DCU

Application of Advanced Liquid Chromatography Mass Spectrometry to Cancer and Biopharmaceutical Proteomics Research

A thesis submitted for the degree of Ph.D.

Michael Henry M.Sc.

January 2019

Supervisors: Dr. Paula Meleady
                 Prof. Martin Clynes

School of Biotechnology, Dublin City University, Glasnevin, Dublin, Ireland

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Degree of PhD is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____

(Michael Henry)

ID Number: 5016144

Date: _____

This thesis is dedicated to my father and mother

Acknowledgements

I would like to thank Dr. Paula Meleady as my supervisor for her guidance and patience throughout this thesis process. Without her, there would be no proteomics group and this research would not have been possible.

I would also like to thank Professor Martin Clynes for giving me an opportunity in 1996 to become a Research Assistant in the N.C.T.C.C. and start me on the path to becoming a research scientist. Professor Clynes' guidance as supervisor and mentor have been invaluable to me over the last 22 years.

I would very much like to thank all the researchers involved in this thesis: Dr. Clair Gallagher, Prashant, Dr. Martin Power, Dr. Kay Reen, and Dr. Dermot O'Sullivan and I would especially like to thank both Dr. Paul Dowling and Dr. Annemarie Larkin. A special mention to Orla Coleman in our proteomics group for putting up with me talking LC-MS/MS and MS data analysis for hours on end.

No research is possible in our Centre without the backbone that is the administrative and technical team. Historically with Joe Carey, Yvonne Reilly, Carol MacNamara and Emer Walsh but a special thanks goes to Gillian Smith and Mairead Callan.

I would like to mention a special thanks to the coffee and lunch gang of Denis and Finbarr.

Family and Friends are the most important people in life. I would like to thank Cora, Shea, Mush, Gareth, Rachel, Dave, John and Denis for being there for me always by going out for dinner, drinking, playing golf, going to gigs and playing live music. My own family of Catherine, Jim and my Mother Maura who have always put up with me and especially Paula for making me feel so welcome in her Family.

Finally, I woud like to thank my Father who I miss so much. I always want to make you proud.

Thesis Outline

Chapter 1.

This chapter gives an overview of current proteomics with emphasis on sample preparation, post-translational modifications, mass spectrometry and proteomic data analysis for cancer and biopharmaceutical proteomic research. This review examines mass spectrometry approaches for differential proteome quantitation. Furthermore, this review highlights recent advances in mass spectrometry for the analysis of proteins and peptides.

Chapter 2.

This chapter describes two protocol methods to analyse clinical patient samples using a resin-based depletion column followed by either protein In-Gel enzymatic digestion or protein In-Solution enzymatic digestion and then analysis by One-Dimensional Reverse Phase Chromatography or Two-Dimensional Strong Cation Exchange (SCX) – Reverse Phase Chromatography (RPC).

Chapter 3.

This chapter provides protocol methods for phosphopeptide enrichment using TiO_2 and IMAC strategies from CHO cells followed by LC-MS analyses (MS2 and MS3) for site-specific determination of phosphorylated serine, threonine or tyrosine residues. The reversible phosphorylation of proteins on serine, threonine, and tyrosine residues is one of the most important post-translational modifications that regulate many biological processes. The phosphoproteome has not been studied in any great detail in recombinant Chinese hamster ovary (CHO) cells to date despite phosphorylation playing a crucial role in regulating many molecular and cellular processes relevant to bioprocess phenotypes including, for example, transcription, translation, growth, apoptosis, and signal transduction.

Chapter 4.

This chapter describes a human serum depletion strategy followed by quantitative label free LC-MS/MS proteomics to deliver an informative portrait of response to therapy of patients with Multiple myeloma. Multiple myeloma (MM) is the second most common blood cancer in adults; it is a neoplasm of terminally differentiated B-cells characterised by a clonal expansion of malignant plasma cells in the bone marrow. As blood measurements are frequently used in the clinic in the assessment of patient health, proteomic workflows for serum/plasma samples to characterise possible disease-related biomarkers with the goal of better/improved diagnostic specificities is of significant importance. 17 MM patient serum samples (9 patients who subsequently responded to treatment with bortezomib / 8 who subsequently failed to respond) were used in this study. In this study, a number of proteins were found to be significantly changed in abundance between responders and non-responders to treatment.

Chapter 5.

Chapter five describes a protein and metabolite expression profiling study in blood using quantitative label free LC-MS/MS and high-throughput antibody-based screening platforms on serum from patients with breast cancer compared to healthy patient controls. Previously identified blood-based breast cancer biomarkers, including cancer antigen 15.3 (CA15-3) are useful in combination with imaging (computed tomography scans, magnetic resonance imaging, X-rays) and physical examination for monitoring tumour burden in advanced breast cancer patients. The aim of this study was to identify a biomarker signature of tumour burden using cancer and non-cancer (healthy controls/non-malignant breast disease) patient samples. Results from both mass spectrometry and high-throughput antibody screening demonstrated that combinations of three candidate biomarkers: Glutamate, 12-Hydroxyeicosatetraenoic acid, Beta-hydroxybutyrate, Factor V and Matrix metalloproteinase-1 with CA15-3, were found to mirror tumour burden when comparing non-malignant breast disease to the different stages of breast cancer.

Chapter 6.

This chapter describes proteomic techniques to identify novel proteins associated with cancer cell invasion/metastasis by identifying the unknown targets of monoclonal antibodies (MAbs) shown to functionally block invasion by cancer cells. Having generated a MAb following immunisation of mice with a clonal population of the pancreatic ductal adenocarcinoma cell line (Mia PaCa-2), the objective was to identify membrane targets involved in cancer invasion using immunoprecipitation and LC-MS/MS with this antibody against the Mia PaCa-2 cell line, a Mia PaCa-2 clone 3 (which displays high levels of invasion *in vitro*) and a squamous lung carcinoma cell line (DLKP-M). The immunoprecipitated proteins were identified as heterodimeric Ku antigen, Ku70/80.

Chapter 7.

Chapter seven describes the application of the methods described in chapter three to CHO cells that have been subjected to temperature reduction during the exponential phase of growth which is a strategy commonly employed by the biopharmaceutical industry to increase product yield. Using two different peptide phosphorylation enrichment strategies (Fe-NTA and TiO_2), we carried out quantitative label free LC-MS/MS to identify site specific differential phosphorylation in CHO cells using the Orbitrap XL Mass Spectrometer. From this analysis we identified 384 differentially expressed phosphopeptides (182 up and 202 down at 31°C) from the TiO_2 enriched samples, and 504 differentially expressed phosphopeptides from the Fe-NTA enriched samples (184 up and 320 down at 31°C). The results suggest significant enrichment of biological processes related to growth, ribosomal biogenesis, and cytoskeleton organization, and molecular functions related to RNA binding, transcription factor activity, and protein serine/threonine kinase activity in the cells following temperature adjustment.

Chapter 8

Chapter eight in this thesis describes the use of immobilised metal affinity chromatography (IMAC) on CHO cells in culture over time for Phosphopeptide enrichment followed by label free LC-MS/MS using the Orbitrap Fusion Mass Spectrometer demonstrating a significant increase in the number of site specific phosphorylation identifications when compared to our previous publication (chapter seven). In total over the various growth phases, we have identified 3,777 differentially expressed unique phosphopeptides from 1,415 differentially expressed unique phosphoproteins. Analysis of the whole cell lysate without phosphopeptide enrichment over the various growth phases revealed the differential expression of 834 unique proteins, with an overlap of 188 proteins between the proteomic and phosphoproteomic analyses. The inclusion of phosphoproteomic data can significantly improve proteome coverage and also give insights into the post-translational level of regulation during cellular growth of recombinant CHO cells.

Chapter 9

Chapter nine in this thesis applied differential quantitative label-free LC-MS/MS proteomic analysis to group clonally-derived cell lines (CDCLs) based on the level of clipping of the Fc-fusion protein they produced. The analysis was carried out over two times points in culture and clones were designated as either having 'high' or 'low' clipping phenotypes. Using quantitative label-free LC-MS/MS analysis we identified 200 differentially expressed proteins between the two experimental groups, and bioinformatic analysis suggested of the data suggests a role for defective protein folding and unfolded protein response.

Research Output

Chapter	Publication title	Publication status	Authors	Contribution
2	Clinical Proteomics: Liquid Chromatography-Mass Spectrometry Purification Systems.	Methods Mol Biol. 2017;1485:375-388	Henry. M. and Meleady. P.	First author, preparation and creation of the published work, specifically writing the initial draft.
3	Phosphopeptide Enrichment and LC-MS/MS Analysis to Study the Phosphoproteome of Recombinant Chinese Hamster Ovary Cells.	Methods Mol Biol. 2017;1603:195-208	Henry. M., Coleman. O., Prashant., Clynes. M. and Meleady. P.	First author, preparation and creation of the published work, specifically writing the initial draft.
4	Novel panel of protein biomarkers to predict response to bortezomib-containing induction regimens in multiple myeloma patients.	BBA Clin. 2017 Jun 7;8:28-34. 10.1016/j.bbacli.2017.05.003	Ting. KR., Henry. M., Meiller. J., Larkin. A., Clynes. M., Meleady. P., Bazou. D., Dowling. P. and O'Gorman. P.	Joint first author, methodology of serum depletion strategies, protein digestion and LC-MS/MS analysis. Conducted the label free differential expression analysis

5	Metabolomic and proteomic analysos of breast cancer patient samples suggests that glutamate and 12-HETE in combination with CA15-3 may be useful biomarkers reflecting tumour burden.	Metabolomics (2015) 11: 620.	Dowling. P., Henry. M., Meleady. P., Clarke. C., Gately. K., O'Byrne. K., Connolly. E., Lynch. V., Ballot. J., Giuseppe. G., Crown. J., Moriarty. M and Clynes. M	Joint first author, methodology of serum depletion strategies, specifically performed the protein digestion and LC-MS/MS analysis. Conducted the label free differential expression analysis
6	7B7: a novel antibody directed against the Ku70/Ku80 heterodimer blocks invasion in pancreatic and lung cancer cells.	Tumour Biol. 2014 Jul;35(7):6983-97	O'Sullivan. D., Henry. M., Joyce. H., Walsh. N., Mc Auley. E., Dowling. P., Swan. N., Moriarty. M., Barnham. P., Clynes. M and Larkin. A.	Joint first author, specifically performed the immunoprecipitation technique used, performed SDS gel analysis, protein in-gel digestions, LC-MS/MS and protein identification data analysis.

7	Differential Phosphoproteomic Analysis of Recombinant Chinese Hamster Ovary Cells Following Temperature Shift.	J Proteome Res. 2017 Jul 7;16(7):2339-2358	Henry. M., Power. M., Kaushik. P., Coleman. O., Clynes. M and Meleady. P.	Joint first author, methodology of enrichment strategies for phosphopeptide enrichment. Specifically perfomed the LC-MS/MS analysis. Conducted the label free differential expression analysis. Preparataion and creation of the initial draft.
8	The expression pattern of the phosphoproteome is significantly changed during the growth phases of recombinant CHO cell culture.	Biotechnol J. 2018 Aug 4:e1700221. doi: 10.1002/biot.201700221	Kaushik. P., Henry. M., Clynes. M. and Meleady. P.	Joint first author, methodology for protein sample scale up. Specifically perfomed LC-MS/MS analysis. Conducted the label free differential expression analysis. Preparation and writing of the initial draft.

9	Clonal variation in Productivity and in Proteolytic clipping of an Fc fusion protein in CHO cells: proteomic analysis suggests a role for defective protein folding and UPR.	J Biotechnol. 2018 May 31;281:21-30. doi: 10.1016/j.jbiotec.2018.05.018.	Henry. M., Gallagher. C., Kelly. RM., Frye. CC., Osborne. MD., Brady. C., Barron. N., Clynes. M and Meleady. P.	First author, methodology for protein digests. LC-MS/MS analysis. Conducted the label free differential expression analysis. Preparation and writing the initial draft.

The role of the candidate here is accurate.

Signed: _Michael Henry_   Date: _28/11/18_   Student

Signed: _Paula Meleady_   Date: _28/11/18_   Principal supervisor

Signed: _Martin Clynes_   Date: _28/11/18_   Secondary supervisor

Signed: _Annemarie_   Date: _28/11/18_   Dr. Annemarie Larkin

Signed: _Paul Dowling_   Date: _28/11/18_   Dr. Paul Dowling

Publications thoughout the duration of this thesis:

1. **"The expression pattern of the phosphoproteome is significantly changed during the growth phases of recombinant CHO cell culture"**. Kaushik. P., Henry. M., Clynes. M and Meleady. P. First published: 04 August 2018 https://doi.org/10.1002/biot.201700221

2. **"Subproteomic profiling of sarcolemma from dystrophic mdx-4cv skeletal muscle".** Murphy S, Zweyer M, Henry M, Meleady P, Mundegar RR, Swandulla D, Ohlendieck K. Data Brief. 2018 Feb 16;17:980-993. doi: 10.1016/j.dib.2018.02.020. eCollection 2018 Apr. PubMed PMID: 29876454; PubMed Central PMCID: PMC5988407.

3. **"Clonal variation in productivity and proteolytic clipping of an Fc-fusion protein in CHO cells: Proteomic analysis suggests a role for defective protein folding and the UPR".** Henry M, Gallagher C, Kelly RM, Frye CC, Osborne MD, Brady CP, Barron N, Clynes M, Meleady P. J Biotechnol. 2018 May 31;281:21-30. doi: 10.1016/j.jbiotec.2018.05.018. [Epub ahead of print] PubMed PMID: 29860056.

4. **"Parallel mRNA, proteomics and miRNA expression analysis in cell line models of the intestine".World J Gastroenterol".** O'Sullivan F, Keenan J, Aherne S, O'Neill F, Clarke C, Henry M, Meleady P, Breen L, Barron N, Clynes M, Horgan K, Doolan P, Murphy R. 2017 Nov 7;23(41):7369-7386. doi:10.3748/wjg.v23.i41.7369. PubMed PMID: 29151691; PubMed Central PMCID:PMC5685843.

5. **"Proteomic analysis of the sarcolemma-enriched fraction from dystrophic mdx-4cv skeletal muscle".** Murphy S, Zweyer M, Henry M, Meleady P, Mundegar RR, Swandulla D, Ohlendieck K. J Proteomics. 2018 Feb 2. pii: S1874-3919(18)30046-0. doi: 10.1016/j.jprot.2018.01.015. [Epub ahead of print] PubMed PMID: 29408692.

6. **"Protein and chemotherapy profiling of extracellular vesicles harvested from therapeutic induced senescent triple negative breast cancer cells"**. Kavanagh EL, Lindsay S, Halasz M, Gubbins LC, Weiner-Gorzel K, Guang MHZ, McGoldrick A, Collins E, Henry M, Blanco-Fernández A, Gorman PO, Fitzpatrick P, Higgins MJ, Dowling P, McCann A. Oncogenesis. 2017 Oct 9;6(10):e388. doi: 10.1038/oncsis.2017.82. PubMed PMID: 28991260; PubMed Central PMCID: PMC5668881.

7. **"A novel inhibitory anti-invasive MAb isolated using phenotypic screening highlights AnxA6 as a functionally relevant target protein in pancreatic cancer".** O'Sullivan D, Dowling P, Joyce H, McAuley E, McCann A, Henry M, McGovern B, Barham P, Kelleher FC, Murphy J, Kennedy S, Swan N, Moriarty M, Clynes M, Larkin A. Br J Cancer. 2017 Oct 24;117(9):1326-1335. doi: 10.1038/bjc.2017.306. Epub 2017 Sep 7. PubMed PMID: 28881357; PubMed Central PMCID: PMC5672937.

8. **"Neutrophil Membrane Cholesterol Content is a Key Factor in Cystic Fibrosis Lung Disease"**. White MM, Geraghty P, Hayes E, Cox S, Leitch W, Alfawaz B, Lavelle GM, McElvaney OJ, Flannery R, Keenan J, Meleady P, Henry M, Clynes M, Gunaratnam C, McElvaney NG, Reeves EP. EBioMedicine. 2017 Sep;23:173-184. doi: 10.1016/j.ebiom.2017.08.013. Epub 2017 Aug 16. PubMed PMID: 28835336; PubMed Central PMCID: PMC5605378.

9. "**Proteomic profiling of the dystrophin complex and membrane fraction from dystrophic mdx muscle reveals decreases in the cytolinker desmoglein and increases in the extracellular matrix stabilizers biglycan and fibronectin**". Murphy S, Brinkmeier H, Krautwald M, Henry M, Meleady P, Ohlendieck K. J Muscle Res Cell Motil. 2017 Apr;38(2):251-268. doi: 10.1007/s10974-017-9478 4. Epub 2017 Aug 12. PubMed PMID: 28803268.

10. **"Novel panel of protein biomarkers to predict response to bortezomib-containing induction regimens in multiple myeloma patients"**. Henry M, Ting KR, Meiller J, Larkin A, Clynes M, Meleady P, Bazou D, Dowling P, O'Gorman P BBA Clin. 2017 Jun 7;8:28-34. doi: 10.1016/j.bbacli.2017.05.003

11. **"A novel inhibitory anti-invasive MAb isolated using phenotypic screening highlights AnxA6 as a functionally relevant target protein in pancreatic cancer"**. O'Sullivan D, Dowling P, Joyce H, McAuley E, McCann A, Henry M, McGovern B, Barham P, Kelleher FC, Murphy J, Kennedy S, Swan N, Moriarty M, Clynes M, Larkin A. Br J Cancer. 2017 Sep 7. doi: 10.1038/bjc.2017.306. [Epub ahead of print] PubMed PMID: 28881357.

12. **"Membrane Cholesterol Content is a Key Factor in Cystic Fibrosis Lung Disease"**. White MM, Geraghty P, Hayes E, Cox S, Leitch W, Alfawaz B, Lavelle GM, McElvaney OJ, Flannery R, Keenan J, Meleady P, Henry M, Clynes M, Gunaratnam C, McElvaney NG, Reeves EP. Neutrophil EBioMedicine. 2017 Sep;23:173-184. doi:10.1016/ j.ebiom.2017.08.013. Epub 2017 Aug 16. PubMed PMID: 28835336.

13. **"Proteomic profiling of the dystrophin complex and membrane fraction from dystrophic mdx muscle reveals decreases in the cytolinker desmoglein and increases in the extracellular matrix stabilizers biglycan and fibronectin"**. Murphy S, Brinkmeier H, Krautwald M, Henry M, Meleady P, Ohlendieck K. J Muscle Res Cell Motil. 2017 Aug 12. doi: 10.1007/s10974-017-9478-4. [Epub ahead of print] PubMed PMID: 28803268.

14. **"Differential Phosphoproteomic Analysis of Recombinant Chinese Hamster Ovary Cells Following Temperature Shift"**. Henry M, Power M, Kaushik P, Coleman O, Clynes M, Meleady P. J Proteome Res. 2017 Jul 7;16(7):2339-2358. doi:10.1021/acs. jproteome.6b00868. Epub 2017 Jun 6. PubMed PMID: 28509555

15. **"Novel panel of protein biomarkers to predict response to bortezomib-containing induction regimens in multiple myeloma patients"**. Henry M, Ting KR, Meiller J, Larkin A, Clynes M, Meleady P, Bazou D, Dowling P, O'Gorman P. BBA Clin. 2017 Jun 7;8:28-34. doi: 10.1016/j.bbacli.2017.05.003. eCollection 2017 Dec. PubMed PMID: 28725572; PubMed Central PMCID: PMC5502697.

16. **"Proteomic profiling of mdx-4cv serum reveals highly elevated levels of the inflammation-induced plasma marker haptoglobin in muscular dystrophy**". Murphy S, Dowling P, Zweyer M, Henry M, Meleady P, Mundegar RR, Swandulla D, Ohlendieck K. Int J Mol Med. 2017 Jun;39(6):1357-1370. doi: 10.3892/ ijmm.2017.2952. Epub 2017 Apr 18. PubMed PMID: 28440464; PubMed Central PMCID: PMC5428965.

17. **"Identification of the metabolic alterations associated with the multidrug resistant phenotype in cancer and their intercellular transfer mediated by extracellular vesicles"**. Lopes-Rodrigues V, Di Luca A, Mleczko J, Meleady P, Henry M, Pesic M, Cabrera D, van Liempd S, Lima RT, O'Connor R, Falcon-Perez JM, Vasconcelos MH. Sci Rep. 2017 Mar 17;7:44541. doi: 10.1038/srep44541. PubMed PMID: 28303926; PubMed Central PMCID: PMC5356019.

18. **"Residual urinary extracellular vesicles in ultracentrifugation supernatants after hydrostatic filtration dialysis enrichment"**. Musante L, Tataruch-Weinert D, Kerjaschki D, Henry M, Meleady P, Holthofer H. J Extracell Vesicles. 2016 Dec 23;6(1):1267896. doi: 10.1080/20013078.2016.1267896. eCollection 2017. PubMed PMID: 28326167; PubMed Central PMCID: PMC5328348.

19. **"Process-relevant concentrations of the leachable bDtBPP impact negatively on CHO cell production characteristics"**. Kelly PS, McSweeney S, Coleman O, Carillo S, Henry M, Chandran D, Kellett A, Bones J, Clynes M, Meleady P, Barron N. Biotechnol Prog. 2016 Nov;32(6):1547-1558. doi: 10.1002/btpr.2345. Epub 2016 Sep 4. PubMed PMID: 27557043.

20. **"Quantitative label-free mass spectrometry analysis of formalin-fixed, paraffin-embedded tissue representing the invasive cutaneous malignant melanoma proteome"**. Dowling P, Moran B, McAuley E, Meleady P, Henry M, Clynes M, McMenamin M, Leonard N, Monks M, Wynne B, Ormond P, Larkin A. Oncol Lett. 2016 Nov;12(5):3296-3304. Epub 2016 Sep 7. PubMed PMID: 27899996; PubMed Central PMCID: PMC5103945.

21. **"Proteomic differences in recombinant CHO cells producing two similar antibody fragments"**. Sommeregger W, Mayrhofer P, Steinfellner W, Reinhart D, Henry M, Clynes M, Meleady P, Kunert R. Biotechnol Bioeng. 2016 Sep;113(9):1902-12. doi: 10.1002/bit.25957. Epub 2016 Mar 16. PubMed PMID: 26913574; PubMed Central PMCID: PMC4985663.

22. **"Proteomic analysis of dystrophin deficiency and associated changes in the aged mdx-4cv heart model of dystrophinopathy-related cardiomyopathy"**. Murphy S, Dowling P, Zweyer M, Mundegar RR, Henry M, Meleady P, Swandulla D, Ohlendieck K. J Proteomics. 2016 Aug 11;145:24-36. doi: 10.1016/ j.jprot.2016.03.011. Epub 2016 Mar 4. PubMed PMID: 26961938.

23. **"The iron-responsive microsomal proteome of Aspergillus fumigatus"**. Moloney NM, Owens RA, Meleady P, Henry M, Dolan SK, Mulvihill E, Clynes M, Doyle S. J Proteomics. 2016 Mar 16;136:99-111. doi: 10.1016/j.jprot.2015.12.025. Epub 2015 Dec 29. PubMed PMID: 26739764.

24. **Multidrug resistant tumour cells shed more microvesicle-like EVs and less exosomes than their drug-sensitive counterpart cells**. Lopes-Rodrigues V, Di Luca A, Sousa D, Seca H, Meleady P, Henry M, Lima RT, O'Connor R, Vasconcelos MH. Biochim Biophys Acta. 2016 Mar;1860(3):618-27. doi: 10.1016/j.bbagen.2015.12.011. Epub 2015 Dec 17. PubMed PMID: 26708992.

25. **"Data supporting the shedding of larger extracellular vesicles by multidrug resistant tumour cells"**. Lopes-Rodrigues V, Di Luca A, Sousa D, Seca H, Meleady P, Henry M, Lima RT, O'Connor R, Vasconcelos MH. Data Brief. 2016 Feb 8;6:1023-7. doi: 10.1016/ j.dib.2016.02.004. eCollection 2016 Mar. PubMed PMID: 26958634; PubMed Central PMCID: PMC4763103.

26. **"Proteomic strategies in the search for novel pancreatic cancer biomarkers and drug targets: recent advances and clinical impact"**.
Coleman O, Henry M, McVey G, Clynes M, Moriarty M, Meleady P. Expert Rev Proteomics. 2016;13(4):383-94. doi: 10.1586/14789450.2016.1167601. Review. PubMed PMID: 26985644.

27. **"Label-free mass spectrometric analysis reveals complex changes in the brain proteome from the mdx-4cv mouse model of Duchenne muscular dystrophy"**. Murphy S, Zweyer M, Henry M, Meleady P, Mundegar RR, Swandulla D, Ohlendieck K. Clin Proteomics. 2015 Nov 23;12:27. doi: 10.1186/s12014-015-9099-0. eCollection 2015. PubMed PMID: 26604869; PubMed Central PMCID: PMC4657206.

28. **"Concurrent Label-Free Mass Spectrometric Analysis of Dystrophin Isoform Dp427 and the Myofibrosis Marker Collagen in Crude Extracts from mdx-4cv Skeletal Muscles"**. Murphy S, Zweyer M, Mundegar RR, Henry M, Meleady P, Swandulla D, Ohlendieck K. Proteomes. 2015 Sep 16;3(3):298-327. doi: 10.3390/ proteomes 3030298. PubMed PMID: 28248273; PubMed Central PMCID: PMC5217383.

29. **"The development of cisplatin resistance in neuroblastoma is accompanied by epithelial to mesenchymaltransition in vitro"**. Piskareva O, Harvey H, Nolan J, Conlon R, Alcock L, Buckley P, Dowling P, Henry M, O'Sullivan F, Bray I, Stallings RL. Cancer Lett. 2015 Aug 10;364(2):142-55. doi: 10.1016/j.canlet.2015.05.004. Epub 2015 May 7. Erratum in: Cancer Lett. 2015 Dec 28;369(2):428. Henry, Michael [added]. PubMed PMID: 25960282.

30. **"Label-free LC-MS analysis of HER2+ breast cancer cell line response to HER2 inhibitor treatment"**. Di Luca A, Henry M, Meleady P, O'Connor R. Daru. 2015 Aug 4;23:40. doi: 10.1186/s40199-015-0120-y. PubMed PMID: 26238995; PubMed Central PMCID: PMC4524286.

31. **"Re-programming CHO cell metabolism using miR-23 tips the balance towards a highly productive phenotype"**. Kelly PS, Breen L, Gallagher C, Kelly S, Henry M, Lao NT, Meleady P, O'Gorman D, Clynes M, Barron N. Biotechnol J. 2015 Jul;10(7):1029-40. doi: 10.1002/biot.201500101. Epub 2015 Jun 24. PubMed PMID: 26097147.

32. **"Label-free mass spectrometric analysis of the mdx-4cv diaphragm identifies the matricellular protein periostin as a potential factor involved in dystrophinopathy-related fibrosis"**. Holland A, Dowling P, Meleady P, Henry M, Zweyer M, Mundegar RR, Swandulla D, Ohlendieck K. Proteomics. 2015 Jul;15(13):2318-31. doi: 10.1002/pmic.201400471. Epub 2015 Apr 1. PubMed PMID: 25737063.

33. **"Comparative Label-Free Mass Spectrometric Analysis of Mildly versus Severely Affected mdx Mouse Skeletal Muscles Identifies Annexin, Lamin, and Vimentin as Universal Dystrophic Markers"**. Holland A, Henry M, Meleady P, Winkler CK, Krautwald M, Brinkmeier H, Ohlendieck K. Molecules. 2015 Jun 19;20(6):11317-44. doi: 10.3390/molecules200611317. PubMed PMID: 26102067.

34. **"Simultaneous Pathoproteomic Evaluation of the Dystrophin-Glycoprotein Complex and Secondary Changes in the mdx-4cv Mouse Model of Duchenne Muscular Dystrophy"**. Murphy S, Henry M, Meleady P, Zweyer M, Mundegar RR, Swandulla D, Ohlendieck K. Biology (Basel). 2015 Jun 10;4(2):397-423. doi: 10.3390/ biology4020397. PubMed PMID: 26067837; PubMed Central PMCID: PMC4498307.

35. **"Metabolomic and proteomic analysos of breast cancer patient samples suggests that glutamate and 12-HETE in combination with CA15-3 may be useful biomarkers reflecting tumour burden"**. Henry M, Dowling P, Meleady P, Clarke C, Gately K, O'Connolly E, Lynch V, Ballot J, Gullo G, Crown J, Moriarty M, , Clynes M. Metabolomics (2015) 11: 620. doi: 10.1007/s11306-014-0723-1

36. **"Abnormal levels of heterogeneous nuclear ribonucleoprotein A2B1 (hnRNPA2B1) in tumour tissue and blood samples from patients diagnosed with lung cancer"**. Dowling P, Pollard D, Larkin A, Henry M, Meleady P, Gately K, O'Byrne K, Barr MP, Lynch V, Ballot J, Crown J, Moriarty M, O'Brien E, Morgan R, Clynes M. Mol Biosyst. 2015 Mar;11(3):743-52. doi: 10.1039/c4mb00384e. Epub 2014 Dec 8. PubMed PMID: 25483567.

37. **"Elevated levels of 14-3-3 proteins, serotonin, gamma enolase and pyruvate kinase identified in clinical samples from patients diagnosed with colorectal cancer"**. Dowling P, Hughes DJ, Larkin AM, Meiller J, Henry M, Meleady P, Lynch V, Pardini B, Naccarati A, Levy M, Vodicka P, Neary P, Clynes M. Clin Chim Acta. 2015 Feb 20;441:133-41. doi: 10.1016/j.cca.2014.12.005. Epub 2014 Dec 23. PubMed PMID: 25540887.

38. **"Transferrin-bound proteins as potential biomarkers for advanced breast cancer patients"**. Dowling P, Palmerini V, <u>Henry M</u>, Meleady P, Lynch V, Ballot J, Gullo G, Crown J, Moriarty M, Clynes M. BBA Clin. 2014 Sep 16;2:24-30. doi: 10.1016/j.bbacli.2014.08.004. eCollection 2014 Dec. PubMed PMID: 26673961; PubMed Central PMCID: PMC4633920.

39. **"A neutrophil intrinsic impairment affecting Rab27a and degranulation in cystic fibrosis is corrected by CFTR potentiator therapy"**. Pohl K, Hayes E, Keenan J, <u>Henry M</u>, Meleady P, Molloy K, Jundi B, Bergin DA, McCarthy C, McElvaney OJ, White MM, Clynes M, Reeves EP, McElvaney NG. Blood. 2014 Aug 14;124(7):999-1009. doi: 10.1182/blood-2014-02-555268. Epub 2014 Jun 16. PubMedPMID: 24934256; PubMed Central PMCID: PMC4133506.

40. **"Identification and functional validation of RAD23B as a potential protein in human breast cancer progression"**. Linge A, Maurya P, Friedrich K, Baretton GB, Kelly S, <u>Henry M</u>, Clynes M, Larkin A, Meleady P. J Proteome Res. 2014 Jul 3;13(7):3212-22. doi: 10.1021/pr4012156. Epub 2014 Jun 19. PubMed PMID: 24897598.

41. **"Intricate effects of primary motor neuronopathy on contractile proteins and metabolic muscle enzymes as revealed by label-free mass spectrometry"**. Holland A, Schmitt-John T, Dowling P, Meleady P, <u>Henry M</u>, Clynes M, Ohlendieck K. Biosci Rep. 2014 Jul 1;34(4). pii: e00119. doi:10.1042/ BSR20140029. PubMed PMID: 24895011; PubMed Central PMCID: PMC4076836.

42. **"7B7: a novel antibody directed against the Ku70/Ku80 heterodimer blocks invasion in pancreatic and lung cancer cells"**. <u>Henry M</u>, O'Sullivan D, Joyce H, Walsh N, Mc Auley E, Dowling P, Swan N, Moriarty M, Barnham P, Clynes M, Larkin A. Tumour Biol. 2014 Jul;35(7):6983-97. doi: 10.1007/s13277-014-1857-5. Epub 2014 Apr 18. PubMed PMID: 24744142.

43. **"PP2A inhibition overcomes acquired resistance to HER2 targeted therapy"**. McDermott MS, Browne BC, Conlon NT, O'Brien NA, Slamon DJ, <u>Henry M</u>, Meleady P, Clynes M, Dowling P, Crown J, O'Donovan N. Mol Cancer. 2014 Jun 24;13:157. doi: 10.1186/1476-4598-13-157. PubMed PMID: 24958351; PubMed Central PMCID: PMC4230643.

44. **"Increased outer arm and core fucose residues on the N-glycans of mutated alpha-1 antitrypsin protein from alpha-1 antitrypsin deficient individuals"**. McCarthy C, Saldova R, O'Brien ME, Bergin DA, Carroll TP, Keenan J, Meleady P, <u>Henry M</u>, Clynes M, Rudd PM, Reeves EP, McElvaney NG. J Proteome Res. 2014 Feb 7;13(2):596-605. doi: 10.1021/pr400752t. Epub 2013 Dec 13. PubMed PMID: 24328305.

Book Chapters

"Clinical Proteomics: Liquid Chromatography-Mass Spectrometry (LC-MS) Purification Systems". <u>Henry M</u>, Meleady P. Methods Mol Biol. 2017;1485:375-388. PubMed PMID: 27730564.

"Filter-Aided Sample Preparation (FASP) for Improved Proteome Analysis of Recombinant Chinese Hamster Ovary Cells". Coleman O, <u>Henry M</u>, Clynes M, Meleady P. Methods Mol Biol. 2017;1603:187-194. doi: 10.1007/978-1-4939-6972-2_12. PubMed PMID: 28493131.

"Phosphopeptide Enrichment and LC-MS/MS Analysis to Study the Phosphoproteome of Recombinant Chinese Hamster Ovary Cells". <u>Henry M</u>, Coleman O, Prashant, Clynes M, Meleady P. Methods Mol Biol. 2017;1603:195-208. doi: 10.1007/978-1-4939-6972-2_13. PubMed PMID: 28493132

Posters/talks presented at international conferences

1. Michael Henry, Martin Power, Prashant, Martin Clynes and Paula Meleady. 'Characterisation of the phosphoproteome to understand and improve bioprocess-relevant phenotypes in recombinant Chinese hamster ovary cells'. Poster and talk presented at ProtemMMX 4.0 Strictly Quantitative, April 5-7 2016, Chester, UK. **Note: This poster was elevated to an early stage researcher oral presentation.**

2. Michael Henry, Martin Power, Prashant, Martin Clynes, Paula Meleady. 'Characterisation of the phosphoproteome to understand and improve bioprocess-relevant phenotypes in recombinant Chinese hamster ovary cells. Poster presented at 17th European Congress on Biotechnology, 3-6th July 2017, Krakow, Poland.

3. European Society for Animal Cell Technology, ESACT 2017 conference. 14-17th May 2017, Lausanne, Switzerland. 'Quantitative phosphoproteomic analysis of CHO cells in response to reduced culture temperature'. Henry M, Power M, Prashant, Coleman O, Clynes M, Meleady P.

Awards

DCU President's Awards for Engagement 2016 -Special award in Enterprise Engagement – Novel LC-MS/MS workflow for proteomic profiling of biological samples.

Table of Contents

Table of Figures

Table of Tables

List of abbreviations:

2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
ACN	Acetonitrile
ANOVA	Analysis of variance
AUC	Area under the curve
CA	Cancer antigen
CDCLs	Clonally-derived cell lines
CHO	Chinese hamster ovary
CHO SEAP	Chinese hamster ovary secreting alkaline phosphatase
CID	Collision Induced Dissociation
CRIGR	*Cricetulus griseus*
DDA	Data dependent acquisition
DDNL	Data dependent neutral loss
DIA	Data independent acquisition
DIGE	Difference in-gel electrophoresis
DTT	DL-Dithiothreitol
EBI	European Bioinformatics institute
ELISA	Enzyme-linked immunosorbent assays
ETD	Electron transfer dissociation
ESI	Electrospray ionisation
FA	Formic acid
Fc	Fragment crystallizable
FDR	False discovery rate
$Fe3^{+}$	Ferric ion
Fe-NTA	Ferric nitrilotriacetate
FT	Fourier Transform
FWHM	Full width half maximum
$Ga3^{+}$	Gallium(^{III}) ion
GO	Gene ontology

HCD	High Collision Dissociation
HCP	Host cell protein
HPLC	High pressure liquid chromatography
KCl	Potassium chloride
Ig	Immunoglobulin
IMAC	Immobilised metal ion affinity chromatography
iTRAQ	Isobaric tag for relative and absolute quantitation
ISS	International scoring system
IT	Ion trap
LC-MS	Liquid chromatography mass spectrometry
LF	Label free
LTQ	Linear trap quadrupole
MAb	Monoclonal antibody
MBC	Metastatic breast cancer
MOAC	Metal oxide immobilized metal ion affinity chromatography
MM	Multiple myeloma
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MS2	Tandem mass spectrometry
MSA	Multistage activation
m/z	Mass-to-charge
NCBI	National Centre of Biotechnology Information
OT	Orbitrap
PANTHER	Protein Analysis Through Evolutionary Relationships
PDAC	Pancreatic ductal adenocarcinoma
PTM	Post translational modification
PPI	Protein-protein interaction
PPM	Parts per million
PRM	Parallel reaction monitoring
PSM	Peptide spectral match

RPC	Reverse phase chromatography
SCX	Strong cation exchange
SIM	Single ion monitoring
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
SWATH	Sequential Windowed Acquisition
SRM	Single reaction monitoring
SUMO	Small ubiquitin-like modifier
TFA	Trifluoroacetic acid
TiO$_2$	Titanium oxide
TMT	Tandem mass tag
UPR	Unfolded protein response
WiSIM	Wide-single ion monitoring
XCorr	Cross-correlation

Abstract

Title: Application of Advanced Liquid Chromatography Mass Spectrometry to Cancer and Biopharmaceutical Proteomics Research

Author: Michael Henry

This thesis focuses on the application of state-of-the-art liquid chromatography mass spectrometry to cancer and biopharmaceutical proteomics research. The goal of cancer proteomic research is to improve early detection, diagnosis and response to treatment by identifying changes in protein expression during cancer progression, which have the potential to be biomarkers or therapeutic targets. The aim of biopharmaceutical proteomics research is to identify proteins involved in desirable recombinant Chinese hamster ovary (CHO) cellular phenotypes which have the potential to be manipulated to improve the efficiency of biotherapeutic production.

The major output from this thesis resulted in the development of world class expertise in the application of protein/peptide mass spectrometry to proteomic research. Highly reproducible HPLC methods were developed which enabled accurate differential expression label-free LC-MS/MS proteomic profiling. Managing the high mass accuracy mass spectrometers led to in-depth identification and characterisation of proteins and their modifications. Meticulous data analyses ensured that all proteomic outputs were of the highest quality.

The scientific output from the thesis resulted in three publications in cancer proteomics and three publications in CHO cell proteomics. In the cancer studies, two manuscripts involved biomarker discovery using serum from multiple myeloma and breast cancer patients, and the third involved the development of an immunoprecipitation and LC-MS/MS technique to identify two novel cancer proteins associated with invasion and metastasis. In the CHO proteomic studies one of the publications resulted in the identification of protein degradation targets that have the potential to be modified using cell line engineering to improve recombinant Fc-fusion protein quality. The final two publications involved the development of phosphorylation enrichment strategies to carry out the first phosphoproteome profiling studies in biopharma research involving recombinant CHO cells. This data significantly improved proteome coverage and provided insights into the post-translational level of regulation during cellular growth of recombinant CHO cells.

