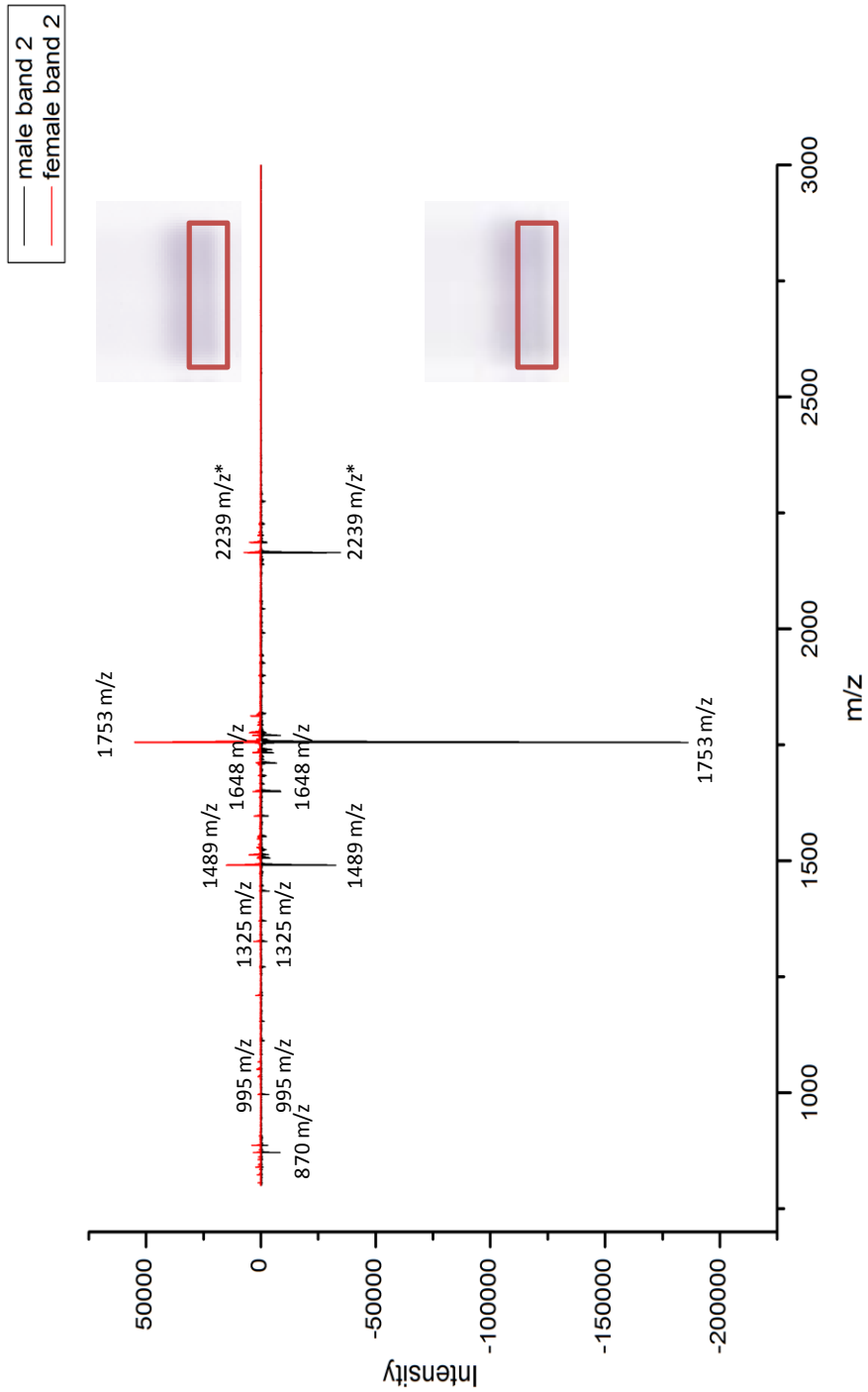


Figure 4.4 Peptide mass fingerprint comparison between male and female *M. minutus* – protein band 2
Small pieces of gel were extracted from the protein bands of interest from the SDS-PAGE analysis of *M. minutus* male and female urine samples. These pieces of gel were destained in 50:50 ACN:NH_4CO_3 before being reduced and alkylated in DTT (10 mM) and iodoacetamide (60 mM) respectively. Following overnight incubation at 37 °C with trypsin, the peptides were collected and mixed 1:1 with α-Cyano-4-hydroxycinnamic acid dissolved in 50% ACN, 0.1% TFA. The mixture (1 µl) was spotted onto a target plate and left to dry at room temperature before being analysed by MALDI-TOF. *trypsin autolysis peak.

Table 4.2. A list of abundant masses from the peptide mass fingerprint analysis of the two protein bands identified by SDS-PAGE.

PMF masses (m/z)	Detected in protein band 1?	Detected in protein band 2?	Observed in both sexes?
842.4	Yes	-	Yes
870.4	-	Yes	Yes
959.6	Yes	-	Yes
995.6	Yes	Yes	Yes
1169.4	Yes	-	Yes
1322.5	Yes	-	Yes
1325.6	-	Yes	Yes
1489.1	-	Yes	Yes
1595.6	Yes	-	Yes
1648.4	-	Yes	Yes
1753.4	Yes	Yes	Yes
1860.4	Yes	-	Yes
2042.7	Yes	-	Yes

4.2.3 Investigating other sources of protein secretion

The concentration of protein in harvest mouse urine was much lower than that observed in other rodents. In a study by Trout (1978), captive harvest mice were observed scent marking certain areas of their habitat, in particular, branches and twigs that were suspended off the ground. To investigate this theory further, Glass rods were placed inside cages of the male and female harvest mice, previously captured from the outdoor enclosure, for approximately two weeks. After two weeks these rods were removed and wiped with a cotton bud soaked in water. The end of the cotton buds were removed and placed into an Eppendorf tube (1.5 ml). The tubes were centrifuged at 2000 rpm for five minutes before the buds were removed and disposed of. Samples of the glass rod washes (5 µl) were then resolved by SDS-PAGE (Figure 4.5). Two protein bands were identified at 16-18 kDa, similar to those observed in the urine samples. The protein concentration of the glass rod samples was much higher than that detected in the urine. Also the origin of these protein bands in the glass rod washes did not appear to be urine as no albumin band was identified on the gel (Figure 4.5). This was confirmed by a creatinine assay. Creatinine is a breakdown product of creatine phosphate which is

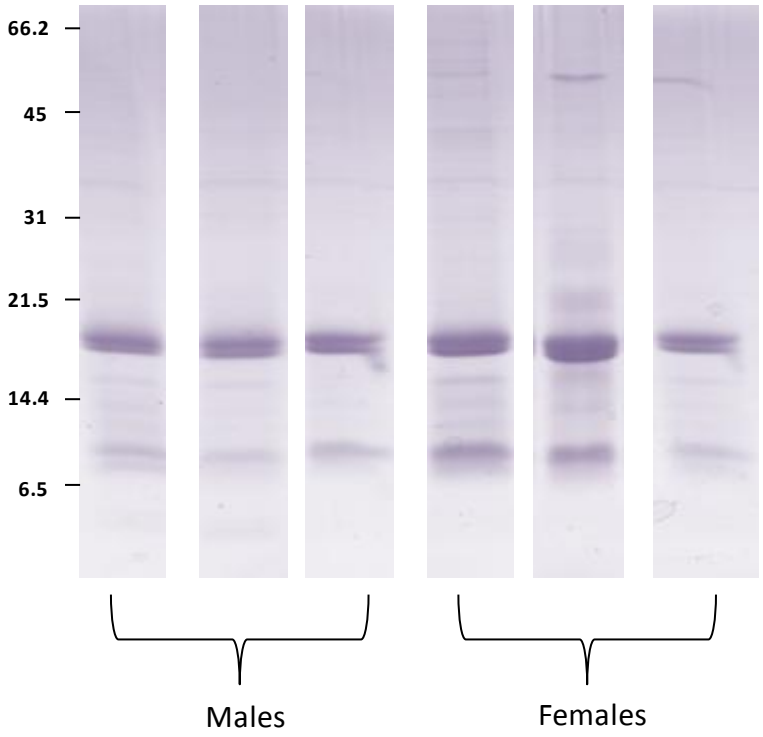


Figure 4.5 SDS-PAGE analysis of Glass rod washes from the cages of male and female harvest mice.

Glass rods from both male and female harvest mice cages were washed with cotton buds soaked in purified water (150 µl). The buds were removed and placed in Eppendorf tubes (1.5 ml) and centrifuged at 2000 rpm for five minutes. A sample of glass rod wash (5 µl) was mixed 1:1 with sample buffer and resolved by SDS-PAGE. Samples were run on a 15% SDS gel. The gel was stained with coomassie blue stain.

used in skeletal muscle contraction and is excreted directly into urine. Only trace amounts of creatinine was detected in the glass rod washes – 0.5-1 µg/ml confirming these proteins were being excreted from multiple areas. In rodents creatinine concentrations in urine are usually much higher typically ranging from 100-500 µg/ml depending on sex. A protein assay confirmed the protein concentration in these rod washes was between 3-4 mg/ml (Figure 4.6).

Further samples were collected from the same set of harvest mice including saliva, paw washes and body washes to establish the origin of the protein secretion. Urine was also collected from the same animals. Paw and body washes were collected in a similar way to the glass rod washes. The paws and stomach were washed with individual cotton buds soaked in water before removing the buds for centrifugation. Saliva was collected using a glass pipette with a small diameter tip and transferred directly into an Eppendorf tube. Approximately 0.5 – 1 µl of saliva was collected from each animal. Purified water (4 µl) was added to each saliva sample to increase the volume for analysis.

Samples of paw (5 µl), body (5 µl), saliva (1 µl) and urine (5 µl) were resolved by SDS-PAGE (Figures 4.7). Two protein bands were identified in all samples around 16-18 kDa. No albumin was identified in the paw and body washes ruling out contamination with urine.

4.2.4 Peptide mass fingerprinting of washes and saliva samples

To investigate the differences between the washes and saliva samples, an in-gel digest of the protein bands of interest (from the SDS-PAGE analysis) was completed. After overnight incubation with trypsin, the digested material was analysed using MALDI-TOF. The PMFs were compared with those collected from the urine analysis (Figure 4.8 and 4.9). No differences were observed confirming that these are most likely the same two proteins.

Samples of gel were also taken and digested with two other proteases (LysC and GluC) that cleave the protein at different sites. These PMFs together with the

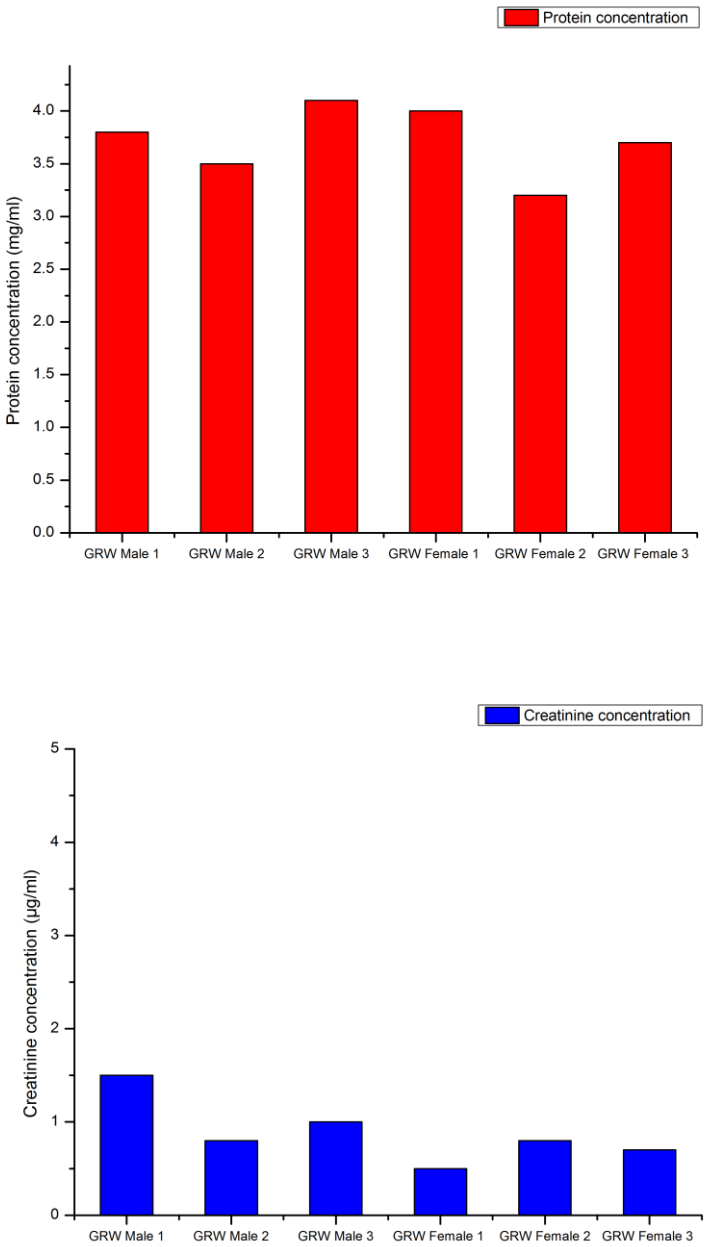


Figure 4.6 Protein and creatinine assays of harvest mouse glass rod washes.
Glass rods from both male and female harvest mice cages were washed with cotton buds soaked in purified water (150 µl). The buds were removed and placed in Eppendorf tubes (1.5 ml) and centrifuged at 2000 rpm for five minutes. Protein assay (top graph). Samples were diluted down in purified water to be in the linear range of the assay. Absorbance readings were measured at 620 nm using a plate reader. Creatinine assay (bottom graph). Creatinine concentration was measured using a creatinine assay kit. A creatinine standard curve was prepared (0-30 µg/ml), these samples did not require a dilution due to the low abundance of creatinine in these samples. Absorbance readings were measured at 570 nm.

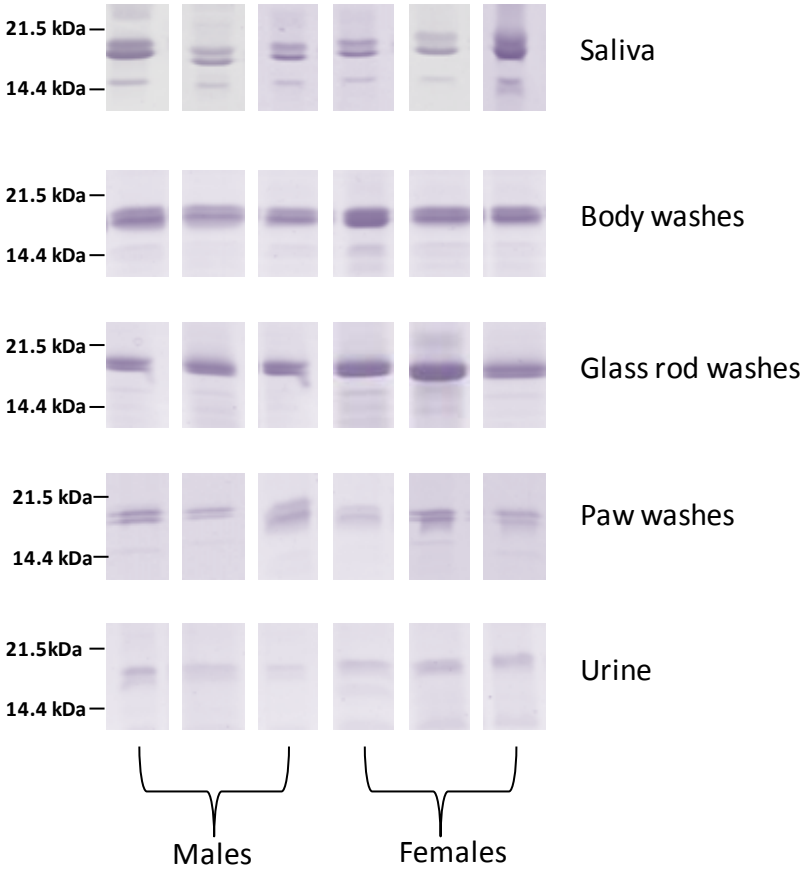


Figure 4.7 SDS-PAGE analysis of saliva, paw washes, body washes, glass rod washes and urine from male and female harvest mice.

Glass rods from both male and female harvest mice cages were washed with cotton buds soaked in purified water (150 µl). The buds were removed and placed in Eppendorf tubes (1.5 ml) and centrifuged at 2000 rpm for five minutes. The paws and stomach of male and female harvest mice were washed with individual cotton buds soaked in water (50 µl) before transferring the buds to Eppendorf tubes (1.5 ml) for centrifugation. Saliva was collected using a glass pipette with a small diameter tip and transferred directly into an Eppendorf tube (0.5 ml). Approximately 0.5 – 1 µl of saliva was collected from each animal. Purified water (4 µl) was added to each saliva sample to increase the volume so analysis could proceed. Samples of paw wash (5 µl), body wash (5 µl), and saliva (1 µl), glass rod wash (5µl) and urine (5 µl) were mixed 1:1 with sample buffer and resolved by SDS-PAGE. Samples were run on a 15% SDS gel. The gel was stained with coomassie blue stain.

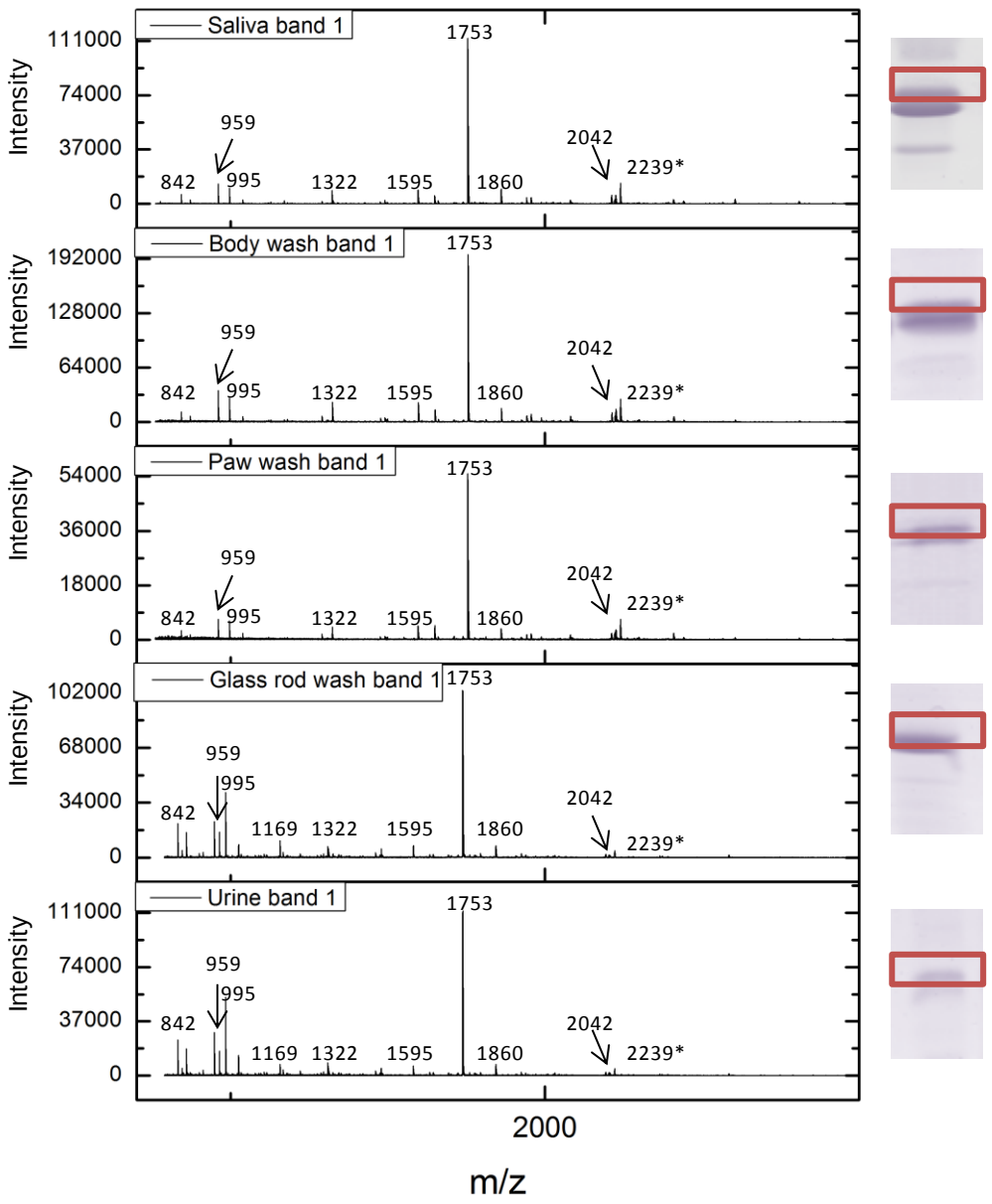


Figure 4.8 Peptide mass fingerprint comparison between paw, body and glass rod washes, saliva and urine from harvest mice – protein band 1 (upper band)
Small pieces of gel were extracted from the protein bands of interest from the SDS-PAGE analysis of harvest mouse paw, body and glass rod washes, and saliva and urine samples. These pieces of gel were destained in 50:50 ACN:NH_4CO_3 before being reduced and alkylated in DTT (10 mM) and iodoacetamide (60 mM) respectively. Following overnight incubation at 37 °C with trypsin, the peptides were collected and mixed 1:1 with α-Cyano-4-hydroxycinnamic acid dissolved in 50% ACN, 0.1% TFA. The mixture (1 µl) was spotted onto a target plate and left to dry at room temperature before being analysed by MALDI-TOF. *trypsin autolysis peak.

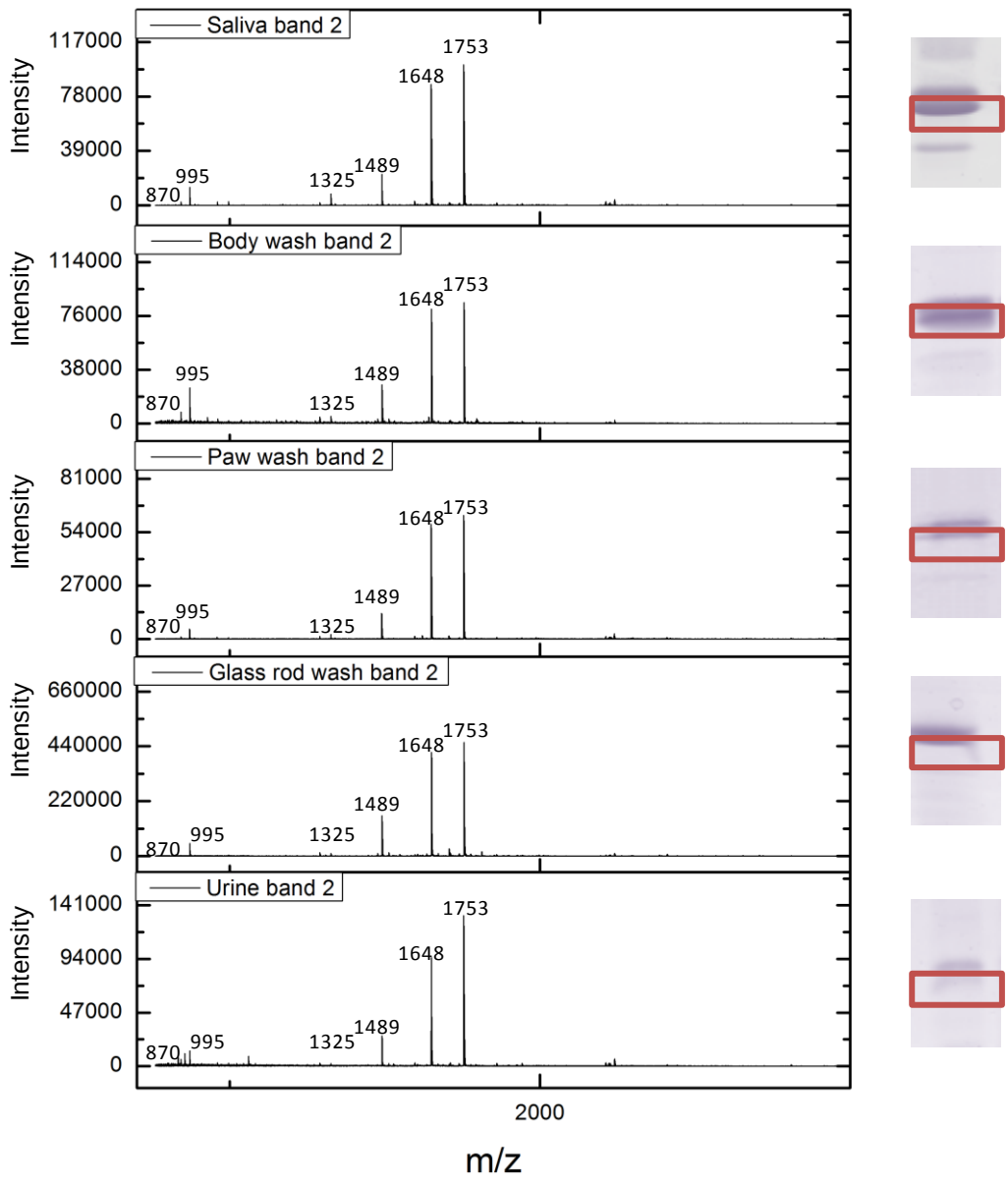


Figure 4.9 Peptide mass fingerprint comparison between paw, body and glass rod washes, saliva and urine from harvest mice – protein band 2 (lower band)
Small pieces of gel were extracted from the protein bands of interest from the SDS-PAGE analysis of harvest mouse paw, body and glass rod washes, saliva and urine samples. These pieces of gel were destained in 50:50 ACN:NH_4CO_3 before being reduced and alkylated in DTT (10 mM) and iodoacetamide (60 mM) respectively. Following overnight incubation at 37 °C with trypsin, the peptides were collected and mixed 1:1 with α-Cyano-4-hydroxycinnamic acid dissolved in 50% ACN, 0.1% TFA. The mixture (1 µl) was spotted onto a target plate and left to dry at room temperature before being analysed by MALDI-TOF. *trypsin autolysis peak.

tryspin PMFs were used to support the sequence evidence found from the *de novo* sequencing analysis.

Table 4.3 A comparison between the abundant masses observed in the original urine PMF and the saliva, glass rod, paw and body washes

PMF masses band 1 -urine	Glass rod wash band 1	Paw wash band 1	Body wash band 1	Saliva band 2
842.4	✓	✓	✓	✓
959.6	✓	✓	✓	✓
995.6	✓	✓	✓	✓
1169.4	✓	✓	✓	✓
1322.5	✓	✓	✓	✓
1595.6	✓	✓	✓	✓
1753.4	✓	✓	✓	✓
1860.4	✓	✓	✓	✓
2042.7	✓	✓	✓	✓
PMF masses band 2 -urine	**Glass rod wash band 2**	**Paw wash band 2**	**Body wash band 2**	**Saliva band 2**
870.4	✓	✓	✓	✓
995.6	✓	✓	✓	✓
1325.6	✓	✓	✓	✓
1489.1	✓	✓	✓	✓
1648.4	✓	✓	✓	✓
1753.4	✓	✓	✓	✓

4.2.5 Determination of an accurate molecular weight

To obtain a more accurate molecular weight, samples of the glass rod washes from three male and three female harvest mice were diluted (5 pmol/µl in 0.1% formic acid) and analysed by electrospray (ESI) mass spectrometry. The glass rod washes were chosen as they contained a higher concentration of protein. Three abundant proteins were identified in both sexes – 16437 Da, 16724 Da and 17888 Da. All animals expressed 16724 Da with varying expression of the other two proteins (Figure 4.10). The variation in expression between individuals may be a result of genetic diversity as the animals originate from the outdoor enclosure where breeding is not controlled so these rodents are not as in-bred as laboratory strains of mouse.

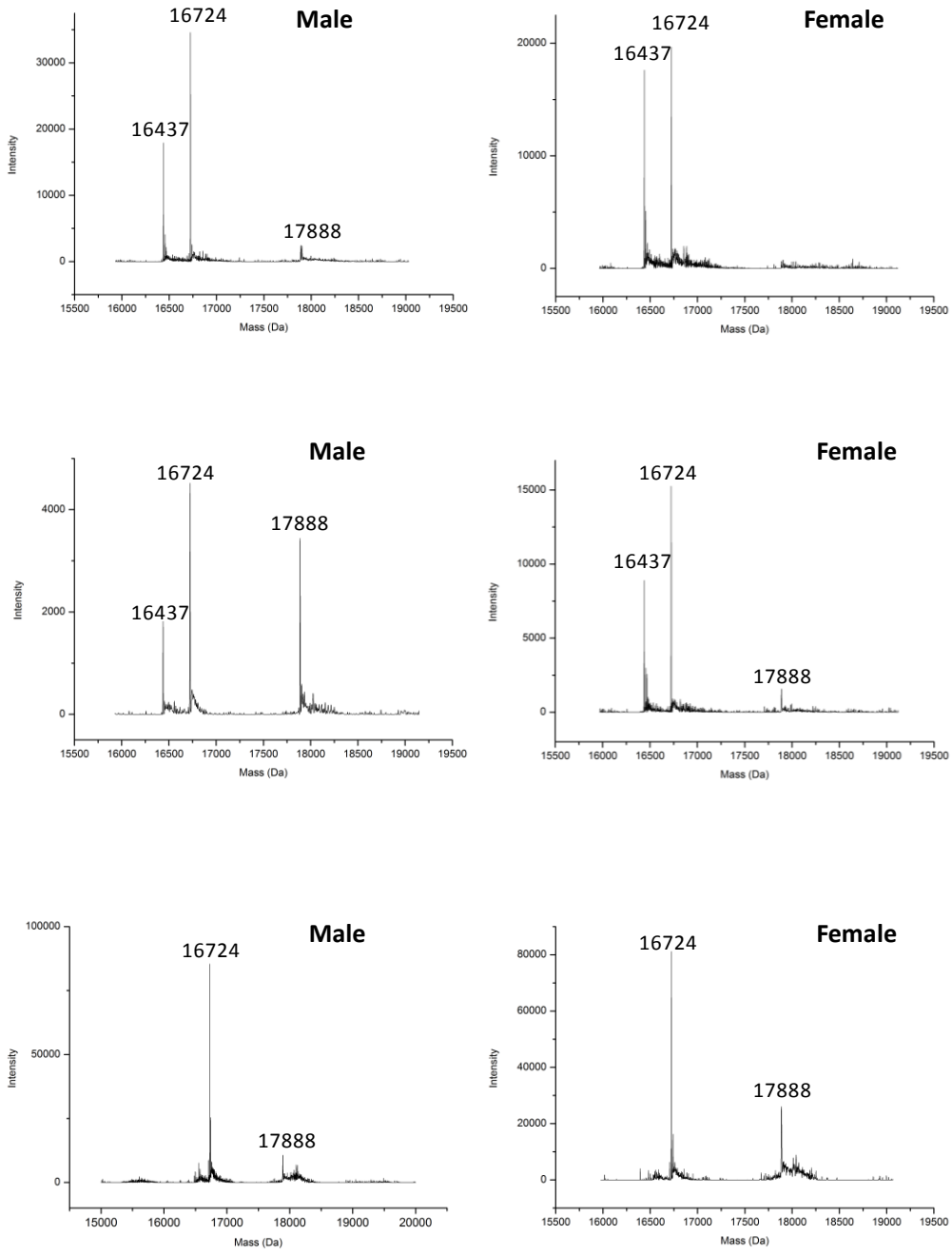


Fig 4.10 Determination of an accurate molecular weight of the protein bands identified by SDS-PAGE in male and female harvest mice.

Glass rod samples from male and female harvest mice were diluted into formic acid (0.1 %) to produce a final concentration of approximately 5 pmol/µl. The samples were then injected onto a C4 desalting trap and masses of proteins present were determined by ESI-MS. Data was processed using maximum entropy software MAX ENT-1 (MassLynx 4.1, Waters).

4.2.6 Discovery analysis

Urine, washes and saliva were digested in-solution using the protocol described in Chapter 2, section 2.5. Following proteolysis overnight with trypsin, the digested material was analysed by LC-MS (Figure 4.11a and 4.11b). The raw data was processed using PEAKS 6 software (Bioinformatics Solutions Inc., Canada).

PEAKS is a bioinformatics software that has *de novo* sequencing, database searching and protein quantification capabilities. Raw data is imported from an LC-MS analysis. This raw data can often contain background noise, redundancy as well as errors due to sample preparation and instrument approximation. PEAKS will use a data refining tool to improve the overall quality of the data. This tool merges MS/MS scans from the same precursor m/z, using retention time to do this. It will also use isotopic envelope patterns to correct the precursor m/z as sometimes instruments will give values that are not the monoisotopic ion. The data refinement tool will also remove low quality spectra and centroid and deconvolute charge and isotopes. The MS/MS data is then *de novo* sequenced using an algorithm and scoring functions that are specific to each mass analyser used. *De novo* sequencing is carried out in the absence of a protein database which is advantageous for identifying novel peptides (Zhang et al., 2003). Once the data has been *de novo* sequenced, a summary of all peptides sequenced is displayed with both a confidence score (as a percentage) for each amino acid in the sequence as well as an overall average local confidence score (as a percentage) for each peptide which is calculated as the total of the residue scores divided by the peptide length. The amino acids are also colour coded to reflect the scoring of the residue for example high scoring peptides are coloured red. This unique feature allows the user to obtain very high confidence sequence tags even in cases where PEAKS can not find the complete sequence with a high confidence level due to poor quality spectra.

It is also possible to include a database search in the set up of the processing method. When a protein database is available, PEAKS can further explain the *de novo* sequences. By comparing the *de novo* sequences with the database, PEAKS

can confirm the database search results, find PTMs, mutations and homologous peptides; as well as output the *de novo* only peptides (Zhang et al., 2011).

Prior to using PEAKS to *de novo* sequence the harvest mouse proteins; it was internally validated using previously sequenced proteins. In this case the well established mouse MUPs were used to test the PEAKS *de novo* abilities. Mouse urine from male B6 laboratory mice was digested and analysed on three different mass spectrometry platforms – Orbitrap, QTOF and an ion trap. The raw data was imported into PEAKS and the data refinement tool initiated. Appropriate error tolerances were set for each mass analyser and the data was *de novo* sequenced. As predicted the data from the Orbitrap produced high scoring data (over 85%) with many of the peptides matching exactly to mouse MUP sequences. This is most likely due to the high sensitivity, mass accuracy and good quality MS/MS fragmentation data provided by the Orbitrap. The other two platforms produced data of a medium quality with very few high scoring peptides. So while PEAKS produced impressive data with the Orbitrap setup, it is also important to choose an appropriate mass spectrometry platform to achieve confident *de novo* sequence data.

For the harvest mouse in-solution digests the PEAKS processing method was set up to *de novo* sequence the peptides followed by a database search using a custom made lipocalin database. Many sequences aligned with odorant binding proteins from mice, rats and bank voles. There were also sequences that aligned with MUPS 4, 5 and 20 (Table 4.4).

Table 4.4 Peptide sequence tags from PEAKS database search. Raw data was processed using PEAKS software. Samples were searched against a lipocalin database for potential matches and sequence tags for peptides that may share high sequence homology to other peptides belonging to lipocalin proteins. PEAKS defaults to leucine for all leucine and isoleucine residues. Using the current LC-MS system, it is not possible to distinguish between the two residues due to their isobaric nature.

Sequence tag	PEAKS de novo score	Protein identification from database	% shared identity
LNGDWFSLLTASEK	93	Rat MUP (*Rattus norvegicus)* MUP 5(*Mus musculus*)	100 80
LEENGSMR	96	Rat MUP (*Rattus norvegicus)* MUP 4 (*Mus musculus*) MUP 5(*Mus musculus*) MUP 20(*Mus musculus*)	100 86 86 86
EPDLSSDLK	98	MUP 20(*Mus musculus*) MUP 5(*Mus musculus*) Rat MUP (*Rattus norvegicus)*	100 100 78
TDYDNYLMFHVTNVK	86	MUP 4(*Mus musculus*) MUP 20(*Mus musculus*)	80 80
CLEAR	89	MUP 20(*Mus musculus*)	100
SVALAADNLNK	97	Aphrodisin (*Cricetulus griseus*)	73
SLTTVTGYVEADGQTYR	85	Odorant binding protein 1a and 1b (*Mus musculus*)	69
EEVEGLMSEVTK	85	Vomeromodulin (*Rattus norvegicus*)	91
LTALAANNADK	98	Odorant binding protein 1 (*Myodes glareolus*)	69
LQEEGPMR	92	Odorant binding protein 2 (*Myodes glareolus*)	78
ELTCEDDCK	94	aphrodisin-like (*Rattus norvegicus)*	67
NQYEGDRNFEPVK	93	Odorant binding protein 2 (*Myodes glareolus*)	73
ATPENLVFYSENVDR	96	Odorant binding protein 1b (*Mus musculus*)	86
LLFVVGK	99	Odorant binding protein 2 (*Myodes glareolus*)	86
TQFEGDNHFAPVK	93	Odorant binding protein 2 (*Myodes glareolus*) Odorant binding protein 3 (*Myodes glareolus*)	83 75
ATPDNLVFYSENLDR	95	Odorant binding protein 1b (*Mus musculus*) Odorant binding protein 3 (*Myodes glareolus*)	73 73
VLFVVGHAPLTPDQR	91	Odorant binding protein 1a (*Mus musculus)*	62

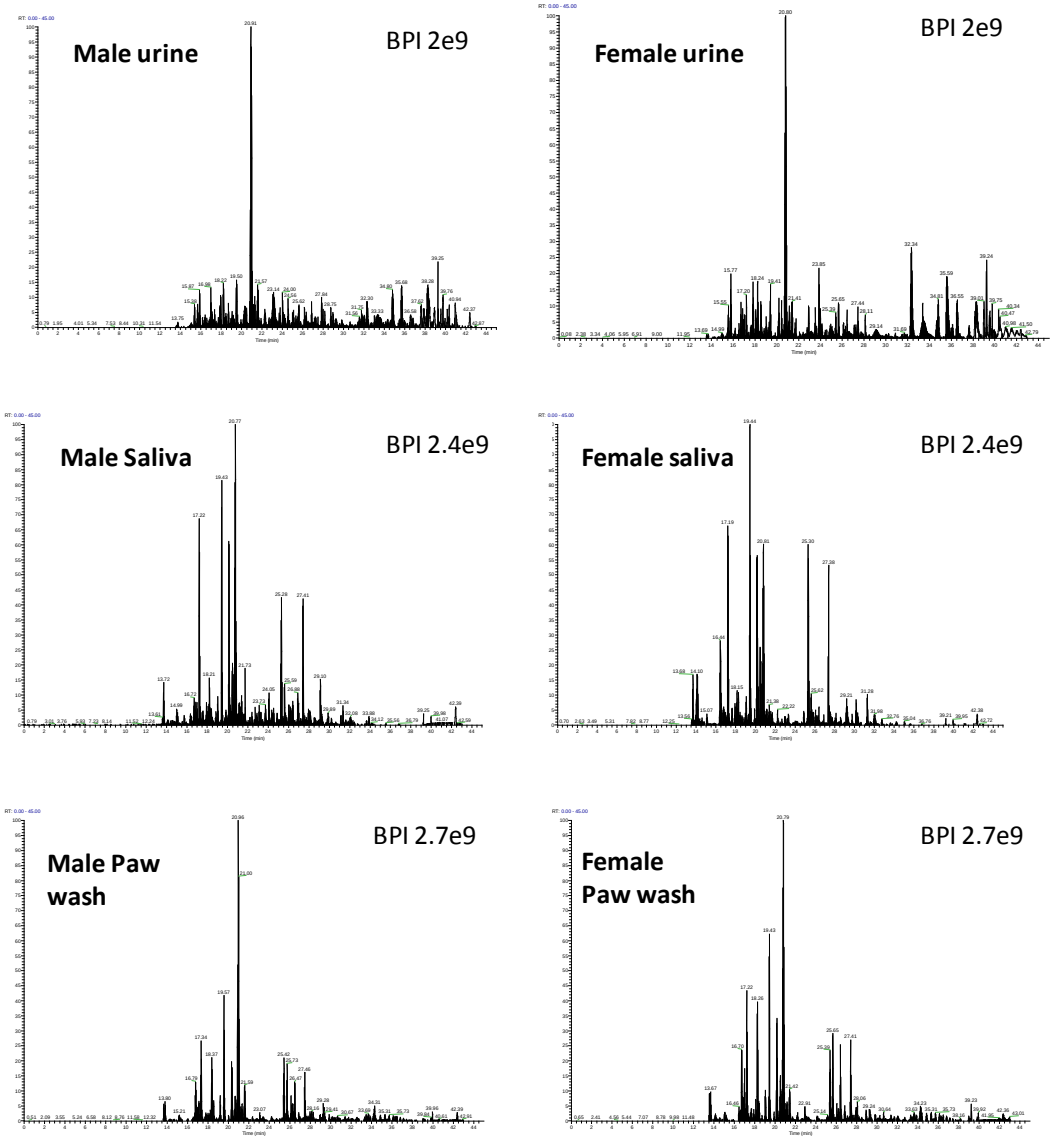


Fig 4.11a Base peak chromatograms from LC-MS discovery run.
Protein (1-5 µg) from male and female urine, paw washes, body washes, glass rod washes and saliva was reduced with DTT (3 mM final concentration) and alkylated with iodoacetamide (9 mM final concentration). Trypsin was added (50:1 substrate:enzyme) and the digests incubated overnight at 37 °C. The samples were acidified with TFA (0.1 % final concentration). Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient. Spectra were acquired between 300-2000 m/z

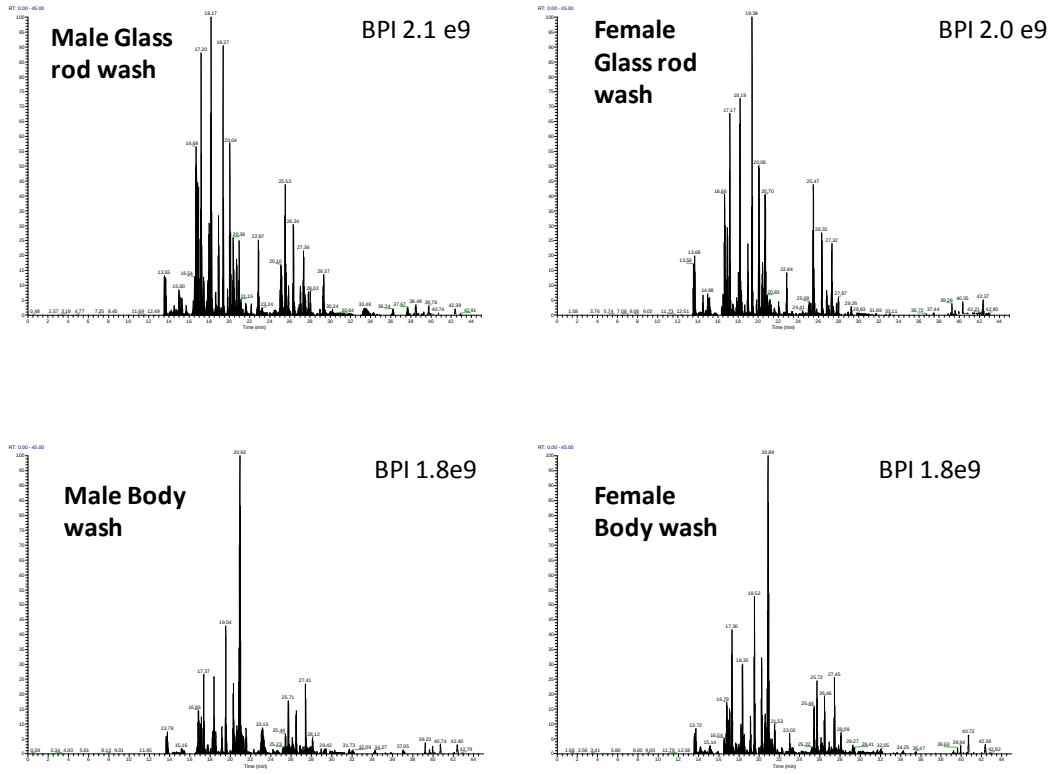


Fig 4.11b Base peak chromatograms from LC-MS discovery run.
Protein (1-5 µg) from male and female urine, paw washes, body washes, glass rod washes and saliva was reduced with DTT (3 mM final concentration) and alkylated with iodoacetamide (9 mM final concentration). Trypsin was added (50:1 substrate:enzyme) and the digests incubated overnight at 37 °C. The samples were acidified with TFA (0.1 % final concentration). Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient. Spectra were acquired between 300-2000 m/z.

4.2.7 Protein purification for *de novo* sequencing analysis

As there was a mixture of proteins present in the samples, a purification step was implemented prior to *de novo* sequencing. Anion exchange chromatography (AEX) was used to separate the proteins into individual fractions. AEX is a non denaturing protein purification technique that separates proteins according to their net charge. Glass rod washes were used for AEX because they contained the highest concentration and amount of protein. The isoelectric point (PI) of the proteins to be separated would normally be used to determine buffers and pH. A pH higher than the PI would give the proteins a negative charge and vice versa for cation exchange. The PI of these proteins was unknown therefore a method that had previously been used to separate lipocalins in the bank vole was used as a starting point.

Samples (100 µl) were injected onto a UNO Q (1 ml) anion exchange column that was equilibrated in 10 mM Hepes pH 8.0 and eluted with a 0-0.5 M NaCl gradient in the same buffer. The chloride ions disturb the ionic interaction between the column resin and negatively charged proteins. The negatively charged proteins are progressively displaced from the resin and eluted from the column. Fractions were collected (1 ml/min) over a 60 minute period. Post analysis, the fractions were split into two tubes, one for SDS-PAGE and one for *de novo* sequencing analysis. Strataclean beads were added to one set of fractions to concentrate the protein solution and analysed by SDS-PAGE. SDS-PAGE analysis of the fractions revealed protein around 16-18 kDa had been eluted in fractions' 12-15 (Figure 4.12). ESI analysis of these fractions identified that proteins 16437 Da and 16724 Da had been successfully separated (Figure 4.13). There was no protein at 17888 Da despite it being present in the starting material.

The absence of 17888 Da was either due to low abundance of this protein or because the AEX conditions did not suit this protein. As the discovery run highlighted the presence of MUPS the AEX conditions were changed to a one previously used to separate MUPS (Robertson *et al.,* 1996) incase this 17888 Da protein was possibly a MUP rather than an odorant binding protein. Samples (100

µl) were injected onto a Mono Q (1 m) anion exchange column that was equilibrated with 50 mM MES buffer pH 5.0 and eluted over a 0-0.2 M NaCl gradient. Fractions were collected (1.5 ml) over an 85 minute period. Strataclean beads were then added to the fractions to concentrate the protein solution. The beads were then analysed by SDS-PAGE. No protein was identified on the gel indicating that the MUPS detected in the discovery run were in low abundance.

4.2.8 *De novo* sequencing analysis

The other half of each fraction containing the protein to be sequenced was split into three aliquots. Strataclean beads were added to each aliquot. The beads were digested using three different proteases - trypsin, endoproteinase LysC and endoproteinase GluC – to produce overlapping sequence information due the specificity of each enzyme. Following overnight proteolysis, the samples were analysed by LC-MS to produce *de novo* sequence data.

Following LC-MS analysis, the raw was processed using PEAKS 6 software. Peptides that were sequenced in PEAKS were then manually BLAST searched (http://blast.ncbi.nlm.nih.gov). BLAST – Basic Local Alignment Search Tool, is a bioinformatics software that identifies sections of similarity between sequences. For each protein there were peptide sequences that shared some homology with odorant binding proteins found in the bank vole and mouse (Table 4.5). The samples were also searched against a lipocalin database and matches and sequence tags for odorant binding proteins were observed (as in the discovery run section 4.2.6) but none for MUPS suggesting the two main proteins that the harvest mouse excretes were most likely to be lipocalins but not MUPS.

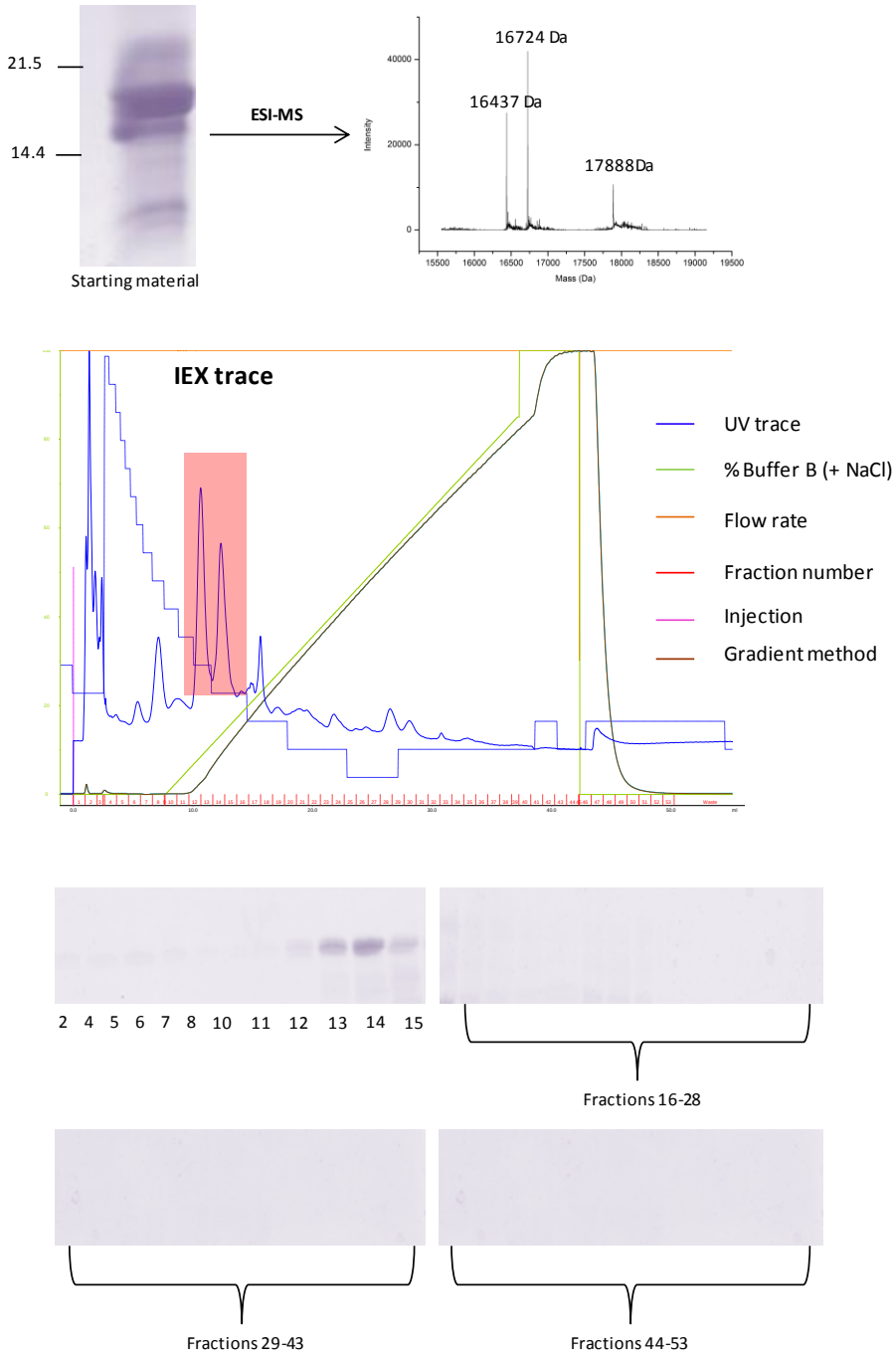
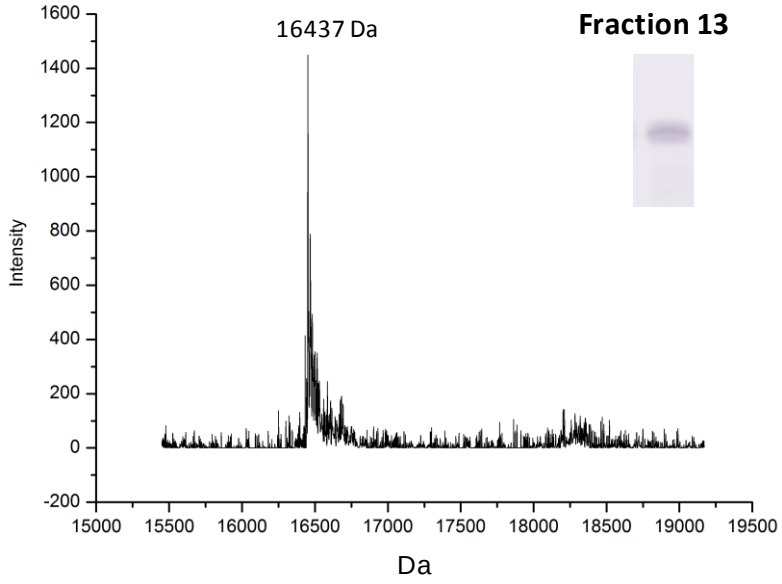


Fig 4.12 Harvest mouse protein purification using anion exchange chromatography.
A glass rod wash from a harvest mouse was initially examined by ESI-MS to check all three proteins to be purified had been expressed. The sample (100 µl) was injected onto a UNO Q (1 ml) anion exchange column that was equilibrated in 10 mM Hepes pH 8.0 and eluted with a 0-0.5 M NaCl gradient. Fractions (1 ml) were collected over a 50 minute period. The fractions were split into two aliquots, one aliquot was treated with strataclean beads to concentrate up any protein in the samples. Samples were vortexed for 2 minutes before being centrifuged at 2000 rpm for two minutes. The supernatant was discarded and the beads were mixed with sample buffer (20 µl) and analysed by SDS-PAGE. Samples were run on a 15% SDS gel. Gels were stained with coomassie blue stain.



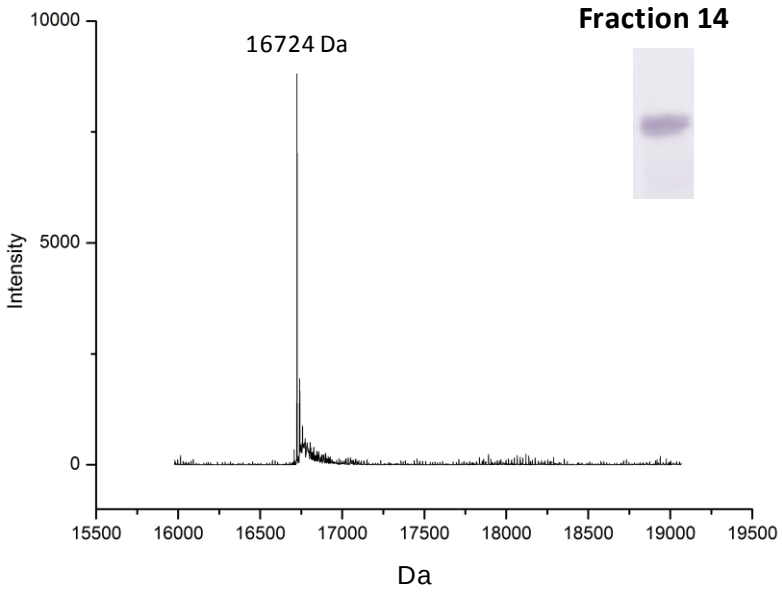


Fig 4.13 Identification of proteins present in AEX fractions.
Following SDS-PAGE analysis, Fractions 13 and 14 were diluted into formic acid (0.1 %) to produce a final concentration of approximately 5 pmol/µl. The samples were then injected onto a C4 desalting trap and masses of proteins present were determined by ESI-MS. Data was processed using maximum entropy software MAX ENT-1 (MassLynx 4.1, Waters).

Table 4.5 BLAST results of harvest mouse protein sequences obtained from LC-MS analysis. Sequences for both proteins were assessed using the blastP algorithm. Search parameters were restricted to rodents.

Harvest mouse Sequence (16724 Da protein)	Protein identification	Score	E value	Sequence homology
SLEGKWK	aphrodisin-like (*Rattus norvegicus)*	23.5	0.11	85%
LTALAANNADK	Odorant binding protein 1 (*Myodes glareolus*)	22.7	55	78%
LQEEGPMR	Odorant binding protein 2 (*Myodes glareolus*)	24.4	16	86%
ELTCEDDCK	aphrodisin-like (*Rattus norvegicus)*	25.7	6.5	67%
NQYEGDRNFEPVK	Odorant binding protein 2 (*Myodes glareolus*)	25.2	14	73%
ATPENLVFYSENVDR	Odorant binding protein 1b (*Mus musculus*)	43.5	8e-08	86%
LLFVVGK	Odorant binding protein 2 (*Myodes glareolus*)	21	186	86
Harvest mouse Sequence (16437 Da protein)	**Protein identification**	**Score**	**E value**	**Sequence homology**
SLEGKWK	aphrodisin-like (*Rattus norvegicus)*	23.5	0.11	85%
SVALAADNLNK	Aphrodisin (*Cricetulus griseus*)	21.8	137	64%
ELTCEDDCKR	aphrodisin-like (*Rattus norvegicus)*	25.7	7.2	67%
TQFEGDNHFAPVK	Odorant binding protein 2 (*Myodes glareolus*) Odorant binding protein 3 (*Myodes glareolus*)	34.6 29.9	0.014 0.44	83% 75%
ATPDNLVFYSENLDR	Odorant binding protein 1b (*Mus musculus*) Odorant binding protein 3 (*Myodes glareolus*)	37.1 35	0.002 0.012	73% 73%
VLFVVGHAPLTPDQR	Odorant binding protein 1a (*Mus musculus*)	28.2	1.8	62%

4.2.9 Determination of the harvest mouse protein sequences

Based on the BLAST results the harvest mouse protein sequences were aligned against OBP 2 (*Myodes glareolus*) for the 16724 Da protein and OBP 3 (*Myodes glareolus*) for the 16437 Da protein (Figures 4.14 and 4.15). Using peptides produced from three different proteases partial sequences for each protein were confirmed. Most sequences were confirmed by at least two cleavage strategies for the 16724 protein. There was less confirmatory evidence for the 16437 Da protein due to the lack of cleavage sites for LysC and GluC. The lack of lysine and glutamic acid residues in certain parts of the protein resulted in only tryptic fragments providing sequence data. The highly conserved lipocalin consensus G-X-W, where X represents any amino acid residue, was observed in both sequences providing confirmation that these proteins belong to the lipocalin family.

To collect further confirmatory sequence information, particularly for protein 16437 Da, a fourth digest (post purification) using endoproteinase AspN (AspN) was prepared. AspN hydrolyses peptide bonds at the N terminal side of aspartic acid and also glutamic acid but at a slower rate. Due to the potential non-specific cleavage, the PEAKS processing method was set to include cleavage at both sites. This was successful as it provided some extra confirmatory sequences and even produced a candidate peptide for the C-terminus of the 16437 Da protein. However in both proteins there is a section (amino acids 50 -65 in Figures 4.14 and 4.15) where little or no sequence data was found at all. This could be due to a number of factors. The first is lack of cleavage sites leading to large peptides that are difficult to fragment and ionise in the mass spectrometer source. Secondly there could be many cleavage sites leading to small peptide fragments being produced, these fragments may be too small to be identified.

The harvest mouse proteins were sequenced with the aid of the peptide mass fingerprints for added confidence (Figure x). PMFs for tryptic and LysC digests were used to match up masses to sequences to confirm the two abundant bands identified by SDS-PAGE were the proteins that had been partially sequenced (Table 4.6). Unfortunately the GluC PMF did not identify any peptides. As many GluC

autolysis peaks were identified, it is likely the protease self-digested preventing efficient digestion. There could potentially be missed cleavages of the harvest mouse peptides but without complete sequence data it is difficult to identify these.

Some of the smaller masses were difficult to detect in the MALDI spectra due to ion suppression caused by the MALDI matrix ions. The LysC PMFs (Figure 4.16) both had an abundant unique ion - 2480 m/z in band 1 (16724 Da) and 2137 m/z in band 2 (16437 Da). This confirmed the 1753 m/z ion in the trypsin PMFs, which was an assumed shared peptide between the two proteins, were two unique sequences even though they share the same mass. Unfortunately only partial sequence information was collected for the 2480 m/z ion There were also some masses in both LysC PMFs that did not have a corresponding sequenced peptide. These could either be missed cleavages or peptides that were difficult to fragment resulting in poor *de novo* sequencing data.

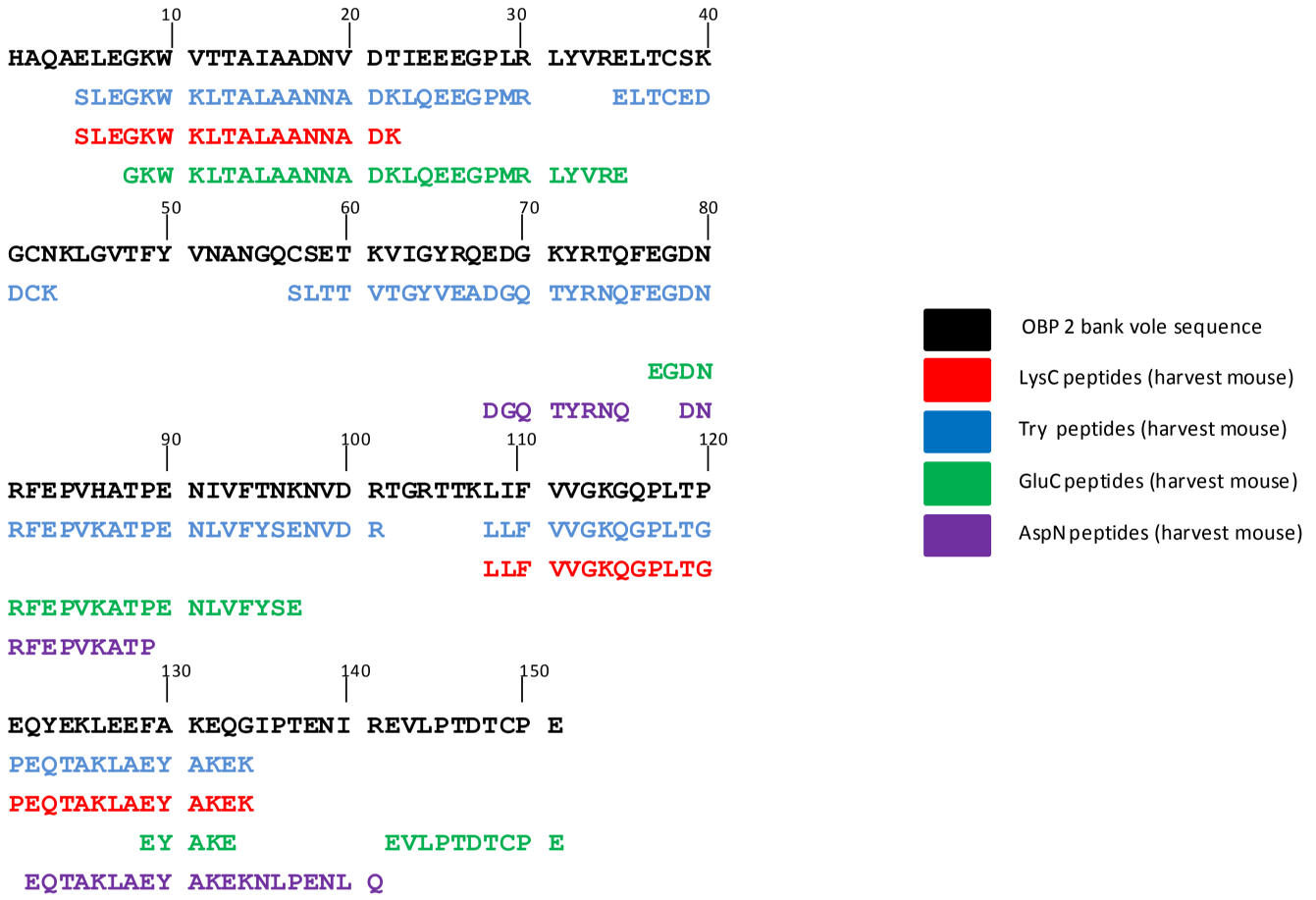


Figure 4.14 Sequence alignment of harvest mouse protein 16724 Da with odorant binding protein 2 (*Myodes glareosin*).

