THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Development of a novel encyclopaedic peptide spectral library using the liquid fraction of sheep blood

Saul Chemonges

Bachelor of Veterinary Medicine

Master of Veterinary Studies

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2017

School of Veterinary Science

Abstract

Proteome measurements derived from tandem mass spectrometry (MS/MS) are a powerful means to detect and quantify known and novel proteins. This methods-based thesis has developed a novel peptide spectral library (PSL) that can be applied to detect multiple proteins within complex MS/MS datasets, including protein markers of inflammation in ovine plasma. Such biomarkers permit better definition of the mammalian response to disease.

The potential use of the PSL is in targeted proteomics applications such as sequential window acquisition of all theoretical fragment ion mass spectra (SWATH)-MS analysis and other targeted proteomics techniques requiring a spectral library for quantitation of proteins found in sheep plasma or serum. A major advantage of SWATH-MS analysis of proteins is that the current alternative, the enzyme-linked immunosorbent assays (ELISA), can only measure a single protein for each kit, has to be validated for each species and is costly and cumbersome.

In light of the field's current status, this thesis sought to address four key objectives: (1) develop a comprehensive method to characterise the ovine circulating acellular proteome; (2) investigate various protein fractionation techniques to enhance protein identification yields from plasma and serum samples; (3) optimise a bioinformatics workflow for constructing a PSL as a tool for identifying proteins and for future proteogenomics uses, and; (4) apply the PSL to samples obtained from healthy and ill sheep to identify candidate markers of inflammation.

To achieve the above objectives, a baseline PSL was established using serum samples from healthy sheep. This PSL was broadened using a range of protein fractionation techniques, including acetone precipitation, partial organic precipitation with acetonitrile (ACN), combinatorial peptide ligand library enrichment (ProteoMiner™, Bio-Rad), as well as off-gel isoelectric focussing. Sample fractions were processed to tryptic peptides and analysed using nano-liquid chromatography electrospray ionisation tandem mass spectrometry (nanoLC-ESI-MS/MS) on an Eksigent nanoLC coupled to a quadrupole time-of-flight (QqTOF) mass spectrometer (TripleTOF 5600+, SCIEX) operated in a data-dependent acquisition (DDA) mode. Data from diseased and archived endotoxin-treated experimental sheep samples, as well as *in silico* predicted and synthesised peptides of five proinflammatory cytokines, namely Interleukin 6 (IL-6), Interleukin 3 (IL-3), Interleukin 1a (IL-1α), Interleukin 1b (IL-1β) and tumour necrosis factor-alpha (TNF-α) were later processed and added to create the encyclopaedic PSL.

For all DDA experiments, MS/MS data were searched against an *Ovis aries* UniProtKB (Universal Protein Resource Consortium Knowledgebase) database using ProteinPilot™ Software (SCIEX)

primarily, and then secondarily using Mascot (Matrix Science) search engine to identify proteins. These protein identifications (IDs) were validated using PeptideShaker (CompOmics, Inc) proteomics informatics software.

Following establishment of the PSL, plasma samples from sheep, before and after endotoxin-treatment, were analysed by data independent acquisition (DIA), namely SWATH-MS. Data were analysed using the SWATH™ MicroApp 2.0 (SCIEX) alongside the PSL to profile proteins in plasma samples in the two cohorts.

The primary ProteinPilot™ search identified 41,288 distinct peptides from 3,195,890 spectra at a false discovery rate (FDR) threshold of 1%. Together with secondary analysis in Mascot and validation in PeptideShaker, these spectra enabled the identification of 398 proteins in the nascent PSL. Using PeptideShaker identification results from DDA experiments, the baseline PSL and the acetone precipitation experiments yielded 133 and 102 protein IDs, respectively. The ACN precipitation workflow resulted in 198 protein IDs. The ProteoMiner™ workflow yielded 305 protein IDs, representing 56.4% increase in protein IDs, compared to undepleted samples. The off-gel experiment yielded 70 protein IDs, 15 (21.4%) of which were attributed to fractionation compared to 55 protein IDs from crude serum. The diseased and endotoxin-treated sheep experiments yielded 183 and 84 protein IDs, respectively, and collectively contributed 80 protein IDs not detected in healthy sheep samples. Validated spectrum annotation matches were obtained from two peptides each of IL-6, IL-3, IL-1α, IL-1β and TNF-α in the PSL from the *in silico* predicted and synthesised proinflammatory cytokines experiment.

The use of SWATH analysis enabled the quantitation of 243 proteins in sheep plasma and also revealed that the samples of the sheep model were non-identical between individuals and yet they were expected to be alike. Forty well-recognised acute phase proteins (APP) were quantitated and 42 other proteins were potentially endotoxin-induced candidate markers of inflammation.

This research has developed the first PSL resource designed for measuring the proteinaceous portion of blood from healthy and diseased sheep. Through its application, the first use of SWATH-MS is reported herein to identify candidate protein markers of inflammation in sheep's plasma. Through this work, the feasibility of simultaneously identifying many protein alterations in a cost-effective manner and with widely available tools has been reinforced, with potential applications in veterinary pathology, animal welfare and in screening laboratory animals before inclusion in experimental groups to minimise differences. In future, the scope of the PSL can be broadened by including proteomic and genomic data from cellular components of blood and other tissues to complement ovine genome annotation efforts.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Peer-reviewed papers

Saul Chemonges. Effect of Intermittent Positive Pressure Ventilation on Depth of Anaesthesia during and after Isoflurane Anaesthesia in Sulphur-Crested Cockatoos *(Cacatua galerita galerita)*. *Veterinary Medicine International*, vol. 2014, Article ID 250523, 7 pages, 2014. doi:10.1155/2014/250523.

David G Platts, Andrew Hilton, Sara Diab, Charles McDonald, Matthew Tunbridge, **Saul Chemonges**, Kimble R Dunster, Kiran Shekar, Darryl J Burstow and John F Fraser. A novel echocardiographic imaging technique, intracatheter echocardiography, to guide veno-venous extracorporeal membrane oxygenation cannulae placement in a validated ovine model. *Intensive Care Medicine Experimental* 2014, 2:2 doi:10.1186/2197-425X-2-2.

David G. Platts, Sara Diab, Kimble R. Dunster, Kiran Shekar, Darryl J. Burstow, Beatrice Sim, Matthew Tunbridge, Charles McDonald, **Saul Chemonges**, Jonathan Chan, John F. Fraser (2014). Feasibility of Perflutren Microsphere Contrast Transthoracic Echocardiography in the Visualisation of Ventricular Endocardium during Venovenous Extracorporeal Membrane Oxygenation in a Validated Ovine Model. *Echocardiography*, 32 3: 548-556. doi:10.1111/echo.12695.

Saul Chemonges, John-Paul Tung and John F Fraser. Proteogenomics of selective susceptibility to endotoxin using circulating acute phase biomarkers and bioassay development in sheep: a review. *Proteome Science* 2014, 12:12 doi:10.1186/1477-5956-12-12.

Chemonges S, Shekar K, Tung J-P, Dunster K, Diab S, Platts DG, Watts RP, Gregory SD, Foley S, Simonova G, McDonald C, Hayes R, Bellpart J, Timms D, Chew MS, Fung YL, Toon M, Maybauer MO, Fraser JF: Optimal management of the critically ill: Anaesthesia, monitoring, data capture and point-of-care technological practices in ovine models of critical care. *BioMed Research International*, vol. 2014, Article ID 468309, 17 pages, 2014. doi:10.1155/2014/468309.

Chemonges, Saul. Suspected selective susceptibility to endotoxin in an ovine model. *Online Journal of Veterinary Research*, 2014, 18 12: 941-963.

Chemonges, Saul. The recognition of LpxC inhibitors as potential antibiotics could revolutionise the management of sepsis in veterinary patients if their unknown biological properties are widely evaluated in suitable animal models. *International Journal of Veterinary Science and Medicine*, 2 2: 99-102. doi:10.1016/j.ijvsm.2014.10.003.

Chemonges, Saul. Profiles of resistance to breathing in spontaneously ventilating anaesthetised

dogs attached to an Ayre's T-piece. *Online Journal of Veterinary Research,* 2015, 19 1: 15-25.

Chemonges S. Learning from critical care management of sheep receiving extra-corporeal membrane oxygenation for smoke-induced acute lung injury as a tool for processing large clinical datasets. *bioRxiv 2016,* bioRxiv 058511; doi: https://doi.org/10.1101/058511.

Chemonges S, Gupta R, Mills PC, Kopp SR, Sadowski P: Characterisation of the circulating acellular proteome of healthy sheep using LC-MS/MS-based proteomics analysis of serum. *Proteome Science* 2017, 15:11. doi:10.1186/s12953-017-0119-z.

Books

Chemonges Saul. Resistance to breathing in anaesthetic breathing systems: a clinical perspective of understanding veterinary anaesthetic breathing circuits using dogs. Saarbrücken, Germany: Lambert Academic Publishing, 2012.

Full conference papers

Saul Chemonges. *Critical care management of sheep receiving extra-corporeal membrane oxygenation due to smoke induced acute lung injury (ECMO S-ALI) and acute sepsis.* Annual Science Week Conference of the Australian & New Zealand College of Veterinary Scientists, Chapter of Anaesthesia, Emergency and Critical Care. QT Gold Coast, Surfers Paradise, 11th July 2013. ***espace****.library.uq.edu.au/view/UQ:328112*

Saul Chemonges. *Contemporary data capture, anaesthesia monitoring and point-of-care Technology in critical care research settings for animal models.* Annual Science Week Conference of the Australian & New Zealand College of Veterinary Scientists, Chapter of Anaesthesia, Emergency and Critical Care. QT Gold Coast, Surfers Paradise, 12 July 2013. ***espace****.library.uq.edu.au/view/UQ:328111*

Saul Chemonges, Rajesh Gupta, Paul Mills, Steven Kopp and Pawel Sadowski. Development of a novel encyclopaedic peptide spectral library using the liquid fraction of sheep blood. Abstract and oral presentation, *Queensland Mass Spectrometry Symposium,* Brisbane, Australia, 15-16 August 2016. ***espace.****library.uq.edu.au/view/UQ:406777*

Conference abstracts

Margaret Passmore, Lin Fung, Kimble Dunster, **Saul Chemonges**, Sara Diab, Robyn Minchinton, Kiran Shekar, John Fraser. *ECMO contributes to lung injury in an ovine model of veno-venous ECMO*. Poster, ANZICS/ACCCN Intensive Care Annual Scientific Meeting (ASM), Hotel Grand Chancellor Hobart, TAS, 17-19 October 2013.

Kiran Shekar, Jason Roberts, Sara Diab, Kimble Dunster, Charles McDonald, **Saul Chemonges**, Gabriela Simonova, Sam Foley, Steven Wallis, David Platts, Lin Fung, Maree Smith, John Fraser. *Extracoproreal membrane oxygenation (ECMO) has more profound influence on ciprofloxacin pharmacokinetics in critically ill sheep when compared with healthy sheep.* Poster, 1st International Conference of the Asia-Pacific Chapter of the Extracorporeal Life Support Organization (APELSO) 2013, 11 - 13 Oct 2013, P4. http://f1000.com/posters/browse/summary/1094577

Margaret Passmore, Lin Fung, Kimble Dunster, **Saul Chemonges**, Sara Diab, Robyn Minchinton, Kiran Shekar, John Fraser. *ECMO contributes to lung injury in an ovine model of veno-venous ECMO.* Poster, 1st International Conference of the Asia-Pacific Chapter of the Extracorporeal Life Support Organization (APELSO) 2013, 11 - 13 Oct 2013.

Chemonges S, Manning M, Mohammed A, Dunster K, DiabS, Choudhary J, Fraser JF. *Blockade of complement mediated inflammation may improve lung transplantation outcomes*. The ANZICS/ACCCN Intensive Care Annual Scientific Meeting (ASM), Hotel Grand Chancellor Hobart, TAS, 17-19 October 2013. http://f1000.com/posters/browse/summary/1094666

Passmore M, Fung YL, Dunster K, **Chemonges S**, Diab S, Minchinton RM, Fraser JF. *ECMO contributes to lung injury in an ovine model of veno venous ECMO.* The ANZICS/ACCCN Intensive Care Annual Scientific Meeting (ASM), Hotel Grand Chancellor Hobart, TAS, 17-19 October 2013.

David Platts, Sara Diab, Charles McDonald, Matthew Tunbridge, **Saul Chemonges**, Kimble Dunster, Kiran Shekar, Darryl Burstow, Daniel Mullany, John Fraser. *A novel echocardiographic imaging technique to guide VV ECMO cannulae placement in a validated ovine model.* O9, P1 Poster, EURO ELSO MEETING 8-11 May, 2013.

David Platts, Sara Diab, Charles McDonald, Matthew Tunbridge, **Saul Chemonges**, Kimble Dunster, Kiran Shekar, Darryl Burstow, Daniel Mullany, John Fraser. *Venovenous ECMO return cannulae continuous high flow and its impact on tricuspid valve geometry and function.* O9, P1 Poster, EURO ELSO 8-11 May, 2013.

David Platts, Andrew Hilton, Sara Diab, Charles McDonald, Matthew Tunbridge, **Saul Chemonges**, Kimble Dunster, Kiran Shekar, Darryl Burstow, John Fraser. *Feasibility of a Novel Echocardiographic Imaging Technique, Intracatheter Echocardiography, to Guide Venovenous Extracorporeal Membrane Oxygenation Cannulae Placement in a Validated Ovine Model.* P2-137 Moderated Poster. 24th Annual Scientic Meeting of the American Society of Echocardiography, ASE June 29- July 2, 2013, Minneapolis Convention Center, MN.

David Platts, Sara Diab, Charles McDonald, Matthew Tunbridge, **Saul Chemonges**, Kimble Dunster, Kiran Shekar, Darryl Burstow, Daniel Mullany, John Fraser. *The Impact of Continuous Flow from Venovenous Extracorporeal Membrane Oxygenation Cannulae on Tricuspid Valve Geometry and Function.* Abstract P2-142, 24th Annual Scientic Meeting of the American Society of Echocardiography, ASE June 29- July 2, 2013, Minneapolis Convention Center, MN.

Chemonges Saul. *Proteogenomic characterisation of circulating acute phase markers and bioassay development in sheep.* In: 2014 Diamantina Health Partners 2nd Annual Forum. Pre-Forum Symposium: Research and Innovation Ideas Awards, Woolloongabba, QLD, Australia, 30 April-2 May, 2014. ***espace****.library.uq.edu.au/view/UQ:329488*

Chemonges Saul and Sadowski Pawel. *Using Skyline to support peptide MRM workflows on Shimadzu LCMS-8050*. In: 62nd American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics, Baltimore, MD, United States, 15-19 June 2014. ***espace****.library.uq.edu.au/view/UQ:347312*

Chemonges Saul, Mills Paul, Kopp Steven and Sadowski Pawel. *Towards a comprehensive targeted proteogenomic assay repository for the liquid fraction of sheep blood.* In: 25th Conference of the ANZSMS and the 6th Conference of the Asia and Oceania Society for Mass Spectrometry Conference (AOMSC), Brisbane, QLD, Australia, 19-22 July 2015.

Chemonges Saul, Gupta Rajesh, Mills Paul, Kopp Steven and Sadowski Pawel. *SWATH enabled distinction of 'proteophenotype' in a sheep model of intensive care.* In: Queensland Mass Spectrometry Symposium, Brisbane, Australia, 15-16 August 2016.

Saul Chemonges, Rajesh Gupta, Paul Mills, Pawel Sadowski and Steven Kopp. *Development of a novel encyclopaedic peptide spectral library using the liquid fraction of sheep blood.* In: Animal Science Olympics at QAAFI, Queensland Bioscience Precinct, The University of Queensland, St. Lucia, Brisbane, Australia, 18 November 2016.

Publications included in this thesis

1. Publication citation – incorporated as Chapters 3 and 4.

Chemonges S, Gupta R, Mills PC, Kopp SR, Sadowski P: Characterisation of the circulating acellular proteome of healthy sheep using LC-MS/MS-based proteomics analysis of serum. *Proteome Science* 2017, 15:11. doi:10.1186/s12953-017-0119-z.

Contributor	Statement of contribution
Chemonges S (Candidate)	Concept and design of the project (90%) Analysis and interpretation of the research data (70%) Drafting and production of significant parts (70%)
Gupta R	Concept and design of the project (0%) Analysis and interpretation of the research data (5%) Drafting and production of significant parts (0%)
Mills PC	Concept and design of the project (0%) Analysis and interpretation of the research data (0%) Drafting and production of significant parts (20%)
Kopp SR	Concept and design of the project (0%) Analysis and interpretation of the research data (0%) Drafting and production of significant parts (10%)
Sadowski P	Concept and design of the project (10%) Analysis and interpretation of the research data (25%) Drafting and production of significant parts (0%)

2. Publication citation – incorporated as APPENDIX 7.0.

Chemonges, Saul. Suspected selective susceptibility to endotoxin in an ovine model. *Online Journal of Veterinary Research,* 2014, 18 12: 941-963.

Contributions by others to the thesis

I was responsible for the concept, design and all experimental data collection of the project, with advisory support and mentorship from Dr Pawel Sadowski. Prof Paul Mills and Dr Steven Kopp provided assistance with the thesis structure.

Non-routine technical work

Mr Vincent Chand assisted in experimental workstation setup processes at the Molecular Genetics Research Facility (MGRF) at Queensland University of Technology (QUT). Mr Rajesh Gupta and Dr Pawel Sadowski assisted in troubleshooting the mass spectrometry instrument and data acquisition at the Proteomics and Small Molecule Mass Spectrometry, Central Analytical Research Facility, QUT.

Analysis and interpretation of research data

I was responsible for data analysis and interpretation of the research data in this thesis.

Drafting significant parts of the work or critically revising it so as to contribute to the interpretation

As principal author, I was responsible for the drafting and writing of the chapters, however advisory input was made by Dr Pawel Sadowski, Prof Paul Mills and Dr Steven Kopp.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

Gratitude is extended to Dr Steven Kopp for kindly accepting to take a leading advisory role for my PhD project and also for providing guidance during my candidature at the School of Veterinary Science, The University of Queensland. I would like to thank Prof Paul Mills for the willingness to be part of my research advisory team and for being my teacher. This research project would not have materialised without the insurmountable contribution, supervisory support, collaboration and expert mentorship of Dr Pawel Sadowski of Queensland University of Technology. In their various capacities, truly helpful people especially the team at the Central Analytical Research Facility, Queensland University of Technology, including Mr Vincent Chand, Dr Kevin Dudley, Dr Rajesh Gupta, Ms Melanie Fitzgerald, Ms Natalie Sukic, Ms Sanjleena Singh and Prof Stephen Blanksby are thanked for enabling me to achieve the goals of my research. I would like to extend my gratitude to Dr John-Paul Tung for kindly allowing me to be part of the collaborative work in my background studies on endotoxin and sepsis at the School of Medicine (SOM), and for kindly facilitating access to archived sheep blood samples at The Australian Red Cross Blood Service. Special thanks are extended to Mr Kimble Dunster and Ms Sara Diab for the technical support in the animal laboratory early in my candidature. I would like to thank all Research High Degree administrative staff at the School of Veterinary Science specifically Ms Annette Winter and Mr Brian Bynon, Prof Judith Greer of SOM and staff of the Graduate School, The University of Queensland for all the help they have accorded me during my time as a student. Special thanks are extended to Prof Michelle Chew and Dr James Broadbent for providing critically valuable feedback during the final stages of writing of the thesis, and to Dr Malini Devadas for editorial assistance.

This research could not have proceeded without financial support from the Australian Government by way of a Postgraduate Award Scholarship through The University of Queensland, advisor research funds and Queensland University of Technology for providing around-the-clock access to their facilities and human resources. I would like to express my gratitude and to say thank you to Linda – the donor of Dr H. George Osborne Research Scholarship through UQ's School of Veterinary Science for kindly awarding me this scholarship that enabled me to complete working on an important aspect this project on a fulltime basis.

Finally, I would like to thank my family, my Wife Marylyn and our children who have tirelessly supported and continue to get productively involved in my work.

Keywords

peptide spectral library, ovine circulating acellular proteome, nanolc-nanoesi-ms/ms, gene ontology, protein pathway analysis, veterinary proteomics, proteogenomics, sequential window acquisition of all theoretical fragment ion spectra mass spectrometry (swath-ms), acute inflammation biomarkers, bioinformatics.

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 060408 Genomics, 50%
ANZSRC code: 060102, Bioinformatics, 25%
ANZSRC code: 070703, Veterinary Diagnosis and Diagnostics, 25%

Fields of Research (FoR) Classification

FoR code: 0707, Veterinary Sciences, 45%
FoR code: 0601, Biochemistry and Cell Biology, 45%
FoR code: 0699, Other Biological Sciences, 10%

Table of Contents

List of Figures & Tables

Figures

Tables

List of Abbreviations used in the thesis

>sp –protein header in Swiss-Prot database
>tr – header for translated nucleotides of the European Molecular Biology Laboratory
°C – Degrees Celsius
µg – microgram
µL – microlitre
µm – micrometre
1D SDS-PAGE – one-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis
20150624_SC_SWATH-MS_ – prefix identifying SWATH data files
2D-DIGE – two-dimensional difference gel electrophoresis
2DE – two-dimensional gel electrophoresis
Å – ångström
A-a – alveolar-arterial oxygen gradient
Ac – acetone precipitated
ACN – acetonitrile
ACN_Gel_Plasma – proteins identified from in-gel digested plasma following acetonitrile precipitation
ACN_Gel_Serum – proteins identified from in-gel digested serum following acetonitrile precipitation
ACN_Sol_Plasma – proteins identified from in-solution digested plasma following acetonitrile precipitation
ACN_Sol_Serum – proteins identified from in-solution digested serum following acetonitrile precipitation
AEC – anion exchange chromatography

AFP – antibody-free depletion
AIMS – accurate inclusion mass screening
ANOVA – analysis of variance
AP – adaptor protein
APA – Australian Postgraduate Award
APP – acetone-precipitated plasma
APPs – acute phase proteins
APS – acetone precipitated serum
AQUA – absolute quantitation using heavy/light isotope ratios against spiked-in heavy labelled peptides
ARCBS – Australian Red Cross Blood Service
BCA – bicinchoninic acid
BiNGO – Biological Networks Gene Ontology
BLAST – Basic Local Alignment Search Tool
BLP – plasma samples drawn before endotoxin treatment
BLS – serum samples drawn before endotoxin treatment
BRF – Biological Research Facility
BSA – bovine serum albumin
BVetMed – Bachelor of Veterinary Medicine
C18 - octadecyl carbon chain-bonded silica
CA – California
CARF – Central Analytical Research Facility
CCO – continuous cardiac output
CD 14 - cluster of differentiation antigen 14
CE – capillary electrophoresis
cGMP - cyclic guanosine monophosphate
CI – cardiac index
CID – collision-induced dissociation
CLA – caseous lymphadenitis

CP – crude plasma

cRAP – common Repository of Adventitious Proteins

CRAPome – Contaminant Repository for Affinity Purification

CS – crude serum

CSIRO – Commonwealth Scientific and Industrial Research Organisation

CSIROS – sheep from the Commonwealth Scientific and Industrial Research Organisation

CSV – comma separated values

C-terminus – carboxyl terminus

Cu-Zn – copper-zinc

dat – extension for generic data file format

DAVID – Database for Annotation, Visualisation and Integrated Discovery

DDA – data-dependent acquisition

DIA – data independent acquisition

DNA – deoxyribonucleic acid

DOI – digital object identifier

D-Score – Search engine-independent score for localising peptide post-translational sites

DTT – DL-Dithiothreitol

EDTA – ethylenediaminetetraacetate

EJV – external jugular vein

ELISA – enzyme-linked immunosorbent assay

EM – electron multiplier

EMBL-EBI – European Molecular Biology Laboratory - European Bioinformatics Institute

ESI – electro-spray ionisation

ESI-QUAD-TOF – electrospray ionisation quadrupole time-of-flight

EZ-Run™ – Protein gel staining solution

FA – formic acid

FASP - filter-aided sample preparation technology

FASTA – text-based format for representing peptide, amino acid or protein sequences.

FDR – false discovery rate

fm, fmol – femtomole

G24PE00 – manufacturer's 24-well protocol for the Agilent 3100 OFFGEL fractionator

GBB – gel loading buffer

GE – General Electric Company

GeLC – sodium dodecyl sulfate polyacrylamide gel electrophoresis

GeLC-MS/MS – sodium dodecyl sulfate polyacrylamide gel electrophoresis followed by tandem mass spectrometry

GN – gene name

GO – Gene Ontology

H – hydrogen

HAP – high abundance protein

HAS – human serum albumin

HILIC – hydrophilic interaction liquid chromatography

HPLC - high performance liquid chromatography

HPPP – Human Plasma Proteome Project

HR – heart rate

HRM - high-resolution multiple reaction monitoring

IAM – iodoacetamide

ICP-MS – inductively-coupled plasma mass spectrometry

ICU – intensive care unit

ID – identification	MAP – mean pulmonary artery pressure
IDMS – isotope dilution mass spectrometry	MCP – microchannel plate
IEF – isoelectric focussing	mg – milligram
IgG – immunoglobulin G	mgf – Mascot generic format
IHBI – Institute of Health and Biomedical Innovation	MGRF – Molecular Genetics Research Facility
	MHC – major histocompatibility complex
IL-1α – Interleukin 1 alpha	miRNAs – micro ribonucleic acids
IL-1β – Interleukin 1 beta	mL – millilitre
IL-3 – Interleukin 3	mm – millimetre
IL-6 – Interleukin 6	mM – millimolar
In-sol – in-solution	MPAP – mean pulmonary artery pressure
IPAS – intact protein analysis system	MRM – multiple reaction monitoring
IT – ion-trap	MS – mass spectrometry
iTRAQ – isobaric tags for relative and absolute quantitation	MS/MS – tandem mass spectrometry
	MudPIT – multi-dimensional protein identification technology
IV – intravenous	MVSt – Master of Veterinary Studies
K [13C6: 15N2] – heavy-labelled lysine (K)-containing ^{13}C and ^{15}N atoms	MWt – molecular weight
kg – kilogram	mzid – extension for mzIdentML format files
LAP – low abundance protein	mzIdentML – data format standard for protein/peptide identifications derived from MS-based proteomics approaches
LC – liquid chromatography	
LC-MS/MS – liquid chromatography tandem mass spectrometry	nanoLC-ESI-MS/MS - nano-liquid chromatography electrospray ionisation tandem mass spectrometry
LC-SWATH-MS – liquid chromatography sequential window acquisition of all theoretical fragment ion spectra mass spectrometry	
LPS – lipopolysaccharide	nanoLC-nanoESI-MS/MS – nano liquid chromatography nano electrospray ionisation tandem mass spectrometry
LTQ-MS – linear ion trap quadrupole mass spectrometry	
m/z – mass-to-charge ratio	NCBI – National Center for Biotechnology Information
MALDI-TOF-MS – matrix assisted laser desorption/ionisation time-of-flight mass spectrometry	NH_4HCO_3 – ammonium bicarbonate
	ns – not significant
	NSW – New South Wales

N-terminus – amino terminus
OH – hydroxyl
OS – organism or species name
P qty – protein quantity
P1P – plasma samples drawn after endotoxin treatment
P1S – serum samples drawn before endotoxin treatment
PANTHER – Protein ANalysis THrough Evolutionary Relationships
PaO_2/FiO_2 – ratio of partial pressure of arterial oxygen and fraction of inspired oxygen
PC – Protein cleavage
PCA – principal component analysis
PCA-DA – principal component analysis and discriminant analysis
PE – evidence of protein existence
PGDipVetClinSt – Postgraduate diploma in Veterinary Clinical Studies
pH – acidity or alkalinity of a solution
PhosphoRS score – probability value for a phosphorylated site of a peptide based on the given MS/MS data
Pi – protease inhibitor
PL – production line
PPIAse - peptidyl-prolyl cis-trans isomerase
PPT – precipitate
PPTion – precipitation
PRIDE – proteomics identifications database
PS – supernatant of plasma after acetone precipitation
PSAQ – protein standard absolute quantitation
PSL – peptide spectral library
PSMs – peptide spectrum matches
PTM – Post-transcriptional modification
PXD – ProteomeXchange dataset
QCAT – concatemer of Q peptides
QLD – Queensland
QqQ – triple quadrupole
QTOF, Q-TOF or QqTOF – quadrupole time-of-flight
Quad HD - four times the definition of standard high definition
QUT – Queensland University of Technology
QUT-MERF – Medical Engineering Research Facility of the Queensland University of Technology
RF – radio frequency
RL – resistant line
ROC – receiver operating characteristic
RP – reverse phase
RP-HPLC – reverse phase high performance liquid chromatography
RT – retention time or room temperature, where appropriate
Rx – treatment
SA – Serum Australis
SC – Saul Chemonges investigator initials for protein sample identification
SCX – strong cation exchange
SD – standard deviation
SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis
SELDI-TOF-MS – surface-enhanced laser desorption ionisation time-of-flight mass spectrometry

SID-LC-MRM-MS – stable isotope dilution multiple reaction monitoring mass spectrometry
SILAC – stable isotope labelling with amino acids in cell culture
SISCAPA – stable isotope standards and capture by anti-peptide antibodies
SOM – School of Medicine, The University of Queensland
SP – serum protein
SRM – selected reaction monitoring
SS – supernatant of serum after acetone precipitation
SUP – supernatant
SV – protein sequence version
SvO_2 – mixed venous oxygen saturation
SVS – School of Veterinary Science, The University of Queensland
SWATH – sequential widow acquisition of all theoretical fragment ion mass spectra
SYSS – saleyard or stockyard-sourced sheep
TAILS – terminal amine isotopic labelling of substrates
TEMED – Tetramethylethylenediamine
TFA – trifluoroacetic acid
TGX™ – Tris-Glycine eXtended precast gels
TIC – total ion chromatogram or current
TIVA – total intravenous anaesthesia
TLC – thin layer chromatography
TMTs – tandem mass tags
TNF-α – tumour necrosis factor-alpha

TOF– time-of-flight
TPP – Trans-Proteomic Pipeline
TPR – temperature, pulse, respiratory rate
TRALI – transfusion-related acute lung injury
trEMBL – translated predicted nucleotides of the European Molecular Biology Laboratory
TripleTOF 5600+ – SCIEX mass spectrometer (instrument) used for sample analysis
Tris – Trisaminomethane
UHPLC-UV – ultra-high performance liquid chromatography coupled with an ultraviolet detector
UK – United Kingdom
UniProtKB – Universal Protein Resource Consortium Knowledgebase
UQ – The University of Queensland
UQ.O – prefix for ex-diagnostic sick sheep serum samples from School of Veterinary Science, The University of Queensland
URL – uniform resource locator
v:v % – volume per volume percent
VH – immunoglobulin heavy chain variable
VIB-UGent - life sciences research institute, University of Gent, Belgium
w/f – workflow
WDR – wide dynamic range
wiff – raw instrument data file format of SCIEX instruments
wiff.scan - raw instrument scan data file format of SCIEX instruments
XIC – extracted ion chromatogram

CHAPTER 1

1.0 Introduction to the thesis

This thesis concerns the development of a platform capable of detecting high and medium abundant proteins and their alterations during illness in the liquid fraction of sheep blood using next-generation tandem mass spectrometry (MS/MS). It describes the construction of a peptide spectral library (PSL) for subsequent application in the sequential window acquisition of all theoretical fragment mass spectra (SWATH) mass spectrometry (MS)(SWATH-MS) technique[1] and proteomics pipeline development for identifying and quantitating proteins reproducibly across samples[1-5]. The proteomics information derived from MS/MS is a powerful tool that can be used to profile known and novel proteins[6-8]. This approach is an attractive alternative to antibody enzyme-linked immunosorbent assay (ELISA) technology, which typically requires a specific kit for each protein of interest and can, as a result, be particularly costly, time consuming and often need to be validated for each different species[9]. In the case of many biological phenomena including inflammation, it is important to provide the investigator with a broader tool to measure the plasma or serum levels of a panel of different proteins. Given the limited number of straightforward methods to simultaneously identify many plasma, serum or even lymph proteins in large domestic animals, the possibility of a robust and cost-effective proteomics approach to do so would be highly advantageous.

Sheep are an important model for biomedical research because their cardiovascular physiology closely matches that of humans[10-12]. It is therefore important to have a better understanding of the proteome in sheep and the proteomic response to changed physiological status to more effectively use this model to study human disease conditions. While there are many parallels between the pathophysiology of acute injuries in humans and sheep, the majority of the research is primarily human-focussed. Moreover, the studies that have used sheep as their models of injury have been sporadic, often with scanty, hastily documented or incomplete information on the detail of the evolution of the model with regard to the mediators of the underlying injury. For example, this thesis was specifically motivated by a review study that showed a lack of consensus on how physiological data was extracted from injured sheep used for scientific purposes in a large number of studies[10], and equivocal and subjective observations emanating from a model of smoke injury in sheep[13] and a model of blood transfusion in sheep with induced endotoxaemia[14]. The incremental aspect of the latter study was futile as subsequent observations suggested a noisy background of preventable varying peri-experimental practices that were previously unaccounted for in the prototypical stages of the original study[15]. The preceding previous studies measured a range of parameters, including some proteins, but this was variable and therefore difficult to compare. The

proteomic approach measures all the proteins and their respective changes during disease, so can better assess and quantify the disease process. It was therefore deemed necessary to provide a robust molecular-level insight into understanding and quantitating the salient markers of injury in sheep. To address these concerns, it was proposed that learning from blood samples derived from injured or sick sheep could form the basis for the development of assays for acute phase proteins (APP) or other disease-associated proteins in the future, on the premise that every injury is accompanied by proteomic alterations in specific biomarkers. As a foundation, a holistic approach was conceived to advance a proteomics method for exploring the liquid fraction of sheep blood, from animals in good health and during acute illness.

The four key objectives of this research project were to: (1) develop a robust and comprehensive method to characterise the circulating acellular proteome in ovine serum using nano liquid chromatography nano electrospray ionisation tandem mass spectrometry (nanoLC-nanoESI-MS/MS)[16], (2) investigate and use various protein fractionation techniques and sample types to enhance protein identification yields from plasma and serum samples, (3) optimise a bioinformatics workflow for constructing a PSL as tool for identifying proteins and for future proteogenomics uses and, (4) develop an approach that can be used to diagnose proteomic changes in plasma indicative of disease in sheep.

The availability of targeted protein data extraction using SWATH-MS analysis was identified as an enabling platform capable of potentially characterising APP in the liquid fraction of sheep blood. As a capstone study, SWATH-MS analysis was used alongside a newly constructed PSL to interrogate archived plasma samples of sheep in which the subjects were injected with endotoxin from *E. coli* to simulate an acute infection[17] and determine if this approach can identify proteins and their alterations during early-phase acute inflammation.

The primary output of this thesis is summarised in Figure 1.0. Two data acquisition strategies were used throughout – data dependent acquisition (DDA) for constructing the PSL and data independent acquisition (DIA) for SWATH-MS[1] biological sample comparison. The SWATH Acquisition MicroApp (SCIEX) was utilised to extract the acquired SWATH-MS data for APP quantitation. Considering that DDA is inherently analyte abundance-biased by only selecting the most intense ions for fragmentation[18], it was necessary to have a set of experiments that captured peptide data derived from various fractionation approaches during PSL construction to ensure comprehensive proteome coverage. On the other hand, SWATH-MS is a 'fragment all' technique which quantitates proteins in an unbiased manner that was employed for validation of the PSL[5]. Simply put, DDA is the traditional shotgun discovery data acquisition, whilst SWATH-MS is a more recent strategy. This thesis utilised both of these strategies.

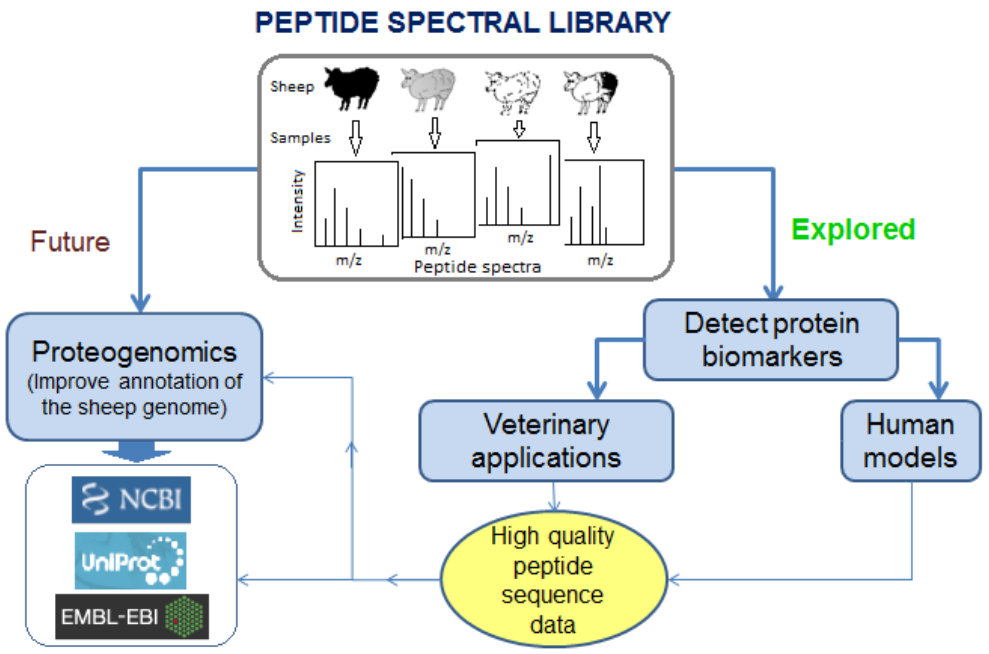


Figure 1.0. The main output of this thesis is a high-quality peptide spectral library (PSL). Thousands of peptide spectral data were acquired using shotgun proteomic analysis of plasma and serum protein samples obtained from several healthy and sick sheep to generate the PSL for archiving in the proteomics identifications (PRIDE) database[19] at the European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI), inclusion in the National Center for Biotechnology Information (NCBI), and the Universal Protein Resource Consortium (UniProt)[20] for annotating the sheep genome. **Key:** *m/z* = mass to charge ratio of individual peptides.

The thesis consists of eight chapters, including this Introduction. The literature review is presented in Chapter 2. Here, the science of MS-based proteomics is introduced, including developments in the human biomedicine field for benchmarking purposes. This is followed by a discussion of the *status quo* of MS assays for the circulating acellular proteome with respect to veterinary species. Based on earlier work[17] and recent reports[16,21,22], there is a consensus that the ability to simultaneously assay many disease- and homeostasis-related proteins would be a welcome advance in diagnosing and monitoring animal diseases in clinical settings. Knowledge gaps were identified in the literature pertaining to exploring the ovine circulating acellular proteome, including the absence of reference MS-based data for sheep[16], proteogenomics tools such as PSL repositories for interrogating plasma and serum proteomes in a targeted manner, optimised bioinformatics workflows, and methods for identifying several candidate circulating APPs simultaneously from controlled experiments. The lack of optimised sample preparation methods, WDR issues and the high cost of running assays were recognised as major impediments in the widespread use of proteomics approaches in veterinary science.

Chapter 3 presents the generic materials and methods used in the proteomics studies applied in this thesis[16]. The generic protocol relies on acetone precipitation[23-26] as the starting fractionation technique for separating proteins from lipids in liquid samples followed by digesting proteins with trypsin into peptides and desalting using C18 material to remove salts. Desalting is a small-scale reverse phase high performance liquid chromatography (RP-HPLC) process whereby salts are eluted while the peptides are concentrated. The process also gets rid of particulate matter (for example, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel pieces). The end result of desalting is enhancement of protein identifications from peptide samples. This is followed by the provision of the details of nanoLC-nanoESI-MS/MS method employed for data acquisition. The chapter forms a common basis for the methodology of Chapter 4 and the modified methods included in the materials and methods of Chapter 5.

Chapter 4 presents a study prototype for characterisation of the circulating acellular proteome in healthy sheep, using generic methods (Chapter 3) applied to serum[16]. The serum proteome of sheep is relatively unknown and is often extrapolated from what is known in cattle. It is a readily available substrate for most experimental and clinical applications as evidenced in recent sheep studies that have applied mass spectrometry proteomics. For these reasons, it was considered necessary to characterise the serum proteome of sheep before using this tool in this species. This chapter therefore represents the feasibility of being able to identify a substantial number of proteins in the liquid fraction of sheep blood. The results of this study represent the baseline upon which alternative methods of sample preparation are based as described in Chapter 5.

Chapter 5 presents strategies for enriching the baseline protein data collected in Chapter 4. This was achieved by using various sample fractionation techniques and sample types to enhance protein identification yields. Protein extraction and identification processes, and comparative studies between serum and plasma proteomes are described in seven experiments covering: (a) a comprehensive analysis of fractions of acetone-precipitated sheep serum and plasma, which evaluates both fractions of acetone precipitation of proteins (this is also necessary in order to establish if there is any benefit associated with using either serum or plasma as the analytical sample for subsequent targeted data extraction); (b) a comprehensive analysis of fractions of partial organic precipitation of sheep serum and plasma proteins using acetonitrile in order to increase the acellular circulating proteome coverage via an innovative pseudo-depletion strategy of the most abundant plasma/serum proteins because ovine antibody-based protein depletion kits are still scarce; (c) the combinatorial peptide ligand library protein enrichment of sheep serum and plasma using a ProteoMiner™ Protein Enrichment kit (Bio-Rad), which equalises the protein abundance and hence increases the chance of identifying more proteins with MS; (d) the off-gel protein

fractionation (isoelectric focussing (IEF)) of serum proteins, which deals with protein sample separation using an Agilent 3100 OFFGEL Fractionator from Agilent Technologies before MS analysis; (e) the extraction of protein data from ex-diagnostic sheep serum; f) the derivation of protein data from serum and plasma samples from endotoxin-treated sheep that had been archived from a different study; and (g) *in silico* prediction of synthetic peptides of five selected proinflammatory cytokines of sheep used to assist in identifying corresponding endogenous cytokines and also as internal standards for quality during protein identification.

Chapter 6 presents an optimised bioinformatics strategy bringing together data from Chapters 4 and 5 to form an encyclopaedic PSL. In this chapter, alternative search algorithms alongside peptide sequence files are used for data mining to validate protein identifications. It furnishes the details of the absolute numbers of peptide spectra and proteins present in the PSL, thereby constituting the foundation for future proteogenomics studies on sheep.

Chapter 7 describes the application of the nascent PSL to determine if it can be used to diagnose proteomic changes indicative of disease in sheep. The PSL together with the SWATH Acquisition MicroApp (SWATH Pipeline) were applied to interrogate plasma samples from a sheep model of intensive care in which the experimental subjects were exposed to an endotoxin[14]. The aim of this clinical capstone study was to detect candidate protein inflammation biomarkers and their alterations in the plasma.

The general discussion, conclusions and future directions of the thesis are presented in Chapter 8, followed by the references and appendices. The discussion looks back over the project to highlight the successes and challenges by acknowledging that the work performed on sheep blood in this thesis is unique. The discussion recognises that the primary output of this method development thesis is the novel encyclopaedic PSL, comprising of high-quality annotated spectra for 398 proteins derived from the circulating acellular proteome of sick and healthy sheep that subsequent researchers may find useful. The work was guided by the goal of identifying protein biomarkers of early-phase acute inflammation because doing so helps in defining the predictors of mammalian response to illness – which translates into understanding resilience to disease. The reason why sheep were used was given as being invaluable production animals that not only contribute to the human food chain, but are also a source of natural wool and used for diverse cultural purposes, including the fulfilling of sacrificial requirements. The several difficulties and costs associated with traditional protein assay methods such as ELISA prompted the development of MS-based proteomic assays as an attractive and potentially viable alternative. In the absence of abundant protein depletion strategies for samples from majority veterinary species, the problem of the wide dynamic range (WDR) between low- and high-abundant proteins can be minimised by combining different

proteomics strategies, such as protein fractionation and chromatography[27]. As an upside, this nascent PSL was applied as a proof of concept capstone study – with promising results – using archived plasma samples from ill sheep with induced endotoxaemia as the model disease state.

The conclusion points out that this project has delivered the much-needed experimental methods benchmarked on studies in humans that can be used for detection of pathology in domestic animals. A vital attribute of the contribution of this work is the optimisation of serum and plasma sample preparation, which is now available for widespread use. This thesis is the first to have developed a novel encyclopaedic PSL for the ovine circulating acellular proteome with the capability of identifying a large number of proteins. The work pioneered the use of a SWATH pipeline on a large scale to interrogate plasma and serum samples of a sheep model of intensive care[14]. It made it possible to distinguish between samples from untreated and endotoxin treated sheep in order to identify hundreds of proteins, including quantitating their alterations in the circulating acellular proteome. This workflow could be improved by well-controlled sample collection preferably from identical subjects under uniform conditions to further enhance the applicability of the present findings for future science.

Included in the future perspectives is the need for a high-capacity trap column and a longer heated analytical column to enable the loading of higher amounts of protein to enhance deeper peptide analysis. Since this work is looking at new methods of protein detection, it will be necessary for future experiments to heavily lean on reproducibility, repeatability and validation experiments. There is a need to have several replicates from the samples of the workflows that were used in this thesis in order to potentially increase the proteome coverage of the PSL. The off-gel fractionation workflow needs to be further optimised and applied for sheep plasma samples as well. By using the already acquired data, the PSL can also be built using an alternative workflow such as the Trans-Proteomic Pipeline (TPP) that permits more control of false discovery rate (FDR) of peptides and proteins[5]. The main future applications of the PSL lies in proteogenomics that will require its broadening by adding genomic an proteomic data from other organs of sheep in order to be more inclusive. There is also need to investigate the effects of intravenous anaesthetic drugs on the extraction of proteins from plasma and serum samples as this was assumed to have been a contributing factor for the low protein identifications in samples obtained sheep anaesthetised by total intravenous anaesthesia (TIVA)[10].

CHAPTER 2

2.0 Literature review

2.1 Proteomic assay development for the circulating acellular proteome in veterinary species

The conventional approaches to analysing proteins in plasma or serum have largely been those that are non-specific and measure total protein content only, such as absorbance and Biuret Test-derived assays[28], or specific antibody-depended methods that detect the amount of a single protein[29-31]. Mass spectrometry-based proteomic assays are emerging as a promising approach that have the advantage of specificity and high throughput of protein analysis[32], but these assays, especially for veterinary species are still limited and need to be developed[16,33].

This review focusses on discovery proteomics (also known as bottom-up or shotgun proteomics), involving protein fractionation, protease digestion, MS/MS fragmentation of peptides and database search, founded on established principles, for analysing the undepleted circulating acellular proteome of non-model mammalian organisms[8,32,34-37]. Non-model organisms are defined as "organisms that have not been selected by the research community for extensive study either for historic reasons, or because they lack the features that make model organisms easy to investigate (e.g. they cannot grow in the laboratory, have a long life cycle, low fecundity or poor genetics)[38]." This definition of model/non-model organisms is disputable as indeed, every animal may make a good model for studies[39]. Mentioning sheep for example, besides its significance as a model for copper poisoning[40], there is also rich literature for it's use in surgery (neurology[41], orthopaedics[42] and cardiology[10]). The overall goal of this review is to provide the theoretical foundation necessary for building a PSL using plasma and serum of sheep to be applied in targeted proteomic data extraction[5] and for future proteogenomics studies. The review begins by outlining proteomics laboratory practices for acquiring reliable results. A discussion on the relevance and challenges of plasma and serum proteomics is provided. This is followed by an account on the difficulties associated with exploring the circulating acellular proteome, protein separation techniques, the basics of MS as an aid in detecting proteins, contemporary proteomics approaches, and bioinformatics strategies, including the use of nominated software to help identify known and novel proteins. The problems associated with studying non-model species and some potential applications of proteomics in veterinary science are also discussed. The conclusion emphasises that there is need develop mass spectrometry-based discovery proteomics methods capable of identifying a wide range of proteins in the circulating acellular proteome in veterinary species.

2.2 The significance of studying the circulating acellular proteome

The circulating acellular proteome occupies an important intersection of proteomics, diagnostics

and medicine[43]. It is partly for this reason that plasma proteins are a major focus in biomarker research, to the extent that a specialised database resource has been established for the Human Plasma Proteome Project (HPPP)[43]. While assaying circulating acellular proteins has been historically crucial in veterinary clinical chemistry[21], the widely available platforms for this purpose, for example handheld refractrometry and automated blood chemistry analysers, are either not sufficient or only provide superficial insight into the circulating acellular proteome. Ceciliani *et al.*[21] state that "proteomics holds the key to unlocking the vision of advancing veterinary pathology and diagnostics", highlighting the need to better characterise the circulating acellular proteome of veterinary species.

Immunoassay-based technologies such as ELISAs are the universal benchmark for validating other protein assay methods[29], because of their high degree of specificity[31]. Protein antibody ELISAs, however, are limited by their inability to scale up[31], the comparatively prohibitive cost of assay development[44], cumbersome multiplexing, long turnaround times, high failure rates[45] and their minimal ability to identify a wide range of proteins[46]. Furthermore, ELISA antibodies can vary in availability and quality[31]. In fact, some MS methods could even be better than ELISAs as a detection method when determining analyte reference ranges[47]. ELISAs face further competition due to the diminished need for high-quality antibodies when using MS proteomic approaches[48]. An attractive alternative to ELISAs is the development of targeted MS/MS-based proteomics approaches that are capable of profiling a wide range of proteins for a fraction of the cost[49], considerable amenability to multiplexing and shorter turnaround times for assays[9] as illustrated in Figure 2.0.

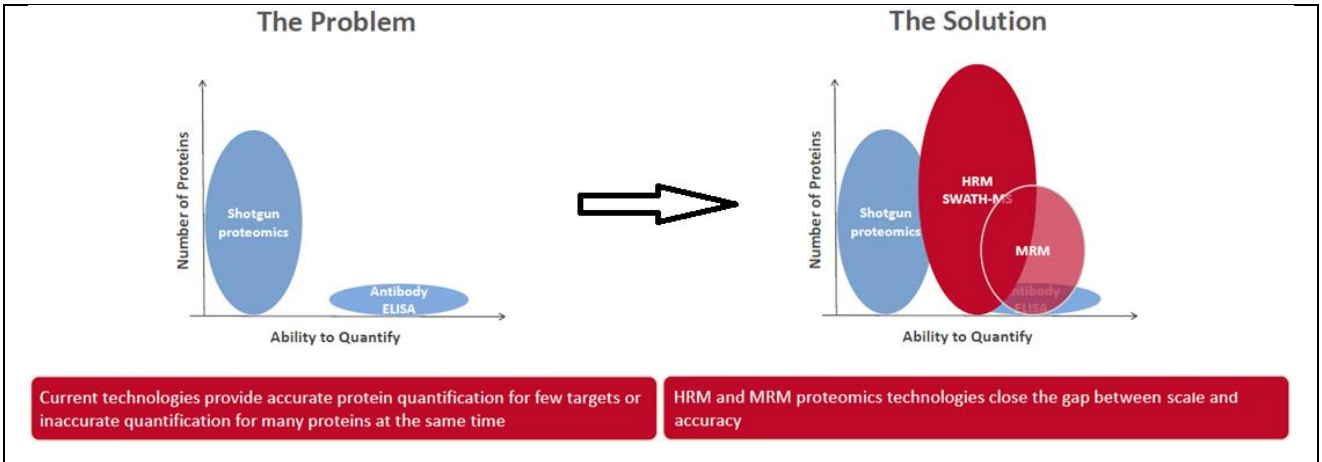


Figure 2.0. Quantitation of proteins: capabilities of mass spectrometry technologies and antibody ELISA. KEY: ELISA – enzyme linked immuno-sorbent assay; MRM –multiple reaction monitoring; HRM – high-resolution multiple reaction monitoring; MS – mass spectrometry; SWATH – sequential window acquisition of all theoretical fragment-ion spectra. *Adapted (with*

written permission of BIOGNOSYS) from Rinner, O. in 5th Berliner LC/MS/MS Symposium, BIOGNOSYS Proteomics - Next Generation, Berlin, Germany, 2013.

2.3 Difficulties in analysing the circulating acellular proteome

Despite numerous efforts aimed at the characterisation of the circulating acellular proteome[27,43], challenges still persist for its analysis[44]. As Pernemalm and Lehtio[49] inferred, "plasma has proven to be the most difficult material to work with using current MS-based proteomics techniques". In studies on humans, for example, the circulating acellular proteome comprises just over 20 highly abundant proteins that constitute 99% of the total proteins[27,43,49]. These proteins include albumin, factor H, apolipoprotein A1, immunoglobulin G, apolipoprotein A2, transferrin, apolipoprotein B, fibrinogen, acid-1-glycoprotein, immunoglobulin A, ceruloplasmin, alpha-2-macroglobulin, complement C4, immunoglobulin M, complement C1q, alpha-1-antitrypsin, immunoglobulin D, complement C3, prealbumin, haptoglobin, plasminogen and thyroxine binding hormone[27]. The remaining 1% lower-abundance proteins, which are biologically important and most sought after, create a WDR; the ratio of lowest to highest abundance protein detectable is up to 10 orders of magnitude[31,49,50].

Sample prefractionation remains crucial to deplete proteins of high abundance in the circulating acellular proteome of humans[51], and there is no evidence to suggest otherwise in veterinary species. It is reasonable to assume that the WDR problem also applies to samples from mammalian veterinary species. Scientific progress in sample preparation and prefractionation continues to yield promising results when coupled with more sensitive instruments for the detection of proteins that may have been nearly impossible to discern previously, while simultaneously addressing WDR[31]. A number of evolving contemporary proteomic approaches, together with some commonly used protein separation and fractionation strategies to address the problem of the WDR, are discussed in subsequent sections of this chapter.

The methods that have been developed for depleting high abundance proteins include protein-based approaches such as immunoaffinity depletion with antibodies[52-54], combinatorial hexa-peptide ligand libraries coupled to beads[55,56] and enrichment of carbonylated proteins by biotinylation of oxidised proteins with biotin hydrazide[57]. A second approach is a peptide-based prefractionation technique that is used to reduce the complexity introduced by digestion of the plasma proteins[54], for example by enrichment of cystein containing peptides using thiol-affinity resin (cysteinyl-peptide enrichment technology)[58]. A third approach is by a combined protein/peptide prefractionation such as glyco-affinity enrichment such as lectin affinity[59,60], hydrazide chemistry[61] and other carbohydrate enrichment for sialylated peptides[62]. A fourth approach to depleting proteins of high

abundance is metal affinity-based fractionation that applies immobilised metal affinity chromatography technology[63,64]. Of special mention is the use of stable isotope standards and capture by anti-peptide antibodies (SISCAPA)[65] – this method has been developed to detect and quantify low-abundance proteins, with promising outcomes. The advantages of applying any of these methods to deplete proteins of high abundance is that that they reduce the complexity of samples by minimising the problem of the signal-to-biological-noise ratio[65], whilst enriching for proteins of low abundance. The downside of applying these systems is the additional costs incurred for antibodies or kits, more time invested for sample analysis and the loss of analytes due to unspecific binding of non-target proteins or peptides[49]. In this project, the combinatorial hexa-peptide ligand libraries coupled to beads (ProteoMiner™) was applied to some samples (Chapter 5). For a detailed insight on this subject, there is an excellent comprehensive review on sample fractionation efforts to deplete highly abundant proteins as a way of dealing with challenges during global analysis of human plasma proteome elsewhere[49].

Due to the considerable complexity of the circulating acellular proteome samples[66], MS instruments that can provide comprehensive analysis of the sample in a single analysis are yet to be developed. It is for this reason that protein fractionation is commonly recommended, by splitting of a proteomic sample into multiple different fractions or pools so that each one can be individually analysed by a separate injection on an LC/MS/MS[67].

Some studies have aimed to deplete high-abundance proteins with concomitant enrichment of low-abundance proteins in horse[68], cattle[69,70] and pig[70] plasma or serum. However, as mentioned earlier in the various methods used, some important associated non-targeted proteins could be lost during the depletion process[71]. Reports by Di Girolamo *et al*.[22,72] and Ceciliani *et al*.[21], provide comprehensive insights into developments in the field of veterinary proteomics. It appears that two dimensional gel electrophoresis (2DE) followed by MALDI-TOF MS is the most-used approach in cattle, sheep, pigs, chicken, horses and even fish[22] for protein characterisation. In recent times however, this method has now been surpassed by newer more capable techniques and taken the opportunity to apply modern methods developed for human proteomic workflows. The modern methods for in-depth proteomic analysis incorporate protein depletion followed by multiple fractionation (10 or more fractions) and then analysis by high-resolution LC/MS/MS. In the current literature review, a few human and animal studies illustrate the developments and strategies used to analyse the circulating acellular proteome – the detected proteins or protein groups are provided in Table 2.0.

An important consideration that should be factored into the analysis of the circulating acellular proteome is biological variation or the differences that occur between individuals that samples are

derived from due to genetic or environmental differences. It is necessary to design experiments to take this variation into account and identify the experimental response of interest for example exploring proteomic changes in plasma after the administration of an endotoxin to sheep. In order to minimise these differences, it would ideally be necessary to have genetically identical individuals with the same physiological status that are raised under uniform environmental conditions. Unlike model animals such as mice, and without cloning, it is relatively difficult to have several genetically identical large animals for inclusion in experiments. In practice having animals of the same age, breed, sex, age and physiological status are commonly used in experiments.

Table 2.0. Some examples of analytical strategies used to explore the circulating acellular proteomes of human and veterinary species.

Strategy for proteome analysis	**Number of proteins identified or monitored**	**Reference**
Various technology platform comparisons for The Human Plasma Proteome Project.	3020, revised to 889	[73,74]
2D-DIGE + LC-MS/MS (Canine haemangiosarcoma).	1 monitored protein biomarker	[75]
Albumin and IgG depletion, 2D-DIGE, MALDI-TOF-MS or LTQ-MS (Swine fever).	10 differentially expressed proteins	[76]
MALDI-TOF, iTRAQ, LC-MS/MS (Bovine mycobacterial infection).	110 proteins	[77]
Depletion + SCX + LC-MS/MS + AIMS.	1105 protein groups; 1.5%, ≥2 peptides	[78]
Peptide N-terminal enrichment using TAILS.	113 proteins; p-value <0.05, ≥1 peptides.	[79]
Combinatorial peptide binding enrichment, 2D-DIGE, MALDI-MS or LC-MS/MS.	15 proteins	[80]
Combinatorial hexapeptide ligand library immobilised on a solid-phase matrix + 2D-DIGE.	29 proteins	[56]
SCX + IPAS + LC-MS/MS.	1709 protein groups; 5 % protein and peptide FDR.	[81]

Strategy for proteome analysis	**Number of proteins identified or monitored**	**Reference**
AEC + 1D SDS-PAGE, SELDI-TOF-MS (Canine lymphoma).	19 biomarker candidate proteins.	82
Various MS/MS data submitted to the proteomics identification database (PRIDE) and human Peptide Atlas using Trans-Proteomic Pipeline[43].	1929 protein groups.	19,83,84
Protein size fractionation, HILIC LC-MS/MS, SCX MudPIT + size exclusion.	1955 proteins; 5% peptide FDR, ≥1 peptides.	85
Protein size fractionation by IEF and GeLC-MS/MS.	1957 protein groups; 0.5% peptide FDR, ≥2 peptides.	86
SCX + intact-protein analysis system (IPAS) + LC-MS/MS.	1961 protein groups; 5% protein and peptide FDR.	87
2D-DIGE + MALDI-TOF-MS, LC-MS/MS (Bovine respiratory disease).	2 protein group changes.	88
Albumin and IgG depletion, 2D-DIGE, LC-MS/MS.	21 differentially expressed proteins	89
Glycoenrichment using hydrazide chemistry[61].	212 protein groups; 1%protein FDR; ≥1 peptides.	90
Peptide N-terminal enrichment.	222 protein groups; 1%peptide FDR, ≥1 peptides.	91
Depletion by affinity chromatography + SCX + RPLC + ESI-MS/MS.	2392 proteins (>94% confidence).	92
Dual-stage column, iTRAQ tagging, offline SCX chromatography, and LC-MS/MS.	689 proteins	93
AEC + SELDI-TOF-MS, LC-MS/MS (Sheep liver fluke infestation).	26 candidate protein biomarkers.	94

Strategy for proteome analysis	**Number of proteins identified or monitored**	**Reference**
AFP, LAP enrichment, HAP depletion, HSA and IgG multi-depletion.	279 protein groups; > 2 peptides, 95% confidence.	[95]
Immunoassay & 2D gel electrophoresis.	289 proteins.	[27]
Immunodepletion, 2-DE-DIGE, LC-MS/MS.	301 proteins.	[51]
2D-DIGE + MALDI-TOF-MS, nLC-MS/MS (Bovine mastitis).	34 proteins.	[96]
Protein size fractionation by depletion + GeLC, depletion + SCX, and depletion only.	342, 251 and 194, respectively; 1% peptide FDR, ≥2 peptides.	[97]
Glycoenrichment using non-lectin-based.	406 protein groups; 1% peptide FDR, ≥2 peptides.	[98]
Glycoenrichment using TiO_2 + 2D-chromatography.	413 protein groups; 1% peptide FDR, ≥1 peptides.	[99]
AEC + SELDI-TOF-MS (Pig reproductive and respiratory syndrome).	47 protein peaks.	[100]
2 different separations techniques + MS.	490 proteins.	[101]
2D-DIGE + MALDI-TOF-MS (Bovine serum & whey).	62 protein groups.	[102]
2D-DIGE + MALDI-TOF/TOF.	15 to 16 mainly APR proteins.	[103]
Serial/tandem depletion.	695 protein groups; 5% peptide FDR, ≥2 peptides.	[104]
SCX + 2D LC-MS/MS, MRM (Equine doping).	70 proteins detected, 49 proteins monitored.	[105]
Glycoenrichment using lectin-based N-glyco FASP technology[106].	800 protein groups; 1% protein and peptide FDR, ≥1 peptides.	[107]

Strategy for proteome analysis	Number of proteins identified or monitored	Reference
Albumin and IgG depletion, LC-MS/MS (Pig foot & mouth disease).	8-9 proteins.	[108]
Two different fractionation methods + MS.	9087 proteins.	[109]
1D-SDS-PAGE, MALDI-TOF-TOF (Sheep scab infestation).	2 major acute phase proteins.	[110]
2-D MALDI-TOF-MS; CE-IT-MS (Validated on sheep lung disease and peri-partum period).	42 medium-high-abundance proteins identified.	[111]
Combinatorial peptide binding enrichment, 2D-DIGE, MALDI-MS/MS (Lambs).	4 differentially expressed proteins.	[112]
Albumin depletion, iTRAQ, SCX LC-MS/MS (Bovine plasma & uterine fluid).	53 proteins FDR <0.10.	[113]
Undepleted, LC-MS/MS; nLC-PC-IDMS-SRM (Chicken ovarian cancer).	3 proteins identified and quantitated.	[114]

Key: 1D SDS-PAGE: one-dimensional sodium dodecyl sulphate polyacrylamide gel electrophoresis; 2D-DIGE:two-dimensional difference gel electrophoresis; AEC: anion-exchange chromatography; AIMS: accurate inclusion mass screening; antibody free depletion: LAP: low abundance protein, HAP: high abundance protein; HSA; human serum albumin and IgG multi-depletion; ESI: electro-spray induction; GeLC: in-gel digestion followed by liquid chromatography; HILIC: hydrophilic interaction liquid chromatography; IEF: iso-electric focussing; IPAS: intact protein analysis system; iTRAQ: isobaric tags for relative and absolute quantitation; LC: liquid chromatography; LTQ: linear ion trap quadrupole; MALDI: matrix assisted laser desorption ionisation; MRM: multiple reaction monitoring; MS/MS: tandem mass spectrometry; MS: mass spectrometry; MudPIT: multi-dimensional protein identification technology; nLC: nano-liquid chromatography; RP: reverse phase; SCX: strong cation exchange; SELDI: surface-enhanced laser desorption ionisation; TAILS: terminal amine isotopic labelling of substrates; TOF: time-of-flight; FASP: filter-aided sample preparation technology; CE: capillary electrophoresis; IT: ion-trap; PC: Protein cleavage; IDMS: isotope dilution mass spectrometry; SRM: selected reaction monitoring.

2.4 Separation strategies for protein analysis

Numerous studies describe universal sample preparation strategies for proteomic studies of non-model organisms[7], such as a recent review by Ceciliani and colleagues[21]. Generally, in discovery proteomics, protein samples are first separated by gel electrophoresis and then digested by an enzyme (usually trypsin) and the peptides are separated by chromatography before MS analysis[21,115], in order to achieve deeper proteome sequence coverage[8], as outlined below.

2.4.1 In-gel separation of proteins

Sample pre-fractionation using methods such as in-gel separation improves sensitivity to peptide detection after digestion during MS analysis[116]. One-dimensional sodium dodecyl sulphate polyacrylamide gel electrophoresis (1D SDS-PAGE), since early improvements[117] based on a glycine-Tris buffer system, is an effective electrophoretic technique for separating proteins[116,118]. This is based on the principle that protein fractionation in gels occurs due to a protein's molecular weight and/or isoelectric point[116]. This technique is a commonly adapted method for making self-casting gels, such as that described by Schägger[118,119] and others[116,120].

In-gel digestion of proteins involves the sample initially being fractionated using 1D SDS-PAGE. The gel is then sliced into bands that are subsequently subjected to in-gel protease enzymatic digestion and the products then analysed by LC-MS/MS[50,116,121-123], preferably after desalting. Despite being widely used, in-gel fractionation is dogged by poor recovery of proteins, poor gel-gel reproducibility and difficulties in being automated[116]. For global proteomic approaches in non-model organisms[36], as is the over-arching subject of this thesis, 1D SDS-PAGE is the experimental fractionation method of choice for discovery proteomics workflows, because it has previously been associated with a high number of protein identifications during LC-MS/MS[116].

2.4.2 Chromatography

Chromatography is used to separate peptides from a protein digest to reduce the complexity of the sample[32,48,124]. The different forms of chromatography are drawn from the same principle, as they all have a stationary phase (solid or liquid supported on a solid) and a mobile phase (a liquid or a gas) [125,126]. The most common column packing material for reversed phase chromatography is C18. Reversed phase chromatography relies on hydrophobic interactions and analytes partitioning between the stationary organic phase such as C18, C8 or biphenyl and the aqueous mobile phase which includes organic modifiers (commonly acetonitrile or methanol).The adsorption of peptides on the surface of the octadecyl carbon chain (C18)-bonded silica as they traverse the column, for instance, depends on hydrogen bonds, van der Waals forces and solubility[125,127,128]. Peptide adsorption can also be varied by changing the solvent, or the temperature or pH of the solvent[126].

There are elegant in-depth reviews focusing on protein and peptide separation chromatographic techniques for further reading in Tang *et al.*[124], Xie *et al.*[129], and Ly and Wasinger[125], for example.

High performance liquid chromatography (HPLC) is based on the observation that the chromatographic resolution increases as the particle size of the packing material decreases. It involves elution of smaller particle-size compounds (for example, peptides) through a pressurised capillary column with increased contact surface area[129-133].Unfortunately, the decrease in packing material particle size also leads to large increases in column backpressure. HPLC columns are routinely packed with 3 μ particles but the use of even smaller particles has led to the development of ultra-high performance liquid chromatography (UHPLC) methods. HPLC is not restricted to capillaries and is most commonly performed using metal columns with 2 mm internal diameter or larger. The work in this thesis was performed with 75 μ internal diameter capillaries using a nano LC. HPLC is faster and allows better separation of peptides in a mixture than unpressurised columns[48] and is amenable to detection methods that are highly automated and extremely sensitive[130]. Currently, two types of HPLC are recognised: normal and the reversed phase variants[130,134]. Normal HPLC uses a non-polar organic solvent and relies on van der Waals forces for adsorption and separation[131]. Reverse phase HPLC (RPLC), on the other hand, uses silica that has been made non-polar and a polar solvent such as methanol[126,134,135] and is renowned for its peptide desalting applications prior to MS analysis[125,136]. In another form of chromatography that uses ion-exchange, peptides are separated, based on the salt content or pH of the mobile phase by way of anion or cation exchange and is often used in combination with other fractionation strategies[125]. Multidimensional LC, for example, uses strong cation exchange (SCX) in combination with RPLC in packed columns, making it very efficient and currently the most commonly used method for peptide separation in bottom-up proteomics approaches[124,129].

The time taken for a particular peptide to travel through an HPLC column to the detector is known as retention time (RT)[126]. This is unique for each peptide and is measured from when a sample is injected into the column to the point at which the display shows a maximum peak detection height in the mass spectrometer[137]. The RT is determined by the material of the stationary phase in the column, composition of the solvent and column temperature[130,131]. Recently, there has been the development of an experimentally derived dimensionless peptide-specific value called indexed retention time (iRT) that allows very accurate RT prediction across many platforms; this has many applications in targeted proteomics[137]. It is preferable to have a high capacity trap column and a long heated analytical column, to enable higher loading of peptides for increased chromatographic resolution and increased limits of detection[138]. While a longer column would improve analytical performance, a typical mass spectrometer detector will only have an effective 3-4 orders of

magnitude of operation and increased loading will have a minimal effect on detecting low abundance proteins. The longer column would allow improved resolution and sampling of individual peptides, helping solve the problem of sample complexity, but this could also be achieved by increased sample fractionation. The peak width of a chromatographic peak is defined as the peak's full width at half maximum[139]. The narrower the peaks means that there's better chromatographic resolution. In practice, this refers to the median of the peak widths for all the identified peptides. Optimum and reproducible LC-MS performance outcomes is when peak widths remain in a narrow range over a multitude of runs[139].

A common method for detecting eluted peptides from an HPLC process is when it is coupled with a mass spectrometer. The output is recorded as peaks, each representing a compound passing through the detector[126,140]. Retention times can be used to identify the peptides present in the solution, provided that pure internal standards are known under identical conditions as in targeted analyses of peptides[140]. The peaks can be used as a way of determining quantities of peptides present in the mixture, with the area under the peak being proportional to the quantity of the protein present.[141] Based on this property, the RT will be the same, irrespective of the concentration of the compound in the mixture. It is therefore possible to calibrate a mass spectrometer so that it can be used to find even minute quantities of a peptide[126,140]. The experiments for this thesis used a Nano-HPLC (Eksigent Ultra 2D, Eksigent) with a Nanospray III ion source (SCIEX) for nano liquid chromatography and nano electrospray ionisation (nanoLC-nanoESI).

2.5 Mass spectrometry for routine detection of proteins

A mass spectrometer is an instrument that separates ionised molecules with different mass-to-charge ratios and determines the amounts of each particle in a mixture[142]. The basic principle of MS involves the vaporisation of a sample, causation of ionisation, acceleration, deflection and detection of ions[143]. The most common types of ion generation for protein anlaysis are matrix-assisted laser desorption/ionisation (MALDI)[144] and electrospray ionisation (ESI)[145]. MALDI and ESI are two forms of soft ionisation that are capable of ionising proteins and peptides while preserving their chemical structure intact. The other ionisation methods are plasma desorption (PD)[146] atmospheric pressure chemical ionisation (APCI), atmospheric pressure photoionisation (APPI)[147], fast atom bombardment (FAB) and electron impact (EI) commonly used for environmental work using GC-MS[148,149].

Mass spectrometry is an important method in analytical proteomics and can be used to identify molecular peptide ions by their characteristic fragmentation patterns in the mass spectrum[32,141,150]. In proteomics, proteins are digested into peptides, dissolved in a polar solvent, ionised and then separated according to mass and charge before being conveyed through a detector that quantifies the

ions by displaying peptide spectra[31]. Most mass spectrometers used for proteomics nowadays are triple quadrupoles, quadrupole time-of-flight (Q-TOF) and Orbitrap instruments[48], but more inventions are likely to surface.

A quadrupole mass analyser comprises of a set of four conducting rods arranged in parallel separated by a middle space, with the opposing pairs of rods electrically connected to each other[151]. Ions are separated based on the stability of their flight paths through an oscillating electric field in the quadrupole[151]. The electric field is created when a radio frequency (RF) voltage is applied between one pair of opposing rods within the quadrupole[151]. A direct current offset voltage is then applied to the other pair of opposing rods[151]. Only ions of a certain mass to charge (*m/z)* value will have a stable flight path through the quadrupole in the resulting electric field, whilst all other ions with unsteady paths will not reach the detector[151]. The RF and direct current voltages can be tuned in a way that allows the quadrupole to act as a mass filter or analyte-specific detector for ions of a particular *m/z*[151]. The analyser can also be operated to scan for a range of *m/z* values by continuously varying the applied voltages[151]. In summary, by applying voltage ramps and RFs, complex magnetic fields are created in quadrupoles that allow ions to be focused and filtered, by magnetic fields.

A triple quadrupole mass spectrometer consists of two quadrupole mass analysers with a collision cell sandwiched between[151]. The first quadrupole mass analyser selects the precursor ions[151]. The selected precursor ions are then fragmented in the collision cell by a process known as collision-induced dissociation (CID) in an inert gas such as argon[151,152] to obtain tandem mass spectra (MS/MS)[153]. Product ion patterns and relative ion abundance can be considerably reproducible if the CID conditions are consistent[151]. The product ions are analysed or selected by the second quadrupole mass analyser, and then passed on to the detector[151]. The precursor and product ion pairs are called mass transitions[151]. When the electric field and collision energy are maintained, only analyte ions having a specified mass transition (precursor/product ion pair) are able to reach the detector, which results in the high specificity of tandem quadrupole mass spectrometric methods[151]. This type of data acquisition is called selected-reaction monitoring (SRM) or multiple-reaction monitoring (MRM) when numerous transitions are monitored during a chromatographic run[151,154].

In mass spectrometry, three methods of ion detection are recognised; these include direct charge detection (Faraday cup detector), image charge detection (inductive detector), and secondary electron generation (electron multiplier (EM) and microchannel plate (MCP) detectors)[155]. Due to the comparatively low sensitivity, direct charge detection is only commonly used in magnetic sector instruments. Because inductive detectors use non-destructive detection method, they are critical to Fourier-transform instruments, for example Orbitrap mass analysers, where signal averaging of

circulating ion cloud is central to the operation of the instrument[155]. Time-of-flight (TOF) ion detectors are designed to have large areas, rapid response times to provide good timing resolution with correspondingly accurate *m/z* determinations and high sensitivity, which are best met by the EM and MCP detectors, based upon their generation of secondary electrons[155,156]. The TOF detector was used in this project.

Peptide fragmentation processes in collision cells – a key technology used in this thesis creates positively charged molecules that are accelerated in a vacuum and deflected by an electromagnet and amplified into an electronic display in the form of a stick diagram[126]. Peptides form molecular ions in the ionisation chamber of the mass spectrometer that can be mass selected and detected[140,157]. When this is coupled with an HPLC output corresponding to when the detector is showing a peak, some ions that are passing through the detector at the time can be diverted (mass selected) and converted into either a mass spectrum, a chromatographic display or both[140]. This gives a fragmentation pattern that can be compared against bioinformatics databases of known patterns[65,141,158,159]. This principle is based on that fact that molecular ions are unstable and tend to break into predictable fragments in a specific reproducible fashion[150] as illustrated in Figure 3.0. Each line on the stick diagram is a mass spectrum representing a different fragment produced when the molecular ion breaks up[126,157]. The tallest line in a molecular ion peak is arbitrarily fractionated into 100 parts called the base peak, representing the commonest fragment formed[126]. The information from the peaks in the spectrum can then be used to derive the peptide sequence[160]. Ideally, the peaks of interest in an MS/MS spectrum are those represented by b- and y-ions, corresponding to the prefix of N-terminal (b-ion) and the suffix of C-terminal (y-ion) fragments, respectively[160]. In practice however, the MS/MS spectrum contains a compound mixture of peptide fragments and uninterpretable 'noise' peaks like those of different ion types such as a-, c-, x-, or z-ions generated when the peptide is not cleaved at an amide bond[160]. Bioinformatic tools enable unknown spectra to be analysed and matched against databases that have libraries for the mass spectra[65,141]. The peak with the highest *m/z* value, represents the relative molecular weight of the peptide, or what is referred to as the precursor ion, which allows it to be identified[65]. The mass of the ion detected is related to the magnetic field used to impel the ions to reach the detector, enabling the number of ions (current) *vs m/z* to be calibrated based on the ^{12}C scale[161]. The relative paths of light to heavy ions in a mass spectrometer give accurate information on the relative masses of isotopes and their relative abundance (proportions) or intensities. That means that the identity of a wide range of compounds can be found without having to know their retention times[126].

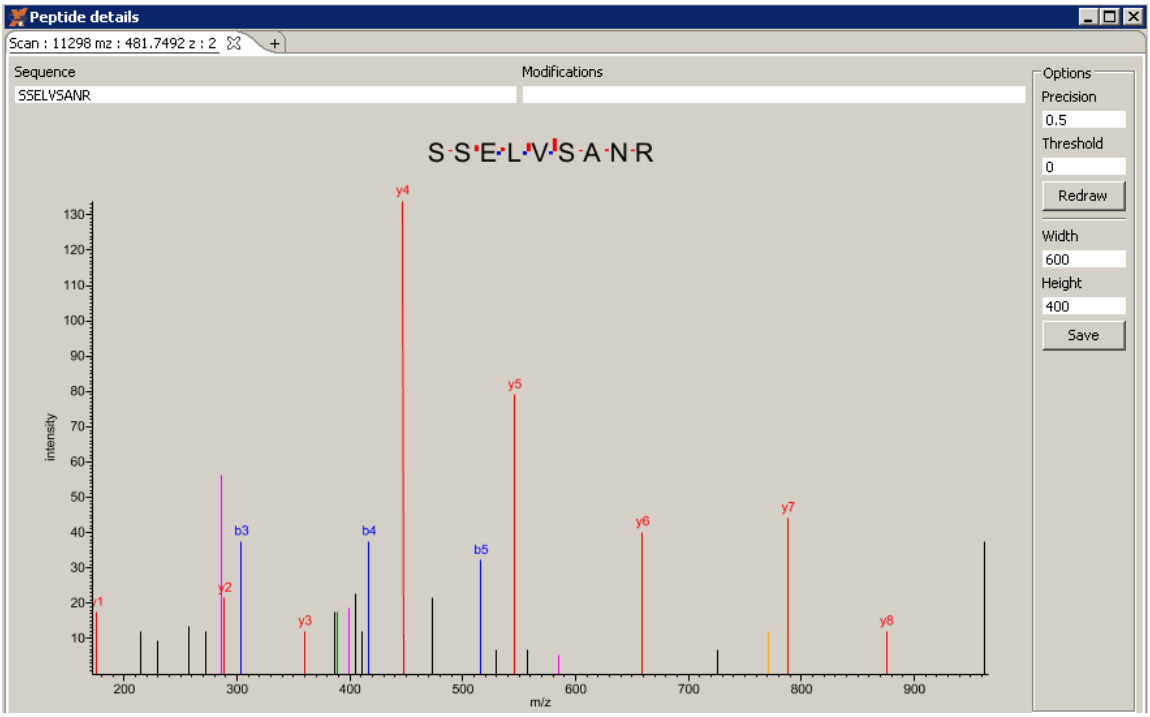


Figure 3.0 Fragmentation pattern of the peptide S-S-E-L-V-S-A-N-R from antithrombin-III (Serpin C1) as displayed by X! Tandem (The Global Proteome Machine Organization; version CYCLONE 2010.12.01.1) [162,163] after a data dependent acquisition (DDA) experiment using a quadrupole time-of-flight instrument (TripleTOF® 5600+, SCIEX) on a tryptic digest of healthy sheep serum. The highest *m/z* value represents the relative molecular weight of the peptide precursor ion. ProteinPilot™ Software 5.0 (SCIEX) using the Paragon™ Algorithm: 5.0.0.0, 4767[164], was initially used to search a custom database in order to generate a calibrated Mascot generic format (.mgf) file that was eventually fed into X!Tandem for analysis. The red sticks represent y ions which extend from the carboxyl or C-terminus of the fragmentation; the blue sticks are b ions representing fragment ions extending from the amino or the N-terminus. The sequential numbers b1-b5 and y1-y8 represent an offset from the previous by the mass of an amino acid.

2.6 Contemporary proteomics methods for samples from veterinary species

In this thesis, it was proposed that a typical proteomic method for a non-model organism be adapted from quantitative proteomics ideas such as advanced by Bantscheff *et al*.[48] and Armengaud *et al*.[7] This comprises sample preparation, protein discovery *(discovery proteomics)*, protein verification using both synthetic and endogenous proteins from biological samples in a targeted fashion *(targeted proteomics)*, and the testing and optimisation of selected bioinformatics workflows[48].

Working with the preceding approach, and once sample preparation workflow is optimised, the proteome can then be characterised using high-resolution MS/MS technology with representative

tissue samples from healthy normal organisms. As pointed out elsewhere[7], unlike model organisms, non-model organisms make difficult experimental organisms[165] as more effort is required because of the absence of well-defined protein databases, but protocol adaptation and method development should be close to identical.

2.6.1 Discovery proteomics

Discovery proteomics is a strategy for global proteome analysis[34,49], using MS technology to detect a large number of proteins from multiple experiments[31]. The compilation of data from multiple experiments is actually one of the defining properties of discovery proteomics. Work for this thesis used a quadrupole time-of-flight instrument (TripleTOF® 5600+ System, SCIEX) that relies on DDA following liquid chromatography (LC) for spectral library construction needed for downstream analysis (Figure 3.1).

2.6.2 Targeted proteomics

Targeted MS proteomics[166] is a technology that involves building an assay using preselected proteins or peptides of interest to provide precise, quantitative and sensitive data important for hypothesis-driven questions, with the aid of a triple quadrupole mass spectrometer[31,65,140]. Albeit distinct from discovery-based proteomics approaches, targeted proteomics complements untargeted shotgun methods[140] in that it can be used for verification purposes[65] by bridging the gap between antibody-based detection (e.g. ELISA) and discovery-based MS, by focussing experiments on important peptides to be fragmented and analysed[31,140].

With advances in high-resolution/high sensitivity MS instrumentation such as Q-TOF, Q-Trap, Orbitrap[167], alongside the inclusion of associated techniques like LC and ESI, it is now possible to identify and quantitate many proteins in the circulating acellular proteome in a single experiment using targeted approaches[65,168-170]. The use of MRM / SRM (defined earlier when discussing quadrupoles in section 2.5) being a highly selective and sensitive technique[115,140], focusses the mass spectrometer to detect only a group of chosen analytes[31] with known fragmentation properties[140] for protein identification and quantitation[43]. This technique is a useful avenue for detecting and verifying assays for low-abundance proteins when using high-resolution mass spectrometry[43,65]. The triple quadrupole mass spectrometer used for MRM is one of the oldest kinds and their MRMs are a “traditional” method when compared to high-resolution mass spectrometry.

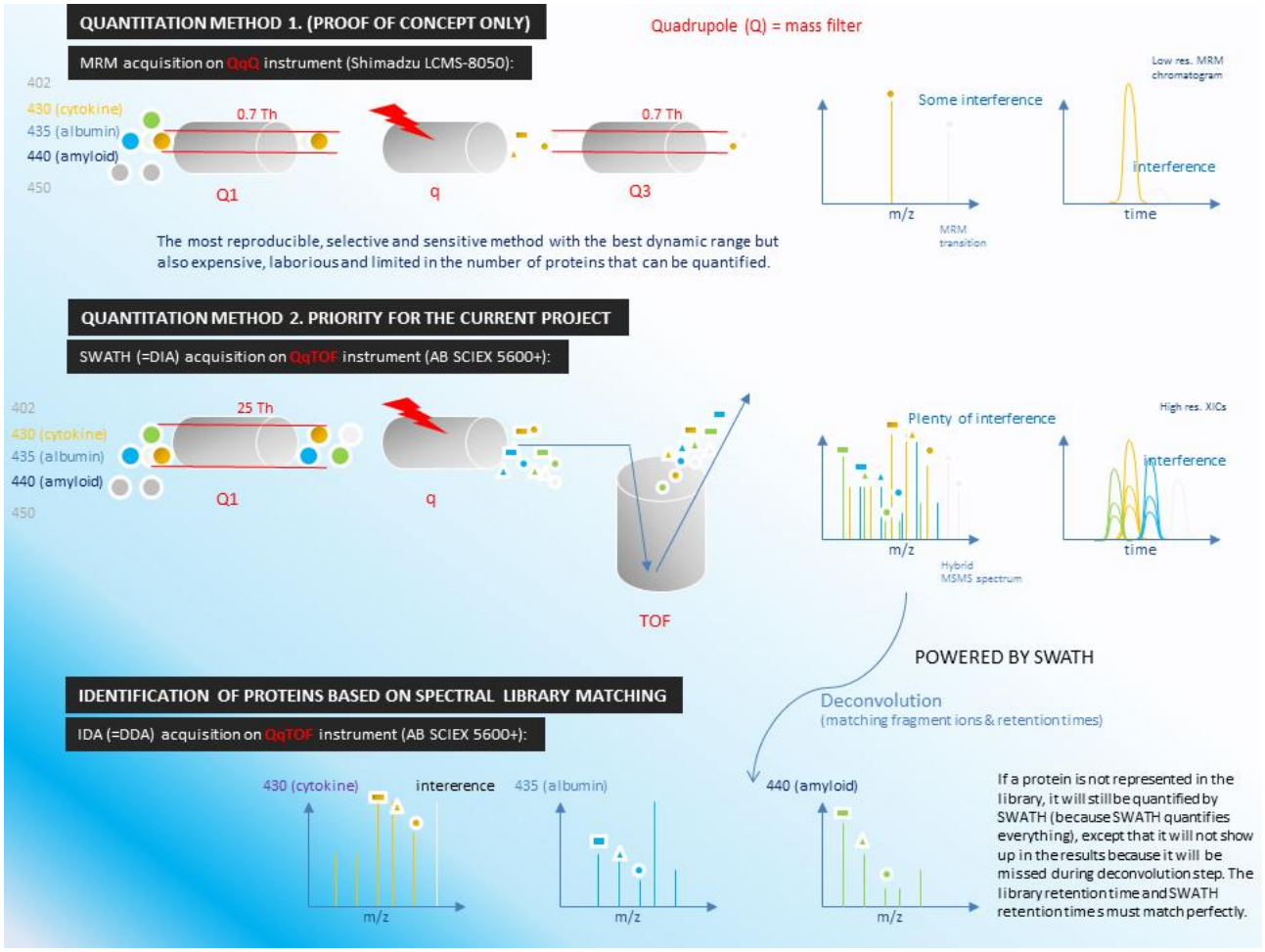


Figure 3.1. Schematic diagram for discovery and targeted proteomics workflows. Protein quantitation method 1 illustrates the principle of the triple quadrupole (QqQ) mass spectrometer commonly used for targeted proteomics. Protein quantitation method 2 illustrates the principle of quadrupole time-of-flight (QTOF or QqTOF). **KEY**: MRM-Multiple reaction monitoring; MS-Mass spectrometry; SWATH-Sequential window acquisition of all theoretical fragment-ion spectra. *Courtesy of, and with permission from, Dr Pawel Sadowski, Queensland University of Technology.*

When a stable isotope dilution multiple reaction monitoring mass spectrometry (SID-LC-MRM-MS) approach on a triple quadrupole (QqQ) instrument[171] is used in conjunction with Skyline software[172], it is possible to verify the most likely protein biomarker candidates for inflammation in the circulating acellular proteome of sheep. There is strong benefit to using MRM-based assays, considering their sensitivity, reproducibility, robustness[31] and superior multiplexing capabilities[173], especially in a multispecies-driven industry such as in veterinary diagnostic pathology in future.

A peptide spectral library (PSL) built from discovery proteomics using a Q-TOF instrument is required for targeted quantitative workflows involving sequential window acquisition of all theoretical fragment ion spectra mass spectrometry (SWATH) or LC-SWATH-MS, a key

technology used in this thesis and MRM. The operation of SWATH is similar to MRM but with much wider isolation windows in Q1 (Figure 3.1) where the increased resolution is provided by the high-resolution TOF detector. These isolation windows are cycled at a sufficient speed to provide MS/MS fragmentation of all analytes present in the sample inside a given mass range during SWATH data aquisition[1]. The SWATH algorithm is written to focus on *m/z* windows instead of ion abundance, to allow fragmenting everything without a priori knowledge of the sample[1].The challenge is then to develop software (such as SWATH™ Acquisition MicroApp, SCIEX) that is capable of analysing and deconvoluting the complex data. Basically, the discovery analysis that has identified a peptides' precursor ion mass, its fragmentation pattern and its retention time is then used to identify specific fragment ions in the correct SWATH window at the correct retention time[1].While the SWATH method is a form of data-independent acquisition (DIA)[174], like MRM, it relies on spectral libraries generated from DDA experiments in order to identify and verify a wider range of proteins. Unlike MRM, where peptide targets are predetermined before data acquisition, SWATH quantifies all peptides by DIA but also allows choosing of protein targets during subsequent data analysis[1]. SWATH acquisition is best suited to target high and medium abundance proteins and is known for providing comprehensive sample coverage with negligible reduction in quantitative precision[175].

The downside of discovery proteomics workflows is the reduced ability to identify proteins of low abundance, the need for extensive protein database searches and the complex bioinformatics processes required to analyse MS data[31,115]. Both SWATH and MRM workflows require a PSL for analyte identification; however, only SWATH requires a library with identical RT parameters as the one used during data acquisition. Consequently, this calls for the need to generate a spectral library exactly the same way as the ultimate SWATH samples, which can be impractical[176]. A strategy to overcome this problem is to use iRT, which allows the RT to be pegged to the peptide (not the platform or data acquisition parameters), thereby allowing fast and accurate RT predication for downstream applications, including SWATH data extraction[137].

2.7 Quantitation of proteins

Peptide and protein quantitation can be achieved by using either label or label-free approaches for absolute or relative quantitation needs[29,48]. The vast majority of quantitation methods are derived applications from cellular proteomics[177]. For quantitation purposes, in-gel digestion appears to affect recovery of individual peptides compared to in-solution digestion approaches[178]. It is for this reason that SWATH and MRM workflows do not currently support in-gel fractionated samples. Some examples for the relative quantitation of peptides include: metabolic labelling with amino acids in cell cultures – e.g. stable isotope labelling by amino acids in cell culture (SILAC)[179];

chemical protein and peptide labelling – e.g., tandem mass tags (TMTs) and isobaric tags for absolute and relative quantitation (iTRAQ)[180,181]; label-free quantitation including LC-MS/MS analysis of peptides such SWATH[48,176]; and MRM (SRM).[48,49] As for absolute quantitation, the methods that have been developed using the stable-isotope-labelled standard approach include absolute quantitation (AQUA)[182], a concatemer of Q peptides (QCAT)[183] protein standard absolute quantitation (PSAQ)[184], absolute SILAC [185] and FlexiQuant[186].

Label-free methods for absolute quantitation of peptides and proteins continue to be developed. These include using peptide-to-spectrum matches (also known as spectrum count)[187], methods employing signal intensity, and inductively-coupled plasma mass spectrometry (ICP-MS)[48]. Label-free methods for peptide and protein quantitation have greater affordability and lack labelling contaminants that could interfere with MS analysis[115]. SWATH processing provides quantitation while confirming the identity of detectable proteins and peptides[175,188].

2.8 Bioinformatics strategies and applications for identifying proteins

2.8.1 Strategies and computer software applications for identifying proteins

Mass spectrometry proteomics heavily relies on sophisticated computational strategies to identify peptides, by matching the acquired MS data against protein databases[6]. The practice is to have the sequence database in FASTA format, with a repository of sequences obtained from reputable sources for example the National Center for Biotechnology Information (NCBI) and the Universal Protein Resource Knowledgebase (UniProtKB) organisation[20,189]. The acquired MS raw data for discovery proteomics workflows is usually processed by the instrument in the first instance into a propriety data format. The data is then processed by instrument-bundled or other propriety or open access software platforms for protein database searches to identify proteins.

There are various types of software that can be used to process raw data from the instrument for searching databases to identify proteins. Some of these software platforms include MS-GF+[190], SEQUEST[191], Andromeda[192], Comet MS[193], MS Amanda[194], MyriMatch[195], Morpheus[196], Tide[197] (TurboSEQUEST + Crux[198]), Open Mass Spectrometry Search Algorithm (OMSSA)[199], ProteinPilot™ software (SCIEX) which uses the Paragon™ algorithm[164], and Mascot (Matrix Science)[200] software. Other tools that could be used are Scaffold software (Proteome Software Inc.)[187] which incorporates X!Tandem (The Global Proteome Machine Organization)[162,163], PeptideProphet™ and ProteinProphet™ bioinformatics algorithms (Institute for Systems Biology, Seattle) to identify proteins[187]; and the Trans-Proteomic Pipeline (TPP)[201] – an open-source bundle of software tools that has been developed for the analysis of MS/MS information[202]. Alternative protein identification search engines continue to evolve that are discussed in depth elsewhere[8].