1   **CHAPTER 1**

Literature Survey

Author: <u>Michael Henry</u>

Literature survey on current proteomics with emphasis on sample preparation, post-translational modifications, mass spectrometry and proteomic data analysis for cancer and biopharmaceutical proteomic research. It examines mass spectrometry approaches for differential proteome quantitation. Furthermore, this review highlights recent advances in mass spectrometry for the analysis of proteins and peptides.

1.1   What is proteomics?

Proteins are involved in nearly all physiological aspects on cellular life. The misregulation of protein expression can result in pathological states like cancer. The use of expression systems utilising mammalian cells for the production of recombinant proteins that are folded correctly, have post-translational modifications and product assembly is also an essential requirement for the biopharmaceutical industry. Proteins are synthesised by translating the information encoded in RNA to a polypeptide chain which takes a specific three dimensional structure. Proteins are subjected to constant turnover and the balance between synthesis and degradation is one of many forms of regulation that are coordinated to achieve a unified cellular response. Accounting for protein turnover holds the potential to reveal crucial relationships of cellular regulation and is an essential component of systems-level models of cell behaviour (Hinkson et al. 2011).

Proteomics is a complex field of study concerned with applying the techniques of analytical chemistry, molecular biology, biochemistry, and genetics to analysing the structure, function, and interactions of the proteins produced by the genome of a particular cell, tissue, or organism. Although proteomic technologies have improved significantly over the intervening twenty years, the basic desire to characterise the proteome of a given cell, tissue or organism has not. Understanding the impact of the proteome in eukaryotes must take into account the fact that proteins are very complex molecules with post-translational modifications, trafficking, subcellular location, interactions and complex formation all contributing to biological function (Gonzalez et al. 2010).

The importance of studying the proteome of a biological sample is demonstrated by the fact that the analysis of mRNA expression profiles is not necessarily a direct reflection of the protein expression profile in the cell (Graves et al. 2002). As a result, many studies have not always shown a good correlation between mRNA and protein expression levels (Ideker et al. 2001, Abbott et al. 1999). The formation of mRNA is step one in a series of biological events which finally results in the synthesis of a protein and whether or not the protein is active (Figure 1-1).
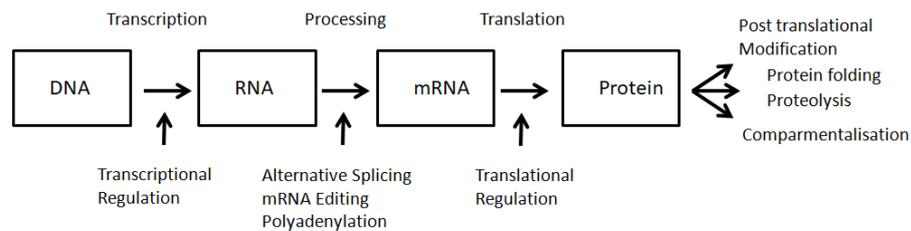
Figure 1-1. An example of how a single gene can give rise to multiple gene products. If RNA is edited or alternatively spliced to a more mature mRNA, multiple protein isoforms can be generated by RNA processing. Proteins can then by regulated by additional mechanisms including protein folding, post translational modifications, etc.

1.1.1 Protein localisation in Eukaryotes

A goal in cell biology is to identify where specific proteins are distributed subcellularly and then to characterise if and how protein localisation changes under different conditions such as response to signals or stress. For example, targeted profiling of the plasma membrane proteome (specifically the proteome exposed on the cell surface) is crucial for the identification of cell surface biomarkers (Smolders et al. 2015). Historically, protein localisation was determined by microscopy or by cell fractionation followed by western blotting techniques; however, both techniques have a low throughput and are also based on predetermined and known components. Using MS based proteomics it is now possible to carry out high throughput protein identification and quantitation using traditional subcellular fractionation methods to study organelle, substructure and/or compartment proteomics from whole cell extracts and tissue samples to measure the distribution of proteins between subcellular fractions (Christoforou et al. 2016, L. J. Foster et al. 2006).

1.1.2 Protein-protein Interactions (PPIs)

Many physiological processes are regulated by protein interaction networks so characterisation can provide valuable information on cell biology. Many proteins carry out their function within cells in the form of protein complexes. MS based proteomics has become the method of choice to comprehensively study PPIs (Smits et al. 2016). There are several strategies available to analyse protein–protein interactions. Target purification and interactor identification is an approach where the protein of interest is purified along with its interactors generally using affinity purification followed by mass spectrometry identification (Ho et al. 2002). However, tubulin and actin often bind non-specifically to

3

affinity chromatography resins thus complicating research toward identifying the cellular targets (Tamura et al. 2003).

Immunoprecipitation using MS provides a sensitive and accurate means of characterising protein complexes (ten Have et al. 2011). Using Immunoprecipitaiton with LC-MS/MS in chapter six we described a proteomic method antibody targeting of cell surface proteins with an established role in cancer invasion as a screening approach to inhibit tumour invasion by a functional monoclonal antibody (O'Sullivan et al. 2014). With the advent of fast and sensitive LC-MS instrumentation, PPIs can be studied at global levels not based on affinity purification of a specific protein of interest to study the co-behaviour of proteins from perturbation experiments (Savitski et al. 2014) or biochemical fractions (Wan et al. 2015). Chemical cross-linking/MS has developed recently into an alternative technique for analysing protein 3D structures, mapping interaction networks and elucidating protein-protein interactions (Häupl et al. 2016). The main advantage of cross-linking MS is that it can be performed under native-like conditions and weak or transient protein interactions are also preserved as covalent bonds are introduced between the interacting molecules. The main structural proteomic approaches use homobifunctional amine-reactive cross linking reagents (Mädler et al. 2009) followed by the protein of interest purification using affinity enrichment (LaCava et al. 2015). The binding between biotin and streptavidin or avidin is one of the strongest known non-covalent biological interactions (Fairhead et al. 2015), and as a result biotinylation is an attractive approach for protein complex purification proteomic studies (de Boer et al. 2003).

1.1.3   Protein post-translational modifications (PTMs)

A limited number of genes can generate a huge level of complexity at the protein level as a result of alternative splicing and post translational modification (PTM). PTMs are essential for most cellular functions such as protein activity, protein degradation and protein-protein interactions. A protein modification can increase/decrease protein activity, it can allow interaction with other proteins, it can allow a protein to localise to a specific part of the cell, it can affect protein folding and affect the rate of turnover, aggregation and degradation. Identifying post translational modifications can provide leads and insights into the function and role of the protein in a biological system. PTMs may result in an addition or removal from an amino acid side chain or protein termini, generated by the cleavage of

signal peptides from proteins or by covalent cross linking between separate protein domains (Walsh et al. 2005).The chemical changes on modified amino acids may form a mass shift which can be measured by mass spectrometry. Tandem mass spectrometry (MS/MS) provides valuable sequence information about modified peptides.

Protein phosphorylation is one of the most common and important post-translational modifications, playing a major role in the control of many biological processes, including cell growth, division, and signalling. Phosphorylation was first described in 1954 (Burnett et al. 1954). This reversible mechanism occurs through protein kinases and consists of the addition of a phosphate group (PO_4) to the polar group R of various amino acids (Ardito et al. 2017). The addition of PO_4 modifies the protein from hydrophobic polar to hydrophilic polar causing the protein to change conformation when interacting with other molecules. A phosphorylated amino acid can also bind molecules, able to interact with other proteins and consequently assemble and detach protein complexes (Bruce et al. 2007). More than one-third of the protein phosphorylation events take place on serine, threonine, and tyrosine amino acid residues (Roskoski et al. 2012). In particular, the phosphorylated residues of serine are 86.4%, followed by residues of threonine 11.8% however only 1.8% of tyrosine residues are phosphorylated (Nishi et al. 2014). Major improvements in the detection of PTMs, primarily by mass spectrometry over the last decade has resulted in an exponential increase in the documented number and type of PTMs that make up the modified eukaryotic proteome. Phosphorlyation can be detected by mass spectrometry by a dominant peak corresponding to the loss of phosphoric acid (H_3PO_4), i.e. a loss of 98 Da from the parent peptide mass. The rate at which PTMs are discovered now far surpasses the rate at which the specific PTM can be experimentally tested for biological function. Thus, a need for effective methods of prioritisation is essential for quantifying the likelihood of a site to be regulatory and/or impactful on biological function (Torres et al. 2016). However, protein phosphorylation is still challenging due to the low abundance of phosphoproteins and often by its low stoichiometry of phosphorylation as well as the extremely high dynamic range and the complexity of the proteome making it nearly impossible to analyse phosphoproteins/peptides directly from the complex proteome by MS without a specific enrichment procedure. Whereas the best enrichment strategies for many PTMs rely on the use of immunoaffinity reagents, for phosphorylated peptides both antibody-based as well as metal affinity enrichment-based methods exist. Phosphorylated peptides can be selectively enriched through interaction of the negatively charged phosphate groups with positively charged metal ions including iron (IMAC), titanium

dioxide, and zirconium dioxide (Fila et al. 2012). Generally, antibodies bind epitopes. Phosphoprotein or phosphopeptide enrichment using immunoprecipitation relies on antibodies raised against phosphorylated amino acids. Phosphoamino acid-selected antibodies against tyrosine-phosphorylated proteins/peptides have been successfully employed as enrichment strategies (Fila et al. 2012) especially as tyrosine phosphorylation is more challenging due to significantly lower levels of tyrosine phosphorylation compared to serine and threonine (Heibeck et al. 2009). Metal affinity chromatography will preferentially identify phosphopeptides present at higher levels (abundance-driven), while antibody-based methods will identify peptides that share the sequence characteristics targeted by the antibody itself (affinity-driven) (Stokes et al. 2015).

Despite the importance of the phosphorylation level of regulation, little work has been carried out on the phosphoproteomic characterisation of Chinese hamster ovary (CHO) cells in bioprocess-relevant conditions. In chapter seven using Orbitrap XL MS we identified 700 differentially expressed CHO phosphopeptides using quantitative label-free LC-MS/MS phosphoproteomic analysis in conjunction with IMAC and TiO_2 phosphopeptide enrichment strategies, following a reduction in temperature from 37 to 31 °C (Henry et al. 2017). In chapter eight using Orbitrap Fusion MS, we carried out an IMAC phosphoproteomic enrichment differential analysis of IgG producing CHO DP12 cells at various phases of growth in serum-free suspension batch culture to characterise dynamic changes to the phosphoproteome with changing culture conditions. This resulted in the identification of 3,777 differentially expressed unique phosphopeptides (Kaushik et al. 2018).

See Figure 1-2 for an example of tandem mass spectrometry (MS/MS) on a phosphorylated peptide DWEDDSDEDMSNFDR to obtain detailed structural information and consequently determine the amino acid sequence with the PTM site modification identified on serine residue 6 (S6-Phospho (79.96 Da) from Mitochondrial import receptor subunit. Identifying peptides from mass spectra is discussed in section 1.2.2.6.
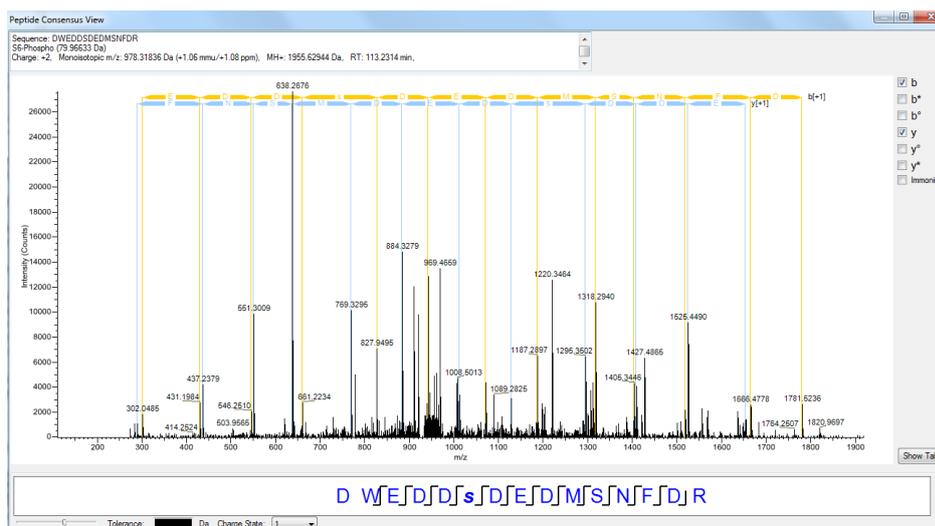
Figure 1-2. An example of an MS/MS from a phosphorylated peptide. The serine 6 residue from the sequence was identified as the phosphorylation site (S6-Phospho (79.96 Da)) from the protein mitochondrial import receptor subunit.

Ubiquitination, a step in the non-lysosomal degradation of proteins, is a crucial post-translational modification in eukaryotic organisms. The structured degradation and turnover of cellular proteins is regulated by the ubiquitin proteasome system (UPS). Most proteins that are critical for cellular regulation and function are targets of the process (Hershko et al. 1996). The ubiquitination process is tightly controlled by three families of enzymes: ubiquitin-activating enzymes (E1s), ubiquitin-conjugating enzymes (E2s), and finally ubiquitin-protein enzymes (E3s) (Morrow et al. 2015). Alterations of protein ubiquitination have been linked to many diseases such as cancer, neurodegenerative diseases, cardiovascular diseases, immunological disorders, and inflammatory diseases (Xu et al. 2013). Ubiquitination sites can be identified by MS through the detection of peptide adducts derived from ubiquitin. The C-terminus of the mature ubiquitin has the amino acid sequence KESTLHLVLRLRGG, in which the last Gly can be conjugated to lysine residues on target proteins. The peptide bonds at the carboxyl side of Arg and Lys can be cleaved by trypsin regardless of whether they are in the target proteins, in the conjugated ubiquitin, or in the free ubiquitin, as long as these amino acids are not modified. When the conjugated ubiquitin is cleaved with trypsin, it leaves two Gly residues on the modified lysine residues, generating a new type of peptide (Xu et al. 2013). See Figure 1-3 for an MS/MS of a ubiquitin modified peptide TLTGKTITLEVEPSDTIENVK with the site modification identified on lysine residue 5 (K5-GG -114.04293 Da) from Ubiquitin-40S ribosomal protein S27a.
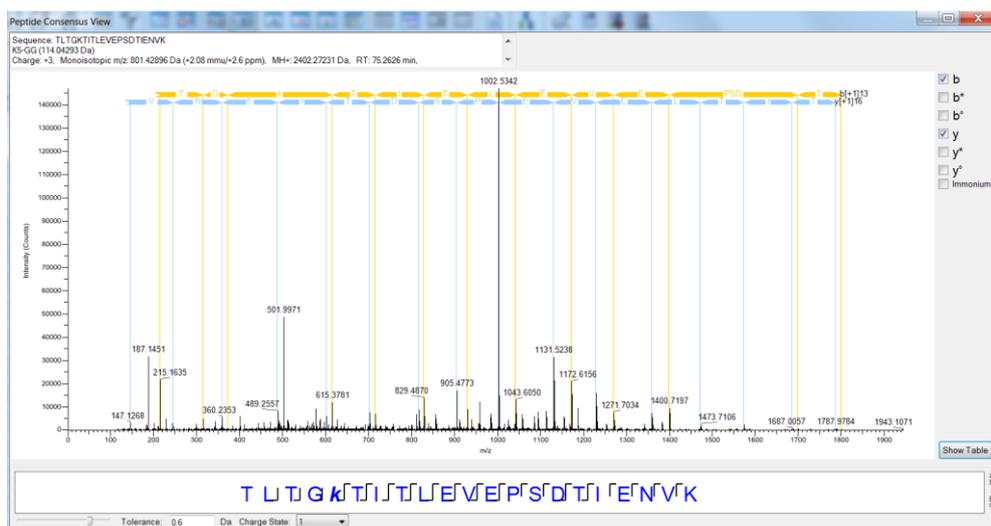
Figure 1-3. An example of an MS/MS from ubiquitinated peptide. The lysine 5 residue was identified as the ubiquitarian site (K5-GG -114.04293 Da) from the protein Ubiquitin-40S ribosomal protein S27a.

SUMO (**Small-Ubiquitin-like Modifier**) is a post-translational modifier of protein substrates at the amino acid lysine residue that conjugates to proteins in response to changes in the cell. SUMO has been found to be involved in mitosis, DNA repair, transcriptional regulation and subcellular localisation (Sula et al. 2012). Sumolyation occurs at certain lysine residues of target proteins via an isopeptide linkage (Knuesel et al. 2005). Ubiquitin has a carboxyl terminal sequence of KGG which allows tryptic digestions of ubiquitin conjugates to produce a tryptic peptide containing a diglycine signature sequence; however the absence of either lysine or arginine in the carboxyl terminus makes SUMO peptide identification with trypsin difficult due to the large tryptic fragments generated (Knuesel et al. 2005). Hsiao et al. (Hsiao et al. 2009) developed a database search tool ("ChopNSpice") to successfully identify SUMO acceptor sites from proteins sumoylated *in vivo* and *in vitro*.

Glycosylation is another common post-translational modification. It is a specific enzymatic process where glycans are attached to proteins or lipids. There are two main types of glycosylation, O-glycosylation where the glycan is attached to a Ser or Thr amino acid residue, and N-glycosylation where the glycan is attached to the Asn residue in the peptide consensus sequence Asn-X-Ser/Thr (where X is any amino acid except Proline). Glycans are involved in a wide range of intracellular, cell to cell and cell to matrix recognition events and are therefore of great biological interest to cancer and biopharmaceutical studies (Morelle et al. 2007). MS strategies for glycoprotein analysis can be divided into top-down analysis or bottom-up. Top-down is the most direct method where determination of the

molecular weight differences of peaks makes it possible to determine the type of glycan modifications on the protein of interest. Bottom-up approach involves characterising glycopeptides obtained following a proteolytic digestion of the glycoprotein (Tsai et al. 2017). Both Top-down and Bottom-up proteomic approaches are described in the application on mass spectrometry section of this thesis (Section 1.2).

Protein **Oxidative** modifications have been known to result in the loss of enzymatic activity, functional alterations, loss of structural integrity, and protein aggregation (Droge et al. 2002). Many diseases result in oxidative damage caused by the activation of the immune system, mitochondrial dysfunction or environmentally-induced oxidative stress (Verrastro et al. 2015). Protein oxidation is commonly measured as a marker of cellular stress and oxidative damage. Oxidation can be detected by mass spectrometry generating a mass increment of 15.9949 Da (Oxygen) to the amino acid residue methionine (Garner et al. 1998).

Protein **Methylation** is another post-translational modification which modulates biological and cellular processes such as protein interactions, transcription, RNA processing and protein dynamics (Afjehi-Sadat et al. 2013). Methylation is catalysed by highly specific methyltransferase enzymes on arginine, lysine, histidine and glutamate. A methyl group generates a mass increase of 14.016 Da on the modified amino acid. Also, peptide ions containing monomethyl arginine and symmetric dimethyl arginine will also generate a neutral loss of 31.04 Da, while a neutral loss of 45.05 Da is indicative of dimethyl arginine (Sidoli et al. 2012).

ε-*N*-acetylation of lysine residues in histones have a fundamental role in transcriptional regulation. ε-*N*-acetylation is predominantly involved in gene activation, due to the property of neutralising a positive charge of amino acid residue lysine thereby weakening histone interactions with DNA (Sidoli et al. 2012). ε-*N*-acetylation of lysine is catalysed by histone acetly transferases (HATs) and the removal is mediated by histone deacetylases (HDACs). A number of HDAC inhibitors induce apoptosis and autophagy, potentially offering new approaches in cancer therapy (Martinet et al. 2011). Acetylation can be detected by mass spectrometry generating a mass increment of 42.011 Da and can also lead to the observation of diagnostic fragment ions at m/z 126.091 and sometimes also m/z 143.118 in the MS/MS spectra of ε-*N*-acetylated peptides (Trelle et al. 2008).

O-sulfation is a post translational modification primarily on tyrosine residues with the addition of sulfate. This modification is limited to transmembrane and secretory proteins having traversed the trans-Golgi network where two tyrosylprotein sulfotransferase enzymes (TPST1 and TPST2) catalyse the transfer of sulfate from adenosine 3'-phosphate 5'-phosphosulfate to tyrosine phenol (Moore et al. 2003). The main function of tyrosine sulfation is the modulation of protein-protein interactions in the extracellular region (Kehoe et al. 2000). Fragmentation of sulfopeptides using CID MS/MS of sulfopeptides results in a neutral loss of sulfur trioxide SO_3 (80Da) (Edelson-Averbukh et al. 2011).

1.1.4   Sample preparation for Proteomics

As the proteome is so complex, there is no one standard method for preparing protein samples for analysis by proteomics. Proteomic protocols vary depending on the sample number, type, experimental goals, and analysis method used. Many factors are considered when designing sample preparation strategies, which can include the source, type, physical properties, sample abundance, complexity and cellular location of the proteins. Workflows that incorporate optimised cellular lysis, subcellular fractionation, depletion of high-abundance proteins or enrichment of select proteins, and mass tagging tools can all contribute to the accurate identification and quantitation of protein samples. Enrichment and/or fractionation steps can be introduced at the protein and/or peptide level if sample complexity needs to be reduced or when a specific subset of proteins/peptides are of interest. In all cases, the quality and reproducibility of sample extraction and preparation can significantly impact the research results. The most common technique to analyse proteins for proteomic studies by mass spectrometry is to digest the protein(s) into smaller peptides with an enzyme and then separate the peptides by reverse phase chromatography into a mass spectrometer by electrospray ionization (ESI) which allows for sequence information to be obtained by carrying out tandem mass spectrometry (MS/MS). Trypsin is by far is the most common enzyme used.

1.1.4.1   Digestion enzymes used for Proteomic analysis
Trypsin cleaves C-terminal to Arginine or Lysine which produces a mixture of peptides containing one basic amino acid residue per peptide. Using ESI these peptides ionize to low charge states (generally 1+, 2+, 3+ and 4+). Peptides are generally separated using HPLC,

ionised and activated to produce informative fragmentation patterns. Using software algorithms, the fragment ion of the peptide is deciphered and matched to its protein. Trypsin is commercially available from numerous vendors, purified from bovine or porcine pancreas or as a heterologous recombinant preparation. Glatter et al. (Glatter et al. 2012) reported that trypsin exhibits lower cleavage efficiency towards lysine than arginine residues and recommend serial digestion with **Lys-C** and trypsin for complex protein digestions to produce fewer missed cleavages. In chapter four and chapter five we describe the dual enzymatic digestion of Lys-C followed by trypsin (Henry et al. 2017, Dowling et al. 2015).

There are many alternatives to trypsin for LC-MS/MS proteomic studies. Trypsin is unable to produce MS identifiable peptides from the C terminus of proteins. C-terminal peptides may have any terminal residue at either side which results in a low charge making them incompatible with LC-MS/MS. As a result, protein C-termini involved in protein integration in membranes, complex formation and protein activity are often underrepresented in tryptic proteomic studies (Tsiatsiani et al. 2015). **Lys-N** is a protease option that cleaves N-terminal to basic amino acids to generate positively charged C-terminal peptides compatible with LC-MS/MS.

Asp-N hydrolyses peptide bonds on the N-terminal side of aspartic and cysteic acid residues. **Glu-C** is a serine protease cleaving at the C-terminal of aspartic acid and glutamic acid residues. **Arg-C** cleaves on the C-terminal of arginine residues (it also includes sites next to proline) along with lysine residues. **Chymotrypsin** is also a serine protease cleaving at the C-terminal of tyrosine, phenylalanine and tryptophan.

Despite the advances using proteomics in cell lysis preparation, peptide preparation, peptide separation, mass spectrometry and database search algorithms protein digestion is still mainly performed using the single enzyme trypsin (Giansanti et al. 2016).

1.1.4.2 Fractionation – Protein/peptide

Despite major improvements in sensitivity and speed of modern mass spectrometers and their high performance liquid chromatography systems, basic loading capacity and ion suppression can still limit the overall coverage of complex proteomic samples. Column overloading can cause poor chromatography such as peak fronting. MS ion suppression can negatively affect precision and accuracy of the instrument. The need to reduce the

complexity or pre-fractionate the sample(s) being analysed if possible and/or if there is sufficient starting material to allow it will always be beneficial to the analysis. There is a wide variety of strategies available to the proteomic researcher for protein and/or peptide fractionation. At the protein level, chromatographic separation methods like ion exchange (IEX), reverse phase, hydrophobic interaction and size exclusion can be employed. In chapter two we describe an off-line strategy for peptide fractionation using Strong Cation Exchange (SCX) prior to phosphopeptide enrichment using both Fe-NTA and TiO_2 methods (Henry et al. 2017b)

1.1.4.3   Depletion and enrichment strategies

Depletion and enrichment strategies can be employed to remove high-abundance proteins of no analytical interest and isolate target proteins/peptides in the sample. It is accepted that proteins in the circulatory system mirror an individual's physiology. Protein levels are generally determined using single-protein immunoassays in hospitals. High-throughput, quantitative analysis using mass-spectrometry-based proteomics of blood, plasma, and serum would be advantageous, however, unfortunately for clinical proteomics, the dynamic range of the protein concentrations of different classes of protein is greater than ten orders of magnitude (Pernemalm et al. 2008, Anderson et al. 2002) and at best the linear dynamic range of a mass spectrometer is five orders of magnitude (A. Kaufmann et al. 2017).

Figure 1-4 demonstrates the reference intervals of 70 proteins found in plasma, and the abundances are displayed on a log scale spanning 12 orders of magnitude. At the high abundance end of the scale is serum albumin (normal concentration range 35-50 mg/ml or 35-50^9 pg/ml) with interleukin 6 at the lower abundance end (normal range 0-5 pg/ml) (Anderson et al. 2002).

The levels of high abundance proteins range from milligram to gram per liter, while those of low abundance proteins are generally less than 1 µg/L (Anderson et al. 2002). It has been suggested that the more interesting, possibly tumour-specific protein biomarkers in plasma/serum may be present at 2-3 orders of magnitude lower than current LC-MS methods allow (B. Kim et al. 2018). Because of these issues, plasma/serum prefractionation methods play critical roles in the reduction of their complexity, allowing an opportunity to explore tissue-derived proteins that leak into the circulation as low abundance serum/plasma proteins. Standard approaches for plasma/serum prefractionation can be

assigned into any of the following approaches; immunodepletion, affinity enrichment, and fractionation. No single approach is perfect, but instead, different approaches may be complementary to each other. Similar challenges of high abundance proteins interfering with the analysis of low abundance proteins is also encountered by those analysing the secretome collected as conditioned medium as to whether the interfering proteins are supplemented serum, cell culture media components and/or recombinant protein secreted by the host cell line. Clabaut et al. (Clabaut et al. 2015) described two methods to study the secretome profiles of human mesenchymal stem cell-derived adipocytes with a combination of cell washing steps and 3 KDa centrifuge filter steps. Kumar et al. (A. Kumar et al. 2015) analysed the supernatant from CHO-K1 culture and carried out LC/LC–MS/MS analysis, allowing for the identification of 3281 different host cell proteins. The CHO-K1 cells were washed six times with 15 mL of PBS and subsequently starved for 12 with serum-free media to reduce high abundance protein interference.
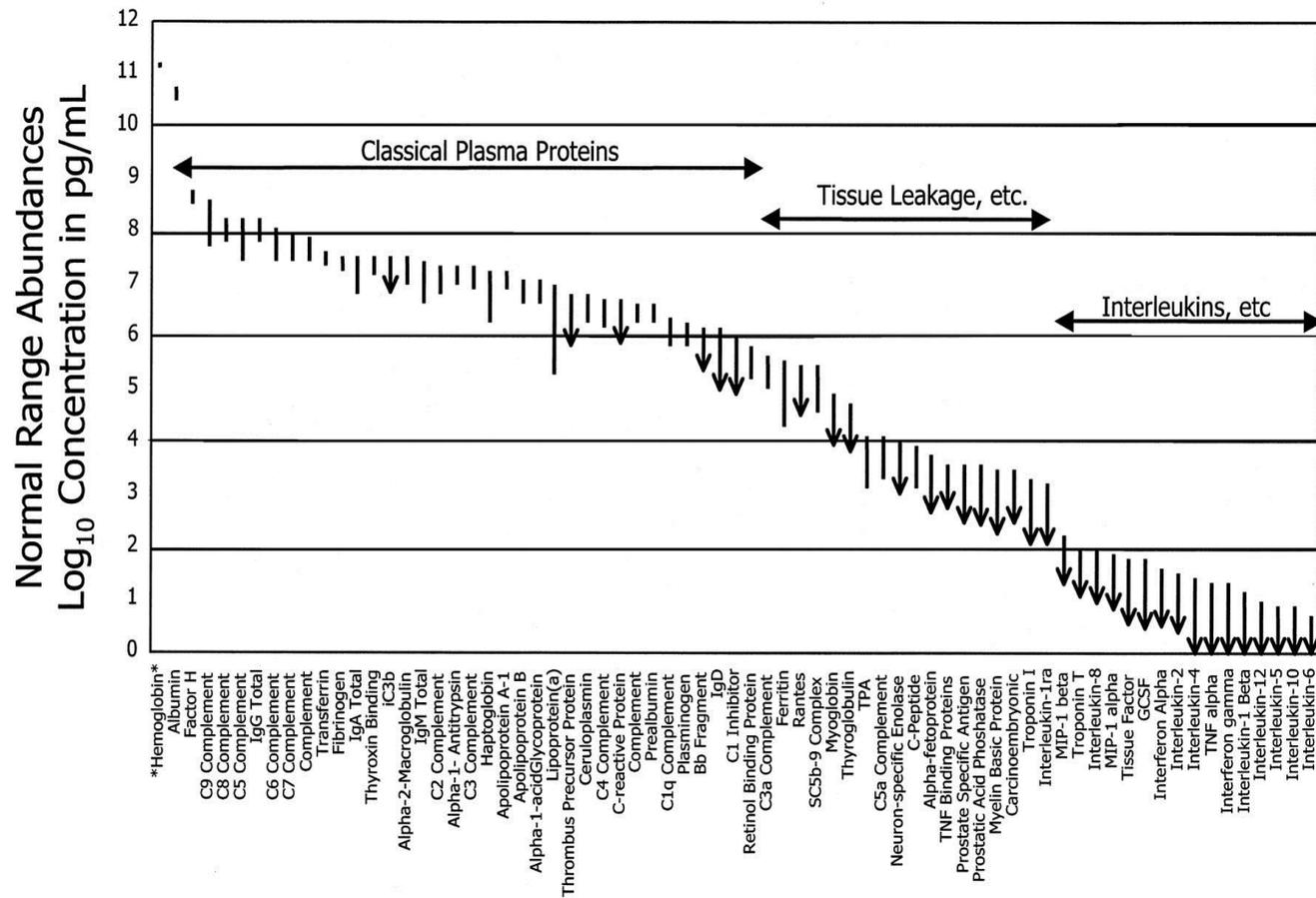
Figure 1-4. Summary of the reference intervals of 70 proteins found in plasma with their abundance displayed on a log scale spanning 12 orders of magnitude. Serum albumin is at the high abundance end while interleukin 6 is at the lower end of the abundance scale where these proteins differ in plasma abundance by a factor of 10^{10} (A. Kaufmann et al. 2017).

1.1.4.4 Immunodepletion for clinical proteomics

Immunodepletion methods generally apply beads linked to antibodies specific towards one or many high abundant proteins in the serum or plasma sample. This capture technique with a stationary phase, allows the collection of depleted proteins in the flow-through separating them from the low abundant proteins. The removal of serum albumin alone results in 50% depletion of the total serum/plasma proteins; for example the Pierce[TM] Top 2 depletion kit removes serum albumin and immunoglobulin G which results in a 75% depletion while a 20 protein depletion resin such as ProteoPrep® 20 depletes 95-99% of the total serum/plasma proteins (Chutipongtanate et al. 2017). See Table 1-1 for summary of depletion strategies commercially available.

Depletion strategy	Proteins removed
Pierce[TM] Top 2 depletion kit	Serum albumin and immunoglobulin G
MARS-6 depletion column	Serum albumin, immunoglobulin G, immunoglobulin A, haptoglobin, α-1 antitrypsin, and transferrin
Sappro® IgY-14 depletion	Serum albumin, IgG, IgA, haptoglobin, α-1 antitrypsin, transferring, IgM, IgD, α-1 macroglobulin, α-1 acid glycoprotein, complement C3, apolipoprotein AI, apolipoprotein AII, and apolipoprotein B
ProteoPrep® 20	Serum albumin, IgG, IgA, haptoglobin, α-1 antitrypsin, transferring, IgM, IgD, α-1 macroglobulin, α-1 acid glycoprotein, complement C3, apolipoprotein AI, apolipoprotein AII, apolipoprotein B, transthyretin, ceruloplasmin, complement C4, complement C1q, prealbumin, and plasminogen

Table 1-1. Commercially available protein depletion strategies for clinical proteomics and the list of proteins that they remove.

Nanotrap particles are hydrogel particles of various sizes ranging from 300 to 3000 nanometers which contain a core made of a cross-linked polymeric networks of N-isopropylacrylamide and co-momers such as allylamine, acrylic acid and N,N'-methylenebisacrylamide. The nanotrap particles are functionalised with charge-based affinity bait using copolymerisation and covalent bind to its shell which act as a capture, concentrate and trap target proteins (Shafagati et al. 2015).

ProteoMiner™ (Bio-Rad Laboratories) is a protein enrichment technology using a large diverse bead-based library of combinatorial peptide ligands which simultaneously dilute high-abundance proteins and concentrate low-abundance proteins (L. Li et al. 2015).

Seong et al. (Seong et al. 2017) used an albumin/IgG removal kit to analyse low abundance proteins in human plasma using ultra-high performance liquid chromatography coupled with highly accurate mass spectrometry to optimise methods for biomarker discovery in human plasma samples for fragmentation pattern of elevated BNP in patients with cardiac disease. Cao et al. (Cao et al. 2015) depleted the 14 highest abundance proteins in human plasma samples from patients following heart failure with a multiple affinity removal system column (MARS) followed by two-dimensional liquid chromatography coupled to a tandem mass spectrometry.

In chapter two, we describe a method for the removal of both albumin and IgG from patient bio-fluids using ProteoPrep® Blue albumin depletion kit from Sigma Aldrich prior to either protein in-gel digestion or in-solution proteomic pipelines (Henry et al. 2017).

In chapter four, we describe a two-step depletion strategy using a combination of the Proteome Purity™ 12 Human Serum Protein Immunodepletion Resin from R & D Systems, United Kingdom to deplete the 12 most abundant proteins (alpha 1- Acid Glycoprotein, alpha 1-Antitrypsin, alpha 2-Macroglobulin, Albumin, Apolipoprotein A-I, Apolipoprotein A-II, Fibrinogen, Haptoglobin, IgA, IgG, IgM, Transferrin) from the serum followed by the addition of Blue Nanotrap particles (RB4VSA) from Ceres to study plasma samples from patients diagnosed with multiple myeloma (9 patients who subsequently responded to treatment/ 8 who subsequently failed to respond to treatment). The aim of the study was to identify a panel of clinically useful biomarkers to predict response to treatment containing bortezomib (Ting et al. 2017) .

In chapter five, we described the use Proteo- Miner™ technology (Bio-Rad Laboratories) to perform serum equalisation on cancer and non-cancer (healthy controls/non-malignant

breast disease) patient samples with the aim to identify a biomarker signature of tumour burden using quantitative profiling by label-free LC–MS/MS analysis (Dowling et al. 2015).

Using MS-based proteomics and combining with immunodepletion of high abundance proteins, as very extensive peptide fractionation, it has been possible to identify >5,000 proteins from plasma samples from patients with myocardial infarction (Keshishian et al. 2015). However; extensive immunodepletion and multi-fractionation is expensive, time consuming, requires large volume of sample, instrument time and analysis time. Geyer et al. (Geyer et al. 2016) described an automated, highly reproducible, 3 h proteomic workflow from blood droplet collection to results to quantitate 1,000-protein plasma proteins with the objective of developing a proteome profile as a proteomic portrait of a person's health state. Post blood collection, the samples were centrifuged to collect plasma and a combination of two immunodepletion kits were used to remove of 20 of the highest abundance plasma proteins.

1.1.4.5 Subcellular fractionation

Subcellular fractionation in tandem with state-of-the-art mass spectrometry-based proteomics represents a powerful tool for expanding the depth of cellular proteome coverage. Subcellular fractionation allows dissection of intracellular organelles, including the isolation of multi-protein complexes from these organelles. As such, many low abundance proteins and a variety of signalling complexes can be enriched, whereas unfractionated whole-cell lysate analyses are dominated by the most abundant proteins. Organelle enrichment can be carried out by centrifugal methods or by chemical kit methods for isolation of membrane, lysosome, peroxisome, nuclei, mitochondrial.

Localisation of organelle proteins by Isotope tagging (LOPIT) (Dunkley et al. 2004) is a proteomic method to determine the subcellular localisation of membrane proteins which uses a combination of biochemical fractionation using density gradient ultracentrifugation with multiplexed quantitative mass spectrometry. This technique was recently redeveloped to enable subcellular localisation of thousands of protein per experiment termed "hyperLOPIT" to include spatial resolution at the suborganelle and large protein complex level (Mulvey et al. 2017).

1.2   Application of mass spectrometry for the analysis of proteins and peptides

1.2.1   Introduction to Mass Spectrometry Technology for Proteomics

One of the most important developments in proteomics has been the development of Mass Spectrometry technology. Proteomic based MS can obtain protein structural information such as protein mass, peptide masses and amino acid sequence information. Using this information, for example, top-down proteomics can analyse whole proteins directly and view protein modifications that can cause mass shifts. Bottom-up proteomics can be used to identify proteins by searching peptide mass and sequence mass information against protein databases.

Proteomic based mass spectrometry requires sample ionisation. The biological sample must be converted into a desolvated ion. The two traditional methods are matrix-assisted laser desorption ionisation (MALDI) or more commonly electrospray ionisation (ESI). Nano ESI coupled with nano high performance liquid chromatography has become the standard LC-MS approach in the majority of proteomic labs. For MALDI ionisation, the sample is mixed uniformly in a large quantity of matrix. When targeted by ultraviolet light (nitrogen laser at 337 nm) the matrix absorbs and converts it to heat energy. The matrix heats rapidly and is vaporized along with the sample. The electrospray ionisation is produced by applying a strong electric field under atmospheric pressure to a liquid passing through a capillary needle. The electric field is obtained by applying a potential difference between the capillary and a counter electrode. This electric field will induce a charge accumulation at the liquid surface at the tip of the capillary which will break the surface to formed highly charged droplets. The charged droplets pass through a heated capillary causing solvent evaporation due to a high electric field in the mass spectrometer. Using ESI, peptides will produce mainly single, double or triple charged ions while large peptide fragments and proteins will produce multiple charges ions. Once a sample is charged, it undergoes separation, deflection and manipulation by the mass spectrometer. The charged samples are deflected at different angles based on their individual masses. Basically, as the charged particle beam passes through the magnetic field, it undergoes separation based on the *m/z* ratio of its particles. The separated particles arrive at different locations on a detector within the mass spectrometer, with each location is translated into a molecular ion peak on the spectrum graph.