Peptide sequences obtained from the PEAKS *de novo* sequencing analysis were BLAST searched and shared high homology with odorant binding protein 2 (*Myodes glareosin*). Peptide sequences from all 3 digests – trypsin , LysC and GluC were aligned with the *M. glareosin* protein. PEAKS defaults to leucine for all leucine and isoleucine residues. Using the current LC-MS system, it is not possible to distinguish between the two residues due to their isobaric nature.

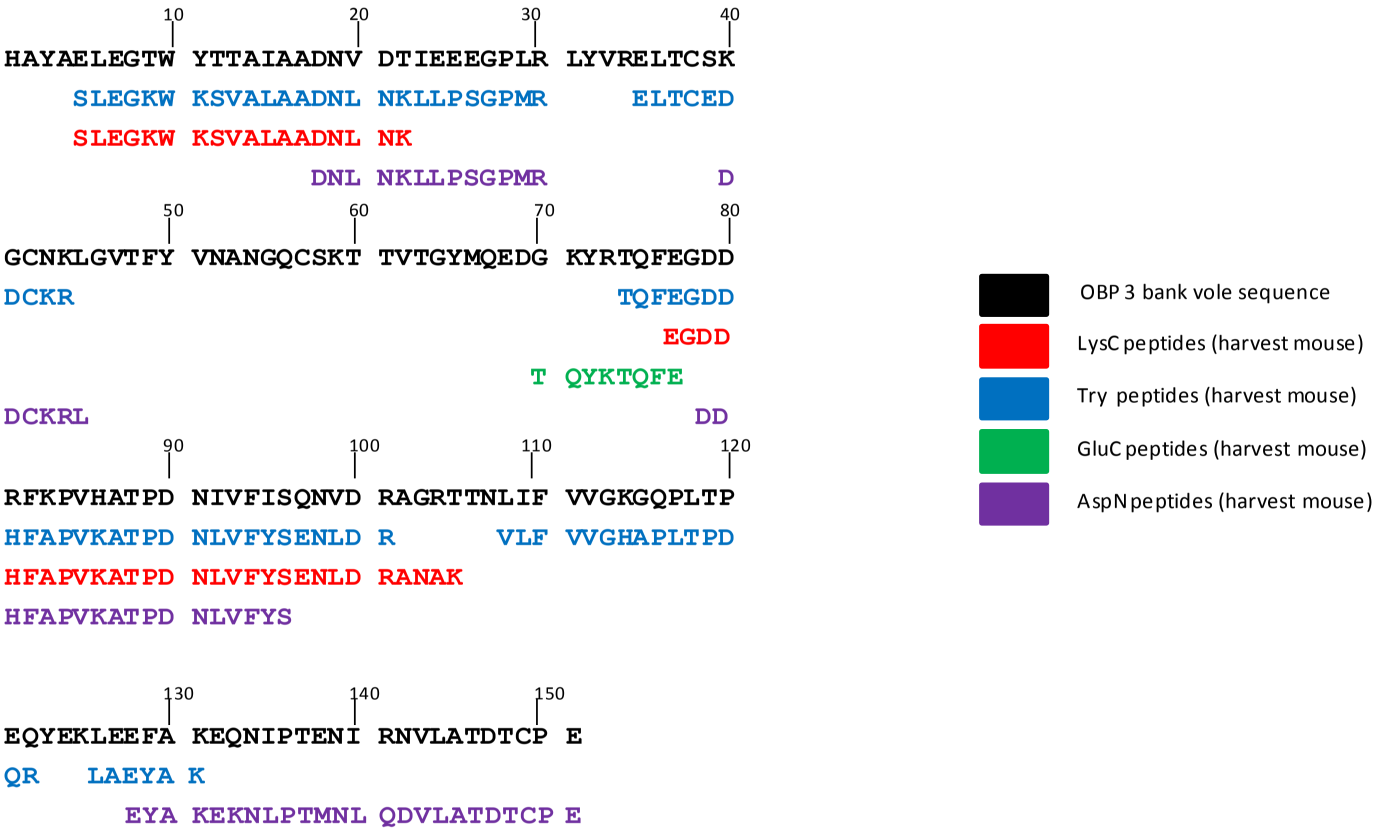


Figure 4.15 Sequence alignment of harvest mouse protein 16437 Da with odorant binding protein 3 (*Myodes glareolus*).
Peptide sequences obtained from the PEAKS *de novo* sequencing analysis were BLAST searched and shared high homology with odorant binding protein 3 (*Myodes glareolus*). Peptide sequences from all 3 digests – trypsin , LysC and GluC were aligned with the *M. glareolus* protein. PEAKS defaults to leucine for all leucine and isoleucine residues. Using the current LC-MS system, it is not possible to distinguish between the two residues due to their isobaric nature..

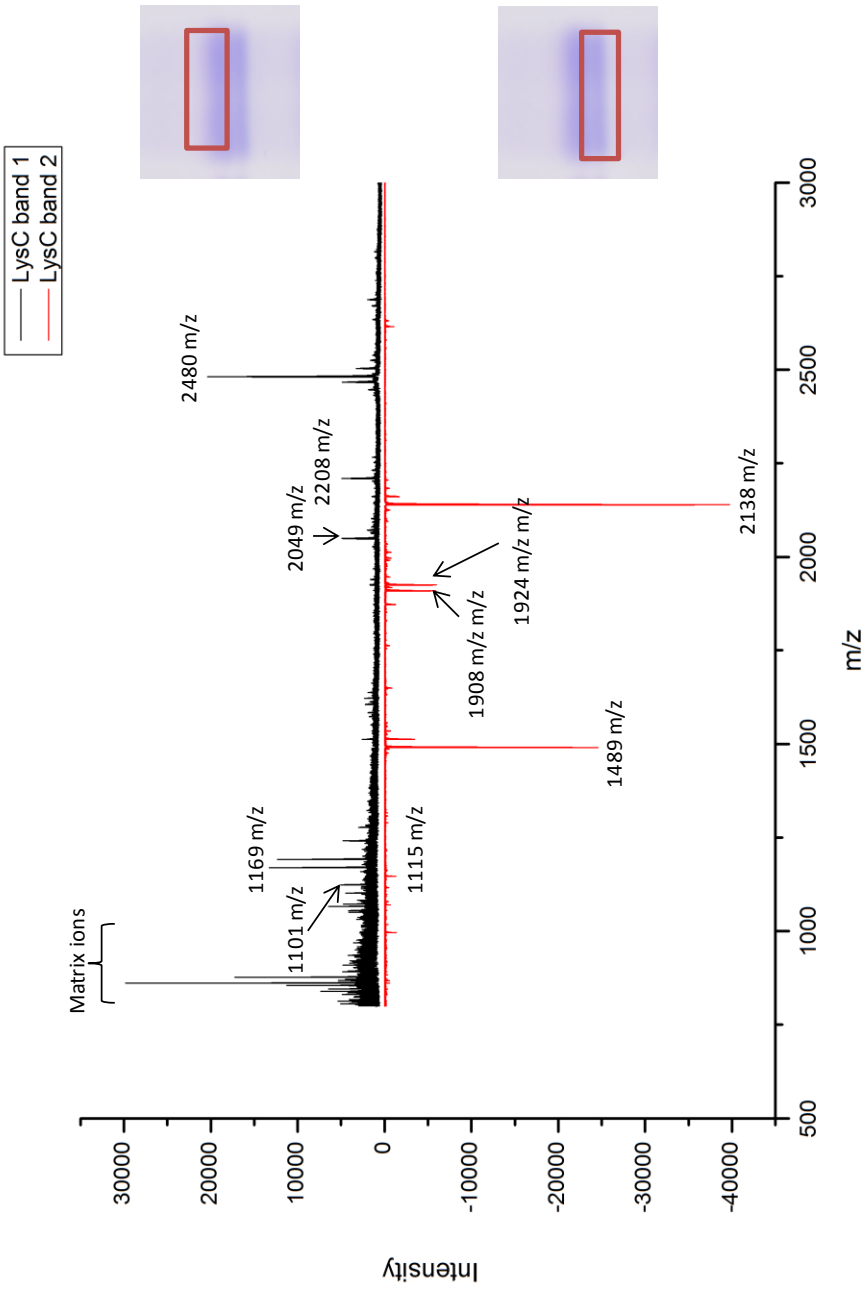


Figure 4.16 LysC Peptide mass fingerprint to compare to the masses from peptides that were *de novo* sequenced. Small pieces of gel were extracted from the protein bands of interest from the SDS-PAGE analysis of *M. minutus* male and female urine samples. These pieces of gel were destained in 50:50 ACN:NH_4CO_3 before being reduced and alkylated in DTT (10 mM) and iodoacetamide (60 mM) respectively. Following overnight incubation at 37 °C with LysC, the peptides were collected and mixed 1:1 with α-Cyano-4-hydroxycinnamic acid dissolved in 50% ACN, 0.1% TFA. The mixture (1 µl) was spotted onto a target plate and left to dry at room temperature before being analysed by MALDI-TOF.

Table 4.6 A comparison between the masses observed by PMF analysis and the sequence data.

Sequence protein 16724 Da	Protease	Corresponding m/z in PMF
LQEEGPMR	Trypsin	959.4
LTALAANNADK	LysC	1101.6
ELTCEDDCK	Trypsin and LysC	1169.4
NQYEGDRNFEPVK	Trypsin	1595.7
ATPENLVFYSENVDR	Trypsin	1753.8
SLTTVTGYVQADGQTYR	Trypsin	1858.9
Sequence protein 16437 Da	**Protease**	**Corresponding m/z in PMF**
SVALAADNLNK	LysC	1115.6
LLPSGPMR	Trypsin	870.5
ELTCEDDCKR	Trypsin	1325.5
TQFEGDNHFAPVK	Trypsin and LysC	1489.7
ATPDNLVFYSENLDR	Trypsin	1753.8
VLFVVGHAPLTPDQR	Trypsin	1648.9
ATPDNLVFYSENLDRANAK	LysC	2138.1

The list of peptides used to piece together the harvest mouse protein sequences are highlighted in tables 4.7 and 4.8. Examples of MS/MS fragmentation spectra are illustrated in figures 4.17-4.24. The rest of the MS/MS product ion spectra are illustrated in supplementary data B.

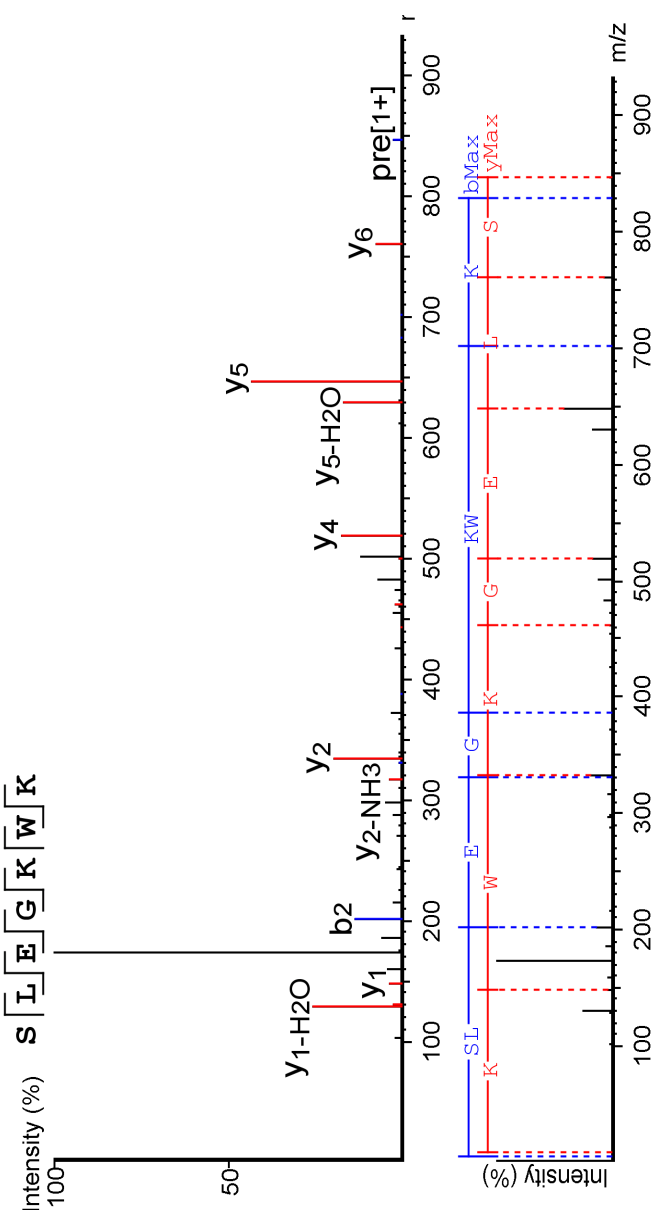


Figure 4.17 *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* tryptic peptide 847 m/z (16724 Da protein)
M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

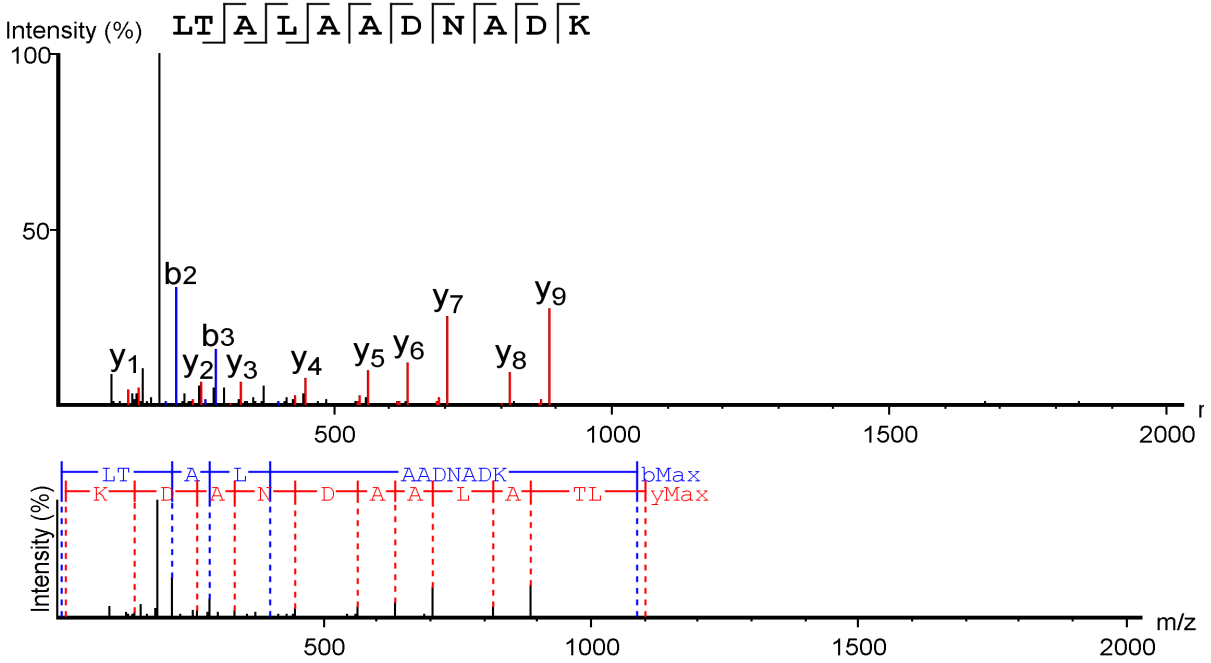


Figure 4.18 *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* LysC peptide 1101 m/z (16724 Da protein)

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

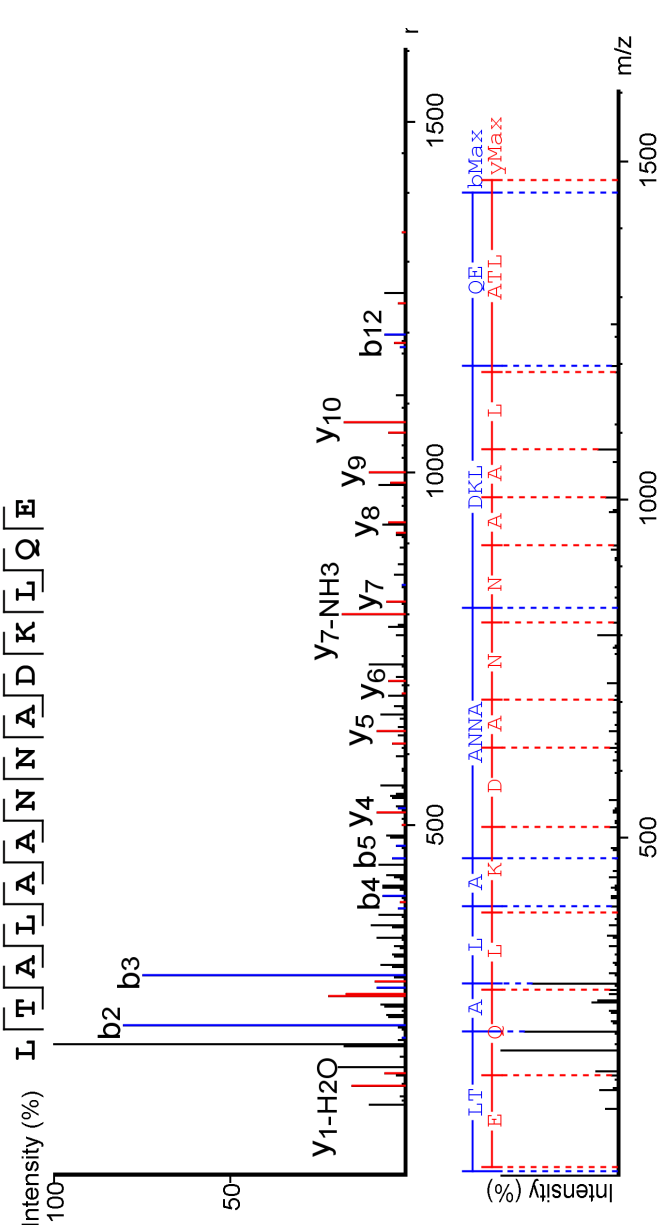


Figure 4.19 *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* GluC peptide 1471 m/z (16724 Da protein)
M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

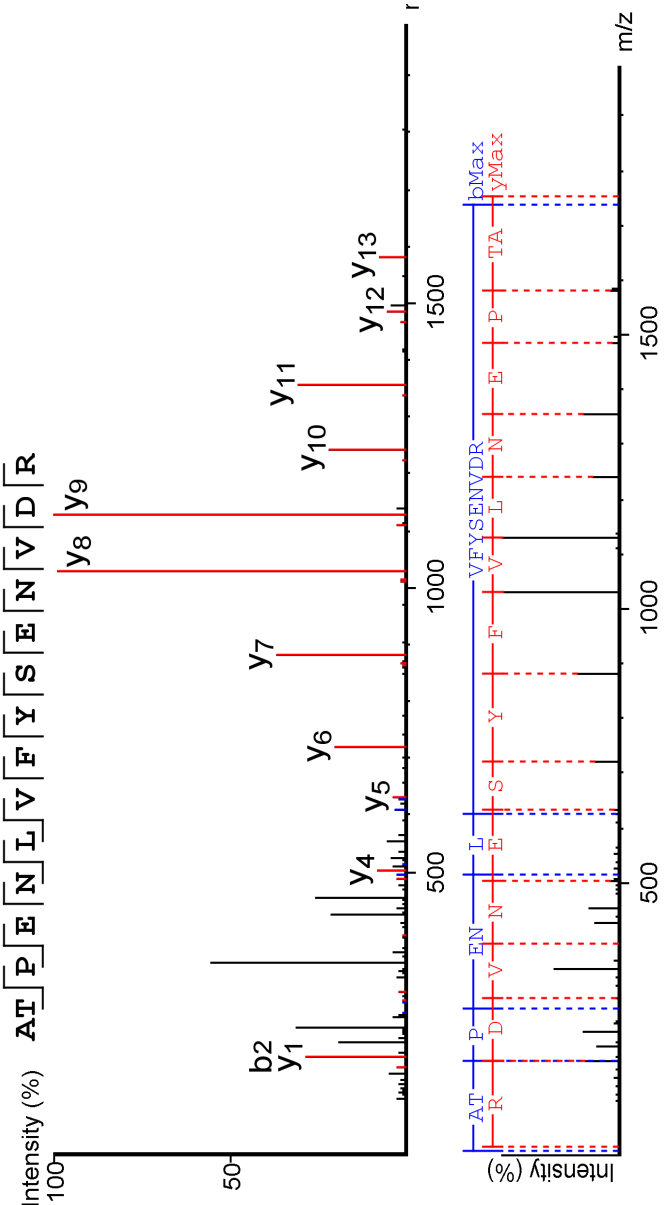


Figure 4.20 *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* tryptic peptide 1753 m/z (16724 Da protein)
M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

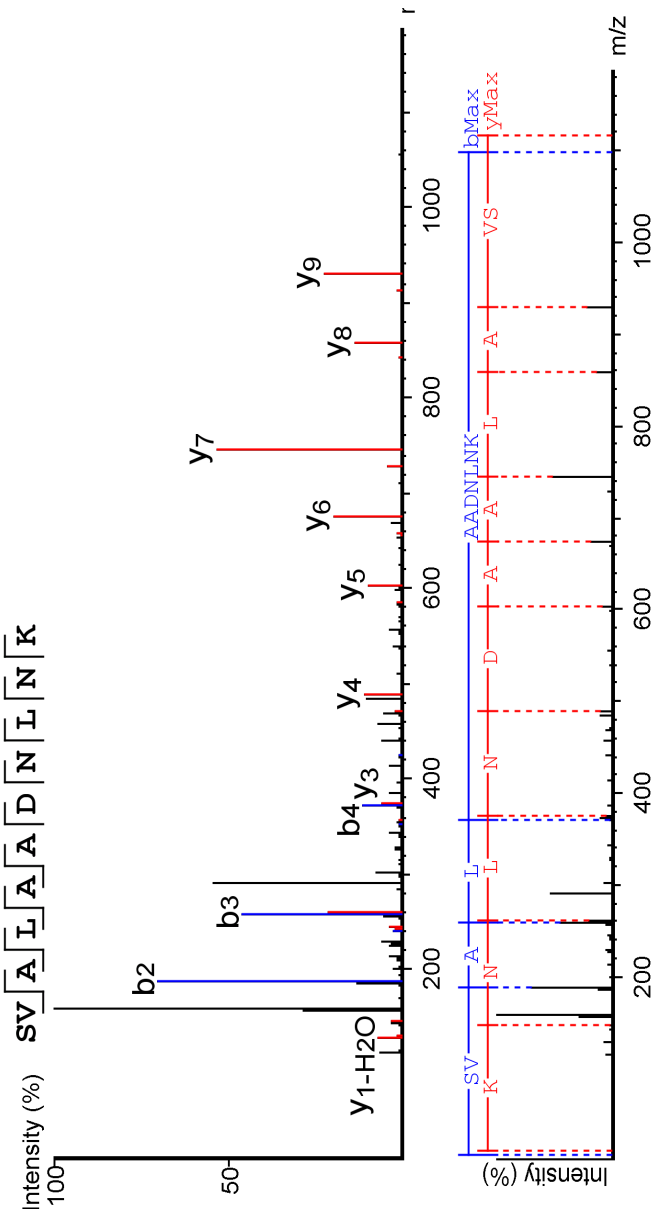


Figure 4.21 *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* tryptic peptide 1114 m/z (16437 Da protein)
M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

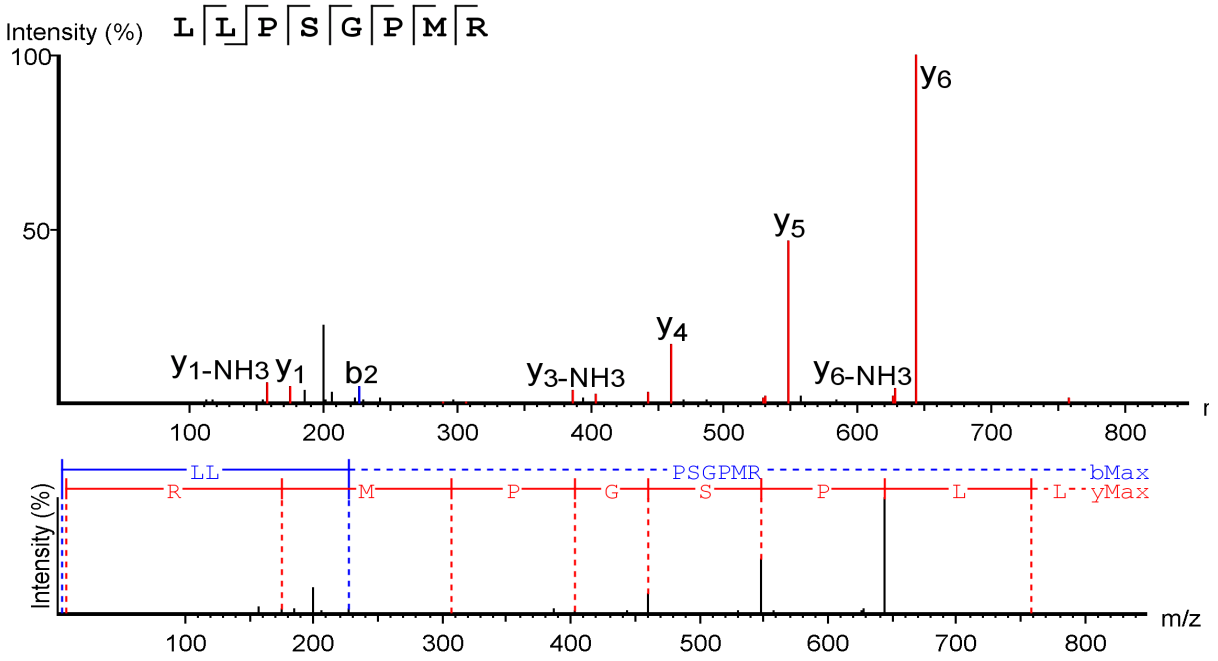


Figure 4.22 *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* tryptic peptide 870 m/z (16437 Da)
M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

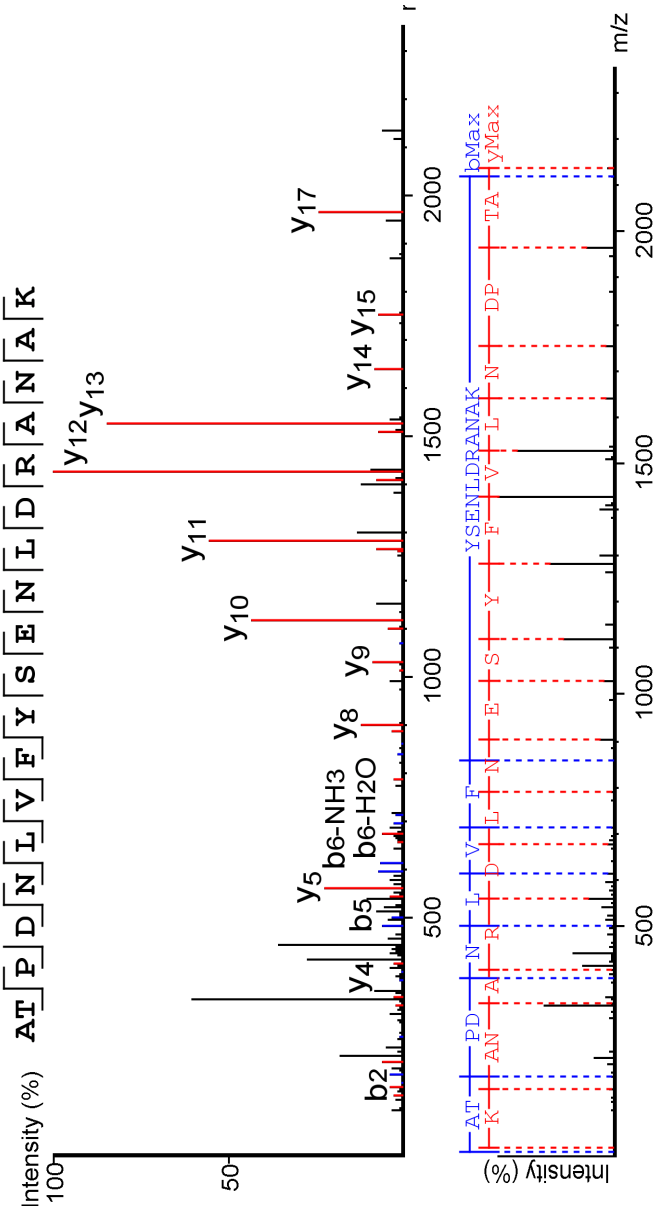


Figure 4.23 *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* LysC peptide 2138 m/z
M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

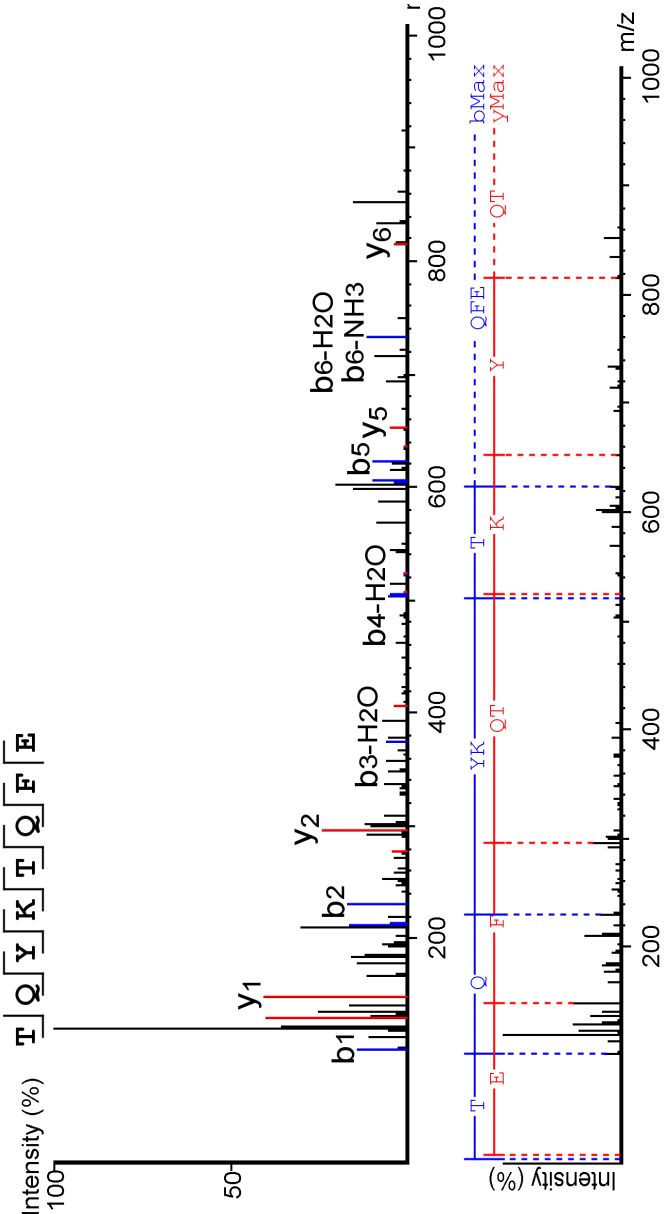


Figure 4.24 *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* GluC peptide 1043 m/z (16437 Da protein)
M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

Table 4.7 A summary of all peptides *de novo* sequenced from the harvest mouse in-solution digest of 16724 Da. The raw data was processed using PEAKS software. A cut off value of 55% for the total confidence level (recommended by PEAKS) was applied to the de novo analysis. Each amino acid was given an individual confidence percentage. The total confidence score was worked out using the mean of the individual scores. PEAKS defaults to leucine for all leucine and isoleucine residues. Using the current LC-MS system, it is not possible to distinguish between the two residues due to their isobaric nature.

Sequence	Protease	Mass (Da)	Individual residue confidence scores (%)	Average PEAKS confidence score (%)
SLEGKWK	Trypsin LysC	846.46	86 87 97 90 90 98 95 97 98 100 98 98 99 96	92 98
LTALAANNADK	Trypsin LysC	1100.58	98 99 100 100 100 100 98 97 97 99 87 97 99 100 100 100 100 99 98 98 99 90	98 98
LQEEGPMR	Trypsin	958.45	99 97 100 99 97 94 92 61	92
ELTCEDDCK	Trypsin	1168.44	96 98 99 100 100 97 100 100 61	94
NQYEGDRNFEPVK	Trypsin	1594.73	95 95 87 94 98 97 94 96 95 98 100 100 59	93
ATPENLVFYSENVDR	Trypsin	1752.83	86 98 100 99 97 99 100 100 95 96 95 91 93 97 92	96
LLFVVGK	Trypsin LysC	774.50	100 99 99 100 100 100 61 99 99 99 100 100 100 58	99 94
QGPLTGPEQTAK	Trypsin	1225.63	86 91 99 99 100 96 91 91 87 92 96 59	91
LAEYAK	Trypsin	693.36	100 100 100 100 100 97 100 100 100 100 100 96	99 99
LAEYAKEK	Trypsin	950.50	100 100 100 99 99 99 99 93	99
GQPLTGPEQTAK	LysC	1125.63	56 51 99 99 99 98 91 91 88 94 96 59	85
EGPMRLYVRE	GluC	1248.62	99 100 100 100 100 96 96 99 59 59	91
GDNRFEPVKATPE	GluC	1458.71	97 97 92 93 93 97 100 99 99 99 99 100 59	94
NLVFYSE	GluC	870.41	90 90 95 97 99 100 59	90
EYAKE	GluC	638.29	99 99 99 100 59	91
LTALAANNADKLQE	GluC	1470.76	90 90 100 100 100 99 97 97 97 99 98 96 98 94	97
DTALVTCPE	GluC	1004.44	85 86 96 98 97 97 99 100 59	91
SLTTVTGYVQADGQTYR	Trypsin	1858.94	60 64 96 99 99 99 99 99 99 97 98 98 69 64 96 98 85	85
DGQTYRNQ	AspN	980.43	83 74 91 94 92 87 90 75	85
DNRFEPVKATP	AspN	1272.61	87 88 85 71 76 88 79 77 82 81 80	81
EQTAKLA	AspN	759.43	95 76 80 79 73 51 69	75
EYAKEKNLPPENLQ	AspN	1671.82	97 92 87 77 90 87 83 92 97 88 97 91 97 86	90

Table 4.8 A summary of all peptides *de novo* sequenced from the harvest mouse in-solution digest of 16437Da. The raw data was processed using PEAKS software. A cut off value of 55% for the total confidence level (recommended by PEAKS) was applied to the de novo analysis. Each amino acid was given an individual confidence percentage. The total confidence score was worked out using the mean of the individual scores. PEAKS defaults to leucine for all leucine and isoleucine residues. Using the current LC-MS system, it is not possible to distinguish between the two residues due to their isobaric nature.

Sequence	Protease	Mass (Da)	Individual residue confidence scores (%)	Average PEAKS confidence score (%)
SLEGKWK	Trypsin LysC	846.46	98 98 99 97 97 99 97	98 98
SVALAADNLNK	Trypsin LysC	1114.60	94 94 99 100 100 99 99 98 98 99 95	98 98
LLPSGPMR	Trypsin	869.48	59 100 100 100 98 97 96 59	89
ELTCEDDCKR	Trypsin	1324.54	89 96 99 100 99 94 98 97 95 95	96
TQFEGDNHFAPVK	Trypsin LysC	1488.70	89 85 99 91 98 98 98 99 100 100 100 100 59	93 94
ATPDNLVFYSENLDR	Trypsin	1752.83	67 96 97 97 97 100 100 100 97 98 96 91 94 97 92	95
VLFVVGHAPLTPDQR	Trypsin	1647.91	74 73 82 92 96 95 97 99 96 94 96 94 92 96 90	91
LAEYAK	Trypsin LysC	693.36	100 100 100 100 100 95 100 100 100 100 100 95	99 99
ATPDNLVFYSENLDRANAK	LysC	2137.04	99 98 93 93 91 100 100 100 97 97 91 76 78 95 90 61 48 65 58	84
TQYKTQFE	GluC	1043.49	88 85 96 85 97 93 98 95	92
DDCKRL	AspN	805.37	87 100 99 90 78 97	
DNLNKLLPSGMPR	AspN	1453.71	78 72 96 96 97 99 98 61 56 56 76 68 81	71
DDHFAPVKATP	AspN	1196.54	62 89 88 95 96 87 90 79 88 97 92	85
DNLVFYS	AspN	856.32	76 70 87 82 73 72 71	76
EYAKEKNLPTMNLQ	AspN	1693.89	99 99 98 95 98 92 87 97 73 74 77 77 96 92	89
DVLATDTCPE	AspN	1119.43	70 75 82 71 54 77 76 92 64 83	74

4.4 Conclusions

Like many other rodent species, harvest mice to excrete proteins belonging to the lipocalin family. These proteins are excreted by both male and females with little sexual dimorphism observed. Both sexes excrete the same proteins (confirmed by ESI-MS and PMF) and the concentration expressed is very similar in each sex (SDS-PAGE). Unlike the majority of other rodents whose primary source of protein excretion is their urine, harvest mice secrete high concentrations of protein in their saliva and possibly their paws with much lower concentrations observed in the urine. The protein identified in the body wash samples was most likely transferred onto the stomach areas during grooming and when climbing up and down the glass rods. The glass rod washes also contained a substantial amount of protein. Urine contamination was ruled out as no albumin was observed during SDS-PAGE analysis and no creatinine was detected in the samples. The origin of the protein secretion was either saliva and the rodents were licking their paws prior to climbing up the rods, or the animals were secreting protein from their paws during the climbing process. Also, when the glass rods were removed from the cages they were heavily coated in a white "sticky" residue. As captive harvest mice have been observed to primarily scent mark on objects suspended above ground, if the primary source of secretion is the paws, the rodents may only secrete during the climbing process which is why the paw washes didn't look to contain as much protein as the other washes and saliva.

Three abundant proteins were identified in both male and female harvest mice by ESI-MS – 16437 Da, 16724 Da and 17888 Da. Unfortunately a purified sample of protein 17888 Da was unable to be collected by AEX and as a consequence this protein was not *de novo* sequenced. There is evidence that the other two proteins are lipocalins as they contain the characteristic conserved lipocalin sequence motif G-X-W. The proteins were sequenced using PEAKS 6 software. Overall the sequence data produced by PEAKS was of high quality with the majority of ALC scores above 90% with a couple of sequences scoring just above 70%. This is a combination of both the unique algorithms used by PEAKS and using the right mass analyser to produce high quality raw data. There were a couple of MS/MS spectra

that upon visual inspection looked difficult to interpret due to the lack of b and y ions. Although PEAKS managed to sequence these with high confidence, care should be taken with the data refinement step where PEAKS removes low quality data as it is possible that too many ions were removed leading to sparse looking spectra. It could be that the parameters were set too high and other peptide spectra were poorly sequenced and therefore discarded because of the 55% cut off.

A database search was also enabled in the PEAKS processing method and there were many matches to mouse MUPs (Figure 4.25). The two dominant proteins secreted by harvest mice are odorant binding proteins with no evidence of MUPs in the AEX fractions so are these MUP sequence matches true? PEAKS database searching uses a series of unique authenticated algorithms to assign a peptide to a protein and then validates the result. PEAKS firstly use the *de novo* sequence tags to find approximate matches in the protein database. All proteins in the database are evaluated according to the sequence tag matches. The top 7000 proteins are used to make a protein shortlist. All peptides in the protein shortlist are used to match MS/MS spectra using a rapid scoring function. The top 512 highest scoring peptide candidates are kept for each MS/MS spectra. A precise scoring function is then used to find the best peptide for each spectrum from the 512 peptides calculated in the peptide shortlisting step. The similarity between the de novo sequence and the database peptide is an important component in the scoring function. The score is also normalized to ensure it can be compared across different spectra (Zhang *et al.,* 2012).

Like most software, PEAKS peptide identification is statistically validated to avoid false positives. The most accepted result validation method is through a false discovery rate (FDR). FDR is defined as the ratio between the false peptide matching spectrums and the total number of peptide matching spectrums above the score threshold. The threshold score is user defined and by adjusting the score thresholds, the result accuracy (FDR) can be traded with the sensitivity (number of reported identifications). Different software equipped with different scoring functions may have significant different trade off efficiencies. A comparison between MASCOT, SEQUEST and PEAKS demonstrated PEAKS performs slightly

MUP 20 (darcin)

```
EEASSMERNF NVEKINGEWY TIMLATDKRE KIEEHGSMRV FVEYIHVLEN SLALKFHIII NEECSEIFLV ADKTEKAGEY
              LNGDWF SLLTASEKRE KLEENGSMR     NLHVLEN SLAFKLHTVV NGECTELFLV ADK

SVTYDGSNTF TILKTDYDNY IMIHLINKKD GETFQLMELY GREPDLSSDI KEKFAQLSEE HGIVRENIID LTNANRCLEA
    DGENTF SLLKTDYDNY LMFHVTNVKD GETFQLMELY GREPDLSSDL K  FLDDYEE HGLLRENLLD LSNANRCLEA

RE
RE
```

MUP 4

```
EEATSKGQNL NVEKINGEWF SILLASDKRE KIEEHGSMRV FVEHIHVLENS LAFKFHTVID GECSEIFLVA DKTEKAGEYS
              LAGNWY SLLLASDKRE KLEEDGSMRV FVEYLHVLGNS LAFKFHTVVN GECTELFLVA DK

VMYDGFNTFT ILKTDYDNYI MFHLINEKDG KTFQLMELYG RKADLNSDIKE KFVKLCEEHG IIKENIIDLT KTNRCLKARE
              TDYDNYL MFHVTNVK    TFQLMELYG R               LCEEHG LLRENLLDLS NANR
```

Figure 4.25 Harvest mouse peptides from the discovery run aligned with MUP 20 and MUP 4.
Harvest mouse peptides from PEAKS database search aligned with mouse MUP sequences. MUP sequences are highlighted in black, harvest mouse sequences that exactly match the MUP sequence are illustrated in red, amino acid differences are highlighted in blue. PEAKS defaults to leucine for all leucine and isoleucine residues. Using the current LC-MS system, it is not possible to distinguish between the two residues due to their isobaric nature.

better when all three software's had an FDR set to exactly the same threshold, PEAKS identified the most peptide matching spectrums (Zhang *et al.,* 2012). PEAKS has an estimate FDR with decoy fusion option which enables search result validation with an enhanced target-decoy approach. Decoy sequences are automatically generated from a target database and are searched by PEAKS. This enables the estimation of the false discovery rate in the analysis report.

Peptides that were *de novo* sequenced were also BLAST searched to find any similarities between sequences. Many sequences shared a high percentage homology with other lipocalins observed in other rodents. For each peptide an E value and score is given. The E value illustrates the number of hits you can expect to see by chance when searching a database of a particular size. It exponentially decreases as the score value increases. The lower the E value the more significant the match is although the length of the sequence is taken into account. Shorter sequences often have higher E values because they have a higher probability of occurring in the database by chance. The score value gives an indication of how good the alignment is; the higher the score, the better the alignment. The score is calculated from a formula that takes into account the alignment of similar or identical residues, as well as any gaps introduced to align the sequences.

Preliminary behavioural studies, carried out by technical staff at the University of Liverpool, Leahurst, testing responses to glass rods show the harvest mice have no attraction to their own odour but respond to scents from unfamiliar rodents. Harvest mice were exposed to a control – two clean glass rods – which they showed little interest in. They were then exposed to a glass rod that had been removed from their cage and contained their own scent and a clean rod. The rodents were attracted to neither rod and didn't spend any more time near their own rod versus the clean rod. The third test was to expose the animal to another two rods - one rod from the cage of another rodent of the same sex and a clean rod. The fourth test was to again expose the rodent to two rods – one from the cage of an unfamiliar rodent of the opposite sex and a clean rod. In both tests, rodents spent a significant amount of time near the rods containing the unfamiliar scent, with slightly stronger responses to male odour rods – males exposed to rods from other

males and females exposed to rods from males. These behavioural tests and the data collected in this chapter, was done using rodents that had been transferred from the outdoor enclosure to cages indoor. It is unknown how this affects the rodents behaviourally and whether or not it effects protein expression of the lipocalin proteins. Although the cages are set up to closely mimic an outdoor habitat, it would be interesting to repeat these behavioural tests on rodents that have just been captured from the enclosure and compare them to rodents who have been living indoors for a number of weeks. Alongside this, collecting urine, saliva and washes would be useful to monitor potential protein expression differences.

What exactly the mice are responding to in the behavioural experiments is unknown. It would be beneficial to complete the *de novo* sequencing of the two abundant proteins so a recombinant form of each could be made for further behavioural studies. Alternative sequencing methods such as electron transfer dissociation (ETD) may result in more sequence coverage. ETD fragmentation causes rapid cleavage of the peptide backbone via the transfer of an electron produced by a radical anion (e.g. Flouranthene) resulting in c and z ions (Figure 1.6) (Syka *et al.,* 2004). One of the advantages of ETD is the ability to fragment larger peptides. In the case of the harvest mouse proteins there was a section (amino acids 54-66, figure 4.14 and amino acids 45-69, figure 4.15) where no sequence data was collected. This could have been due to a lack of protease cleavage sites resulting in large peptides that would be difficult to fragment using CID. ETD fragmentation could therefore provide sequence information for these parts of the protein. Alternatively top-down ETD using the intact protein (the glass rod wash fractions) could also be used to gain further sequence information.

The numbers of wild harvest mice have fluctuated quite dramatically over the last 10 years. Predation, harsh winters and the rise in the housing development projects have caused numbers to drop to a threatened status on some occasions. Various short term conservation projects including the recycling of tennis balls, to use as harvest mouse nests, after the Wimbledon tennis championship in 2001 have seen numbers of wild harvest rise. It is anticipated that this work will contribute to

long term conservation projects in a similar manner to the pest control projects discussed in chapter 3. By understanding the behaviour and communication between these rodents, it may be possible to manipulate their behaviours to encourage them to live in safer habitats in the wild. Also, a greater understanding into how they reproduce would be advantageous so if numbers should fall again it will be possible to promote successful reproduction strategies.

Chapter 5: Seasonal expression of urinary proteins in the male mouse lemur (*Microcebus*)

5.1 Introduction

Mouse lemurs (*Microcebus)* are small nocturnal primates native to Madagascar. With a total length of approximately 11 inches, they are the world's smallest primate. At present there has been 19 species of mouse lemur identified (Mittermeier et al, 2010; Radespiel *et al.,* 2012). Each species vary little in their physical features such as their size and phenotypic traits. This chapter will focus on two of the mouse lemur species – *Microcebus murinus* and *Microcebus lehilahytsara.*

Mouse lemurs are social animals although they prefer to forage alone (Bearder, 1987). Their social pattern can vary depending on gender and season. Mouse lemurs regularly interact with their conspecifics and establish steady home ranges which often overlap (Radespiel 2000; Weidt *et al.,* 2004). Male home ranges are often larger than females and often increase in size during the mating season as a possible strategy to improve mating success (Schmelting, 2000; Schmelting *et al*, 2000). Females prefer to form stable restricted matrilineal sleeping groups (Radespiel *et al.,* 2001; Lutermann *et al.,* 2006; Jurges *et al.,* 2013) while males are frequently found sleeping alone (Radespiel *et al.,* 1998; Schmelting, 2000) although during the reproductive season can be found in female nesting sites.

The mouse lemur mating system can be described as multi male/multi female (Fietz, 1999). They have defined breeding seasons, the onset of which is triggered by seasonal changes and the length of daylight. Female promiscuity leads to sperm competition in males, (Karr and Pitnick, 1999) who often establish dominance hierarchies prior to the beginning of the reproductive season (Perret, 1992). Physiological changes can be seen in males to cope with this evolutionary pressure and improve their chances of reproductive success. In *M. murinus* changes such as increased body mass and testis size have been observed, changes which can be

subsequently reversed in subordinate males after exposure to urine from a dominant male (Perret and Schilling 1987; Perret and Schilling, 1995). As females are the dominant sex they ultimately decide whether mating will take place (Radespiel and Zimmermann, 2001). Females will accept or refuse to mate depending on her reproductive interests. At present reasons for female mate choice are poorly understood.

In mice and rats chemical communication is well documented (Beynon and Hurst, 2004; Robertson *et al.,* 2007; Beynon *et al.,* 2007; Roberts *et al.,* 2010; Roberts *et al.,* 2012). Both have functional vomeronasal receptors, VR1 and VR2, for detecting volatiles and non volatiles respectively (Krieger *et al*, 1999; Sugai *et al.,* 2006). VR1 genes are found in most mammals with large variation and diversity between taxonomy (Grus *et al.,* 2005; Grus *et al.,* 2007). Until recently, intact VR2 genes were thought to be limited to rodents and marsupials. However two intact VR2 genes have now been identified in the gray mouse lemur (*M. murinus*) with expression established in the vomeronasal organ (Hohenbrink *et al.,* 2012). This is particularly interesting as VR2 receptors in rodents bind non-volatiles such as MUPS which are used to communicate a variety of information such as health, relatedness and reproductive status (Beynon and Hurst, 2003).

5.2 Aims and objectives

This chapter will focus on examining the urine content of *Microcebus murinus* and *Microcebus lehilahytsara* with the aim of identifying and characterising any potential proteins that may be used in scent communication. The objectives of the study were:

- To observe both male and female mouse lemur urine in 2 species of mouse lemur - *Microcebus murinus* and *Microcebus lehilahytsara.* As mouse lemurs have a specific breeding season, urine samples were taken both in and out of season.
- Identify any differences between species, sexes and season.

- Characterise and sequence proteins of interest using mass spectrometric techniques.

5.3 Results and discussion

5.3.1 Identification of a sex-specific protein in male mouse lemurs (*Microcebus*)

Urine samples were collected from captive male and female mouse lemurs (*M. murinus and M. lehilahytsara*) during breeding and non-breeding season. The urine (10 µl) was analysed by 1D SDS-PAGE. In both species, a single protein band at approximately 10 kDa was identified in some of the male mouse lemur urine samples collected during the reproductive season (Figure 5.1, Figure 5.2). This protein was not present in the samples collected out of season. No dominant protein was identified in the female mouse lemur urine samples.

Protein and creatinine measurements were taken to assess the concentration of protein in the urine of the male mouse lemurs. Protein concentration varies with the volume of urine excreted. To correct for urine dilution, creatinine levels were also measured. Creatinine is a breakdown product of creatine phosphate which is used in skeletal muscle contraction. The daily production of creatinine is dependent on muscle mass which fluctuates very little so the amount of creatinine produced remains fairly constant. Therefore measuring the protein: creatinine ratio provides an appropriate correction for urine dilution. The protein concentration varied between males, ranging from 0.6 mg/ml – 1.6 mg/ml (Figure 5.3 and Figure 5.4).

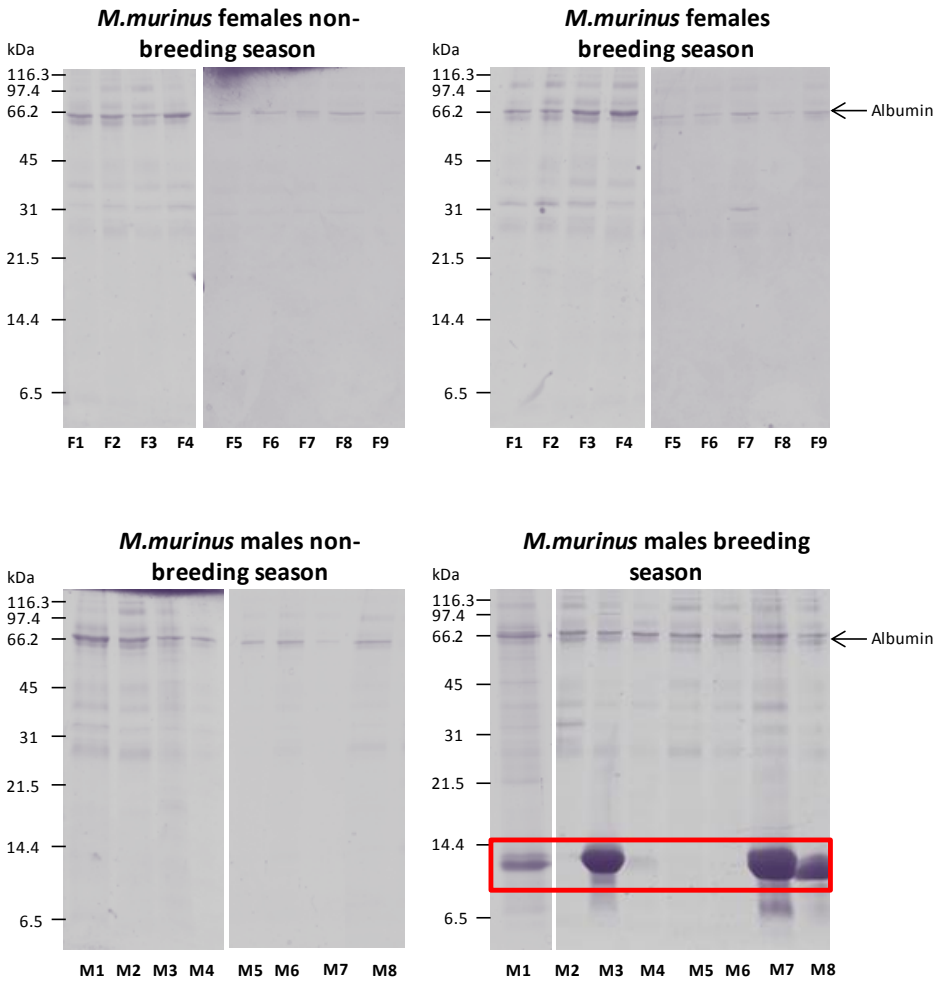


Fig 5.1 SDS-PAGE analysis of male and female *Microcebus murinus* urine samples. Urine samples were taken from both male and female mouse lemurs during both the reproductive and non-reproductive season. Urine (10 µl) was mixed 1:1 with samples buffer and run on a 15% SDS gel. Gels were stained with coomassie blue stain. The red box highlights a protein expressed in certain male mouse lemurs during the breeding season. Creatinine is only 113 Da so is not present on the gel.

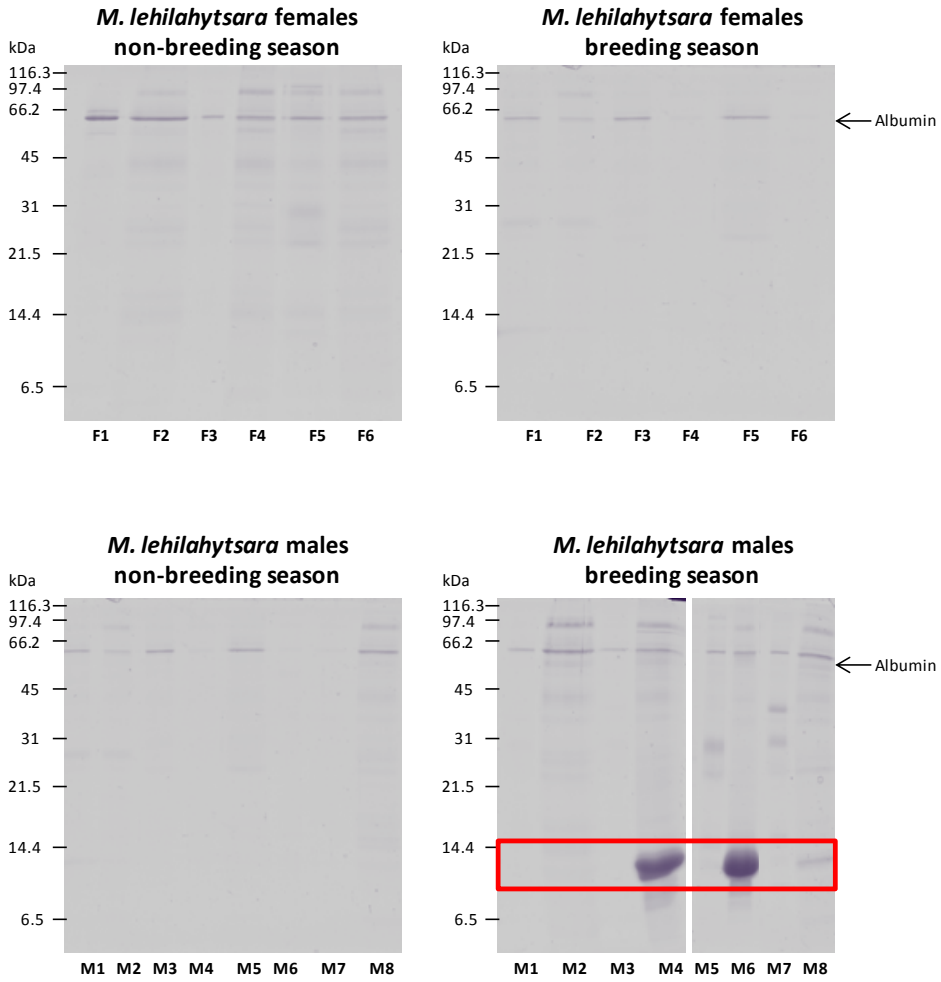


Fig 5.2 SDS-PAGE analysis of male and female *Microcebus lehilahytsara* urine samples.
Urine samples were taken from both male and female mouse lemurs during both the reproductive and non-reproductive season. Urine (10 µl) was mixed 1:1 with samples buffer and run on a 15% SDS gel. Gels were stained with coomassie blue stain. The red box highlights a protein expressed in certain male mouse lemurs during the breeding season.

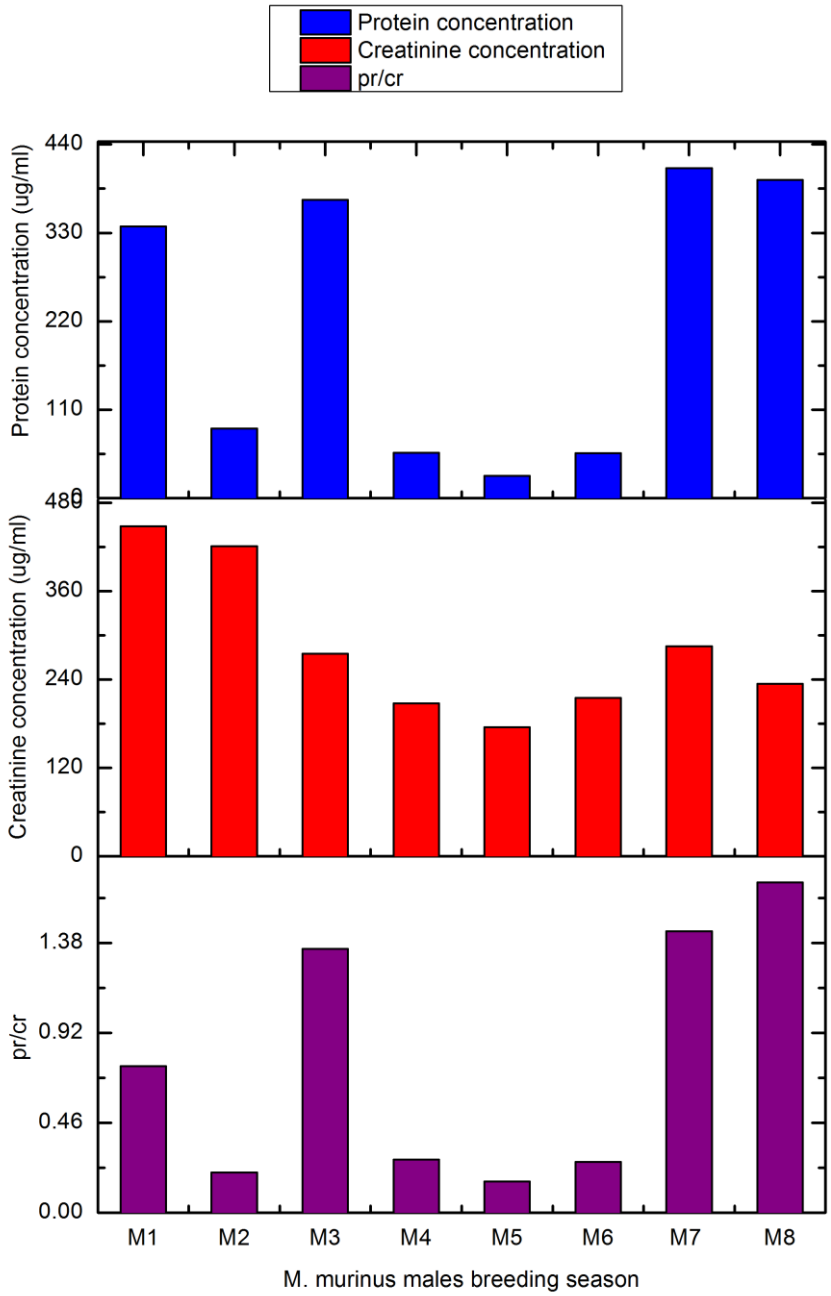


Figure 5.3 Determination of protein concentration in *M. murinus* males during the breeding season.

Protein concentration was determined using a Coomassie Plus protein assay kit. Bovine serum albumin was used to prepare a standard curve (0-50 µg/ml). Samples were diluted down in purified water to be in the linear range of the assay. Absorbance readings were measured at 620 nm using a plate reader. Creatinine concentration was measured using a creatinine assay kit. A creatinine standard curve was prepared (0-30 µg/ml) and samples diluted down in purified water to be in the linear range of the assay. Absorbance readings were measured at 570 mm. The protein: creatinine ratio was then calculated to correct for urine dilution.