To identify peptides, the experimentally-generated mass spectra are matched against protein sequence databases, generating a qualitative list of potential matches, ranked by confidence, of proteins that may be present in the sample[21]. Identifying peptides, however, is not without challenges, and some remedies have been proposed[6]. Briefly, four cardinal attributes have been identified as being crucial to the successful identification of peptides (Table 2.1). In addition to these, there is also need for computational strategies for building and using customised protein sequence databases, overcoming the challenge of false positive identifications in proteomics, reliable prediction of splice-junctions, and correction of sampling bias[6], just to name a few.

Table 2.1. Determinants of peptide identification in proteomics studies.*

No.	Attribute
1	Finding the right balance between completeness and optimum size of protein sequence database. The more comprehensive it is, the larger the database size will be, leading to the increase in the time required to perform searches.
2	Applying multiple database search tools to increase sensitivity and specificity associated with a given peptide identification strategy. Different search tools may be able to assign spectra to peptides that could have been missed by another search tool.
3	Accessing computational time and resources necessary for data processing, factoring in false discovery rate (FDR) estimation or analysis. Proteomics data acquisition, processing and data analysis require considerable investment in computing in terms of hardware, software and considerable patience in order to process the data.
4	Having the ability to interpret the data in biological context. There is need to have adequate human resources with appropriate training to relate proteomics findings for potential application to real-life situations.

* Drawn from references [8] and [6].

In general, the majority of proteomic searching tools are largely unintegrated and remain rather labour intensive[159]. One strategy to address the manual task problem is to derive peptide sequences using *de novo* sequencing and homology searching of MS/MS spectra, without the need for a sequence database[203]. An excellent lesson on the salient computational and user-level aspects on *de novo* sequencing of peptides has been covered in detail by Chong and Leong[204]. Meanwhile, integrated proteomic search tools continue to be developed that can perform all functions, from MS/MS matching to genome annotation, on a single platform[8]. One platform that minimises the human interface by promoting automation and with the capability of running on personal computers

is a software tool called Peppy[159]. Another example is the Proteomic Mapping Pipeline, which has easily-customisable proteomic search capabilities for non-model organisms[205].

Given the preceding background and together with an optimised bioinformatics workflow, an established proteomics method could be applied to interrogate the circulating acellular proteome in mammalian veterinary species[17].

2.8.2 Gene ontology (GO) enrichment and protein pathway analyses of identified proteins

Gene ontology (GO) annotation and pathway analysis of proteins is a bioinformatics approach used after identifying proteins to assess the biological relevance of the protein data[206]. This is necessary as it complements protein quantitation for example in the outcomes from protein fold changes during biomarker studies[2,207-209]. Several tools are publicly available for GO enrichments analysis, each using unique algorithms such as g:Profiler[210], BiNGO[211] that displays data in Cytoscape[212], Ontologizer[213], GOrilla[214] and DAVID (The Database for Annotation, Visualization and Integrated Discovery) gene functional classification tool[215], to name only a few. The application of protein pathway analytical tools facilitates the visualisation of interactions between genes, proteins and other biological molecules to reveal biological pathways and to generate global canonical pathway protein interactions[216,217] that could be fundamental when using animal models to study human disease, for instance, because of the conserved nature of pathways[218]. Examples of such pathway tools include the commercial Ingenuity Pathway Knowledge Base[31,33,219] and the open-source freely available PANTHER (Protein ANalysis THrough Evolutionary Relationships) software suite[220].

2.9 Validation of proteomics results

The large number of spectra generated by MS proteomics experiments to be matched with peptide sequences and subsequent protein identification requires validation in order to minimise false identifications. Validation of the data can be performed both at peptide and protein levels[221]. This may involve manual inspection, but this is cumbersome and subjective, or by using automated analysis software[221]. A useful strategy is to use multiple software that implement probabilistic approaches[221] to analyse the same dataset and select spectral identifications that are shared among the various software. It is important to remember that the analysis software only just matches patterns between theoretical and experimental data. The optimal form of validation would be to synthesise every peptide identified and compare the synthetic form to the hypothetical data. Clearly this is not possible for all, but a few of the most important peptides to illustrate that the analysis is working. Protein-level validation may take the form of technical and biological replication of the experiment and compare the protein identifications[221], by using the emerging high resolution MRM (HR-MRM) method or the gold-standard specific antbody techniques (ELISA and Western

blotting)[222].

2.10 Studying non-model mammalian organisms in proteomics

The proteomics approach for studying non-model organisms is based on the principle of the genetic theory. Genes encoded in DNA are transcribed into RNA that is then translated by ribosomes into proteins. The (accurate) detection of an endogenous protein in the circulating acellular proteome is therefore a confirmation that there is a genetic code for it[223,224]. The availability of accompanying genome data is central to the ability to detect proteins. Except for animal species commonly used in translational research whose genomes have been largely sequenced (such the mouse) or those of veterinary importance (such as the dog, horse, ox, chicken and pig), there is a relative scarcity of proteomic data for other animals[21]. Complete genomic data is important because many studies on animal models of disease are founded on the premise that every form of disease insult is associated with characteristic changes in protein expression[17,225-229]. For animals whose genomes have only been recently sequenced, such as sheep[230], coding errors are still be present, and there may be many gaps that need to be annotated. Another problem of working with non-model organisms is that not enough is known about the protein change responses to a range of physiological and pathological events. This problem is confounded further by the impact of breed differences on response to stimuli in many veterinary species.

As proteomics heavily relies on complete genomic sequencing data, some approaches used in non-model animals include optimising data analysis steps, combining available database information of the target species with related species, testing different software tools and different search parameters[6,8] and assuming that the results will be similar to that of humans. The co-ordinated sharing of proteomic data through publicly available repositories such as the ProteomeXchange consortium[231] is invaluable, as it allows constant querying of proteomic data on an open access platform.

2.11 Potential applications of proteomics in veterinary science lies in proteogenomics

Proteogenomics has been described as a platform that can be used to establish a draft protein database of any organism, by providing an avenue for studying difficult experimental organisms[7]. A truly comprehensive proteogenomics project is characterised by attributes that encompass genomic and proteomics efforts with the end result of annotating the genome[6,8]. High-resolution MS techniques for peptide detection enable the prediction of open reading frames of an organism, by correlating MS data with genome sequence information based on the expressed protein evidence[8,232] from protein samples. This means that, although time-consuming, there is now capacity to identify a plethora of proteins with potential biotechnological applications in production and companion

animals. Proteomic data can also be used for annotation of genomes of most animals that are yet to be fully sequenced[8,232].

After identifying proteins, quantitating them can be useful in diagnosing pathological aberrations and monitoring disease progression[43]. However, comprehensive MS-based methods to estimate serum or plasma levels of various proteins in veterinary species are still lacking. Using a spectral counting approach[84] to build a database to determine normal protein levels for animals could prove useful for veterinarians and researchers in future.

2.12 Conclusions

Proteomics is a promising tool for veterinary science applications. Method development and validated experiments are needed, preferably those benchmarking on studies in humans, to elucidate important findings, including the detection of subclinical disease in domestic animals – an application that could improve animal welfare. Resilience to disease is an important trait; however, the relative paucity of detailed information on the immune response in many veterinary species, especially production animals, means that there may be sub-optimal scope for selecting breeding stock based on immunological resilience. If specific and reliable markers of immunological resilience could be identified, such markers could be used to identify individuals that might be at increased risk of infectious disease, and to predict resilient individuals. Such markers could then be used to study the heritability of resilience, with the eventual aim of developing a screening panel to select resilient breeding stock. Furthermore, a panel of markers could also be used as a surveillance tool, to monitor the potential deleterious immunological effects of selecting for performance traits such as reproduction or growth rate. Such a strategy would help to inform optimal balanced genetic development in flocks. While the outlook appears bright, there is need to optimise sample collection/preparation methods, address WDR issues and reduce the cost of running assays, to allow the widespread use of proteomics in domestic animal science[21] and to uncover protein expression and characterisation by building and exploiting the capabilities of PSLs in this era of proteogenomics.

CHAPTER 3

3.0 Materials and methods

3.1 Introduction

This chapter presents the generic materials and methods used in proteomics studies applied in this thesis. The generic or universal approach of protein analysis relies on acetone precipitation of proteins in order to remove lipids mainly[233]. This is followed by in-solution and in-gel digestion of proteins into peptides with trypsin and desalting of peptides prior to LC-MS/MS analysis. The effect of desalting of peptides prior to analysis and its reproducibility as a critical element of all proteomic experiments was also evaluated. Desalting of peptides is even more crucial when using label-free proteomics approaches[176]. The details of nanoLC-nanoESI-MS/MS method employed are also provided. Considered together, the procedures covered herein form a common basis for the adapted proteomics methods used in the entire thesis.

3.2 Ethics statement

The studies included in this thesis were conducted in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes[234]. Animal tissue samples for method development were acquired following approval from Queensland University of Technology (QUT) Confirmation number: 1400000591, QV reference number: 46375 and Ethics number TRIM 10/8428 issued to Serum Australis Pty Ltd (http://www.serumaustralis.com.au), for bleeding live sheep and supply of blood products. Tissue from live animal experiments had animal ethics approval obtained from the University Animal Ethics Committee of QUT (reference 0800000555), which was ratified by The University of Queensland (UQ).

3.3 Reagents used for sample preparation

10x Tris/Glycine/SDS Buffer (Bio-Rad, Inc., cat. No. 161-0732), Acetonitrile (LC/MS grade, suitable for UHPLC-UV; Fisher Scientific, Optima®, cat. No. A955-4), Acrylamide/Bis-acrylamide (Sigma-Aldrich, cat. No. A2917), Ammonium acetate (≥ 98%; Sigma-Aldrich, A1542; Lot # SLBBF4804V), Ammonium bicarbonate (NH_4HCO_3) ReagentPlus®, ≥99.0%; Sigma-Aldrich, cat. No. A6141), Ammonium Chloride (AJAX Chemicals, UNIVAR®), Ammonium persulfate (Bio-Rad, cat. No. 161-0700), Bicinchoninic acid (BCA Protein Assay Kit; Pierce™, cat. No. 23225), Bromophenol Blue (Bio-Rad, cat. No. 161-0404), Disodium ethylenediaminetetraacetate dihydrate (EDTA ≥99% pure, Bio-Rad, cat. No. 161-0729); DL-Dithiothreitol (DTT, ≥98% (TLC), ≥99.0% (titration); Sigma-Aldrich, cat. No. D0632),

Formic acid (≥99.5 %, LC/MS grade; Fisher Scientific, Optima®, cat. No. A117-50), Formic acid in

acetonitrile (0.1% (vol/vol), LC/MS grade; Fisher Scientific, Optima®, cat. No. LS120-4), Formic acid in water (0.1% (vol/vol), LC/MS grade; Fisher Scientific, Optima®, cat. No. LS118-4), Glycerol (for molecular biology ≥99%; Sigma-Aldrich, cat. No. G5516), Hydrochloric acid (BioReagent for molecular biology 36.5-38.0%; Sigma-Aldrich, cat No. H1758); Iodoacetamide (Crystalline HPLC, ≥99%; Sigma-Aldrich, Lot # SLBC7561V, PCode: 1001691393); Mass Spectrometry Safe Protease and Phosphatase Inhibitor Cocktail (MSSAFE, Sigma-Aldrich), Protease inhibitor cocktail tablets (cOmplete™, Mini, EDTA-free; Roche, Lot No. 10051300), Protein gel staining solution (EZ-Run™, Fisher Scientific; cat. No. BP3620), Sodiun dodecyl sulfate ≥ 98.5% (Sigma-Aldrich, L4509, Batch # 046K0085), Tetramethylethylenediamine (TEMED; Bio-Rad, cat. No. 161-0800), Trifluoroacetic acid (CHROMASOLV® for HPLC, ≥99.0%; Sigma-Aldrich, cat. No. 302031), Trifluoroacetic acid in water (0.1% (vol/vol), LC/MS grade; Fisher Scientific, Optima®, cat. No. LS119-212), Trisaminomethane (Tris ≥99.8%; Bio-Rad, Inc., cat. No. 161-0716), Trypsin (sequencing grade modified; Promega, cat. No. V5117), Urea (Bio-Rad, cat. No. 161-0730), Liquid chromatography–mass spectrometry (LC-MS) water (LC/MS grade, suitable for UHPLC; Fisher Scientific, Optima®, cat. No. W6-4), Deionised water (Millipore), Albumin standard 2 mg/mL (Bovine Serum Albumin Standard, Pierce™, Thermo Fisher Scientific), ExcelGel SDS Buffer Strips (anode and cathode) (GE Healthcare Sciences), and IPG Buffer pH 4–7 (GE Healthcare Sciences).

3.4 Equipment used in the experiments

Quadrupole time-of-flight mass spectrometer (TripleTOF® 5600+ System, SCIEX) with Nanospray III ion source (SCIEX), Nano-HPLC (Eksigent Ultra-2D, Eksigent Technologies, Dublin, CA), Centrifuge (Eppendorf, cat. No. 5424R), Mixer (Eppendorf Thermomixer® Comfort), Vortex mixer (Ratek, Model VM1), pH meter (Cardy Twin pH Meter, Spectrum Technologies), Spectrophotometer (NanoDrop 2000; Thermo Scientific), and Offgel Electrophoresis System (3100 OFFGEL Fractionator, Agilent Technologies). SpeedVac Concentrator (Christ® cat. No. RVC 2-33 IR).

3.5 Consumables used for the experiments

Protein Enrichment Kit (ProteoMiner™ Small-Capacity Protein Enrichment Kit, Bio-Rad cat. Nos. 163-3006 and 163-3008), Polyacrylamide gels (Mini-PROTEAN® TGX™ Precast Gels, Bio-Rad, cat. No. 4561084), Electrospray emitters (New Objective, cat. No. FS360-20-10-N-20-C12), NanoLC analytical column (ChromXP C18 3 µm, 150mm × 75 µm 120 Å, Eksigent Technologies, Dublin, CA), NanoLC trap cartridge (ChromXP C18CL 5 µm, 10 mm × 0.3 mm 120 Å, Eksigent Technologies, Dublin, CA), Protein low binding tubes (volume 0.5 and 1.5 ml; Eppendorf, cat. nos. 022431064 and 022431081), 10 µL (0.6 µL bed of chromatography media) desalting pipette tips

(Millipore® ZipTips C^{18}, Sigma-Aldrich, cat. No. Z720046), 100µL bed desalting pipette tips (Pierce™ C^{18} Tips, Thermo Fisher Scientific, cat. No. 87784), 750 µL Volume polypropylene autosampler vials (12 x 32mm Snap Neck Vial, Waters, cat. No. 186005224).

3.6 Sample preparation for protein analysis

Frozen sheep serum samples were thawed on ice and then centrifuged at 13,000 *g* at 4 ^{o}C for 20 min as previously described[16]. The sediment and top layer comprising mainly of lipids and suds were discarded, retaining the supernatant. The protein concentration of the supernatant was determined with bicinchoninic acid (BCA) protein assay kit (BCA Protein Assay Kit, Pierce™) according to the manufacturer's instructions using a spectrophotometer (NanoDrop 2000, Thermo Scientific). The supernatant was then either directly analysed or enriched by acetone precipitation of proteins. In some experiments, a protease inhibitor cocktail tablet (cOmplete, Roche) was added into the sample after thawing, according to the manufacturer's instructions in order to evaluate if there was any difference in protein yields when a protease inhibitor was present in the samples than when it was not[235].

3.7 Generic universal sample delipidation by acetone precipitation of proteins

Proteins in serum were precipitated by adding 4 × (v:v %) of cold (-20 °C) acetone and then incubating at -20 °C for 16 h, prior to centrifugation at 4,000 *g* for 2 min. The supernatant was discarded. This method was described as delipidation rather than fractionation as the proteins were collected in only a single pool (pellet) at the end of the process. The pellet was washed with cold acetone and the suspension was centrifuged at 4,000 *g* for 5 min at 4 °C. The supernatant was discarded and this procedure was repeated one more time. The pellet was then dissolved in freshly prepared 8 M urea in 25 mM NH_4HCO_3 (Sigma-Aldrich) buffer. The mixture was centrifuged at 4,000 g for 5 min at 4 °C, the supernatant was kept and the insoluble sediment was discarded. The protein concentration of the supernatant was determined using the BCA method[28]. It is important to note that although this is a standard method, the use of urea can interfere with downstream processes including tryptic digests and gel electrophoresis resulting in poorly resolved gels.

3.8 Generic in-gel protein fractionation and digestion workflow

3.8.1 1D SDS-PAGE

The universal 1D SDS-PAGE procedure adapted in this thesis for protein separation (also referred to as in-gel separation of proteins) was based on its established description[117] and subsequent refinements[118,119,236,237]. Briefly, the resolving self-cast 10% SDS-PAGE gels were prepared by mixing 3.85 mL of deionised water (Millipore) with 2 mL of 40% acrylamide / 2% bisacrylamide mix (final concentrations – 10% acrylamide / 0.5% bisacrylamide), 2 mL of 1.5M tris (pH 8.8), 80

µL of 10% SDS, 80 µL of 10% ammonium persulfate and 8 µL of Tetramethylethylenediamine (TEMED, Bio-Rad), which was then poured into a casting chamber (Mini-PROTEAN, Bio-Rad) and left to set after overlaying the gel with isopropanol. The isopropanol was then removed and the stacking 4% SDS-PAGE gels were prepared by mixing 3.15 mL of deionised water (Millipore) with 0.5 mL of 40% acrylamide / 2% bisacrylamide mix (final concentrations – 5% acrylamide / 0.25% bisacrylamide), 2 mL of 0.5 M tris (pH 6.8), 50 µL of 10% SDS, 50 µL of 10% ammonium persulfate and 5 µL of TEMED, and then poured on top of the resolving gel followed by insertion of plastic comb to mould the loading wells. The gels were left to polymerise for 12 h at RT before loading the wells with known quantities of protein samples. A stock solution of 10× running buffer was prepared by mixing 30 g of Tris, 144 g of glycine, 10 g of SDS and made up to 1L solution using de-ionised water (Millipore) before storage at RT. A 5× sample buffer stock was prepared by mixing 1.75 mL of 0.5M Tris (pH 6.8), 4.5 mL of glycerol, 0.5 g of SDS, 1 drop of 2.5mg/mL solution of bromophenol blue. The solution was brought up to 10 mL, then aliquoted into 500 µL and stored at -20 °C until use.

To load the gel, a desirable quantity of protein in the sample was aimed at approximately 0.1 µg/µL for each stained protein band[118] in either self-cast or precast polyacrylamide gel (Mini-PROTEAN® TGX™ Precast 4-15% Gels, Bio-Rad). A protein standard (Precision Plus Protein™ Dual Xtra, Bio-Rad) was loaded into to the first well of every gel for the estimation of the protein molecular weight. Protein concentrations of bovine serum albumin (BSA) standard (Pierce) and the samples were adjusted so that suitable amounts of protein could be loaded onto the gel with freshly prepared 20mM solution of dithiothreitol (DTT) in liquid chromatography mass spectrometry (LC-MS) water with 1× SDS sample buffer. Gel loading preparations were made in 1,500 µL (or 500 µL, where appropriate) Eppendorf tubes and then vortexed briefly before being centrifuged at 4,000 *g* for 6 s. The tubes containing the sample buffer and sample were heated at 56 °C for 3 min to unfold proteins prior to loading onto the gel and electrophoretic separation of proteins at a fixed voltage (150-180V) for approximately 45 min in a vertical electrophoresis chamber (Mini-PROTEAN Electrophoresis Cell, Bio-Rad) powered by PowerPac™ Universal power supply from Bio-Rad.

The gels were stained with Coomassie brilliant blue (EZ-Run™, Protein Gel Staining Solution, Fisher Scientific) according to the manufacturer's instructions and then photographed using a handheld camera (5.7-inch Quad HD Super AMOLED®, Samsung; or New 8-megapixel iSight camera with 1.5µ pixels with Optical image stabilisation, iPhone 6, Apple Inc.).

Gel bands were excised into a clean 1,500 µL Eppendorf tube and de-stained using 50% acetonitrile (ACN) (Optima®, Fisher Scientific) in 25mM NH_4HCO_3 accompanied by agitation at 750 rpm for 20 min at RT. This procedure was repeatedly and alternated with washing the gel bands with 25

mM NH_4HCO_3 buffer. Once de-stained, final washing of the gel bands was performed using LC-MS grade water followed by incubation for 20 min at RT. The water was discarded and the gel bands were cut into approximately 1 mm^3 pieces using a 10 µL pipette tip. Gel bands were dehydrated by adding 100% ACN and agitating at 750 rpm for 10 min at RT prior to drying in a vacuum centrifuge (SpeedVac Concentrator Christ® cat. No. RVC 2-33 IR), for 10 min.

In-gel proteins were reduced in order to break disulphide bonds and alkylated to prevent the bonds re-forming as originally described elsewhere[238]. Briefly, freshly prepared 10 mM DTT (Sigma-Aldrich) in 25 mM NH_4HCO_3 buffer was added sufficiently to cover the vacuum dried gel pieces and agitated at 750 rpm for 45 min at 56 °C. Freshly prepared 55 mM iodoacetamide (IAM) (Sigma-Aldrich) in 25 mM NH_4HCO_3 was added to the sample so the volume added was twice that of the DTT added in the previous step and agitated for 30 min at RT in the dark. The reagents were washed off with 25 mM NH_4HCO_3 buffer with agitation for 5 min at RT, before centrifuging briefly and discarding the supernatant. Gel bands were then dehydrated using 100% ACN and agitated at 1400 rpm for 10 min at RT. The entire supernatant was discarded prior to drying the gel pieces in a vacuum centrifuge (SpeedVac Concentrator Christ® cat. No. RVC 2-33 IR), for 20 min.

Vacuum-dried gel pieces were incubated on ice for 5 min before adding 0.005 µg/µL solution of freshly prepared ice-cold working solution of trypsin (Trypsin Gold, Mass Spectrometry Grade, Promega) in 50 mM NH_4HCO_3 buffer enough trypsin buffer to cover the dry gel pieces[239] and left incubating for a further 30 min until the entire enzyme solution had entered the gel pieces. Gel pieces were then covered in 50 mM NH_4HCO_3 buffer and left to incubate for 16 hrs at 37 °C on an agitator at 300 rpm. Digestion was stopped by adding 100 µL of 5% formic acid (FA) (Sigma-Aldrich). Peptide extraction was performed by agitating the gel pieces at 1,000 rpm for 15 min at RT. The peptide-containing supernatant was collected into a clean 0.5 ml low binding Eppendorf tube. Gel pieces were further washed by adding 5% FA in 50% ACN and agitating at1000 rpm for 15 min, before collecting the supernatant. Gel bands were further extracted by adding 100% ACN and agitation at 1000 rpm for 15 min at RT. The entire supernatant was collected and then completely vacuum-dried prior to reconstitution in 10 µL of 0.1% trifluoroacetic acid (TFA) (Sigma-Aldrich) / 2% ACN followed by desalting of peptides.

3.9 Generic in-solution digestion of proteins workflow

The method adapted here was based on the one established by Villén and Gygi[123]. Briefly, a known quantity of serum or plasma protein sample was thawed and kept on ice at 4 °C was reduced using 20 mM DTT (to achieve 5 mM final concentration) and the mixture was incubated for 1h at room temperature (RT). The mixture was alkylated using 55 mM IAM (to achieve 14 mM final concentration) followed by incubation for 20 min in the dark at RT. Alkylation was quenched using

DTT followed by further incubation for 5 min in the dark. The mixture was diluted with 25mM NH_4HCO_3 buffer followed by adding aqueous 70 mM $CaCl_2$ (to achieve 10 mM final concentration). Trypsin (Promega) was added at enzyme to substrate (protein concentration of sample) ratio of 1:50. The contents were incubated for 16 h at 37 °C with gentle agitation and then cooled to RT. Digestion was stopped using 10% TFA before vacuum concentrating the contents to dryness. The dried peptides were reconstituted in aqueous 0.1% TFA/2%ACN before desalting.

3.10 Generic desalting of tryptic peptide digests

It is often necessary to remove salts and particulate matter including excess of trypsin from peptide digests prior to analysis to prevent blockage of nanoLC columns and also to reduce noise artefacts of MS spectra[240-242]. Desalting of tryptic peptide digests was optimised and performed using either octadecyl carbon chain (C18) pipette tips (ZipTip® Pipette Tips, Millipore, or Pierce C18 Tips, Thermo Fisher Scientific) depending on the filter capacity according to manufacturer's instructions. Briefly, the desalting pipette tip was conditioned using a solution of 50% ACN/0.05% trifluoroacetic acid (TFA) in LC-MS grade water and then equilibrated with 2% ACN/ 0.1% TFA in LC-MS water. After carefully and gently pipetting the entire sample up and down the membrane for at least 10 times, the membrane was washed with 2% ACN/0.1% TFA in LC-MS water. The peptides were eluted using 70% ACN/0.1% TFA in LC-MS water, vacuum dried and then reconstituted in 10 µL of 2% ACN/0.1% FA in LC-MS grade water and transferred into a polypropylene autosampler vial for nanoLC-nanoESI-MS/MS analysis.

3.11 nanoLC-nanoESI-MS/MS

3.11.1 Chromatography

Peptide spectral data from approximately 400 ng – 1 µg of injected tryptic peptides per sample were generated using nanoLC-nanoESI-MS/MS on a TripleTOF® 5600+ instrument. Peptides were separated by performing reversed-phase chromatography using an Eksigent ekspert™ nanoLC 400 System directly coupled to the MS/MS instrument. The LC platform was setup in a trap and elute configuration with a 10 mm × 0.3 mm trap cartridge packed with ChromXP C18CL 5 µm 120 Å material and a 150 mm × 75 µm analytical column packed with ChromXP C18 3 µm 120 Å (Eksigent Technologies, Dublin, CA). The mobile phase solvents were composed of mobile phase A: water/0.1 % FA; mobile phase B: ACN/0.1% FA; and mobile phase C: water/2% ACN/0.1% FA. Trapping was performed in mobile phase C for 5 min at 5 µL/min followed by an elution configuration across a 90 min gradient using mobile phases A and B at a conserved flowrate of 300 nL/min. The proportions of both solvents were adjusted at specified time-points of 0, 60, 65, 70, 79, 80 and 90 min corresponding to 98, 60, 35, 5, 5, 98 and 98 % of solvent A, and 2, 40, 65, 95, 95, 2

and 2 % of solvent B, respectively. To minimise retention time drift, the analytical column was maintained at 40°C.

3.11.2 Data dependent acquisition (DDA)

After chromatography, peptides were injected into the instrument and analysed by DDA. The DDA mode of the instrument was set to obtain high resolution (30,000) TOF-MS scans over a mass range of 350-1350 *m/z*, followed by up to 40 (top 40) high sensitivity MS/MS scans of the most abundant peptide ions per cycle. The selection criteria for the peptide ions included intensity greater than 150 cps and charge state of 2-5. The dynamic exclusion duration was set at 12 s to account for the difference in chromatographic peak width. Each survey (TOF-MS) scan lasted 0.25 s and the product ion (MS/MS) scan lasted 0.05 s resulting in a total cycle time of 2.3 s. The ions were fragmented in the collision cell using rolling collision energy, and collision energy spread (CES) was set to 5. The collected peptide ion fragmentation spectra were stored in .wiff format (SCIEX).

3.11.2.1 DDA data processing

3.11.2.2 Overview of DDA data processing strategy

Because of the genome of sheep being incompletely sequenced or annotated, proteins were identified by matching tryptic peptides against a composite protein sequence database of sheep, goat and ox using ProteinPilot™ in the first instance in order to capture homologous sequences. The inclusion of protein sequences from related species is a helpful strategy when exploring and establishing foundation proteogenomics data to identify known or novel genes of the non-model study subject — in this case sheep[6-8,20,189,243,244]. Mascot [200] search was subsequently conducted using a sheep only sequence database to identify high-scoring proteins and PeptideShaker [245] to verify protein identifications from the primary search data.

3.11.2.3 Primary protein sequence database searching

The acquired MS/MS data from the instrument were extracted and annotated with amino acid sequences from a custom built database using the Paragon™ Algorithm[164] (ProteinPilot™ Software 5.0, Revision Number: 4769, SCIEX). The custom database (55,108 sequences; 34,001,891 residues) with added common contaminants was assembled in FASTA format downloaded on 15 April, 2015 from a repository of non-redundant and predicted protein sequences of *Ovis aries*, *Bos taurus* and *Capra hircus* sourced from NCBI (National Center for Biotechnology Information – http://www.ncbi.nlm.nih.gov/protein/). Another *Ovis aries*-only custom database (27,393 sequences, 13,114,569 residues) with added contaminants from the common Repository of Adventitious Proteins, cRAP (http://www.thegpm.org/crap/), was assembled in FASTA format on 26 July, 2016 from UniProtKB (Universal Protein Resource Knowledgebase –

http://www.uniprot.org/). For ProteinPilot™ searches, the following settings were selected: Sample type: Identification; Cys Alkylation: Iodoacetamide; Digestion: Trypsin; Instrument: TripleTOF5600; Special Factors: Urea denaturation; Species: None; Search effort: Thorough ID; ID Focus: Amino acid substitution; Results Quality: Detected protein threshold [Unused ProtScore (Conf)] ≥ 0.05 with false discovery rate (FDR) selected. Annotations were only retrieved from UniProt during composite file searches (i.e. when a group of files were searched concurrently). The automatically generated Excel (Microsoft® Excel™ 2010, Microsoft Corporation) spreadsheet report from ProteinPilot™ output was manually inspected and then meticulously curated to filter out contaminants and protein identifications with 0 (zero) unused confidence scores. The protein-level FDR analysis report generated automatically in Excel by ProteinPilot™ was used to determine identification thresholds. FDR is the number of significant peptide matches in the decoy results divided by the number of significant matches in the target results. The report showed proteins identified at critical FDR, estimated FDR and numeric receiver operating characteristic (ROC) plots of the data from DDA experiments. The identification threshold was determined as the highest number of proteins identified corresponding to 1% critical FDR on the Global FDR fit in the report. Only proteins identified at FDR ≤1% with ≥ 2 high-scoring peptides were considered for protein lists and for comparative analysis in the first instance.

The .group file derived from ProteinPilot™ composite search of all samples analysed by DDA were used as a reference PSL with SWATH™ Acquisition MicroApp version 2.0 (SCIEX) data extraction. The resulting library file contained the following information required for targeted data extraction: UniProt accessions, stripped peptide sequences, modified peptide sequences, Q1 and Q3 ion detection, retention times, relative intensities, precursor charges, fragment types, scores, confidence and decoy results.

The .group file data in ProteinPilot™ were exported as calibrated Mascot generic format (.mgf) and mzIdentML (.mzid) format files. The .mgfs were further reformatted by an in-house Perl script (www.perl.org) or mgf repair tool (SCIEX) prior to loading via a Daemon application to Mascot search engine (Matrix Science, London, UK; version 2.5.1)[200]. Mascot was set up to search the same custom database that was used in ProteinPilot™ with the following search parameters: Type of search: MS/MS Ion Search; Enzyme: Trypsin; Fixed modifications: Carbamidomethyl (C); variable modifications: Deamidated (NQ), Oxidation (M); Mass values: Monoisotopic; Protein mass: Unrestricted; Peptide mass tolerance: ± 10 ppm; Fragment mass tolerance: ± 0.01 Da; Max missed cleavages: 1; Instrument type: ESI-QUAD-TOF, and the auto-decoy search option was selected. Peak list and identification data from the search were exported in a .dat format for further processing. Protein lists were exported in csv format for immediate data evaluation and curation to

remove contaminants in Excel spreadsheet. Only proteins identified with 2 or more high-scoring peptides were included for further evaluation.

3.11.2.4 Secondary protein sequence database search and protein identification

The .mgf and .dat or .mzIdentML files (from ProteinPilot™ after proprietary source code editing – *see Chapter 6*) were also loaded for protein identification using PeptideShaker Software[245]. Peak lists obtained from MS/MS spectra were identified using Mascot[200]. Protein identification was conducted against a concatenated target/decoy[246] version of the *Ovis arie*s (27284, 99.5%) complement of the UniProtKB, 27,411 (target) sequences. The decoy sequences were created by reversing the target sequences in SearchGUI. The identification settings were as follows: Trypsin with a maximum of 1 missed cleavages; 10.0 ppm as MS1 and 0.5 Da as MS2 tolerances; fixed modifications: Carbamidomethylation of C (+57.021464 Da), variable modifications: Deamidation of N (+0.984016 Da), Deamidation of Q (+0.984016 Da), Oxidation of M (+15.994915 Da), Pyrolidone from E (--18.010565 Da) and Pyrolidone from Q (--17.026549 Da), fixed modifications during refinement procedure: Carbamidomethylation of C (+57.021464 Da). All algorithm-specific settings are listed in the Certificate of Analysis available in the data files.

Peptides and proteins were inferred from the spectrum identification results using PeptideShaker version 1.13.0[245]. Peptide Spectrum Matches (PSMs), peptides and proteins were validated at a 1.0% FDR estimated using the decoy hit distribution. All validation thresholds are listed in the Certificate of Analysis available in the data files. Post-translational modification localisations were scored using the D-score[247] and the phosphoRS score[248] with a threshold of 95.0 as implemented in the compomics-utilities package[249]. A phosphoRS score above this threshold was considered as a confident localisation. Protein identification reports were exported in .xlsx format for evaluation and curation in Excel spreadsheet. Only proteins identified with 2 or more verified peptides were included for further evaluation.

3.11.3 Data independent acquisition (DIA)

Eluted peptides were subjected to cyclic data independent acquisition (DIA) using a generic SWATH-MS™ acquisition method (SCIEX) based on its earlier description[1]. The instrument was operated using a mass range of 100 msec for the survey scan (MS), followed by performing MS/MS on all precursors in a cyclic manner using an accumulation time of 0.1 s per SWATH-MS window (36 windows total, each 26 *m/z* units wide) for a total cycle time of 3.75 s. The above parameters allowed the collection of at least 6 data points for each chromatographic peak of a peptide to ensure a reasonably accurate quantitation. It should be noted that a typical quantitation requires approximately 14 data points over a peak to define it accurately, although this is probably not

pratical with a slower method like SWATH-MS. The lower peak definition is associated with increased variability of quantitation.

3.11.3.1 DIA data processing

The generated raw .wiff instrument files were concurrently imported into PeakView® v.2.2 Software (SCIEX) with SWATH-MS™ Acquisition MicroApp 2.0 (SCIEX) incorporated for spectral alignment and targeted data extraction alongside a loaded PSL that was earlier constructed generated from a composite of DDA experiments using sheep serum and plasma acquired on the same instrument.

PeakView Software® performed the mining of extracted ion chromatograms (XICs) and determined peak areas for the entire sample set. The XICs of fragment ions from MS/MS spectra of targeted proteins and peptides were generated by the SWATH-MS™ Acquisition MicroApp, which also integrated the peak areas from the SWATH-MS™ data files.

The parameters for SWATH-MS™ data extraction were set as follows: up to five peptides per protein, up to five transitions per peptide, peptide confidence level of 99%, include shared peptides, and extracted ion chromatograph (XIC) width of 50 ppm. The processed data were amenable to exportation as quantitative output for the peak area under the intensity curve for individual ions, the summed intensity of individual ions for a given peptide, and the summed intensity of peptides for a given protein, in .txt format.

3.12 Statistical analysis of the processed data

3.12.1 Identified protein lists

Proteins lists from DDA experiments were presented in spreadsheet and charts (Microsoft® Excel™ 2010, Microsoft Corporation) and then exported for analysis and visualised using Venny 2.1[250], jvenn[251] or BioVenn[252] software version 2007 – 2017, where appropriate.

Peptide and protein peak areas from DIA experiments were automatically passed to MarkerView™ Software version 1.3 (SCIEX) for visualisation, normalisation and statistical analysis. Protein identification entries were exported into Excel™ spreadsheet, which facilitated inspection and comparative analysis using BioVenn software.

3.12.2 Gene ontology (GO) – term and protein pathway analysis

The proteins identified by Mascot searches were subjected to gene ontology (GO) analysis using Protein ANalysis THrough Evolutionary Relationships (PANTHER) classification tool[220]. In the PANTHER tool, the gene entries were analysed by aligning them to *Bos taurus,* as the closest organism analogous to sheep because *Ovis aries* entries were not available. The results of GO-term

analysis of molecular function, biological process, cellular component, protein class and pathway analysis of the detected proteins were displayed in Excel™ charts.

CHAPTER 4

4.0 Characterisation of the circulating acellular proteome of healthy sheep using nanoLC-nanoESI-MS/MS analysis of serum

4.1 Abstract

Unlike humans, there is currently no publicly available reference mass spectrometry-based circulating acellular proteome data for sheep, limiting the analysis and interpretation of a range of physiological changes and disease states. The objective of this Chapter was to develop a robust and comprehensive method to characterise the circulating acellular proteome in ovine serum.

Serum samples from healthy sheep were subjected to shotgun proteomic analysis using nanoLC-nanoESI-MS/MS on a TripleTOF® 5600+, SCIEX instrument. Proteins were identified using ProteinPilot™ and Mascot software based on a minimum of two unmodified highly scoring unique peptides per protein at a false discovery rate (FDR) of 1% by searching a subset of the UniProtKB database. PeptideShaker searches were used to validate protein identifications from ProteinPilot™ and Mascot.

ProteinPilot™ and Mascot identified 245 and 379 protein groups (IDs), respectively, and PeptideShaker validated 133 protein IDs from the entire dataset. Since Mascot software is considered the industry standard and identified the most proteins, these were analysed using PANTHER classification tool revealing the association of 349 genes with 127 protein pathway hits.

These results demonstrated for the first time the feasibility of characterising the ovine circulating acellular proteome using nanoLC-nanoESI-MS/MS. This peptide spectral data contributes to a protein library that can be used to identify a wide range of proteins in ovine serum.

4.2 Introduction

The well-defined serum proteome of humans permits analysis and interpretation of a range of physiology changes and disease states[101,253], but not that of sheep in comparison . To date, the serum proteome of sheep is largely extrapolated from cattle, which can be inaccurate despite a 97% similarity in protein coding sequences[254] and different promoters driving the expression of specific proteins[255]. Characterisation of the serum proteome of sheep would therefore be useful to quantify disease in this species.

Sheep are a major production species, providing meat and wool, plus are used in a range of biotechnological and translational studies[256-260]. Despite this, relatively little is known about the responses of sheep to a variety of pathophysiological events. There is therefore a need to comprehensively characterise the proteins in ovine serum for better quantitative assessment of

disease processes. Blood is relatively easily collected from sheep[22,111,261-263], but comparatively only a small number of proteins have been identified, limiting the capacity to assess disease[111,264]. One problem to date is that protein sample preparation in published studies on sheep have been inadequate and have generally ignored the full conventions for reporting identified proteins from samples[265,266]. Consequently, data are lacking on optimised sample preparation approaches for shotgun proteomics workflows using more than one protein sequence search engine to explore the circulating acellular proteome of sheep. For example, the number of proteins identified by single laboratories using gel fractionation followed by MS from human plasma has been in the region of nearly 300 protein IDs)[27]. In 2005, LC-MS/MS data from multiple sample preparation techniques and protein sequence search engines for HPPP from 18 laboratories worldwide collectively identified 3,020 plasma proteins based on a minimum of 2-high-scoring peptides[73,267]. This number of protein IDs from HPPP studies was subsequently revised to 889[73,74]. A study that used RP-HPLC and LC-ESI-MS/MS to analyse and define the human baseline plasma proteome identified 200 proteins[216]. More recently, protein expression profiles of human plasma proteins using 1D SDS-PAGE coupled with nanoLC-ESI-MS/MS in a single laboratory identified 253 proteins after desalting of the peptides[268]. A similar approach to that used in the preceding study was considered attractive to be used in exploring the circulating acellular proteome of sheep.

In this chapter, nanoLC-nanoESI-MS/MS was used to analyse peptide samples derived from healthy sheep following 1D SDS-PAGE and in-solution digestion.

4.3 Methods

4.3.1 Animal care, sample collection, storage and preparation

Serum samples of approximately 2 year-old healthy adult female Merino sheep (*n*=6) with ear tag identification numbers 473, 413, 463, 471, 476 and 478 belonging to an experimental colony at QUT and the Australian Red Cross Blood Service (ARCBS) were obtained for the development and optimisation of a comprehensive proteomic approach for interrogating the circulating acellular proteome. The sheep were reared according to established standard operating procedures, described elsewhere[10], in order to minimise variables that are known to modify the proteome such as breed, and physiological status for example stress, season, lactation, nutrition, health, pregnancy and oestrus status. Sample aliquots of 500 µL were stored in 1.5 mL Eppendorf tubes at -80 °C at the ARCBS, Brisbane. The samples were transferred to the wet laboratory at the Molecular Genetics Research Facility (MGRF) within Central Analytical Research Facility (CARF), QUT for processing. Processed samples were analysed at the Proteomics and Small Molecule Mass Spectrometry laboratory at CARF, QUT.

4.3.2 Experimental layout and data collection

In order to characterise the serum proteome of sheep, two universal sample preparation strategies described in Chapter 3 for shotgun proteome analysis were employed in three paired sets of experiments (first, second and third), using in-gel (1D SDS-PAGE) and in-solution protein digestion of serum samples. This was followed by peptide analysis by nanoLC-nanoESI-MS/MS.

4.3.2.1 1D SDS-PAGE of normal sheep serum workflow

As a pilot study, an acetone precipitated serum sample obtained from one sheep (Sheep ID 473) was processed and subjected to 1D SDS-PAGE to ascertain the feasibility of obtaining protein identification data as a basis for constructing a peptide spectral library in future (first in-gel digestion). In order to determine the optimum amount of serum protein to load, 2, 10 and 22 µg of protein were run in separate wells of the same gel. To determine the amount of protein that needed to be loaded on a gel for protein bands to be visualised after using EZ-Run protein stain, 250, 500 and 2500 fmol of bovine serum albumin (BSA) protein were loaded in separate wells of another gel and run.

In order to increase the protein coverage, a fraction of acetone precipitated serum sample from Sheep ID 473 was subjected to 1D SDS-PAGE in two gels run concurrently (second in-gel digestion). One gel was loaded with 50 µg and 100 µg of protein in adjacent lanes and the second gel was also loaded with 50 µg, 100 µg and 50 µg in adjacent lanes.

In all the 1D SDS-PAGE experiments above (first and second in-gel digestions), individual gel lanes were cut out and divided into 12 portions in a ladder-like fashion and individually digested so that the sum total of the proteins in the sample were submitted for analysis, even portions of gel that did not appear to contain proteins. This approach enabled the analysis of the entire proteome spread over a larger number of injections onto the MS instrument.

In order to determine the effect of the quantity of protein loaded, acetone precipitation and a protease inhibitor on protein coverage, pooled serum samples from six healthy sheep (Sheep IDs 413, 463, 471,473, 476 and 478) were processed and subjected to 1D SDS-PAGE in three gels (third in-gel digestion). The samples utilised consisted of crude protein (200 µg and 100 µg) in one gel with each lane cut out and divided into 6 portions and individually digested. One lane in a second gel loaded with 100 µg of crude serum with a protease inhibitor had 6 portions of gel cut and digested, another lane loaded with 100 µg of acetone precipitated serum protein with a protease inhibitor had 12 portions of gel cut and digested and a third lane loaded with 100 µg of crude serum without a protease inhibitor had 6 portions of gel cut and digested. A third gel was loaded and run identically as the second gel; the first lane loaded with 100 µg of crude serum with a protease

inhibitor had 14 portions of gel cut and digested, the second lane loaded with 100 µg of acetone precipitated serum protein had 12 portions of gel cut and digested and a third lane loaded with 100 µg of crude serum without a protease inhibitor had 6 portions of gel cut and digested.

4.3.2.2 In-solution digestion of sheep serum workflow

As a pilot study, 10 µg of acetone precipitated serum sample obtained from one sheep was subjected to in-solution digestion to ascertain the feasibility of obtaining protein identification data as a basis for protein quantitation in future (first in-solution digestion). In order to determine the effect of using unfractionated sample on protein coverage, a fraction of 20 µg of crude serum sample from the sheep used in the first in-gel digestion was subjected to in-solution digestion and analysed (second in-solution digestion). A third experiment utilised 100 µg of pooled crude serum samples from all six sheep (Sheep IDs 473, 413, 463, 471, 476 and 478) for in-solution digestion in order to determine the effect of using a higher quantity of protein substrate on protein coverage (third in-solution digestion).

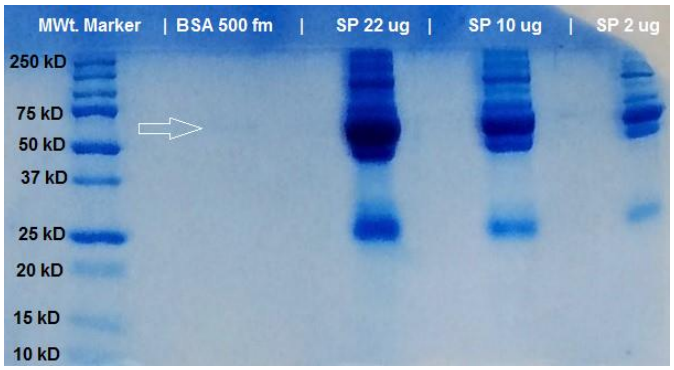
4.3.3 Data archiving

The mass spectrometry data along with the identification results were deposited to ProteomeXchange Consortium[231] via the proteomics identifications (PRIDE) partner repository[269] with the dataset identifiers PXD004989 and DOI: 10.6019/PXD004989. The URL is https://www.ebi.ac.uk/pride/archive/projects/PXD004989 and can be accessed with username reviewer99399@ebi.ac.uk and password QBFFTGzl.

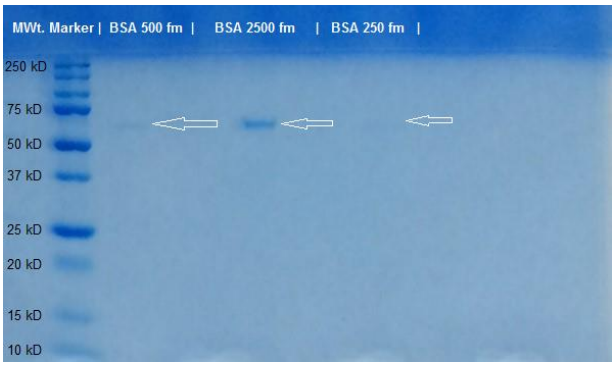
4.4 Results

4.4.1 1D SDS-PAGE

The results of the first, second and third in-gel digestion workflows are presented in Figures 4.0, 4.1 and 4.2, respectively. The details of the individual gels are provided in the figure captions. Except for Gel B in Figure 4.0, the protein sample lanes of all the other gels were subjected to in-gel digestion followed by nanoLC-nanoESI-MS/MS to identify proteins. Protein IDs were obtained using ProteinPilot™ to search a UniProtKB composite database of *Ovis aries, Bos taurus and Capra hircus* with a results quality of FDR ≤1%; ≥ 2 peptides for a protein to be considered confidently identified as the highest scoring member of the protein group. The Pro Group™ Algorithm in ProteinPilot™ assigned one protein the best confidence possible (unused score) among protein isoforms, which enabled protein subset differentiation, as well the suppression of false positives for protein-grouping analysis[270]. The results were therefore based on protein group identifications presented as protein identifications (IDs).

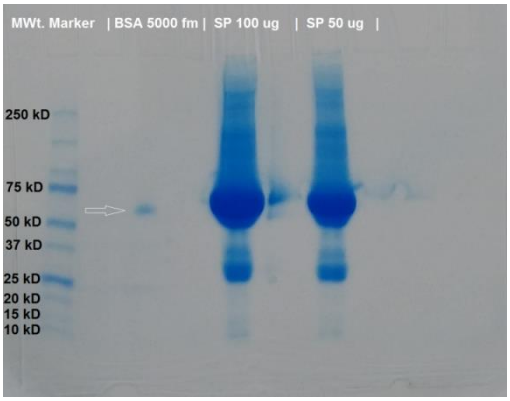


↑. Gel A

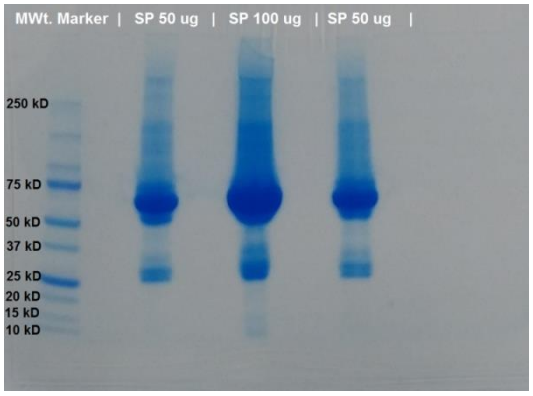


↑. Gel B

Figure 4.0. Coomassie-stained 1D SDS-PAGE gels used in first in-gel digestion. Fractions of acetone precipitated serum protein from one healthy sheep (Sheep ID 473) were loaded alongside bovine serum albumin (BSA) in Gel A. Gel A suffered a handling artefact to the top right corner of the gel. The leftmost well of both gels were loaded with 4 µL of a protein molecular weight standard (Precision Plus Protein™ Dual Xtra, Bio-Rad Laboratories). One well in Gel A was loaded with 500 fm of BSA standard; other three wells were loaded with 22 µg, 10 µg and 2 µg each of sheep serum protein sample, that yielded 120, 19 and 41 protein identifications, respectively. After the molecular weight standard, other three in Gel B were loaded with 500 fm, 2500 fm and 250 fm each of BSA standard. Arrows show BSA standard. **Key**: kD = kiloDalton; BSA= bovine serum albumin; sheep SP= serum protein; ug=µg; MWt= Molecular weight marker fm=fmol.

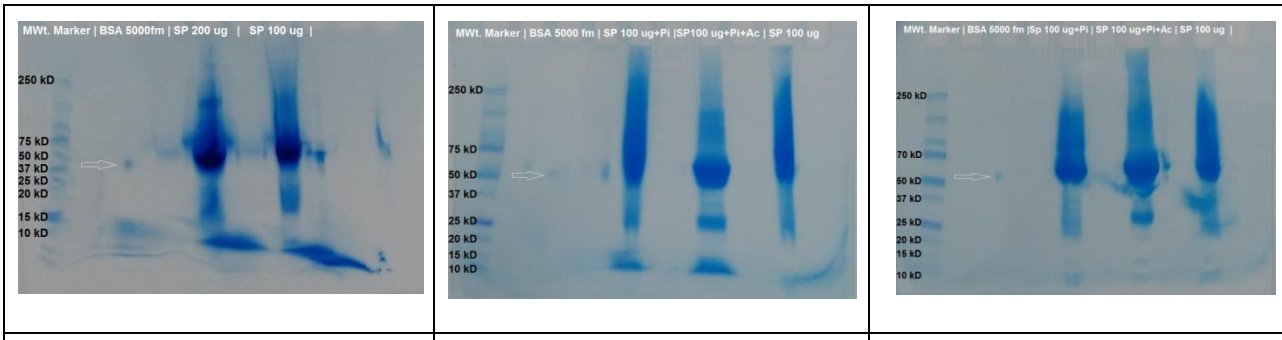


↑. Gel A



↑. Gel B

↑. **Figure 4.1.** Coomassie-stained 1D SDS-PAGE gels used in the second in-gel digestion. Fractions of acetone precipitated sheep serum protein samples from one healthy sheep (Sheep ID 473) were used. One well in Gel A was loaded with BSA standard (arrow); two other wells had 100 µg and 50 µg of protein each that yielded 151 and 127 protein IDs, respectively. Three wells in Gel B were loaded with 50 µg, 100 µg and 50 µg each of protein that yielded 144, 156 and 141 protein IDs, respectively. **Key**: kD = kiloDalton; BSA= bovine serum albumin; sheep SP = serum protein; ug = µg; MWt = Molecular weight; fm = fmol. The arrow in gel A points at the BSA standard band.



↑. Gel A	↑. Gel B	↑. Gel C

Figure 4.2. Coomassie-stained 1D SDS-PAGE gels used for the third in-gel digestion. Fractions of pooled serum protein samples from six healthy sheep were used. Gel A was loaded with 200 µg and 100 µg of crude serum separately yielding 40 and 38 protein IDs, respectively. Gels B and C were loaded identically with three-100 µg of serum protein that were treated as follows: one well had crude serum with a protease inhibitor (SP 100 µg+Pi), the second had acetone precipitated serum and a protease inhibitor (SP 100 µg+Pi+Ac) and the third had crude serum only (SP 100 µg), each category yielded 162, 143 and 114 protein IDs, respectively. **Key**: kD = kiloDalton; BSA= bovine serum albumin (white arrow); sheep SP= serum protein; ug = µg; MWt = Molecular weight; fm = fmol; Ac = acetone precipitated serum; Pi = protease inhibitor (cOmplete™, Roche).

The protein ID results of the first, second and third in-gel and in-solution digestions are summarised in Table 4.0. The Pro Group™ Algorithm in ProteinPilot™ assigned one protein the best confidence possible (unused confidence) among protein isoforms, which enabled protein subset differentiation, reduction of redundant protein IDs as well the suppression of false positives for protein-grouping analysis[270]. The results were therefore based on protein group identifications presented as protein identifications (IDs).

In the present set of experiments, proteins were identified by using peptide signatures to search custom-built protein sequence databases. Protein ID confidence was determined by the number of proteins that were assuredly accepted as correct, having been identified by two or more high-scoring peptides[271,272]. Overall, a total of 267 confident and unique protein groups were identified using ProteinPilot™ by searching a composite UniProtKB database after combining all the three in-gel digestion workflows (first, second and third in-gel digestions) from a total quantity of 1,284 µg of serum protein obtained from six healthy sheep. The UniProtKB entries for the identified proteins are presented in Appendix 4.0.

Table 4.0. The number of proteins identified by ProteinPilot™ Software from in-gel and in-solution digestion of healthy sheep serum samples by searching a composite database.

Experiment →	**First digestion**		**Second digestion**		**Third digestion**	
Digestion type	In-gel	In-sol	In-gel	In-sol	In-gel	In-sol
Serum protein source	Ac	Ac	Ac	Crude	Ac + Crude	Crude
Total quantity of protein analysed	34 µg	10 µg	350 µg	20 µg	900 µg	100 µg
Number of protein IDs	120	25	241	100	182	32
Key: In-sol = In-solution; Ac = Acetone precipitated; IDs = Identifications						

4.4.2 In-solution digestion

A composite ProteinPilot™ search of all the three in-solution digestion workflow samples comprising of 130 µg of serum protein yielded a total of 102 protein IDs. The UniProtKB entries for these proteins are presented in Appendix 4.1.

A comparison between the protein ID list derived from combined first, second and third in-gel digestion (in-gel digestion workflow) and that of combined first, second and third in-solution digestion (in-solution digestion workflow) in BioVenn Software[252] which collectively yielded 284 protein IDs is presented in Figure 4.3. The UniProtKB entries of the 17 proteins that were exclusive to the in-solution digestion workflow (i.e. proteins were not detected by in-gel workflow) are A0A0F6QNP7, W5PSQ7, W5QH45, W5NQW9, G5E604, W5PZF0, W5NWX6, Q1KZF3, W5PJZ2, W5QDP8, W5PDR7, W5PN97, W5PXI6, F1N3Q7, C6ZP49, G3N346 and Q3SYR8.

A combined ProteinPilot search of the pilot data from one sheep and the additional data from five sheep for both in-gel and in-solution digestion workflows using a composite *Ovis aries, Bos taurus and Capra hircus* database yielded an overall outcome of 274 protein IDs. Based on comparison with previous studies and protein database resources[17,20,110,111,189,263,264,273-280], there were 67 known, 207 novel and 83 disease-associated serum proteins identified using this composite database whose UniProtKB entries are presented in Appendix 4.2, Appendix 4.3 and Appendix 4.4, respectively. The known proteins are those that have been cited in the literature and also have a confirmed status in UniProtKB. Novel proteins constitute those that previously appeared as predicted and proteins

that had hitherto been inferred by homology. Disease-associated proteins refer to proteins that are expressed or alter during pathology in sheep and other species.

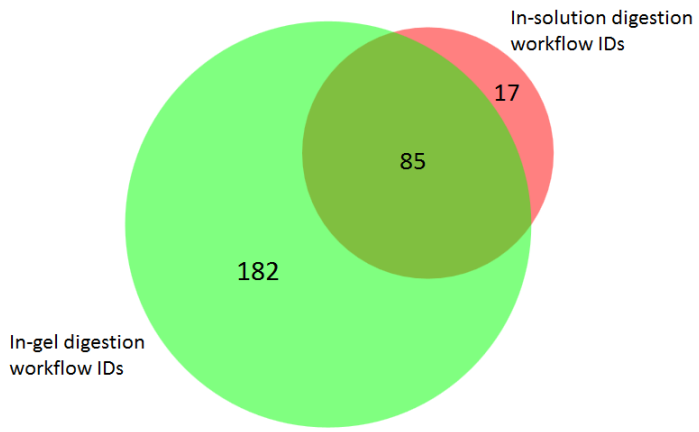


Figure 4.3. Comparison of lists of protein identifications (IDs) derived from in-solution versus in-gel digestion using BioVenn Software. Protein identifications were made by searching a composite database of *Bos taurus, Ovis aries* and *Capra hircus* using ProteinPilot™ Software. Only 17 protein IDs were exclusive to the in-solution workflow compared to 182 protein IDs exclusive to in-gel workflow.

4.4.3 Combined protein identifications from 1D SDS-PAGE and in-solution digestion of serum using ProteinPilot™ and Mascot database search engines and PeptideShaker search

Protein yields from a composite search of all the sample data from the three workflows (first, second and third in-gel and in-solution digestion) using a sheep-only UniProtKB database were as follows: ProteinPilot: 245 IDs; Mascot: 379 IDs and PeptideShaker: 133 IDs. Again, based on comparison with previous studies and protein database resources[17,20,110,111,189,263,264,273-280], and using the 379 Mascot protein IDs, there were 77 known, 302 novel and 83 disease-associated serum proteins identified using this sheep only database as listed in Appendix 4.5, 4.6 and 4.7, respectively.

The 379 protein IDs from Mascot search were used as a benchmark for further downstream analysis. Every sheep protein ID made in Mascot was mapped to a distinct gene. Of all the 379 protein IDs made by searching the sheep only UniProtKB database, only 70 proteins had been annotated based on sequence similarity to other species, whilst 305 proteins were uncharacterised. Of the 70 annotated proteins, only annexin A2 (P14639), serum albumin (P12303), transthyretin (B3SV56), nuclear receptor subfamily 1 group D member 1 (A2SW69) and insulin-like growth factor-binding protein 2a (Q29400) had been reviewed and therefore included in the Swiss-Prot subset of UniProtKB. The unreviewed, but named proteins are presented in Table 4.1.

Table 4.1. Unreviewed but named proteins in UniProtKB identified in serum of healthy sheep

14-3-3 Protein sigma	Gelsolin isoform B
Adaptor protein complex subunit beta	Glutathione peroxidase
Adenylyl cyclase-associated protein	Glyceraldehyde-3-phosphate dehydrogenase
Adiponectin	Growth hormone receptor variant h
Alpha-1-acid glycoprotein	Histone H2a
Alpha-1-antitrypsin transcript variant 1	Histone H2b
Alpha-2-HS-glycoprotein	Histone H3
Alpha-mannosidase	Histone H4
Amine oxidase	Importin subunit alpha
Angiotensinogen	Large tumour suppressor-like 1 protein
Antithrombin-III	L-Lactate dehydrogenase
Apolipoprotein E	MHC Class II antigen
Arginase	Monocyte differentiation antigen cd14
Aspartate aminotransferase	Olfactory receptor
Beta-a globin chain	Oxysterol-binding protein
Carbonic anhydrase 2	Pentaxin (pentraxin)
Carboxypeptidase	Peptidyl-prolyl cis-trans isomerase (PPIase)
Centromere protein C	Phosphodiesterase
Ceruloplasmin	Plasminogen
CGMP-dependent protein kinase	Polypeptide N-acetylgalactosaminyltransferase
Chitinase-3-like protein 1	Proteasome subunit alpha type
Clusterin	Proteasome subunit beta type
Coagulation factor IX	Protein-serine/threonine kinase
Condensin complex subunit 2	Protein-tyrosine-phosphatase
Conglutinin 1	Prothrombin
Corneodesmosome protein	Superoxide dismutase [Cu-Zn]
C-X-C motif chemokine	Thyroxine-binding globulin
Dipeptidase	Transaldolase
DNA polymerase	Tubulin alpha chain
Factor H	Tubulin beta chain
Fibrinogen alpha chain	Uricase
Fibulin-1	VH region chain
Fructose-1,6-bisphosphatase 1	

4.4.4 GO-term analysis of proteins identified in serum of healthy sheep

The 379 proteins identified by a composite Mascot search of the first, second and third in-gel and in-solution digestion of serum proteins from healthy sheep were subjected to GO-term analysis using PANTHER classification tool[220]. In the PANTHER tool, the gene entries were analysed by aligning them to *Bos taurus*, as the closest organism analogous to sheep because *Ovis aries* entries were not available. The PANTHER analysis resulted into 349 bovine aligned gene entries as listed in Appendix 4.8.

The results of GO-term analysis of molecular function, biological process, cellular component, protein class and pathway analysis of the detected proteins are provided in Figure 4.4. Looking at the molecular function domain of the proteins alone based on the GO-term results (Panel A), catalytic activity was dominant of the 264 function hits. From the protein IDs that had names, at least 27 of them were specifically classified as enzymes from protein database searches. It is evident from these results that there is a hierarchy in the biological processes of the 586 process hits (Panel B). The cellular component GO domain (Panel C) for serum from healthy sheep had 214 hits in total. The protein class GO domain (Panel D) had 386 class hits, with enzyme modulation topping the list. Among the 49 prominent protein pathways that were displayed in PANTHER from the analysed genes, 14 were represented by over 3.0% contribution to the revealed pathway pool (Panel E).

4.5 Discussion and conclusion

This chapter reports the development of a proteomics baseline profile of healthy sheep serum by analysing peptides derived from in-solution digestion and 1D SDS-PAGE using nanoLC-nanoESI-MS/MS. The major outcome was that 379 proteins were identified, compared for example to 42 proteins from serum of sheep with mild respiratory disease during peripartum period[111] and a single protein (serum amyloid A) in sheep with scrapie[264]. Both of these cited earlier sheep studies used two dimensional (2-DE) surface enhanced laser desorption/ionisation time of flight mass spectrometry (SELDI-TOF MS) and LC-MS/MS.

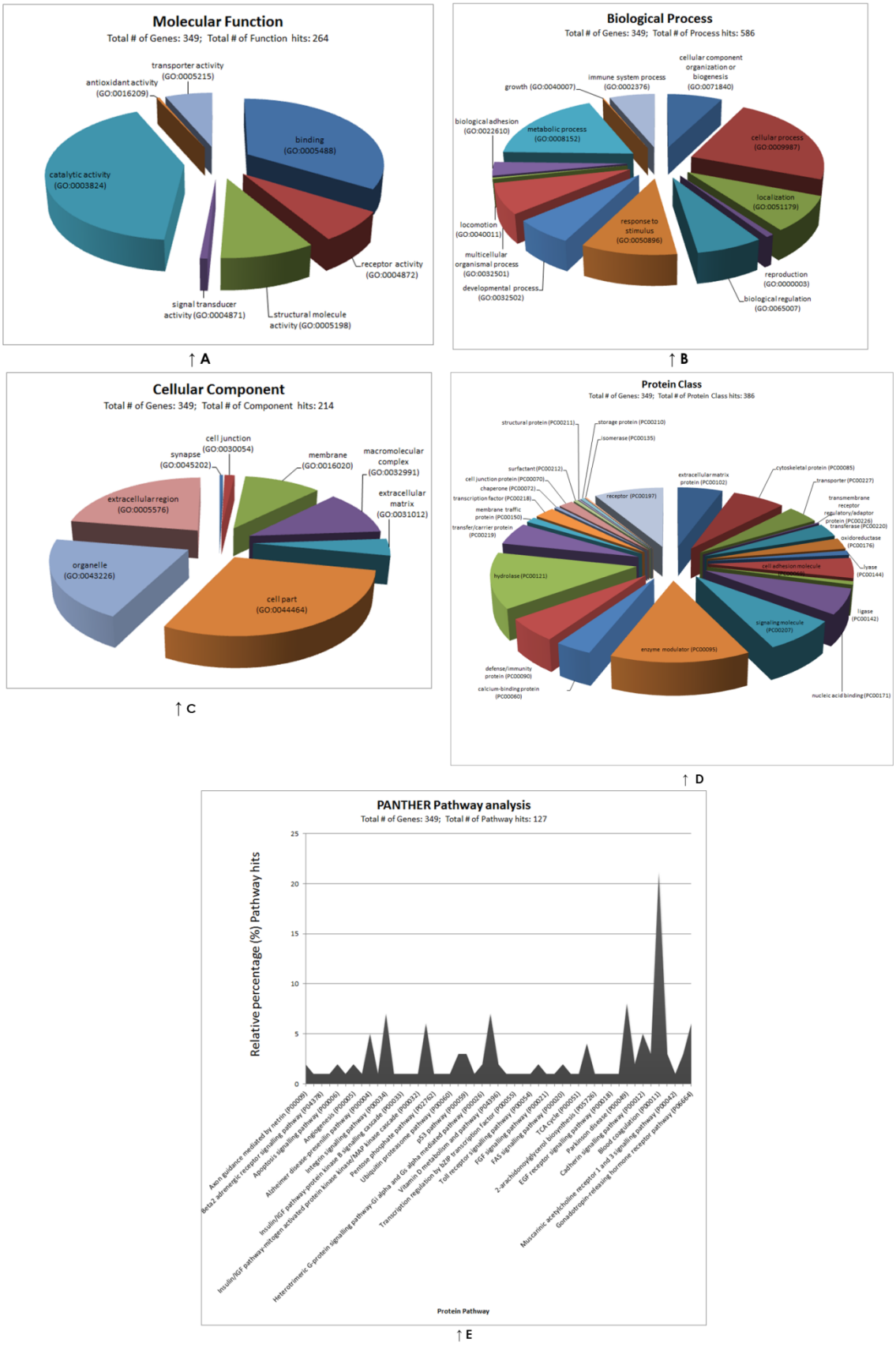


Figure 4.4. Gene Ontology (GO) and pathway analysis of Mascot protein IDs in healthy sheep serum. The GO – term domains are: A–molecular function, B–biological function, C– cellular component, and D–protein class. E–pathway analysis.

In species other than sheep, for example in various studies that analysed human sera, up to 490[101], 695[104], 800[107], 889[73,74], 1,105[78], 1,709[81], 1,929[19,83,84], 1,961[87], 1,955[85], 1,957[86], 2,392[92] and 9,085[109] proteins were identified using multidimensional separation coupled with MS. This suggests that with improved sample fractionation and adequate protein sequence database resources, there is scope to identify more proteins in the circulating proteome of sheep. There is also a report that assessed three different lots of foetal bovine serum by NanoLC-MS/MS analysis in which 79, 90, and 91 proteins were identified[281]. The preceding study recognised that there is variability in the protein content of different lots of foetal bovine serum – a commonly used growth medium for cell cultures, which affects the consistency of cell growth. The lot with a higher number of protein IDs was associated with higher cell growth rate[281]. Identification of these proteins is important clinically to determining health or altered physiology, such as stress[111].

The use of 1D SDS-PAGE in this study facilitated serum protein samples to be fractionated to reduce protein complexity prior to nanoLC-nanoESI-MS/MS analysis[282]. The first in-gel digestion experiment enabled the determination of the quantity of protein from samples and the amount of the BSA standard that needed to be loaded onto the gel to ensure that protein bands were visible and clearly defined (Figure 4.0). Loading a larger quantity of protein onto the gel was necessary to discover as many proteins as possible using DDA[1]. However, the 2 µg lane yielded 41 protein IDs in the first in-gel digestion (Figure 4.0), while the 10 µg-lane yielded 20 protein IDs and the 22 µg lane yielded 121 protein IDs. The 10 µg lane was analysed initially and the 2 µg and 22 µg lanes were analysed 6 weeks later once the extractions had been optimised and the instrument tuned.

The second in-gel digestion (Figure 4.1) increased the protein coverage by loading more protein into the gel wells using a fraction of the acetone precipitated serum sample used in the 1st in-gel digestion. The 100 µg (2 replicates) and 50 µg (3 replicates) protein loads in the 2nd in-gel digestion workflow yielded comparable numbers of protein IDs for each of the loaded quantity of protein. This suggests that reproducibility of the amount of protein loaded into the gel lanes had been achieved[282]. The second in-gel digestion was an improvement of the 1st in-gel digestion by having replicates and having increased quantities of loaded protein per lane, using the same serum sample of 1st in-gel digestion from Sheep ID 473.

The 1D SDS-PAGE preparation of one gel in the third in-gel digestion had a number of visual artefacts (Figure 4.2). The distortion in the 10-15kD region of Gel A could have been attributed to a defect in the gel possibly due to inconsistency in gel polymerisation creating artefact bands[117], overloading and/or the presence of a pocket between the gel and the cassette housing that allowed the protein samples to leak out the gel[283]. This could have also contributed to the low number of protein yields made from this gel, compared to the 100 µg × 2 lanes in Gels B and C. This could

be because considerably less protein was extracted from the gels in some samples than others. A couple of variables were also introduced in this experiment, in addition to the quantity of proteins loaded on to the gel wells as planned. The analysis of fractionated crude serum that had a protease inhibitor yielded a higher number of protein IDs, compared to the acetone precipitated sample that also had a protease inhibitor. This suggests that a considerable number of proteins were present in the acetone precipitation supernatant that was discarded. The discardment of the supernatant from acetone precipitation is a routine practice during generic or universal sample preparation for proteomic analysis[24].

As for the in-solution digestion workflow, the number of protein identifications from analysing 100 µg of crude serum protein was low when compared with 20 µg. The sample for the first in-solution digestion using 10 µg of acetone precipitated serum that was drawn from one healthy pilot sheep yielded even a lower number of protein IDs. This sample was prepared and analysed at the same time as the 10 µg sample of the first in-gel digestion discussed earlier. Protein detection was therefore likely to have been affected by unoptimised experimental processes at the time prior to running on the MS instrument. The number of protein IDs from second in-solution digestion using 20 µg of crude serum from the same sheep was considered substantial and comparable to those of other studies[77,79,96,100,102,105,111,113]. Unexpectedly however, the third in-solution that utilised 100 µg of pooled crude serum from six sheep under the same experimental conditions much lower yield of protein IDs. It is thought that this result was possibly due to the inhibition of trypsin by the presence of intravenous agents in the pooled sample from the anaesthetic cocktail used to anaesthetise the sheep, as this was not the case with the pilot sheep sample in which the sheep was not anaesthetised during sample collection.

BioVenn Software[252] was utilised for visualisation of the data presented in Figure 4.3. This tool enabled the comparison of a protein identification list derived from in-gel digestion with that from in-solution digestion by displaying the data in an area-proportional Venn diagram. It showed protein IDs that were exclusive to in-solution and in-gel, and those common between the two digestions. Of the 17 protein IDs that were exclusive to in-solution digestion workflow, five were mapped to the ox, two to the goat and the remaining 10 IDs were for sheep. Despite having known genes, the vast majority of the identified proteins were either uncharacterised or unreviewed in UniProtKB. Another interesting observation was that the combined list of 284 protein IDs from in-gel and in-solution digestion displayed in BioVenn Software was marginally higher than the 274 IDs from a composite ProteinPilot™ search of the same datasets. It is likely that the subsequent composite ProteinPilot™ search helped to further group proteins, thereby improving the confidence of protein IDs by minimising false protein identifications – a known challenge when searching a multi-species

protein database to identify proteins.

The data showed different numbers of protein identifications from different samples which was not entirely unexpected as repeat injections of the same sample will give slightly different results every time due to the stochastic nature of the peak picking process. In that sense, the individual data presented may have not been adequetly useful. It can be argued that perhaps not enough experiments were performed to judge the true variability of the different sample preparation methods, given the potential variability of the acquisitions. It is necessary to point out the level of potential variability here, since this could prove critical in Chapter 7 that deals with label-free quantitation. On the other hand, repeat injections are not necessarily considered critical for the purposes of constructing a spectral library, given the inherent stochastic nature of the peak picking process of DDA experiments. Nevertheless, prior to conducting the experiments serious attempts were made to assess how reproducible sample preparation and data acquisition were, in particular from in-solution digests of BSA standards and replicate analyses of ß-galactosidase from *E. coli* for in-house quality control of the instrument. This was done in order to assess the satisfactory reproducibility of repeat injections of individual samples (data not shown).

The comparatively conservative number of protein IDs made by PeptideShaker search is because the protein entries were identified using only validated unique peptides. This stringent feature of using only unique peptides for protein identification is not readily obvious in either ProteinPilot™ or Mascot in a user-friendly manner, whose protein ID entries were based on at least two high-scoring peptides per protein only, on the assumption that the peptides were unique to the protein. For this reason, PeptideShaker was selected as the most appropriate protein ID validation software that was used for the ultimate construction of the PSL of the circulating acellular proteome of sheep in Chapter 6.

In this preliminary study however, the results from Mascot search were embraced and utilised for further analysis because this software platform has been used in previous studies on sheep[111] and it is widely used by the proteomics community. Mascot software is considered the industry standard, because it implements a vast array of applications necessary for protein identification[284]. As of September, 2016, the 379 protein IDs complete with UniProtKB accessions was probably the highest number of sheep serum proteins to date using nanoLC-nanoESI-MS/MS. This study can therefore be considered the first to provide a comprehensive MS/MS protein sequence data of serum proteins of normal sheep and by contributing to the efforts of annotating genes and charactering sheep proteins. Despite most of the protein IDs not being characterised in UniProtKB, their mapping to known genes and the available mass spectrometry-derived peptide sequence data alongside verification on more than one software platforms, constitute strong supportive evidence

that the identified proteins do exist. The downside of the Mascot search is that it does not provide a user-friendly protein sequence output that can be readily tabulated as in the case of ProteinPilot™ IDs. For this reason, only protein names and UniProtKB entries were utilised mostly for the purposes the present study. The different search engines produced different results by searching the same database possibly because of the disparate manner in how their algorithms work. In addition, the sheep genome has not been fully defined compared to that of humans or mice by which these software programs were benchmarked upon.

Regarding GO-term analysis, the significance of many of the enzymes that dominated catalytic activity in the molecular function domain (Figure 4.4 A), remains to be documented in sheep, but the functions of some are known. For example, adenylyl cyclase-associated protein regulates cofilin function, the actin cytoskeleton, and cell adhesion[285]. Alpha-mannosidase participates in glycoprotein synthesis and endoplasmic reticulum quality control[286]. It has been reported to be downregulated in locoweed *(Oxytropis sericea)* in sheep[287,288], for example. The functions of other identified enzymes drawn from[189,289-318] are provided in Appendix 4.9.

Serum samples of healthy adult female Merino sheep were utilised for this chapter. It is quite possible that a relatively low representation of the growth process domain in the biological process GO-term was because the serum samples were derived from adult sheep. Also, the cellular component fractions could possibly vary depending on the physiological status of the sheep – which remains yet to be determined and documented. It can be argued that hormonal changes and the influence of age contribute to observations of serum proteome profiles and this should be accounted for. For instance, studies in sheep have shown that diurnal variations in metabolic and stress-responsive hormones do occur[319].

In the present study, there were mechanisms in place to mitigate the effects of stress on the laboratory sheep. The sheep were reared together and acclimatised to their housing and handling by people as a standard management practice prior to blood sampling[10,15]. Also, there was no variation in calorie intake because feed was supplemented as required[10,15] in order to mitigate the well-established phenomenon of seasonal weight loss – a well-established major nutritional stress factor in sheep[320]. During agistment, there were wethers that belonged to other experiments of the research group, but there were no entire males to cause 'ram effect' that could have caused surges in reproductive hormones[321], for example. Nevertheless, gonadotropic activity would have occurred naturally in the ewes to cause hormonal changes[322], perhaps even with a synchronised hypothalamic-pituitary-ovarian axis in all the ewes, as this phenomenon is known to occur naturally[323]. All the sheep were approximately 2 years old and were therefore, practically in the same metabolic and physiological state during blood sampling. Also, the sheep belonged to an

ovine model of blood transfusion[14], so most preventable adverse attributes had been catered for.

It can be argued that it would have been better to apply the GO classification to the larger data set presented later in the thesis, rather than to the relatively smaller number of identifications in this chapter. On the contrary however, this prototype study is discrete in that it contains only protein identifications derived from analysing serum of healthy sheep. This can be considered to represent the baseline proteome. Later datasets contain protein IDs derived from analysing both serum and plasma, and also include data from naturally ill sheep and sheep with experimentally induced endotoxaemia, which would make it difficult to separately evaluate the GO classification profiles of protein IDs derived only from healthy individuals.

The knowledge from this prototype study has illuminated a considerable number of bovine-aligned gene entries associated with protein pathways that can be valuably exploited by animal model studies using sheep serum as their analyte. A downside of the present study is that no males were represented in the dataset. Future studies should take into account hormonal changes and be gender and age inclusive in order to capture broad aspects of the proteome that could have been missed.

Secondly, all of the experiments were DDA discovery experiments of single injections using 90 min (1.5 h) LC gradients. It can be argued that this may not have given optimal analytical depth, but it was adequate for feasibility studies to enable the construction of a peptide spectral library and the best match for the SWATH experiments that are detailed in Chapter 7. As long as method optimisation processes have been accomplished, there is no contraindication for not using single injections for building a spectral library. A longer LC gradient would have allowed for gaining a deep analysis of protein samples. However, from other published studies using the SCIEX workflow as in this thesis, a 90 min gradient has been popularly used, especially when building spectral libraries to be used by SWATH on the same instrument. And also, a longer LC gradient can be a very slow process, given how SWATH works. In future, a longer gradient and the use of repeat injections with exclusion lists should be included in the methods.

Thirdly in hindsight, if the analysis and discussion provided scope for inclusion of information on the number of individual peptides identified and the % of MS/MS spectra that led to peptide identifications, then the true value of this data and its analysis would be apparent in order to provide opportunities for further future method optimisation. These data can be mined to provide much greater detail on how efficient the acquisition process was and add some interesting discussion points. For example, information on the width of the chromatography peaks is useful and can be obtained from XIC of the MS survey scan. It would have been ideal to provide a more detailed comparison of the results, for example which proteins were common to which search engines, how many individual peptides were identified per protein. The downside of this is that presenting these

data here would exceed the word count requirements for this thesis. Nevertheless, this information is publicly available at https://tinyurl.com/y7697fde and can also be accessed and have the analysis reproduced from the PRIDE archive link provided in the methods section as per the recommended publication norms of proteomics datasets.

In conclusion, this chapter has demonstrated for the first time that it is feasible to identify several hundred sheep serum proteins using a traditional sample preparation approach followed by nanoLC-nanoESI-MS/MS analysis. By utilising the PANTHER tool, this serum-derived prototype of the ovine circulating acellular proteome revealed the association of 349 genes with 127 protein pathway hits. When used with protein quantitative data that includes plasma samples, these findings have the potential to be applied as the foundation for establishing the baseline normal ovine serum proteome for comparison with samples from sick sheep. Following on in Chapter 5 are descriptions of plasma and serum sample preparation methods that enhance peptide extraction and improve protein identifications.

CHAPTER 5

5.0 Strategies for enhancing peptide extraction from the liquid fraction of sheep blood for protein identification by nanoLC-nanoESI-MS/MS analysis

5.1 Abstract

Separation of proteins in a plasma or serum sample is a critical component of comprehensive proteomic analysis. In this chapter, several approaches to separating proteins were investigated. The protein yields from using plasma or serum as the analyte protein sample source were also compared. The protein fractionation techniques used included comprehensive acetone precipitation, partial organic precipitation with ACN, combinatorial peptide ligand library protein enrichment and off-gel fractionation. All fractionated samples were also subjected to 1D SDS-PAGE.

The protein fractions from the samples were then subjected to shotgun proteomics analysis and the resulting proteins were identified, as previously described in Chapter 3. Using Mascot IDs as the standard, combinatorial peptide ligand library protein enrichment yielded the most protein IDs, 439, which represented 207 additional protein IDs to those from undepleted samples. ACN precipitation workflow resulted in 376 protein IDs, with plasma contributing the highest number of exclusive protein IDs. Acetone precipitation yielded 136 protein IDs, with serum fractions yielding a higher number of protein IDs than plasma did. The least sensitive standalone technique was off-gel fractionation, which yielded 84 protein IDs, compared to 55 protein IDs from unfractionated serum.

These results represent the first comprehensive comparative analysis of various protein sample fractionation techniques for separation of proteins from the liquid fraction of sheep blood. Combinatorial peptide ligand library protein enrichment of plasma and serum samples was associated with the highest number of protein ID yields.

5.2 Introduction

Protein-level fractionation prior to subsequent separation using chromatography and analysis of the resulting tryptic peptides is a well-recognised strategy to enhance protein ID yields in shotgun proteomics workflows[67,116,324-327]. This chapter describes four protein sample preparation strategies to enhance peptide extraction from protein samples derived from the liquid fraction (the circulating acellular proteome) of sheep blood for protein identification, in addition to 1D SDS-PAGE. The paragraphs that follow provide some background to the application of these strategies.

Acetone precipitation of proteins in solution is a commonly used procedure to fractionate plasma and serum analytes prior to proteomic investigations[328]. Like other organic solvents, acetone causes precipitation through hydrophobic aggregation, by changing the solvation of the protein with

water[23]. The efficiency of any precipitating agent depends on the physicochemical characteristics of the protein[329]. In most cases, the supernatant fraction after precipitation is seldom required and is discarded[330]. The precipitate, containing concentrated proteins, is retained to be used for downstream proteomic analysis. It is therefore noteworthy that only a few previous studies have examined the usually discarded supernatant fraction[330]. As an aim of the present study, then, the comprehensive analysis of fractions of acetone-precipitated sheep plasma and serum would provide new insights into the evaluation of the efficiency of acetone precipitation. There is a downside to acetone precipitation in that it has been known to modify certain peptides that become evident after subsequent proteolysis of the precipitate fraction, which could compromise proteomics outcomes[331]. In addition, depending on the structure of individual proteins, precipitation is known to alter the protein composition of samples[329]. The intent to explore in detail what might still be left in the supernatant was a nested aim within the broader aim of examining the acetone-precipitation technique in comparison with other techniques.

Partial organic precipitation of proteins using ACN[332-336], followed by analysis of both fractions, is an innovative approach used to increase acellular circulating proteome coverage via a pseudo-depletion strategy of the most abundant proteins[337]. This approach is attractive because optimised ovine antibody-based protein depletion kits are still difficult to source. For instance, the ACN precipitation workflow of the method previously described by Mostovenko et al.[337] and that of Bluemlein and Ralser[338], which evaluated the effects of pH and protein concentration in human plasma samples resulting in substantial protein yields, could be used to explore samples from sheep. In one of the previous studies[337], a higher number of protein IDs was achieved following ACN precipitation at a pH of 5. A higher pH of 9 was not associated with any benefit with respect to depletion of abundant proteins, with the added disadvantage that the pellets were difficult to resuspend[337]. By adapting these earlier studies[337,338], it would be possible to ensure a deeper coverage of the circulating acellular proteome of sheep in the present study.

On the basis of recent studies, combinatorial peptide ligand library protein enrichment (commercially available as the ProteoMiner™ kit) is poised to become a promising and vital inclusion in sample preparation workflows for proteomic analysis[56,72,80,339]. Although there are many studies of this enrichment approach performed on human plasma, its application on samples from blood fractions of domestic animals, especially sheep, are poorly described. In a study that compared ovine and bovine milk whey proteomes by MS analysis, the use of ProteoMiner™ enrichment resulted in increased proteome coverage, in comparison to undepleted samples[339]. As an aim of this chapter, it was considered necessary to evaluate the effect of depletion of highly abundant proteins in the circulating acellular proteome of sheep using ProteoMiner™, prior to

nanoLC-nanoESI-MS/MS analysis.

Another approach to protein separation is off-gel fractionation (isoelectric focussing)[340]. This is a composite of in-gel and in-solution methods of protein sample separation, whereby proteins are not trapped in-gel, but are recovered in solution before downstream analysis, such as mass spectrometry[116,340]. In principle, this method takes advantage of an immobilised pH gradient strip, together with a pI gradient, for analyte separation[116]. This approach has been widely used in fractionating human peptide samples[341] and in proteins[116,340]. The method has been suggested as suitable for discovery protein analysis of human plasma[340] and has been used, for example, in a study that compared an animal cell lysate and its tryptic peptide digests[342]. The use of off-gel fractionation has recently been reported in the sample preparation of meat products[343] and sheep milk whey[339], but reports on its use on the liquid fraction of blood are not available. Moreover, except for one report detailing its use on protein analysis of buffalo and goat meat[344], the inclusion of 1D SDS-PAGE in off-gel workflows is not commonly described in the literature. One experiment in this chapter aimed to apply off-gel fractionation to undigested sheep serum and to subject the fractions to 1D SDS-PAGE, with a view to increasing protein coverage and to visualise protein bands from the fractions.

In light of the preceding context, a set of experiments was carried out that used four different protein fractionation approaches applied across several hundred samples, aiming collectively to ensure comprehensive proteome coverage and compare their protein yields.

5.3 Methods

5.3.1 A comprehensive analysis of fractions of acetone-precipitated sheep plasma and serum

Freshly collected pooled plasma in acid citrate dextrose[345,346] and serum samples from 20 healthy adult sheep were obtained from Serum Australis Pty Ltd under refrigeration conditions at 4^{o}C (Table 5.0). In order to characterise proteins, the crude analytes (untreated plasma and serum), acetone precipitates and supernatant fractions of plasma and serum samples were subjected to 1D SDS-PAGE (for protein band visualisation only – these gels were not submitted for MS analysis) and in-solution digestion (for protein analysis), as described in the generic methods in Chapter 3. A separate precast gel was used for plasma and serum 1D SDS-PAGE workflow (Figure 5.0). Proteins from the in-solution acetone-precipitation workflow were identified using ProteinPilot™ software by searching a composite (sheep, ox and goat) NCBI database in the first instance. This was done for comparative purposes only to benefit investigators who prefer performing searches using the NCBI database. Proteins were then definitively identified using ProteinPilot, Mascot and PeptideShaker, by searching a sheep-only UniProtKB database for inclusion in the PSL.

Table 5.0. Details of healthy adult sheep that provided plasma and serum samples obtained from Serum Australis (SA).

No.	Sheep ID	Sex	Body Weight	No.	Sheep ID	Sex	Body Weight
	Plasma			**Serum**			
1	30	Male	70 kg	11	147	Male	70 kg
2	20	Male	65 kg	12	133	Female	70 kg
3	186	Male	65 kg	13	117	Female	65 kg
4	162	Female	60 kg	14	92	Male	70 kg
5	127	Female	60 kg	15	NT	Male	70 kg
6	22	Male	60 kg	16	50	Female	65 kg
7	123	Female	65 kg	17	198	Male	60 kg
8	135	Female	65 kg	18	128	Female	60 kg
9	119	Female	65 kg	19	184	Male	60 kg
10	107	Female	60 kg	20	137	Female	65 kg

5.3.2 A comprehensive analysis of fractions of partial organic precipitation of sheep plasma and serum proteins using acetonitrile

A tripartite experiment was designed to determine the effect of pH, protein dilution and partial organic precipitation, using ACN on pooled plasma and pooled serum of healthy sheep, on protein ID yields (Figure 5.0). Samples (500 µL) of plasma and serum were centrifuged at 17,500 *g* at 4°C for 1 minute and then six aliquots (50 µL) were made from each of the supernatants. The pH of the pooled sheep serum supernatant was 7.6 (control), while that of two additional aliquots was adjusted to 3.3 or 8.6 by titrating with acetic acid or by adding 16 drops of 200 mM NH_4HCO_3 buffer, respectively. For the pooled sheep plasma, the pH was 7.72 (control), while that of two additional aliquots was adjusted to 3.01 by titrating with acetic acid, and to 8.86 by adding 13 drops of 200 mM NH_4HCO_3 buffer directly to the sample aliquots, respectively. To investigate the effect of protein concentration on isolation and detection, three other aliquots were diluted 1:10 (v:v %) with 100 mM ammonium acetate buffer to achieve corresponding pHs of the plasma and serum workflows.

For protein precipitation, ACN was mixed with the serum or plasma samples 1:1 (v:v %). The samples were vortexed vigorously three times at 1,400 *g* for 5 seconds, and then incubated for 10 minutes on a vortexing platform at 400 *g* at RT. Vortexing and sonication steps were repeated twice before the samples were centrifuged at 17,500 *g* at 4°C for 10 minutes. After precipitation, the supernatants were collected into 1,500 µL Eppendorf tubes. The pellets and supernatants were

separated and vacuum dried at 37°C in a SpeedVac Concentrator (Christ® cat. No. RVC 2-33 IR). The dried precipitates and supernatant fractions were vigorously vortexed and sonicated in 100 µL and 30 µL of 6 M urea dissolved in 25 mM NH_4HCO_3 buffer, respectively. The protein concentration was then determined using the BCA protein assay method. Each sample corresponding to a given pH and dilution was aliquoted in 20 µg quantities in duplicate for in-gel digestion (a) and in-solution digestion (b), as illustrated in Figure 5.3. The samples were then subjected to generic 1D SDS-PAGE and in-solution digestion prior to nanoLC-nanoESI-MS/MS.

To determine the reproducibility of the in-gel experiments and to compare ACN-precipitated samples under different pH conditions, 20 µg of crude serum (control) was loaded alongside 20 µg × 4 lanes of acetone-precipitated serum and 20 µg × 4 lanes of serum supernatant fractions in one gel. Similarly, 20 µg of crude plasma (control) was loaded alongside 20 µg × 4 lanes of acetone-precipitated plasma and 20 µg × 4 lanes of plasma supernatant fractions in another gel. Both gels were subjected to generic 1D SDS-PAGE.

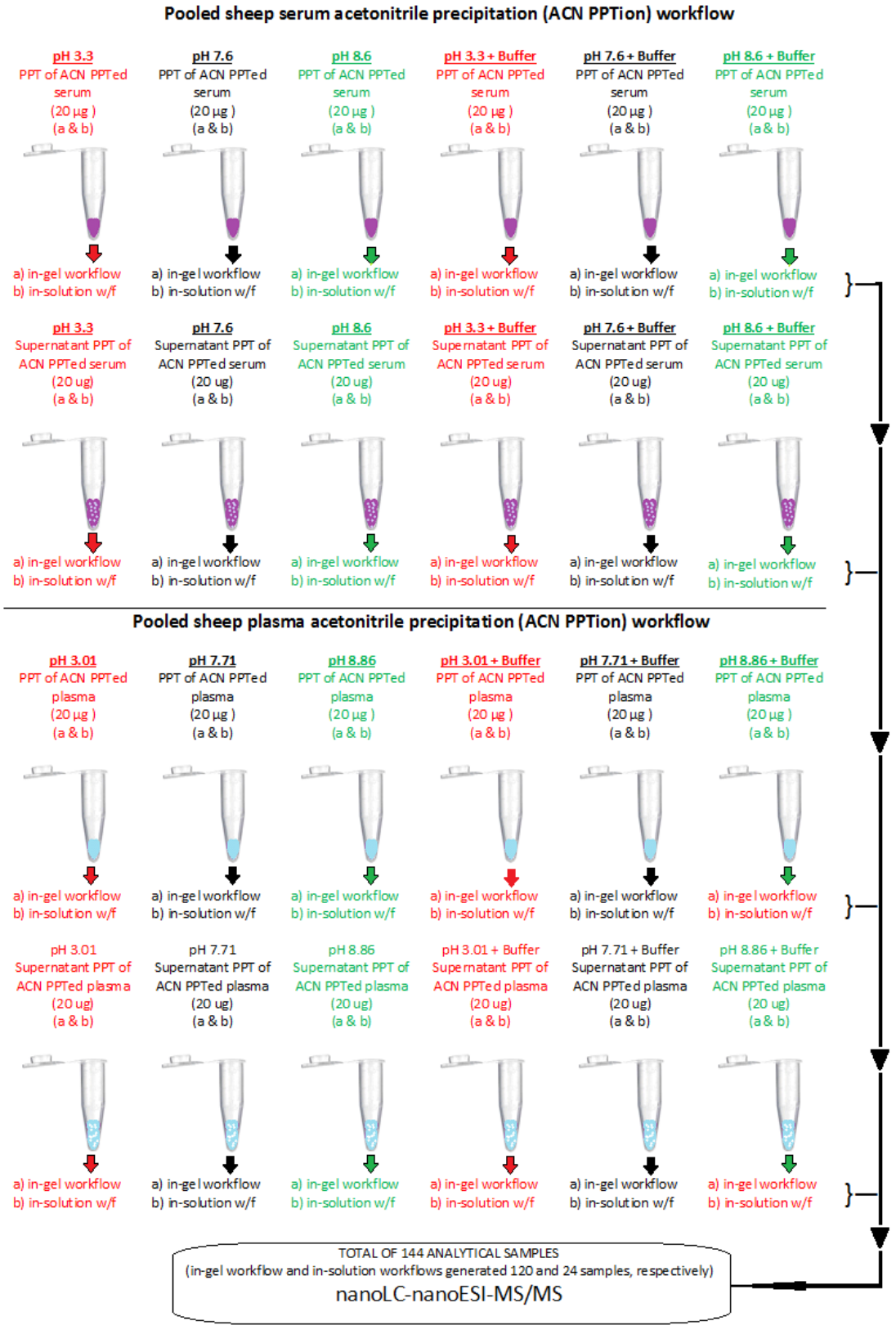


Figure 5.0. Experimental design for acetonitrile precipitation of proteins in serum and plasma from healthy sheep for in-depth proteome coverage for the PSL. **Key:** ACN = acetonitrile; PPT = precipitate; PPTion = precipitation; w/f = workflow.