Nanotechnology devices have the potential to significantly expand the capabilities of proteomics, for example, by addressing the current limitations in selectively reaching a target protein *in vivo* through physical and biological barriers and detecting low abundance targets. In the last decade, the sensitivity of analysis and accuracy of results for protein identification by MS have increased by several orders of magnitude.

Protein identification is obtained by amino acid sequencing, known as tandem mass spectrometry (MS/MS), which is used to fragment a specific peptide or protein which can then be used to deduce the amino acid sequence.

Shotgun proteomics (Bottom up), is the term used to describe the enzymatic digestion of proteins to produce peptides which are ionised by a mass spectrometer. The ionised peptides are analysed by their mass-to-charge (m/z) ratio and precursors (peptides). These can then be selected for fragmentation and have the m/z fragmented peptides determined. Top down proteomics is the term used to describe the introduction of intact proteins into the mass spectrometer and where fragmentation can allow for protein sequence coverage.

The majority of proteomic experiments involve the bottom-up approach. With the sheer numbers and large dynamic range of peptide abundances involved, these proteomic approaches require ultra-fast peptide detection and fragmentation with extreme sensitivity which will reduce measurement times, increase and improve protein identified sequence coverage to allow for in-depth proteomic analysis.

There are many instrument formats available for proteomic researchers (Domon et al. 2006). However linear ion trap technology (Syka et al. 2004) coupled with Fourier Transform (FT) mass spectrometry has become the most popular platform for proteomics as it combines high resolution capabilities of the FT along with the robustness, speed and sensitivity of the ion trap. With FTMS, masses are represented by frequencies and because frequencies can be measured with high accuracy, FTMS offers very high mass measurement accuracy. Mass accuracy, with regards to mass spectrometry is the ratio of the m/z measurement error to the true m/z and is generally described in parts per million (ppm). The mass resolving power (resolution) of an instrument is its ability to distinguish two peaks with slightly different m/z values and their mass difference is generally described as full width at half maximum (FWHM). Accurate mass is an extremely powerful filter to confirm the identity of a compound and for identification of an unknown (Makarov et al. 2006).

Martin et al. (Martin et al. 2000) described the first implementation of an ion cyclotron resonance instrument coupled with a 7T magnet to operate as a high resolution instrument. However; these instruments were very large and expensive to maintain. Syka et al. (Syka et al. 2004) described the design and performance of a prototype high performance hybrid mass spectrometer which consisted of a linear quadrupole ion trap (QLT) coupled to a Fourier transform ion cyclotron resonance mass analyzer (FTMS). A few years later, Makarov developed the Orbitrap^{TM} analyser coupled to a linear ion trap providing proteomic researchers with a somewhat affordable, very small and extremely powerful analyser (Hu et al. 2005).

The MS proteomics facility in the NICB Dublin City University has used Orbitrap MS technology (Thermo Fisher) for the last decade with the Orbitrap XL and now more recently with the Orbitrap Fusion Tribrid. Chapters 2, 3, 4, 5, 6, 7, and 9 used OrbiTrap XL mass spectrometry analysis and Chapter 8 used the Orbitrap Fusion Tribid mass spectrometer.

1.2.1.1 Bottom Up/Shotgun Proteomics

All proteins of interest are extracted and digested with a protease or combination of proteases to produce peptides and subjecting them to a "shotgun" analysis, also referred to as "bottom up" analysis (Kelleher et al. 1999). The key steps for this approach are, (1) the protein sample needs to be digested into short peptides (which can be separated by liquid chromatography), and (2) when introduced into a mass spectrometer, they are fragmented to generate sequence information by matching by computational methods to a protein database to yield a protein match. The general approach is deep coverage or as much information as possible for a defined set of proteins/peptide (targeted MS). Shotgun MS proteomics approach requires no prior knowledge of the peptides present so the MS instrument can be run in a data dependent or independent mode. Peptides obtained from the enzymatic digestion are separated by liquid chromatography systems coupled to mass spectrometers (LC-MS). During the analysis, eluting peptides from the chromatography are selected to a predefined criteria (e.g. peptide signal rising above the noise in a full scan mass spectrum) and fragmented, producing tandem (MS/MS) mass spectra known as data-dependent acquisition or all eluting peptides with a defined m/z window are selected for fragmentation know as data-independent acquisition.

1.2.1.2 Data-dependent acquisition (DDA)

DDA is the most widely used approach in shotgun proteomics (Kalli et al. 2013) and generally selects the most intense peptide ion from the full MS scan and selects it for fragmentation. The criteria for MS and MS^n events are defined by the user. Examples of the parameters to be selected for full MS acquisition are mass resolution, selection of monoisotopic precursors, automatic gain control (AGC) target value for the ion population and full MS maximum injection time. For tandem mass spectrometry events, parameters can be mass resolution, minimal signal threshold, maximum ion injection times, number of MS^n events and dynamic exclusion times. See Figure 1-6 for an example of a peptide intensity threshold value to trigger an MS/MS scan. Peptides labelled 1 through to 7 have passed a threshold intensity value and will trigger an MS/MS event.

The nature and complexity of the sample being analysed by the mass spectrometer can have a major effect on the quality of protein/peptide identification rates. The common mode for MS acquisition in most labs is the highest number of identifications from a high peptide load. MS optimisation is usually time consuming and requires many injections of the same sample. It has become the normal approach for MS experts with limited time and too many samples to "pick and stick", i.e. pick a method and stick with it for most samples as there is not enough time nor in most cases enough, sample to optimise. Unfortunately, peptide identifications can be the casualty for sample and time saving.

One of the main advantages of ion trap mass spectrometry is the ability to carry out multiple MS/MS fragmentation steps on a single precursor molecule (e.g. a peptide) and its product ions (e.g. fragmented peptide) to create a pattern of product-precursor ion relationship to produce extensive structural information. With LC/MS ion trap experiments, optimisation of filling the ion trap and detecting the ions within the chromatographic time period of an eluting peptide is critical for successful confident identifications. These events are described as analytical cycle time and in a high-performance ion trap can be divided into 4 events.

     1. Ion injection

     2. Automatic Gain control (pre scan)

     3. Isolation and activation of the peptide in the trap

     4. Mass analysis by scanning ions out of the trap

Ion injection time is determined by the rate that ions enter the trap while mass analysis is a function of scan speed and its selected mass range.

With DDA methods, time spent on optimising/balancing MS injection times and scan rate times can drastically improve peptide identification success rates depending on the complexity (i.e. number of peptide) of the sample being analysed. Short max ion times and fast ion scan rates (35 ms) should be used for complex peptide samples while longer max ion injection times (300 ms) with slower scan rates used for less complex samples. Using an Orbitrap Fusion Tribrid Mass Spectrometer, we analysed two samples of various concentrations and complexity and compared fast scan rates and slower scan rates. The high concentration/complex sample was a 1 µg peptide injection from a rat thyroid cell lysate sample and we compared a maximum ion time of 35 ms and a 300 ms maximum ion time. The fast scan time method resulted in 99,502 MS/MS scans, with 18,599 unique peptides resulting in the identification of 3,508 rat proteins. The slower scan times resulted in 76,989 MS/MS scans and only 15,121 unique peptides resulting in only 3,119 proteins being identified (See Figure 1-5.). From this data, short max ion time is the best for complex sample analysis as more MS/MS were acquired resulting in more unique peptides being identified.

The low concentration/complex sample was a 0.1 µg peptide fraction from a human serum neutrophil sample that had a membrane protein enrichment. We compared the maximum ion time of 35 ms and 300 ms maximum ion time methods. The fast scan times resulted in 38,773 MS/MS scans, but only 713 unique peptides from 333 identified human proteins. The slower scan times resulted in 22,661 MS/MS scans but had 2150 unique peptides from 686 proteins identified (Figure 1-5). We found longer max ion time is the better for low concentration complex sample analysis as better quality MS/MS using longer ion time is more important than a maximum possible MS/MS count.

Number of Peptide IDs vs Concentration

	High Concentration Complex Sample	Low Concentration Complex Sample
■ 300 ms	15121	2150
■ 35 ms	18599	713
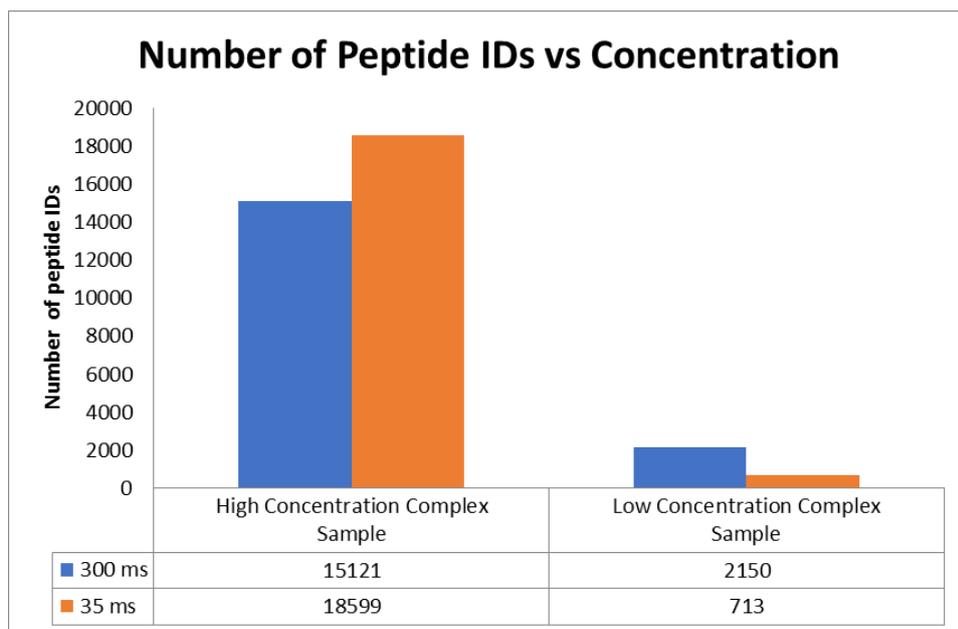
Figure 1-5. Comparison of how ion trap method parameters can drastically affect the number of confident peptide identification results. At high concentration of sample, fast ion injection time is optimal while at low concentrations of sample, longer injection times are optimal.

1.2.1.3 Data Independent Acquisition (DIA)

An alternative to DDA mass spectrometry is DIA. The peptide selection criteria with DDA can favour high abundant/intensity peptides. DIA methods select all peptide ions within a given mass range and above the detection limit of the LC-MS from a sample and are fragmented independently of their intensity.

DIA, theoretically allows for the identification and quantification of all peptide precursors; however, the MS/MS spectra generated is composed of the fragmented ions from all the different peptides precursors unlike traditional DDA where the precursor peptide mass is directly linked to the MS/MS fragmentation. The acquisition workflow involves cyclical recording across the chromatographic time frame of consecutive full MS survey scans and MS/MS fragment ion spectra for all peptides in a predetermined isolation window. The isolation window is described as the m/z range that the MS detector uses to isolate the precursor peptide. Then, using reference to pre-recorded spectral libraries, targeted data extraction is carried out on both the full MS and MS/MS. See Figure 1-7 for an example of data independent acquisition below where reaction monitoring with 30 x 20 m/z wide windows = 500 m/z will measure all the fragment ions for all precursors (peptides) between 400 and 900 m/z. In DIA analysis, the resulting fragment ion spectra is highly multiplexed

(no direct relationship between the precursor ion and its fragmented ions) and as a result, interpretation and interrogation of the data generated can be challenging.

Peptide identification from DIA data can be achieved using reference spectral libraries for the targeted extraction of quantitative information of the peptides included in these libraries (Schubert et al. 2015) using tools such as Spectronaut, OpenSWATH, Skyline, or PeakView. Bruderer et al. (Bruderer et al. 2017) and Ritz et al. (Ritz et al. 2017) used DIA to characterise peptides presented by human leukocyte antigen (HLA) class I molecules and reported they could confidently identify more peptides than the DDA method on the same instrument with 100 times less sample. Either method can generate large numbers of MS/MS spectra which require automated search engines capable of identifying peptides (and quantifying if necessary).

Another approach using DIA MS data can be a targeted approach known as Sequential Windowed Acquisition of All theoretical Fragment Ion Mass Spectra (SWATH) (Schubert et al. 2015) which uses a targeted extraction of specific ion fragmented spectra for identification and quantification.

SWATH MS aims to complement traditional shotgun mass spectrometry techniques and SRM methods (Gillet et al. 2012). An alternative approach to SWATH is WiSIM-DIA which is a combination of conventional DIA with a wide-SIM (wide selected-ion monitoring) windows which partition the precursor m/z space thus producing high quality precursor ion chromatograms (Koopmans et al. 2018). An example of WiSIM-DIA workflow on an Orbitrap Fusion instrument can be seen in Figure 1-8.

Minimal signal threshold for MS/MS selection
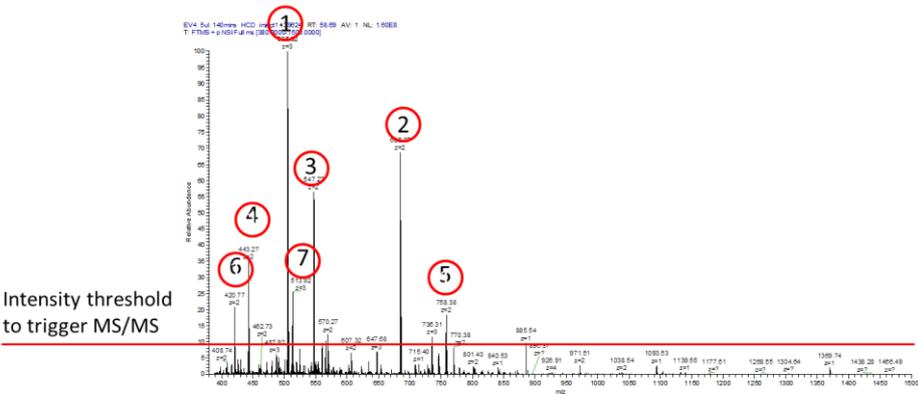
Figure 1-6. Data dependent acquisition method where all peptides measured above a predefined value/intensity are selected for MS/MS.

Comprehensive MS and MS/MS sampling

Figure 1-7. Data independent acquisition method where all peptides within the given mass range (m/z) are selected for MS/MS.
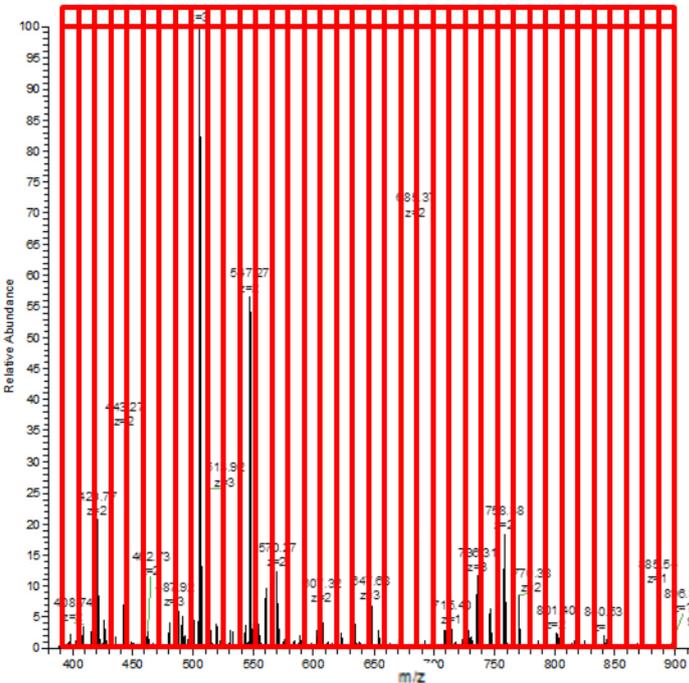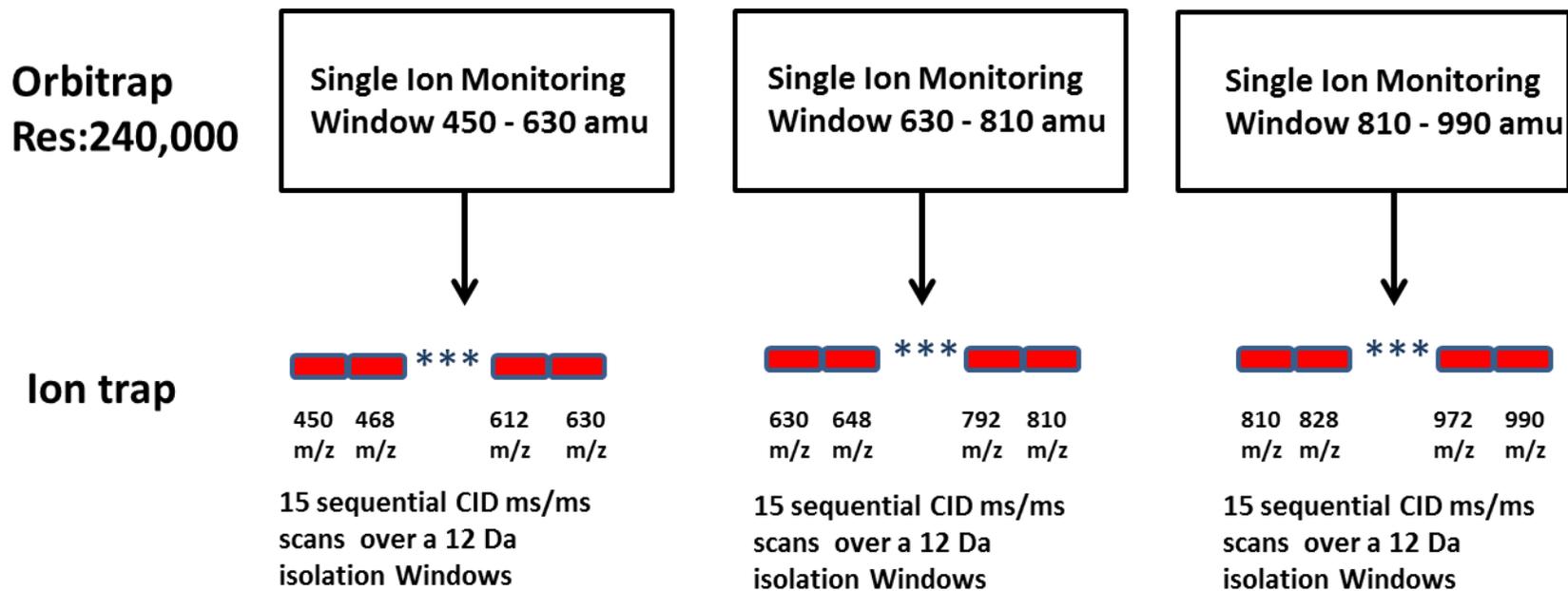
Figure 1-8. WiSIM DIA on the Orbitrap fusion. Scanning at 240,000 resolution in the Orbitrap with wide isolation windows of 180 amu (which covers 450-990 m/z) while in parallel to each SIM scan, 15 CID MS/MS scans are carried out in the ion trap with 12 amu isolation windows.

1.2.2 MS/MS Fragmentation Strategies

In an observed MS/MS spectrum, the type of ions observed will depend on a number of factors including the charge state of the peptide, its primary sequence, how the energy was introduced etc. etc. Fragment ions can only be detected by the analyser in a mass spectrometer if they carry at least one charge. If the retained charge is on the N-terminal fragment, the ion is classed as either a, b or c however, if the charge is retained on the C-terminal, then the ion is described as either x, y or z. (See Roesptroff nomenclature in Figure 1-9) (Roepstroff et al. 1984).

1.2.2.1 Collision Induced Dissociation (CID)

CID is traditionally one of the most common fragmentation processes for tandem mass spectrometry to dissociate peptide ions for sequence analysis. Generally, the preferred site of cleavage of the peptide ion is its amide bond on the peptide backbone. Following collision induced dissociation, the amide bond on the peptide backbone will fragment to produce a series of –y ions and –b ions (see Figure 1-9). CID fragmentation was used in chapters 2, 3, 4, 5, 6, 7, 8 and 9 in this thesis as the standard fragmentation technique for proteomic analysis of peptides.
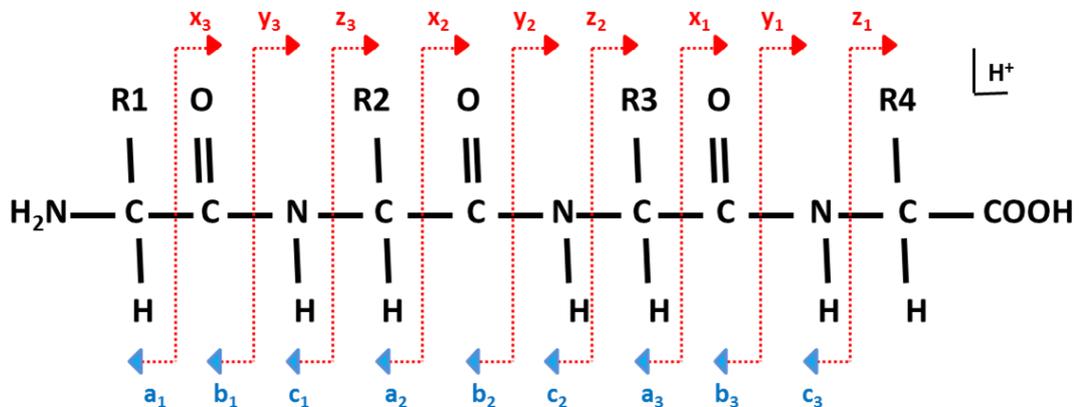
Figure 1-9. Roepstroff Nomenclature demonstrating the fragment ions formed from the peptide backbone cleavage of a protonated peptide. Fragment ions retaining a positive charge on the carboxy terminus are termed –x, -y and –z type ions while the fragment ions retained on the positive charge on the amino terminus are termed –a, -b and –c ions.

1.2.2.2　High Collision Dissociation (HCD)

HCD is a fragmentation option on Orbitrap analysers (Thermo Scientific). HCD is a higher energy dissociation compared to CID, and as a result enables a wider range of fragmentation pathways and has no low-mass cut-off (Jedrychowski et al. 2011). Historically, HCD resulted in higher quality MS/MS spectra, however; traditionally it is a slower fragmentation strategy as spectral acquisition involves Fourier transform detection. Newer Orbitrap mass spectrometers like the Orbitrap Fusion instruments allow for HCD fragmentation to also take place in the ion trap too and thus have comparable acquisition speed to CID ion trap (Hebert et al. 2014).

1.2.2.3　Neutral loss M3

Data dependent neutral loss (DDNL) MS3 methods consist of additional fragmentation of the product of the precursor neutral loss in the form of an MS3 scan. This technique is useful if there is a predominant loss in MS2 data that does not inform the user much about the structure of the compound (typical of phosphorylated peptides). To avoid this extra scan, if sufficient information is contained in the MS/MS scan, a trigger based on a dominant peak in the MS/MS spectrum can be set for a certain intensity and/or a certain mass (Villén et al. 2008). When a phosphorylated peptide is fragmented by CID or HCD, a dominant peak corresponding to the loss of phosphoric acid (H_3PO_4), i.e. a loss of 98 Da from the parent peptide mass, is observed in the tandem mass spectrum. This loss of phosphoric acid generally results in the peptide's backbone bond staying intact, thus the resulting spectra produces a poor sequence spectrum. In chapter 7 we describe a data-depended neutral loss MS (MS3) strategy for phosphopeptide analysis (Henry et al. 2017).

1.2.2.4　Electron transfer dissociation (ETD)

ETD is an ion/ion chemistry MS method to fragment peptides. Fragmentation occurs by transferring an electron from a radical anion to a protonated peptide. This induces fragmentation of the peptide backbone, causing cleavage of the Cα-N bond which creates complementary –c and –z type ions (Mikesh et al. 2006). This is a soft/low energy fragmentation method which can preserve PTMs. ETD can also be utilised to fragment large peptide (greater than 20 amino acids) and high charge state peptide fragments which generally do not fragment well with collision dissociation fragmentation methods. Fluoranthene has emerged as one of the most widely used reagent anions for ETD (M. Kim et al. 2012).

1.2.2.5 Multistage activation (MSA)

MSA is fragmentation strategy possible on hybrid Mass Spectrometers and more recently as CID scans on the Orbitrap Fusion MS platform. MSA is useful when a precursor has a predominant loss that does not inform the user much about the structure of the compound which happens often in phosphorylated peptides. MSA is a pseudo-MS3 scan where the precursor is fragmented as per usual and then a subsequent specified mass below that precursor is fragmented. All of the ions are then detected in a single scan. MSA assumes that the major loss will be occurring in all cases, so the instrument does not spend time scanning out to see if that fragment is there first. MSA does have a time penalty relative to CID MS2, about 10 milli-seconds, as it is carrying out an additional activation. However, it is much quicker than performing true MS3 (neutral loss) analysis where you would first do the MS2 followed by detection then re-isolation and fragmentation for the MS3 experiment. We used MSA fragmentation in chapter eight for MS phosphopeptide analysis (Kaushik et al. 2018).

1.2.2.6 Identifying peptides from mass spectra

Peptide identification from mass spectra data still remains one of the main bottlenecks in proteomic experiments. The first important step in proteomic data processing is the correct assignment of the MS/MS spectrum generated from a peptide ion. Identifications can be classified into 2 main categories:

a. Database searching

b. *De novo* sequencing

Database searching is the most frequently used peptide identification pipeline for large-scale proteomic studies (Kong et al. 2017). Database searching, results in a peptide identification by either correlating the MS/MS fragment ion spectra with theoretical spectra generated from a protein sequence database or by correlating the MS/MS fragment ion spectra with a spectral library of MS/MS spectra identified in previous experiments. A spectral library is meticulously complied from previously observed and identified MS/MS spectra. See Figure 1-10 of an example using SEQUEST search package to match a candidate fragmented peptide to the theoretical spectrum generated from the Chinese hamster ovary FASTA protein database (UniProtKB unreviewed TrEMBL Chinese hamster protein database). Theoretical b- and y–ions amino acid sequence masses (highlighted in red and blue respectively) are confidently matched to observed spectrum

masses corresponding to the amino acid sequence IFQIHTSR from the protein 26S protease regulatory subunit 4 [*Cricetulus griseus*].   See Figure 1-11 of an example using SEQUEST to match the same candidate fragmented peptide to the Chinese hamster mass spectra library from chemdata website which contains 74,509 distinct peptides from 158,026 spectra. The top spectrum is the candidate fragmented peptides and the bottom (upside down) spectrum is the confidently identified/match from the spectral library also corresponding to the amino acid sequence IFQIHTSR from the protein 26S protease regulatory subunit 4 [*Cricetulus griseus*].

#1	b+	b²⁺	Seq.	y+	y²⁺	#2
1	114.09134	57.54931	I			8
2	261.15975	131.08352	F	888.46863	444.73795	7
3	389.21833	195.11280	Q	741.40021	371.20374	6
4	502.30240	251.65484	I	613.34163	307.17446	5
5	639.36131	320.18429	H	500.25757	250.63242	4
6	740.40899	370.70813	T	363.19866	182.10297	3
7	827.44101	414.22415	S	262.15098	131.57913	2
8			R	175.11895	88.06311	1
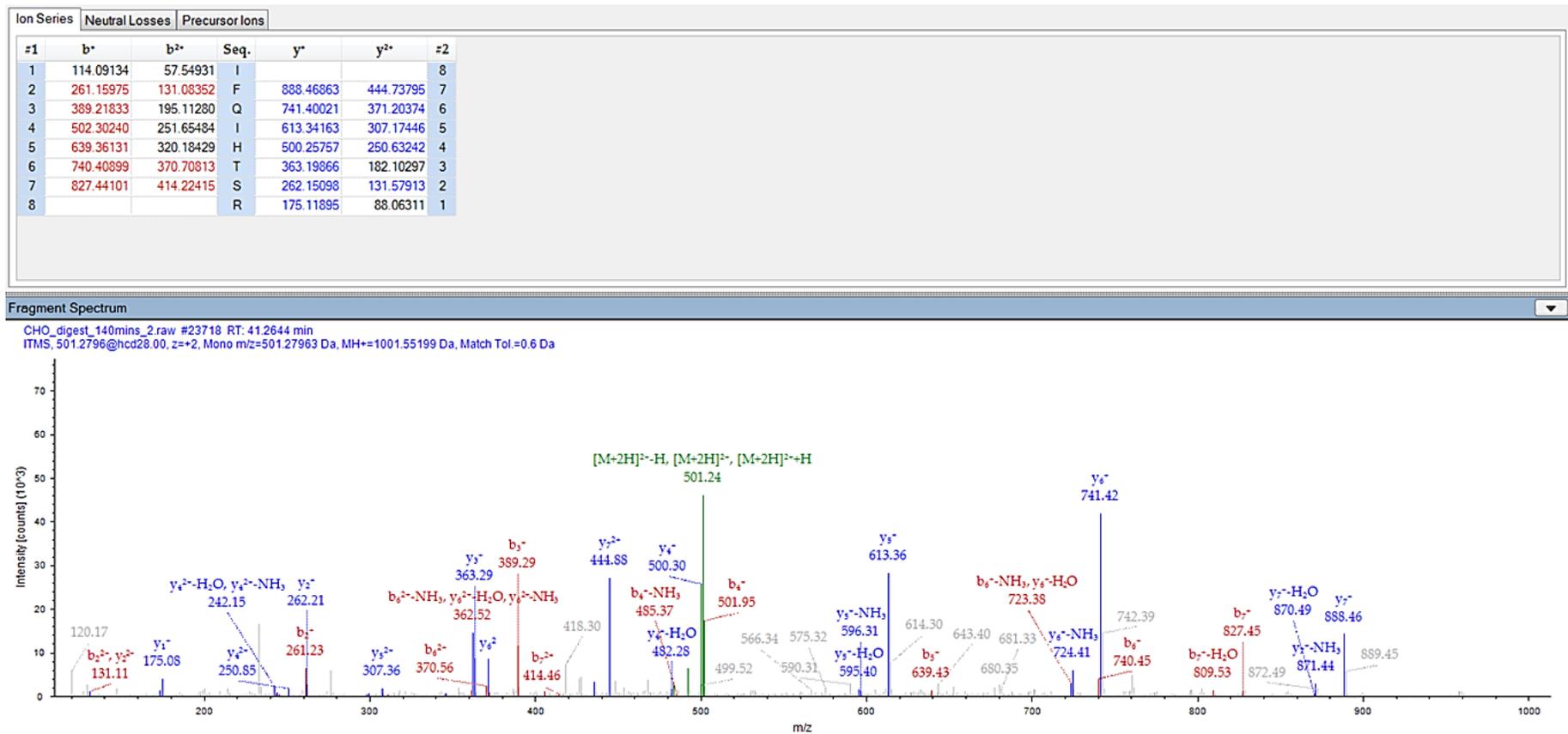
Figure 1-10. HCD MS/MS fragment ion spectra from a Chinese hamster ovary cell line digest analysed on the Orbitrap Fusion and the data matched to the Chinese hamster ovary fasta protein database (UniProtKB unreviewed TrEMBL Chinese hamster protein database). The MS/MS spectra of b- and y-ions matched confidently with the amino acid sequence IFQIHTSR was matched to 26S protease regulatory subunit 4 [Cricetulus griseus].
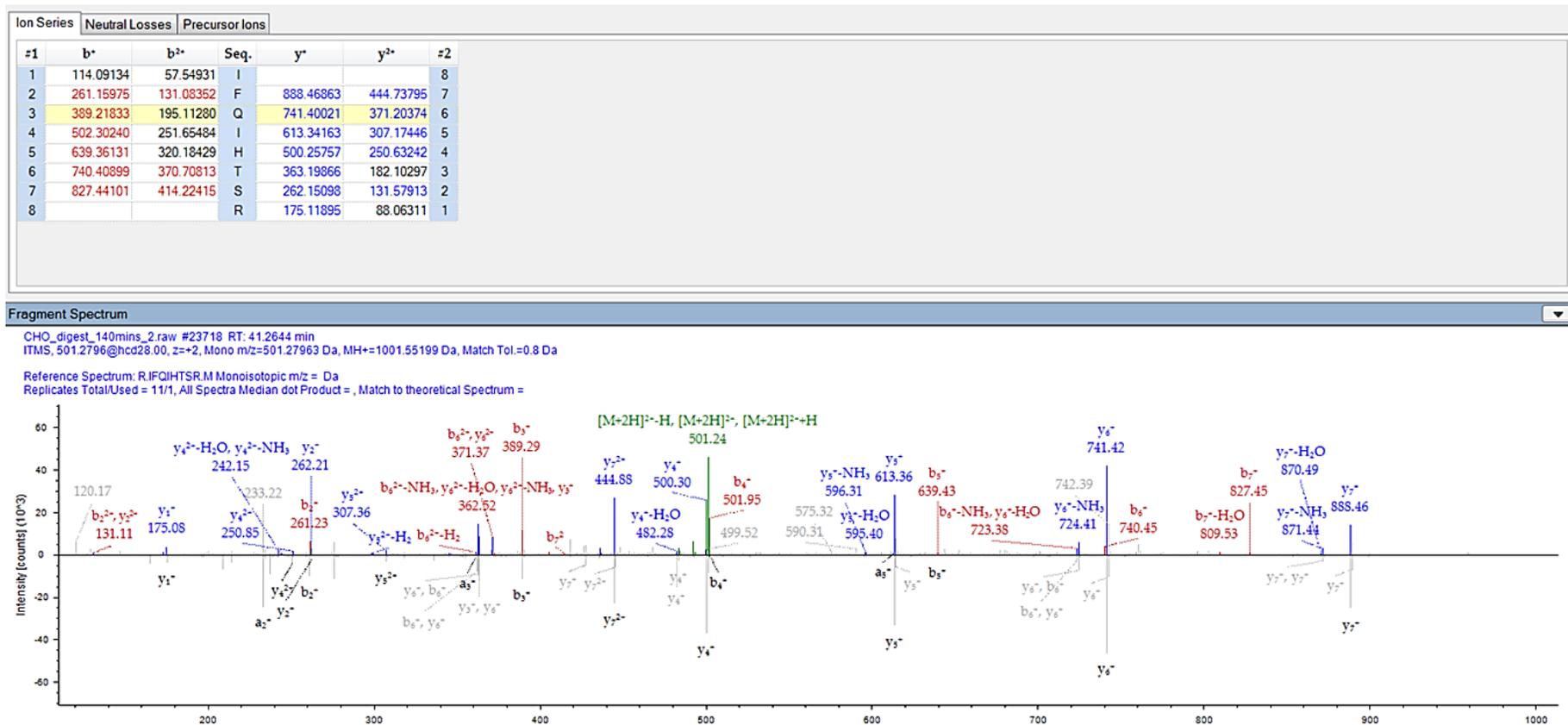
Figure 1-11. An MS/MS fragment ion spectra from a Chinese hamster ovary cell line digest analysed on the Orbitrap Fusion and the data matched to the reference spectrum from Chinese hamster ovary fasta protein Chinese hamster mass spectra library (https://chemdata.nist/gov/) There was a strong match of the candidate peptide (top) to the reference spectrum (bottom) IFQIHTSR from the HCD generated MS/MS spectrum identified as 26S protease regulatory subunit 4 [*Cricetulus griseus*].

De novo sequencing approach for peptide identification is a process that derives a peptide's amino acid sequence using the MS/MS fragment ion spectra without the assistance of a sequence database. It can also be used if the sequence contains modified peptides or polymorphisms. *De novo* sequencing is computationally heavy and requires high resolution fragment ion spectra so generally is not compatible with large scale proteomic studies (Kong et al. 2017).

When using protein databases for peptide identification, the MS/MS spectrum is correlated against the constructed theoretical spectra as long as a certain set of database criteria have been satisfied. One of the earliest software packages was SEQUEST (Eng et al. 2008) and more recently succeeded by SEQUEST HT. The main scoring function of SEQUEST is X_{corr} (Eng et al. 2008) which assigns a similarity score to a given pairing of an observed MS/MS spectrum and to that of a theoretical spectrum of a candidate peptide. The main two reasons that SEQUEST has stood the test of time is X_{corr} cross-correlation has proved to be an excellent discriminator in the presence of noise peaks and is easily integrated into fully automated processing pipelines to determine which spectra are confidently identified and then assembled protein inferences from the peptide/ion spectrum matches (Tabb et al. 2015). We used SEQUEST in chapters 4, 6, 7, 8 and 9 in this thesis.

MASCOT software package (MATRIX SCIENCE) uses a probability modelling algorithm and protein database searching. Experimental peptide and fragmented ion masses are matched to ones generated *in-silico* from a database. We used MASCOT in chapter five and we used MASCOT in conjunction with SEQUEST in chapter 4 in this thesis.

Although search engines are relatively effective at identifying peptides with a defined measure of reliability, their localisation of site/s of modification is often arbitrary and unreliable (Chalkley et al. 2012). Historically, CID fragmentation is the method of choice of peptide fragmentation; however alterative complementary configurations are now possible like ETD and HCD. Localisation of modification amino acid sites using fragmentation information is not straightforward (Chalkley et al. 2012) as a site can only be assigned if fragment ions surrounding the site are present which can be especially difficult if multiple sites are modified, e.g. multiple phosphorylatable residues being adjacent and single peptides having multiple modifications (Collins et al. 2014). There are various evaluation tools available in conjunction with the database search engines to help modification site assignment including PhosRS (Taus et al. 2011), Ascore (Bailey et al. 2009, Beausoleil et al.

2006) and SloMo (Bailey et al. 2009). In chapters 7 and 8 in this thesis we used PhosRS for site specific phosphorylation assignment following SEQUEST confident identification.

Figure 1-12 demonstrates SEQUEST search results from some of the fragmentation options available on the Orbitrap Fusion carried out on the phosphorylated peptide KVELs*ESEEDKGSK.

Fragmentation was carried out using multiple dissociation techniques of HCD, MSA, ETD and neutral loss CID MS3 strategies to demonstrate the versatility of the next generation Orbitrap fusion tribrid MS platform especially for phosphorylated peptides. SEQUEST X_{corr} cross-correlation scores for this particular phosphorylated peptide were: ETD 5.64, CID MS2 1.91 and the neutral loss CID MS3 4.14, HCD 3.78 and MSA 4.37.