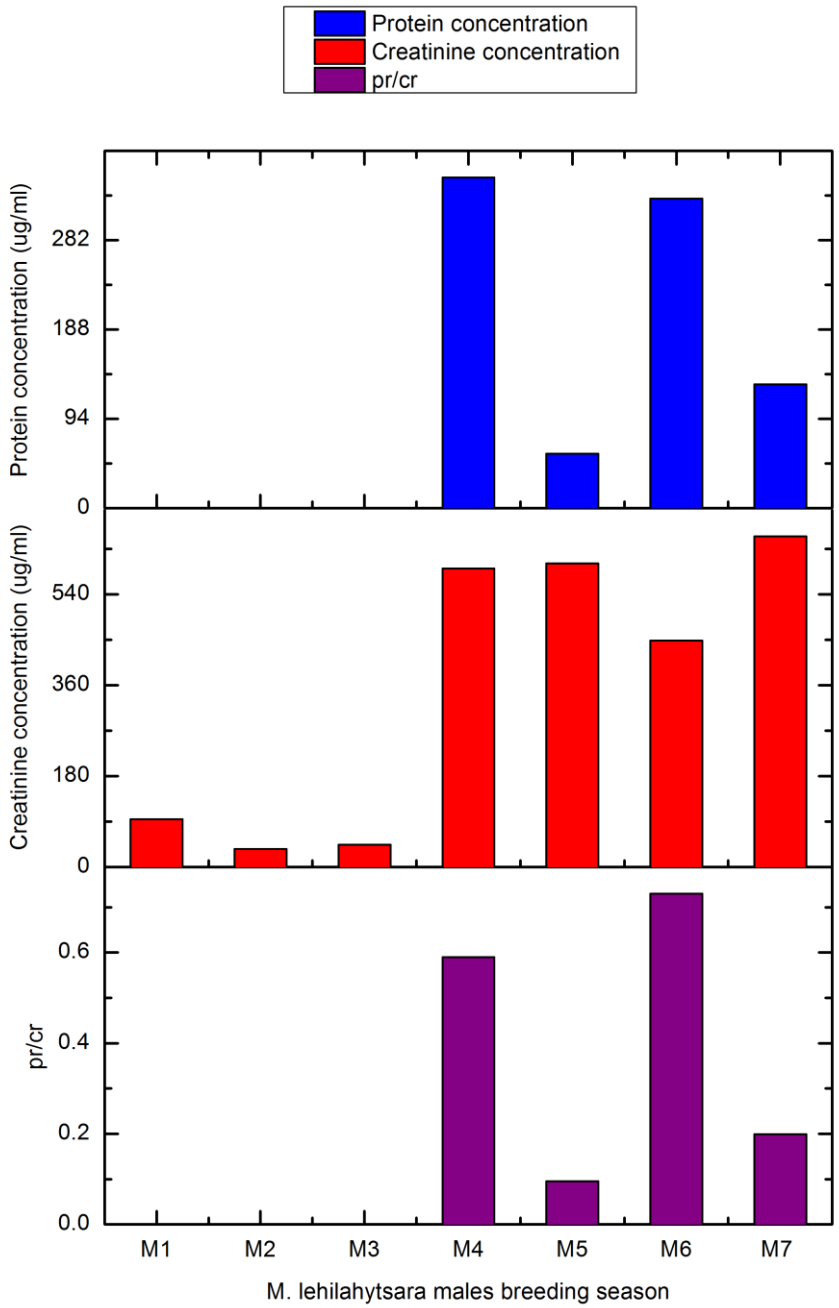


Figure 5.4 Determination of protein concentration in *M. lehilahytsara* males during the breeding season.

Protein concentration was determined using a Coomassie Plus protein assay kit. Bovine serum albumin was used to prepare a standard curve (0-50 µg/ml). Samples were diluted down in purified water to be in the linear range of the assay. Absorbance readings were measured at 620 nm using a plate reader. Creatinine concentration was measured using a creatinine assay kit. A creatinine standard curve was prepared (0-30 µg/ml) and samples diluted down in purified water to be in the linear range of the assay. Absorbance readings were measured at 570 mm. The protein:creatinine ratio was then calculated to correct for urine dilution.

5.3.2 Determination of an accurate molecular weight

To obtain a more accurate molecular weight, urine from the male mouse lemurs' was analysed by electrospray (ESI) mass spectrometry. A protein with a molecular weight of 9388 Da was identified in the male *M. murinus* urine samples (Figure 5.5). A second peak 16 Da heavier was also detected. This was either a second protein or a modification on one of the amino acids (methionine) that make up the protein. This would have to be confirmed by *de novo* sequencing. The masses were consistent in each mouse lemur. Analysis of the male *M .lehilahytsara* urine revealed a mass with a slightly different molecular weight of 9418 Da. Again this mass was consistent in each of the male *M. lehilahytsara* samples and had a second peak present in the chromatogram 16 Da heavier (Figure 5.6).

5.3.3 Peptide mass fingerprinting

To investigate the differences between the two species, an in-gel digest of the protein bands of interest (from the SDS-PAGE analysis) was completed. Following overnight incubation with trypsin the digests were analysed by MALDI-TOF to produce a peptide mass fingerprint (Figure 5.7). The peptide masses produced were compared to protein databases (swissprot) and statistically analysed to see if there were any matches. No matches were identified for the protein band in either species.

A comparison of the two species highlighted a peptide (991 m/z) in the *M. lehilahytsara* samples that was 30Da heavier than a peptide seen in *M. murinus* samples (961 m/z) which was consistent with the mass difference in the intact mass spectra (5.3.2). Another peptide, 2979 m/z, which had a +16 Da equivalent at 2995 m/z, was also identified (Figure 5.8). Many of the other major peptides were the same in each species suggesting that it was the same protein except for some amino acid mutations in the 961/991 m/z peptide to account for the 30Da difference.

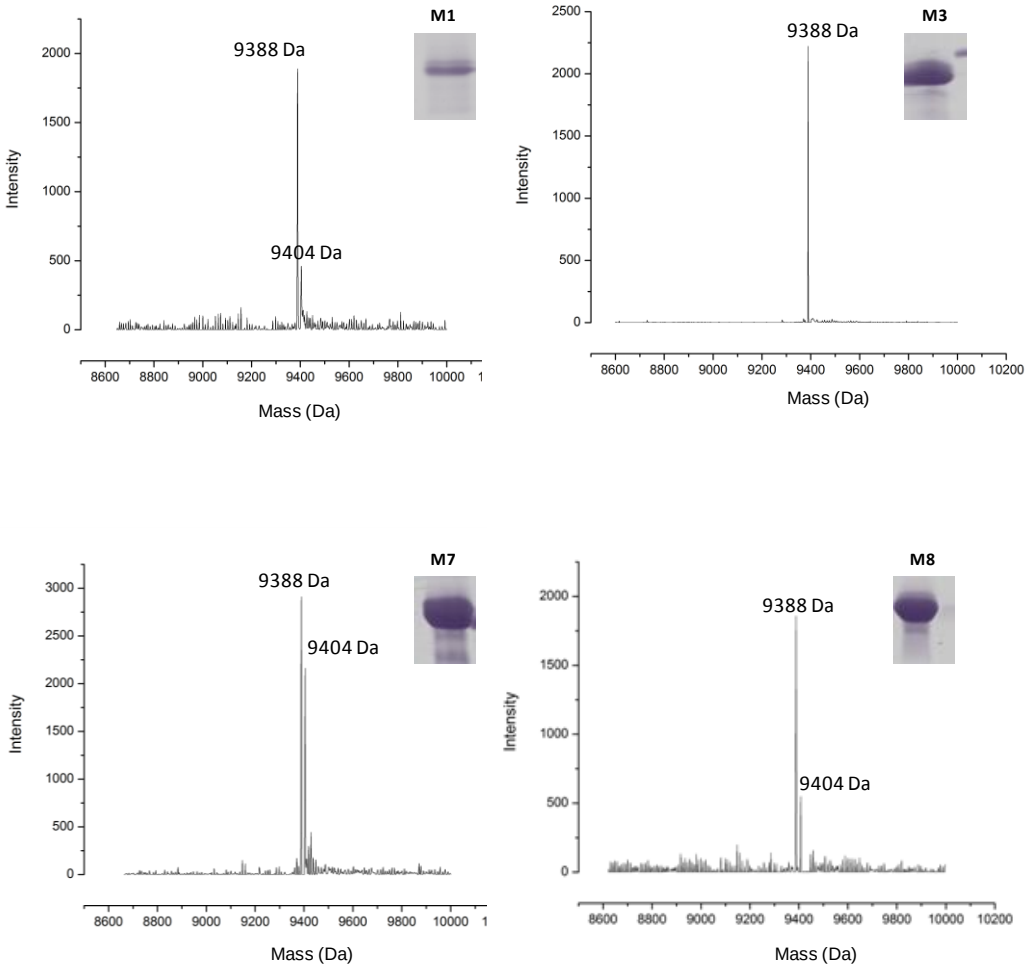


Fig 5.5 Determination of the molecular weight of the male specific urinary protein in the *Microcebus murinus* urine samples.
Urine samples containing substantial amounts of protein (from SDS-PAGE analysis) were diluted into formic acid (0.1 %) to produce a final concentration of approximately 5 pmol/µl. The samples were then injected onto a C4 desalting trap and masses of proteins present were determined by ESI-MS. Data was processed using maximum entropy software MAX ENT-1 (MassLynx 4.1, Waters).

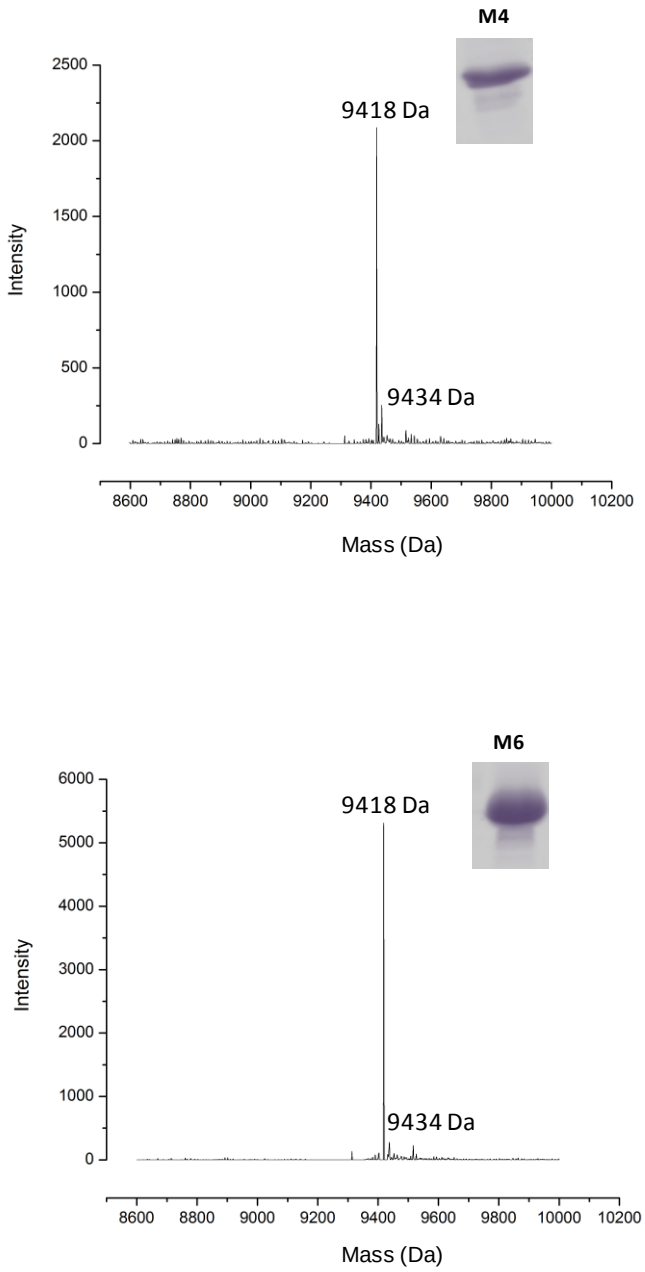


Fig 5.6 Determination of the molecular weight of the male specific urinary protein in the *Microcebus lehilahytsara* urine samples.
Urine samples containing substantial amounts of protein (from SDS-PAGE analysis) were diluted into formic acid (0.1 %) to produce a final concentration of approximately 5 pmol/µl. The samples were then injected onto a C4 desalting trap and masses of proteins present were determined by ESI-MS. Data was processed using maximum entropy software MAX ENT-1 (MassLynx 4.1, Waters).

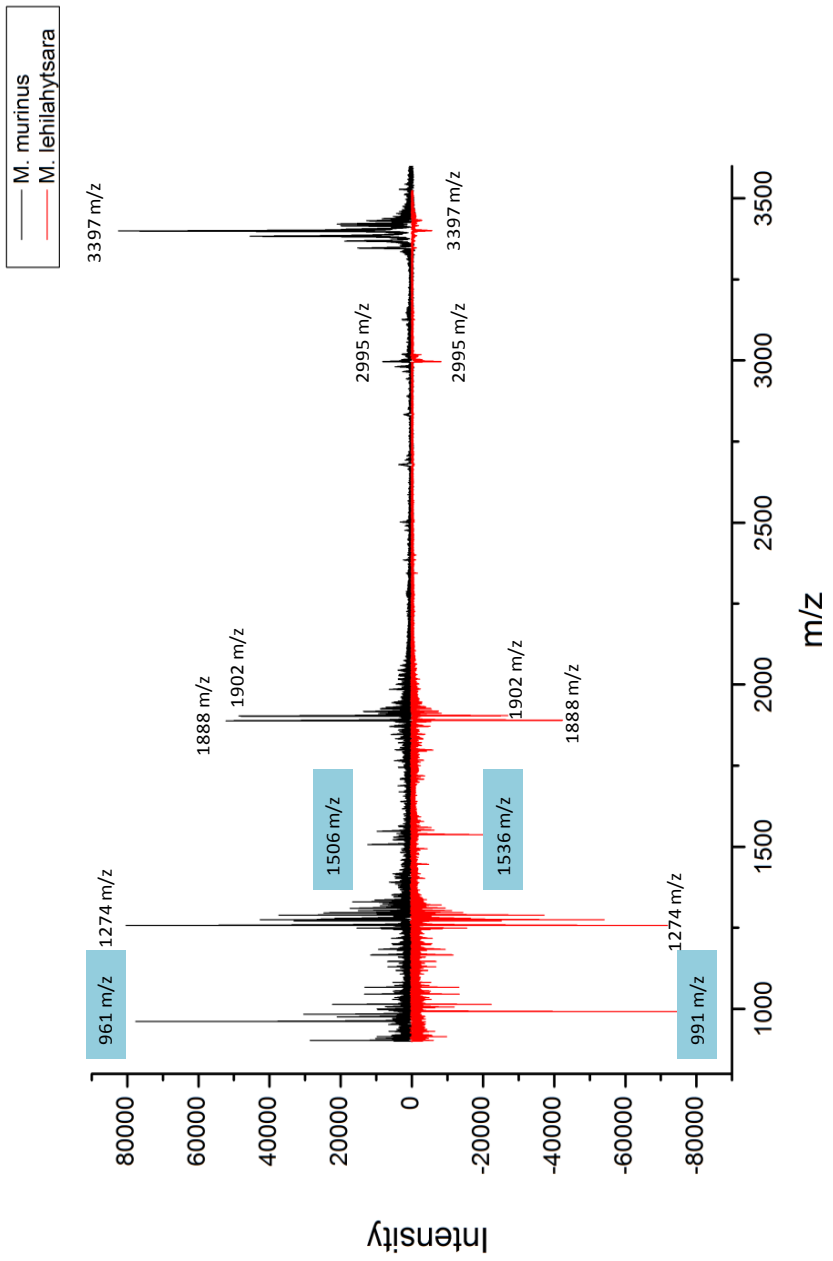


Fig 5.7 Peptide mass fingerprinting.
Small pieces of gel was extracted from the protein bands of interest from the SDS-PAGE analysis of *M. murinus* and *M. lehilahytsara* male urine samples. These pieces of gel were destained in 50:50 ACN:NH_4CO_3 before being reduced and alkylated in DTT (10 mM) and iodoacetamide (60 mM) respectively. Following overnight incubation at 37 °C with trypsin, the peptides were collected and mixed 1:1 with α-Cyano-4-hydroxycinnamic acid dissolved in 50% ACN, 0.1% TFA. The mixture (1 µl) was spotted onto a target plate and left to dry at room temperature before being analysed by MALDI-TOF. The 30 Da difference between the masses identified in each species was localised to one peptide – 961 m/z in the *M. murinus* samples and 991 m/z in the *M. lehilahytsara* samples (highlighted in pink). There is a partial missed cleavage of this peptide at 1506 m/z (*M. murinus)* and 1536 m/z (*M. lehilahytsara)* which was later proved by de novo sequence analysis (section 5.2.4).

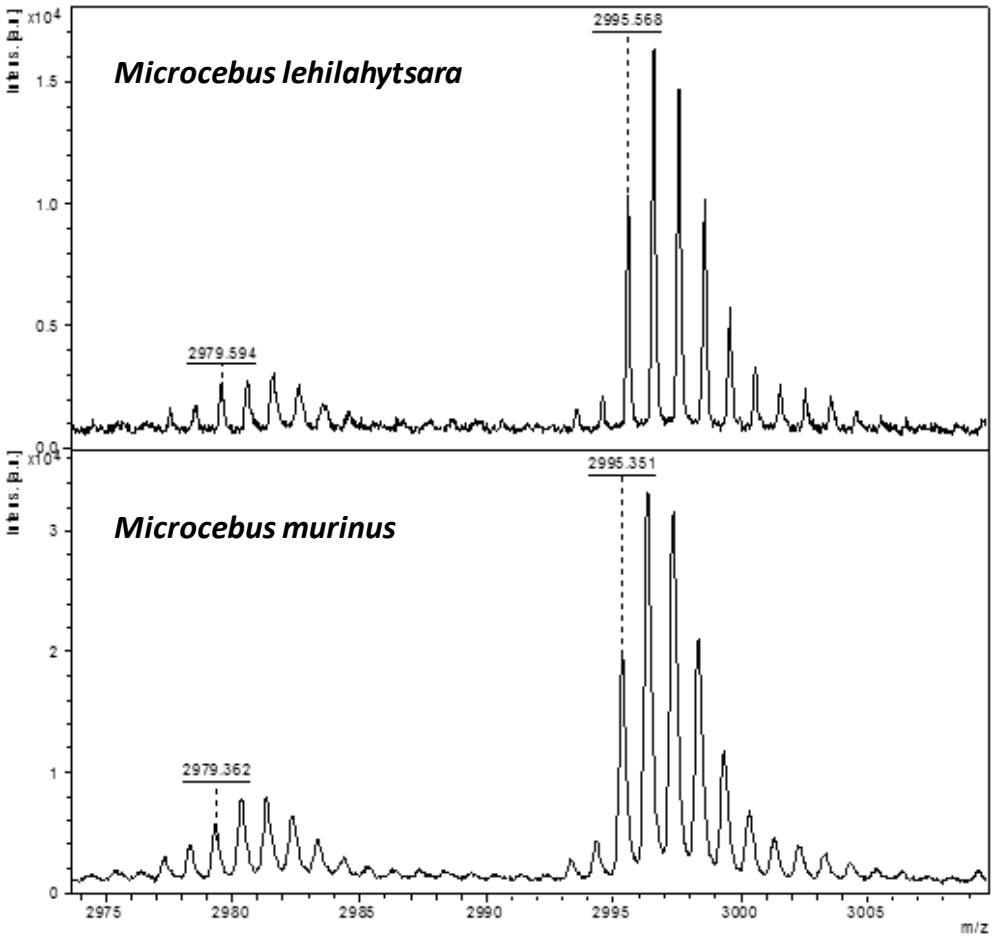


Fig 5.8 Identification of the +16 Da adduct observed in the ESI-MS data.
Small pieces of gel was extracted from the protein bands of interest from the SDS-PAGE analysis of *M. murinus* and *M. lehilahytsara* male urine samples. These pieces of gel were destained in 50:50 ACN:NH_4CO_3 before being reduced and alkylated in DTT (10 mM) and iodoacetamide (60 mM) respectively. Following overnight incubation at 37 °C with trypsin, the peptides were collected and mixed 1:1 with α-Cyano-4-hydroxycinnamic acid dissolved in 50% ACN, 0.1% TFA. The mixture (1 µl) was spotted onto a target plate and left to dry at room temperature before being analysed by MALDI-TOF. A peptide was detected at 2979 m/z with a +16 Da adduct at 2995 m/z in both species of mouse lemur.

Samples of the gel band were also taken and digested with two other proteases (endoproteinase LysC and endoproteinase GluC) that cleave the protein at different sites. These PMFs together with the trypsin ones were used to support the sequence evidence found from the *de novo* sequencing analysis.

5.3.4 *De novo* sequencing analysis

To obtain amino acid sequence information, urine (containing the protein of interest) from both species was digested with three different proteases – trypsin, LysC and GluC – to produce overlapping sequence information due the specificity of each enzyme. The digested urine was analysed by LC-MSMS and the raw data was *de novo* sequenced using PEAKS 6 software for proteomics (Table 5.1).

Table 5.1 Peptide sequences identified from PEAKS *de novo* analysis. Raw MSMS data from each digest was analysed using PEAKS software for proteomics.

Sequence	Species	Protease	Mass (Da)
WGNCPAEK	*M. murinus*	Trypsin LysC	960.41 960.41
VKGGKEKWGNCPTE	*M. lehilahytsara*	GluC	1588.76
WGNCPTEK	*M. lehilahytsara*	Trypsin LysC	990.41 990.41
SGPSQCHSDNDCPGDKK	*M. murinus* *M. lehilahytsara*	Trypsin LysC	1887.74 1887.74
CCFLHCSYK	*M. murinus* *M. lehilahytsara*	Trypsin LysC	1273.50
CVSPER	*M. murinus* *M. lehilahytsara*	Trypsin	746.33
CVSPERNRK	*M. murinus*	LysC	1144.57
EGLGQMAPVLE	*M. murinus*	GluC	1158.50
TWNVGQVGQE	*M. murinus* *M. lehilahytsara*	GluC	1116.52
QGAPDTWNVPVADTWNVGQVGQEASPQK	*M. murinus*	LysC	2978.41

The peptide sequences were BLAST searched (http://blast.ncbi.nlm.nih.gov). A number of sequences showed high similarity to a Whey Acidic Protein (WAP) identified in the ring tailed lemur (*Lemur catta*), (Table 5.2).

Table 5.2 BLAST results of mouse lemur sequences obtained from LC-MS analysis. Sequences for both species of lemur were assessed using the blastP algorithm. High sequence homology with WAP 4 – disulphide core domain 12 (*Lemur catta*) was observed with a number of peptides. Search parameters were restricted to mammals.

Mouse lemur sequence	Protein identification	Score	E value	Sequence Identity
WGNCPAEK	WAP 4 –disulphide core domain 12 (*Lemur catta*)	31.2	3.9	100%
WGNCPTEK	WAP 4 –disulphide core domain 12 (*Lemur catta*)	31.2	3.9	100%
SGPSQCHSDNDCGPDKK	WAP 4 –disulphide core domain 12 (*Lemur catta*)	41.4	0.004	81%
CCFLHCYSK	WAP 4 –disulphide core domain 12 (*Lemur catta)*	20.2	0.29	66%
CVSPER	WAP 4 –disulphide core domain 12 (*Lemur catta)*	20.6	0.16	83%
VKGGKEKWGNCPTE	WAP 4 –disulphide core domain 12 (*Lemur catta)*	48.6	2E-10	93%
EGLGQMAPVLE	WAP 4 –disulphide core domain 12 (*Lemur catta)*	28.6	0.82	80%
TWNVGQVGQE	WAP 4 –disulphide core domain 12 (*Lemur catta)*			
QGAPDTWNVPVADTWNVGQVGQEASPQK	WAP 4 –disulphide core domain 12 (*Lemur catta)*	38.8	0.001	64%

5.3.5 Determination of the mouse lemur protein sequence

The *L. catta* protein, full name WAP 4 –disulphide core domain 12 (WFDC12), was therefore used to align further MS/MS sequences (Figure 5.9). The *L. catta* protein was used to construct the mouse lemur sequence as many of the peptides share high homology to this protein. The first section of the protein (amino acids 1-55 on figure 5.9) was relatively straightforward to sequence and align with the *L. catta* protein. These peptides both ionised and fragmented well in the LC-MS analysis making sequencing easier. No sequence information was discovered that could be aligned with the middle section of the ring tailed lemur protein (greyed out on figure 5.9). This is unsurprising giving the intact mass identified the mouse lemur

Fig 5.9 Sequence alignment of the mouse lemur peptide sequences with WAP 4 –disulphide core domain 12 (*Lemur catta).*

Peptide sequences obtained from the PEAKS de novo sequencing analysis were BLAST searched and shared high homology with WAP 4 –disulphide core domain 12 *(Lemur catta*). Peptide sequences from all 4 digests – trypsin , LysC, GluC and AspN were aligned with the *L. catta* protein. No evidence was found to suggest the middles section if the *L. catta* protein (greyed out) exists in the mouse lemur protein. The last section of the protein, amino acids 97-125, was pieced together using a combination of proteases. The sections of peptides highlighted in orange were the parts of the peptide that scored poorly in PEAKS and were therefore discarded when putting the mouse lemur sequence together. The letter X symbolises the 30 Da difference between species (section 5.2.6). PEAKS 6 defaults to leucine for all leucine and isoleucine residues. Using the current LC-MS system, it is not possible to distinguish between the two residues due to their isobaric nature.

protein to be approximately 9.4 kDa compared to the *L. catta* protein which is 13.2 kDa. The C terminus of the peptide was thought to be peptide 2979 m/z previously identified in the PMF as having a possible +16 Da modification. This was confirmed by subtracting away the sequenced peptides from the intact mass to give 2978 Da. This is a large peptide and would struggle to ionise and fragment leading to poor *de novo* data. This peptide was identified in one of the *M. murinus* LysC digests and PEAKS did attempt to sequence it - QGAPDTWNVPVADTWNVGQVGQEASPQK. The first section of the sequence QGAPDTWNVPVA had lower scores for the individual amino acids in PEAKS as the raw spectra was difficult to interpret. The last part of the sequence DTWNVGQVGQEASPQK had improved fragmentation and was therefore easier to sequence leading to increased confidence scores by PEAKS. It also shared high homology with the *L. catta* protein. A partial piece of the second part of sequence was confirmed in a GluC digest TWNVGQVGQE. GluC appears to have cleaved at an aspartic acid residue which is possible as GluC can also cleave at aspartic acid residues at a rate of 100-300 times slower than at glutamic acid residues. To confirm this and also gather sequence data for the poorly sequenced section of 2979 m/z an additional digest using endoproteinase AspN (Asp-N) was completed. The raw LC-MS data was analysed by PEAKS proteomics software (Table 5.3).

Table 5.3. AspN sequences to support C terminal sequence data collected from digests with alternative proteases. Raw MSMS data from each digest was analysed using PEAKS software for proteomics.

Peptide	Species	Protease	Mass (Da)
EGLGQMAPVPQGA	*M. murinus*	AspN	1253.60
EGLGQMAPVPGAQ	*M. lehilahytsara*	AspN	1253.60
DTWNVGQVG	*M. murinus* *M. lehilahytsara*	AspN	1102.50
EASPQKEWS	*M. murinus* *M. lehilahytsara*	AspN	1060.48

The peptide at 1103.50 m/z confirmed the residue to be an aspartic acid. Cleavage at the N terminal side of glutamic acid identified a second peptide at 1061.58 m/z. Both of these peptides confirmed the high scoring section of the peptide 2979 m/z observed in the LysC digest – DTWNVGQVGQEASPQK.

A third peptide 1126.54 m/z was also observed in the AspN digest and shared high homology with the *L. catta* protein. There was a slight difference between species, EGLGQMAPVPQGA in the *M. murinus* digest and EGLGQMAPVPGAQ in the *M. lehilahytsara* digest. The confidence scores were much higher in the *M. murinus* digest so this was the peptide used to assemble the sequence. It also agreed with the high scoring section of the peptide 1159.50 m/z in the GluC digest - EGLGQMAPVLE. The last two residues were given a slightly lower confidence values by PEAKS and did not align with the *L. catta* protein. The rest of sequence scored highly and did align with the *L. catta* protein.

This peptide would account for the first section of the LysC peptide 2979 m/z that was poorly sequenced. As there was an internal aspartate residue in the 2979 m/z peptide it could be cleaved into two smaller peptides by AspN making fragmentation more effective. Using a combination of three sequences from the three different proteases the final sequence of the C terminus was determined as GLGQMAPVPQGADTWNVGQVGQEASPQKEWS. There is also a methionine residue present in this 2979 m/z peptide which would explain the additional peak detected at 16Da heavier on the intact mass spectrum in both species. The final mouse lemur sequence is illustrated in figure 5.10. A summary of all peptides used to determine the mouse lemur sequences are highlighted in Table 5.4 and 5.5 and examples of MS/MS spectra are illustrated in figures 5.11-5.18. The remaining MS/MS spectra can be found in supplementary data C.

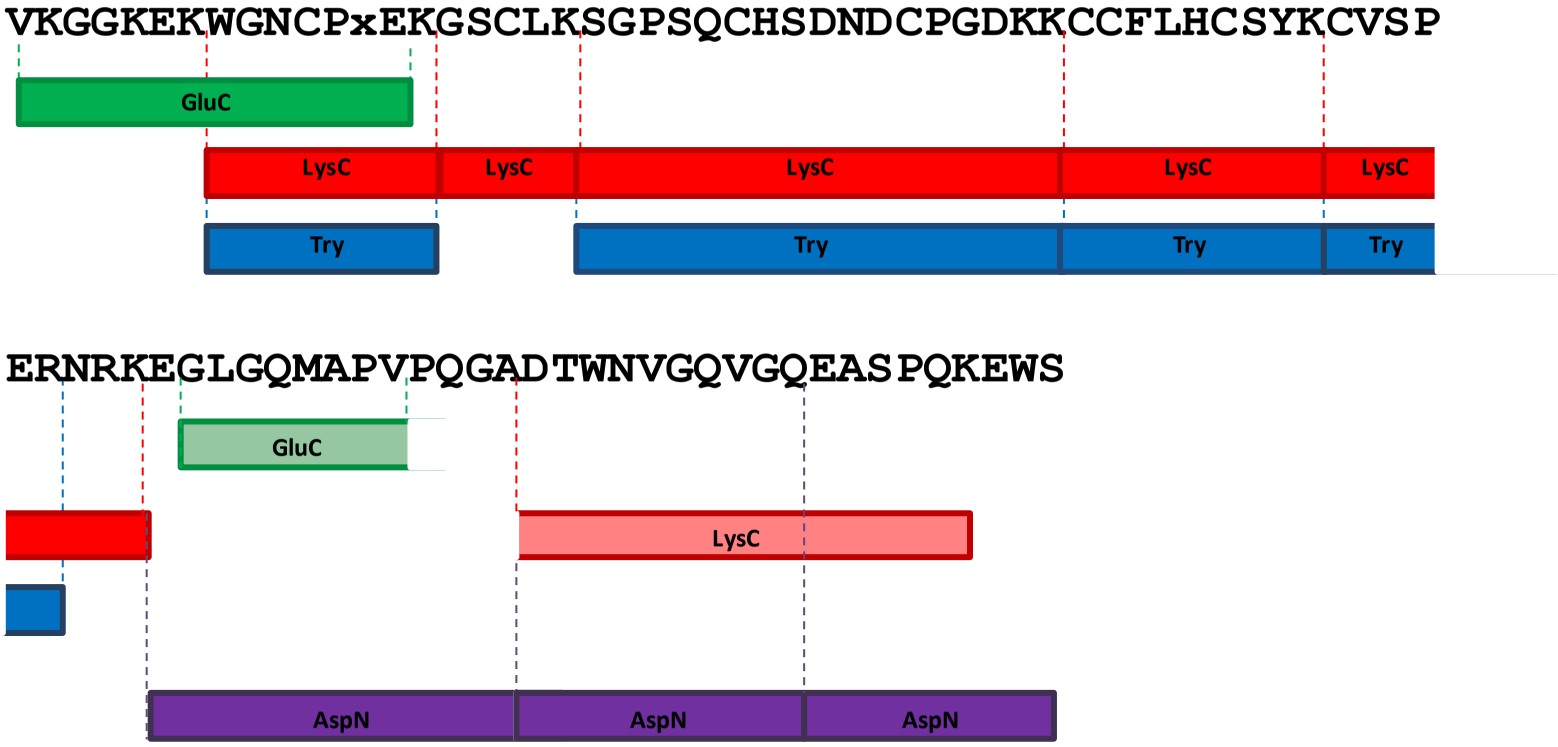


Fig 5.10 Peptide map of the final mouse lemur sequence.
Most of the sequence can be confirmed by two different proteases. The solid, bold coloured blocks are where the sequence was confirmed using full peptide sequences. Some peptides had sections of high scoring residues and sections of low scoring amino acids in the PEAKS analysis depending on how good the raw data was. The high scoring sections of these peptides have been aligned with the sequence on this map and are highlighted using a fainter colour to reflect the fact it is a partial sequence.

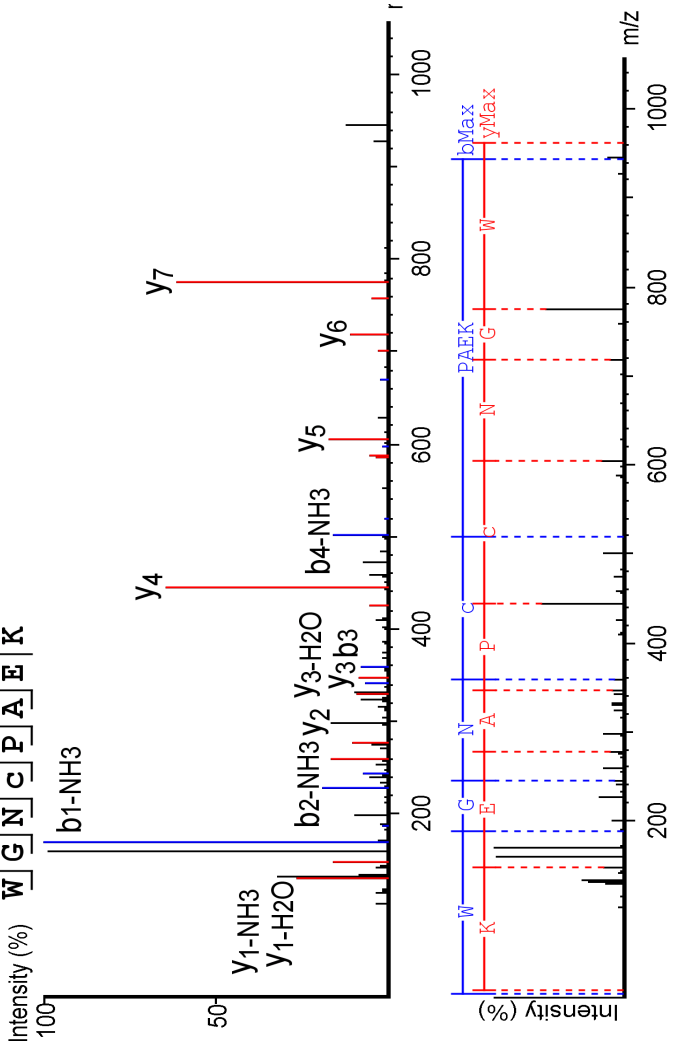


Figure 5.11 *De novo* sequence analysis of the processed MS/MS spectra of *M. murinus* tryptic peptide 961 m/z
M. murinus urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

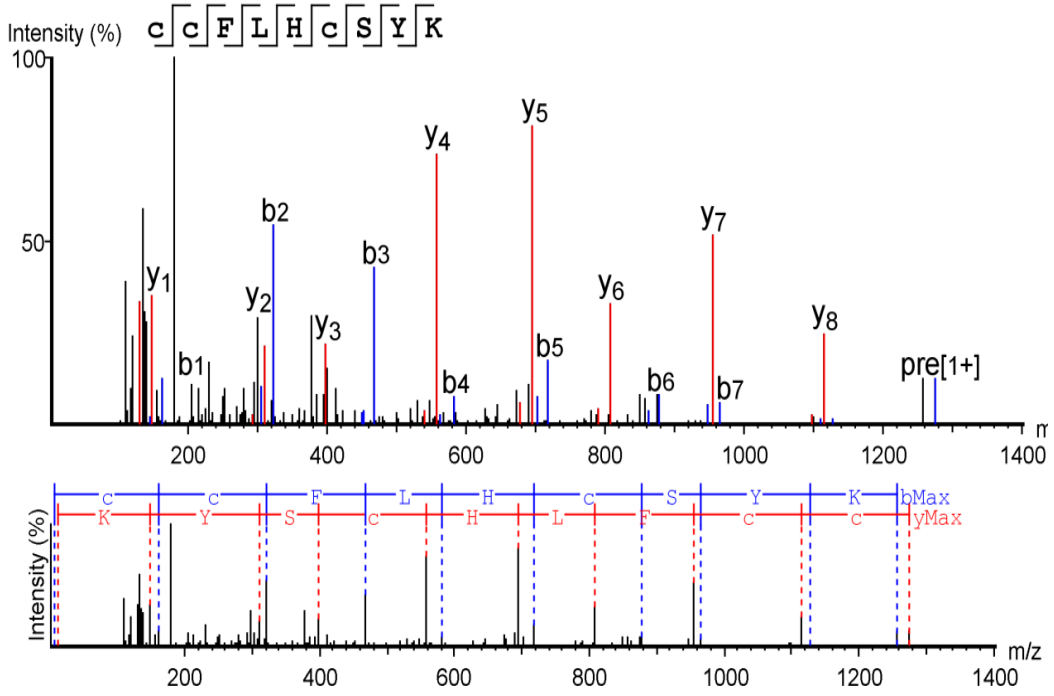


Figure 5.12 *De novo* sequence analysis of the processed MS/MS spectra of *M. murinus* LysC peptide 1274 m/z.
M. murinus urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

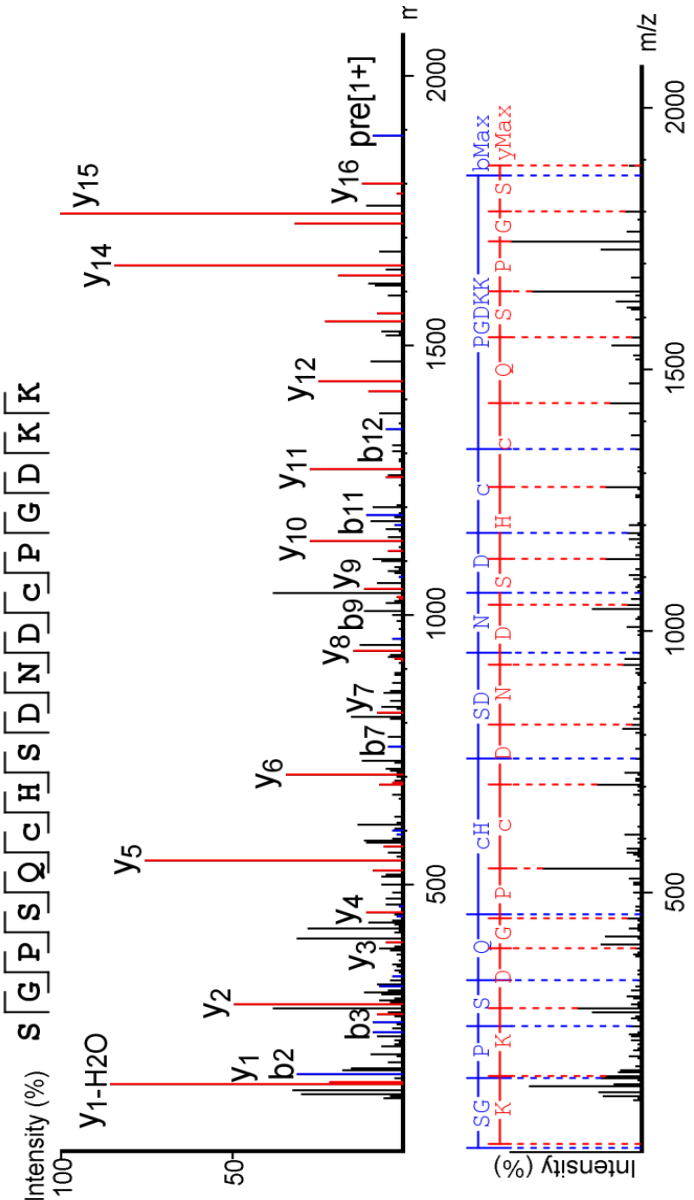


Figure 5.13 ***De novo*** **sequence analysis of the processed MS/MS spectra of *M. Murinus* tryptic peptide 1888 m/z.**
M. Murinus urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

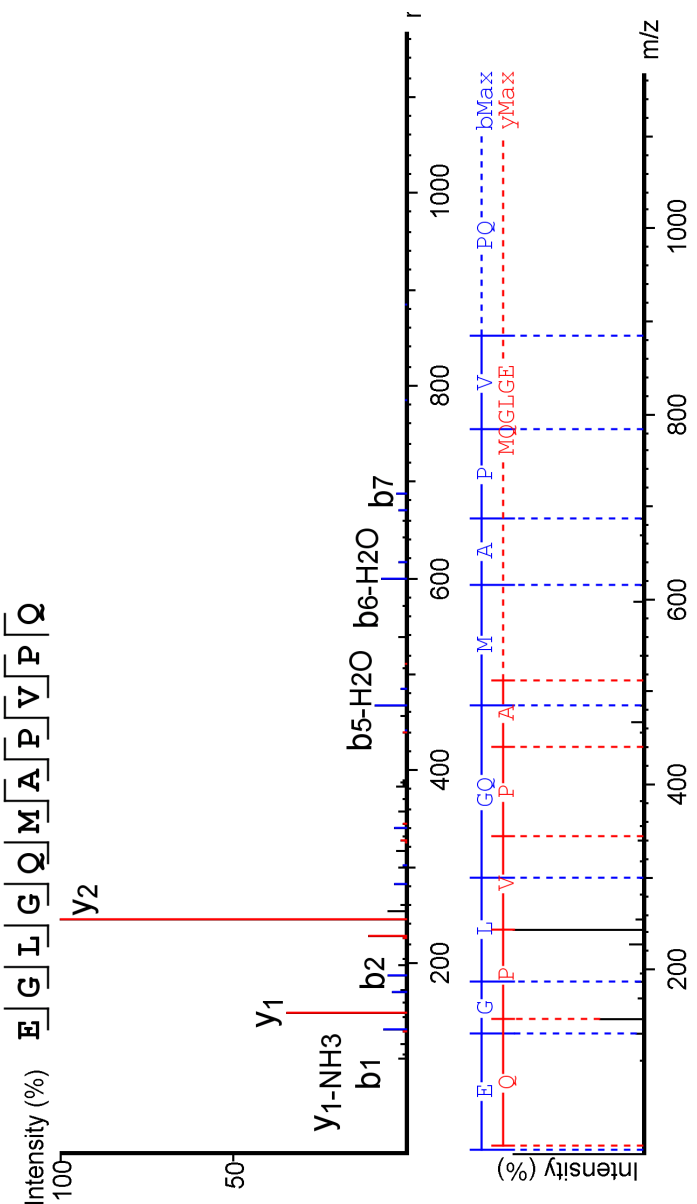


Figure 5.14 *De novo* sequence analysis of the processed MS/MS spectra of *M. Murinus* AspN peptide 1125 m/z.
M. Murinus urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

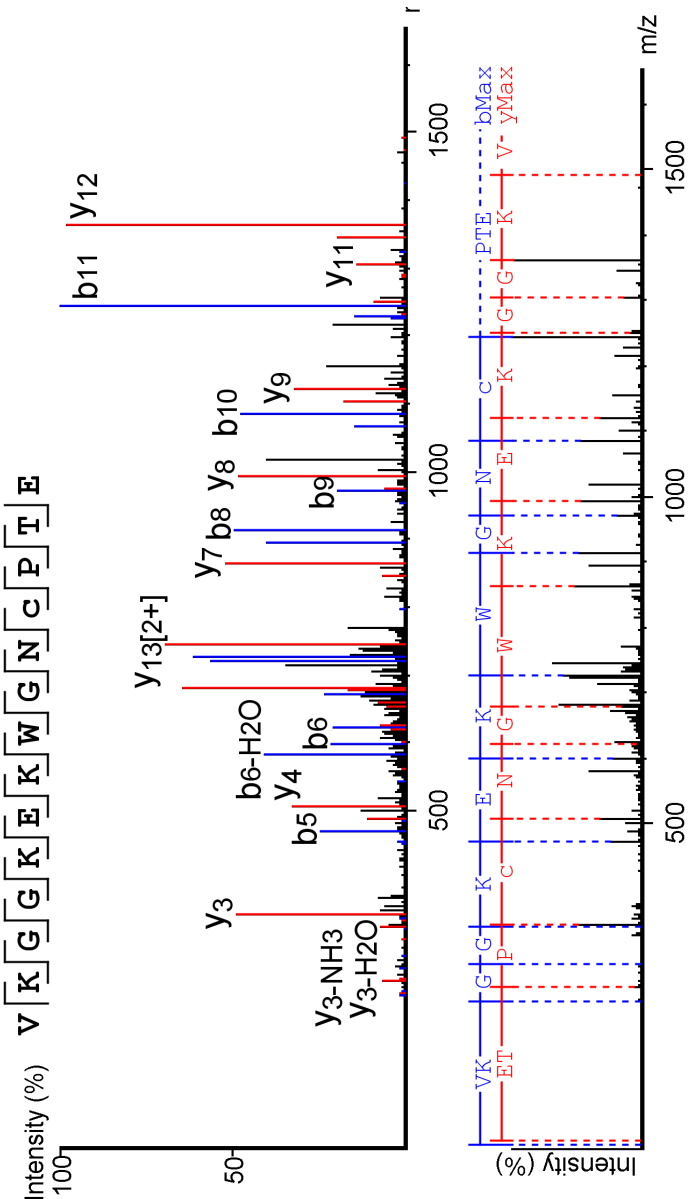


Figure 5.15 ***De novo*** **sequence analysis of the processed MS/MS spectra of *M. lehilahytsara* GluC peptide 1588 m/z.**
M. lehilahytsara urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

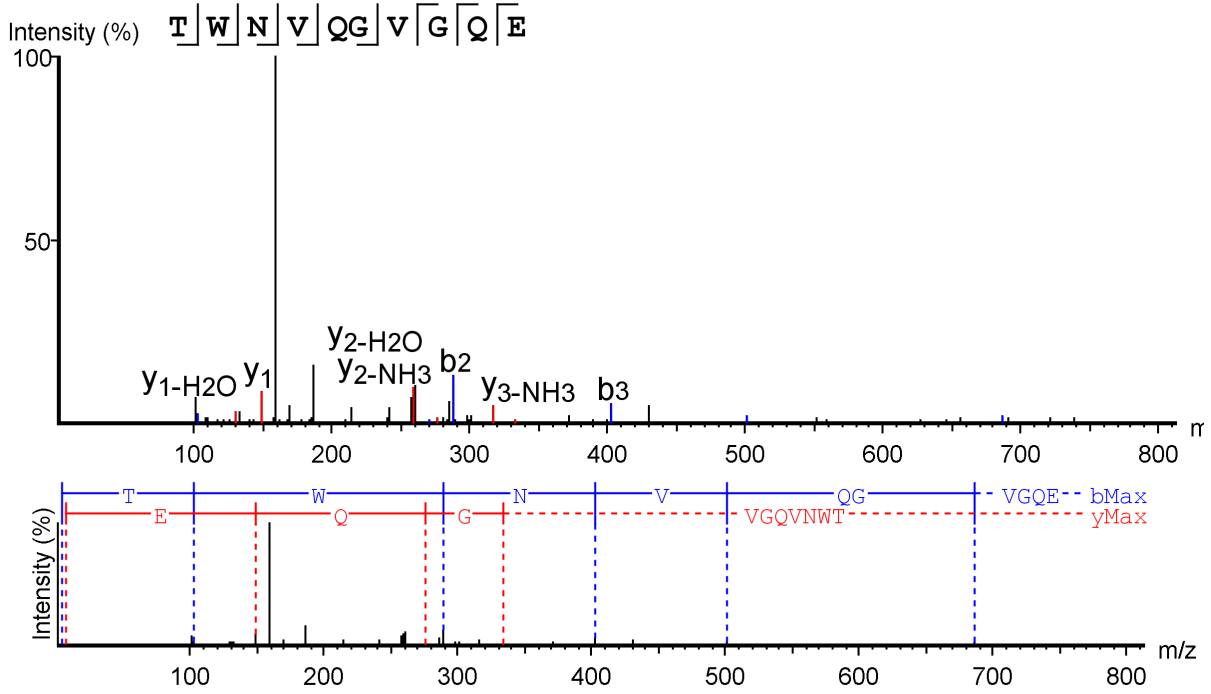


Figure 5.16 *De novo* sequence analysis of the processed MS/MS spectra of *M. lehilahytsara* GluC peptide 1116 m/z.
M. lehilahytsara urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

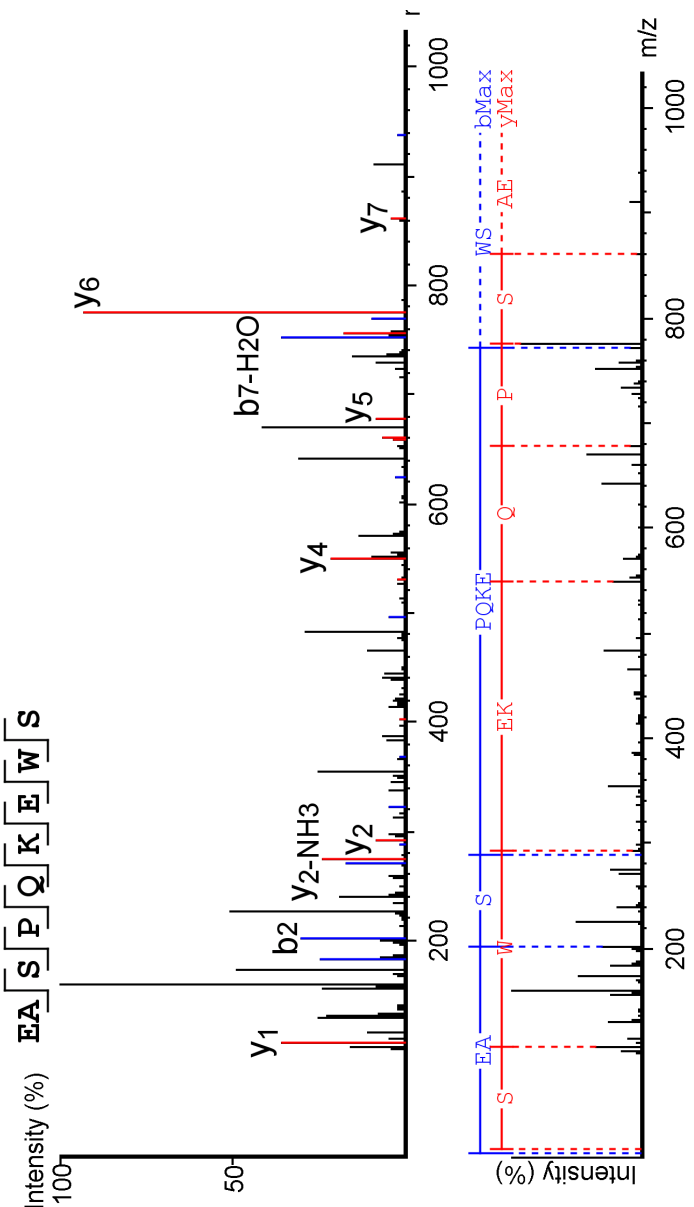


Figure 5.17 *De novo* sequence analysis of the processed MS/MS spectra of *M. lehilahytsara* AspN peptide 1060 m/z.
M. lehilahytsara urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

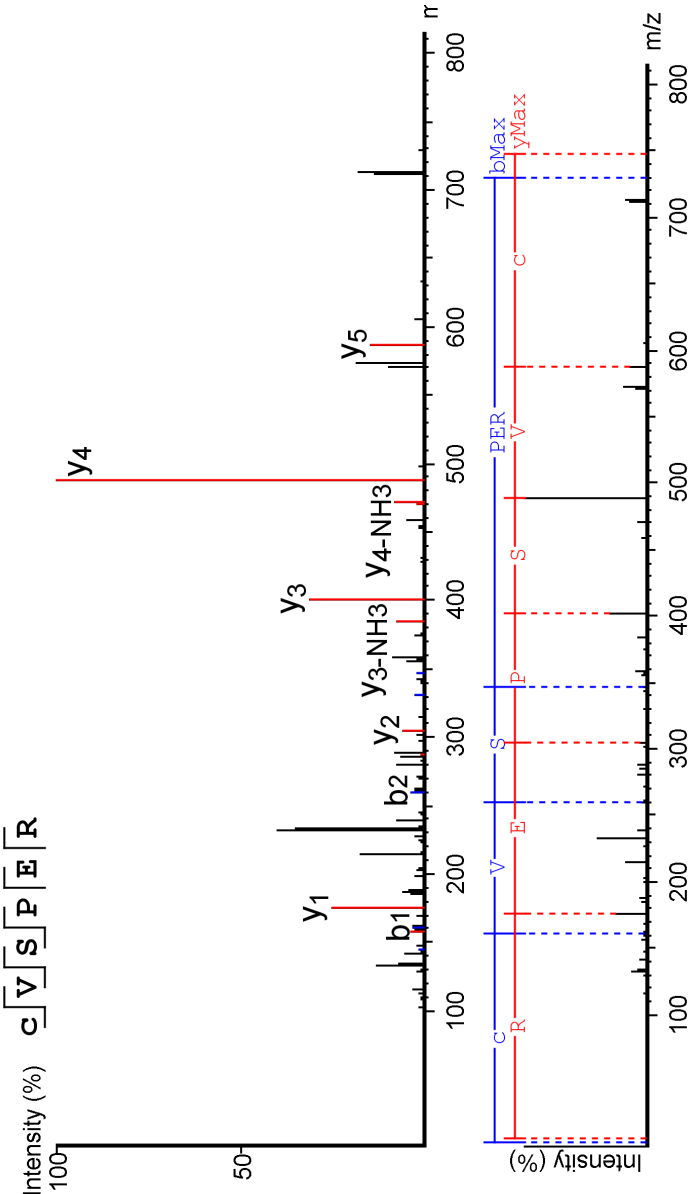


Figure 5.18 *De novo* sequence analysis of the processed MS/MS spectra of *M. lehilahytsara* tryptic peptide 746 m/z.
M. lehilahytsara urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).

Table 5.4 A summary of all peptides *de novo* sequenced from the *M. murinus* in-solution digests. The raw data was processed using PEAKS software. A cut off value of 55% for the total confidence level (recommended by PEAKS) was applied to the de novo analysis. Each amino acid was given an individual confidence percentage. The total confidence score was worked out using the mean of the individual scores.

Sequence	Species	Protease(s)	Mass (Da)	Individual residue confidence scores (%)	Total PEAKS confidence score (%)
WGNCPAEK	*M. mur*	Trypsin	960.41	98, 98, 98, 99, 95, 96, 99, 89	96
		LysC	960.41	99, 98, 99, 99, 96, 97, 99, 90	97
CCFLHCSYK	*M. mur*	Trypsin	1273.50	98, 99, 99, 99, 99, 99, 99, 99, 88	98
		LysC	1273.50	99, 98, 99, 99, 96, 97, 99, 90, 99	97
CVSPER	*M. mur*	Trypsin	746.33	99, 99, 100, 100, 100, 70	94
CVSPERNRK	*M. mur*	LysC	1144.57	86, 85, 97, 88, 77, 83, 78, 67, 67	81
EGLGQMAPVLE	*M. mur*	GluC	1158.50	95, 94, 97, 92, 88, 97, 99, 99, 99, 33, 33	84
TWNVGQVGQE	*M. mur*	GluC	1116.52	90, 91, 69, 89, 75, 72, 88, 90, 92, 77	83
QGAPDTWNVPVADTWNVGQVGQEASPQK	*M. mur*	LysC	2978.41	49, 41, 42, 42, 54, 66, 56, 52, 65, 87, 66, 86, 89, 90, 86, 80, 95, 84, 85, 87, 97, 98, 91, 92, 88	74
SGPSQCHSDNDCPGDKK	*M. mur*	Trypsin	1887.74	95, 100, 100, 95, 93, 100, 92, 99, 99, 97, 98, 100, 97, 94, 95, 99, 93	97
		LysC	1887.74	97, 97, 96, 86, 85, 100, 93, 99, 99, 95, 99, 100, 98, 96, 98, 100, 96	96
EGLGQMAPVPQGA	*M. mur*	AspN	1253.61	96, 73, 87, 78, 64, 92, 97, 96, 98, 87, 83,76,88	86
DTWNVGQVG	*M. mur*	AspN	1102.50	87, 88, 90, 97, 93, 95, 96, 88, 84	91
EASPQKEWS	*M. mur*	AspN	1060.48	98, 97, 93, 89, 75, 85, 96, 93, 86	90

Table 5.5 A summary of all peptides de novo sequenced from the *M. lehilahytsara* in-solution digests. The raw data was processed using PEAKS software. A cut off value of 55% for the total confidence level (recommended by PEAKS) was applied to the de novo analysis. Each amino acid was given an individual confidence percentage. The total confidence score was worked out using the mean of the individual scores.

Sequence	Species	Protease(s)	Mass (Da)	Individual residue confidence scores (%)	Total PEAKS confidence score (%)
VKGGKEKWGNCPTE	*M. lehi*	GluC	1588.76	71, 87, 79, 83, 95, 99, 99, 97, 92, 96, 98, 95, 93, 86	90
WGNCPTEK	*M. lehi*	Trypsin LysC	991.47 991.47	92, 97, 97, 93, 75, 84, 97, 95 97, 96, 98, 97, 91, 88, 96, 72	92 92
CCFLHCSYK	*M. lehi*	Trypsin LysC	1273.50 1273.50	93, 98, 99, 99, 99, 99, 99, 99, 68 92, 97, 99, 99, 99, 98, 98, 95, 68	95 94
SGPSQCHSDNDCPG DKK	*M. lehi*	Trypsin LysC	1887.74 1887.74	94, 94, 95, 87, 84, 100, 91, 99, 97, 94, 99, 100, 96, 92, 93, 99, 94 89, 97, 96, 82, 81, 100, 94, 100, 99, 97, 99, 100, 98, 97, 98, 100, 96	95 95
CVSPER	*M. lehi*	Trypsin	746.33	99, 99, 100, 99, 99, 55	91
EGLGQMAPVPGAQ	*M. lehi*	AspN	1253.61	96, 73, 87, 78, 64, 92, 97, 96, 98, 87, 55,56,58	79
EASPQKEWS	*M. lehi*	AspN	1060.48	98, 97, 93, 88, 75, 86, 96, 93, 86	90
TWNVGQVGQE	*M. lehi*	GluC	1116.52	90, 91, 69, 89, 75, 72, 88, 90, 92, 77	83

The final mouse lemur sequence was input into a software tool that gives the masses of expected peptides in the digest. This would further support the *de novo* analysis and confirm that the middle section of the *L. catta* protein was not present in the mouse lemur protein. These theoretical sequences were then matched up to the PMFs (section 5.2.3) (Figure 5.19a, 5.19b, 5.19c and 5.19d). The software tool was set to allow 3 missed cleavages –sites were the enzyme has not cleaved. Most of the peptides on the PMF were matched up to the theoretical peptides produced by the digest software tool. The exception was the N terminal peptide VKGGKEK

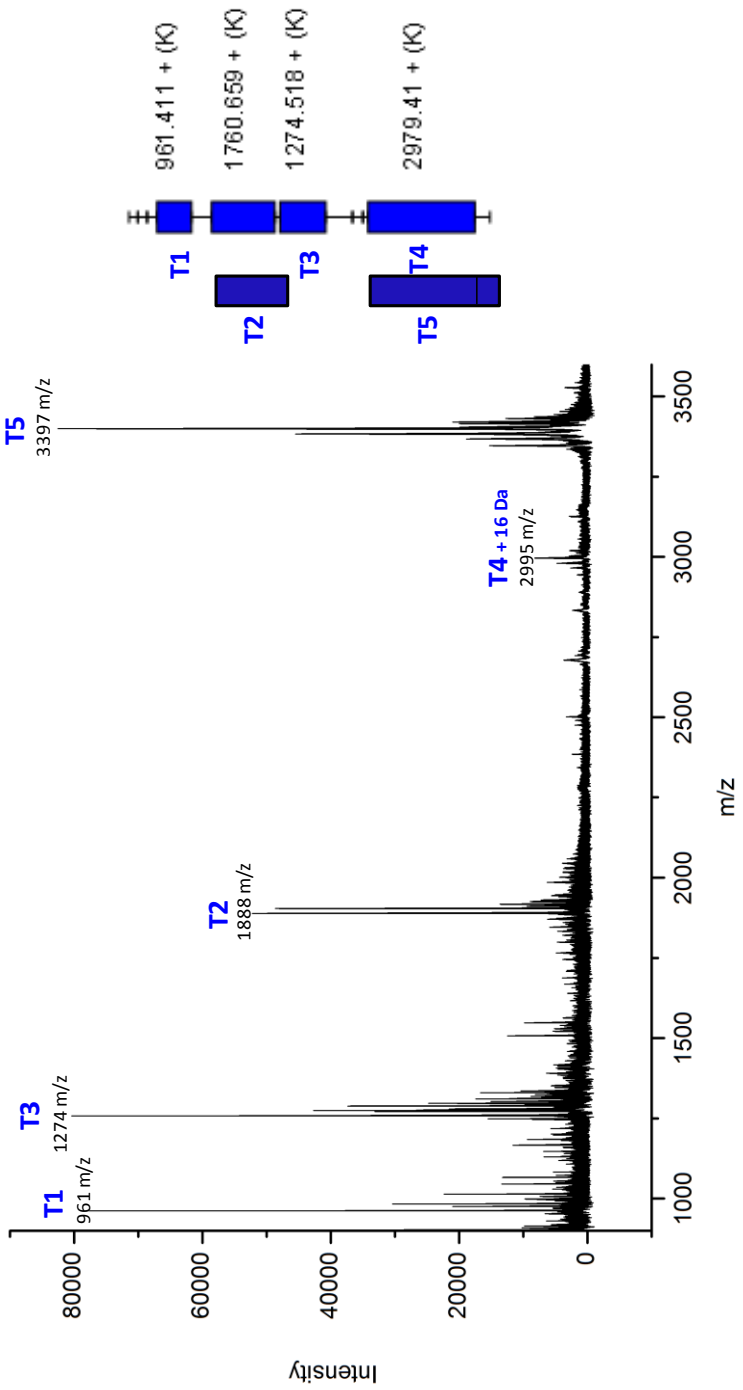


Fig 5.19a Trypsin PMF comparison with theoretical digest of mouse lemur protein.
To further support the *de novo* sequence analysis the final mouse lemur protein was input into a digest tool and theoretically digested with trypsin. This was then matched up to a trypsin PMF. .Peptides below 800 Da are not shown on the peptide map on the left as they were difficult to identify on MALDI due to matrix suppression. Peptides T2 and T5 represent missed cleavages. The 2995 m/z peptide was identified as T4 +16 Da to reflect the oxidation of the methionine residue in the 2979 m/z peptide. The PMF from *M. murinus* was used in this example.