5.3.3 Combinatorial peptide ligand library enrichment of sheep plasma and serum

A protein enrichment kit (ProteoMiner™, Bio-Rad Laboratories) was used to prepare pooled samples of plasma and serum from healthy sheep according to the manufacturer's instructions, with an additional adapted step for elution (Figure 5.1). Briefly, a spin column was placed in a capless collection tube and centrifuged at 1,000 *g* for 30–60 seconds to remove the storage solution supplied with the kit. The collected material was then discarded. The bottom cap of the column was replaced, followed by adding 600 µL of wash buffer. The column was rotated end-to-end several times over a 5-minute period. The bottom cap was removed and the column was placed in a capless collection tube and centrifuged at 1,000 *g* for 30–60 seconds to remove the buffer, followed by discarding the collected material. The preceding procedure was repeated up to this point, followed by removing the cap, placing the column in a capless collection tube and then centrifuging at 1,000 *g* for 30–60 seconds to remove the wash buffer. The collected material was discarded and the bottom cap on the spin column was replaced. At this point, the column contained 100 µL of settled beads, ready for sample binding. The elution process was repeated six times before pooling the collected elutes for desalting and downstream analysis. One more elution cycle (bead elute) for in-solution digestion was carried out to identify any proteins left behind by the first standard elution cycle.

5.3.4 Off-gel fractionation of serum proteins

In this set of serum sample preparation experiments, proteins were separated using the OFFGEL Fractionator (Agilent 3100, Agilent Technologies). All fractionation procedures used the default 24 cm fractionation program, as described in the manufacturer's instructions, except for the use of glycerol. For 24 separations, 400 µg of protein from crude serum samples SC449–SC472 (C1–C24), and 400 µg protein from acetone-precipitated serum samples SC473–SC496 (A1–A24) pooled from ten healthy sheep were brought to 3.6 mL with carrier ampholytes (GE Healthcare Life Sciences), in line with the manufacturer's specifications (0.96% v/v final concentration). After the immobilised pH gradient strips were swollen, the samples were loaded and the wells were sealed as per the standard protocol. The pre-defined 24–well focussing protocol (OG24PE00) was then run until 60 kilovolt hours was achieved, at which time focussed proteins were held with 10 µA current. From each of the 24 fractions of 400 µg samples of crude (SC449-SC472) and acetone-precipitated serum (SC473–SC496) from the fractionator, 100 µL was subjected to in-solution digestion. Five microlitres of solution from each well was recovered for in-gel digestion and subsequently merged into two samples, SC665 (crude serum) and SC666 (acetone-precipitated serum), prior to analysis.

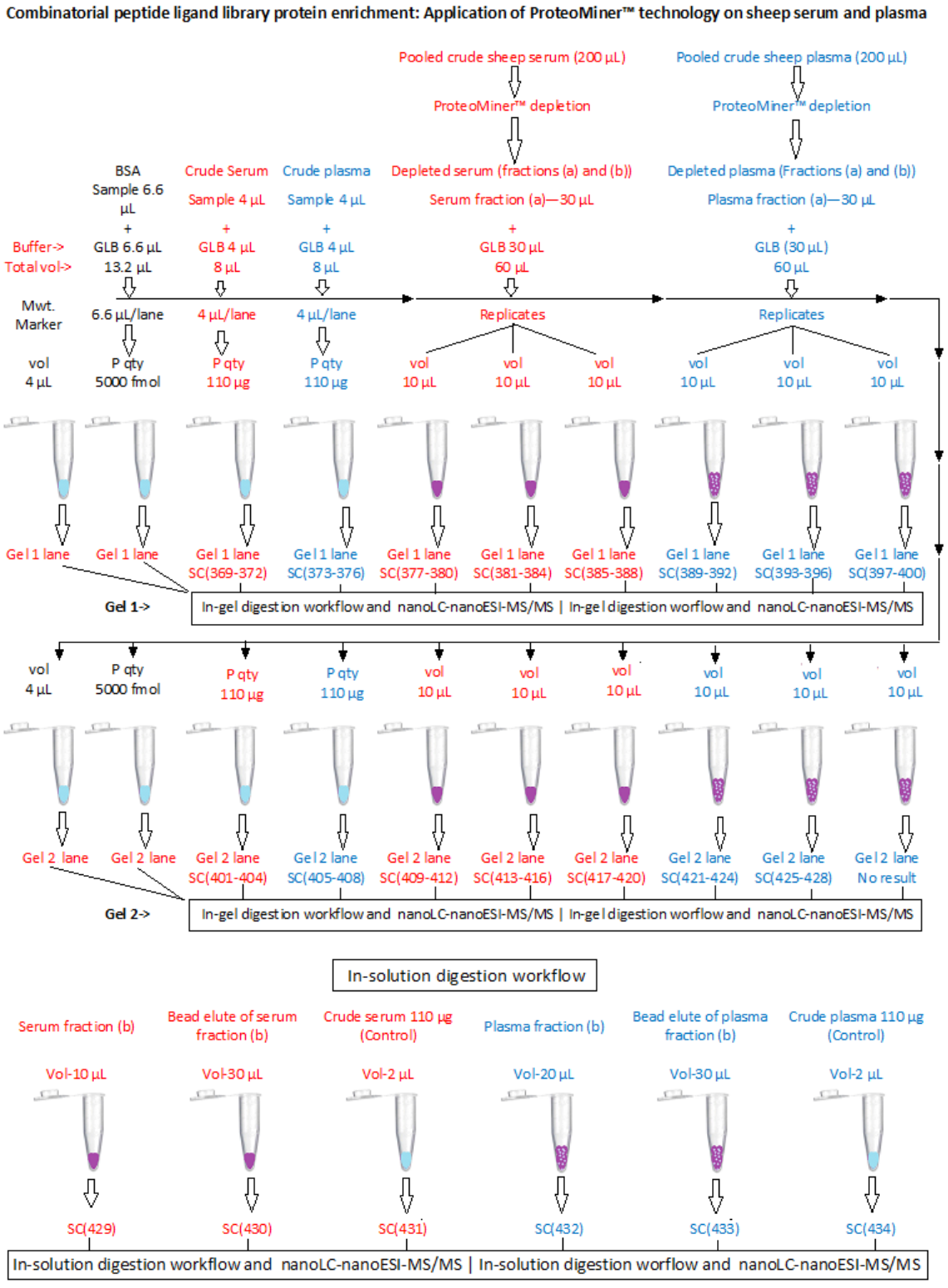


Figure 5.1. Experimental design for combinatorial peptide ligand library protein enrichment sheep plasma and serum prior to in-gel and in-solution digestion of proteins preceding nanoLC-nanoESI-MS/MS. **Key:** GLB = gel loading buffer; vol = volume; MWt. = molecular weight; P qty = protein quantity; SC = investigator initials followed by sample number or batch of samples in brackets.

5.4 Results

5.4.1 A comprehensive analysis of fractions of acetone-precipitated sheep plasma and serum

5.4.1.1 1D SDS-PAGE

Coomassie-stained 1D SDS-PAGE preparations of acetone precipitation and fractionation studies of plasma and serum are presented in Figure 5.2. Note the differences between precipitate and supernatant bands, as well as the band differences in the lower molecular mass regions between plasma and serum fractions.

Representative individual gel lanes for each precipitate/supernatant fraction pairs, crude plasma and serum protein loadings depicted in Figure 5.2 A and B were spliced and drawn together for visual comparison (Figure 5.2 C). Overall, the acetone-precipitated fractions had the most intense gel bands. Plasma fractions had visible bands in the 10kD region (red rectangular box in APP and CP gel lanes) that were less obvious in the serum lanes. Acetone-precipitated serum lanes had visible bands in the 15 kD region (brown rectangular box in the APS gel lane) that were apparent in other gel lanes. The serum supernatant band was more visible in the 37kD region than was the plasma supernatant.

5.4.1.2 In-solution digestion

For protein quantification, in-gel digestion appears to affect recovery of individual peptides moreso than in-solution digestion[178]. For this reason, only in-solution samples of this acetone-precipitation workflow were further analysed for protein identification.

5.4.1.3 Composite ox, goat and sheep NCBI protein database search results in ProteinPilot™

Protein IDs from the acetone-precipitation workflow are highlighted in Figure 5.3. There were 23 out of 142 protein IDs exclusive to the serum supernatant and 16 out of 125 protein IDs exclusive to the plasma supernatant. Of the combined 39 protein IDs in supernatants of both plasma and serum, only six were common to both supernatants. Overall, the analysis of both fractions of plasma and serum – and their respective crude analytes – resulted in 154 protein IDs.

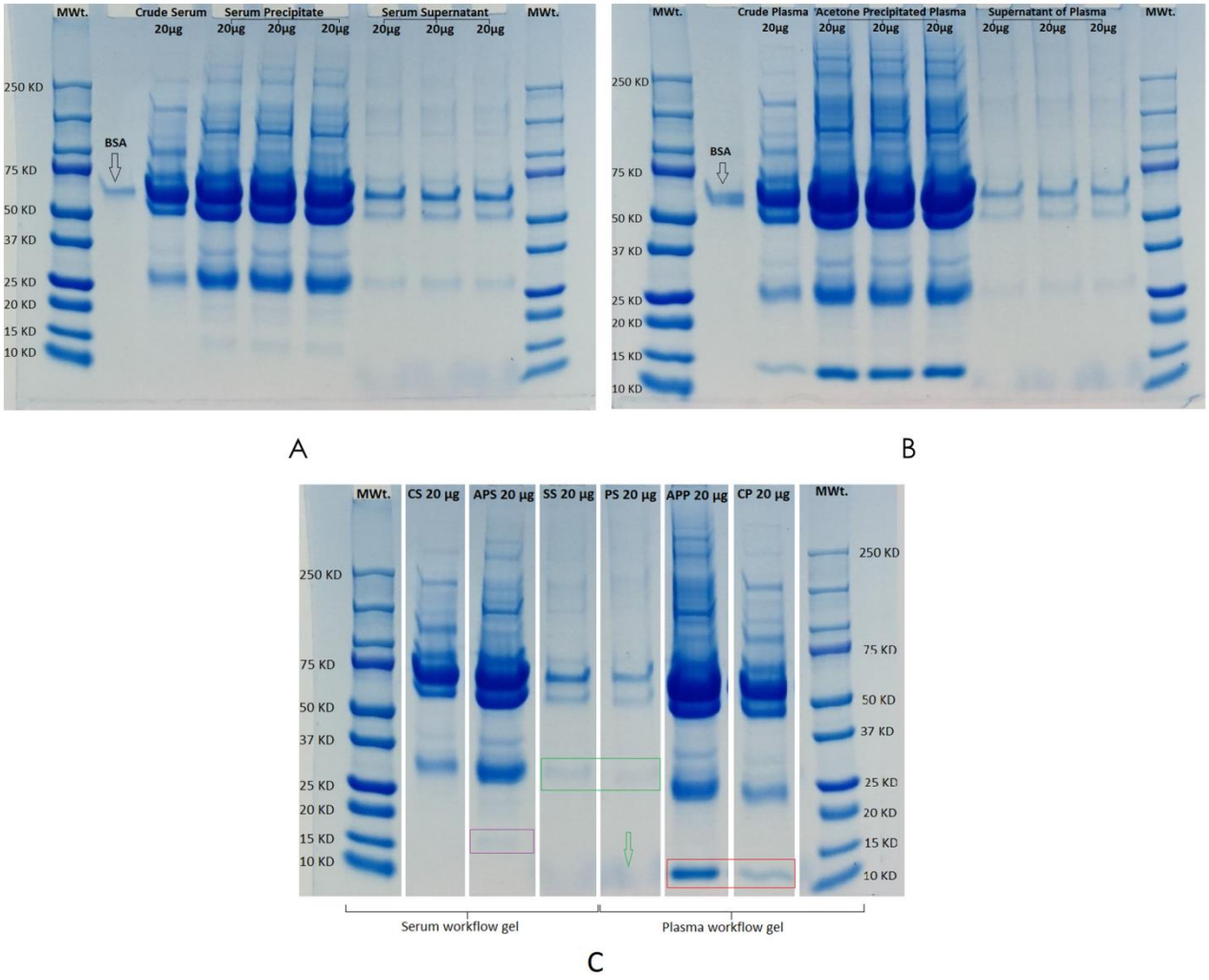
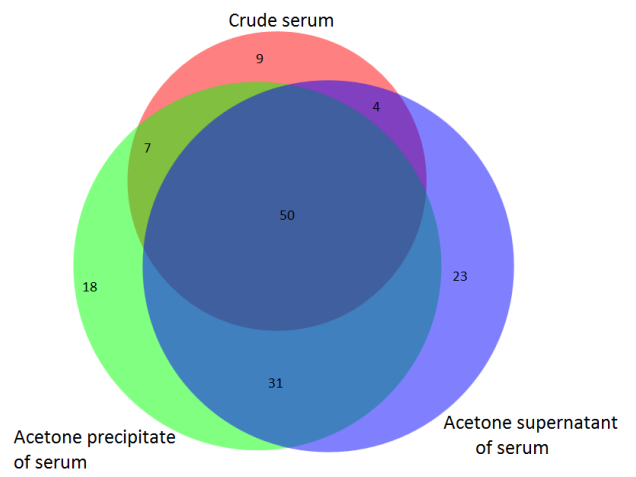
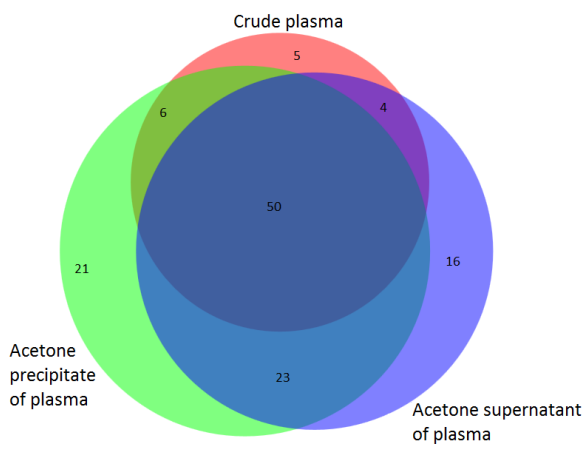


Figure 5.2. 1D SDS-PAGE preparations of acetone precipitation and fractionation studies using pooled serum (A) and plasma (B) samples from 20 healthy adult sheep. A comparison of gel bands of lanes spliced from gel (A) and gel (B) are illustrated in (C). The left-most and right-most wells of both gels A and B were loaded with 4 µL of a protein molecular weight standard (Precision Plus Protein™ Dual Xtra, Bio-Rad Laboratories). The second well from the right of each gel was loaded with a BSA standard (5,000 fmol). Other wells were loaded with 20 µg of protein samples each, as follows: Gel A – crude serum, serum precipitate of acetone × 3 replicates, and serum supernatant from acetone precipitation × 3 replicates; Gel B – crude plasma, plasma precipitate of acetone × 3 replicates, and plasma supernatant from acetone precipitation × 3 replicates. **KEY:** MWt. = molecular weight marker; CS = crude serum; APS = acetone-precipitated serum; SS = supernatant of serum after acetone precipitation; PS = supernatant of plasma after acetone precipitation; APP = acetone-precipitated plasma; CP = crude plasma.





A **B**

Figure 5.3. Protein identifications (IDs) in ProteinPilot™ software using a composite ox, goat and sheep NCBI protein database following acetone precipitation and analysis of both the precipitate and supernatant fractions of plasma and serum from healthy sheep. Substantial numbers of protein IDs were present in the supernatant fractions. **Key:** A = serum workflow; B = plasma workflow.

5.4.1.4 UniProtKB sheep protein database search results for ProteinPilot, PeptideShaker and Mascot

The results of searching a sheep-only UniProtKB database are presented in Figure 5.4. Panel A shows the comparison of protein IDs in crude plasma and serum with their respective acetone-precipitated fractions using ProteinPilot, PeptideShaker and Mascot. ProteinPilot yielded the most protein IDs across the board, followed by Mascot and then PeptideShaker. The protein numbers were comparable in acetone-precipitated factions of plasma and serum, but not in supernatants. More proteins were identified from crude plasma than from crude serum, and this difference was particularly marked when PeptideShaker was used. Acetone-precipitated serum fractions yielded a higher number of protein IDs than did plasma, regardless of which search engine was employed.

Protein IDs from PeptideShaker were compared (Figure 5.4 B–G). To highlight the differences, entries from UniProtKB are provided for protein IDs that were exclusive to individual fractions only. Panel B shows the results of the comparison of IDs in crude serum, crude plasma, acetone precipitate of serum proteins (Serum PPT), acetone precipitate of plasma proteins (Plasma PPT) and their respective supernatants (Serum SUP and Plasma SUP) in samples from healthy sheep. The entire acetone-precipitation workflow experiment yielded 105 validated protein ID hits in PeptideShaker, 20 of which were collectively exclusive to crude serum, Serum PPT, Plasma PPT, Serum SUP and Plasma SUP (Figure 5.4 B).

There were 60 protein IDs made from crude serum and crude plasma, with the latter yielding more proteins that were exclusive (Figure 5.4 C). Of these 60 IDs, 6 proteins (P68116, C8BKD1, W5QDP8, W5NXM1, W5QA07, B3GS77) were exclusive to crude serum and 19 (W5NQ46, P02075, W5PHP7, W5P812, W5Q5A6, A5YBU9, W5NUJ7, P32262, W5PHI7, W5PID9, F2YQ13, W5P1J8, W5PFC9, W5QH50, W5NXP3, W5Q2E1, W5PJZ1, W5PW21, W5PGT6) were exclusive to crude plasma. Acetone-fractionated serum yielded 91 protein IDs. Of these, 4 (4.4%) proteins were exclusive to the precipitate of serum and 27 (29.7%) were exclusive to the serum supernatant.

Unlike for serum fractions, as described above, acetone-fractionated plasma yielded 82 IDs, with the supernatant yielding 20.7% unique IDs and the precipitate fraction 17.1% unique IDs (Figure 5.4 E). A comparison of the acetone precipitates of plasma and serum resulted in comparable numbers of protein IDs (Figure 5.4 F). There were 79 protein IDs in the combined Serum PPT and Plasma PPT. Of these 79 IDs, 14 (17.7%) protein IDs were exclusive to the serum precipitate, while 15 (19%) protein IDs were exclusive to the plasma precipitate.

Acetone-precipitation supernatant fractions of both plasma and serum had a combined total of 96 protein IDs, 28 (29.2%) of which were exclusive to the serum supernatant fraction, while 9 (9.4%) were exclusive to the plasma supernatant (Figure 5.4 G).The entire acetone-precipitation workflow experiment yielded 136 protein IDs in Mascot.

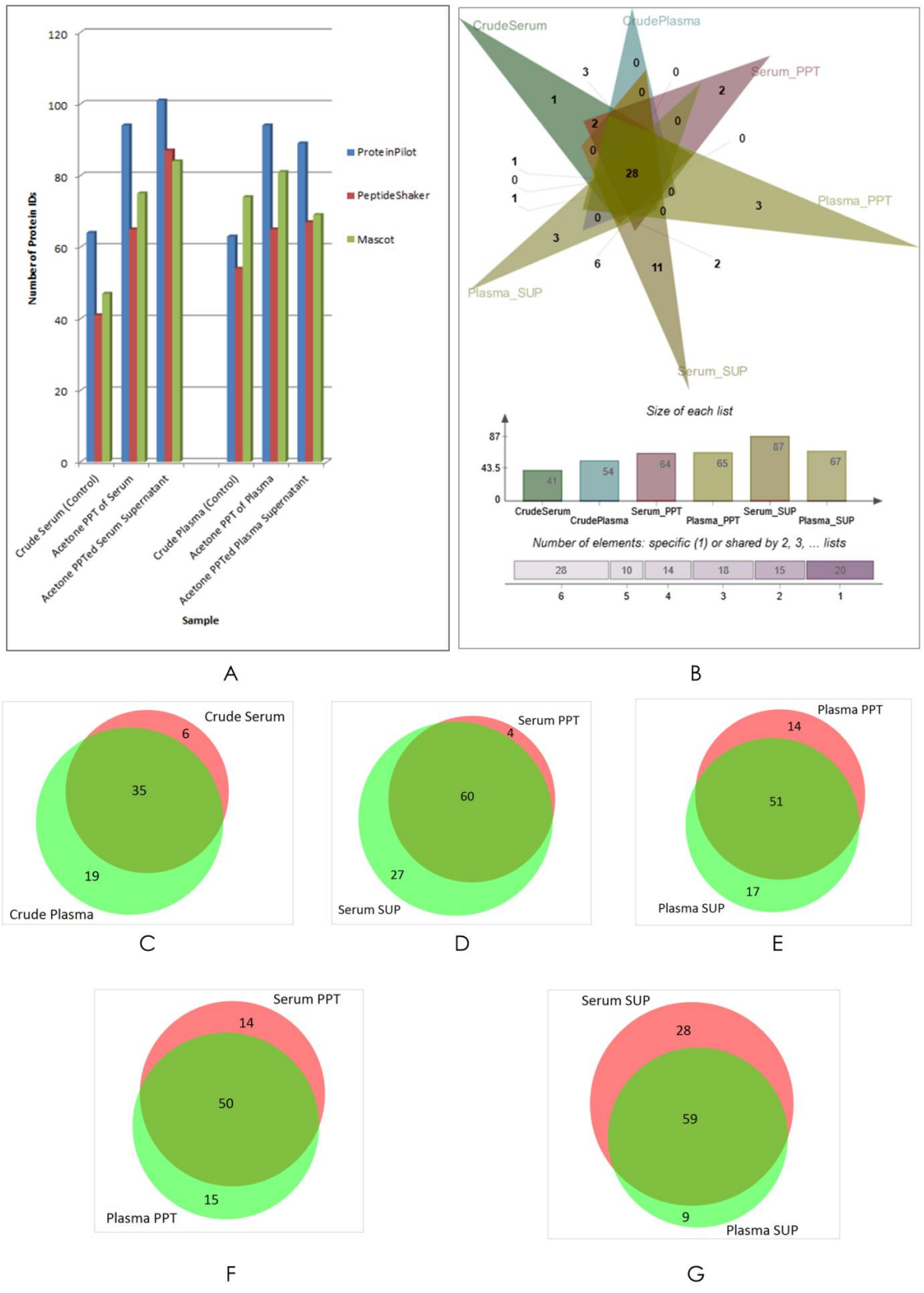


Figure 5.4. Comparison of protein identifications (IDs) from the acetone-precipitation workflow (A), and PeptideShaker IDs of the respective fractions of acetone precipitation (B, C, D, E, F and G). **Key:** IDs - identifications; PPT = precipitate; PPTed = precipitated; SUP = supernatant.

5.4.2 A comprehensive analysis of fractions of partial organic precipitation of sheep plasma and serum proteins using acetonitrile

5.4.2.1 1D SDS-PAGE

The analysis of the fractions of plasma and serum derived from ACN precipitation subjected to 1D SDS-PAGE are presented in Figure 5.5.

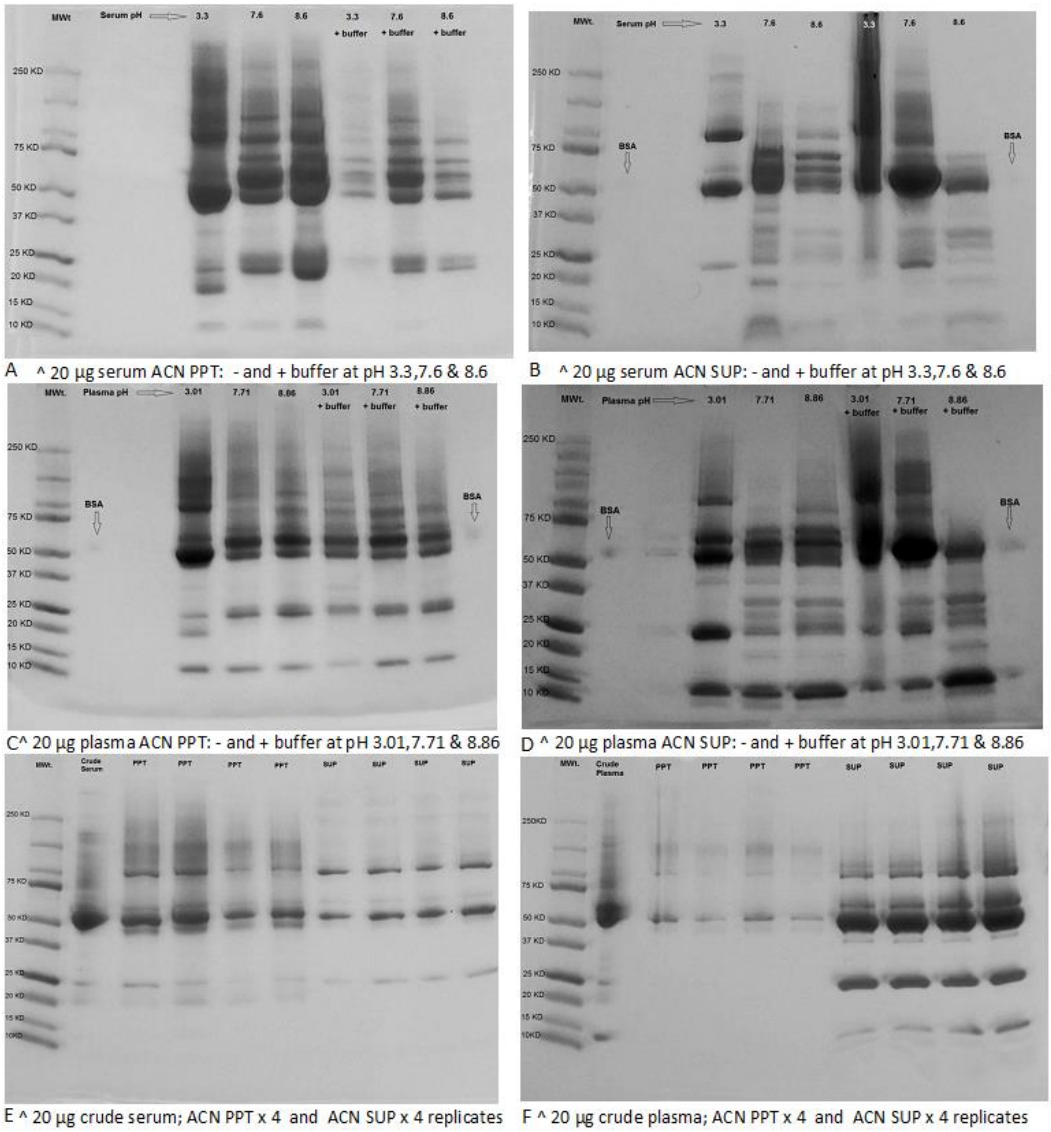


Figure 5.5. 1D SDS-PAGE images of crude and ACN-precipitated fractions of plasma and serum from healthy sheep under different pH conditions and protein concentrations. Each gel lane in A, B, C, D, E and F (except the BSA lane) was loaded with 20 µg of fractionated sample. Crude serum (E) or plasma (F) was loaded in the second lane as a control. **Key:** MWt. = molecular weight; KD = kilodalton; ^ = gel image above; + = with; – = without; ACN = acetonitrile; PPT = precipitate of acetonitrile precipitation; BSA = bovine serum albumin standard (5,000 fmol); SUP = supernatant of acetonitrile precipitation.

There were distinct differences in protein bands between ACN precipitates, supernatants, protein concentrations, pH and crude samples of plasma and serum (panels A, B, C and D). Comparative precipitate and supernatant bands were reversed in appearance between serum (panel E) and plasma (panel F) in the reproducibility experiments.

The gel lanes from 1D SDS-PAGE of ACN precipitation workflow for serum (A and B) and plasma (C and D) in Figure 5.5 were analysed for proteins using ProteinPilot, PeptideShaker and Mascot search engines (Figure 5.6). The numbers of protein IDs detected are shown in panel A for serum and panel B for plasma. The protein IDs in C and D were those of combined ACN precipitates of plasma and serum fractions, respectively, at three different pH conditions and protein concentrations, using the three protein search engines.

5.4.2.2 In-solution digestion

The protein IDs in panels E and F in Figure 5.6 were those of in-solution digestion of ACN-precipitated sheep plasma and serum fractions, respectively. A pH of 8.6 favoured protein identification in both plasma and serum. The protein IDs shown in G and H were generated from the Mascot search engine for in-solution digestion only, and a composite of in-solution and in-gel digestion of combined ACN-precipitated sheep plasma and serum fractions (PPT and SUP), respectively, at different pH conditions and protein concentrations. There were subtle differences in the number of proteins identified between ACN precipitates, supernatants, protein concentrations, pH and crude samples of plasma and serum. Overall, the number of protein IDs in plasma was marginally higher than those in serum.

The number of protein IDs generated by Mascot of a composite of 1D SDS-PAGE of ACN precipitation workflow samples of sheep serum (ACN_Gel_Serum) and plasma (ACN_Gel_Plasma) are shown in panel I. Of the 338 protein IDs from the combined in-gel experiments of plasma and serum, the majority were from plasma. Similarly, the protein IDs from a composite of in-solution digestion of ACN precipitation workflow samples of sheep serum (ACN_Sol_Serum) and plasma (ACN_Sol_Plasma) are shown in panel J. The analysis of in-solution digested samples resulted in a total of 185 protein IDs. The number of protein IDs of a composite of 1D SDS-PAGE of ACN_Gel_Serum, ACN_Gel_Plasma, ACN_Sol_Serum and ACN_Sol_Plasma are shown in K. Overall, the entire ACN precipitation workflow resulted in 376 protein IDs. Plasma yielded the largest number (30.3%) of exclusive protein IDs.

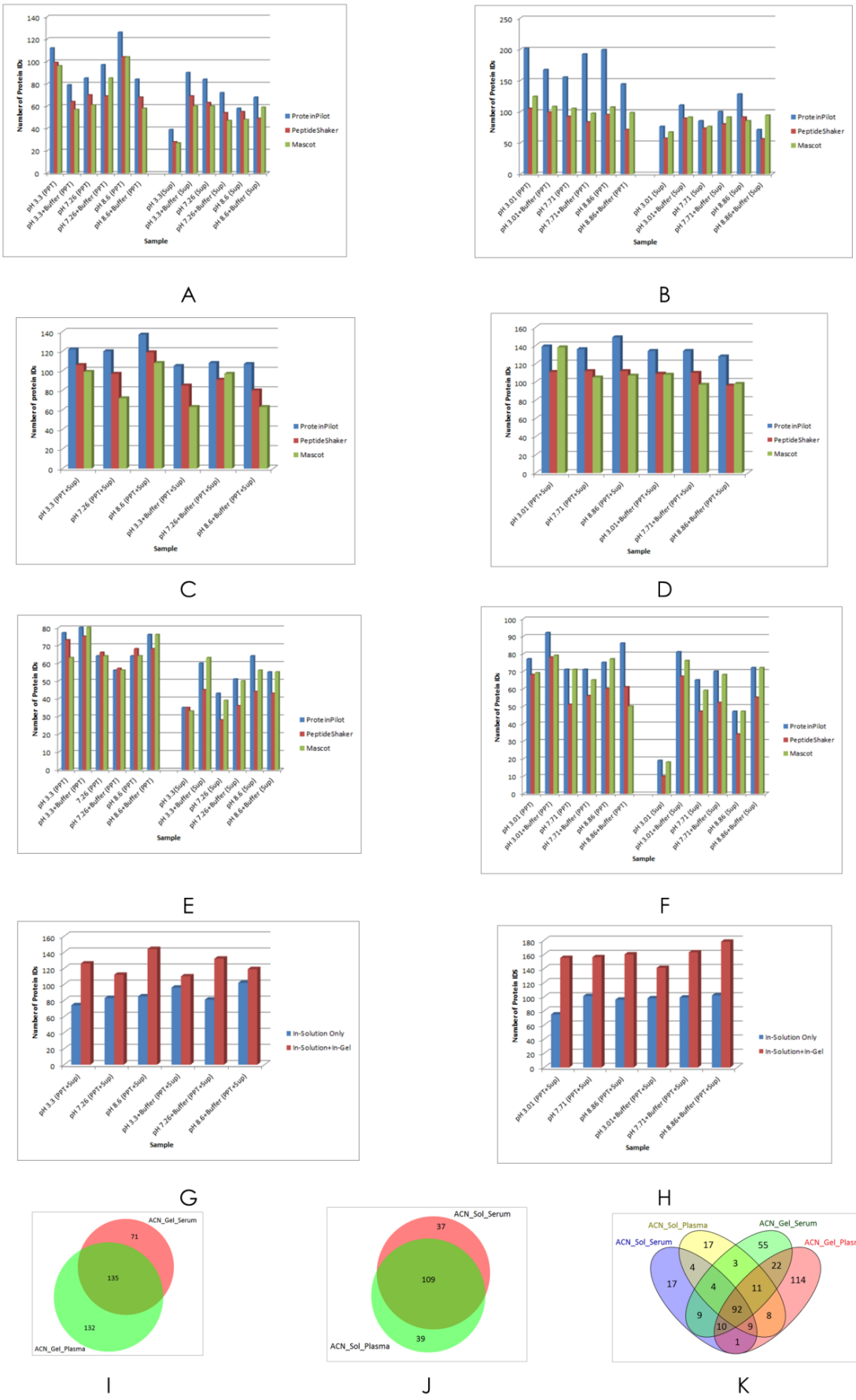


Figure 5.6. Protein identifications and comparisons of ACN precipitation workflow. **Key:** PPT =

precipitate; SUP = supernatant; ACN = acetonitrile, ACN_Gel_Serum = proteins identified from in-gel digested serum; ACN_Gel_Plasma = proteins identified from in-gel digested plasma; ACN_Sol_Serum = proteins identified from in-solution digested serum; ACN_Sol_Plasma = proteins identified from in-solution digested plasma.

5.4.3 Combinatorial peptide ligand library protein enrichment of sheep plasma and serum

5.4.3.1 1D SDS-PAGE

Figure 5.7 shows the application of ProteoMiner™ abundant protein depletion technology to sheep plasma and serum. The images of the 1D SDS-PAGE workflow following ProteoMiner™ depletions are presented in Figures 5.7 A and B. The annotated images of Coomassie-stained 1D SDS-PAGE gels of crude and ProteoMiner™ depleted fractions of plasma and serum from healthy sheep are represented by panels A and B, respectively. Crude serum and crude plasma were loaded using 110 µg of protein per lane. Depleted samples mixed at 1:1 (v:v %) with gel loading buffer were loaded at 10 µL/lane for ProteoMiner™ plasma and serum elutes. Although both gels A and B were loaded identically and under uniform conditions, including electrophoresis in the same tank, there were noticeable gel-related aberrations in the stained gel outcomes. The MWt. marker migrated into the 2nd gel lane, while the BSA band shifted to the MWt. lane, visible under the 75 kD mark in gel B. The 2nd lane of depleted plasma in gel A was faint, while the 3rd lane of depleted plasma in gel B was missing entirely. These gel-related aberrations affected the visual outcome of some protein bands. All the loaded lanes (MWt. marker, BSA, crude serum, crude plasma, three replicates of each of crude plasma and serum) were represented in gel A, except that the second and third plasma lanes appeared faint and distorted, respectively. The gel lanes for plasma and serum for the two gels were analysed for proteins using three different search engines, and the results are presented in Figures 5.7 C and D. The distortion of the 3rd plasma lane did not impact on the number of protein IDs, but the faint 2nd plasma lane had a comparatively low number of protein IDs (Figure 5.7 C). In gel B, there was marked distortion in appearance of the bands of MWt. marker and BSA lanes; the MWt. marker drifted into the BSA lane and the 3rd depleted plasma lane was missing altogether, due to a defect in the outer gel well that caused the sample to be lost (Figure 5.7 D). Except for some horizontal stretching in appearance in the 71 kD region for crude serum and crude plasma lanes, and a vertical streak in the 3rd depleted serum lane, the rest of the bands in all the lanes were comparable with those of gel A. The slightly more intense vertical streak in the 3rd depleted serum lane of gel B was possibly caused by a larger aggregation of sample settling in a pocket in an uneven gel well.

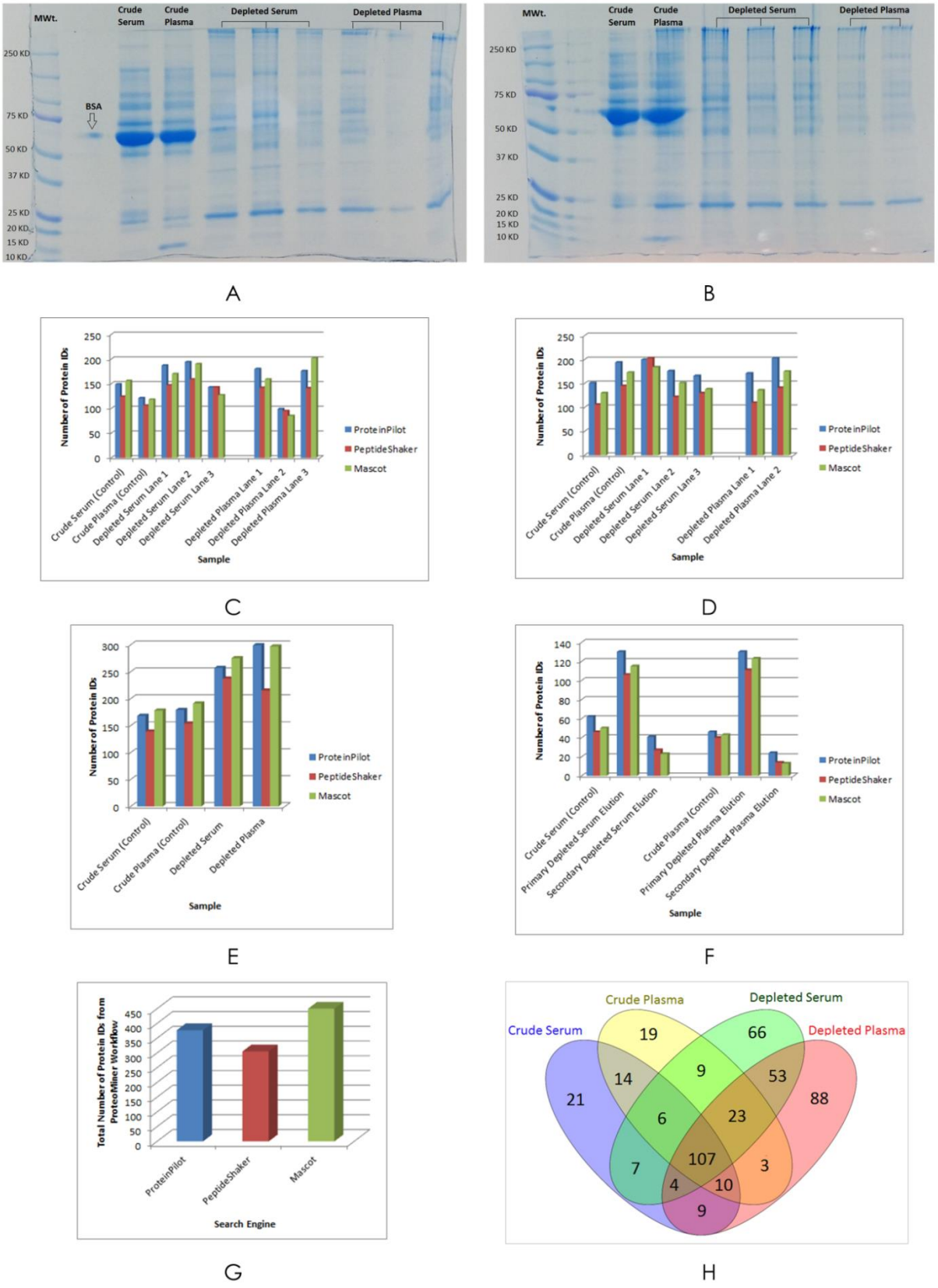


Figure 5.7. Evaluation of ProteoMiner™ using plasma and serum of sheep. **Key:** A and B = 1D SDS-PAGE workflow; C = protein identification (ID) results of plasma; D = protein ID results of serum; E = comparison of crude plasma and serum protein IDs; F = comparison of crude and

depleted analytes with protein IDs made from secondary elution; G = Number of protein IDs in ProteinPilot, PeptideShaker and Mascot; H = comparison of Mascot protein IDs of crude and depleted plasmas and serum.

Proteins were identified using ProteinPilot, PeptideShaker and Mascot (Figure 5.7 C, D and E). Protein IDs were compared between 1D SDS-PAGE and in-solution sample digestion, and depleted samples were compared to controls (crude serum and crude plasma). The protein IDs for gels A and B are shown in panels C and D, respectively. The combined protein IDs for gels A and B are presented in panel E.

5.4.3.2 In-solution digestion

Protein IDs from the in-solution workflow are shown in Figure 5.7 F. Depleted plasma and serum had a higher number of protein IDs than did the crude respective undepleted analytes. There were substantial numbers of proteins identified in the secondary elution, with IDs from serum being marginally higher than IDs from plasma. For in-solution digested samples, ProteinPilot yielded the most protein IDs, with Mascot performing less well in this regard. Crude serum, primary depleted serum elution and secondary depleted serum elution had 62, 130 and 41 ProteinPilot IDs, respectively. Crude plasma, primary depleted plasma elution and secondary depleted plasma elution had 46, 130 and 24 ProteinPilot IDs, respectively.

A comparison of the protein IDs in all three search engines is provided in panel G. The distribution of protein IDs made in Mascot across all samples is shown in panel H. Overall, Mascot was superior to ProteinPilot and PeptideShaker in identifying proteins after combining in-gel and in-solution digestion results (Figure 5.7 G). Using ProteoMiner™ resulted in 439 protein IDs from the entire workflow. These IDs comprised 207 additional protein IDs (66 + 53 + 88 IDs) to the undepleted samples, of which 15% (66 IDs) and 20% (88 IDs) were exclusive to depleted serum and depleted plasma, respectively (Figure 5.7 H).

5.4.4 Off-gel fractionation of serum proteins

5.4.4.1 1D SDS-PAGE

Panels A to F of Figure 5.8 represent the results of Coomassie-stained 1D SDS-PAGE preparations of 24 protein solution harvests from off-gel fractionation of 400 µg of pooled crude serum (CS1–CS8, A; CS9–CS16, B; CS17–CS24, C) and pooled acetone-precipitated serum protein (AS1–AS8, D; AS9–AS16, E; AS17–AS24, F). Prior to nanoLC-nanoESI-MS/MS, CS1–CS8, CS9–CS16 and CS17–CS24 gel lane digests were pooled into one sample (SC665) and AS1–AS8, AS9–AS16 and AS17–AS2 gel lanes digests were pooled into another sample (SC666). The combined pooled

samples (SC665 and SC666) resulted in 57 protein IDs in Mascot.

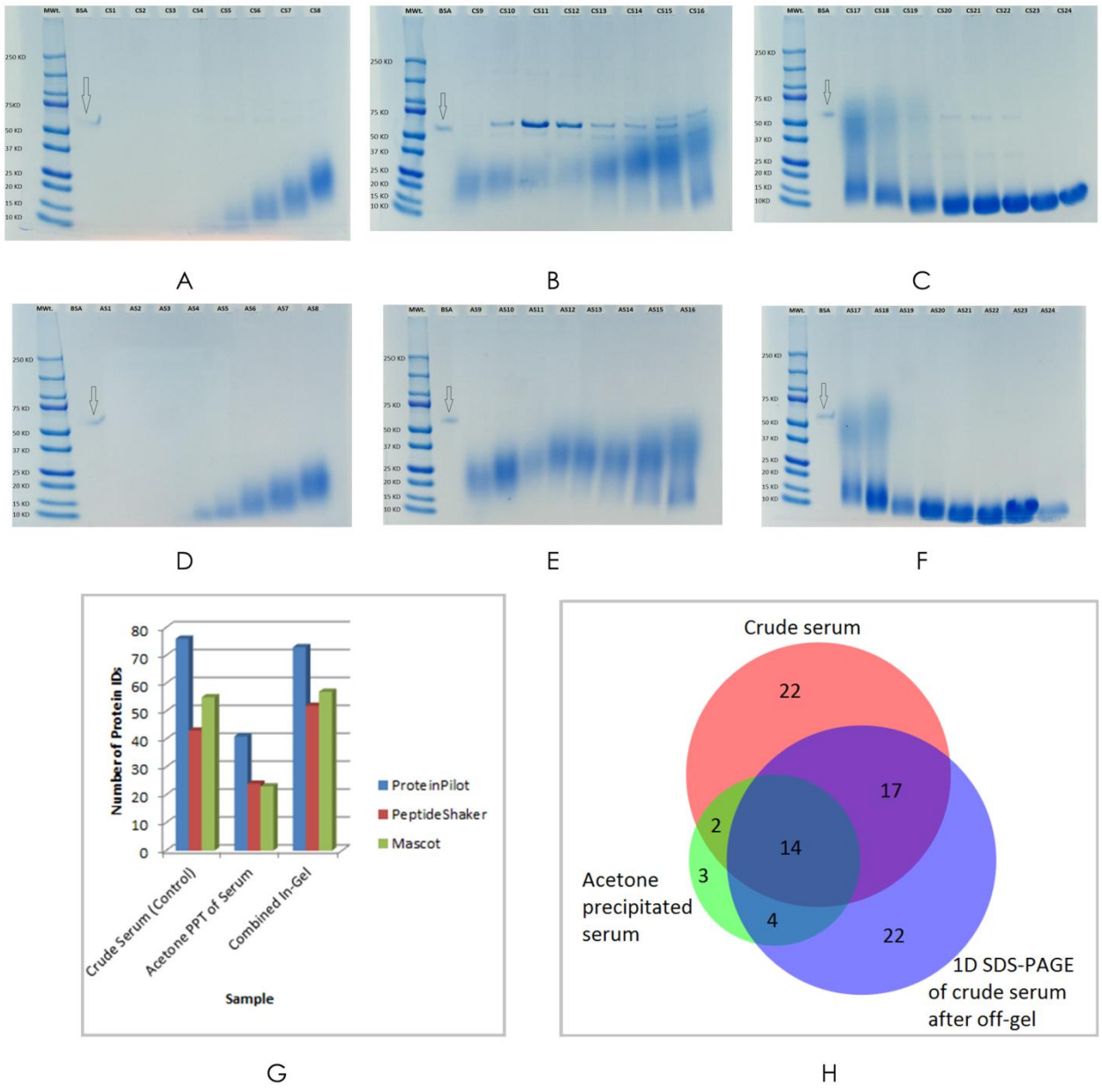


Figure 5.8. Off-gel fractionation workflow using sheep serum. Coomassie-stained 1D SDS-PAGE preparations of 24 protein solution harvests from off-gel fractionation of 400 µg of pooled crude serum A = CS1–CS8; B = CS9–CS16; C = CS17–CS24; and pooled acetone-precipitated serum protein D = AS1–AS8; E = AS9–AS16; F = AS17–AS24; G = Comparisons of protein ID yields from ProteinPilot, PeptideShaker and Mascot in the entire off-gel workflow; H = Comparison of Mascot protein IDs from the contributing workflow elements. **Key:** BSA = bovine serum albumin standard (5,000 fmol) arrow; CS = crude serum; AS acetone-precipitated serum; IDs = identifications; PPT = precipitate.

5.4.4.2 In-solution digestion

The individual off-gel fractions (CS1–CS24 and AS1–AS24) of in-solution digests were merged into the respective sample categories. Comparisons of protein ID yields from ProteinPilot, PeptideShaker and Mascot in the entire off-gel workflow are presented in Figure 5.8 G. Using Mascot as an example because of this software's popular standing in the industry, this workflow contributed 29 (34.5%) protein IDs (3 + 4 + 22 IDs) compared to crude serum (control). Eighty-four proteins were identified in this entire workflow (Figure 5.8 H).

The ultimate goal of this thesis was to construct a PSL using only high-quality data from validated peptide IDs derived from PeptideShaker search as detailed in Chapter 6. Comparisons of protein IDs for the different sample preparation methods using this software platform are illustrated in Figure 5.9. The 133 protein IDs reported in Chapter 4 were essentially from an acetone precipitation workflow, and 7 of these protein IDs were unique to this workflow. The ACN precipitation workflow yielded 198 protein IDs, 37 of which were unique to this workflow. ProteoMiner™ fractionation yielded protein 305 IDs, 145 of which were unique to this workflow, and the Off-Gel fractionation yielded 70 protein IDs, 3 of which were unique this workflow. Of the fractionated samples, it is evident that the ACN precipitation workflow produced novel protein IDs that were not also discovered through ProteoMiner™.

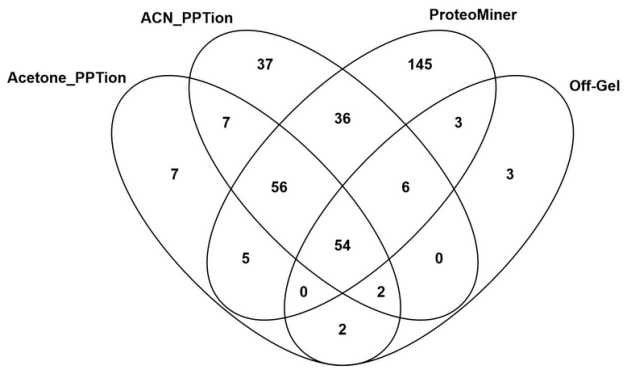


Figure 5.9. Comparison of protein ID yields using PeptideShaker software from different sample preparation techniques of sheep plasma and serum. Acetone precipitation (Acetone_PPT) was compared with acetonitrile precitipation (ACN_PPTion), ProteoMiner and Off-Gel workflows. ProteoMiner enrichment produced the highest number of protein IDs. ACN_PPTion, produced novel protein IDs that were not discovered through ProteoMiner™.

5.5 Discussion and conclusion

The overall goal of this chapter was to explore strategies for enhancing peptide extraction from sheep plasma and serum samples in order to identify more proteins and to determine which workflow gave the most yields. The results of this chapter produced a large body of data that has significantly increased the protein identification and analytical depth from what was achieved in

Chapter 4. A considerable effort went into comparing the different techniques using the UniProtKB sheep-only database by applying the analytical strategy described in the methods in order to acquire useful data to build a PSL from the circulating acellular proteome of sheep. The discussion that follows covers the observations that were made in the four approaches described.

In the present study, the output from the acetone-precipitation experiment represents the first comprehensive examination of sample (serum versus plasma) and technique (acetone-precipitation) variables on downstream protein identification yields in samples derived from the liquid fraction of sheep blood. This set of observations is important in that it will assist investigators who wish to choose either plasma or serum for their studies to make informed decisions on which medium to use if considering acetone precipitation in the sample preparation pipeline.

It is now evident that supernatant fractions of acetone-precipitated samples contain a considerable number of proteins that are not captured in the other fractions. The proteins that remained in the supernant fractions were likely to be those that failed to aggregate by the known hydrophopic mechanism induced by organic precipitation[23,329]. The distinct gel band enrichment regions from acetone preciptation of these samples (for example, the 10–15 kD region of plasma precipitate, the 17 kD region of serum precipitates and the 30 kD region of serum supernatant in 1D SDS-PAGE preparations; Figure 5.0) are interesting findings that could be useful in studies targeting proteins in these molecular weight regions. The enriched 10–15 kD region band in the plasma fractions could be α- and β-globulin[347], suggesting that there may have been some degree of haemolysis in the plasma, or possibly that this band is a characteristic unique to plasma. A common protein in the 30 kD region in human blood is peroxiredoxin[347], and this could be the case for sheep blood too. These highlighted differences between plasma and serum samples have not been reported previously.

A composite NCBI protein sequence database containing ox, goat and sheep was also used in this workflow, in conjunction with the ProteinPilot search engine, to capture homologous sequences. It was necessary to have protein identification data from this database because of the wide use of its non-redundant protein sequences in several other studies[348]. Multispecies protein sequence databases are known to be associated with increased false positive protein identifications, and in some cases yield spurious results that make protein ID difficult, due to highly homologous sequences and isoform variations of proteins[348]. However, the use of related-species composite protein sequence databases is widely advocated, especially for non-model organisms whose genomes have not yet been fully sequenced[7]. This is because the traditional approach for non-model organism proteomics is to have a protein sequence database derived from annotated genome(s), either from the non-model organism itself (sheep in this case), from the most closely related organism that has been sequenced or from a composite of closely related organisms, to be able to

interpret the data[7]. The UniProtKB database was chosen for the definitive protein searches because its use is preferred and supported by protein search engines for downstream processes, such as PeptideShaker[245] and Scaffold[187].

From these repeatable results, it is evident that the serum supernatant from acetone preciptation yields a comparatively higher number of protein IDs (30%) exclusive to this fraction than does the plasma supernantant (Figure 5.4 G). The higher number from serum confirms that the fractionation of samples reduces the complexity of samples and improves the protein number available for identification[337,349], especially after taking into consideration that serum is a fraction of plasma. The depletion of clotting factor proteins, including high-abundance fibrinogen, possibly played a role in identifying more proteins in serum than in plasma. This is because the removal of any high-abundance proteins improves protein detection yields[18]. These results are the first known studies with regard to acetone precipitation of sheep plasma and serum to show that supernatant fractions contain a considerable number of proteins. Proteins in supernatant fractions are missed during routine proteomic analysis, yet these proteins could be potentially important biomarkers, particularly in many animal models whose proteomes have not been fully characterised.

The ACN precipitation experimental workflow was adapted from earlier studies[337,338]. In the present study, a higher number of protein IDs was achieved from both the plasma supernatant and pellet at a higher pH of 8.86 than in other pH conditions (Figure 5.5 H). This observation contrasts with that of an earlier study in which ACN treatment in a higher pH did not improve protein yields[337]. The novelty here was the simplification of the mode of adding optimally standardised quantities of acetic acid and ammonium bicarbonate solutions to adjust the pH of the samples. In addition, the 1D SDS-PAGE results were superior in that they exhibited characteristic differences in protein bands. For example, the reversal in the appearance of ACN-treated plasma and serum and their respective supernatants in the reproducibility experiments (Figure 5.5 E and F) have not been previously reported. The protein bands in the ACN precipitate fractions of serum were more pronounced than were the supernatant fractions. By contrast, the protein bands in the ACN precipitate fractions of plasma were lighter than were the supernatant fractions. This suggests that treatment with ACN enriches a wider spectrum of protein species in the plasma supernatant fraction than in the serum supernatant, based on the band appearance of the 1D SDS-PAGE preparations. Conversely, treatment with ACN enriches a wider spectrum of protein species in the serum precipitate fraction than in the plasma precipitate fraction, based on the band appearance of the gels. It would have been interesting to analyse and compare the protein yields from the gel lanes and bands; this is a consideration for future studies.

While plasma produced the most protein IDs overall in the ACN precipitation workflow, it was serum at pH 8.6 that stood out prominently in showing the combined effect of pH and sample dilution with buffer, especially in the in-gel digestion workflow, on protein IDs. In the 1D SDS-PAGE experiments, plasma yielded twice as many (39.1%) exclusive protein IDs as did serum (21%), whereas the number of exclusive protein IDs was comparable for plasma and serum in the in-solution workflow. This observation suggests that it is important to consider 1D SDS-PAGE when designing DDA protein discovery workflows. If the goal of the experiment is to identify as many proteins as possible (for example, in the generation of a PSL), then the inclusion of 1D SDS-PAGE of plasma is suggested, because this is associated with a wider protein coverage. Therefore, within pH limits that are yet to be defined, the overall conclusion from this experiment is that with ACN precipitation, a higher pH, coupled with sample dilution, appears to favour protein IDs when both fractions of ACN precipitation are analysed.

Taking into consideration the preceeding discussion, it can be advanced that ACN precipitation workflow was more successful with numerous small proteins becoming more abundant in the supernatants (Figure 5.5). The most abundant proteins (albumin and immunoglobulins) still dominated the gels, indicating that the dynamic range problem had not been properly resolved. The data acquisition in this chapter was not only devoted to the analysis of proteins that had been successfully identified in Chapter 4 that utilised only serum derived from healthy experimental female sheep colony belonging to QUT and the ARCB in Brisbane. The samples used were sourced from healthy commercially farmed sheep from Serum Australis Ltd. in NSW and it involved the analysis of both serum and plasma from male and female sheep as described in the methods. This therefore represented a different set of physiological conditions that could have influenced the composition of the circulating acellular proteome.

The combinatorial peptide ligand library protein enrichment experiments evaluated a commercial kit called ProteoMiner™, using sheep plasma and serum as the protein source. This is because this kit has been reported as having the capabilities of capturing and facilitating the detection of low-abundance proteins in complex samples[72]. The findings of ProteoMiner™ use in the present study were consistent with the increase in protein IDs observed in one other study that used it on ovine and bovine whey[339], meaning that this approach of enrichment improves protein ID outcomes. An important consideration in this workflow is that the standard method was modified by including an additional elution cycle for the beads using the buffer in order to determine if substantial proteins were left behind during the standard elution recommended by the manufacturer of the kit.

Overall, the effect of ProteoMiner™ was evident in the band appearance of 1D SDS-PAGE workflow and this was reflected in the number of protein IDs (Figures 5.7 C, D, E and F). There

was noticeable enrichment of a band in the 25 kD region for the depleted fractions of plasma and serum (Figures 5.7 A and B). The 15 kD region band present in the crude plasma lanes were barely visible in the depleted fractions in both gels. These differential observations are important in that they could be used to target proteins of interest if depletion or enrichment is required, depending on the experimental goal. Observations from the in-solution digestion workflow mirrored those of in-gel digestion. Most importantly, a considerable number of proteins were present in the beads after the standard elution recommended by the manufacturer of ProteoMiner™.

In this ProteoMiner™ workflow, depleted plasma was a superior substrate to serum with respect to the number of proteins identified following tryptic digestion, as evidenced by ProteinPilot and Mascot search engine searches. This observation should not be interpreted as plasma being superior to serum or vice-versa; rather, it simply means that ProteoMiner™ was better-optimised for plasma than for serum, as per the manufacturer's recommendation that it should be used on plasma samples.

The off-gel fractionation workflow was evaluated using crude and acetone-precipitated sheep serum as well as 1D SPD-PAGE. Despite being smudged, the Coomassie staining of the 1D SDS-PAGE gel bands had a characteristic pattern depending on the off-gel fraction analysed (Figures 5.8 A to F). The contents of the buffers, including the relatively high urea concentration in the fractions commonly used as per the protocol of this workflow, possibly contributed to the diffuse appearance of the lane boundaries and smudging. Nevertheless, there were distinct protein bands in the 71 kDa region for crude serum gels from fractions CS10–CS22. For example, one particular band was most intense in the C11 fraction and then had faded gradually by the CS22 fraction. There were 1–2 other faint bands around the same 71 kDa region in fractions CS13–CS16. These 71 kDa-region bands were not present in the gels loaded with acetone-precipitated serum fractions. This suggests that acetone precipitation played a role in depleting the albumin fraction responsible for the appearance of bands in this region. Even though 1D SDS-PAGE data were not shown in the study by Maheswarappa and colleagues[344], a single band in the 71 kDa region was observed in the present study.

The low number of proteins identified following in-gel digestion of the off-gel fractionated samples was reflected in the poorly resolved gel bands. Again, it is possible that the high urea content and the buffers could have interfered with tryptic digestion of the recovered intact proteins. For this reason, the samples were pooled together for protein database searching. This was also the case with in-solution digestion samples, for which the raw data sample files were pooled for protein sequence database searching. In addition, off-gel fractionation workflow was considered exploratory, with a view to including this workflow in the proteomics workflows of the QUT-

CARF laboratory. After this initial experiment, the next step would have been to evaluate peptides derived from the digestion of crude and acetone-precipitated serum by off-gel fractionation prior to nanoLC-nanoESI-MS/MS; however, this was outside the scope of the current work. For comparison, a similar evaluation would also be done for plasma. Nonetheless, this workflow contributed peptide data that enabled the identification of 84 proteins in Mascot (Figure 5.8 H) and 70 protein IDs in PeptideShaker (Figure 5.9) for the purposes of inclusion in the PSL in Chapter 6. In a poorly defined serum proteome such that of sheep, 70 high quality protein IDs is considered substantial as compared to an earlier study[111]. From information available in UniProtKB, the three proteins that were unique to the Off-Gel workflow in PeptideShaker identifications were W5P627 – an uncharacterised protein pointing to GSN gene with possible molecular functions of actin, calcium and myosin II binding; P68116 – a characterised fibrinogen beta chain protein derived from FGB gene with known biological processes that include adaptive immune response, blood coagulation and innate immune response; and W5PZI0 – an uncharacterised protein pointing to LOC101113728 gene locus which is possibly clusterin that is known to be a chaperone protein in human literature. Considering that an interest of this thesis was about characterising the proteome of the liquid fraction of sheep blood, with specific emphasis on the application of the PSL to identify proteins that take part in inflammation (Chapter 7), the identification of a characterised protein involved in blood coagulation that was not picked up by other workflows, albeit only one, was in itself, a significant addition to the PSL. This means that there should now be sufficient peptide spectral data to enable the development of an MRM assay for identifying ovine fibrinogen beta chain protein (P68116) in future. To illustrate this observation further, cattle for example, currently have nearly 6,000 characterised proteins in UniProtKB as opposed to only 457 sheep proteins. By taking this close species comparison into consideration, it was necessary to retain every unique protein ID derived from this workflow to contribute to the PSL, without dismissing the entire experiment, despite the sub-optimal overall number of protein ID yields that could have been otherwise interpreted as a negative experimental outcome, as compared to protein ID yields from ProteoMiner™ and ACN workflows. It is also important to note that there are currently no other comparable studies have used this workflow on sheep serum to be benchmarked on.

This chapter was probably the most successful chapter in the entire thesis but there could still be reservations regarding the strategy that was adopted. The ideal strategy should have been the depletion of abundant proteins followed by a high-level fractionation strategy. Presently, there are no antibody based protein depletion products that have been produced for sheep. However, a previous generation of plasma depletion products based on cibachrom blue and protein A/G are available (ProteoPrep® Immunoaffinity Albumin & IgG Depletion Kit, (PROT-IA)). At the outset,

this project implemented the analysis of the undepleted circulating acellular proteome of sheep in the foundation study in Chapter 4 in order to have a baseline serum proteome data for comparison with fractionated protein samples. Therefore, this chapter was the cradle of experimenting on depletion and fractionation of the circulating acellular proteome samples of sheep using several known effective strategies that have been used in other studies, including horses[68] and cattle[69,70] and pigs[70]. This ambition could be explored further in future if there is access to sufficient financial resources to purchase these kits, which in part, could explain the missed opportunity for this thesis.

In summary, the experiments described in this chapter have provided the first in-depth insight into alternative sheep plasma and serum sample preparation approaches with the goal of protein identification by nanoLC-nanoESI-MS/MS that improved on traditional methods. In addition to using different protein sequence database search algorithms to ID proteins, each of the four experimental workflows added a novel dimension of sample preparation. The acetone-precipitation experiment added previously unreported protein ID data from supernatant fractions. The ACN experiments provided protein data due to the effects of higher pH and sample dilution, as well as analysing both fractions of ACN precipitation as factors that favour protein IDs from samples. The additional elution step for ProteoMiner™ showed that a considerable number of proteins were present in the beads after the standard elution recommended by the manufacturer. The off-gel fractionation experiments added data from 1D SDS-PAGE. In terms of protein yields using Mascot IDs as the standard, ProteoMiner™ enrichment contributed the highest number, followed by ACN precipitation, acetone precipitation and off-gel fractionation, in that order. Based on the observations made in this chapter, it is suggested that ACN precipitation and ProteoMiner™ workflows being the methods that worked best, should be preferentially included in future sample preparation workflows. Following on, in Chapter 6 is a description of the construction of the PSL from data drawn from this chapter and Chapter 4, which also results in the determination of the number of proteins from a composite of all the DDA experiments.

CHAPTER 6

6.0 Bioinformatics strategy for assembling an encyclopaedic peptide spectral library derived from plasma and serum samples of sheep

6.1 Abstract

This chapter describes the bioinformatics approach used for the construction of a PSL using peptide data derived from healthy sheep in Chapters 4 and 5. Data from diseased and endotoxin-treated sheep, as well as synthesised *in silico* predicted peptides of some proinflammatory cytokines, were analysed and added to generate a comprehensive PSL. The objective was to form a PSL that can be applied in targeted proteomics workflows to simultaneously analyse several proteins.

The first step in the identification of proteins using bioinformatics is the requirement for a protein sequence database for search algorithms to utilise. The PSL was assembled by searching MS/MS data from a composite of 501 input .mgf files derived from DDA experiments against a generated custom *Ovis aries* UniProtKB database using ProteinPilot™, resulting in the identification of 563 proteins and 41,288 distinct peptides from 3,195,890 spectra at 1% FDR. The same dataset was searched using Mascot search engine and validated using PeptideShaker software which identified 398 proteins. ProteinPilot™, Mascot and PeptideShaker, collectively identified 1,103 proteins from spectral data derived from both plasma and serum samples. Only the high-quality data from PeptideShaker identifications were retained for construction of the PSL.

These results represent the first encyclopaedic PSL constructed entirely from the liquid fraction of sheep blood, with data drawn from healthy and sick individuals. If implemented, this PSL could be of practical significance to veterinary science in future for identifying proteins of sheep and closely related species such as cattle and goats. The functions of these proteins and how they relate to each other can then help in understanding various physiological perturbations and to the scientific community seeking to use sheep as a model for studying human disease.

6.2 Introduction

This chapter describes a bioinformatics strategy for the construction of a PSL from the liquid fraction of sheep blood (the circulating acellular proteome) and its enrichment from DDA data. Bioinformatics refers to the use of various computer applications to organise and understand the molecular information in a biological context[350]. The generation of this PSL was motivated by the lack of such a tool for exploring large numbers of plasma and serum proteins. The overarching objective was to construct a comprehensive peptide spectral repository that could be applied to wide-ranging downstream MS/MS protein assays.

The previous chapters have identified a large number of peptide sequences and these now need to be associated with specific proteins. Once validated, a spectral library can permit identification of proteins in subsequent plasma or serum samples by using software tools capable searching MS/MS data to rapidly identify proteins. Spectral libraries can be also be utilised for validating pepetide and protein ID results of searches performed using tools such as ProteinPilot and X!Tandem.

Spectral library search strategy has emerged as a promising alternative for peptide identification, in which MS/MS spectra are directly compared against a reference library of confidently assigned spectra[351]. This approach relies heavily on a comprehensive set of known spectra that have been assigned to amino acid sequences[351]. The PSL is central to the discovery of novel peptides and proteins – a previously unobserved spectrum can be identified by comparing it to all prospective hits in the library for the most similar match[352]. This makes proteomics information derived from MS/MS a powerful tool for profiling proteins[6-8,351]. A well curated PSL can also improve the identification of known and novel proteins, including those with post-translation modifications (PTM) because spectra of unmodified peptides are in principle similar to the spectra of PTM-modified peptides[353].

This chapter brings together DDA data from Chapters 4 and 5 that were derived from analysing healthy sheep plasma and serum samples. The PSL was also enriched with data derived from analysing samples from sick sheep. Having data from healthy and diseased sheep is considered essential to assemble an inclusive and comprehensive PSL that can be a reference tool or databank for future targeted proteomic investigations, including those where pathophysiology is a major focus[351,354] – which is key to the quality and usefulness of the PSL.

Peptide data from plasma and serum of endotoxin-treated sheep from earlier studies[14,15] was also acquired. The inclusion of this sample set was crucial, because the capstone study (Chapter 7) employed the PSL to evaluate proteome profiles of sheep plasma samples of sheep exposed to endotoxin[14,15]. As a preparatory step, it was essential to acquire tryptic peptide data from samples of endotoxaemic sheep for inclusion in the PSL for downstream SWATH applications of the PSL.

Finally for the validation of the approach to protein identification and enrichment of this PSL, *in silico* predicted synthetic peptides of selected proinflammatory cytokines of sheep were injected into the instrument, analysed and added. Cytokines are regulators of mammalian responses to injury[17,355]. In human medicine, proinflammatory cytokines such as IL-1 and TNF-α are known to escalate the pathological effects[355], and this is expected to apply to sheep as well. Many cytokines have low abundancy in circulation and often escape identification by global proteomics approaches[356]. The detection of cytokines in the plasma and serum of veterinary patients has traditionally relied on species-specific antibody ELISAs as evidenced in studies in dogs,[357,358]

cats,[359] cattle,[360,361] and sheep[362], but these are not amenable to multiplexing[356]. Owing to the relatively considerable expense of using individual ELISAs for each protein assay in veterinary species, an attractive alternative is to generate spectral data from predicted unique peptides of selected sheep proinflammatory cytokines. This strategy makes the PSL potentially useful in simultaneously quantifying the levels of many cytokines in a single assay in future, by using experimentally derived spectral data not currently available in MS/MS repositories[356].

To assemble the PSL, all DDA data were analysed in ProteinPilot™ in the first instance. The handling of protein identification information after being derived from raw data using the Paragon™ algorithm[164] in ProteinPilot™ is detailed in this chapter. Data were exported for secondary protein identification using Mascot and X!Tandem search engines and the results of the searches conducted using the above three search engines were validated by PeptideShaker bioinformatic software. By using an established target-decoy approach, PeptideShaker by default, enables it to to merge peptide/spectrum matches of multiple search engines thereby giving it a better edge of improved confidence and sensitivity in peptide and protein ID outcomes as compared to using a single search platform[363].

6.3 Methods

6.3.1 Generation of a custom sheep UniProtKB protein sequence database

A step-by-step method that was used to generate a sheep-only protein sequence file from UniProtKB in FASTA format is illustrated in APPENDIX 6.0. This database provided protein sequence information that was utilised by search engines and protein ID validation software to assign peptide signals to spectra.

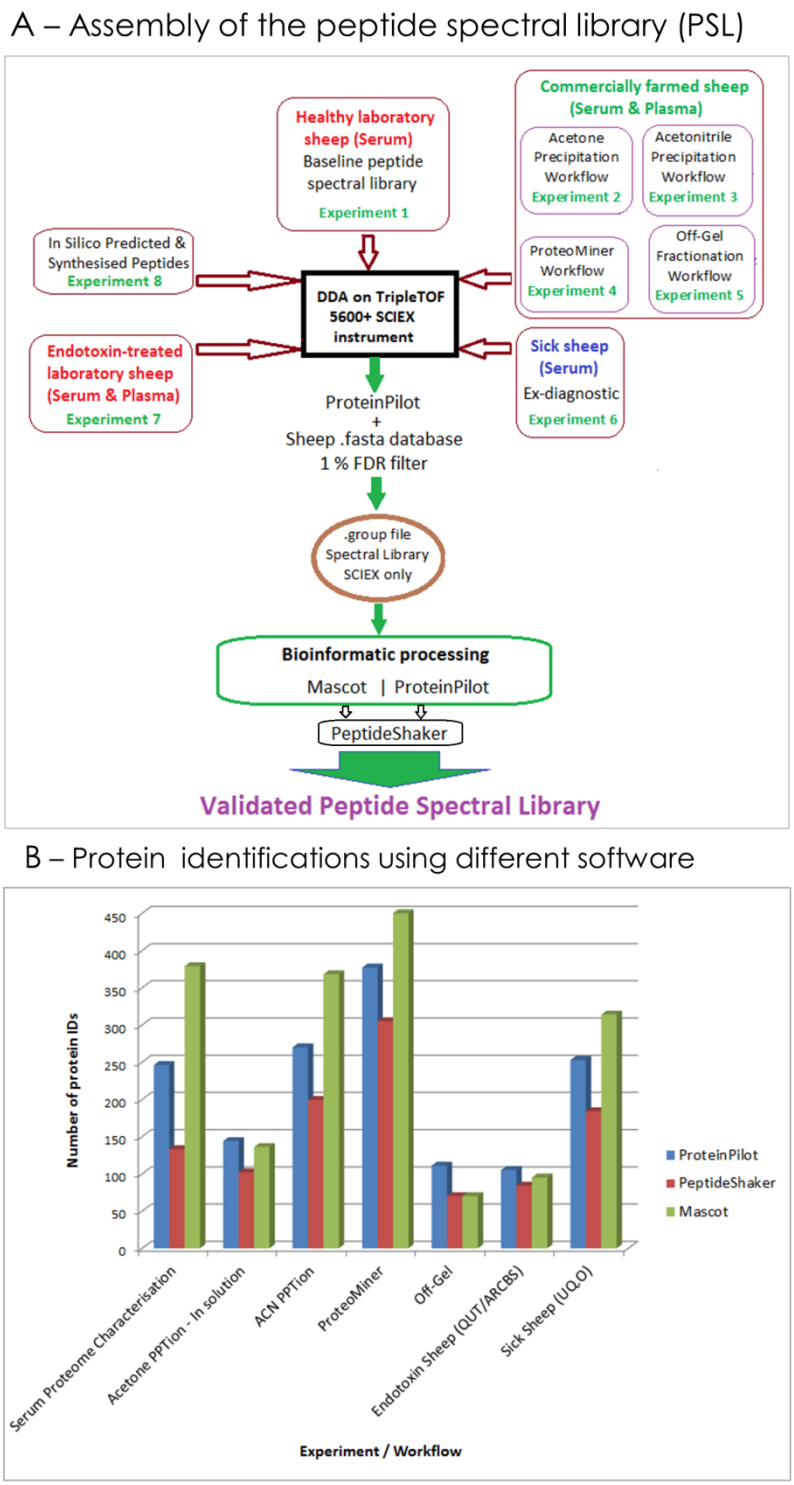
6.3.2 Data sources for the assembly of the peptide spectral library

The encyclopaedic PSL was assembled by combining DDA data derived from a composite of eight nanoLC-nanoESI-MS/MS experiments. In the first instance, ProteinPilot™ was used to search data from each experiment separately against the generated sheep protein sequence database. Data from the resulting individual group files were exported as .mgf files which were inturn searched concurrently in ProteinPilot™ software to generate a composite .group file, a corresponding .mgf file and a protein list. The individual .mgf files were also searched using Mascot search engine to generate a composite protein list and corresponding .dat files. PeptideShaker was then used to validate protein hits as illustrated in Figure 6.0, and detailed below.

6.3.2.1 Peptide data from healthy sheep plasma and serum samples

The foundation data for constructing the PSL was obtained from serum proteome characterisation[16]

in Chapter 4 which also represents the baseline PSL (Experiment 1) in Figure 6.0. Additional data were drawn from Chapter 5 in which there was enhanced peptide extraction from both plasma and serum samples (Experiments 2, 3, 4 and 5).



A – Schematic diagram of the experimental workflow for the derivation of data for the PSL using a Q-TOF (TipleTOF 5600+, SCIEX) instrument in a data dependent acquisition (DDA) mode. The PSL data were derived from eight experiments comprising of Experiment 1) a baseline PSL based on serum, Experiment 2) analysis of fractions of acetone precipitated sheep serum and plasma, Experiment 3) analysis of fractions of partial organic precipitation of sheep serum and plasma proteins using acetonitrile, Experiment 4) combinatorial peptide ligand library protein enrichment of sheep serum and plasma, Experiment 5) off-gel fractionation of serum proteins, Experiment 6) analysis of ex-diagnostic sheep serum Experiment 7) analysis of serum and plasma of endotoxin treated sheep, and Experiment 8) in silico predicted and synthesised unique peptides of five known acute phase proteins (cytokines).

B – The numbers of protein identifications from ProteinPilot, PeptideShaker and Mascot protein sequence database search engines

Figure 6.0 Experimental workflow – A; A summary of the number of proteins identified from the individual experiments in Panel A that contributed to the peptide spectral library – B.

6.3.2.2 Peptide data from sick sheep serum

Data from serum samples of sick sheep with a range of different conditions were obtained from UQ's School of Veterinary Science clinical pathology laboratory (Experiment 6). The samples were subjected to in-solution and in-gel digestion using generic methods as described Chapter 3 to yield peptides for analysis as presented in APPENDIX 6.1. There were 21 proteins that were identified in Chapter 4, but not in sick sheep serum suggesting that these proteins probably diminish during illness.

6.3.2.3 Peptide data from plasma and serum of endotoxin-treated sheep

Data from plasma and serum samples of sheep treated with endotoxin based on earlier observations[15] were added to the PSL (Experiment 7). The samples were subjected to 1D SDS-PAGE and in-solution digestion and analysed to extract peptide data as detailed in APPENDIX 6.2.

6.3.2.4 Data from *in silico* predicted synthetic peptides of selected proinflammatory cytokines

Peptide data of five candidate proinflammatory cytokines of sheep were selected as targets, based on reports in the literature[17,291,364-367], for inlcusion in the PSL (Experiment 8), by using an established approach[368,369]. Homologous peptide sequences between *Bos taurus* and *Ovis aries* of IL-6, IL-3, IL-1α, IL-1β and TNF-α were obtained from UniProtKB/Swiss-Prot or the non-redundant protein database of NCBI and imported into Skyline software[172]. The description of obtaining data from the selected peptides is presented in APPENDIX 6.3.

6.3.3 Validation of protein identifications in the PSL

6.3.3.1 ProteinPilot™ search

The .mgf files derived from the primary searches of the experiments in Figure 6.0 A were subjected to a composite ProteinPilot™ search using a sheep-only UniProtKB database as described in Chapter 3 to generate a single .group file – which is in itself, a PSL when using the SCIEX SWATH protein analysis pipeline, and a protein list.

6.3.3.2 Mascot search

The .mgf files from the primary search of all the experiments illustrated in Figure 6.0 A were submitted to Mascot Server v2.5.1 for searching using a Daemon client. The concatenated target/decoy database was configured to be compliant with PeptideShaker[245] as well, prior to being uploaded to the Mascot Server. The Mascot searches were conducted as described in Chapter 3 using the parameters shown in Figure 6.1. The same UniProtKB protein sequence database used by the ProteinPilot™ was used in these searches.

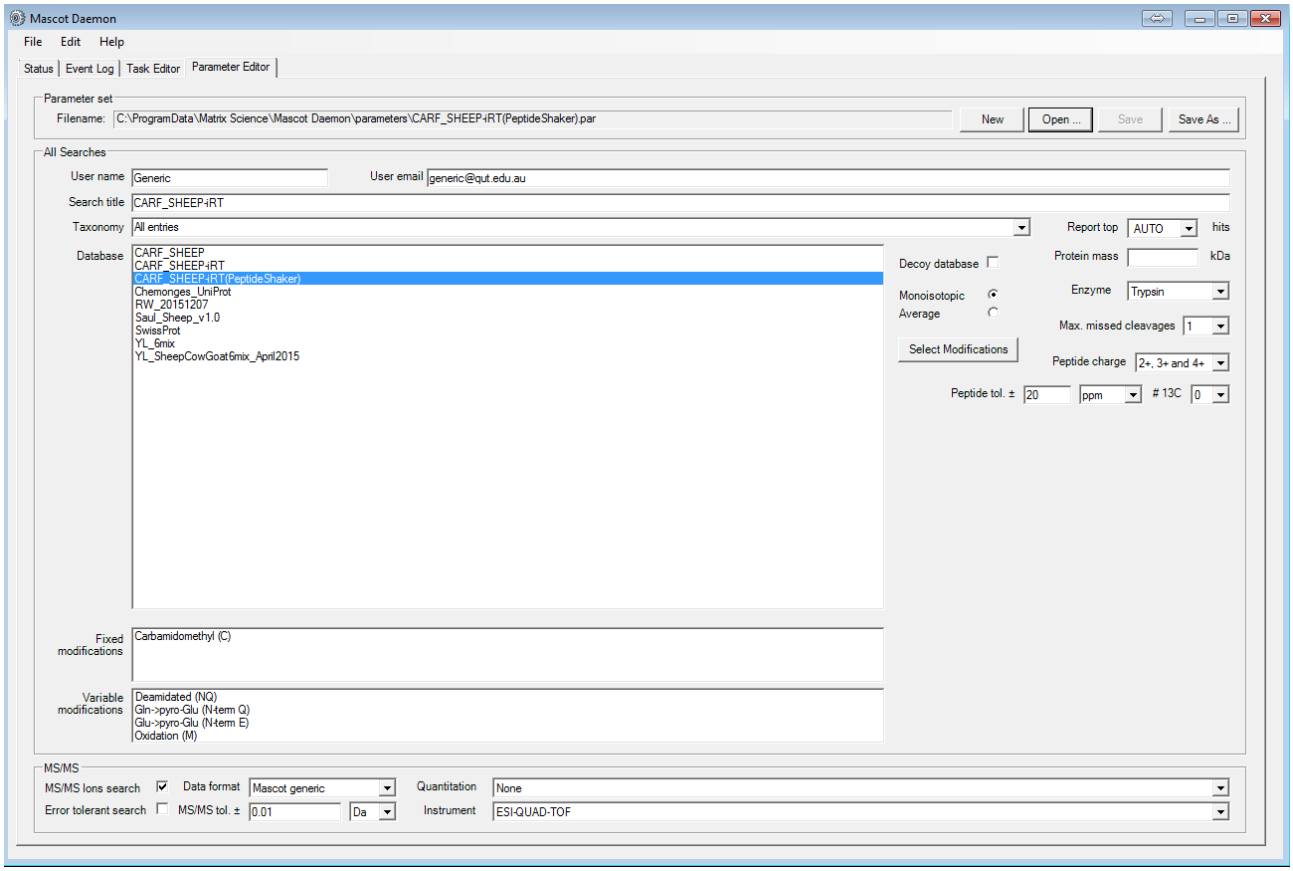


Figure 6.1. Mascot Daemon user interface showing submission parameters for searches in Mascot Server v2.5.1 using a concatenated target/decoy database that was configured to be compliant with PeptideShaker searches as well. Note that the Decoy database box was unselected – a requirement for the results output to be compatible with downstream PeptideShaker searches.

6.3.3.3 PeptideShaker search

The mzIdentML files from the primary searches in ProteinPilot™ in Figure 6.0 A and .dat files from Mascot searches were loaded into PeptideShaker to validate the protein IDs using the same UniProtKB database used by ProteinPilot™, by applying the same search parameters and protein identification criteria as described in Chapter 3. Native mzIdentML (.mzid) format files created in ProteinPilot™ were not supported in PeptideShaker. This required code substitution by way of editing the mzIdentML files in order for the downstream file processing in PeptideShaker to function (Figure 6.2). The protein ID results were exported into Microsoft® Excel™ for evaluation.

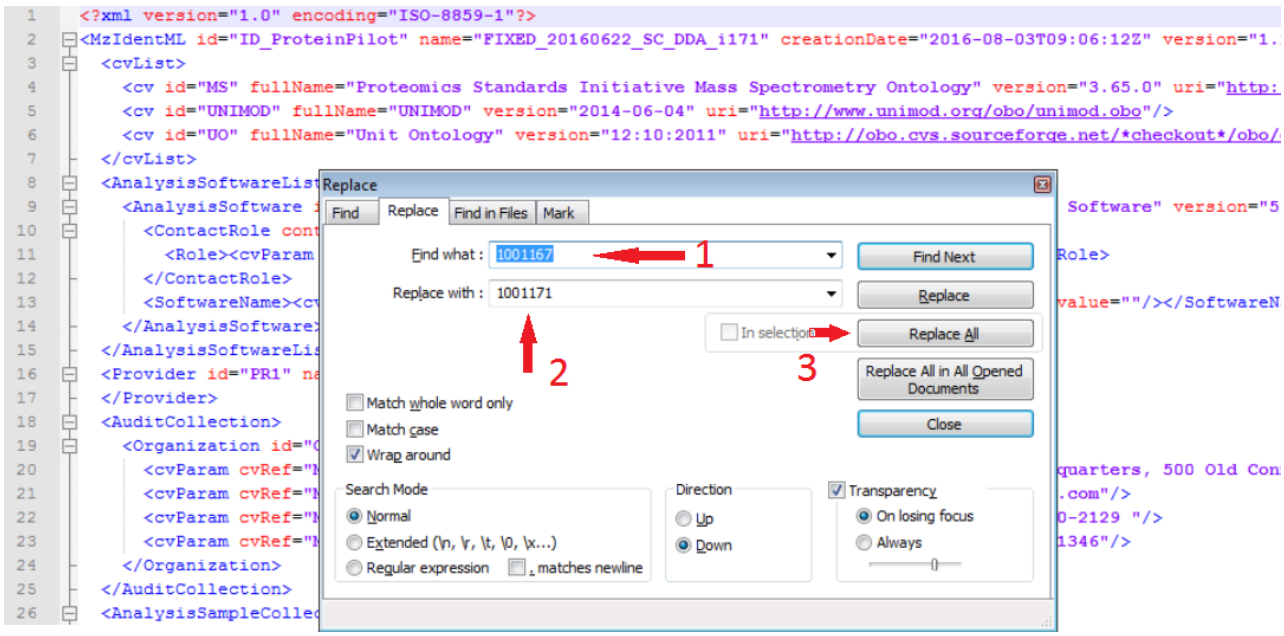


Figure 6.2. Fixing native mzIdentML file to work in PeptideShaker using Notepad++. According to a technical brief from SCIEX, PeptideShaker required the MS:1001171 score flag to identify peptide spectrum matches (PSM) scores in mzIdentML (*.mzid) files, meanwhile the native mzIdentML from ProteinPilot has MS:1001167. Therefore, MS:1001167 was replaced with MS:1001171 in all the mzIdentML files prior to loading into PeptideShaker. The mzIdentML files were opened in either Notepad++ editor (notepad-plus-plus.org) or (EditPro editor (Alentum Software Ltd) for larger files). In Notepad++, the score flag replacement was performed by going to 'Search' menu (or Ctrl+H) to activate the 'Replace' dialog above, followed by the steps 1, 2 and 3 (red arrows) and the files were saved and ready to be processed in PeptideShaker.

6.3.3.4 Analysis of synthetic peptides to validate protein identifications and search parameters

Synthetic peptide standards of five cytokines (two unique peptides per cytokine) were spiked into a subset of samples and analysed on the instrument and then appended to the library to confirm the validity of search parameters and the bioinformatics pipeline, for quality control of DDA experiments as described in APPENDIX 6.3. No peptides were missing in the results generated from all the three search platforms used, including PeptideShaker at 1% FDR, which confirmed the validity of bioinformatics pipeline and its parameters.

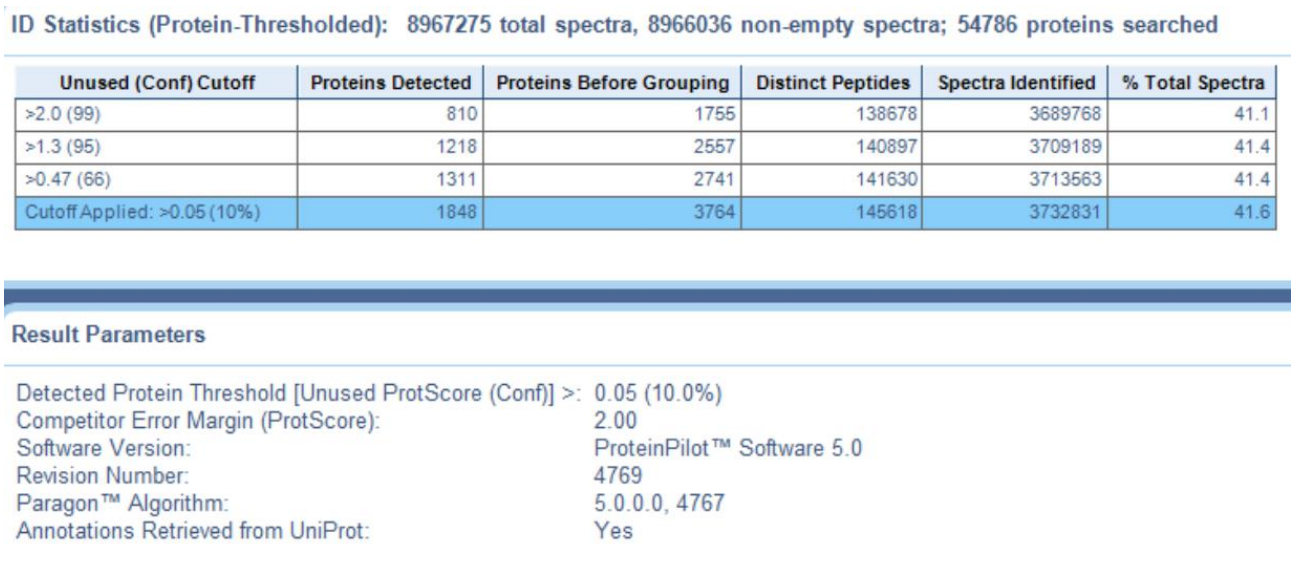
6.4 Results

6.4.1 Generation of a sheep-only custom UniProtKB protein sequence database

The custom concatenated target/decoy protein sequence database that was generated consisted of 54,786 sequences; 26,229,144 residues.

6.4.2 ProteinPilot™ search results

The protein identification statistics from searching all the input files from eight DDA experiments illustrated in Figure 6.0 in ProteinPilot™ are presented in Figure 6.3.



ID Statistics (Protein-Thresholded): 8967275 total spectra, 8966036 non-empty spectra; 54786 proteins searched

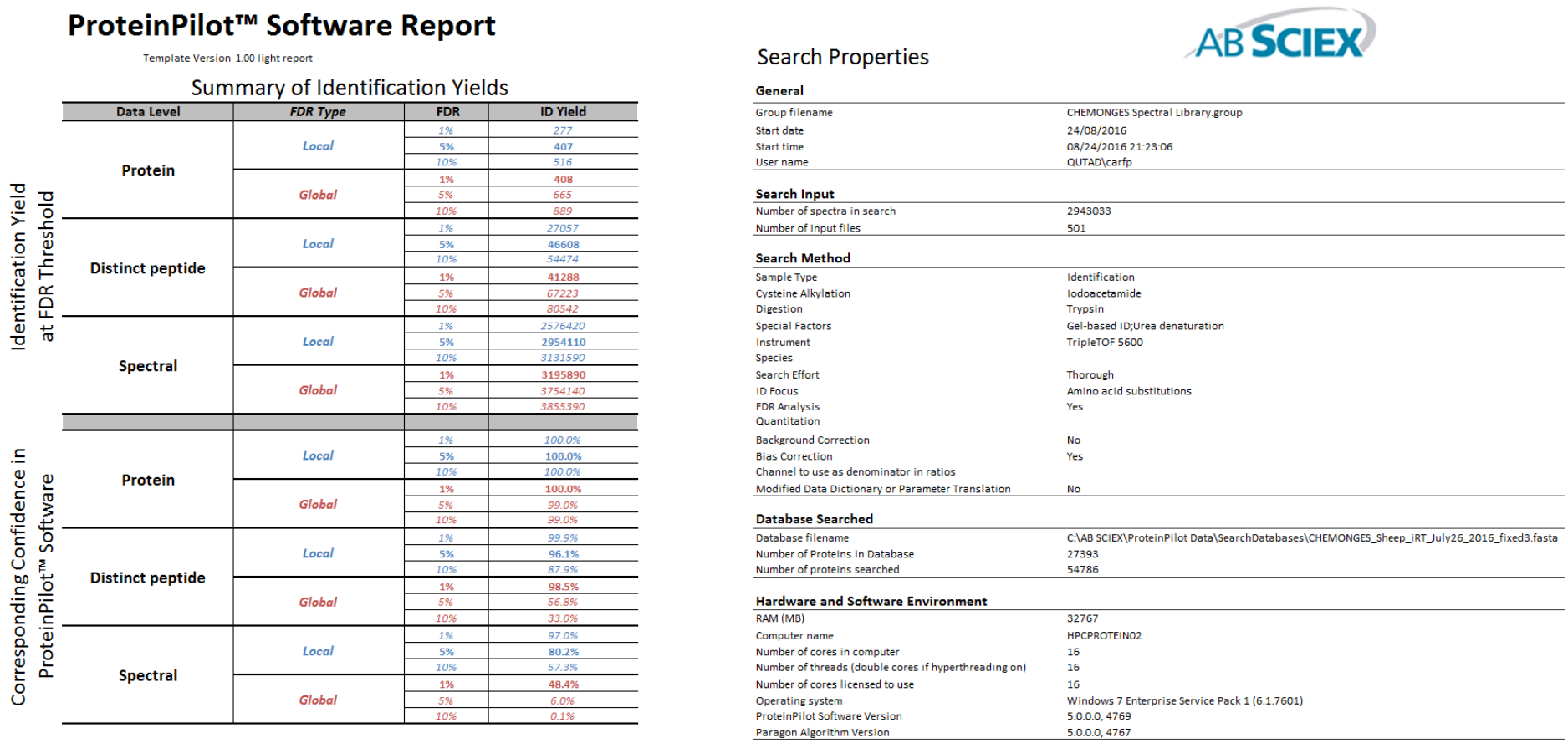
Unused (Conf) Cutoff	Proteins Detected	Proteins Before Grouping	Distinct Peptides	Spectra Identified	% Total Spectra
>2.0 (99)	810	1755	138678	3689768	41.1
>1.3 (95)	1218	2557	140897	3709189	41.4
>0.47 (66)	1311	2741	141630	3713563	41.4
Cutoff Applied: >0.05 (10%)	1848	3764	145618	3732831	41.6

Result Parameters

Detected Protein Threshold [Unused ProtScore (Conf)] >:	0.05 (10.0%)
Competitor Error Margin (ProtScore):	2.00
Software Version:	ProteinPilot™ Software 5.0
Revision Number:	4769
Paragon™ Algorithm:	5.0.0.0, 4767
Annotations Retrieved from UniProt:	Yes

Figure 6.3. Screenshot of protein identification statistics in ProteinPilot™ after searching 501 .mgf files from eight experimental workflows using a UniProtKB database for the assembly of the peptide spectral library. These results represent a crude overview that includes decoy protein hits and contaminants such as keratin. A decoy protein is a protein identified against a decoy (nonsense) protein sequence database. A contaminant protein is a protein that did not originate from the sample but was introduced during sample processing.