Although the Orbitrap Tribrid MS instrument has multiple fragmentation options, we found the MSA strategy yielded the highest number of confident phosphorylated site specific identifications. We compared a 1 µg phosphopeptide enriched sample separated over 2 h with MSA yielding 7,148 confident phosphorylated peptides compared to HCD with 6,626 confident phosphorylated peptides. We found ETD fragmentation a slower fragmentation method on this instrument and it generated 50 % less MS/MS spectrums and approximately 50 % less confident identifications.
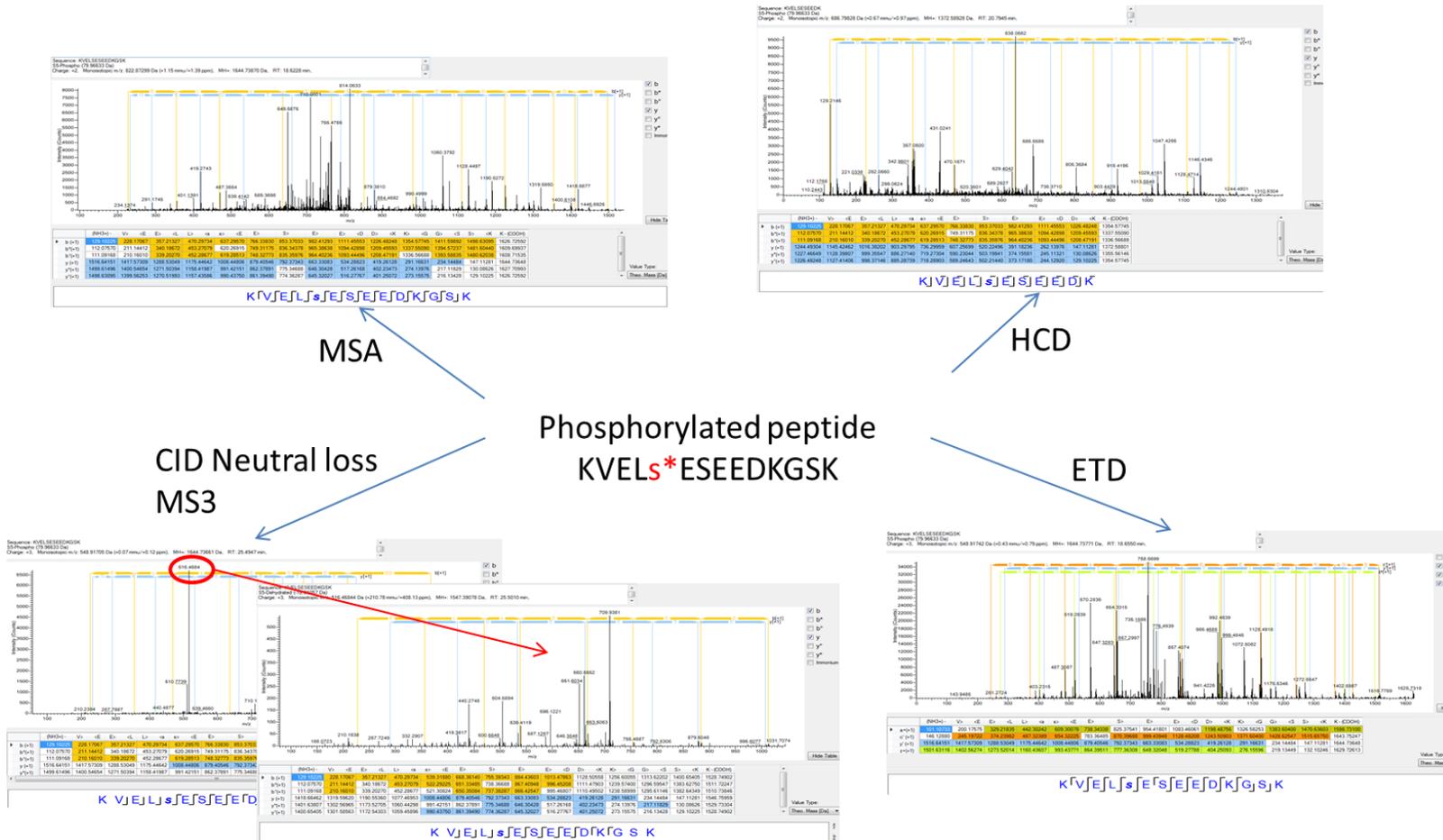
Figure 1-12. Phosphorylated peptide KVELs*ESEEDKGSK selected fragmentation using the various options available on the Orbitrap Fusion Tribrid MS. The strategies demonstrated were MSA, HCD, ETD and neutral loss MS3 CID each showing different fragment ion patterns when search using SEQUEST.

35

1.2.3 Top Down proteomics

The term Top-Down originated as a means to distinguish protein analysis/identification of intact protein isoforms or proteomforms by mass spectrometry (Smith et al. 2013) from the commonly used "bottom up" mass spectrometry approach.

Structural proteomics is the process of high-throughput characterisation of the three-dimensional structure of protein complexes or the proteins present in a specific cellular organelle (Manjasetty et al. 2012). Combining protein chemistry and modern mass spectrometry has paved the way for structural analysis of proteins and protein complexes as these techniques require less protein than the conventional structural techniques like NMR and X-Ray crystallography. Characterisation of changes in protein confirmation and protein-protein interactions can be studied using a combination of crosslinking (F. Liu et al. 2018) and/or hydrogen/deuterium exchange (HDX) with mass spectrometry (F. Liu et al. 2018). Orbitrap MS with resolution capabilities >500,000 allow for sub 1 Da distinction between intact proteofroms (McAlister et al. 2008). A disadvantage of direct intact protein analysis using mass spectrometry and electrospray ionisation is the spectra generated from the protein species with high charge states in a relatively small m/z window (Campbell et al. 2010). However, Orbitrap MS in conjunction with liquid chromatography separation can improve this issue.

The MAbPac^{TM} RP column from Thermo Scientific is a high-performance reverse-phase column suited to MS analysis and is suitable for characterisation of monoclonal antibody, fragments, variants, antibody drug conjugates, and we have used it for Fc-fusion protein variant separations. The MAbPac RP chromatography column can separate similar proteins and allow for intact mass analysis on Fc-fusion molecules. Analysis by SDS PAGE gel electrophoresis showed no visible difference in the profile of a Fc-fusion protein secreted from two CHO cell lines Figure 1-13 A and B when using a total protein stain (Coomassie blue) (Figure 1-13 A) (Henry et al. 2018). Using MAbPac RP over a 15 m separation time and UV at 214 nm the fragmented pattern of the different variants can be observed (Figure 1-13 B). The MAbPac RP separation uses water and acetonitrile and is therefore compatible with ion trap mass spectrometry. The resolution provided by the Orbitrap mass analyser on this Fc-fusion protein can easily be isotopically resolved at a resolution setting of 100,000. The charge envelope obtained for the biotherapeutic Fc-fusion protein analysed under denaturing conditions with the most abundant charge state detected at m/z 1,305.662 and

the deconvoluted monoisotopic mass using Magtran software (Grundy Research Group) at 46,808 Da (expected mass of the Fc-fusion protein). (Figure 1-13 C).
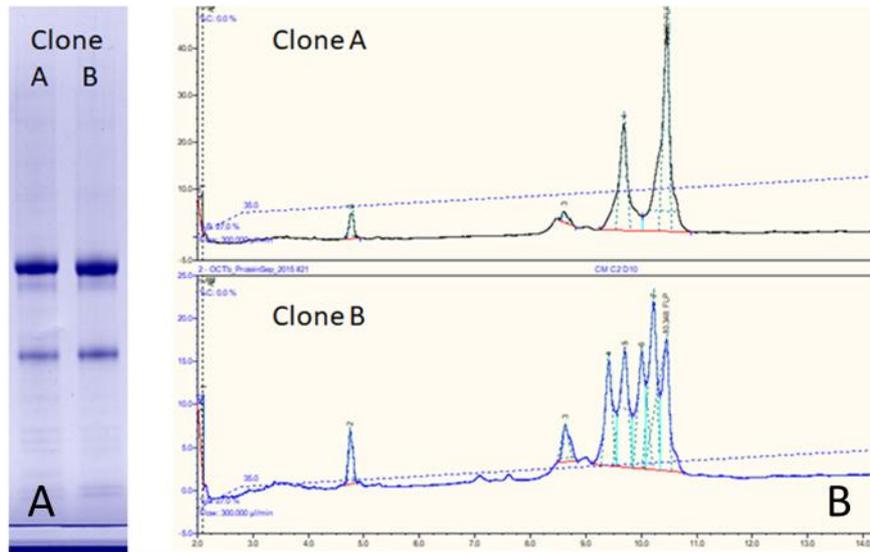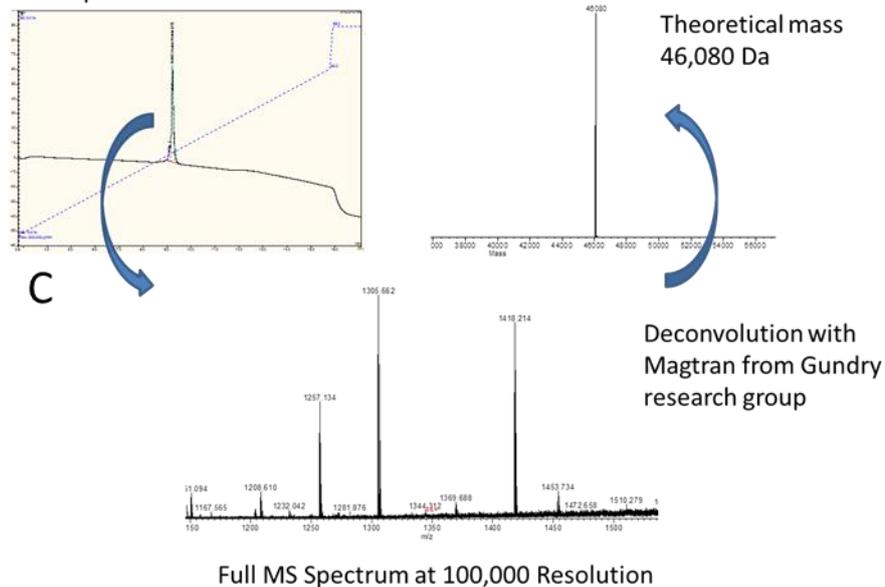
Figure 1-13. Comparison of an Fc-fusion protein profile from CHO cells to compare variants using 1D SDS gel stained with Coomassie blue and a HPLC MAbPac RP separation in conjunction with UV detection at 214 nm. Figure A demonstrates SDS Gel electrophoresis was unable to differentiate Fc-fusion protein profiles from two different CHO cell line samples which differed in their degree of product "clipping" (Henry et al., 2018). Figure B demonstrates reverse phased separation and UV analysis using MAbPAC chromatography resulted in the visualisation of additional clipped variants between the two samples. Figure C demonstrates Top Down proteomic analysis of the Fc-fusion protein after an LC separation into an Orbitrap XL mass spectrometer scanning at 100,000 resolution with a deconvoluted monoisotopic mass using Magtran software (Gundry Research Group) of 46,808 Da (expected mass of the Fc-fusion protein).

1.2.4 Database Utilisation

The two main repositories for sequence databases are the European Bioinformatics institute (www.ebi.ac.uk) and the National Centre of Biotechnology Information (www.ncbi.nlm.nih.gov). All mass spectrometry data search algorithms use the Fasta format and support searching of databases of protein sequences (Cottrell et al. 2011).

The success of the Human Genome Project making available the human genome sequence and the genomes of numerous other species revolutionised the study of biology. Instead of investigating a single gene or a small set of genes, whole ensembles of genes could be studied simultaneously. The exact same principle allowed for the study of proteins in a comprehensive manner using mass spectrometric analysis comparing mass information against database patterns and masses. In 2008, a draft of the complete human proteome was released from UniProtKB/Swiss-Prot containing approximately 20,000 putative human protein-coding genes representing only one UniProtKB/SwissProt identifying keyword (https://www.uniprot.org/help/human_proteome). Close to 40% of the 20,000 entries contain manually annotated isoforms representing over 22,000 additional sequences. Unfortunately, UniProtKB/SwissProt still only contains 236 reviewed sequences for Chinese hamster *Cricetulus griseus* (June 10th, 2018), historically requiring CHO proteomic researchers to reply on homology-based identifications from MS data (Meleady et al. 2012).

In 2010 with the emergence of affordable sequencing techniques like Roche 454 pyrosequencing and the Illumina/Solexa platform allowed for parallel sequencing at an unprecedented level (J. Zhang et al. 2011). Becker et al. (J. Becker et al. 2011) used pyrosequencing technology on the expressed cDNA from a wide range of CHO cell lines which resulted in the first set of unassembled genomic sequence data of CHO-K1 (Hammond et al. 2011). Shortly afterwards, Xu et al. (X. Xu et al. 2011a) assembled the draft genome for CHO K1 and made the database available from (Ref NCBI). Bielefeld-BOKU also created a CHO cell protein database in 2011 (J. Becker et al. 2011) based on 30,578 generated unfiltered isotigs and made the database available (see Bielefeld-Boku reference). Now two Chinese hamster ovary cell protein databases were available for CHO cell proteomic research and Meleady et al. (Meleady et al. 2012a) published a paper demonstrating a 50% increase in the number of confident proteins identified from CHO cell proteins by mass spectrometry when compared to the traditional mammalian homology-based searches.

1.2.5 Bioinformatic analysis of Proteomic data

The output from either shotgun proteomic approaches or more targeted methods generally result in a long list of identified factors including probability score, spectral match scores and possible quantitative values. To understand and interpret this data and to generate a hypothesis based on, for example, the response of the proteome to a challenge or condition, the protein list needs to filtered and classified. For functional analysis of a large protein list it is important to correctly match the protein name generated from the MS search algorithm against the protein database result to a unique identifier. Gene names have been standardised. However, protein names can differ between databases (example BB CHO vs. NCBI *CRIGR* data). There are web based algorithms to match protein names to corresponding gene names (example PICR or Cronos); however, Ensembl can accept UniProtKB protein identifiers as inputs. Generally, the first step for any functional interpretation of proteomic data is to match the protein identifier with its associated gene ontology (GO) term (Ashburner et al. 2000). The three main terms that genes are associated to hierarchically clustered function terms to describe the "biological process", "molecular function" and "cellular component". Depending on the proteomic fasta protein database used, some database search algorithms such as SEQUEST in Proteome Discoverer, MaxQuant and X!Tandem have implemented a GO-term association step. However, if the fasta protein database is not fully annotated with a corresponding GO term, the closest GO-term from a related protein can be located using a BLAST similarity search using the peptide identified. Online tools are also available to CHO researchers when protein accession number is available to convert the protein identification to a gene symbol. GOCHO or bioDBnet are examples of such online tools.

Following GO-term annotation, the gene names associated with the protein list can be analysed by GO-term enrichment analysis which will compare the abundance of the specific GO-terms with the abundance of the data list or in the organism of interest (Malik et al. 2010). A *p*-value can be calculated to determine and overrepresentation of a specific GO-term to demonstrate functions that are significantly enriched in one sample over another. Two common web-based software tools available are DAVID and Babeleomics resources (Jiao et al. 2012, Medina et al. 2010). We used DAVID for GO-term enrichment analysis in chapter 7, 8 and 9 in this thesis. In chapter 9 the Gene Ontology analysis showed a significant enrichment of biological processes and molecular functions related to protein folding, response to unfolded protein and protein translation (Henry et al. 2018).

Biological molecular pathways can also be analysed using the gene names from the large-scale MS proteomic data generated using pathway analysis tools that are open-access databases or commercial databases. Examples are, Protein Analysis Through Evolutionary Relationships (Ref: PANTHER) classification system, Search Tool for the Retrieval of Interacting Genes/Proteins (Ref: STRING), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Ingenuity Pathways Analysis (Ref: IPA). In chapter 9 we used STRING functional interaction network analysis.

1.2.6 Mass Spectrometry Methods for Differential Proteomics

Differential expression proteomics is the quantitative study of protein expression between samples that differ by some variable. Commonly in this approach, protein expression of the entire proteome or of sub-proteomes between samples is compared. Using this approach proteomic researchers aim to identify novel proteins and pathways involved in specific diseases or associated with specific desirable cell phenotypes. The main approaches for differential proteomics typically involve chemical labelling or label free analysis and mass spectrometry.

Differential expression proteomics is the quantitative study of protein expression between samples that differ by some variable. Commonly in this approach, protein expression of the entire proteome or of sub-proteomes between samples is compared. Using this approach proteomic researchers aim to identify novel proteins and pathways involved in specific diseases or associated with specific desirable cell phenotypes.

1.2.6.1 Label-free differential expression Proteomics

Mass spectrometry measurements of biological samples by direct quantitation of the signal response of their derived peptides was established over ten years ago. Bondarenko et al. (Bondarenko et al. 2002) and Wang et al. (Wang et al. 2005) demonstrated a linear relationship between protein abundance and peptide peak area. Liu et al. (J. Liu et al. 2004) showed that spectral sampling was an accurate reflection on relative abundance with a linear correlation over a 2 order of magnitude linear dynamic range. Since then, numerous strategies and algorithms implementing these themes of peptide ion signal peak intensity and peptide spectral matches (PSM) spectral counting – based quantitation have been published (Al Shweiki et al. 2017).

Quantitative label-free discovery proteomics on peptide samples allows the researcher to relatively quantify and identify thousands of proteins from complex samples as long as the instrumentation can provide highly reproducible chromatography (as sample to sample comparison is required) and also high-resolution, accurate mass spectrometry. LC-MS runs need to be aligned for comparative expression profiling between multiple different runs/samples. The two most popular software packages for label free quantitation by mass spectrometry are Progenesis QI for Proteomics and Maxquant LFQ (MQ). Maxquant is freely available from the Max Planck Institute of Biochemistry in Martinsreid, Germany (Cox et al. 2014). The software quantifies proteins across samples using the maximum (pairwise) peptide ratio information extracted peptide ion signal intensity (Al Shweiki et al. 2017) which are then normalised by minimising the overall fold change all every peptide across all samples prior to normalisation. Progenesis® QIP which is marketed by Waters also quantifies each protein based on its peptide ion signal peak intensity.

An example of a label free pipeline using Progenesis QI pipeline where a total ion chromatogram view is used to align to LC-MS runs for label free analysis can be seen in Figure 1-14. LC-MS runs are represented as m/z vs. retention time overlaid where the ion map is visualised as an alignment target in pink and the ion map to be aligned to in green (Figure 1-14A). Once the ion maps are aligned correctly they turn black (Figure 1-14B) and the total ion chromatogram is aligned (Figure 1-14C). Following alignment, peak picking and ion abundance quantification is carried out in each LC-MS run and the data is normalised (Figure 1-14D). We use analysis of variance (ANOVA) with a p-value of 0.05 or greater to select peptides to submit for identification between the experimental groups being compared.

Spectral peak intensities are the most common approach where the relative abundance of the same peptide in different samples is measured across consecutive LC-MS runs (under the assumption that the measurements are performed under identical conditions). In theory there is no limit to the number of samples that can be analysed. Normalisation is performed based on the median of the peptide ion intensities against selected reference spectra. The unique peptide ion intensity for peptides from specific protein are summed to generate an abundance value. Using one-way ANOVA, p-values can be calculated based on transformed values. Supporting MS/MS data provides peptide/protein identifications. MaxQuant is a set of algorithms for peak detection and scoring/identifying peptides. Taking

raw mass spectral data MaxQuant assembles a 3D image from MS spectra over m/z retention times and ion intensity (Schaab et al. 2012).

In this thesis we used label free quantitation by LC-MS/MS analysis on human patient samples in both chapter 4 (Ting et al. 2017) and chapter 5 (Dowling et al. 2015). We also used label free quantitation by LC-MS/MS to study phosphopeptide enriched Chinese hamster ovary cell samples in both chapter 7 (Henry et al. 2017a) and chapter eight (Kaushik et al. 2018) and finally in chapter 9 we studied intracellular CHO cell proteins in by label free quantitation (Henry et al. 2018).
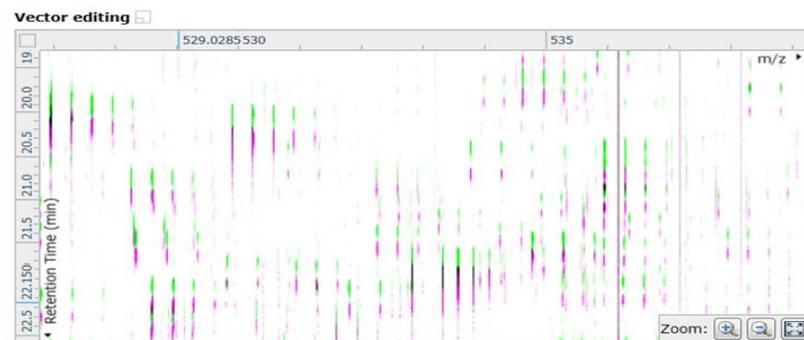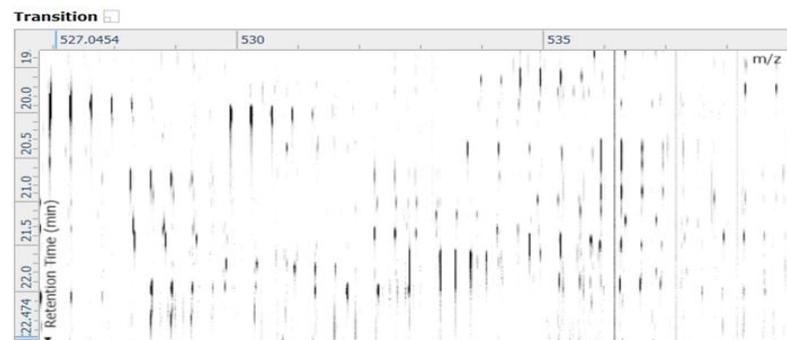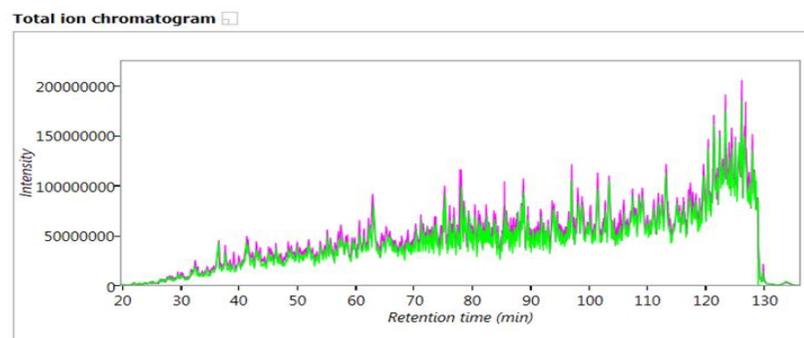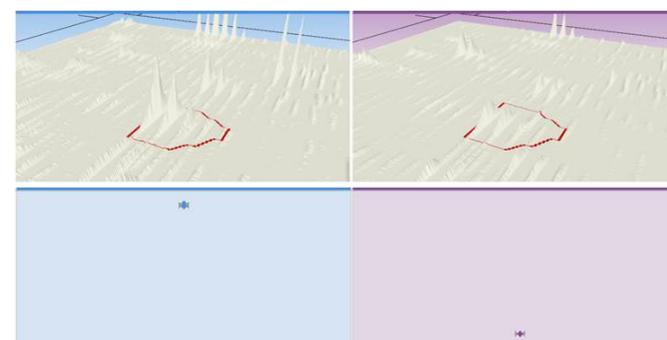
Figure 1-14. Progenesis QI pipeline. A+B: Represent ion map overlapping where peptides are matched between samples. In this figure x-axis represents the m/z and the y-axis represents the retention time. Peptides are visualised as an alignment target in pink and the ion map to be aligned to in green C: Total Ion Chromatogram view of the finished aligned run. D: ion abundance quantification and statistical analysis.

1.2.6.2 Labels - Stable isotopes Isotopic tags/Isobaric mass tags

Isobaric chemical tags in conjunction with MS/MS strategies allow for quantification and identification of many different pooled samples in the one experiment. The isobaric tags are very small chemical molecules with identical structure that covalently bind to the free amino termini of lysine residues of peptides thus allowing labelling peptide samples in a given experiment (see Figure 1-15). Following MS/MS of labelled peptides, each isobaric tag produces a unique reporter ion signature thus making relative quantitation possible when compared to the other reporter ions intensities in the same MS/MS spectrum.

The two main labels used by proteomic researchers are tandem mass tag (TMT) and isobaric tag for relative and absolute quantitation (iTRAQ). iTRAQ chemical labelling technique allows multiplexing of up to eight different samples on the peptide level (Pichler et al. 2010).

The TMT 10-plex reagent set contains ten different isobaric compounds with the same mass and chemical structure (i.e. isotopomeric) composed of an amine-reactive NHS-ester group, a spacer arm and a mass reporter. The reagent set enables up to ten different peptide samples prepared from cells or tissues to be labelled in parallel and then combined for analysis. For each sample, a unique reporter mass (i.e., TMT10 126-131Da) in the low-mass region of the high-resolution MS/MS spectrum is used to measure relative protein expression levels during peptide fragmentation and tandem mass spectrometry (Figure 1-15). McAlister et al. (G. McAlister et al. 2012) used TMT labelling to examine 8 different colorectal cancer cell lines in culture using complex basic pH fractionation and low pH fractionation (24 fractions in total) into a Velos Orbitrap Elite (Thermo Scientific) mass spectrometer running a multinotch MS^3 methods. Hughes et al. (Hughes et al. 2017) further demonstrated improved sensitivity and MS^3 isolation strategies for the use of Isobaric Tags and the latest Orbitrap Fusion mass spectrometers. They performed an in-depth examination of the various acquisition methods available on the Orbitrap Fusion MS that yielded benefits in terms of proteome coverage and validated method configurations and parameters for isobaric tagging on current generation Orbitrap instruments. Currently, up to 11-plex labelled experiments are now possible with TMT.

Figure 1-15. Thermo Scientific offer the TMT 10plex^{Tm} Isobaric Mass Tag amine reactive system allowing up to 10 peptide samples to be labelled in parallel with TMT reagents and then combined. The labelled samples can be fractionated and then analysed by high resolution mass spectrometry before the data analysis allows for reporter ion relative abundance quantification and peptide identification. The relative abundances of the target peptides from the different labelled samples are measure by comparing the reporter ions generated on MS/MS fragmentation of the different mass tags.

1.2.6.3 Mass Spectrometry for Quantitative Targeted Proteomics

Desiderio et al. (Desiderio et al. 1983) first published targeted Mass Spectrometry assay for peptide quantitation and used isotopically labelled peptides as internal standards. Later the term Multiple Reaction Monitoring (MRM) was coined by Kusmierz et al. (Kusmierz et al. 1990) who compared spectra of labelled peptide samples to unlabelled in human tissue extracts. More recent approaches include Selective Reaction Monitoring (SRM) which allow the researcher to focus on the quantitation of a number of proteins of interest to perform either relative or absolute quantitation with very high specificity and sensitivity (Catenacci et al. 2014). Targeted proteomics requires monitoring of proteotypic peptides (PTPs) with a unique amino acid sequence to identify a specific protein of interest in the sample proteome being analysed by mass spectrometry (Adeola et al. 2016). The selection of the PTP is essential as targeted proteomics only involves monitoring the known precursor peptide and the fragmented ions generated from that peptide i.e. the protein(s) under investigation are known beforehand. Targeted approaches generally start with candidate proteins from a hypothesis based approach from proteomic data generated from results of extensive fractionation methods. Using SRM it is imperative to use peptides that are specific to the candidate protein of interest and give sufficient signal intensity. In this way the experiment can be qualitatively and quantitatively validated in the samples being analysed. Liquid chromatography separation times, ionization and fragmentation setting must be set prior to the experiment starting. As the target peptides are known prior to analysis, MS conditions can be set up to improve selectivity and dynamic range of the instrumentation compared to conventional shotgun proteomic methods (Shi et al. 2016). Targeted MS assays also allow for multiplexing exceeding those possible by standard immunological assays.

The classical targeted approach uses triple quadrupole mass spectrometers using the first and third quadrupoles as m/z filters and the second quadrupole as a collision cell to fragment the peptide(s) (Picotti et al. 2013). The parent and product ions are described as transitions which are further analysed in the third quadrupole to filter specific ions thus improving selectivity and sensitivity (Peterson et al. 2012). Absolute quantitation involves spiking synthetic peptides into samples of interest at known concentrations and following mass spectrometry on the sample, the precursor and fragmented peaks of the synthetic peptide are compared to that of the natural occurring peptide in the sample. Using labelled synthetic peptides, it is possible to validate biomarkers by targeted proteomics (Gallien et al. 2011). The stable isotope-labelled/stable isotope-dilution (SIL/SID) has

become a gold standard for SRM protein quantitation where synthetic peptides are labelled with stable isotopes at the C-terminal Arginine and Lysine residues ($^{13}C_6$ $^{15}N_4$ for Arg and $^{13}C_6$ $^{15}N_2$ for Lys) that are chemically identical to the light native peptide of interest (AQUA peptide) (Gerber et al. 2003). Targeted approaches like these can be difficult to develop and maintain and this is true when methods require complex retention time scheduling which are heavily reliant on reproducible chromatography and can often require constant optimisation of conditions. Newer instrumentation has allowed the move away from the traditional triple quadrupole mass spectrometers with the advent of hybrid mass spectrometer instruments including the Thermo Scientific QExactive range and Orbitrap Fusion Tribrid instruments which have the speed, selectivity and sensitivity to allow for targeted proteomics. These instruments can be run as shotgun proteomics instruments and as target proteomic instruments.

Parallel reaction monitoring (PRM) is another labelled peptide targeted proteomic technique using specifically Orbitrap mass spectrometer (Gallien et al. 2015).The workflow requires the addition of heavy peptides standards into the analytical sample of interest, with sequences that are analogous to endogenous peptides of interest. The heavy isotopes peptide standard differs in mass from the endogenous forms of the peptides; however, the retention times match the endogenous peptides exactly. In samples where the target peptide may be present in very low concentrations, once the specific heavy peptide is detected the instrument automatically triggers high-quality MS/MS analysis on the expected location of the endogenous form of the peptide.

Güzel et al. (Güzel et al. 2018) studying proteomic alterations in early stage cervical cancer used PRM to quantify *MCM3*, *CEACAM5*, *S100P* and *ICAM1* in digests of whole tissue lysates using stable isotope-labelled (SIL) peptides. Sandow et al. (Sandow et al. 2018) detected and identified the urinary biomarkers MSLN and CA125 in urine from ovarian cancer patients using PRM and observed their abundance increase in malignancy consistent with literature (Hellstrom et al. 2011). Rauniyar et al. (Rauniyar et al. 2018) also used PRM to quantitate multiple biomarkers in various diseases, e.g. 143 biomarkers in a kidney disease study.

1.3 Overview of Orbitrap MS instrumentation used in this thesis

1.3.1 LTQ Orbitrap XL

The Orbitrap was commercially available in 2005 and gained popularity with proteomic groups due to its high resolving power. The LTQ Orbitrap XL is a hybrid instrument that has two distinct mass analysers. The linear ion trap front end (LTQ) and the Orbitrap allowing for combined tandem mass spectrometry in the ion trap (LTQ) with the high resolution and high mass accuracy capabilities of the FT Orbitrap.

The curved linear traps (C-trap) focuses ions in narrow areas to be ejected as compact bursts of ions which is achieved by altering the trapping fields of the electrodes at the front and rear of the instrument. The presence of a collision gas reduces the kinetic energy of the ions within the trap. Curved linear traps provide an efficient way to feed ions into the orbitrap. The curved linear ion trap allows ions to be injected into the orbitrap at a position offset from its equator and in fact, by doing this, ions will perform coherent axial oscillations almost immediately after injection without the need for any additional excitation; this type of ion injection is known as "fast injection"(Badman et al. 1999).

Ions within the orbitrap have three different frequencies of oscillation: axial, radial and rotational. In order to detect frequencies of oscillations, the motion of ions needs to be coherent. Coherence in the axial direction can be induced by different ways (Hu et al. 2005). The Fourier Transform of the signal can be used to produce the required mass spectrum. This mode of detection yields the highest mass resolution that the orbitrap can achieve. The Fourier Transform mode measures coherent oscillations in the axial direction (image current detection). See Figure 1-16 for an overview of the Orbitrap XL schematics.
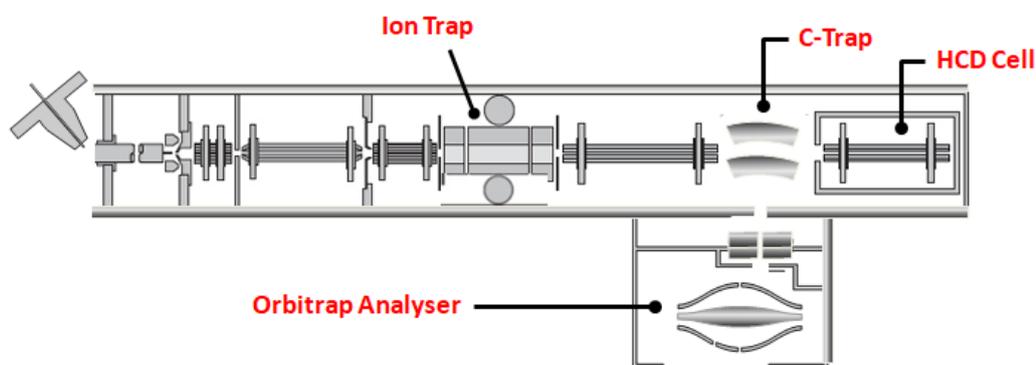
Figure 1-16. Schematic of the Orbitrap XL mass spectrometer

We used the LTQ Orbitrap XL in chapters two, four, five, seven and nine in this thesis. The instrument was operated for all "bottom up" approaches at high resolution in the Orbitrap mass analyser (m/z 400-1200) with a resolution of 30,000 at m/z 400 and an automatic gain control (AGC) target of 10^6. The 3 most intense ion were selected for CID MS/MS fragmentation in the linear ion trap at a threshold of 1,000, an AGC of 100,000, a spray voltage of 1.8 kV, normalised collision energy of 32% at an activation of q=0.25 and an activation time of 35 ms. Detection in the linear ion trap of previously selected ions were dynamically excluded for 40 seconds.

Figure 1-17 shows an example of a 1 µg CHO K1 total lysate sample prepared using the FASP method (Coleman et al. 2017) and separated over 3 h into an Orbitrap XL. Using the parameters described above, 18,908 MS/MS were acquired, and following SEQUEST algorithm 7,073 peptides were matched to 1,502 CHO proteins. The top 3 abundant peptides (highlighted in red) from the full MS scan are selected for MS/MS.

Figure 1-18 shows an example of a CHO K1 phosphopeptide-enriched sample using IMAC separated over 3 h into an Orbitrap XL using a neutral loss MS3 data dependent method. Neutral loss of phosphoric acid in ion trap fragmentation can result in reduced intensity of sequence information so additional MS/MS scan of the dominant peak can increase the quality of the spectra for sequence information. Using this method on the Orbitrap XL we identified 1,634 site specific phosphorylated peptides. A phosphopeptide with a mass of 787.8 m/z (z=2) highlighted in red was selected for MS/MS resulting in a dominant peak at 738.8 m/z corresponding to the loss of phosphoric acid in the MS/MS scan which then triggered an additional fragmentation of the product the precursor neutral loss in the form of an MS3 scan (MS/MS/MS) providing additional sequence information.
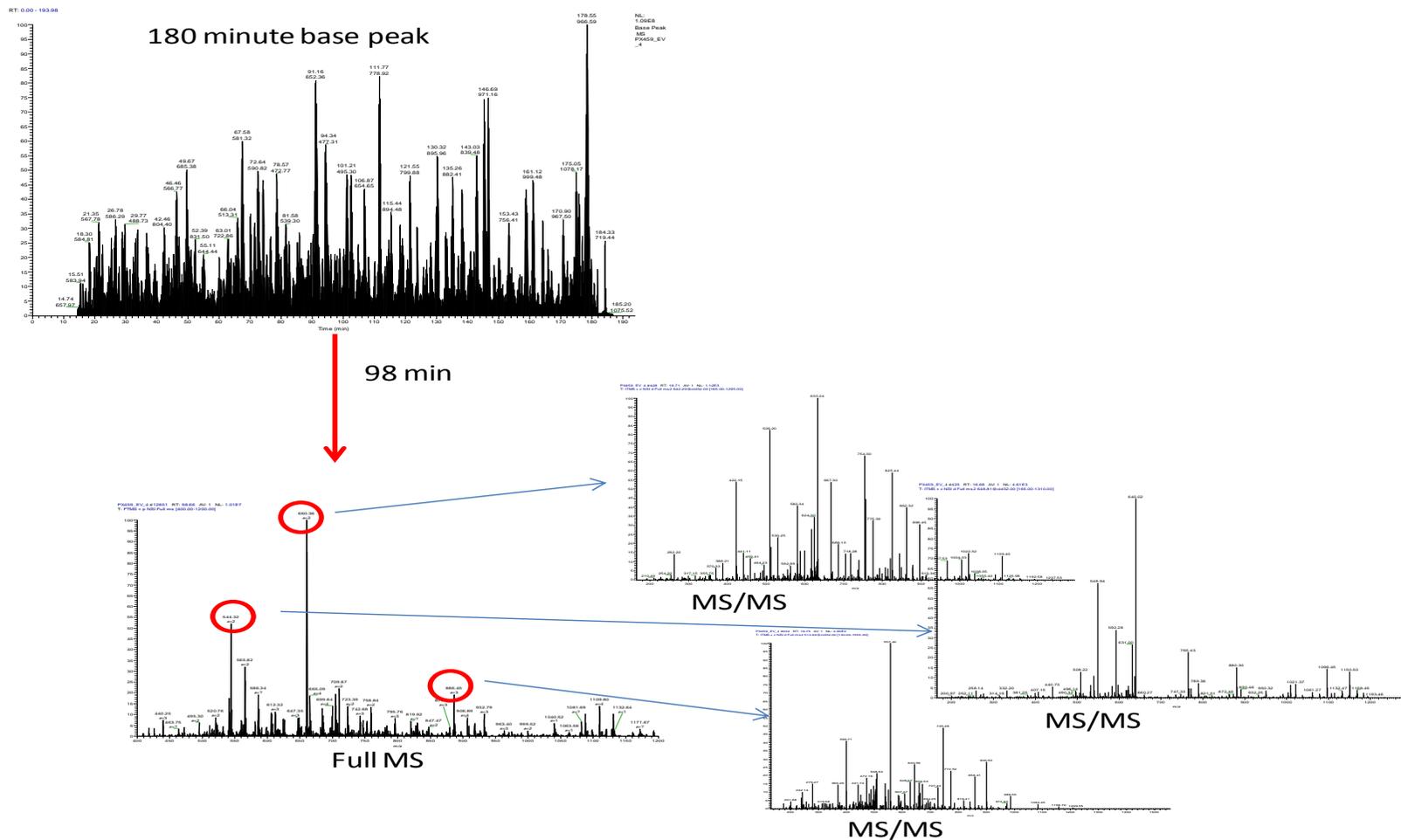
Figure 1-17. Base peak chromatogram of a CHO K1 lysate digestion separated for three hours into an Orbitrap XL. Full MS spectrum at 30,000 resolution and the top three peptides based on their intensities were selected for MS/MS analysis.