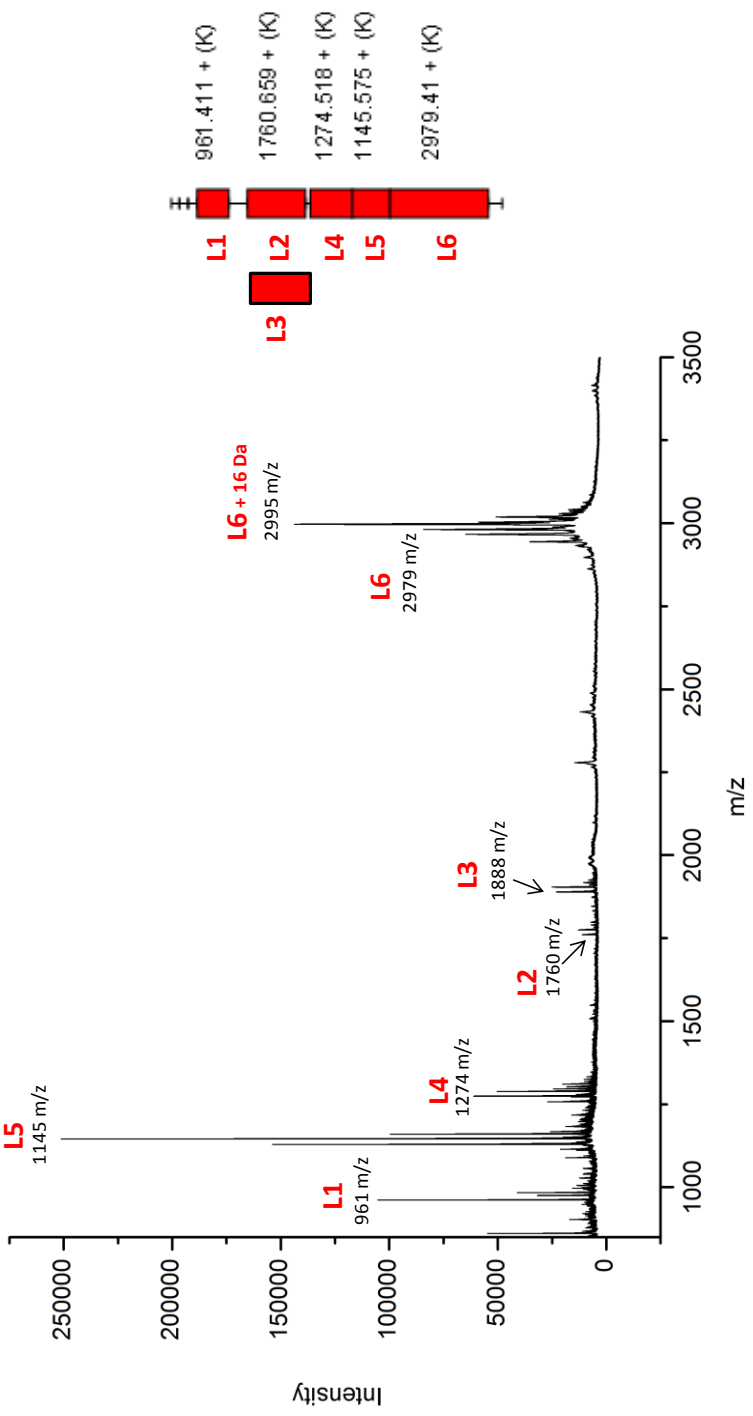


Fig 5.19b LysC PMF comparison with theoretical digest of mouse lemur protein.
To further support the *de novo* sequence analysis the final mouse lemur protein was input into a digest tool and theoretically digested with LysC. This was then matched up to a LysC PMF. Peptides below 800 Da are not shown on the peptide map on the left as they were difficult to identify on MALDI due to matrix suppression. Peptide L3 represents a missed cleavage. The 2995 m/z peptide was identified as L6 +16 Da to reflect the oxidation of the methionine residue in the 2979 m/z peptide. The PMF from *m. murinus* was used in this example.

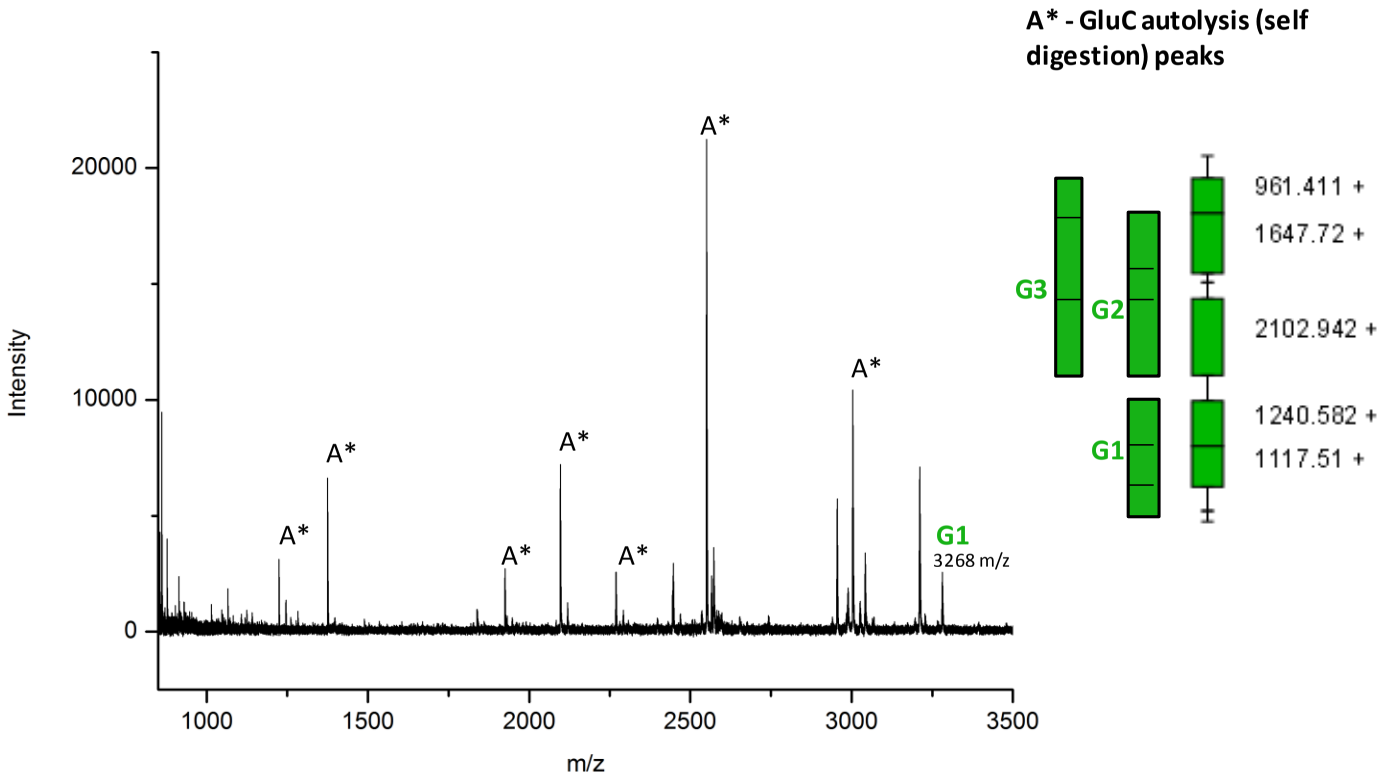


Fig 5.19c GluC PMF comparison with theoretical digest of mouse lemur protein.
To further support the *de novo* sequence analysis the final mouse lemur protein was input into a digest tool and theoretically digested with GluC. This was then matched up to a GluC PMF. Peptides below 800 Da are not shown on the peptide map on the left as they were difficult to identify on MALDI due to matrix suppression. Unfortunately the GluC self digested quite rapidly leading to quite extensive missed cleavages. The mass range of the MALDI was further extended to identify these (Figure 5.19d).

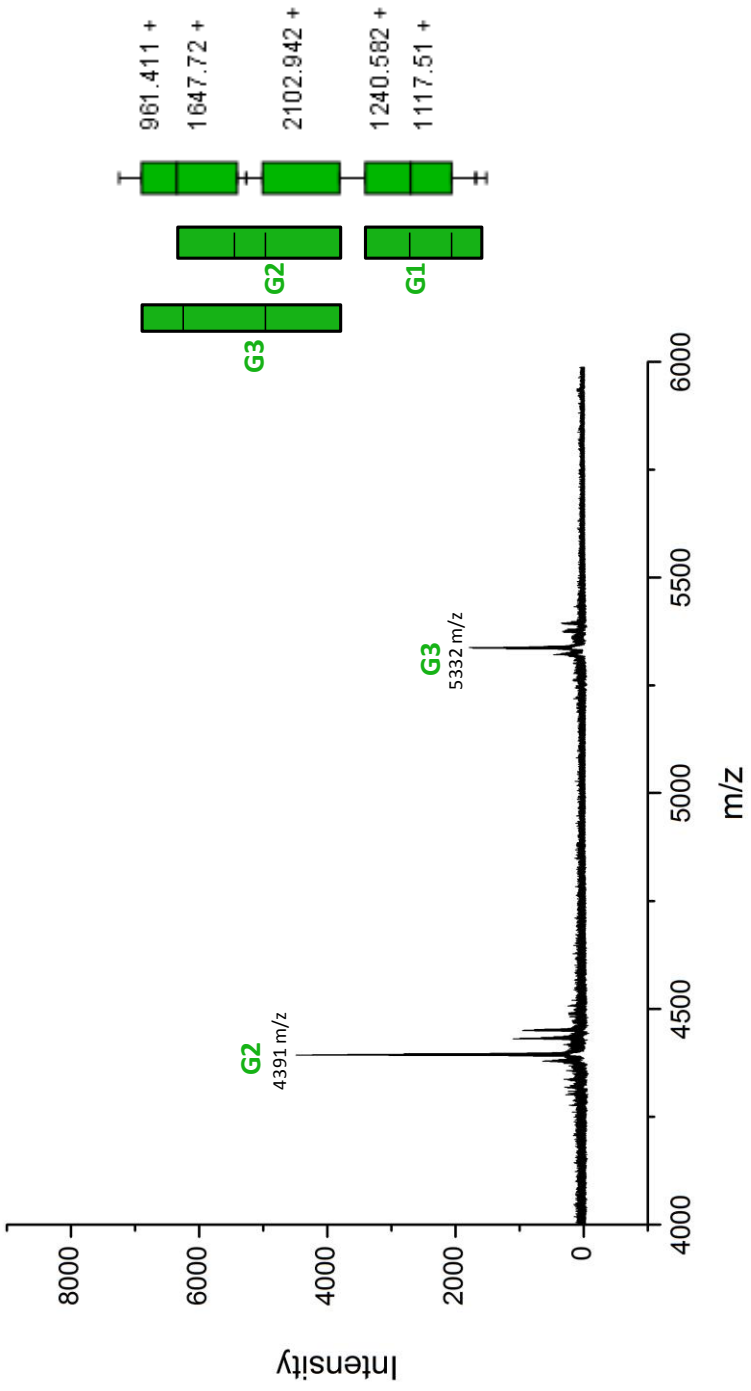


Fig 5.19d GluC PMF comparison with theoretical digest of mouse lemur protein using an extended mass range for detection.
The mass range was extended up to 6000 Da on the MALDI-TOF instrument to further identify missed cleavages at 4391 m/z and 5332 m/z.

which was below the mass limit for this MALDI method. Masses below 800 Da are normally suppressed by matrix ions and therefore difficult to identify. It was however detected as a missed cleavage in the *M. lehilahytsara* GluC LC-MSMS digests (Table 5.5). The summary of the PMF peptides aligned with the mouse lemur sequence is demonstrated in figure 5.20. The total mass of the mouse lemur sequence in each species was also calculated and is in agreement with the intact mass data.

Following the *de novo* sequence analysis and determination of the mouse lemur sequence, genome data for WFDC 12 in *M. murinus* was released. Using the information on the Ensembl database (which obtains its data from Genbank) a predicted amino acid sequence encoded by WFDC 12 *M. murinus* was predicted. Also the part of the *L. catta* protein that is not found in the mouse lemur protein does have a corresponding nucleotide sequence in mouse lemur, and use of an alternative splice site can account for the shorter protein found in urine (Figure 5.21). The predicted *M. murinus* sequence is in agreement with the *de novo* analysis providing further supporting evidence that the final sequence is correct. Once the sequence was confirmed, a model of the protein was constructed using Pymol visualisation software (Schrodinger, Inc) (Figure 5.23).

5.3.6 Sequence differences between species

Sequence analysis of the *M. lehilahytsara* samples identified one single amino acid change that explained the 30 Da increase in molecular weight. As predicted by the peptide mass fingerprint, this sequence change was observed in peptide 991 m/z (961m/z in *M. murinus*). The 991m/z peptide was sequenced as `WGNCPTEK` with the 961 m/z peptide sequenced as `WGNCPAEK` in the *M. murinus* samples. The Ala – Thr substitution accounted for the 30 Da difference in mass. There were no other amino acid changes identified between species – the rest of the protein sequence was identical (Figure 5.22).

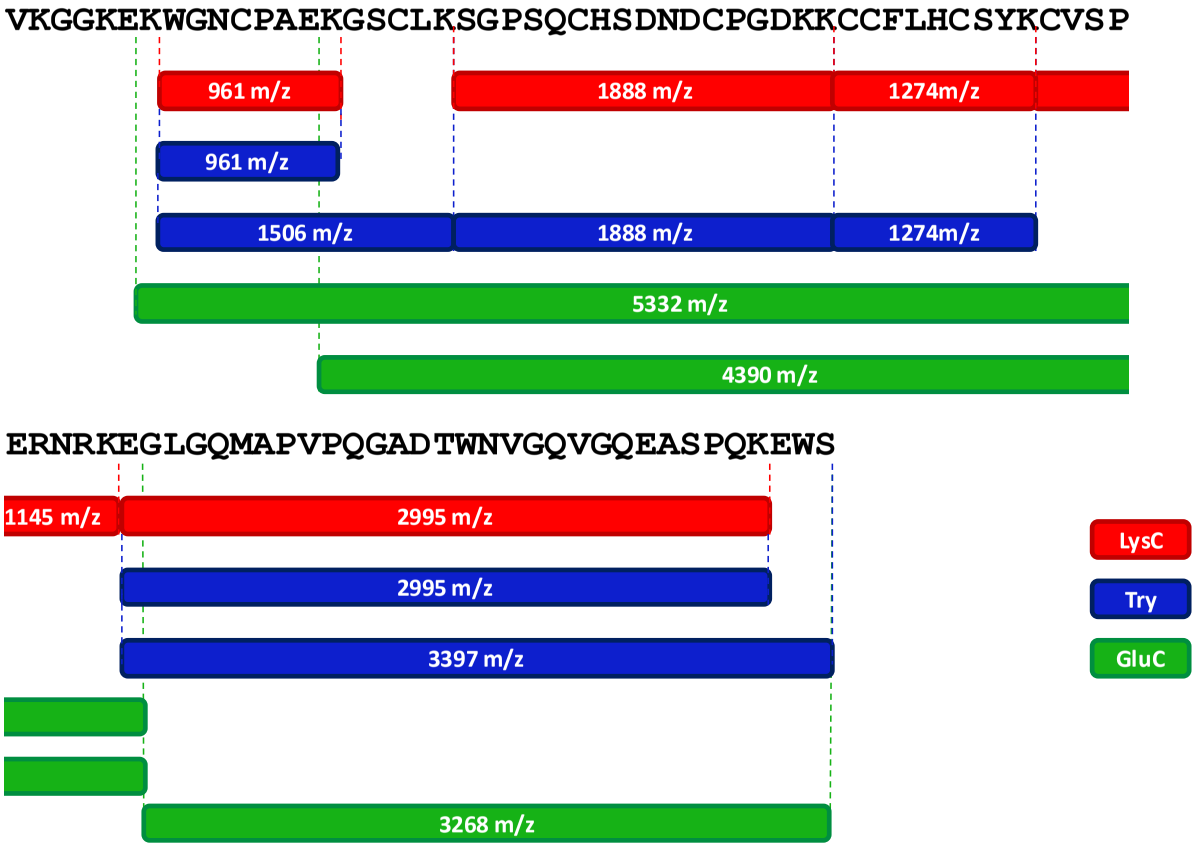


Fig 5.20 Alignment of mouse lemur sequence using peptide masses from PMFs.
The masses observed in the PMFs provided coverage for the majority of the protein. They also overlapped with each other providing further confidence in the sequence analysis. The exception was the N terminal peptide VKGGKEK which was below the mass limit for this MALDI method.

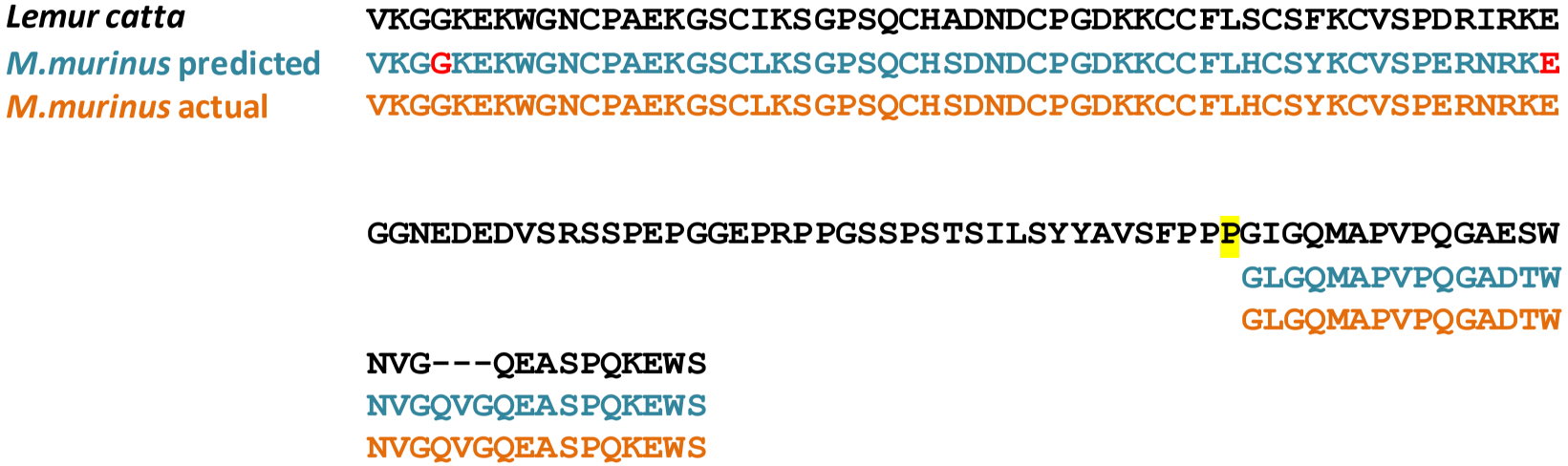


Fig 5.21 Alignment of the final mouse lemur sequence with the genome data from both *L. catta* and *M. murinus*.
The predicted *M. murinus* amino acid sequence encoded by the WFDC12 locus from the *M. murinus* Ensembl sequence (blue) aligned with the sequence derived from the *de novo* analysis (orange). This genome data confirms the de novo analysis and also provides further evidence that the middle section of the *L. catta* WDFC12 protein does not exist in the mouse lemur sequence. The exon boundaries are denoted in red – exon 1 encodes the signalling peptide and exon 2 encodes the WFDC domain. The proline residue highlighted in yellow is an alternative splice site for the start of exon3 in the *M. murinus* Ensembl sequence

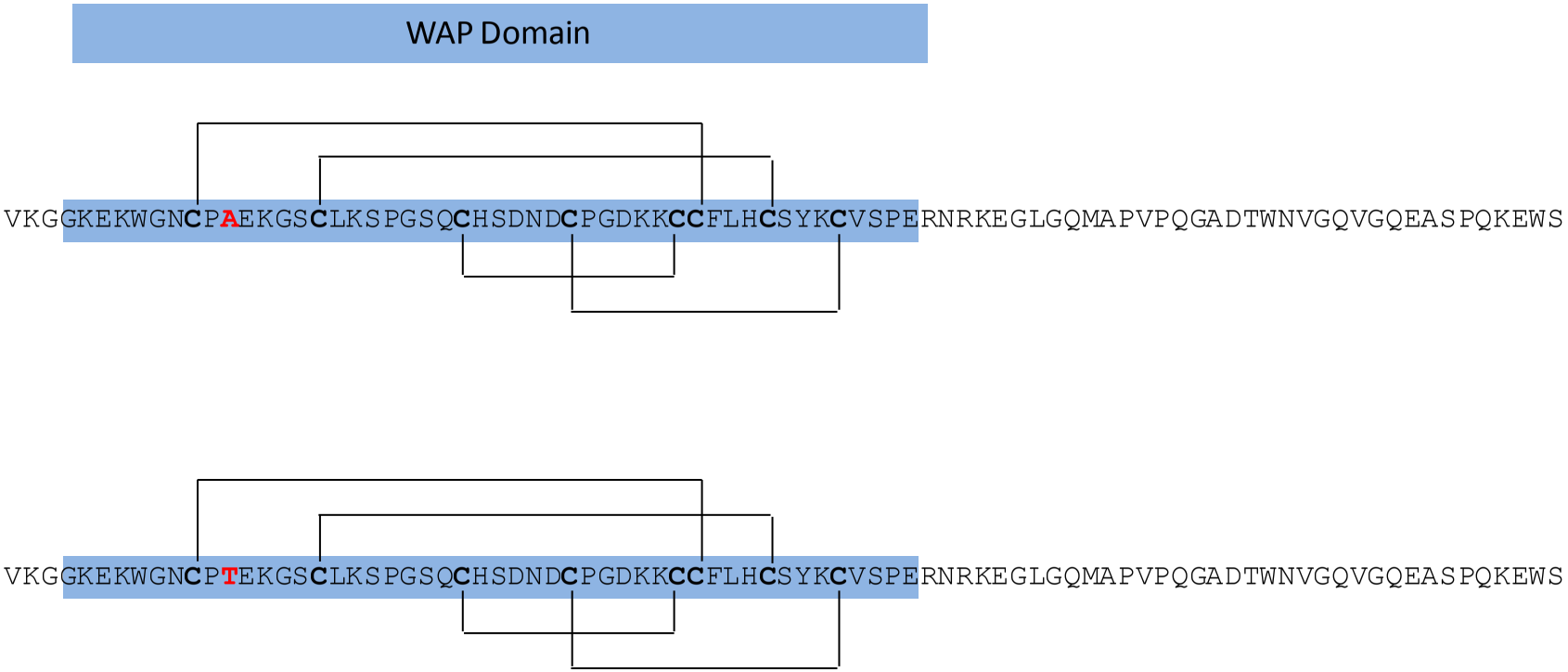


Fig 5.22 Final sequences of the mouse lemur WFDC 12 protein.
The mouse lemur protein has been identified as a WFDC 12 protein. The top sequence represents the sequence found in *M. murinus* and the bottom sequence represents the *M. lehilahytsara* urinary protein. The sequence difference is highlighted in red. The disulphide bond positions in the mouse lemur WAP protein were deduced from those seen in the *L. catta* WFDC protein.

Fig 5.23 Structural analysis of the mouse lemur protein.
A model of the final mouse lemur protein. The tertiary structure was generated using PyMOL molecular visualisation software (Schrodinger, Inc).

5.3.7 Identification of a novel Whey acidic protein

WAP proteins were first identified in the whey fraction of mouse milk (Hennighausen and Sippel, 1982). Rat, camel, rabbit and pig milk were also found to contain a considerable amount of WAP protein (Campbell *et al.,* 1984; Beg *et al.,* 1986; Devinoy *et al.,* 1988; Simpson *et al.,* 1998). These WAP proteins were found to contain disulphide rich domains of approximately 40-50 amino acids. These domains shared limited sequence identity except for 8 characteristically-spaced cysteine residues forming disulphide bonds (Hennighausen and Sippel, 1982). These structural domains were termed four disulphide core domains (FDC) (Drenth *et al.,* 1980). This protein family was therefore named Whey acidic protein four disulphide core proteins (WFDC).

Despite the name not all WFDC proteins are found in milk (Ranganathan *et al*, 1999). Many WFDC have been discovered across all lineages and all share very limited sequence homology except for the highly conserved cysteine region. They are allocated into sub groups depending upon biological function and tissue expression. Biological functions include antibacterial and antifungal action, protease inhibition, tumour suppression and anti-inflammatory activity (Sallenave *et al.,* 1994; Hiemstra *et al.,* 1996; Larsen *et al.,* 1998; McAlhany *et al.,* 2003; Hagiwara *et al.,* 2003; Clauss *et al.,* 2005; Williams *et al.,* 2006; Moreau *et al.,* 2008).

The mouse lemur protein has been identified as WFDC 12. WFDC12 has been studied in several primates and is known to be expressed in the prostate as well as the skin, lungs and oesophagus (Hagiwara *et al.,* 2003). In humans, the WFDC locus contains genes that encode seminal proteins semenogelin 1 and 2 (SEMG1 and SEMG2) which are essential in male reproduction (Lundwall and Clauss, 2011). SEMG proteins are highly expressed in the seminal vesicles and make up over half of the human ejaculate. Post ejaculation these proteins cross link to form a gel matrix that encases ejaculated spermatozoa and trapping it in the female reproductive tract. A protease named prostate-specific antigen (PSA) then breaks down this gel matrix to allow motility of the spermatozoa to return. In contrast to monoandrous mating where the ejaculate is “loose” in texture, in multi-male/multi female mating

systems the male ejaculate forms a rigid solid copulatory plug. Also the rate of SEMG2 evolution is thought to correlate with female promiscuity and semen coagulation which is thought to be related to post-copulatory sperm competition (Doris *et al.,* 2004). In most primates, including mouse lemurs, the WFDC 12 gene in particular resides on the same centromeric sublocus as the genes encoding these reproductive proteins (Hurle *et al.,* 2007). In a study by Hurle *et al.,* 2007, evidence of positive selection on WFDC12 was observed during primate evolution which suggests this gene may be involved in sexual selection.

5.3.8 Potential functions of the mouse lemur WDFC 12 protein

As many WFDC proteins have protease inhibition effects protease inhibition of the mouse lemur protein was investigated. Many WAP proteins inhibit trypsin, chymotrypsin and elastase (McCrudden *et al.,* 2007) so potential trypsin inhibition properties of the mouse lemur protein were examined. Trypsin activity was previously investigated in chapter 3 to assess whether MUPs where forming inhibitory products that made them resistant to complete proteolysis by trypsin (see section 3.3.2). An alternative approach to using the spectrophotometric method described here could have been to use a similar set-up illustrated in chapter 3 by replacing the MUP protein with the mouse lemur protein. If the mouse lemur protein inhibited trypsin then the digestion reaction would not go to completion. However this would depend on the rate of inhibition, something which can be calculated using a spectrophotometric assay. If the protein was a slow inhibitor this may be difficult to detect by SDS-PAGE as the majority of the protein would be digested and any that has not been may be too low to visualise on the gel.

A stock solution of trypsin was prepared (200 µg/ml in 50 mM Tris-HCl pH 8.0 + 10 mM $CaCl_2$) and incubated at 37 °C for 60 minutes. At 10 minute intervals an aliquot was removed and mixed with a trypsin substrate N_α-Benzoyl-L-arginine 4-nitroanilide (BAPNA, 0.5 mM in 50 mM Tris-HCl pH 8.0 + 10 mM $CaCl_2$) to produce a final concentration of 20 µg/ml trypsin. Trypsin recognises BAPNA as a substrate and cleaves at arginine to release the 4 nitroanaline which turns the solution

yellow. The absorbance (405 nm) was measured every minute over a 10 minute period. The absorbance readings were plotted against time to assess if incubating trypsin at 37 °C caused a reduction in activity over time (Figure 5.24a). This was not the case and trypsin activity remained stable over the 60 minute incubation period.

A solution containing the mouse lemur protein and trypsin was prepared (200 µg/ml of each in 50 mM Tris-HCl pH 8.0 + 10 mM $CaCl_2$) and incubated at 37 °C for 60 minutes. At 10 minute intervals an aliquot was removed and mixed with BAPNA (0.5 mM in 50 mM Tris-HCl pH 8.0 + 10 mM $CaCl_2$) to produce a final concentration of 20 µg/ml of each. The absorbance (405 nm) was measured every minute over a 10 minute period. The absorbance readings were plotted against time to assess if incubating trypsin with mouse lemur WFDC protein inhibited trypsin activity over time (Figure 5.24b). If the mouse lemur protein was an inhibitor of trypsin the absorbance readings would decrease or reach a plateau as the amount of p-nitroanaline produced slows down or halts completely depending on the strength of inhibition. No inhibition of trypsin was identified. The experiment was repeated again, the only difference was the concentration of mouse lemur protein was doubled (40 µg/ml final concentration). Still no inhibition of trypsin was identified (Figure 5.24c).

These results were not surprising as the key amino acid required in the trypsin inhibition process is a Lys at position 18, which is a serine in the mouse lemur protein (Cechova and Muszynska, 1970). A WFDC protein named eppin is an androgen dependant epididymal protease that plays an important role in human male reproduction and fertility. Functions of eppin include modulation of PSA activity, antimicrobial action and inhibition of sperm motility by binding to SEMG1 (Yenugu *et al.,* 2004; O'Rand *et al.,* 2011). Eppin also lacks a Lys residue at position 18 in its FDC domain and therefore does not inhibit trypsin. It does however inhibit elastase (McCrudden *et al.,* 2007). It is possible that the mouse lemur protein may share some of the functions of eppin as they are both male specific and there is evidence of the mouse lemur protein having a role in sexual selection and reproduction (Hurle *et al.,* 2007). As eppin inhibits elastase, the mouse lemur protein was also examined to see if it too reduced or halted the activity elastase.

For investigating elastase inhibition by the mouse lemur protein, the same experimental process was applied as with the trypsin inhibition experiment. An elastase substrate N-Succinyl-ALA-ALA-ALA p-Nitroanilide was used instead of BAPNA. Again no inhibition was identified (Figure 5.25).

The lack of protease inhibition is not very concerning. It is thought that elastase inhibition by eppin is due to an additional domain it possesses – kunitz domain which are normally found in proteins that are responsible for inhibiting the activity of protein degrading enzymes. Furthermore isolation of the WFDC domain in eppin shows antimicrobial effects against *E. coli*. A study by Donpudsa et al., 2010 tested two recombinant crustin proteins for protease inhibition and antimicrobial effects as both proteins contained a WFDC domain. Both proteins did not inhibit protease activity but did exert antimicrobial activity through a bactericidal effect. Antimicrobial activity of the mouse lemur protein is yet to be investigated.

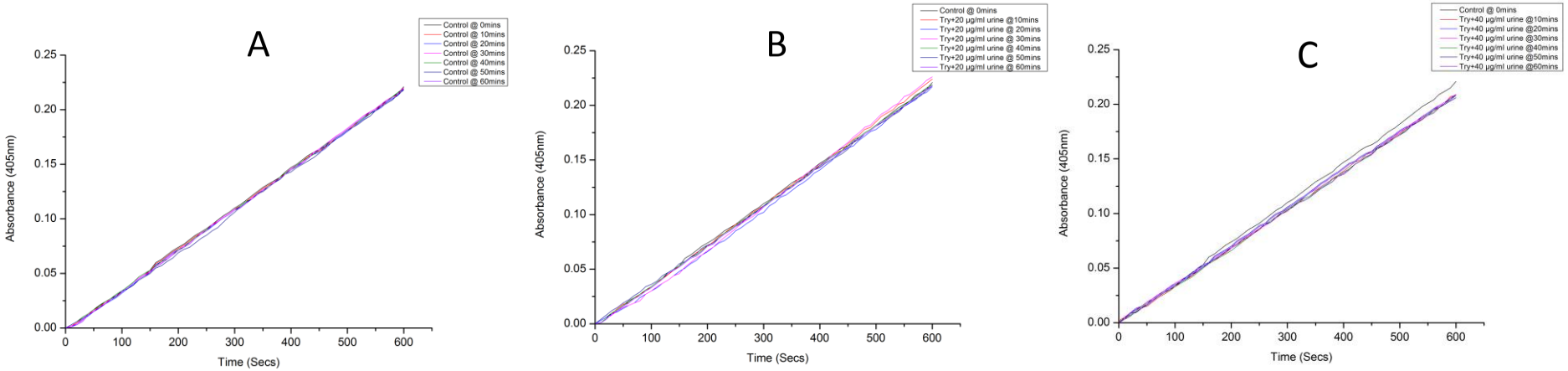


Fig 5.24 Assessment of trypsin inhibition following incubation with the mouse lemur WFDC protein.

Graph A. A sample containing trypsin (diluted in 50 mM Tris-HCL pH8.0 + 10 mM $CaCl_2$) was incubated at 37 °C for 60 minutes. An aliquot was removed every 10 minutes and mixed with a trypsin substrate BAPNA (0.5 mM in 50 mM Tris-HCL pH8.0 + 10 mM $CaCl_2$) to produce final concentrations of 20 µg/ml trypsin. The absorbance (405 nm) measured at every minute over a 10 minute period. The absorbance readings were plotted against time and degree of degradation of trypsin activity compared. Graph B. Male mouse lemur urine containing the WFDC 12 protein was incubated with trypsin for 60 minutes. After 10 minute intervals an aliquot of the urine+trypsin was removed and mixed with BAPNA (0.5 mM in 50 mM Tris-HCL pH8.0 + 10 mM $CaCl_2$) to produce final concentrations of 20 µg/ml of each. The absorbance (405 nm) was measured every minute over a 10 minute period. The absorbance readings were plotted against time and degree of inhibition of trypsin activity compared. Graph C. The concentration of male mouse lemur was doubled and incubated with trypsin for 60 minutes. After 10 minute intervals an aliquot of the urine+trypsin was removed and mixed with BAPNA (0.5 mM in 50 mM Tris-HCL pH8.0 + 10 mM $CaCl_2$) to produce final concentrations of 20 µg/ml trypsin and 40 µg/ml WFDC 12 protein . The absorbance (405 nm) was measured every minute over a 10 minute period. The absorbance readings were plotted against time and degree of inhibition of trypsin activity compared.

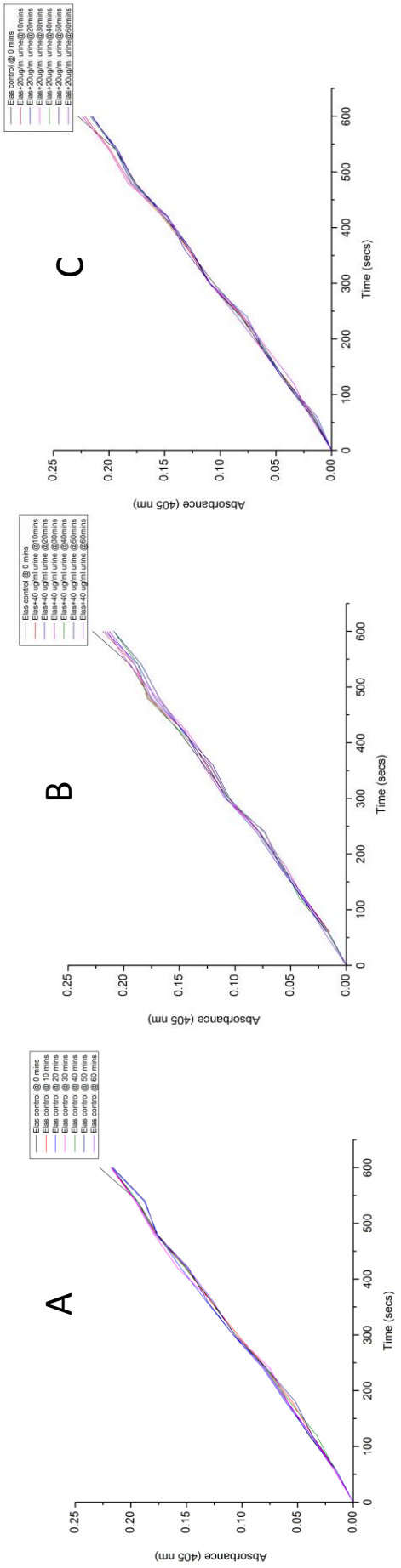


Fig 5.25 Assessment of elastase inhibition following incubation with the mouse lemur WFDC protein.
Graph A. A sample containing elastase (diluted in 50 mM Tris-HCl pH8.0 + 10 mM $CaCl_2$) was incubated at 37 °C for 60 minutes. An aliquot was removed every 10 minutes and mixed with an elastase substrate N-Succinyl-ALA-ALA-ALA p-Nitroanilide (0.5 mM in 50 mM Tris-HCL pH8.0 + 10 mM $CaCl_2$) to produce final concentrations of 20 µg/ml elastase. The absorbance (405 nm) measured at every minute over a 10 minute period. The absorbance readings were plotted against time and degree of degradation of elastase activity compared. Graph B. Male mouse lemur urine containing the WFDC 12 protein was incubated with elastase for 60 minutes. After 10 minute intervals an aliquot of the urine+elastase was removed and mixed with N-Succinyl-ALA-ALA-ALA p-Nitroanilide (0.5 mM in 50 mM Tris-HCL pH8.0 + 10 mM $CaCl_2$) to produce final concentrations of 20 µg/ml of each. The absorbance (405 nm) was measured every minute over a 10 minute period. The absorbance readings were plotted against time and degree of inhibition of elastase activity compared. Graph C. The concentration of male mouse lemur was doubled and incubated with elastase for 60 minutes. After 10 minute intervals an aliquot of the urine+elastase was removed and mixed with N-Succinyl-ALA-ALA-ALA p-Nitroanilide (0.5 mM in 50 mM Tris-HCL pH8.0 + 10 mM $CaCl_2$) to produce final concentrations of 20 µg/ml elastase and 40 µg/ml WFDC 12 protein . The absorbance (405 nm) was measured every minute over a 10 minute period. The absorbance readings were plotted against time and degree of inhibition of elastase activity compared.

5.4 Conclusions

Using advanced proteomic techniques, it was possible to fully identify and characterise the mouse lemur urinary protein prior to obtaining genomic information. This protein has been identified as a WFDC protein. This protein was observed in certain male mouse lemurs from two different species *Microcebus murinus* and *Microcebus lehilahytsara* and is only expressed during the reproductive season. Only one amino acid mutation is present to differentiate between the two species. The de novo sequencing approach was similar to that undertaken in the harvest mouse chapter. PEAKS produced high scoring confidence levels for all sequences with the majority of scoring over 90%.

Despite having two functional MUP genes, this protein is not a MUP and is in fact very different to the lipocalins observed in many rodent species. WAP proteins have a variety of biological functions, some of which were explored in this chapter. Protease inhibition properties of this protein were investigated and they were shown to not affect protease activity. As many other WAP proteins have antimicrobial properties, including those that do not display protease inhibition, the mouse lemur protein should also be examined to see if antimicrobial activity is one of its functions.

Assessing the link between chemical signalling and the expression of this protein was not in the scope of this project. The captive mouse lemurs live in triads and unfortunately not all males that live together were sampled. Also only one sample of urine was provided for each mouse lemur sampled.

It would be beneficial to not only sample all males, but also take a number of samples just before breeding season commences and throughout the breeding season itself. If protein expression was related to male dominance then increased concentrations may be observed just prior to the breeding season when dominance hierarchies are established. Also, by sampling all males who live together, if expression is dominance related we might expect to see just one member of the group expressing the protein. Taking samples over a number of breeding seasons may confirm the dominance theory if the same male continue to express the

protein. Also it may potentially reveal new dominance hierarchies being formed as the existing alpha males get older and are no longer classed as a threat to younger maturing males.

If the expression of this protein peaks during the season when actual mating is taking place, then the protein may play a role in sperm competition and possible serve as an attractant for females as they are the dominant sex. Alternatively, it could be a protein that has antimicrobial properties to protect sperm in the reproductive tract and has escaped the filtration step due to its small size resulting in the presence of the protein in urine or the primary origin of the protein may not be from urine but from the seminal fluid. It is not known if the mouse lemurs had mated prior to sampling. If the protein originates from seminal fluid, then residual amounts could have been left in the urinary tract and consequently ended up in the urine sample which reinforces the need for multiple samples collections. It would therefore be advantageous to examine seminal fluid for the presence of this WAP protein.

Chapter 6: General conclusions

The aim of this thesis was to investigate the protein content of scent marks using advanced proteomic techniques. With significant developments in the mass spectrometry field in particular, it was possible to achieve the main aims and objectives set out at the beginning of this thesis (section 1.7). The newer generation TOF analyser with improved dynamic range and sensitivity, enabled the absolute quantification of MUP isoforms in B6 laboratory mice using a QconCAT method. Previously, MUP quantification was limited to ESI-MS and it was only possible to relatively quantify the major isoforms. While the QconCAT method was successful, further testing of this method using more biological and technical replicates is fundamental to ensure robustness and reproducibility of the method, particularly surrounding the digestion of the MUP proteins. Future MUP quant studies should focus on trying to improve the digestion or alternatively, look at re-designing the QconCAT to contain flanking regions or the use of alternative proteases. There would no doubt be similar problems when it comes to unique peptides but enzyme cleavage sites may be less challenging to cleave, improving digestion efficiency. Peptides could also be chosen on how well they digest so there can be confidence in complete digestion of at least the peptides chosen for quantification. Other options for absolute quantification such as AQUA peptides and PSAQ standards would see the same issues arise as with the current QconCAT – limited unique peptides and digestion discrepancies. A PSAQ method could potentially result in a more accurate method as quantification is done using multiple peptides but this would be very time consuming and costly. A PSAQ standard would have to be produced for each MUP and a pre-fractionation method developed to separate each MUP isoform prior to digestion.

In the past *de novo* sequencing of proteins could be a very time consuming and manual task. The introduction of mass analysers such as the Orbitrap and the improvements made in software means *de novo* sequencing is now both quicker and more automated. In chapters two and three of this thesis, proteins secreted by the harvest mouse and mouse lemur were *de novo* sequenced prior to genomic

data becoming available. The high resolution, sensitivity and efficient fragmentation (HCD) of the QExactive mass analyser resulted in high quality sequence data being obtained for both species. PEAKS 6 software proved to be reliable and accurate for *de novo* sequencing although it is good practice to manually inspect MS/MS spectra as the software is only as good as the raw fragmentation data supplied for interpretation.

Three proteins were identified in the harvest mouse and *de novo* data was obtained for two of these proteins. These proteins were identified as being from the lipocalin family. Unlike mice and rats, there appeared to be no sexual dimorphism between males and females. This is similar to roborovskin, a protein secreted by Roborovski hamsters that also show no expression differences between sexes. Three abundant proteins were identified in both sexes by ESI-MS but the isoform pattern was different between individuals. As the harvest mice originate from an outdoor enclosure they are not as inbred as the laboratory strains of mice so the variation could be due to genetic diversity. SDS-PAGE analysis revealed two abundant protein bands and after extensive mass spectrometry analysis it appears that harvest mice express large quantities of odorant binding proteins and it is these proteins that are most likely used to convey information although further behavioural studies would have to confirm this. The third mass (17888 Da) was not able to be characterised as this protein was not observed by AEX chromatography and could therefore not be isolated and characterised. The most likely cause of this was there simply was not enough of the protein present in the sample. Capturing a larger sample size of harvest mice from the outdoor enclosure could result in obtaining a rodent who expresses much larger quantities of this particular protein allowing for identification and characterisation.

A discovery run on the harvest mouse protein secretion not only identified odorant binding proteins but also MUPS. The database and BLAST matches were to the peripheral MUPS – 5, 4 and 20. The peripheral MUP genes are older and more stable and it is thought the central MUPs in mice are a result of gene duplications and divergence of the peripheral genes. So in the case of the harvest mouse if they do have peripheral MUP genes then results would suggest these genes have maybe

not diverged as there was no evidence of central MUPs in the harvest mouse. Genomic data would be necessary to confirm this.

The primary source of the protein secretion in the harvest mouse was also different to other rodents. The majority of rodents secrete a large amount (mg/ml) of protein into their urine and this is used as their primary source of communication. In the case of the harvest mouse urinary protein abundance was low and the primary source of secretion is either the paws and/or saliva. There are two theories – one is that the paws secrete the protein during climbing, an activity harvest do routinely too reach nests, food etc, which is why the SDS-PAGE of the paw washes identified low abundant bands compared to the glass rod washes, which the rodents climb up. The second theory is that the protein is primarily excreted in saliva and the rodent licks their paws before climbing. Many peptides were matched to mouse MUPs 4 and 5 in the discovery run which are not observed in mouse urine but have been observed in mouse saliva. It would be beneficial to complete the sequencing of the two lipocalins so that more in-depth behavioural studies can occur and identify individual roles for the individual proteins. It would also be interesting to observe any potential volatiles that may be bound to these proteins as MUPs in rats and mice bind a number of small molecules which have many roles in chemical signalling.

The final chapter of this thesis examined protein expression in two species of mouse lemur *Microcebus murinus* and *Microcebus lehilahytsara.* The protein was identified as a male specific WAP protein that was only expressed in the breeding season with only a single amino acid mutation between species. WAP proteins have a variety of functions and are expressed across all lineages. As the protein has been fully characterised and now has genomic data to support the *de novo* sequence analysis, the next step is to find out the biological function. Behavioural studies were not in the scope of this project but after the identification of the WAP several hypotheses have been raised.

One theory is this protein is not involved in chemical communication at all and may just serve as an antimicrobial that protects the sperm as it enters the female

reproductive tract. One of the functions of eppin is bind to the surface of SEMG1 and performs a protective shield around the spermatozoa once the ejaculate has entered the recipient female. This would account for the mouse lemur protein being male specific and also up-regulated during the breeding season.

The second theory is that this protein is involved in scent communication. The protein is very different to the MUP and other lipocalin proteins deposited by rodents. Male mouse lemurs are known to establish dominance over each other at the beginning of the breeding season so it is possible this protein plays a role in sperm competition. Physiological effects including decreased body mass and testosterone levels have been observed in subordinate males post exposure to dominant male urine (Perret and Schilling 1987; Perret and Schilling, 1995). Alternatively, this protein, like eppin, may be androgen dependant. If the males that are not expressing this protein are subordinate males then their testosterone levels will be lower and protein expression down regulated. This theory would have to be investigated further and establish which of the mouse lemurs sampled live together and which ones are at the top of the dominance hierarchy.

As females are the dominant sex they ultimately decide who to mate with. The protein may serve as an attractant for females similar to darcin in male mice which causes females to become sexual attracted to individual males (Roberts et al., 2010). A study using captive gray mouse lemurs by Radespiel *et al.,* 2002 found no correlation between dominance and reproductive success and that maybe female mate choice and sperm competition play more of a central role. In a separate study by Aujard 1997 the removal of the VNO in sexually experienced males caused a dramatic decline in intermale aggression but did not impair successful mating or testosterone levels.

To gain further insight into the potential function of this protein, comprehensive behavioural studies will need to be established. A recombinant form of this mouse lemur protein would assist in answering questions surrounding possible use in chemical signalling. Recombinant forms of mouse MUPS have previously provided a detailed analysis into both intrasexual and intersexual scent communication.

The MUP quantification work will be used in an ongoing project on pest control in developing countries. Rodents are not the only creatures classed as "pests" and it is possible that other laboratories investigating pest control strategies can gain some valuable insight from observations and data collected from the rodent pest control project. For example, some insects can be described as "pests" due to their destructive nature, damage to crops and homes, and for carrying fatal diseases. Pheromones have been observed in insects such as termites and mosquitoes. Mosquitoes are well known for carrying the potentially fatal disease malaria. It is the females who are adapted to have the ability to feed on humans and other animals for blood to give her the nutrients she needs to produce eggs. Females also transmit a pheromone to attract males when they are ready to mate (Pitts *et al.,* 2014). Using strategies developed in the MUP rodent control project, gaining further insight into mosquito (and other insect) behaviour, particularly on the protein chemistry level, could dramatically reduce the number of malaria cases.

Animal welfare is another area of research that the MUP quantification studies will benefit. As discussed in chapter 3, animal welfare projects monitor the wellbeing of laboratory rodents and being able to identify what MUPs cause aggression between conspecifics will be advantageous to these projects. However animal welfare is not just restricted to the welfare of laboratory rodents but to other animals in captivity such as zoo animals. Zoos are fundamental in preventing the extinction of many endangered species and take part in worldwide breeding programmes. The well being of these animals is crucial for the success of these breeding programmes. Females will not feel comfortable producing young if they are not living a place they feel secure in. Monitoring chemical signals, in particular those related to stress and reproduction, may provide indicators to how relaxed an animal feels in its current habitat.

It is expected that the identification and characterisation of the harvest mouse and mouse lemur proteins will contribute towards conservation studies. At present the numbers of harvest mice in the wild has increased quite significantly and are no longer classed as endangered for the time being. Long term conservation projects will be developed using the data from this thesis to prevent the numbers

dramatically reducing as they have done in the past. There are also other research groups investigating conservation strategies for other endangered species such as the Syrian hamster, which is nearly extinct in the wild, who will be able to use similar strategies to preserve these species. The mouse lemur conservation status is also currently classed as vulnerable in the wild with numbers falling due to deforestation of their natural habitat. Using data from this project, behavioural studies will be carried out to determine if this protein is related to chemical signalling and if so conservation projects will begin to hopefully increase the number of wild mouse lemurs. Again, if successful, this approach could be used by other research groups and charities focusing on conserving vulnerable species.

Many animal welfare and conservation projects are set up based on information provided by behavioural laboratories so to set up projects using protein chemistry data will be quite unique. Having protein chemistry data to complement behavioural data will enhance the understanding of current complex behavioural observations. It also highlights the importance of collaboration between biochemistry and behavioural labs in the approach to chemical communication. This partnership makes understanding the role pheromones in the complex social behaviour of the animal kingdom more achievable.

Chapter 7: References

Aebersold R & Mann M. (2003). Mass spectrometry-based proteomics. *Nature* **422,** 198-207.

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ. (1990). Basic local alignment search tool. *Journal of molecular biology* **215,** 403-410.

Archunan G & Ramesh Kumar K. (2013). 1-iodoundecane, an estrus indicating urinary chemo signal in bovine (Bos taurus). *Journal of Veterinary Science and Technology* **3**.

Armstrong SD, Robertson DH, Cheetham SA, Hurst JL & Beynon RJ. (2005). Structural and functional differences in isoforms of mouse major urinary proteins: a male-specific protein that preferentially binds a male pheromone. *The Biochemical journal* **391,** 343-350.

Arteaga L, Bautista A, Martínez-Gómez M, Nicolás L & Hudson R. (2008). Scent marking, dominance and serum testosterone levels in male domestic rabbits. *Physiology & Behavior* **94,** 510-515.

Atherton JC. (2006). Regulation of fluid and electrolyte balance by the kidney. *Anaesthesia & Intensive Care Medicine* **7,** 227-233.

Atherton JC. (2012). Role of the kidney in acid–base balance. *Anaesthesia & Intensive Care Medicine* **13,** 299-301.

Atkeson TD, Marchinton RL & Miller KV. (1988). Vocalizations of white-tailed deer. *American Midland Naturalist* **120,** 194-200.

Aujard F. (1997). Effect of vomeronasal organ removal on male socio-sexual responses to female in a prosimian primate (Microcebus murinus). *Physiol Behav* **62,** 1003-1008.

Bacchini A, Gaetani E & Cavaggioni A. (1992). Pheromone binding proteins of the mouse, Mus musculus. *Experientia* **48,** 419-421.

Bailey CJ. (2011). Renal glucose reabsorption inhibitors to treat diabetes. *Trends Pharmacol Sci* **32,** 63-71.

Beg OU, von Bahr-Lindstrom H, Zaidi ZH & Jornvall H. (1986). A camel milk whey protein rich in half-cystine. Primary structure, assessment of variations, internal repeat

patterns, and relationships with neurophysin and other active polypeptides. *European journal of biochemistry / FEBS* **159,** 195-201.

Bertsch A, Leinenbach A, Pervukhin A, Lubeck M, Hartmer R, Baessmann C, Elnakady YA, Muller R, Bocker S, Huber CG & Kohlbacher O. (2009). De novo peptide sequencing by tandem MS using complementary CID and electron transfer dissociation. *Electrophoresis* **30,** 3736-3747.

Beynon R, Hurst J, Turton M, Robertson DL, Armstrong S, Cheetham S, Simpson D, MacNicoll A & Humphries R. (2008). Urinary Lipocalins in Rodenta:is there a Generic Model? In *Chemical Signals in Vertebrates 11*, ed. Hurst J, Beynon R, Roberts SC & Wyatt T, pp. 37-49. Springer New York.

Beynon RJ & Hurst JL. (2003). Multiple roles of major urinary proteins in the house mouse, Mus domesticus. *Biochemical Society transactions* **31,** 142-146.

Beynon RJ & Hurst JL. (2004). Urinary proteins and the modulation of chemical scents in mice and rats. *Peptides* **25,** 1553-1563.

Beynon RJ, Veggerby C, Payne CE, Robertson DH, Gaskell SJ, Humphries RE & Hurst JL. (2002). Polymorphism in major urinary proteins: molecular heterogeneity in a wild mouse population. *Journal of chemical ecology* **28,** 1429-1446.

Bian X, Liu D, Zeng H, Zhang G, Wei R & Hou R. (2013). Exposure to odors of rivals enhances sexual motivation in male giant pandas. *PloS one* **8,** e69889.

Bischoff R, Lepage P, Jaquinod M, Cauet G, Acker-Klein M, Clesse D, Laporte M, Bayol A, Van Dorsselaer A & Roitsch C. (1993). Sequence-specific deamidation: isolation and biochemical characterization of succinimide intermediates of recombinant hirudin. *Biochemistry* **32,** 725-734.

Bocskei Z, Findlay JB, North AC, Phillips SE, Somers WS, Wright CE, Lionetti C, Tirindelli R & Cavaggioni A. (1991). Crystallization of and preliminary X-ray data for the mouse major urinary protein and rat alpha-2u globulin. *Journal of molecular biology* **218,** 699-701.

Bocskei Z, Groom CR, Flower DR, Wright CE, Phillips SE, Cavaggioni A, Findlay JB & North AC. (1992). Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. *Nature* **360,** 186-188.

Bodmer RE. (1991). Book review: Lemurs of Madagascar and the Comoros: the IUCN red data book. By C. Harcourt & J. Thornback. IUCN, Cambridge. 1990. 240 pp. ISBN 2 88032 957 4. Price: £18·00, $36·00. *Biological Conservation* **57,** 236.

Booth KK & Katz LS. (2000). Role of the vomeronasal organ in neonatal offspring recognition in sheep. *Biology of reproduction* **63,** 953-958.

Brashares JS & Arcese P. (1999). Scent marking in a territorial African antelope: I. The maintenance of borders between male oribi. *Animal behaviour* **57,** 1-10.

Brennan PA & Zufall F. (2006). Pheromonal communication in vertebrates. *Nature* **444,** 308-315.

Briand L, Blon F, Trotier D & Pernollet JC. (2004). Natural ligands of hamster aphrodisin. *Chemical senses* **29,** 425-430.

Briscoe BK, Lewis MA & Parrish SE. (2002). Home Range Formation in Wolves Due to Scent Marking. *Bulletin of Mathematical Biology* **64,** 261-284.

Brouette-Lahlou I, Godinot F & Vernet-Maury E. (1999). The mother rat's vomeronasal organ is involved in detection of dodecyl propionate, the pup's preputial gland pheromone. *Physiol Behav* **66,** 427-436.

Brown RE & Macdonald DW. (1985). *Social odours in mammals edited by Richard E. Brown and David W. MacDonald*. Oxford: Clarendon Press, 1985.

Bruce HM. (1960). A block to pregnancy in the mouse caused by proximity of strange males. *Journal of reproduction and fertility* **1,** 96-103.

Brun V, Dupuis A, Adrait A, Marcellin M, Thomas D, Court M, Vandenesch F & Garin J. (2007). Isotope-labeled protein standards: toward absolute quantitative proteomics. *Molecular & cellular proteomics : MCP* **6,** 2139-2149.

Brun V, Masselon C, Garin J & Dupuis A. (2009). Isotope dilution strategies for absolute quantitative proteomics. *Journal of proteomics* **72,** 740-749.

Burgener N, Dehnhard M, Hofer H & East ML. (2009). Does anal gland scent signal identity in the spotted hyaena? *Animal behaviour* **77,** 707-715.

Caligioni CS. (2009). Assessing reproductive status/stages in mice. *Current protocols in neuroscience / editorial board, Jacqueline N Crawley [et al]* **Appendix 4,** Appendix 4I.

Campbell SM, Rosen JM, Hennighausen LG, Strech-Jurk U & Sippel AE. (1984). Comparison of the whey acidic protein genes of the rat and mouse. *Nucleic acids research* **12,** 8685-8697.

Cate Ct, Bruins WS, den Ouden J, Egberts T, Neevel H, Spierings M, van der Burg K & Brokerhof AW. (2009). Tinbergen revisited: a replication and extension of experiments on the beak colour preferences of herring gull chicks. *Animal behaviour* **77,** 795-802.

Cavaggioni A & Mucignat-Caretta C. (2000). Major urinary proteins, alpha(2U)-globulins and aphrodisin. *Biochimica et biophysica acta* **1482,** 218-228.

Cechova D & Muszynska G. (1970). Role of lysine 18 in active center of cow colostrum trypsin inhibitor. *FEBS letters* **8,** 84-86.

Chamero P, Marton TF, Logan DW, Flanagan K, Cruz JR, Saghatelian A, Cravatt BF & Stowers L. (2007). Identification of protein pheromones that promote aggressive behaviour. *Nature* **450,** 899-902.

Cheetham SA, Smith AL, Armstrong SD, Beynon RJ & Hurst JL. (2009). Limited variation in the major urinary proteins of laboratory mice. *Physiol Behav* **96,** 253-261.

Cheetham SA, Thom MD, Jury F, Ollier WE, Beynon RJ & Hurst JL. (2007). The genetic basis of individual-recognition signals in the mouse. *Current biology : CB* **17,** 1771-1777.

Chelius D & Bondarenko PV. (2002). Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *Journal of proteome research* **1,** 317-323.

Chen J, Wang M & Turko IV. (2013). Quantification of amyloid precursor protein isoforms using quantification concatamer internal standard. *Analytical chemistry* **85,** 303-307.

Clancy AN, Macrides F, Singer AG & Agosta WC. (1984). Male hamster copulatory responses to a high molecular weight fraction of vaginal discharge: effects of vomeronasal organ removal. *Physiol Behav* **33,** 653-660.

Clauss A, Lilja H & Lundwall A. (2005). The evolution of a genetic locus encoding small serine proteinase inhibitors. *Biochemical and biophysical research communications* **333,** 383-389.

Cohen-Tannoudji J, Lavenet C, Locatelli A, Tillet Y & Signoret JP. (1989). Non-involvement of the accessory olfactory system in the LH response of anoestrous ewes to male odour. *Journal of reproduction and fertility* **86,** 135-144.

Cottrell JS. (2011). Protein identification using MS/MS data. *Journal of proteomics* **74,** 1842-1851.

Darwish Marie A, Veggerby C, Robertson DH, Gaskell SJ, Hubbard SJ, Martinsen L, Hurst JL & Beynon RJ. (2001). Effect of polymorphisms on ligand binding by mouse major urinary proteins. *Protein science : a publication of the Protein Society* **10,** 411-417.

David Smith JL, McDougal C & Miquelle D. (1989). Scent marking in free-ranging tigers,Panthera tigris. *Animal behaviour* **37, Part 1,** 1-10.

Delgadillo JA, Gelez H, Ungerfeld R, Hawken PA & Martin GB. (2009). The 'male effect' in sheep and goats--revisiting the dogmas. *Behavioural brain research* **200,** 304-314.

Devinoy E, Hubert C, Schaerer E, Houdebine LM & Kraehenbuhl JP. (1988). Sequence of the rabbit whey acidic protein cDNA. *Nucleic acids research* **16,** 8180.

Donpudsa S, Soderhall I, Rimphanitchayakit V, Cerenius L, Tassanakajon A & Soderhall K. (2010). Proteinase inhibitory activities of two two-domain Kazal proteinase inhibitors from the freshwater crayfish Pacifastacus leniusculus and the importance of the P(2) position in proteinase inhibitory activity. *Fish & shellfish immunology* **29,** 716-723.

Dorries KM, Adkins-Regan E & Halpern BP. (1995). Olfactory sensitivity to the pheromone, androstenone, is sexually dimorphic in the pig. *Physiol Behav* **57,** 255-259.

Dorries KM, Adkins-Regan E & Halpern BP. (1997). Sensitivity and behavioral responses to the pheromone androstenone are not mediated by the vomeronasal organ in domestic pigs. *Brain, behavior and evolution* **49,** 53-62.

Dorus S, Evans PD, Wyckoff GJ, Choi SS & Lahn BT. (2004). Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nature genetics* **36,** 1326-1329.

Drea CM, Vignieri SN, Kim HS, Weldele ML & Glickman SE. (2002). Responses to olfactory stimuli in spotted hyenas (Crocuta crocuta): II. Discrimination of conspecific scent. *Journal of comparative psychology (Washington, DC : 1983)* **116,** 342-349.

Drenth J, Low BW, Richardson JS & Wright CS. (1980). The toxin-agglutinin fold. A new group of small protein structures organized around a four-disulfide core. *The Journal of biological chemistry* **255,** 2652-2655.

Drickamer LC. (1977). Delay of sexual maturation in female house mice by exposure to grouped females or urine from grouped females. *Journal of reproduction and fertility* **51,** 77-81.

Drickamer LC. (1986). Effects of urine from females in oestrus on puberty in female mice. *Journal of reproduction and fertility* **77,** 613-622.

Dulac C & Axel R. (1995). A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83,** 195-206.

Dulac C & Torello AT. (2003). Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nature reviews Neuroscience* **4,** 551-562.

Eyers CE, Lawless C, Wedge DC, Lau KW, Gaskell SJ & Hubbard SJ. (2011). CONSeQuence: prediction of reference peptides for absolute quantitative proteomics using consensus machine learning approaches. *Molecular & cellular proteomics : MCP* **10,** M110.003384.

Eyers CE, Simpson DM, Wong SC, Beynon RJ & Gaskell SJ. (2008). QCAL--a novel standard for assessing instrument conditions for proteome analysis. *Journal of the American Society for Mass Spectrometry* **19,** 1275-1280.

Fay FH. (1982). Ecology and Biology of the Pacific Walrus, Odobenus rosmarus divergens Illiger. *North American Fauna***,** 1-279.

Fietz J. (1999). Mating system of Microcebus murinus. *American journal of primatology* **48,** 127-133.

Finlayson JS, Asofsky R, Potter M & Runner CC. (1965). Major urinary protein complex of normal mice: origin. *Science (New York, NY)* **149,** 981-982.

Finlayson JS & Baumann CA. (1957). Protein-bound sterols in rodent urine. *The American journal of physiology* **190,** 297-302.

Flower DR. (1996). The lipocalin protein family: structure and function. *The Biochemical journal* **318 (Pt 1),** 1-14.

Flower DR, North AC & Attwood TK. (1993). Structure and sequence relationships in the lipocalins and related proteins. *Protein science : a publication of the Protein Society* **2,** 753-761.

Flower DR, North AC & Sansom CE. (2000). The lipocalin protein family: structural and sequence overview. *Biochimica et biophysica acta* **1482,** 9-24.

Freeman-Gallant CR, Meguerdichian M, Wheelwright NT & Sollecito SV. (2003). Social pairing and female mating fidelity predicted by restriction fragment length polymorphism similarity at the major histocompatibility complex in a songbird. *Molecular ecology* **12,** 3077-3083.

Friedman DB, Hoving S & Westermeier R. (2009). Isoelectric focusing and two-dimensional gel electrophoresis. *Methods in enzymology* **463,** 515-540.

Geiger T & Clarke S. (1987). Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *The Journal of biological chemistry* **262,** 785-794.

Gosling LM. (1982). A Reassessment of the Function of Scent Marking in Territories. *Zeitschrift für Tierpsychologie* **60,** 89-118.

Gosling LM & McKay HV. (1990). Competitor assessment by scent matching: an experimental test. *Behav Ecol Sociobiol* **26,** 415-420.

Gosling LM & Roberts SC. (2001a). Scent-marking by male mammals: Cheat-proof signals to competitors and mates. In *Advances in the Study of Behavior*, ed. Peter J. B. Slater JSRCTS & Timothy JR, pp. 169-217. Academic Press.

Gosling LM & Roberts SC. (2001b). Testing ideas about the function of scent marks in territories from spatial patterns. *Animal behaviour* **62,** F7-F10.