Within ProteinPilot™ software, reports were automatically created in spreadsheet format (Microsoft® Excel™) that included the protein search summary as illustrated in Figure 6.4.
The exported protein list was inspected and curated as described in the generic methods in Chapter 3. Briefly, proteins with zero confidence, decoy (reversed) protein identifications and hits from the cRAP database (http://www.thegpm.org/crap/) were removed (Figure 6.5) by using a similar method as previously described in another earlier study[13]. Only proteins with 2 or more high-scoring peptides were considered for further analysis.
Data were also exported in mzIdentML and .mgf formats to be used by Mascot and PeptideShaker searches for spectral and protein ID validation.



ProteinPilot™ Software Report

Template Version 1.00 light report

Summary of Identification Yields

Identification Yield at FDR Threshold

Data Level	FDR Type	FDR	ID Yield
Protein	Local	1%	277
		5%	407
		10%	516
	Global	1%	408
		5%	665
		10%	889
Distinct peptide	Local	1%	27057
		5%	46608
		10%	54474
	Global	1%	41288
		5%	67223
		10%	80542
Spectral	Local	1%	2576420
		5%	2954110
		10%	3131590
	Global	1%	3195890
		5%	3754140
		10%	3855390

Corresponding Confidence in ProteinPilot™ Software

Protein	Local	1%	100.0%
		5%	100.0%
		10%	100.0%
	Global	1%	100.0%
		5%	99.0%
		10%	99.0%
Distinct peptide	Local	1%	99.9%
		5%	96.1%
		10%	87.9%
	Global	1%	98.5%
		5%	56.8%
		10%	33.0%
Spectral	Local	1%	97.0%
		5%	80.2%
		10%	57.3%
	Global	1%	48.4%
		5%	6.0%
		10%	0.1%

Search Properties

General

Group filename	CHEMONGES Spectral Library.group
Start date	24/08/2016
Start time	08/24/2016 21:23:06
User name	QUTAD\carfp

Search Input

Number of spectra in search	2943033
Number of input files	501

Search Method

Sample Type	Identification
Cysteine Alkylation	Iodoacetamide
Digestion	Trypsin
Special Factors	Gel-based ID;Urea denaturation
Instrument	TripleTOF 5600
Species	
Search Effort	Thorough
ID Focus	Amino acid substitutions
FDR Analysis	Yes
Quantitation	
Background Correction	No
Bias Correction	Yes
Channel to use as denominator in ratios	
Modified Data Dictionary or Parameter Translation	No

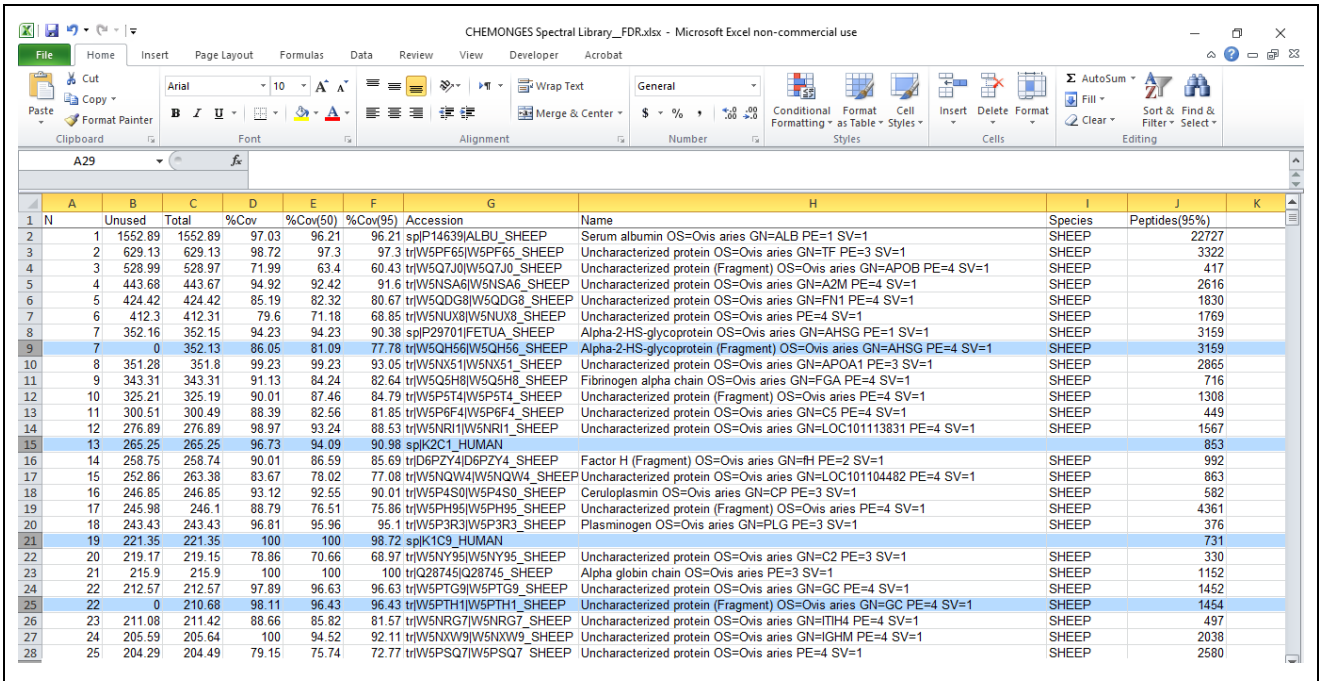
Database Searched

Database filename	C:\AB SCIEX\ProteinPilot Data\SearchDatabases\CHEMONGES_Sheep_iRT_July26_2016_fixed3.fasta
Number of Proteins in Database	27393
Number of proteins searched	54786

Hardware and Software Environment

RAM (MB)	32767
Computer name	HPCPROTEIN02
Number of cores in computer	16
Number of threads (double cores if hyperthreading on)	16
Number of cores licensed to use	16
Operating system	Windows 7 Enterprise Service Pack 1 (6.1.7601)
ProteinPilot Software Version	5.0.0.0, 4769
Paragon Algorithm Version	5.0.0.0, 4767

Figure 6.4 Protein search summary report in ProteinPilot™ Software showing identification yields and database search properties of assembled data from sheep serum and plasma of eight data dependent acquisition (DDA) experiments using a TripleTOF 5600 (SCIEX) instrument for the construction of a peptide spectral library (PSL). The protein identification yield at 1% global false discovery rate (FDR) was considered further. This is an automatically generated report.



CHEMONGES Spectral Library_FDR.xlsx - Microsoft Excel non-commercial use

N	Unused	Total	%Cov	%Cov(50)	%Cov(95)	Accession	Name	Species	Peptides(95%)
1	1552.89	1552.89	97.03	96.21	96.21	sp|P14639|ALBU_SHEEP	Serum albumin OS=Ovis aries GN=ALB PE=1 SV=1	SHEEP	22727
2	629.13	629.13	98.72	97.3	97.3	tr|W5PF65|W5PF65_SHEEP	Uncharacterized protein OS=Ovis aries GN=TF PE=3 SV=1	SHEEP	3322
3	528.99	528.97	71.99	63.4	60.43	tr|W5Q7J0|W5Q7J0_SHEEP	Uncharacterized protein (Fragment) OS=Ovis aries GN=APOB PE=4 SV=1	SHEEP	417
4	443.68	443.67	94.92	92.42	91.6	tr|W5NSA6|W5NSA6_SHEEP	Uncharacterized protein OS=Ovis aries GN=A2M PE=4 SV=1	SHEEP	2616
5	424.42	424.42	85.19	82.32	80.67	tr|W5QDG8|W5QDG8_SHEEP	Uncharacterized protein OS=Ovis aries GN=FN1 PE=4 SV=1	SHEEP	1830
6	412.3	412.31	79.6	71.18	68.85	tr|W5NUX8|W5NUX8_SHEEP	Uncharacterized protein OS=Ovis aries PE=4 SV=1	SHEEP	1769
7	352.16	352.15	94.23	94.23	90.38	sp|P29701|FETUA_SHEEP	Alpha-2-HS-glycoprotein OS=Ovis aries GN=AHSG PE=1 SV=1	SHEEP	3159
7	0	352.13	86.05	81.09	77.78	tr|W5QH56|W5QH56_SHEEP	Alpha-2-HS-glycoprotein (Fragment) OS=Ovis aries GN=AHSG PE=4 SV=1	SHEEP	3159
8	351.28	351.8	99.23	99.23	93.05	tr|W5NX51|W5NX51_SHEEP	Uncharacterized protein OS=Ovis aries GN=APOA1 PE=3 SV=1	SHEEP	2865
9	343.31	343.31	91.13	84.24	82.64	tr|W5Q5H8|W5Q5H8_SHEEP	Fibrinogen alpha chain OS=Ovis aries GN=FGA PE=4 SV=1	SHEEP	716
10	325.21	325.19	90.01	87.46	84.79	tr|W5P5T4|W5P5T4_SHEEP	Uncharacterized protein (Fragment) OS=Ovis aries PE=4 SV=1	SHEEP	1308
11	300.51	300.49	88.39	82.56	81.85	tr|W5P6F4|W5P6F4_SHEEP	Uncharacterized protein OS=Ovis aries GN=C5 PE=4 SV=1	SHEEP	449
12	276.89	276.89	98.97	93.24	88.53	tr|W5NRI1|W5NRI1_SHEEP	Uncharacterized protein OS=Ovis aries GN=LOC101113831 PE=4 SV=1	SHEEP	1567
13	265.25	265.25	96.73	94.09	90.98	sp|K2C1_HUMAN			853
14	258.75	258.74	90.01	86.59	85.69	tr|D6PZY4|D6PZY4_SHEEP	Factor H (Fragment) OS=Ovis aries GN=fH PE=2 SV=1	SHEEP	992
15	252.86	263.38	83.67	78.02	77.08	tr|W5NQW4|W5NQW4_SHEEP	Uncharacterized protein OS=Ovis aries GN=LOC101104482 PE=4 SV=1	SHEEP	863
16	246.85	246.85	93.12	92.55	90.01	tr|W5P4S0|W5P4S0_SHEEP	Ceruloplasmin OS=Ovis aries GN=CP PE=3 SV=1	SHEEP	582
17	245.98	246.1	88.79	76.51	75.86	tr|W5PH95|W5PH95_SHEEP	Uncharacterized protein (Fragment) OS=Ovis aries PE=4 SV=1	SHEEP	4361
18	243.43	243.43	96.81	95.96	95.1	tr|W5P3R3|W5P3R3_SHEEP	Plasminogen OS=Ovis aries GN=PLG PE=3 SV=1	SHEEP	376
19	221.35	221.35	100	100	98.72	sp|K1C9_HUMAN			731
20	219.17	219.15	78.86	70.66	68.97	tr|W5NY95|W5NY95_SHEEP	Uncharacterized protein OS=Ovis aries GN=C2 PE=3 SV=1	SHEEP	330
21	215.9	215.9	100	100	100	tr|Q28745|Q28745_SHEEP	Alpha globin chain OS=Ovis aries PE=3 SV=1	SHEEP	1152
22	212.57	212.57	97.89	96.63	96.63	tr|W5PTG9|W5PTG9_SHEEP	Uncharacterized protein OS=Ovis aries GN=GC PE=4 SV=1	SHEEP	1452
22	0	210.68	98.11	96.43	96.43	tr|W5PTH1|W5PTH1_SHEEP	Uncharacterized protein (Fragment) OS=Ovis aries GN=GC PE=4 SV=1	SHEEP	1454
23	211.08	211.42	88.66	85.82	81.57	tr|W5NRG7|W5NRG7_SHEEP	Uncharacterized protein OS=Ovis aries GN=ITIH4 PE=4 SV=1	SHEEP	497
24	205.59	205.64	100	94.52	92.11	tr|W5NXW9|W5NXW9_SHEEP	Uncharacterized protein OS=Ovis aries GN=IGHM PE=4 SV=1	SHEEP	2038
25	204.29	204.49	79.15	75.74	72.77	tr|W5PSQ7|W5PSQ7_SHEEP	Uncharacterized protein OS=Ovis aries PE=4 SV=1	SHEEP	2580

Figure 6.5. Inspection of a protein list output from ProteinPilot™ Software for the peptide spectral library showing the removal of proteins (light blue rows) with zero (0) unused confidence and hits from the common Repository of Adventitious Proteins (cRAP) database (http://www.thegpm.org/crap/).

After curating the protein list manually, ProteinPilot™ yielded 564 protein IDs from the entire PSL based on maximum FDR of 1% and a minimum of 2 high-scoring distinct peptides per protein for identification. The full list of these protein IDs is presented in APPENDIX 6.4. Only 25 of the 564 (4.4%) sheep proteins identified by ProteinPilot had been reviewed in UniProtKB. The UniProtKB entry numbers of the reviewed proteins are P14639, P29701, Q9XT27, P12303, P02083, Q7M2U8, P00922, A2SW69, P49920, O77642, Q28579, P02190, P11839, Q29400, Q06435, P67976, P62262, Q5MIB5, P09670, P68253, P49929, Q9XSM0, P23383, P50413 and P62297.

6.4.3 Mascot search results

After Mascot searches, protein identification lists at significance threshold of $p<0.05$ which equates to 1% FDR were exported via the report builder option in csv format for the immediate evaluation of the protein IDs as shown in Figure 6.6. The csv files were opened in spreadsheet (Microsoft®, Excel™) for inspection and removal of cRAP database proteins before merging the output files. Only proteins identified by at least 2 or more high-scoring and unique peptides were considered for further analysis. Also, .dat files were exported for further processing in PeptideShaker workflow as required. After merging all the results from the Mascot searches, a total of 830 proteins were identified. The UniProtKB entries for the identified proteins are presented in APPENDIX 6.5.

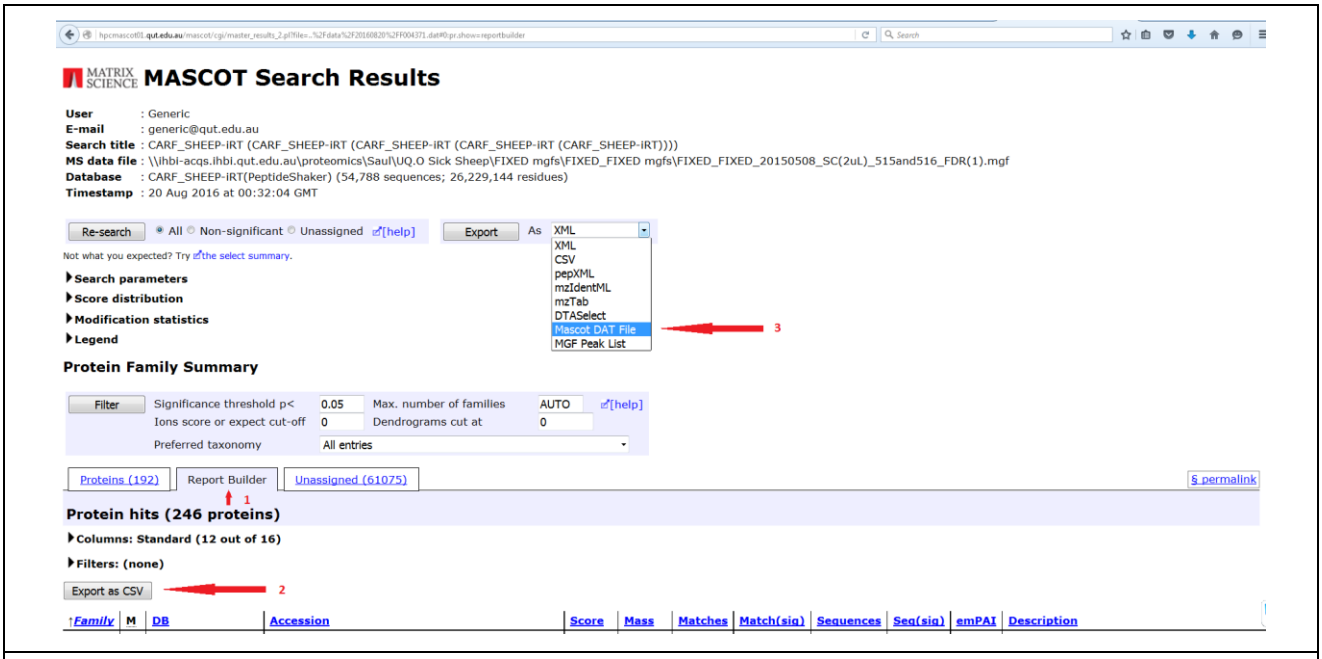


Figure 6.6. Mascot search results user interface showing the procedure (red arrows) for exporting protein lists in csv and .dat file formats. Using a significance threshold of p<0.05 (in the format controls) or target decoy of 1% FDR (in decoy search summary) to generate protein IDs lists, the Report Builder tab was clicked (1), followed by Export as CSV tab (2) directly into Microsoft® Excel™ for evaluation. The Export tab was clicked again to export Mascot .dat file (3) and .mgf peak list file to be utilised by PeptideShaker as required.

6.4.4 PeptideShaker search results

The protein lists from the searches in PeptideShaker of the individual experiments in Figure 6.0 were merged in Microsoft® Excel™. A total of 398 proteins were identified based on 1% FDR and 2 high-scoring validated unique peptides per protein. The full list of UniProtKB entries for the identified proteins is presented in APPENDIX 6.6.

6.4.5 Consensus protein identifications from ProteinPilot, Mascot and PeptideShaker searches

The protein lists from ProteinPilot™, Mascot and PeptideShaker searches were compared as illustrated in Figure 6.7. The UniProtKB entries of the cumulative 1,103 proteins identified in the PSL by the three search engines are presented in APPENDIX 6.7.

6.4.6 Data archive for the PSL

The MS data, with validated spectra along with the identification results were exported from PeptideShaker and deposited to ProteomeXchange Consortium[231] via the proteomics identifications (PRIDE) partner repository[269] with the dataset identifiers PXD005002 and DOI: 10.6019/PXD005002. The URL is https://www.ebi.ac.uk/pride/archive/projects/PXD005002 and can be accessed with username reviewer82838@ebi.ac.uk and password woBane0z.

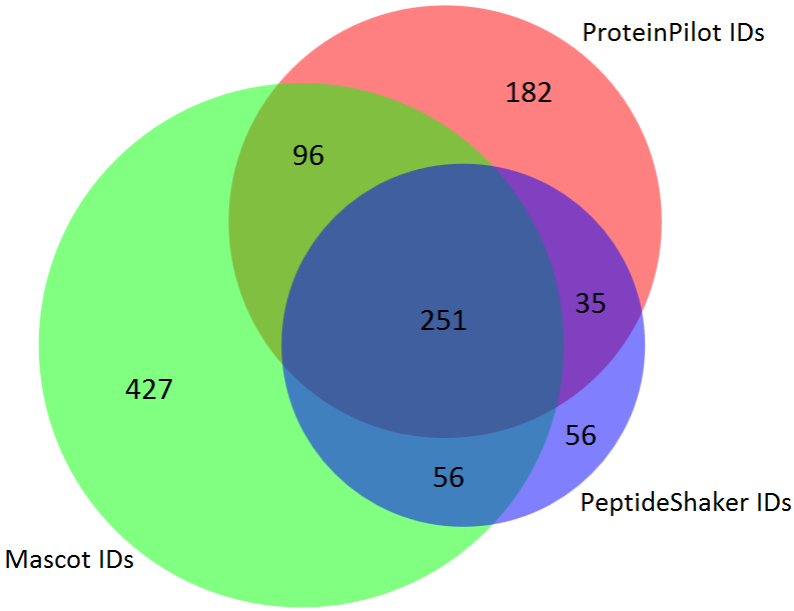


Figure 6.7. Comparison of the results of ProteinPilot™, Mascot and PeptideShaker searches showing protein identification yields using data from sheep serum and plasma of eight data dependent acquisition (DDA) experiments using a TripleTOF 5600 (SCIEX) instrument for the construction of the peptide spectral library (PSL). A total number of 1,103 proteins were identified collectively using all the three software platforms by searching subset of a sheep-only UniProtKB protein sequence database. Only the 398 validated high-quality protein identifications from PeptideShaker were retained for the PSL.

6.5. Discussion and conclusion

The initial part of this chapter presented details for the generation of a custom sheep-only UniProtKB protein sequence database by providing a step-by-step approach. Readers will find this useful, because many studies either do not provide explicit details for creating their in-house protein sequence databases for conducting protein searches, or it is probably assumed that it is common knowledge. The resulting database had fundamental use in the various protein sequence search algorithms employed in this thesis. As of October 2016, only 455 (1.67%) sheep proteins in the entire UniProtKB protein sequence database of 27,293 proteins had been reviewed (APPENDIX 6.0, Step No. 3). This was a small number of proteins compared, for example, to the bovine protein sequence database that had 6,870 out of 32,159 (21.4%) reviewed proteins[20]. Apart from having a reviewed status, the number of characterised sheep proteins was even lower. Despite the largely unreviewed and uncharacterised sheep proteins in UniProtKB, the utilisation of the principle of conserved genes between related species enables the use of the reasonable assumption that the known proteins and genes in the ox for example, are analogous to those of sheep[20,254,370].

Both canonical and isoform protein sequences were selected for inclusion into the database during its generation. The preceding strategy of protein sequence inclusion was to broaden the database due to the exploratory or method-development nature of the project of this thesis. In addition, given that the sheep genome annotation is still a work in progress[230,273], information on sheep protein functionality is still largely deficient[339] and it is therefore reasonable to utilise as much related protein data available as possible to identify sheep proteins. The use of databases containing not only the canonical, but also known isoforms of proteins has a small impact on the number of reported proteins[371]. Protein isoforms are however known to complicate protein inference during downstream bioinformatics analysis, as noted by the developers of PeptideShaker[245].

The reviewed proteins belong to the Swiss-Prot subset of UniProtKB and their protein identification tags have the suffix '>sp' in the FASTA database[20,189]. All the unreviewed proteins belong to the TrEMBL subset of the UniProtKB, and have the suffix '>tr' in the FASTA database. Each protein in this chapter was identified by their UniProtKB subset tag (sp/tr) followed by the entry number, accession number, protein name (where available), organisms species (OS), gene (GN) evidence of protein existence (PE) and sequence version (SV)[20,189]. According to UniProtKB, there are five levels of evidence for PE; 1 = protein level means the protein has been detected by mass spectrometry and/or ELISA, 2 = detection of protein transcript, 3 = inference from homology, 4 = predicted protein, and 5 = uncertain[20]. For example, the notation for Transforming growth factor beta-1 FASTA is >tr|W5PC52|W5PC52_SHEEP Transforming growth factor beta-1 OS=*Ovis aries*, GN=TGFB1 PE=3 SV=1. Similarly, the notation for Tumour necrosis factor FASTA is >sp|P23383|TNFA_SHEEP Tumour necrosis factor OS=*Ovis aries* GN=TNF PE=2 SV=2.

It was necessary to append FASTA sequences of protein contaminants in the database to enable them to be identified and filtered out during the generation of protein lists for performing data analysis[372]. Here, cRAP FASTA sequences (http://www.thegpm.org/crap/index.html) were used. There are also other repositories whose contaminants could have been used, however, these are focused mainly on human and yeast samples for example the CRAPome expandable repository of contaminants[372].

The second part of this chapter concerned the validation of protein identifications in the PSL using alternative software platforms and also by analysing synthetic peptide standards. Mascot and PeptideShaker bioinformatic platforms were also used to analyse the same dataset. The PSL was enriched with peptide data obtained from sick sheep serum (APPENDIX 6.1). This workflow that used ex-diagnostic sick sheep serum captured a different protein set to that from healthy sheep, thereby supporting the hypothesis that there will be proteins not captured from healthy sheep. According to Mascot results, this workflow, that included sick sheep and healthy sheep (control),

contributed peptide data that enabled the identification of 314 proteins for enriching the PSL .

Peptide data was also obtained from archived plasma and serum samples of endotoxin-treated sheep from a published model for transfusion-related acute lung injury (TRALI)[14] and from samples of the incremental arm of that study[15] (APPENDIX 6.2). Peptide recovery from these samples was lower than expected compared to the other experiments discussed here that also contributed data to enrich the PSL. It was for this reason that this experimental set was repeated in order to rule out the influence of other factors such as sample preparation. It is probable that there were substances in the samples that interfered with peptide extraction, considering that the samples were obtained while the sheep were receiving total intravenous anaesthesia with a cocktail of anaesthetics[10], which is an important finding in this thesis. The substances in the anaesthetic cocktail comprised of alfaxalone, ketamine, fentanyl and butorphanol, which may have affected protein extraction, possibly by inhibiting trypsin. A possible hypothesis is that these interfering substances could have possibly even caused protein proteins to alter structurally in a manner that prevented the ability of trypsin to function normally to enhance protein digestion[373]. The presence of these substances in the samples may have co-precipitated with the proteins during acetone precipitation, thereby causing the poor definition of protein bands in most of the 1D SDS-PAGE experiments (APPENDIX 6.2). Another suggested possible hypothesis to explain the low number of proteins extracted from this workflow is that the anaesthetic agents could have interfered with the BCA protein assay (a simple assay but more prone to interferences than other protein assays), leading to an overestimation of protein concentration in the samples. If less protein was being prepared for digestion, then the amount of peptides detected in the samples would be less. But then it can also be argued that this was unlikely to have been the case because 1D SDS-PAGE gel images clearly showed that there were sufficient quantities of protein loaded in the gels, given that every gel had a visible BSA control band that had far less amount of protein (5,000 fm) loaded. Considering that this study utilised archived samples from an unrelated study conducted many years ago by the ARCBS/QUT and UQ's School of Medicine, there was no opportunity to have control samples from non-anaesthetised sheep. Consequently, experiments should be performed in future that include spiking the suspected anaesthetic agents into identical plasma samples to determine if this observed effect is reproducible. Also, it would be interesting to have both a positive control, such as a molecular marker band from the same gel as the samples, and a negative control such a blank gel piece, analysed together to rule out any technical issues[373] prior to determining the presence of interfering substances in the samples. Alternative digestion strategies such as FASP digestion[233], could also be used to remove interfering substances from the samples. Overall, this endotoxin-treated sheep workflow contributed peptide data of at least 100 protein IDs for PSL enrichment.

Considered collectively, the sick sheep and end endotoxaemic sheep workflows added 80 proteins to the PSL pool as shown in Figure A6.2(b) (APPENDIX 6.2). This therefore underscores the rationale for having added data from non-healthy sheep to contribute to the PSL.

Constructing an effective PSL that can potentially have capabilities of being applied to identify several proteins simultaneously necessitates having information of certain known peptides of targeted proteins; for example, by selecting only unique peptide sequences to a single cytokine and also peptide fragmentation information, a suitable data acquisition strategy and instrumentation. The representative results of one out of two top-ranking unique peptides each from five cytokines of sheep that were predicted *in silico* presented in APPENDIX 6.3 were full MS/MS mass spectra acquired on a TripleTOF 5600+ instrument identifying each of the unique peptides of the selected cytokine present in the PSL. The collected full MS/MS scans produced high-quality MS/MS spectra for the PSL. In addition, the labelled peptides will provide an additional robust internal standard for the peptides during future use of the library.

The goal of the *in silico* prediction of cytokine peptides was to generate peptide fragmentation data for inclusion in the PSL and to validate protein identifications. During inflammation or infection, members of the proinflammatory cytokine families of IL-1, IL6 and TNF-α are released from macrophages and monocytes, which in turn stimulate the acute phase response, leading to a chain of events in the inflammatory cascade, while IL-3, which belongs to colony stimulating factor family, is known to have overlapping functions that collectively modulate inflammation[17]. Cytokines are biological mediators of physiological responses to disease and the immune system[17,374,375]. They recognise and modulate the response of an organism to infection by regulating activation, replication, chemotaxis and apoptosis of immune cells[17].

The MS/MS spectra of the *in silico* predicted peptides were ideal as they all resulted in validated spectrum annotation matches by being recognised as unique peptides to the selected cytokines. Having the ability to identify endogenous low-abundance selected cytokines (for example, in samples derived from sheep plasma and serum using the PSL alongside targeted proteomics bioinformatics software platforms)[356], is a substantial achievement of the present work. The presence of spectral data for the selected peptides makes the library considerably valuable, as it can be used to identify these selected important markers of inflammation using targeted proteomics approaches that could potentially rival antibody ELISA technology in future.

An important feature of this chapter was the description of the fine details of the assembly of the PSL that illustrated the validation of 398 proteins by PeptideShaker from a pool of 1,103 proteins collectively identitied by other software. This strategy brought together data of interesting proteins from the three individual search platforms that resulted in a total of 438 proteins that were identified

by at least two search engines. Of these protein identifications, 251 were highly confident hits as they were identified by all three search engines. The other 187 protein identifications were moderate confidence hits. The remaining 665 protein low confidence hits that were identified by a single search engine only (Figure 6.7). Whilst the merging of protein identifications of several search platforms does not always increase the number of proteins identified, this strategy is supported because it increases the number of peptides per protein[371]. It is however quite possible that other search engines could identify further subsets of proteins in the nascent PSL in addition to what was identified by ProteinPilot, Mascot and validated by PeptideShaker. When interpreting the results from multiple protein identification search platforms, it was important to ensure that protein hits were truly unique other than from the redundant information in the FASTA file. The datasets were also processed using correct input parameters and by applying FDR statistics.

In conclusion, the bioinformatics strategy and data assembly processes described in this chapter constitutes the first encyclopaedic PSL constructed entirely from the liquid fraction of sheep blood, with peptide data drawn from healthy and sick animals. The use of more than one search engine enabled the identification and validation of an unprecedented number of proteins in the PSL. If implemented, this PSL could be of practical significance to veterinary science in understanding various physiological perturbations involving proteins in the liquid fraction of blood, and to the scientific community seeking to use sheep as a model for studying human disease. This nascent PSL was needed for protein quantitation using SWATH-MS in Chapter 7 and it could also be useful for proteogenomics applications in future.

CHAPTER 7

7.0 Application of the peptide spectral library for identifying plasma proteins involved in early-phase acute systemic inflammation

7.1 Abstract

Identifying proteins involved in systemic inflammation in plasma is fundamental to diagnosis and monitoring response to treatment. The availability of targeted protein data extraction using SWATH-MS analysis provides a prospect for developing a platform capable of detecting numerous proteins in plasma simultaneously, many of which are not routinely measured. In this chapter, SWATH-MS analysis was used to evaluate the effectiveness of the PSL developed in Chapter 6 to characterise the response in sheep to endotoxaemia. Plasma samples were available from a study in which sheep were injected with endotoxin (lipopolysaccharide (LPS)) from *Escherichia coli* and blood collected (before the plasma separated) before and after 75 min after LPS exposure. The proteins in these plasma samples were identified by LC-SWATH-MS, with 243 proteins able to be quantified in plasma from healthy and endotoxaemic sheep. Forty-one of these proteins are known to be involved in the inflammatory process, while a further 42 proteins may also be involved in early-phase acute systemic inflammation. This is the first time SWATH-MS analysis has been used to distinguish proteins that may be used to diagnose and characterise early-phase acute systemic inflammation in sheep.

7.2 Introduction

Sheep have an important role in food, fibre and scientific research as models for human disease[10,376,377], yet there is only limited information about inflammation-related plasma proteins in this species. There are some reports that have proposed methods for identifying circulating proteins involved in early-phase acute systemic inflammation in sheep or related species using SWATH-MS[17]. Having the ability to profile multiple proteins involved in inflammation is fundamental in understanding acute pathophysiology in animals[17,22,110,111,378,379].

Prior to this, ovine antibody immunoassays have been used for protein detection in plasma samples[21,22,380,381]. However, this is cumbersome, relies on a different kit for each protein (and therefore is expensive) and often has to be validated for the species of interest. A proteomics approach using a PSL permits large numbers of proteins to be simultaneously identified and quantified, significantly increasing the ability to monitor changes in pathophysiology[382]. In this chapter, a subset of the novel PSL developed as part of this thesis in Chapter 6 was used to identify proteins and potentially their alterations in plasma of sheep during early-phase experimentally induced endotoxaemia. Here, SWATH-MS[383,384] approach using SWATH™ MicroApp Software

(SCIEX) for data extraction alongside the PSL, was applied to interrogate a small number of archived plasma samples from a separate study[14,385] in which sheep were exposed to acute *Escherichia coli* lipopolysaccharide (LPS) endotoxaemia as the disease model in order to achieve a comprehensive analysis of proteins[1,383] and to determine if plasma proteins involved in early-phase acute inflammation can be detected. The intent was to use these proteins to distinguish between normal sheep and sheep with endotoxaemia as a proof of concept. As a prospect, this approach could potentially complement antibody-based methods for protein detection and quantitation[188]. Additionaly, the MS/MS data will be much welcome in aiding the characterisation of sheep proteins in publicly available protein sequence repositories for future biotechnological applications.

7.3 Methods

7.3.1 Background of the experimental sheep selection and endotoxin treatment procedures

The plasma samples used for this study were obtained from a primary study whose animal ethics clearance details are found in the generic methods section in Chapter 3. The experimental sheep selection and procedures were as previously described[15] and included here as APPENDIX 7.0. In brief, the primary study used LPS to prime the immune system after allocating sheep to receive an infusion containing LPS from *E. coli* serotype O55:B5 (Sigma-Aldrich, Castle Hill, NSW, Australia)[14]. The samples were obtained opportunistically from archived samples stored during the study. Following on here is the proteomics analysis of the archived plasma samples from healthy and endotoxaemic sheep experiments of that study[14] as cited earlier[15].

7.3.2 Sample selection, preparation and nanoLC-nanoESI-MS/MS analysis.

The primary studies from which the plasma samples were derived from did not have MS-based proteomics analysis as part of the experimental protocol[14,385]. Six sheep per group (3 groups in total) were sampled from all the sheep enrolled in the original studies[14,385], based on the observations of the report in APPENDIX 7.0. This number of samples was determined with the consideration of acquiring representative samples across the cohorts of the different sheep traits (Wool production, Parasite resistance and the unknown sheep trait (SYSS))[15]. The samples comprised of pre- and post-endotoxin treatment time-points (Table 7.0). The samples were taken at baseline and after 75 min of LPS endotoxaemia[15]. Sample aliquots containing 100 µg of plasma protein were prepared for protein analysis by following the generic in-solution digestion of proteins workflow and desalted prior to nanoLC-nanoESI-MS/MS analysis on a TripleTOF® 5600 instrument using SWATH-MS™ acquisition method as described in the in Chapter 3.

Table 7.0. Sheep plasma samples used in SWATH-MS analysis experiments.

Sheep ID and Trait	Plasma Samples (Lab No./[protein](µg/µL) vol. (µL))	
	Before LPS Rx	**After LPS Rx**
90 (Wool)	SC569 (86.4) 1.3	SC570 (77.6) 1.3
97 (Wool)	SC573 (68.8) 1.5	SC574 (74.4) 1.3
100 (Wool)	SC577 (68) 1.5	SC578 (76.8) 1.3
101 (Wool)	SC581 (57.6) 1.7	SC582 (88) 1.1
102 (Wool)	SC585 (82.4) 1.2	SC586 (84.8) 1.2
103 (Wool)	SC589 (73.6) 1.4	SC590 (88) 1.1
66 (Parasite)	SC593 (72) 1.4	SC594 (65.6) 1.5
67 (Parasite)	SC597 (84) 1.2	SC598 (80) 1.3
71 (Parasite)	SC601 (74.4) 1.3	SC602 (76.8) 1.3
72 (Parasite)	SC605 (67.2) 1.5	SC606 (72) 1.4
74 (Parasite)	SC609 (73.6) 1.4	SC610 (69.6) 1.4
75 (Parasite)	SC613 (56) 1.8	SC614 (85.6) 1.2
602 (SYSS)	SC617 (84) 1.2	SC618 (69.6) 1.4
710 (SYSS)	SC621 (76) 1.3	SC622 (77.6) 1.3
719 (SYSS)	SC625 (83.2) 1.2	SC626 (84.8) 1.2
729 (SYSS)	SC629 (82.7) 1.2	SC630 (67.2) 1.5
814 (SYSS)	SC633 (94.4) 1.1	SC634 (87.2) 1.1
815 (SYSS)	SC637 (90.4) 1.1	SC638 (58.4) 1.7

Key: Table headers represent sheep identities (ID), production traits, laboratory sample numbers (Lab No.), protein concentration [protein] (µg/µL) in parentheses after acetone precipitation and dissolution of the protein pellet in 8 M urea/25 mM NH4HCO3 and volume (vol) (µL) that contained 100 µg of plasma protein from 18 sheep in groups of six, based on strain/experimental category showing samples taken before and after E. coli LPS endotoxin treatment (Rx).

7.3.3 Set-up of the SWATH-MS analysis experiment and workflow

The plasma samples representing the time-points before and after treatment with endotoxin of each individual sheep in Table 7.0 of the primary experiment together with the summary of the workflow are provided in Figure 7.0.

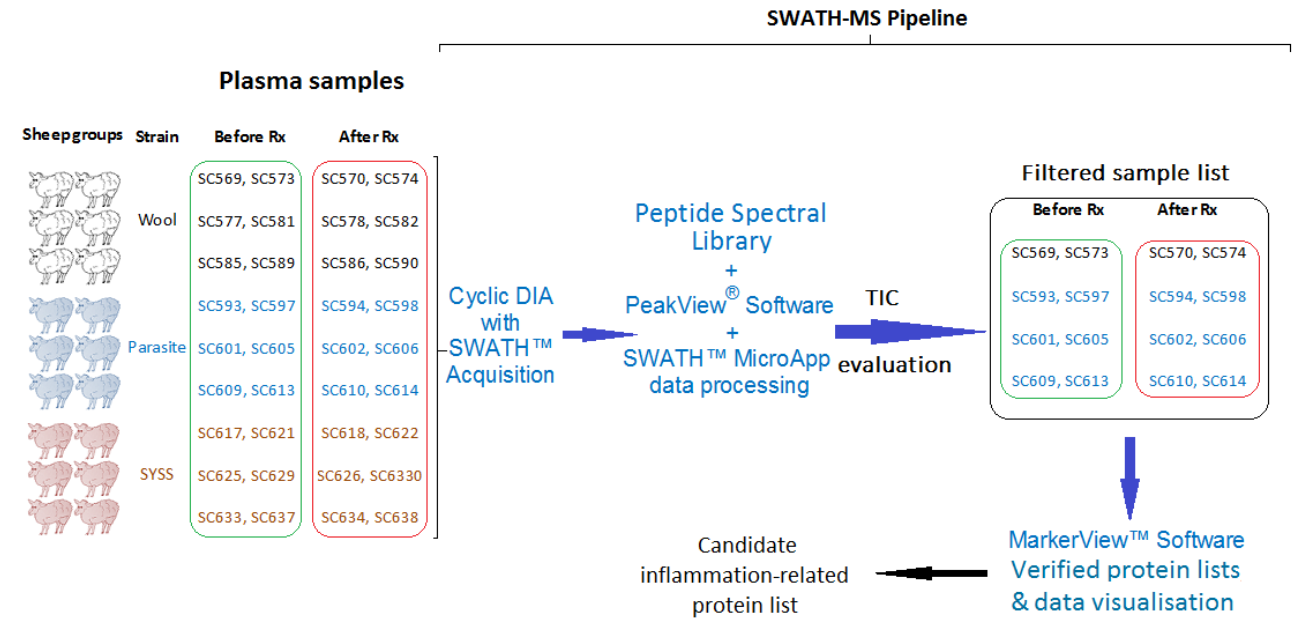


Figure 7.0. Identification of candidate early-phase inflammation-related proteins in plasma samples of endotoxaemic sheep. Samples were obtained from 18 adult merino ewes that were grouped based on strain cohort: Wool production (Wool), Parasite resistance (Parasite) and Stockyard sourced sheep of unknown strain (SYSS) from a study in which the subjects were treated with *E. coli* lipopolysaccharide (LPS) endotoxin[15]. The samples represented baseline time point i.e. the one before treatment (Before Rx) and the other after treatment (After Rx). Each of the samples were individually analysed using the SWATH-MS pipeline. Only eight paired samples met the criteria for further comparative analysis for protein quantitation during this proof of concept study. The individual raw data files (.wiff and .wiff.scan) appear in QUT Proteo dive archive with SC in each sample name replaced by the prefix 20150624_SC_SWATH-MS_.

This study aimed to analyse plasma samples from 18 experimental sheep representing one time point before and the other after endotoxin treatment which would have resulted in 36 samples in all. In the end however, only 16 samples derived from 8 sheep were used owing to considerably low peptide signals due to possible effects of anaesthetics (as discussed in Chapter 6) in a set of samples for an entire cohort of sheep and the TIC characteristics in some of the samples, leading to their exclusion from further analysis for protein identification (see Figure 7.0). The aberrant TICs could have been attributed to inadvertent contamination or to some extent due to batch effects[386,387] during sample collection in the primary study. The use of TICs to evaluate the similarity of analytical

peptide samples based on reproducible overlapping morphology of the traces is simple and straightforward and it has previously been used in SWATH-MS studies[188]. An alternative objective approach would have been to extract known peptides from the samples and evaluate their covariance in order to select the data files for further analysis. With further exclusion of samples that were considered outliers after PCA-DA, five pairs (10 samples) were used to the definitive protein quantitation.

7.3.4 Peptide spectral library clean-up and SWATH-MS analysis parameters

The PSL was imported into PeakView® using the SWATH-MS™ MicroApp in the first instance as described elsewhere[388]. This was followed by loading the .wiff files for SWATH-MS processing. The library was inspected for contaminants and decoy (reversed) proteins and unselected if they were found. A decoy protein is a protein identified against a decoy (nonsense) protein sequence database. A contaminant protein (such as keratin) is a protein that did not originate from the sample but was introduced during sample processing. Peptide retention times were recalibrated by selecting peptides identified with the best (cleanest) six transitions in their XICs (Figure 7.1) and mass spectrum showing MS/MS hits in the PSL (pink sticks) and extracted ions from SWATH data (blue sticks) (Figure 7.2) from proteins common across all samples and then applied to generate a calibration curve (APPENDIX 7.1). This was followed by entering the processing settings for filtering peptides. The SWATH-MS™ MicroApp is a plugin to PeakView® software that is embedded in the "Quantitation" tab on the menu. Navigating to, and clicking on SWATH-MS processing in the dropdown Quantitation menu activated the screen that is illustrated in APPENDIX 7.2.

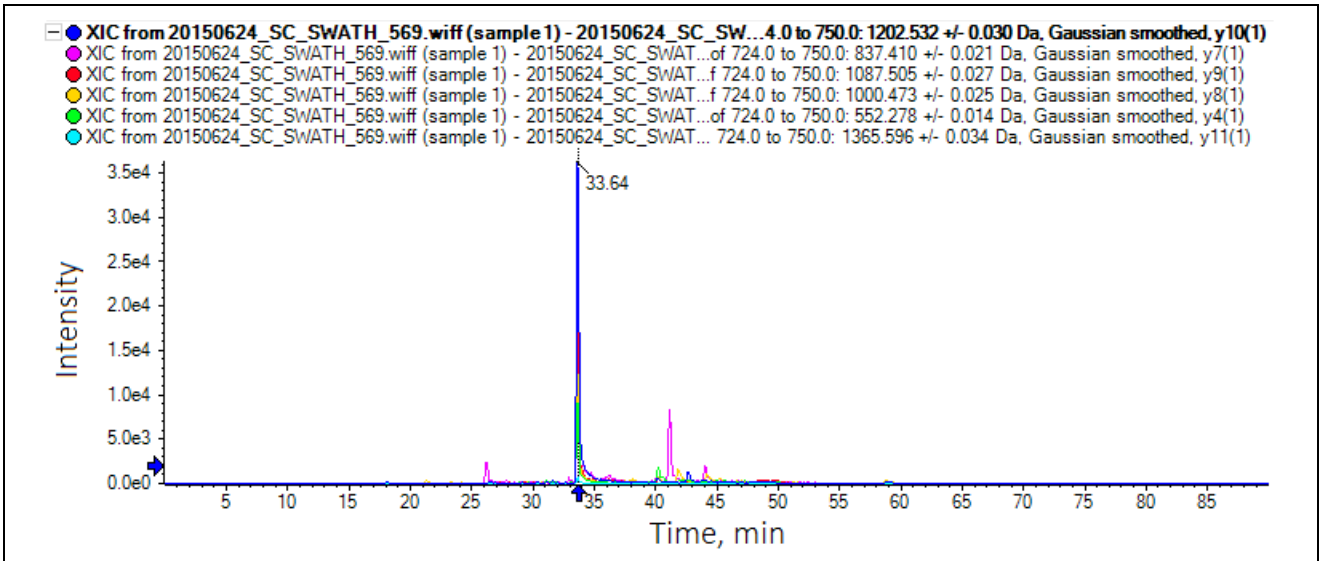


Figure 7.1. Example of XICs of a selected identified calibration peptide TYDSYLGDDYVR.

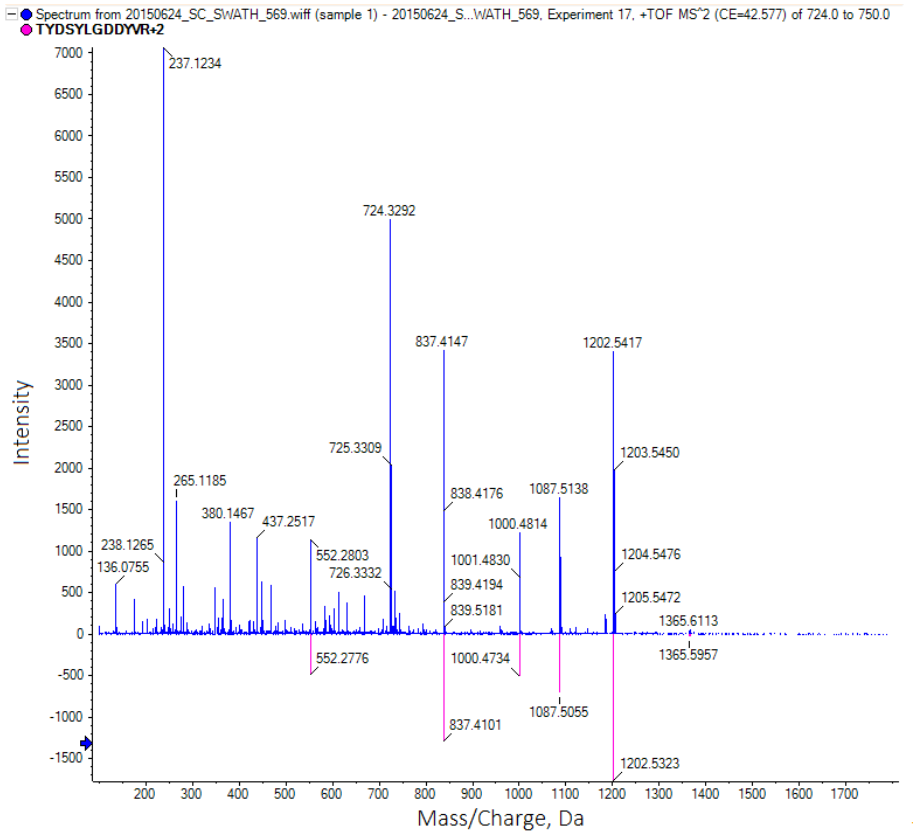


Figure 7.2. Tandem mass spectrum – MS/MS (pink sticks) of the doubly charged calibration peptide – TYDSYLGDDYVR that was used to extract corresponding ions from SWATH data (blue sticks)

7.3.5 Data analysis

Peptide and protein peak areas were automatically imported into MarkerView™ Software version 1.3 (SCIEX) for visualisation. Prior to further analysis, data were normalised using Median Peak Ratios feature. Each sample was normalised to the raw file SC_SWATH-MS_569.wiff that was derived from analysing the healthy sheep sample SC569 (reference sample) for the peaks which the samples had in common. The underlying assumption for this normalisation approach was that all samples contained many 'background' peaks in common with the respective reference sample[389]. For each sample, the ratios of the peak areas (for the sample and the reference sample) were calculated for all peaks which were greater than 1% of the largest peak. The median ratio from the previous step was then used as the scaling factor for each sample. In the event of less than ten peaks in common between the two samples, normalisation was not performed and the scale factor was set to 1.0.

After normalisation of the data, supervised principal component analysis also known as PCA discriminant analysis (PCA-DA) was performed. This approach was chosen because of prior knowledge of how the samples were treated (i.e. before and after endotoxin treatment) to define and

improve sample separation in the resulting scores plots. Data were displayed as charts. The PCA-DA scores plot displayed the scores for the first principal component against the second. Similarly the PCA-DA loadings plot displayed loadings for the first principal component against the second.

In order to identify proteins that potentially altered during endotoxaemia, peak areas of proteins constituted from individual peptides in the samples before, and after endotoxin treatment were compared using the paired *t*-test tool in MarkerView™ Software after automatically excluding outliers. Protein identification entries were exported in .xlsx format (Microsoft® Excel™) into spreadsheet. Data were displayed in the default MarkerView™ Software spreadsheet output with 18 protein-level categories as headings, namely: row (numerical identification), index (protein abundance hierarchy), peak name (UniProtKB protein accession number), *m/z* (not applicable in this dataset), retention time (not applicable in this dataset), protein group (protein name and status in UniProtKB), use (application of the protein data for statistical comparison – true), the conventional paired *t*-test outputs (*t*-value, *p*-value, mean 1, mean 2, median 1, median 2, sigma 1, sigma 2), delta (observed numerical change in protein concentration), fold change (a standardised measure of quantitative change in protein concentration, and log (fold change). The spreadsheet data format enabled user-level inspection, further curation and annotation of the protein IDs. Protein sequences of identified uncharacterised or unnamed proteins were retrieved from UniProtKB database and then analysed individually and named using the Basic Local Alignment Search Tool (BLAST) in NCBI database. For the scope and purpose of this thesis as a proof of concept study, only filtered data that included the UniProtKB accession number, gene name, NCBI name, protein status, UniProtKB name and fold change of individual proteins after treatment of the sheep with endotoxin from *E. coli* were presented in the results.

7.4 Results

7.4.1 Outcome of sample comparison and protein data

PeakView® Software was used for an initial examination of consistency and quality of collected data. The summed peptide signals (total ion chromatograms, TICs) of samples as displayed in this software are presented in Figure 7.3.

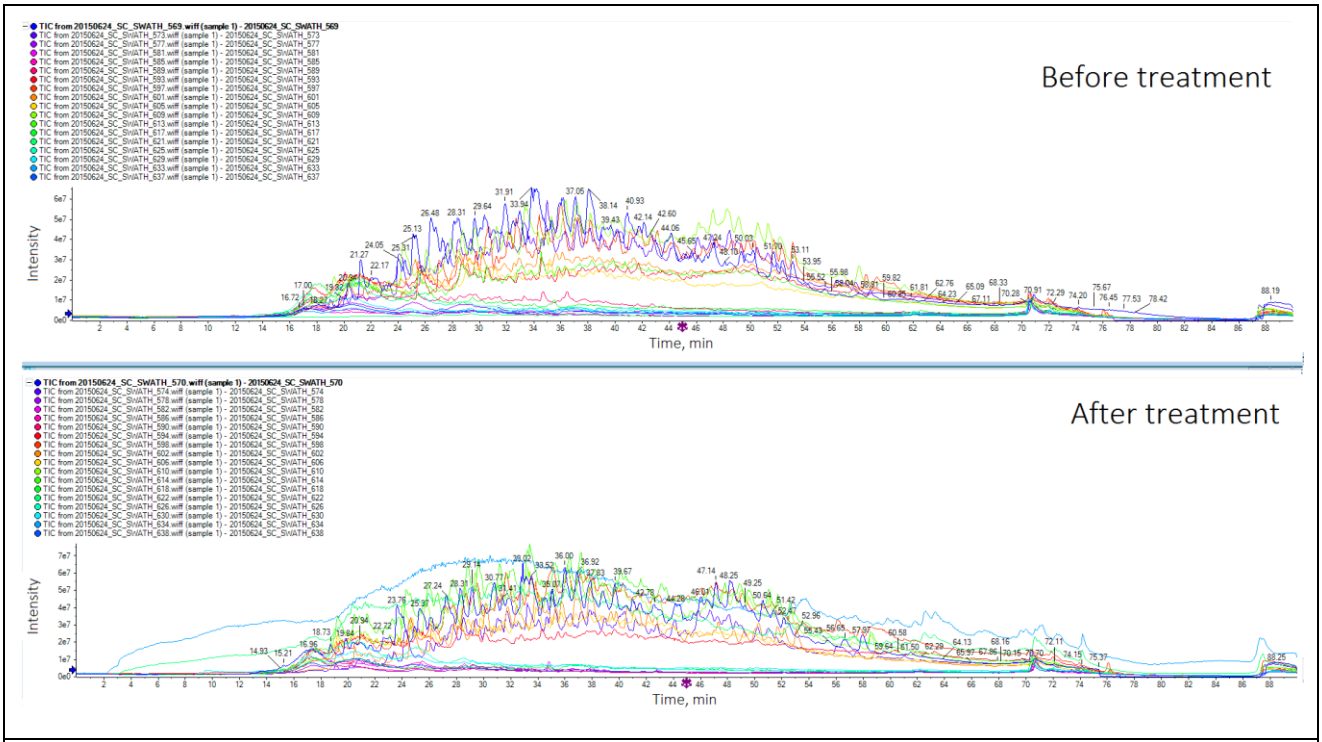


Figure 7.3. Total ion chromatograms (TICs) of peptides representing before (upper panel) and after (lower panel) endotoxin treatment in sheep plasma samples. There were 18 paired samples for comparison represented in both panels.

After all the raw data sample files were examined in the graphical interface of PeakView® software, it became apparent that there was considerable variability in the appearance of the TICs whereas more consistent, almost overlapping TIC signals were expected since the samples were derived from an original study that assumed that the study subjects and therefor the samples, had similar characteristics[14]. An interesting observation during the inspection of TIC morphology data was that most of the TICs with comparatively low peptide intensities that peaked at the hydrophilic end of the TIC were those of the SYSS cohort of sheep (Figure 7.3). Based on the preceding observation and the background that SYSS sheep had a different anaesthetic protocol[15], these data files and other files whose TICs differed considerably from the rest were excluded from further analysis. This resulted in the final analysis of eight paired samples (Figure 7.4).

In order to determine if there were any associations or correlation between the samples, the remaining 16 samples (8 pairs) that met the inclusion criteria based on TIC characteristics and uniform anaesthetic protocol were subjected to supervised PCA discriminant analysis (PCA-DA) in MarkerView™ Software as illustrated in Figure 7.5.

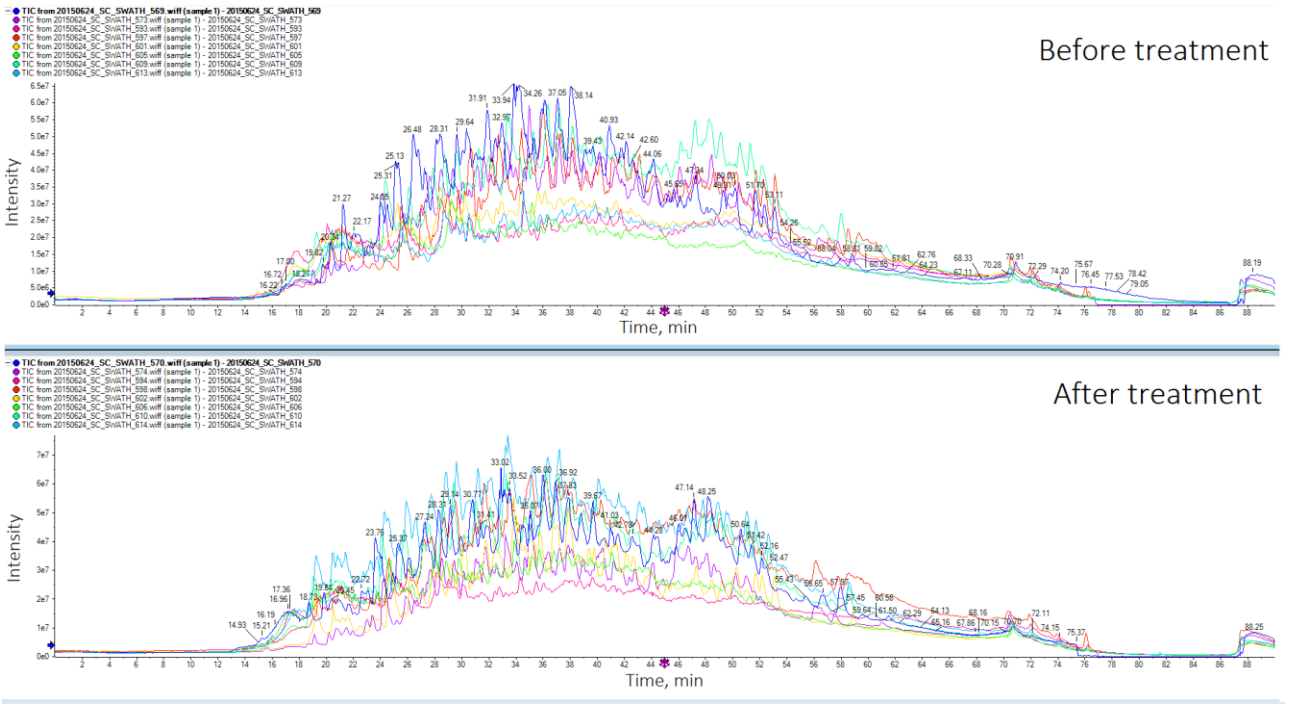
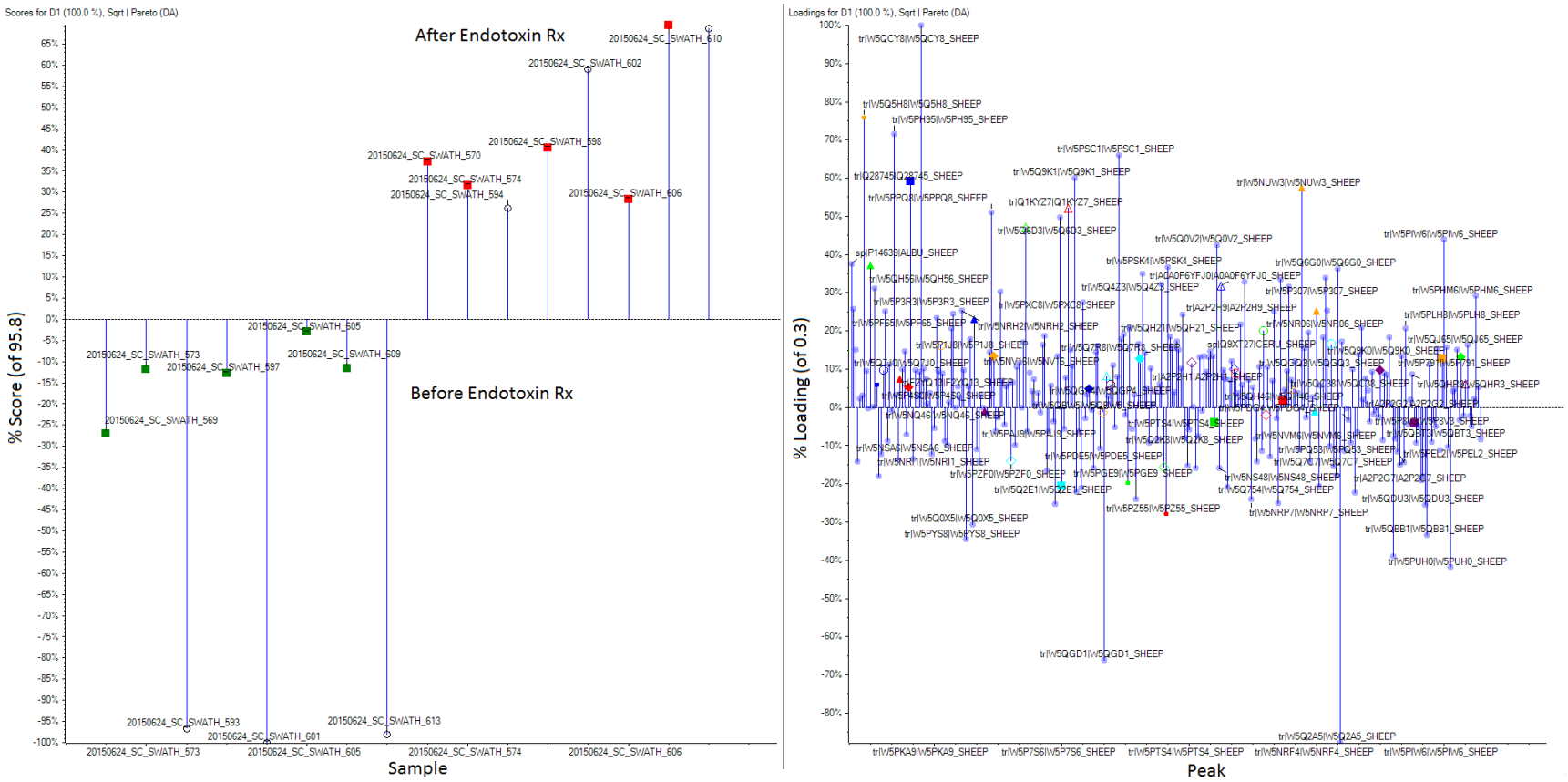


Figure 7.4. Total ion chromatograms (TICs) of peptides representing before (upper panel) and after (lower panel) endotoxin treatment in sheep plasma samples after filtering out aberrant and those with low peptide intensities. There were 8 paired samples for comparison represented in both panels.



A **B**

Figure 7.5. Supervised PCA of 16 SWATH-MS-analysed plasma samples and protein data in MarkerView® software after normalisation. The samples comprised of eight treatment pairs representing before and after endotoxin treatment (Rx) time-points. The score plot (A) summarises the relationship between the samples and the loadings plot (B) illustrates the protein peak patterns identified in the samples in the score plot. The data points on the PCA-DA scores plot were colour-coded showing before and after endotoxin treatment. The blank data points (while circles) were considered to be outliers based on plot characteristics. In the loadings plot, the peaks for after endotoxin treatment were denser that those before treatment.

7.4.2 Identification of proteins and their potential alterations in plasma of endotoxaemic sheep

Up to 243 proteins were identified and quantitated by SWATH-MS analysis by comparing plasma samples from healthy sheep (i.e. before treatment with endotoxin) and sheep with endotoxaemia. The signals of high-scoring individual peptides were extracted which enabled the identification of proteins and the comparison of their peak areas, for example a peptide of serotransferrin (Figure 7.6) and pentaxin (C-reactive protein) (Figure 7.7).

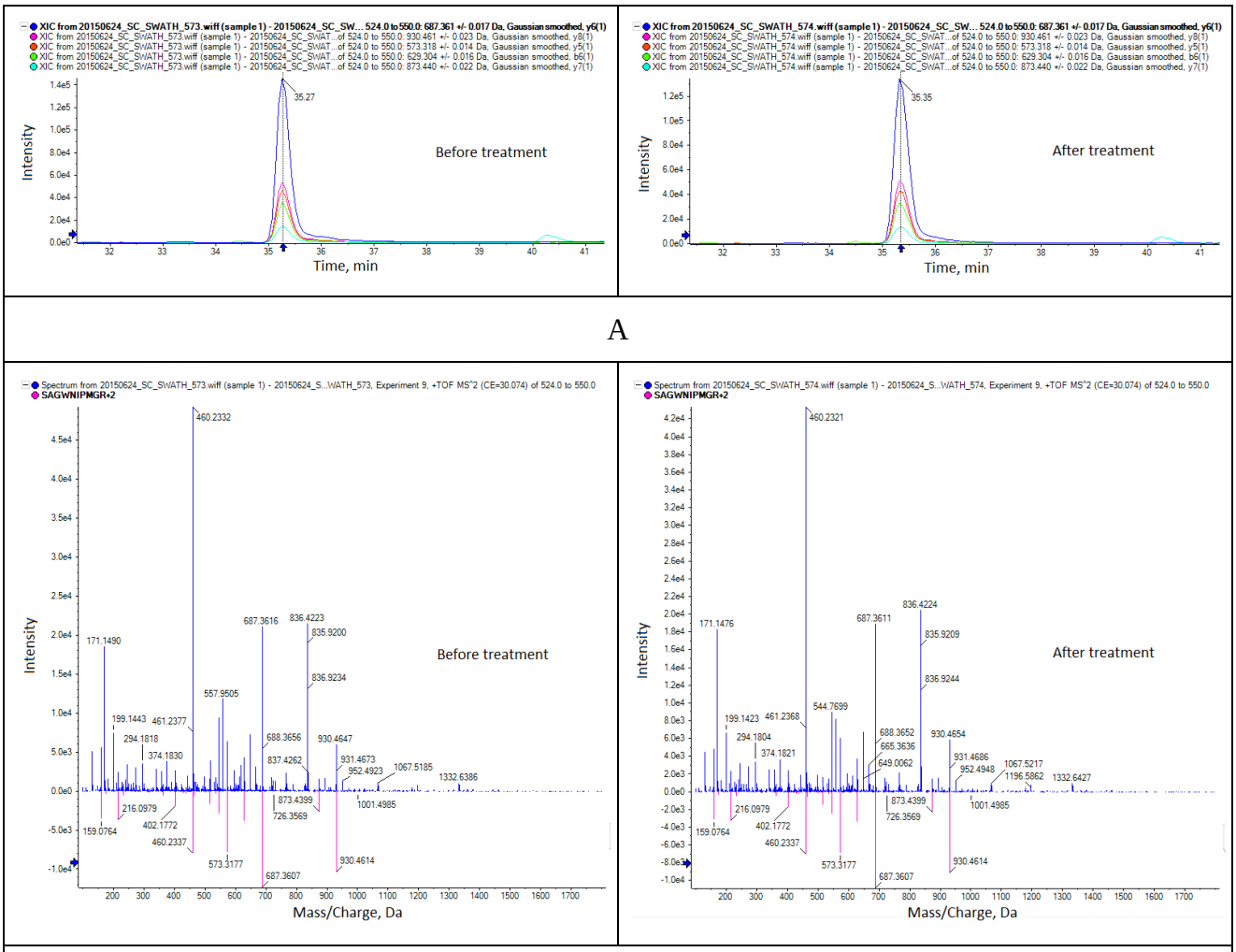
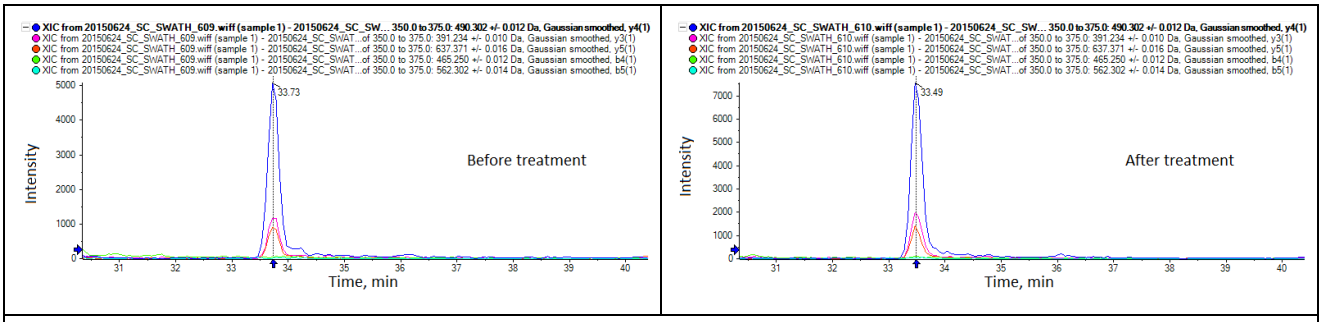
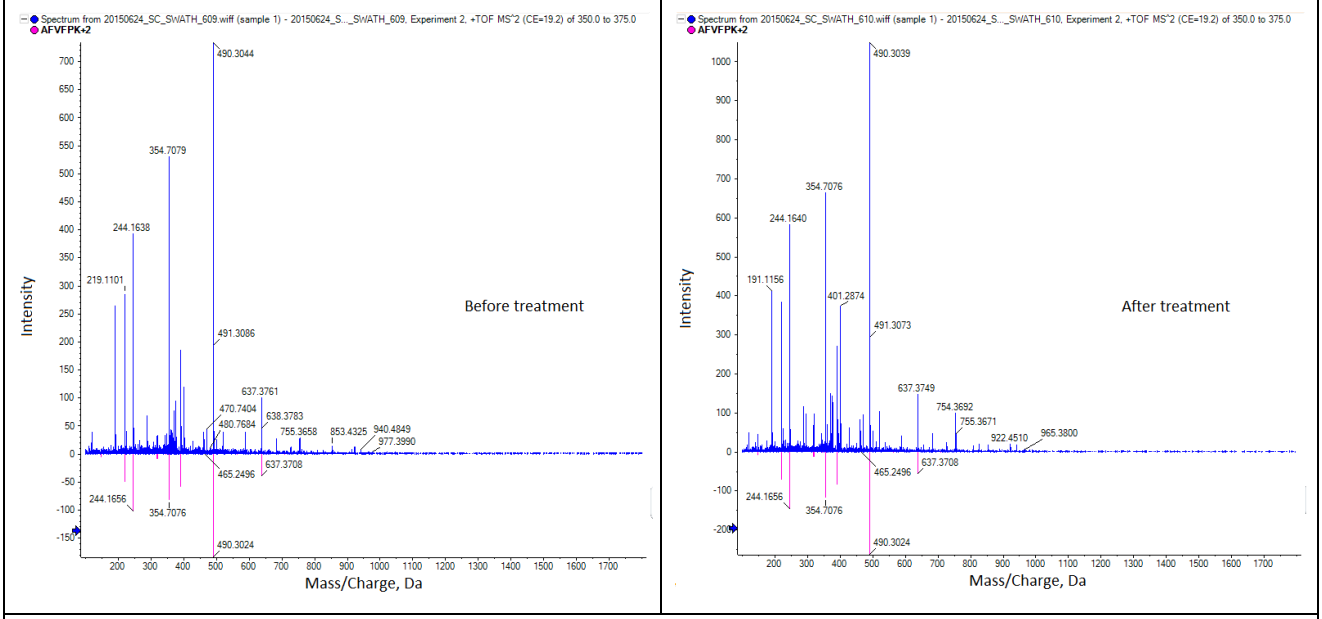


Figure 7.6. Quantitation of serotransferrin. Panel A shows an example of XICs of SAGWNIPMGR extracted from SWATH using five fragment ions specific to this particular selected peptide of serotransferrin in sheep plasma before and after treatment with endotoxin. Panel B shows a comparison of SWATH signal (blue) for one particular sample to library MS/MS signal (pink), which confirms the ID of the peptide. Serotransferrin is known to decrease during inflammation.



A



B

Figure 7.7. Quantitation of C-Reactive protein. Panel A shows an example of XICs of AFVFPK - a selected identified peptide of pentaxin (C-Reactive protein) in sheep plasma before and after treatment with endotoxin. Panel B shows the corresponding mass spectrums of this doubly charged AFVFPK peptide in Panel A which confirms the ID of the peptide. C-Reactive protein is known to increase during inflammation.

The full details of proteins that were quantitated by SWATH-MS analysis alongside their fold-chage values are listed in Table A7.5 in APPENDIX 7.3. For the purposes of this proof of concept study, within the limitations of only 75 minutes of induced endotoxaemia in sheep, any protein that had a fold change above or below 1.0 based on *t*-test results was considered to have the tendency of protein alteration criteria during endotoxaemia as compared to healthy status, i.e. before the sheep were treated with endotoxin. It is therefore reasonable to suggest that known and potentially inflammation-related proteins showed the most alteration during endotoxaemia and had peak area ratios far greater or far less than 1.0.

Some 40 well-recognised proteins involved in systemic inflammation and their corresponding gene

names in parentheses that were quantitated include albumin (ALB), serotransferrin (TF), amine oxidase (AOC3), C-X-C motif chemokine (PF4), alpha-2-glycoprotein 1, zinc-binding (AZGP1), angiotensinogen (AGT), apolipoprotein M (APOM), apolipoprotein B (APOB), fibrinogen alpha (FGA), fibrinogen beta chain (FGB), alpha-1-antiproteinase precursor (SERPINA1), ceruloplasmin (CP), L-lactate dehydrogenase B chain (LDHB), alpha-1-microglobulin/bikunin precursor (AMBP), apolipoprotein A-1 (APOA1), Alpha-1-acid glycoprotein (ORM1), antithrombin-III (SERPINC1), haptoglobin (LOC101102413), alpha-1-macroglobulin (A2M), Serum amyloid A4 protein (LOC101120613), apolipoprotein A4 (APOA4), apolipoprotein F (APOF), Pentaxin (CRP), apolipoprotein H (APOH), apolipoprotein C3 (APOC3), Plasminogen (PLG), apolipoprotein A2 (APOA2), lipopolysaccharide binding protein (LBP), Alpha-1B-glycoprotein (A1BG), mannan binding lectin serine peptidase 2 (MASP2), apolipoprotein D (APOD), alpha 2-HS glycoprotein (AHSG), retinol-binding protein 4 (RBP4), beta-2-microglobulin (B2M), monocyte differentiation antigen (CD14), inter-alpha-trypsin inhibitor heavy chain 1 (ITIH1), inter-alpha-trypsin inhibitor heavy chain 2 (ITIH2), inter-alpha-trypsin inhibitor heavy chain 3 (ITIH3), inter-alpha-trypsin inhibitor heavy chain 4 (ITIH4) and apolipoprotein E (APOE).

In addition to proteins that are already well-recognised to be involved in acute systemic inflammation, there were 42 other proteins that showed considerable alteration during early-phase acute endotoxaemia either by increasing or decreasing – these proteins could potentially be recognised as markers of inflammation in future. These candidate proteins extracted from Table A7.5 in APPENDIX 7.3, with their corresponding gene names in parentheses followed by their fold change values that occurred during early-phase acute endotoxaemia were: tRNA methyltransferase 11 homolog (TRMT11) 16.2, family with sequence similarity 105 member A (FAM105A) 15.9, sentrin-specific protease 8-like (LOC101116576) 6.2, lysine methyltransferase 2A (KMT2A) 5.1, family with sequence similarity 13 member B (FAM13B) 4.4, T cell receptor delta variable 2 (TRDV2) 3.1, elongation factor 1-alpha 1-like (EEF1A1)2.8, Titin (TTN) 2.7, Histone cluster 2 H2A family member b (HIST2H2AB) 2.6, ceramide synthase 4 (CERS4) 2.6, platelet basic protein (PPBP) 2.5, angiopoietin like 6 (ANGPTL6) 2.5, serpin family A member 5 (SERPINA5) 2.5, leukocyte cell derived chemotaxin 2 (LECT2) 2.4, proline rich 14(PRR14) 2.3, glucose-6-phosphate isomerase (GPI) 2.3, antigen WC1.1 (WC11) 2.2, casein alpha s1 (CSN1S1) 2.2, insulin like growth factor 2 (IGF2) 2.2, leucine rich repeat LGI family member 2 (LGI2) 2.2 choline O-acetyltransferase (CHAT) 2.2, pentraxin 3 (PTX3) 2.1, proteasome subunit alpha 2 (PSMA2) 2.1, osteoglycin (OGN) 2.1, enolase 1 ENO1 2.0, cystatin C (CST3)1.9, short palate, lung and nasal epithelium carcinoma-associated protein 2B-like (LOC101107619) 1.9, zinc finger protein 512B (ZNF512B) 0.6, inter-alpha-trypsin inhibitor heavy chain 3 (ITIH3) 0.6 myeloblastin (PRTN3) 0.6,

alpha globin chain (HBA1) 0.6, junction plakoglobin (JUP) 0.5, haemoglobin, beta (HBBA) 0.5, complement factor D (CFD) 0.5, insulin-like growth factor binding protein acid labile subunit (IGFALS) 0.5, myosin light chain 1 (MYL1) 0.5, lysozyme C, milk isozyme (LOC101102714) 0.5, fibulin 5 (FBLN5) 0.5, biotinidase (BTD) 0.4, Histone H2B type 1 (LOC101108086) 0.4, quiescin sulfhydryl oxidase 1 (QSOX1) 0.4, Beta-C globin (LOC100134870) 0.4 and alanyl aminopeptidase, membrane (ANPEP) 0.3.

7.5 Discussion and conclusions

This study used a sensitive method previously reported in human studies to analyse and identify plasma proteins that alter during acute endotoxaemia (i.e. biomarkers of acute inflammation). The ultimate goal was to develop a pipeline, using a PSL, as a proof of concept that is able to support quantitation a long list of proteins in sheep plasma that could be fished in future studies for acute inflammation biomarker candidates.

This chapter also discusses problems with the analytical samples that were used, for example the possible presence of intravenous anaesthetic drug residues in the samples that could have interfered with protein extraction from plasma samples. It outlines the potential future applications of this research effort as a considerable addition to the quantitative profiling of proteins involved in early-phase acute inflammation and their alterations.

The plasma samples used represented the early phase of induced endotoxaemia in experimental sheep of archived samples from a previously published ovine model of intensive care on acute pulmonary injury[14]. A subset of the nascent PSL assembled in Chapter 6 was used alongside the SWATH-MS approach to interrogate the samples to identify proteins and their alterations during endotoxaemia. As label-free quantitation is not a trivial process, inherent analytical challenges were identified and optimisation experiments were performed to ensure that any alterations identified awere not due to technical variation. Validation of the adopted analytical procedure was performed prior to the actual application of the method to evaluate the plasma samples. Replicate LPS treated samples were analysed by injecting every sample three times and they produced equivalent results and did not themselves show up significant changes (data not shown).

Changes in plasma proteins are useful to study in animal models of disease involving inflammation[17,390,391] and can assist in defining the course of disease processes and the response to treatment[17,378]. It should be noted that in clinical practice, normally relative changes in plasma protein concentration are commonly considered, although proteomics also permits comparison of absolute changes. Because such proteins are used to inform (or select) medical interventions, they are useful yardsticks in monitoring disease pathophysiology if their profiles are determined[17,392] and

having the ability to do so *en masse* in future if the current approach ever gets to be applied clinically would therefore be paramount. The advantage of this would be that the analysis of many proteins simultaneously helps to understand more physiological perturbations compared to when only individual or small numbers of biomarkers are monitored.

According to reports in the literature, the number of proteins recognised to be involved in systemic inflammation by being either up-or downregulated is in the lower hundreds[17,378,390,391]. The findings from the present study on sheep plasma suggest that in addition to recognised APPs that were induced by LPS, a substantial number of other proteins tended to suggest altered intensities during endotoxaemia. In this study, circulating albumin – which is known to be a negative APP was increased, albeit by only 1.1 fold. During inflammation albumin usually decreases because it is mobilised by the liver to synthesise other proteins involved in the immune response[393]. A possible hypothesis to explain the observed subtle increase here is that there was a momentary upregulation of albumin during early phase acute inflammation before perhaps subsequently decreasing, considering that acute endotoxaemia had been established in the sheep for only 75 minutes.

The major APPs in sheep are haptoglobin and serum amyloid A, while alpha 1β- glycoprotein and C-reactive protein are regarded as moderate APPs[17,390]. In the present study, serum amyloid A-4 (LOC101120613) increased 1.4 times in response to LPS, whereas haptoglobin (LOC101102413) – which still remains to be unequivocally characterised by MS-based proteomics in sheep, increased by 1.2 times and alpha 1β- glycoprotein increased by 1.2 times. It should be noted, however, that different diseases may induce different acute inflammation protein profiles[378], meaning that the current model of endotoxaemia may not be representative of all acute systemic inflammation disease states. Furthermore, the current study collected samples only 75 minutes after endotoxin administration, so this could be considered as only being representative of an early phase of inflammation.

The proteomic profiling of acute inflammation profiling using MS/MS is still an emerging field in the veterinary science[390] and further work is required to determine the biological or clinical relevance[17]. A major advantage of the current (proteomic) approach is that many of the protein changes following LPS treatment are mediated by cytokines[17,394 374,375,395,396] and these, including C-X-C motif chemokine (PF4), pro-platelet basic protein (PPBP), leukocyte cell-derived chemotaxin-2 (LECT2), and possibly interleukin-1 beta (IL1b) can also be measured, but they could have been below limits of detection or beyond the capabilities of the technology that was used, due to the considerable WDR created by abundant proteins.

This study had limitations, including the fact that the samples were collected opportunistically from a specific group of sheep initially for two desirable animal production traits: Endoparasite resistance

(Parasite) and wool production (Wool). These traits may have influenced the extent of response or the specific proteins altered in response to treatment of the sheep with endotoxin. Repeating this study with other types of sheep and disease processes would therefore be prudent.

Another limitation was that protease inhibitors were not added to samples to minimise endogenous protease activity. Furthermore, the sheep were anaesthetised using total intravenous anaesthesia whereby chemical residues in blood may have interfered with protein extraction as compared to if inhalation anaesthesia were used. It can also be argued that there could have been technical problems in the acquisition of the data because of the disqualification of nearly half the dataset due to lower signals. While these problematic changes were sufficiently large to be noted, there may have been more subtle differences that were in other samples that may have been significant but went undetected. It is also important to note that the plasma samples that were utilised in this study were for a single time-point only, representing protein alterations in samples drawn after 75 min from the commencement infusion of LPS after instrumentation (tracheostomy for ventilation and invansive cannulation of blood vessels) of the sheep and during anaesthesia. This may have provided limited information about protein changes before instrumentation or after this period. For example in other studies, serum amyloid A and haptoglobin have been monitored over longer periods, up to days after the initial challenge[397,398]. It would therefore have been interesting to have more blood collections over the 12 – 24 hr period following LPS administration.

In this study, a number of proteins were identified as having altered concentrations after LPS treatment. The data were presented in a comprehensible manner to satisfy the objectives of this proof of concept study that utilised archived plasma samples and this was within the scope of the thesis as explained in the methods above. Each protein had up to 5 different peptides that were measured as proxies for the protein. Unfortunately, because of the protein-level output of the results, the use of SWATH™ Acquisition MicroApp in conjunction with MarkerView™ Software pipeline from SCIEX employed in this study did not permit the user to evaluate the average change in concentration or standard deviations of individual peptides. Nevertheless, the data should allow other investigators or reviewers to objectively assess its value in its current form and also after the data are made available in a public archive, especially having reported protein alterations as fold change values, representing before and after treatment of the sheep with LPS from *E. coli*.

Much as the functionality of the PSL has been demonstrated, it is important to note that the results of the identified protein alterations during acute endotoxaemia have not been validated and should be best considered as preliminary or proof of concept only. Given that the study was carried out using plasma, the possible validation strategies are by antibody enzymatic analysis, western blotting and possibly HR-MRM as discussed earlier in Chapter 2. There are currently very limited validated

antibody and enzymatic assays for sheep and therefore, validating these results would be a life-long study and is out of scope for this thesis. However, in a limited number of cases such as serum amyloid, some cytokines and haptoglobin, quantitative assays are available and routinely used. In future, these could be used and compare the results with ELISA or colorimetric assays, in the case of haptoglobin. Proteins that do change due to LPS have previously been documented elsewhere[17], so this could be used to compare with the proteins detected in this thesis.

In conclusion, this is the first time a SWATH-MS analysis pipeline has been used to investigate plasma samples of a large animal model to quantitate a large number of proteins simultaneously. Up to 243 proteins were quantitated, with 40 proteins identified as APPs and a further 42 proteins potentially related to early-phase systemic inflammation. These findings have major implications in veterinary pathology and animal welfare[17], particularly when considering that this technology could be further developed to be more reliable and useful to rival the current gold standard for protein identification – immunoassays, plus that it can also analyse proteins for which there is not kit available.

CHAPTER 8

8.0 General discussion, conclusions and future directions

8.1 General discussion

Key points:

- The development of a peptide spectral library from ovine blood is considered an excellent area of research. Given that the technology is now relatively well developed through the analysis of other species, the project was largely an analytical challenge. The scale of the work required to perform a comprehensive analysis of the proteins present in the ovine circulating acellular proteome is such that it is compatible with an original contribution, and the information gathered adds to the sum of knowledge known about human plasma/serum proteome analysis.
- The primary output of this method development thesis was the compilation of a novel encyclopaedic PSL, comprising high-quality spectra of 398 proteins derived from plasma and serum of sick and healthy sheep, using shotgun proteomics. Headway was made in optimising a bioinformatics approach to validate the developed proteomics package. This PSL can be used for targeted protein data extraction, using strategies such as SWATH-MS assays for the identification of a wide range of proteins[259,399] and proteogenomics applications. The work performed in this project was unique in that it was performed on sheep blood and its real value is in having the highest quality data that subsequent researchers may find useful.
- The work in this thesis aimed to identify proteins involved in early-phase acute inflammation because doing so helps in defining the predictors of mammalian response to illness[17]. This in turn translates into understanding resilience to disease, and in monitoring progression of disease. The project explored a timely molecular approach to understanding injury in sheep by focussing on proteomic characterisation of potential circulating acute phase markers. This was achieved by applying a subset of the nascent PSL on samples from healthy and ill sheep to detect candidate markers of inflammation after 75 minutes of exposure to endotoxin. To this end, a considerable number of candidate markers of inflammation were identified simultaneously for the first time in the liquid fraction sheep blood, using SWATH quantitation.
- Several difficulties and costs associated with traditional protein assay methods, such as ELISA, make MS-based proteomic assays an attractive alternative. In the absence of abundant protein depletion strategies for samples from a majority of veterinary species, the problem of the WDR between low- and high-abundant proteins can be minimised by combining different proteomic strategies, such as protein fractionation and chromatography[27]. It is evident that working with

non-model organisms calls for the need to optimise bioinformatics steps in order to minimise false positive protein identification and ensure a high degree of assay sensitivity. On the basis of currently available protein detection approaches, the SWATH technique appears to be preferable for quantifying known and novel proteins on a large scale[1,3-5,17]. A high-capacity trap column and a long (e.g. 50 cm) heated analytical column or changing nanoflow column to capillary flow columns, for example, would enable the loading of higher amounts of protein to improve chromatographic resolution and enhance the detection of more proteins by the MS instrument. In addition, MS-based proteomic studies of non-model organisms require considerable time-consuming computing resources because of limited or incomplete protein sequence databases and the need for manual curation of the output data (human interface).

- This proteomics approach using the nascent PSL, which has room for further optimisation, was applied in a proof of concept study – with promising results – using ill sheep with induced endotoxaemia as the model disease state. The PSL is considered to be one of the most developed and comprehensive to be derived from the circulating acellular proteome sheep and has several potential future applications.
- Overall, four objectives were addressed in this thesis, namely (1) the development a feasible proteomic method to characterise the ovine acellular circulating proteome, (2) construction of a PSL repository for serum and plasma, (3) optimisation of a bioinformatics approach to validate the developed proteomics package and (4) application of the method to potentially detect candidate markers of inflammation in plasma.
- In this chapter, an opportunity is provided to look back over the project and highlight the successes and challenges of the methods that were described in the thesis and what their outcomes mean on a broader scale. Further, the discussion relates to how the number of proteins identified from the various strategies compare with other animal studies and those of humans in the literature by describing what was done to ID proteins and explains why these steps enhanced the ID of proteins. An explanation is also given of how the software programs for identifying the proteins were selected. The discussion then moves to the practical application of the methods for identifying proteins in sheep plasma and an investigation into how the relative expression of the proteins could be used to detect alterations in pathophysiology. The optimal workflow for the analysis of plasma proteins in sheep blood is discussed as comprising of depletion, effective fractionation and analysis on high-resolution MS instrument. An account on what would have been done differently in hindsight (lessons learnt) is provided, including experiments that should be performed to discover more about the inflammatory response in sheep given the limited observations from SWATH experiments in Chapter 7. Later, it is

proposed that proteomics could be used to investigate species differences in response to disease and why some individuals are more or less susceptible to disease insult.

Despite the progress that has been made in the field of human plasma and serum proteomics, a review of the literature in Chapter 2 of this thesis highlighted the existence of many knowledge gaps in our understanding of the circulating acellular proteome of veterinary species. Considerable work still remains to be done in defining various proteomes of veterinary species, as most are non-model organisms[7,38]. This is important because MS-based proteomics approaches are heavily reliant on the extent of sequencing and annotation of the genome of the species under study[67,400,401]. Some well-established aspects of circulating acellular proteomics in veterinary species, with references to human studies, were introduced and discussed in the literature review. The arguments presented there were focussed entirely on the developments relevant to the scope of this thesis, such as the well-established challenges posed by the WDR encountered when analysing proteins derived from shotgun proteomics pipelines and suggested approaches for dealing with it. This background was sufficient to enable the formulation of an appropriate choice and development of the ultimate methodology for the thesis project.

The methodology developed for this thesis was founded on generic methods used in plasma and serum proteomic studies that were described in Chapter 3[16]. It relied upon commonly used processes of sample preparation, such as acetone precipitation and desalting, for the clean-up of proteins prior to analysis. Because proteomic sample preparation is tedious and error-prone but crucial for the generation of reliable results[402], it was considered necessary to benchmark to and build this initial aspect of the project upon existing works in order to obtain comparable and reproducible results[403] prior to the innovations applied in the subsequent chapters. Hence, this thesis focussed upon sample preparation strategies to enhance protein identification in plasma or serum for construction of the PSL and its subsequent application in the proof of concept in Chapter 7.

Desalting of peptides prior to MS analysis was considered a priority step during sample preparation because of the large number of samples (over 900) processed for this project. The effect of desalting of peptides prior to analysis and its reproducibility as a critical element of the entire proteomic method was evaluated. Early observations identified a substantial workflow bottleneck regarding the manual desalting of many samples reproducibly during method optimisation stage. Chapter 3 workflow therefore represented a fundamental keystone methodology for the generation of the draft reference ovine serum-based baseline acellular circulating proteome described in Chapter 4[16].

The observations made in Chapter 4 highlight the feasibility of establishing a reference serum proteome for sheep, thereby addressing the first objective of the thesis[16]. Previously, there were limited data and no studies exploring an optimum MS-proteomics approach suitable for veterinary

species. Once the feasibility of establishing a baseline PSL was realised, this ambition was broadened to further explore the circulating acellular proteome and additional enrichment and enhancement of its contents. This was accomplished by an array of strategic protein sample preparation approaches described in Chapter 5.