180 minute base peak

42 min

Peptide 787.8 m/z (z=2) selected for MS/MS
A dominant peak at 738.8 m/z corresponding to
the loss of phosphoric acid in the MS/MS scan which
triggers an additional fragmentation of the product
the precursor neutral loss in the form of an MS3 scan

Full MS
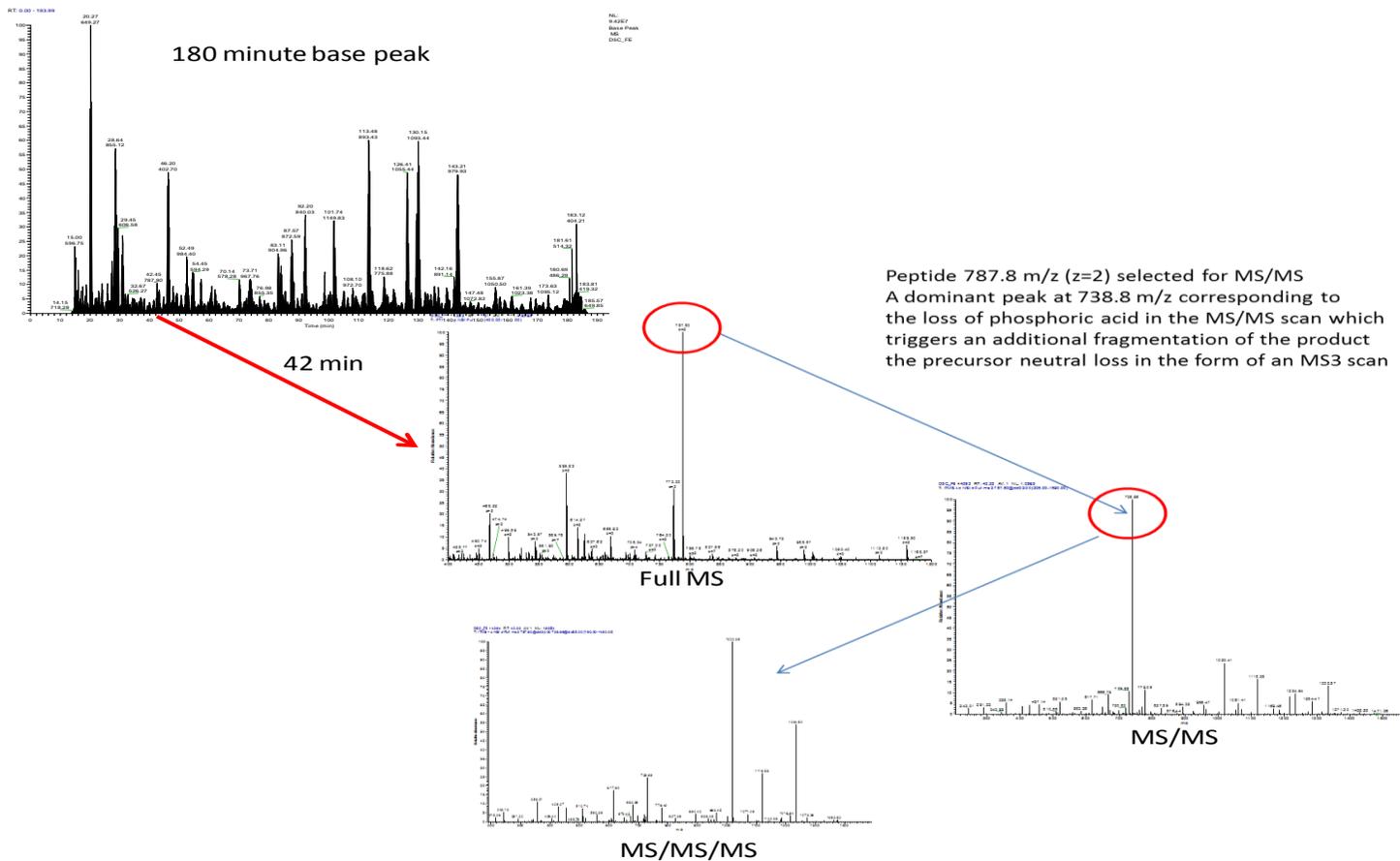
MS/MS

MS/MS/MS

Figure 1-18. Phosphopeptide with a mass of 787.8 m/z (z=2) selected for MS/MS. The resulting spectral pattern had a dominant peak with a mass of 738.8 m/z corresponding to the loss of phosphoric acid from the parent mass. This measured mass loss triggers an additional MS/MS of the product in the form of an MS3 scan generating additional fragmentation information.

1.3.2   Orbitrap Fusion Tribrid

The Orbitrap Fusion Tribrid mass spectrometer combines the best of quadrupole, orbitrap and linear ion trap analysis in one instrument. See Figure 1-19 for a schematic overview of the Orbitrap Fusion Tribrid. Multiple fragmentation techniques are possible with CID, HCD, ETD and EThCD at any stage of MS^n with the subsequent mass analysis in either low resolution ion trap or high resolution orbitrap. Next generation ion sources and ion optics have increased sensitivity and speed of MS and MS/MS acquisition. The ability to carry out ultrahigh resolution at 500,000 at 200 m/z enables molecular weight determination of intact proteins.

1.3.2.1   Improvements in ion manipulation with the Orbitrap Fusion Tribrid

The active beam guide (ABG) is a bent flatapole lens ion guide which now reduces noise by preventing neutrals and high-velocity clusters from entering the quadrupole. An axial field applied along the length of the rods improves robustness by eliminating the effects of local charging (see Figure 1-19).

The quadrupole mass filter is used to select a specific mass range for transmission to the ion optics and detectors downstream. Its high ion transmission at isolation widths down to 0.4 amu improves sensitivity and selectivity. Radio Frequencies (RF) and DC voltages and the selected values in the quadrupole mass filter determine the range of m/z charge ratios allowed to be transmitted. The quadrupole RF amplitude and DC voltage are set to specific values which then allow only ions of a certain m/z range to be maintained within bounded oscillations as their velocity carries them through the mass filter.

The ion-routing multipole (IRM) has multiple functions. It accumulates ions and route them to either the linear ion trap or the Orbitrap mass analyser. It also performs HCD at any fragmentation stage. The IRM allows parallel analysis, a function only found in the Orbitrap Fusion Tribrid mass spectrometer.

Dual-Pressure Linear Ion Trap. A high-pressure cell for MS^n precursor ion isolation from 0.2 amu up to 600 amu and both Collision-induced dissociation (CID) and electron-transfer dissociation (ETD). A low-pressure cell allows for an improved scan speed, resolving power, and mass accuracy.

The Orbitrap Fusion Tribrid uses the new Ultra-High-Field Orbitrap which is 1.5 times smaller than the previous orbitrap with resolving power up to 500,000 FWHM.
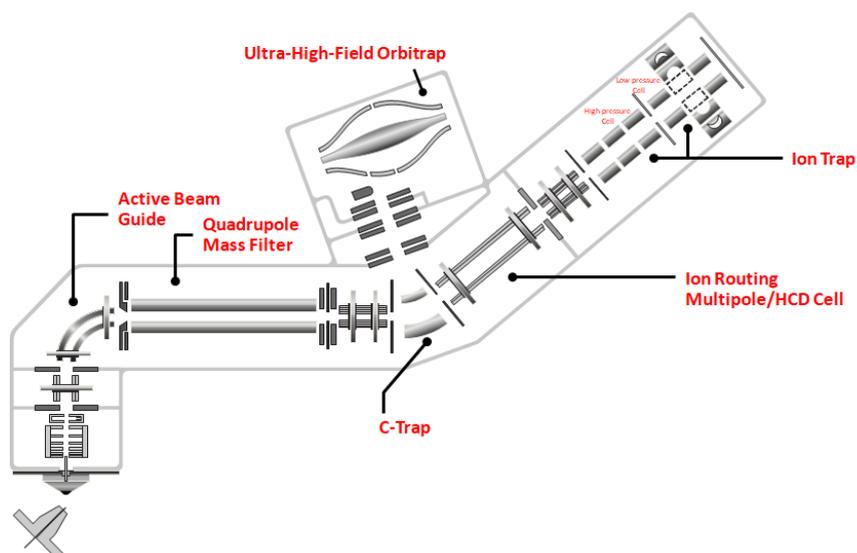
Figure 1-19. Schematic of the Orbitrap Fusion Tribrid Mass Spectrometer with the new hardware implementations. The active beam guide for noise reducing. The Quadrupole mass filter to select specific mass range for transmission selectivity and improved sensitivity. The ion-routing multipole for parallel analysis. Accumulates ions for either HCD fragmentation or routing to the ion trap.

1.3.2.2 Proteomic analysis using the Orbitrap Fusion Tribrid mass spectrometer

A typical operation of an Orbitrap Fusion Tribrid for "bottom up" analysis is described in Chapter nine of this thesis. Basically, the instrument is run at high resolution in the Orbitrap mass analyser (m/z 375-1500) with a resolution of 120,000 at m/z 200 and an automatic gain control (AGC) target of $4.0e^5$ and an intensity threshold of $5.0e^3$. For MS/MS in a 3 second window, the AGC target was $2.0e^4$ with as many MS/MS with the HCD collision energy set to 28 and a maximum injection time of 35 ms. The quadrupole isolation window was 1.6 m/z and a spray voltage of 1.8kV.

See Figure 1-20 for an example of a 1 µg CHO K1 total lysate sample prepared using the FASP method (Coleman et al. 2017) and separated over 140 mins into an Orbitrap Fusion Tribrid. Using the parameters above, 111,850 MS/MS were acquired and following SEQUEST algorithm 30,933 peptides were matched to 4,822 CHO proteins.
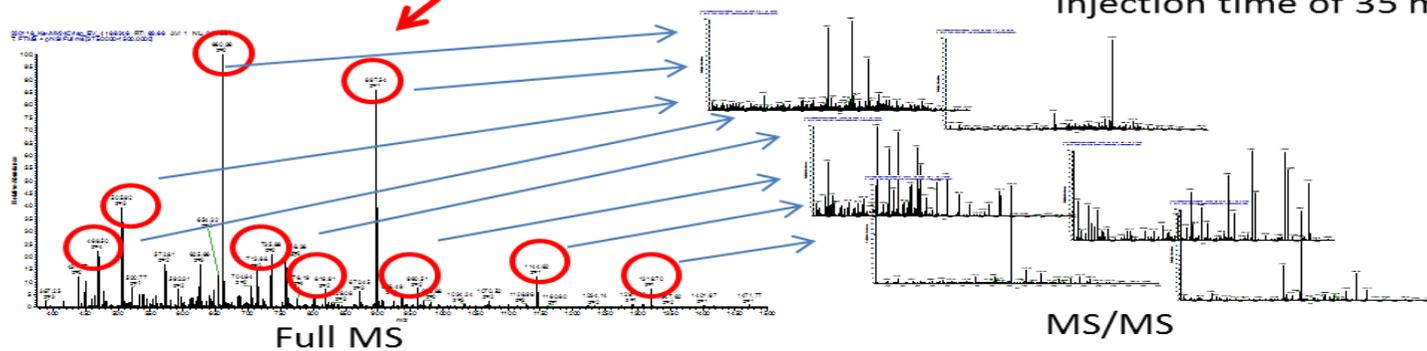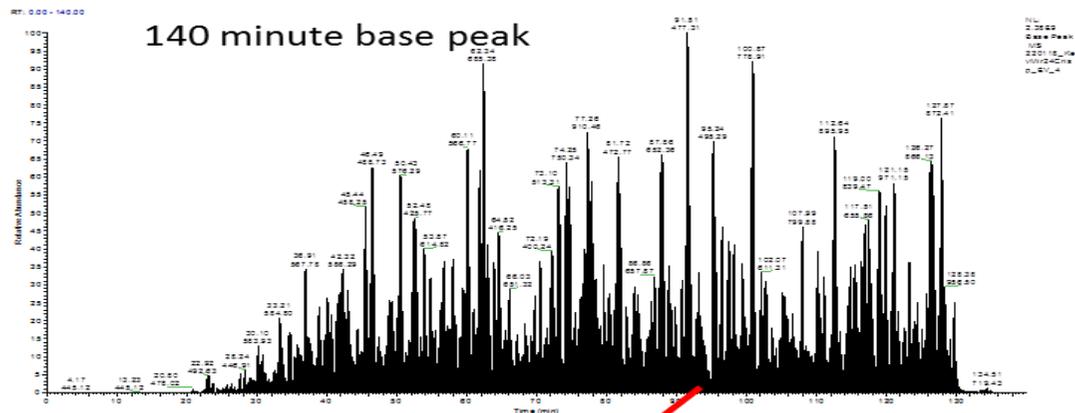
Figure 1-20. Base peak chromatogram of a CHO K1 lysate digestion separated over a 140 min RPC separation into an Orbitrap Fusion Tribrid instrument. Full MS spectrum at 120,000 resolution and MS/MS was set up to acquire in the ion trap as many scans over a 3 second window with a maximum injection time of 35ms for each MS/MS.
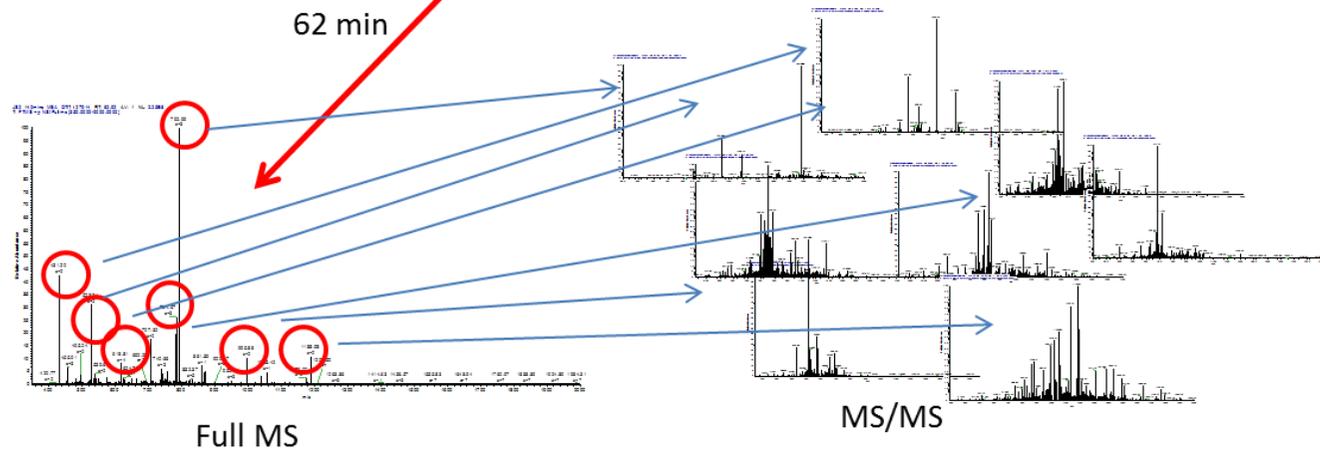
Figure 1-21. Base peak chromatogram of a CHO SEAP IMAC phosphopeptide enriched sample separated over a 120 m RPC separation into an Orbitrap Fusion Tribrid instrument. Full MS spectrum was set 120,000 resolution and MS/MS fragmentation was carried out in the ion trap part of the instrument using MSA.

Figure 1-21 shows an example of a CHO K1 phosphopeptide-enriched sample using IMAC separated over 2 h into an Orbitrap Fusion using MSA fragmentation. MSA methods assumes that the peptide has a predominant loss associated with phosphorylated peptides. The MSA pseudo-MS3 scan fragments the peptide and then carries out a subsequent specified mass below that precursor is fragmented; in this case -98 Da for phosphate. All of the ions are then detected in a single scan. It is assumes the major loss will be occurring in all cases, so the instrument does not spend time scanning out to see if that fragment is there first. We identified 6,079 phosphopeptides from a single 1 µg phosphopeptide enriched sample from CHO SEAP cells.

1.3.3 Comparison of LTQ Orbitrap XL vs Orbitrap Fusion Tribrid proteomic analysis

Table 1-2 outlines a brief comparison of Orbitrap XL and Orbitrap Fusion instruments main features (Scan rates/speed refers to the number of spectra per unit time that be generated. Scan rate can be defined as the spectral generation rate in hertz).

Table 1-3 and Figure 1-22 outline a comparison of the Orbitrap XL and Orbitrap Fusion Tribrid instrumentation and a comparison of the same CHO KI sample analysed on both instruments.

	Mass Resolution	Mass Accuracy	Scan rate	Fragmentation
Orbitrap Fusion	15,000–500,000 (FWHM) at m/z 200	<3ppm RMS using External calibration	Orbitrap MSn up to 18 Hz Ion trap MSn up to 22-25 Hz	CID, MSA, HCD, ETD, ETciD, EThcD
Orbitrap XL	7,500-100,000 at m/z 400	<3ppm RMS using External calibration	Orbitrap MSn up to 20 Hz Ion trap MSn up to 8-12 Hz	CID, HCD and ETD (optional).

Table 1-2. Comparison of Orbitrap XL and Orbitrap Fusion Mass Spectrometry features. (Scan rates/speed refers to the number of spectra per unit time that be generated. Scan rate can be defined as the spectral generation rate in hertz).

Instrument	Separation	Full MS / MS^2	Fragmentation	Data dependent settings
Orbitrap Fusion	Gradient 5 – 30% B in 120 min	120K / Ion trap (low res)	HCD - 28%	Max MS/MS within 3 seconds of Full MS
Orbitrap XL	Gradient 5 – 30% B in 180 min	30K / Ion trap (low res)	CID - 32%	3 MS/MS per Full Scan

Instrument	Sample	Proteins	Peptides	PSM	No. of MS^2	Proteins ID (min. 2 peptide)
Orbitrap Fusion	Chinese Hamster Ovary cell digest	4822	30,933	40,459	111,850	3641
Orbitrap XL	Chinese Hamster Ovary cell digest	1502	7073	8843	18,908	1045

Instrument	Separation	Full MS / MS^2	Fragmentation	Data dependent settings
Orbitrap Fusion	Gradient 5 – 30% B in 120 min	120K / Ion trap (low res)	MSA - 28%	Max MS/MS within 3 seconds of Full MS with MSA merged spectra

Instrument	Sample	Proteins	Phospho Peptides ptmRS score >75%	PSM	No. of MS^2
Orbitrap Fusion	Chinese Hamster Ovary cell Phospeptide enriched sample	2879	6079	11,580	86,831
Orbitrap XL	Chinese Hamster Ovary cell Phospeptide enriched sample	876	1634	2317	10,981

Table 1-3. Separation and Mass spectrometry setting comparing Orbitrap XL and Orbitrap Fusion analysis of a Chinese Hamster Ovary digest sample and a Phosphopeptide enriched sample.
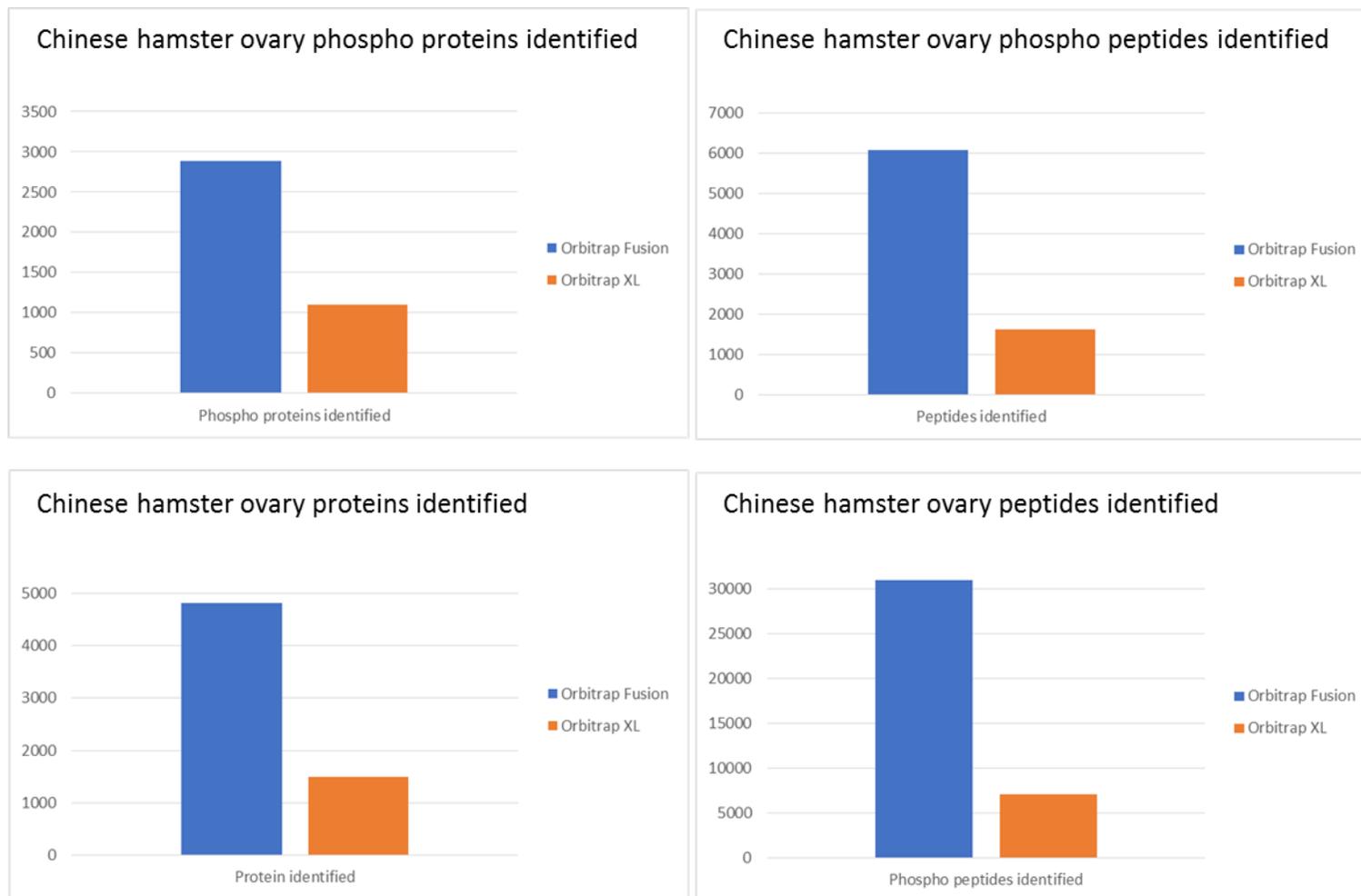
Figure 1-22. A comparison of Orbitrap fusion Tribrid MS data to Orbitrap XL MS data using the same Chinese hamster ovary tryptically digested protein sample. The Orbitrap fusion Tribrid instrument confidently identified over four times as many peptides translating into confidently over three and a half time more proteins.

1.4   An overview of Chinese hamster ovary cell proteomics

The Chinese Hamster was brought into the United States in 1948. In 1957, Theodore Puck obtained a Chinese hamster from a laboratory at the Boston Cancer Research Foundation, where he and his colleagues used it to derive the original CHO cell line (Puck et al. 1957).The CHO cell line generates billions in sales from the generation of therapeutic monoclonal antibodies and recombinant proteins for the biotechnology industry. Establishment of stable transfected CHO cell lines is the basis of the production of therapeutic proteins for the biotechnology industry. Once a cell line is developed, its productivity and expression profile is optimised either by culture medium and production processes to maximize high quality yield of the product. These optimised biotechnology industry standards for cell line development are costly and time consuming. CHO cells are, however, immortalised cells, characterised by a high degree of genetic and phenotypic diversity. Varying profiles make a universal approach for cell line development and optimisation very difficult (Wurm et al. 2013). A better understanding of the biological and cellular characteristics of the CHO proteome will help improve productivity levels and quality.

There have been a number of proteomic studies carried out to try and understand CHO cells in order to identify targets that can be used in engineering targets and to improved understanding of the cellular and molecular biology of these cells. These studies have been carried out using gel based (e.g. 2D gels) and gel-free approaches.

Using 2D-DIGE, Meleady et al. (Meleady et al. 2011a) carried out proteomic profiling on two matched pairs of monoclonal antibody producing CHO cell lines that differed in their ability to sustain productivity over a 10 day fed batch culture which resulted in the identification of 89 differentially expressed proteins. They concluded that these proteins may have an important role in sustaining the important phenotype trait of high productivity. Table 1-4 outlines an overview list of recent CHO cell proteomic applications using 2D gels approaches.

An alternative strategy to gel based "classical proteomics" involves the use of in-solution protein extract digestion, peptide fractionation by liquid chromatography (single or multidimensional) coupled with electrospray ionisation tandem mass spectrometry. This gel free approach can also be quantitative using labels (e.g. iTRAQ, TMT) or label free MS techniques. Meleady et al. (Meleady et al. 2012) used quantitative label free LC-MS/MS

profiling to study CHO protein abundance following over expression of miR-7 in CHO SEAP cells. From the analysis they found 93 proteins decreased in expression levels and 74 increased following over expression of miR-7 in CHO SEAP cells compared the null cells. Sommeregger et al. (Sommeregger et al. 2016) used quantitative label free LC-MS/MS to investigate the intracellular proteome difference in CHO cells producing two similar antibody fragments. They concluded that not only the levels of the specific antibody secretion but the antibody itself had a large impact on the cell proteome.

During the manufacturing process, some proteins derived from the host cells are introduced into the downstream process. It is essential that these host cell proteins (HCPs) impurities are cleared (to acceptable levels) as they can have a detrimental effect on product quality and also patient exposure to a biotherapeutic drug containing HCPs can induce anti-drug antibodies and anti HCP antibodies (Bracewell et al. 2015). Regulatory guidance mandates the control of Host cell protein (HCP) concentration in the production of recombinant therapeutic proteins (Valente et al. 2018). The HCP content from cultured CHO cells is generally measured using Enzyme Link ImmunoSorbent Assays (ELISA) and 2D-PAGE (Hogwood et al. 2017). Using western blotting of 2D-PAGE gels and estimation of HCP clearance or presence can be achieved. As western blotting confirms and determines relative abundance of the specific known protein(s) in a 2D gel, mass spectrometry and In-gel digestion can be employed to identify unknown proteins in the 2D gel. The recovery of extracellular CHO HCP for proteomics analysis can be challenging due to the relatively low protein concentrations and complex composition of media (Valente et al. 2015).

The regulatory trend towards deeper characterisation of HCP impurities has led to greater use of mass spectrometry (MS) to corroborate and supplement traditional immunoassays. Liquid chromatography coupled to mass spectrometry (LC-MS) has become, over the last few years, the technology of choice for protein analysis. The improvement in MS-based proteomics provides the identification and quantification of thousands of proteins in a single run. As a result of this LC-MS based approaches are becoming increasingly more commonly used as orthogonal and powerful alternative methods to ELISA by enabling identification, monitoring and quantitation of low level HCPs (Valente et al. 2018).

Park et al. (J. H. Park et al. 2017) analysed HCPs from fed-batch cultures and batch cultures using gel free MS proteomics. They identified 2,145 HCPs in the fed-batch culture and 1934 proteins in the batch culture from 36 LC-MS/MS runs using a QExactive Mass spectrometer. Kumar et al. (A. Kumar et al. 2015) used LC-LC-MS/MS analysis to identify host cell proteins

(HCPs) in conditioned medium from CHO K1 cells in culture with a view to evaluate their immunogenicity and help assess process and purification methods used in the production of biologic drugs. Walker et al. (Walker et al. 2017) examined up to 20 samples per day looking at CHO HCP composition and quantitation using 1D LC-MS/MS with SWATH DIA MS. Using a one hr method they could detect HCPs at the ng/mg range. Table 1-5 outlines a list recent CHO cell proteomic publications using HCP analysis.

In this thesis, I have used gel-free quantitative CHO cell proteomics to understand the intracellular events that could play a role in determining why cell lines from the same cell line development project can vary in the extent of Fc-fusion protein clipping using differential quantitative label-free LC-MS/MS proteomic analysis (Henry et al. 2018). Our group were the first in the field to publish in CHO phosphoproteomic analysis using IMAC and TiO$_2$ phosphopeptide enrichment strategies in conjunction with differential quantitative label-free LC-MS/MS proteomic analysis to study cellular processes in CHO cells following a reduction of culture temperature (temperature shift) (Henry et al. 2017a). We substantially improved our understanding of phosphoproteomic CHO cell processes with a follow up study using state-of-the-art mass spectrometry to analyse IgG producing CHO cells at various phases of growth (Henry et al. 2018)**.** Table 1-6 outlines an overview of gel-free CHO cell proteomic approaches.

Study/Comparison	Technique	Key Finding	Author
CHO cell comparison of stable transfection to express a heavy chain monoclonal antibody	2D-DIGE LC-MS/MS	Demonstrated proteins associated with cell stress via up-regulated metabolism, depleted nutrients and accumulated waste metabolites.	(Blondeel et al. 2016)
Review of principles of 2D-PAGE and 2D-DIGE on recombinant Chinese hamster ovary cells	2D-DIGE	Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and two-dimensional difference gel electrophoresis (2D-DIGE) continue to be one of the most versatile and widely used techniques to study the proteome of a biological system.	(Meleady 2018)
2D-DIGE profiling analysis of two matched pairs of monoclonal antibody-producing CHO cell lines that differ in their ability to sustain productivity over a 10 day fed-batch culture	2D-DIGE and LC-MS/MS	89 distinct differentially expressed proteins identified with 12 proteins (AKRIB8, ANXA1, ANXA4, EIF3I, G6PD, HSPA8, HSP90B1, HSPD1, NUDC, PGAM1, RUVBL1 and CNN3) that were differentially expressed in the same direction that may have an important role in sustaining high productivity of recombinant protein over the duration of a fed-batch bioprocess culture.	(Meleady et al. 2011a)
4 CHO DUKK Mab Secreting cells designated fast and slow growing	2D-DIGE and LC-MS/MS	58 proteins identified as differentially expressed playing an important role in the regulation of cell growth during log phase and as potential candidates for CHO cell line engineering	(Doolan et al. 2010)
CHO MUC1 IgG2a grown in 4 different perfusion cultivation systems	2D DIGE and MALDI-TOF MS	Demonstrated consequences of glucose limitation for the protein expression machinery and could play a role in higher recombinant protein production	(Wingens et al. 2015)

Comparative differential study using rCHO cells producing and antibody in serum free medium.	2D Gels and LC-MS/MS	A greater understanding of the effect of hydrolysates on CHO intracellular events.	(J. Y. Kim et al. 2012)
Proteomic analysis of CHO cells undergoing apoptosis during prolonged cultivation.	2D DIGE and LC-MS/MS	A foundation for optimising cell line specific cultivation processes and prolonging longevity although the study did not identify direct apoptosis regulating proteins.	(Wei et al. 2011)
Utilisation of CHO specific sequence database and mass spectrometry proteomics	2D gels and LC-MS/MS	Increased by 40-50% total number of proteins identified from 2D gels using MS based proteomics.	(Meleady et al. 2012b)

Table 1-4 Compiles a list of recent CHO cell proteomic applications using 2D gels for intracellular analysis

Study/Comparison	Technique	Key Finding	Author
Label Free study looking at Clonal variation in productivity and proteolytic clipping of an Fc-fusion protein in CHO cells	LC-MS/MS label free Profiling	This study identified protein targets that could be modified using cell line engineering approaches to improve the quality of recombinant Fc-fusion protein production in the biopharmaceutical industry Proteomic analysis suggested a role for defective protein folding and the UPR	(Henry et al. 2018)
Large-scale label free differential phosphoproteomic study of IgG producing CHO DP12 cells at various phases of growth.	LC-MS/MS label free Profiling	This study identified 3,777 differentially expressed unique phosphopeptides from 1,415 differentially expressed unique phosphoproteins. CHO phosphoproteomic data gives significant insights into the post-translational level of regulation during cellular growth of recombinant CHO cells.	(Henry et al. 2018)
CHO K1, CHO S and CHO DG44 cells grown in batch culture in two different types of basal media.	Multiple Reactions Monitoring LC-MS	The monitoring of protein biomarkers for the early prediction of cell stress and death is a valuable tool for process characterization and efficient biomanufacturing control	(Albrecht et al. 2018)
Large scale profiling of CHO-S and CHO-DG44	LC-MS/MS profiling	11,801 unique CHO proteins identified	(Heffner et al. 2017)
Phosphopeptide enrichment quantitative analysis of CHO SEAP cells following temperature reduction	Label Free LC-MS/MS profiling	700 differentially expressed phosphopeptides were identified relating to growth, ribosomal biogenesis and cytoskeleton organisation.	(Henry et al. 2017a)
Comparative study of CHO cell lines secreting two similar antibody fragments	Label Free LC-MS/MS profiling	They concluded that not only the levels of the specific antibody secretion but the antibody itself had a large impact on the cell proteome	(Sommeregger et al. 2016)

Filter-Aided Sample Preparation (FASP) for Improved Proteome Analysis of Recombinant Chinese Hamster Ovary Cells**.**	LC-MS/MS profiling	A Filter Aided Sample Preparation (FASP) protocol for the extraction of proteins from CHO cells for proteomic studies using ultrafiltration devices where the membrane pores are small enough to allow contaminating detergents to pass through, while proteins are too large and are retained and concentrated in the filter unit.	(Coleman et al. 2017)
iTRAQ peptide quantitation of HCPs from antibody producing CHO-S cells.	Labelled LC-MS/MS profiling	iTRAQ labelled quantitative analysis for CHO HCP analysis as an application for process development knowledge of critical HCPs.	(Chiverton et al. 2016)
iTRAQ peptide quantitation of supernatant from three transfected CHO cell lines.	Labelled LC-MS/MS profiling	There was a 60% protein identification overlap in the proteins identified between the transfected CHO clones.	(Zhu et al. 2016)
Re-programmed CHO K1 cells using miR-23	Label Free LC-MS/MS profiling	Label-free proteomic analysis uncovered various potential novel targets of miR-23 including LE1 and IDH1, both implicated in oxidative metabolism and mitochondrial activity.	(Kelly et al. 2015)
Three CHO cell lines secreting monoclonal antibody exposed to suboptimal copper levels.	Label Free LC-MS/MS profiling	Label-free proteomic analysis uncovered under copper deficient conditions, a substantial reduction in the protein levels of the multiple subunits of cytochrome c oxidase of the mitochondrial electron transport chain.	(S. Kang et al. 2014)
Identify CHO proteins from the secretome, cellular proteome and glycoproteome	LC-MS/MS Profiling	6164 grouped CHO proteins were identified from the combined glycoproteome and proteome analysis.	(Baycin-Hizal et al. 2012a)
Metabolomic profile to eludicate characteristics of high mAb producing CHO cells	LC-MS/MS Profiling	High producing CHO cell lines had elevated levels of specific metabolites for regulation of their redox status	(Chong et al. 2012)

Utilisation of CHO specific sequence database and mass spectrometry proteomics	LC-MS/MS Profiling	Increased by 40-50% total number of proteins identified from 2D gels using MS based proteomics.	(Meleady et al. 2012a)
Comparative study of CHO SEAP cells over expressing miR-7	Label Free LC-MS/MS profiling	93 proteins decreased in expression levels and 74 increased following over expression of miR-7 in CHO SEAP cells compared the null cells involved in protein folding and secretion.	(Meleady et al. 2012)
Study role of microRNA in regulation of CHO cells growth comparing clones with different growth rates	Multi-Omics qPCR, Microarray and label free quantitative MS	GO analysis on 285 differentially expressed proteins identified biological processes associated with proliferation including mRNA processing and translation	(Clarke et al. 2012)
iTRAQ peptide labelling of a CHO cell line producing a monoclonal antibody	Labelled 2D- LC-MS/MS profiling	Provided a dynamic view of protein expression throughout the fed batch culture providing useful information of the biological processes driving mammalian cell culture.	(Carlage et al. 2012)

Table 1-5. Compiles a list of recent CHO cell proteomic applications using gel-free approaches

Study/Comparison	Technique	Key Finding	Author
CHO cells HCP	2D PAGE	2D PAGE with western blotting can be used for the analysis of HCP content during bioprocessing culturing of CHO cells.	(Hogwood et al. 2014)
CHO HCP levels in a purified drug substance protein P1	2D DIGE and LC-MS/MS	Identified CHO catalase subunit in purified protein P1.	(Ahluwalia et al. 2017)
CHO cells HCP clearance/presence can be achieved by comparing 2D-PAGE	2D PAGE	Characterization of Host Cell Proteins (HCPs) in CHO Cell Bioprocesses	(Hogwood et al. 2017)
Compared HCP profiles using three CHO cell lines and 2D PAGE	2D PAGE and LC-MS/MS identification	More similar than different: Host cell protein production using three null CHO cell lines	(Yuk et al. 2015)
Used 2D-PAGE to visualize antibody species in the eluate and HCPs to screen cell lines and monitor and identify HCPs.	2D PAGE separation and LC-MS/MS identification	The dynamics of the CHO host cell protein profile during clarification and protein A capture in a platform antibody purification process	(Hogwood et al. 2014)
CHO HCP in batch-fed and batch cultures	LC-MS/MS Profiling	They identified 2145 HCPs in the fed-batch culture and 1934 proteins in the batch culture.	(J. H. Park et al. 2017)

A study to understand the effects of GAA overexpression on cell lysosomal phenotype and host cell protein release	Zyograph gels and LC-MS/MS	The overexpression of a lysosomal therapeutic recombinant protein impacts on the lysosomal phenotype of a CHO cell line.	(Migani et al. 2017)
CHO HCP 1hr Profiling	LC-MS/MS DIA method	Detect CHO HCPs at the ng/mg level using 1h analysis	(Walker et al. 2017)
A comparison of CHO HCP subpopulations associated with 15 different mAbs during protein A chromatography	LC-MS/MS	Characterization of the co-elution of host cell proteins with monoclonal antibodies during protein A purification	(Q. Zhang et al. 2016)
An advanced 2D LCS MS/MS method to demonstrate that the HCP enrichment for 2 different mAB products	LC-MS/MS	Development of advanced host cell protein enrichment and detection strategies to enable process relevant spike challenge studies	(Soderquist et al. 2015)
Quantitative definition and monitoring of the HCP proteome using iTRAQ–a study of an industrial mAb producing CHO-S cell line	LC-MS/MS	Used iTRAQ LC-MS/MS to analyze samples from a CHO culture harvest (HCCF). The data provides the basis for application of iTRAQ for process development based upon knowledge of critical HCPs.	(Chiverton et al. 2016)

Table 1-6. Compiles a list of recent CHO cell HCP analysis for proteomic applications

1.5   Conclusion

The application of strict sample preparation techniques, coupled with highly reproducible nano-HPLC methods, state-of-the-art mass spectrometry and data analytics will advance the field of proteomics research generating confident reproducible data.

Managing well, high mass accuracy mass spectrometers will lead to in-depth identification and characterisation of proteins and their modifications along with meticulous data analyses to ensure that all proteomic outputs are of the highest quality.

This will ensure cancer proteomic research can produce reliable data to improve early detection, diagnosis and response to treatment by accurately identifying changes in protein expression during cancer progression, which have the potential to be biomarkers or therapeutic targets.

When applied to biopharmaceutical proteomics research, we can confidently identify proteins involved in desirable recombinant Chinese hamster ovary (CHO) cellular phenotypes which have the potential to be manipulated (e.g. through cell line engineering approaches) to improve the efficiency of biotherapeutic production.

Over the last decade, the proteomic analytic hardware equipment involved in identifying proteins from biological samples has reached a level of sophistication where researchers need and want to identify thousands of confident identifications; however, the bioinformatic interpretation and data processing is struggling to keep up and find a better biological understanding from the data generated.

1.6   References

Abbott, A. (1999). A post-genomic challenge: Learning to read patterns of protein synthesis. Nature, 402(6763), 715.

Adeola, H. A., Calder, B., Soares, N. C., Kaestner, L., Blackburn, J. M., & Zerbini, L. F. (2016). In silico verification and parallel reaction monitoring prevalidation of potential prostate cancer biomarkers. Future Oncology, 12(1), 43-57.

Aebersold, R., & Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. Nature, 537(7620), 347.