Gosling LM & Stone RD. (1990). Mutual avoidance by European moles *Talpa europaea.* In *Chemical Signals in Vertebrates 5*

ed. .W. Macdonald DMl-S, and S.E. Natynczuk., pp. 367-377. Oxford University Press, Oxford and New York.

Gregoire C, Rosinski-Chupin I, Rabillon J, Alzari PM, David B & Dandeu JP. (1996). cDNA cloning and sequencing reveal the major horse allergen Equ c1 to be a glycoprotein member of the lipocalin superfamily. *The Journal of biological chemistry* **271,** 32951-32959.

Grus WE, Shi P, Zhang YP & Zhang J. (2005). Dramatic variation of the vomeronasal pheromone receptor gene repertoire among five orders of placental and marsupial mammals. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 5767-5772.

Grus WE & Zhang J. (2004). Rapid turnover and species-specificity of vomeronasal pheromone receptor genes in mice and rats. *Gene* **340,** 303-312.

Gutierrez-Garcia AG, Contreras CM, Mendoza-Lopez MR, Garcia-Barradas O & Cruz-Sanchez JS. (2007). Urine from stressed rats increases immobility in receptor rats forced to swim: role of 2-heptanone. *Physiol Behav* **91,** 166-172.

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH & Aebersold R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology* **17,** 994-999.

Hagey L & MacDonald E. (2003). Chemical cues identify gender and individuality in Giant pandas (Ailuropoda melanoleuca). *Journal of chemical ecology* **29,** 1479-1488.

Hagiwara K, Kikuchi T, Endo Y, Huqun, Usui K, Takahashi M, Shibata N, Kusakabe T, Xin H, Hoshi S, Miki M, Inooka N, Tokue Y & Nukiwa T. (2003). Mouse SWAM1 and SWAM2 are antibacterial proteins composed of a single whey acidic protein motif. *Journal of immunology (Baltimore, Md : 1950)* **170,** 1973-1979.

Harder JD, Jackson LM & Koester DC. (2008). Behavioral and reproductive responses of female opossums to volatile and nonvolatile components of male suprasternal gland secretion. *Hormones and Behavior* **54,** 741-747.

He J, Ma L, Kim S, Nakai J & Yu CR. (2008). Encoding gender and individual information in the mouse vomeronasal organ. *Science (New York, NY)* **320,** 535-538.

Hedstrom L. (2002). Serine protease mechanism and specificity. *Chemical reviews* **102,** 4501-4524.

Hennighausen LG & Sippel AE. (1982). Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins. *Nucleic acids research* **10,** 2677-2684.

Hiemstra PS, Maassen RJ, Stolk J, Heinzel-Wieland R, Steffens GJ & Dijkman JH. (1996). Antibacterial activity of antileukoprotease. *Infection and immunity* **64,** 4520-4524.

Hohenbrink P, Mundy NI, Zimmermann E & Radespiel U. (2013). First evidence for functional vomeronasal 2 receptor genes in primates. *Biology letters* **9,** 20121006.

Hu Q, Noll RJ, Li H, Makarov A, Hardman M & Graham Cooks R. (2005). The Orbitrap: a new mass spectrometer. *Journal of mass spectrometry : JMS* **40,** 430-443.

Hudson R & Distel H. (1986). Pheromonal release of suckling in rabbits does not depend on the vomeronasal organ. *Physiol Behav* **37,** 123-128.

Humphries RE, Robertson DH, Beynon RJ & Hurst JL. (1999). Unravelling the chemical basis of competitive scent marking in house mice. *Animal behaviour* **58,** 1177-1190.

Hurle B, Swanson W & Green ED. (2007). Comparative sequence analyses reveal rapid and divergent evolutionary changes of the WFDC locus in the primate lineage. *Genome research* **17,** 276-286.

Hurst JL. (1990). Urine marking in populations of wild house mice Mus domesticus rutty. II. Communication between females. *Animal behaviour* **40,** 223-232.

Hurst JL. (1993). The priming effects of urine substrate marks on interactions between male house mice, Mus musculus domesticus Schwarz & Schwarz. *Animal behaviour* **45,** 55-81.

Hurst JL. (2009). Female recognition and assessment of males through scent. *Behavioural brain research* **200,** 295-303.

Hurst JL & Beynon RJ. (2004). Scent wars: the chemobiology of competitive signalling in mice. *BioEssays : news and reviews in molecular, cellular and developmental biology* **26,** 1288-1298.

Hurst JL, Payne CE, Nevison CM, Marie AD, Humphries RE, Robertson DH, Cavaggioni A & Beynon RJ. (2001). Individual recognition in mice mediated by major urinary proteins. *Nature* **414,** 631-634.

Hurst JL, Robertson DHL, Tolladay U & Beynon RJ. (1998). Proteins in urine scent marks of male house mice extend the longevity of olfactory signals. *Animal behaviour* **55,** 1289-1297.

Ilmonen P, Penn DJ, Damjanovich K, Morrison L, Ghotbi L & Potts WK. (2007). Major histocompatibility complex heterozygosity reduces fitness in experimentally infected mice. *Genetics* **176,** 2501-2508.

Ishiwaka R, Kinoshita Y, Satou H, Kakihara H & Masuda Y. (2010). Overwintering in nests on the ground in the harvest mouse. *Landscape Ecol Eng* **6,** 335-342.

Ishiwaka R & Mori T. (1999). Early development of climbing skills in harvest mice. *Animal behaviour* **58,** 203-209.

Janeway CA, Jr., Mamula MJ & Rudensky A. (1993). Rules for peptide presentation by MHC class II molecules. *International reviews of immunology* **10,** 301-311.

Jang T, Singer AG & O'Connell RJ. (2001). Induction of c-fos in hamster accessory olfactory bulbs by natural and cloned aphrodisin. *Neuroreport* **12,** 449-452.

Jemiolo B, Harvey S & Novotny M. (1986). Promotion of the Whitten effect in female mice by synthetic analogs of male urinary constituents. *Proceedings of the National Academy of Sciences of the United States of America* **83,** 4576-4579.

Johnson D, al-Shawi R & Bishop JO. (1995). Sexual dimorphism and growth hormone induction of murine pheromone-binding proteins. *Journal of molecular endocrinology* **14,** 21-34.

Johnson RP. (1973). Scent marking in mammals. *Animal behaviour* **21,** 521-535.

Johnston RE, Chiang G & Tung C. (1994). The information in scent over-marks of golden hamsters. *Animal behaviour* **48,** 323-330.

Jordan WC & Bruford MW. (1998). New perspectives on mate choice and the MHC. *Heredity* **81 (Pt 3),** 239-245.

Jurges V, Kitzler J, Zingg R & Radespiel U. (2013). First insights into the social organisation of Goodman's mouse lemur (Microcebus lehilahytsara)--testing predictions from socio-ecological hypotheses in the Masoala hall of Zurich Zoo. *Folia primatologica; international journal of primatology* **84,** 32-48.

Kardasz S. (2009). The function of the nephron and the formation of urine. *Anaesthesia & Intensive Care Medicine* **10,** 265-270.

Karlson P & Luscher M. (1959). Pheromones': a new term for a class of biologically active substances. *Nature* **183,** 55-56.

Karr TL & Pitnick S. (1999). Sperm competition: defining the rules of engagement. *Current biology : CB* **9,** R787-790.

Kettenbach AN, Rush J & Gerber SA. (2011). Absolute quantification of protein and post-translational modification abundance with stable isotope-labeled synthetic peptides. *Nature protocols* **6,** 175-186.

Kimoto H, Haga S, Sato K & Touhara K. (2005). Sex-specific peptides from exocrine glands stimulate mouse vomeronasal sensory neurons. *Nature* **437,** 898-901.

Kimoto H, Sato K, Nodari F, Haga S, Holy TE & Touhara K. (2007). Sex- and strain-specific expression and vomeronasal activity of mouse ESP family peptides. *Current biology : CB* **17,** 1879-1884.

Kirkpatrick DS, Gerber SA & Gygi SP. (2005). The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods (San Diego, Calif)* **35,** 265-273.

Kito K, Ota K, Fujita T & Ito T. (2007). A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes. *Journal of proteome research* **6,** 792-800.

Klein J & Figueroa F. (1986). Evolution of the major histocompatibility complex. *Critical reviews in immunology* **6,** 295-386.

Klemm WR, Sherry CJ, Sis RF & Morris DL. (1984). Electrographic recording from bovine vomeronasal capsule under spontaneous and stimulated conditions. *Brain research bulletin* **12,** 275-282.

Knopf JL, Gallagher JF & Held WA. (1983). Differential, multihormonal regulation of the mouse major urinary protein gene family in the liver. *Molecular and cellular biology* **3,** 2232-2240.

Kober M, Trillmich F & Naguib M. (2007). Vocal mother–pup communication in guinea pigs: effects of call familiarity and female reproductive state. *Animal behaviour* **73,** 917-925.

Kocher T, Holzmann J, Mazanek M, Taus T, Mohring T, Ammerer G, Pichler P & Mechtler K. (2010). Optimized Quantification and Identification on LTQ Orbitrap using Peptide Labeling with Isobaric Tags. *Journal of Biomolecular Techniques : JBT* **21,** S61.

Konieczny A, Morgenstern JP, Bizinkauskas CB, Lilley CH, Brauer AW, Bond JF, Aalberse RC, Wallner BP & Kasaian MT. (1997). The major dog allergens, Can f 1 and Can f 2, are salivary lipocalin proteins: cloning and immunological characterization of the recombinant forms. *Immunology* **92,** 577-586.

Krauter K, Leinwand L, D'Eustachio P, Ruddle F & Darnell JE, Jr. (1982). Structural genes of the mouse major urinary protein are on chromosome 4. *The Journal of cell biology* **94,** 414-417.

Krieger J, Schmitt A, Lobel D, Gudermann T, Schultz G, Breer H & Boekhoff I. (1999). Selective activation of G protein subtypes in the vomeronasal organ upon stimulation with urine-derived compounds. *The Journal of biological chemistry* **274,** 4655-4662.

Kroner C, Breer H, Singer AG & O'Connell RJ. (1996). Pheromone-induced second messenger signaling in the hamster vomeronasal organ. *Neuroreport* **7,** 2989-2992.

Kumar KR, Archunan G, Jeyaraman R & Narasimhan S. (2000). Chemical characterization of bovine urine with special reference to oestrus. *Veterinary research communications* **24,** 445-454.

Lacey JC, Beynon RJ & Hurst JL. (2007). The importance of exposure to other male scents in determining competitive behaviour among inbred male mice. *Applied Animal Behaviour Science* **104,** 130-142.

Laemmli UK. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680-685.

Lange V, Picotti P, Domon B & Aebersold R. (2008). Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular systems biology* **4,** 222.

Larina IM, Pastushkova LK, Kireev KS & Grigoriev AI. (2013). Formation of the urine proteome of healthy humans. *Hum Physiol* **39,** 147-161.

Larsen M, Ressler SJ, Lu B, Gerdes MJ, McBride L, Dang TD & Rowley DR. (1998). Molecular cloning and expression of ps20 growth inhibitor. A novel WAP-type "four-disulfide core" domain protein expressed in smooth muscle. *The Journal of biological chemistry* **273,** 4574-4584.

Lebert D, Dupuis A, Garin J, Bruley C & Brun V. (2011). Production and use of stable isotope-labeled proteins for absolute quantitative proteomics. *Methods in molecular biology (Clifton, NJ)* **753,** 93-115.

Leinders-Zufall T, Brennan P, Widmayer P, S PC, Maul-Pavicic A, Jager M, Li XH, Breer H, Zufall F & Boehm T. (2004). MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science (New York, NY)* **306,** 1033-1037.

Levy F, Kendrick KM, Goode JA, Guevara-Guzman R & Keverne EB. (1995). Oxytocin and vasopressin release in the olfactory bulb of parturient ewes: changes with maternal experience and effects on acetylcholine, gamma-aminobutyric acid, glutamate and noradrenaline release. *Brain research* **669,** 197-206.

Lin DY, Zhang SZ, Block E & Katz LC. (2005). Encoding social signals in the mouse main olfactory bulb. *Nature* **434,** 470-477.

Loebel D, Scaloni A, Paolini S, Fini C, Ferrara L, Breer H & Pelosi P. (2000). Cloning, post-translational modifications, heterologous expression and ligand-binding of boar salivary lipocalin. *The Biochemical journal* **350 Pt 2,** 369-379.

Logan DW, Marton TF & Stowers L. (2008). Species specificity in major urinary proteins by parallel evolution. *PloS one* **3,** e3280.

Lucke C, Franzoni L, Abbate F, Lohr F, Ferrari E, Sorbi RT, Ruterjans H & Spisni A. (1999). Solution structure of a recombinant mouse major urinary protein. *European journal of biochemistry / FEBS* **266,** 1210-1218.

Lundgren DH, Hwang SI, Wu L & Han DK. (2010). Role of spectral counting in quantitative proteomics. *Expert review of proteomics* **7,** 39-53.

Lundwall A & Clauss A. (2011). Genes encoding WFDC- and Kunitz-type protease inhibitor domains: are they related? *Biochemical Society transactions* **39,** 1398-1402.

Lutermann H, Schmelting B, Radespiel U, Ehresmann P & Zimmermann E. (2006). The role of survival for the evolution of female philopatry in a solitary forager, the grey mouse lemur (Microcebus murinus). *Proceedings Biological sciences / The Royal Society* **273,** 2527-2533.

Ma B & Johnson R. (2012). De novo sequencing and homology searching. *Molecular & cellular proteomics : MCP* **11,** O111.014902.

Ma W, Miao Z & Novotny MV. (1998). Role of the adrenal gland and adrenal-mediated chemosignals in suppression of estrus in the house mouse: the lee-boot effect revisited. *Biology of reproduction* **59,** 1317-1320.

Magert HJ, Cieslak A, Alkan O, Luscher B, Kauffels W & Forssmann WG. (1999). The golden hamster aphrodisin gene. Structure, expression in parotid glands of female animals, and comparison with a similar murine gene. *The Journal of biological chemistry* **274,** 444-450.

Marchese S, Pes D, Scaloni A, Carbone V & Pelosi P. (1998). Lipocalins of boar salivary glands binding odours and pheromones. *European journal of biochemistry / FEBS* **252,** 563-568.

Marcondes FK, Bianchi FJ & Tanno AP. (2002). Determination of the estrous cycle phases of rats: some helpful considerations. *Brazilian journal of biology = Revista brasleira de biologia* **62,** 609-614.

Mateo JM. (2006). Development of individually distinct recognition cues. *Developmental psychobiology* **48,** 508-519.

Matsunami H & Buck LB. (1997). A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* **90,** 775-784.

McAlhany SJ, Ressler SJ, Larsen M, Tuxhorn JA, Yang F, Dang TD & Rowley DR. (2003). Promotion of angiogenesis by ps20 in the differential reactive stroma prostate cancer xenograft model. *Cancer research* **63,** 5859-5865.

McCrudden MT, Dafforn TR, Houston DF, Turkington PT & Timson DJ. (2008). Functional domains of the human epididymal protease inhibitor, eppin. *The FEBS journal* **275,** 1742-1750.

Mombaerts P. (2004). Genes and ligands for odorant, vomeronasal and taste receptors. *Nature reviews Neuroscience* **5,** 263-278.

Moreau T, Baranger K, Dade S, Dallet-Choisy S, Guyot N & Zani ML. (2008). Multifaceted roles of human elafin and secretory leukocyte proteinase inhibitor (SLPI), two serine protease inhibitors of the chelonianin family. *Biochimie* **90,** 284-295.

Mucignat-Caretta C, Caretta A & Cavaggioni A. (1995). Acceleration of puberty onset in female mice by male urinary proteins. *The Journal of physiology* **486 (Pt 2),** 517-522.

Mucignat-Caretta C, Cavaggioni A & Caretta A. (2004). Male urinary chemosignals differentially affect aggressive behavior in male mice. *Journal of chemical ecology* **30,** 777-791.

Mucignat Caretta C, Caretta A & Cavaggioni A. (1995). Pheromonally accelerated puberty is enhanced by previous experience of the same stimulus. *Physiol Behav* **57,** 901-903.

Mudge JM, Armstrong SD, McLaren K, Beynon RJ, Hurst JL, Nicholson C, Robertson DH, Wilming LG & Harrow JL. (2008). Dynamic instability of the major urinary protein gene family revealed by genomic and phenotypic comparisons between C57 and 129 strain mice. *Genome biology* **9,** R91.

Murata K, Tamogami S, Itou M, Ohkubo Y, Wakabayashi Y, Watanabe H, Okamura H, Takeuchi Y & Mori Y. (2014). Identification of an Olfactory Signal Molecule that Activates the Central Regulator of Reproduction in Goats. *Current Biology* **24,** 681-686.

Mykytowycz R. (1965). Further observations on the territorial function and histology of the submandibular cutaneous (chin) glands in the rabbit, Oryctolagus cuniculus (L.). *Animal behaviour* **13,** 400-412.

Nanavati D, Gucek M, Milne JL, Subramaniam S & Markey SP. (2008). Stoichiometry and absolute quantification of proteins with mass spectrometry using fluorescent and

isotope-labeled concatenated peptide standards. *Molecular & cellular proteomics : MCP* **7,** 442-447.

Naresh CN, Hayen A, Weening A, Craig JC & Chadban SJ. (2013). Day-to-day variability in spot urine albumin-creatinine ratio. *American journal of kidney diseases : the official journal of the National Kidney Foundation* **62,** 1095-1101.

Nekaris KAI. (2011). Lemurs of Madagascar (3rd edition) by Russell A. Mittermeier, Edward E. Louis Jr., Matthew Richardson, Christoph Schwitzer, Olivier Langrand, Anthony B. Rylands et al.. (2010), pp. 762, Conservation International, Arlington, USA. ISBN 9781934151235 (pbk), USD 55.00. *Oryx* **45,** 458-459.

Neuhaus OW. (1986). Renal reabsorption of low molecular weight proteins in adult male rats: alpha 2u-globulin. *Proceedings of the Society for Experimental Biology and Medicine Society for Experimental Biology and Medicine (New York, NY)* **182,** 531-539.

Nevison CM, Barnard CJ, Beynon RJ & Hurst JL. (2000). The consequences of inbreeding for recognizing competitors. *Proceedings Biological sciences / The Royal Society* **267,** 687-694.

Nodari F, Hsu FF, Fu X, Holekamp TF, Kao LF, Turk J & Holy TE. (2008). Sulfated steroids as natural ligands of mouse pheromone-sensing neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **28,** 6407-6418.

Novotny M, Harvey S, Jemiolo B & Alberts J. (1985). Synthetic pheromones that promote inter-male aggression in mice. *Proceedings of the National Academy of Sciences of the United States of America* **82,** 2059-2061.

Novotny M, Jemiolo B, Harvey S, Wiesler D & Marchlewska-Koj A. (1986). Adrenal-mediated endogenous metabolites inhibit puberty in female mice. *Science (New York, NY)* **231,** 722-725.

Novotny MV. (2003). Pheromones, binding proteins and receptor responses in rodents. *Biochemical Society transactions* **31,** 117-122.

Novotny MV, Jemiolo B, Wiesler D, Ma W, Harvey S, Xu F, Xie TM & Carmack M. (1999). A unique urinary constituent, 6-hydroxy-6-methyl-3-heptanone, is a pheromone that accelerates puberty in female mice. *Chemistry & biology* **6,** 377-383.

O'Rand MG, Widgren EE, Hamil KG, Silva EJ & Richardson RT. (2011). Functional studies of eppin. *Biochemical Society transactions* **39,** 1447-1449.

Ohno S. (1970). *Evolution by gene duplication by Susumu Ohno*. Berlin: Springer, 1970.

OlsEn KH, Grahn M, Lohm J & Langefors A. (1998). MHC and kin discrimination in juvenile Arctic charr, Salvelinus alpinus (L.). *Animal behaviour* **56,** 319-327.

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A & Mann M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics : MCP* **1,** 376-386.

Ong SE & Mann M. (2006). A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature protocols* **1,** 2650-2660.

Ong SE & Mann M. (2007). Stable isotope labeling by amino acids in cell culture for quantitative proteomics. *Methods in molecular biology (Clifton, NJ)* **359,** 37-52.

Papes F, Logan DW & Stowers L. (2010). The vomeronasal organ mediates interspecies defensive behaviors through detection of protein pheromone homologs. *Cell* **141,** 692-703.

Parkening TA, Collins TJ & Smith ER. (1982). Plasma and pituitary concentrations of LH, FSH, and prolactin in aging C57BL/6 mice at various times of the estrous cycle. *Neurobiology of aging* **3,** 31-35.

Penn D & Potts W. (1998). MHC-disassortative mating preferences reversed by cross-fostering. *Proceedings Biological sciences / The Royal Society* **265,** 1299-1306.

Perkins DN, Pappin DJ, Creasy DM & Cottrell JS. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20,** 3551-3567.

Perret M. (1992). Environmental and social determinants of sexual function in the male lesser mouse lemur (Microcebus murinus). *Folia primatologica; international journal of primatology* **59,** 1-25.

Perret M & Schilling A. (1987). Intermale sexual effect elicited by volatile urinary ether extract inMicrocebus murinus (Prosimian, Primates). *Journal of chemical ecology* **13,** 495-507.

Perret M & Schilling A. (1995). Sexual responses to urinary chemosignals depend on photoperiod in a male primate. *Physiol Behav* **58,** 633-639.

Peters RP & Mech LD. (1975). Scent-marking in wolves. *American scientist* **63,** 628-637.

Petrulis A. (2013). Chemosignals, hormones and mammalian reproduction. *Hormones and Behavior* **63,** 723-741.

Picotti P & Aebersold R. (2012). Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature methods* **9,** 555-566.

Picotti P, Rinner O, Stallmach R, Dautel F, Farrah T, Domon B, Wenschuh H & Aebersold R. (2010). High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nature methods* **7,** 43-46.

Pomerantz SM & Clemens LG. (1981). Ultrasonic vocalizations in male deer mice (Peromyscus maniculatus bairdi): Their role in male sexual behavior. *Physiology & Behavior* **27,** 869-872.

Pratt JM, Simpson DM, Doherty MK, Rivers J, Gaskell SJ & Beynon RJ. (2006). Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nature protocols* **1,** 1029-1043.

Radespiel U. (2000). Sociality in the gray mouse lemur (Microcebus murinus) in northwestern Madagascar. *American journal of primatology* **51,** 21-40.

Radespiel U, Cepok S, Zietemann V & Zimmermann E. (1998). Sex-specific usage patterns of sleeping sites in grey mouse lemurs (Microcebus murinus) in northwestern Madagascar. *American journal of primatology* **46,** 77-84.

Radespiel U, Dal Secco V, Drögemüller C, Braune P, Labes E & Zimmermann E. (2002). Sexual selection, multiple mating and paternity in grey mouse lemurs, Microcebus murinus. *Animal behaviour* **63,** 259-268.

Radespiel U, Ratsimbazafy JH, Rasoloharijaona S, Raveloson H, Andriaholinirina N, Rakotondravony R, Randrianarison RM & Randrianambinina B. (2012). First indications of a highland specialist among mouse lemurs (Microcebus spp.) and evidence for a new mouse lemur species from eastern Madagascar. *Primates; journal of primatology* **53,** 157-170.

Radespiel U & Zimmermann E. (2001). Female dominance in captive gray mouse lemurs (Microcebus murinus). *American journal of primatology* **54,** 181-192.

Rajagopal T, Archunan G, Geraldine P & Balasundaram C. (2010). Assessment of dominance hierarchy through urine scent marking and its chemical constituents in male blackbuck Antelope cervicapra, a critically endangered species. *Behavioural processes* **85,** 58-67.

Rajanarayanan S & Archunan G. (2011). Identification of urinary sex pheromones in female buffaloes and their influence on bull reproductive behaviour. *Research in veterinary science* **91,** 301-305.

Ramm SA, Cheetham SA & Hurst JL. (2008). Encoding choosiness: female attraction requires prior physical contact with individual male scents in mice. *Proceedings Biological sciences / The Royal Society* **275,** 1727-1735.

Ranganathan S, Simpson KJ, Shaw DC & Nicholas KR. (1999). The whey acidic protein family: a new signature motif and three-dimensional structure by comparative modeling. *Journal of molecular graphics & modelling* **17,** 106-113, 134-106.

Rasmussen LE, Lee TD, Zhang A, Roelofs WL & Daves GD, Jr. (1997). Purification, identification, concentration and bioactivity of (Z)-7-dodecen-1-yl acetate: sex pheromone of the female Asian elephant, Elephas maximus. *Chemical senses* **22,** 417-437.

Rich TJ & Hurst JL. (1998). Scent marks as reliable signals of the competitive ability of mates. *Animal behaviour* **56,** 727-735.

Rivers J, McDonald L, Edwards IJ & Beynon RJ. (2008). Asparagine deamidation and the role of higher order protein structure. *Journal of proteome research* **7,** 921-927.

Rivers J, Simpson DM, Robertson DH, Gaskell SJ & Beynon RJ. (2007). Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. *Molecular & cellular proteomics : MCP* **6,** 1416-1427.

Roberts SA, Davidson AJ, McLean L, Beynon RJ & Hurst JL. (2012). Pheromonal induction of spatial learning in mice. *Science (New York, NY)* **338,** 1462-1465.

Roberts SA, Simpson DM, Armstrong SD, Davidson AJ, Robertson DH, McLean L, Beynon RJ & Hurst JL. (2010). Darcin: a male pheromone that stimulates female memory and sexual attraction to an individual male's odour. *BMC biology* **8,** 75.

Roberts SC, Gosling LM, Carter V & Petrie M. (2008). MHC-correlated odour preferences in humans and the use of oral contraceptives. *Proceedings Biological sciences / The Royal Society* **275,** 2715-2722.

Robertson DH, Beynon RJ & Evershed RP. (1993). Extraction, characterization, and binding analysis of two pheromonally active ligands associated with major urinary protein of house mouse (Mus musculus). *Journal of chemical ecology* **19,** 1405-1416.

Robertson DH, Cox KA, Gaskell SJ, Evershed RP & Beynon RJ. (1996). Molecular heterogeneity in the Major Urinary Proteins of the house mouse Mus musculus. *The Biochemical journal* **316 (Pt 1),** 265-272.

Robertson DH, Hurst JL, Bolgar MS, Gaskell SJ & Beynon RJ. (1997). Molecular heterogeneity of urinary proteins in wild house mouse populations. *Rapid communications in mass spectrometry : RCM* **11,** 786-790.

Robertson DH, Hurst JL, Searle JB, Gunduz I & Beynon RJ. (2007). Characterization and comparison of major urinary proteins from the house mouse, Mus musculus domesticus, and the aboriginal mouse, Mus macedonicus. *Journal of chemical ecology* **33,** 613-630.

Robertson DL, Marie A, Veggerby C, Hurst J & Beynon R. (2001). Characteristics of Ligand Binding and Release by Major Urinary Proteins. In *Chemical Signals in Vertebrates 9*, ed. Marchlewska-Koj A, Lepri J & Müller-Schwarze D, pp. 169-176. Springer US.

Robinson AB & Rudd CJ. (1974). Deamidation of glutaminyl and asparaginyl residues in peptides and proteins. *Current topics in cellular regulation* **8,** 247-295.

Robinson NE, Robinson AB & Merrifield RB. (2001). Mass spectrometric evaluation of synthetic peptides as primary structure models for peptide and protein deamidation. *The journal of peptide research : official journal of the American Peptide Society* **57,** 483-493.

Rodriguez I, Del Punta K, Rothman A, Ishii T & Mombaerts P. (2002). Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors. *Nature neuroscience* **5,** 134-140.

Rodriguez I & Mombaerts P. (2002). Novel human vomeronasal receptor-like genes reveal species-specific families. *Current biology : CB* **12,** R409-411.

Ronco C. (2007). CRRT protects the kidney during acute renal failure. *The International journal of artificial organs* **30,** 279-280.

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A & Pappin DJ. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & cellular proteomics : MCP* **3,** 1154-1169.

Rouquier S, Blancher A & Giorgi D. (2000). The olfactory receptor gene repertoire in primates and mouse: evidence for reduction of the functional fraction in primates.

Proceedings of the National Academy of Sciences of the United States of America **97,** 2870-2874.

Rouquier S & Giorgi D. (2007). Olfactory receptor gene repertoires in mammals. *Mutation research* **616,** 95-102.

Roy AK & Neuhaus OW. (1966). Identification of rat urinary proteins by zone and immunoelectrophoresis. *Proceedings of the Society for Experimental Biology and Medicine Society for Experimental Biology and Medicine (New York, NY)* **121,** 894-899.

Rubenstein DA, Yin W & Frame MD. (2012). Chapter 12 - Flow Through the Kidney. In *Biofluid Mechanics*, ed. Rubenstein DA, Yin W & Frame MD, pp. 325-345. Academic Press, Boston.

Ruemke P & Thung PJ. (1964). IMMUNOLOGICAL STUDIES ON THE SEX-DEPENDENT PREALBUMIN IN MOUSE URINE AND ON ITS OCCURRENCE IN THE SERUM. *Acta endocrinologica* **47,** 156-164.

S H & R T. (1991). Genus *Micromys*. In *Tha Hanbook of British Mammals*

3edn, ed. Corbet G & S H, pp. 233–239. Wiley-Blackwell Oxford.

Saarelainen S, Taivainen A, Rytkonen-Nissinen M, Auriola S, Immonen A, Mantyjarvi R, Rautiainen J, Kinnunen T & Virtanen T. (2004). Assessment of recombinant dog allergens Can f 1 and Can f 2 for the diagnosis of dog allergy. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **34,** 1576-1582.

Sallenave JM, Shulmann J, Crossley J, Jordana M & Gauldie J. (1994). Regulation of secretory leukocyte proteinase inhibitor (SLPI) and elastase-specific inhibitor (ESI/elafin) in human airway epithelial cells by cytokines and neutrophilic enzymes. *American journal of respiratory cell and molecular biology* **11,** 733-741.

Schilling A, Perret M & Predine J. (1984). Sexual inhibition in a prosimian primate: a pheromone-like effect. *The Journal of endocrinology* **102,** 143-151.

Schmelting B, Zimmermann E, Berke O, Bruford MW & Radespiel U. (2007). Experience-dependent recapture rates and reproductive success in male grey mouse lemurs (Microcebus murinus). *American journal of physical anthropology* **133,** 743-752.

Schwensow N, Eberle M & Sommer S. (2008). Compatibility counts: MHC-associated mate choice in a wild promiscuous primate. *Proceedings Biological sciences / The Royal Society* **275,** 555-564.

Sherborne AL, Thom MD, Paterson S, Jury F, Ollier WE, Stockley P, Beynon RJ & Hurst JL. (2007). The genetic basis of inbreeding avoidance in house mice. *Current biology : CB* **17,** 2061-2066.

Siepen JA, Keevil EJ, Knight D & Hubbard SJ. (2007). Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *Journal of proteome research* **6,** 399-408.

Sillero-Zubiri C. (2004). Wolves: Behavior, Ecology and Conservation edited by L. David Mech and Luigi Boitani (2003), xvii + 448 pp., The University of Chicago Press, Chicago, USA. ISBN 0 226 51696 2 (hbk), $49.00. *Oryx* **38,** 234-235.

Simpson KJ, Bird P, Shaw D & Nicholas K. (1998). Molecular characterisation and hormone-dependent expression of the porcine whey acidic protein gene. *Journal of molecular endocrinology* **20,** 27-35.

Singer AG, Macrides F, Clancy AN & Agosta WC. (1986). Purification and analysis of a proteinaceous aphrodisiac pheromone from hamster vaginal discharge. *The Journal of biological chemistry* **261,** 13323-13326.

Singh N, Agrawal S & Rastogi AK. (1997). Infectious diseases and immunity: special reference to major histocompatibility complex. *Emerging infectious diseases* **3,** 41-49.

Smith W, Butler AJ, Hazell LA, Chapman MD, Pomes A, Nickels DG & Thomas WR. (2004). Fel d 4, a cat lipocalin allergen. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **34,** 1732-1738.

Spiegelberg T & Bishop JO. (1988). Tissue-specific gene expression in mouse hepatocytes cultured in growth-restricting medium. *Molecular and cellular biology* **8,** 3338-3344.

Stirling I, Calvert W & Spencer C. (1987). Evidence of stereotyped underwater vocalizations of male Atlantic walruses (Odobenus rosmarus rosmarus). *Canadian Journal of Zoology* **65,** 2311-2321.

Stopka P, Janotova K & Heyrovsky D. (2007). The advertisement role of major urinary proteins in mice. *Physiol Behav* **91,** 667-670.

Stopkova R, Zdrahal Z, Ryba S, Sedo O, Sandera M & Stopka P. (2010). Novel OBP genes similar to hamster Aphrodisin in the bank vole, Myodes glareolus. *BMC genomics* **11,** 45.

Stuart-Fox D & Moussalli A. (2008). Selection for Social Signalling Drives the Evolution of Chameleon Colour Change. *PLoS Biol* **6,** e25.

Sugai T, Yoshimura H, Kato N & Onoda N. (2006). Component-dependent urine responses in the rat accessory olfactory bulb. *Neuroreport* **17,** 1663-1667.

Sun L & MÜLler-Schwarze D. (1998). Beaver response to recurrent alien scents: scent fence or scent match? *Animal behaviour* **55,** 1529-1536.

Swaisgood RR, Lindburg DG & Zhou X. (1999). Giant pandas discriminate individual differences in conspecific scent. *Animal behaviour* **57,** 1045-1053.

Swaisgood RR, Lindburg DG, Zhou X & Owen MA. (2000). The effects of sex, reproductive condition and context on discrimination of conspecific odours by giant pandas. *Animal behaviour* **60,** 227-237.

Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J & Hunt DF. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 9528-9533.

TD W. (2003). Pheromones and animal behaviour: communication by smell and taste.

Thom MD, Stockley P, Jury F, Ollier WE, Beynon RJ & Hurst JL. (2008). The direct assessment of genetic heterozygosity through scent in the mouse. *Current biology : CB* **18,** 619-623.

Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Johnstone R, Mohammed AK & Hamon C. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical chemistry* **75,** 1895-1904.

Thompson RN, McMillon R, Napier A & Wekesa KS. (2007). Pregnancy block by MHC class I peptides is mediated via the production of inositol 1,4,5-trisphosphate in the mouse vomeronasal organ. *The Journal of experimental biology* **210,** 1406-1412.

Timm DE, Baker LJ, Mueller H, Zidek L & Novotny MV. (2001). Structural basis of pheromone binding to mouse major urinary protein (MUP-I). *Protein science : a publication of the Protein Society* **10,** 997-1004.

Tinbergen N & Perdeck AC. (1951). On the Stimulus Situation Releasing the Begging Response in the Newly Hatched Herring Gull Chick (Larus Argentatus Argentatus Pont.). *Behaviour* **3,** 1-39.

Towbin H, Ozbey O & Zingel O. (2001). An immunoblotting method for high-resolution isoelectric focusing of protein isoforms on immobilized pH gradients. *Electrophoresis* **22,** 1887-1893.

Trout RC. (1978). A review of studies on captive Harvest mice (Micromys minutus (Pallas)). *Mammal Review* **8,** 159-175.

Turton MJ, Robertson DH, Smith JR, Hurst JL & Beynon RJ. (2010). Roborovskin, a lipocalin in the urine of the Roborovski hamster, Phodopus roborovskii. *Chemical senses* **35,** 675-684.

Van Der Lee S & Boot LM. (1955). Spontaneous pseudopregnancy in mice. *Acta physiologica et pharmacologica Neerlandica* **4,** 442-444.

Vandenbergh JG. (1969). Male odor accelerates female sexual maturation in mice. *Endocrinology* **84,** 658-660.

Von Schantz T, Wittzell H, Göransson G & Grahn M. (1997). Mate Choice, Male Condition-Dependent Ornamentation and MHC in the Pheasant. *Hereditas* **127,** 133-140.

von Schantz T, Wittzell H, Goransson G, Grahn M & Persson K. (1996). MHC genotype and male ornamentation: genetic evidence for the Hamilton-Zuk model. *Proceedings Biological sciences / The Royal Society* **263,** 265-271.

Walmer DK, Wrona MA, Hughes CL & Nelson KG. (1992). Lactoferrin expression in the mouse reproductive tract during the natural estrous cycle: correlation with circulating estradiol and progesterone. *Endocrinology* **131,** 1458-1466.

Weidt A, Hagenah N, Randrianambinina B, Radespiel U & Zimmermann E. (2004). Social organization of the golden brown mouse lemur (Microcebus ravelobensis). *American journal of physical anthropology* **123,** 40-51.

Whitten WK. (1958). Modification of the oestrous cycle of the mouse by external stimuli associated with the male; changes in the oestrous cycle determined by vaginal smears. *The Journal of endocrinology* **17,** 307-313.

Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, Yan JX, Gooley AA, Hughes G, Humphery-Smith I, Williams KL & Hochstrasser DF. (1996). From proteins to

proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Bio/technology (Nature Publishing Company)* **14,** 61-65.

Williams SE, Brown TI, Roghanian A & Sallenave JM. (2006). SLPI and elafin: one glove, many fingers. *Clinical science (London, England : 1979)* **110,** 21-35.

Wisniewski JR, Zougman A, Nagaraj N & Mann M. (2009). Universal sample preparation method for proteome analysis. *Nature methods* **6,** 359-362.

Wittig I & Schagger H. (2008). Features and applications of blue-native and clear-native electrophoresis. *Proteomics* **8,** 3974-3990.

Wong JW & Cagney G. (2010). An overview of label-free quantitation methods in proteomics by mass spectrometry. *Methods in molecular biology (Clifton, NJ)* **604,** 273-283.

Wu C, Robertson DH, Hubbard SJ, Gaskell SJ & Beynon RJ. (1999). Proteolysis of native proteins. Trapping of a reaction intermediate. *The Journal of biological chemistry* **274,** 1108-1115.

Wyatt TD. (2009). Pheromones and Animal Behaviour: Communication by Smell and Taste. *Pragmatics and Cognition* **17,** 482-490.

Yenugu S, Richardson RT, Sivashanmugam P, Wang Z, O'Rand M G, French FS & Hall SH. (2004). Antimicrobial activity of human EPPIN, an androgen-regulated, sperm-bound protein with a whey acidic protein motif. *Biology of reproduction* **71,** 1484-1490.

Young JM, Kambere M, Trask BJ & Lane RP. (2005). Divergent V1R repertoires in five species: Amplification in rodents, decimation in primates, and a surprisingly small repertoire in dogs. *Genome research* **15,** 231-240.

Yu YQ, Gilar M & Gebler JC. (2004). A complete peptide mapping of membrane proteins: a novel surfactant aiding the enzymatic digestion of bacteriorhodopsin. *Rapid communications in mass spectrometry : RCM* **18,** 711-715.

Yu YQ, Gilar M, Lee PJ, Bouvier ES & Gebler JC. (2003). Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Analytical chemistry* **75,** 6023-6028.

Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, Zhang W, Zhang Z, Lajoie GA & Ma B. (2012). PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & cellular proteomics : MCP* **11,** M111.010587.

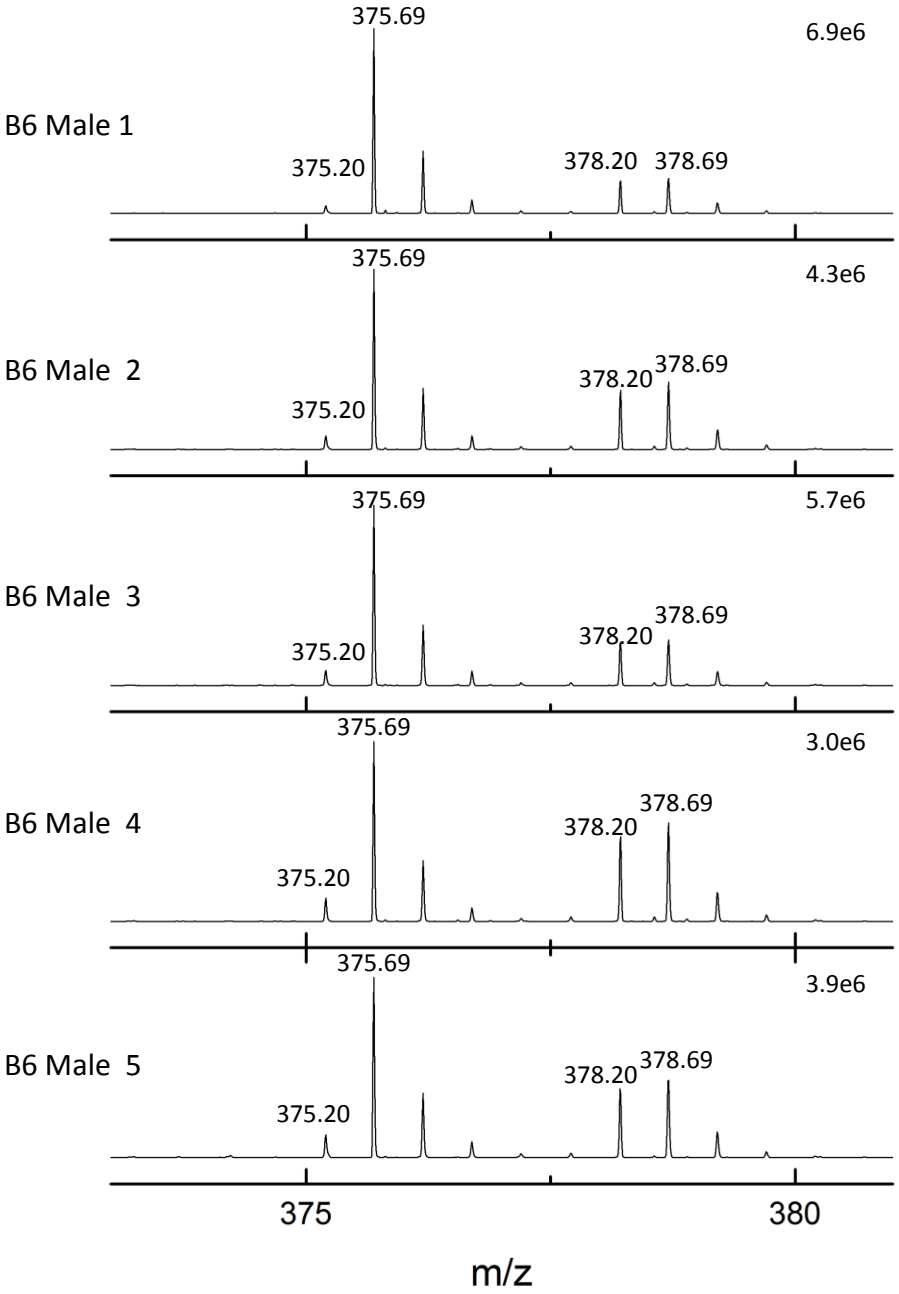
Zhang JX, Sun L, Zhang JH & Feng ZY. (2008). Sex- and gonad-affecting scent compounds and 3 male pheromones in the rat. *Chemical senses* **33,** 611-621.

Zhang X & Firestein S. (2002). The olfactory receptor gene superfamily of the mouse. *Nature neuroscience* **5,** 124-133.

Zhang X, Rogers M, Tian H, Zhang X, Zou DJ, Liu J, Ma M, Shepherd GM & Firestein SJ. (2004). High-throughput microarray detection of olfactory receptor gene expression in the mouse. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 14168-14173.

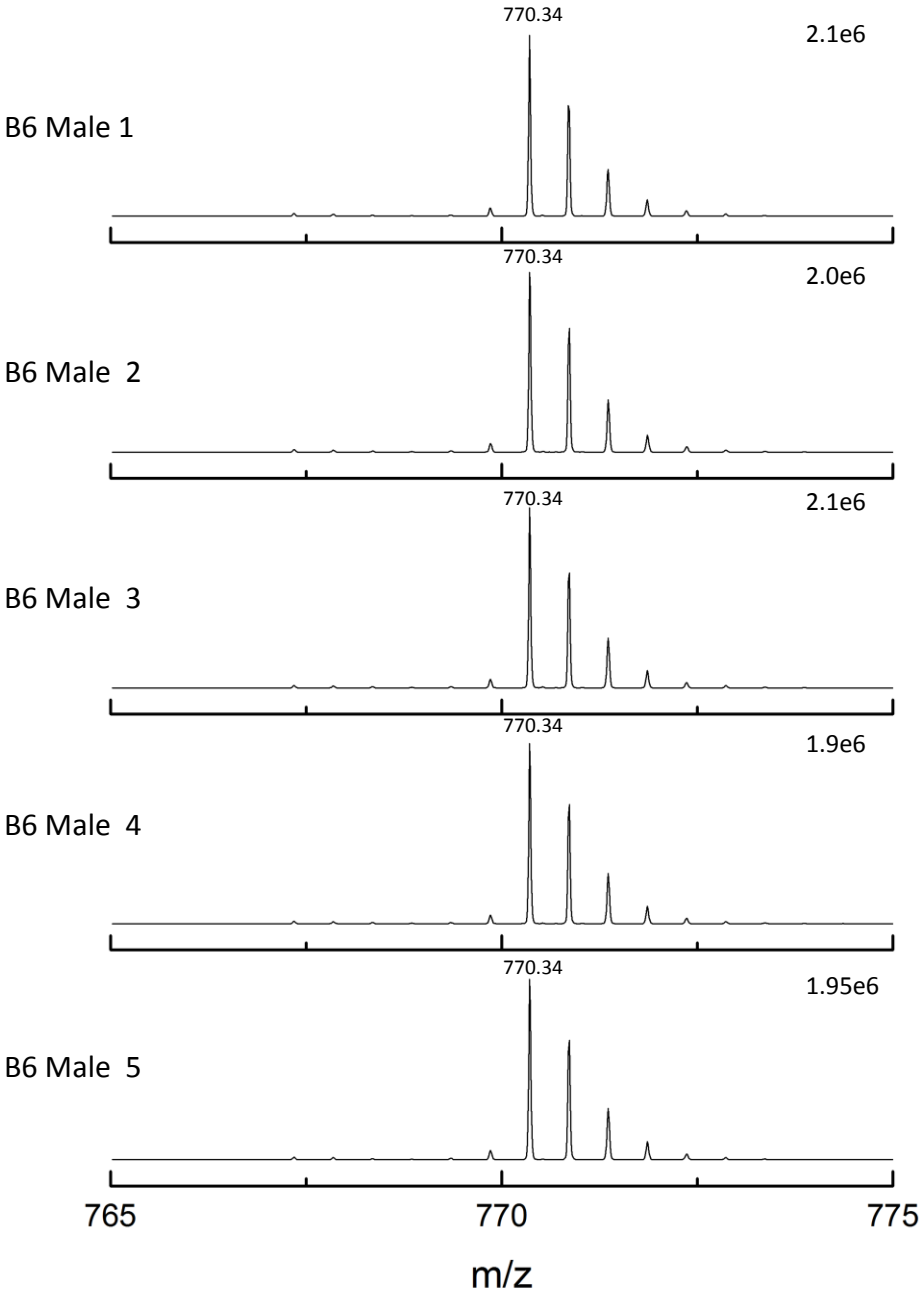
Zidek L, Stone MJ, Lato SM, Pagel MD, Miao Z, Ellington AD & Novotny MV. (1999). NMR mapping of the recombinant mouse major urinary protein I binding site occupied by the pheromone 2-sec-butyl-4,5-dihydrothiazole. *Biochemistry* **38,** 9850-9861.

Zufall F & Leinders-Zufall T. (2007). Mammalian pheromone sensing. *Current opinion in neurobiology* **17,** 483-489.

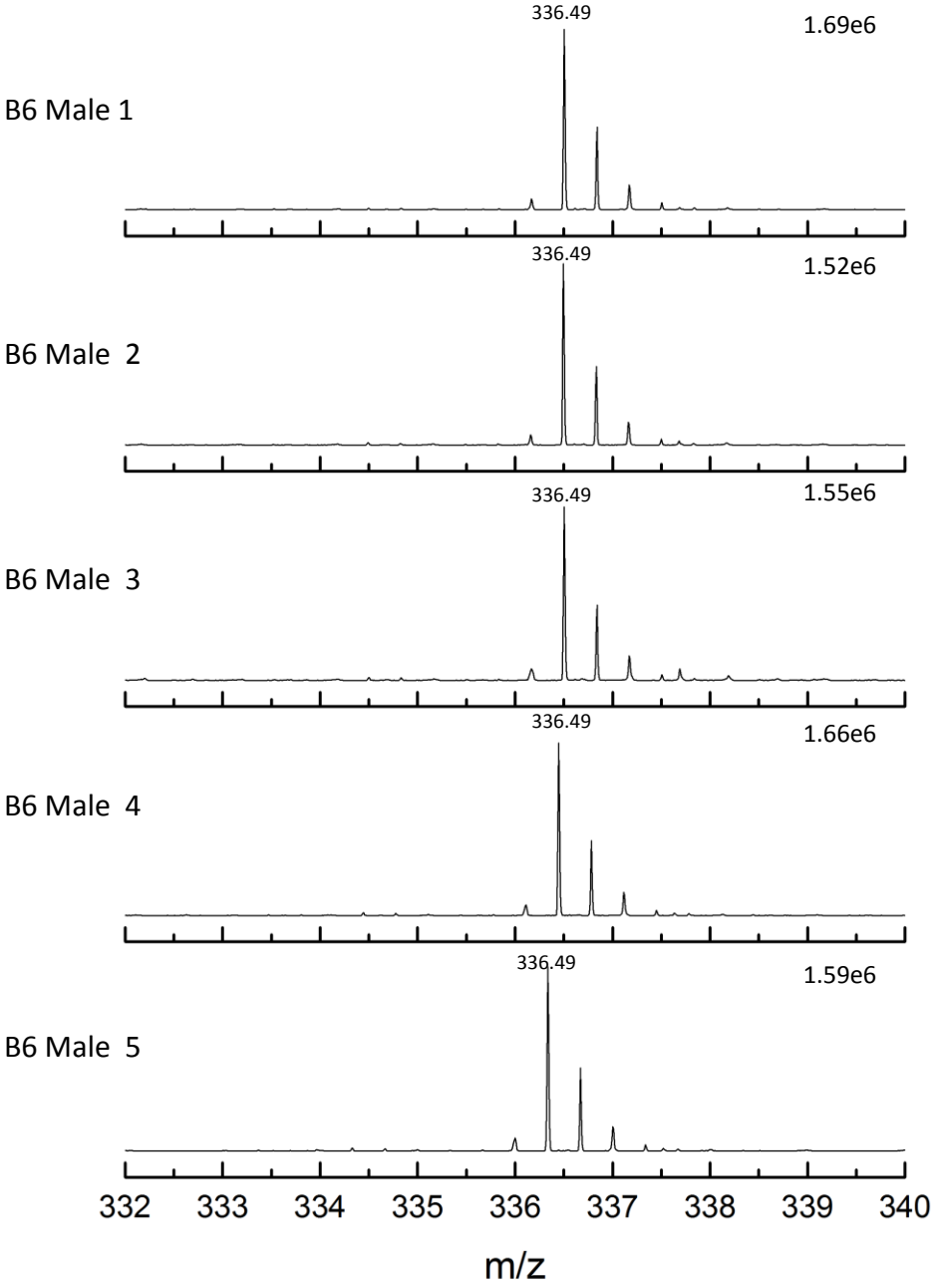


Supplementary data A. LC-MS analysis of individual male C57BL/6 mice. Q peptides 1 and 3.

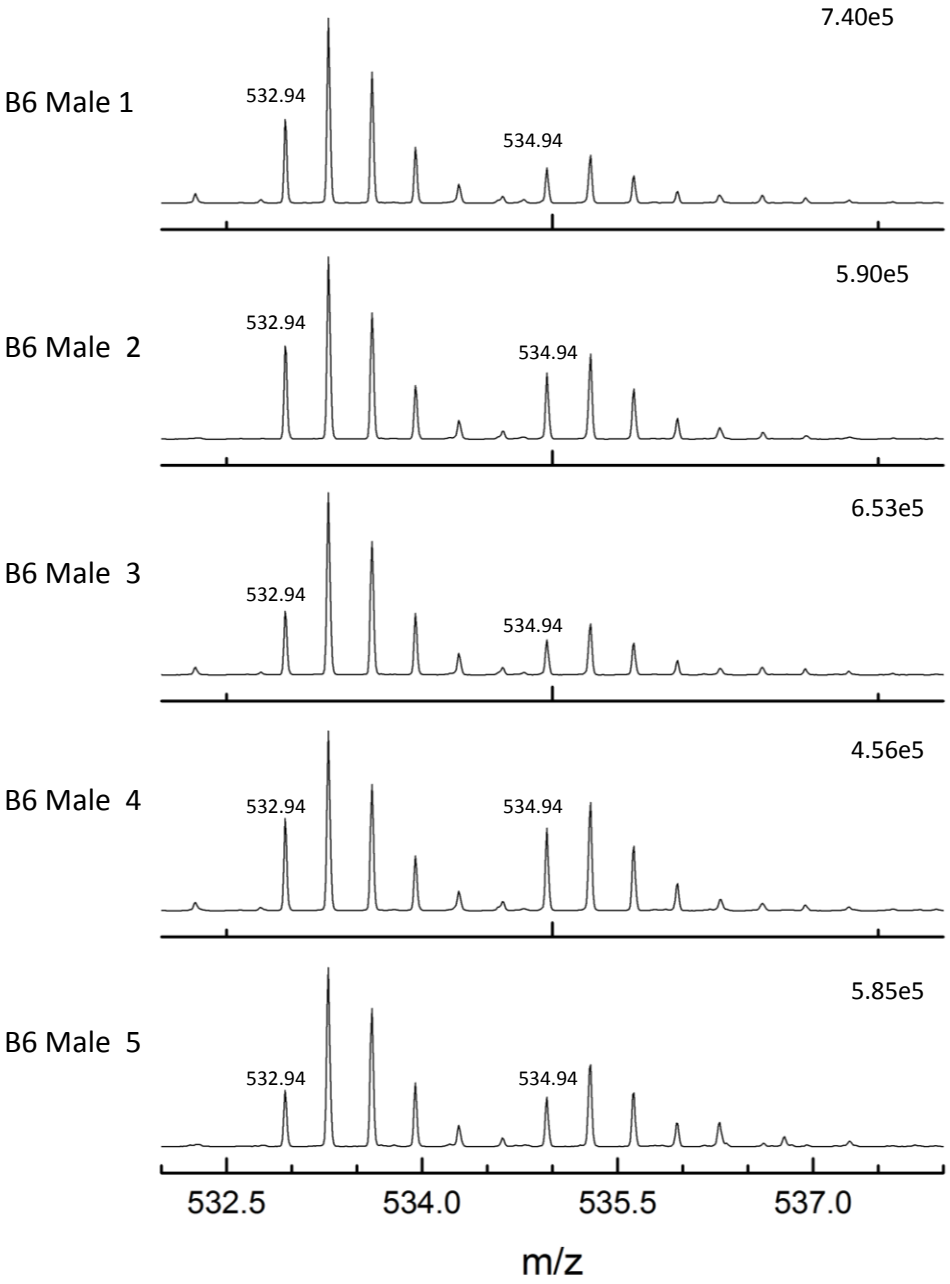
A known amount of QconCAT was added to five individual male urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier.



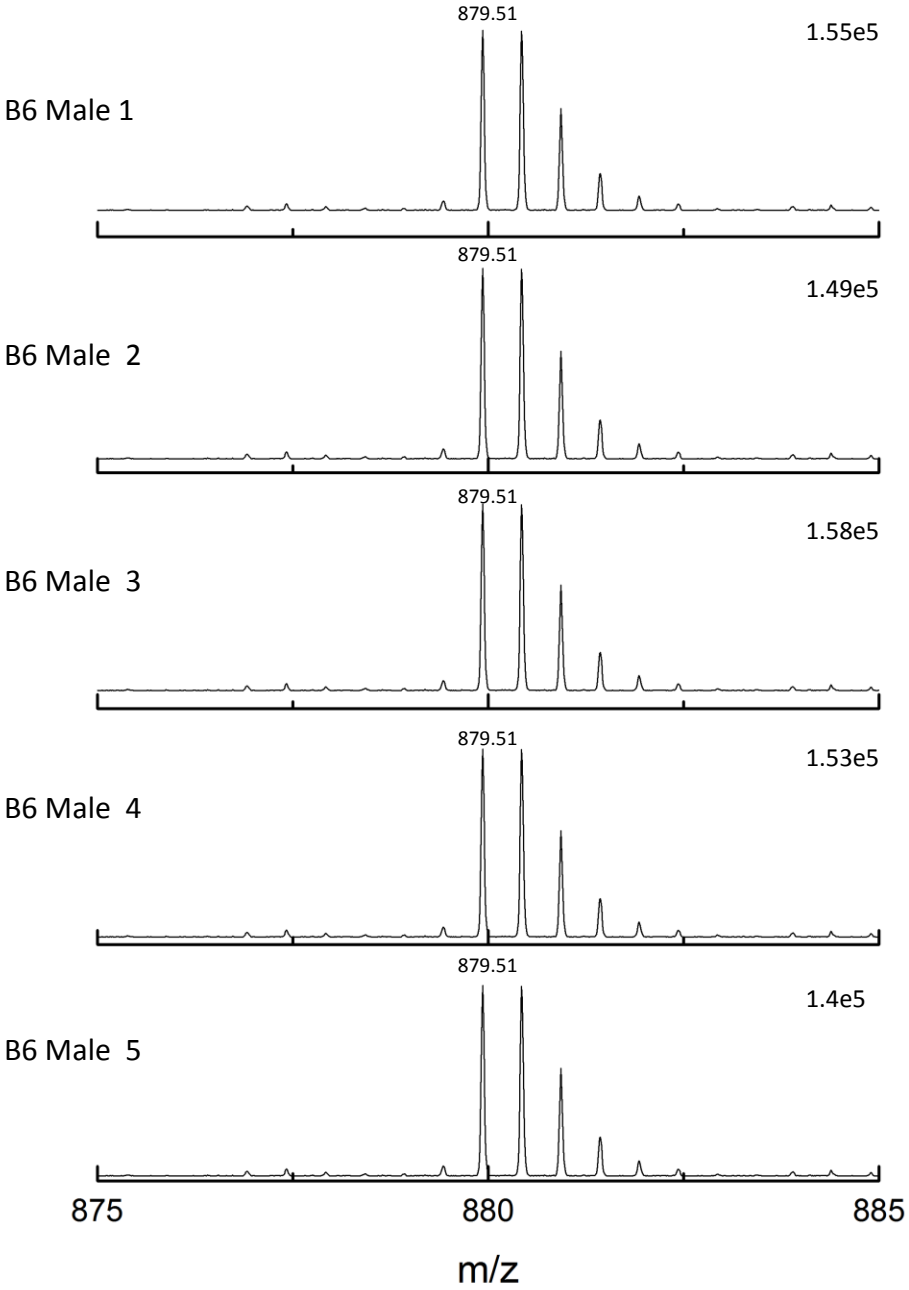
Supplementary data A. LC-MS analysis of individual male C57BL/6 mice. Q peptide 2. A known amount of QconCAT was added to five individual male urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier. No "light" analyte was detected in these five individual B6 male mice.



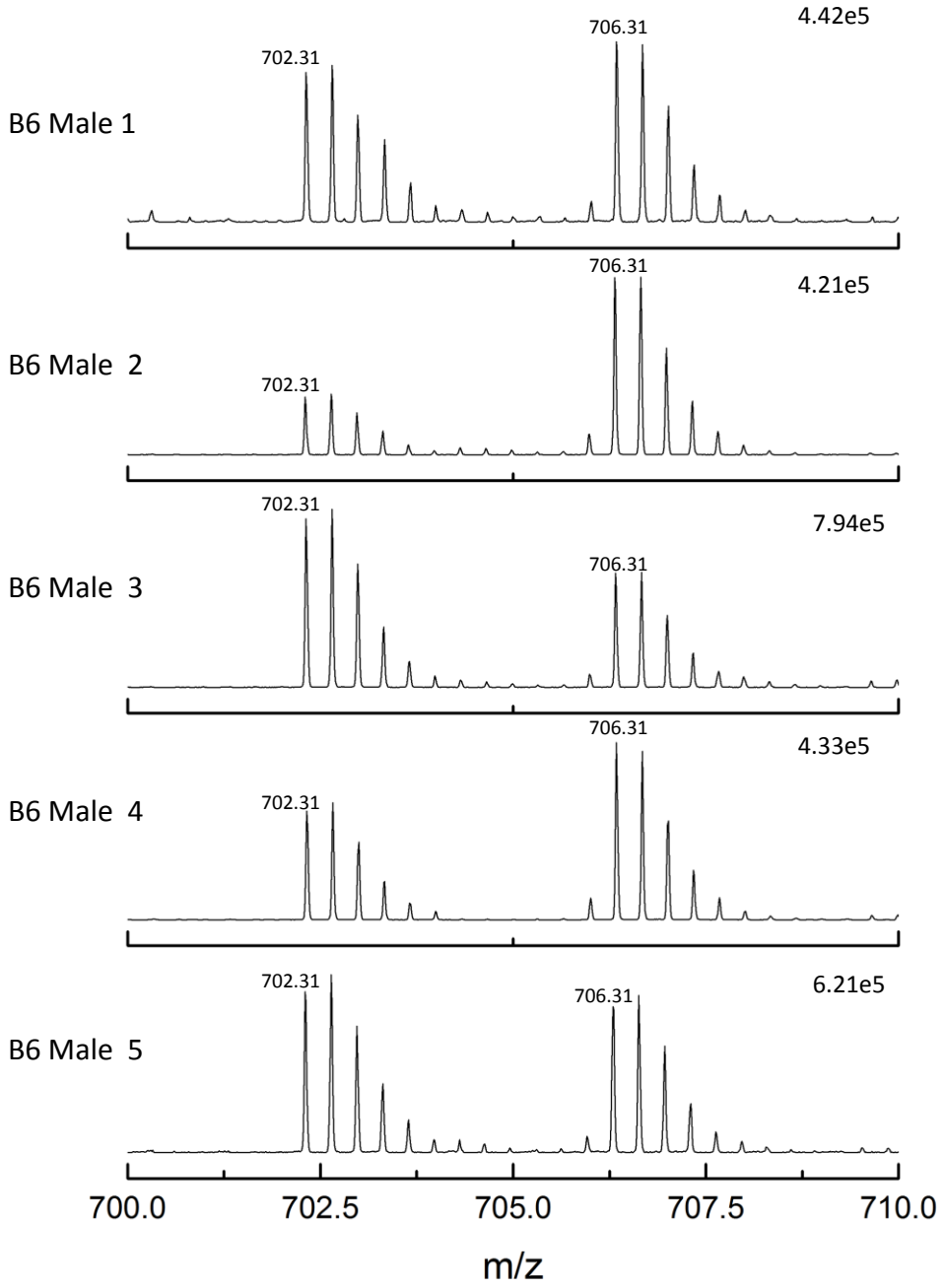
Supplementary data A. LC-MS analysis of individual male C57BL/6 mice. Q peptide 4. A known amount of QconCAT was added to five individual male urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier. No "light" analyte was detected in these five individual B6 male mice.