The characterisation of the baseline circulating acellular proteome was accomplished by adaptation of two universal sample preparation strategies for proteome analysis[233], namely, by separation and tryptic digestion of in-gel trapped[404,405] and in-solution proteins[35,123]. The peptides derived from the protein digestions were then subjected to separation by chromatography before being MS/MS analysed. The MS/MS data were then processed using multiple vendor-specified algorithms bundled in a number of bioinformatics software programs to aid in identifying proteins from the tryptic peptides. Consequently, the baseline acellular circulating proteome was characterised following the development and optimisation of proteomic methods using serum samples obtained from healthy laboratory sheep through discovery proteomics experiments that included in-gel and in-solution digestion workflows. Taking into consideration the conventions for reporting detected proteins[265,266], the platform used in this study showed the capability of simultaneously detecting hundreds of key serum proteins using sheep as a model.

Pathway analysis is used to reveal the association between identified proteins and their known pathway interactions[216]. The Ingenuity Pathway Knowledge Base, for example, displays interactions between genes, proteins and other biological molecules to reveal biological pathways and to generate global canonical pathway protein interactions[216,217]. By utilising the PANTHER pathway analysis tool[220] in this thesis, the serum-derived prototype of the ovine circulating acellular proteome revealed the association of 349 genes with 127 protein pathway hits[16]. The viability of this workflow was accomplished using generic methods, which cemented the first step in the construction of the PSL. These results were considered as a prototype for the normal ovine circulating acellular proteome based on serum that could then be used as a reference benchmark tool for testing protein alterations during physiological perturbations, such as during illness, in any ovine model. A similar approach for determining the baseline proteome could be applied to other veterinary species. The advantage of understanding better how animals such as sheep respond to inflammation with long-term benefits has long been recognised[17].

In Chapter 5, the assorted sample preparation strategies facilitated the capture of as much incremental peptide information as possible under different conditions, to enable the subsequent construction of a novel encyclopaedic PSL in Chapter 6 – the primary thrust of this thesis by conglomerating data from seven experiments to enrich the baseline proteome established in Chapter 4. Considering biotechnical work on human plasma as the benchmark[406], and by drawing parallels

to it, this work has demonstrated that it is possible to construct an in-depth PSL for characterising the circulating acellular proteome of a non-model species using serum and plasma samples from, for example, healthy and sick sheep. In order to fulfil the requirements of the second objective of the thesis, the prime aspects of its construction are highlighted in the discussion that follows.

Acetone precipitation of proteins is a well-established sample preparation procedure[24] that is usually performed to rid samples of substances in the supernatant that interfere with protein analysis[25,26]. However, there are few studies that have investigated the supernatant fraction that is usually routinely discarded, even though some proteins are lost in the process[407]. For instance, one particular study observed that a substantial number of proteins were present in the supernatant of acetone precipitated rat brain homogenate that were not present in the precipitate[407]. These findings are consistent with those of the present study, where considerable numbers of proteins were present in the supernatant after acetone precipitation of sheep serum and plasma, some of which were unique to this fraction. For the purposes of building a PSL, it is suggested that both fractions of acetone precipitation samples be analysed in future. For example, there is a possibility of proteins with specific qualities i.e. those that fail to precipitate by known mechanisms[23,329], that aggregate in the supernatant and that could be important for biomarker studies. It is therefore reasonable to assert that some previous studies have been biased by excluding families of proteins with particular chemistry that did not favour precipitation with acetone. This is a very important finding of the present work.

The rationale for partial organic precipitation using ACN during sample preparation was to reduce the footprint of abundant proteins, such as albumin, globulins and lipoproteins, thereby enriching the less abundant protein species in the supernatant[332,337,408,409]. This approach was attractive for this thesis because it has been trialled in samples from several species, including humans[337], with promising results of improved protein identification, but data for sheep were lacking. The findings of the experiments from this thesis support the use of ACN precipitation for enhanced protein identification capacity. A notable observation in the present study which differed from an earlier report[337] is that more protein detections were made in buffer-diluted ACN treated samples with a higher pH than with a lower pH. This could be because a higher pH favours the digestion of protein by trypsin into tryptic peptides[410]. Overall, ACN precipitation workflow improved protein detection considerably compared to untreated samples derived from sheep plasma and serum and should be considered for inclusion in pipelines that seek to deepen the proteome coverage of their DDA experiments.

Combinatorial peptide ligand library plasma treatment has been known to enlarge the visibility of medium to low abundant proteins and has potential in biomarker research[411]. This technology has

made headway in preparing samples derived from animals; for example, it was used to deplete plasma samples of cattle during biomarker studies for Johne's disease[412] and investigation of type 2 diabetes in mice[413]. It has been utilised in a number of human study pipelines, such as in identification of candidate biomarkers for hepatitis B–associated liver cirrhosis in humans[414] and other conditions[415-417], all with improved numbers of proteins identified. This thesis thus utilised ProteoMiner™ technology to acquire protein data for sheep serum and plasma for the first time and was by far the most successful experiment.

Off-gel fractionation was trialled in the present study, with the goal of gaining additional information on the analytes – compared to other fraction methods – because it has been reported that it performs better than SDS-PAGE when analysing human plasma samples[418]. Other than a recent study on sheep whey, there are currently no other studies that have used this fractionation approach on sheep serum or plasma. The advantage of off-gel fractionation is that crude protein samples can be analysed, as well as peptides. In this thesis, only serum was used but future studies could also acquire data from plasma for inclusion in the PSL to cover proteins that were not in serum.

In principle, PSLs are to be derived from a wide range of observed high-quality MS/MS spectra from DDA experiments that are as inclusive as possible for future investigations when mining the library[5]. For this reason, it was imperative to include peptide data from the serum of ill sheep (both naturally sick and experimentally induced endotoxaemia) in order to establish a comprehensive library that would capture proteins that might only be upregulated during systemic illness. The endgame here was to have a library that focussed upon the inflammation/systemic states angle – this was more the emphasis than exploring more general biology.

The PSL contains peptide sequences of five pro-inflammatory cytokines that were synthesised and added. These synthetic cytokines acted both as quality control internal standards to ensure that data were processed correctly for protein identification and will in future hopefully aid in detecting endogenous cytokines that are usually in low abundance. In addition, the library is considerably more extensive than any publicly available animal PSL, such as those present in the PRIDE Archive[19], and it is potentially versatile due to its homology with bovine protein sequences. The *in silico* predicted and synthesised peptides for the five cytokines were in fact selected on the basis of their homology between sheep and cattle. The PSL could therefore be adapted as a building block to complement existing and future efforts for construction of a plasma- and serum-derived bovine PSL. Although a synthetic library can result in additional proteins hits and be very useful, it is significantly more expensive and more time-consuming to construct because it requires *in silico* prediction of the best and high-scoring unique peptide candidates. An impediment associated with

combining libraries derived from samples of different species, although beyond the scope of this thesis, is the uncertainty concerning how bioinformatic platforms, such as PeakView, Skyline or similar, interpret homologous peptides. This would inevitably call for the non-trivial task of manually going through large peptide lists in order to retain only high-scoring peptide and protein homologues while minimising redundant sequences in the library. Even with a single species PSL such as the one constructed for sheep in this thesis, it is still necessary to mine it using multiple search platforms in order to establish its protein and peptide content. The spectra of peptides from the five synthesised cytokines were visible in the graphical interface of PeptideShaker, confirming their presense in the PSL. Given the strigency in validating protein IDs, it is for this reason that a bioinformatics strategy using this software platform was utilised in this thesis in order to retain only high-quality data for the PSL.

This being a method development project, several optimisation experiments were necessary to realise the goal of being able to identify proteins isolated from the circulating acellular proteome of sheep. This was achieved by the trialling of various alternative approaches for proteome analysis and bioinformatics tools, as detailed in Chapter 6. The choice of the software platforms used in this thesis was determined by (1) the requirement of the primary instrument bundle – in the case of ProteinPilot™ and TripleTOF 5600+ instrument; (2) availability of a high performance computing host for the Mascot Server at QUT; (3) cost and quality of the software (PeptideShaker is free, open source and has the ability to interpret data from different search engines[245]); and 4) human resource expertise and supervisory support.

One software package with capabilities for validating protein identifications on a single platform that was trialled is Scaffold Software[187] *(data not included)*. It enabled the comparison of different sample datasets and search engines[187]. Within Scaffold, X! Tandem[162,163] protein identifications were compared to Mascot[200] identifications. In addition, Scaffold intrinsically used peptide prophet and protein prophet algorithms to identify proteins[187,419]. Scaffold takes into consideration the universal determination of FDR when using decoy databases for protein sequence searches[420]. Protein inference in Scaffold depends on the basis of spectral counting prior to the assembly of the identified peptides[421]. This same approach has previously been used in protein identification and validation studies of uterine luminal fluid during early pregnancy in sheep[370]. Later in the project, the size of the library exceeded the capacity of the available computing resources at QUT to fully utilise this potentially useful bioinformatics tool; nevertheless, some notable observations were made. Scaffold identified 377 proteins in the PSL compared to 564 protein IDs made by ProteinPilot, 830 IDs made by Mascot and 398 IDs made by PeptideShaker, as reported in Chapter 6. In terms of the number of protein IDs, the performance of Scaffold was therefore comparable to

that of Mascot but since Mascot is widely considered the quasi-industry standard and the most widely used protein sequence search engine[284], it would be reasonable to benchmark Scaffold searches to those of Mascot.

The findings from the baseline proteome study (Chapter 4)[16], represented the identification of medium to high abundance protein IDs in serum within the capabilities and limitations of the available software platforms at the time. It would have been more desirable to derive quantitative data from the analysed samples of that study. The use of Scaffold would have determined the differences in protein identifications and relative quantities between normal and sick sheep serum samples from this dataset, for example. Some research efforts have used a similar approach to evaluate sheep serum proteome, such as the one that investigated scrapie[264] and another that investigated respiratory disease[111]. The study on scrapie utilised SELDI-TOF-MS and LC-MS/MS and was able to detect elevated levels of serum amyloid in infected sheep compared to normal sheep. The earlier study was able to distinguish the scrapie group from the healthy control sheep[264]. In the latter study, two-dimensional electrophoresis (2-DE) map analysis followed by SELDI-TOF was used to compare healthy sheep serum profiles with those of sheep with respiratory disease, which detected alterations in at least six known acute phase proteins[111]. In comparison, the approach of the present study has the potential to analyse several proteins simultaneously. The currently available reports of proteomics studies on the liquid fraction of blood with this same approach are few in number, as most reports are focussed on other body fluids of ruminants[370,112,293,339,379,422-424]. The present study is therefore a valuable addition to literature on the plasma and serum of sheep.

One advantage of using Scaffold is that it is possible to browse or view protein alterations that are comparatively up- or downregulated or those that do not show any changes between normal and diseased serum states, for example. It is also possible to determine fold changes using the statistical analytical tool, which can be useful in determining candidate biomarkers. The level of significant protein alterations can also be determined using this platform. The stated features that Scaffold has are currently lacking in the ProteinPilot™, Mascot and PeptideShaker software platforms that were used in this thesis. The utility of a single bioinformatic analytical platform to validate proteins is an area for future research endeavours.

The concurrent use of ProteinPilot™, Mascot and PeptideShaker to interrogate the same dataset of PSL samples facilitated an optimum evaluation of the number of proteins in the PSL for this particular project. The three tools collectively identified over 1000 proteins in the PSL. Whenever computing, financial and time resources permit, it makes reasonable sense to utilise various software platforms to analyse a given PSL, because each tool has additional peculiar ways (algorithms) of identifying proteins that may not be shared by others. This is because each platform

employs a different technique for spectrum identification[425]. Considering all efforts to improve protein identifications from samples, it is recommended to combine results from multiple search strategies[425]. It is therefore quite possible that if other software tools are used to analyse the present PSL derived from the ovine circulating acellular proteome, the collective number of identified proteins could even be higher. The difference in the ability of the different software packages to ID proteins is a significant finding of this thesis. The proteins missed or detected by specific software platforms is therefore dependent on the algorithms used by the different packages.

This nascent PSL was applied in a proof of concept capstone SWATH-MS study using archived plasma samples from sheep with induced endotoxaemia in Chapter 7 with promising results. The key outcomes of the capstone study was the quantitation of 243 proteins in plasma of endotoxaemic sheep which includes 40 established APPs and at least additional 42 proteins that appear to correlate with early-phase acute systemic inflammation. This illustrated the relevance and practical applications of the PSL relating to detection of plasma proteins, thereby fulfilling the fourth objective of this thesis. The other downstream relevance of Chapter 7 is the possibility of its application to observe any differences between the normal and disease states in terms of proteins found. It can however be argued that the experiment performed in Chapter 7 was limited in part because it used archived samples collected several years ago from a completely different study on sheep that was not designed for MS proteomics for the purpose of evaluating the PSL only. In order to discover more about the inflammatory response in sheep, future experiments should be designed with MS proteomics objective in mind, and with appropriate controls, use of identical sheep, appropriate sample collection, current sample preparation methods, well-researched analytical strategy, together with a comprehensive and consistent data analysis process. However this is a methods thesis, with the method development having greater prominence than clinical outcomes. Elaborating on clinical outcomes is an appropriate way to explore the potential applications of the developed method, which is in line with the current trend of protein biomarker studies.

Sheep are an excellent model for translational research, particularly since the sheep genome is yet to be completely sequenced[230]. Furthermore, the relatively low cost per head and ease of handling make sheep an attractive potential model for translational studies. An important next step in capturing this potential is the establishment of a proteogenomics study model to accurately define a sheep's response to stimuli, such as acute injury, before sheep can be used successfully as a model for other species, including humans. Aside from the obvious wider benefits to society, there is a clear link between biomedical potential and opportunity for industry, because supply of sheep to translational study investigators would represent an additional marketing avenue for the producer.

8.1.1 Lessons learnt

A successful proteomic analysis requires the adoption of a well-researched analytical strategy at the beginning of the project, along with a comprehensive and consistent data analysis process. Mass spectrometry run time is often limited by demand for the required technology, as well by as the cost of accessing it. It is vital that a solid analytical framework is also supported by adequate access to trained personnel with appropriate expertise and interest in veterinary proteomics. Successful proteomics workflows are therefore dependent upon a range of important factors, and it is vital that project teams collectively consider these factors and plan accordingly.

As canvassed in the review of literature in this thesis, the optimum strategy for in depth proteomic analysis of plasma proteins requires that the following conditions are met: a) depletion of abundant plasma proteins, b) efficient fractionation of the proteins present in the sample into many different pools, and c) the analysis of these samples on a modern high-resolution mass spectrometer. The utilisation of sheep blood in this project was met with challenges, particularly because specific abundant protein depletion kits that have been developed for analysing human and mouse blood could not be used for depletion of these proteins from sheep plasma. There were, however, other options that were not explored in this project due to unprecedented financial and time limitations, despite them being used in other animals[68,69]. For example, the use of the ProteoMiner™ kit provided an effective abundant protein depletion strategy, although it could only be used for a relatively small part of the overall project workflow due to cost and time considerations. Once depletion is accomplished, the sample is then split into multiple fractions for individual analysis. This was achieved, in part, by the use of 1D SDS-PAGE, whereby the gel lanes were divided into several fractions and analysed individually to achieve adequate fractionation. An attempt was also made using the Agilent OFFGEL Fractionator, however, this experiment was unsuccessful and this platform was not pursued further. Another available option that was not explored in this thesis was the use of offline SCX fractionation. By comparing and contrasting ACN precipitation results with the gel output from the ProteoMiner™ bead system (Figure 5.7), it is evident that this latter system led to the greatest number of protein identifications. Perhaps this system process might have been explored in greater detail by varying the sample loading quantity. It is possible that a greater quantity of sample loaded might enhance the overall workflow by permitting the identification of more proteins.

During the early stages of the project, a predicted approach to peptide identification, incorporating amino-acid substitution with reduced identification stringency, coupled with the utility of multiple databases and search engines was used to maximise the likelihood of protein identifications. Such an explorative strategy casts a wide net, and from the outset seeks to achieve sensitivity at a cost to

specificity. It was therefore essential that the downstream workflow sought to validate preliminary results through the use of synthesised peptide standards. After this, a more stringent data processing strategy that included detecting spiked-in synthetic peptides was employed to minimise the chances of false identifications. The strategy incorporated interrogation of a sheep-only UniProtKB database. The rationale of using multiple search engines to search a single database underpinned an important hypothesis for this work; that the search engine itself represents a significant variable in the overall proteomics workflow. Employing a range of different search engines allowed for informed comparison, and highlighted strengths and weaknesses of each different search engine. As the goal of this project was to develop a PSL for use by future researchers, there was the obligation to include only those peptides that had been identified under the most rigorous conditions, and to highlight which combination of techniques, platforms and search engines yielded the highest quality output. For this reason, only identifications from PeptideShaker compiled from all experiments were retained for assembling the PSL.

Spectra from synthetic peptides of cytokines were developed to be used for SISCAPA-MRM-MS (https://espace.library.uq.edu.au/view/UQ:347312 – *data not shown*). This approach was later abandoned due to time constraints and scope of the PhD project and therefore, spectra were simply added to the PSL to enhance its analytical depth. SWATH technology has advanced the proteomics frontier, because it facilitates the rapid quantitation of a large number of proteins. However, it is only useful for quantitation of more abundant proteins, largely due to wide dynamic range constraints encountered by the technology. Currently, it is known that unenriched protein samples pose particular difficulties when using the SWATH platform, notably where quantitation is sought for proteins that exist below 10 ng/ml. The five endogenous cytokines that were emulated by means of synthetic peptides exist at concentrations well below this, which explains why none of them were identified by DDA without enrichment. The background to the synthesis of these cytokines was led by the informed understanding that their concentrations are expected to elevate following induced acute endotoxaemia. It was intended to include these cytokines in the overall analysis; however this aspiration did not take into consideration the major dynamic range limitations of the current SWATH technology, and therefore such a goal could not be achieved in the scope of this thesis. Nevertheless, these synthetic cytokine peptides were useful for in-house purposes only in determining if SWATH data processing parameters were correct. It is also important to note that endogenous cytokines (as well as many other proteins present in the PSL) were not detected using SWATH which was expected, given their low abundance. Theoretically, mass spectrometry can be used to detect plasma protein alterations during illness, however it is true that the most important and specific changes generally occur in proteins of low abundance. This therefore warrants their

enrichment to enable detection.

8.2 Conclusions

Proteomic analysis of the circulating acellular proteome, and modern mass spectrometry data acquisition approaches such as SWATH-MS, comprise a promising field that has a wide variety of applications in veterinary science. This project has delivered the foundation of the much-needed experiments benchmarked on studies in humans that can be used for detection of pathology in domestic animals. A vital attribute of the contribution of this work is the optimisation of serum and plasma sample preparation, which is now available for widespread use.

A platform capable of identifying several hundred sheep serum proteins using nanoLC-nanoESI-MS/MS using serum from healthy sheep has now been developed. This approach can now be used with protein quantitative data to potentially establish baseline proteomes of healthy domestic animals to be applied in comparison with samples from sick individuals and most importantly, it is a useful tool for quantifying proteins in closely related mammals (e.g. cattle and goats).

This thesis is probably the first to have developed a novel encyclopaedic PSL for the ovine circulating acellular proteome with the capability of identifying a large number of proteins. If implemented, this proteogenomic tool could be of practical significance for veterinary science in understanding physiological perturbations in the liquid fraction of blood, and for the scientific community seeking to use sheep or related species as a model for studying human disease. The use of multiple protein sequence database search engines enabled the identification and validation of proteins in the PSL constructed from the liquid fraction of sheep blood.

This work pioneered the use of a SWATH-MS pipeline on a large scale to interrogate archived plasma samples of a sheep model of intensive care. It made it possible to distinguish between samples from untreated and endotoxin treated sheep in order to quantitate several proteins, including their alterations in the circulating acellular proteome. It also confirmed that the samples of the sheep model were non-identical. The several proteins quantitated during induced endotoxaemia suggest the potential practical outcome of this approach, by showing proteins that were potentially upregulated or downregulated during early-phase acute inflammation.

In summary, this project has developed a novel encyclopaedic PSL using the liquid fraction of sheep blood. It can be asserted that this work has clearly pioneered the basis for the exploration of the circulating acellular proteome of sheep – a non-model organism – although there are still many proteins that remain to be characterised.

8.3 Future directions

The research questions that formed the premise of this thesis arose from previous work (*see* Introduction to the thesis). Similarly, the findings of the experiments conducted and described in the thesis have generated new research ideas.

In continuing this line of research, it will be necessary to take note of the existence of certain experimental factors when interpreting the results. For example, the LC-MS conditions used for the thesis were standard, based on the protocols developed for the instrument that largely ran peptide samples from humans and model organisms. At this stage, it is not possible to ascertain if peptide samples from sheep plasma and serum would require different instrument settings. Overall, the thesis focussed on optimisation of sample preparation and set a foundation for the development of a bioinformatics pipeline for data analysis on the assumption that all peptides can be analysed similarly, irrespective of the species of origin. In future, a high-capacity trap column and a longer heated analytical column (50 cm, for example) or a combination of a capillary flow rates and capillary columns would enable the loading of higher amounts of protein for deeper peptide analysis. This could be utilised together with variable SWATH windows to collect cleaner signals with reduced interference and a more sensitive MS platform, such as the TripleTOF 6600 (SCIEX) instrument, that has a better dynamic range. Further optimisation of the digestion procedure (for example introducing FASP digestion instead of in-solution digestion) could provide more reproducible data and improved protein coverage. On the bioinformatics front, an application of *de novo* sequencing approaches could potentially enable annotating more spectra and to permit more proteins to be quantified.

The experiments that were associated with the majority of protein identifications during DDA runs were ACN precipitation and ProteoMiner™ workflows. There is a need to have several replicates from the samples of these workflows in order to potentially increase proteome coverage for inclusion in the PSL. The off-gel fractionation workflow needs to be further optimised and applied for sheep plasma samples as well.

The PSL of the ovine circulating acellular proteome was built using a workflow based on the commercial tools developed by SCIEX. This workflow had limited avenues for user interface to permit the editing of critical factors such as FDR determination and automatic filtering of cRAP and other unwanted protein entries, except in the SWATH MicroApp user interface. In order to further enhance and make the PSL more versatile, it would be reasonable to process the data using an alternative workflow like that of the established TPP and OpenSWATH workflow[5].

There is need to perform GO analysis using the high-quality protein IDs in the PSL. I deally, this

would be followed by a discussion on how the results could include certain classes of proteins present in plasma/serum and possibly the failure to detect proteins that are already known to be present most likely due to dynamic range limitations.

Because the current PSL was built using only the liquid fraction of sheep blood, it will be prudent to broaden its capabilities by including data from cellular components of blood and other tissues. Representative tissues for genomic and proteomic analysis could be obtained from the nervous (brain), musculoskeletal (muscles and bone), integument, lymphatic (lymph nodes), respiratory (lungs), endocrine (adrenals, thyroid, pituitary) or gastro-intestinal tissues (intestines, pancreas, liver) and gonads of sheep. This will enable the drafting of an inclusive protein database. Genomic information derived from this composite of experiments alongside the proteomics data would be fundamental in complementing the efforts of the ovine genome annotation. Also, future experiments should be designed with the consideration of SWATH-MS proteomics analysis pipeline in mind in order to achieve reproducibility in the samples as evident from bovine samples (unrelated, but reliant on the methods and findings of this thesis) analysed at QUT Laboratory (Sadowski, P., *Personal Communication*). For example, automated robotic liquid handling is proposed to perform certain steps of the workflow during sample preparation that could improve reproducibility by eliminating errors that could be introduced by humans. The ultimate immediate future application for advancing the scope of the findings of these additional proteomics experiments has been identified to be in proteogenomics studies.

The foundation of the present work has been exploratory in nature by looking at innovative methods of protein detection. It will therefore be necessary for future experiments to be designed to repeat, reproduce and validate what has been observed thus far. Most importantly, there is a need to have several replicates from the samples of the workflows used in this thesis in order to potentially increase the proteome coverage of the PSL.

List of References

1. Gillet, L. C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R. & Aebersold, R. Targeted data extraction of the ms/ms spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Molecular and Cellular Proteomics* **11**, O111 016717, doi:10.1074/mcp.O111.016717 (2012).

2. Liu, Y., Buil, A., Collins, B. C., Gillet, L. C. J., Blum, L. C., Cheng, L.-Y., Vitek, O., Mouritsen, J., Lachance, G., Spector, T. D., Dermitzakis, E. T. & Aebersold, R. Quantitative variability of 342 plasma proteins in a human twin population. *Molecular Systems Biology* **11**, 786-n/a, doi:10.15252/msb.20145728 (2015).

3. Liu, Y., Hüttenhain, R., Surinova, S., Gillet, L. C. J., Mouritsen, J., Brunner, R., Navarro, P. & Aebersold, R. Quantitative measurements of n-linked glycoproteins in human plasma by swath-ms. *Proteomics* **13**, 1247-1256, doi:10.1002/pmic.201200417 (2013).

4. Röst, H. L., Rosenberger, G., Navarro, P., Gillet, L., Miladinović, S. M., Schubert, O. T., Wolski, W., Collins, B. C., Malmström, J., Malmström, L., Aebersold, R., Lunds tekniska högskola, L. T. H., Department of, I., Institutionen för, I., Lunds, u., Institutioner vid, L. T. H., Departments at, L. T. H., Lund, U. & Faculty of Engineering, L. T. H. a. L. U. Openswath enables automated, targeted analysis of data-independent acquisition ms data. *Nature Biotechnology* **32**, 219, doi:10.1038/nbt.2841 (2014).

5. Schubert, O. T., Gillet, L. C., Collins, B. C., Navarro, P., Rosenberger, G., Wolski, W. E., Lam, H., Amodei, D., Mallick, P., MacLean, B. & Aebersold, R. Building high-quality assay libraries for targeted analysis of swath ms data. *Nature Protocols* **10**, 426-441, doi:10.1038/nprot.2015.015 (2015).

6. Nesvizhskii, A. I. Proteogenomics: Concepts, applications and computational strategies. *Nature Methods* **11**, 1114-1125, doi:10.1038/nmeth.3144 (2014).

7. Armengaud, J., Trapp, J., Pible, O., Geffard, O., Chaumot, A. & Hartmann, E. M. Non-model organisms, a species endangered by proteogenomics. *Journal of Proteomics* **105**, 5, doi:10.1016/j.jprot.2014.01.007 (2014).

8. Renuse, S., Chaerkady, R. & Pandey, A. Proteogenomics. *Proteomics* **11**, 620-630, doi:10.1002/pmic.201000615 (2011).

9. Rinner, O. in *5th Berliner LC /MS/MS Symposium* (BIOGNOSYS Proteomics - Next Generation, Berlin, Germany, 2013).

10. Chemonges, S., Shekar, K., Tung, J. P., Dunster, K. R., Diab, S., Platts, D., Watts, R. P., Gregory, S. D., Foley, S., Simonova, G., McDonald, C., Hayes, R., Bellpart, J., Timms, D., Chew, M., Fung, Y. L., Toon, M., Maybauer, M. O. & Fraser, J. F. Optimal management of the critically ill: Anaesthesia, monitoring, data capture, and point-of-care technological practices in ovine models of critical care. *Biomed Res Int* **2014**, 468309, doi:10.1155/2014/468309 (2014).

11. DiVincenti, L., Westcott, R. & Lee, C. Sheep (ovis aries) as a model for cardiovascular surgery and management before, during, and after cardiopulmonary bypass. *Journal of the American Association for Laboratory Animal Science : JAALAS* **53**, 439-448 (2014).

12. Leroux, A. A., Moonen, M. L., Pierard, L. A., Kolh, P. & Amory, H. Animal models of mitral regurgitation induced by mitral valve chordae tendineae rupture. *Journal of Heart Valve Disease* **21**, 416-423 (2012).

13. Chemonges, S. Learning from critical care management of sheep receiving extra-corporeal membrane oxygenation for smoke-induced acute lung injury as a tool for processing large clinical datasets. *bioRxiv*, doi:10.1101/058511 (2016).

14. Tung, J. P., Fung, Y. L., Nataatmadja, M., Colebourne, K. I., Esmaeel, H. M., Wilson, K., Barnett, A. G., Wood, P., Silliman, C. C. & Fraser, J. F. A novel in vivo ovine model of transfusion-related acute lung injury (trali). *Vox Sanguinis* **100**, 219-230, doi:10.1111/j.1423-0410.2010.01381.x (2011).

15. Chemonges , S. Suspected selective susceptibility to endotoxin in an ovine model. *Online Journal of Veterinary Research* **18**, 941-963 (2014).

16. Chemonges, S., Gupta, R., Mills, P. C., Kopp, S. R. & Sadowski, P. Characterisation of the circulating acellular proteome of healthy sheep using lc-ms/ms-based proteomics analysis of serum. *Proteome Science* **15**, 11, doi:10.1186/s12953-017-0119-z (2017).

17. Chemonges, S., Tung, J. P. & Fraser, J. F. Proteogenomics of selective susceptibility to endotoxin using circulating acute phase biomarkers and bioassay development in sheep: A review. *Proteome Sci* **12**, 12, doi:10.1186/1477-5956-12-12 (2014).

18. Liu, T., Qian, W.-J., Mottaz, H. M., Gritsenko, M. A., Norbeck, A. D., Moore, R. J., Purvine, S. O., Camp, D. G. & Smith, R. D. Evaluation of multi-protein immunoaffinity subtraction for plasma proteomics and candidate biomarker discovery using mass spectrometry. *Molecular & cellular proteomics : MCP* **5**, 2167-2174, doi:10.1074/mcp.T600039-MCP200 (2006).

19. Vizcaino, J. A., Cote, R. G., Csordas, A., Dianes, J. A., Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Perez-Riverol, Y., Reisinger, F., Rios, D., Wang, R. & Hermjakob, H. The proteomics identifications (pride) database and associated tools: Status in 2013. *Nucleic Acids Research* **41**, D1063-1069, doi:10.1093/nar/gks1262 (2013).

20. Consortium, T. U. Uniprot: A hub for protein information. *Nucleic Acids Research* **43**, D204-D212, doi:10.1093/nar/gku989 (2015).

21. Ceciliani, F., Eckersall, D., Burchmore, R. & Lecchi, C. Proteomics in veterinary medicine: Applications and trends in disease pathogenesis and diagnostics. *Veterinary Pathology* **51**, 351-362, doi:10.1177/0300985813502819 (2014).

22. Di Girolamo, F., D'Amato, A., Lante, I., Signore, F., Muraca, M. & Putignani, L. Farm animal serum proteomics and impact on human health. *International journal of molecular sciences* **15**, 15396-15411, doi:10.3390/ijms150915396 (2014).

23. Englard, S. & Seifter, S. in *Methods in enzymology* Vol. Volume 182 (ed P. Deutscher Murray) 285-300 (Academic Press, 1990).

24. Wu, X., Xiong, E., Wang, W., Scali, M. & Cresti, M. Universal sample preparation method integrating trichloroacetic acid/acetone precipitation with phenol extraction for crop proteomic analysis. *Nat. Protocols* **9**, 362-374, doi:10.1038/nprot.2014.022 (2014).

25. Antonioli, P., Bachi, A., Fasoli, E. & Righetti, P. G. Efficient removal of DNA from proteomic samples prior to two-dimensional map analysis. *Journal of Chromatography A* **1216**, 3606-3612, doi:10.1016/j.chroma.2008.11.053 (2009).

26. Garcia-Rodriguez, S., Castilla, S. A., Machado, A. & Ayala, A. Comparison of methods for sample preparation of individual rat cerebrospinal fluid samples prior to two-dimensional polyacrylamide gel electrophoresis. *Biotechnology Letters* **25**, 1899-1903, doi:10.1023/B:BILE.0000003979.92664.11 (2003).

27. Anderson, N. L. & Anderson, N. G. The human plasma proteome: History, character, and diagnostic prospects. *Molecular and Cellular Proteomics* **1**, 845-867 (2002).

28. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J. & Klenk, D. C. Measurement of protein using bicinchoninic acid. *Analytical Biochemistry* **150**, 76-85 (1985).

29. Wasinger, V. C., Zeng, M. & Yau, Y. Current status and advances in quantitative proteomic mass spectrometry. *International journal of proteomics* **2013**, 180605-180612, doi:10.1155/2013/180605 (2013).

30. Leng, S. X., McElhaney, J. E., Walston, J. D., Xie, D., Fedarko, N. S. & Kuchel, G. A. Elisa and multiplex technologies for cytokine measurement in inflammation and aging research. *The journals of gerontology. Series A, Biological sciences and medical sciences* **63**, 879-884 (2008).

31. Marx, V. Targeted proteomics. *Nature Methods* **10**, 19-22 (2013).

32. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207, doi:10.1038/nature01511 (2003).

33. Zhong, L., Taylor, D. L. & Whittington, R. J. Proteomic profiling of ovine serum by seldi-tof ms: Optimisation, reproducibility and feasibility of biomarker discovery using routinely collected samples. *Comparative Immunology Microbiology and Infectious Diseases* **33**, doi:10.1016/j.cimid.2008.07.009 (2010).

34. Domon, B. & Aebersold, R. Options and considerations when selecting a quantitative proteomics strategy. *Nature Biotechnology* **28**, 710-721, doi:10.1038/nbt.1661 (2010).

35. Bondar, O. P., Barnidge, D. R., Klee, E. W., Davis, B. J. & Klee, G. G. Lc-ms/ms quantification of zn-alpha 2 glycoprotein: A potential serum biomarker for prostate cancer. *Clinical Chemistry* **53**, 673-678, doi:10.1373/clinchem.2006.079681 (2007).

36. Nawy, T. Non-model organisms. *Nature Methods* **9**, 37-37, doi:10.1038/nmeth.1824 (2012).

37. Zhong, L., Taylor, D., Begg, D. J. & Whittington, R. J. Biomarker discovery for ovine paratuberculosis (johne's disease) by proteomic serum profiling. *Comparative Immunology Microbiology and Infectious Diseases* **34** (2011).

38. Nature_Publishing_Group. *Non-model organisms*, http://www.nature.com/subjects/non-model-organisms (2016).

39. Sullivan, W. The institute for the study of non–model organisms and other fantasies. *Molecular Biology of the Cell* **26**, 387-389, doi:10.1091/mbc.E14-03-0814 (2015).

40. Fuentealba, I. C. & Aburto, E. M. Animal models of copper-associated liver disease. *Comparative Hepatology* **2**, 5-5, doi:10.1186/1476-5926-2-5 (2003).

41. Back, S. A., Riddle, A., Dean, J. & Hohimer, A. R. The instrumented fetal sheep as a model of cerebral white matter injury in the premature infant. *Neurotherapeutics* **9**, 359-370, doi:10.1007/s13311-012-0108-y (2012).

42. Martini, L., Fini, M., Giavaresi, G. & Giardino, R. Sheep model in orthopedic research: A literature review. *Comparative Medicine* **51**, 292-299 (2001).

43. Nanjappa, V., Thomas, J. K., Marimuthu, A., Muthusamy, B., Radhakrishnan, A., Sharma, R., Ahmad Khan, A., Balakrishnan, L., Sahasrabuddhe, N. A., Kumar, S., Jhaveri, B. N., Sheth, K. V., Kumar Khatana, R., Shaw, P. G., Srikanth, S. M., Mathur, P. P., Shankar, S., Nagaraja, D., Christopher, R., Mathivanan, S. *et al.* Plasma proteome database as a resource for proteomics research: 2014 update. *Nucleic Acids Research* **42**, D959-965, doi:10.1093/nar/gkt1251 (2014).

44. Parker, C. E. & Borchers, C. H. Mass spectrometry based biomarker discovery, verification, and validation--quality assurance and control of protein biomarker assays. *Molecular Oncology* **8**, 840-858, doi:10.1016/j.molonc.2014.03.006 (2014).

45. Whiteaker, J. R., Lin, C., Kennedy, J., Hou, L., Trute, M., Sokal, I., Yan, P., Schoenherr, R. M., Zhao, L., Voytovich, U. J., Kelly-Spratt, K. S., Krasnoselsky, A., Gafken, P. R., Hogan, J. M., Jones, L. A., Wang, P., Amon, L., Chodosh, L. A., Nelson, P. S., McIntosh, M. W. *et al.* A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nature Biotechnology* **29**, 625-634, doi:10.1038/nbt.1900 (2011).

46. Anderson, N. L. The clinical plasma proteome: A survey of clinical assays for proteins in plasma and serum. *Clinical Chemistry* **56**, 177-185, doi:10.1373/clinchem.2009.126706 (2010).

47. Williams, D. K. & Muddiman, D. C. Absolute quantification of c-reactive protein in human plasma derived from patients with epithelial ovarian cancer utilizing protein cleavage isotope dilution mass spectrometry. *Journal of Proteome Research* **8**, 1085-1090, doi:10.1021/pr800922p (2009).

48. Bantscheff, M., Lemeer, S., Savitski, M. M. & Kuster, B. Quantitative mass spectrometry in proteomics: Critical review update from 2007 to the present. *Analytical and Bioanalytical Chemistry* **404**, 939-965, doi:10.1007/s00216-012-6203-4 (2012).

49. Pernemalm, M. & Lehtio, J. Mass spectrometry-based plasma proteomics: State of the art and future outlook. *Expert Rev Proteomics* **11**, 431-448, doi:10.1586/14789450.2014.901157 (2014).

50. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols* **1**, 2856-2860, doi:10.1038/nprot.2006.468 (2006).

51. Smith, M. P. W., Banks, R. E., Wood, S. L., Zougman, A., Ho, J. T. C., Peng, J., Jackson, D., Cairns, D. A., Lewington, A. J. P. & Selby, P. J. A systematic analysis of the effects of increasing degrees of serum immunodepletion in terms of depth of coverage and other key aspects in top-down and bottom-up proteomic analyses. *Proteomics* **11**, 2222-2235, doi:10.1002/pmic.201100005 (2011).

52. Pieper, R., Su, Q., Gatlin, C. L., Huang, S. T., Anderson, N. L. & Steiner, S. Multi-component immunoaffinity subtraction chromatography: An innovative step towards a comprehensive survey of the human plasma proteome. *Proteomics* **3**, 422-432, doi:10.1002/pmic.200390057 (2003).

53. Tu, C., Rudnick, P. A., Martinez, M. Y., Cheek, K. L., Stein, S. E., Slebos, R. J. C. & Liebler, D. C. Depletion of abundant plasma proteins and limitations of plasma proteomics. *Journal of Proteome Research* **9**, 4982-4991, doi:10.1021/pr100646w (2010).

54. Pernemalm, M., Lewensohn, R. & Lehtio, J. Affinity prefractionation for ms-based plasma proteomics. *Proteomics* **9**, 1420-1427, doi:10.1002/pmic.200800377 (2009).

55. Thulasiraman, V., Lin, S., Gheorghiu, L., Lathrop, J., Lomas, L., Hammond, D. & Boschetti, E. Reduction of the concentration difference of proteins in biological liquids using a library of combinatorial ligands. *Electrophoresis* **26**, 3561-3571, doi:10.1002/elps.200500147 (2005).

56. Hagiwara, T., Saito, Y., Nakamura, Y., Tomonaga, T., Murakami, Y. & Kondo, T. Combined use of a solid-phase hexapeptide ligand library with liquid chromatography and two-dimensional difference gel electrophoresis for intact plasma proteomics. *International journal of proteomics* **2011**, 739615, doi:10.1155/2011/739615 (2011).

57. Mirzaei, H., Baena, B., Barbas, C. & Regnier, F. Identification of oxidized proteins in rat plasma using avidin chromatography and tandem mass spectrometry. *Proteomics* **8**, 1516-1527, doi:10.1002/pmic.200700363 (2008).

58. Liu, T., Qian, W. J., Strittmatter, E. F., Camp, D. G., 2nd, Anderson, G. A., Thrall, B. D. & Smith, R. D. High-throughput comparative proteome analysis using a quantitative cysteinyl-peptide enrichment technology. *Analytical Chemistry* **76**, 5345-5353, doi:10.1021/ac049485q (2004).

59. Zhao, J., Simeone, D. M., Heidt, D., Anderson, M. A. & Lubman, D. M. Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: Application to pancreatic cancer serum. *Journal of Proteome Research* **5**, 1792-1802, doi:10.1021/pr060034r (2006).

60. Yang, Z. & Hancock, W. S. Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *Journal of Chromatography A* **1053**, 79-88 (2004).

61. Zhang, H., Li, X. J., Martin, D. B. & Aebersold, R. Identification and quantification of n-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nature Biotechnology* **21**, 660-666, doi:10.1038/nbt827 (2003).

62. Larsen, M. R., Jensen, S. S., Jakobsen, L. A. & Heegaard, N. H. Exploring the sialiome using titanium dioxide chromatography and mass spectrometry. *Molecular and Cellular Proteomics* **6**, 1778-1787, doi:10.1074/mcp.M700086-MCP200 (2007).

63. Banks, R. E., Stanley, A. J., Cairns, D. A., Barrett, J. H., Clarke, P., Thompson, D. & Selby, P. J. Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. *Clinical Chemistry* **51**, 1637-1649, doi:10.1373/clinchem.2005.051417 (2005).

64. Cheng, A. J., Chen, L. C., Chien, K. Y., Chen, Y. J., Chang, J. T., Wang, H. M., Liao, C. T. & Chen, I. H. Oral cancer plasma tumor marker identified with bead-based affinity-fractionated proteomic technology. *Clinical Chemistry* **51**, 2236-2244, doi:10.1373/clinchem.2005.052324 (2005).

65. Gillette, M. A. & Carr, S. A. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nature Methods* **10**, 28, doi:10.1038/nmeth.2309 (2013).

66. Dayon, L. & Kussmann, M. Proteomics of human plasma: A critical comparison of analytical workflows in terms of effort, throughput and outcome. *EuPA Open Proteomics* **1**, 8-16, doi:https://doi.org/10.1016/j.euprot.2013.08.001 (2013).

67. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C. & Yates, J. R. Protein analysis by shotgun/bottom-up proteomics. *Chemical Reviews* **113**, 2343-2394, doi:10.1021/cr3003533 (2013).

68. Olver, C. S., Webb, T. L., Long, L. J., Scherman, H. & Prenni, J. E. Comparison of methods for depletion of albumin and igg from equine serum. *Veterinary Clinical Pathology* **39**, 337-345, doi:10.1111/j.1939-165X.2010.00241.x (2010).

69. Faulkner, S., Elia, G., Hillard, M., O'Boyle, P., Dunn, M. & Morris, D. Immunodepletion of albumin and immunoglobulin g from bovine plasma. *Proteomics* **11**, 2329-2335, doi:10.1002/pmic.201000364 (2011).

70. Marco-Ramell, A. & Bassols, A. Enrichment of low-abundance proteins from bovine and porcine serum samples for proteomic studies. *Research in Veterinary Science* **89**, 340-343, doi:10.1016/j.rvsc.2010.03.019 (2010).

71. Bellei, E., Bergamini, S., Monari, E., Fantoni, L. I., Cuoghi, A., Ozben, T. & Tomasi, A. High-abundance proteins depletion for serum proteomic analysis: Concomitant removal of non-targeted proteins. *Amino Acids* **40**, 145-156, doi:10.1007/s00726-010-0628-x (2011).

72. Di Girolamo, F., Righetti, P. G., Soste, M., Feng, Y. & Picotti, P. Reproducibility of combinatorial peptide ligand libraries for proteome capture evaluated by selected reaction monitoring. *Journal of Proteomics* **89**, 215-226, doi:10.1016/j.jprot.2013.05.037 (2013).

73. Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Hermjakob, H., Apweiler, R., Haab, B. B., Simpson, R. J., Eddes, J. S., Kapp, E. A., Moritz, R. L., Chan, D. W., Rai, A. J., Admon, A., Aebersold, R., Eng, J., Hancock, W. S., Hefta, S. A., Meyer, H. *et al.* Overview of the hupo plasma proteome project: Results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **5**, 3226-3245, doi:10.1002/pmic.200500358 (2005).

74. States, D. J., Omenn, G. S., Blackwell, T. W., Fermin, D., Eng, J., Speicher, D. W. & Hanash, S. M. Challenges in deriving high-confidence protein identifications from data gathered by a hupo plasma proteome collaborative study. *Nature Biotechnology* **24**, 333-338, doi:10.1038/nbt1183 (2006).

75. Kirby, G. M., Mackay, A., Grant, A., Woods, P., McEwen, B., Khanna, C., Macri, J., Hayes, M. A. & Stalker, M. Concentration of lipocalin region of collagen xxvii alpha 1 in the serum of dogs with hemangiosarcoma. *Journal of veterinary internal medicine / American College of Veterinary Internal Medicine* **25**, 497-503, doi:10.1111/j.1939-1676.2011.0709.x (2011).

76. Sun, J.-f., Shi, Z.-x., Guo, H.-c., Li, S. & Tu, C.-c. Proteomic analysis of swine serum following highly virulent classical swine fever virus infection. *Virology Journal* **8**, 107-107, doi:10.1186/1743-422X-8-107 (2011).

77. Seth, M., Lamont, E. A., Janagama, H. K., Widdel, A., Vulchanova, L., Stabel, J. R., Waters, W. R., Palmer, M. V. & Sreevatsan, S. Biomarker discovery in subclinical mycobacterial infections of cattle. *PloS One* **4**, e5478, doi:10.1371/journal.pone.0005478 (2009).

78. Addona, T. A., Shi, X., Keshishian, H., Mani, D. R., Burgess, M., Gillette, M. A., Clauser, K. R., Shen, D., Lewis, G. D., Farrell, L. A., Fifer, M. A., Sabatine, M. S., Gerszten, R. E. & Carr, S. A. A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease. *Nature Biotechnology* **29**, 635-643, doi:10.1038/nbt.1899 (2011).

79. Guryca, V., Lamerz, J., Ducret, A. & Cutler, P. Qualitative improvement and quantitative assessment of n-terminomics. *Proteomics* **12**, 1207-1216, doi:10.1002/pmic.201100430 (2012).

80. Marco-Ramell, A., Arroyo, L., Saco, Y., García-Heredia, A., Camps, J., Fina, M., Piedrafita, J. & Bassols, A. Proteomic analysis reveals oxidative stress response as the main adaptative physiological mechanism in cows under different production systems. *Journal of Proteomics* **75**, 4399-4411, doi:10.1016/j.jprot.2012.04.002 (2012).

81. Vander Lugt, M. T., Braun, T. M., Hanash, S., Ritz, J., Ho, V. T., Antin, J. H., Zhang, Q., Wong, C. H., Wang, H., Chin, A., Gomez, A., Harris, A. C., Levine, J. E., Choi, S. W., Couriel, D., Reddy, P., Ferrara, J. L. & Paczesny, S. St2 as a marker for risk of therapy-resistant graft-versus-host disease and death. *New England Journal of Medicine* **369**, 529-539, doi:10.1056/NEJMoa1213299 (2013).

82. Ratcliffe, L., Mian, S., Slater, K., King, H., Napolitano, M., Aucoin, D. & Mobasheri, A. Proteomic identification and profiling of canine lymphoma patients. *Veterinary and Comparative Oncology* **7**, 92 (2009).

83. Farrah, T., Deutsch, E. W. & Aebersold, R. Using the human plasma peptideatlas to study human plasma proteins. *Methods in Molecular Biology* **728**, 349-374, doi:10.1007/978-1-61779-068-3_23 (2011).

84. Farrah, T., Deutsch, E. W., Omenn, G. S., Campbell, D. S., Sun, Z., Bletz, J. A., Mallick, P., Katz, J. E., Malmstrom, J., Ossola, R., Watts, J. D., Lin, B., Zhang, H., Moritz, R. L. & Aebersold, R. A high-confidence human plasma proteome reference set with estimated concentrations in peptideatlas. *Molecular and Cellular Proteomics* **10**, M110 006353, doi:10.1074/mcp.M110.006353 (2011).

85. Garbis, S. D., Roumeliotis, T. I., Tyritzis, S. I., Zorpas, K. M., Pavlakis, K. & Constantinides, C. A. A novel multidimensional protein identification technology approach combining protein size exclusion prefractionation, peptide zwitterion-ion hydrophilic interaction chromatography, and nano-ultraperformance rp chromatography/nesi-ms2 for the in-depth analysis of the serum proteome and phosphoproteome: Application to clinical sera derived from humans with benign prostate hyperplasia. *Analytical Chemistry* **83**, 708-718, doi:10.1021/ac102075d (2011).

86. Beer, L. A., Wang, H., Tang, H. Y., Cao, Z., Chang-Wong, T., Tanyi, J. L., Zhang, R., Liu, Q. & Speicher, D. W. Identification of multiple novel protein biomarkers shed by human serous ovarian tumors into the blood of immunocompromised mice and verified in patient sera. *PloS One* **8**, e60129, doi:10.1371/journal.pone.0060129 (2013).

87. Paczesny, S., Braun, T. M., Levine, J. E., Hogan, J., Crawford, J., Coffing, B., Olsen, S., Choi, S. W., Wang, H., Faca, V., Pitteri, S., Zhang, Q., Chin, A., Kitko, C., Mineishi, S., Yanik, G., Peres, E., Hanauer, D., Wang, Y., Reddy, P. *et al.* Elafin is a biomarker of graft-versus-host disease of the skin. *Science Translational Medicine* **2**, 13ra12, doi:10.1126/scitranslmed.3000406 (2010).

88. Aich, P., Babiuk, L. A., Potter, A. A. & Griebel, P. Biomarkers for prediction of bovine respiratory disease outcome. *Omics : a journal of integrative biology* **13**, 199-209, doi:10.1089/omi.2009.0012 (2009).

89. Steelman, S. M. & Chowdhary, B. P. Plasma proteomics shows an elevation of the anti-inflammatory protein apoa-iv in chronic equine laminitis. *BMC Veterinary Research* **8**, 179-179, doi:10.1186/1746-6148-8-179 (2012).

90. Cima, I., Schiess, R., Wild, P., Kaelin, M., Schuffler, P., Lange, V., Picotti, P., Ossola, R., Templeton, A., Schubert, O., Fuchs, T., Leippold, T., Wyler, S., Zehetner, J., Jochum, W., Buhmann, J., Cerny, T., Moch, H., Gillessen, S., Aebersold, R. *et al.* Cancer genetics-guided discovery of serum biomarker signatures for diagnosis and prognosis of prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 3342-3347, doi:10.1073/pnas.1013699108 (2011).

91. Wildes, D. & Wells, J. A. Sampling the n-terminal proteome of human blood. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 4561-4566, doi:10.1073/pnas.0914495107 (2010).

92. Shen, Y., Kim, J., Strittmatter, E. F., Jacobs, J. M., Camp, D. G., 2nd, Fang, R., Tolie, N., Moore, R. J. & Smith, R. D. Characterization of the human blood plasma proteome. *Proteomics* **5**, 4034-4045, doi:10.1002/pmic.200401246 (2005).

93. Cao, Z., Yende, S., Kellum, J. A. & Robinson, R. A. S. Additions to the human plasma proteome via a tandem mars depletion itraq-based workflow. *International journal of proteomics* **2013**, 654356-654358, doi:10.1155/2013/654356 (2013).

94. Rioux, M. C., Carmona, C., Acosta, D., Ward, B., Ndao, M., Gibbs, B. F., Bennett, H. P. & Spithill, T. W. Discovery and validation of serum biomarkers expressed over the first twelve weeks of fasciola hepatica infection in sheep. *International Journal for Parasitology* **38**, 123-136, doi:10.1016/j.ijpara.2007.07.017 (2008).

95. Millioni, R., Tolin, S., Puricelli, L., Sbrignadello, S., Fadini, G. P., Tessari, P. & Arrigoni, G. High abundance proteins depletion vs low abundance proteins enrichment: Comparison of methods to reduce the plasma proteome complexity. *PloS One* **6**, e19603, doi:10.1371/journal.pone.0019603 (2011).

96. Turk, R., Piras, C., Kovačić, M., Samardžija, M., Ahmed, H., De Canio, M., Urbani, A., Meštrić, Z. F., Soggiu, A., Bonizzi, L. & Roncada, P. Proteomics of inflammatory and oxidative stress response in cows with subclinical and clinical mastitis. *Journal of Proteomics* **75**, 4412-4428, doi:10.1016/j.jprot.2012.05.021 (2012).

97. Gautam, P., Nair, S. C., Ramamoorthy, K., Swamy, C. V. & Nagaraj, R. Analysis of human blood plasma proteome from ten healthy volunteers from indian population. *PloS One* **8**, e72584, doi:10.1371/journal.pone.0072584 (2013).

98. Zhou, H., Froehlich, J. W., Briscoe, A. C. & Lee, R. S. The glycofilter: A simple and comprehensive sample preparation platform for proteomics, n-glycomics and glycosylation site assignment. *Molecular and Cellular Proteomics* **12**, 2981-2991, doi:10.1074/mcp.M113.027953 (2013).

99. Zhao, X., Ma, C., Han, H., Jiang, J., Tian, F., Wang, J., Ying, W. & Qian, X. Comparison and optimization of strategies for a more profound profiling of the sialylated n-glycoproteomics in human plasma using metal oxide enrichment. *Analytical and Bioanalytical Chemistry* **405**, 5519-5529, doi:10.1007/s00216-013-6971-5 (2013).

100. Genini, S., Paternoster, T., Costa, A., Botti, S., Luini, M. V., Caprera, A. & Giuffra, E. Identification of serum proteomic biomarkers for early porcine reproductive and respiratory syndrome (prrs) infection. *Proteome science* **10**, 48-48, doi:10.1186/1477-5956-10-48

(2012).

101. Adkins, J. N., Varnum, S. M., Auberry, K. J., Moore, R. J., Angell, N. H., Smith, R. D., Springer, D. L. & Pounds, J. G. Toward a human blood serum proteome: Analysis by multidimensional separation coupled with mass spectrometry. *Molecular and Cellular Proteomics* **1**, 947-955 (2002).

102. Alonso-Fauste, I., Andrés, M., Iturralde, M., Lampreave, F., Gallart, J. & Álava, M. A. Proteomic characterization by 2-de in bovine serum and whey from healthy and mastitis affected farm animals. *Journal of Proteomics* **75**, 3015, doi:10.1016/j.jprot.2011.11.035 (2011).

103. De Felice, C., Cortelazzo, A., Signorini, C., Guerranti, R., Leoncini, S., Pecorelli, A., Durand, T., Galano, J.-M., Oger, C., Zollo, G., Montomoli, B., Landi, C., Bini, L., Valacchi, G., Ciccoli, L. & Hayek, J. Effects of ω-3 polyunsaturated fatty acids on plasma proteome in rett syndrome. *Mediators of Inflammation* **2013**, 723269-723269, doi:10.1155/2013/723269 (2013).

104. Qian, W. J., Kaleta, D. T., Petritis, B. O., Jiang, H., Liu, T., Zhang, X., Mottaz, H. M., Varnum, S. M., Camp, D. G., 2nd, Huang, L., Fang, X., Zhang, W. W. & Smith, R. D. Enhanced detection of low abundance human plasma proteins using a tandem igy12-supermix immunoaffinity separation strategy. *Molecular and Cellular Proteomics* **7**, 1963-1973, doi:10.1074/mcp.M800008-MCP200 (2008).

105. Barton, C., Beck, P., Kay, R., Teale, P. & Roberts, J. Multiplexed lc-ms/ms analysis of horse plasma proteins to study doping in sport. *Proteomics* **9**, 3058-3065, doi:10.1002/pmic.200800737 (2009).

106. Zielinska, D. F., Gnad, F., Wisniewski, J. R. & Mann, M. Precision mapping of an in vivo n-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897-907, doi:10.1016/j.cell.2010.04.012 (2010).

107. Boersema, P. J., Geiger, T., Wisniewski, J. R. & Mann, M. Quantification of the n-glycosylated secretome by super-silac during breast cancer progression and in human blood samples. *Molecular and Cellular Proteomics* **12**, 158-171, doi:10.1074/mcp.M112.023614 (2013).

108. Liu, Y., Zhang, K., Zheng, H., Shang, Y., Guo, J., Tian, H., Lu, G., Jin, Y., He, J., Cai, X. & Liu, X. Proteomics analysis of porcine serum proteins by lc-ms/ms after foot-and-mouth disease virus (fmdv) infection. *The Journal of veterinary medical science / the Japanese*

Society of Veterinary Science **73**, 1569, doi:10.1292/jvms.11-0019 (2011).

109. Liu, X., Valentine, S. J., Plasencia, M. D., Trimpin, S., Naylor, S. & Clemmer, D. E. Mapping the human plasma proteome by scx-lc-ims-ms. *Journal of the American Society for Mass Spectrometry* **18**, 1249-1264, doi:10.1016/j.jasms.2007.04.012 (2007).

110. Wells, B., Innocent, G. T., Eckersall, P. D., McCulloch, E., Nisbet, A. J. & Burgess, S. T. Two major ruminant acute phase proteins, haptoglobin and serum amyloid a, as serum biomarkers during active sheep scab infestation. *Veterinary Research* **44**, 103, doi:10.1186/1297-9716-44-103 (2013).

111. Chiaradia, E., Avellini, L., Tartaglia, M., Gaiti, A., Just, I., Scoppetta, F., Czentnar, Z. & Pich, A. Proteomic evaluation of sheep serum proteins. *BMC Veterinary Research* **8**, 66, doi:10.1186/1746-6148-8-66 (2012).

112. Hernandez-Castellano, L. E., Almeida, A. M., Ventosa, M., Coelho, A. V., Castro, N. & Arguello, A. The effect of colostrum intake on blood plasma proteome profile in newborn lambs: Low abundance proteins. *BMC Veterinary Research* **10**, 85, doi:10.1186/1746-6148-10-85 (2014).

113. Faulkner, S., Elia, G., Mullen, M. P., O'Boyle, P., Dunn, M. J. & Morris, D. A comparison of the bovine uterine and plasma proteome using itraq proteomics. *Proteomics* **12**, 2014-2023, doi:10.1002/pmic.201100609 (2012).

114. Andrews Kingon, G. L., Petitte, J. N., Muddiman, D. C. & Hawkridge, A. M. Multi-peptide nlc-pc-idms-srm-based assay for the quantification of biomarkers in the chicken ovarian cancer model. *Methods* **61**, 323-330, doi:10.1016/j.ymeth.2013.04.004 (2013).

115. Matafora, V., Bachi, A. & Capasso, G. Genomics and proteomics: How long do we need to reach clinical results? *Blood Purification* **36**, 7, doi:10.1159/000350578 (2013).

116. Jafari, M., Primo, V., Smejkal, G. B., Moskovets, E. V., Kuo, W. P. & Ivanov, A. R. Comparison of in-gel protein separation techniques commonly used for fractionation in mass spectrometry-based proteomic profiling. *Electrophoresis* **33**, 2516-2526, doi:10.1002/elps.201200031 (2012).

117. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage t4. *Nature* **227**, 680-685 (1970).

118. Schägger, H. Tricine-sds-page. *Nature Protocols* **1**, 16-22, doi:10.1038/nprot.2006.4 (2006).

119. Schägger, H. Electrophoretic isolation of membrane proteins from acrylamide gels. *Applied Biochemistry and Biotechnology* **48**, 185-203, doi:10.1007/BF02788741 (1994).

120. Lößner, C., Wee, S., Ler, S. G., Li, R. H. X., Carney, T., Blackstock, W. & Gunaratne, J. Expanding the zebrafish embryo proteome using multiple fractionation approaches and tandem mass spectrometry. *Proteomics* **12**, 1879-1882, doi:10.1002/pmic.201100576 (2012).

121. Markus, S., Marie-Anne, H. & Bernhard, K. Profiling core proteomes of human cell lines by one-dimensional page and liquid chromatography-tandem mass spectrometry. *Molecular & Cellular Proteomics* **2**, 1297-1305, doi:10.1074/mcp.M300087-MCP200 (2003).

122. Simpson, R. J. *Purifying proteins for proteomics: A laboratory manual*. (Cold Spring Harbor Laboratory Press, 2004).

123. Villén, J. & Gygi, S. P. The scx/imac enrichment approach for global phosphorylation analysis by mass spectrometry. *Nature Protocols* **3**, 1630-1638, doi:10.1038/nprot.2008.150 (2008).

124. Tang, J., Gao, M., Deng, C. & Zhang, X. Recent development of multi-dimensional chromatography strategies in proteome research. *Journal of Chromatography. B: Analytical Technologies in the Biomedical and Life Sciences* **866**, 123-132, doi:10.1016/j.jchromb.2008.01.029 (2008).

125. Ly, L. & Wasinger, V. C. Protein and peptide fractionation, enrichment and depletion: Tools for the complex proteome. *Proteomics* **11**, 513-534, doi:10.1002/pmic.201000394 (2011).

126. Clark, J. *Chemguide: Helping you to understand chemistry*, http://www.chemguide.co.uk/ (2014).

127. Delle Site, A. Factors affecting sorption of organic compounds in natural sorbent/water systems and sorption coefficients for selected pollutants. A review. *Journal of Physical and Chemical Reference Data* **30**, 187, doi:10.1063/1.1347984 (2001).

128. Gray, J. J. The interaction of proteins with solid surfaces. *Current Opinion in Structural Biology* **14**, 110-115, doi:10.1016/j.sbi.2003.12.001 (2004).

129. Xie, F., Smith, R. D. & Shen, Y. Advanced proteomic liquid chromatography. *Journal of Chromatography A* **1261**, 78-90, doi:10.1016/j.chroma.2012.06.098 (2012).

130. Hyung, S. W., Kim, M. S., Mun, D. G., Lee, H. & Lee, S. W. The effect and potential of using a temperature controlled separation column with ultra-high pressure microcapillary liquid chromatography/tandem mass spectrometry on proteomic analysis. *Analyst* **136**, 2100-2105, doi:10.1039/c0an00724b (2011).

131. Chen, Y., Mant, C. T. & Hodges, R. S. Temperature selectivity effects in reversed-phase liquid chromatography due to conformation differences between helical and non-helical peptides. *Journal of Chromatography A* **1010**, 45-61 (2003).

132. Hartmann, E., Chen, Y., Mant, C. T., Jungbauer, A. & Hodges, R. S. Comparison of reversed-phase liquid chromatography and hydrophilic interaction/cation-exchange chromatography for the separation of amphipathic alpha-helical peptides with l- and d-amino acid substitutions in the hydrophilic face. *Journal of Chromatography A* **1009**, 61-71 (2003).

133. Mant, C. T., Chen, Y. & Hodges, R. S. Temperature profiling of polypeptides in reversed-phase liquid chromatography. I. Monitoring of dimerization and unfolding of amphipathic alpha-helical peptides. *Journal of Chromatography A* **1009**, 29-43 (2003).

134. Nagaraj, N., Kulak, N. A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O. & Mann, M. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra hplc runs on a bench top orbitrap. *Molecular and Cellular Proteomics* **11**, M111 013722, doi:10.1074/mcp.M111.013722 (2012).

135. Queiroz, J. A., Tomaz, C. T. & Cabral, J. M. S. Hydrophobic interaction chromatography of proteins. *Journal of Biotechnology* **87**, 143-159, doi:10.1016/S0168-1656(01)00237-1 (2001).

136. Liu, H., Lin, D. & Yates, J. R., 3rd. Multidimensional separations for protein/peptide analysis in the post-genomic era. *Biotechniques* **32**, 898, 900, 902 passim (2002).

137. Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., MacCoss, M. J. & Rinner, O. Using irt, a normalized retention time for more targeted measurement of peptides. *Proteomics* **12**, 1111-1121, doi:10.1002/pmic.201100463 (2012).

138. Vehus, T., Roberg-Larsen, H., Waaler, J., Aslaksen, S., Krauss, S., Wilson, S. R. & Lundanes, E. Versatile, sensitive liquid chromatography mass spectrometry - implementation of 10 mum ot columns suitable for small molecules, peptides and proteins. *Scientific Reports* **6**, 37507, doi:10.1038/srep37507 (2016).

139. Hsieh, E. J., Bereman, M. S., Durand, S., Valaskovic, G. A. & MacCoss, M. J. Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow lc-ms/ms of complex proteomic samples. *Journal of the American Society for Mass Spectrometry* **24**, 148-153, doi:10.1007/s13361-012-0508-6 (2013).

140. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: Workflows, potential, pitfalls and future directions. *Nature Methods* **9**, 555, doi:10.1038/nmeth.2015 (2012).

141. Bruce, C., Stone, K., Gulcicek, E. & Williams, K. Proteomics and the analysis of proteomic data: 2013 overview of current protein-profiling technologies. *Curr Protoc Bioinformatics* **Chapter 13**, Unit 13 21, doi:10.1002/0471250953.bi1321s41 (2013).

142. Smythe, W. R., Rumbaugh, L. H. & West, S. S. A high intensity mass-spectrometer. *Physical Review* **45**, 724-727, doi:10.1103/PhysRev.45.724 (1934).

143. Chang, S. Y. & Yeung, E. S. Laser vaporization/ionization interface for capillary electrophoresis-time-of-flight mass spectrometry. *Analytical Chemistry* **69**, 2251-2257, doi:10.1021/ac961071c (1997).

144. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry* **60**, 2299-2301 (1988).

145. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).

146. Macfarlane, R. D. & Torgerson, D. F. Californium-252 plasma desorption mass spectroscopy. *Science* **191**, 920-925 (1976).

147. Robb, D. B., Covey, T. R. & Bruins, A. P. Atmospheric pressure photoionization: An ionization method for liquid chromatography-mass spectrometry. *Analytical Chemistry* **72**, 3653-3659 (2000).

148. Dettmer, K., Aronov, P. A. & Hammock, B. D. Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews* **26**, 51-78, doi:10.1002/mas.20108 (2007).

149. Hanley, L. & Zimmermann, R. Light and molecular ions: The emergence of vacuum uv single-photon ionization in ms. *Analytical Chemistry* **81**, 4174-4182, doi:10.1021/ac8013675 (2009).

150. Biemann, K. Contributions of mass spectrometry to peptide and protein structure. *Biomedical and Environmental Mass Spectrometry* **16**, 99-111 (1988).

151. Clarke, W. in *Mass spectrometry for the clinical laboratory* 1-15 (Academic Press, 2017).

152. Wells, J. M. & McLuckey, S. A. Collision-induced dissociation (cid) of peptides and proteins. *Methods in Enzymology* **402**, 148-185, doi:10.1016/s0076-6879(05)02005-7 (2005).

153. Kind, T. & Fiehn, O. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanalytical Reviews* **2**, 23-60, doi:10.1007/s12566-010-0015-9 (2010).

154. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: A tutorial. *Molecular Systems Biology* **4**, 222-222, doi:10.1038/msb.2008.61 (2008).

155. Liu, R., Li, Q. & Smith, L. M. Detection of large ions in time-of-flight mass spectrometry: Effects of ion mass and acceleration voltage on microchannel plate detector response. *Journal of the American Society for Mass Spectrometry* **25**, 1374-1383, doi:10.1007/s13361-014-0903-2 (2014).

156. Chen, X., Westphall, M. S. & Smith, L. M. Mass spectrometric analysis of DNA mixtures: Instrumental effects responsible for decreased sensitivity with increasing mass. *Analytical Chemistry* **75**, 5944-5952, doi:10.1021/ac030127h (2003).

157. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry* **11**, 601, doi:10.1002/bms.1200111109 (1984).

158. Karpievitch, Y. V., Polpitiya, A. D., Anderson, G. A., Smith, R. D. & Dabney, A. R. Liquid chromatography mass spectrometry-based proteomics: Biological and technological aspects. *The annals of applied statistics* **4**, 1797-1823, doi:10.1214/10-AOAS341 (2010).

159. Risk, B. A., Spitzer, W. J. & Giddings, M. C. Peppy: Proteogenomic search software. *Journal of Proteome Research* **12**, 3019-3025, doi:10.1021/pr400208w (2013).

160. Cleveland, J. P. & Rose, J. R. Identification of b-/y-ions in ms/ms spectra using a two stage neural network. *Proteome Science* **11**, S4-S4, doi:10.1186/1477-5956-11-S1-S4 (2013).

161. Valeja, S. G., Kaiser, N. K., Xian, F., Hendrickson, C. L., Rouse, J. C. & Marshall, A. G. Unit mass baseline resolution for an intact 148 kda therapeutic monoclonal antibody by fourier transform ion cyclotron resonance mass spectrometry. *Analytical Chemistry* **83**, 8391 (2011).

162. Craig, R. & Beavis, R. C. A method for reducing the time required to match protein sequences with tandem mass spectra. *Rapid communications in mass spectrometry : RCM* **17**, 2310-2316, doi:10.1002/rcm.1198 (2003).

163. Craig, R. & Beavis, R. C. Tandem: Matching proteins with tandem mass spectra. *Bioinformatics (Oxford, England)* **20**, 1466-1467, doi:10.1093/bioinformatics/bth092 (2004).

164. Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P., Hunter, C. L., Nuwaysir, L. M. & Schaeffer, D. A. The paragon algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Molecular & Cellular Proteomics* **6**, 1638-1655, doi:10.1074/mcp.T600050-MCP200 (2007).

165. Tagu, D., Colbourne, J. K. & Nègre, N. Genomic data integration for ecological and evolutionary traits in non-model organisms. *BMC Genomics* **15**, 490-490, doi:10.1186/1471-2164-15-490 (2014).

166. Surinova, S., Hüttenhain, R., Chang, C.-Y., Espona, L., Vitek, O. & Aebersold, R. Automated selected reaction monitoring data analysis workflow for large-scale targeted proteomic studies. *Nature Protocols* **8**, 1602, doi:10.1038/nprot.2013.091 (2013).

167. Liebler, D. C. & Zimmerman, L. J. Targeted quantitation of proteins by mass spectrometry. *Biochemistry* **52**, 3797-3806, doi:10.1021/bi400110b (2013).

168. Domanski, D., Percy, A. J., Yang, J., Chambers, A. G., Hill, J. S., Freue, G. V. & Borchers, C. H. Mrm-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. *Proteomics* **12**, 1222-1243, doi:10.1002/pmic.201100568 (2012).

169. Huttenhain, R., Soste, M., Selevsek, N., Rost, H., Sethi, A., Carapito, C., Farrah, T., Deutsch, E. W., Kusebauch, U., Moritz, R. L., Nimeus-Malmstrom, E., Rinner, O. & Aebersold, R. Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Science Translational Medicine* **4**, 142ra194, doi:10.1126/scitranslmed.3003989 (2012).

170. Percy, A. J., Chambers, A. G., Yang, J. & Borchers, C. H. Multiplexed mrm-based quantitation of candidate cancer biomarker proteins in undepleted and non-enriched human plasma. *Proteomics* **13**, 2202-2215, doi:10.1002/pmic.201200316 (2013).

171. Abbatiello, S. E., Mani, D. R., Schilling, B., Maclean, B., Zimmerman, L. J., Feng, X., Cusack, M. P., Sedransk, N., Hall, S. C., Addona, T., Allen, S., Dodder, N. G., Ghosh, M., Held, J. M., Hedrick, V., Inerowicz, H. D., Jackson, A., Keshishian, H., Kim, J. W., Lyssand, J. S. *et al.* Design, implementation and multisite evaluation of a system suitability protocol for the quantitative assessment of instrument performance in liquid chromatography-multiple reaction monitoring-ms (lc-mrm-ms). *Molecular & cellular proteomics : MCP* **12**, 2623-2639, doi:10.1074/mcp.M112.027078 (2013).

172. MacLean, B., MacCoss, M. J., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L. & Liebler, D. C. Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968, doi:10.1093/bioinformatics/btq054 (2010).

173. Stergachis, A. B., MacLean, B., Lee, K., Stamatoyannopoulos, J. A. & MacCoss, M. J. Rapid empirical discovery of optimal peptides for targeted proteomics. *Nature Methods* **8**, 1041-1043, doi:10.1038/nmeth.1770 (2011).

174. Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature Methods* **1**, 39-45, doi:10.1038/nmeth705 (2004).

175. Mollah, S. & Hunter, C. Discovery to Validation: Transition from SWATH® Acquisition to Targeted MRM Analysis for Quantitative Proteomics Pipeline. Publication number: RUO-MKT-02-2758-A, AB Sciex Pte. Ltd., SCIEX, USA (2015).

176. Vowinckel, J., Capuano, F., Campbell, K., Deery, M. J., Lilley, K. S. & Ralser, M. The beauty of being (label)-free: Sample preparation methods for swath-ms and next-generation targeted proteomics. *F1000Research* **2**, 272, doi:10.12688/f1000research.2-272.v2 (2013).

177. Sadowski, P. G., Groen, A. J., Dupree, P. & Lilley, K. S. Sub-cellular localization of membrane proteins. *Proteomics* **8**, 3991-4011, doi:10.1002/pmic.200800217 (2008).

178. Havlis, J. & Shevchenko, A. Absolute quantification of proteins in solutions and in polyacrylamide gels by mass spectrometry. *Analytical Chemistry [H.W. Wilson - GS]* **76**, 3029 (2004).

179. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. & Mann, M. Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. *Molecular and Cellular Proteomics* **1**, 376-386 (2002).

180. Karp, N. A., Huber, W., Sadowski, P. G., Charles, P. D., Hester, S. V. & Lilley, K. S. Addressing accuracy and precision issues in itraq quantitation. *Molecular & cellular proteomics : MCP* **9**, 1885-1897, doi:10.1074/mcp.M900628-MCP200 (2010).

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

182. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem ms. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 6940-6945, doi:10.1073/pnas.0832254100 (2003).

183. Beynon, R. J., Doherty, M. K., Pratt, J. M. & Gaskell, S. J. Multiplexed absolute quantification in proteomics using artificial qcat proteins of concatenated signature peptides. *Nature Methods* **2**, 587-589, doi:10.1038/nmeth774 (2005).

184. Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., Vandenesch, F. & Garin, J. Isotope-labeled protein standards: Toward absolute quantitative proteomics. *Molecular and Cellular Proteomics* **6**, 2139-2149, doi:10.1074/mcp.M700163-MCP200 (2007).

185. Hanke, S., Besir, H., Oesterhelt, D. & Mann, M. Absolute silac for accurate quantitation of proteins in complex mixtures down to the attomole level. *Journal of Proteome Research* **7**, 1118-1130, doi:10.1021/pr7007175 (2008).

186. Singh, S., Springer, M., Steen, J., Kirschner, M. W. & Steen, H. Flexiquant: A novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides. *Journal of Proteome Research* **8**, 2201-2210, doi:10.1021/pr800654s (2009).

187. Searle, B. C. Scaffold: A bioinformatic tool for validating ms/ms-based proteomic studies. *Proteomics* **10**, 1265-1269, doi:10.1002/pmic.200900437 (2010).

188. Sidoli, S., Lin, S., Xiong, L., Bhanu, N. V., Karch, K. R., Johansen, E., Hunter, C., Mollah, S. & Garcia, B. A. Sequential window acquisition of all theoretical mass spectra (swath) analysis for characterization and quantification of histone post-translational modifications. *Molecular and Cellular Proteomics* **14**, 2420-2428, doi:10.1074/mcp.O114.046102 (2015).

189. Magrane, M. & Consortium, U. Uniprot knowledgebase: A hub of integrated protein data. *Database* **2011**, doi:10.1093/database/bar009 (2011).

190. Kim, S. & Pevzner, P. A. Universal database search tool for proteomics. *Nature communications* **5**, 5277-5277, doi:10.1038/ncomms6277 (2014).

191. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **5**, 976-989, doi:10.1016/1044-0305(94)80016-2 (1994).

192. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V. & Mann, M. Andromeda: A peptide search engine integrated into the maxquant environment. *Journal of Proteome Research* **10**, 1794-1805, doi:10.1021/pr101065j (2011).

193. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: An open-source ms/ms sequence database search tool. *Proteomics* **13**, 22-24, doi:10.1002/pmic.201200439 (2013).

194. Dorfer, V., Pichler, P., Stranzl, T., Stadlmann, J., Taus, T., Winkler, S. & Mechtler, K. Ms amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *Journal of Proteome Research* **13**, 3679-3684, doi:10.1021/pr500202e (2014).

195. Tabb, D. L., Fernando, C. G. & Chambers, M. C. Myrimatch: Highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *Journal of Proteome Research* **6**, 654-661, doi:10.1021/pr0604054 (2007).

196. Wenger, C. D. & Coon, J. J. A proteomics search algorithm specifically designed for high-resolution tandem mass spectra. *Journal of Proteome Research* **12**, 1377-1386, doi:10.1021/pr301024c (2013).

197. Diament, B. & Noble, W. S. Faster sequest searching for peptide identification from tandem mass spectra. *Journal of Proteome Research* **10**, 3871-3879, doi:10.1021/pr101196n (2011).

198. Park, C. Y., Klammer, A. A., Kall, L., MacCoss, M. J. & Noble, W. S. Rapid and accurate peptide identification from tandem mass spectra. *Journal of Proteome Research* **7**, 3022-3027, doi:10.1021/pr800127y (2008).

199. Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W. & Bryant, S. H. Open mass spectrometry search algorithm. *Journal of Proteome Research* **3**, 958-964, doi:10.1021/pr0499491 (2004).

200. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567, doi:10.1002/(sici)1522-2683(19991201)20:18<3551::aid-elps3551>3.0.co;2-2 (1999).

201. Keller, A., Eng, J., Zhang, N., Li, X.-j. & Aebersold, R. A uniform proteomics ms/ms analysis platform utilizing open xml file formats. *Molecular Systems Biology* **1**, 2005.0017-2005.0017, doi:10.1038/msb4100024 (2005).

202. Deutsch, E. W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B., Eng, J. K., Martin, D. B., Nesvizhskii, A. & Aebersold, R. A guided tour of the trans-proteomic pipeline. *Proteomics* **10**, 1150-1159,

doi:10.1002/pmic.200900375 (2010).

203. Ma, B. & Johnson, R. De novo sequencing and homology searching. *Molecular and Cellular Proteomics* **11**, O111 014902, doi:10.1074/mcp.O111.014902 (2012).

204. Chong, K. F. & Leong, H. W. Tutorial on de novo peptide sequencing using ms/ms mass spectrometry. *Journal of Bioinformatics and Computational Biology* **10**, 1231002, doi:10.1142/S0219720012310026 (2012).

205. Sanders, W. S., Wang, N., Bridges, S. M., Malone, B. M., Dandass, Y. S., McCarthy, F. M., Nanduri, B., Lawrence, M. L. & Burgess, S. C. The proteogenomic mapping tool. *BMC Bioinformatics* **12**, 115-115, doi:10.1186/1471-2105-12-115 (2011).

206. Dalman, M. R., Deeter, A., Nimishakavi, G. & Duan, Z.-H. Fold change and p-value cutoffs significantly alter microarray interpretations. *BMC Bioinformatics* **13**, S11-S11, doi:10.1186/1471-2105-13-S2-S11 (2012).

207. Zhang, F. & Chen, J. Y. Discovery of pathway biomarkers from coupled proteomics and systems biology methods. *BMC Genomics* **11**, S12-S12, doi:10.1186/1471-2164-11-S2-S12 (2010).

208. Myers, J. S., von Lersner, A. K., Robbins, C. J. & Sang, Q.-X. A. Differentially expressed genes and signature pathways of human prostate cancer. *PloS One* **10**, e0145322, doi:10.1371/journal.pone.0145322 (2015).

209. Li, C., Tan, X. F., Lim, T. K., Lin, Q. & Gong, Z. Comprehensive and quantitative proteomic analyses of zebrafish plasma reveals conserved protein profiles between genders and between zebrafish and human. *Scientific Reports* **6**, 24329, doi:10.1038/srep24329 (2016).

210. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. G:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research* **35**, W193-W200, doi:10.1093/nar/gkm226 (2007).

211. Maere, S., Heymans, K. & Kuiper, M. Bingo: A cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448-3449, doi:10.1093/bioinformatics/bti551 (2005).

212. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

213. Bauer, S., Grossmann, S., Vingron, M. & Robinson, P. N. Ontologizer 2.0—a multifunctional tool for go term enrichment analysis and data exploration. *Bioinformatics* **24**, 1650-1651, doi:10.1093/bioinformatics/btn250 (2008).

214. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. Gorilla: A tool for discovery and visualization of enriched go terms in ranked gene lists. *BMC Bioinformatics* **10**, 48, doi:10.1186/1471-2105-10-48 (2009).

215. Huang, D. W., Sherman, B. T., Tan, Q., Collins, J. R., Alvord, W. G., Roayaei, J., Stephens, R., Baseler, M. W., Lane, H. C. & Lempicki, R. A. The david gene functional classification tool: A novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biology* **8**, R183-R183, doi:10.1186/gb-2007-8-9-r183 (2007).

216. Cai, X. W., Shedden, K. A., Yuan, S. H., Davis, M. A., Xu, L. Y., Xie, C. Y., Fu, X. L., Lawrence, T. S., Lubman, D. M. & Kong, F. M. Baseline plasma proteomic analysis to identify biomarkers that predict radiation-induced lung toxicity in patients receiving radiation for non-small cell lung cancer. *Journal of Thoracic Oncology* **6**, 1073-1078, doi:10.1097/JTO.0b013e3182152ba6 (2011).

217. Uriarte, S. M., Powell, D. W., Luerman, G. C., Merchant, M. L., Cummins, T. D., Jog, N. R., Ward, R. A. & McLeish, K. R. Comparison of proteins expressed on secretory vesicle membranes and plasma membranes of human neutrophils. *Journal of Immunology* **180**, 5575-5581 (2008).

218. van Noort, V., Snel, B. & Huynen, M. A. Predicting gene function by conserved co-expression. *Trends in Genetics* **19**, 238-242, doi:10.1016/s0168-9525(03)00056-8 (2003).

219. Krämer, A., Green, J., Pollard, J. & Tugendreich, S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* **30**, 523-530, doi:10.1093/bioinformatics/btt703 (2014).

220. Mi, H., Poudel, S., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Panther version 10: Expanded protein families and functions, and analysis tools. *Nucleic Acids Research* **44**, D336-342, doi:10.1093/nar/gkv1194 (2016).

221. Nesvizhskii, A. I., Vitek, O. & Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature Methods* **4**, 787-797, doi:10.1038/nmeth1088 (2007).

222. Humphries, J. M., Penno, M. A. S., Weiland, F., Klingler-Hoffmann, M., Zuber, A., Boussioutas, A., Ernst, M. & Hoffmann, P. Identification and validation of novel candidate protein biomarkers for the detection of human gastric cancer. *Biochimica et Biophysica Acta*

(BBA) - Proteins and Proteomics **1844**, 1051-1058, doi:https://doi.org/10.1016/j.bbapap.2014.01.018 (2014).

223. Vanderperre, B., Lucier, J. F., Bissonnette, C., Motard, J., Tremblay, G., Vanderperre, S., Wisztorski, M., Salzet, M., Boisvert, F. M. & Roucou, X. Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PloS One* **8**, e70698, doi:10.1371/journal.pone.0070698 (2013).

224. Blakeley, P., Siepen, J. A., Lawless, C. & Hubbard, S. J. Investigating protein isoforms via proteomics: A feasibility study. *Proteomics* **10**, 1127-1140, doi:10.1002/pmic.200900445 (2010).

225. Hudson, L. J., Bevan, S., Wotherspoon, G., Gentry, C., Fox, A. & Winter, J. Vr1 protein expression increases in undamaged drg neurons after partial nerve injury. *European Journal of Neuroscience* **13**, 2105-2114, doi:10.1046/j.0953-816x.2001.01591.x (2001).

226. Singleton, K. D., Serkova, N., Beckey, V. E. & Wischmeyer, P. E. Glutamine attenuates lung injury and improves survival after sepsis: Role of enhanced heat shock protein expression. *Critical Care Medicine* **33**, 1206-1213 (2005).

227. Dragunow, M., Beilharz, E., Sirimanne, E., Lawlor, P., Williams, C., Bravo, R. & Gluckman, P. Immediate-early gene protein expression in neurons undergoing delayed death, but not necrosis, following hypoxic-ischaemic injury to the young rat brain. *Molecular Brain Research* **25**, 19-33, doi:http://dx.doi.org/10.1016/0169-328X(94)90274-7 (1994).

228. Frey, E., Regenfelder, F., Sussmann, P., Zumstein, M., Gerber, C., Born, W. & Fuchs, B. Adipogenic and myogenic gene expression in rotator cuff muscle of the sheep after tendon tear. *Journal of Orthopaedic Research* **27**, 504-509, doi:10.1002/jor.20695 (2009).

229. Serrano, C., Bolea, R., Lyahyai, J., Filali, H., Varona, L., Marcos-Carcavilla, A., Acin, C., Calvo, J. H., Serrano, M., Badiola, J. J., Zaragoza, P. & Martin-Burriel, I. Changes in hsp gene and protein expression in natural scrapie with brain damage. *Veterinary Research* **42**, 13, doi:10.1186/1297-9716-42-13 (2011).

230. Jiang, Y., Zhang, W., Stanton, J.-A., Brauning, R., Barris, W. C., Hourlier, T., Aken, B. L., Searle, S. M. J., Adelson, D. L., Bian, C., Cam, G. R., Xie, M., Chen, Y., Cheng, S., DeSilva, U., Dixen, K., Dong, Y., Fan, G., Franklin, I. R., Fu, S. *et al.* The sheep genome illuminates biology of the rumen and lipid metabolism. *Science (New York, N.Y.)* **344**, 1168, doi:10.1126/science.1252806 (2014).

231. Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dianes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P. A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H. J., Albar, J. P., Martinez-Bartolome, S. *et al.* Proteomexchange provides globally coordinated proteomics data submission and dissemination. *Nature Biotechnology* **32**, 223-226, doi:10.1038/nbt.2839 (2014).

232. Jaffe, J. D., Berg, H. C. & Church, G. M. Proteogenomic mapping as a complementary method to perform genome annotation. *Proteomics* **4**, 59-77, doi:10.1002/pmic.200300511 (2004).

233. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nature Methods* **6**, 359-362, doi:10.1038/nmeth.1322 (2009).

234. National, Health, and, Medical, Research & Council. *Australian code of practice for the care and use of animals for scientific purposes*, https://www.nhmrc.gov.au/guidelines/publications/ea28 (2013).

235. Clifton, J., Huang, F., Rucevic, M., Cao, L., Hixson, D. & Josic, D. Protease inhibitors as possible pitfalls in proteomic analyses of complex biological samples. *Journal of Proteomics* **74**, 935-941, doi:10.1016/j.jprot.2011.02.010 (2011).

236. Schägger, H. & von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kda. *Analytical Biochemistry* **166**, 368-379, doi:10.1016/0003-2697(87)90587-2 (1987).

237. Brunelle, J. L. & Green, R. One-dimensional sds-polyacrylamide gel electrophoresis (1d sds-page). *Methods in Enzymology* **541**, 151-159, doi:10.1016/b978-0-12-420119-4.00012-4 (2014).

238. Herbert, B., Galvani, M., Hamdan, M., Olivieri, E., MacCarthy, J., Pedersen, S. & Righetti, P. G. Reduction of alkylation of proteins in preparation of two-dimensional map analysis: Why, when, and how? *Electrophoresis* **22**, 2046-2057, doi:10.1002/1522-2683(200106)22:10<2046::AID-ELPS2046>3.0.CO;2-C (2001).

239. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protocols* **1**, 2856-2860 (2007).

240. Yu, Y., Smith, M. & Pieper, R. A spinnable and automatable stagetip for high throughput peptide desalting and proteomics. (2014).

241. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using stagetips. *Nat. Protocols* **2**, 1896-1906 (2007).

242. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Meth* **11**, 319-324, doi:10.1038/nmeth.2834 (2014).

243. Trapp, J., Geffard, O., Imbert, G., Gaillard, J.-C., Davin, A.-H., Chaumot, A. & Armengaud, J. Proteogenomics of gammarus fossarum to document the reproductive system of amphipods. *Molecular & Cellular Proteomics : MCP* **13**, 3612-3625, doi:10.1074/mcp.M114.038851 (2014).

244. Sigdel, T. K. & Sarwal, M. M. The proteogenomic path towards biomarker discovery. *Pediatric Transplantation* **12**, 737-747, doi:10.1111/j.1399-3046.2008.01018.x (2008).

245. Vaudel, M., Burkhart, J. M., Zahedi, R. P., Oveland, E., Berven, F. S., Sickmann, A., Martens, L. & Barsnes, H. Peptideshaker enables reanalysis of ms-derived proteomics data sets. *Nature Biotechnology* **33**, 22-24, doi:10.1038/nbt.3109 (2015).

246. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for mass spectrometry-based proteomics. *Methods in Molecular Biology* **604**, 55-71, doi:10.1007/978-1-60761-444-9_5 (2010).

247. Vaudel, M., Breiter, D., Beck, F., Rahnenfuhrer, J., Martens, L. & Zahedi, R. P. D-score: A search engine independent md-score. *Proteomics* **13**, 1036-1041, doi:10.1002/pmic.201200408 (2013).

248. Taus, T., Kocher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C. & Mechtler, K. Universal and confident phosphorylation site localization using phosphors. *Journal of Proteome Research* **10**, 5354-5362, doi:10.1021/pr200611n (2011).

249. Barsnes, H., Vaudel, M., Colaert, N., Helsens, K., Sickmann, A., Berven, F. S. & Martens, L. Compomics-utilities: An open-source java library for computational proteomics. *BMC Bioinformatics* **12**, 70, doi:10.1186/1471-2105-12-70 (2011).

250. Venny. An interactive tool for comparing lists with venn's diagrams (2007-2015).

251. Bardou, P., Mariette, J., Escudie, F., Djemiel, C. & Klopp, C. Jvenn: An interactive venn diagram viewer. *BMC Bioinformatics* **15**, 293-293, doi:10.1186/1471-2105-15-293 (2014).

252. Hulsen, T., de Vlieg, J. & Alkema, W. Biovenn - a web application for the comparison and visualization of biological lists using area-proportional venn diagrams. *BMC Genomics* **9**,

488, doi:10.1186/1471-2164-9-488 (2008).

253. Pieper, R., Gatlin, C. L., Makusky, A. J., Russo, P. S., Schatz, C. R., Miller, S. S., Su, Q., McGrath, A. M., Estock, M. A., Parmar, P. P., Zhao, M., Huang, S. T., Zhou, J., Wang, F., Esquer-Blasco, R., Anderson, N. L., Taylor, J. & Steiner, S. The human serum proteome: Display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins. *Proteomics* **3**, 1345-1364, doi:10.1002/pmic.200300449 (2003).

254. Kijas, J. W., Menzies, M. & Ingham, A. Sequence diversity and rates of molecular evolution between sheep and cattle genes. *Animal Genetics* **37**, 171-174, doi:10.1111/j.1365-2052.2005.01399.x (2006).

255. Vanselow, J., Furbass, R., Rehbock, F., Klautschek, G. & Schwerin, M. Cattle and sheep use different promoters to direct the expression of the aromatase cytochrome p450 encoding gene, cyp19, during pregnancy. *Domestic Animal Endocrinology* **27**, 99-114, doi:10.1016/j.domaniend.2004.01.008 (2004).

256. Walters, E. M., Agca, Y., Ganjam, V. & Evans, T. Animal models got you puzzled?: Think pig. *Annals of the New York Academy of Sciences* **1245**, 63-64, doi:10.1111/j.1749-6632.2011.06345.x (2011).

257. Wang, S., Liu, Y., Fang, D. & Shi, S. The miniature pig: A useful large animal model for dental and orofacial research. *Oral Diseases* **13**, 530-537, doi:10.1111/j.1601-0825.2006.01337.x (2007).

258. Dehoux, J. P. & Gianello, P. The importance of large animal models in transplantation. *Frontiers in Bioscience* **12**, 4864-4880 (2007).

259. Casal, M. & Haskins, M. Large animal models and gene therapy. *European Journal of Human Genetics* **14**, 266-272 (2005).

260. Schaefer, A., Schneeberger, Y., Stenzig, J., Biermann, D., Jelinek, M., Reichenspurner, H., Eschenhagen, T., Ehmke, H. & Schwoerer, A. P. A new animal model for investigation of mechanical unloading in hypertrophic and failing hearts: Combination of transverse aortic constriction and heterotopic heart transplantation. *PloS One* **11**, e0148259, doi:10.1371/journal.pone.0148259 (2016).

261. Meling, S., Kvalheim, O. M., Arneberg, R., Bardsen, K., Hjelle, A. & Ulvund, M. J. Investigation of serum protein profiles in scrapie infected sheep by means of seldi-tof-ms and multivariate data analysis. *BMC Research Notes* **6**, 466, doi:10.1186/1756-0500-6-466

(2013).

262. Batxelli-Molina, I., Salvetat, N., Andreoletti, O., Guerrier, L., Vicat, G., Molina, F. & Mourton-Gilles, C. Ovine serum biomarkers of early and late phase scrapie. *BMC Veterinary Research* **6**, 49, doi:10.1186/1746-6148-6-49 (2010).

263. Zhong, L., Taylor, D., Begg, D. J. & Whittington, R. J. Biomarker discovery for ovine paratuberculosis (johne's disease) by proteomic serum profiling. *Comparative Immunology Microbiology and Infectious Diseases* **34**, 315-326, doi:10.1016/j.cimid.2011.03.001 (2011).