Afjehi-Sadat, L., & Garcia, B. A. (2013). Comprehending dynamic protein methylation with mass spectrometry. Current Opinion in Chemical Biology, 17(1), 12-19.

Ahluwalia, D., Dhillon, H., Slaney, T., Song, H., Boux, H., Mehta, S., et al. (2017). Identification of a host cell protein impurity in therapeutic protein, P1. Journal of Pharmaceutical and Biomedical Analysis, 141, 32-38.

Al Shweiki, M. R., Mönchgesang, S., Majovsky, P., Thieme, D., Trutschel, D., & Hoehenwarter, W. (2017). Assessment of label-free quantification in discovery proteomics and impact of technological factors and natural variability of protein abundance. Journal of Proteome Research, 16(4), 1410-1424.

Albrecht, S., Kaisermayer, C., Reinhart, D., Ambrose, M., Kunert, R., Lindeberg, A., et al. (2018). Multiple reaction monitoring targeted LC-MS analysis of potential cell death marker proteins for increased bioprocess control. Analytical and Bioanalytical Chemistry, 410(13), 3197-3207.

Anderson, N. L., & Anderson, N. G. (2002). The human plasma proteome: History, character, and diagnostic prospects. Molecular & Cellular Proteomics: MCP, 1(11), 845-867.

Ardito, F., Giuliani, M., Perrone, D., Troiano, G., & Lo Muzio, L. (2017). The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy. International Journal of Molecular Medicine, 40(2), 271-280.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: Tool for the unification of biology. Nature Genetics, 25(1), 25.

Babelomics - http://www.babelomics.org

Badman, E. R., Patterson, G. E., Wells, J. M., Santini, R. E., & Cooks, R. G. (1999). Differential non-destructive image current detection in a fourier transform quadrupole ion trap. Journal of Mass Spectrometry, 34(8), 889-894.

Baik, J. Y., & Lee, G. M. (2010). A DIGE approach for the assessment of differential expression of the CHO proteome under sodium butyrate addition: Effect of Bcl-xL overexpression. Biotechnology and Bioengineering, 105(2), 358-367.

Bailey, C. M., Sweet, S. M., Cunningham, D. L., Zeller, M., Heath, J. K., & Cooper, H. J. (2009). SLoMo: Automated site localization of modifications from ETD/ECD mass spectra. Journal of Proteome Research, 8(4), 1965-1971.

Bandaranayake, A. D., & Almo, S. C. (2014). Recent advances in mammalian protein production. FEBS Letters, 588(2), 253-260.

Baycin-Hizal, D., Tabb, D. L., Chaerkady, R., Chen, L., Lewis, N. E., Nagarajan, H., et al. (2012). Proteomic analysis of chinese hamster ovary cells. Journal of Proteome Research, 11(11), 5265-5276.

Beausoleil, S. A., Villén, J., Gerber, S. A., Rush, J., & Gygi, S. P. (2006). A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nature Biotechnology, 24(10), 1285.

Becker, J., Hackl, M., Rupp, O., Jakobi, T., Schneider, J., Szczepanowski, R., et al. (2011). Unraveling the chinese hamster ovary cell line transcriptome by next-generation sequencing. Journal of Biotechnology, 156(3), 227-235.

Bielefeld-Boku - ftp://ftp.cebitec.unibielfeld.de/pub/supplements/BB-CHO/

BioDBnet - http://biodbnet-abcc.ncifcrf.gov/

Blondeel, E. J., Ho, R., Schulze, S., Sokolenko, S., Guillemette, S. R., Slivac, I., et al. (2016). An omics approach to rational feed: Enhancing growth in CHO cultures with NMR metabolomics and 2D-DIGE proteomics. Journal of Biotechnology, 234, 127-138.

Bondarenko, P. V., Chelius, D., & Shaler, T. A. (2002). Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography– tandem mass spectrometry. Analytical Chemistry, 74(18), 4741-4749.

Bracewell, D. G., Francis, R., & Smales, C. M. (2015). The future of host cell protein (HCP) identification during process development and manufacturing linked to a risk-based management for their control. Biotechnology and Bioengineering, 112(9), 1727-1737.

Bruce, A., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2007). Molecular biology of the cell 5th edn (new york: Garland science).

Bruderer, T., Varesio, E., & Hopfgartner, G. (2017). The use of LC predicted retention times to extend metabolites identification with SWATH data acquisition. Journal of Chromatography B, 1071, 3-10.

Bunai, K., & Yamane, K. (2005). Effectiveness and limitation of two-dimensional gel electrophoresis in bacterial membrane protein proteomics and perspectives. Journal of Chromatography B, 815(1-2), 227-236.

BURNETT, G., & KENNEDY, E. P. (1954). The enzymatic phosphorylation of proteins. The Journal of Biological Chemistry, 211(2), 969-980.

Campbell, D. I., Ferreira, C. R., Eberlin, L. S., & Cooks, R. G. (2012). Improved spatial resolution in the imaging of biological tissue using desorption electrospray ionization. Analytical and Bioanalytical Chemistry, 404(2), 389-398.

Cao, T. H., Quinn, P. A., Sandhu, J. K., Voors, A. A., Lang, C. C., Parry, H. M., et al. (2015). Identification of novel biomarkers in plasma for prediction of treatment response in patients with heart failure. The Lancet, 385, S26.

Carlage, T., Kshirsagar, R., Zang, L., Janakiraman, V., Hincapie, M., Lyubarskaya, Y., et al. (2012). Analysis of dynamic changes in the proteome of a Bcl-XL overexpressing chinese hamster ovary cell culture during exponential and stationary phases. Biotechnology Progress, 28(3), 814-823.

Catenacci, D. V., Liao, W., Thyparambil, S., Henderson, L., Xu, P., Zhao, L., et al. (2014). Absolute quantitation of met using mass spectrometry for clinical application: Assay precision, stability, and correlation with MET gene amplification in FFPE tumor tissue. PloS One, 9(7), e100586.

Chalkley, R. J., & Clauser, K. R. (2012). Modification site localization scoring: Strategies and performance. Molecular & Cellular Proteomics: MCP, 11(5), 3-14. doi:10.1074/mcp.R111.015305 [doi]

Chemdata - https://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:cdownload

Chiverton, L. M., Evans, C., Pandhal, J., Landels, A. R., Rees, B. J., Levison, P. R., et al. (2016). Quantitative definition and monitoring of the host cell protein proteome using iTRAQ–a study of an industrial mAb producing CHO-S cell line. Biotechnology Journal, 11(8), 1014-1024.

Choe, L. H., Aggarwal, K., Franck, Z., & Lee, K. H. (2005). A comparison of the consistency of proteome quantitation using two-dimensional electrophoresis and shotgun isobaric tagging in escherichia coli cells. Electrophoresis, 26(12), 2437-2449.

Chong, W. P. K., Thng, S. H., Hiu, A. P., Lee, D., Chan, E. C. Y., & Ho, Y. S. (2012). LC-MS-based metabolic characterization of high monoclonal antibody-producing chinese hamster ovary cells. Biotechnology and Bioengineering, 109(12), 3103-3111.

Christoforou, A., Mulvey, C. M., Breckels, L. M., Geladaki, A., Hurrell, T., Hayward, P. C., et al. (2016). A draft map of the mouse pluripotent stem cell spatial proteome. Nature Communications, 7, 8992. doi:10.1038/ncomms9992 [doi]

Chutipongtanate, S., Chatchen, S., & Svasti, J. (2017). Plasma prefractionation methods for proteomic analysis and perspectives in clinical applications. PROTEOMICS-Clinical Applications,

Clabaut, A., Grare, C., Léger, T., Hardouin, P., & Broux, O. (2015). Variations of secretome profiles according to conditioned medium preparation: The example of human mesenchymal stem cell-derived adipocytes. Electrophoresis, 36(20), 2587-2593.

Clarke, C., Henry, M., Doolan, P., Kelly, S., Aherne, S., Sanchez, N., et al. (2012). Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. BMC Genomics, 13(1), 656.

Coleman, O., Henry, M., Clynes, M., & Meleady, P. (2017). Filter-aided sample preparation (FASP) for improved proteome analysis of recombinant chinese hamster ovary cells. Heterologous protein production in CHO cells (pp. 187-194) Springer.

Collins, M. O., Wright, J. C., Jones, M., Rayner, J. C., & Choudhary, J. S. (2014). Confident and sensitive phosphoproteomics using combinations of collision induced dissociation and electron transfer dissociation. Journal of Proteomics, 103, 1-14.

Cottrell, J. S. (2011). Protein identification using MS/MS data. Journal of Proteomics, 74(10), 1842-1851.

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., & Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Molecular & Cellular Proteomics: MCP, 13(9), 2513-2526. doi:10.1074/mcp.M113.031591 [doi]

DAVID - https://david.ncifcrf.gov/

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., et al. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proceedings of the National Academy of Sciences of the United States of America, 100(13), 7480-7485. doi:10.1073/pnas.1332608100 [doi]

Desiderio, D. M., & Kai, M. (1983). Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. Biological Mass Spectrometry, 10(8), 471-479.

Domon, B., & Aebersold, R. (2006). Mass spectrometry and protein analysis. Science (New York, N.Y.), 312(5771), 212-217. doi:312/5771/212 [pii]

Doolan, P., Meleady, P., Barron, N., Henry, M., Gallagher, R., Gammell, P., et al. (2010). Microarray and proteomics expression profiling identifies several candidates, including the valosin-containing protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines. Biotechnology and Bioengineering, 106(1), 42-56.

Dowling, P., Henry, M., Meleady, P., Clarke, C., Gately, K., O'Byrne, K., et al. (2015). Metabolomic and proteomic analysis of breast cancer patient samples suggests that glutamate and 12-HETE in combination with CA15-3 may be useful biomarkers reflecting tumour burden. Metabolomics, 11(3), 620-635.

Droge, W. (2002). Free radicals in the physiological control of cell function. Physiological Reviews, 82(1), 47-95.

Dunkley, T. P., Watson, R., Griffin, J. L., Dupree, P., & Lilley, K. S. (2004). Localization of organelle proteins by isotope tagging (LOPIT). Molecular & Cellular Proteomics: MCP, 3(11), 1128-1134. doi:10.1074/mcp.T400009-MCP200 [doi]

Edelson-Averbukh, M., Shevchenko, A., Pipkorn, R., & Lehmann, W. D. (2011). Discrimination between peptide O-sulfo-and O-phosphotyrosine residues by negative ion mode electrospray tandem mass spectrometry. Journal of the American Society for Mass Spectrometry, 22(12), 2256-2268.

Eng, J. K., Fischer, B., Grossmann, J., & MacCoss, M. J. (2008). A fast SEQUEST cross correlation algorithm. Journal of Proteome Research, 7(10), 4598-4602.

Fairhead, M., & Howarth, M. (2015). Site-specific biotinylation of purified proteins using BirA. Site-specific protein labeling (pp. 171-184) Springer.

Fila, J., & Honys, D. (2012). Enrichment techniques employed in phosphoproteomics. Amino Acids, 43(3), 1025-1047. doi:10.1007/s00726-011-1111-z [doi]

Foster, L. J., de Hoog, C. L., Zhang, Y., Zhang, Y., Xie, X., Mootha, V. K., et al. (2006). A mammalian organelle map by protein correlation profiling. Cell, 125(1), 187-199.

Gallien, S., Duriez, E., & Domon, B. (2011). Selected reaction monitoring applied to proteomics. Journal of Mass Spectrometry, 46(3), 298-312.

Gallien, S., Kim, S. Y., & Domon, B. (2015). Large-scale targeted proteomics using internal standard triggered-parallel reaction monitoring (IS-PRM). Molecular & Cellular Proteomics: MCP, 14(6), 1630-1644. doi:10.1074/mcp.O114.043968 [doi]

Garner, B., Witting, P. K., Waldeck, A. R., Christison, J. K., Raftery, M., & Stocker, R. (1998). Oxidation of high density lipoproteins. I. formation of methionine sulfoxide in apolipoproteins AI and AII is an early event that accompanies lipid peroxidation and can be enhanced by alpha-tocopherol. The Journal of Biological Chemistry, 273(11), 6080-6087.

Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., & Gygi, S. P. (2003). 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(12), 6940-6945. doi:10.1073/pnas.0832254100 [doi]

Geyer, P. E., Kulak, N. A., Pichler, G., Holdt, L. M., Teupser, D., & Mann, M. (2016). Plasma proteome profiling to assess human health and disease. Cell Systems, 2(3), 185-195.

Giansanti, P., Tsiatsiani, L., Low, T. Y., & Heck, A. J. (2016). Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. Nature Protocols, 11(5), 993.

Glatter, T., Ludwig, C., Ahrné, E., Aebersold, R., Heck, A. J., & Schmidt, A. (2012). Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem lys-C/trypsin proteolysis over trypsin digestion. Journal of Proteome Research, 11(11), 5145-5156.

GoCHO - http://ebdrub.biosustain.dtu.dk/gocho

GO - http://www.geneontology.org

Gonzalez-Fernandez, R., Prats, E., & Jorrin-Novo, J. V. (2010). Proteomics of plant pathogenic fungi. Journal of Biomedicine & Biotechnology, 2010, 932527. doi:10.1155/2010/932527 [doi]

Graves, P. R., & Haystead, T. A. (2002). Molecular biologist's guide to proteomics. Microbiology and Molecular Biology Reviews: MMBR, 66(1), 39-63; table of contents.

Güzel, C., Govorukhina, N. I., Wisman, G. B. A., Stingl, C., Dekker, L. J., Klip, H. G., et al. (2018). Proteomic alterations in early stage cervical cancer. Oncotarget, 9(26), 18128.

Häupl, B., Ihling, C. H., & Sinz, A. (2016). Protein interaction network of human protein kinase D2 revealed by chemical cross-linking/mass spectrometry. Journal of Proteome Research, 15(10), 3686-3699.

Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphall, M. S., et al. (2014). The one hour yeast proteome. Molecular & Cellular Proteomics: MCP, 13(1), 339-347. doi:10.1074/mcp.M113.034769 [doi]

Heffner, K. M., Hizal, D. B., Yerganian, G. S., Kumar, A., Can, O., O'Meally, R., et al. (2017). Lessons from the hamster: Cricetulus griseus tissue and CHO cell line proteome comparison. Journal of Proteome Research, 16(10), 3672-3687.

Heibeck, T. H., Ding, S., Opresko, L. K., Zhao, R., Schepmoes, A. A., Yang, F., et al. (2009). An extensive survey of tyrosine phosphorylation revealing new sites in human mammary epithelial cells. Journal of Proteome Research, 8(8), 3852-3861.

Hellstrom, I., & Hellstrom, K. E. (2011). Two new biomarkers, mesothelin and HE4, for diagnosis of ovarian carcinoma. Expert Opinion on Medical Diagnostics, 5(3), 227-240.

Henry, M., Coleman, O., Clynes, M., & Meleady, P. (2017). Phosphopeptide enrichment and LC-MS/MS analysis to study the phosphoproteome of recombinant chinese hamster ovary cells. Heterologous Protein Production in CHO Cells: Methods and Protocols, , 195-208.

Henry, M., Gallagher, C., Kelly, R. M., Frye, C. C., Osborne, M. D., Brady, C. P., et al. (2018). Clonal variation in productivity and proteolytic clipping of an fc-fusion protein in CHO cells: Proteomic analysis suggests a role for defective protein folding and the UPR. Journal of Biotechnology,

Henry, M., & Meleady, P. (2017). Clinical proteomics: Liquid Chromatography–Mass spectrometry (LC–MS) purification systems. Protein chromatography (pp. 375-388) Springer.

Henry, M., Power, M., Kaushik, P., Coleman, O., Clynes, M., & Meleady, P. (2017). Differential phosphoproteomic analysis of recombinant chinese hamster ovary cells following temperature shift. Journal of Proteome Research,

Hershko, A. (1996). Lessons from the discovery ofthe ubiquitin system. Trends in Biochemical Sciences, 21(11), 445-449.

Hinkson, I. V., & Elias, J. E. (2011). The dynamic state of protein turnover: It's about time. Trends in Cell Biology, 21(5), 293-303.

Hjerpe, R., Aillet, F., Lopitz-Otsoa, F., Lang, V., England, P., & Rodriguez, M. S. (2009). Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities. EMBO Reports, 10(11), 1250-1258. doi:10.1038/embor.2009.192 [doi]

Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S., et al. (2002). Systematic identification of protein complexes in saccharomyces cerevisiae by mass spectrometry. Nature, 415(6868), 180.

Hogwood, C. E., Bracewell, D. G., & Smales, C. M. (2014). Measurement and control of host cell proteins (HCPs) in CHO cell bioprocesses. Current Opinion in Biotechnology, 30, 153-160.

Hogwood, C. E., Chiverton, L. M., & Mark Smales, C. (2017). Characterization of host cell proteins (HCPs) in CHO cell bioprocesses. Heterologous Protein Production in CHO Cells: Methods and Protocols, , 243-250.

Hsiao, H. H., Meulmeester, E., Frank, B. T., Melchior, F., & Urlaub, H. (2009). "ChopNSpice," a mass spectrometric approach that allows identification of endogenous small ubiquitin-like modifier-conjugated peptides. Molecular & Cellular Proteomics: MCP, 8(12), 2664-2675. doi:10.1074/mcp.M900087-MCP200 [doi]

Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., & Graham Cooks, R. (2005). The orbitrap: A new mass spectrometer. Journal of Mass Spectrometry, 40(4), 430-443.

Hughes, C. S., Spicer, V., Krokhin, O. V., & Morin, G. B. (2017). Investigating acquisition performance on the orbitrap fusion when using tandem MS/MS/MS scanning with isobaric tags. Journal of Proteome Research, 16(5), 1839-1846.

Hunt, D. F., Shabanowitz, J., McIver, R. T., Hunter, R. L., & Syka, J. E. (1985). Ionization and mass analysis of nonvolatile compounds by particle bombardment-quadrupole-fourier transform mass spectrometry. Analytical Chemistry, 57(3), 765-768.

Hunter, T. (1995). Protein kinases and phosphatases: The yin and yang of protein phosphorylation and signaling. Cell, 80(2), 225-236.

Ingenuity pathway analysis - https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis

Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., et al. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science (New York, N.Y.), 292(5518), 929-934. doi:10.1126/science.292.5518.929 [doi]

Jayapal, K. P., Sui, S., Philp, R. J., Kok, Y., Yap, M. G., Griffin, T. J., et al. (2010). Multitagging proteomic strategy to estimate protein turnover rates in dynamic systems. Journal of Proteome Research, 9(5), 2087-2097.

Jedrychowski, M. P., Huttlin, E. L., Haas, W., Sowa, M. E., Rad, R., & Gygi, S. P. (2011). Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. Molecular & Cellular Proteomics: MCP, 10(12), M111.009910. doi:10.1074/mcp.M111.009910 [doi]

Jiao, X., Sherman, B. T., Huang, D. W., Stephens, R., Baseler, M. W., Lane, H. C., et al. (2012). DAVID-WS: A stateful web service to facilitate gene/protein list analysis. Bioinformatics, 28(13), 1805-1806.

Kalli, A., Smith, G. T., Sweredoski, M. J., & Hess, S. (2013). Evaluation and optimization of mass spectrometric settings during data-dependent acquisition mode: Focus on LTQ-orbitrap mass analyzers. Journal of Proteome Research, 12(7), 3071-3086.

Kang, S., Xiao, G., Ren, D., Zhang, Z., Le, N., Trentalange, M., et al. (2014). Proteomics analysis of altered cellular metabolism induced by insufficient copper level. Journal of Biotechnology, 189, 15-26.

Kaufmann, A., & Walker, S. (2017). Comparison of linear intrascan and interscan dynamic ranges of orbitrap and ion-mobility time-of-flight mass spectrometers. Rapid Communications in Mass Spectrometry, 31(22), 1915-1926.

KEGG - https://www.genome.jp/kegg/

Kehoe, J. W., & Bertozzi, C. R. (2000). Tyrosine sulfation: A modulator of extracellular protein–protein interactions. Chemistry & Biology, 7(3), R57-R61.

Kelleher, N. L., Lin, H. Y., Valaskovic, G. A., Aaserud, D. J., Fridriksson, E. K., & McLafferty, F. W. (1999). Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. Journal of the American Chemical Society, 121(4), 806-812.

Kelly, P. S., Breen, L., Gallagher, C., Kelly, S., Henry, M., Lao, N. T., et al. (2015). Re-programming CHO cell metabolism using miR-23 tips the balance towards a highly productive phenotype. Biotechnology Journal, 10(7), 1029-1040.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The human genome browser at UCSC. Genome Research, 12(6), 996-1006. doi:10.1101/gr.229102 [doi]

Keshishian, H., Burgess, M. W., Gillette, M. A., Mertins, P., Clauser, K. R., Mani, D. R., et al. (2015). Multiplexed, quantitative workflow for sensitive biomarker discovery in plasma yields novel candidates for early myocardial injury. Molecular & Cellular Proteomics: MCP, 14(9), 2375-2393. doi:10.1074/mcp.M114.046813 [doi]

Kieselbach, T., & Oscarsson, J. (2017). Dataset of the proteome of purified outer membrane vesicles from the human pathogen aggregatibacter actinomycetemcomintans. Data in Brief, 10, 426-431.

Kim, B., Araujo, R., Howard, M., Magni, R., Liotta, L., & Luchini, A. (2018). Affinity enrichment for mass spectrometry: Improving the yield of low abundance biomarkers. Expert Review of Proteomics, (just-accepted)

Kim, J. Y., Kim, Y., & Lee, G. M. (2012). CHO cells in biotechnology for production of recombinant proteins: Current state and further potential. Applied Microbiology and Biotechnology, 93(3), 917-930.

Kim, M., & Pandey, A. (2012). Electron transfer dissociation mass spectrometry in proteomics. Proteomics, 12(4-5), 530-542.

Knuesel, M., Cheung, H. T., Hamady, M., Barthel, K. K., & Liu, X. (2005). A method of mapping protein sumoylation sites by mass spectrometry using a modified small ubiquitin-like modifier 1 (SUMO-1) and a computational program. Molecular & Cellular Proteomic : MCP, 4(10), 1626-1636. doi:T500011-MCP200 [pii]

Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., & Nesvizhskii, A. I. (2017). MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. Nature Methods, 14(5), 513.

Koopmans, F., Ho, J. T., Smit, A. B., & Li, K. W. (2018). Comparative analyses of data independent acquisition mass spectrometric approaches: DIA, WiSIM-DIA, and untargeted DIA. Proteomics, 18(1), 1700304.

Kumar, A., Baycin-Hizal, D., Wolozny, D., Pedersen, L. E., Lewis, N. E., Heffner, K., et al. (2015). Elucidation of the CHO super-ome (CHO-SO) by proteoinformatics. Journal of Proteome Research, 14(11), 4687-4703.

Kusmierz, J. J., Sumrada, R., & Desiderio, D. M. (1990). Fast atom bombardment mass spectrometric quantitative analysis of methionine-enkephalin in human pituitary tissues. Analytical Chemistry, 62(21), 2395-2400.

LaCava, J., Molloy, K. R., Taylor, M. S., Domanski, M., Chait, B. T., & Rout, M. P. (2015). Affinity proteomics to study endogenous protein complexes: Pointers, pitfalls, preferences and perspectives. Biotechniques, 58(3), 103-119. doi:10.2144/000114262 [doi]

LASKOWSKI, M., & WU, F. C. (1953). Temporary inhibition of trypsin. The Journal of Biological Chemistry, 204(2), 797-805.

Leitner, A. (2010). Phosphopeptide enrichment using metal oxide affinity chromatography. TrAC Trends in Analytical Chemistry, 29(2), 177-185.

Li, L. (2015). Dynamic range compression with ProteoMiner™: Principles and examples. Proteomic profiling (pp. 99-107) Springer.

Liu, J., Chi, Y., Jiang, G., Tai, C., Peng, J., & Hu, J. (2004). Ionic liquid-based liquid-phase microextraction, a new sample enrichment procedure for liquid chromatography. Journal of Chromatography A, 1026(1-2), 143-147.

Liu, F., Lossl, P., Rabbitts, B. M., Balaban, R. S., & Heck, A. J. R. (2018). The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. Molecular & Cellular Proteomics : MCP, 17(2), 216-232. doi:10.1074/mcp.RA117.000470 [doi]

Mädler, S., Bich, C., Touboul, D., & Zenobi, R. (2009). Chemical cross-linking with NHS esters: A systematic study on amino acid reactivities. Journal of Mass Spectrometry, 44(5), 694-706.

Makarov, A., Denisov, E., Lange, O., & Horning, S. (2006). Dynamic range of mass accuracy in LTQ orbitrap hybrid mass spectrometer. Journal of the American Society for Mass Spectrometry, 17(7), 977-982.

Malafaia, C. B., Guerra, M. L., Silva, T. D., Paiva, P. M., Souza, E. B., Correia, M. T., et al. (2015). Selection of a protein solubilization method suitable for phytopathogenic bacteria: A proteomics approach. Proteome Science, 13(1), 5.

Malik, R., Dulla, K., Nigg, E. A., & Körner, R. (2010). From proteome lists to biological impact–tools and strategies for the analysis of large MS data sets. Proteomics, 10(6), 1270-1283.

Manjasetty, B. A., Büssow, K., Panjikar, S., & Turnbull, A. P. (2012). Current methods in structural proteomics and its applications in biological sciences. 3 Biotech, 2(2), 89-113.

Martin, S. E., Shabanowitz, J., Hunt, D. F., & Marto, J. A. (2000). Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. Analytical Chemistry, 72(18), 4266-4274.

Martinet, N., & Bertrand, P. (2011). Interpreting clinical assays for histone deacetylase inhibitors. Cancer Management and Research, 3, 117-141. doi:10.2147/CMR.S9661 [doi]

Matsumoto, M., Hatakeyama, S., Oyamada, K., Oda, Y., Nishimura, T., & Nakayama, K. I. (2005). Large-scale analysis of the human ubiquitin-related proteome. Proteomics, 5(16), 4145-4151.

McAlister, G., Jedrychowski, M., Yu, Y., Ting, L., Huttlin, E., Rad, R., et al. (2012). Isolating multiple MS2 fragments using waveforms with multiple frequency notches improves MS3 sensitivity~ 8 fold over standard MS3-based TMT methods. Paper presented at the 60th ASMS Conference on Mass Spectrometry and Allied Topics, Vancouver, Canada,

McAlister, G. C., Berggren, W. T., Griep-Raming, J., Horning, S., Makarov, A., Phanstiel, D., et al. (2008). A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. Journal of Proteome Research, 7(8), 3127-3136.

Medina, I., Carbonell, J. ½., Pulido, L., Madeira, S. C., Goetz, S., Conesa, A., et al. (2010). Babelomics: An integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. Nucleic Acids Research, 38(suppl_2), W210-W213.

MASCOT - www.matrixscience.com

Meierhofer, D., Wang, X., Huang, L., & Kaiser, P. (2008). Quantitative analysis of global ubiquitination in HeLa cells by mass spectrometry. Journal of Proteome Research, 7(10), 4566-4576.

Meleady, P. (2018). Two-dimensional gel electrophoresis and 2D-DIGE. Difference gel electrophoresis (pp. 3-14) Springer.

Meleady, P., Doolan, P., Henry, M., Barron, N., Keenan, J., O'Sullivan, F., et al. (2011). Sustained productivity in recombinant chinese hamster ovary (CHO) cell lines: Proteome analysis of the molecular basis for a process-related phenotype. BMC Biotechnology, 11(1), 78.

Meleady, P., Gallagher, M., Clarke, C., Henry, M., Sanchez, N., Barron, N., et al. (2012). Impact of miR-7 over-expression on the proteome of chinese hamster ovary cells. Journal of Biotechnology, 160(3-4), 251-262.

Meleady, P., Hoffrogge, R., Henry, M., Rupp, O., Bort, J. H., Clarke, C., et al. (2012). Utilization and evaluation of CHO-specific sequence databases for mass spectrometry based proteomics. Biotechnology and Bioengineering, 109(6), 1386-1394.

Micallef, J., Dharsee, M., Chen, J., Ackloo, S., Evans, K., Qiu, L., et al. (2010). Applying mass spectrometry based proteomic technology to advance the understanding of multiple myeloma. Journal of Hematology & Oncology, 3(1), 13.

Migani, D., Smales, C. M., & Bracewell, D. G. (2017). Effects of lysosomal biotherapeutic recombinant protein expression on cell stress and protease and general host cell protein release in chinese hamster ovary cells. Biotechnology Progress, 33(3), 666-676.

Mikesh, L. M., Ueberheide, B., Chi, A., Coon, J. J., Syka, J. E., Shabanowitz, J., et al. (2006). The utility of ETD mass spectrometry in proteomic analysis. Biochimica Et Biophysica Acta (BBA)-Proteins and Proteomics, 1764(12), 1811-1822.

Moore, K. L. (2003). The biology and enzymology of protein tyrosine O-sulfation. The Journal of Biological Chemistry, 278(27), 24243-24246. doi:10.1074/jbc.R300008200 [doi]

Morelle, W., & Michalski, J. (2007). Analysis of protein glycosylation by mass spectrometry. Nature Protocols, 2(7), 1585.

Morrow, J. K., Lin, H., Sun, S., & Zhang, S. (2015). Targeting ubiquitination for cancer therapies. Future Medicinal Chemistry, 7(17), 2333-2350.

Mulvey, C. M., Breckels, L. M., Geladaki, A., Britovšek, N. K., Nightingale, D. J., Christoforou, A., et al. (2017). Using hyperLOPIT to perform high-resolution mapping of the spatial proteome. Nature Protocols, 12(6), 1110.

NCBI - ftp://ftp.ncbi.nlm.nih.gov/blast/db/fasta

Nishi, H., Shaytan, A., & Panchenko, A. R. (2014). Physicochemical mechanisms of protein regulation by phosphorylation. Frontiers in Genetics, 5, 270.

O'Sullivan, D., Henry, M., Joyce, H., Walsh, N., Mc Auley, E., Dowling, P., et al. (2014). 7B7: A novel antibody directed against the Ku70/Ku80 heterodimer blocks invasion in pancreatic and lung cancer cells. Tumor Biology, 35(7), 6983-6997.

O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. The Journal of Biological Chemistry, 250(10), 4007-4021.

Osula, O., Swatkoski, S., & Cotter, R. J. (2012). Identification of protein SUMOylation sites by mass spectrometry using combined microwave-assisted aspartic acid cleavage and tryptic digestion. Journal of Mass Spectrometry, 47(5), 644-654.

Pan, Y., & Konermann, L. (2010). Membrane protein structural insights from chemical labeling and mass spectrometry. Analyst, 135(6), 1191-1200.

Pandey, A., Fernandez, M. M., Steen, H., Blagoev, B., Nielsen, M. M., Roche, S., et al. (2000). Identification of a novel immunoreceptor tyrosine-based activation motif-containing molecule, STAM2, by mass spectrometry and its involvement in growth factor and cytokine receptor signaling pathways. The Journal of Biological Chemistry, 275(49), 38633-38639. doi:10.1074/jbc.M007849200 [doi]

PANTHER - (http://www.pantherdb.org/)

Park, J. H., Jin, J. H., Ji, I. J., An, H. J., Kim, J. W., & Lee, G. M. (2017). Proteomic analysis of host cell protein dynamics in the supernatant of Fc-Fusion Protein-Producing CHO DG44 and DUKX-B11 cell lines in batch and Fed-Batch cultures. Biotechnology and Bioengineering,

Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C., Cheng, D., Marsischky, G., et al. (2003). A proteomics approach to understanding protein ubiquitination. Nature Biotechnology, 21(8), 921.

Pernemalm, M., Orre, L. M., Lengqvist, J., Wikström, P., Lewensohn, R., & Lehtiö, J. (2008). Evaluation of three principally different intact protein prefractionation methods for plasma biomarker discovery. Journal of Proteome Research, 7(7), 2712-2722.

Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., & Coon, J. J. (2012). Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. Molecular & Cellular Proteomics : MCP, 11(11), 1475-1488. doi:10.1074/mcp.O112.020131 [doi]

Pichler, P., Köcher, T., Holzmann, J., Mazanek, M., Taus, T., Ammerer, G., et al. (2010). Peptide labeling with isobaric tags yields higher identification rates using iTRAQ 4-plex compared to TMT 6-plex and iTRAQ 8-plex on LTQ orbitrap. Analytical Chemistry, 82(15), 6549-6558.

Picotti, P., Clément-Ziza, M., Lam, H., Campbell, D. S., Schmidt, A., Deutsch, E. W., et al. (2013). A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. Nature, 494(7436), 266.

PUCK, T. T., CIECIURA, S. J., & FISHER, H. W. (1957). Clonal growth in vitro of human cells with fibroblastic morphology; comparison of growth and genetic characteristics of single epithelioid and fibroblast-like cells from a variety of human organs. The Journal of Experimental Medicine, 106(1), 145-158.

Rauniyar, N. (2015). Parallel reaction monitoring: A targeted experiment performed using high resolution and high mass accuracy mass spectrometry. International Journal of Molecular Sciences, 16(12), 28566-28581.

Rauniyar, N., Yu, X., Cantley, J., Voss, E. Z., Belcher, J., Colangelo, C. M., et al. (2018). Quantification of urinary protein biomarkers of autosomal dominant polycystic kidney disease by parallel reaction monitoring. Proteomics.Clinical Applications, , e1700157. doi:10.1002/prca.201700157 [doi]

Ritz, D., Kinzi, J., Neri, D., & Fugmann, T. (2017). Data-Independent acquisition of HLA class I peptidomes on the Q exactive mass spectrometer platform. Proteomics, 17(19)

Roepstorff, P., & Fohlman, J. (1984). Letter to the editors. Biological Mass Spectrometry, 11(11), 601-601.

Roskoski Jr, R. (2012). ERK1/2 MAP kinases: Structure, function, and regulation. Pharmacological Research, 66(2), 105-143.

Sandow, J. J., Rainczuk, A., Infusini, G., Makanji, M., Bilandzic, M., Wilson, A. L., et al. (2018). Discovery and validation of novel protein biomarkers in ovarian cancer patient urine. PROTEOMICS–Clinical Applications, 12(3), 1700135.

Savitski, M. M., Reinhard, F. B., Franken, H., Werner, T., Savitski, M. F., Eberhard, D., et al. (2014). Tracking cancer drugs in living cells by thermal profiling of the proteome. Science (New York, N.Y.), 346(6205), 1255784. doi:10.1126/science.1255784 [doi]

Schaab, C., Geiger, T., Stoehr, G., Cox, J., & Mann, M. (2012). Analysis of high accuracy, quantitative proteomics data in the MaxQB database. Molecular & Cellular Proteomics : MCP, 11(3), M111.014068. doi:10.1074/mcp.M111.014068 [doi]

Schubert, O. T., Gillet, L. C., Collins, B. C., Navarro, P., Rosenberger, G., Wolski, W. E., et al. (2015). Building high-quality assay libraries for targeted analysis of SWATH MS data. Nature Protocols, 10(3), 426.

Scigelova, M., Hornshaw, M., Giannakopulos, A., & Makarov, A. (2011). Fourier transform mass spectrometry. Molecular & Cellular Proteomics : MCP, 10(7), M111.009431. doi:10.1074/mcp.M111.009431 [doi]

Seong, Y., Yoo, Y. S., Akter, H., & Kang, M. (2017). Sample preparation for detection of low abundance proteins in human plasma using ultra-high performance liquid chromatography coupled with highly accurate mass spectrometry. Journal of Chromatography B, 1060, 272-280.

Shafagati, N., Lundberg, L., Baer, A., Patanarut, A., Fite, K., Lepene, B., et al. (2015). The use of nanotrap particles in the enhanced detection of rift valley fever virus nucleoprotein. PloS One, 10(5), e0128215.

Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., et al. (1996). Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. Proceedings of the National Academy of Sciences of the United States of America, 93(25), 14440-14445.

Shi, T., Song, E., Nie, S., Rodland, K. D., Liu, T., Qian, W., et al. (2016). Advances in targeted proteomics and applications to biomedical research. Proteomics, 16(15-16), 2160-2182.

Sidoli, S., Cheng, L., & Jensen, O. N. (2012). Proteomics in chromatin biology and epigenetics: Elucidation of post-translational modifications of histone proteins by mass spectrometry. Journal of Proteomics, 75(12), 3419-3433.

Smith, L. M., Kelleher, N. L., Linial, M., Goodlett, D., Langridge-Smith, P., Goo, Y. A., et al. (2013). Proteoform: A single term describing protein complexity. Nature Methods, 10(3), 186.

Smits, A. H., & Vermeulen, M. (2016). Characterizing protein–protein interactions using mass spectrometry: Challenges and opportunities. Trends in Biotechnology, 34(10), 825-834.

Smolders, K., Lombaert, N., Valkenborg, D., Baggerman, G., & Arckens, L. (2015). An effective plasma membrane proteomics approach for small tissue samples. Scientific Reports, 5, 10917.

Soderquist, R. G., Trumbo, M., Hart, R. A., Zhang, Q., & Flynn, G. C. (2015). Development of advanced host cell protein enrichment and detection strategies to enable process relevant spike challenge studies. Biotechnology Progress, 31(4), 983-989.

Sommeregger, W., Mayrhofer, P., Steinfellner, W., Reinhart, D., Henry, M., Clynes, M., et al. (2016). Proteomic differences in recombinant CHO cells producing two similar antibody fragments. Biotechnology and Bioengineering, 113(9), 1902-1912.

Stokes, M. P., Farnsworth, C. L., Gu, H., Jia, X., Worsfold, C. R., Yang, V., et al. (2015). Complementary PTM profiling of drug response in human gastric carcinoma by immunoaffinity and IMAC methods with total proteome analysis. Proteomes, 3(3), 160-183.

STRING - https://string-db.org/

Syka, J. E., Marto, J. A., Bai, D. L., Horning, S., Senko, M. W., Schwartz, J. C., et al. (2004). Novel linear quadrupole ion trap/FT mass spectrometer: Performance characterization and use in the comparative analysis of histone H3 post-translational modifications. Journal of Proteome Research, 3(3), 621-626.

Tabb, D. L. (2015). The SEQUEST family tree. Journal of the American Society for Mass Spectrometry, 26(11), 1814-1819.

Tamura, T., Terada, T., & Tanaka, A. (2003). A quantitative analysis and chemical approach for the reduction of nonspecific binding proteins on affinity resins. Bioconjugate Chemistry, 14(6), 1222-1230.

Tannu, N. S., & Hemby, S. E. (2006). Two-dimensional fluorescence difference gel electrophoresis for comparative proteomics profiling. Nature Protocols, 1(4), 1732.

Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., et al. (2011). Universal and confident phosphorylation site localization using phosphoRS. Journal of Proteome Research, 10(12), 5354-5362.

ten Have, S., Boulon, S., Ahmad, Y., & Lamond, A. I. (2011). Mass spectrometry-based immuno-precipitation proteomics–The user's guide. Proteomics, 11(6), 1153-1159.

Torres, M. P., Dewhurst, H., & Sundararaman, N. (2016). Proteome-wide structural analysis of PTM hotspots reveals regulatory elements predicted to impact biological function and disease. Molecular & Cellular Proteomics : MCP, 15(11), 3513-3528. doi:M116.062331 [pii]

Trelle, M. B., & Jensen, O. N. (2008). Utility of immonium ions for assignment of ε-N-acetyllysine-containing peptides by tandem mass spectrometry. Analytical Chemistry, 80(9), 3422-3430.