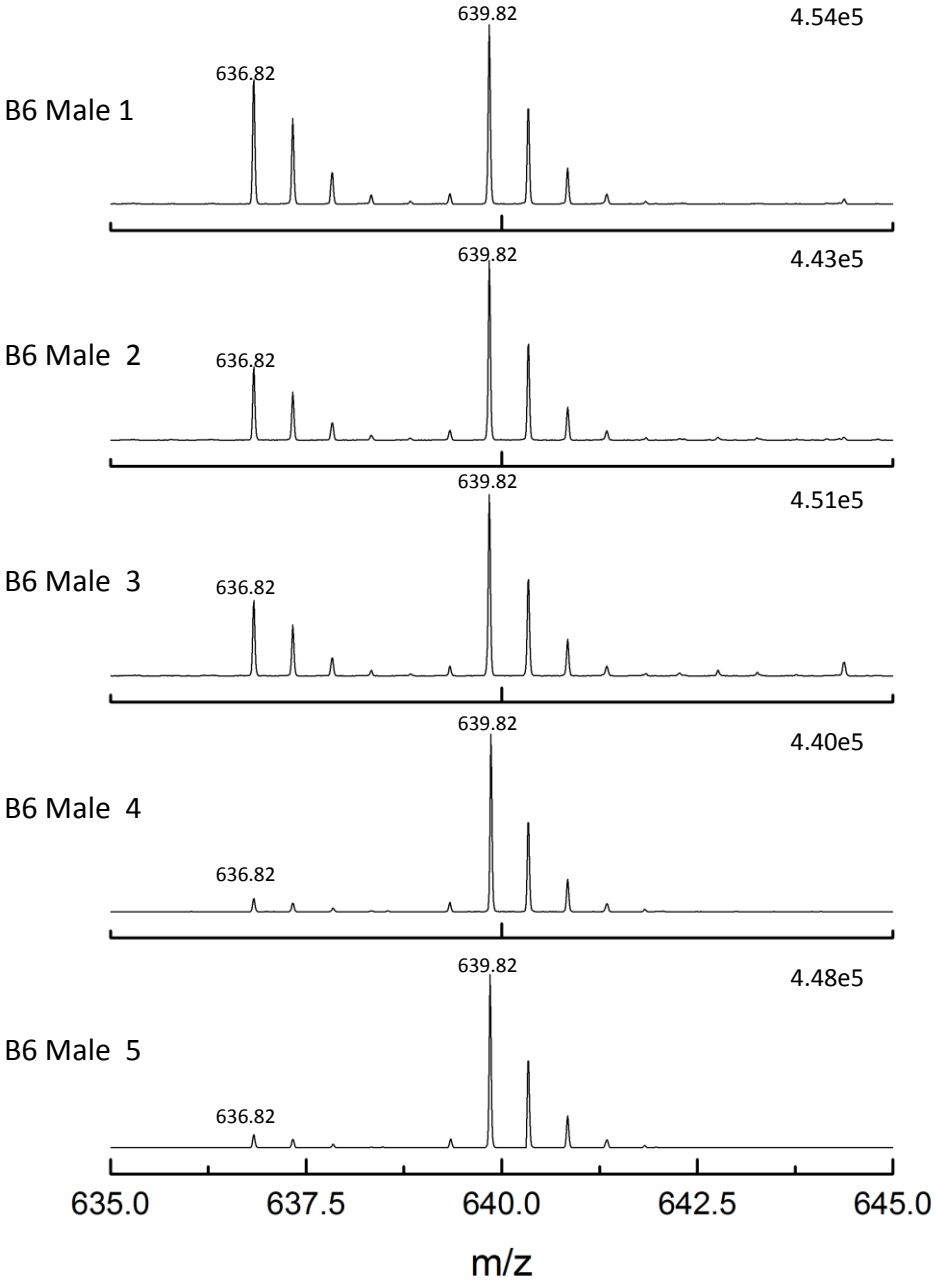
Supplementary data A. LC-MS analysis of individual male C57BL/6 mice. Q peptide 5. A known amount of QconCAT was added to five individual male urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier.



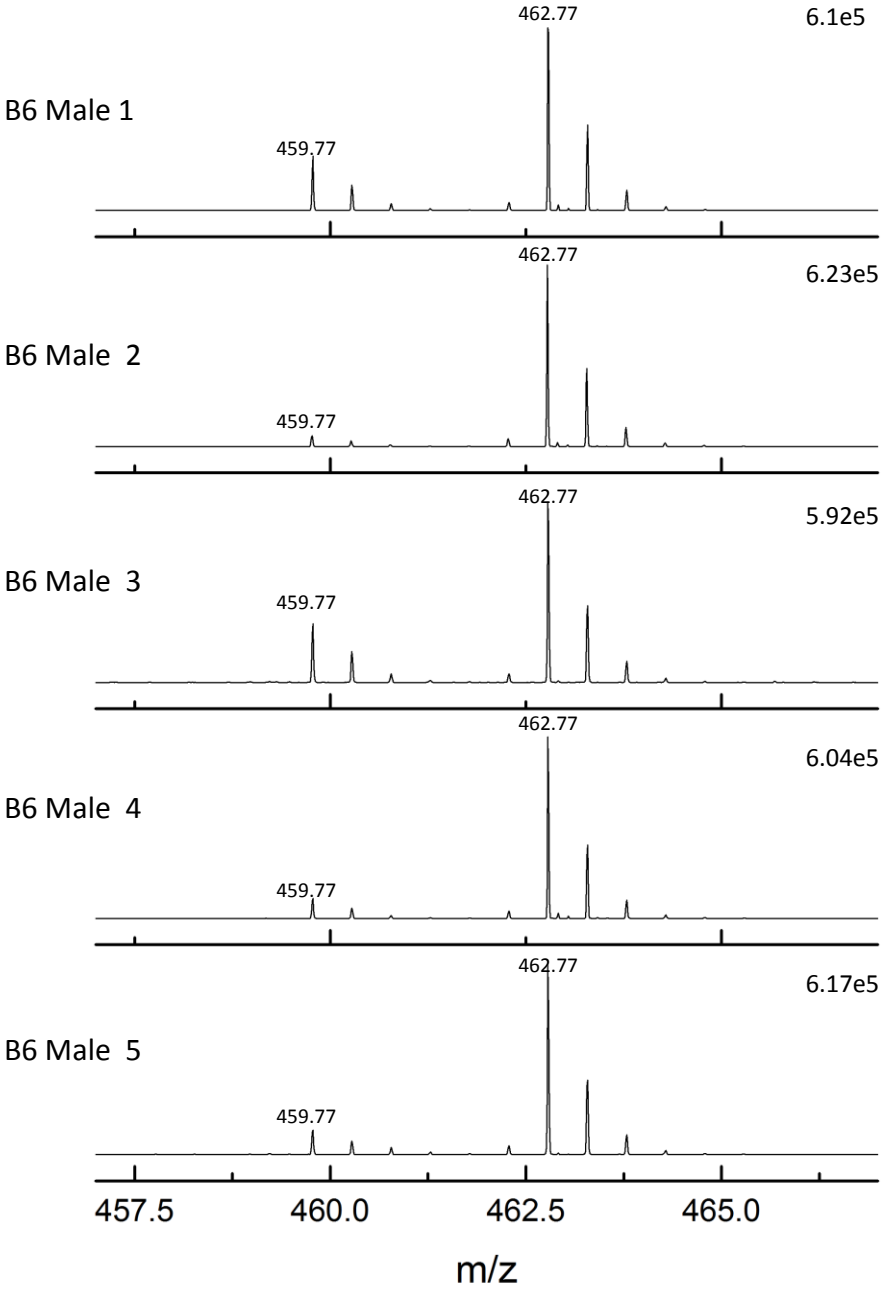
Supplementary data A. LC-MS analysis of individual male C57BL/6 mice. Q peptide 6. A known amount of QconCAT was added to five individual male urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier. No "light" analyte was detected in these five individual B6 male mice.



Supplementary data A. LC-MS analysis of individual male C57BL/6 mice. Q peptide 7. A known amount of QconCAT was added to five individual male urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 12 Da heavier. This set of peptides are from a LysC digest so the heavy peptide is 12 Da heavier than the light due to the internal labelled arginine residue.

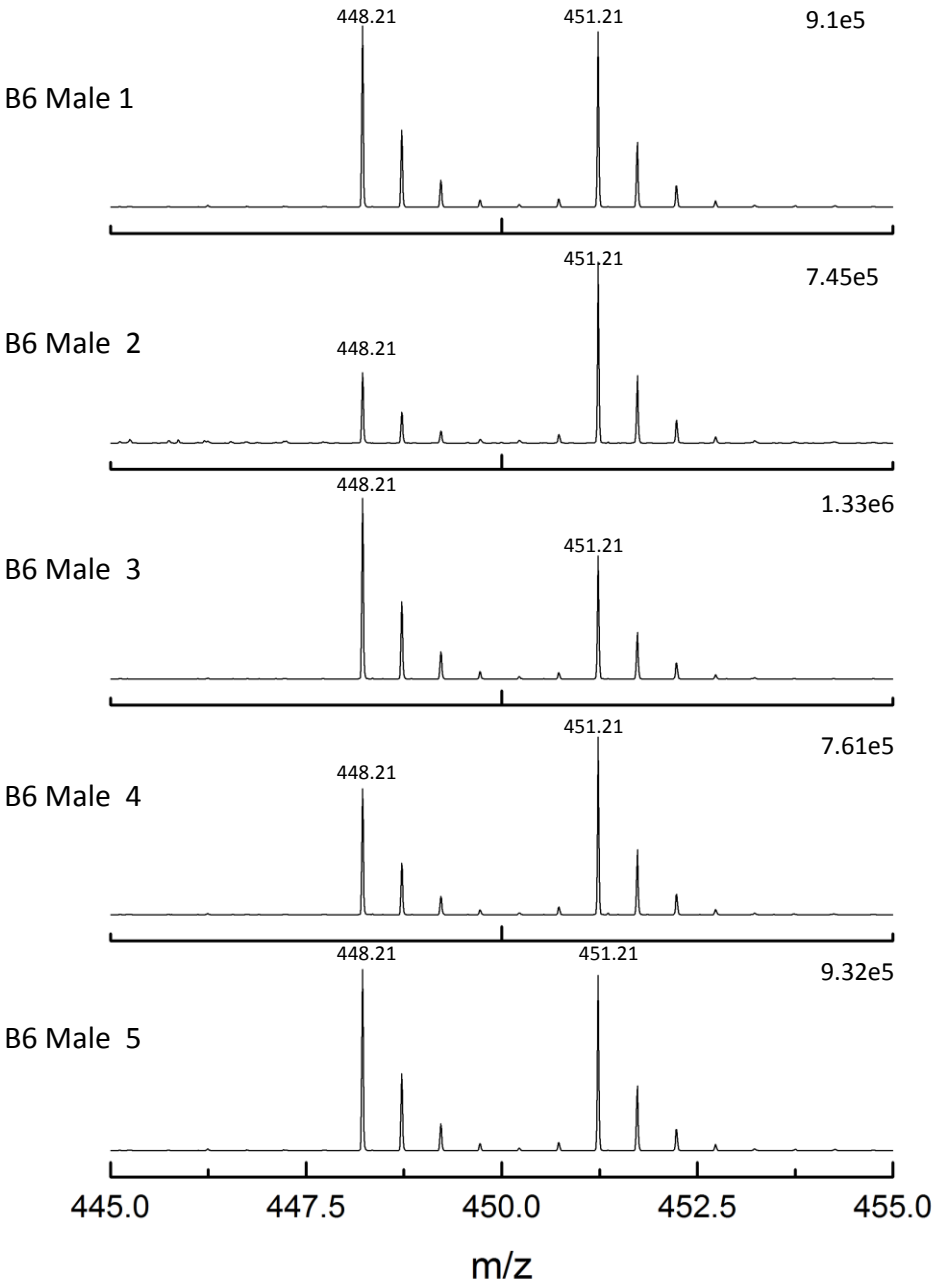


Supplementary data A. LC-MS analysis of individual male C57BL/6 mice. Q peptide 8. A known amount of QconCAT was added to five individual male urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier.



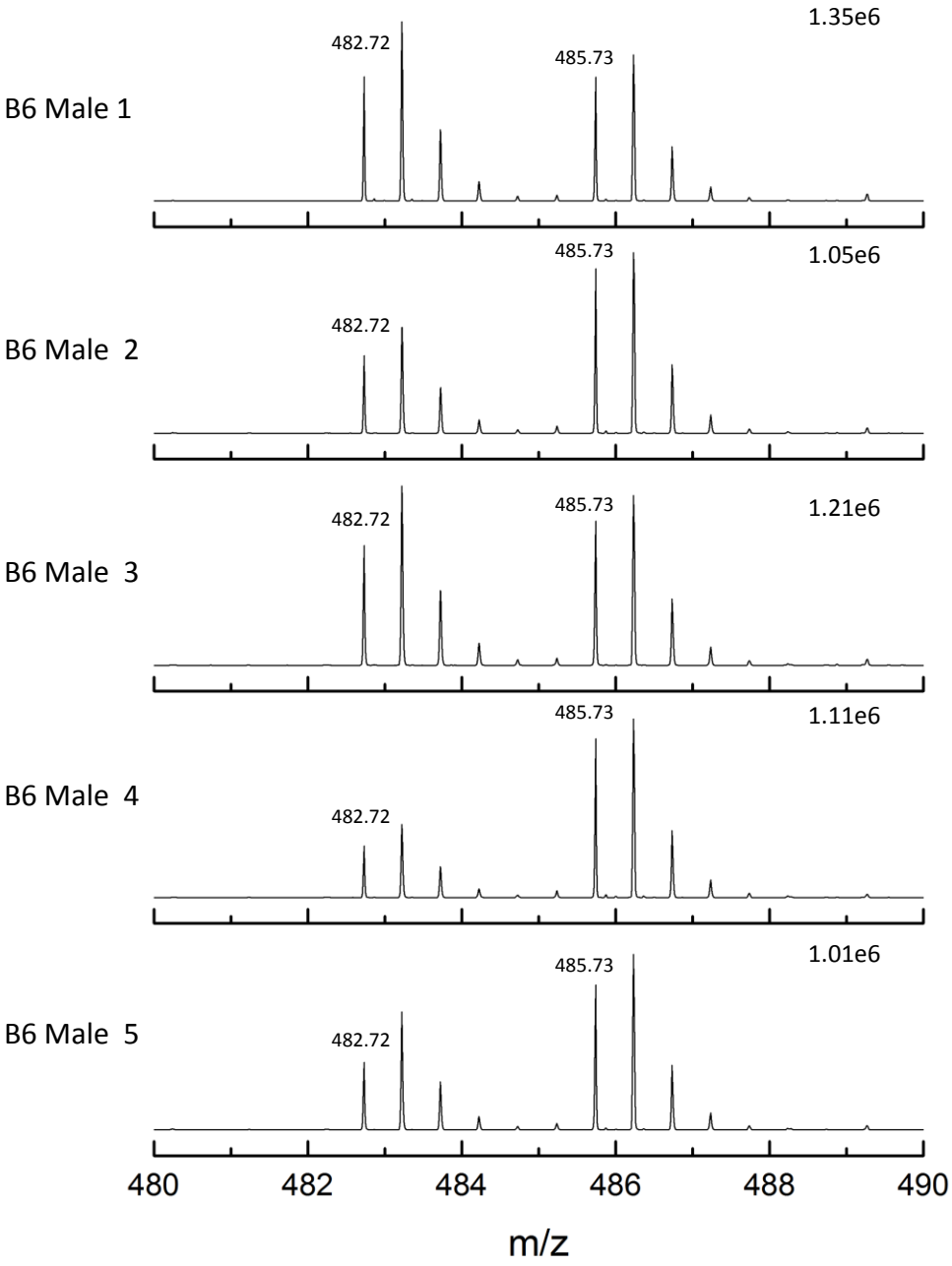
Supplementary data A. LC-MS analysis of individual male C57BL/6 mice. Q peptide 11.

A known amount of QconCAT was added to five individual male urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier.



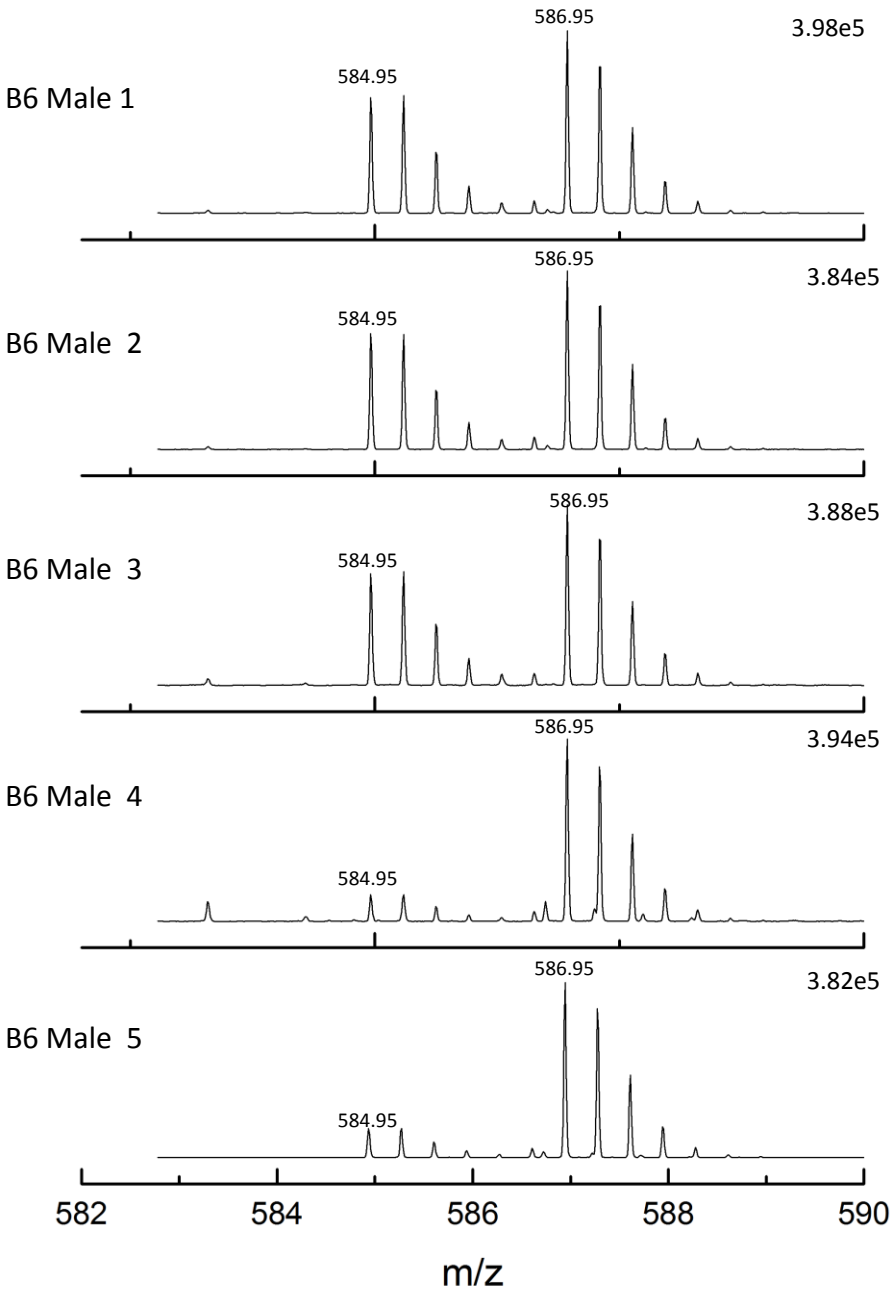
Supplementary data A. LC-MS analysis of individual male C57BL/6 mice. Q peptide 12.

A known amount of QconCAT was added to five individual male urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier.



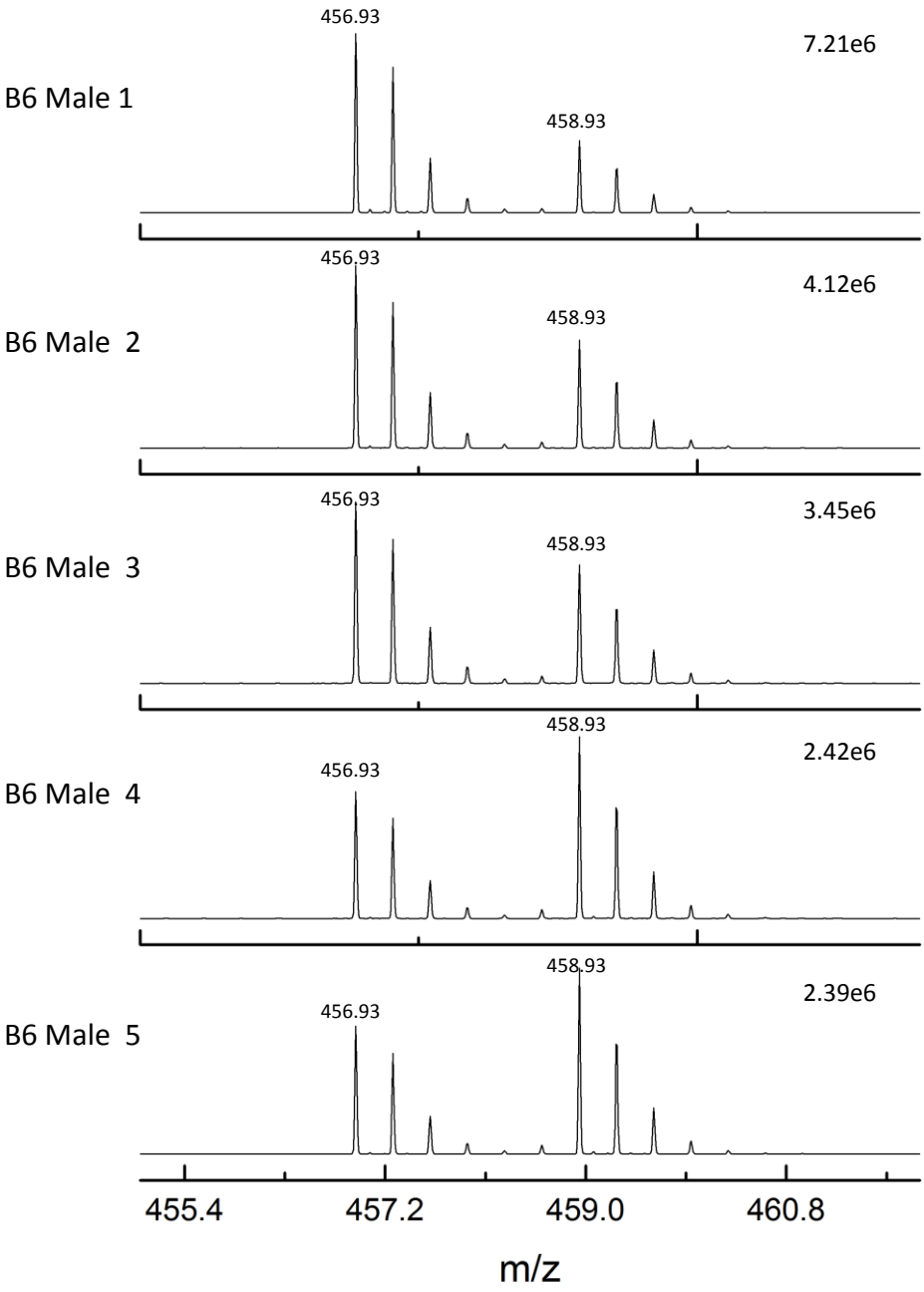
Supplementary data A. LC-MS analysis of individual male C57BL/6 mice. Q peptide 13.

A known amount of QconCAT was added to five individual male urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier.



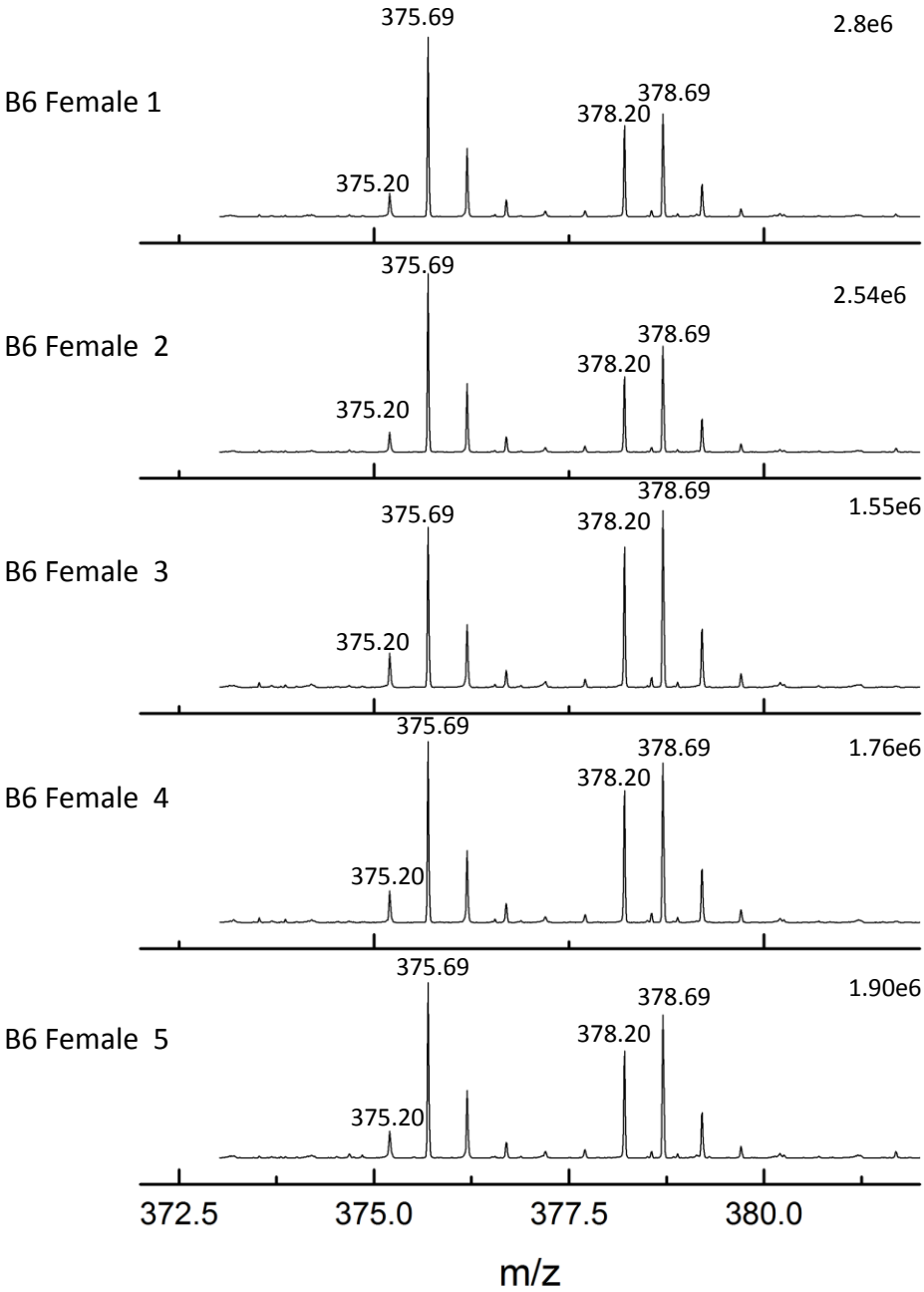
Supplementary data A. LC-MS analysis of individual male C57BL/6 mice. Q peptide 14.

A known amount of QconCAT was added to five individual male urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier.



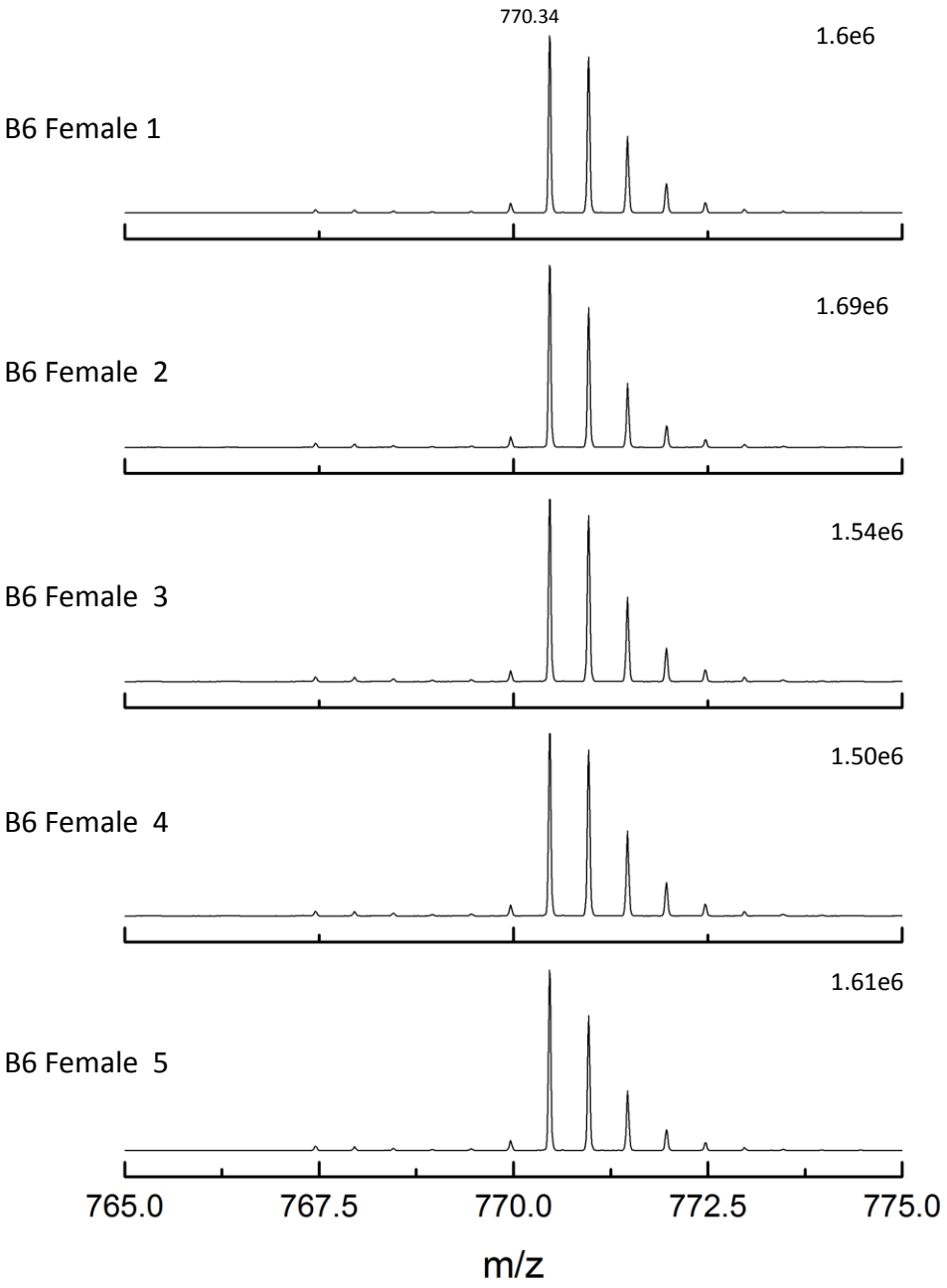
Supplementary data A. LC-MS analysis of individual male C57BL/6 mice. Q peptide 15.

A known amount of QconCAT was added to five individual male urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier.

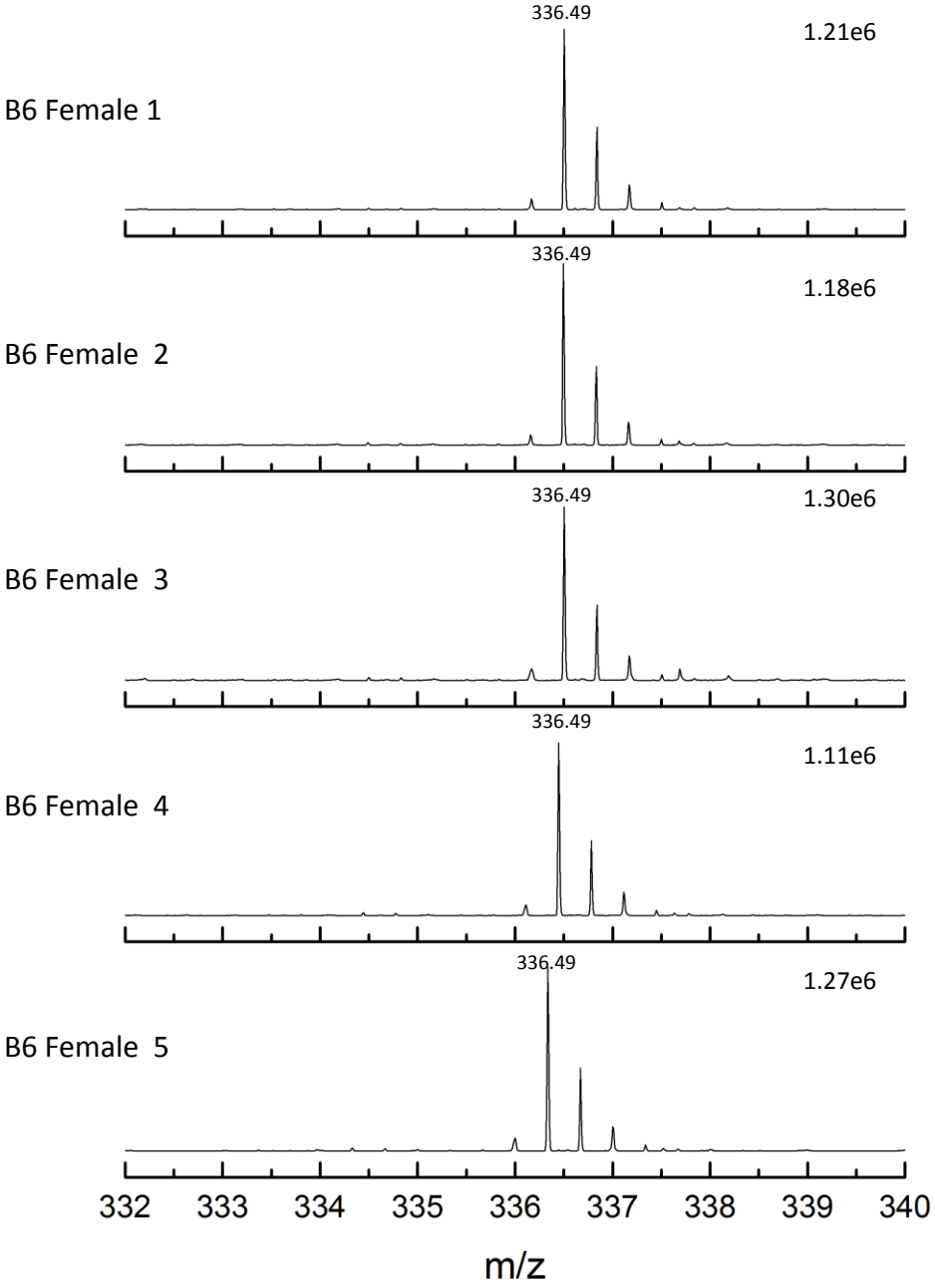


Supplementary data A. LC-MS analysis of individual female C57BL/6 mice. Q peptides 1 and 3.

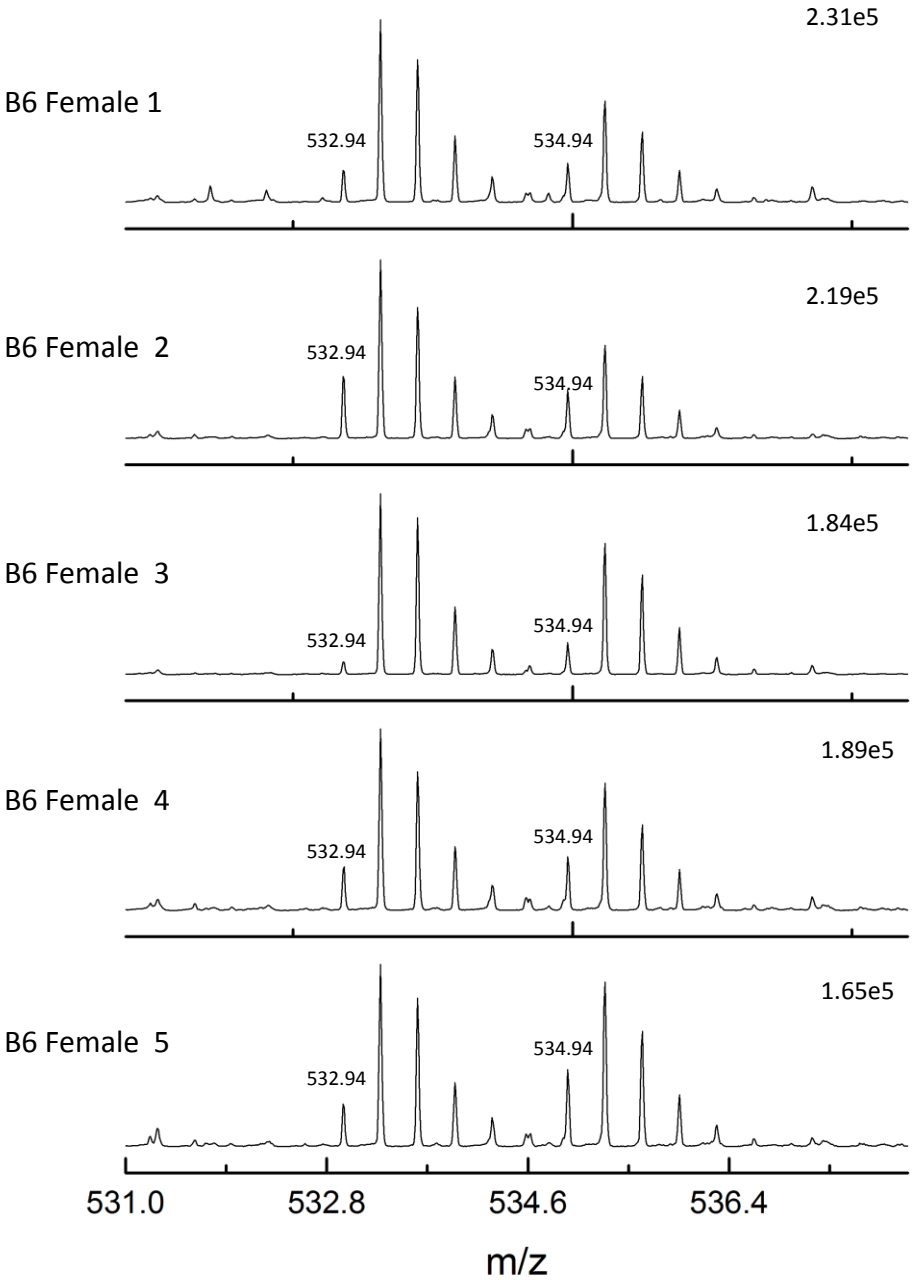
A known amount of QconCAT was added to five individual female urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier.



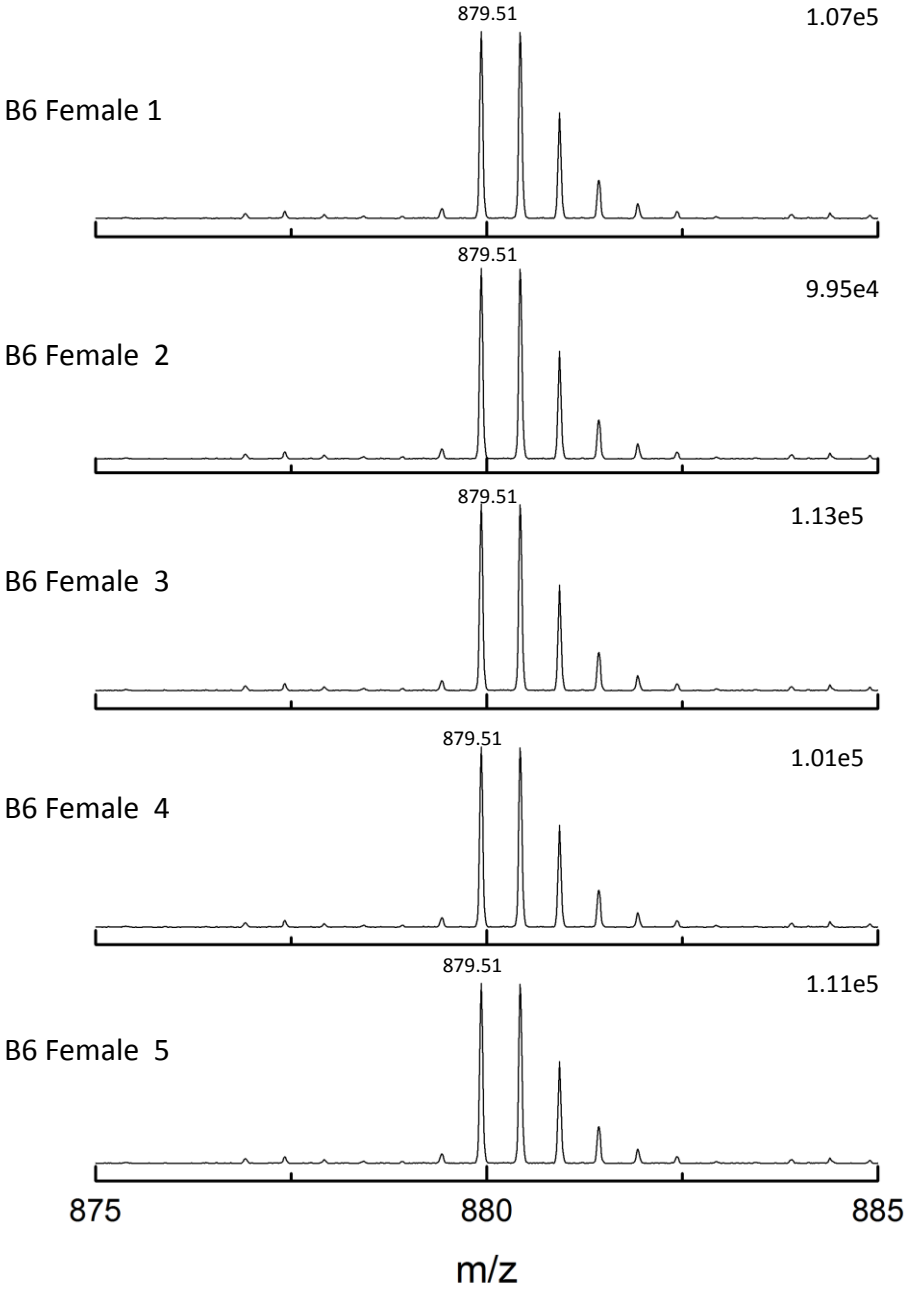
Supplementary data A. LC-MS analysis of individual female C57BL/6 mice. Q peptide 2.
A known amount of QconCAT was added to five individual female urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier. No "light" analyte was detected in these five individual B6 female mice.



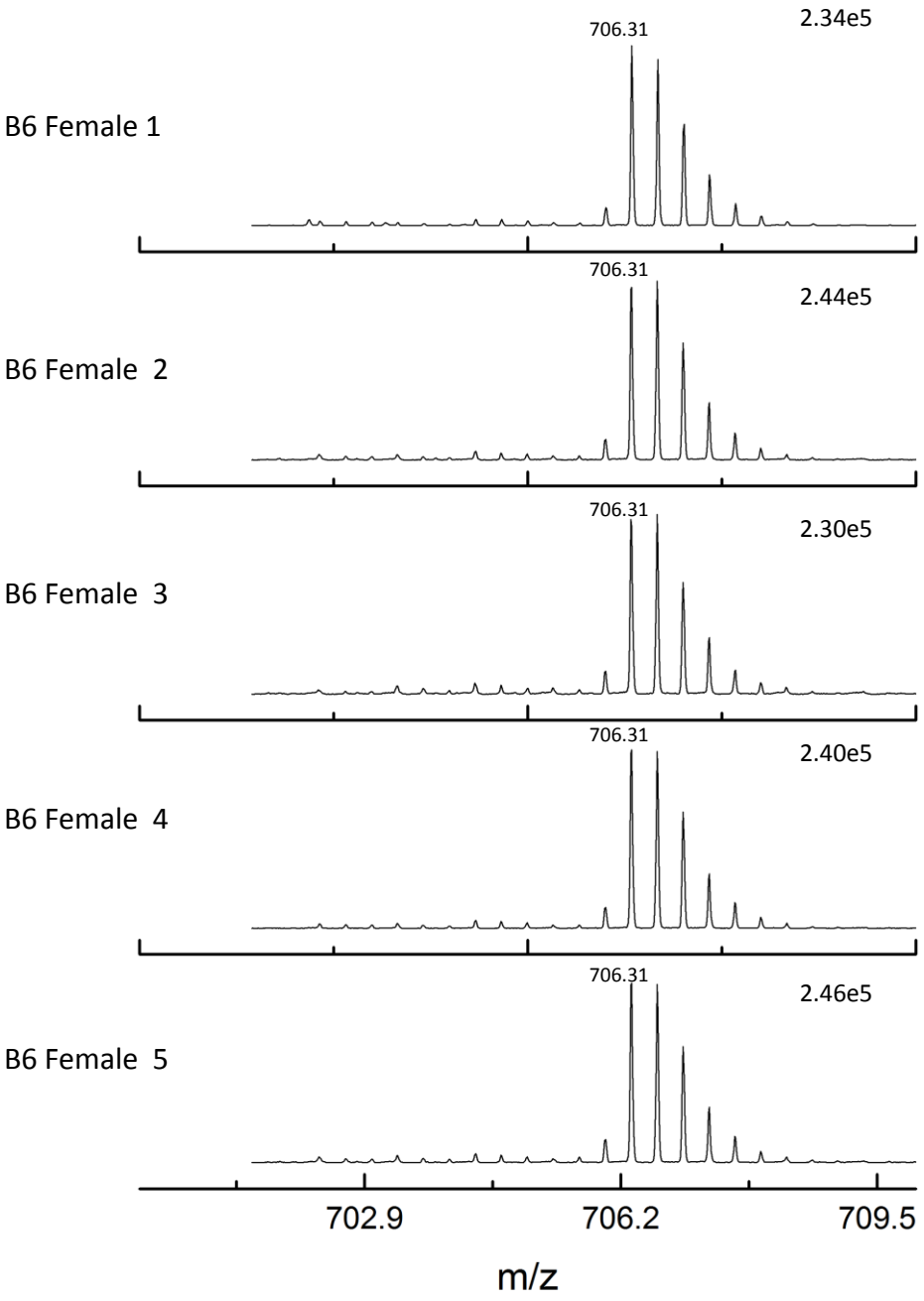
Supplementary data A. LC-MS analysis of individual female C57BL/6 mice. Q peptide 4.
A known amount of QconCAT was added to five individual female urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier. No "light" analyte was detected in these five individual B6 female mice.



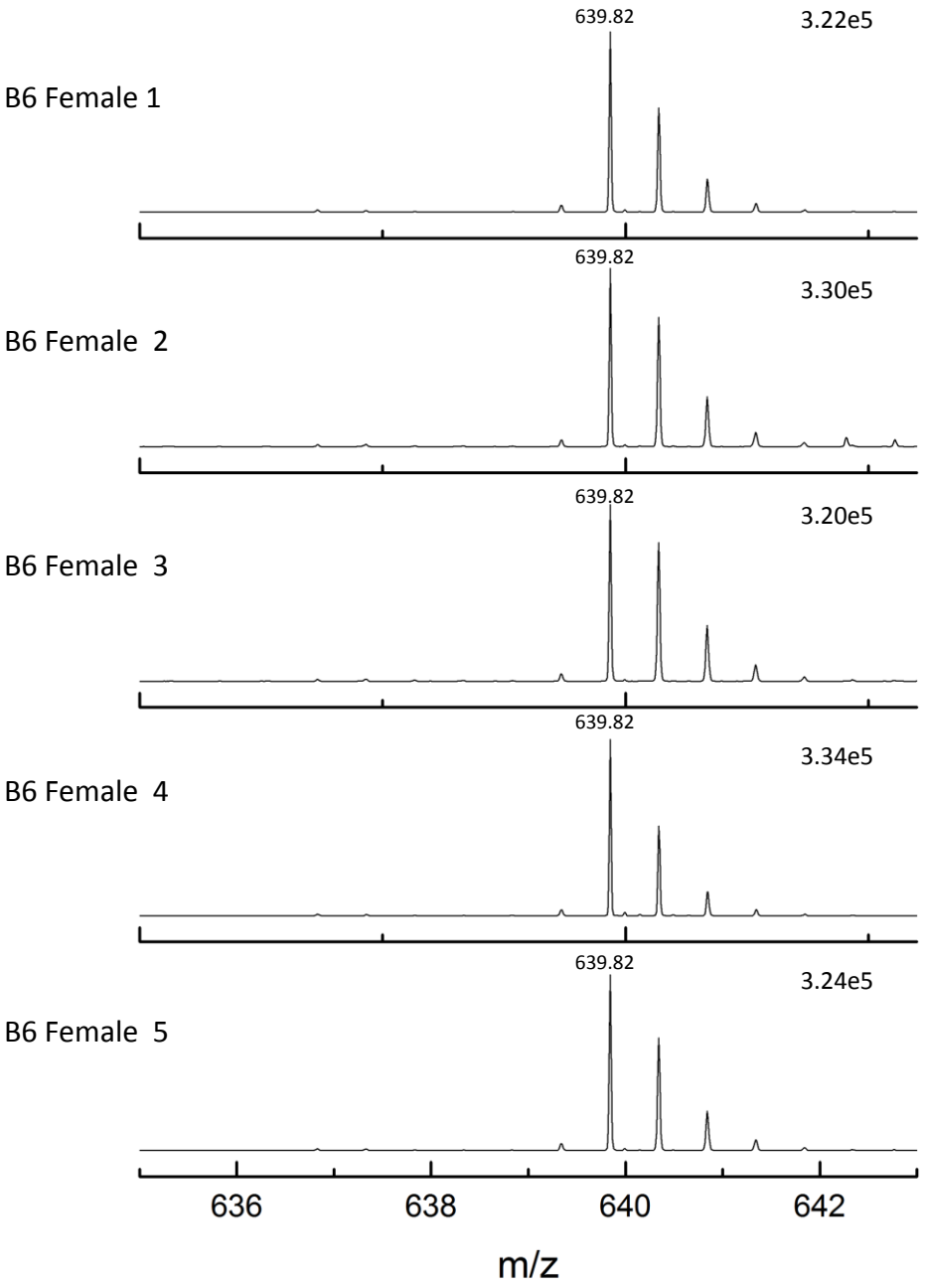
Supplementary data A. LC-MS analysis of individual female C57BL/6 mice. Q peptide 5.
A known amount of QconCAT was added to five individual female urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier.



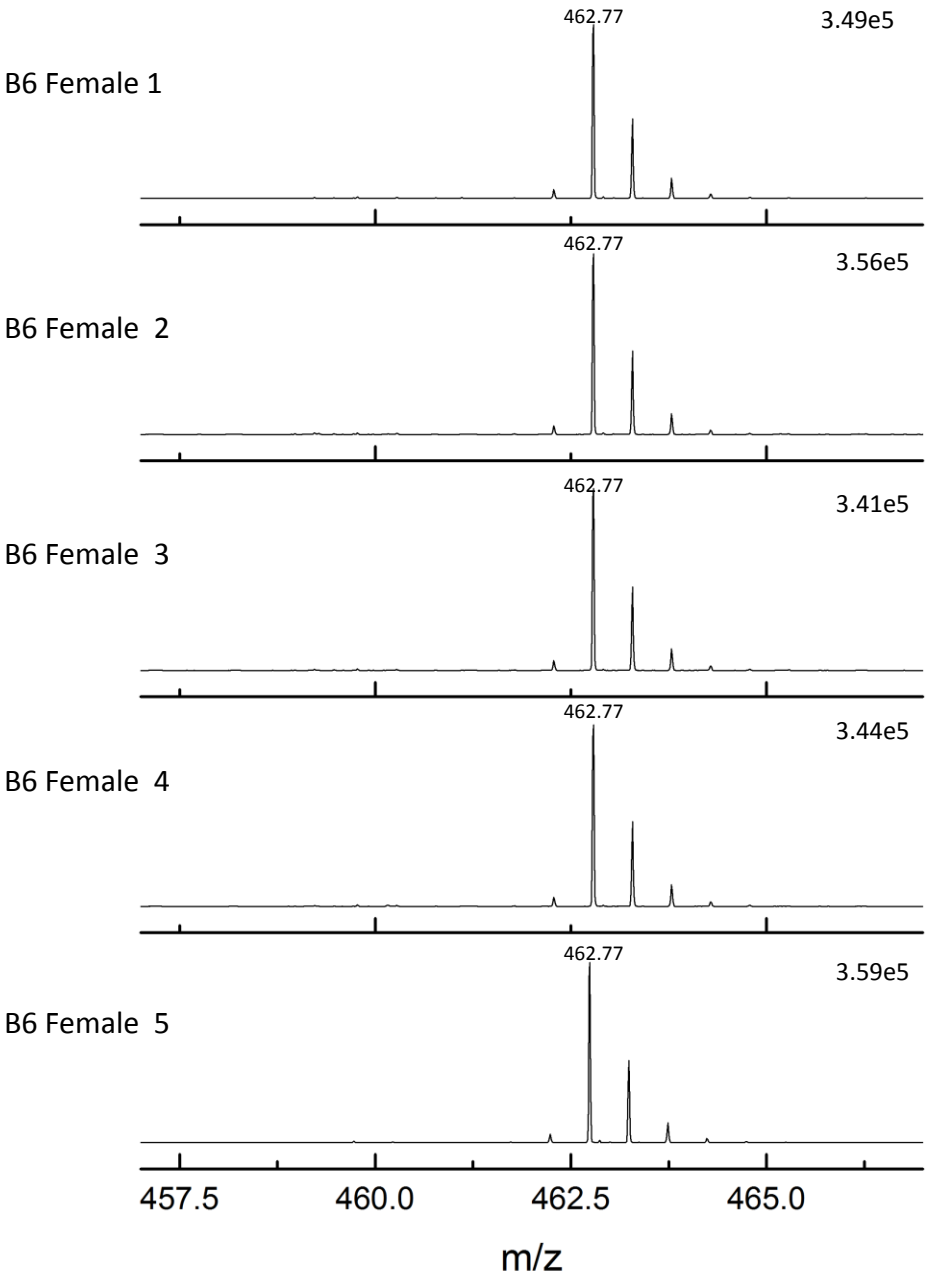
Supplementary data A. LC-MS analysis of individual female C57BL/6 mice. Q peptide 6.
A known amount of QconCAT was added to five individual female urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier.



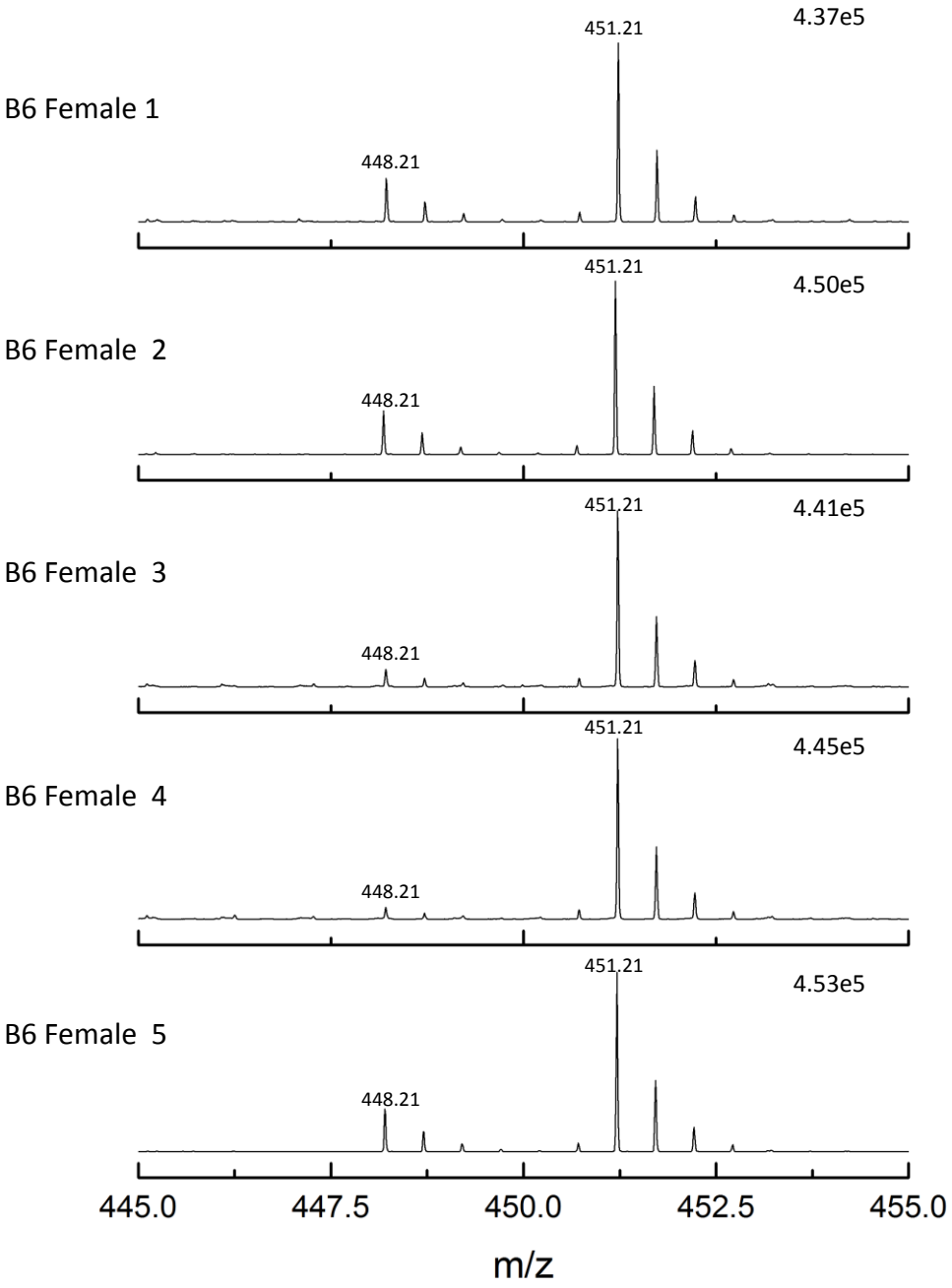
Supplementary data A. LC-MS analysis of individual female C57BL/6 mice. Q peptide 7.
A known amount of QconCAT was added to five individual female urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier. No "light" analyte was detected in these five individual B6 female mice.



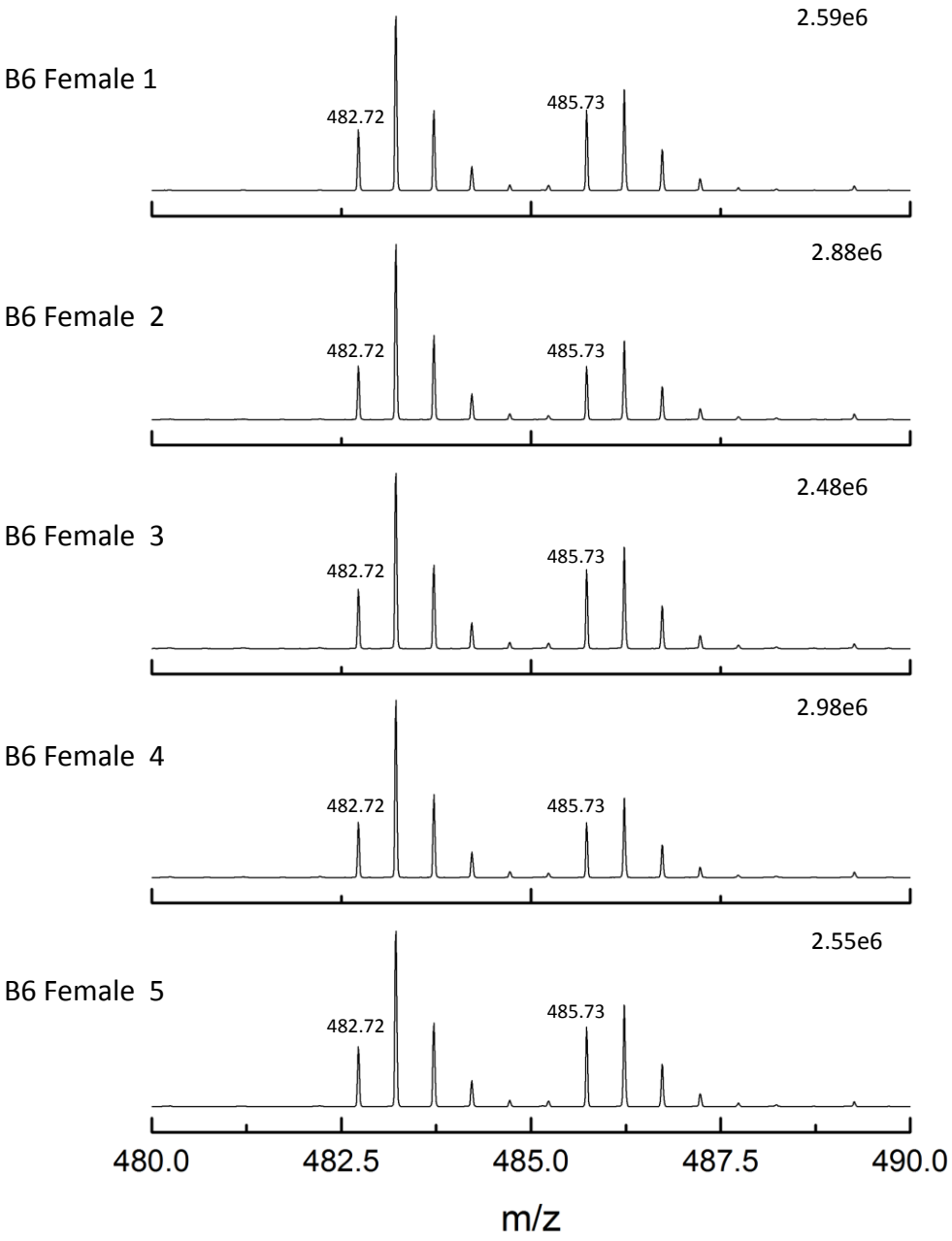
Supplementary data A. LC-MS analysis of individual female C57BL/6 mice. Q peptide 8. A known amount of QconCAT was added to five individual female urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier. No "light" analyte was detected in these five individual B6 female mice.