264. Meling, S., Kvalheim, O. M., Arneberg, R., Bardsen, K., Hjelle, A. & Ulvund, M. J. Investigation of serum protein profiles in scrapie infected sheep by means of seldi-tof-ms and multivariate data analysis. *BMC Research Notes* **6**, 466, doi:10.1186/1756-0500-6-466 (2013).

265. Bradshaw, R. A., Burlingame, A. L., Carr, S. & Aebersold, R. Reporting protein identification data: The next generation of guidelines. *Molecular & cellular proteomics : MCP* **5**, 787-788, doi:10.1074/mcp.E600005-MCP200 (2006).

266. Seymour, S. L., Farrah, T., Binz, P.-A., Chalkley, R. J., Cottrell, J. S., Searle, B. C., Tabb, D. L., Vizcaíno, J. A., Prieto, G., Uszkoreit, J., Eisenacher, M., Martínez-Bartolomé, S., Ghali, F. & Jones, A. R. A standardized framing for reporting protein identifications in mzidentml 1.2. *Proteomics* **14**, 2389-2399, doi:10.1002/pmic.201400080 (2014).

267. Omenn, G. S. Plasma proteomics, the human proteome project, and cancer-associated alternative splice variant proteins(,). *Biochimica et Biophysica Acta* **1844**, 866-873, doi:10.1016/j.bbapap.2013.10.016 (2014).

268. Clement, C. C., Aphkhazava, D., Nieves, E., Callaway, M., Olszewski, W., Rotzschke, O. & Santambrogio, L. Protein expression profiles of human lymph and plasma mapped by 2d-dige and 1d sds–page coupled with nanolc–esi–ms/ms bottom-up proteomics. *Journal of Proteomics* **78**, 172-187, doi:10.1016/j.jprot.2012.11.013 (2013).

269. Martens, L., Hermjakob, H., Jones, P., Adamski, M., Taylor, C., States, D., Gevaert, K., Vandekerckhove, J. & Apweiler, R. Pride: The proteomics identifications database. *Proteomics* **5**, 3537-3545, doi:10.1002/pmic.200401303 (2005).

270. Seymour, S. L. & Hunter, C. J. 1-5 (AB Sciex Pte. Ltd., USA, 2015).

271. Spivak, M., Weston, J., Tomazela, D., MacCoss, M. J. & Noble, W. S. Direct maximization of protein identifications from tandem mass spectra. *Molecular and Cellular Proteomics* **11**, M111 012161, doi:10.1074/mcp.M111.012161 (2012).

272. States, D. J., Omenn, G. S., Blackwell, T. W., Fermin, D., Eng, J., Speicher, D. W. & Hanash, S. M. Challenges in deriving high-confidence protein identifications from data gathered by a hupo plasma proteome collaborative study. *Nat Biotech* **24**, 333-338 (2006).

273. International Sheep Genomics, C., Archibald, A. L., Cockett, N. E., Dalrymple, B. P., Faraut, T., Kijas, J. W., Maddox, J. F., McEwan, J. C., Hutton Oddy, V., Raadsma, H. W., Wade, C., Wang, J., Wang, W. & Xun, X. The sheep genome reference sequence: A work in progress. *Animal Genetics* **41**, 449-453, doi:10.1111/j.1365-2052.2010.02100.x (2010).

274. Bannach, O., Birkmann, E., Reinartz, E., Jaeger, K.-E., Langeveld, J. P. M., Rohwer, R. G., Gregori, L., Terry, L. A., Willbold, D. & Riesner, D. Detection of prion protein particles in blood plasma of scrapie infected sheep. *PloS One* **7**, e36620, doi:10.1371/journal.pone.0036620 (2012).

275. Katunguka-Rwakishaya, E. Influence of trypanosoma congolense infection on some blood inorganic and protein constituents in sheep. *Revue d'Elevage et de Medecine Veterinaire des Pays Tropicaux* **49**, 311-314 (1996).

276. Meling, S., Bardsen, K. & Ulvund, M. J. Presence of an acute phase response in sheep with clinical classical scrapie. *BMC Veterinary Research* **8**, doi:10.1186/1746-6148-8-113 (2012).

277. Ersdal, C., Jørgensen, H. J. & Lie, K. I. Acute and chronic erysipelothrix rhusiopathiae infection in lambs. *Veterinary Pathology* **52**, 635-643, doi:10.1177/0300985814556187 (2014).

278. Aytekin, I., Aksit, H., Sait, A., Kaya, F., Aksit, D., Gokmen, M. & Baca, A. U. Evaluation of oxidative stress via total antioxidant status, sialic acid, malondialdehyde and rt-pcr findings in sheep affected with bluetongue. *Veterinary Record Open* **2**, e000054, doi:10.1136/vetreco-2014-000054 (2015).

279. Burgess, S. T. G., Nunn, F., Nath, M., Frew, D., Wells, B., Marr, E. J., Huntley, J. F., McNeilly, T. N. & Nisbet, A. J. A recombinant subunit vaccine for the control of ovine psoroptic mange (sheep scab). *Veterinary Research* **47**, 26, doi:10.1186/s13567-016-0315-3 (2016).

280. McDonald, C. I., Fung, Y. L., Shekar, K., Diab, S. D., Dunster, K. R., Passmore, M. R., Foley, S. R., Simonova, G., Platts, D. & Fraser, J. F. The impact of acute lung injury, ecmo and transfusion on oxidative stress and plasma selenium levels in an ovine model. *Journal of Trace Elements in Medicine and Biology* **30**, 4-10, doi:10.1016/j.jtemb.2015.01.004 (2015).

281. Zheng, X., Baker, H., Hancock, W. S., Fawaz, F., McCaman, M. & Pungor, E., Jr. Proteomic analysis for the assessment of different lots of fetal bovine serum as a raw material for cell culture. Part iv. Application of proteomics to the manufacture of biological drugs. *Biotechnology Progress* **22**, 1294-1300, doi:10.1021/bp060121o (2006).

282. Zhang, G., Fenyö, D. & Neubert, T. A. Use of DNA ladders for reproducible protein fractionation by sds-page for quantitative proteomics. *Journal of Proteome Research* **7**, 678-686, doi:10.1021/pr700601y (2008).

283. Gallagher, S. R. One-dimensional sds gel electrophoresis of proteins. *Current Protocols in Molecular Biology* **Chapter 10**, Unit 10 12A, doi:10.1002/0471142727.mb1002as75 (2006).

284. Deutsch, E. W. File formats commonly used in mass spectrometry proteomics. *Molecular & Cellular Proteomics : MCP* **11**, 1612-1621, doi:10.1074/mcp.R112.019695 (2012).

285. Zhang, H., Ghai, P., Wu, H., Wang, C., Field, J. & Zhou, G. L. Mammalian adenylyl cyclase-associated protein 1 (cap1) regulates cofilin function, the actin cytoskeleton, and cell adhesion. *Journal of Biological Chemistry* **288**, 20966-20977, doi:10.1074/jbc.M113.484535 (2013).

286. Herscovics, A. Structure and function of class i alpha 1,2-mannosidases involved in glycoprotein synthesis and endoplasmic reticulum quality control. *Biochimie* **83**, 757-762 (2001).

287. Stegelmeier, B. L., James, L. F., Panter, K. E., Gardner, D. R., Pfister, J. A., Ralphs, M. H. & Molyneux, R. J. Dose response of sheep poisoned with locoweed (oxytropis sericea). *Journal of Veterinary Diagnostic Investigation* **11**, 448-456 (1999).

288. Stegelmeier, B. L., James, L. F., Panter, K. E. & Molyneux, R. J. Tissue and serum swainsonine concentrations in sheep ingesting astragalus lentiginosus (locoweed). *Veterinary and Human Toxicology* **37**, 336-339 (1995).

289. Yu, P. H., Davis, B. A., Boulton, A. A. & Zuo, D. M. Deamination of aliphatic amines by type b monoamine oxidase and semicarbazide-sensitive amine oxidase; pharmacological implications. *Journal of Neural Transmission. Supplementum* **41**, 397-406 (1994).

290. Sousse, L. E., Yamamoto, Y., Enkhbaatar, P., Rehberg, S. W., Wells, S. M., Leonard, S., Traber, M. G., Yu, Y. M., Cox, R. A., Hawkins, H. K., Traber, L. D., Herndon, D. N. & Traber, D. L. Acute lung injury-induced collagen deposition is associated with elevated asymmetric dimethylarginine and arginase activity. *Shock* **35**, 282-288,

doi:10.1097/SHK.0b013e3181fddd82 (2011).

291. Sousse, L. E., Jonkam, C. C., Traber, D. L., Hawkins, H. K., Rehberg, S. W., Traber, L. D., Herndon, D. N. & Enkhbaatar, P. Pseudomonas aeruginosa is associated with increased lung cytokines and asymmetric dimethylarginine compared with methicillin-resistant staphylococcus aureus. *Shock* **36**, 466-470, doi:10.1097/SHK.0b013e3182336b45 (2011).

292. Alemu, P., Forsyth, G. W. & Searcy, G. P. A comparison of parameters used to assess liver damage in sheep treated with carbon tetrachloride. *Canadian Journal of Comparative Medicine* **41**, 420-427 (1977).

293. Katsoulos, P. D., Christodoulopoulos, G., Minas, A., Karatzia, M. A., Pourliotis, K. & Kritas, S. K. The role of lactate dehydrogenase, alkaline phosphatase and aspartate aminotransferase in the diagnosis of subclinical intramammary infections in dairy sheep and goats. *Journal of Dairy Research* **77**, 107-111, doi:10.1017/s0022029909990410 (2010).

294. Bodeker, D., Oppelland, G. & Holler, H. Involvement of carbonic anhydrase in ammonia flux across rumen mucosa in vitro. *Experimental Physiology* **77**, 517-519 (1992).

295. Leonhard-Marek, S., Gabel, G. & Martens, H. Effects of short chain fatty acids and carbon dioxide on magnesium transport across sheep rumen epithelium. *Experimental Physiology* **83**, 155-164 (1998).

296. Petrera, A., Kern, U., Linz, D., Gomez-Auli, A., Hohl, M., Gassenhuber, J., Sadowski, T. & Schilling, O. Proteomic profiling of cardiomyocyte-specific cathepsin a overexpression links cathepsin a to the oxidative stress response. *Journal of Proteome Research* **15**, 3188-3195, doi:10.1021/acs.jproteome.6b00413 (2016).

297. Kostadinov, S., Shah, B. A., Alroy, J. & Phornphutkul, C. A case of galactosialidosis with novel mutations of the protective protein/cathepsin a gene: Diagnosis prompted by trophoblast vacuolization on placental examination. *Pediatric and Developmental Pathology* **17**, 474-477, doi:10.2350/14-05-1500-cr.1 (2014).

298. Hofmann, F., Feil, R., Kleppisch, T. & Schlossmann, J. Function of cgmp-dependent protein kinases as revealed by gene deletion. *Physiological Reviews* **86**, 1-23, doi:10.1152/physrev.00015.2005 (2006).

299. Bohr, S., Patel, S. J., Vasko, R., Shen, K., Golberg, A., Berthiaume, F. & Yarmush, M. L. The role of chi3l1 (chitinase-3-like-1) in the pathogenesis of infections in burns in a mouse model. *PloS One* **10**, e0140440, doi:10.1371/journal.pone.0140440 (2015).

300. Habib, G. M., Shi, Z. Z., Cuevas, A. A. & Lieberman, M. W. Identification of two additional members of the membrane-bound dipeptidase family. *FASEB Journal* **17**, 1313-1315, doi:10.1096/fj.02-0899fje (2003).

301. Bebenek, K. & Kunkel, T. A. Functions of DNA polymerases. *Advances in Protein Chemistry* **69**, 137-165, doi:10.1016/s0065-3233(04)69005-x (2004).

302. Marcus, F., Rittenhouse, J., Gontero, B. & Harrsch, P. B. Function, structure and evolution of fructose-1,6-bisphosphatase. *Archivos de Biología y Medicina Experimentales* **20**, 371-378 (1987).

303. Visinoni, S., Khalid, N. F., Joannides, C. N., Shulkes, A., Yim, M., Whitehead, J., Tiganis, T., Lamont, B. J., Favaloro, J. M., Proietto, J., Andrikopoulos, S. & Fam, B. C. The role of liver fructose-1,6-bisphosphatase in regulating appetite and adiposity. *Diabetes* **61**, 1122-1132, doi:10.2337/db11-1511 (2012).

304. Woolliams, J. A., Wiener, G., Anderson, P. H. & McMurray, C. H. Variation in the activities of glutathione peroxidase and superoxide dismutase and in the concentration of copper in the blood in various breed crosses of sheep. *Research in Veterinary Science* **34**, 253-256 (1983).

305. Carlile, G. W., Chalmers-Redman, R. M. E., Tatton, N. A., Pong, A., Borden, K. L. B. & Tatton, W. G. Reduced apoptosis after nerve growth factor and serum withdrawal: Conversion of tetrameric glyceraldehyde-3-phosphate dehydrogenase to a dimer. *Molecular Pharmacology* **57**, 2-12 (2000).

306. Yugueros, J., Temprano, A., Berzal, B., Sánchez, M. a., Hernanz, C., Luengo, J. M. a. & Naharro, G. Glyceraldehyde-3-phosphate dehydrogenase-encoding gene as a useful taxonomic tool for staphylococcusspp. *Journal of Clinical Microbiology* **38**, 4351-4355 (2000).

307. Beavo, J. A. Cyclic nucleotide phosphodiesterases: Functional implications of multiple isoforms. *Physiological Reviews* **75**, 725-748 (1995).

308. Fritz, T. A., Hurley, J. H., Trinh, L. B., Shiloach, J. & Tabak, L. A. The beginnings of mucin biosynthesis: The crystal structure of udp-galnac:Polypeptide alpha-n-acetylgalactosaminyltransferase-t1. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15307-15312, doi:10.1073/pnas.0405657101 (2004).

309. Tenno, M., Kezdy, F. J., Elhammer, A. P. & Kurosaka, A. Function of the lectin domain of polypeptide n-acetylgalactosaminyltransferase 1. *Biochemical and Biophysical Research*

Communications **298**, 755-759 (2002).

310. Ramadoss, J., Liao, W. X., Morschauser, T. J., Lopez, G. E., Patankar, M. S., Chen, D. B. & Magness, R. R. Endothelial caveolar hub regulation of adenosine triphosphate-induced endothelial nitric oxide synthase subcellular partitioning and domain-specific phosphorylation. *Hypertension* **59**, 1052-1059, doi:10.1161/hypertensionaha.111.189498 (2012).

311. Filali, H., Martín-Burriel, I., Harders, F., Varona, L., Hedman, C., Mediano, D. R., Monzón, M., Bossers, A., Badiola, J. J. & Bolea, R. Gene expression profiling of mesenteric lymph nodes from sheep with natural scrapie. *BMC Genomics* **15**, 1-17, doi:10.1186/1471-2164-15-59 (2014).

312. Yamashita, H., Kotani, T., Park, J.-h., Murata, Y., Okazawa, H., Ohnishi, H., Ku, Y. & Matozaki, T. Role of the protein tyrosine phosphatase shp2 in homeostasis of the intestinal epithelium. *PloS One* **9**, e92904, doi:10.1371/journal.pone.0092904 (2014).

313. Gurung, R. B., Begg, D. J., Purdie, A. C., Bach, H. & Whittington, R. J. Immunoreactivity of protein tyrosine phosphatase a (ptpa) in sera from sheep infected with mycobacterium avium subspecies paratuberculosis. *Veterinary Immunology and Immunopathology* **160**, 129-132, doi:10.1016/j.vetimm.2014.03.013 (2014).

314. Steel, E. G., Witzel, D. A. & Blanks, A. Acquired coagulation factor x activity deficiency connected with hymenoxys odorata dc (compositae), bitterweed poisoning in sheep. *American Journal of Veterinary Research* **37**, 1383-1386 (1976).

315. Tillman, P., Carson, S. N. & Talken, L. Platelet function and coagulation parameters in sheep during experimental vascular surgery. *Laboratory Animal Science* **31**, 263-267 (1981).

316. Filippovich, I., Sorokina, N., St Pierre, L., Flight, S., de Jersey, J., Perry, N., Masci, P. P. & Lavin, M. F. Cloning and functional expression of venom prothrombin activator protease from pseudonaja textilis with whole blood procoagulant activity. *British Journal of Haematology* **131**, 237-246, doi:10.1111/j.1365-2141.2005.05744.x (2005).

317. Klee, L. & Zand, R. Probable epitopes: Relationships between myelin basic protein antigenic determinants and viral and bacterial proteins. *Neuroinformatics* **2**, 59-70, doi:10.1385/ni:2:1:059 (2004).

318. Bongaerts, G. P. & Vogels, G. D. Mechanism of uricase action. *Biochimica et Biophysica Acta* **567**, 295-308 (1979).

319. Rietema, S. E., Blackberry, M. A., Maloney, S. K., Martin, G. B., Hawken, P. A. & Blache, D. Twenty-four-hour profiles of metabolic and stress hormones in sheep selected for a calm or nervous temperament. *Domestic Animal Endocrinology* **53**, 78-87, doi:10.1016/j.domaniend.2015.05.005 (2015).

320. Almeida, A. M., Palhinhas, R. G., Kilminster, T., Scanlon, T., van Harten, S., Milton, J., Blache, D., Greeff, J., Oldham, C., Coelho, A. V. & Cardoso, L. A. The effect of weight loss on the muscle proteome in the damara, dorper and australian merino ovine breeds. *PloS One* **11**, e0146367, doi:10.1371/journal.pone.0146367 (2016).

321. Knight, T. W., Lindsay, D. R. & Oldham, C. M. Proceedings: The influence of rams on the fertility of the ewe. *Journal of Reproduction and Fertility* **43**, 377-378 (1975).

322. Brown, H. M., Fabre Nys, C., Cognie, J. & Scaramuzzi, R. J. Short oestrous cycles in sheep during anoestrus involve defects in progesterone biosynthesis and luteal neovascularisation. *Reproduction* **147**, 357-367, doi:10.1530/rep-13-0400 (2014).

323. Mitchell, L. M., King, M. E., Aitken, R. P., Gebbie, F. E. & Wallace, J. M. Ovulation, fertilization and lambing rates, and peripheral progesterone concentrations, in ewes inseminated at a natural oestrus during november or february. *Journal of Reproduction and Fertility* **115**, 133-140 (1999).

324. Freeman, W. M. & Hemby, S. E. Proteomics for protein expression profiling in neuroscience. *Neurochemical Research* **29**, 1065-1081 (2004).

325. Wu, W. W., Wang, G., Baek, S. J. & Shen, R.-F. Comparative study of three proteomic quantitative methods, dige, cicat, and itraq, using 2d gel- or lc−maldi tof/tof. *Journal of Proteome Research* **5**, 651-658, doi:10.1021/pr050405o (2006).

326. Gan, C. S., Reardon, K. F. & Wright, P. C. Comparison of protein and peptide prefractionation methods for the shotgun proteomic analysis of synechocystis sp. Pcc 6803. *Proteomics* **5**, 2468-2478, doi:10.1002/pmic.200401266 (2005).

327. Doucette, A. A., Tran, J. C., Wall, M. J. & Fitzsimmons, S. Intact proteome fractionation strategies compatible with mass spectrometry. *Expert Rev Proteomics* **8**, 787-800, doi:10.1586/epr.11.67 (2011).

328. Crowell, A. M., Wall, M. J. & Doucette, A. A. Maximizing recovery of water-soluble proteins through acetone precipitation. *Analytica Chimica Acta* **796**, 48-54, doi:10.1016/j.aca.2013.08.005 (2013).

329. Zellner, M., Winkler, W., Hayden, H., Diestinger, M., Eliasen, M., Gesslbauer, B., Miller, I., Chang, M., Kungl, A., Roth, E. & Oehler, R. Quantitative validation of different protein precipitation methods in proteome analysis of blood platelets. *Electrophoresis* **26**, 2481-2489, doi:10.1002/elps.200410262 (2005).

330. Fic, E., Kedracka-Krok, S., Jankowska, U., Pirog, A. & Dziedzicka-Wasylewska, M. Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis. *Electrophoresis* **31**, 3573-3579, doi:10.1002/elps.201000197 (2010).

331. Simpson, D. M. & Beynon, R. J. Acetone precipitation of proteins and the modification of peptides. *Journal of Proteome Research* **9**, 444-450, doi:10.1021/pr900806x (2010).

332. Polson, C., Sarkar, P., Incledon, B., Raguvaran, V. & Grant, R. Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography-tandem mass spectrometry. *Journal of Chromatography. B: Analytical Technologies in the Biomedical and Life Sciences* **785**, 263-275 (2003).

333. Zea-Ponce, Y. & Laruelle, M. Protein precipitation: An expedient procedure for the routine analysis of the plasma metabolites of [123i]ibzm. *Nuclear Medicine and Biology* **26**, 811-814 (1999).

334. Alshammari, T. M., Al-Hassan, A. A., Hadda, T. B. & Aljofan, M. Comparison of different serum sample extraction methods and their suitability for mass spectrometry analysis. *Saudi Pharmaceutical Journal* **23**, 689-697, doi:http://doi.org/10.1016/j.jsps.2015.01.023 (2015).

335. Hao, R., Adoligbe, C., Jiang, B., Zhao, X., Gui, L., Qu, K., Wu, S. & Zan, L. An optimized trichloroacetic acid/acetone precipitation method for two-dimensional gel electrophoresis analysis of qinchuan cattle longissimus dorsi muscle containing high proportion of marbling. *PloS One* **10**, e0124723, doi:10.1371/journal.pone.0124723 (2015).

336. Kay, R., Barton, C., Ratcliffe, L., Matharoo-Ball, B., Brown, P., Roberts, J., Teale, P. & Creaser, C. Enrichment of low molecular weight serum proteins using acetonitrile precipitation for mass spectrometry based proteomic analysis. *Rapid Communications in Mass Spectrometry* **22**, 3255-3260, doi:10.1002/rcm.3729 (2008).

337. Mostovenko, E., Scott, H. C., Klychnikov, O., Dalebout, H., Deelder, A. M. & Palmblad, M. Protein fractionation for quantitative plasma proteomics by semi-selective precipitation. *J Proteomics Bioinform* **5**, 217-221, doi:10.4172/jpb.1000239 (2012).

338. Bluemlein, K. & Ralser, M. Monitoring protein expression in whole-cell extracts by targeted label- and standard-free lc-ms/ms. *Nat. Protocols* **6**, 859-869. (2011).

339. Ha, M., Sabherwal, M., Duncan, E., Stevens, S., Stockwell, P., McConnell, M., Bekhit Ael, D. & Carne, A. In-depth characterization of sheep (ovis aries) milk whey proteome and comparison with cow (bos taurus). *PloS One* **10**, e0139774, doi:10.1371/journal.pone.0139774 (2015).

340. Heller, M., Michel, P. E., Morier, P., Crettaz, D., Wenz, C., Tissot, J. D., Reymond, F. & Rossier, J. S. Two-stage off-gel isoelectric focusing: Protein followed by peptide fractionation and application to proteome analysis of human plasma. *Electrophoresis* **26**, 1174-1188, doi:10.1002/elps.200410106 (2005).

341. Franchin, C., Pivato, M., Rattazzi, M., Arrigoni, G. & Millioni, R. Offgel fractionation of peptides: Where really is your sample? *Journal of Chromatography A*, doi:10.1016/j.chroma.2014.05.077 (2014).

342. Magdeldin, S., Elguoshy, A., Yoshida, Y., Hirao, Y., Xu, B., Zhang, Y., Yamamoto, K., Takimoto, H., Fujinaka, H., Kinoshita, N. & Yamamoto, T. Complementary protein and peptide offgel fractionation for high-throughput proteomic analysis. *Analytical Chemistry* **87**, 8481-8488, doi:10.1021/acs.analchem.5b01911 (2015).

343. Von Bargen, C., Brockmeyer, J. & Humpf, H.-U. Meat authentication: A new hplc-ms/ms based method for the fast and sensitive detection of horse and pork in highly processed food. *Journal of Agricultural and Food Chemistry* **62**, 9428, doi:10.1021/jf503468t (2014).

344. Maheswarappa, N. B., Rani, K. U., Kumar, Y. P., Kulkarni, V. V. & Rapole, S. Proteomic based approach for characterizing 4-hydroxy-2-nonenal induced oxidation of buffalo (bubalus bubalis) and goat (capra hircus) meat myoglobins. *Proteome Science* **14**, 18, doi:10.1186/s12953-016-0108-7 (2016).

345. Handel, M., Winkler, J., Hornlein, R. F., Northoff, H., Heeg, P. & Sell, S. Influence of acid-citrate-dextrose anticoagulant on blood quality in retransfusion systems after total knee arthroplasty. *Archives of Orthopaedic and Trauma Surgery* **122**, 269-273, doi:10.1007/s00402-001-0375-2 (2002).

346. Gibson, J. G., Gregory, C. B. & Button, L. N. Citrate-phosphate-dextrose solution for preservation of human blood: A further report. *Transfusion* **1**, 280-287, doi:10.1111/j.1537-2995.1961.tb00059.x (1961).

347. Murphy, S. C., Samuel, B. U., Harrison, T., Speicher, K. D., Speicher, D. W., Reid, M. E., Prohaska, R., Low, P. S., Tanner, M. J., Mohandas, N. & Haldar, K. Erythrocyte detergent-resistant membrane proteins: Their characterization and selective uptake during malarial

infection. *Blood* **103**, 1920-1928, doi:10.1182/blood-2003-09-3165 (2004).

348. Ma, Z.-Q., Dasari, S., Chambers, M. C., Litton, M. D., Sobecki, S. M., Zimmerman, L. J., Halvey, P. J., Schilling, B., Drake, P. M., Gibson, B. W. & Tabb, D. L. Idpicker 2.0: Improved protein assembly with high discrimination peptide identification filtering. *Journal of Proteome Research* **8**, 3872-3881, doi:10.1021/pr900360j (2009).

349. Shin, J., Kim, G., Kabir, M. H., Park, S. J., Lee, S. T. & Lee, C. Use of composite protein database including search result sequences for mass spectrometric analysis of cell secretome. *PloS One* **10**, e0121692, doi:10.1371/journal.pone.0121692 (2015).

350. Luscombe, N. M., Greenbaum, D. & Gerstein, M. What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine* **40**, 346-358 (2001).

351. Lam, H. & Aebersold, R. Building and searching tandem mass (ms/ms) spectral libraries for peptide identification in proteomics. *Methods* **54**, 424-431, doi:10.1016/j.ymeth.2011.01.007 (2011).

352. Wang, J., Pérez-Santiago, J., Katz, J. E., Mallick, P. & Bandeira, N. Peptide identification from mixture tandem mass spectra. *Molecular & Cellular Proteomics : MCP* **9**, 1476-1485, doi:10.1074/mcp.M000136-MCP201 (2010).

353. Chick, J. M., Kolippakkam, D., Nusinow, D. P., Zhai, B., Rad, R., Huttlin, E. L. & Gygi, S. P. An ultra-tolerant database search reveals that a myriad of modified peptides contributes to unassigned spectra in shotgun proteomics. *Nature Biotechnology* **33**, 743-749, doi:10.1038/nbt.3267 (2015).

354. Lam, H., Deutsch, E. W., Eddes, J. S., Eng, J. K., Stein, S. E. & Aebersold, R. Building consensus spectral libraries for peptide identification in proteomics. *Nat Meth* **5**, 873-875, (2008).

355. Dinarello, C. A. Proinflammatory cytokines. *Chest* **118**, 503-508 (2000).

356. Hewel, J. A., Liu, J., Onishi, K., Fong, V., Chandran, S., Olsen, J. B., Pogoutse, O., Schutkowski, M., Wenschuh, H., Winkler, D. F., Eckler, L., Zandstra, P. W. & Emili, A. Synthetic peptide arrays for pathway-level protein monitoring by liquid chromatography-tandem mass spectrometry. *Molecular and Cellular Proteomics* **9**, 2460-2473, doi:10.1074/mcp.M900456-MCP200 (2010).

357. Sawyere, D. M., Lanz, O. I., Dahlgren, L. A., Barry, S. L., Nichols, A. C. & Werre, S. R. Cytokine and growth factor concentrations in canine autologous conditioned serum.

Veterinary Surgery **45**, 582-586, doi:10.1111/vsu.12506 (2016).

358. Goddard, A., Leisewitz, A. L., Kjelgaard-Hansen, M., Kristensen, A. T. & Schoeman, J. P. Excessive pro-inflammatory serum cytokine concentrations in virulent canine babesiosis. *PloS One* **11**, e0150113, doi:10.1371/journal.pone.0150113 (2016).

359. Gruen, M. E., Messenger, K. M., Thomson, A. E., Griffith, E. H., Paradise, H., Vaden, S. & Lascelles, B. D. A comparison of serum and plasma cytokine values using a multiplexed assay in cats. *Veterinary Immunology and Immunopathology* **182**, 69-73, doi:10.1016/j.vetimm.2016.10.003 (2016).

360. Koets, A. P., de Schwartz, N., Tooten, P., Kankofer, M., Broekhuijsen-Davies, J. M., Rutten, V. P., van Leengoed, L. A., Taverne, M. A. & Gruys, E. Release of proinflammatory cytokines related to luteolysis and the periparturient acute phase response in prostaglandin-induced parturition in cows. *Theriogenology* **49**, 797-812, doi:10.1016/s0093-691x(98)00029-6 (1998).

361. Kasimanickam, R. K., Kasimanickam, V. R., Olsen, J. R., Jeffress, E. J., Moore, D. A. & Kastelic, J. P. Associations among serum pro- and anti-inflammatory cytokines, metabolic mediators, body condition, and uterine disease in postpartum dairy cows. *Reproductive Biology and Endocrinology : RB&E* **11**, 103-103, doi:10.1186/1477-7827-11-103 (2013).

362. Yarim, G. F., Karahan, S. & Nisbet, C. Elevated plasma levels of interleukin 1 beta, tumour necrosis factor alpha and monocyte chemotactic protein 1 are associated with pregnancy toxaemia in ewes. *Veterinary Research Communications* **31**, 565-573, doi:10.1007/s11259-007-3551-1 (2007).

363. Vaudel, M., Burkhart, J. M., Zahedi, R. P., Oveland, E., Berven, F. S., Sickmann, A., Martens, L. & Barsnes, H. Peptideshaker enables reanalysis of ms-derived proteomics data sets. *Nature Biotechnology* **33** (2015).

364. Roh, S. G., Suzuki, Y., Gotoh, T., Tatsumi, R. & Katoh, K. Physiological roles of adipokines, hepatokines, and myokines in ruminants. *Asian-Australasian Journal of Animal Sciences* **29**, 1-15, doi:10.5713/ajas.16.0001R (2016).

365. van Miert, A. S. Pro-inflammatory cytokines in a ruminant model: Pathophysiological, pharmacological, and therapeutic aspects. *Veterinary Quarterly* **17**, 41-50, doi:10.1080/01652176.1995.9694530 (1995).

366. Pepin, M., Seow, H. F., Corner, L., Rothel, J. S., Hodgson, A. L. & Wood, P. R. Cytokine gene expression in sheep following experimental infection with various strains of

corynebacterium pseudotuberculosis differing in virulence. *Veterinary Research* **28**, 149-163 (1997).

367. Sloane, P. J., Elsasser, T. H., Spath, J. A., Jr., Albertine, K. H. & Gee, M. H. Plasma tumor necrosis factor-alpha during long-term endotoxemia in awake sheep. *J Appl Physiol (1985)* **73**, 1831-1837 (1992).

368. Carr, S. A., Abbatiello, S. E., Ackermann, B. L., Borchers, C., Domon, B., Deutsch, E. W., Grant, R. P., Hoofnagle, A. N., Hüttenhain, R., Koomen, J. M., Liebler, D. C., Liu, T., MacLean, B., Mani, D. R., Mansfield, E., Neubert, H., Paulovich, A. G., Reiter, L., Vitek, O., Aebersold, R. *et al.* Targeted peptide measurements in biology and medicine: Best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Molecular & Cellular Proteomics : MCP* **13**, 907-917, doi:10.1074/mcp.M113.036095 (2014).

369. Schilling, B., MacLean, B., Held, J. M., Sahu, A. K., Rardin, M. J., Sorensen, D. J., Peters, T., Wolfe, A. J., Hunter, C. L., MacCoss, M. J. & Gibson, B. W. Multiplexed, scheduled, high-resolution parallel reaction monitoring on a full scan qqtof instrument with integrated data-dependent and targeted mass spectrometric workflows. *Analytical Chemistry* **87**, 10222-10229, doi:10.1021/acs.analchem.5b02983 (2015).

370. Koch, J. M., Ramadoss, J. & Magness, R. R. Proteomic profile of uterine luminal fluid from early pregnant ewes. *Journal of Proteome Research* **9**, 3878-3885, doi:10.1021/pr100096b (2010).

371. Audain, E., Uszkoreit, J., Sachsenberg, T., Pfeuffer, J., Liang, X., Hermjakob, H., Sanchez, A., Eisenacher, M., Reinert, K., Tabb, D. L., Kohlbacher, O. & Perez-Riverol, Y. In-depth analysis of protein inference algorithms using multiple search engines and well-defined metrics. *Journal of Proteomics* **150**, 170-182, doi:10.1016/j.jprot.2016.08.002 (2017).

372. Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J. P., St-Denis, N. A., Li, T., Miteva, Y. V., Hauri, S., Sardiu, M. E., Low, T. Y., Halim, V. A., Bagshaw, R. D., Hubner, N. C., Al-Hakim, A., Bouchard, A., Faubert, D., Fermin, D., Dunham, W. H., Goudreault, M., Lin, Z. Y. *et al.* The crapome: A contaminant repository for affinity purification-mass spectrometry data. *Nature Methods* **10**, 730-736, doi:10.1038/nmeth.2557 (2013).

373. Gundry, R. L., White, M. Y., Murray, C. I., Kane, L. A., Fu, Q., Stanley, B. A. & Van Eyk, J. E. Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **CHAPTER 10**, Unit10.25-Unit10.25,

doi:10.1002/0471142727.mb1025s88 (2009).

374. McInnes, C. J. Current research on ovine cytokines. *British Veterinary Journal* **149**, 371-386, doi:10.1016/S0007-1935(05)80076-4 (1993).

375. Budhia, S., Haring, L. F., McConnell, I. & Blacklaws, B. A. Quantitation of ovine cytokine mrna by real-time rt-pcr. *Journal of Immunological Methods* **309**, 160-172, doi:10.1016/j.jim.2005.12.006 (2006).

376. Pinnapureddy, A. R., Stayner, C., McEwan, J., Baddeley, O., Forman, J. & Eccles, M. R. Large animal models of rare genetic disorders: Sheep as phenotypically relevant models of human genetic disease. *Orphanet Journal of Rare Diseases* **10**, 107, doi:10.1186/s13023-015-0327-5 (2015).

377. Traber, D. L. Animal models: The sheep. *Critical Care Medicine* **28**, 591-592 (2000).

378. Ceciliani, F., Ceron, J. J., Eckersall, P. D. & Sauerwein, H. Acute phase proteins in ruminants. *Journal of Proteomics* **75**, 4207-4231, doi:10.1016/j.jprot.2012.04.004 (2012).

379. Alonso-Fauste, I., Andres, M., Iturralde, M., Lampreave, F., Gallart, J. & Alava, M. A. Proteomic characterization by 2-de in bovine serum and whey from healthy and mastitis affected farm animals. *Journal of Proteomics* **75**, 3015-3030, doi:10.1016/j.jprot.2011.11.035 (2012).

380. Binns, S. H., Green, L. E. & Bailey, M. Development and validation of an elisa to detect antibodies to corynebacterium pseudotuberculosis in ovine sera. *Veterinary Microbiology* **123**, 169-179, doi:10.1016/j.vetmic.2007.02.015 (2007).

381. Saini, P. K., Riaz, M., Webert, D. W., Eckersall, P. D., Young, C. R., Stanker, L. H., Chakrabarti, E. & Judkins, J. C. Development of a simple enzyme immunoassay for blood haptoglobin concentration in cattle and its application in improving food safety. *American Journal of Veterinary Research* **59**, 1101-1107 (1998).

382. Grange, R. D., Thompson, J. P. & Lambert, D. G. Radioimmunoassay, enzyme and non-enzyme-based immunoassays. *British Journal of Anaesthesia* **112**, 213-216, doi:10.1093/bja/aet293 (2014).

383. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R. & Aebersold, R. Targeted data extraction of the ms/ms spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics : MCP* **11**, O111.016717, doi:10.1074/mcp.O111.016717 (2012).

384. Hopfgartner, G., Tonoli, D. & Varesio, E. High-resolution mass spectrometry for integrated qualitative and quantitative analysis of pharmaceuticals in biological matrices. *Analytical and Bioanalytical Chemistry* **402**, 2587-2596, doi:10.1007/s00216-011-5641-8 (2012).

385. Tung, J.-P. *Development of an ovine model of transfusion-related acute lung injury (trali)* Dissertation/Thesis thesis, University of Queensland, (2012).

386. Gregori, J., Villarreal, L., Mendez, O., Sanchez, A., Baselga, J. & Villanueva, J. Batch effects correction improves the sensitivity of significance tests in spectral counting-based comparative discovery proteomics. *Journal of Proteomics* **75**, 3938-3951, doi:10.1016/j.jprot.2012.05.005 (2012).

387. Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K. & Irizarry, R. A. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics* **11**, 733-739, doi:10.1038/nrg2825 (2010).

388. Holewinski, R. J., Parker, S. J., Matlock, A. D., Venkatraman, V. & Van Eyk, J. E. in *Quantitative proteomics by mass spectrometry* (ed Salvatore Sechi) 265-279 (Springer New York, 2016).

389. Kultima, K., Nilsson, A., Scholz, B., Rossbach, U. L., Fälth, M. & Andrén, P. E. Development and evaluation of normalization methods for label-free relative quantification of endogenous peptides. *Molecular & Cellular Proteomics : MCP* **8**, 2285-2295, doi:10.1074/mcp.M800514-MCP200 (2009).

390. Cray, C., Zaias, J. & Altman, N. H. Acute phase response in animals: A review. *Comparative Medicine* **59**, 517-526 (2009).

391. Jain, S., Gautam, V. & Naseem, S. Acute-phase proteins: As diagnostic tool. *Journal of Pharmacy & Bioallied Sciences* **3**, 118-127, doi:10.4103/0975-7406.76489 (2011).

392. Dayer, E., Dayer, J. M. & Roux-Lombard, P. Primer: The practical use of biological markers of rheumatic and systemic inflammatory diseases. *Nature Clinical Practice: Rheumatology* **3**, 512-520, doi:10.1038/ncprheum0572 (2007).

393. Vincent, J.-L., Dubois, M.-J., Navickis, R. J. & Wilkes, M. M. Hypoalbuminemia in acute illness: Is there a rationale for intervention?: A meta-analysis of cohort studies and controlled trials. *Annals of Surgery* **237**, 319-334, doi:10.1097/01.SLA.0000055547.93484.87 (2003).

394. Lohuis, J. A., Verheijden, J. H., Burvenich, C. & van Miert, A. S. Pathophysiological effects of endotoxins in ruminants. 1. Changes in body temperature and reticulo-rumen motility, and the effect of repeated administration. *Veterinary Quarterly* **10**, 109-116, doi:10.1080/01652176.1988.9694157 (1988).

395. Salvesen, Ø., Reiten, M. R., Heegaard, P. M. H., Tranulis, M. A., Espenes, A., Skovgaard, K. & Ersdal, C. Activation of innate immune genes in caprine blood leukocytes after systemic endotoxin challenge. *BMC Veterinary Research* **12**, 241, doi:10.1186/s12917-016-0870-x (2016).

396. Barr, J. in *Journal of Veterinary Emergency and Critical Care.* 1 (Wiley Subscription Services, Inc).

397. Winter, P., Fuchs, K., Walshe, K. & Colditz, I. G. Serum amyloid a in the serum and milk of ewes with mastitis induced experimentally with staphylococcus epidermidis. *Veterinary Record* **152**, 558-562 (2003).

398. Wells, B., Innocent, G. T., Eckersall, P. D., McCulloch, E., Nisbet, A. J. & Burgess, S. T. Two major ruminant acute phase proteins, haptoglobin and serum amyloid a, as serum biomarkers during active sheep scab infestation. *Veterinary Research* **44**, 103, doi:10.1186/1297-9716-44-103 (2013).

399. Liu, Y., Hüttenhain, R., Collins, B. & Aebersold, R. Mass spectrometric protein maps for biomarker discovery and clinical research. *Expert Review of Molecular Diagnostics* **13**, 811-825, doi:10.1586/14737159.2013.845089 (2013).

400. Konstantinidis, K. T., Serres, M. H., Romine, M. F., Rodrigues, J. L. M., Auchtung, J., McCue, L.-A., Lipton, M. S., Obraztsova, A., Giometti, C. S., Nealson, K. H., Fredrickson, J. K. & Tiedje, J. M. Comparative systems biology across an evolutionary gradient within the shewanella genus. *Proceedings of the National Academy of Sciences* **106**, 15909-15914, doi:10.1073/pnas.0902000106 (2009).

401. Allen, E. E. & Banfield, J. F. Community genomics in microbial ecology and evolution. *Nat Rev Micro* **3**, 489-498 (2005).

402. Luque-Garcia, J. L. & Neubert, T. A. Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry. *Journal of Chromatography A* **1153**, 259-276, doi:10.1016/j.chroma.2006.11.054 (2007).

403. Rai, A. J., Gelfand, C. A., Haywood, B. C., Warunek, D. J., Yi, J., Schuchard, M. D., Mehigh, R. J., Cockrill, S. L., Scott, G. B., Tammen, H., Schulz-Knappe, P., Speicher, D.

W., Vitzthum, F., Haab, B. B., Siest, G. & Chan, D. W. Hupo plasma proteome project specimen collection and handling: Towards the standardization of parameters for plasma proteome samples. *Proteomics* **5**, 3262-3277, doi:10.1002/pmic.200401245 (2005).

404. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Analytical Chemistry* **68**, 850-858, doi:10.1021/ac950914h (1996).

405. Shevchenko, A. & Shevchenko, A. Evaluation of the efficiency of in-gel digestion of proteins by peptide isotopic labeling and maldi mass spectrometry. *Analytical Biochemistry* **296**, 279-283, doi:10.1006/abio.2001.5321 (2001).

406. Crawford, A. S., Schuchard, M. D., Melm, C. D., Chapman, H. A., Wildsmith, J., Ray, K. B., Mehigh, R. J., Chen, D. E. & Scott, G. B. I. Novel proteoprep® 20 immunoaffinity depletion resin for human plasma. (Sigma-Aldrich, St. Louis, Missouri, USA, 2006).

407. Fic, E., Kedracka-Krok, S., Jankowska, U., Pirog, A. & Dziedzicka-Wasylewska, M. Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis. *Electrophoresis* **31**, 3573-3579, doi:10.1002/elps.201000197 (2010).

408. Chertov, O., Biragyn, A., Kwak, L. W., Simpson, J. T., Boronina, T., Hoang, V. M., Prieto, D. A., Conrads, T. P., Veenstra, T. D. & Fisher, R. J. Organic solvent extraction of proteins and peptides from serum as an effective sample preparation for detection and identification of biomarkers by mass spectrometry. *Proteomics* **4**, 1195-1203, doi:10.1002/pmic.200300677 (2004).

409. Chertov, O., Simpson, J. T., Biragyn, A., Conrads, T. P., Veenstra, T. D. & Fisher, R. J. Enrichment of low-molecular-weight proteins from biofluids for biomarker discovery. *Expert Rev Proteomics* **2**, 139-145, doi:10.1586/14789450.2.1.139 (2005).

410. Switzar, L., Giera, M. & Niessen, W. M. A. Protein digestion: An overview of the available techniques and recent developments. *Journal of Proteome Research* **12**, 1067-1077, doi:10.1021/pr301201x (2013).

411. Beseme, O., Fertin, M., Drobecq, H., Amouyel, P. & Pinet, F. Combinatorial peptide ligand library plasma treatment: Advantages for accessing low-abundance proteins. *Electrophoresis* **31**, 2697-2704, doi:10.1002/elps.201000188 (2010).

412. You, Q., Verschoor, C. P., Pant, S. D., Macri, J., Kirby, G. M. & Karrow, N. A. Proteomic analysis of plasma from holstein cows testing positive for mycobacterium avium subsp. Paratuberculosis (map). *Veterinary Immunology and Immunopathology* **148**, 243-251,

doi:10.1016/j.vetimm.2012.05.002 (2012).

413. von Toerne, C., Kahle, M., Schafer, A., Ispiryan, R., Blindert, M., Hrabe De Angelis, M., Neschen, S., Ueffing, M. & Hauck, S. M. Apoe, mbl2, and psp plasma protein levels correlate with diabetic phenotype in nzo mice--an optimized rapid workflow for srm-based quantification. *Journal of Proteome Research* **12**, 1331-1343, doi:10.1021/pr3009836 (2013).

414. Marrocco, C., Rinalducci, S., Mohamadkhani, A., D'Amici, G. M. & Zolla, L. Plasma gelsolin protein: A candidate biomarker for hepatitis b-associated liver cirrhosis identified by proteomic approach. *Blood Transfusion. Trasfusione del Sangue* **8 Suppl 3**, s105-112, doi:10.2450/2010.017s (2010).

415. Ferrín, G., Rodríguez-Perálvarez, M., Aguilar-Melero, P., Ranchal, I., Llamoza, C., Linares, C. I., González-Rubio, S., Muntané, J., Briceño, J., López-Cillero, P., Montero-Álvarez, J. L. & de la Mata, M. Plasma protein biomarkers of hepatocellular carcinoma in hcv-infected alcoholic patients with cirrhosis. *PloS One* **10**, e0118527, doi:10.1371/journal.pone.0118527 (2015).

416. Chen, L., Zhai, L., Li, Y., Li, N., Zhang, C., Ping, L., Chang, L., Wu, J., Li, X., Shi, D. & Xu, P. Development of gel-filter method for high enrichment of low-molecular weight proteins from serum. *PloS One* **10**, e0115862, doi:10.1371/journal.pone.0115862 (2015).

417. Ruis-Gonzalez, M. D., Canete, M. D., Gomez-Chaparro, J. L., Abril, N., Canete, R. & Lopez-Barea, J. Alterations of protein expression in serum of infants with intrauterine growth restriction and different gestational ages. *Journal of Proteomics* **119**, 169-182, doi:10.1016/j.jprot.2015.02.003 (2015).

418. Cao, Z., Tang, H.-Y., Wang, H., Liu, Q. & Speicher, D. W. Systematic comparison of fractionation methods for in-depth analysis of plasma proteomes. *Journal of Proteome Research* **11**, 3090-3100, doi:10.1021/pr201068b (2012).

419. Sikdar, S., Gill, R. & Datta, S. Improving protein identification from tandem mass spectrometry data by one-step methods and integrating data from other platforms. *Brief Bioinform* **17**, 262-269, doi:10.1093/bib/bbv043 (2016).

420. Kall, L., Storey, J. D., MacCoss, M. J. & Noble, W. S. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *Journal of Proteome Research* **7**, 29-34, doi:10.1021/pr700600n (2008).

421. Huang, T., Wang, J., Yu, W. & He, Z. Protein inference: A review. *Brief Bioinform* **13**, 586-614, doi:10.1093/bib/bbs004 (2012).

422. Thomas, F. C., Waterston, M., Hastie, P., Parkin, T., Haining, H. & Eckersall, P. D. The major acute phase proteins of bovine milk in a commercial dairy herd. *BMC Veterinary Research* **11**, 1-10, doi:10.1186/s12917-015-0533-3 (2015).

423. Bhasme, P. C., Kurjogi, M. M., Sanakal, R. D., Kaliwal, R. B. & Kaliwal, B. B. In silico characterization of putative drug targets in staphylococcus saprophyticus, causing bovine mastitis. *Bioinformation* **9**, 339-344, doi:10.6026/97320630009339 (2013).

424. Ulutas, P. A. & Ozpinar, A. Effect of mannheimia (pasteurella) haemolytica infection on acute-phase proteins and some mineral levels in colostrum-breast milk-fed or colostrum-breast milk-deprived sheep. *Veterinary Research Communications* **30**, 485-495, doi:10.1007/s11259-006-3246-z (2006).

425. Shteynberg, D., Nesvizhskii, A. I., Moritz, R. L. & Deutsch, E. W. Combining results of multiple search engines in proteomics. *Molecular & Cellular Proteomics : MCP* **12**, 2383-2393, doi:10.1074/mcp.R113.027797 (2013).

426. Merrifield, R. B. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *Journal of the American Chemical Society* **85**, 2149-2154, doi:10.1021/ja00897a025 (1963).

Appendices

APPENDIX 4.0. The UniProtKB accession numbers of 267 proteins from a ProteinPilot search of a composite database search of *Bos taurus, Ovis aries and Capra hircus* database of combined first, second and third in-gel digestion samples.

P14639, W5Q7J0, W5PF65, Q2UVX4, W5NSA6, W5P6F4, W5NUX8, W5P4S5, D6PZY4, W5PH95, W5NY95, W5QDG8, W5P3R3, W5NX51, W5PTG9, W5NQW4, W5NRG7, W5P627, W5PZS7, W5NPK5, W5QAB1, W5P0Q4, C8BKD1, W5PTR4, W5NXW9, W5PW21, W5Q124, W5P5I3, W5NSH2, W5P1J8, W5PGT6, W5P8R7, W5PH81, W5PBY2, Q1RMN8, W5QH50, W5PFC9, W5QH46, W5PHP7, W5NY46, W5QH56, W5QH54, W5NTW3, E1BH06, E1BKT9, W5NWM2, W5PXX3, W5P101, W5PID9, W5Q0X5, W5NYJ9, W5PXU6, W5PXI3, W5PJZ1, W5PDR5, W5Q268, W5NRI1, Q8SPJ1, W5PI61, W5Q9A2, I6W7A2, W5P812, W5PKA9, W5P338, W5QI29, W5P3J3, W5Q750, W5Q4Q3, Q3LRQ1, W5PHI7, W5PXC8, W5PE53, W5PGT9, W5NRR7, W5PZI1, Q29437, W5P7S6, W5PDE5, W5PD71, W5QGG0, W5PEI4, G3N0V0, W5Q0L2, Q1KYZ7, W5PHP8, P12303, W5QFP0, W5QGP4, Q06AV9, G3MZE0, W5P4C6, B5B304, W5NXP3, Q28745, W5NRH2, W5Q0R1, W5PFJ0, Q2KJF1, Q28161, W5PPQ8, W5PYG2, W5QAA3, W5QGD1, A2P2H9, G5E5V0, Q32PJ2, A6QP30, F1MIW8, V6F9A2, W5P2Q8, W5PDS4, W5Q2E1, W5QGB5, W5P229, Q32S29, F1N514, P04272, A2I7N1, Q09TE3, W5PLL2, A8YXZ2, I3WAE6, A6QPP2, W5NYF4, Q3SWW8, W5NX96, W5Q3K6, W5PD84, W5Q961, W5PDQ9, A6QM09, W5QAR2, E1BFN5, F1MLW8, W5QBW5, F1MLP3, W5PDP6, G8JL00, A4IFI0, W5Q0V2, W5P880, W5NXJ3, W5PVL4, W5PAB5, X4ZFS1, W5Q5A6, F1MH40, E1BNR0, W5Q9D5, Q29439, H6WVW6, W5QE21, W5PJ97, W5QA54, W5PTU7, D6PX64, W5PIJ5, A7MAZ5, W5NZ47, W5PDN1, Q0P5K6, W5PGS4, W5P3H8, W5PSM6, E1BI98, W5Q3R0, F1MCF8, W5PZH5, E1BB91, W5NVM6, W5QH21, O46544, E1BF59, W5QAA1, W5PG63, Q9XT27, A6NBZ0, G5E5V1, P17690, W5NUW3, W5Q4Z3, W5PVH9, G3MXG6, W5Q038, Q9MZS8, W5PJ66, W5PB04, W5Q517, W5PD62, W5P2V6, G8JKZ8, W5PB46, W5NS93, G3MWI3, W5PMY0, W5PW62, A5D9D2, W5QIK8, F1N3M0, W5PZM1, W5PTL2, W5PB07, W5Q5H8, W5P1X9, W5QFL0, P00974, W5NPN4, E1BBS9, E1BD43, W5NUU8, W5PC09, W5PIC9, W5NVT0, W5PK04, W5NQ91, Q29RQ1, W5NTD9, I7CT57, W5NQP5, W5NZQ2, W5QCX2, E1B9D7, Q2KI85, G3X6N3, F2YQ13, Q28085, P34955, F1MMP5, E1B726, Q95121, Q3ZEJ6, G3X6I0, W5QFH1, W5Q5N5, Q6LBN7, A5YBU8, A5PJT7, Q5GN72, Q05B55, W5PAL4, W5NWU4, E1BC09, E1B9F8, W5QJ27, F6RAG5, W5PQH2, W5QID7, W5Q5T2, A5D7A2, W5P7M5, Q9TTA5, W5PIA1, A7MBB0, W5PB00, Q32T06, W5Q7R8, Q3ZCH5, E1BFG1, W5Q7Z7, E1BLN6, W5QHN0, Q1A2D1, W5P9U4, E1BCV0, W5P964, C5ISA2 and W5Q505.

APPENDIX 4.1. The UniProtKB accession numbers of 102 proteins from a ProteinPilot search of a composite database search of *Bos taurus, Ovis aries and Capra hircus* database of the first, second and third in-solution digestion samples.

P14639, W5PF65, W5NSA6, W5NX51, W5PTG9, A0A0F6QNP7, W5P3R3, W5QH56, W5PSQ7, W5PZS7, W5P4S5, W5P101, W5NXW9, W5Q268, W5PH95, W5QAB1, C8BKD1, D6PZY4, W5QH45, W5PJ97, W5NY95, W5NRG7, W5PTR4, W5NWM2, W5PHP7, W5P5I3, W5P0Q4, W5P812, W5PID9, W5PZI1, W5PBY2, W5NQW9, W5Q0X5, Q28745, W5P1J8, W5NY46, W5NRI1, W5NTW3, G5E604, W5QDG8, W5NPK5, W5QH50, W5NRH2, W5PXC8, Q2KJF1, W5PZF0, W5NWX6, Q1KZF3, P12303, W5PPQ8, W5Q9A2, W5QH54, W5P6F4, W5PH81, W5P627, W5PW21, W5PXU6, V6F9A2, W5PJZ2, W5P8R7, W5QDP8, W5PDR7, W5PD71, W5NSH2, W5PGT9, W5PN97, W5PXI6, W5QGP4, F1N3Q7, W5QI29, F1MLW8, E1BFN5, W5PGT6, W5NYJ9, G3MZE0, W5NZ47, W5Q124, O46544, W5NXP3, W5Q750, W5Q4Q3, F1N514, W5Q961, W5PTL2, W5PIC9, W5PI61, W5P3J3, W5PDE5, W5Q2E1, P17690, W5QFP0, Q9XT27, C6ZP49, Q1RMN8, Q95121, W5PHI7, G3X6I0, W5PAL4, F1MCF8, G3N346, Q3ZCH5 and Q3SYR8.

APPENDIX 4.2. The UniProtKB accession numbers of 67 previously known proteins identified from a ProteinPilot search of a composite database search of *Bos taurus, Ovis aries* and *Capra hircus* database of the combined first, second and third in-gel and in-solution digestion samples.

P14639, Q2UVX4, D6PZY4, C8BKD1, E1BH06, E1BKT9, Q8SPJ1, Q1RMN8, I6W7A2, Q3LRQ1, Q29437, P12303, G3N0V0, Q06AV9, Q2KJF1, B5B304, Q28161, A2P2H9, Q32PJ2, A6QP30, P04272, A8YXZ2, Q09TE3, I3WAE6, Q3SWW8, A6QPP2, F1MCF8, A6H7J7, X4ZFS1, H6WVW6, E1BNR0, F1MZ96, D6PX64, A7MAZ5, Q32T06, E1BI98, Q0P5K6, P00974, E1BB91, O46544, E1BF59, Q3MHN5, Q9XT27, A6NBZ0, P17690, Q9MZS8, G8JKZ8, A5D9D2, F1N3M0, Q3ZCH5, E1BD43, Q2KI85, G3X6N3, F2YQ13, Q28085, P34955, Q95121, Q6LBN7, A5PJT7, Q5GN72, Q3SYR8, Q05B55, A5D7A2, Q9TTA5, A7MBB0, A4IFI0 and C5ISA2.

APPENDIX 4.3. The UniProtKB accession numbers of 207 novel proteins identified from a ProteinPilot search of a composite database search of Bos taurus, Ovis aries and Capra hircus database of the combined first, second and third in-gel and in-solution digestion samples.

W5PF65, W5Q7J0, W5NSA6, W5P6F4, W5P3R3, W5NUX8, W5P4S5, W5PTG9, W5NX51, W5NY95, W5QDG8, W5PH95, W5NQW4, W5NRG7, W5P627, W5QH56, W5PZS7, W5QAB1, W5NPK5, W5PTR4, W5P0Q4, W5NXW9, W5P5I3, W5PW21, W5Q124, W5NSH2, W5NWM2,

W5QH50, W5P1J8, W5QH45, W5PBY2, W5P8R7, W5PGT6, W5PHP7, W5P101, W5PH81, W5NY46, W5PFC9, W5NTW3, W5QH54, W5PID9, W5Q268, W5NYJ9, W5PXX3, W5Q0X5, W5PJZ1, W5PXI3, W5PXU6, W5NRI1, W5PDR5, W5P812, W5PJ97, W5PI61, W5Q9A2, W5QI29, W5P3J3, W5Q4Q3, W5Q750, W5PXC8, W5P338, W5PKA9, W5PZI1, W5PGT9, W5PHI7, W5PE53, W5PD71, Q1KYZ7, W5NRR7, Q28745, W5PEI4, W5QFP0, W5P7S6, W5PDE5, W5QGP4, W5QGG0, W5Q0L2, W5PHP8, W5PPQ8, W5P4C6, W5NXP3, W5NRH2, W5Q0R1, G3MZE0, W5PFJ0, E1BFN5, W5Q2E1, W5PYG2, W5QAA3, F1MIW8, G5E5V0, W5QGD1, F1N514, V6F9A2, W5PDS4, W5P2Q8, W5Q961, W5QGB5, W5P229, Q32S29, W5PZF0, W5PLL2, W5NWX6, W5NZ47, W5NYF4, W5NX96, W5Q3K6, W5PD84, W5PDQ9, A2I7N2, W5QAR2, W5QBW5, F1MLP3, W5PDP6, G8JL00, W5Q0V2, W5P880, W5PN97, W5NXJ3, W5PVL4, W5PAB5, W5QE21, W5Q5A6, W5Q9D5, Q29439, W5QA54, W5PTU7, W5NS93, W5PIJ5, W5PDN1, W5PGS4, W5P3H8, F1MLW8, W5PSM6, W5Q3R0, W5NVM6, W5QH21, W5PW62, W5PZH5, W5QAA1, W5PAL4, W5PG63, G5E5V1, W5NUW3, W5Q4Z3, W5PVH9, G3MXG6, W5PXI0, W5Q038, W5PJ66, W5PB04, W5Q517, W5PTL2, W5PD62, W5P2V6, W5PMY0, W5PB46, G3MWI3, W5QIK8, W5PB07, W5Q5H8, W5P1X9, W5QFL0, W5NPN4, E1BBS9, W5NUU8, W5PC09, W5PIC9, W5NVT0, W5NQ91, F1N045, W5NTD9, W5PIA1, W5NZQ2, W5NQP5, W5QCX2, E1B9D7, F1MMP5, E1B726, G3X6I0, C6ZP49, W5QFH1, W5Q5N5, A5YBU8, W5QDP8, W5NWU4, E1BC09, E1B9F8, W5QJ27, F6RAG5, W5PQH2, W5QID7, W5Q5T2, W5P7M5, W5PZM1, W5PK04, W5PB00, W5Q7R8, E1BFG1, W5Q7Z7, E1BLN6, W5QHN0, Q1A2D1, W5P9U4, E1BCV0, W5P964, W5Q505 and W5PDG4.

APPENDIX 4.4. The UniProtKB accession numbers of 83 disease-associated proteins identified from a ProteinPilot search of a composite database search of *Bos taurus, Ovis aries* and *Capra hircus* database of the combined first, second and third in-gel and in-solution digestion samples.

P14639, W5PF65, W5Q7J0, W5P4S5, W5NX51, W5NRG7, W5P627, W5QH56, W5PZS7, W5QAB1, W5PTR4, C8BKD1, W5P0Q4, W5NXW9, W5PW21, W5P1J8, W5NTW3, W5QH54, W5PJZ1, W5PXU6, I6W7A2, W5Q4Q3, W5PXC8, W5PD71, Q1KYZ7, W5NRR7, Q28745, W5PEI4, W5QFP0, Q29437, P12303, W5P7S6, W5QGP4, W5Q0L2, Q06AV9, Q2KJF1, G5E5V0, A2P2H9, W5QGD1, V6F9A2, W5PDS4, P04272, W5PZF0, Q09TE3, W5PLL2, I3WAE6, Q3SWW8, W5PDQ9, A2I7N2, A6QPP2, W5QBW5, W5PVL4, W5Q5A6, W5PTU7, W5NS93, W5P3H8, W5PZH5, Q9XT27, P17690, W5NUW3, W5Q4Z3, Q9MZS8, G8JKZ8, W5PB07, W5Q5H8, W5P1X9, Q3ZCH5, E1BD43, W5PC09, W5PIA1, W5NQP5, G3X6N3, P34955, F1MMP5, E1B726, Q5GN72, Q3SYR8, Q05B55, E1BC09, W5PZM1, Q1A2D1, W5P9U4 and W5P964.

APPENDIX 4.5. The UniProtKB accession numbers of 77 previously known serum proteins identified from a Mascot search of an *Ovis aries* protein sequence database of the combined first, second and third in-gel and in-solution digestion samples.

W5QAA1, W5QG16, A0A0M4KDI9, W5P7S6, I1WXR3, W5QH56, W5PS45, W5P1J8, W5NYJ9, A2SW69, W5PTR4, W5P1Q0, W5PI61, W5PWY0, W5NUG0, Q1KYZ7, W5PTU7, W5PAM4, W5PA54, W5P4S0, W5QCP0, W5NTD9, W5PZI1, W5PLL0, W5PWU1, I0CF13, D4P8S5, W5PZF0, D6PX64, W5NZB4, W5Q828, D6PZY4, W5Q5H8, W5QDP8, W5PFR8, F2YQ13, W5PDJ6, W5PDG3, A0MPT5, A8DR93, W5PHK5, W5QAH2, W5QFL0, W5PHH3, W5PGW2, B3GS77, Q29400, U3N1L1, W5QGD1, A0A0P0QND2, Q06AV9, B3SV56, W5Q9H4, W5Q7C0, W5PD71, K4P1S5, W5QGW8, W5P3R3, W5PDT4, W5NS93, W5PC09, W5P9U4, W5QD30, W5QIY3, W5QE19, W5P987, W5Q2D7, C8BKD1, P14639, W5NQP5, W5NRR7, W5P7X3, P12303, C5ISA2, W5PPT6, W5PWL1 and A2P2H1.

APPENDIX 4.6. The UniProtKB accession numbers of 302 novel serum proteins identified from a Mascot search of an *Ovis aries* protein sequence database of the combined first, second and third in-gel and in-solution digestion samples.

W5QFE4, W5NX51, W5QFP0, W5PJ97, W5NWM2, W5NPN4, W5PDC8, W5QBW5, W5PF65, W5QDH9, W5PXC8, W5QGG0, W5QGP4, W5PQL6, W5Q4Q3, W5PYG2, W5PGT9, W5P812, W5Q4W5, W5NTL7, W5PWW5, W5PV54, W5PDS4, W5Q2U7, W5Q754, W5PX97, W5PW21, W5P6U4, W5QAB1, W5P5I0, W5PAB5, W5PBY0, W5NSH2, W5P880, W5Q4P0, W5Q3K6, W5PZY7, W5PXU6, W5PKV4, W5QC26, W5P2U2, W5NVN1, W5QH54, W5NS65, W5P691, W5Q749, W5QAX3, W5P9B0, W5Q268, W5Q2E1, W5PXI0, W5NSA6, W5PDE5, W5P9V7, W5NRI1, W5PID9, W5PDR5, W5P8R7, W5PXI3, W5P0Q4, W5Q5C2, W5PH81, W5P4C6, W5PGS0, W5QJ00, W5P336, W5NRG7, W5PIK2, W5QCP9, W5QAA3, W5P101, W5P6F4, W5PPN5, W5PEI4, W5NY95, W5QFK2, W5P229, W5Q9D5, W5QHH3, W5PGT6, W5PG63, W5PPQ8, W5PKA9, W5PVG5, W5QAR2, W5Q224, W5P3J3, W5PJZ1, W5PVL4, W5PXX3, W5Q0R1, W5NXW9, W5QH21, W5NY46, W5QJ69, W5PD12, W5NYG1, W5QI29, W5P3H8, W5QH45, W5P7L5, W5Q0L2, W5PFJ0, W5PTG9, W5Q038, W5QH06, W5PIJ5, W5Q5A6, W5PVX0, W5NVC1, W5P0B2, W5PT48, W5NX89, W5P041, W5PD15, W5PYQ3, W5NY68, W5PW62, W5Q1W4, W5NUL7, W5PEL1, W5P1Y8, W5QA54, W5PG04, W5PMY0, W5Q9B7, W5PSM6, W5NYF4, W5P895, W5PG55, W5P3A7, W5QCH5, W5NXR3, W5Q2R6, W5PDQ3, W5P3G6, W5PTL2, W5QDW7, W5PJE4, W5P9W4, W5PAE2, W5Q910, W5QBN7, W5QF95, W5QBT8, W5PZ27, W5NRV1, W5P0P1, W5NV45, W5Q950, W5PG50, W5NVT0, W5PN97, W5P988, W5QFP2, W5PB04, W5P640, W5PKN2, W5PLJ9, W5PD62, W5PDQ6, W5PDY2, W5PEJ5, W5NZQ2, W5QG19, W5PXN1, W5PQM7, W5PJP9, W5PQ53, W5PYL5, W5QIW7,

W5Q8Y3, W5PWT9, W5PDN1, W5P808, W5QIK8, W5NX74, W5Q9L9, W5PH55, W5Q5F0, W5NSI7, W5PJ85, W5PD75, W5P8Y7, W5P143, W5NT24, W5NTQ9, W5PL69, W5QET9, W5Q370, W5Q9P0, W5PA78, W5P092, W5PHI3, W5Q620, W5PPJ4, W5PVG2, W5PZH5, W5P9V5, W5PZK2, W5PG36, W5QC28, W5P2V3, W5PP29, W5NUU7, W5PIG7, W5P9A6, W5QBV7, W5PKK4, W5P1D3, W5PAJ9, W5Q564, W5PZJ0, W5P1N7, W5PTE9, W5QE21, W5QIM3, W5PEY4, W5P985, W5QD80, W5PX18, W5P4X7, W5PK74, W5Q0N1, W5PLC4, W5PQ63, W5PKW9, W5P9C3, W5QI75, W5PV57, W5NW67, W5PQH3, W5Q5W3, W5Q5R2, W5NSZ2, W5Q517, W5PJ66, W5PZW5, W5QH43, W5Q6L3, W5QHC2, W5Q5P5, W5P4R6, W5PQH2, W5NT03, W5PC99, W5Q7J0, W5Q7R8, W5QH50, W5P5T4, W5Q7T8, W5PE53, W5PIC9, W5Q7Z7, W5PHP8, W5Q4B1, W5PDQ9, W5PVH9, W5Q961, W5PES2, W5Q4Z3, W5PDP6, W5NX96, W5Q9A2, W5NQJ8, W5P248, W5P2I4, W5PD84, W5NSV3, W5NWX6, W5NPK4, W5NQ85, W5PME8, W5P2L1, W5NXV2, W5PGS4, W5QHD7, W5P964, W5PUG1, W5P149, W5PZ65, W5PHA9, W5P2Y4, W5NZ47, W5QCX2, W5P8T8, W5PY75, W5PB46, W5PUJ4, W5Q0V2, W5Q9K6, W5PNU9, W5NXI6, W5PA64, W5NQM5, W5NSH6, W5PXV7, W5PPT3, W5P1X7, W5Q0U9, W5PGZ8 and W5P5V0.

APPENDIX 4.7. The UniProtKB accession numbers of 83 disease-associated serum proteins identified from a Mascot search of an *Ovis aries* protein sequence database of the combined first, second and third in-gel and in-solution digestion samples.

P14639, W5PF65, W5Q7J0, W5P4S5, W5NX51, W5NRG7, W5P627, W5QH56, W5PZS7, W5QAB1, W5PTR4, C8BKD1, W5P0Q4, W5NXW9, W5PW21, W5P1J8, W5NTW3, W5QH54, W5PJZ1, W5PXU6, I6W7A2, W5Q4Q3, W5PXC8, W5PD71, Q1KYZ7, W5NRR7, Q28745, W5PEI4, W5QFP0, Q29437, P12303, W5P7S6, W5QGP4, W5Q0L2, Q06AV9, Q2KJF1, G5E5V0, A2P2H9, W5QGD1, V6F9A2, W5PDS4, P04272, W5PZF0, Q09TE3, W5PLL2, I3WAE6, Q3SWW8, W5PDQ9, A2I7N2, A6QPP2, W5QBW5, W5PVL4, W5Q5A6, W5PTU7, W5NS93, W5P3H8, W5PZH5, Q9XT27, P17690, W5NUW3, W5Q4Z3, Q9MZS8, G8JKZ8, W5PB07, W5Q5H8, W5P1X9, Q3ZCH5, E1BD43, W5PC09, W5PIA1, W5NQP5, G3X6N3, P34955, F1MMP5, E1B726, Q5GN72, Q3SYR8, Q05B55, E1BC09, W5PZM1, Q1A2D1, W5P9U4 and W5P964.

APPENDIX 4.8. List of 349 bovine aligned gene entries derived from inputting 379 *Ovis aries* protein entries in the PANTHER classification tool.

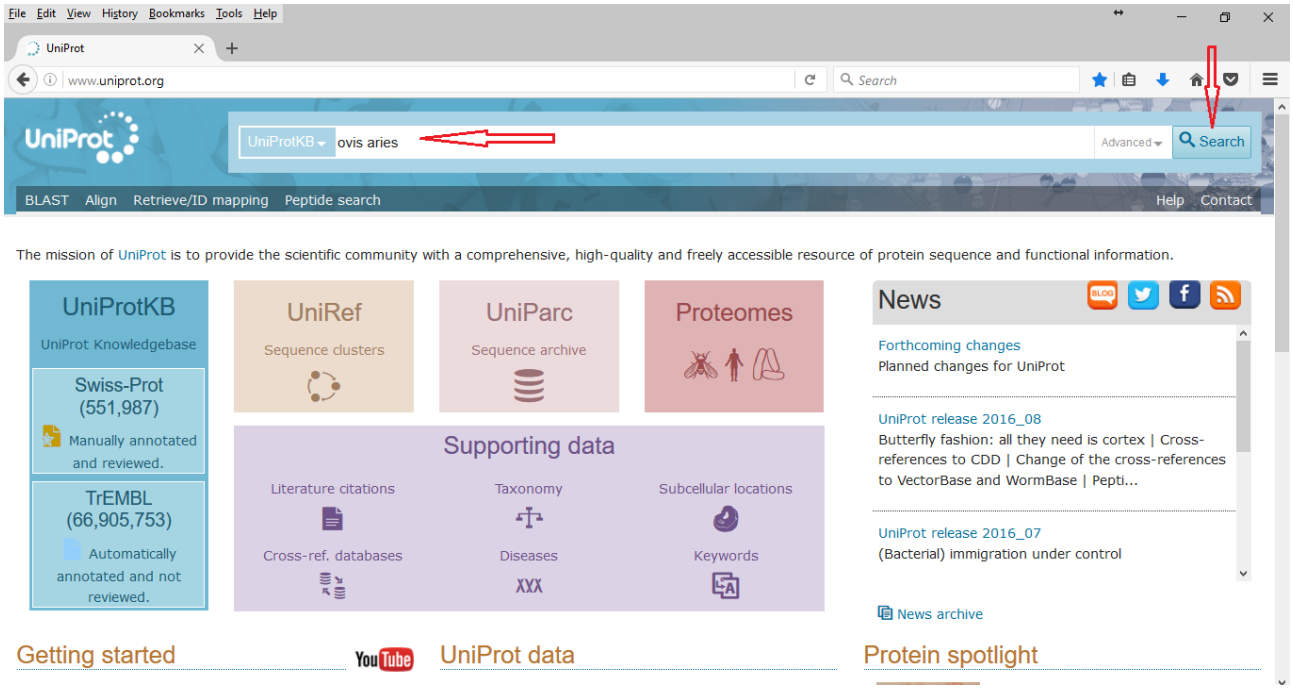
CA2, MYCBP2, F2, LRG1, OAS2, MAN2B1, KCTD19, PSMA4, PSMB2, KLKB1, CNTRL, F10, PLXNA3, APOH, CPB2, SERPING1, HGFAC, VTN, GPX3, fH, PLG, SRSF7, C1S, PLA1A, F12, EXOC3L1, VDAC1, RRP1B, XPO1, TAF2, TALDO1, JUP, SELENBP1, VCL, BRIX1, C7, RPS27A, GC, TTC28, LRP2, CTSA, TMEM62, C2CD2L, CGN1, HABP2, TF, PDE3A, A2M,

ITIH4, FABP5, IGFBP5, CCHCR1, VDAC2, DOPEY2, BLMH, ZNF804A, PHYHIPL, ALKBH1, MACF1, DNM1, CDH13, ITIH2, LCP1, AHSG, PSMA1, HPX, ABCA13, ATRN, NCAPH, C6, AVL9, KPNA5, ITIH3, NFATC2, SERPIND1, TTR, FSIP2, ECM1, AFG3L2, NR1D1, DHX37, FAM72A, HSPA8, LUM, CFD, CTTN, POGZ, CPN1, SMCHD1, ORM1,OMD, INTS4, MBL2, SYNCRIP, PRG4, SKA1, TNXB, LRRC15, F13B, RBP4, EME1, IQSEC2, SETBP1, GHR, MYO7A, TGM3, CWC22, NSL1, VWF, HSP90AA1,HSPCA, DSP, C9, SEPP1, LAP3, AHNAK, TUBA4A, NOX1, FZD6, F13A1, PSMA5, AMBP, OTOF, CASP14, GGH, PF4, ASB2, SFN, PCOLCE, SERPINF2, SMURF1, ADIPOQ, ERCC6, IDE, KIF27, CLEC3B, CEP104, C1QB, C8B, PLTP, PEPD, CTNNBL1, ANPEP, MYO1B, TBC1D23, CDH5, BTD, MGAM, PLEKHM1, NEK5, APC2, SERPINA5, MASP1, ADGRF2, AP1B1, GOT2, COL6A3, GC, PLEKHG1, SYNGAP1, CHMP4C, MROH6, C1QA, C8A, THBS1, MST1, THBS4, RASGRF1, FGG, C11orf63, DYM, SHANK1, A1BG, LARP6, TUBB, Lats1, SAMM50, HEPHL1, IFT172, PROS1, IGFBP2, LRRC9, APOA1, EPG5, C1QC, ANXA2,ANX2, SERPINA7, IFT140, F5, IGHM, BTBD9, ABHD14B, AZGP1, HRG, SERPINF1, POLA1, ZMYM4, PRKG2, TBC1D5, DPEP2, PHKA1, NTN4, PSMA6, PPIA, MST1, CCAR2, APOD, Crisp3,CRISP3, TBC1D32, ALOX5, LRRK2, NCAM1, C1R, FGA, COL6A1, CFI, UHRF1BP1L, WDR47, CHI3L1, LDHB, CDSN, MBIP, FBP1, SNX25, SERPINC1, TF, APOF, DSG1, TRMT11, TXNL4B, LRRC17, ACTG1, TICRR, CDK5R1, AP1B1, RYR1, AASDH, C1orf204, GALNT13, FBXO3, PTPRC, TRPM4, ALMS1, GSN, HIST1H1D, IGF2R, OSMR, ARG1, FETUB, APOB, ZNF638, ENO1, VPS9D1, CDH19, DSG4, CAP1, APOC3, HCFC1R1, SERPINB13, PRG4, CRABP2, DSC3, C2, CP, DNAH17, IGFALS, PROC, MASP2, TGM1, CRP, AFM, INPP5B, AGT, CSMD1, OMD, SAFB2, WHSC1L1, KNG1, BTAF1, TCN2, HIVEP1, ANKRD11, MOB3B, KLHL8, F9, ITGA6, CD14, COLEC11, PSMA7, FCGBP, ZNF248, SART3, APOE, DSC1, FBLN1, CABYR, PDE4DIP, FNBP1L, G6PC, SUSD1, DNAH2, PLA2G7, SERPINB5, PLEC, C5, NUP155, ALB, MANSC1, IGHE, FAM78B, CCDC89, PGLYRP2, FLG2, GAPDH, ADAM15, TTN, PKP1, TFRC, ICOSLG, DRP2, LBP, CFP, GPLD1, CCDC177, XIRP2, HIST1H2AD, PON1, GOLIM4, PIGR, PSMA3, PKN3, SERPINB12, SERPINA1, BST1, ANK1, APOA4, LMNA, VNN1, BROX, NACA, APOM, SERPINA10, KIAA1751, OSBPL6, GRK7, ABCC5, SHBG, CD44, RND3, APOA2 and C4BPA.

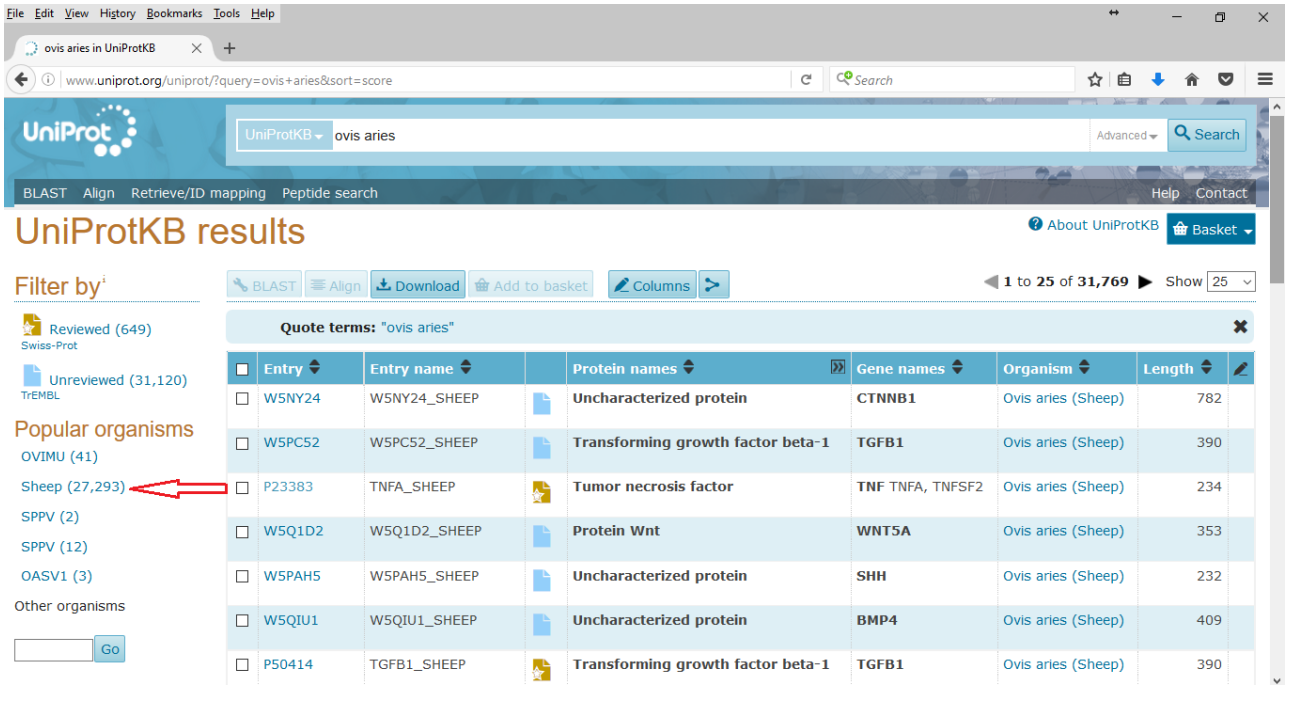
APPENDIX 6.0. Generating a UniProtKB sheep-only protein FASTA sequence database

Figure A6.0. Steps for generating a UniProtKB sheep-only protein FASTA sequence database

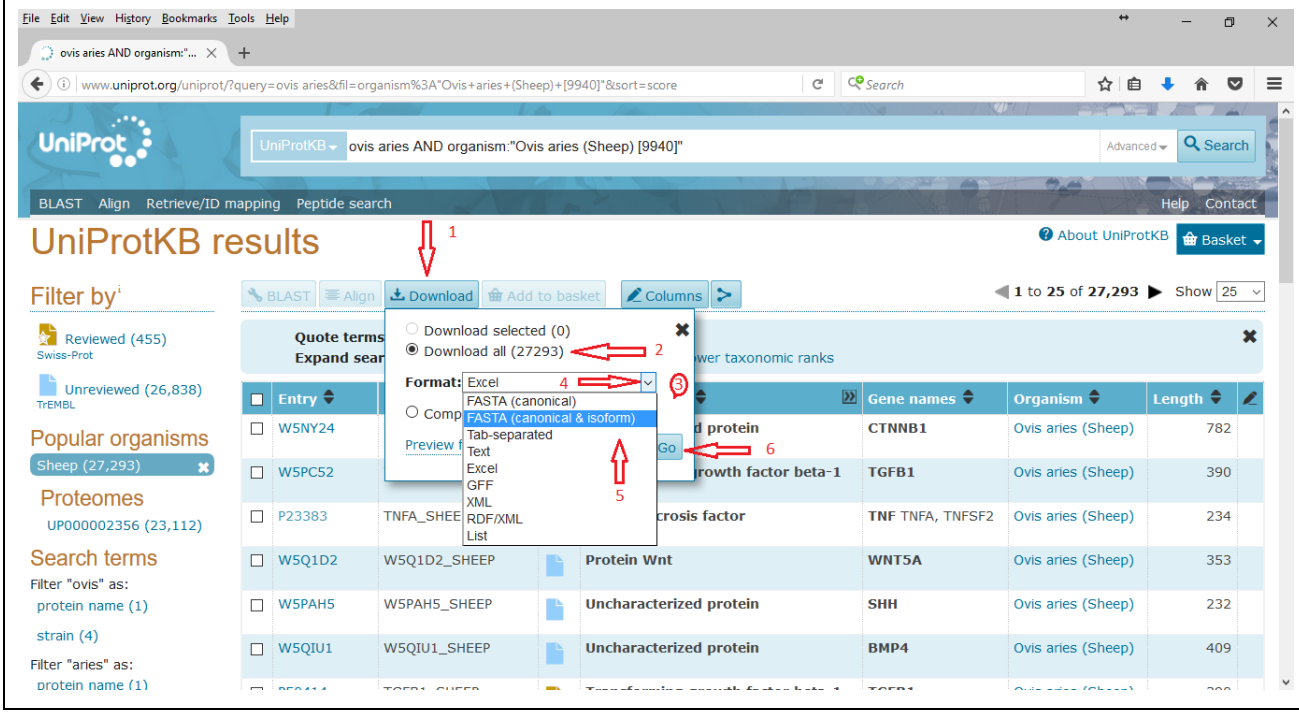
Step 1. Navigate to www.uniprot.org and type *Ovis aries* and then search (red arrows).



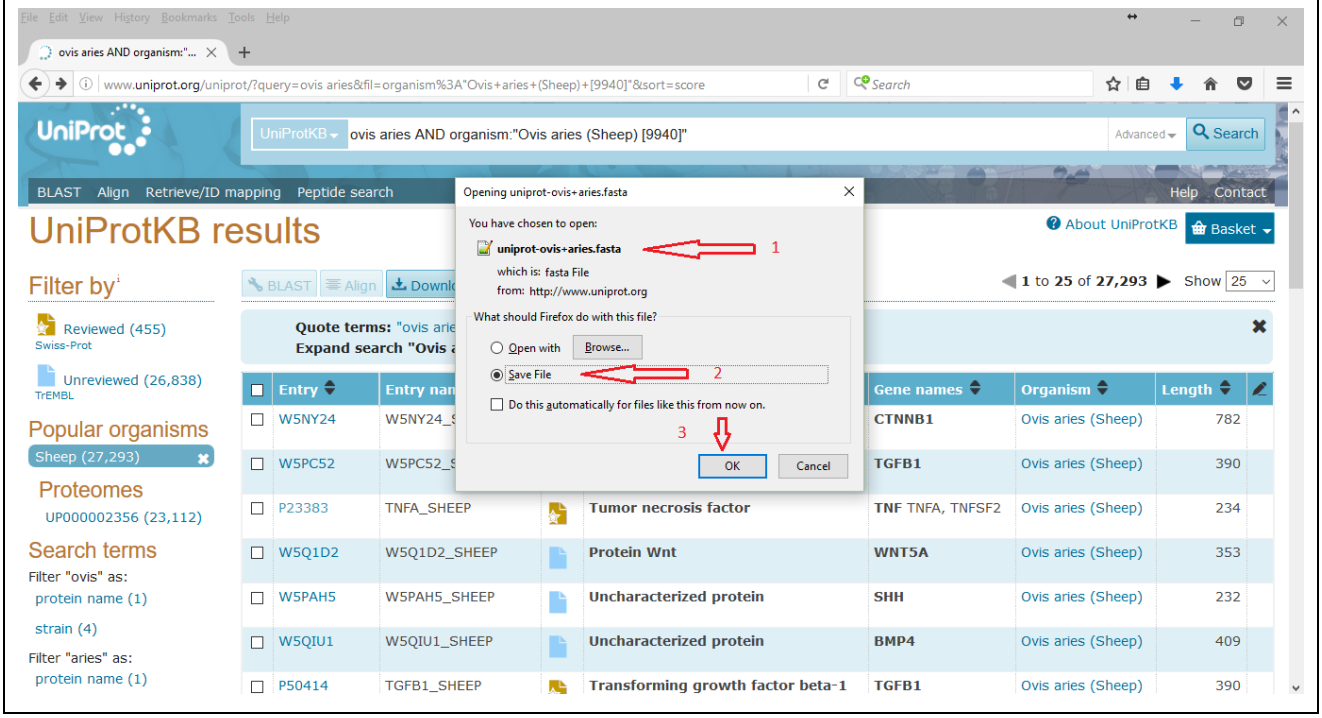
Step 2. Under popular organisms on the left side of the screen, select Sheep (red arrow).



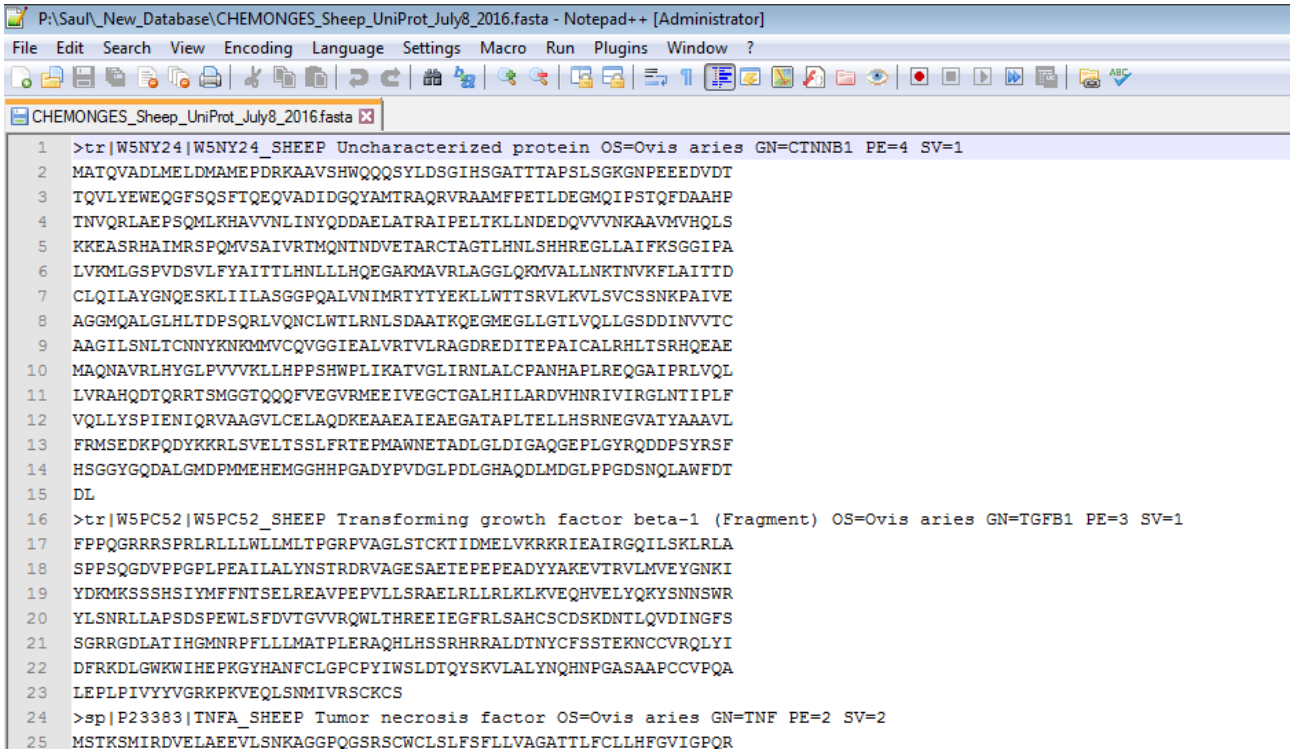
Step 3. Click Download (1), Download all (2), Uncompressed (3), Format (4), FASTA (Canonical & isoform)(6), and then Go (6).



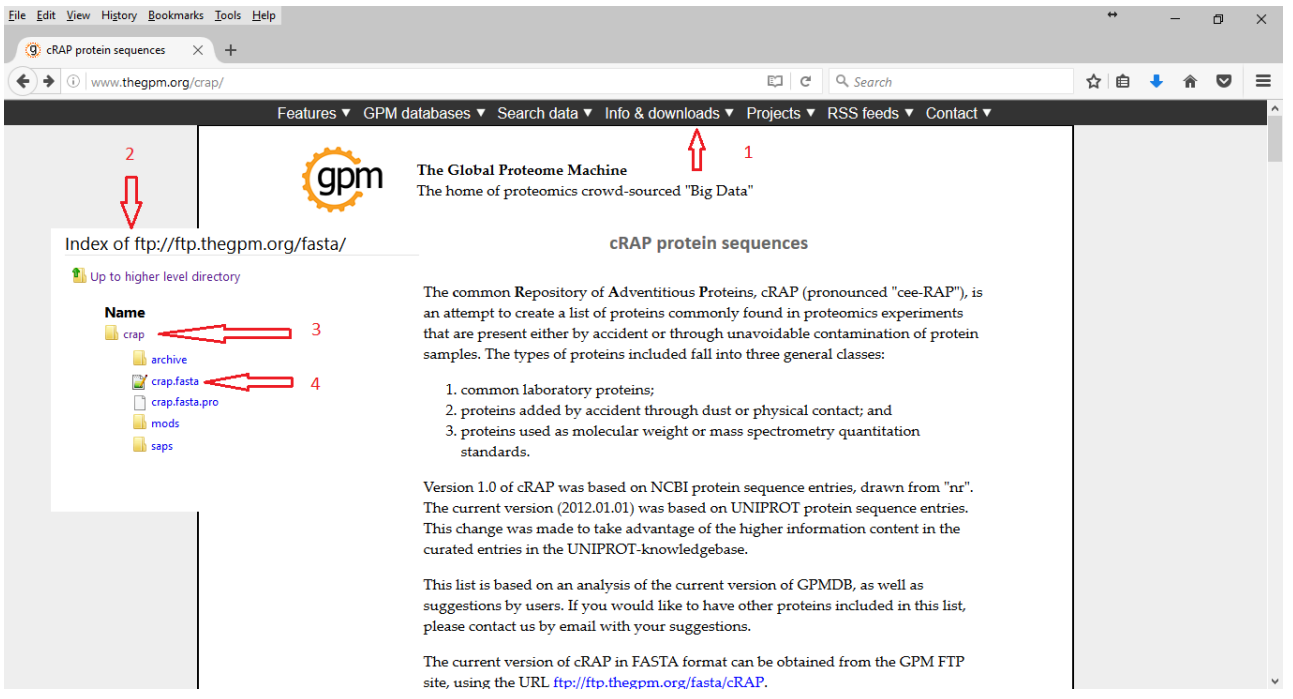
Step 4. Saving the downloaded sequence fasta file: Take note of the file name (1), Save File, click OK (3).