Tsai, P., & Chen, S. (2017). A brief review of bioinformatics tools for glycosylation analysis by mass spectrometry. Mass Spectrometry, 6(2), S0064-S0064.

Tsiatsiani, L., & Heck, A. J. (2015). Proteomics beyond trypsin. The FEBS Journal, 282(14), 2612-2626.

Umar, A., Dalebout, J. C., Timmermans, A. M., Foekens, J. A., & Luider, T. M. (2005). Method optimisation for peptide profiling of microdissected breast carcinoma tissue by matrix-assisted laser desorption/ionisation-time of flight and matrix-assisted laser desorption/ionisation-time of flight/time of flight-mass spectrometry. Proteomics, 5(10), 2680-2688.

Valente, K. N., Lenhoff, A. M., & Lee, K. H. (2015). Expression of difficult-to-remove host cell protein impurities during extended chinese hamster ovary cell culture and their impact on continuous bioprocessing. Biotechnology and Bioengineering, 112(6), 1232-1242.

Valente, K. N., Levy, N. E., Lee, K. H., & Lenhoff, A. M. (2018). Applications of proteomic methods for CHO host cell protein characterization in biopharmaceutical manufacturing. Current Opinion in Biotechnology, 53, 144-150.

Verrastro, I., Pasha, S., Jensen, K. T., Pitt, A. R., & Spickett, C. M. (2015). Mass spectrometry-based methods for identifying oxidized proteins in disease: Advances and challenges. Biomolecules, 5(2), 378-411.

Villén, J., Beausoleil, S. A., & Gygi, S. P. (2008). Evaluation of the utility of neutral-loss-dependent MS3 strategies in large-scale phosphorylation analysis. Proteomics, 8(21), 4444-4452.

Walker, D. E., Yang, F., Carver, J., Joe, K., Michels, D. A., & Yu, X. C. (2017). A modular and adaptive mass spectrometry-based platform for support of bioprocess development toward optimal host cell protein clearance. Paper presented at the Mabs, , 9. (4) pp. 654-663.

Walsh, C. T., Garneau-Tsodikova, S., & Gatto, G. J. (2005). Protein posttranslational modifications: The chemistry of proteome diversifications. Angewandte Chemie International Edition, 44(45), 7342-7372.

Walsh, G. (2014). Biopharmaceutical benchmarks 2014. Nature Biotechnology, 32(10), 992-1000.

Wan, S., Kelly, P. M., Mahon, E., Stöckmann, H., Rudd, P. M., Caruso, F., et al. (2015). The "sweet" side of the protein corona: Effects of glycosylation on nanoparticle–cell interactions. ACS Nano, 9(2), 2157-2166.

Wang, W. U., Chen, C., Lin, K. H., Fang, Y., & Lieber, C. M. (2005). Label-free detection of small-molecule-protein interactions by using nanowire nanosensors. Proceedings of the National Academy of Sciences of the United States of America, 102(9), 3208-3212. doi:0406368102 [pii]

Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., et al. (1995). Progress with gene-product mapping of the mollicutes: Mycoplasma genitalium. Electrophoresis, 16(1), 1090-1094.

Wei, Y. C., Naderi, S., Meshram, M., Budman, H., Scharer, J. M., Ingalls, B. P., et al. (2011). Proteomics analysis of chinese hamster ovary cells undergoing apoptosis during prolonged cultivation. Cytotechnology, 63(6), 663-677.

Wingens, M., Gätgens, J., Schmidt, A., Albaum, S. P., Büntemeyer, H., Noll, T., et al. (2015). 2D-DIGE screening of high-productive CHO cells under glucose limitation—Basic changes in the proteome equipment and hints for epigenetic effects. Journal of Biotechnology, 201, 86-97.

Wiśniewski, J. R., Zougman, A., & Mann, M. (2009). Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. Journal of Proteome Research, 8(12), 5674-5678.

Wurm, F. M. (2013). CHO quasispecies—implications for manufacturing processes. Processes, 1(3), 296-311.

Xu, G., & Jaffrey, S. R. (2013). Proteomic identification of protein ubiquitination events. Biotechnology and Genetic Engineering Reviews, 29(1), 73-109.

Xu, X., Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., et al. (2011). The genomic sequence of the chinese hamster ovary (CHO)-K1 cell line. Nature Biotechnology, 29(8), 735-741.

Yan, J. X., Devenish, A. T., Wait, R., Stone, T., Lewis, S., & Fowler, S. (2002). Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of escherichia coli. Proteomics, 2(12), 1682-1698.

Yang, X., Wang, Z., Guo, L., Zhu, Z., & Zhang, Y. (2017). Proteome-wide analysis of N-glycosylation stoichiometry using SWATH technology. Journal of Proteome Research, 16(10), 3830-3840.

Ye, H., Hill, J., Gucinski, A. C., Boyne, M. T., & Buhse, L. F. (2015). Direct site-specific glycoform identification and quantitative comparison of glycoprotein therapeutics: Imiglucerase and velaglucerase alfa. The AAPS Journal, 17(2), 405-415.

Yeung, Y., Nieves, E., Angeletti, R. H., & Stanley, E. R. (2008). Removal of detergents from protein digests for mass spectrometry analysis. Analytical Biochemistry, 382(2), 135-137.

Yuk, I. H., Nishihara, J., Walker Jr, D., Huang, E., Gunawan, F., Subramanian, J., et al. (2015). More similar than different: Host cell protein production using three null CHO cell lines. Biotechnology and Bioengineering, 112(10), 2068-2083.

Zhang, J., Chiodini, R., Badr, A., & Zhang, G. (2011). The impact of next-generation sequencing on genomics. Journal of Genetics and Genomics, 38(3), 95-109.

Zhang, N., Chen, R., Young, N., Wishart, D., Winter, P., Weiner, J. H., et al. (2007). Comparison of SDS-and methanol-assisted protein solubilization and digestion methods for escherichia coli membrane proteome analysis by 2-D LC-MS/MS. Proteomics, 7(4), 484-493.

Zhang, Q., Goetze, A. M., Cui, H., Wylie, J., Tillotson, B., Hewig, A., et al. (2016). Characterization of the co-elution of host cell proteins with monoclonal antibodies during protein A purification. Biotechnology Progress, 32(3), 708-717.

Zhu, G., Sun, L., Albanetti, T., Linkous, T., Larkin, C., Schoner, R., et al. (2016). Quantitative analysis of the supernatant from host and transfected CHO cells using iTRAQ 8-plex technique. Biotechnology and Bioengineering, 113(10), 2140-2148.

Clinical Proteomics
-Liquid Chromatography-
Mass Spectrometry (LC/MS)
Purification Systems

Authors: <u>Michael Henry</u> and Paula Meleady

Michael Henry and Paula Meleady conceived the presented idea for this chapter. Michael Henry was principally responsible for the design of the methodology and the preparation, creation and presentation of this chapter, specifically writing the initial draft.

Abstract

Liquid Chromatography/Mass Spectrometry (LC/MS) has become a routine powerful technology in clinical proteomic studies for protein identification, protein characterisation and the discovery of biomarkers. In this chapter, we describe two protocol methods to analyse clinical patient samples using a resin-based depletion column followed by either protein In-Gel enzymatic digestion or protein In-Solution enzymatic digestion and then analysis by One-Dimensional Reverse Phase Chromatography or Two-Dimensional Strong Cation Exchange (SCX) – Reverse Phase Chromatography (RPC).

Key words: Clinical proteomics, immuno-depletion, protein separation, peptide separation, protein identification, protein digestion, high resolution tandem mass spectrometry.

2.1 Introduction

Liquid chromatography-mass spectrometry (LC-MS) is an analytical technique that combines the physical separation capabilities of liquid chromatography with the mass analysis capabilities of mass spectrometry and is a fundamental tool in clinical proteomics (Aebersold and Mann 2003). Monitoring the protein expression pattern in mammalian cells and from clinical patient samples by proteomic technologies offers opportunities to discover potentially new biomarkers for the early detection and diagnosis of diseases including cancer (Linge et al. 2014). There are numerous protein and peptide fractionation techniques employed including traditional chromatography (Brunoro et al. 2015), immunodepletion (Martin-Lorenzo et al. 2014), subcellular fractionation (Ohmine et al. 2015), nanoparticle enrichment (Magni et al. 2014) and ProteoMiner™ technology (combinatorial library of hexapeptide ligands coupled to beads) (Monari et al. 2011) for protein expression profiling of clinical samples. Using these techniques, it is possible to carry out relative or absolute LC-MS quantitation. Relative quantitative LC-MS proteomic analysis uses labels to compare protein or peptide abundances between samples. Stable Isotope Labelling with Amino Acids in Cell Culture (SILAC) (Stebbing et al. 2015) as well as Isotope-Coded Affinity Tags (ICAT) (Shiio and Aebersold 2006) and isobaric tags like iTRAQ (Crabb et al. 2015) and TMT (Sinclair and Timms 2013) are commonly used proteomic labelling techniques used in quantitative proteomics. Absolute quantitation can also be applied to LC-MS with spiked known concentrations of labelled synthetic peptides (Capuano et al. 2011). A label-free approach can also be employed for both relative and absolute quantitation using LC-MS (Holland et al. 2015, Linge et al. 2014). These proteomic tools have been used for the differential analysis of various types of clinical biological samples including tissue (Holland et al. 2015), cell line models (Dowling, Pollard et al. 2015), saliva (Dowling et al. 2008), plasma (Dowling, Hughes et al. 2015), serum (Dowling et al. 2015), urine (Nijenhuis et al. 2015), and blood (Dowling et al. 2014), to better understand the molecular basis of the pathogenesis of disease, and also in the validation and characterisation of disease-associated proteins. There have been many advances in the use of LC-MS in the proteomic analysis of clinical samples but many challenges still remain including the dynamic range of proteins present in biological samples such as serum or plasma where potential biomarkers of interest may be present at very low concentrations thus requiring the depletion of large serum proteins, and also the sensitivity and reproducibility of the LC and MS instrumentation (Hakimi

et al. 2014). Albumin and IgG are two major protein components in serum/plasma contributing to approximately 80% of the total protein concentration. Removal of albumin and IgG can allow for the visualisation and possible identification of co-migrating proteins on one-dimensional (1D) and/or two-dimensional (2D) polyacrylamide gel electrophoresis (PAGE). Removal of these abundant proteins will also make possible the detection of low abundant proteins present as the in-spectrum dynamic range of most mass spectrometers is not compatible with the dynamic range of protein concentrations found in plasma/serum samples that have not been depleted (Tu et al. 2010).

In this chapter we describe LC-MS based approaches for analysing proteins from clinical patient bio-fluid samples that has been immuno-depleted of their high abundance serum proteins to allow for the detection of medium and low abundant proteins present in the sample. The first approach involves visualising a depleted serum sample on a Coomassie-blue stained SDS gel and In-gel digestion of the separated proteins using first-dimension reversed-phase liquid chromatography mass spectrometry (1D-RPC LC-MS). The second approach involves on-line two-dimensional liquid chromatography using strong cation exchange separation for the first-dimension fractionation followed by reversed-phase separation for the second-dimension (2D-SCX-RPC LC-MS) of the same serum sample after an In-solution protein digestion is carried out.

2.2 Materials and Methods

2.2.1 Immuno-depletion from serum sample

- ProteoPrep® Blue Albumin Depletion Kit (see Note 1).
- Serum sample from blood prepared and stored using standardised procedures (Dowling et al. 2014).
- Protein Assay: e.g. Quick StartTM Bradford Protein Assay Bio-Rad). Bovine Serum Albumin Standard (2 mg/mL) (Bio-Rad). Prepare standards using a 2 mg/mL stock of Bovine Serum Albumin (BSA) at the final concentrations: 1000, 750, 500, 250, 125 and 0 mg/mL.
- Microplate Spectrophotometer, e.g. Multiskan^{TM} GO (Thermo Fisher Scientific).
- Tube Rotator (e.g. SB2, Stuart).

2.2.2 Protein Precipitation

- Cold (-20 C) Acetone.
- Resuspension buffer: a solution containing 6 M Urea, 2 M thiourea, 10 mM Tris.
- Refrigerated high speed centrifuge (e.g. Hettich Mikro 200R).

2.2.3 Sample Preparation from 1D PAGE Gels (In-Gel Digestion)

- Gel Stain, e.g. Brilliant Blue G – Colloidal (Sigma) at 1X concentration (Note 2)
- LC-MS grade Water, LC-MS grade Methanol, LC-MS grade Acetonitrile, LC-MS grade formic acid (see Note 3).
- Ammonium Bicarbonate solutions: 100 mM and 50 mM. Prepare fresh on the day of use.
- Solution containing ammonium bicarbonate (10 mM) in 10% (v/v) acetonitrile. Prepare fresh on the day of use.
- 50 mM acetic acid solution.

- Trypsin solution. Use sequence grade trypsin (see Note 4). Prepare 12.5 ng/µL trypsin in 10 mM ammonium bicarbonate containing 10% (v/v) acetonitrile. Resuspend the trypsin at 1 µg/mL in 50 mM acetic acid (100 µg vial resuspended in 100 µL of 50 mM acetic acid buffer). Remove the required amount of this 80X trypsin stock needed and refreeze the unused portion in 10 µL aliquots and store at -20 ºC. Make the 1X trypsin shortly before use by diluting in 10 mM ammonium bicarbonate containing 10% (v/v) acetonitrile to a concentration of 12.5 ng/µL (1:80 (v/v)).

- ProteaseMAX™ Surfactant Trypsin Enhancer solution: Resuspend a 100 µg vial of ProteaseMAX™ (Promega) with 100 µL of 50 mM ammonium bicarbonate to give a 1% solution (see Note 5). Freeze 10 µL aliquots and store at -20 °C.

- Solution containing DTT (25 mM) in ammonium bicarbonate (50 mM). Make shortly before use.

- Solution containing iodoacetamide (55 mM) in ammonium bicarbonate (50 mM). Make shortly before use.

- TFA Trypsin stopping solution: 50% Trifluoroacetic acid (TFA) solution. Use a fume hood to prepare this solution.

- C18 - 10 µL ZipTips® (Millipore)

- Wetting Solution: 100% Acetonitrile.

- Equilibration Buffer (Washing Solution): 2% Acetonitrile containing 0.1% TFA.

- Elution Buffer: 50% Acetonitrile containing 0.1% TFA.

- Vacuum centrifuge (e.g. SpeedVac™).

2.2.4   One Dimensional Reverse Phase Chromatography

- Solvent A: 2% acetonitrile in LC-MS grade water containing 0.1% formic acid. Prepare 100 mL. Use a fume hood to prepare this solution.

- Solvent B: 80% Acetonitrile, 20% LC-MS grade water, 0.8% formic acid. Prepare 100 mL. Use a fume hood to prepare this solution.

- Sample loading buffer 1: 2% acetonitrile in LC-MS grade water containing 0.1% Trifluoroacetic acid (TFA). Prepare 1000 mL. Use a fume hood to prepare this solution.

- NanoLC system: e.g. Ultimate® 3000 RSLCnanoLC system (Thermo Fisher Scientific) or a similar system.

- Mass Spectrometer: e.g. LTQ Orbitrap XL (Thermo Fisher Scientific) or a similar system.

- Column: Acclaim®PepMap100 75 µm x 50 cm, nanoViper C18, 3 µm, 100Å (Thermo Fisher Scientific).

- Trap column: PepMap C18 trap cartridge (300 µm x 5 mm) (LC Thermo Fisher Scientific).

2.2.5   Sample Preparation from Protein In-solution

- Trypsin solution. See (see 2.2.3).
- ProteaseMAX^{TM} (Promega) solution: (see 2.2.3).
- Solution containing DTT (0.5 M) in ammonium bicarbonate (50 mM). Make shortly before use.
- Solution containing iodoacetamide (0.55 M) in ammonium bicarbonate (50 mM). Make shortly before use.
- TFA Trypsin stopping solution (see 2.2.3).

2.2.6   Two Dimensional Strong Cation Exchange – Reverse Phase Chromatography

- Solvent A, Solvent B, (2. 2.4).
- Sample loading buffer 2: 5% acetonitrile in LC-MS grade water containing 0.025% Trifluoroacetic acid (TFA). Prepare 1000 mL.
- 1 mM NaCl solution prepared in 5% Acetonitrile, 0.025% TFA.
- 2 mM NaCl solution prepared in 5% Acetonitrile, 0.025% TFA.
- 5 mM NaCl solution prepared in 5% Acetonitrile, 0.025% TFA.
- 10 mM NaCl solution prepared in 5% Acetonitrile, 0.025% TFA.
- 20 mM NaCl solution prepared in 5% Acetonitrile, 0.025% TFA.
- 50 mM NaCl solution prepared in 5% Acetonitrile, 0.025% TFA.
- 100 mM NaCl solution prepared in 5% Acetonitrile, 0.025% TFA.

- 200 mM NaCl solution prepared in 5% Acetonitrile, 0.025% TFA.

- 500 mM NaCl solution prepared in 5% Acetonitrile, 0.025% TFA.

- 1 M NaCl solution prepared in 5% Acetonitrile, 0.025% TFA.

- SCX Column:BioX- SCX, 5 μm (500 μm x 15 mm) (Thermo Fisher Scientific).

- Column: Acclaim®PepMap100 75 μm x 50 cm, nanoViper C18, 3μm, 100Å (Thermo Fisher Scientific).

- Trap column: PepMap C18 trap cartridge (300 μm x 5 mm) (Thermo Fisher Scientific).

- NanoLC system: e.g. Ultimate® 3000 RSLCnanoLC system (Thermo Fisher Scientific) or a similar system.

- Mass Spectrometer: e.g. LTQ Orbitrap XL (Thermo Fisher Scientific) or a similar system.

2.3   Methods

2.3.1   Immuno-depletion of serum sample

1.   Transfer 0.4 mL of the ProteoPrep Blue albumin and IgG depletion resin per sample to a spin column and centrifuge at 8,000 xg for 10 s.

2.   Wash the resin with 0.4 mL of equilibration buffer and centrifuge at 8,000 xg for 10 s.

3.   Add 50 μL of serum to the resin and incubate at room temperature rotating for 10 min using a tube rotator.

4.   Reapply the eluate back onto the resin for a further 5 min rotating at room temperature before centrifuging again at 8,000 xg for 10 s. Transfer the supernatant to a fresh tube.

5.   Wash the resin with 100 μL of equilibration buffer, centrifuge at 8,000xg for 10 s and pool the sample with the first collection (150 μL in total).

6.   Determine the protein concentration prior to separation by SDS PAGE. Pipette 5 μL of standard and unknown samples into individual wells of a microplate in triplicate. Add 250 μL of Bradford dye reagent to each sample and mix. Incubate at room temperature for a minimum of 5 min and a maximum of 1 h.

Set spectrophotometer to 595 nm and zero against the blank samples. (See Note 6).

7. Run samples on a standard 1D SDS-PAGE gel and stain with coomassie blue, e.g. Brilliant Blue G – Colloidal.

8. Shows a Coomassie blue stained SDS-PAGE gel comparing a 10 µg serum sample before and after albumin and IgG depletion, and shows the reduction in the amounts of highly abundant IgG and albumin.

2.3.2 Protein In-Gel Digestion protocol

1. Excise the protein bands of interest from the stained gel (see Note 7) using a sterile polypropylene pasteur pipette (see Note 8) and transfer the cored gel spots to a 1.7 mL microcentrifuge tube.

2. Wash each gel plug with 200 µL of LC-MS grade water.

3. Destain the gel plug with 100 µL of Methanol:50 mM Ammonium Bicarbonate (1:1 vol:vol) to remove the stain.

4. Dehydrate the gel plug with 100 µL Acetonitrile. Once the gel slice has dehydrated remove the supernatant and dry in a vacuum centrifuge (e.g. SpeedVac^{TM}) until the sample is dry.

5. Reduce the sample with 100 µL of 25 mM DTT in 50 mM ammonium bicarbonate for 20 min at 56 °C (see Note 9).

6. Alkylate the sample with 100 µL of 55 mM iodoacetamide in 50 mM ammonium bicarbonate at room temperature in the dark for 20 min.

7. Dehydrate and dry the sample again as outlined in 3.2.4.

8. Digest the sample by rehydrating with 20 µL of a 12.5 ng/µL trypsin (see Note 10) containing 0.01% ProteaseMAX for 10 min. Overlay the digesting sample with 30 µL of 50 mM ammonium bicarbonate and gently mix.

9. Incubate at 50 °C for 1 h or 37 °C for 2-4 h.

10. Centrifuge at 12,000 xg for 10 s to collect any condensate.

11. Transfer the digested peptides to a fresh tube and acidify the solution with the addition of TFA Trypsin stopping solution to a final concentration of 0.5%. If the amount of digested protein is sufficiently high, the digest can be analysed

directly by LC-MS or it can be stored at -20 °C (if it is not going to be analysed straight away).

12.   However, if the peptide abundance is low, the peptide sample can be concentrated and purified with 10 µL C18 Millipore ZipTips®. Wet the C18 ZipTip by aspirating and dispensing 15 µL of wetting solution x 3 times through the ZipTip and discarding the solution each time (see Note 11). Wet the ZipTip again by aspirating and dispensing 15 µL of elution buffer x 3 times discarding the solution each time. Equilibrate the ZipTip by aspirating and dispensing 15 µL of equilibration buffer x 3 times discarding the solution each time. Bind the low abundant peptide sample to the ZipTip by aspirating and dispensing the sample through the ZipTip without discarding the sample. After binding is completed, wash the sample on the Zip Tip by aspirating and dispensing 15 µL of equilibration buffer solution x 3 times. Elute the sample with 10 µL of elution buffer into a clean microcentrifuge tube and repeat this step. As the elution buffer contains acetonitrile, the sample must be evaporated by vacuum centrifugation (e.g. Speed Vac^{TM}) prior to LC-MS analysis. If the digest is not going to be analysed straight away, freeze and store at –20 °C. If going straight to LC-MS analysis, rehydrate the sample in 20 µL of 2% acetonitrile containing 1% formic acid.

2.3.3   One dimensional (1D) reverse phase chromatography using a 60 minute separation time.

1.   Carry out nano-flow chromatography with a suitable LC such as the Ultimate® 3000 RSLC nanoLC system which is used here. The LC is configured, as shown in Figure 2, using an Acclaim®PepMap100 75 µm x 50 cm column and a nanoViper C18, 3µm, 100Å and PepMap C18 trap cartridge (300 µm x 5 mm).

2.   Prior to analysis, equilibrate the columns in Solvent A for 10 min and set the column temperatures to 35 °C using a column oven.

3. Using a 20 μL loop, pick up a 5 μL injection of sample at a flow rate of 25 μL/min using Sample loading buffer 1 and desalt for 3 min through the PepMap C18 trap cartridge.

4. Separate the peptides using a 60 min linear gradient elution using Solvent A and Solvent B starting with 2% Solvent B to 90% B over 60 min at a flow rate of 350 nL/min. See Note 12 for details of longer separation times used for complex peptide separations. Figure 2-3 shows the peptides obtained from a typical 1D LC-MS experimental run from a 1D PAGE gel band over a 60 minute gradient from a patient serum sample after immuno-depletion.

5. Use a stainless steel emitter (see Note 13) for nano-electrospraying the samples into the mass spectrometer with the ion spray voltage set at 1.5 kV.

6. Use a tandem mass spectrometer (MS/MS) to analyse the peptides eluting from the nano-flow chromatography system. We use an LTQ OrbitrapXL™ Hybrid Ion Trap-Orbitrap Mass Spectrometer for this analysis. Operate the MS analysis in a data dependent mode so that the top 10 most abundant precursors in each MS scan are selected for MS/MS using a collision energy of 32%, precursor isolation is set at a width of 2 Da, intensity threshold for precursor selection of 1,000). Select the following global parameters; precursor ions between m/z 400-1200 at a resolution of 30,000 collected in profile mode. The m/z values for charge state =1 and >4 are rejected. Enable dynamic exclusion with a repeat count of 1 and set the exclusion duration to 40 seconds.

7. Use computer software search engine(s) to analyse data from tandem mass-spectrometry samples to match mass spectra with peptide sequences. The search engine(s) will match the tandem mass spectra with peptide sequences from a protein sequence database and use the identified peptides to infer the protein content of the sample. Through Proteome Discoverer 1.4 (Thermo Fisher Scientific), we use a combination of the search engines MASCOT (www.matrixscience.com) and SEQUEST (Thermo Fisher Scientific) against the UniProt/SwissProt fasta sequence database using the homo sapien (human)

taxomony when searching the data from patient samples. Set cysteine carbamidomethylation as a fixed modification and set oxidation of methionine as a variable modification. For database analysis using OrbitrapXL data, set a precursor mass tolerance of 10 ppm and an MS/MS tolerance at 0.6 Da using full trypsin specificity and allow for 2 missed cleavages.

8. We use strict scoring criteria for our protein identification search engine results (Omenn et al. 2005). (See Note 14).

2.3.4 Protein In-Solution Digestion protocol

1. Precipitate the protein sample (50-200 µg) if it contains detergent or you are not sure of the composition of the buffer with 4 volumes of pre-chilled -20 °C acetone. Cool the required volume of acetone to -20 ºC for 20 min beforehand. Vortex the sample and incubate for 60 min at -20 ºC. Centrifuge at 12,000 xg for 10 min at 4 °C to pellet the protein. Resuspend the pellet in resuspension buffer (see 2.2.2) at a minimum concentration of 1.2 mg/mL

2. Determine protein concentration using Bradford Protein Assay as outlined in 3.1.6.

3. Digest 10 µg of protein (ensure final concentration of urea in the solution is <1 M) (see Note 15).

4. Add 50 mM Ammonium Bicarbonate to a final volume of 50 µL.

5. Add 0.5 µL of 0.5 M DTT and incubate at 56 ºC for 20 min.

6. Add 1.35 µL of 0.55 M Iodoacetamide and incubate at room temperature in the dark for 20 min.

7. Add 1 µg of trypsin and 0.5 µL of 1% ProteaseMAX and digest at 37 °C for a minimum of 3 h.

8. Centrifuge at 12,000 xg for 10 s to collect any condensate and to stop the trypsin activity by the addition of TFA Trypsin stopping solution to a final concentration of 0.5%.

9. If the digest is not going to be analysed straight away, freeze and store at −20 °C.

2.3.5 Two dimensional (2D) strong cation exchange (SCX) reverse phase chromatography (RPC) for complex protein mixtures

1. We carry out nano-flow chromatography using an Ultimate® 3000 RSLC nanoLC system, using an Acclaim®PepMap100 75µm x 50cm, nanoViper C18, 3µm, 100Å, a PepMap C18 trap cartridge (300 µm x 5 mm) and a BioX-SCX (500 µm x 15 mm) configured as outlined in Figure 4.

2. Using a 100 µL loop, pick up a 10 µL injection of peptide sample at a flow rate of 25 µL/min with Sample loading buffer 2 to load onto the BioX-SCX trap for 5 min.

3. Elute the peptides stepwise from the SCX trap through the different salt steps of 1, 2, 5, 10, 20, 50, 100, 200, 500, 1000 mM.

4. Carry out a linear gradient elution using Solvent A and Solvent B starting with 2% buffer B to 90% B over 40 min at a flow rate of 350 nL/min.

5. Figure 5 shows base peak chromatograms for the salt injections on the SCX trap. The fractions are overlaid to show different peptide profiles following the separation on the reverse-phase column.

6. Mass spectrometer parameters are outlined in 2.3.6.

7. Data analysis is outlined in 2.3.7.

2.4   Notes

1. There are many commercially available immunodepletion kits and columns that can remove high abundant proteins in serum such as IgG, serum albumin, etc. Examples include ProteoPrep® and Seppro® (SIGMA), ProteoSpin^{TM} (NORGEN BioTek Corp), Proteome Purify (R&D Systems), Top2 and Top12 Depletion Spin Columns (Pierce), Multiple Affinity Removal (MARS) range (Agilent). These kits and columns are not just confined to use on serum but can be used for immuno-depletion of other biological fluids (e.g. plasma, vitreous fluid, wound exudate, etc.).

2. Proteins separated by either 1D or 2D PAGE for protein identification are normally visualised by staining with dyes such as Coomassie Brilliant Blue (CBB) R250 and silver stain (gluteraldehyde-free), or fluorescence-based methods such as Sypro Ruby, Deep Purple, etc. For mass spectrometry analysis, the silver stain method should not use glutaraldehyde.

3. All solvents and water used must be LC-MS grade. All chemicals must be of the highest purity.

4. The most common enzyme used for peptide generation for LC-MS analysis is trypsin which has been modified by reductive alkylation to suppress trypsin autolysis. Trypsin is a robust enzyme and digestions can be performed under denaturing conditions. We use modified porcine trypsin (Promega) in our laboratory.

5. We use ProteaseMax^{TM} Surfactant Trypsin Enhancer as it removes the need to extract peptides following In-gel digestion. We also find there is improved peptide recovery from gels compared to when not using it. When we carry out protein In-solutions digestions we find protein solubilisation is improved. ProteaseMaxTM Surfactant Trypsin Enhancer degrades over the course of the digestion, so samples are ready for LC-MS without the need to remove the detergent.

6. To ensure precision in determining protein concentrations when using a Bradford protein assay, the relative standard deviation (RSD) for each triplicate sample UV value must be very low <1%.

7. When working with SDS-PAGE gels, gloves must be worn at all times to minimise keratin and dust contamination which can affect your LC-MS results. When excising protein from gels this work should be done in a laminar flow cabinet to minimise the possibility of any hair, dust or skin flakes contaminating the sample with unwanted keratins.

8. The use of polypropylene tubes and tips helps to minimise protein loss by adsorption.

9. Reduction and alkylation is essential for samples prepared by 1D SDS-PAGE separation even when reducing agent is already present in the loading buffer.

10. There are enzymes other than trypsin that can be used for protein digestion which have different cleavage specificities, such as Asp-N and Glu-C.

11. Avoid introducing any air bubbles into the ZipTip at any stage of the process.

12. Complex peptide mixtures can require long separation times. We use 120, 180 or 300 min separation times depending on the complexity of the sample.

13. We use stainless steel emitters to maintain a stable spray into our Mass Spectrometer however stainless steel emitters are not suitable when looking at phosphorylated peptides so fused silica emitters are recommended instead.

14. Peptides with either ion scores of 40 or more (from Mascot) and peptides with XCorr scores >1.9 for singly charged ions, >2.2 for doubly charged ions and >3.75 for triply charged ions or more (from Sequest HT) are accepted as high confidence identifications Proteins with 2 peptides matched are generally accepted as identified.

15. Trypsin activity is inhibited by urea at concentrations >1M.

2.5 Figure Legends

Figure 2-1. Non-depleted (left) vs Serum/IgG depleted sample (right) from a serum sample separated on a Coomassie stained SDS-PAGE gel. The region highlighted in red shows proteins no longer masked by very high levels of albumin and IgG. Non-depleted (left) vs Serum/IgG depleted sample (right) from a serum sample separated on a Coomassie stained SDS-PAGE gel. The region highlighted in red shows proteins no longer masked by very high levels of albumin and IgG.

Figure 2-2. Schematic using a 10 port valve for the fluidic connections for one dimensional (1D) reverse phase chromatography for simple protein mixtures.Schematic using a 10 port valve for the fluidic connections for one dimensional (1D) reverse phase chromatography for simple protein mixtures.

Figure 2-3. Mass spectrometry trace showing the total ion intensity from all the mass spectra recorded during a 1D reverse phase LC-MS using a 60 min reverse phase separation, shown as a function of elution time. The sample shown is a serum protein digest from a 1D gel band.Mass spectrometry trace showing the total ion intensity from all the mass spectra recorded during a 1D reverse phase LC-MS using a 60 min reverse phase separation, shown as a function of elution time. The sample shown is a serum protein digest from a 1D gel band.

Figure 2-4. **Schematic using two 10 port valves for the fluidic connections for two dimensional (2D) LC salt plug experiment.** Schematic using two 10 port valves for the fluidic connections for two dimensional (2D) LC salt plug experiment.

Figure 2-5. Mass spectrometry traces overlaid from 10 salt plug fractions generated from a 2D-LC run. Base peak chromatograms for the salt injections on the SCX trap fractions are overlaid to show different peptide profiles following separation on the reverse-phase column. Each fraction is designated by a different colour chromatogram. Mass spectrometry traces overlaid from 10 salt plug fractions generated from a 2D-LC run. Base peak chromatograms for the salt injections on the SCX trap, fractions are overlaid to show different peptide profiles following separation on the reverse-phase column. Each salt fraction is designated by a different colour chromatogram.
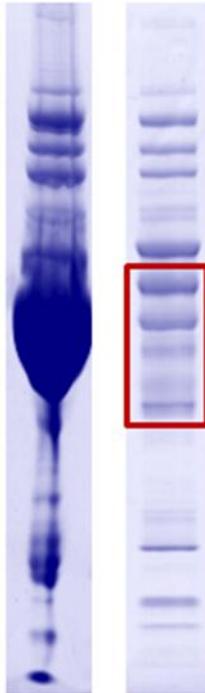
Figure 1

Figure 2-1. Non-depleted (left) vs Serum/IgG depleted sample (right) from a serum sample separated on a Coomassie stained SDS-PAGE gel. The region highlighted in red shows proteins no longer masked by very high levels of albumin and IgG.
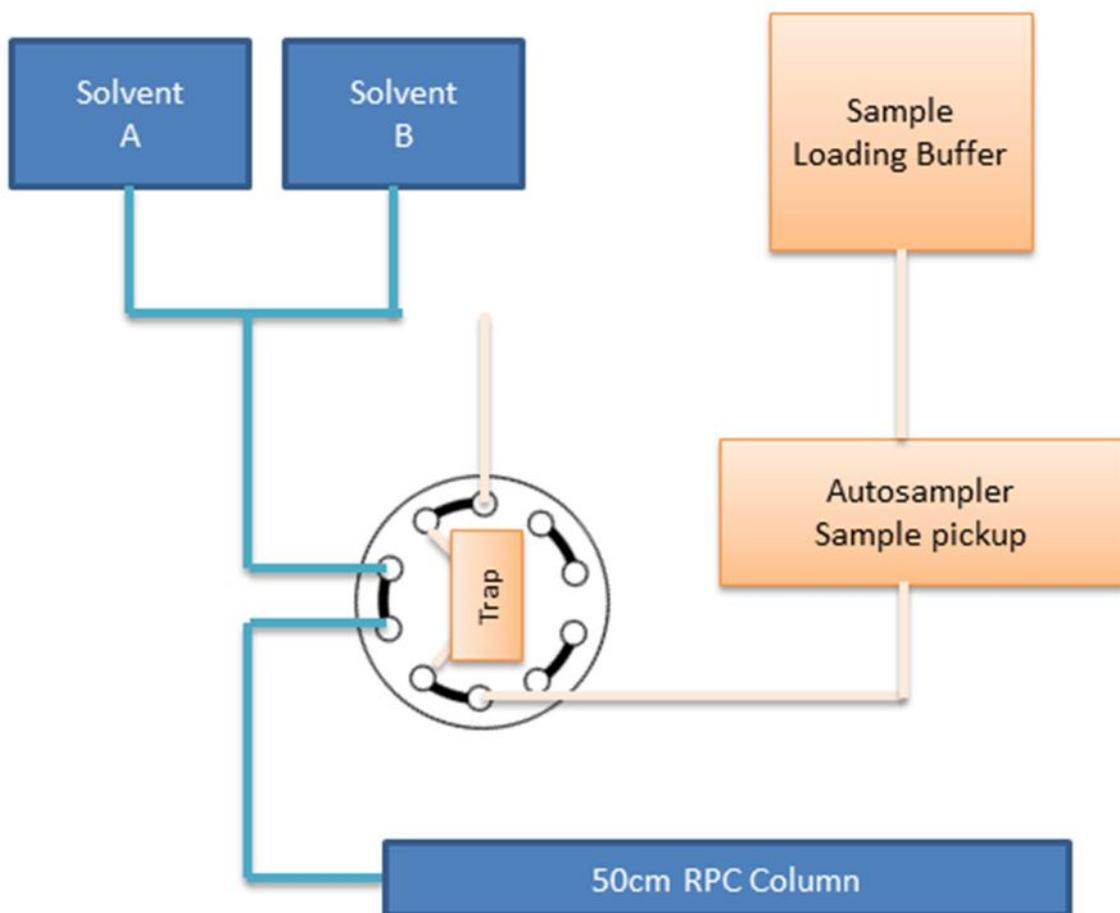
Figure 2

Figure 2-2. Schematic using a 10 port valve for the fluidic connections for one dimensional (1D) reverse phase chromatography for simple protein mixtures.
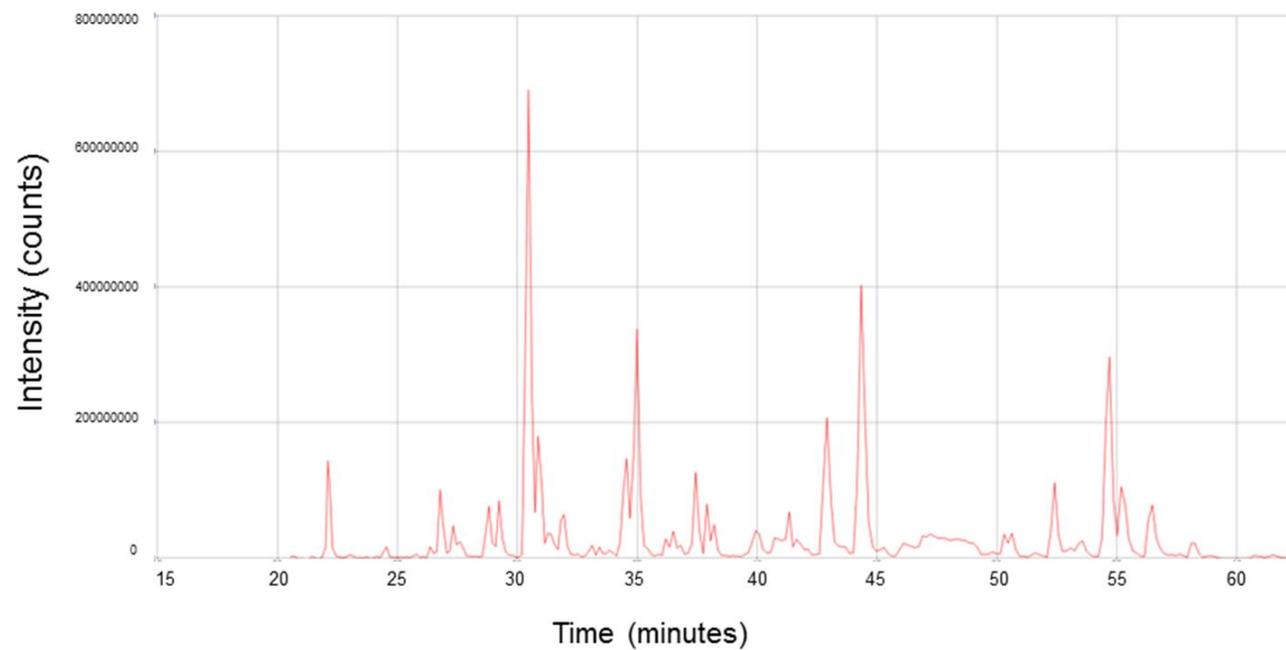
Figure 3