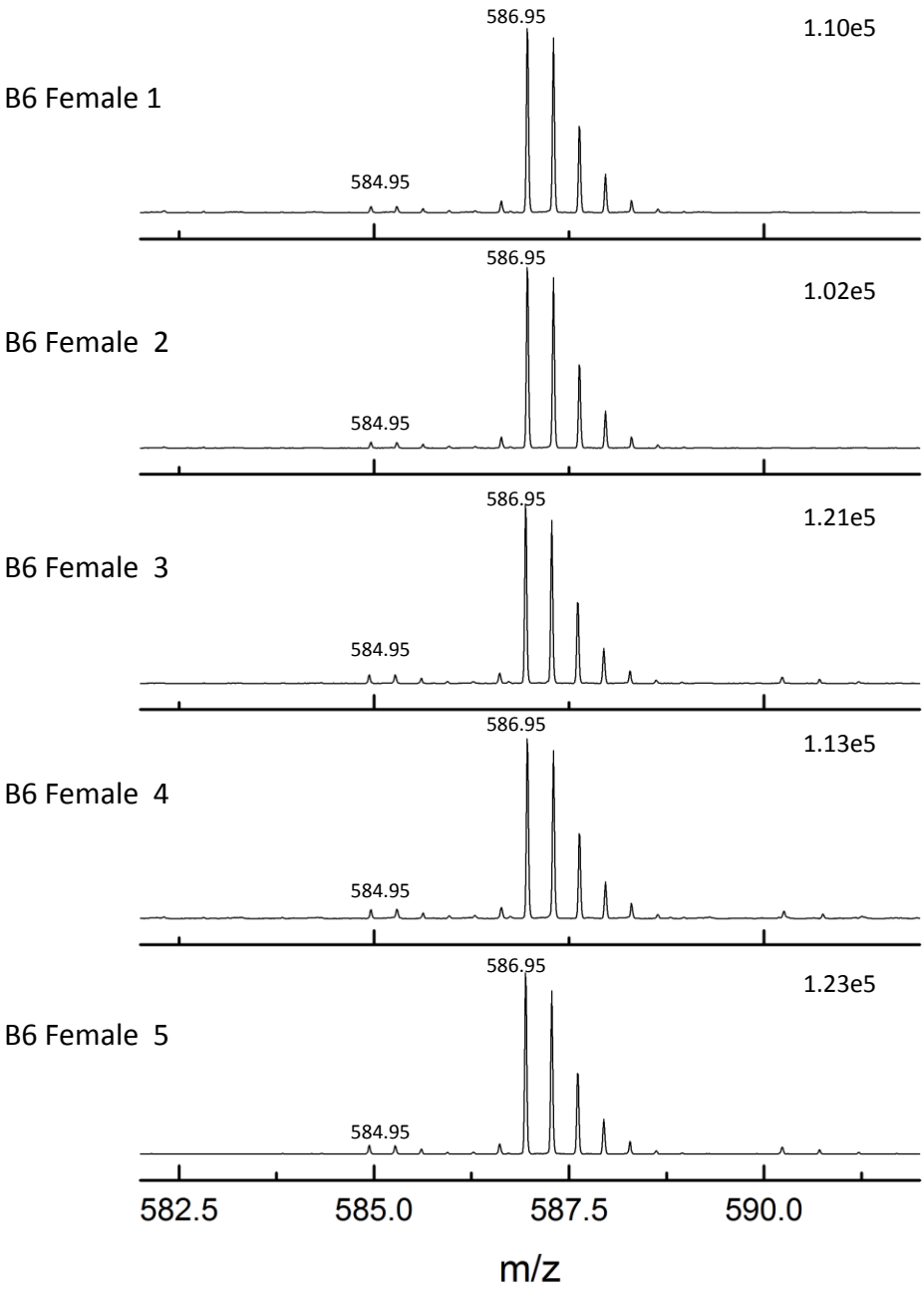
Supplementary data A. LC-MS analysis of individual female C57BL/6 mice. Q peptide 11. A known amount of QconCAT was added to five individual female urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier. No "light" analyte was detected in these five individual B6 female mice.



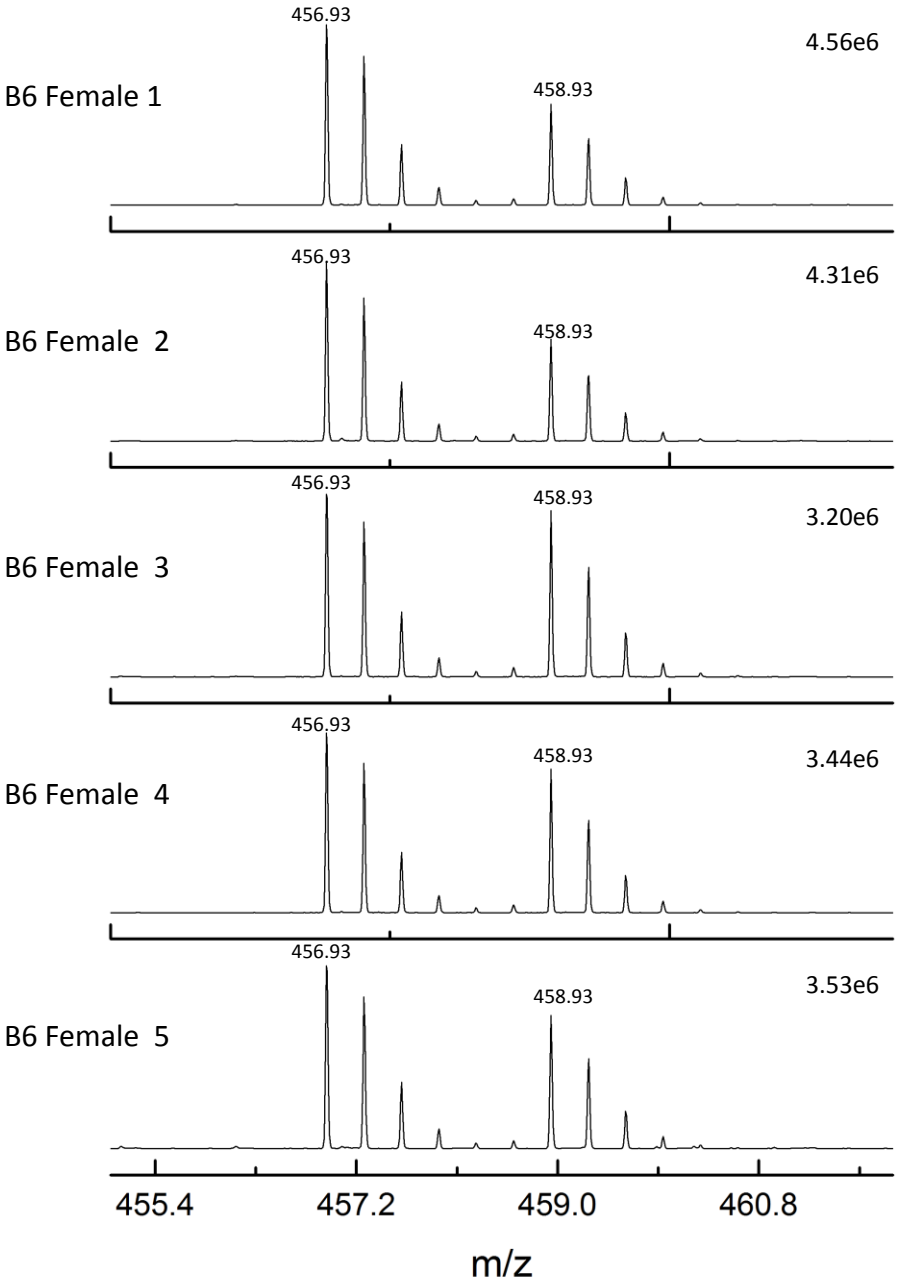
Supplementary data A. LC-MS analysis of individual female C57BL/6 mice. Q peptide 12.
A known amount of QconCAT was added to five individual female urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier.



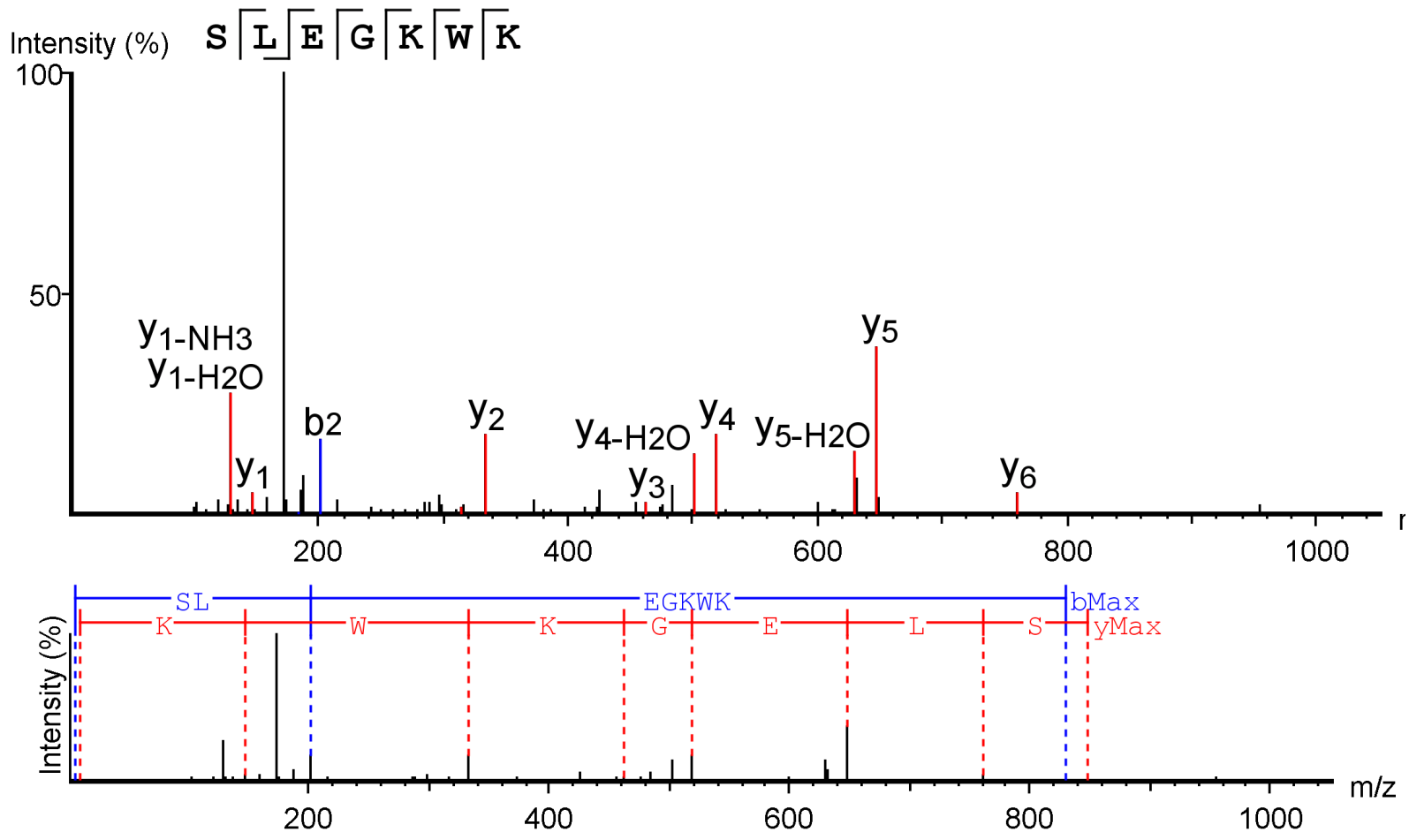
Supplementary data A. LC-MS analysis of individual female C57BL/6 mice. Q peptide 13. A known amount of QconCAT was added to five individual female urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier.



Supplementary data A. LC-MS analysis of individual female C57BL/6 mice. Q peptide 14.
A known amount of QconCAT was added to five individual female urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier..

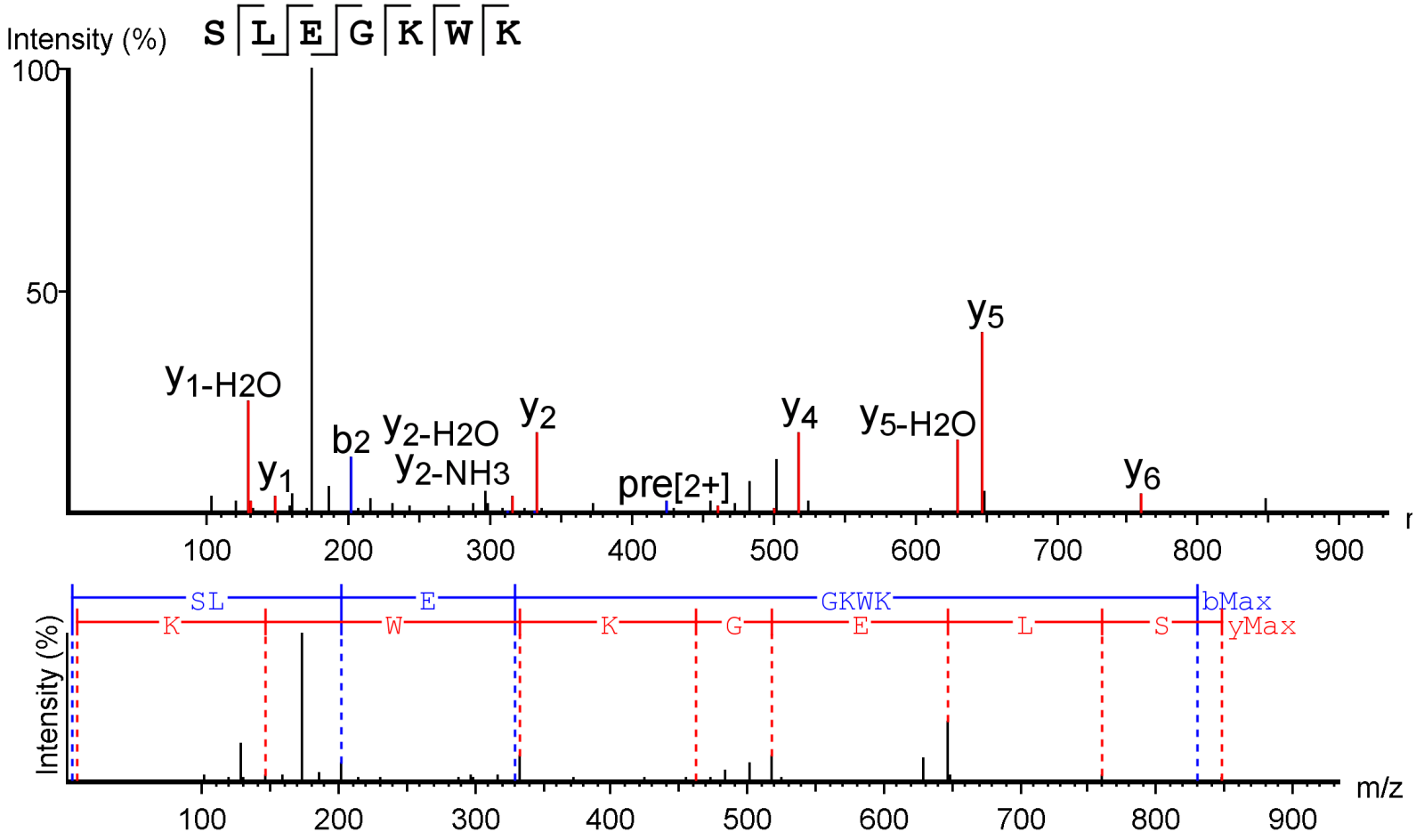


Supplementary data A. LC-MS analysis of individual female C57BL/6 mice. Q peptide 15.
A known amount of QconCAT was added to five individual female urine samples and digested using the protocol optimised in section 3.3.2. The digested material was analysed by LC-MS. The peptide pairs consist of the "light" analyte peptide and the corresponding "heavy" Q peptide 6 Da heavier.



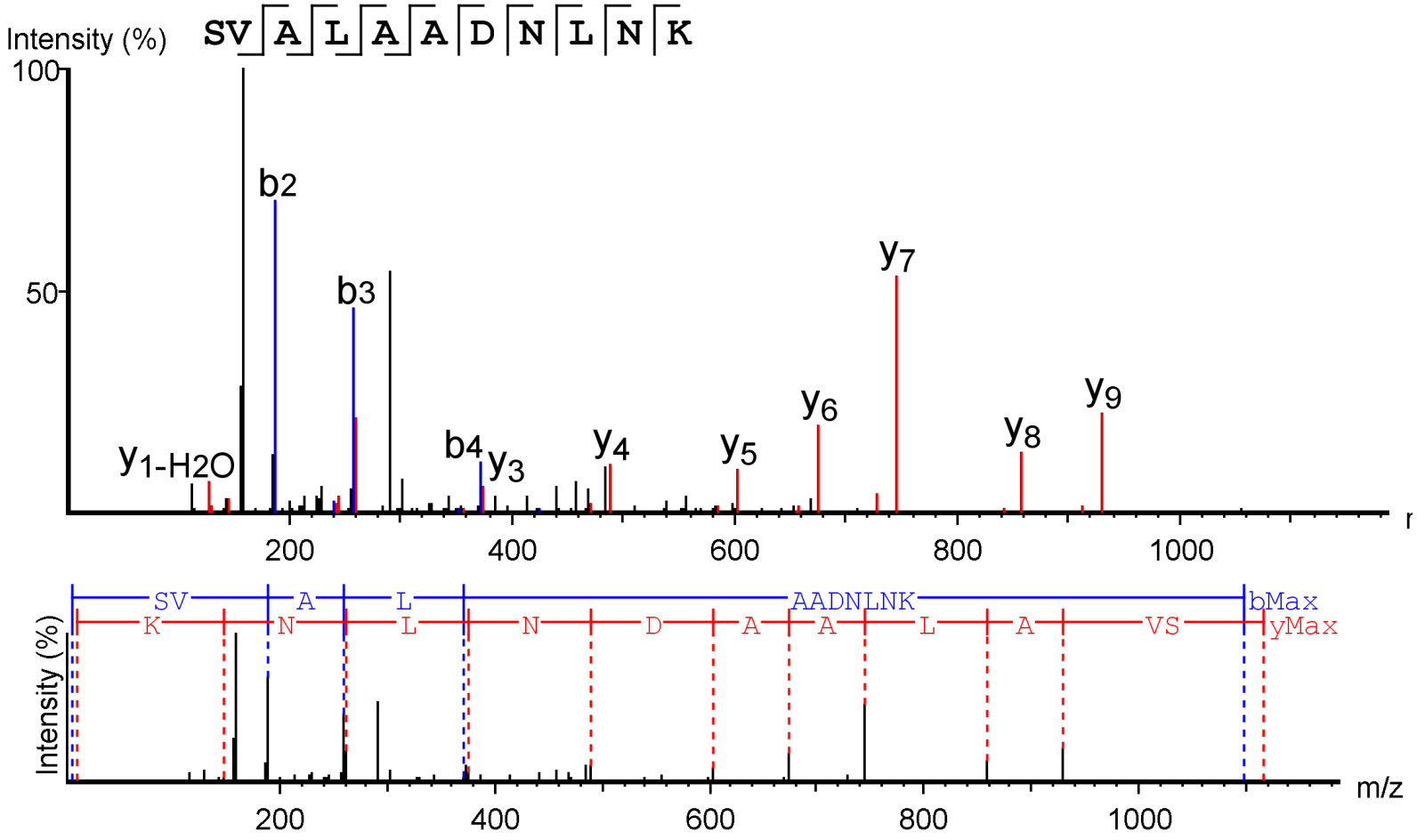
Supplementary Data B. *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* Try peptide 847 m/z

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS ®software (Bioinformatics Solutions Inc, Canada).



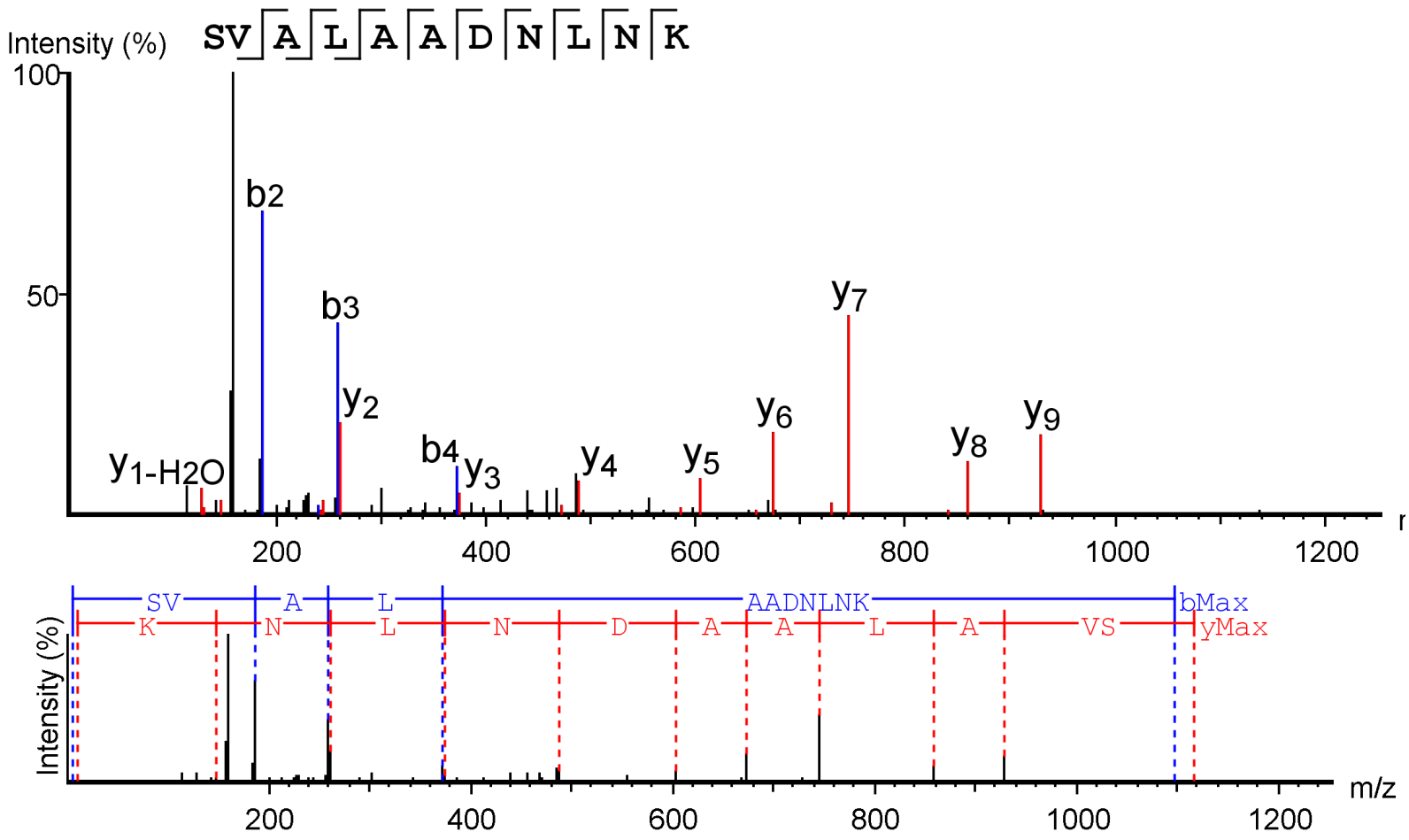
Supplementary Data B. *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* LysC peptide 847 m/z

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS ®software (Bioinformatics Solutions Inc, Canada).



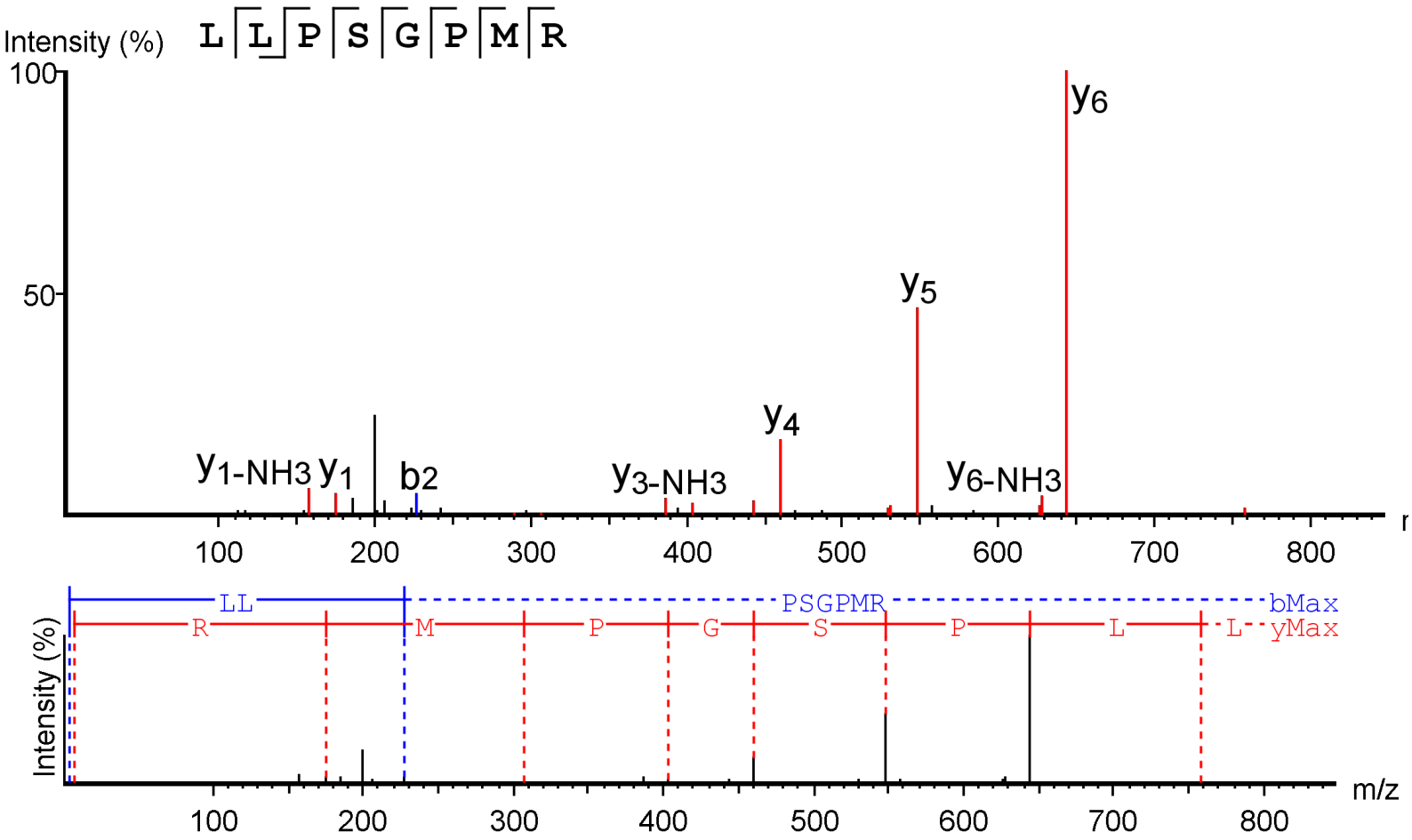
Supplementary Data B. *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* tryptic peptide 1114 m/z

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS ®software (Bioinformatics Solutions Inc, Canada).



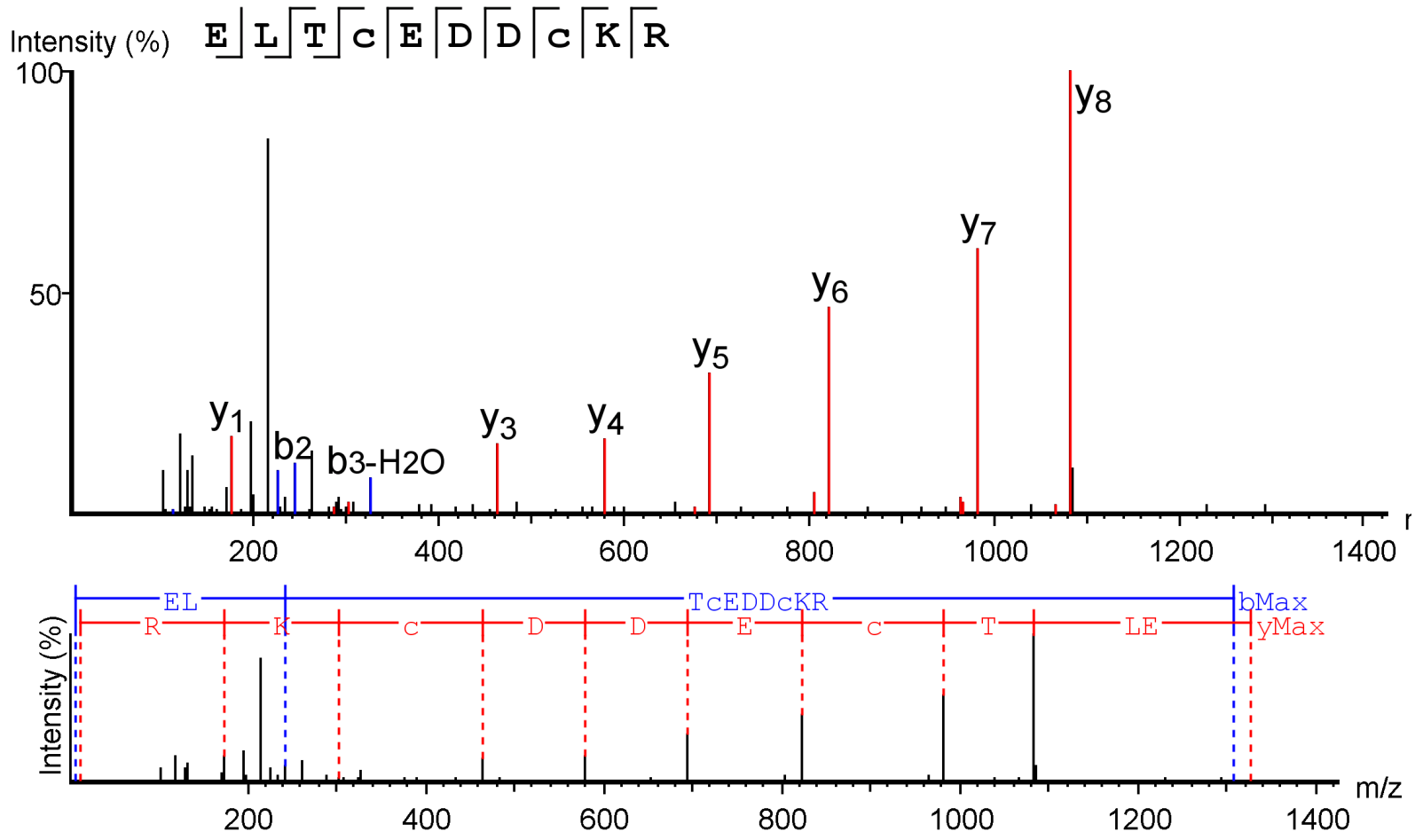
Supplementary Data B. *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* LysC peptide 1114 m/z

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS ®software (Bioinformatics Solutions Inc, Canada).



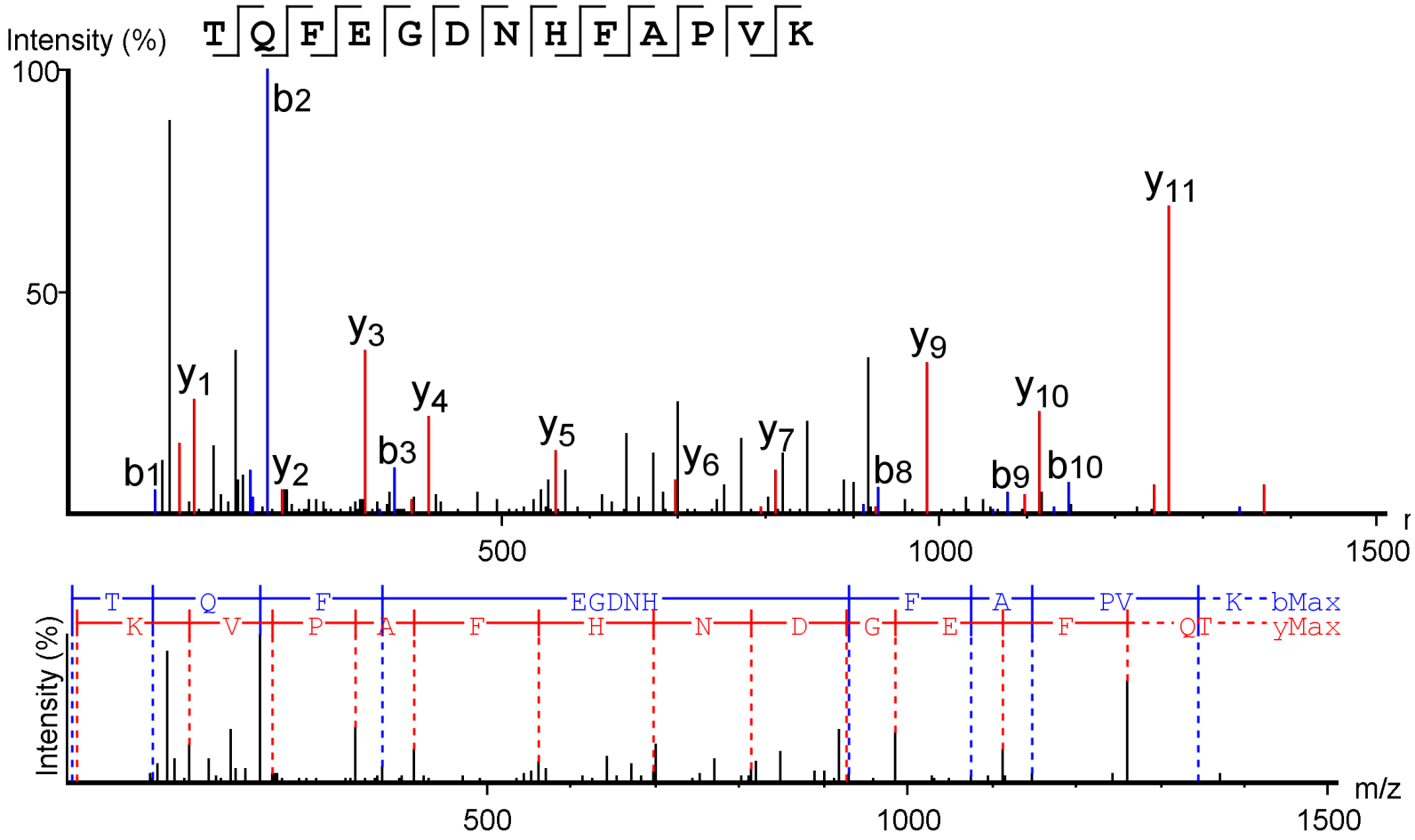
Supplementary Data B. *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* tryptic peptide 870 m/z

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS ®software (Bioinformatics Solutions Inc, Canada).



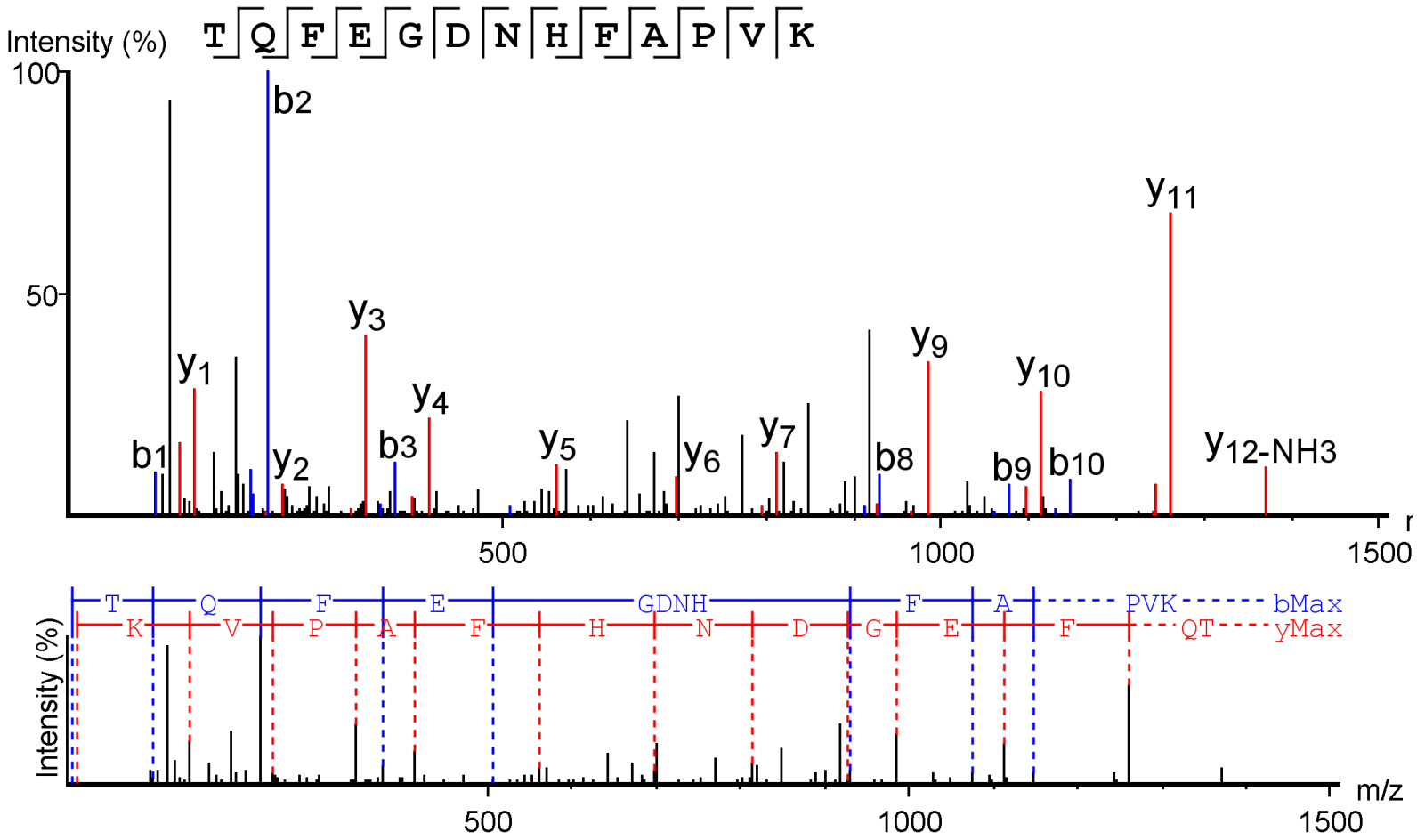
Supplementary Data B. *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* tryptic peptide 1325 m/z

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS ®software (Bioinformatics Solutions Inc, Canada).



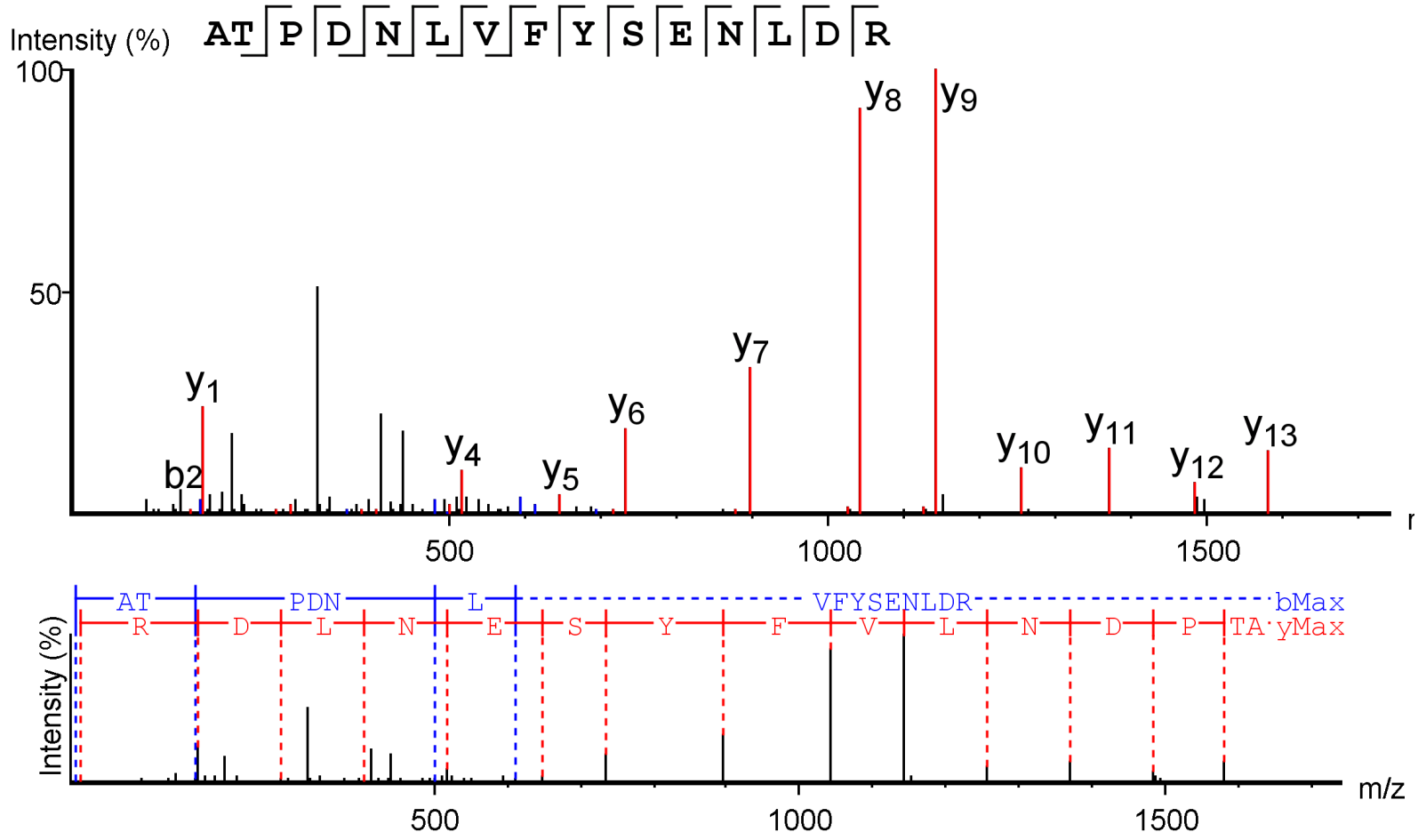
Supplementary Data B. *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* tryptic peptide 1489 m/z

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS ®software (Bioinformatics Solutions Inc, Canada).



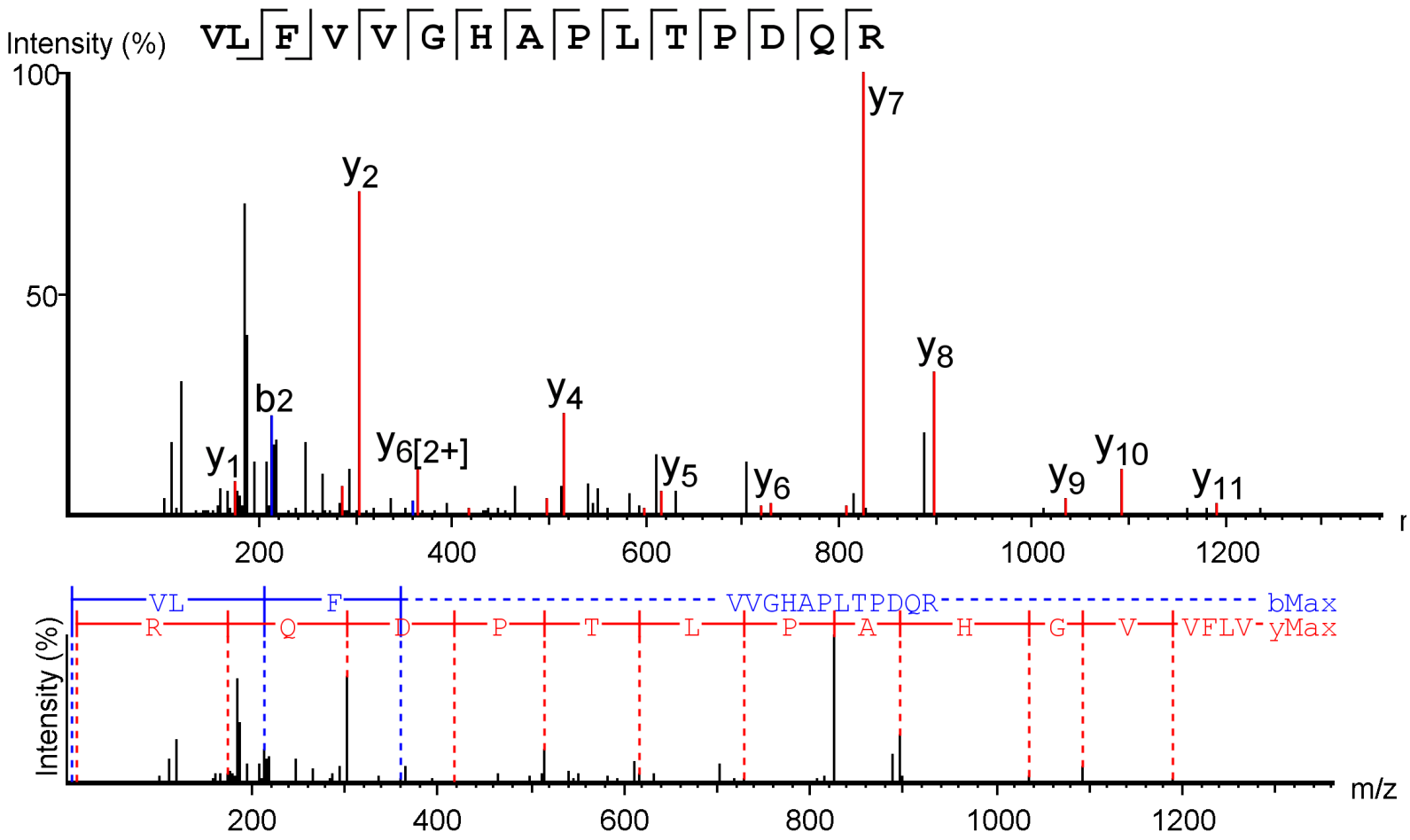
Supplementary Data B. *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* LysC peptide 1489 m/z

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS ®software (Bioinformatics Solutions Inc, Canada).



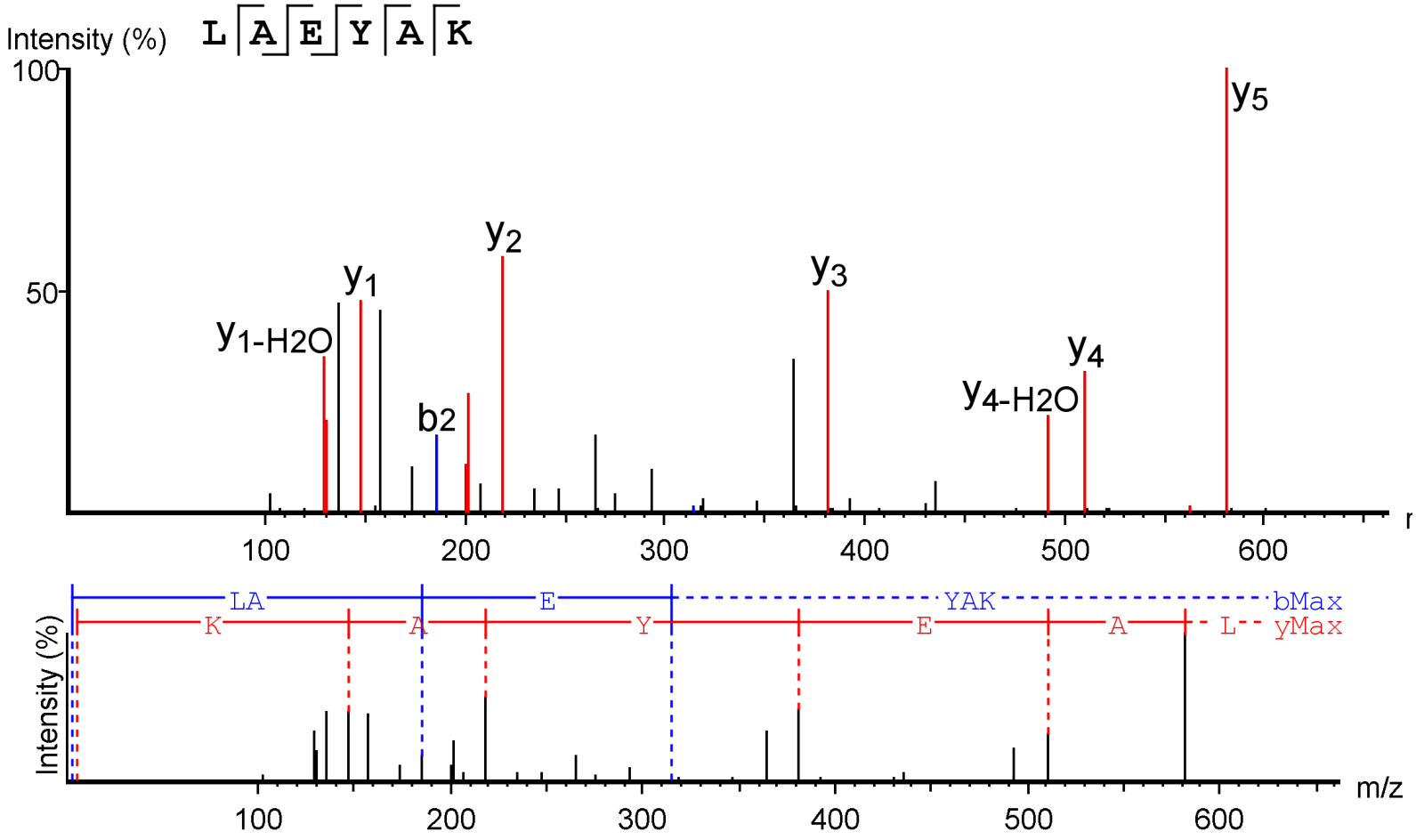
Supplementary Data B. *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* tryptic peptide 1753 m/z

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS ®software (Bioinformatics Solutions Inc, Canada).



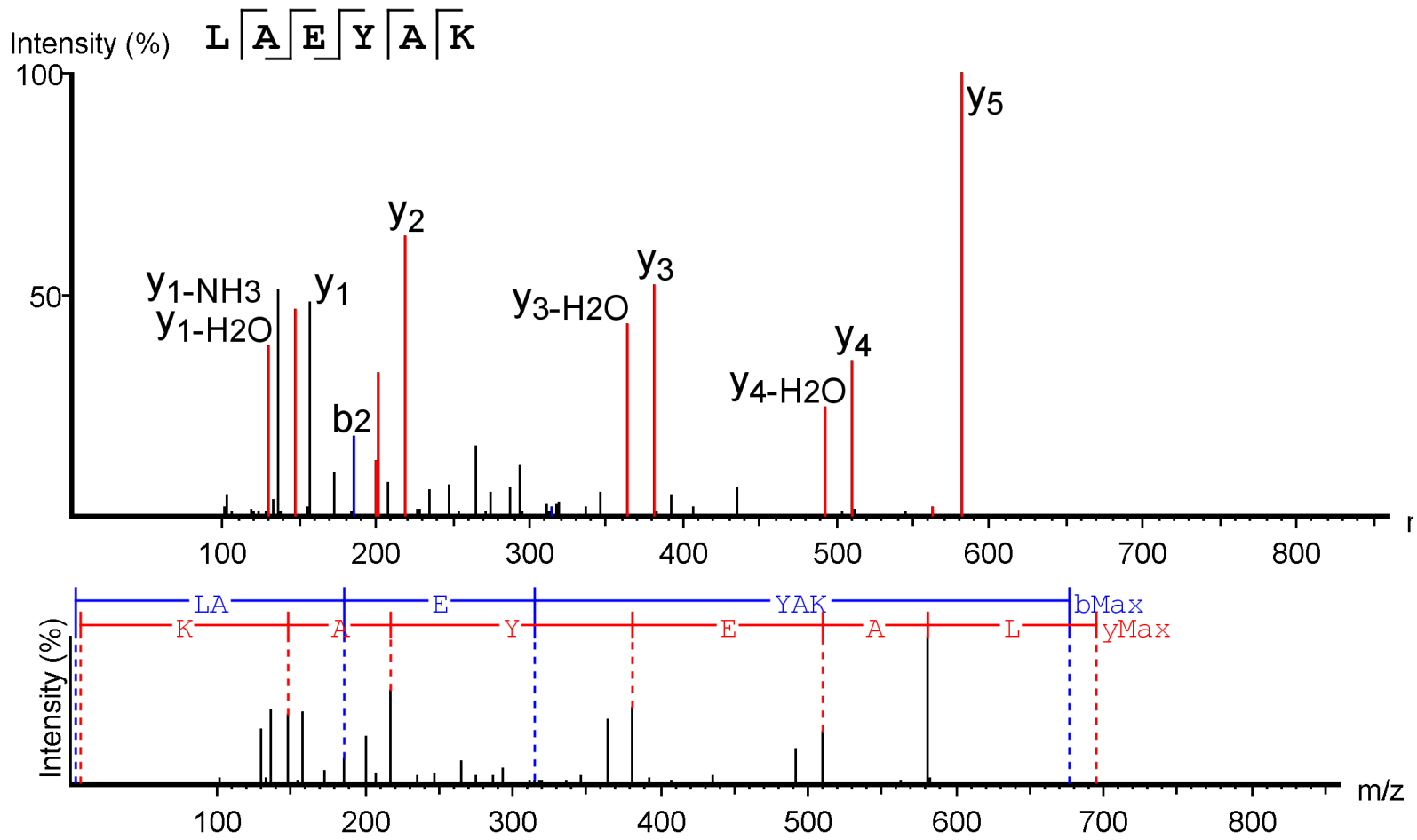
Supplementary Data B. *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* tryptic peptide 1649 m/z

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS ®software (Bioinformatics Solutions Inc, Canada).



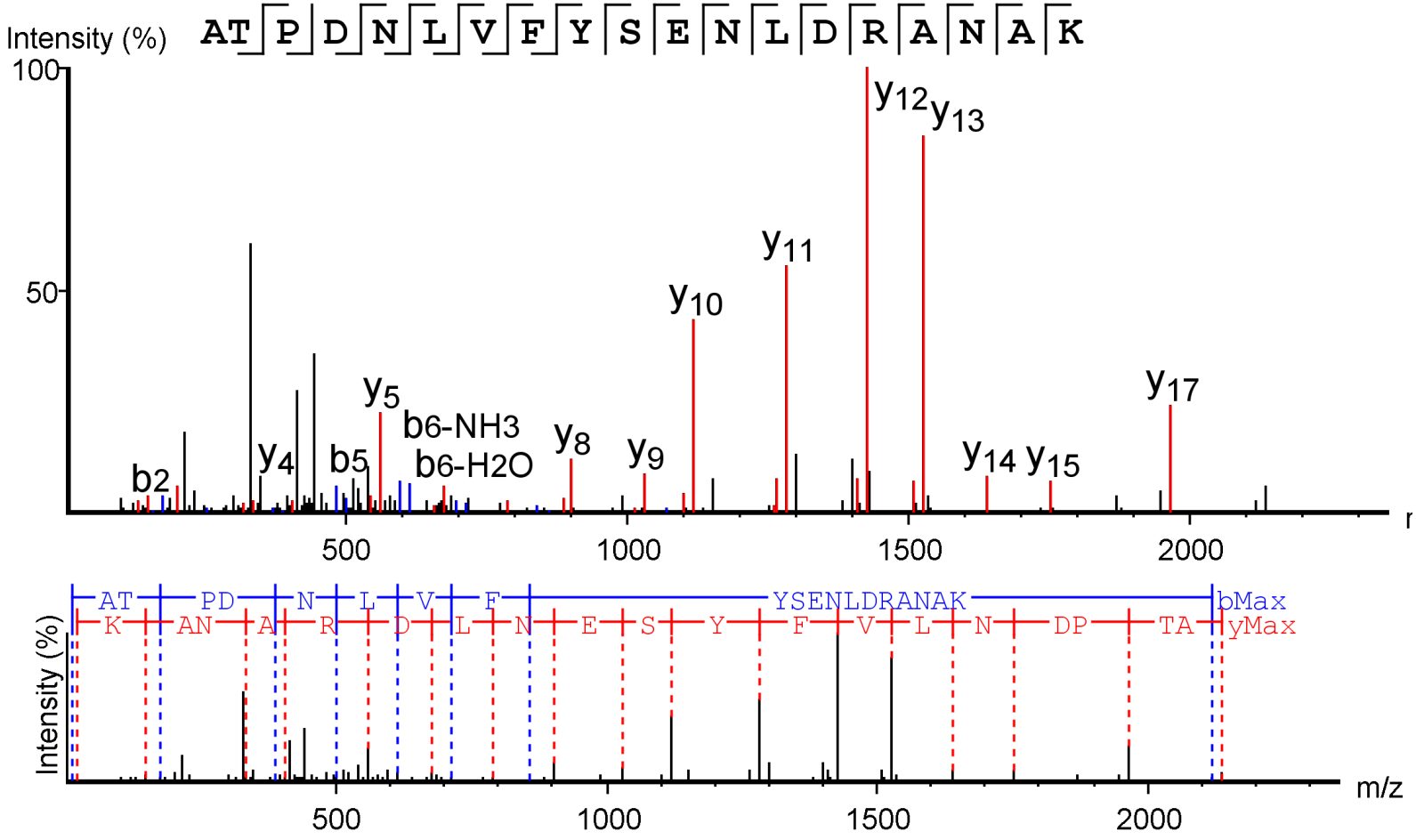
Supplementary Data B. *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* tryptic peptide 643 m/z

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS ®software (Bioinformatics Solutions Inc, Canada).



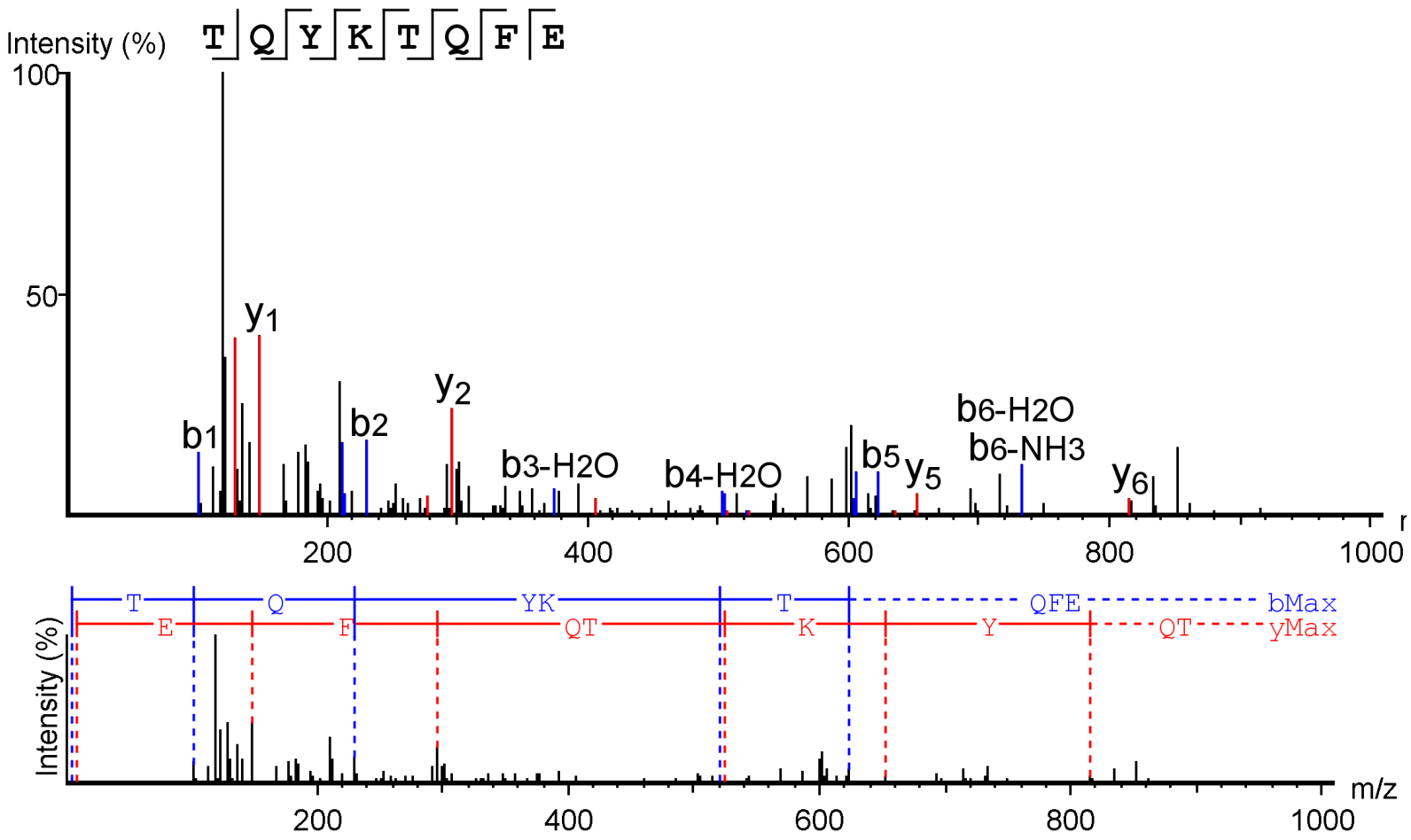
Supplementary Data B. *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* tryptic peptide 643 m/z

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS ®software (Bioinformatics Solutions Inc, Canada).



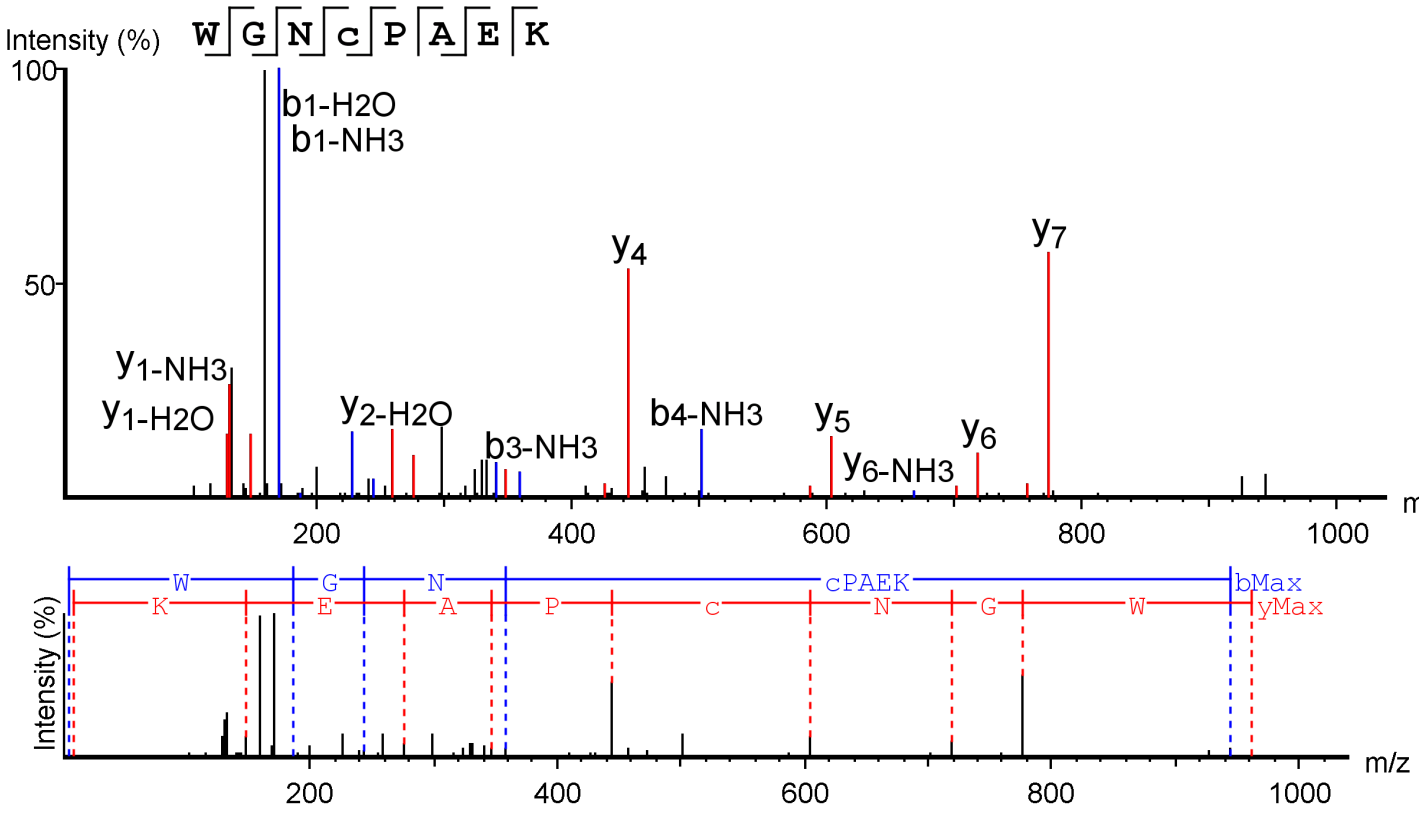
Supplementary Data B. *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* LysC peptide 2138 m/z

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS ®software (Bioinformatics Solutions Inc, Canada).