Step 5. Open the FASTA file in a text code editor such as Notepad++ or EditPad and inspect.



Step 6. Download and append common Repository of Adventitious Proteins (crap.fasta) sequences to the open FASTA file from http://www.thegpm.org/crap/, by following the arrows 1, 2, 3 and 4. Add indexed retention time (iRT) peptide sequences and save the FASTA file. Create a concatenated target/decoy database from this FASTA file using PeptideShaker.



APPENDIX 6.1. Extraction of protein data from ex-diagnostic sheep serum

Ex-diagnostic frozen serum samples from nine sheep that presented with a range of different conditions were obtained from UQ's School of Veterinary Science clinical pathology laboratory (UQ.O) (Table 6.1). These samples (SC502–SC510), a pooled variant representing all samples (SC511) and a pooled variant representing healthy free-range sheep samples (SC512) as a control, were subjected to in-gel digestion and in-solution using generic methods as described Chapter 3.

Table A6.1. Sick sheep serum samples for the derivation of peptide data for PSL enrichment.

UQ.O ID	UQ lab no	Unique Lab sample ID	Crude serum protein (µg/µL)	Reconstituted acetone-precipitated serum protein (µg/uL)
1	66943	SC502	65	76
2	67737	SC503	57	64
3	68327	SC504	78	72
4	68354	SC505	78	64
5	68364	SC506	56	80
6	68447	SC507	47	76
7	68494	SC508	69	68
8	70192	SC509	69	80
9	70802	SC510	-	80
10	Pooled (1-9)	SC511	-	76
11	Pooled SA	SC512	-	15

Key: UQ; The University of Queensland; UQ Lab; UQ School of Veterinary Science clinical pathology laboratory; UQ.O; serum samples from UQ Lab; SA: serum samples from healthy sheep from Serum Australis Pty Ltd.

The results of Coomassie-stained 1D SDS-PAGE gels used for acetone-precipitated serum samples from sick and healthy sheep are presented in Figure A6.1 A and B, and protein IDs are presented in panel C. The gels in panel A and panel B were used for fractionating reconstituted acetone-precipitated pooled serum from healthy sheep (SA11) and serum from sick sheep (UQ.O 1 to UQ.O 10). The two gels had a lane each loaded with 4 µL of a molecular weight (MWt.) marker and BSA standard (arrows). Sample lanes were loaded with 200 µg of protein. Note the differences in appearance of the bands for each sample and the replicates of UQ.O 10 in the two gels.

Panel C of Figure A6.1 shows the protein IDs from in-gel digestion of sick sheep serum samples (SC502–SC510, SC511 [pooled]) and a control in ProteinPilot, PeptideShaker and Mascot search

engines; meanwhile, the protein IDs for in-solution digestion workflow are shown in panel D. The number of protein IDs from in-gel digestion was marginally higher than those yielded by in-solution digestion.

The results of protein IDs for in-solution digestion are presented in Figure A6.1 D. The combined number of protein IDs in the Mascot search engine from in-gel and in-solution digestion from the entire workflow are provided in Figure A6.1E.

The total number of combined protein IDs yielded by Mascot from in-gel and in-solution digestion of serum samples from sick sheep (SC502–SC511 [UQ.O1-UQ.O9]) and a pooled control sample (SC511) from healthy sheep are shown in Figure A6.1 E. Except for two samples (SC502 and SC506), the number of protein IDs were comparable for all other samples. The overall outcome of this entire workflow, using the Mascot platform, was the identification of 314 proteins.

It was evident from the results of 1D SDS-PAGE that profiles of the individual samples had distinct differences based on the appearance of protein bands in the individual case lanes compared to the control sample from healthy sheep (Figure A6.1 A and B). The pooled sick sheep serum sample replicates (UQ.O10) were comparable within the gels, but there was an inter-gel difference in the appearance of the bands. Despite the gels being prepared and run identically together in the same tank, the protein bands of the UQ.O10 sample in gel A appeared dense and diffuse, and yet the bands were well separated in the three replicate lanes in gel B. Intrinsic gel properties or a variation in the physical nature of the sample during loading could have contributed to this visual difference (for example, gel A was loaded before gel B). This illustrates a known challenge of 1D SDS-PAGE: inter-gel reproducibility[116].

The numbers of protein IDs from in-gel and in-solution workflows from sick sheep serum samples were comparable, averaging approximately 125 IDs in the combined results of each individual sheep (Figure A6.1 C and D). ProteinPilot™ had the highest number of IDs, followed by Mascot and then PeptideShaker.

The UniProtKB accession numbers of the 183 protein IDs from PeptideShaker for inclusion in the PSL were W5Q7J0, W5NSA6, W5NX51, W5PF65, W5Q5H8, W5NRI1, W5P0Q4, P14639, W5NRG7, W5NQ46, W5P6F4, W5NWM2, W5NQW4, W5QDG7, W5Q124, W5PW21, P32262, Q9XT27, W5Q5A6, W5PZI1, P02075, W5PJ97, W5PHP7, A5YBU9, F2YQ13, W5PTG9, W5NXW9, W5PI61, W5NTW3, W5NY46, W5PJZ1, W5PFC9, Q1A2D1, P20757, W5NSH2, P29701, W5QAB1, W5P1J8, W5PXC8, W5P5I0, W5Q9A2, W5Q0X5, W5PKA9, P12303, W5PID9, W5PHI7, W5PGT6, W5PXI3, W5PD71, W5P7S6, W5Q9D5, B6UV62, W5QFP0, W5P101, W5QBW5, W5P3R3, W5PGT9, W5Q0L2, W5PTU7, P52210, W5P1X9, W5QH46, W5QH54, W5Q961, P02190, C8BKD1, W5PE53, W5QI29, D6PZY4, W5NWX6, W5QH50,

W5NXP3, W5Q4Q3, W5PYG2, P42819, W5NZ47, W5PHP8, W5NPK5, W5NRH2, W5P812, W5PDJ6, C8BKC5, W5QGD1, W5Q0R1, W5PIA1, A2SW69, Q1KYZ7, W5PVL4, W5PJR0, W5Q749, W5Q2E1, W5NTT7, W5PDS4, W5PDE5, W5P336, P50450, W5PH81, W5P3J3, W5Q268, W5NS74, O46544, W5QA64, W5NXP6, W5PBY0, W5PUC1, Q29439, W5QA54, W5NYF4, C5IS96, W5Q4Z3, W5PJG0, W5QGP4, W5QA07, W5NUJ7, A2P2H1, W5NUU7, B0BL71, A0A0U1YZ59, W5PJ69, W5Q3K6, W5NXM6, W5PDQ9, W5PEI4, W5QDF3, W5PDR5, W5NXJ3, W5P880, W5NX96, W5PZH5, W5PPQ8, W5Q7Z7, W5Q7R8, P68056, W5QDP8, W5PN97, W5PS88, W5PJR5, W5PDP6, W5P5W9, D6PX64, W5P9J8, W5PG63, W5NQP5, W5PV54, W5PLL0, P49920, D7RIF5, W5P4C6, W5PLC9, Q4TVY4, W5PUJ4, W5PJ66, W5NXM1, W5NR06, W5NU00, W5Q2U7, W5PI92, W5PAJ9, W5PLQ1, W5NQ21, W5PD62, A0A077JGJ6, P11839, W5P481, W5Q7T8, C5IJA0, W5NYA7, H9CJU6, W5PD84, P0C276, W5QFQ0, W5PK04, A9YUY8, W5NYA1, W5NV73, W5QBV7, C0LQH2, K4P494, W5P5K9, W5PPG6, W5P323, A0A0H3V7A0 and W5Q723.

This workflow added 71 proteins that were not identified in Chapter 4 – the analysis of the circulating acellular proteome of healthy sheep, apart from 112 protein IDs that were common between the two workflows. The UniProtKB accession numbers of the sick sheep-associated proteins are W5Q5H8, W5NQ46, P02075, F2YQ13, Q1A2D1, B6UV62, P52210, W5P1X9, W5QH46, P02190, P42819, W5PDJ6, C8BKC5, W5PIA1, W5PJR0, W5NTT7, W5NS74, W5QA64, W5NXP6, W5PUC1, W5QA54, C5IS96, W5PJG0, W5QA07, W5NUU7, B0BL71, W5PJ69, W5NXM6, W5QDF3, W5PDR5, W5PZH5, P68056, W5PN97, W5PS88, W5PJR5, W5P5W9, W5P9J8, W5PG63, W5NQP5, P49920, D7RIF5, W5PLC9, Q4TVY4, W5PUJ4, W5NR06, W5NU00, W5Q2U7, W5PI92, W5PAJ9, W5PLQ1, W5NQ21, W5PD62, A0A077JGJ6, P11839, W5P481, W5Q7T8, C5IJA0, W5NYA7, H9CJU6, W5QFQ0, W5PK04, A9YUY8, W5NV73, W5QBV7, C0LQH2, K4P494, W5P5K9, W5PPG6, W5P323, A0A0H3V7A0 and W5Q723.

Also, there were 21 proteins that were identified in Chapter 4, but not in sick sheep serum suggesting that these proteins probably diminish during illness. The UniProtKB accession numbers of these proteins were W5P8R7, W5PXU6, W5PXX3, W5P4S0, W5P9B0, W5PHH3, B5B304, W5QGG0, W5QAA3, W5PIJ5, W5PIC9, W5PVH9, W5PWE9, W5NVM6, W5PTE9, P30035, W5P229, W5QH45, W5P3H8, W5P2V3 and W5QAR2.

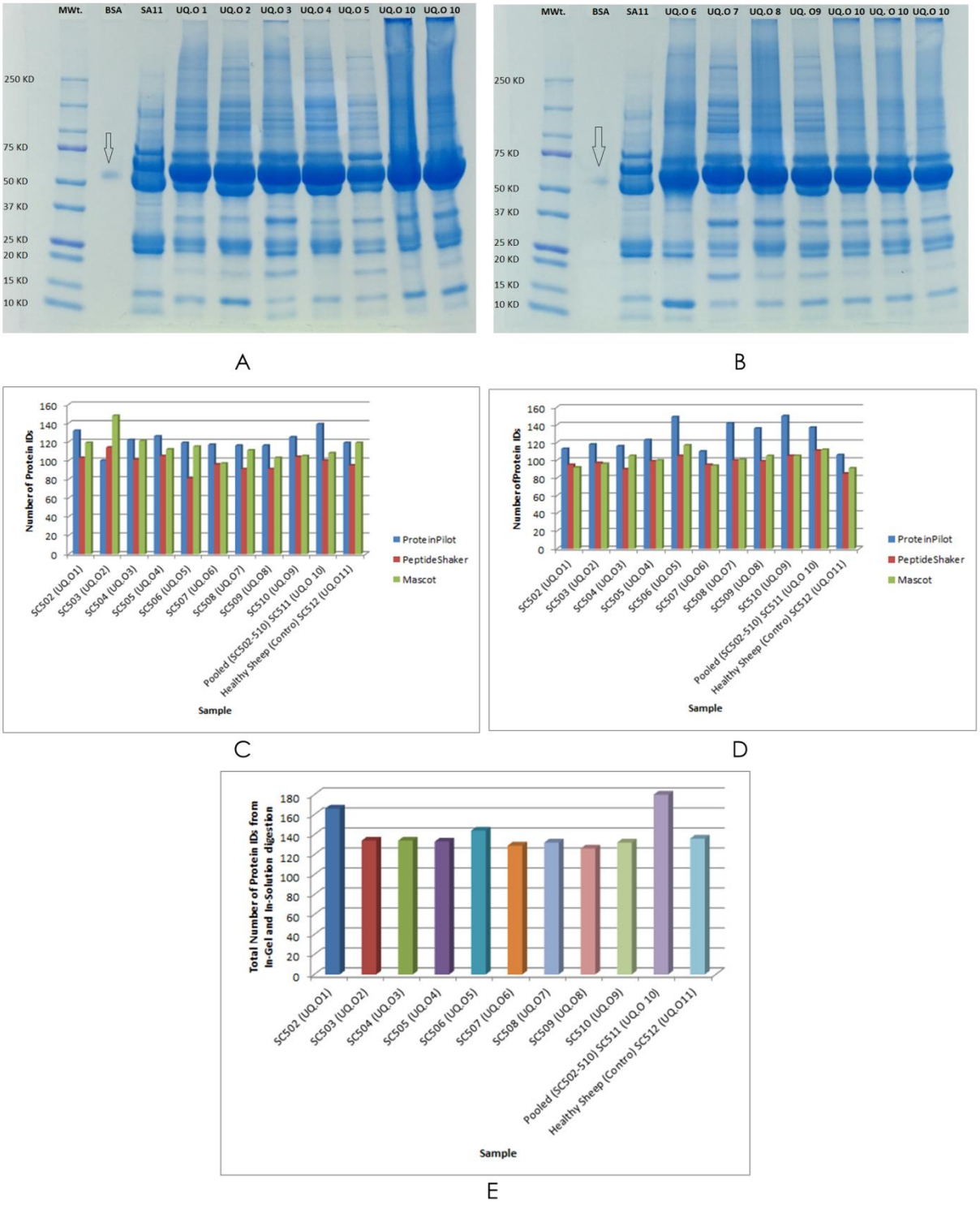


Figure A6.1. Extraction of protein data from ex-diagnostic sheep serum. A and B = Coomassie-stained 1D SDS-PAGE preparations for fractionating acetone-precipitated pooled serum from healthy sheep (SA11) and serum from sick sheep (UQ.O 1 to UQ.O 10); C = protein identifications (IDs) from in-gel digestion; D = protein IDs from in-solution digestion; E = total number of protein IDs from in-solution and in-gel digestion of the individual, pooled and healthy control samples.

APPENDIX 6.2. Derivation of peptide data from plasma and serum of endotoxin-treated sheep

Plasma and serum samples of sheep that were treated with endotoxin were used for this experiment (Table A6.2). The samples were categorised based on trait/strain or experimental group of the sheep model of intensive care based on earlier findings[15]. The samples were subjected to 1D SDS-PAGE and in-solution digestion as described in the generic methods in Chapter 3.

Table A6.2. Plasma and serum samples pooled from 24 sheep in groups of six showing samples taken before and after endotoxin treatment for in-solution and in-gel workflows. †

Experiment/Samples→		**B4 E. Rx Plasma**	**After E. Rx Plasma**	**B4 E. Rx Serum**	**After E. Rx Serum**
↓Sheep trait/Strain	**Workflow**				
Wool production	In-solution	SC537/717	SC538/718	SC539/719	SC540/720
	In-gel	SC553/734	SC554/735	SC555/736	SC556/737
Parasite resistance	In-solution	SC541/721	SC542/722	SC543/723	SC544/724
	In-gel	SC557/738	SC558/739	SC559/740	SC560/741
SYSS	In-solution	SC545/725	SC546/726	SC547/727	SC548*/728
	In-gel	SC561/742	SC562/743	SC563/744	SC564/745
Saline (endotoxin control)	In-solution	SC549*/729	SC550/730	SC551/731	SC552/732
	In-gel	SC565/746	SC566/747	SC567/748	SC568/749

† The experiments were repeated due to sub-optimal protein recoveries, thus the double sample numbers; for example, SC537/717. **Key:** B4 = before; E. = endotoxin; Rx = treatment; SYSS = stockyard sourced sheep; SC = unique laboratory sample identification (investigator initials followed by sample serial number); * = No peptide peaks detected.

Coomassie-stained 1D SDS-PAGE gels of pooled plasma and serum samples from 24 sheep from a sheep model of intensive care are presented in panels A, B, C and D of Figure A6.2(a). The gel grouping was based on the sheep trait/strain or experimental groups (4 groups × 6 sheep): A –wool production, B – parasite resistance, C – SYSS (unknown strain/trait), D – SYSS saline (sham challenge; unknown strain/trait). The results represent samples drawn before and after treating the sheep with endotoxin (A, B & C), and experimental controls that received saline (D). Sample lanes were loaded with 200 µg of protein in duplicate. Note the differences in P1P lanes (gel A), BLP lanes (gel C) and BLS lanes (gel D). Serum lanes had an intense band in the 15 kD region that were

less distinct in plasma lanes. Protein yields from 1D SDS-PAGE are shown in panel E, which shows that protein extraction from the first 1D SDS-PAGE was better than that from the second.

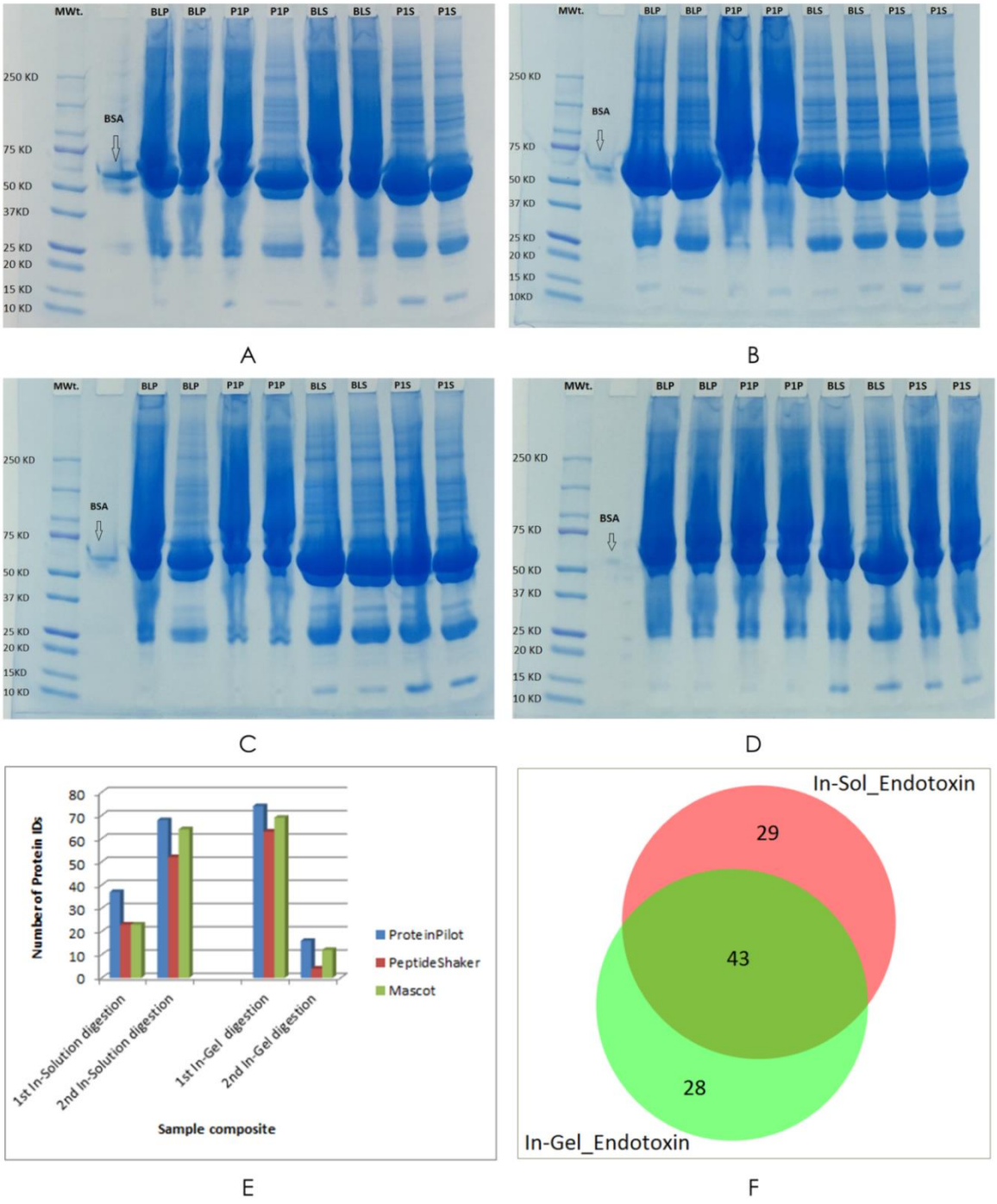


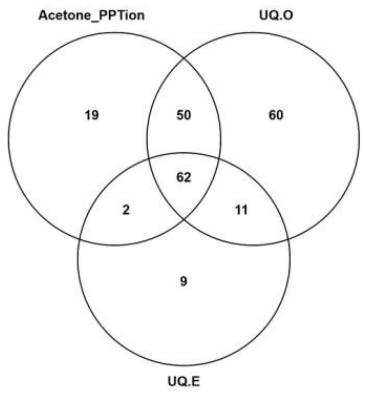
Figure A6.2(a). Protein data of endotoxin-exposed sheep workflow. Coomassie-stained 1D SDS-PAGE gels of pooled plasma and serum samples from 24 sheep from a sheep model of intensive care based on strain: A = wool production; B = parasite resistance; C = stock yard sourced sheep (SYSS)(unknown strain/trait); D = SYSS saline (sham challenge); E = protein IDs from in-solution workflow of the first (SC537–SC552) and second (SC717–SC732) digestions of samples from

sheep exposed to endotoxin in ProteinPilot, PeptideShaker and Mascot; F = Mascot protein IDs from the entire workflow of endotoxin-treated sheep. **Key:** Rx = treatment; BLP = before endotoxin Rx (plasma); P1P = after endotoxin Rx (plasma); BLS = before endotoxin Rx (serum); P1S = after endotoxin Rx (serum); IDs = identifications; In-Sol_Endotoxin = protein IDs from combined 1st and 2nd in-solution digestion of endotoxin workflow; In-Gel_Endotoxin = protein IDs from combined 1st and 2nd in-gel digestions of the endotoxin workflow.

Figure A6.2(a) E shows the number of protein IDs from the in-solution workflow of the first (SC537–SC552) and second (SC717–SC732) digestions of samples from sheep exposed to endotoxin in ProteinPilot, PeptideShaker and Mascot. Protein yields from the second in-solution digestion were better than the first. Mascot identified 100 proteins in the combined in-solution (In-Sol_Endotoxin) and 1D SDS-PAGE (In-Gel_Endotoxin) from the entire workflow of endotoxin-treated sheep, including controls, as shown in Figure A6.2(a) F, and PeptideShaker identified 84 proteins.

The 84 PeptideShaker protein IDs in UniProtKB from this workflow for inclusion in the PSL were W5Q5H8, P14639, P29701, W5NSA6, W5PF65, W5NRI1, W5QDG7, W5NX51, W5NQ46, Q9XT27, W5Q5A6, F2YQ13, W5Q7J0, W5PW21, W5PHP7, W5NRG7, P32262, A5YBU9, P12303, W5PTG9, W5P1J8, W5NTW3, W5NQW4, W5PZS7, W5NU00, W5P6F4, W5QAB1, P21621, W5PJ97, W5NXW9, W5P3R3, W5NSH2, W5NWM2, W5PJZ1, W5NRH2, W5PID9, W5NY46, O46544, W5PPQ8, W5Q124, P20757, W5P0Q4, W5Q4Q3, Q1A2D1, W5P101, W5QH46, W5PZI0, P42819, A2P2H1, W5PDE5, W5Q0X5, W5PXC8, W5NPK5, W5QH50, C8BKD1, D6PZY4, W5PD71, W5PFC9, W5QH54, W5PG63, W5PHI7, B6UV62, Q1KYZ7, W5Q1R5, W5NXP3, P29455, Q28579, P30035, W5QAR2, W5PVL4, W5PE53, W5Q0L2, W5NXM1, W5QA64, W5NYA1, Q7M2U8, W5NXJ3, W5PGT6, W5QFP0, W5P0H0, W5PJG0, W5NZ47, W5QI29 and P23383. Of these protein IDs, only 63 were common between proteins identified in Chapter 4 that were derived from evaluating the circulating acellular proteome of healthy sheep, and those from endotoxaemic sheep. As compared with protein IDs in Chapter 4, a total of 69 protein IDs were not identified in endotoxaemic blood, but there were additional 20 protein IDs generated from this workflow. The UniProtKB accession numbers of these 20 proteins are W5Q5H8, W5NQ46, F2YQ13, W5PZS7, W5NU00, P21621, Q1A2D1, W5QH46, W5PZI0, P42819, W5PG63, B6UV62, W5Q1R5, P29455, Q28579, W5QA64, Q7M2U8, W5P0H0, W5PJG0 and P23383.

Considered collectively, the sick sheep and endotoxaemic sheep workflows contributed 80 proteins to the PSL pool (Figure A6.2(b)).

	**Figure A6.2(b).** Comparison of protein IDs from the analysis of the circulating acellular proteome of healthy sheep compared with IDs drawn from sick sheep (UQ.O) and sheep treated with endotoxin from *E. coli* (UQ.E). The UQ.O and UQ.E workflows collectively contributed 80 protein IDs to the PSL that were not detected in healthy sheep (Chapter 4).

APPENDIX 6.3. Generation of *in silico* predicted synthetic peptides of selected proinflammatory cytokines

Briefly, complete tryptic peptide sequences of homologous peptide sequences between *Bos taurus* and *Ovis aries* of IL-6, IL-3, IL-1α, IL-1β and TNF-α suitable for Fmoc synthesis or solid phase peptide synthesis[426] were predicted by *in silico* digestion with the aid of Skyline software[172]. The basic selection criteria was that peptide candidates were allowed a maximum of one missed tryptic cleavage site while excluding reactive cysteine and methionine residues and histidine where possible, in order to avoid higher charge states during electrospray ionisation. The length of peptides was limited to 8 - 15 amino acids to assure reasonable yield in peptide synthesis. Two top-ranking unique peptides per protein with capabilities of being detected by mass spectrometry that met this basic selection criteria were chosen as targets for proteomics experiments with Skyline. Each of the 10 target peptides was then synthesised by Mimotopes Pty Ltd, (Clayton, Victoria, Australia) in automated synthesisers using the mild Fmoc chemistry method. The peptides were synthesised as unlabelled and K [13C6: 15N2]-labelled by introducing stable isotope labels: heavy lysine (K)-containing ^{13}C and ^{15}N atoms were incorporated at the C-terminal tryptic residue (Table A6.3). The heavy and light peptides were mixed in equimolar proportions and 1000 fmol were spiked into 16 samples (SC717-SC732) or analysed directly in one sample (SC733) using the TripleTOF 5600+ instrument in a DDA mode as decribed in Chapter 3, to collect fragment ion spectra (full MS/MS scans) as illustrated by the examples in Figure A6.3.

Table A6.3. Unlabelled and K[13C6: 15N2]-labelled H- and -OH termini synthetic peptides of five sheep cytokines.

Peptide	Hydrophobicity index	Molecular Weight	Length	Peptide sequence K*= (1042) labelled	Cytokine
1	0.295	914.1	8	LLLTTPEK	Interleukin 6
2	0.037	1108.2	9	MQSSNEWVK	
3	-0.020	1473.6	14	ITPSPEGSLNSDEK	Interleukin 3
4	0.214	1625.8	15	AFMTFATDTFGSDSK	
5	0.074	1245.4	11	FMSLDTSETSK	Interleukin 1a
6	0.055	1267.4	11	SAHYSFQSNVK	
7	0.352	1202.5	11	QVVSVIVAMEK	Interleukin 1b
8	0.049	1541.7	14	GDTPTLQLEEVDPK	
9	0.154	909.1	8	IAVSYQTK	TNF-α
10	0.344	857.1	8	VNILSAIK	
11	0.295	922.2	8	LLLTTPEK*	Interleukin 6
12	0.037	1116.3	9	MQSSNEWVK*	
13	-0.020	1481.6	14	ITPSPEGSLNSDEK*	Interleukin 3
14	0.214	1633.8	15	AFMTFATDTFGSDSK*	
15	0.074	1253.4	11	FMSLDTSETSK*	Interleukin 1a
16	0.055	1275.4	11	SAHYSFQSNVK*	
17	0.352	1210.5	11	QVVSVIVAMEK*	Interleukin 1b
18	0.049	1549.7	14	GDTPTLQLEEVDPK*	
19	0.154	917.1	8	IAVSYQTK*	TNF-α
20	0.344	865.1	8	VNILSAIK*	

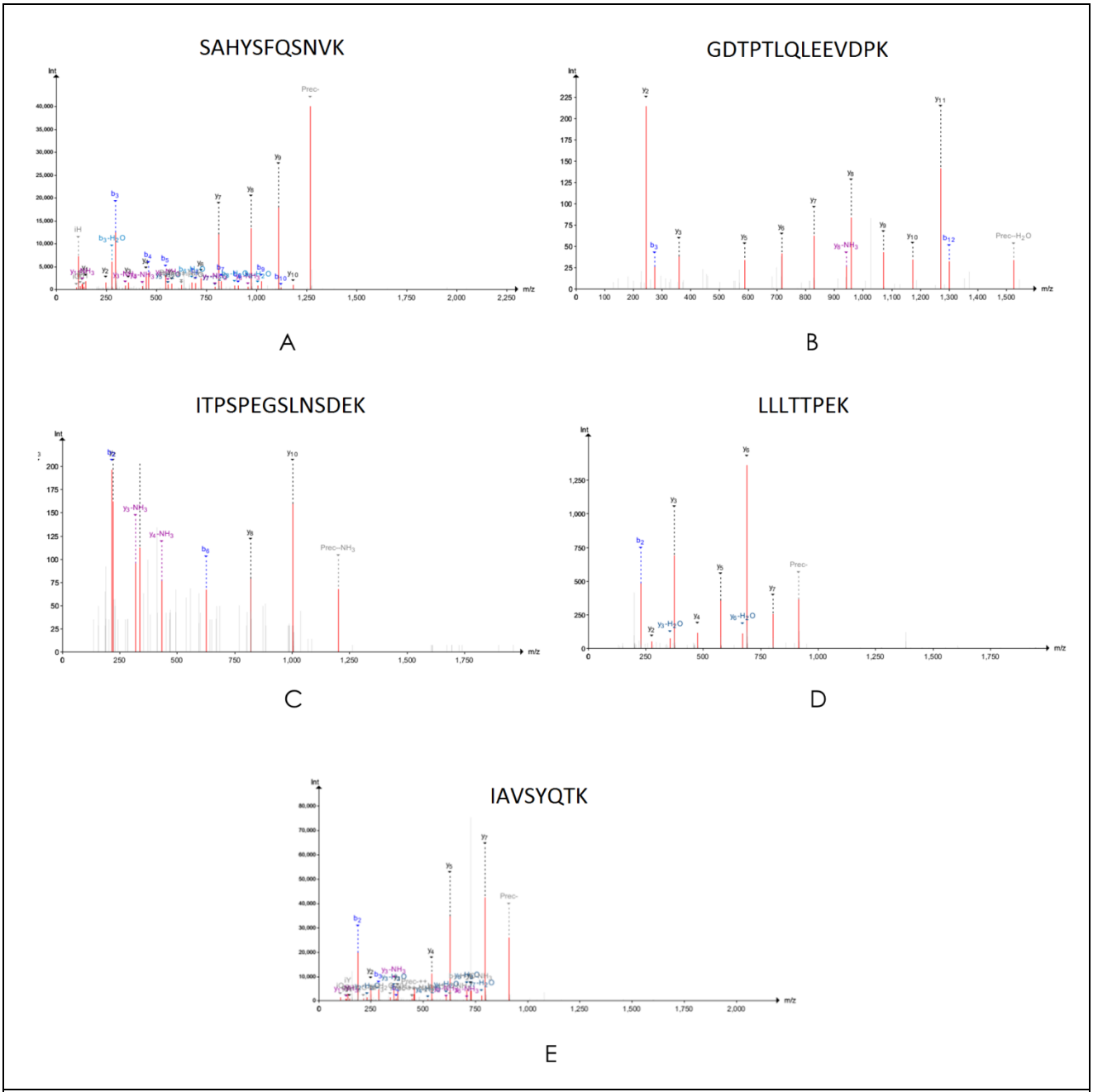


Figure A6.3. Representative MS/MS spectra of one unique *in silico* predicted and synthesised peptide each of five proinflammatory cytokines present in the peptide spectral library as displayed in PeptideShaker Software.

APPENDIX 6.4. The UniProtKB entries of 564 proteins identified by ProteinPilot™ Software in the peptide spectral library.

P14639, W5PWE9, W5PH95, W5PF65, P29701, W5NX51, W5NSA6, W5PSQ7, W5NXW9, W5QDG7, W5NRI1, W5PTG9, W5NPK5, W5QCY8, Q28743, W5P0Q4, Q1A2D1, D6PZY4, W5NQW4, W5PMR1, O46544, A2P2G4, W5PJ97, I1WXR3, W5PZS7, W5Q5H8, W5PBY0, W5NQ46, Q9XT27, W5P4S0, W5Q5A6, A6NBZ0, W5NRG7, W5NUJ7, W5QAB1, W5PHI7,

A0A0F6YFJ0, W5NY46, W5NWM2, W5Q7J0, P12303, W5Q0X5, W5P3R3, W5PID9, W5NRH2, P02083, W5NY95, F2YQ13, W5PTR4, C8BKD1, W5NVM6, W5P627, W5P101, W5NXP6, W5NZ47, W5PW21, W5P1J8, O02762, W5PHP7, W5PZF0, W5NTW3, W5NSH2, W5PFC9, W5NXM1, W5QAN8, W5Q268, W5QDP8, Q7M2U8, W5Q6D3, W5PI61, W5PGT6, W5PH81, W5Q6U2, W5QFP0, W5PKA9, W5P812, W5PFP1, W5Q6A0, W5PPQ8, W5Q7Z7, W5P5I0, W5P3J3, W5QGP4, W5QI29, W5PD71, W5NWX6, A0A0F6VY37, W5NUW3, W5NVB2, W5P338, W5NYJ9, W5QH54, W5NXP3, W5QH45, W5PXI3, W5NWU4, W5Q9A2, W5NXJ3, B7TJ06, W5PG63, W5PDS4, W5PE53, W5PDR5, W5PXC8, W5QH50, W5PTL2, Q29439, W5Q2K7, D7RIF5, W5Q4Q3, W5Q505, W5Q9D5, W5P7S6, W5NVG2, W5PN97, W5NS48, W5QA07, B5B304, W5Q961, W5Q2U8, W5PDQ9, W5PDP6, W5PGT9, W5PTU7, W5PAJ9, W5PDJ6, W5QA64, W5PZK7, W5Q6G0, P00922, W5QBW5, W5PXX3, W5Q0L2, W5Q754, W5P880, W5Q749, W5PCA0, W5PVL4, W5NPI9, W5PHP8, W5Q3K6, W5PJG0, W5Q9K1, W5PZ55, W5Q2E1, C8BKC5, W5NZH3, W5P988, W5PD62, W5Q3I7, W5PIA1, W5Q517, W5P4C6, W5P9V5, D7R7V6, W5PIN8, W5P229, W5PD84, W5PPT6, W5NRR7, C5ISA2, A2SW69, W5PEI4, W5PAB5, P49920, W5QAW7, A5YBU8, W5QGD1, O77642, Q06AV9, W5PYG2, W5QC41, W5QIK8, W5Q0R1, C5IS96, W5PI92, W5P9B0, W5P1X9, W5NYF4, W5P1W2, W5PJR0, W5PDN1, W5QFR6, M4WG34, W5NUS6, K4P494, W5P195, W5PZH5, W5PTS4, W5QGG0, W5QCY7, Q28579, A0A077JGJ6, W5PLL2, P02190, W5PSM6, W5QH21, W5PQD8, W5PJ66, P11839, W5QBD7, Q7JFW9, W5PB04, W5NUI0, W5PFJ0, C7EDS5, W5PB46, W5NX96, W5NPN4, A0A0U1YZ59, W5P2V3, W5NU34, W5PGS4, W5NV45, W5NRV6, W5NTG5, W5PLB7, W5PEY4, Q29400, W5P5W9, W5Q4Z3, W5QAA3, W5PHA3, W5P9J8, W5Q7T8, W5Q5L6, W5PJ69, Q06435, W5NQP5, W5NTT7, C0LQH2, W5P4W8, W5Q2U7, W5Q038, M4WGF1, W5P430, D6BJI0, W5PV69, W5NXN8, W5QET9, W5P915, W5PUC1, W5Q0V2, W5PVM3, W5QA54, W5PUU8, W5PYX7, W5PJR5, P67976, W5PBS4, W5NY50, D6PX64, W5NX91, W5Q224, W5PV74, W5PH03, W5NSF6, W5QFL0, W5QFR8, W5PC09, P62262, W5PTR5, W5P3A7, W5PLF8, D5HKJ4, W5P565, W5PQ96, W5PES2, W5PQH2, W5NSR4, W5PH45, W5PD87, Q5MIB5, P09670, W5NXM6, W5PVH9, W5Q0Y3, W5PZT3, W5NZQ2, W5Q723, W5NSZ2, W5Q9K6, W5NS65, A9YUY8, W5QDM2, W5PSZ5, W5P082, W5NYP8, W5PCP3, W5P9W4, W5NWH4, W5PDG4, W5NZW5, P68253, B0LRN3, W5PTZ9, W5P448, W5QEL6, W5P663, W5QJ27, W5NUU7, W5Q7C7, W5NS93, W5PSC8, W5NRA9, W5NTB3, W5PLV2, W5QE46, B3GS77, W5PIC9, W5PTE9, W5QCP9, W5P5V0, W5PEM0, W5PG55, W5QIV5, W5PWY0, W5QIW1, W5QE21, W5P8T8, W5QJB0, W5QAI5, W5PQK6, W5Q8B1, W5PPG6, W5QIG9, P49929, W5PZ65, W5PDF4, W5PBM4, W5QBV7, C5IJA0, W5PB07, W5PIG7, W5P481, W5P094, W5NZU3, W5PJC2, A8DR93, W5PMH6, W5P323, W5P044, W5PKC9, W5P3Q3, W5PJA0, W5QD30, W5P9U4, W5NT47, W5QGV5,

W5P3H8, W5P491, W5P620, W5PBN5, W5QHS2, W5P671, W5QGX2, W5P9V7, W5PEP7, W5PQA8, W5P0B2, W5PIN6, W5P0I5, W5NQ91, W5Q2A8, W5PZJ5, W5P311, W5NUP9, W5PAM4, W5PPN8, W5QBT3, W5NV73, W5P742, W5PG95, W5PYQ3, W5PF71, W5PS88, W5PIJ5, W5NR63, W5P5K9, W5PG36, W5NYA7, W5P060, W5Q9P7, W5PER5, W5Q3T5, W5P352, W5Q6D6, W5PKQ5, W5PDD0, W5PZ59, W5NTD9, W5PJV7, W5P644, W5PR13, W5PA19, W5QE19, Q4LBE5, W5P1N0, W5P326, W5PL69, W5P0X0, W5PIC7, W5P7M5, W5QG17, W5Q0G8, W5Q809, W5P731, W5NUL9, W5PMA2, W5P0T8, W5NQK0, W5NZ75, W5QBR9, W5P5F4, W5NRL0, W5QID7, W5PTL7, W5NXZ1, W5PEB0, W5Q1R4, W5PA64, W5P2Y4, W5PTU9, Q9XSM0, W5PHT8, W5QF71, W5Q097, W5PXH5, W5Q5D7, W5Q9L9, W5QIM9, W5Q6B0, W5PQ53, W5NPH5, W5P1Y8, W5Q0L1, W5PCI1, W5PEL1, W5Q8J9, W5P7G7, W5QDT2, W5NW13, A5YBV0, P23383, W5PGZ8, W5P2M3, W5NY22, W5PRG8, W5PEJ5, W5P7B1, W5NUA9, W5PXL4, W5PMY0, W5PNM5, W5QGV4, W5NW78, W5Q0G0, W5NS18, W5PSL9, W5QI70, W5QIY3, W5NZJ3, W5NX11, W5Q670, W5QGM7, W5QIL2, W5P689, W5PYR8, W5QD37, W5Q759, W5NXK9, W5PNP1, W5PMI8, W5PLP6, B2LU28, W5PNI5, W5Q8Z1, W5QEV3, W5PAT6, W5NS44, W5P964, W5PGS0, W5PXB4, W5Q254, W5PP57, W5PBX6, W5PTS2, W5Q5D8, W5QGN4, W5PE89, W5P908, W5QAJ7, W5PBH6, W5P8V3, P50413, W5PVY5, W5PGA9, W5PU57, W5QAK3, W5PQT7, W5PVW3, W5PIW6, W5NYJ2, W5P5V2, W5QBV3, W5QGN6, W5PVX3, W5Q6L3, W5Q3D0, W5Q0U9, W5PCG5, W5QDG4, W5PM28, W5PHX1, W5P8G7, W5NWD0, W5P7E0, W5PUU0, W5QF94, W5QBE7, W5PW91, W5PKF3, W5PJH9, Q7M355, W5QG48, W5NR00, W5NSE9, W5NU20, W5PS94, W5NR56, W5PD49, W5PKZ0, W5P8Y7, W5NS27, W5NQD5, W5QH53, W5P2S9, W5P6T2, W5PAJ2, W5Q540, W5PWG6, W5PVG2, W5PGC2, W5PM15, W5PBC2, W5NSA5, W5QGB7, W5QBD4, W5PKQ7, W5PVV7, W5PWQ1, W5PDQ0, W5PTZ3, W5PCK5, W5P578, W5PBU5, W5PLC9, P62297, W5P0H0, W5NYM6, W5P3D1, W5PKF8, W5P691 and W5PMS1.

APPENDIX 6.5. The UniProtKB entries of 830 proteins identified by Mascot in the peptide spectral library.

W5Q7J0, P14639, W5NSA6, W5PF65, W5P6F4, W5Q754, W5NUX8, W5P5T4, W5NY95, W5P4S0, D6PZY4, W5P3R3, W5NRI1, W5NRG7, W5PTG9, F2YQ13, W5Q7Z7, W5NX51, W5QAB1, W5PW21, W5P8R7, C8BKD1, W5PTR4, W5NSH2, W5PGT6, W5P0Q4, W5P5I0, I1WXR3, W5PH81, W5PBY0, W5QH45, W5Q2U7, W5QH50, W5NXW9, W5P1J8, W5NWM2, W5NY46, W5QH54, W5PXX3, W5QFK2, W5NYJ9, W5PXI3, W5Q5C2, W5QH56, W5P101, W5PXU6, W5PI61, W5P336, W5Q268, W5PID9, W5PDR5, W5Q7R8, W5PFJ0, W5QI29, W5P2U2, W5PJZ1, W5P812, W5P3J3, W5Q749, W5PEI4, W5PKA9, W5QGG0, W5Q4Q3,

W5PZI1, W5Q9A2, W5P248, W5PXC8, W5QAX3, W5PHP8, W5QFP0, W5QH06, W5PQL6, W5PE53, W5Q0L2, W5PGT9, W5NRR7, W5Q4B1, W5P691, Q1KYZ7, W5PD71, W5Q9D5, W5P7S6, Q06AV9, W5Q038, W5Q224, W5QFE4, P12303, W5PDE5, W5QDP8, W5QCP9, W5P6U4, Q28743, W5P229, W5P880, W5PWW5, W5QGP4, W5Q2E1, W5P4C6, W5PV54, A2SW69, W5Q4P0, W5PPQ8, W5PJ97, W5Q0R1, W5PPN5, W5QJ00, W5NVN1, W5P2I4, W5QHH3, W5Q961, W5PVL4, W5QAH2, W5QAA3, W5PHH3, W5NX96, W5QGD1, W5PD84, W5Q7T8, W5P3H8, W5PDC8, W5PIK2, W5QC26, A2P2H1, W5Q3K6, W5PDP6, W5PIC9, W5PZY7, W5PVH9, W5P9B0, W5QFL0, W5NS65, W5PAB5, W5PG63, W5PGS0, W5PD12, W5QCP0, W5P9V7, W5Q4W5, W5QJ69, W5QDH9, W5PKV4, W5NTL7, W5NYG1, W5NQJ8, W5P7L5, W5PDQ9, W5Q4Z3, W5PYG2, W5PDS4, W5PLL0, W5QBW5, W5QH21, W5QAR2, W5PZF0, W5PVG5, I0CF13, W5NPN4, W5PX97, W5PXI0, W5PES2, W5NTD9, W5PJ85, W5PQM7, W5PEL1, W5NSV3, W5PJE4, W5NUU7, W5P640, W5PP29, W5PWT9, W5Q2D7, W5PXN1, W5Q1W4, W5Q620, W5PYL5, W5PAE2, W5PME8, W5PJP9, W5PG04, W5PG50, W5PZK2, W5P0B2, W5PAM4, W5PY75, W5PKN2, A0MPT5, W5NZ47, A0A0M4KDI9, W5NYF4, W5PTU7, W5Q0V2, W5PSM6, W5PB46, W5P2V3, W5Q5A6, W5PB04, W5PPT6, W5PZ65, W5QG19, W5QCX2, W5P149, W5P895, W5PC09, W5NS93, W5P964, W5PYQ3, W5NZQ2, W5PG36, W5QAA1, W5PN97, W5PS45, W5PTL2, W5P9W4, W5NTQ9, W5PDY2, W5P3A7, W5NUG0, W5NQ85, W5PDQ6, W5Q9L9, W5NXR3, W5PEJ5, W5Q2R6, W5NY68, W5Q828, W5PVG2, W5QDW7, W5P092, W5Q9P0, W5QBT8, W5NX74, W5PHI3, W5P041, W5PA54, W5PIG7, W5Q8Y3, W5Q5F0, W5NT24, W5NWX6, W5PIJ5, W5QIK8, W5PGW2, W5NQP5, D6PX64, W5NXI6, W5Q5H8, W5NVT0, W5P988, W5QIW7, W5P9V5, W5PGS4, W5P2Y4, W5P9U4, W5QF95, W5PVX0, W5Q9K6, W5PDN1, W5PD15, K4P1S5, W5NPK4, W5QFP2, W5QET9, W5PMY0, W5P8Y7, W5PHA9, W5Q910, W5PUG1, W5PM86, W5QHD7, B2ZA84, W5PZ27, W5Q9B7, W5PA64, W5PW62, W5PZH5, W5P8T8, W5PD75, W5P143, W5NV45, W5PL69, W5PUJ4, W5PWU1, W5NUL7, W5Q370, W5Q7C0, W5PWL1, W5PLJ9, W5P0P1, W5PQ53, W5PDQ3, W5PH55, W5PDG3, W5P808, W5QC28, W5PT48, W5QBN7, W5PPJ4, W5Q950, W5P1Y8, B3SV56, W5P3G6, W5QGW8, W5NX89, W5QCH5, W5NSI7, U3N1L1, W5PA78, W5NRV1, W5PNU9, W5NVC1, W5P1Q0, Q7M371, W5P2L1, W5PHK5, W5Q9H4, W5NXV2, W5P9A6, W5QA54, W5PG55, W5PD62, W5P9C3, W5PAJ9, W5PJ66, W5PTE9, W5NSZ2, W5QD30, W5Q517, W5P4X7, W5P7X3, W5QHC2, W5QBV7, W5QE21, W5QIY3, W5QG16, W5PKW9, W5PQH2, W5PXV7, Q29400, W5PWY0, A8DR93, W5QH43, W5PPT3, W5P1D3, W5P4R6, D0VWZ0, W5P987, W5PV57, B3GS77, W5PGZ8, W5Q6L3, W5PQH3, W5QD80, D4P8S5, W5P5V0, W5P985, W5Q0N1, W5NZB4, W5NT03, W5Q5W3, W5PX18, A0A0P0QND2, W5QIM3, W5PZJ0, C5ISA2, W5QI75, W5PC99, W5NQM5, W5PDJ6, W5Q5P5, W5P1N7, W5PLC4, W5PK74, W5Q564, W5QE19, W5P1X7, W5Q0U9, W5PKK4,

W5PQ63, W5NW67, W5PDT4, W5Q5R2, W5NSH6, Q5XUI1, W5PEY4, W5PZW5, W5PFR8, W5NQ46, Q1KYZ6, W5PHP7, W5NTW3, Q28745, P02075, W5QDG7, W5NQW4, O46544, A5YBU9, W5NPK5, W5NRH2, W5P1L7, P30035, W5PIA1, W5PZ55, P68116, C8BKC5, W5Q9H1, W5QA07, O02762, W5P3Q3, W5NY78, K4P494, B5B304, W5QA64, B0BL71, W5Q5V4, W5NSC7, W5Q293, W5NXP3, W5PHI7, A5YBU8, A2P2G1, W5Q3I7, W5PXZ3, P68214, A3QP67, W5NYG7, W5QHS2, W5QD06, W5Q0D7, W5PME2, W5Q124, W5PWE9, Q9XT27, W5PZS7, W5P627, W5NU00, Q7M2U8, W5Q0X5, W5NUJ7, Q1A2D1, W5NQ45, W5NVB2, W5P060, Q29439, P09670, P49920, W5NR06, C5IS96, W5Q369, W5P9J8, W5Q226, A4ZVY6, W5NUW3, W5PXE2, W5P620, W5PBX1, W5P9L8, W5PTB4, W5PKH2, W5NYP8, W5PBN5, P0C276, W5NTB3, W5Q2A5, W5PI10, W5PD87, W5NR63, W5PEC1, W5P881, W5PHG0, W5Q0Y3, W5NQ51, W5PNI5, W5NQ38, W5QF60, W5NUS6, W5Q367, W5NTL6, B0LRN3, W5P5K9, W5P4K7, W5P8L9, W5NVM6, W5PI92, W5PFI7, W5Q5D7, W5PP57, W5P8S5, W5NYA7, W5P671, W5PPK7, W5P2D1, C0LQH2, Q28893, W5PE96, W5NZW9, W5Q0C0, W5PQQ8, W5Q1N4, W5Q3J3, W5PGV3, W5Q5I4, W5PLU0, W5PZM1, W5QJ31, W5NQB8, W5Q5V9, W5P596, W5QEF4, W5Q7H9, W5PZQ8, W5NPR6, W5PKP8, W5PJ43, W5P4E6, W5Q9T3, W5PSI1, W5PD73, R4R2H5, D6BJI0, W5PMS1, W5NQ08, W5PNY6, W5PTU9, W5Q417, W5NYR7, W5P0C1, P32237, W5PY00, W5Q4E1, W5P078, W5NVG4, W5PY15, W5P4S8, W5QGN6, W5PZR6, H2DGR2, W5QA34, W5Q8B1, D5HKJ4, W5NVW9, D0VWY9, W5PLN7, W5NSR4, W5Q8S4, W5PZU0, W5PFY4, W5NRZ7, W5PIN4, W5P094, W5PAI7, W5NST9, W5PM69, W5PL67, W5PJR0, W5PCA0, W5Q4U7, A2VBC7, I3RWJ0, W5P673, W5PF26, Q9XSM0, B9VGZ6, W5NVY5, W5P6L3, W5PL06, W5PMF4, W5P736, W5PIU1, W5PFD3, W5NSB9, W5QJA2, W5NU34, W5NY22, W5PV93, P50413, W5Q9T2, W5Q7I2, W5PH95, W5PTS4, W5PZI0, W5Q6R7, W5PLB7, W5NQW9, W5Q9K1, W5NVG2, W5QH46, W5NZH3, W5QC41, W5QCY7, W5Q0F3, W5NS74, W5PEL7, W5P195, W5QC38, Q5MIB5, C7EDS5, W5NTJ3, A0A077JGJ6, W5PE90, P22793, W5Q5L6, W5NTE3, W5PT09, W5NPI9, W5QIV5, W5PF71, W5PV74, W5PAG8, W5NY50, W5PDF4, W5PN84, W5QDM2, P62262, W5NUY7, W5PH03, W5PVM3, W5PSC8, W5PQK6, W5PPG6, Q1RPQ7, W5PMH6, W5Q2K6, W5PT68, W5QFP3, W5PT38, W5PXV3, W5PJC2, W5NQ22, H9CJU6, W5NT47, W5PHT8, W5NU40, W5PZT3, D3IU20, W5PLQ1, W5NSF6, W5QE46, W5PK67, W5NRP7, W5QCK9, W5PGT0, W5PSM2, W5PSP1, W5PWH2, W5QB36, W5PKG7, W5P2R3, W5P5W9, W5QE14, W5PLV2, W5PC67, W5QDT2, W5PQD8, W5PXA3, W5P915, W5PBG2, K4P231, W5NYX4, W5PPZ8, W5PIF0, W5NXZ4, W5PBM4, W5QBD7, W5Q670, W5PBX6, W5QIL2, W5PSX0, W5P657, W5P1Y5, W5QA91, W5NX11, W5PEM0, W5NRF4, W5QIG9, W5Q3N2, W5PMZ6, W5P0X0, W5PCQ3, W5QDP7, W5PB37, W5P300, W5NRS4, W5NW52, W5QGM9, W5PIV4, W5PRK9, W5NX57, W5PGV9, W5Q6W6, W5QGQ0, W5NZR2, W5NXK6, W5P7W5,

W5NZU3, W5NQ21, W5PZC0, W5PHE6, W5PN22, W5P0W5, B5DC89, W5PI70, W5QIW1, W5NZJ3, D7RH37, W5PEP7, W5P1U6, W5PYK0, W5PZ59, W5QCG2, W5PFW3, M4T8F1, W5QGM7, W5PNG2, W5PCR1, W5Q9P7, W5PI34, W5Q0L1, W5P1T7, W5Q043, W5QA42, W5NW86, W5PLB4, W5PJA0, W5PB07, W5NY84, W5PBY1, W5PBS4, W5Q334, W5P377, W5Q5H7, W5PVU5, W5PD49, W5QDT9, W5PWR1, W5NRA9, Q28884, W5PTS0, W5PM28, W5QF71, W5PLB6, C5IJA0, W5QDN3, Q7M355, W5Q6C4, W5PAI3, W5PA65, W5PK97, W5PYT3, W5PR03, W5P148, A0A0A7ETA6, W5PHI8, W5PW49, W5P7I2, W5P393, W5Q1J8, W5Q8M7, W5Q5C7, W5QI89, M4WG34, W5PR13, W5PG12, W5PX41, W5NYZ0, W5Q6W3, W5Q1R5, W5PJE0, W5NYX8, P02083, W5P1W2, W5P1X9, P42819, P02190, W5PUC1, W5QDF3, W5Q723, A4ZVY8, W5PJ69, W5NXM6, W5Q5L3, P67976, W5NTT7, D7RIF5, A2I7L0, W5NXJ3, W5P082, A0A0U1YZ59, W5PS88, W5PGC9, P07846, W5QAH1, W5NU63, W5PKC9, W5QFQ0, W5P481, A2P2I3, W5NVS8, W5Q684, W5PJG0, W5P323, P17607, A9YUY8, W5P656, A0A0H3V7A0, W5QD47, W5P044, W5QJ62, W5QAM3, W5PLC9, W5QIV1, W5PIW6, W5PYI1, W5PVY5, W5NQ83, W5QDH3, W5PZ95, W5PX84, Q30DP7, W5PIF6, W5NQ91, W5NTZ6, W5PGG9, W5Q4B0, W5NRC0, W5PSB1, W5QDR0, W5QGQ3, W5Q627, W5P525 and W5NYJ2.

APPENDIX 6.6. The UniProtKB entries of 398 proteins identified by PeptideShaker in the Peptide spectral library.

W5NSA6, W5Q7J0, W5NX51, W5PF65, W5NRI1, W5P6F4, W5NRG7, W5NSH2, W5PHP7, W5NQW4, W5QDG7, W5NWM2, W5NY46, W5PW21, W5PTG9, W5NTW3, W5Q124, W5P0Q4, W5PJZ1, W5QAB1, P32262, W5NXW9, P20757, W5PID9, W5Q7Z7, W5PGT6, W5P5I0, W5Q4Q3, W5P1J8, W5QH54, W5QH50, W5PFC9, W5Q0X5, W5P8R7, W5PXC8, W5PKA9, W5P101, P29701, C8BKD1, W5Q749, W5PJ97, W5Q9A2, W5Q7R8, W5PD71, W5NRH2, W5PDR5, W5PE53, W5Q0L2, W5QI29, W5NPK5, W5PGT9, P12303, W5PXI3, W5P3J3, W5QFP0, P50450, W5PHI7, A2P2I0, W5PH81, W5P336, W5Q2E1, W5P7S6, W5QGG0, W5NXM1, W5NXP3, W5Q9D5, W5P9B0, W5PD84, A2SW69, W5Q0R1, W5PV54, W5P812, W5Q2U7, W5Q268, W5P4S0, W5PPQ8, O46544, W5PDQ9, W5QDP8, W5PHP8, W5PHH3, D6PZY4, W5PEI4, Q29439, W5PDE5, W5QGD1, W5P229, W5Q3K6, W5PXX3, W5PDS4, W5NWX6, W5QGP4, Q1KYZ7, W5PAB5, W5Q5H8, W5NUU7, W5Q0V2, W5Q4Z3, W5PBY0, W5PYG2, W5NS65, W5QE21, W5QCP9, W5NYF4, W5Q961, W5PDP6, W5NQP5, A0A0U1YZ59, W5NX96, W5P627, W5QIK8, W5Q5A6, W5NXP6, W5P2V3, W5NXJ3, W5QBW5, W5P060, W5Q038, W5QAR2, W5PQH0, W5P323, W5NS93, W5P880, W5P640, B6EBS6, W5P988, O77642, W5P9U4, P14639, W5QIM3, W5PIJ5, W5P4C6, W5P2Y4, W5PLF8,

W5PDN1, W5Q6A0, W5NUJ7, W5P3R3, P02075, W5NQ46, W5PZI1, W5PIA1, W5QH45, A5YBU9, B6UV62, W5P9V5, W5NZ47, P30035, W5QA64, Q7M2U8, P68056, P00922, W5QA07, B3GS77, W5PZ55, W5PJ66, W5NU00, W5PDJ6, W5NYA1, W5QA54, B0BL71, F2YQ13, W5PSQ7, Q9XT27, W5PTU7, W5PI61, C8BKC5, W5QH46, A2P2H1, W5P9J8, W5PN97, W5PLL0, P49920, C5IS96, W5PD87, P0C276, W5PG63, W5PD62, W5P430, W5PZS7, D7RIF5, W5PI92, W5PAJ9, W5NYA7, W5P671, W5NU34, W5NTB3, W5P5K9, D6PX64, W5PJR0, W5PES2, W5PSM6, W5QDR0, W5Q0Y3, W5PIW6, A3QP67, Q1A2D1, W5QBV7, W5NR63, W5Q2A5, W5PVL4, W5Q419, W5PJH9, P50413, W5Q0G0, W5PM28, W5NVT0, W5QA34, W5PQS3, W5NUS6, W5PZA0, W5P8T8, Q9XSM0, W5PEW5, W5PWU7, W5Q0U9, B5B304, W5NR06, K4P494, W5QE46, W5PLQ1, W5NQ08, W5QCX2, W5PMS1, W5Q2N2, W5NSZ2, W5NZ71, W5QEC3, W5NWL0, P68116, W5PZI0, P52210, W5PTS4, W5QDG8, W5NQW9, W5Q687, W5NVG2, W5Q9K1, W5Q6L8, W5QC41, W5NTJ3, W5NZH3, W5Q6D3, W5PB46, C5ISA2, W5NS74, W5Q517, Q70TH4, W5Q0F3, W5Q3I7, Q5MIB5, W5PEY4, D7R7V6, W5PV74, W5NS48, W5Q5L6, W5PZH5, W5QH21, W5PN84, W5NPN4, W5PT38, W5NY50, W5PDF4, W5P195, W5PH03, W5PEL7, W5PMH6, W5QET9, W5PF71, W5PLV2, W5QDM2, W5PLB7, W5Q224, A0A077JGJ6, W5PVM3, W5PSC8, W5QA42, P22793, W5Q6E8, W5PUG1, A9YUY8, W5PTR5, W5NV45, W5PZ59, W5NZQ2, W5NUI0, W5PPT6, W5PJC2, P79360, W5PQD8, W5QD30, W5QB36, W5PB07, A9P323, W5PBM4, W5Q1C7, W5PZT3, P68253, P11839, K4P231, P29361, W5Q5D7, W5NSF6, W5PMY0, A5YBU8, W5NSB0, W5PZ65, W5Q888, W5PWU4, W5P0E1, W5QHL7, W5PUJ4, W5PEM0, W5Q5Z3, Q6TMG6, W5NQ21, W5QEL6, W5QDT2, W5P352, W5QIW1, W5QAJ7, W5PG36, W5PEP7, W5NSR5, W5QGM7, W5NRL0, P62262, W5PLN8, W5QIL2, W5PTE9, W5PBS4, W5QE19, W5QEV3, W5PZF7, W5PB04, W5PJA0, W5QIY3, W5P9W4, W5PTL2, P50415, W5P7G7, W5PS94, W5PSZ4, W5QHS2, W5P8W6, W5NZU3, W5NRA9, W5PHI4, W5QDS5, W5PDG5, P68251, C8BKD8, W5NX11, W5QAI5, W5PFI7, W5QFS0, W5NS09, W5NXZ1, W5PYQ3, W5P0X0, W5PW91, W5Q972, W5PH51, W5PDV8, W5P5S5, W5PTS2, W5NY78, C5IJA0, P21621, P42819, W5Q1R5, P29455, Q28579, W5P0H0, W5PJG0, P23383, W5P1X9, P02190, W5NTT7, W5PUC1, W5PJ69, W5NXM6, W5QDF3, W5PS88, W5PJR5, W5P5W9, W5PLC9, Q4TVY4, W5P481, W5Q7T8, H9CJU6, W5QFQ0, W5PK04, W5NV73, C0LQH2, W5PPG6, A0A0H3V7A0, and W5Q723.

APPENDIX 6.7. The UniProtKB entries of 1,103 proteins in the peptide spectral library from a combined identification by ProteinPilot, Mascot and PeptideShaker.

P14639, W5PF65, W5NX51, W5NSA6, W5NXW9, W5QDG7, W5NRI1, W5PTG9, W5NPK5, W5P0Q4, Q1A2D1, D6PZY4, W5NQW4, O46544, W5PJ97, W5PZS7, W5Q5H8, W5PBY0, W5NQ46, Q9XT27, W5P4S0, W5Q5A6, W5NRG7, W5NUJ7, W5QAB1, W5PHI7, W5NY46, W5NWM2, W5Q7J0, P12303, W5Q0X5, W5P3R3, W5PID9, W5NRH2, F2YQ13, C8BKD1, W5P627, W5P101, W5NZ47, W5PW21, W5P1J8, W5PHP7, W5NTW3, W5NSH2, W5Q268, W5QDP8, Q7M2U8, W5PI61, W5PGT6, W5PH81, W5QFP0, W5PKA9, W5P812, W5PPQ8, W5Q7Z7, W5P5I0, W5P3J3, W5QGP4, W5QI29, W5PD71, W5NWX6, W5QH54, W5NXP3, W5QH45, W5PXI3, W5Q9A2, W5NXJ3, W5PG63, W5PDS4, W5PE53, W5PDR5, W5PXC8, W5QH50, W5PTL2, Q29439, D7RIF5, W5Q4Q3, W5Q9D5, W5P7S6, W5NVG2, W5PN97, W5QA07, B5B304, W5Q961, W5PDQ9, W5PDP6, W5PGT9, W5PTU7, W5PAJ9, W5PDJ6, W5QA64, W5QBW5, W5PXX3, W5Q0L2, W5P880, W5Q749, W5PVL4, W5PHP8, W5Q3K6, W5PJG0, W5Q9K1, W5PZ55, W5Q2E1, C8BKC5, W5NZH3, W5P988, W5PD62, W5Q3I7, W5PIA1, W5Q517, W5P4C6, W5P9V5, W5P229, W5PD84, W5PPT6, C5ISA2, A2SW69, W5PEI4, W5PAB5, P49920, A5YBU8, W5QGD1, W5PYG2, W5QC41, W5QIK8, W5Q0R1, C5IS96, W5PI92, W5P9B0, W5P1X9, W5NYF4, W5PJR0, W5PDN1, W5NUS6, K4P494, W5P195, W5PZH5, W5PTS4, W5QGG0, A0A077JGJ6, P02190, W5PSM6, W5QH21, W5PQD8, W5PJ66, W5PB04, W5PB46, W5NX96, W5NPN4, A0A0U1YZ59, W5P2V3, W5NU34, W5NV45, W5PLB7, W5PEY4, W5P5W9, W5Q4Z3, W5P9J8, W5Q7T8, W5Q5L6, W5PJ69, W5NQP5, W5NTT7, C0LQH2, W5Q2U7, W5Q038, W5QET9, W5PUC1, W5Q0V2, W5PVM3, W5QA54, W5PBS4, W5NY50, D6PX64, W5Q224, W5PV74, W5PH03, W5NSF6, P62262, W5PES2, W5PD87, Q5MIB5, W5NXM6, W5Q0Y3, W5PZT3, W5NZQ2, W5Q723, W5NSZ2, W5NS65, A9YUY8, W5QDM2, W5P9W4, W5NUU7, W5NS93, W5PSC8, W5NRA9, W5NTB3, W5PLV2, W5QE46, B3GS77, W5PTE9, W5QCP9, W5PEM0, W5QIW1, W5QE21, W5P8T8, W5PPG6, W5PZ65, W5PDF4, W5PBM4, W5QBV7, C5IJA0, W5PB07, W5P481, W5NZU3, W5PJC2, W5PMH6, W5P323, W5PJA0, W5QD30, W5P9U4, W5QHS2, W5P671, W5PEP7, W5PYQ3, W5PF71, W5PS88, W5PIJ5, W5NR63, W5P5K9, W5PG36, W5NYA7, W5P060, W5PZ59, W5QE19, W5P0X0, W5P2Y4, Q9XSM0, W5Q5D7, W5QDT2, W5PMY0, W5QIY3, W5NX11, W5QGM7, W5QIL2, P50413, W5PIW6, W5Q0U9, W5PM28, W5PLC9, W5PMS1, W5P6F4, W5P8R7, W5P336, W5Q7R8, W5PJZ1, W5PZI1, Q1KYZ7, W5PDE5, W5PV54, W5PHH3, A2P2H1, W5PLL0, W5QAR2, W5P640, W5QCX2, W5NVT0, W5PUG1, W5PUJ4, W5QIM3, P02075, A5YBU9, P30035, P68116, W5NY78, B0BL71, A3QP67, W5Q124, W5NU00, W5NR06, P0C276, W5Q2A5, W5PFI7, W5NQ08, W5QA34, W5PZI0, W5NQW9, W5QH46,

W5Q0F3, W5NS74, W5PEL7, W5NTJ3, P22793, W5PN84, W5PT38, H9CJU6, W5PLQ1, W5QB36, K4P231, W5NQ21, W5QA42, W5Q1R5, P42819, W5QDF3, W5QFQ0, A0A0H3V7A0, W5QDR0, P29701, W5PSQ7, W5NXP6, W5PFC9, W5NXM1, W5Q6D3, W5Q6A0, W5NS48, P00922, D7R7V6, O77642, Q28579, P11839, W5NUI0, W5P430, W5PJR5, W5PTR5, W5PLF8, P68253, W5QEL6, W5QAI5, W5NV73, W5P352, W5NRL0, W5NXZ1, W5P7G7, P23383, W5Q0G0, W5QEV3, W5PTS2, W5QAJ7, W5PW91, W5PJH9, W5PS94, W5P0H0, W5PWE9, W5PH95, Q28743, I1WXR3, P02083, W5NY95, W5PTR4, W5NVM6, O02762, W5PZF0, W5NUW3, W5NVB2, W5NYJ9, W5Q754, W5PCA0, W5NPI9, W5NRR7, Q06AV9, W5P1W2, M4WG34, W5QCY7, W5QBD7, W5PFJ0, C7EDS5, W5PGS4, Q29400, W5QAA3, D6BJI0, W5P915, P67976, W5QFL0, W5PC09, W5P3A7, D5HKJ4, W5PQH2, W5NSR4, P09670, W5PVH9, W5Q9K6, W5P082, W5NYP8, B0LRN3, W5PIC9, W5P5V0, W5PG55, W5QIV5, W5PWY0, W5PQK6, W5Q8B1, W5QIG9, W5PIG7, W5P094, A8DR93, W5P044, W5PKC9, W5P3Q3, W5NT47, W5P3H8, W5P620, W5PBN5, W5P9V7, W5P0B2, W5NQ91, W5PAM4, W5Q9P7, W5NTD9, W5PR13, W5PL69, W5PA64, W5PTU9, W5PHT8, W5QF71, W5Q9L9, W5PQ53, W5P1Y8, W5Q0L1, W5PEL1, W5PGZ8, W5NY22, W5PEJ5, W5NZJ3, W5Q670, W5PNI5, W5P964, W5PGS0, W5PP57, W5PBX6, W5PVY5, W5NYJ2, W5QGN6, W5Q6L3, Q7M355, W5PD49, W5P8Y7, W5PVG2, W5P691, P32262, P20757, P50450, A2P2I0, W5PQH0, B6EBS6, B6UV62, P68056, W5NYA1, W5Q419, W5PQS3, W5PZA0, W5PEW5, W5PWU7, W5Q2N2, W5NZ71, W5QEC3, W5NWL0, P52210, W5QDG8, W5Q687, W5Q6L8, Q70TH4, W5Q6E8, P79360, A9P323, W5Q1C7, P29361, W5NSB0, W5Q888, W5PWU4, W5P0E1, W5QHL7, W5Q5Z3, Q6TMG6, W5NSR5, W5PLN8, W5PZF7, P50415, W5PSZ4, W5P8W6, W5PHI4, W5QDS5, W5PDG5, P68251, C8BKD8, W5QFS0, W5NS09, W5Q972, W5PH51, W5PDV8, W5P5S5, P21621, P29455, Q4TVY4, W5PK04, W5NUX8, W5P5T4, W5QFK2, W5Q5C2, W5QH56, W5PXU6, W5P2U2, W5P248, W5QAX3, W5QH06, W5PQL6, W5Q4B1, W5QFE4, W5P6U4, W5PWW5, W5Q4P0, W5PPN5, W5QJ00, W5NVN1, W5P2I4, W5QHH3, W5QAH2, W5PDC8, W5PIK2, W5QC26, W5PZY7, W5PD12, W5QCP0, W5Q4W5, W5QJ69, W5QDH9, W5PKV4, W5NTL7, W5NYG1, W5NQJ8, W5P7L5, W5PVG5, I0CF13, W5PX97, W5PXI0, W5PJ85, W5PQM7, W5NSV3, W5PJE4, W5PP29, W5PWT9, W5Q2D7, W5PXN1, W5Q1W4, W5Q620, W5PYL5, W5PAE2, W5PME8, W5PJP9, W5PG04, W5PG50, W5PZK2, W5PY75, W5PKN2, A0MPT5, A0A0M4KDI9, W5QG19, W5P149, W5P895, W5QAA1, W5PS45, W5NTQ9, W5PDY2, W5NUG0, W5NQ85, W5PDQ6, W5NXR3, W5Q2R6, W5NY68, W5Q828, W5QDW7, W5P092, W5Q9P0, W5QBT8, W5NX74, W5PHI3, W5P041, W5PA54, W5Q8Y3, W5Q5F0, W5NT24, W5PGW2, W5NXI6, W5QIW7, W5QF95, W5PVX0, W5PD15, K4P1S5, W5NPK4, W5QFP2, W5PHA9, W5Q910, W5PM86, W5QHD7, B2ZA84, W5PZ27, W5Q9B7, W5PW62, W5PD75, W5P143, W5PWU1, W5NUL7, W5Q370, W5Q7C0, W5PWL1,

W5PLJ9, W5P0P1, W5PDQ3, W5PH55, W5PDG3, W5P808, W5QC28, W5PT48, W5QBN7, W5PPJ4, W5Q950, B3SV56, W5P3G6, W5QGW8, W5NX89, W5QCH5, W5NSI7, U3N1L1, W5PA78, W5NRV1, W5PNU9, W5NVC1, W5P1Q0, Q7M371, W5P2L1, W5PHK5, W5Q9H4, W5NXV2, W5P9A6, W5P9C3, W5P4X7, W5P7X3, W5QHC2, W5QG16, W5PKW9, W5PXV7, W5QH43, W5PPT3, W5P1D3, W5P4R6, D0VWZ0, W5P987, W5PV57, W5PQH3, W5QD80, D4P8S5, W5P985, W5Q0N1, W5NZB4, W5NT03, W5Q5W3, W5PX18, A0A0P0QND2, W5PZJ0, W5QI75, W5PC99, W5NQM5, W5Q5P5, W5P1N7, W5PLC4, W5PK74, W5Q564, W5P1X7, W5PKK4, W5PQ63, W5NW67, W5PDT4, W5Q5R2, W5NSH6, Q5XUI1, W5PZW5, W5PFR8, Q1KYZ6, Q28745, W5P1L7, W5Q9H1, W5Q5V4, W5NSC7, W5Q293, A2P2G1, W5PXZ3, P68214, W5NYG7, W5QD06, W5Q0D7, W5PME2, W5NQ45, W5Q369, W5Q226, A4ZVY6, W5PXE2, W5PBX1, W5P9L8, W5PTB4, W5PKH2, W5PI10, W5PEC1, W5P881, W5PHG0, W5NQ51, W5NQ38, W5QF60, W5Q367, W5NTL6, W5P4K7, W5P8L9, W5P8S5, W5PPK7, W5P2D1, Q28893, W5PE96, W5NZW9, W5Q0C0, W5PQQ8, W5Q1N4, W5Q3J3, W5PGV3, W5Q5I4, W5PLU0, W5PZM1, W5QJ31, W5NQB8, W5Q5V9, W5P596, W5QEF4, W5Q7H9, W5PZQ8, W5NPR6, W5PKP8, W5PJ43, W5P4E6, W5Q9T3, W5PSI1, W5PD73, R4R2H5, W5PNY6, W5Q417, W5NYR7, W5P0C1, P32237, W5PY00, W5Q4E1, W5P078, W5NVG4, W5PY15, W5P4S8, W5PZR6, H2DGR2, W5NVW9, D0VWY9, W5PLN7, W5Q8S4, W5PZU0, W5PFY4, W5NRZ7, W5PIN4, W5PAI7, W5NST9, W5PM69, W5PL67, W5Q4U7, A2VBC7, I3RWJ0, W5P673, W5PF26, B9VGZ6, W5NVY5, W5P6L3, W5PL06, W5PMF4, W5P736, W5PIU1, W5PFD3, W5NSB9, W5QJA2, W5PV93, W5Q9T2, W5Q7I2, W5Q6R7, W5QC38, W5PE90, W5NTE3, W5PT09, W5PAG8, W5NUY7, Q1RPQ7, W5Q2K6, W5PT68, W5QFP3, W5PXV3, W5NQ22, W5NU40, D3IU20, W5PK67, W5NRP7, W5QCK9, W5PGT0, W5PSM2, W5PSP1, W5PWH2, W5PKG7, W5P2R3, W5QE14, W5PC67, W5PXA3, W5PBG2, W5NYX4, W5PPZ8, W5PIF0, W5NXZ4, W5PSX0, W5P657, W5P1Y5, W5QA91, W5NRF4, W5Q3N2, W5PMZ6, W5PCQ3, W5QDP7, W5PB37, W5P300, W5NRS4, W5NW52, W5QGM9, W5PIV4, W5PRK9, W5NX57, W5PGV9, W5Q6W6, W5QGQ0, W5NZR2, W5NXK6, W5P7W5, W5PZC0, W5PHE6, W5PN22, W5P0W5, B5DC89, W5PI70, D7RH37, W5P1U6, W5PYK0, W5QCG2, W5PFW3, M4T8F1, W5PNG2, W5PCR1, W5PI34, W5P1T7, W5Q043, W5NW86, W5PLB4, W5NY84, W5PBY1, W5Q334, W5P377, W5Q5H7, W5PVU5, W5QDT9, W5PWR1, Q28884, W5PTS0, W5PLB6, W5QDN3, W5Q6C4, W5PAI3, W5PA65, W5PK97, W5PYT3, W5PR03, W5P148, A0A0A7ETA6, W5PHI8, W5PW49, W5P7I2, W5P393, W5Q1J8, W5Q8M7, W5Q5C7, W5QI89, W5PG12, W5PX41, W5NYZ0, W5Q6W3, W5PJE0, W5NYX8, A4ZVY8, W5Q5L3, A2I7L0, W5PGC9, P07846, W5QAH1, W5NU63, A2P2I3, W5NVS8, W5Q684, P17607, W5P656, W5QD47, W5QJ62, W5QAM3, W5QIV1, W5PYI1, W5NQ83, W5QDH3, W5PZ95, W5PX84, Q30DP7, W5PIF6, W5NTZ6, W5PGG9, W5Q4B0, W5NRC0, W5PSB1,

W5QGQ3, W5Q627, W5P525, W5QCY8, W5PMR1, A2P2G4, A6NBZ0, A0A0F6YFJ0, W5QAN8, W5Q6U2, W5PFP1, A0A0F6VY37, W5P338, W5NWU4, B7TJ06, W5Q2K7, W5Q505, W5Q2U8, W5PZK7, W5Q6G0, W5PIN8, W5QAW7, W5QFR6, W5PLL2, Q7JFW9, W5NRV6, W5NTG5, W5PHA3, Q06435, W5P4W8, M4WGF1, W5PV69, W5NXN8, W5PUU8, W5PYX7, W5NX91, W5QFR8, W5P565, W5PQ96, W5PH45, W5PSZ5, W5PCP3, W5NWH4, W5PDG4, W5NZW5, W5PTZ9, W5P448, W5P663, W5QJ27, W5Q7C7, W5QJB0, P49929, W5QGV5, W5P491, W5QGX2, W5PQA8, W5PIN6, W5P0I5, W5Q2A8, W5PZJ5, W5P311, W5NUP9, W5PPN8, W5QBT3, W5P742, W5PG95, W5PER5, W5Q3T5, W5Q6D6, W5PKQ5, W5PDD0, W5PJV7, W5P644, W5PA19, Q4LBE5, W5P1N0, W5P326, W5PIC7, W5P7M5, W5QG17, W5Q0G8, W5Q809, W5P731, W5NUL9, W5PMA2, W5P0T8, W5NQK0, W5NZ75, W5QBR9, W5P5F4, W5QID7, W5PTL7, W5PEB0, W5Q1R4, W5Q097, W5PXH5, W5QIM9, W5Q6B0, W5NPH5, W5PCI1, W5Q8J9, W5NW13, A5YBV0, W5P2M3, W5PRG8, W5P7B1, W5NUA9, W5PXL4, W5PNM5, W5QGV4, W5NW78, W5NS18, W5PSL9, W5QI70, W5P689, W5PYR8, W5QD37, W5Q759, W5NXK9, W5PNP1, W5PMI8, W5PLP6, B2LU28, W5Q8Z1, W5PAT6, W5NS44, W5PXB4, W5Q254, W5Q5D8, W5QGN4, W5PE89, W5P908, W5PBH6, W5P8V3, W5PGA9, W5PU57, W5QAK3, W5PQT7, W5PVW3, W5P5V2, W5QBV3, W5PVX3, W5Q3D0, W5PCG5, W5QDG4, W5PHX1, W5P8G7, W5NWD0, W5P7E0, W5PUU0, W5QF94, W5QBE7, W5PKF3, W5QG48, W5NR00, W5NSE9, W5NU20, W5NR56, W5PKZ0, W5NS27, W5NQD5, W5QH53, W5P2S9, W5P6T2, W5PAJ2, W5Q540, W5PWG6, W5PGC2, W5PM15, W5PBC2, W5NSA5, W5QGB7, W5QBD4, W5PKQ7, W5PVV7, W5PWQ1, W5PDQ0, W5PTZ3, W5PCK5, W5P578, W5PBU5, P62297, W5NYM6, W5P3D1 and W5PKF8.

APPENDIX 7.0. Suspected selective susceptibility to endotoxin in an ovine model

Accepted Author Manuscript

Online Journal of Veterinary Research, 2014, 18 12: 941-963.

Saul CHEMONGES

BVetMed, PGDipVetClinSt, MVSt, MANZCVS

Institutional affiliations:

1. *School of Veterinary Science, The University of Queensland, Gatton, QLD 4343, Australia.*
2. *Central Analytical Research Facility, Institute for Future Environments, Queensland University of Technology, Gardens Point, George Street, Brisbane QLD 4001, Australia*

Email: *s.chemonges@uq.edu.au*

Abstract

Recent observations show that some sheep appear to be more susceptible to the effects of *Escherichia coli* lipopolysaccharide (LPS) endotoxin than others despite being the same age and breed. This incidental observation made during a study with a translational benefit to human emergency medicine prompted further investigations focused on the background of the base model (sheep) and peri-experimental practices. A predetermined dose of 15 µg/kg of LPS from *E. coli* serotype O55:B5 was planned to be infused into a number of sheep to prime their immune system prior to blood transfusion studies. Some sheep subsequently received a lesser dose of LPS due to unexpected heighted susceptibility to endotoxin. It was hypothesised that genetic, environmental or managemental practices could have been contributing factors to this observation. Sheep that were more robust to endotoxin were raised in open pasture, whereas the more susceptible ones were essentially from a controlled scientific breeding colony. Epigenetic factors need to be considered when designing protracted large animal experiments as these aspects can influence the host's response to endotoxin challenge. It is suggested that an understanding of the proteogenomics of serum or plasma could help to understand LPS morbidity in sheep and similar pathology in other mammals, including humans.

KEY-WORDS: Selective resistance to endotoxin; Sheep; *Escherichia coli*; Lipopolysaccharide (LPS); disease challenge; Large animal models; serum and plasma proteogenomics.

Introduction

Recent reports indicate that some sheep appear to be more susceptible to the effects *Escherichia coli* lipopolysaccharide (LPS) endotoxin challenge than others, regardless of having the same characteristics[1,2]. This unexpected observation was made during the continuation of primary studies that had different objectives and involved simulation of infection in sheep prior to blood transfusion

using LPS[1-3] in a laboratory environment. Given this background, it is reasonable to hypothesise that genetic, environmental, managemental and peri-experimental practices could have been contributing factors to the observation apparent selective susceptibility of sheep to endotoxin.

Endotoxin is commonly used to mimic infection in experimental animals. Whilst there are guidelines for scientists whose work involves the use of animal models in research establishments around the world, the basic understanding of the factors associated with predisposition to, and/or simulation of infection in animals may be short of emphasis in the overall scheme of peri-experimental processes, particularly when large animals are involved. Infection is when a microorganism invades a susceptible host, starts to multiply and gets established[4]. It takes time for signs of the infection to appear after a pathogen enters the host's body. The pathogen has to overcome the host's defences, before it can be considered established. Therefore, the injection of endotoxin especially directly into blood vessels may not be an accurate model for infection[1]. Even if infected, the host may still resist the development and establishment of an infection the by way of active immunity. Immunity may be inherited or due to antibody stimulation via vaccine or previous exposure to the disease, leading to freedom from clinical signs of disease after challenge[5].

The aim of this paper is to report observations made based on a sheep model that appeared to show differential resilience to LPS endotoxin alongside controls during priming of the immune system with LPS from *E. coli* in a protracted primary study with a different trajectory of objectives in translational medicine. It also reports findings on studies on the background, source, genetic selection information and management practices of the sheep, with the expectation of using archived samples for proteogenomic studies in future. Some early observations on the base model from the primary study have been published[3], however, incremental data from the primary study are beyond the scope of the present report. An understanding of the apparent selective susceptibility or morbidity to LPS in sheep may help to understand or predict similar pathology in other mammals, including humans. The knowledge could also provide useful insights that may be applicable to other food animal producers by way of predicting disease susceptibility of their livestock.

Materials and methods

Ethical approval

Animal ethics approval for the sheep studies used in the manuscript was obtained from the University Animal Ethics Committee of the Queensland University of Technology (QUT) – reference 0800000555 and ratified by The University of Queensland. The studies were conducted in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes[6].

Study design

This was an observational experimental study based on incidental findings of a primary study with different objectives using sheep that showed selective resistance to the effects of endotoxin. The investigative component of the present study was based on the premise that the archived serum and plasma samples from experimental sheep of the primary study, could be useful in providing clues to the observed differential susceptibility to endotoxin with the help of targeting circulating acute phase proteins (APP) and micro ribonucleic acids (miRNAs) using proteogenomic approaches. Whilst it is conceivable that improper sample storage may lead to protein degradation, however, as of yet, there are no studies that have investigated the effects that deep frozen samples at -80^{o}C storage and handling could have on the outcome of downstream processes using mass spectrometry for example[7].

A pseudo-randomisation protocol was used to allocate animals into experimental groups in the primary study as previously described[3,8]. The incremental part of the primary study was continued under the same protocol. In brief, for each experiment, sheep were selected at random and then assigned to a pre-determined experiment according to a planned schedule and proportional numbers of animals in the groups. Each group was allocated sheep throughout the entire period in which the study was conducted in order to minimise the effects of any seasonal variation of the different batches. The sheep were then subjected to pre-experimental complete veterinary clinical examination that included body temperature, pulse, respiratory (TPR) and body weight parameter checks as previously described[2]. In brief, the sheep were restrained in a sling cage and the ventral aspect of the neck is shaved to facilitate venous vascular access.

Experimental animals and procedures

Background of animals

An animal bio-data request was made to the Biological Research Facility (BRF) housed within the Medical Engineering Research Facility of the Queensland University of Technology (QUT-MERF). The information request was essential in order to trace the background and source of the experimental sheep enrolled for the primary studies, based on ear tag number identification of the sheep prior to agistment between January 2009 and May 2013. Detailed documentation including strain, breed, genetic line information and management practices (vaccination, drenching, mulesing, docking history and other relevant standard operating procedures as appropriate) was tracked and processed including further records from the sheep vendors and BRF as appropriate. This information was necessary in order to investigate the relative importance of genetics, managemental and environmental factors that could help to piece together and explain the variation in LPS susceptibility in experimental sheep of the primary study. A field visit to the commercial sheep

agistment property that QUT-MERF was the lessee was made in order to capture relevant farm information and document on the pasture, housing and management practices.

Animal selection

Animal selection and experimental procedures for the primary study have previously been described [1,3]. In brief, batches of approximately 2-year old healthy adult Merino ewes *(Ovis aries)* were agisted as a flock in an open pasture farm to be used as part of, and a continuation of previously described primary studies[3]. The sheep were handled as per standard operating procedures that have been described in detail elsewhere[2] and briefly outlined below.

Housing and husbandry

All the sheep were reared at a farm with improved pastures, natural shade from trees with free access to water before being transferred to a purpose-built animal experimental laboratory of the Queensland University of Technology[2].

Pre-anaesthetic care

Within two weeks of experiments, animals were housed at a purpose built animal facility and managed as previously described[2,3] . Briefly, the sheep were fed proprietary sheep pellets, lucern and had free access to water. Shelter was provided in built concrete-floored sheds in which the sheep had free access. Shade was also provided by large trees in the paddocks and the sheep interacted freely with each other. Animals were fasted overnight (for approximately 24 h) with free access to drinking water until two hours before the procedure.

Anaesthesia, monitoring and assessment of LPS priming

The sheep were anaesthetised and maintained under ICU conditions. The management of anaesthesia, mechanical ventilation, supplemental oxygen, hydration and infusion have been described[2,3,8]. Briefly, a central venous line was placed in the external jugular vein (EJV) of the awake sheep. This catheter was used for intravenous (IV) administration of the pre-medication, induction agents, drugs, fluids and total intravenous anaesthesia (TIVA). The sheep were pre-medicated with midazolam (0.5 mg/kg) and buprenorphine (0.01mg/kg) by slow IV injection. An 8Fr sheath was placed in the opposite EJV for subsequent placement of a pulmonary artery catheter. General anaesthesia was induced with alfaxalone (3 mg/kg) IV. In early experiments[3,9], anaesthesia was induced maintained by intravenous ketamine/midazolam anaesthesia supplemented with butorphanol if indicated. In later experiments[8], anaesthesia was maintained with ketamine, midazolam, fentanyl and alfaxalone combination. Animals were placed in right lateral recumbency and monitored as reported earlier[2,3,8]. For the purposes of this report, only the first 2 h and 15 min after completion of instrumentation of the sheep are relevant.

An hour of stabilisation was allowed after which baseline monitoring parameters were noted, and then 40 min thereafter, LPS priming events and monitoring followed as illustrated in Figure A7.1.

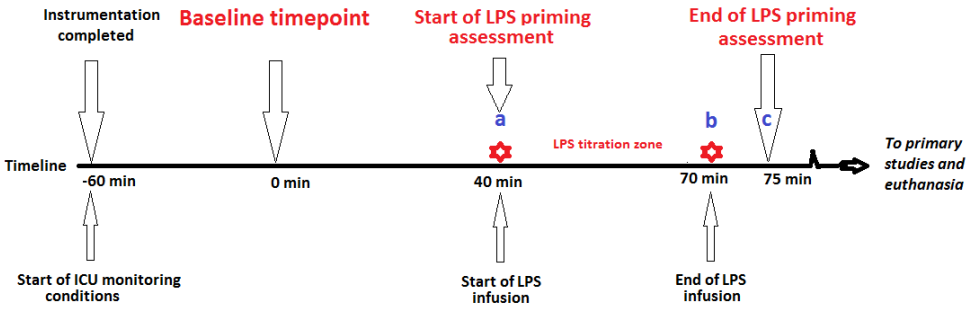


Figure A7.1. Schematic of *Escherichia coli* lipopolysaccharide (LPS) endotoxin-treated sheep showing LPS assessment time points. The sheep were anaesthetised, instrumented and monitored under intensive care unit (ICU) to conditions. After an hour and 40 mins, a saline solution containing 15µg/kg of LPS was infused at 0.5 µg/kg/min (a) over a period of 30 min (b), or as it turned out unexpectedly in the case of sheep that were more susceptible to LPS, the infusion was stopped in the LPS titration zone between (a) and (b) due to early onset of signs of LPS endotoxaemia. After completion of the LPS infusion, the sheep were monitored to a predetermined time-point (c). The sheep then proceeded to the primary study beyond the scope of this study.

To study the effects of LPS, 76 sheep were allocated to receive 30 ml infusion containing LPS as previously described[3,8,9]. In brief, a 15 µg/kg dose of the endotoxin LPS from *Escherichia coli* serotype O55:B5 (Sigma-Aldrich, Castle Hill, NSW, Australia) was chosen and infused at 0.5 µg/kg/min to prime the sheep's immune system before proceeding with the requirements of the primary study. The monitoring of the systemic effects of LPS (endotoxaemia) was based on a predetermined assessment criterion of priming guided by global haemodynamics, alveolar–arterial gradient (A–a gradient) of oxygen and the ratio of partial pressure arterial oxygen and fraction of inspired oxygen (PaO_2/FiO_2 ratio) as described previously[3,8]. Briefly, relative to baseline readings, endotoxaemia was confirmed by >50% increase in mean pulmonary artery pressure (MAP) and at least one of the following: >10% decrease in mean arterial pressure, >10% increase in heart rate (HR), >10% decrease in cardiac index (CI) >10% decrease in oxygenation saturation from mixed venous blood (SvO_2) and increase of A-a gradient or decrease in PaO_2/FiO_2 ratio[3]. If signs of severe endotoxaemia developed such as MAP < 50% of baseline and or a gradient of <10 mmHg between MAP and MPAP being a sign of impending haemodynamic collapse, before the completion of the calculated LPS dose, the infusion, was stopped and the animal was monitored. The assessment of endotoxaemia and monitoring data during the early cohort of sheep has been published with the base model[3]. If the animal decompensated further, euthanasia was indicated for as previously

reported[2]. After monitoring the effects of LPS the sheep proceeded for the requirements of the primary study outside the scope of this report. The sheep were subsequently euthanized thereafter and managed as previously described[2].

Statistical methods

One hundred and five sheep were categorised into two main groups based on outcomes of susceptibility to LPS i.e. the more LPS susceptible *n=41* (titrated LPS dose group), and the less susceptible *n= 64* (fixed LPS dose group). Animals that received placebo as the priming event, those that were used for optimisation of experimental processes and those that experienced adverse effects or those that were used as untreated controls were excluded from analysis. Body weight data of the two groups of sheep were checked for normality using D'Agostino & Pearson omnibus normality test. The mean, median, standard deviations of the weights of sheep were determined and tabulated. The titrated LPS dose group were further categorised into two sub-groups based on desirable production traits from the findings of the breeding backgrounds of the sheep. LPS dose differences between the titrated LPS, and fixed LPS groups of sheep were compared. LPS doses of the two sub-groups of the titrated LPS group of sheep were compared against each other, and against the sheep of the unknown trait (less susceptible group) using unpaired two-tailed t-test and one-way analysis of variance (ANOVA). All *p*-values were two-sided and less than 0.05 was considered statistically significant. All statistical calculations were performed using GraphPad PRISM 6 software (GraphPad Software, La Jolla, CA).

Results

Animal backgrounds, selection and study design

A total of 105 Merino ewes were enrolled in an ongoing primary study between January 2009 and May 2012 (Table A7.1). The animals were procured in 6 batches during this period from two different vendors corresponding to the dates of the experiments of *27 Jan to 18 Jun 2009; 06 Oct to 10 Dec 2009; 02 Feb to 30 Mar 2010; 16 Aug 2011 to 06 Mar 2012; 23 Jul 2012 to 28 Nov 2012;* and *07 Mar 2012 to 21 May 2013*. The last two batches comprising 41 sheep were sourced from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) FD McMaster Laboratory (Armidale, NSW, Australia). The CSIRO sheep (CSIROS) were agisted jointly as a flock in an open pasture commercial farm at Mt Cotton on the outskirts of South-East Brisbane with 64 fine-wool resident Merino ewes that had earlier been sourced from a sale yard vendor (Sale yard sourced sheep [SYSS]) at a commercial livestock market (The Australian Livestock Markets Association Inc., Warwick, QLD, Australia). All the sheep were to be used as part of, and a continuation of previously described primary studies[3,9]. It was found that the CSIROS group had two production traits namely an endoparasite resistant line (Parasite RL), and a wool production line (Wool PL).

The Parasite RL sheep were identified by ear tag numbers below 200 (A0***) and the Wool PL sheep tag numbers were the ones over 3 thousand (A3***) (Table 1). After the randomisation processes for the primary study, of the 105 enrolled sheep, 76 were allocated to be treated (primed) with LPS. The LPS primed animals constituted the study subjects for this report (Figure A7.2). Agisted together were also several other sheep of mixed ages, gender (wethers) and breeds such as Border Leicester cross breeds belonging to other projects of the animal facility.

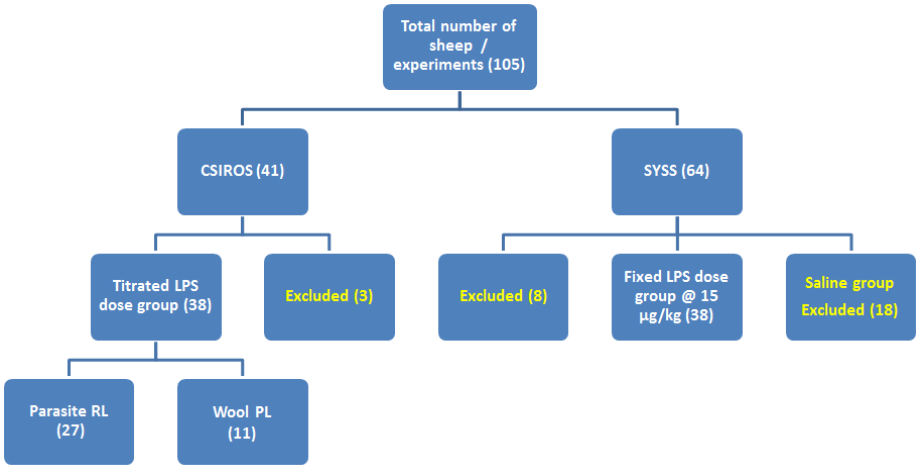


Figure A7.2. The total number of Merino (105) ewes that were enrolled to study the priming effects of *Escherichia coli* lipopolysaccharide (LPS) between 27/01/2009 and 21/05/2013. Further background studies showed that the ewes were purchased from two vendors: Commonwealth Scientific and Industrial Research Organisation (CSIRO) FD McMaster Laboratory (Armidale, NSW, Australia)(CSIROS), and The Australian Livestock Markets Association Inc., (Warwick, QLD, Australia) (SYSS). For this study, only 76 ewes that received LPS priming were included. After 3 ewes were excluded (1 had no weight recorded, 1 had no LPS dose recorded and 1 suffered an adverse event) of CSIROS, were ewes that had been selected for two production traits comprising of the internal parasite resistant line (Parasite RL [n=27]), and the wool production line (Wool PL [11]). The sheep that received saline placebo (n=18), those that were used for optimisation of experimental procedures (n=4), those that had adverse events (n=1), those with missing data (n=2) and untreated sheep used as controls (n=4) were excluded from data analysis for this report.

Housing and husbandry

Husbandry practices prior to agistment

There were certain standard husbandry operations that took place every year for CSIROS. All animals were grazed in paddocks on natural grass with improved pastures and not housed indoors. They were supplemented with grain in drought conditions when and if necessary. The animals were vaccinated against caseous lymphadenitis (CLA) and clostridial diseases according to standard

practices and vaccine manufacturer's instructions. Mobs of sheep had monthly faecal monitoring for gastrointestinal worm eggs. Samples were taken at random from a few sheep in the mob. Then the whole mob was drenched with proprietary anthelmintics if it was necessary. The sheep were all shorn once a year, as weaners in September and as adults in July. The animals were crutched in February to March. The animals were monitored for external parasites and treated if indicated. All treatments were mob based and not individual animal based. Husbandry operations for the SYSS were not available.

Husbandry practices during agistment

The commercial farm where the sheep were agisted had a dog-proof perimeter wire fencing, improved pastures, natural shade from trees and the animals had free access to fresh drinking water (Figure A7.3). At the time of agistment of the last two batches of sheep (CSIROS), pasture feed was supplemented with processed proprietary sheep pellets (RIVERINA HOBBY FARM PELLETS, West End, QLD, Australia) when required once every couple of days. The farm had a purpose built shed with raised grated wooden floors to allow free drainage of urine and droppings (Figure A7.4). Sheep handling infrastructure including, mustering pens, crush, weighing sling, sheering gear and a loading rump were available (Figure A7.5).

Figure A7.3. Part of a paddock of a commercial farm that sheep were agisted between 27/01/2009 and 21/05/2013 on behalf of Queensland University of Technology Medical Engineering Facility (QUT-MERF) for experimental studies.

Note the improved pasture in the foreground and the dog-proof fencing to the right of the photograph to protect the sheep from predatory dog attacks.

Figure A7.4. A purpose built mustering pen at a commercial farm that was used by Queensland University of Technology Medical Engineering Facility (QUT-MERF) for agistment of experimental sheep. Note the raised wooden grate flooring to allow free drainage of urine and droppings.

Figure A7.5. An ewe suspended on a sling while being weighed at purpose built shed at a commercial farm that was used by Queensland University of Technology Medical Engineering Facility (QUT-MERF) for agistment of experimental sheep. Other sheep handling infrastructure at the property (not pictured) included a crush, sheering gear and a loading rump.

The husbandry operations at the farm included routine drenching with triple combination drench as per veterinary and manufacturer's advice for the control and treatment of internal parasites and for selenium and cobalt supplementation for sheep. The triple drench (TRIGUARD®, Merial, NSW, Australia) contained (Abamectin 1.0 g/L, Oxfendazole 22.7 g/L, Levamisole hydrochloride 33.9 g/L, Sodium selenate (selenium) 0.5 g/L and Cobalt disodium EDTA 2.2 g/L. Ectoparasites were controlled with Amitraz 125g/L (TAKTIC® EC, Virbac Animal Health, Milpera, NSW, Australia) diluted at 2ml / L of water and sprayed onto the sheep according to manufacturer's recommendations. The sheep were also regularly treated with ivermectin (IVOMEC Injection®, Merial, NSW, Australia), for added protection against sucking lice and gastrointestinal worms. The sheep had regular veterinary checks, hoof care and were shorn at the end of winter. At the time of the field visit to the farm, one sheep was found to have a paralysis tick *(ixodes holocyclus)* on the neck. After removal of the tick, the sheep was treated conservatively and recovered fully within 3 days.

When the sheep were transferred to holding pens at animal facility (usually two weeks prior to experiments), they were fed high quality lucern chaff. The last two batches of sheep (CSIROS) were fed pellets at 100-300g/head/day in addition to chaff. This heightened level of nutrition was not associated with any recorded health issues.

Anaesthesia and experimental outcomes of LPS priming

In the experiments involving the first four batches of sheep, the animals were premedicated with butorphanol (0.5 mg/kg iv) and midazolam (0.5 mg/kg iv). General anaesthesia was induced with ketamine (5 mg/kg iv) and maintained by ketamine (8 mg/kg/hr iv)/midazolam (0.7 mg/kg/hr iv). Anaesthesia was supplemented with boluses of butorphanol titrated to effect as required. In the last two batches of sheep, the animals were premedicated with buprenorphine (0.01mg/kg iv) and

midazolam (0.5 mg/kg iv). Anaesthesia was induced with alfaxalone (3 mg/kg iv) and maintained with ketamine (up to 5 mg/kg/hr iv)/, midazolam (up to 0.5 mg/kg/hr iv), fentanyl (5 µg/kg/hr) and alfaxalone (4-6 mg/kg/hr) combination. From the induction of anaesthesia to the baseline time point of the primary study (Figure A7.1), all the seep received Hartmann's Solution at 15 mL/kg/hr iv, and then dropped to 1 mL/kg/hr iv thereafter. Of the 76 sheep that were challenged with LPS, animals that exhibited higher susceptibility to, and those that exhibited less susceptibility to the effects of LPS resulted in being equally split into two groups. There was a significant difference (*p = 0.02*) in their normally distributed weights between the groups (Figure A7.6). Table A7.2 shows the descriptive statistics of the two groups of sheep based on susceptibility to LPS.

Table A7.2. Weight groups of sheep that exhibited selective susceptibility to *E.coli* LPS.

	LPS dose group of sheep by weight (kg)	
Weight Parameter	Fixed LPS dose	Titrated LPS dose
Number of sheep in group	38	38
Minimum	25.7	30.3
Maximum	50	53.5
Median	36	39.5
Mean	36.3	39.5
Standard Deviation	6.1	5.6
Standard Error of Mean	0.98	0.93
D'Agostino & Pearson omnibus normality test		
Passed normality test (α=0.05)?	Yes	Yes
P value summary	Ns	Ns

Legend: There was a significant difference (p = 0.0211, two tailed t test) between the weights of sheep that got a fixed LPS dose and those that got a titrated LPS dose.

There was a significant difference ($p<0.001$) in mean LPS dose required for priming the immune system between the two groups of sheep (Table A7.3). Table A7.4 shows the two groups of sheep based on animal production traits. The more susceptible sheep required considerably small doses of LPS based on defined criteria of haemodynamic instability for priming the immune system compared to the less LPS susceptible (more resistant) sheep that received a fixed dose of LPS (Figure A7.7). The Parasite RL sheep showed marginally higher resistance to LPS challenge compared to the Wool PL sheep, and SYSS showed even considerably higher and unmatched resistance to LPS (Figure A7.8).

Table A7.3. Sheep that exhibited selective susceptibility to *E. coli* LPS.

Group based on LPS Susceptibility	No. of Sheep	Weight (Mean± SD)(kg)	Total LPS dose (µg/kg)
Fixed LPS dose	38	36.31 ± 6.07	15
Titrated LPS dose	38	39.51 ± 5.75	3.4±3.6

Legend: Sheep that received a fixed dose LPS to prime their immune system were less susceptible to effects of endotoxin than those that required LPS to be titrated. SD=Standard deviation.

Regarding archived samples for future use were samples of serum, plasma and formalin fixed /mounted tissue samples for the SYSS. In addition, CSIROS archived tissue samples preserved for deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) analysis which SYSS did not have.

Table A7.4. Merino ewes from the same mob with two production traits (endoparasite resistance and wool production) treated with a titrated dose of *E. coli* LPS.

	Sheep selection trait	
Parameter	*Endoparasite resistance*	*Wool production*
Number of sheep *(n)*	27	11
Weight (kg) ± SD (Range)	38 ± 5.6 (23.2 - 30.3)	43 ± 4.5 (35.5 - 49.8)
LPS dose (µg/kg) ± SD (Range)	4.2 ± 4.0 (1.12 - 15)	1.3 ± 0.2 (1.02 – 1.69)

Legend: Sheep selected for endoparasite resistance and were more tolerant to the effects of endotoxin than those selected for wool production. SD=Standard deviation.

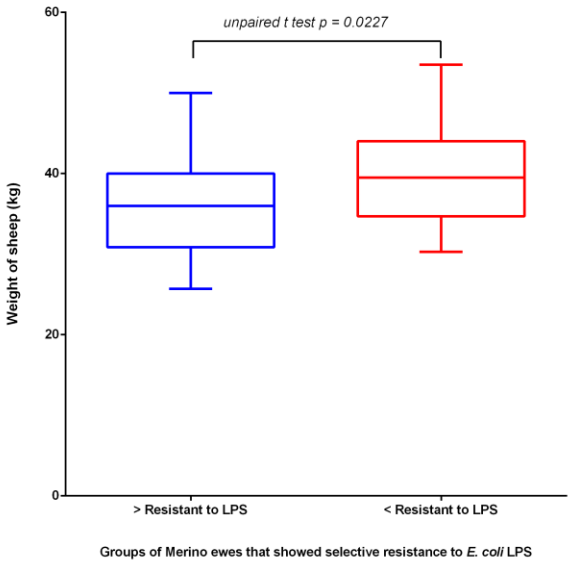


Figure A7.6. Merino ewes from the same mob that showed selective resistance to *Escherichia coli* lipopolysaccharide (LPS) endotoxin challenge. The sheep that were less resistant to the

effects of LPS (< resistant) were sourced from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) FD McMaster Laboratory (Armidale, NSW, Australia)(CSIROS). The sheep that were more resistant to the effects of LPS (> resistant) were sourced from The Australian Livestock Markets Association Inc., (Warwick, QLD, Australia) (SYSS). There was a significant difference ($p<0.001$) between CSIROS and SYSS on the amount to LPS required to cause endotoxaemia.

Discussion

There is growing need and recognition for using large animals for purposes of research and teaching[2]. This study was a retrospective, observational assessment of experimental animals and presents associations suggested to be involved with lipopolysaccharide susceptibility/resistance in experimental sheep from previously published transfusion model research on acute pulmonary injury. The first thing that comes into the minds of readers about this paper is that although it deals with an interesting topic, there are concerns because of lack of any genetic or physiological (haematological, immunological or biochemical) data that would fully support the observations that some sheep appear to be more susceptible to the effects of *E. coli* LPS endotoxin than others despite being the same age and breed. The observations made in this report were made with those background concerns in mind, but would then raise the question of how the base model of the primary study[3] was validated without the need for a complete biochemical panel or genetic analysis of the ovine model. Obviously, having the necessary data to support all observations is good science; however it is not always possible due to inherent challenges facing retrospective studies and logistical constraints. This then calls for the question that what lessons can be learnt from the validated base model referred to in this study?

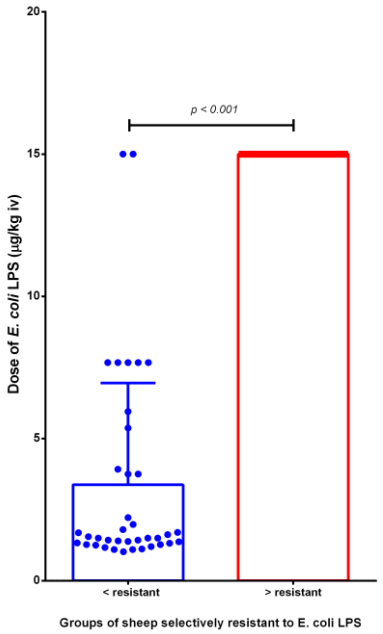


Figure A7.7. Merino ewes from the same mob showing selective resistance to *Escherichia coli* lipopolysaccharide (LPS) endotoxin. There was a significant difference ($p<0.001$) in mean

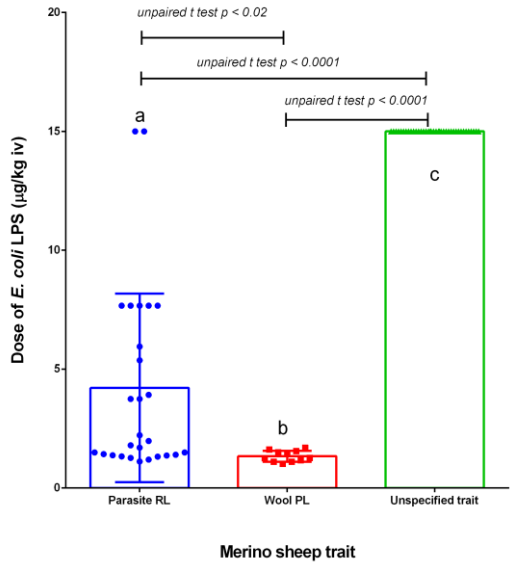


Figure A7.8. Outcomes of Merino ewes with different selection traits challenged with *Escherichia coli* lipopolysaccharide (LPS) endotoxin. The sheep that were sourced from the Commonwealth Scientific and Industrial Research Organisation FD McMaster Laboratory (Armidale, NSW, Australia) (CSIROS) were found to be of two selection traits: Endoparasite resistant line (Parasite RL) and Wool production line (Wool PL).

The Parasite RL (a) appeared to have an edge (*p<0.02*) of resistance to LPS challenge over the Wool PL (b): Sheep selected for wool production. The sheep of an unknown trait (SYSS) that were sourced from The Australian Livestock Markets Association Inc., (Warwick, QLD, Australia) showed considerable resistance to LPS challenge (c) compared to a CSIROS (a, b). There were significant differences in LPS dose (Mean ± SD) between a and c (*p<0.0001*), and b and c (*p<0.0001*).

This report provides a snapshot of the effects of *E. coli* LPS endotoxin injected into batches of sheep that were procured over a study period of nearly 3½ years. From the observations of the apparent selective susceptibility of the sheep to *E. coli* LPS, it would seem that the primary study required "approximately 2 year-old healthy female sheep" to be supplied by the host animal facility over a protracted period of time for the purposes of continuing studies based on an earlier model[3].

The primary study had taken steps to minimise the effects of seasonal variation by allocating the sheep to the various arms of the experiment through randomisation. As it turned out, the possibility of having sheep from different backgrounds and having sheep of different traits may not have been considered and factored into the methods. It is now evident multiple confounding factors may have contributed to the observed variability of the effects of endotoxin in sheep from the same mob.

The husbandry practices of CSIROS were well documented. The vendor had records for this cohort of sheep and standard operating procedures were in place. This group of sheep were essentially from a controlled breeding colony, which explains why information on the two desirable animal production traits of endoparasite resistance (Parasite RL) and wool production line (Wool PL) were traceable. From the observations so far, it appears that the Parasite RL were more resistant to the effects of LPS than the Wool PL. On the other hand, it can however be argued that perhaps the variability could have been due to un-optimised LPS dose, considering that the CSIROS response to LPS was much different, compared to that of SYSS, and that the experimental team was still readjusting to the drastic response of CSIROS to endotoxin.

Regarding feeding, Australia is known to have a history of many climate related challenges that affects many agricultural industries as the latter are climate reliant[10]. Grazing sheep on natural grass with improved pastures certainly makes a great deal of economic sense and it is associated with optimal productivity[10-12]. Grain supplementation during drought conditions is a well-recognised practise in sheep husbandry despite the associated shortcomings such as such as lactic acidosis (grain poisoning) and enterotoxaemia[13-17]. During supplementation, feed is provided to animals in order to improve their performance and, in some cases to allow the pasture to improve[18].

As with animal health, caseous lymphadenitis (CLA) is a disease of immense economic importance in sheep production worldwide. In Australia, sheep are routinely vaccinated against CLA[19-23]. It is also a recommended practice to vaccinate sheep against clostridial diseases such as botulism, black leg, black disease, tetanus and enterotoxaemia[24-27]. Without pre-purchase records of SYSS, it is not possible to comment on any husbandry practices related to that cohort of sheep, except for known sheep growing practises in Australia. It is reasonable however to assume that since this group of sheep was grown on open pasture and having been at a sale yard, it is quite possible that they had some form of active immunity that could, at least in part, explain the relative resistance to LPS as compared to CSIROS.

In sheep husbandry, in addition to disease, a major cause of stock losses is attributable to predation by feral or wild canines if there are no control measures in place[28-31]. The commonest native wild predators for sheep in Australia are dingoes. The property at which the sheep were agisted in this report, had dog-proof fencing – which may have worked to keep away grown up dingoes as well.

Perhaps as an added measure to protect sheep in similar establishments, donkeys could be kept alongside the sheep as recent reports have suggested that donkeys can be very effective deterrents to wild dogs and dingoes when reared together with sheep[32-34]. The addition of donkeys in a colony of research sheep would probably attract additional layers of concerns of their welfare, safety around people and their "non-native or feral" status as donkeys are considered introduced pests in Australia.

The control of ectoparasites was based on established principles of management targeting specific parasites, to prevent animal welfare issues and ill health[35-37]. The targeted external parasites were flies, ticks, lice and mites. Even with prevention strategies in place, paralysis ticks in Queensland could still pose a risk to animals as in that one asymptomatic sheep that had a paralysis tick.

During the pre-anaesthetic period, interestingly, the heightened level of nutrition at the animal holding facility leading up to the experiments for CSIROS was not associated with any metabolic disturbances that are usually seen when sheep are suddenly introduced to a high carbohydrate source. This may suggest that this cohort of sheep were probably accustomed to a high plane of feeding prior to agistment. As the sheep were fasted for 24h with free access to water until approximately 2 hours prior to anaesthetic induction[2], it is quite possible that CSIROS were not accustomed to their food being withheld for that long. Except for pre-anaesthetic veterinary physical examination, pre-anaesthetic or pre-interventional laboratory investigations were not included in the primary study. Baseline samples for the primary study were collected after induction of anaesthesia and after invasive instrumentation (Figure A7.1). Without a pre-anaesthetic laboratory panel, it is not possible to objectively determine the metabolic status or reference point of the experimental animal prior to any interventions for downstream studies that routinely require pre-interventional parameters. It can be argued that CSIROS were probably in a physiological state of negative energy balance prior to the experiments – a predisposition that may have contributed to the sheep to be comparatively susceptible to the effects of LPS.

The preparation of sheep for anaesthesia and physiological monitoring to enable the assessment of LPS priming and experimental outcomes were well optimised. However, there were concerns about the alterations in the protocol for premedication and general anaesthesia in the course of the experiments. These alterations may have influenced the outcome of LPS assessment. For example, the individual roles of ketamine and alfaxalone as anaesthetics on the effects of LPS in sheep are unknown. Ketamine has long been known to increase the vascular tone in sheep[38]; it is therefore plausible to assert that the type of anaesthetic may have contributed to the observed selective susceptibility to LPS in sheep as the LPS endotoxaemia assessment was heavily reliant on the assessment of haemodynamic parameters.

There were some shortcomings worth noting in the study. Based on the findings of this report, caution ought to be exercised before arriving at the conclusion that there was selective resistance to LPS endotoxin in sheep with similar characteristics. This is because certain procedures were not standardised. Albeit the sheep were of the same breed, they had different backgrounds, environmental conditions, vendors and production traits. There was also variability in the anaesthetic protocol as expounded in the preceding section. Other than the variability noted in the experimental subjects, there were no published studies on basic serum/plasma biochemistry on the validation studies of the model for which the primary study[3] was based on. The confounding factors of time, age, transportation, stress, and interventions prior to sampling may not have been fully accounted for. The study was gender biased - no males were recruited. Perhaps the most crucial set back is that there was no pre-anaesthetic laboratory work factored into the primary study. In veterinary practice, sampling for baseline parameters prior to any intervention is widely considered as a standard operating procedure. Samples for RNA and DNA analysis for future studies were not uniformly collected across the board. The archived samples may therefore not be suitable for planned comprehensive comparative proteogenomic studies of SYSS and CSIROS because no pre-interventional samples that could act individual subjects' controls were collected. Only archived serum and plasma samples could be of use in future comparative studies for all the sheep. Haemodynamic (MPAP, MAP, SvO2, CCO), respiratory ([A-a] gradient, PaO_2/FiO_2 ratio, arterial blood gas analysis), haematology, serum/plasma biochemistry and urinalysis findings are not available for reasons beyond the purpose of this report.

Finally, as a future perspective, in order to fully determine if selective susceptibility to *E. coli* endotoxin in sheep is apparent, all the sources of variation identified and discussed in this report need to be taken into consideration and minimised. This research could benefit from more concrete mechanistic support for genetic, epigenetic, immune, and/or other factor involvement. It is important to strongly consider exploring potential innate immune factors to support the assertion of selective susceptibility to endotoxin. Baseline samples and parameters should be taken before any form of intervention is done to represent the "normal" state or of the experimental animal. The analysis of baseline samples will therefore act as a benchmark for comparison with outcomes of interventions downstream. The present discussion is heavily focused on experimental animal use in research as these factors are proposed as being crucial in the development of optimum large animal experimental methods for better study outcomes. Detailed attention to mechanisms of variable immune recognition of LPS should also be taken into consideration by way of further studies. With this in mind, studies are planned to develop a method that will first establish the normal circulating acellular proteome, followed by validation of the method using samples from LPS primed sheep,

potentially identifying acute phase acute phase proteins (APP) as recently published[1].

Conclusions

In conclusion, cautiously taking all observations into perspective, some merino sheep appear to be more susceptible to the effects of *E. coli* LPS endotoxin than others despite being the same age and breed. It should be noted that the observations that have been advanced in the current study are potentially interesting physiological and/or pathophysiological findings that warrant subsequent exploration of genetic, epigenetic, innate immune system involvement. Genetic and physiological data are needed in order to fully back these observations – calling for further studies and the inclusion of such data to the methods of protracted large animal models with potential translational applications. Sheep brought up under hardy conditions may be more resistant to the effects of endotoxin than those grown under more optimal conditions. It appears that sheep that have the endoparasite resistance trait may also be resistant to *E. coli* LPS compared to sheep with the wool production trait. There are a variety of known mitigating environmental and unknown epigenetic factors that need to be taken into consideration when designing protracted large animal experiments as they can influence the host's outcomes in regards to exposure to endotoxin. These observations could be relevant in understanding the relative importance of phenotype and managemental differences, and could set a foundation for further proteogenomic studies on plasma and serum to elucidate the apparent variation in LPS susceptibility in sheep. Individual genetic characteristics may also contribute in defining and predicting response to an infection challenge in sheep. The clues in understanding of the mechanism of the selection to LPS morbidity in sheep could be due to changes in plasma/serum proteins, whose elucidation may help to understand similar pathology in other mammals, including humans.

Acknowledgements

The primary study was funded by the Australian Red Cross Blood Service via a grant to the Chief Investigator and Co-Investigators. The author is thankful to the investigators of the original study for their general supervision through existing adjunct arrangements with The University of Queensland at the Prince Charles Hospital Campus and for help in providing access to part of the historical data (SYSS) included in this report. The author was supported by the Institute of Health and Biomedical Innovation (IHBI) of the Queensland University of Technology and an Australian Postgraduate Award (APA) Research High Degree (PhD) scholarship through the School of Veterinary Science (SVS), The University of Queensland. Gratitude is extended to the staff of Queensland University Medical Engineering Facility (QUT-MERF) where the experiments were conducted, and the Commonwealth Scientific and Industrial Research Organisation (CSIRO) FD

McMaster Laboratory (Armidale, NSW, Australia) for kindly providing information on the sheep they supplied to QUT-MERF.

This study formed the foundation of current research on proteogenomic characterisation of circulating acute phase markers and bioassay development in sheep whose research advisory team includes Dr Pawel Sadowski (CARF-QUT), Prof Paul Mills (SVS, UQ) & Dr Steven Kopp (SVS, UQ).

Disclosure

The author declares that there is no conflict of interests regarding the publication of this paper.

References

1. Chemonges, S., Tung, J. P. & Fraser, J. F. Proteogenomics of selective susceptibility to endotoxin using circulating acute phase biomarkers and bioassay development in sheep: A review. *Proteome Science* 12, 12, doi:10.1186/1477-5956-12-12 (2014).
2. Chemonges, S., Shekar, K., Tung, J. P., Dunster, K. R., Diab, S., Platts, D., Watts, R. P., Gregory, S. D., Foley, S., Simonova, G., McDonald, C., Hayes, R., Bellpart, J., Timms, D., Chew, M., Fung, Y. L., Toon, M., Maybauer, M. O. & Fraser, J. F. Optimal management of the critically ill: Anaesthesia, monitoring, data capture, and point-of-care technological practices in ovine models of critical care. *Biomedical Research International* 2014, 468309, doi:10.1155/2014/468309 (2014).
3. Tung, J. P., Fung, Y. L., Nataatmadja, M., Colebourne, K. I., Esmaeel, H. M., Wilson, K., Barnett, A. G., Wood, P., Silliman, C. C. & Fraser, J. F. A novel in vivo ovine model of transfusion-related acute lung injury (trali). *Vox Sanguinis* 100, 219-230, doi:10.1111/j.1423-0410.2010.01381.x (2011).
4. Thomson, P. D. & Smith, D. J. What is infection? *The American Journal of Surgery* 167, S7-S11, doi:10.1016/0002-9610(94)90003-5 (1994).
5. Raadsma, H. W., Gray, G. D. & Woolaston, R. R. Breeding for disease resistance in merino sheep in australia. *Revue Scientifique et Technique* 17, 315-328 (1998).
6. National, Health, and, Medical, Research & Council. *Australian code of practice for the care and use of animals for scientific purposes*, <https://www.nhmrc.gov.au/guidelines/publications/ea28> (2013).
7. Mitchell, B. L., Yasui, Y., Li, C. I., Fitzpatrick, A. L. & Lampe, P. D. Impact of freeze-thaw cycles and storage time on plasma samples used in mass spectrometry based biomarker

discovery projects. *Cancer Informatics* 1, 98-104 (2005).

8. Tung, J. P., Fraser, J. F., Nataatmadja, M., Colebourne, K. I., Barnett, A. G., Glenister, K. M., Zhou, A. Y., Wood, P., Silliman, C. C. & Fung, Y. L. Age of blood and recipient factors determine the severity of transfusion-related acute lung injury (TRALI). *Critical Care (London, England)* 16, R19, doi:10.1186/cc11178 (2012).
9. Tung, J-P. *Development of an ovine model of transfusion-related acute lung injury (TRALI)* Dissertation/Thesis thesis, University of Queensland, (2012).
10. Andrew, M. Drought, dust and despair. *Historian*, 14 (2011).
11. Cattle, S. R. & Southorn, N. J. Macroporosity of pasture topsoils after three years of set-stocked and rotational grazing by sheep. *Australian Journal of Soil Research* 48, 43, doi:10.1071/SR09004 (2010).
12. Norman, H. C., Wilmot, M. G., Thomas, D. T., Barrett-Lennard, E. G. & Masters, D. G. Sheep production, plant growth and nutritive value of a saltbush-based pasture system subject to rotational grazing or set stocking. *Small Ruminant Research* 91, 103-109, doi:10.1016/j.smallrumres.2009.11.022 (2010).
13. McGregor, B. A. & Whiting, C. J. Grain excretion by goats fed whole or processed cereals with various roughages. *Small Ruminant Research* 115, 21, doi:10.1016/j.smallrumres.2013.08.009 (2013).
14. Dixon, R. M., Karda, W., Hosking, B. J. & Egan, A. R. Effects of oilseed meals and grain–urea supplements fed infrequently on digestion in sheep. *Animal Feed Science and Technology* 110, 95-110, doi:10.1016/S0377-8401(03)00205-0 (2003).
15. Bush, R. D., Toribio, J. A. & Windsor, P. A. The impact of malnutrition and other causes of losses of adult sheep in 12 flocks during drought. *Australian Veterinary Journal* 84, 254-254, doi:10.1111/j.1751-0813.2006.00002.x (2006).
16. Alcock, D. J. & Hegarty, R. S. Potential effects of animal management and genetic improvement on enteric methane emissions, emissions intensity and productivity of sheep enterprises at cowra, australia. *Animal Feed Science and Technology* 166, 749-760, doi:10.1016/j.anifeedsci.2011.04.053 (2011).
17. Dove, H. & McMullen, K. G. Diet selection, herbage intake and liveweight gain in young sheep grazing dual-purpose wheats and sheep responses to mineral supplements. *Animal Production Science* 49, 749, doi:10.1071/an09009 (2009).
18. Holst, P., White A. *Primefacts* (Department of Primary Industries, State of New South Wales, 2007).
19. Zavoshti, F. R., Khoojine, A. B. S., Helan, J. A., Hassanzadeh, B. & Heydari, A. A. Frequency of caseous lymphadenitis (cla) in sheep slaughtered in an abattoir in tabriz:

Comparison of bacterial culture and pathological study. *Comparative Clinical Pathology* 21, 667-671, doi:10.1007/s00580-010-1154-7 (2012).

20. Windsor, P. A. Control of caseous lymphadenitis. *The Veterinary clinics of North America. Food animal practice* 27, 193-202, doi:10.1016/j.cvfa.2010.10.019 (2011).
21. O'Reilly, K. M., Medley, G. F. & Green, L. E. The control of corynebacterium pseudotuberculosis infection in sheep flocks: A mathematical model of the impact of vaccination, serological testing, clinical examination and lancing of abscesses. *Preventive Veterinary Medicine* 95, 115-126, doi:10.1016/j.prevetmed.2010.02.012 (2010).
22. Guimarães, A. S., Gouveia, A. M. G., Carmo, F. B., Heinemann, M. B., Portela, R. W. D., Meyer, R., Lage, A. P., Seyffert, N., Miyoshi, A. & Azevedo, V. High sero-prevalence of caseous lymphadenitis identified in slaughterhouse samples as a consequence of deficiencies in sheep farm management in the state of minas gerais, brazil. *BMC Veterinary Research* 7, 68-68, doi:10.1186/1746-6148-7-68 (2011).
23. Alessandro de Sá, G., Aurora Maria Guimarães, G., Filipe Borges do, C., Rebeca Barbosa, P., Nubia, S., Dayana, R., Andrey Pereira, L., Marcos Bryan, H., Anderson, M. & Vasco, A. Caseous lymphadenitis: Epidemiology, diagnosis, and control. *The Institute of Integrative Omics and Applied Biotechnology Journal* 2, 33-43 (2011).
24. Ellis, T. M., Masters, A. M., Sutherland, S. S., Carson, J. M. & Gregory, A. R. Variation in cultural, morphological, biochemical properties and infectivity of australian isolates of dermatophilus congolensis. *Veterinary Microbiology* 38, 81-102, doi:10.1016/0378-1135(93)90077-K (1993).
25. Dagleish, M. P., Benavides, J. & Chianini, F. Immunohistochemical diagnosis of infectious diseases of sheep. *Small Ruminant Research* 92, 19-35, doi:10.1016/j.smallrumres.2010.04.003 (2010).
26. Uzal, F. A. Diagnosis of clostridium perfringens intestinal infections in sheep and goats. *Anaerobe* 10, 135-143, doi:10.1016/j.anaerobe.2003.08.005 (2004).
27. Lewis, C. J. Control of important clostridial diseases of sheep. *Veterinary Clinics of North America: Food Animal Practice* 27, 121-126, doi:10.1016/j.cvfa.2010.10.009 (2011).
28. Allen, B. L. & West, P. Influence of dingoes on sheep distribution in australia. *Australian Veterinary Journal* 91, 261-267, doi:10.1111/avj.12075 (2013).
29. Dennis, S. M. Predators and perinatal mortality of lambs in western australia. *Australian Veterinary Journal* 45, 6-9, doi:10.1111/j.1751-0813.1969.tb01854.x (1969).
30. King, A. J., Wilson, A. M., Wilshin, S. D., Lowe, J., Haddadi, H., Hailes, S. & Morton, A. J. Selfish-herd behaviour of sheep under threat. *Current Biology* 22, R561-562, doi:10.1016/j.cub.2012.05.008 (2012).

31. Morris, S. T. Economics of sheep production. *Small Ruminant Research* 86, 59-62, doi:10.1016/j.smallrumres.2009.09.019 (2009).
32. Ross, K. & Stephen, B. in *Wall Street Journal* 11 (Dow Jones & Company Inc, Brussels, 2013).
33. Phinjo Gombu, T. S. in *Toronto Star* A.8 (Torstar Syndication Services, a Division of Toronto Star Newspapers Limited, Toronto, Ont, 1990).
34. Hayward, M. *Reintroduction of top-order predators*. Vol. no. 5. (Wiley-Blackwell, 2009).
35. Windsor, P. A. & Lomax, S. Addressing welfare concerns in control of ovine cutaneous myiosis in sheep in australia. *Small Ruminant Research* 110, 165, doi:10.1016/j.smallrumres.2012.11.027 (2012).
36. Plant, J. W. & Lewis, C. J. Treatment and control of ectoparasites in sheep. *The Veterinary clinics of North America. Food animal practice* 27, 203-212, doi:10.1016/j.cvfa.2010.10.012 (2011).
37. Levot, G. Unstable pyrethroid resistance in sheep body lice bovicola ovis (schrank), (phthiraptera: Trichodectidae) and its implications for lice control on sheep. *Veterinary Parasitology* 185, 274-278, doi:10.1016/j.vetpar.2011.09.036 (2012).
38. Craft, J. B., Jr., Coaldrake, L. A., Yonekura, M. L., Dao, S. D., Co, E. G., Roizen, M. F., Mazel, P., Gilman, R., Shokes, L. & Trevor, A. J. Ketamine, catecholamines, and uterine tone in pregnant ewes. *American Journal of Obstetrics and Gynecology* 146, 429-434 (1983).

Table A7.1. Ewes used by Queensland University of Technology for a project from January 2009 - May 2013.

Date	Experiment	Sheep	Weight (kg)	Treatment	LPS dose (µg/kg)
27-Jan-09	**1**	**582**	**40**	**LPS @ 8 ug/kg/min**	**240**
10-Feb-09	**2**	**571**	**54**	**LPS @ 1 ug/kg/min**	**30**
03-Mar-09	3	619	33	LPS @ 0.5 ug/kg/min	15
17-Mar-09	4	625	40	LPS @ 0.5 ug/kg/min	15
18-Mar-09	5	613	34	LPS @ 0.5 ug/kg/min	15
19-Mar-09	6	617	40	LPS @ 0.5 ug/kg/min	15
24-Mar-09	7	618	43	LPS @ 0.5 ug/kg/min	15
25-Mar-08	**8**	**624**	**34**	**N/A**	
26-Mar-09	**9**	**605**	**48**	**N/A**	
07-Apr-09	10	612	37	LPS @ 0.5 ug/kg/min	15
08-Apr-09	11	608	30	LPS @ 0.5 ug/kg/min	15
09-Apr-09	**12**	**622**	**45**	**saline (30mL)**	**0**
14-Apr-09	13	602	37	LPS @ 0.5 ug/kg/min	15
16-Apr-09	14	620	46	LPS @ 0.5 ug/kg/min	15
21-Apr-09	**15**	**621**	**42**	**saline (30mL)**	**0**
23-Apr-09	**16**	**611**	**39**	**saline (30mL)**	**0**
28-Apr-09	**17**	**665**	**41**	**saline (30mL)**	**0**
30-Apr-09	**18**	**677**	**44**	**saline (30mL)**	**0**
05-May-09	**19**	**673**	**46**	**saline (30mL)**	**0**
07-May-09	**20**	**706**	**40**	**saline (30mL)**	**0**
12-May-09	**21**	**702**	**47**	**N/A**	
14-May-09	**22**	**703**	**33**	**saline (30mL)**	**0**
02-Jun-09	23	698	37	LPS @ 0.5 ug/kg/min	15
03-Jun-09	24	697	36	LPS @ 0.5 ug/kg/min	15
04-Jun-09	**25**	**699**	‡	**N/A**	
09-Jun-09	26	701	45	LPS @ 0.5 ug/kg/min	15
11-Jun-09	27	745	36	LPS @ 0.5 ug/kg/min	15
16-Jun-09	28	719	39	LPS @ 0.5 ug/kg/min	15
18-Jun-09	29	710	43	LPS @ 0.5 ug/kg/min	15
06-Oct-09	30	729	38	LPS @ 0.5 ug/kg/min	15

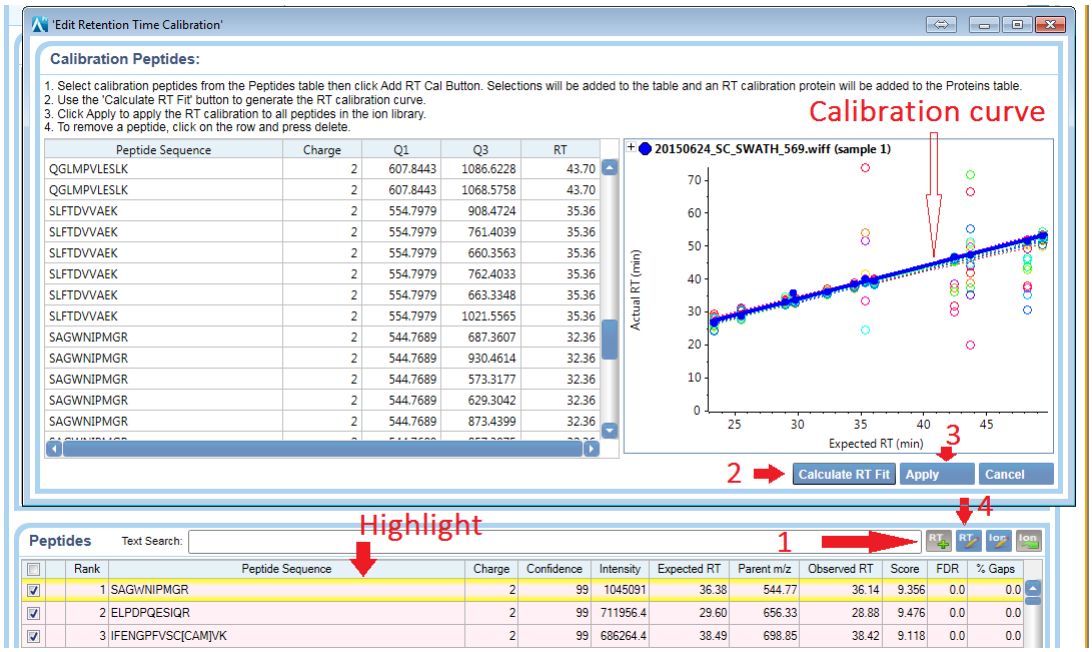
Date	Experiment	Sheep	Weight (kg)	Treatment	LPS dose (µg/kg)
27-Oct-09	**31**	**704**	**40**	**N/A**	
29-Oct-09	**32**	**720**	**40**	**saline (30mL)**	**0**
10-Nov-09	33	711	37	LPS @ 0.5 ug/kg/min	15
12-Nov-09	**34**	**742**	**46**	**saline (30mL)**	**0**
19-Nov-09	**35**	**667**	**39.2**	**saline (30mL)**	**0**
24-Nov-09	36	708	33	LPS @ 0.5 ug/kg/min	15
26-Nov-09	**37**	**713**	**40**	**N/A**	
01-Dec-09	38	670	43	LPS @ 0.5 ug/kg/min	15
03-Dec-09	39	671	45	LPS @ 0.5 ug/kg/min	15
08-Dec-09	40	714	37.5	LPS @ 0.5 ug/kg/min	15
10-Dec-09	**41**	**764**	**54**	**saline (30mL)**	**0**
2-Feb-10	**42**	**773**	**37**	**saline (30mL)**	**0**
4-Feb-10	**43**	**792**	**38.4**	**saline (30mL)**	**0**
9-Feb-10	**44**	**789**	**46**	**saline (30mL)**	**0**
11-Feb-10	**45**	**781**	**39**	**saline (30mL)**	**0**
16-Feb-10	46	786	35.5	LPS @ 0.5 ug/kg/min	15
23-Feb-10	**47**	**780**	**33**	**saline (30mL)**	**0**
25-Feb-10	48	A252	36	LPS @ 0.5 ug/kg/min	15
2-Mar-10	49	794	31	LPS @ 0.5 ug/kg/min	15
4-Mar-10	50	778	38	LPS @ 0.5 ug/kg/min	15
9-Mar-10	51	805	50	LPS @ 0.5 ug/kg/min	15
11-Mar-10	52	820	36	LPS @ 0.5 ug/kg/min	15
16-Mar-10	53	815	48	LPS @ 0.5 ug/kg/min	15
18-Mar-10	54	814	34	LPS @ 0.5 ug/kg/min	15
23-Mar-10	**55**	**816**	**41**	**saline (30mL)**	**0**
30-Mar-10	56	828	26	LPS @ 0.5 ug/kg/min	15
16-Aug-11	57	1100	30.5	LPS @ 0.5 ug/kg/min	15
18-Aug-11	58	1101	30.1	LPS @ 0.5 ug/kg/min	15
23-Aug-11	59	1098	29.2	LPS @ 0.5 ug/kg/min	15
30-Aug-11	60	1067	33.7	LPS @ 0.5 ug/kg/min	15
27-Feb-12	61	1190	29.2	LPS @ 0.5 ug/kg/min	15
27-Feb-12	62	1186	27.9	LPS @ 0.5 ug/kg/min	15

Date	Experiment	Sheep	Weight (kg)	Treatment	LPS dose (µg/kg)
6-Mar-12	63	1195	25.7	LPS @ 0.5 ug/kg/min	15
6-Mar-12	64	1192	29.6	LPS @ 0.5 ug/kg/min	15
23-Jul-12	‡	**09A0066**	**29.3**	**N/A**	**Terminated**
14-Aug-12	65	09A0079	32	LPS titration	15
14-Aug-12	66	09A0016	30.3	LPS titration	15
16-Aug-12	67	09A0126	31.8	LPS titration	7.67
16-Aug-12	68	09A0092	31	LPS titration	3.75
7-Sep-12	69	09A0113	34.9	LPS titration	1.5
7-Sep-12	70	09A0100	33.7	LPS titration	3.75
13-Sep-12	71	09A0035	36.4	LPS titration	7.67
13-Sep-12	72	09A0021	38.7	LPS titration	7.67
20-Sep-12	73	09A0074	34.1	LPS titration	7.67
20-Sep-12	74	09A0037	36.4	LPS titration	7.67
17-Oct-12	75	09A0140	33.6	LPS titration	5.95
17-Oct-12	76	09A0102	33.5	LPS titration	5.37
24-Oct-12	77	09A0039	38.2	LPS titration	3.92
24-Oct-12	78	09A0072	40.5	LPS titration	2.22
31-Oct-12	79	09A0064	40.5	LPS titration	1.98
31-Oct-12	80	09A0076	38.8	LPS titration	1.8
1-Nov-12	81	09A0071	43.7	LPS titration	1.37
1-Nov-12	82	09A0073	43	LPS titration	1.4
8-Nov-12	83	09A0096	39.8	LPS titration	1.5
-Nov-12	84	09A0078	36.2	LPS titration	1.38
9-Nov-12	85	09A0093	35	LPS titration	1.43
9-Nov-12	86	09A0071	33.2	LPS titration	1.2
27-Nov-12	87	09A0114	41.2	LPS titration	1.7
27-Nov-12	88	09A0116	45	LPS titration	1.33
28-Nov-12	89	09A3023	39.5	LPS titration	1.27
28-Nov-12	90	09A3032	40	LPS titration	1.25
7-Mar-13	91	09A3061	39.5	LPS titration	1.43
7-Mar-13	**92**	**1187**	‡	**LPS titration**	**1.43**
17-Apr-13	93	09A3192	39.8	LPS titration	1.5

Date	Experiment	Sheep	Weight (kg)	Treatment	LPS dose (µg/kg)
17-Apr-13	94	09A3084	43.1	LPS titration	1.62
18-Apr-13	95	09A3157	35.5	LPS titration	1.69
18-Apr-13	**96**	**09A3153**	**32**	**LPS titration**	‡
9-May-13	97	09A3190	45.1	LPS titration	1.55
9-May-13	98	09A0127	45.6	LPS titration	1.32
10-May-13	99	09A0117	47.1	LPS titration	1.27
10-May-13	100	09A3063	46.9	LPS titration	1.17
20-May-13	101	09A3159	45.5	LPS titration	1.1
20-May-13	102	09A3204	49.8	LPS titration	1.1
21-May-13	103	09A3021	48.9	LPS titration	1.02
21-May-13	104	09A0079	53.5	LPS titration	1.12

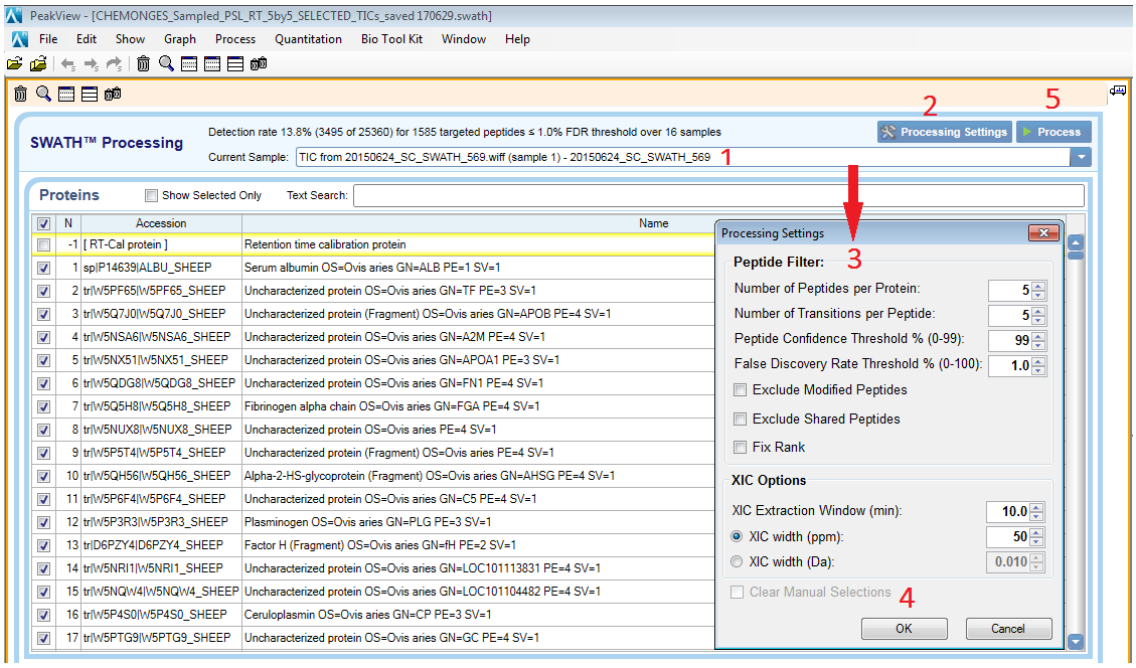
Legend: Experiment details (date of experiment, experiment number, sheep identification and treatment) of ewes (n = 105) that were agisted as a mob at a commercial farm for Queensland University of Technology Medical Engineering Facility (QUT-MERF). The sheep were purchased from 2 separate vendors: The Australian Livestock Markets Association Inc., (Warwick, QLD, Australia) – for experiments 1-64, and 92; and the Commonwealth Scientific and Industrial Research Organisation (CSIRO) FD McMaster Laboratory (Armidale, NSW, Australia) – for the terminated experiment, experiments 65-91 and 93-104. Key: LPS = Lipopolysaccharide from Escherichia coli; N/A = No treatment given: control experiments; ‡ = missing data. Experiments 1 & 2 were used for LPS optimisation experiments. All experiments in bold were excluded from the analysis of the results of the present report.

APPENDIX 7.1. Steps for generating a retention time (RT) calibration curve



Highlight and select calibration peptides, followed by the steps 1 and 2 and apply it – 3. The RT calibration curve can be accessed by clicking 4 in the main home screen.

APPENDIX 7.2. Entering SWATH-MS processing settings



How SWATH-MS processing settings (red arrow) for the PSL in PeakView® Software user interface were entered. Sample .wiff files 1 were loaded, followed by steps 2, 3, 4 and 5 above.

APPENDIX 7.3 List of proteins that were quantitated by SWATH-MS analysis in plasma of endotoxaemic sheep

Table A7.5. The UniProtKB accession numbers, gene names, NCBI names, protein status, UniProtKB names and fold change values of 243 sheep plasma proteins that altered during *E. coli* lipopolysaccharide-induced endotoxaemia. Fold change represents how the quantity of protein changed from before and after 75 minutes of acute endotoxaemia based on their protein peak area comparisons that were processed in MarkerView™ Software (SCIEX).

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5PD76_SHEEP	TRMT11	tRNA methyltransferase 11 homolog	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=TRMT11 PE=4 SV=1	16.2
W5PUH0_SHEEP	FAM105A	Family with sequence similarity 105 member A	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=FAM105A PE=4 SV=1	15.9
W5QJ60_SHEEP	LOC101116576	Sentrin-specific protease 8-like	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=LOC101116576 PE=4 SV=1	6.2
W5QGD1_SHEEP	LDHB	L-lactate dehydrogenase	Predicted	L-lactate dehydrogenase OS=Ovis aries GN=LDHB PE=3 SV=1	6.0
W5PJD1_SHEEP	KMT2A	Lysine methyltransferase 2A	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=KMT2A PE=4 SV=1	5.1
W5Q2A5_SHEEP	FAM13B	Family with sequence similarity 13 member B	Predicted	Uncharacterised protein OS=Ovis aries GN=FAM13B PE=4 SV=1	4.4

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5QDU3_SHEEP	TRDV2	T cell receptor delta variable 2	Predicted	Uncharacterised protein OS=Ovis aries GN=TRDV2 PE=4 SV=1	3.1
W5PHP2_SHEEP	EEF1A1	Elongation factor 1-alpha 1-like	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries PE=4 SV=1	2.8
W5Q754_SHEEP	TTN	Titin	Predicted	Uncharacterised protein OS=Ovis aries GN=TTN PE=4 SV=1	2.7
W5PV69_SHEEP	HIST2H2AB	Histone cluster 2 H2A family member b	Predicted	Histone H2A OS=Ovis aries GN=HIST2H2AB PE=3 SV=1	2.6
W5P084_SHEEP	CERS4	Ceramide synthase 4	Predicted	Uncharacterised protein OS=Ovis aries GN=CERS4 PE=4 SV=1	2.6
W5PZ55_SHEEP	PPBP	Platelet basic protein	Predicted	C-X-C motif chemokine OS=Ovis aries GN=PPBP PE=3 SV=1	2.5
W5PZT2_SHEEP	ANGPTL6	Angiopoietin like 6	Predicted	Uncharacterised protein OS=Ovis aries GN=ANGPTL6 PE=4 SV=1	2.5
W5Q0L2_SHEEP	SERPINA5	Serpin family A member 5	Predicted	Uncharacterised protein OS=Ovis aries GN=SERPINA5 PE=3 SV=1	2.5
W5PZT3_SHEEP	LECT2	Leukocyte cell derived chemotaxin 2	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=LECT2 PE=4 SV=1	2.4
W5PA83_SHEEP	PRR14	proline rich 14	Predicted	Uncharacterised protein OS=Ovis aries GN=PRR14 PE=4 SV=1	2.3

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5P323_SHEEP	GPI	glucose-6-phosphate isomerase	Predicted	Glucose-6-phosphate isomerase OS=Ovis aries GN=GPI PE=3 SV=1	2.3
W5NYP1_SHEEP	WC11	Antigen WC1.1	Predicted	Uncharacterised protein OS=Ovis aries PE=4 SV=1	2.2
G3LUQ4_SHEEP	CSN1S1	Casein alpha s1	Predicted	Alpha s1 casein OS=Ovis aries GN=CSN1S1 PE=4 SV=1	2.2
A2P2G7_SHEEP	VH	VH region	Predicted	VH region (Fragment) OS=Ovis aries GN=VH PE=2 SV=1	2.2
W5P082_SHEEP	IGF2	Insulin like growth factor 2	Predicted	Insulin-like growth factor II OS=Ovis aries GN=IGF2 PE=3 SV=1	2.2
W5PKA9_SHEEP	F5	F5 coagulation factor V	Predicted	Uncharacterised protein OS=Ovis aries GN=F5 PE=3 SV=1	2.2
W5P8V3_SHEEP	LGI2	Leucine rich repeat LGI family member 2	Predicted	Uncharacterised protein OS=Ovis aries GN=LGI2 PE=4 SV=1	2.2
W5NVP3_SHEEP	CHAT	choline O-acetyltransferase	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=CHAT PE=3 SV=1	2.2
W5NXM6_SHEEP	PTX3	Pentraxin 3	Predicted	Uncharacterised protein OS=Ovis aries GN=PTX3 PE=4 SV=1	2.1
W5Q7C7_SHEEP	PSMA2	proteasome subunit alpha 2	Predicted	Proteasome subunit alpha type OS=Ovis aries GN=PSMA2 PE=3 SV=1	2.1

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5PEL2_SHEEP	OGN	Osteoglycin	Predicted	Uncharacterised protein OS=Ovis aries GN=OGN PE=4 SV=1	2.1
W5PIG6_SHEEP	ENO1	Enolase 1	Predicted	Uncharacterised protein OS=Ovis aries GN=ENO1 PE=3 SV=1	2.0
K4P494_SHEEP	CST3	Cystatin C	Predicted	Cystatin OS=Ovis aries GN=CST3 PE=2 SV=1	1.9
W5P915_SHEEP	LOC101107619	Short palate, lung and nasal epithelium carcinoma-associated protein 2B-like	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101107619 PE=4 SV=1	1.9
W5Q0X5_SHEEP	LOC101115576	Serpin A3-5	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101115576 PE=3 SV=1	1.9
W5PGE9_SHEEP	IGHA	immunoglobulin alpha heavy chain	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries PE=4 SV=1	1.9
W5P293_SHEEP	CPEB4	cytoplasmic polyadenylation element binding protein 4	Predicted	Uncharacterised protein OS=Ovis aries GN=CPEB4 PE=4 SV=1	1.9
W5Q0R1_SHEEP	SHBG	Sex hormone binding globulin	Predicted	Uncharacterised protein OS=Ovis aries GN=SHBG PE=4 SV=1	1.9
W5NS74_SHEEP	ACAA1	Acetyl-CoA acyltransferase 1	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=ACAA1 PE=3 SV=1	1.9

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5Q2E1_SHEEP	LUM	Lumican	Predicted	Uncharacterised protein OS=Ovis aries GN=LUM PE=4 SV=1	1.9
W5NVM6_SHEEP	LOC101119889	Ig heavy chain V region PJ14-like	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101119889 PE=4 SV=1	1.8
W5PEI4_SHEEP	KLKB1	Kallikrein B1	Predicted	Uncharacterised protein OS=Ovis aries GN=KLKB1 PE=3 SV=1	1.8
B5B304_SHEEP	CFH	Complement factor H	Predicted	Complement factor H (Fragment) OS=Ovis aries GN=CFH PE=2 SV=1	1.8
W5PAJ9_SHEEP	APOM	Apolipoprotein M	Predicted	Uncharacterised protein OS=Ovis aries GN=APOM PE=4 SV=1	1.8
W5Q124_SHEEP	LOC101119509	Serpin A3-8	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101119509 PE=3 SV=1	1.7
W5P481_SHEEP	COL1A1	Collagen type I alpha 1 chain	Predicted	Uncharacterised protein OS=Ovis aries GN=COL1A1 PE=4 SV=1	1.7
W5PDG3_SHEEP	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	Predicted	Glyceraldehyde-3-phosphate dehydrogenase OS=Ovis aries GN=GAPDH PE=3 SV=1	1.7
W5QBB1_SHEEP	C8G	Complement component C8 gamma chain, partial	Predicted	Uncharacterised protein OS=Ovis aries PE=4 SV=1	1.7

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5P880_SHEEP	PRG4	Proteoglycan 4	Predicted	Uncharacterised protein OS=Ovis aries GN=PRG4 PE=4 SV=1	1.7
W5PHI7_SHEEP	LOC101116892	Serpin A3-1-like	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101116892 PE=3 SV=1	1.7
W5Q868_SHEEP	RFX4	Regulatory factor X4	Predicted	Uncharacterised protein OS=Ovis aries GN=RFX4 PE=4 SV=1	1.6
W5NSA6_SHEEP	A2M	Alpha-2-macroglobulin	Predicted	Uncharacterised protein OS=Ovis aries GN=A2M PE=4 SV=1	1.6
W5P4C6_SHEEP	F12	Coagulation factor XII	Predicted	Uncharacterised protein OS=Ovis aries GN=F12 PE=3 SV=1	1.6
W5QH54_SHEEP	FETUB	Fetuin B	Predicted	Uncharacterised protein OS=Ovis aries GN=FETUB PE=4 SV=1	1.6
W5PJZ1_SHEEP	SERPING1	Serpin family G member 1	Predicted	Uncharacterised protein OS=Ovis aries GN=SERPING1 PE=3 SV=1	1.6
W5NXW9_SHEEP	IGHM	Transcription factor binding to IGHM enhancer 3	Predicted	Uncharacterised protein OS=Ovis aries GN=IGHM PE=4 SV=1	1.6
W5PF69_SHEEP	C2CD4C	C2 calcium dependent domain containing 4C	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=C2CD4C PE=4 SV=1	1.6
F2YQ13_SHEEP	GSN	Gelsolin	Predicted	Gelsolin isoform b OS=Ovis aries GN=GSN PE=2 SV=1	1.6

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5PZF0_SHEEP	PF4	Platelet factor 4	Predicted	C-X-C motif chemokine OS=Ovis aries GN=PF4 PE=3 SV=1	1.6
W5NQW4_SHEEP	LOC101104482	alpha-1-macroglobulin-like	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101104482 PE=4 SV=1	1.6
W5Q5A6_SHEEP	FGG	Fibrinogen gamma chain	Predicted	Uncharacterised protein OS=Ovis aries GN=FGG PE=4 SV=1	1.5
W5P8R7_SHEEP	FCGBP	Fc fragment of IgG binding protein	Predicted	Uncharacterised protein OS=Ovis aries GN=FCGBP PE=4 SV=1	1.5
O46544_SHEEP	C3	Complement C3	Predicted	Complement component C3 (Fragment) OS=Ovis aries GN=C3 PE=2 SV=1	1.5
W5PJ97_SHEEP	APOA2	Apolipoprotein A2	Predicted	Uncharacterised protein OS=Ovis aries GN=APOA2 PE=4 SV=1	1.5
W5PQK6_SHEEP	TLN1	Talin 1	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=TLN1 PE=4 SV=1	1.5
W5PXI3_SHEEP	AFM	Afamin	Predicted	Uncharacterised protein OS=Ovis aries GN=AFM PE=4 SV=1	1.5
W5PBY0_SHEEP	C4BPA	Complement component 4 binding protein alpha	Predicted	Uncharacterised protein OS=Ovis aries GN=C4BPA PE=4 SV=1	1.5
W5PVL4_SHEEP	MBL2	Mannose-binding lectin (protein C) 2, soluble	Predicted	Uncharacterised protein OS=Ovis aries GN=MBL2 PE=4 SV=1	1.5

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5NSF6_SHEEP	RARRES2	Retinoic acid receptor responder 2	Predicted	Uncharacterised protein OS=Ovis aries GN=RARRES2 PE=4 SV=1	1.5
W5PNS6_SHEEP	TDRD7	Tudor domain containing 7	Predicted	Uncharacterised protein OS=Ovis aries GN=TDRD7 PE=4 SV=1	1.5
W5P663_SHEEP	ENO3	Enolase 3	Predicted	Uncharacterised protein OS=Ovis aries GN=ENO3 PE=3 SV=1	1.5
W5PID9_SHEEP	C9	Complement C9	Predicted	Uncharacterised protein OS=Ovis aries GN=C9 PE=4 SV=1	1.5
Q06AV9_SHEEP	CD14	CD14 molecule	Predicted	Monocyte differentiation antigen CD14 OS=Ovis aries GN=CD14 PE=2 SV=1	1.4
W5QHZ5_SHEEP	LOC106990772	Ig kappa chain - sheep	Predicted	Uncharacterised protein OS=Ovis aries PE=4 SV=1	1.4
W5NWX6_SHEEP	APOC3	Apolipoprotein C3	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=APOC3 PE=4 SV=1	1.4
W5NX51_SHEEP	APOA1	Apolipoprotein A-1	Predicted	Uncharacterised protein OS=Ovis aries GN=APOA1 PE=3 SV=1	1.4
W5P7S6_SHEEP	ORM1	Orosomucoid 1	Predicted	Alpha-1-acid glycoprotein OS=Ovis aries GN=ORM1 PE=3 SV=1	1.4
W5PJR0_SHEEP	LOC101120613	Serum amyloid A-4 protein	Predicted	Serum amyloid A protein OS=Ovis aries GN=LOC101120613 PE=3 SV=1	1.4

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5NV14_SHEEP	IGL	immunoglobulin V lambda chain	Predicted	Uncharacterised protein OS=Ovis aries PE=4 SV=1	1.4
W5Q4Q3_SHEEP	SERPIND1	serpin family D member 1	Predicted	Uncharacterised protein OS=Ovis aries GN=SERPIND1 PE=3 SV=1	1.4
W5PQI4_SHEEP	EDEM2	ER degradation enhancing alpha-mannosidase like protein 2	Predicted	alpha-1,2-Mannosidase OS=Ovis aries GN=EDEM2 PE=3 SV=1	1.4
W5PCA0_SHEEP	ALDOB	Aldolase, fructose-bisphosphate B	Predicted	Fructose-bisphosphate aldolase OS=Ovis aries GN=ALDOB PE=3 SV=1	1.3
W5PCG1_SHEEP	TTC39C	Tetratricopeptide repeat domain 39C	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=TTC39C PE=4 SV=1	1.3
W5NY46_SHEEP	PON1	Paraoxonase 1	Predicted	Uncharacterised protein OS=Ovis aries GN=PON1 PE=4 SV=1	1.3
W5NQ46_SHEEP	FGB	Fibrinogen beta chain	Predicted	Fibrinogen beta chain OS=Ovis aries GN=FGB PE=4 SV=1	1.3
W5PYG2_SHEEP	VNN1	Vanin 1	Predicted	Uncharacterised protein OS=Ovis aries GN=VNN1 PE=4 SV=1	1.3
W5QA64_SHEEP	LOC101119895	Protein HP-20 homolog	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101119895 PE=4 SV=1	1.3
W5PD62_SHEEP	CPB2	Carboxypeptidase B2	Predicted	Uncharacterised protein OS=Ovis aries GN=CPB2 PE=4 SV=1	1.3

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5P5I0_SHEEP	CFI	Complement factor I	Predicted	Uncharacterised protein OS=Ovis aries GN=CFI PE=3 SV=1	1.3
W5P6F4_SHEEP	C5	Complement C5	Predicted	Uncharacterised protein OS=Ovis aries GN=C5 PE=4 SV=1	1.3
W5PDR5_SHEEP	C8A	Complement C8 alpha chain	Predicted	Uncharacterised protein OS=Ovis aries GN=C8A PE=4 SV=1	1.3
W5QCP9_SHEEP	COL6A3	Collagen type VI alpha 3 chain	Predicted	Uncharacterised protein OS=Ovis aries GN=COL6A3 PE=4 SV=1	1.3
W5NTT7_SHEEP	COL1A2	Collagen type I alpha 2 chain	Predicted	Uncharacterised protein OS=Ovis aries GN=COL1A2 PE=4 SV=1	1.3
W5PD71_SHEEP	CRP	Pentaxin	Predicted	Pentaxin OS=Ovis aries GN=CRP PE=3 SV=1	1.3
W5PTR4_SHEEP	SERPINC1	Serpin family C member 1	Predicted	Antithrombin-III OS=Ovis aries GN=SERPINC1 PE=3 SV=1	1.3
A6NBZ0_SHEEP	C3	Complement component C3d	Predicted	Complement component 3d (Fragment) OS=Ovis aries PE=2 SV=1	1.3
W5P101_SHEEP	A1BG	Alpha-1-B glycoprotein	Predicted	Uncharacterised protein OS=Ovis aries GN=A1BG PE=4 SV=1	1.3
W5PDG4_SHEEP	HMCN2	Hemicentin 2	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=HMCN2 PE=4 SV=1	1.2

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5PFP1_SHEEP	LTF	Lactotransferrin	Predicted	Uncharacterised protein OS=Ovis aries GN=LTF PE=3 SV=1	1.2
A2P2H6_SHEEP	VH	Ig mu heavy chain V region precursor	Predicted	VH region (Fragment) OS=Ovis aries GN=VH PE=2 SV=1	1.2
W5PDE5_SHEEP	LOC101120001	Zona pellucida sperm-binding protein 3 receptor-like	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101120001 PE=4 SV=1	1.2
W5PLQ1_SHEEP	LOC101111190	Hepatitis A virus cellular receptor 1-like	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101111190 PE=4 SV=1	1.2
W5NYF4_SHEEP	PGLYRP2	Peptidoglycan recognition protein 2	Predicted	Uncharacterised protein OS=Ovis aries GN=PGLYRP2 PE=4 SV=1	1.2
W5PFC9_SHEEP	LOC101117129	inhibitor of carbonic anhydrase-like	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=LOC101117129 PE=4 SV=1	1.2
W5Q9D5_SHEEP	VTN	Vitronectin	Predicted	Uncharacterised protein OS=Ovis aries GN=VTN PE=4 SV=1	1.2
W5QH45_SHEEP	KNG1	Kininogen 1	Predicted	Uncharacterised protein OS=Ovis aries GN=KNG1 PE=4 SV=1	1.2
W5PDJ6_SHEEP	GPX3	Glutathione peroxidase 3	Predicted	Glutathione peroxidase (Fragment) OS=Ovis aries GN=GPX3 PE=3 SV=1	1.2

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5QH50_SHEEP	HRG	Histidine rich glycoprotein	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=HRG PE=4 SV=1	1.2
W5QAB1_SHEEP	HPX	Hemopexin	Predicted	Uncharacterised protein OS=Ovis aries GN=HPX PE=4 SV=1	1.2
APOE_SHEEP	APOE	Apolipoprotein E	Exists	Apolipoprotein E OS=Ovis aries GN=APOE PE=2 SV=1	1.2
D6PZY4_SHEEP	fH	Fumarate hydratase	Predicted	Factor H (Fragment) OS=Ovis aries GN=fH PE=2 SV=1	1.2
TTHY_SHEEP	TTR	Transthyretin	Exists	Transthyretin OS=Ovis aries GN=TTR PE=2 SV=1	1.2
W5Q6C4_SHEEP	GCLM	Glutamate-cysteine ligase modifier subunit	Predicted	Uncharacterised protein OS=Ovis aries GN=GCLM PE=4 SV=1	1.2
W5NY95_SHEEP	C2	Complement C2	Predicted	Uncharacterised protein OS=Ovis aries GN=C2 PE=3 SV=1	1.2
W5PF71_SHEEP	KPNB1	Karyopherin subunit beta 1	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=KPNB1 PE=4 SV=1	1.2
W5PG63_SHEEP	VWF	von Willebrand factor	Predicted	Uncharacterised protein OS=Ovis aries GN=VWF PE=4 SV=1	1.2

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5QJ65_SHEEP	GALNT16	Polypeptide N-acetylgalactosaminyltransferase 16	Predicted	Polypeptide N-acetylgalactosaminyltransferase (Fragment) OS=Ovis aries GN=GALNT16 PE=3 SV=1	1.2
W5PXX3_SHEEP	F13B	Coagulation factor XIII B chain	Predicted	Uncharacterised protein OS=Ovis aries GN=F13B PE=4 SV=1	1.2
W5NRR7_SHEEP	SERPINA7	Serpin family A member 7	Predicted	Thyroxine-binding globulin OS=Ovis aries GN=SERPINA7 PE=3 SV=1	1.2
W5PGZ8_SHEEP	APOF	Apolipoprotein F	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=APOF PE=4 SV=1	1.2
W5PVM3_SHEEP	MYOC	Myocilin	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=MYOC PE=4 SV=1	1.2
W5P3B1_SHEEP	PNLDC1	PARN like, ribonuclease domain containing 1	Predicted	Uncharacterised protein OS=Ovis aries GN=PNLDC1 PE=4 SV=1	1.2
W5PF65_SHEEP	TF	Serotransferrin	Predicted	Uncharacterised protein OS=Ovis aries GN=TF PE=3 SV=1	1.2
W5QBW5_SHEEP	LBP	Lipopolysaccharide binding protein	Predicted	Uncharacterised protein OS=Ovis aries GN=LBP PE=4 SV=1	1.2
W5NUX8_SHEEP	C4	Complement C4 precursor	Predicted	Uncharacterised protein OS=Ovis aries PE=4 SV=1	1.2

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5PSQ7_SHEEP	IGL	Immunoglobulin lambda light chain	Predicted	Uncharacterised protein OS=Ovis aries PE=4 SV=1	1.2
W5QA54_SHEEP	LOC101119629	Protein HP-25 homolog 2	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101119629 PE=4 SV=1	1.2
I1WXR3_SHEEP	SERPINA1	Serpin family A member 1	Predicted	Alpha-1-antitrypsin transcript variant 1 OS=Ovis aries GN=SERPINA1 PE=2 SV=1	1.1
W5PLC2_SHEEP	CSN2	casein beta	Predicted	Beta-casein OS=Ovis aries GN=CSN2 PE=3 SV=1	1.1
W5PZG5_SHEEP	OCRL	OCRL, inositol polyphosphate-5-phosphatase	Predicted	Uncharacterised protein OS=Ovis aries GN=OCRL PE=4 SV=1	1.1
W5NTW3_SHEEP	ITIH1	Inter-alpha-trypsin inhibitor heavy chain 1	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=ITIH1 PE=4 SV=1	1.1
W5PHU4_SHEEP	UBAP1	Ubiquitin associated protein 1	Predicted	Uncharacterised protein OS=Ovis aries GN=UBAP1 PE=4 SV=1	1.1
W5P812_SHEEP	AMBP	Alpha-1-microglobulin/bikunin precursor	Predicted	Uncharacterised protein OS=Ovis aries GN=AMBP PE=4 SV=1	1.1
W5PH45_SHEEP	NUP205	Nucleoporin 205	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=NUP205 PE=4 SV=1	1.1

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
ALBU_SHEEP	ALB	Albumin	Exists	Serum albumin OS=Ovis aries GN=ALB PE=1 SV=1	1.1
W5PGT9_SHEEP	IGHE	Immunoglobulin epsilon-chain	Predicted	Uncharacterised protein OS=Ovis aries GN=IGHE PE=4 SV=1	1.1
X4ZFS1_SHEEP	AdipoQ	Adiponectin, C1Q and collagen domain containing	Predicted	Adiponectin OS=Ovis aries GN=AdipoQ PE=2 SV=1	1.1
W5PW21_SHEEP	ITIH2	Inter-alpha-trypsin inhibitor heavy chain 2	Predicted	Uncharacterised protein OS=Ovis aries GN=ITIH2 PE=4 SV=1	1.1
W5QC38_SHEEP	LOC101119975	Tubulin alpha-1A chain	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101119975 PE=1 SV=1	1.1
Q29439_SHEEP	C4	Complement 4	Predicted	Complement component C4 (Fragment) OS=Ovis aries GN=C4 PE=4 SV=1	1.1
W5P3J3_SHEEP	C1S	Complement C1s	Predicted	Uncharacterised protein OS=Ovis aries GN=C1S PE=3 SV=1	1.1
W5Q9A2_SHEEP	AZGP1	Alpha-2-glycoprotein 1, zinc-binding	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=AZGP1 PE=3 SV=1	1.1
W5PDN3_SHEEP	SPTBN2	Spectrin beta chain, non-erythrocytic 2	Predicted	Uncharacterised protein OS=Ovis aries GN=SPTBN2 PE=4 SV=1	1.1
W5PHP8_SHEEP	LRG1	Leucine rich alpha-2-glycoprotein 1	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=LRG1 PE=4 SV=1	1.1

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5PUE9_SHEEP	ZSCAN2	Zinc finger and SCAN domain containing 2	Predicted	Uncharacterised protein OS=Ovis aries GN=ZSCAN2 PE=4 SV=1	1.1
W5NZ47_SHEEP	RBP4	Retinol binding protein 4	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=RBP4 PE=4 SV=1	1.1
W5Q002_SHEEP	PLAG1	PLAG1 zinc finger	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=PLAG1 PE=4 SV=1	1.1
W5Q5D7_SHEEP	CCDC18	Coiled-coil domain containing 18	Predicted	Uncharacterised protein OS=Ovis aries GN=CCDC18 PE=4 SV=1	1.1
C8BKD1_SHEEP	F2	Coagulation factor II, thrombin	Predicted	Prothrombin OS=Ovis aries GN=F2 PE=2 SV=1	1.1
W5QI29_SHEEP	ECM1	Extracellular matrix protein 1	Predicted	Uncharacterised protein OS=Ovis aries GN=ECM1 PE=4 SV=1	1.1
W5QHZ8_SHEEP	IGK	Immunoglobulin kappa-4 light chain variable region	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries PE=4 SV=1	1.1
ANXA2_SHEEP	ANXA2	Annexin A2	Exists	Annexin A2 OS=Ovis aries GN=ANXA2 PE=1 SV=1	1.1
CERU_SHEEP	CP	Ceruloplasmin	Exists	Ceruloplasmin OS=Ovis aries GN=CP PE=2 SV=1	1.1
W5PN97_SHEEP	CLEC3B	C-type lectin domain family 3 member B	Predicted	Uncharacterised protein OS=Ovis aries GN=CLEC3B PE=4 SV=1	1.1

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5PSH4_SHEEP	LOC105614527	immunoglobulin lambda-6b light chain variable region [Ovis aries]	Predicted	Uncharacterised protein OS=Ovis aries PE=4 SV=1	1.1
W5PDQ9_SHEEP	C1QC	complement C1q C chain	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=C1QC PE=4 SV=1	1.1
W5PTS4_SHEEP	RNASE4	ribonuclease A family member 4	Predicted	Uncharacterised protein OS=Ovis aries GN=RNASE4 PE=3 SV=1	1.1
W5PIC9_SHEEP	BST1	bone marrow stromal cell antigen 1	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=BST1 PE=4 SV=1	1.1
W5NYJ9_SHEEP	AGT	Angiotensinogen (Fragment)	Predicted	Angiotensinogen (Fragment) OS=Ovis aries GN=AGT PE=3 SV=1	1.1
W5P791_SHEEP	ASXL3	Additional sex combs like 3, transcriptional regulator	Predicted	Uncharacterised protein OS=Ovis aries GN=ASXL3 PE=4 SV=1	1.1
W5P1J8_SHEEP	AOC3	Primary amine oxidase, lung isozyme	Predicted	Amine oxidase OS=Ovis aries GN=LOC101113086 PE=3 SV=1	1.1
W5P0Q4_SHEEP	LOC101102413	Haptoglobin	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101102413 PE=3 SV=1	1.1
W5Q961_SHEEP	LOC101107947	Apolipoprotein F-like	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=LOC101107947 PE=4 SV=1	1.0

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5PXI0_SHEEP	SHANK1	SH3 and multiple ankyrin repeat domains 1	Predicted	Uncharacterised protein OS=Ovis aries GN=SHANK1 PE=4 SV=1	1.0
W5QDP8_SHEEP	FBLN1	Fibulin 1	Predicted	Fibulin-1 OS=Ovis aries GN=FBLN1 PE=3 SV=1	1.0
W5NX96_SHEEP	ATRN	Attractin	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=ATRN PE=4 SV=1	1.0
W5PD84_SHEEP	F10	Coagulation factor X	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=F10 PE=3 SV=1	1.0
W5Q268_SHEEP	APOH	Apolipoprotein H	Predicted	Uncharacterised protein OS=Ovis aries GN=APOH PE=4 SV=1	1.0
W5PE53_SHEEP	C8B	complement C8 beta chain	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=C8B PE=4 SV=1	1.0
W5QDG8_SHEEP	FN1	Fibronectin 1	Predicted	Uncharacterised protein OS=Ovis aries GN=FN1 PE=4 SV=1	1.0
W5QAA1_SHEEP	SFN	Stratifin	Predicted	14-3-3 protein sigma OS=Ovis aries GN=SFN PE=3 SV=1	1.0
W5PTL2_SHEEP	CFP	Complement factor properdin	Predicted	Uncharacterised protein OS=Ovis aries GN=CFP PE=4 SV=1	1.0
W5P3Q3_SHEEP	LOC100101238	Regakine 1-like protein	Predicted	C-C motif chemokine OS=Ovis aries GN=LOC100101238 PE=3 SV=1	1.0

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5NX95_SHEEP	IGL	Ig lambda chain V-I region NIG-64	Predicted	Uncharacterised protein OS=Ovis aries PE=4 SV=1	1.0
W5P1X9_SHEEP	ALDOA	Aldolase, fructose-bisphosphate A	Predicted	Fructose-bisphosphate aldolase OS=Ovis aries GN=ALDOA PE=3 SV=1	1.0
A4ZVY9_SHEEP	B2M	Beta-2-microglobulin	Predicted	Beta-2-microglobulin OS=Ovis aries GN=B2M PE=2 SV=1	1.0
W5Q2T6_SHEEP	IGF1	Insulin like growth factor 1	Predicted	Insulin-like growth factor I OS=Ovis aries GN=IGF1 PE=3 SV=1	1.0
W5Q9W2_SHEEP	IGL	Immunoglobulin lambda-2b light chain variable region	Predicted	Uncharacterised protein OS=Ovis aries PE=4 SV=1	1.0
W5QH56_SHEEP	AHSG	Alpha-2-HS-glycoprotein (Fragment)	Predicted	Alpha-2-HS-glycoprotein (Fragment) OS=Ovis aries GN=AHSG PE=4 SV=1	1.0
W5P9V5_SHEEP	PIGR	Polymeric immunoglobulin receptor	Predicted	Uncharacterised protein OS=Ovis aries GN=PIGR PE=4 SV=1	1.0
W5PHP7_SHEEP	LOC101117146	Serpin A3-7-like	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101117146 PE=3 SV=1	1.0
W5NRG7_SHEEP	ITIH4	Inter-alpha-trypsin inhibitor heavy chain family member 4	Predicted	Uncharacterised protein OS=Ovis aries GN=ITIH4 PE=4 SV=1	1.0
MYG_SHEEP	MB	Myoglobin	Exists	Myoglobin OS=Ovis aries GN=MB PE=1 SV=2	1.0

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5PXU6_SHEEP	SERPINF1	Serpin family F member 1	Predicted	Uncharacterised protein OS=Ovis aries GN=SERPINF1 PE=3 SV=1	1.0
W5QGP4_SHEEP	APOD	Apolipoprotein D	Predicted	Uncharacterised protein OS=Ovis aries GN=APOD PE=3 SV=1	1.0
W5NWM2_SHEEP	APOA4	Apolipoprotein A4	Predicted	Uncharacterised protein OS=Ovis aries GN=APOA4 PE=3 SV=1	1.0
IBP4_SHEEP	IGFBP4	Insulin like growth factor binding protein 4	Predicted	Insulin-like growth factor-binding protein 4 OS=Ovis aries GN=IGFBP4 PE=1 SV=1	1.0
Q1A2D1_SHEEP	HBBK	Beta-K globin chain	Predicted	Beta-K globin chain OS=Ovis aries GN=HBBK PE=3 SV=1	1.0
W5QA07_SHEEP	LOC101119384	Protein HP-25 homolog 1	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101119384 PE=4 SV=1	1.0
W5PH95_SHEEP	IGH1	Immunoglobulin gamma-1 chain	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries PE=4 SV=1	1.0
B7SH79_SHEEP	NFIL3	Nuclear factor, interleukin 3 regulated	Predicted	Nuclear factor interleukin-3-regulated protein OS=Ovis aries GN=NFIL3 PE=2 SV=1	1.0
W5Q517_SHEEP	PCOLCE	Procollagen C-endopeptidase enhancer	Predicted	Uncharacterised protein OS=Ovis aries GN=PCOLCE PE=4 SV=1	1.0

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5P1W2_SHEEP	FOLR3	Folate receptor alpha	Predicted	Uncharacterised protein OS=Ovis aries GN=FOLR3 PE=4 SV=1	1.0
W5PQ53_SHEEP	ICOSLG	inducible T-cell costimulator ligand	Predicted	Uncharacterised protein OS=Ovis aries GN=ICOSLG PE=4 SV=1	0.9
W5NU33_SHEEP	TREML1	triggering receptor expressed on myeloid cells like 1	Predicted	Uncharacterised protein OS=Ovis aries GN=TREML1 PE=4 SV=1	0.9
W5QBV7_SHEEP	CD44	CD44 molecule (Indian blood group)	Predicted	Uncharacterised protein OS=Ovis aries GN=CD44 PE=4 SV=1	0.9
W5PTG9_SHEEP	GC	GC, vitamin D binding protein	Predicted	Uncharacterised protein OS=Ovis aries GN=GC PE=4 SV=1	0.9
W5Q5H8_SHEEP	FGA	Fibrinogen alpha chain	Predicted	Fibrinogen alpha chain OS=Ovis aries GN=FGA PE=4 SV=1	0.9
W5PH81_SHEEP	C7	Complement C7	Predicted	Uncharacterised protein OS=Ovis aries GN=C7 PE=4 SV=1	0.9
W5P3R3_SHEEP	PLG	Plasminogen	Predicted	Plasminogen OS=Ovis aries GN=PLG PE=3 SV=1	0.9
W5PZI1_SHEEP	LOC101113728	Clusterin	Predicted	Clusterin OS=Ovis aries GN=LOC101113728 PE=3 SV=1	0.9
W5Q7Z7_SHEEP	DSP	Desmoplakin	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=DSP PE=4 SV=1	0.9

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5PJA0_SHEEP	CUTA	CutA divalent cation tolerance homolog	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=CUTA PE=4 SV=1	0.9
W5PSM6_SHEEP	HABP2	Hyaluronan binding protein 2	Predicted	Uncharacterised protein OS=Ovis aries GN=HABP2 PE=3 SV=1	0.9
W5NV16_SHEEP	IGH	Immunoglobulin heavy chain precursor	Predicted	Uncharacterised protein OS=Ovis aries PE=4 SV=1	0.9
W5PPQ8_SHEEP	JCHAIN	Joining chain of multimeric IgA and IgM	Predicted	Uncharacterised protein OS=Ovis aries GN=JCHAIN PE=4 SV=1	0.9
W5QI15_SHEEP	IGK	Immunoglobulin kappa-1 light chain variable region	Predicted	Uncharacterised protein OS=Ovis aries PE=4 SV=1	0.8
W5P229_SHEEP	MASP2	Mannan binding lectin serine peptidase 2	Predicted	Uncharacterised protein OS=Ovis aries GN=MASP2 PE=3 SV=1	0.8
W5PTU7_SHEEP	CA2	Carbonic anhydrase 2	Predicted	Carbonic anhydrase 2 (Fragment) OS=Ovis aries GN=CA2 PE=4 SV=1	0.8
CBG_SHEEP	SERPINA6	Serpin family A member 6	Predicted	Corticosteroid-binding globulin OS=Ovis aries GN=SERPINA6 PE=2 SV=1	0.8
W5QHR3_SHEEP	RMND5A	Required for meiotic nuclear division 5 homolog A	Predicted	Uncharacterised protein OS=Ovis aries GN=RMND5A PE=4 SV=1	0.8
W5Q9K0_SHEEP	TTC25	Tetratricopeptide repeat domain 25	Predicted	Uncharacterised protein OS=Ovis aries GN=TTC25 PE=4 SV=1	0.8

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5QGG0_SHEEP	TFRC	Transferrin receptor	Predicted	Uncharacterised protein OS=Ovis aries GN=TFRC PE=4 SV=1	0.8
C0LQH2_SHEEP	IGFBP	Insulin like growth factor binding protein 2	Predicted	Insulin-like growth factor-binding protein-3 OS=Ovis aries GN=IGFBP-3 PE=2 SV=1	0.8
W5Q9P7_SHEEP	CFL1	Cofilin 1	Predicted	Cofilin-1 (Fragment) OS=Ovis aries GN=CFL1 PE=3 SV=1	0.8
W5PGT6_SHEEP	C6	Complement C6	Predicted	Uncharacterised protein OS=Ovis aries GN=C6 PE=4 SV=1	0.8
W5PLL2_SHEEP	F9	Coagulation factor IX	Predicted	Coagulation factor IX OS=Ovis aries GN=F9 PE=3 SV=1	0.8
HBBF_SHEEP	LOC101106199	Hemoglobin fetal subunit beta	Exists	Hemoglobin fetal subunit beta OS=Ovis aries PE=1 SV=1	0.8
W5PVH9_SHEEP	HGFAC	HGF activator	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=HGFAC PE=3 SV=1	0.8
W5QH21_SHEEP	MASP1	Mannan binding lectin serine peptidase 1	Predicted	Uncharacterised protein OS=Ovis aries GN=MASP1 PE=3 SV=1	0.8
W5PB46_SHEEP	PLTP	Phospholipid transfer protein	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=PLTP PE=4 SV=1	0.8

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5PAL4_SHEEP	LOC105612802	CD5 antigen-like isoform X1	Predicted	Uncharacterised protein OS=Ovis aries PE=4 SV=1	0.7
W5PLH8_SHEEP	KAZN	Kazrin, periplakin interacting protein	Predicted	Uncharacterised protein OS=Ovis aries GN=KAZN PE=4 SV=1	0.7
W5Q7J0_SHEEP	APOB	Apolipoprotein B	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=APOB PE=4 SV=1	0.7
W5PSK4_SHEEP	IGL	Immunoglobulin lambda-like polypeptide 1	Predicted	Uncharacterised protein OS=Ovis aries PE=4 SV=1	0.7
W5PTS2_SHEEP	MAMDC2	MAM domain containing 2	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=MAMDC2 PE=4 SV=1	0.7
W5Q6N9_SHEEP	AP3B1	Adaptor related protein complex 3 beta 1 subunit	Predicted	AP-3 complex subunit beta OS=Ovis aries GN=AP3B1 PE=3 SV=1	0.7
W5PXC8_SHEEP	SERPINF2	Serpin family F member 2	Predicted	Uncharacterised protein OS=Ovis aries GN=SERPINF2 PE=3 SV=1	0.7
W5QCY7_SHEEP	SPP2	Secreted phosphoprotein 2	Predicted	Secreted phosphoprotein 24 OS=Ovis aries GN=SPP2 PE=4 SV=1	0.7
D7RIF5_SHEEP	ACTB	Actin beta	Predicted	Beta-actin variant 2 OS=Ovis aries GN=ACTB PE=2 SV=1	0.6

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5PBX6_SHEEP	ARHGEF17	Rho guanine nucleotide exchange factor 17	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=ARHGEF17 PE=4 SV=1	0.6
W5PHM6_SHEEP	ZNF512B	Zinc finger protein 512B	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=ZNF512B PE=4 SV=1	0.6
W5NSH2_SHEEP	ITIH3	Inter-alpha-trypsin inhibitor heavy chain 3	Predicted	Uncharacterised protein OS=Ovis aries GN=ITIH3 PE=4 SV=1	0.6
W5PIW6_SHEEP	PRTN3	Myeloblastin	Predicted	Uncharacterised protein OS=Ovis aries GN=PRTN3 PE=3 SV=1	0.6
Q28745_SHEEP	HBA1	Alpha globin chain	Predicted	Alpha globin chain OS=Ovis aries PE=3 SV=1	0.6
W5Q7R8_SHEEP	JUP	Junction plakoglobin	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=JUP PE=4 SV=1	0.5
Q1KYZ7_SHEEP	HBBA	Haemoglobin, beta	Predicted	Beta-A globin chain OS=Ovis aries GN=HBBA PE=3 SV=1	0.5
W5PJ66_SHEEP	CFD	Complement factor D	Predicted	Uncharacterised protein OS=Ovis aries GN=CFD PE=3 SV=1	0.5
W5Q4Z3_SHEEP	IGFALS	Insulin like growth factor binding protein acid labile subunit	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=IGFALS PE=4 SV=1	0.5

UniProtKB Accession	Gene Name	NCBI Name	Protein Status	UniProtKB Name	Fold Change
W5QD16_SHEEP	MYL1	Myosin light chain 1	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=MYL1 PE=4 SV=1	0.5
W5QGQ3_SHEEP	LOC101102714	Lysozyme C, milk isozyme	Predicted	Uncharacterised protein OS=Ovis aries GN=LOC101102714 PE=3 SV=1	0.5
W5PSC8_SHEEP	FBLN5	Fibulin 5	Predicted	Uncharacterised protein OS=Ovis aries GN=FBLN5 PE=4 SV=1	0.5
W5Q0V2_SHEEP	BTD	Biotinidase	Predicted	Uncharacterised protein (Fragment) OS=Ovis aries GN=BTD PE=4 SV=1	0.4
W5NR06_SHEEP	LOC101108086	Histone H2B type 1	Predicted	Histone H2B OS=Ovis aries GN=LOC101108086 PE=3 SV=1	0.4
W5Q9K1_SHEEP	QSOX1	Quiescin sulfhydryl oxidase 1	Predicted	Sulfhydryl oxidase OS=Ovis aries GN=QSOX1 PE=4 SV=1	0.4
A0A0F6YFJ0_SHEEP	LOC100134870	Beta-C globin	Predicted	Beta-C globin OS=Ovis aries PE=3 SV=1	0.4
W5PQH2_SHEEP	ANPEP	Alanyl aminopeptidase, membrane	Predicted	Uncharacterised protein OS=Ovis aries GN=ANPEP PE=4 SV=1	0.3