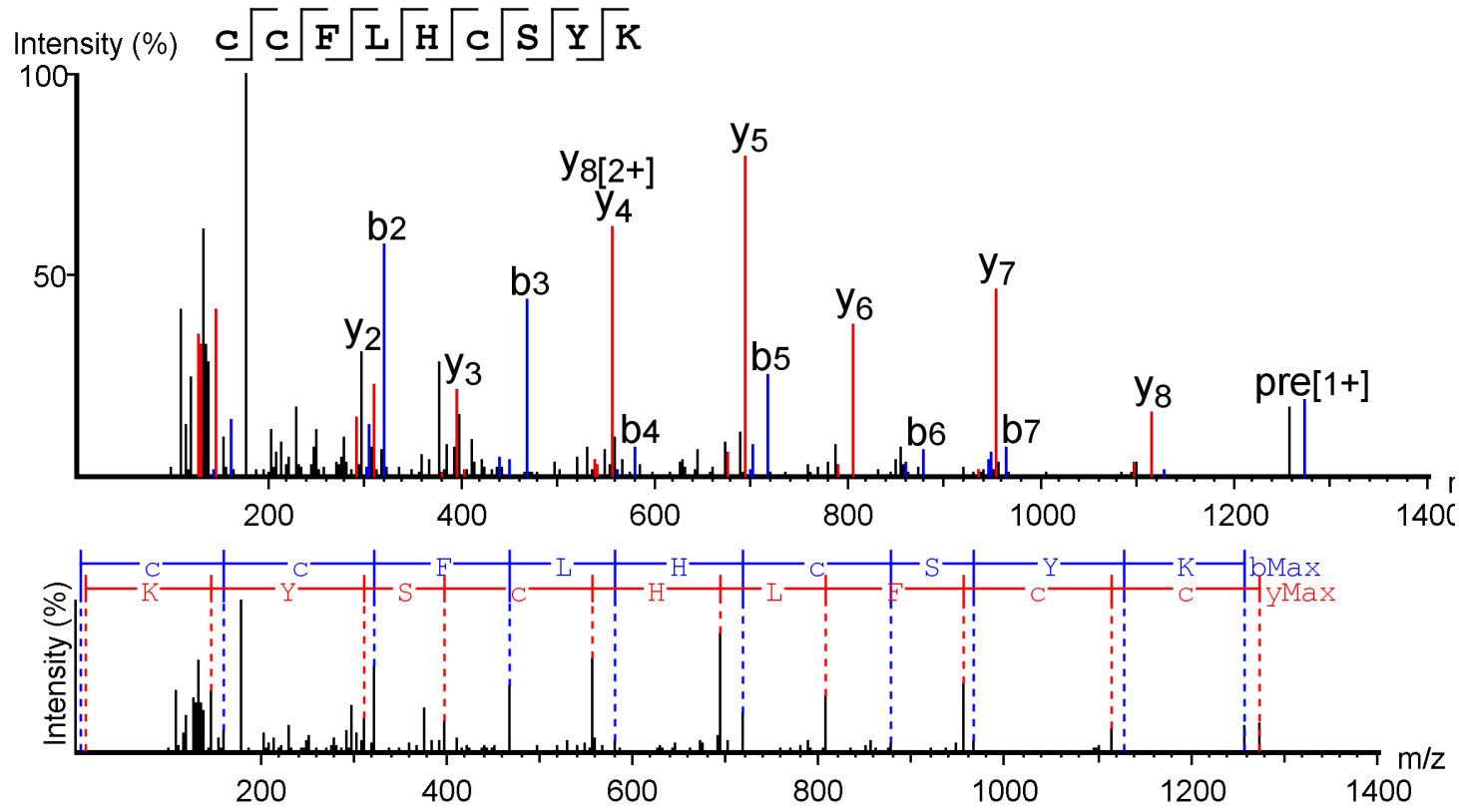
Supplementary Data B. *De novo* sequence analysis of the processed MS/MS spectra of *M. minutus* GluC peptide 1043 m/z

M. minutus glass rod samples containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 chromatography system. The samples were injected (typically equivalent to 500fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS ®software (Bioinformatics Solutions Inc, Canada).



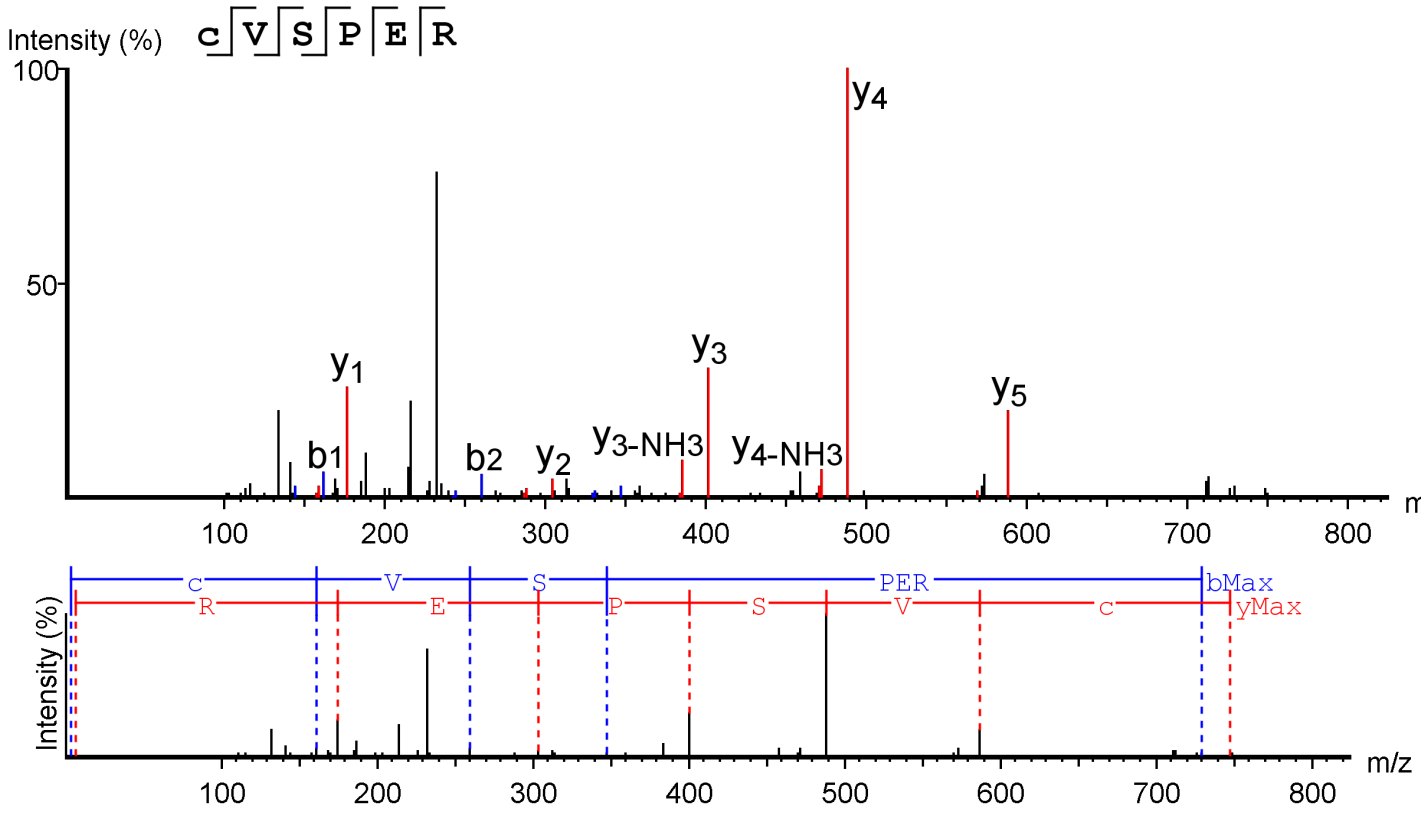
Supplementary data C. ***De novo*** **sequence analysis of the processed MS/MS spectra of** ***M. murinus*** **LysC peptide 961 m/z.**

M. murinus urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



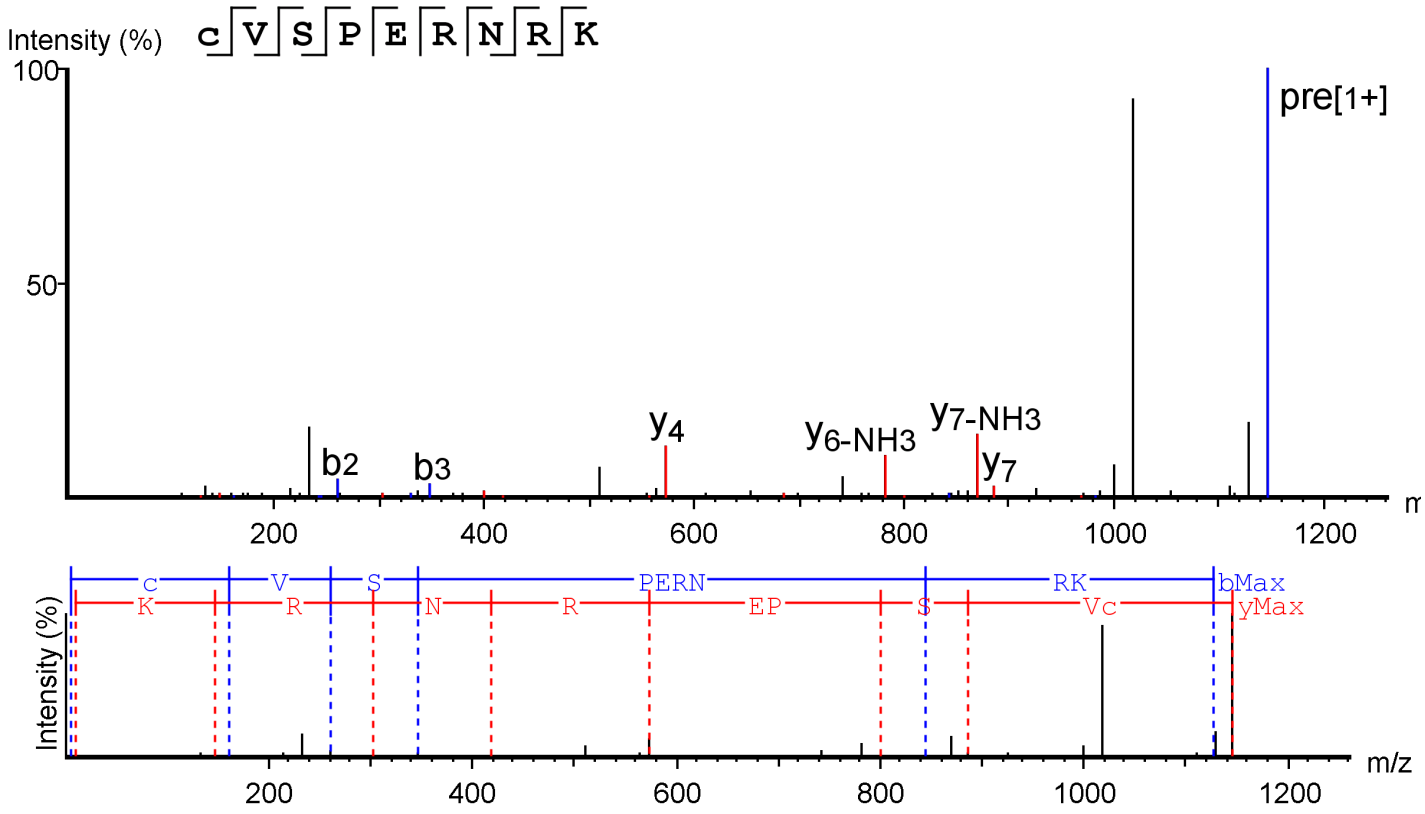
Supplementary data C. *De novo* sequence analysis of the processed MS/MS spectra of *M. murinus* tryptic peptide 1274 m/z.

M. murinus urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



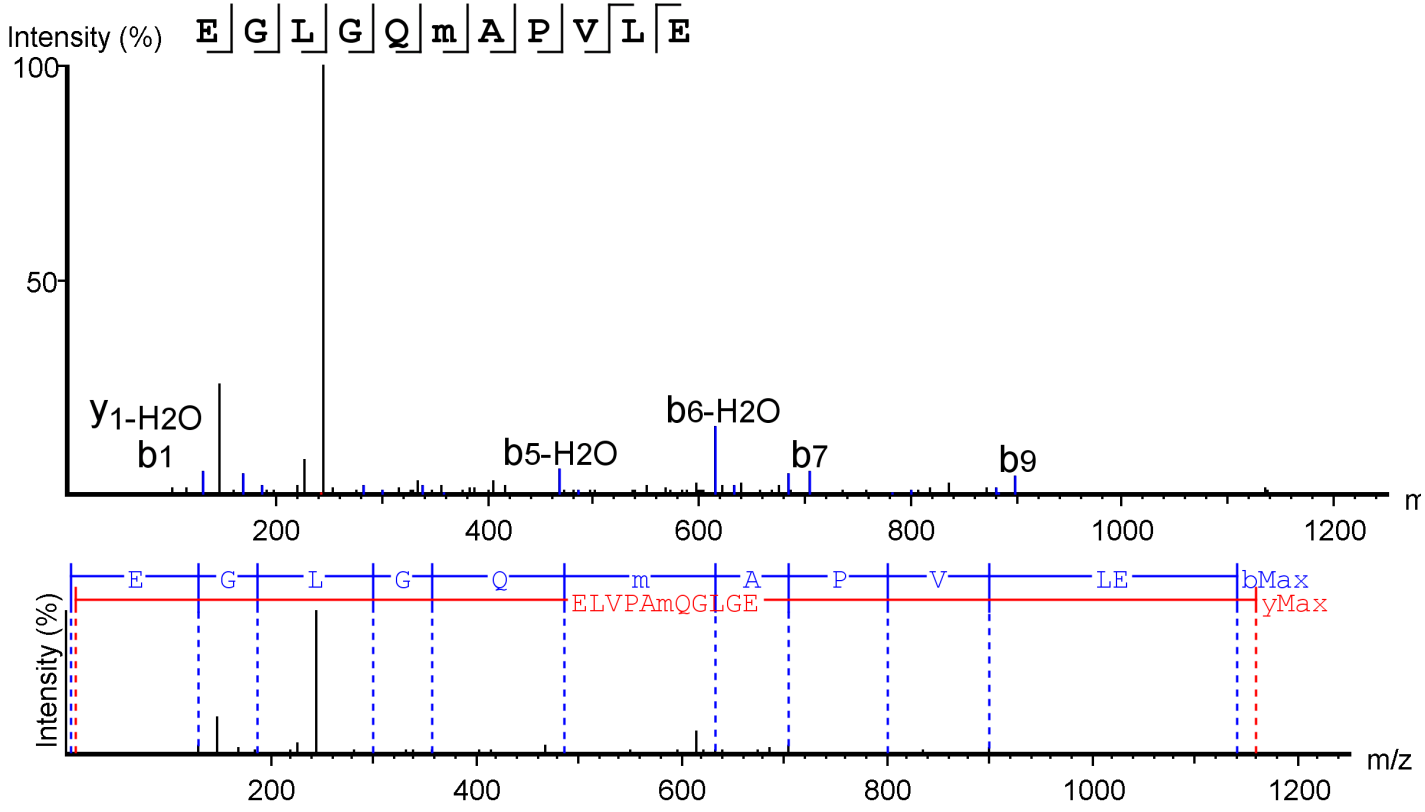
Supplementary data C. ***De novo*** **sequence analysis of the processed MS/MS spectra of** ***M. murinus*** **tryptic peptide 746 m/z.**

M. murinus urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



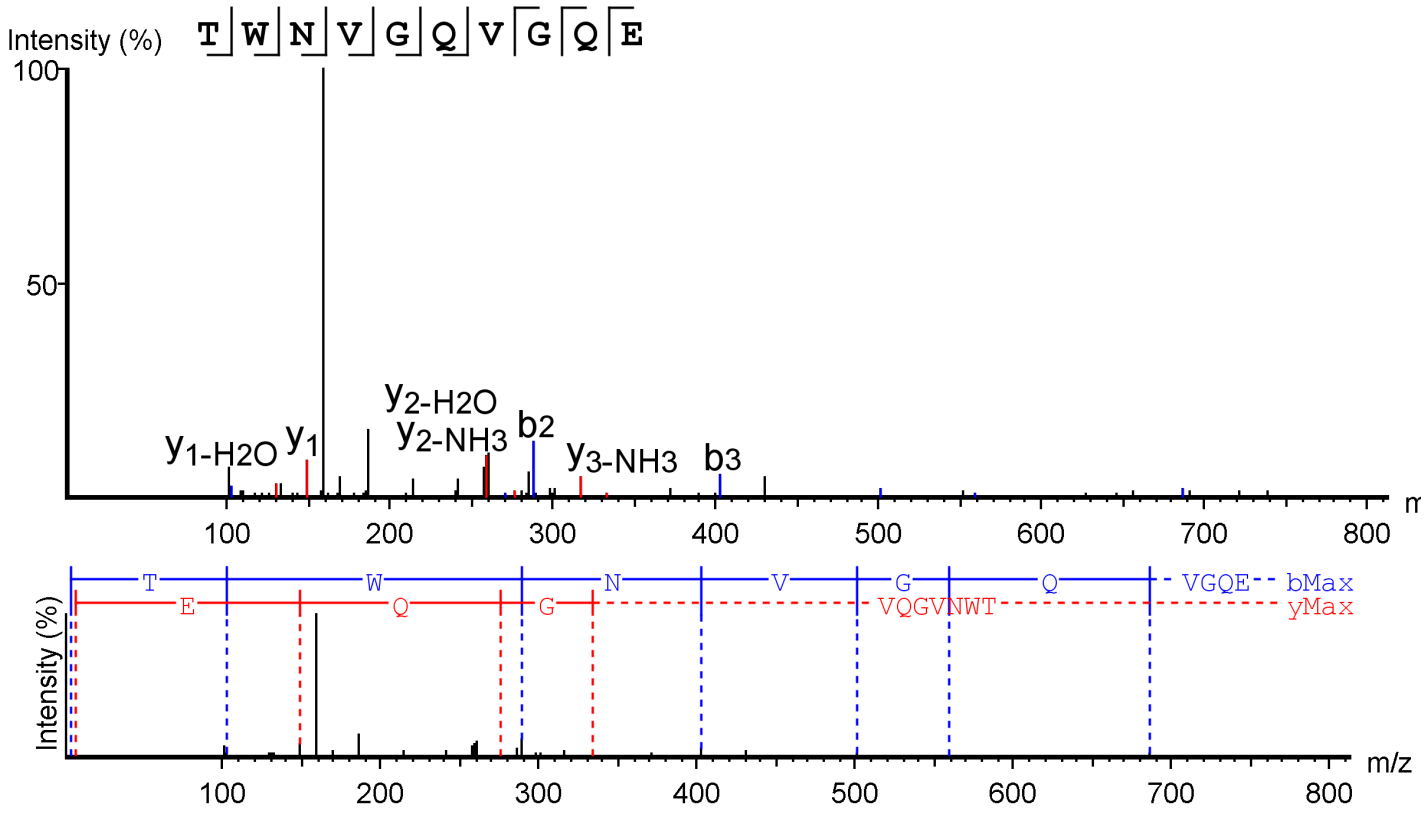
Supplementary data C. ***De novo*** **sequence analysis of the processed MS/MS spectra of** ***M. murinus*** **LysC peptide 1145 m/z.**

M. murinus urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



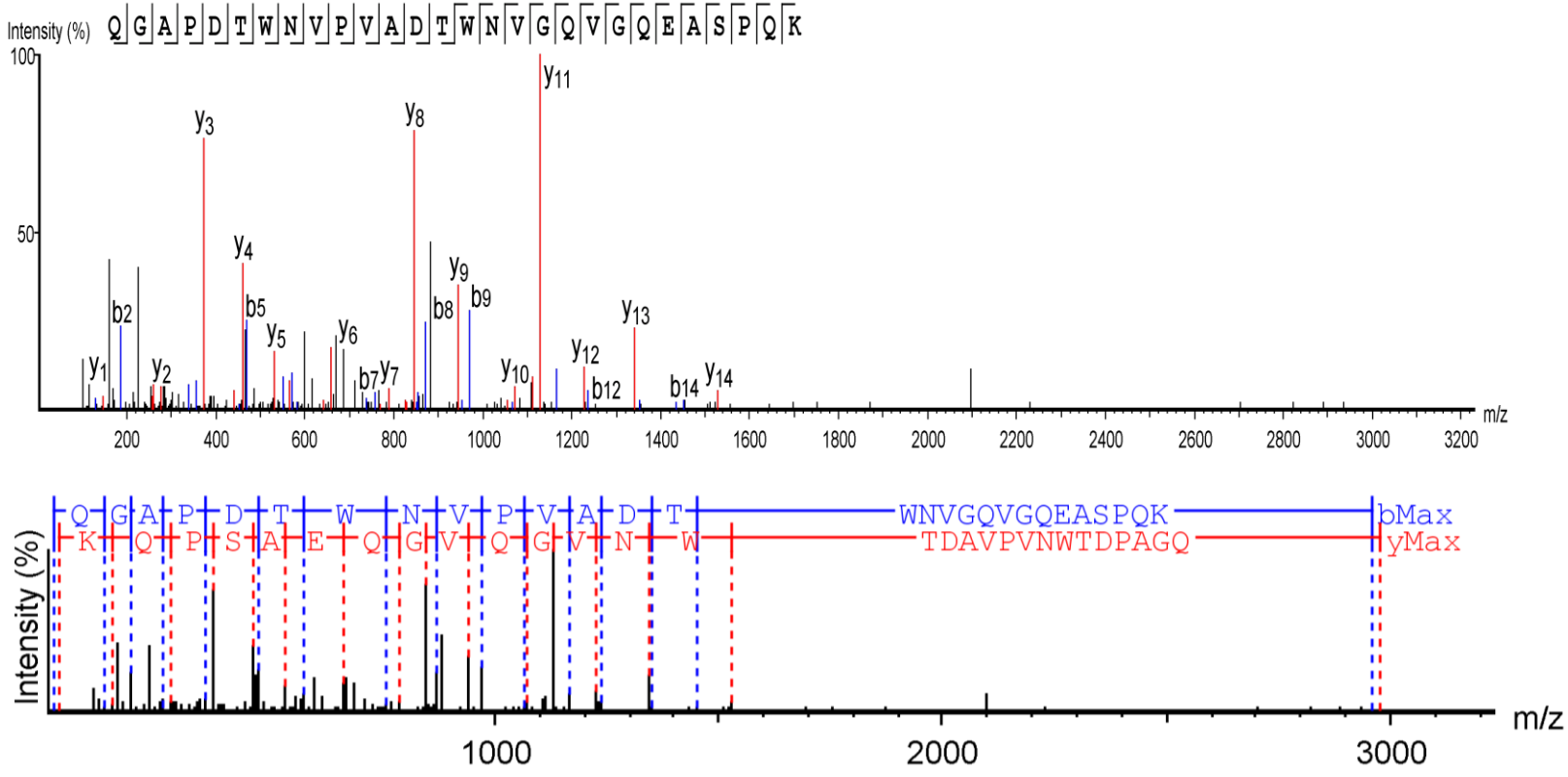
Supplementary data C. ***De novo*** **sequence analysis of the processed MS/MS spectra of** ***M. murinus*** **GluC peptide 1158 m/z.**

M. murinus urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



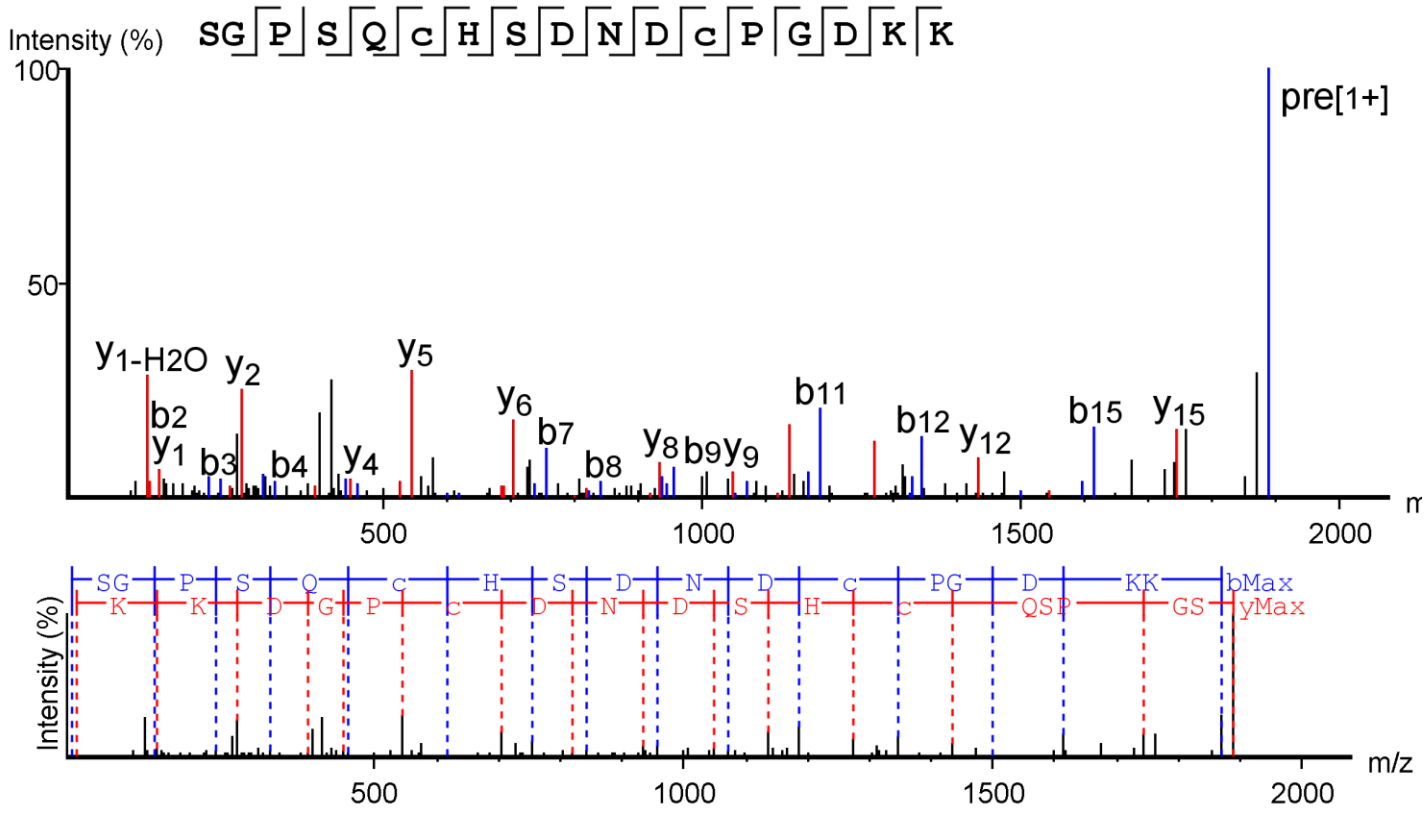
Supplementary data C. ***De novo*** **sequence analysis of the processed MS/MS spectra of** ***M. murinus*** **GluC peptide 1116 m/z.**

M. murinus urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



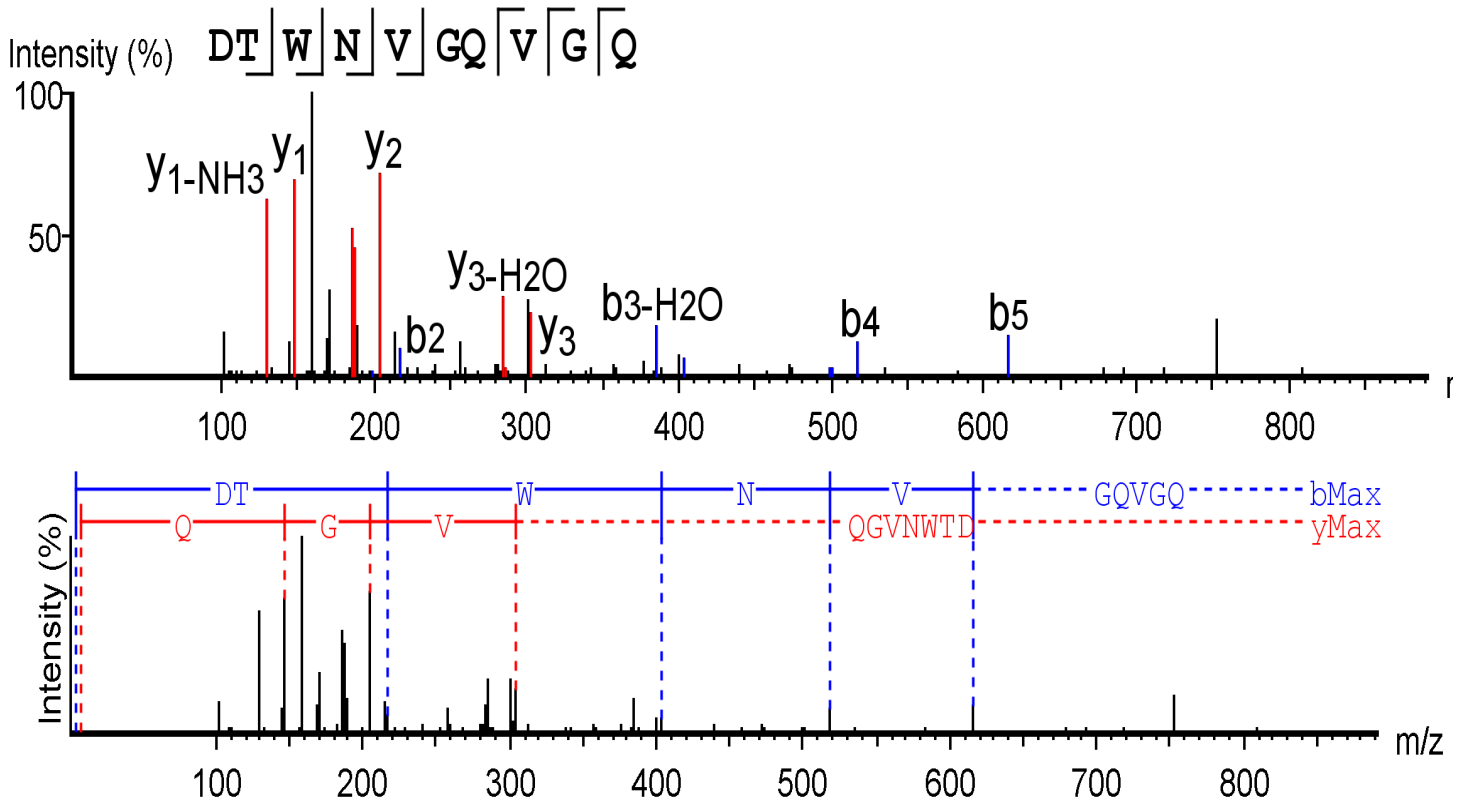
Supplementary data C. ***De novo*** **sequence analysis of the processed MS/MS spectra of** ***M. Murinus*** **LysC peptide 2979 m/z.**

M. Murinus urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



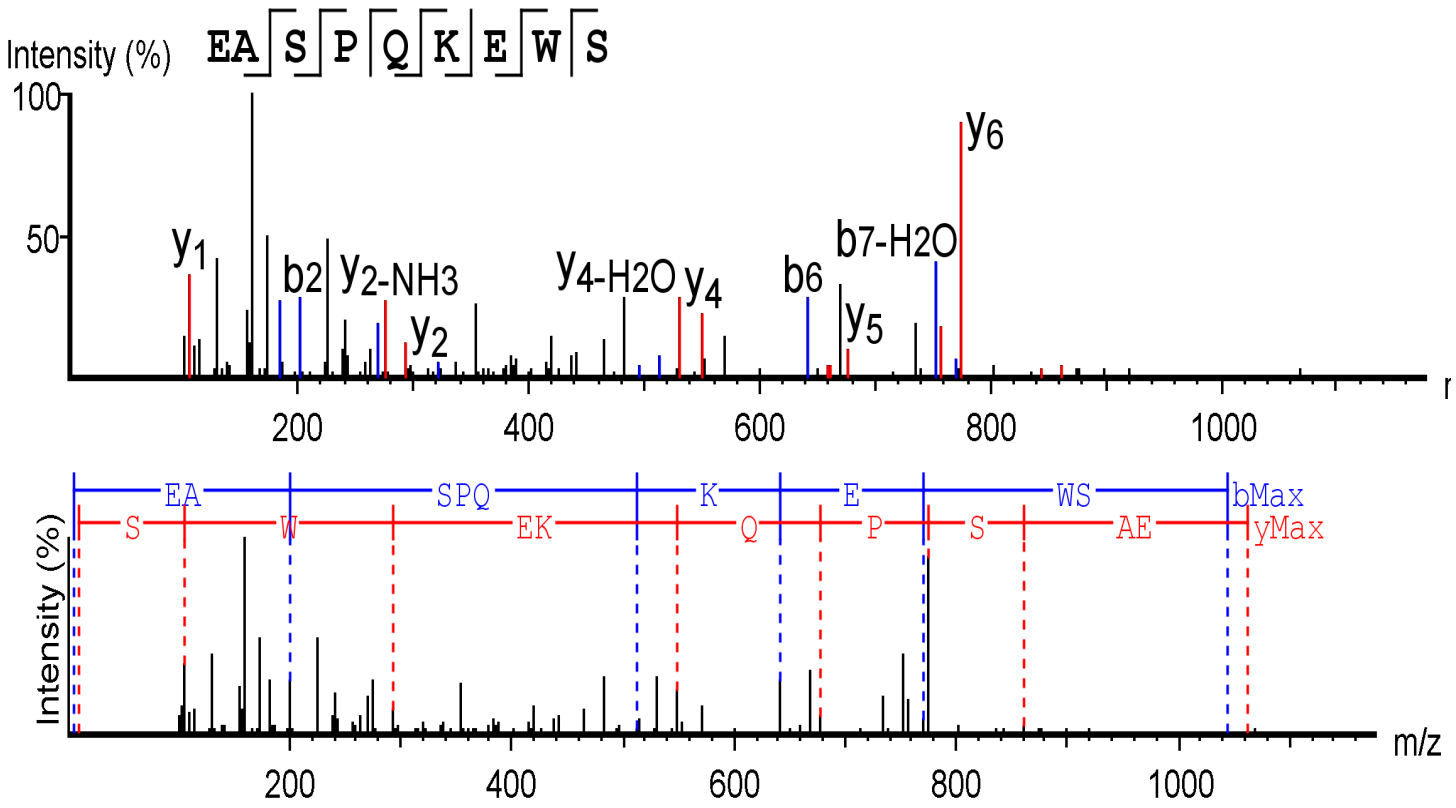
Supplementary data C. ***De novo*** **sequence analysis of the processed MS/MS spectra of** ***M. Murinus*** **LysC peptide 1888 m/z.**

M. Murinus urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



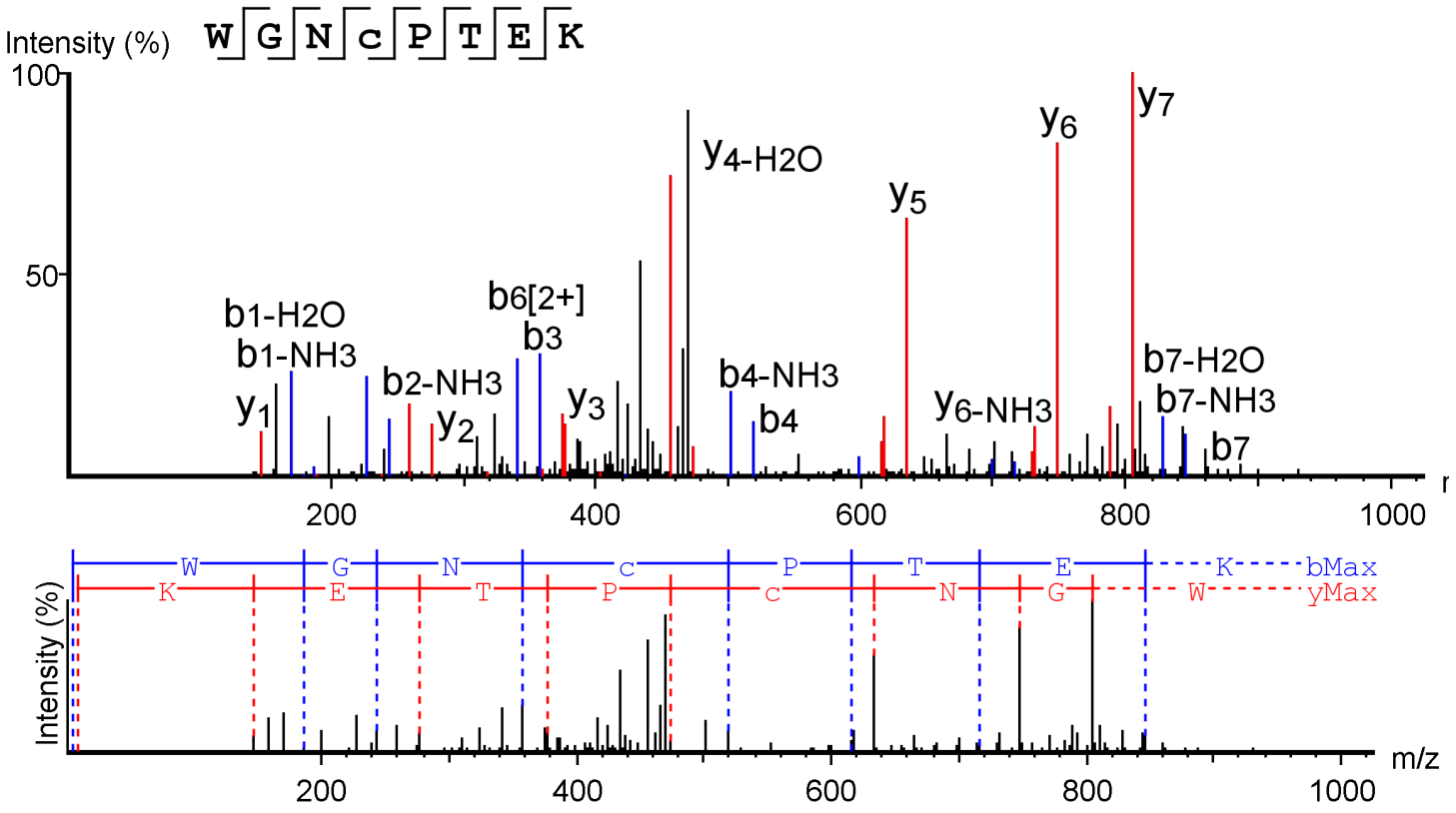
Supplementary data C. ***De novo*** **sequence analysis of the processed MS/MS spectra of** ***M. Murinus*** **AspN peptide 1102 m/z.**

M. Murinus urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



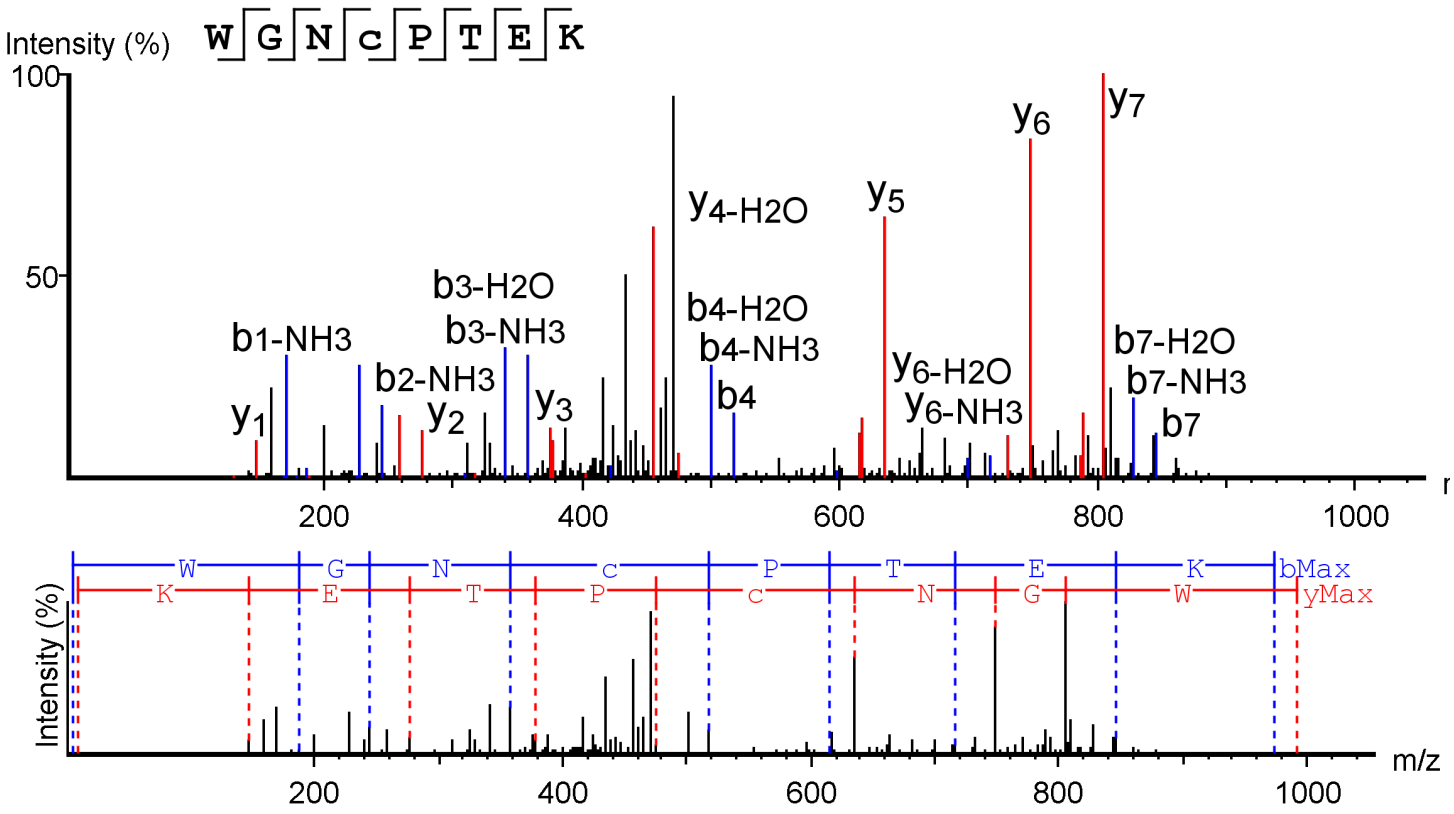
Supplementary data C. ***De novo*** **sequence analysis of the processed MS/MS spectra of** ***M. Murinus*** **AspN peptide 1060 m/z.**

M. Murinus urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



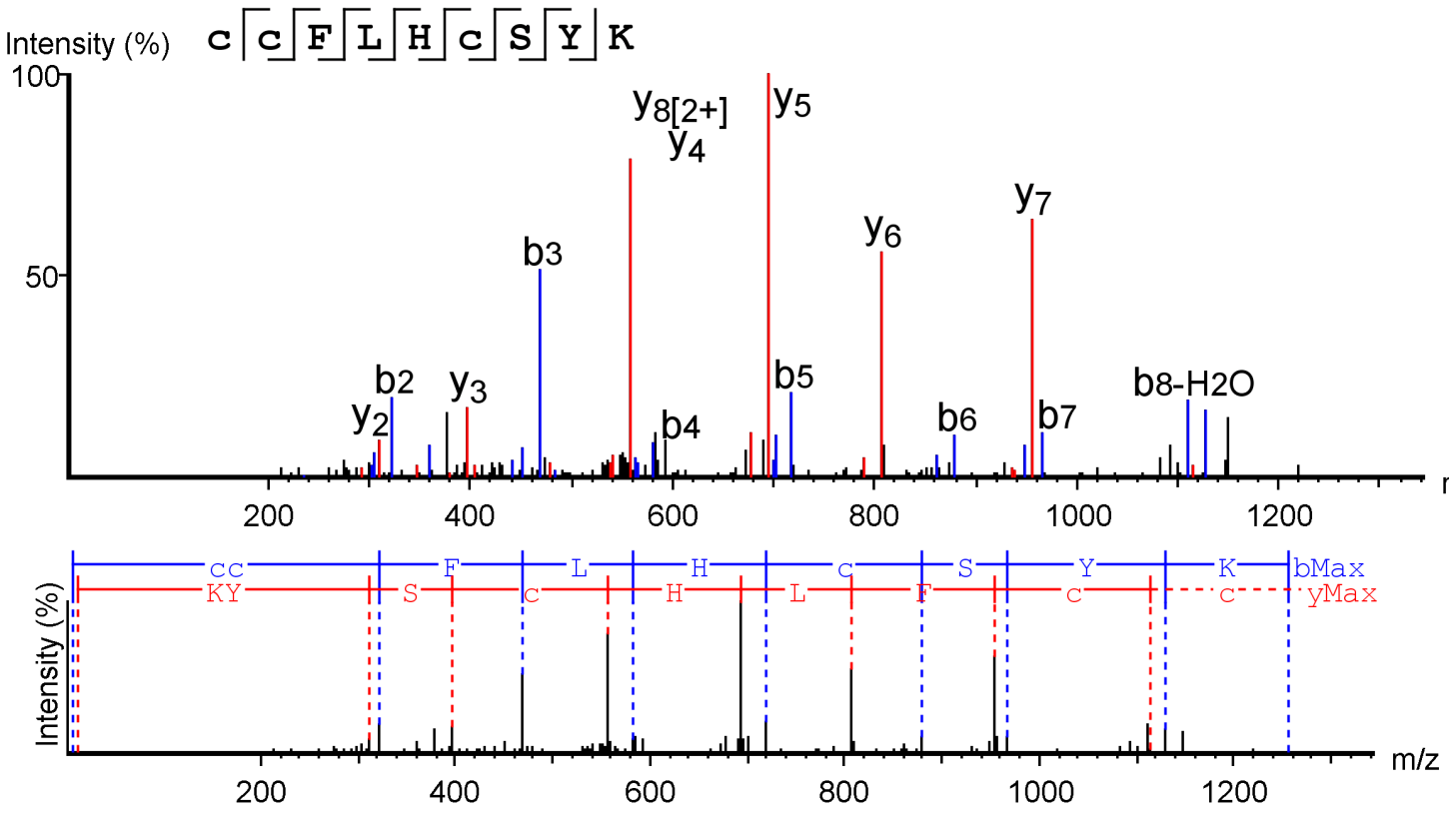
Supplementary data C. ***De novo*** **sequence analysis of the processed MS/MS spectra of** ***M. lehilahytsara*** **tryptic peptide 991 m/z.**

M. lehilahytsara urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



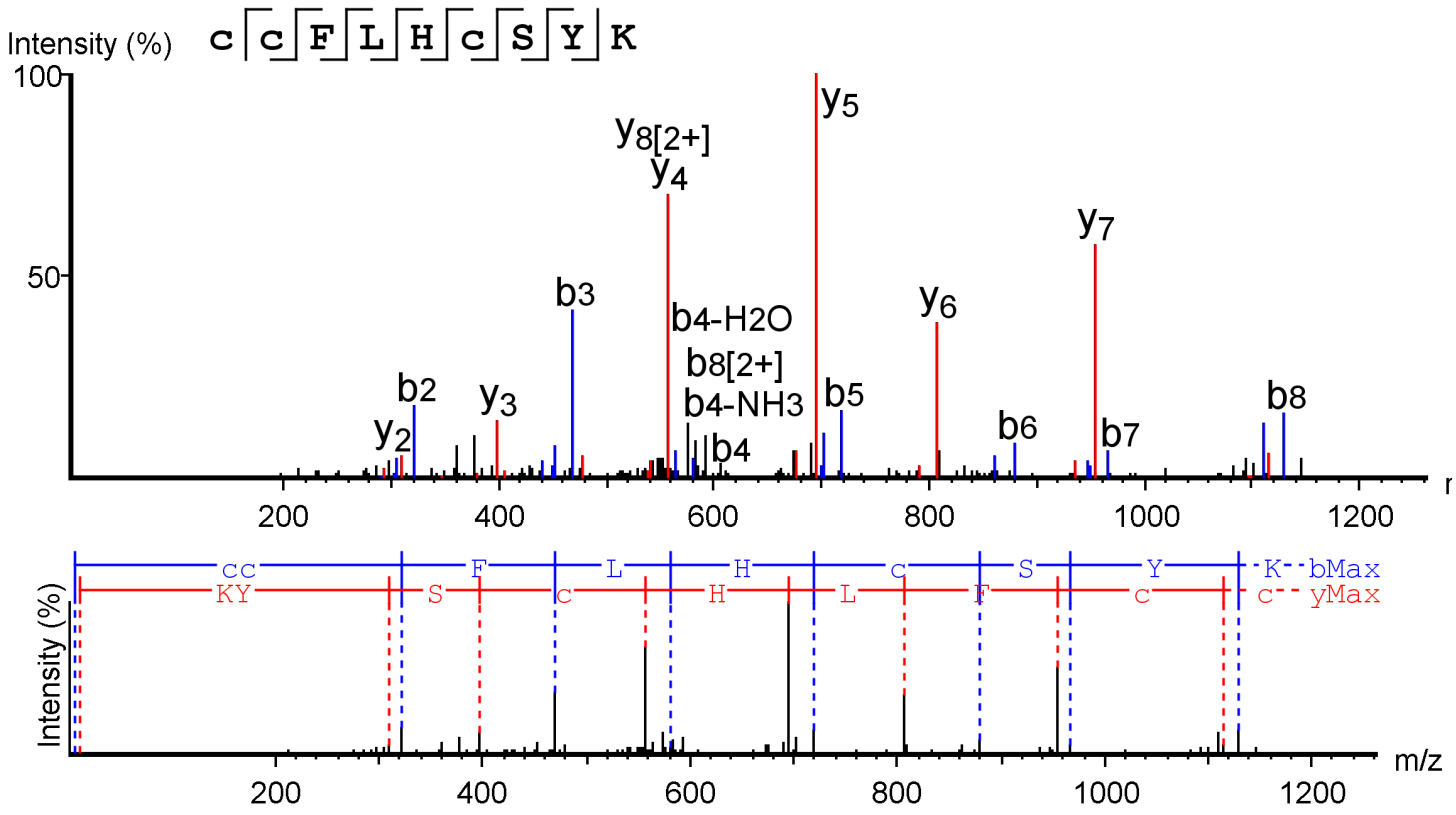
Supplementary data C. ***De novo*** **sequence analysis of the processed MS/MS spectra of** ***M. lehilahytsara*** **LysC peptide 991 m/z.**

M. lehilahytsara urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



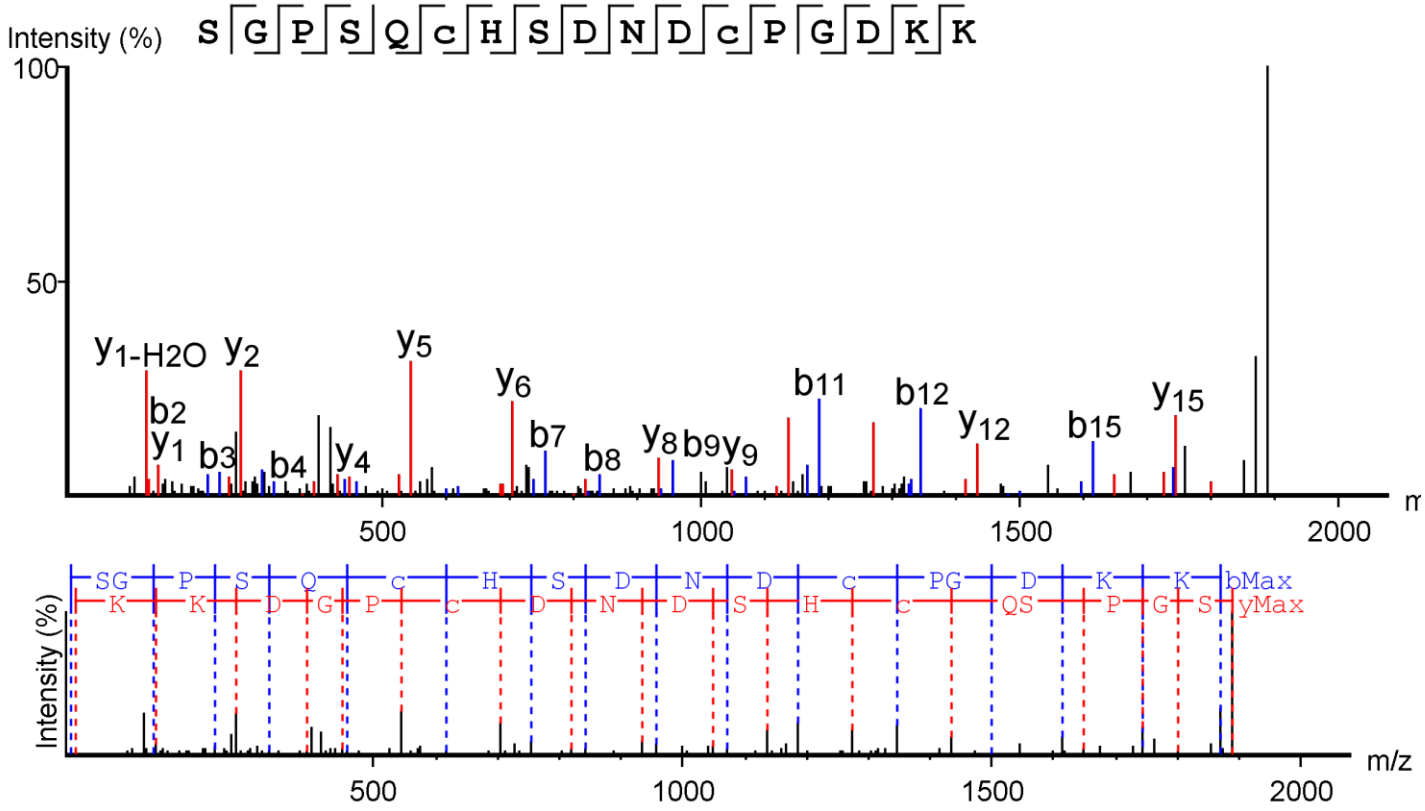
Supplementary data C. *De novo* sequence analysis of the processed MS/MS spectra of *M. lehilahytsara* tryptic peptide 1274 m/z.

M. lehilahytsara urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



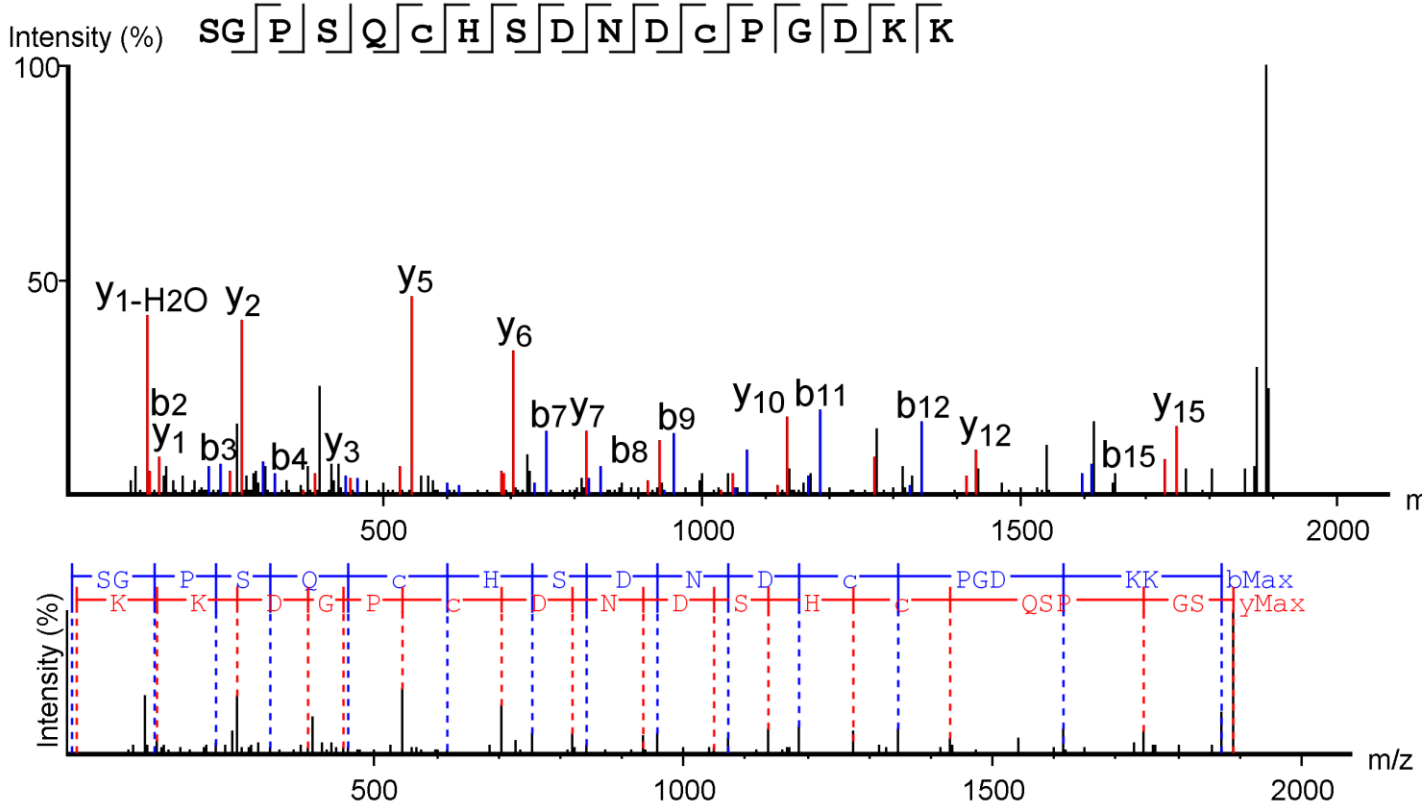
Supplementary data C. ***De novo*** **sequence analysis of the processed MS/MS spectra of** ***M. lehilahytsara*** **LysC peptide 1274 m/z.**

M. lehilahytsara urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



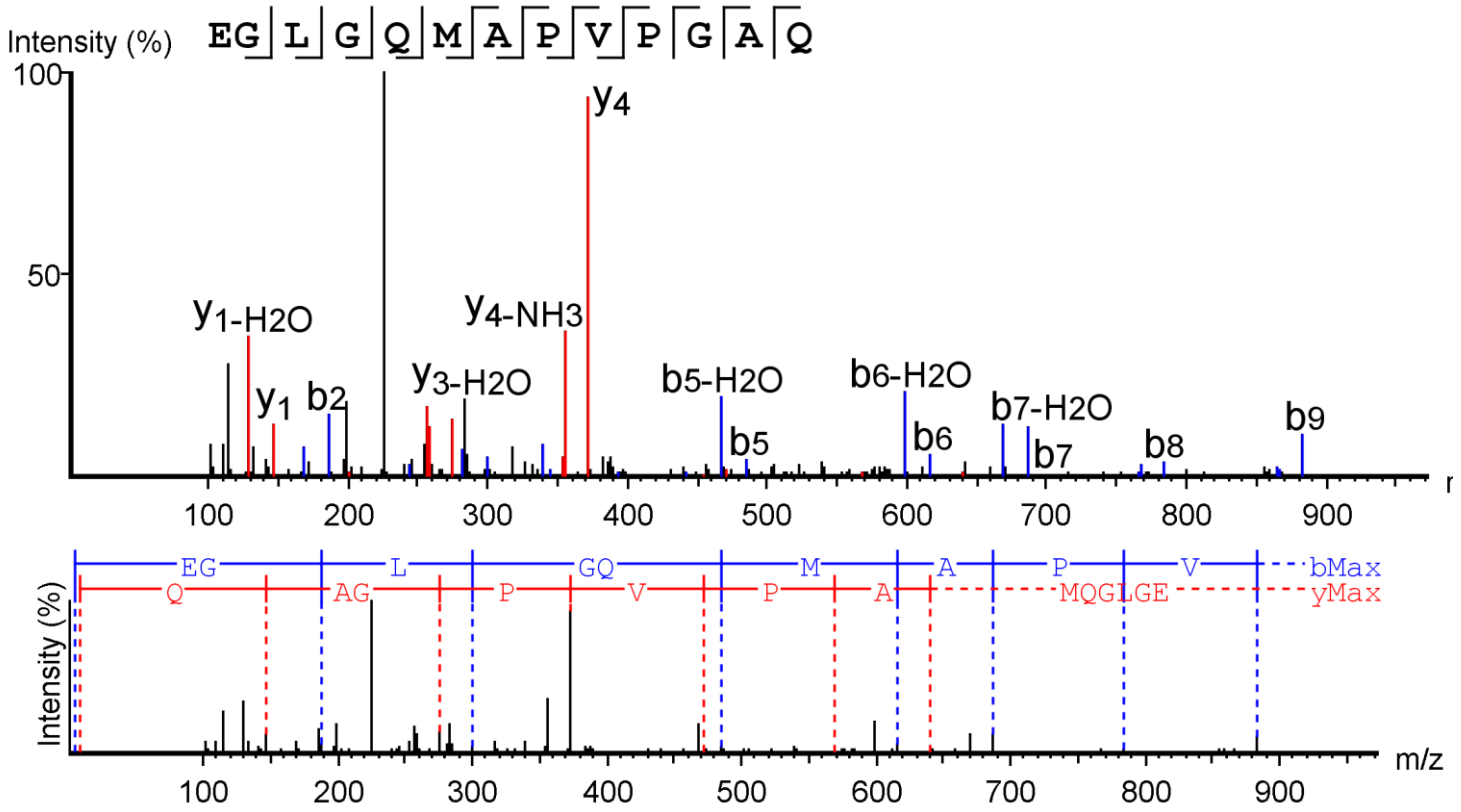
Supplementary data C. *De novo* sequence analysis of the processed MS/MS spectra of *M. lehilahytsara* tryptic peptide 1888 m/z.

M. lehilahytsara urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



Supplementary data C. ***De novo*** **sequence analysis of the processed MS/MS spectra of** ***M. lehilahytsara*** **LysC peptide 1888 m/z.**

M. lehilahytsara urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).



Supplementary data C. ***De novo*** **sequence analysis of the processed MS/MS spectra of** ***M. lehilahytsara*** **AspN peptide 1125 m/z.**

M. lehilahytsara urine containing the protein of interest was digested using the in-solution digest protocol listed in the methods section. Peptides from the in-solution proteolysis were analysed using a Thermo Scientific QExactive mass spectrometer coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 nano chromatography system. The samples were injected (typically equivalent to 500 fmol protein) onto a reversed phase column and were eluted over a 1 h acetonitrile gradient . Spectra were acquired between 300-2000m/z. Raw data was processed using PEAKS 6 ®software (Bioinformatics Solutions Inc, Canada).