Figure 2-3. Mass spectrometry trace showing the total ion intensity from all the mass spectra recorded during a 1D reverse phase LC-MS using a 60 min reverse phase separation, shown as a function of elution time. The sample shown is a serum protein digest from a 1D gel band.
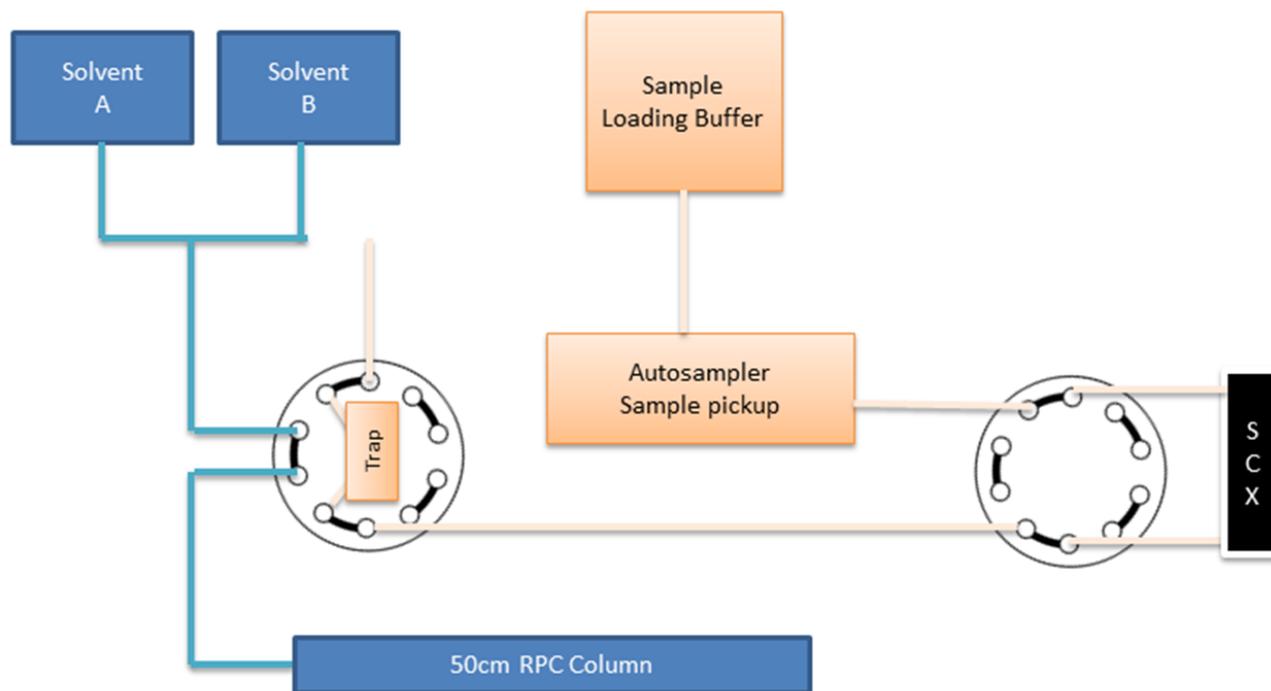
Figure 4

Figure 2-4. Schematic using two 10 port valves for the fluidic connections for two dimensional (2D) LC salt plug experiment.
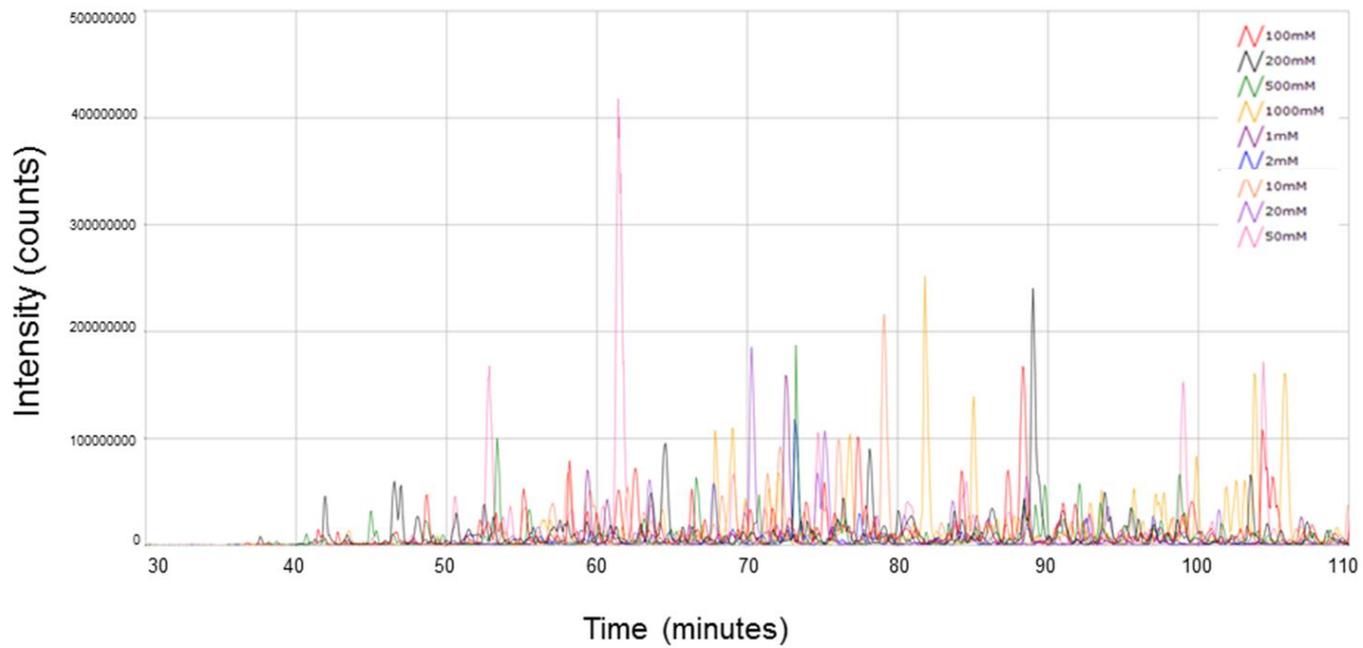
Figure 5

.

Figure 2-5. Mass spectrometry traces overlaid from 10 salt plug fractions generated from a 2D-LC run. Base peak chromatograms for the salt injections on the SCX trap fractions are overlaid to show different peptide profiles following separation on the reverse-phase column. Each fraction is designated by a different colour chromatogram.

2.6   References

Aebersold, R. and Mann, M. 2003. Mass spectrometry-based proteomics. *Nature,* 422(6928), pp.198.

Brunoro, G.V.F., Carvalho, P.C., da Silva Ferreira, André Teixeira, Perales, J., Valente, R.H., de Moura Gallo, Claudia Vitória, Pagnoncelli, D. and da Costa Neves-Ferreira, Ana Gisele. 2015. Proteomic profiling of nipple aspirate fluid (NAF): Exploring the complementarity of different peptide fractionation strategies. *Journal of Proteomics,* 117pp.86-94.

Capuano, F., Bond, N.J., Gatto, L., Beaudoin, F., Napier, J.A., Benvenuto, E., Lilley, K.S. and Baschieri, S. 2011. LC-MS/MS methods for absolute quantification and identification of proteins associated with chimeric plant oil bodies. *Analytical Chemistry,* 83(24), pp.9267-9272.

Crabb, J.W., Hu, B., Crabb, J.S., Triozzi, P., Saunthararajah, Y., Tubbs, R. and Singh, A.D. 2015. iTRAQ quantitative proteomic comparison of metastatic and non-metastatic uveal melanoma tumors. *PloS One,* 10(8), pp.e0135543.

Dowling, P., -, W.R., -, M.P., -, H.M., -, C.A. and -, C.M. - Analysis of the saliva proteome from patients with head and neck squamous cell carcinoma reveals differences in abundance levels of proteins associated with tumour progression and metastasis. - *J Proteomics.2008 Jul 21;71(2):168-75.Doi: 10.1016/j.Jprot.2008.04.004.Epub 2008 may 1.,*

Dowling, P., Hayes, C., Ting, K.R., Hameed, A., Meiller, J., Mitsiades, C., Anderson, K.C., Clynes, M., Clarke, C. and Richardson, P. 2014. Identification of proteins found to be significantly altered when comparing the serum proteome from multiple myeloma patients with varying degrees of bone disease. *BMC Genomics,* 15(1), pp.904.

Dowling, P., Hughes, D.J., Larkin, A.M., Meiller, J., Henry, M., Meleady, P., Lynch, V., Pardini, B., Naccarati, A. and Levy, M. 2015. Elevated levels of 14-3-3 proteins, serotonin, gamma enolase and pyruvate kinase identified in clinical samples from patients diagnosed with colorectal cancer. *Clinica Chimica Acta,* 441pp.133-141.

Dowling, P., Pollard, D., Larkin, A., Henry, M., Meleady, P., Gately, K., O'Byrne, K., Barr, M.P., Lynch, V. and Ballot, J. 2015. Abnormal levels of heterogeneous nuclear ribonucleoprotein A2B1 (hnRNPA2B1) in tumour tissue and blood samples from patients diagnosed with lung cancer. *Molecular BioSystems,* 11(3), pp.743-752.

Dowling, P., Wormald, R., Meleady, P., Henry, M., Curran, A. and Clynes, M. 2008. Analysis of the saliva proteome from patients with head and neck squamous cell carcinoma reveals differences in abundance levels of proteins associated with tumour progression and metastasis. *Journal of Proteomics,* 71(2), pp.168-175.

Hakimi, A., Auluck, J., Jones, G.D., Ng, L.L. and Jones, D.J. 2014. Assessment of reproducibility in depletion and enrichment workflows for plasma proteomics using label-free quantitative data-independent LC-MS. *Proteomics,* 14(1), pp.4-13.

Holland, A., Henry, M., Meleady, P., Winkler, C.K., Krautwald, M., Brinkmeier, H. and Ohlendieck, K. 2015. Comparative label-free mass spectrometric analysis of mildly versus severely affected mdx mouse skeletal muscles identifies annexin, lamin, and vimentin as universal dystrophic markers. *Molecules,* 20(6), pp.11317-11344.

Linge, A., Maurya, P., Friedrich, K., Baretton, G.B., Kelly, S., Henry, M., Clynes, M., Larkin, A. and Meleady, P. 2014. Identification and functional validation of RAD23B as a potential protein in human breast cancer progression. *Journal of Proteome Research,* 13(7), pp.3212-3222.

Magni, R., Espina, B.H., Liotta, L.A., Luchini, A. and Espina, V. 2014. Hydrogel nanoparticle harvesting of plasma or urine for detecting low abundance proteins. *Journal of Visualized Experiments : JoVE,* (90):e51789. doi(90), pp.e51789.

Martin-Lorenzo, M., Gonzalez-Calero, L., Zubiri, I., Diaz-Payno, P.J., Sanz-Maroto, A., Posada-Ayala, M., Ortiz, A., Vivanco, F. and Alvarez-Llamas, G. 2014. Urine 2 DE proteome analysis in healthy condition and kidney disease. *Electrophoresis,* 35(18), pp.2634-2641.

Monari, E., Casali, C., Cuoghi, A., Nesci, J., Bellei, E., Bergamini, S., Fantoni, L.I., Natali, P., Morandi, U. and Tomasi, A. 2011. Enriched sera protein profiling for detection of non-small cell lung cancer biomarkers. *Proteome Science,* 9(1), pp.55.

Nijenhuis, C., Lucas, L., Rosing, H., Schellens, J., Beijnen, J., Gorman, S., Burke, S., Campbell, D., Chapple, M. and Yousey, T. 2015. Validation of high-performance liquid chromatography–tandem mass spectrometry assays quantifying omacetaxine mepesuccinate and its 4'-des-methyl and cephalotaxine metabolites in human plasma and urine. *Journal of Chromatography B,* 1002pp.152-159.

Ohmine, K., Kawaguchi, K., Ohtsuki, S., Motoi, F., Ohtsuka, H., Kamiie, J., Abe, T., Unno, M. and Terasaki, T. 2015. Quantitative targeted proteomics of pancreatic cancer: Deoxycytidine kinase protein level correlates to progression-free survival of patients receiving gemcitabine treatment. *Molecular Pharmaceutics,* 12(9), pp.3282-3291.

Shiio, Y. and Aebersold, R. 2006. Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. *Nature Protocols,* 1(1), pp.139.

Sinclair, J. and Timms, J.F. 2013. Proteomic profiling of ovarian cancer models using TMT-LC-MS/MS *IN:* Anonymous*Ovarian Cancer.* Springer, pp.271-284.

Stebbing, J., Zhang, H., Xu, Y., Grothey, A., Ajuh, P., Angelopoulos, N. and Giamas, G. 2015. Characterization of the tyrosine kinase-regulated proteome in breast cancer by combined use of RNA interference (RNAi) and stable isotope labeling with amino acids in cell culture (SILAC) quantitative proteomics. *Molecular & Cellular Proteomics : MCP,* 14(9), pp.2479-2492.

Tu, C., Rudnick, P.A., Martinez, M.Y., Cheek, K.L., Stein, S.E., Slebos, R.J. and Liebler, D.C. 2010. Depletion of abundant plasma proteins and limitations of plasma proteomics. *Journal of Proteome Research,* 9(10), pp.4982-4991.

Phosphopeptide enrichment and LC-MS/MS analysis to study the phosphoproteome of recombinant Chinese hamster ovary cells

Authors: Michael Henry, Orla Coleman, Prashant Kaushik, Martin Clynes and Paula Meleady

Michael Henry and Paula Meleady conceived the presented idea for this chapter. Michael Henry was principally responsible for the design of the methodology and the preparation, creation and presentation of this chapter, specifically writing the initial draft.

Abstract

The reversible phosphorylation of proteins on serine, threonine and tyrosine residues is one of the most important post-translational modifications that regulate many biological processes. The phosphoproteome has not been studied in any great detail in recombinant Chinese hamster ovary (CHO) cells to date despite phosphorylation playing a crucial role in regulating many molecular and cellular processes relevant to bioprocess phenotypes including, for example, transcription, translation, growth, apoptosis and signal transduction. In this chapter we provide a protocol for phosphoproteomic analysis of Chinese hamster ovary cells using phosphopeptide enrichment with metal oxide affinity chromatography (MOAC) and immobilized metal affinity chromatography (IMAC) techniques, followed by site-specific identification of phosphorylated residues using LC-MS (MS2 and MS3) strategies.

Key words: Phosphoproteomics, phosphopeptide enrichment, site specific phosphorylation, Chinese hamster ovary cells.

3.1   Introduction

The Chinese hamster ovary (CHO) cell is the most commonly used mammalian host cell line for the production of recombinant biotherapeutics (Walsh 2014). This dominance is likely to continue for the foreseeable future due to their ability to correctly fold proteins that are compatible with human use, their ability to produce proteins with human-like post-translational modifications, their safety record, and their track record in industry (Wurm 2004). There has been considerable success in developing high-producing CHO cell culture processes through media optimisation and bioreactor design (Prentice, Ehrenfels and Sisk 2007, Altamirano et al. 2000); however, the bottlenecks in the cellular machinery for the efficient production of recombinant proteins are poorly understood. In order to improve the efficiency of production with the overall aim to reduce the costs of therapeutics a fundamental understanding of the biology underpinning growth and productivity of CHO cells is required. Systems biology approaches (including transcriptomics, proteomics, metabolomics, etc.) are being increasingly used to characterise recombinant CHO cells and will be instrumental in metabolic and cell engineering interventions to enhance CHO cell lines for more efficient production of therapeutics (Gutierrez and Lewis 2015, Kildegaard et al. 2013).

In the past CHO systems biology has been hampered by the lack of genomic data on CHO cells. The recently sequenced genomes of CHO cells (Kaas et al. 2015, Lewis et al. 2013, Brinkrolf et al. 2013, Xu et al. 2011) have opened up a new era involving analysis of CHO cells on a systems biology level. Examining the transcriptome can provide insight into the regulation of cellular processes but mRNA levels are not necessarily correlated with protein expression levels for their protein products (Baycin-Hizal et al. 2012b, Guo et al. 2008). In addition, proteins are subjected to a variety of post-translational modifications (PTMs) that affect biological activity; phosphorylation being one of most widely studied PTMs, especially in disease states such as cancer. The vast majority of 'omic based studies in CHO cells to date have missed out on the post-translational level of regulation despite their pivotal role in the regulation of growth, cell cycle arrest, apoptosis, transcription, signal transduction, etc. and hence are likely to be central to understanding and controlling bioprocess-relevant phenotypes. The phosphoproteome has not been assessed in any detail in CHO (A. Farrell et al. 2014) though some research has been done in CHO cells, based on knowledge of phosphorylation events in cancer cells described in the literature.

PTMs increase chemical diversity and complexity of the proteome, and can determine protein function by altering activity, cellular location, turnover, and interaction with other proteins (Choudhary and

Mann 2010). The complexity of the cellular processes that rely upon phosphorylation is demonstrated in that approximately 30% of all eukaryotic proteins can become phosphorylated (Olsen et al. 2010, Cohen 2001), though some of these phosphorylation events may be physiologically irrelevant (Dephoure et al. 2013). Almost half of the tyrosine kinases of the human 'kinome' are implicated in human cancers as well as numerous serine-threonine kinases, including Akt and mTOR (Martini et al. 2014). Phosphatases such as PTEN and PP2A are also known to play important roles in cancer (Hopkins et al. 2014, A. S. Farrell et al. 2014). Phosphorylation of nuclear proteins has been widely recognised as a key chemical modification necessary for the control of mRNA transcription (Whitmarsh and Davis 2000)*.*

While some proteins are constitutively phosphorylated, most are only transiently phosphorylated following key cellular cues, thus altering the chemical state of a protein in subtle ways that are not easily detected using standard profiling techniques (Simon and Cravatt 2008). Phosphorylation sites in most proteins are sub-stoichiometric and hence are often present in low abundance compared to their native counterparts. As a result phosphopeptide enrichment using immobilized metal ion (Fe^{3+} or Ga^{3+}) affinity chromatography (IMAC) and metal oxide affinity chromatography (MOAC) using TiO_2 or ZrO_2 prior to LC-MS/MS analysis is widely used in phosphoproteomic studies (Angel et al. 2012). Both these materials have different binding specificities, resulting in increased phosphoproteome coverage (Thingholm et al. 2008).

In this chapter we describe methods for phosphopeptide enrichment using TiO_2 and IMAC strategies from CHO cells followed by LC-MS analyses (MS2 and MS3) for site-specific determination of phosphorylated serine, threonine or tyrosine residues. The methods described are also amenable to differential phosphoproteomic studies.

3.2 Materials

3.2.1 Equipment

- Pierce® Fe-NTA (IMAC) Phosphopeptide Enrichment Kit (Thermo Fisher Scientific).
- Pierce® TiO$_2$ Phosphopeptide Enrichment Kit (Thermo Fisher Scientific).
- Pierce™ Graphite Spin Columns (Thermo Fisher Scientific).
- Pierce™ Strong Cation Exchange (SCX) Spin Column, Mini (Thermo Fisher Scientific).
- Microcentrifuge capable of spinning 1.7 mL microcentrifuge tubes at 14,000 x*g* at 4°C.
- Microcentrifuge polypropylene 1.7 mL collection tubes.
- Sonicating probe.
- Vortex.
- Fume Hood.
- Microplate Spectrophotometer reading 96-well plate format at 595 nm.
- Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific).
- Dry Heating Block.
- Vacuum Evaporator/Lyophiliser (e.g. SpeedVac™, Thermo Fisher Scientific).
- Ultimate® 3000 RSLCnanoLC system (Thermo Fisher Scientific).
- LTQ Orbitrap XL (Thermo Fisher Scientific).
- SilicaTip™ Standard Coating Tubing OD/ID 360/20µm Tip, ID 10 µm, length 5cm (New Objective).
- Analytical column: Acclaim®PepMap100 75µm x 50cm, nanoViper C18, 3µm, 100Å (Thermo Fisher Scientific).
- Trap column: C18 PepMap, 300 µm ID × 5 mm, 5 µm particle size, 100 Å pore size; Thermo Fisher Scientific).
- Xcalibur software, version 2.0.7 (Thermo Fisher Scientific).
- Proteome Discover 2.1 (Thermo Fisher Scientific) with SEQUEST HT and phospho*RS* 3.1 (Taus et al. 2011a) and suitable *Cricetulus griseus* protein database in fasta format (Meleady et al. 2012a)*.*

3.2.2 Reagents

Ensure that all solvents and water are LC-MS grade.

- DL-Dithiothreitol (DTT).

- Iodoacetamide.

- Urea.

- LC-MS grade water.

- HEPES.

- Sequencing-grade modified trypsin (e.g. Promega, Thermo Fisher Scientific, Sigma Aldrich).

- ProteaseMax^{TM} Surfactant Trypsin Enhancer (Promega).

- Trifluoroacetic acid (TFA) Use a fume hood when preparing solutions with this acid.

- Acetonitrile (ACN).

- Ammonium Bicarbonate.

- Halt™ Protease Inhibitor Cocktail (Thermo Fisher Scientific).

- Halt™ Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific).

- Ammonium Hydroxide (NH$_4$OH) ACS Reagent (Sigma Aldrich).

- Potassium Phosphate Monobasic (KH$_2$PO$_4$).

- Potassium Chloride (KCl).

3.2.3 Cell Lysis, Protein Extraction and Protein Quantification

- 50 mM Ammonium Bicarbonate (must be freshly prepared and used within 24 hours).

- 0.5 M DTT in 50 mM Ammonium Bicarbonate (must be freshly prepared).

- 0.55 M Iodoacetamide in 50 mM Ammonium Bicarbonate (must be freshly prepared).

- Trypsin solution: Use sequence grade trypsin (*see* **Note 1**).

- ProteaseMax^{TM} Surfactant Enhancer (*see* **Note 2**).

3.2.4 Protein Digestion

- 50 mM Ammonium Bicarbonate (must be freshly prepared and used within 24 hours).
- 0.5 M DTT in 50 mM Ammonium Bicarbonate (must be freshly prepared).
- 0.55 M Iodoacetamide in 50 mM Ammonium Bicarbonate (must be freshly prepared).
- Trypsin solution: Use sequence grade trypsin (*see* **Note 1**).
- ProteaseMax^{TM} Surfactant Enhancer (*see* **Note 2**).

3.2.5 Peptide Fractionation Using Strong Cation Exchange (SCX) spin cartridges

- Diluent solution: To prepare 1 L add 1.36 g of KH_2PO_4 and 250 mL of acetonitrile to 750 mL of LC-MS grade water and adjust the pH to 3.0.
- Prepare the following 10 mL Potassium Chloride (KCl) elution buffers as outlined in Table 3-1 using the 10 mM KH_2PO_4 in 25% acetonitrile pH 3.0 diluent.

Elution Buffer	KCL (g)	10mM KH$_2$PO$_4$ in 25% ACN (mL)	KCl concentration (mM)
Elution buffer step 1	0.007	10	10
Elution buffer step 2	0.037	10	50
Elution buffer step 3	0.075	10	100
Elution buffer step 4	0.373	10	500

Table 3-1. *Overview of preparation of KCl elution buffers.*

3.2.6   Fe-NTA Phosphopeptide Enrichment Buffer Preparation

Wash buffer A and B must be prepared prior to starting the protocol. 300 µL of each wash buffer is required per spin column i.e. sample.

- Wash buffer A: Add 150 µL of the 2X wash buffer stock supplied in the kit to 150 µL of LC-MS grade water.

- Wash buffer B: Add 150 µL of the 2X wash buffer stock supplied in the kit and 30 µL of acetonitrile to 120 µL of LC-MS grade water.

3.2.7   TiO$_2$ Phosphopeptide Enrichment Buffer Preparation

The following buffer preparations are for the enrichment of 1-6 samples, if more samples are being processed scale up the buffer volumes accordingly or see manufacturer's protocol.

- Buffer A: Add 401 µL of ACN and 2 µL of TFA to 100 µL of LC-MS water.

- Buffer B: Add 400 µL of 90% lactic acid (provided in the kit) to 1 mL of Buffer A.

- Elution Buffer 1: Add 20 µL of 30% Ammonium Hydroxide to 380 µL of LC-MS water.

- Elution Buffer 2: Add 20 µL of Pyrrolidine (provided in the kit) to 380 µL of LC-MS water.

3.2.8   Peptide Purification with Pierce™ Graphite Spin Columns

- Column preparation buffer (1 M NH$_4$OH): Add 0.688 mL of Ammonium hydroxide solution ACS reagent (28 -30%) to 2.5 mL of LC-MS water and adjust the final volume of the solution to 10 mL with LC-MS water. Store at 4°C until required.

- Equilibration and rinse solution: 1% TFA. Add 100 µL of TFA to 10 mL of LC-MS water.

- Elution solution: 0.1% TFA, 70% ACN. Add 100 µL of TFA to 3 mL of LC-MS water containing 7 mL of ACN.

3.2.9   Reverse Phase Chromatography Buffers

- Solvent A: 2% ACN in LC-MS grade water containing 0.1% formic acid. Prepare 100 mL. Use a fume hood to prepare this solution.

- Solvent B: 80% ACN, 20% LC-MS grade water containing 0.8% formic acid. Prepare 100 mL.

3.3 Methods

3.3.1 Cell Lysis

1. Harvest cells and wash with pre-chilled 50 mM HEPES buffer three times. Centrifuge at 1000 x*g* for 5 min and remove the supernatant. Snap freeze the cell pellet in a microcentrifuge tube using liquid nitrogen (take extreme caution when doing this). Store at -80°C until cell lysis is performed.

2. Lyse cell pellets corresponding to 1 x 10^7 cells with 1 mL of lysis buffer containing 1x Halt™ Protease Inhibitor cocktail and 1x Halt™ Protease Inhibitor cocktail. Suspend the cell pellet in the lysis buffer and mix the solution by vigorous vortexing for 5 min.

3. Disrupt the cells using a sonicating probe while maintaining the lysate at 4^oC.

4. Centrifuge the sample at 14,000 x*g* for 15 min at 4^oC. Collect supernatant into a fresh microcentrifuge tube.

5. Determine protein concentration using a BCA assay according to manufacturer's instructions.

6. Transfer 500 µg of protein lysate aliquots into microcentrifuge tubes and freeze at -80ºC until required.

7. Prior to digestion, dilute sample with 50 mM ammonium bicarbonate to achieve a final urea concentration of 1 M.

3.3.2 In-Solution Protein Digestion

1. Adjust the volume of 500 µg protein lysate to 84 µL with 50 mM ammonium bicarbonate.

2. Add 1 µL of 0.5 M DTT and heat the lysate at 56°C for 20 min to reduce the protein.

3. Add 2.7 µL of 0.55 M iodoacetamide and incubate at room temperature in the dark for 20 min to allow for alkylation.

4. Add 12.5 µl of 1 µg/µL sequence-grade trypsin (1:40 enzyme:protein).

5. Digest at 37°C for 3 h in a dry heating block under gentle rotation

6. Centrifuge at 14,000 x*g* at 20°C for 20 min and transfer the supernatant to a new microcentrifuge tube.

7. Add TFA to a final concentration of 0.1% to inactivate the trypsin.

8. Proceed directly to peptide concentration and desalting **(subsection 3.3.4)** or store the sample at -80°C until required.

3.3.3   Peptide Concentration and Desalting with Graphite Spin Columns

1. Remove the top and bottom cap from the graphite spin column (*see* **Note 3**) and place the column in a 1.7 mL collection tube. Centrifuge at 2000 x*g* for 1 min and discard the flow-through.

2. Prime the Graphite spin column using 100 µL of 1 M NH$_4$OH. Centrifuge at 2000 x*g* for 1 min and discard the flow-through.

3. Activate the graphite spin column using 100 µL of ACN. Centrifuge at 2000 x*g* for 1 min and discard the flow-through.

4. Apply the peptide sample from **subsection 3.3.2** to the graphite resin and allow sample binding for 10 min with occasional vortexing.

5. Centrifuge at 1000 x*g* for 3 min and discard the flow-through.

6. Wash the graphite spin column using 100 µL of the rinse solution, centrifuge at 2000 x*g* for 1 min and discard rinse wash.

7. Repeat **step 6**.

8. Place the graphite spin column in a new collection tube and elute the peptide sample using 100 µL of elution buffer into a new microcentrifuge tube.

9. Repeat step 8 three more times to give a final volume of 400 µL.

10. Concentrate the peptide sample to dryness in a SpeedVac vacuum dryer.

11. Freeze peptide sample at -80°C or proceed directly to Strong Cation Exchange preparation.

3.3.4   Strong Cation Exchange (SCX) using Pierce™ Mini Spin Columns

1. Condition a SCX spin column with 400 μL of 10 mM KH_2PO_4 in 25% ACN pH 3.0 and centrifuge at 2,000 x*g* for 5 min. Discard the flow-through

2. Resuspend the dried peptides from the final step of **subsection 3.3.3** in 200 μL of 10 mM KH_2PO_4 in 25% ACN pH 3.0. Apply this to the conditioned SCX column and centrifuge at 2,000 x*g* for 5 min.

3. Add 200 μL of elution buffer step 1 (see **subsection 3.2.5**) and centrifuge at 2,000 x*g* for 5 min and collect the flow-through.

4. Repeat this procedure to elute the peptides in a step-wise manner using elution buffer step 1 through to elution buffer step 4, collecting each flow-through generating 5 samples (initial flow through, 10, 50, 100 and 500 mM salt fractions).

5. Eluted peptides are desalted and concentrated again using Graphite spin columns (go to **subsection 3.3.3**).

6. Following graphite spin column desalting and concentration, the SCX fractions peptide samples can either be phosphopeptide enriched using Fe-NTA (IMAC) (go to **subsection 3.3.5**) or TiO_2 (go to **subsection 3.3.6**).

3.3.5   IMAC Phosphopeptide Enrichment

1. Resuspend the dried sample from the SCX fractionation protocol in 200 μL of Binding buffer. Add the sample to a Fe-NTA spin column and incubate for 20 min at room temperature on a rotator.

2. Remove bottom cap from the column and place the column into a microcentrifuge tube and centrifuge at 1,000 x*g* for 1 min. Transfer the column to a new tube.

3. Add 100 μL of wash buffer A to the column and pipette well to mix contents. Centrifuge the tube at 1,000 x*g* for 1 min. Transfer column to a new tube.

4. Repeat step 3.

5. Add 100 μL of wash buffer B to the column and pipette well to mix contents. Centrifuge the tube at 1,000 x*g* for 1 min. Transfer the column to a new tube.

6. Repeat step 5.

7. Equilibrate the column by adding 100 μL of LC-MS water to the column and mix contents by pipetting. Centrifuge the column at 1,000 xg for 1 min. Transfer column to a new tube.

8. Elute the phosphopeptides by adding 50 μL of Elution buffer to the column and incubate for 5 min at room temperature. Centrifuge the column at 1,000 xg for 1 minute and retain the eluate (flow-through).

9. Repeat step 8 and pool the elution fractions.

10. Acidify the elution fraction by adding 200 μL of 2.5% TFA to the sample. Proceed to **subsection 3.3.7**.

3.3.6   TiO$_2$ Phosphopeptide Enrichment

1. Place a centrifuge column adaptor in a collection tube and insert a TiO$_2$ spin tip into the adaptor. Add 20 μL of Buffer A and centrifuge at 3,000 xg for 2 min. Discard the flow-through.

2. Add 20 μL of Buffer B and centrifuge at 3,000 xg for 2 min. Discard the flow-through.

3. Resuspend the dried sample from the SCX fractionation protocol in 150 μL of Buffer B. Add the sample to the spin tip and centrifuge at 1,000 xg for 10 min.

4. Collect the flow-through sample and reapply it to the spin tip and centrifuge at 1,000 xg for 10 min.

5. Wash the column by adding 20 μL of Buffer B and centrifuging at 3,000 xg for 2 min.

6. Wash the column by adding 20 μL of Buffer A and centrifuge at 3,000 xg for 2 min. Repeat this step twice so that the column has been washed three times.

7. Place the spin tip in a new collection tube and add 50 μL of Elution Buffer 1. Centrifuge at 1,000 xg for 5 min.

8. Using the same collection tube, add 50 μL of Elution Buffer 2. Centrifuge at 1,000 xg for 5 min.

9. Acidify the elution fraction by adding 100 μL of 2.5% TFA to the sample. Proceed to **subsection 3.3.7**.

3.3.7   LC-MS/MS analysis

1. Resuspend the dried phosphopeptide samples in 25 µL of LC-MS grade water with 0.1% TFA and 2% ACN.

2. Carry out nano LC–MS/MS analysis using, for example, an Ultimate 3000 RSLCnano system coupled to a hybrid linear ion trap/Orbitrap mass spectrometer. Use SilicaTip™ Standard Coating Tubing as emitter tips for nano electrospray as the formation of phosphopeptide metal ion complexes can cause a mass spectrum signal intensity decrease of protonated phosphopeptides (Liu et al. 2005)*. (*See* **Note 4**).

3. Load each sample onto a C18 trap column. Desalt the sample for 3 min using a flow rate of 25 µL/min in 0.1% TFA containing 2% ACN (Loading Buffer). Switch the trap column online with the analytical column using a column oven at 35 °C and elute the peptides with the following binary gradients of: Mobile Phase Buffer A and Mobile phase buffer B: 0–25% solvent B in 120 min and 25–50% solvent B in a further 60 min, where solvent A consisted of 2% acetonitrile (ACN) and 0.1% formic acid in water and solvent B consisted of 80% ACN and 0.08% formic acid in water. Set the column flow rate to 300 nL/min. Acquire the data with Xcalibur software.

4. Operate the LTQ Orbitrap XL in data-dependent neutral loss mode and externally calibrate the instrument. Acquire survey MS scans in the Orbitrap in the 400–1800 m/z range with the resolution set to a value of 30,000 at m/z 400. Fragment up to three of the most intense ions (1+, 2+ and 3+) per scan using CID in the linear ion trap followed by a Neutral loss MS3 event. MS3 is activated if a neutral loss of phosphoric acid is detected indicated by a -98, -49 or -32.7 Da loss from the parent precursor. Enable dynamic exclusion with a repeat count of 1, repeat duration of 30 s, exclusion list size of 500 and exclusion duration of 40 s. Set the minimum signal threshold to 3000. Collect all tandem mass spectra using a normalised collision energy of 32%, an isolation window of 2 m/z with an activation time of 30 ms.

3.3.8 Mass Spectrometry Data Searches for Phosphorylation Analysis

1. Search MS spectra using Proteome Discoverer 2.1 against Sequest HT with phospho*RS* 3.1 against a suitable *Cricetulus griseus* protein fasta database (*see* **Note 5**).

2. The processing workflow used through Proteome Discoverer 2.1 is shown in Figure 3.1 (*See* **Note 6**).

3. Scan Event Filter(Baycin-Hizal et al. 2012a) and (Guo et al. 2008) (Figure 3.1) selects for MS2 and MS3 data respectively.

4. From Scan Event Filter(Baycin-Hizal et al. 2012a) into Search algorithm SEQUEST HT **(2)** (Figure 3.1), apply the following search parameters for all MS2 data for protein identification: (i) set peptide mass tolerance to 20 ppm, (ii) set MS/MS mass tolerance to 0.6 Da, (iii) allow up to two missed cleavages, (iv) set carbamidomethylation as a fixed modification (i.e. Carboxymethyl (C) means that all calculations will use 161 Da as the mass of cysteine). (v) set methionine oxidation as a variable modification (where the methionine residue may or may not be oxidised (Met +16 Da)), and (vi) set serine, threonine and tyrosine phosphorylation as variable modifications (where the serine, threonine, and tyrosine residues may have the addition of a phosphate group HPO_3 (Ser +80 Da, Thr +80 Da, Tyr +80 Da)). (Figure 2a and Figure.2b).

5. From Scan Event Filter (Guo et al. 2008) into Search algorithm SEQUEST HT (Cohen 2001)(see Figure 3-1), set the following search parameters for all MS3 (neutral loss) data for protein identification: (i) set peptide mass tolerance to 0.6 Da (ii) set MS/MS mass tolerance to 0.6 Da, (iii) allow up to two missed cleavages, (iv) set carbamidomethylation as a fixed modification (v) set methionine oxidation as variable modification (vi) set serine, threonine and tyrosine phosphorylation as variable modifications (vii) and set neutral loss of phosphoric acid H_3PO_4 (-98 Da) from phosphorylated serine, threonine and tyrosine phosphorylation as variable modifications. (See Figure 3-2 c).

6. Process the output from SEQUEST through Percolator (a semi-supervised learning and decoy database search strategy to distinguish between correct and incorrect identifications accepting a 1% FDR)(Spivak et al. 2009).

7. Accept only peptides with XCorr scores >1.5 for singly charged ions, >2.0 for doubly charged ions and >3.0 for triply charged ions.

8. Determine the phosphorylation site location in the peptide sequence through phospho*RS* (ptmRS) in Proteome Discoverer 2.1 (Taus et al. 2011a)*.* Apply a site probability cut off score of 75% or greater for S, T or Y amino acids (Alpert, Hudecz and Mechtler 2015, Roitinger et al. 2015, Taus et al. 2011b).

3.4   Notes

1. Use sequencing-grade or mass spectrometry grade trypsin. Reconstitute the lyophilised trypsin in 50 mM acetic acid to give a final concentration of 1 µg/µL and store aliquots at -20°C until required. In our lab we use Promega's Trypsin Gold Mass Spectrometry grade or Pierce™ Trypsin Protease MS grade for protein digestion.

2. Add 100 µL of 50 mM ammonium bicarbonate to a 1 mg vial of ProteaseMax™ Surfactant to a give a 1% solution. Store aliquots at -20ºC until required. Ammonium Bicarbonate must be freshly prepared and used within 24 h.

3. For peptide purification use graphite spin columns instead of C18 as graphite binds hydrophilic peptides more efficiently than C18.

4. The emitter tip connecting the LC to the MS should not be stainless steel. Applying a voltage across a stainless emitter tip will cause irreversible binding of phosphopeptides.

5. For protein/peptides identification using Chinese Hamster Ovary cells we recommend using either protein fasta databases from Bielefeld BOKU CHO expressed Protein Database (BBCHO)(Meleady et al. 2012b) or from *Cricetulus griseus* fasta sequences available from https://www.ncbi.nlm.nih.gov/protein.

6. Using spectrum selector filter in Proteome Discoverer to separate MS2 and MS3 data allows for significantly shorter search times as MS2 streams can be searched with variable phosphorylation modifications on S, T and Y, while MS3 streams can be searched with variable phosphorylation modifications on S, T, and Y along with dehydration as a variable modification on S, T and Y. (Figure 3-2**)**.

3.5   Figure Legends

Figure 3-1 Processing workflow for MS2 and MS3 data using Proteome Discoverer 2.1.

Figure 3-2 MS/MS of DGQVINTSQHDDLE showing sequence ladder matching glutamine (Q) to serine (S) and has a mass difference of 87 Da confirming unmodified serine. (B) MS/MS of DGQVINTS*QHDDL showing a sequence ladder matching glutamine (Q) to serine (S) and has a mass difference of 167 Da (addition of phosphate group (HPO3) confirming phosphorylated serine (+80 ΔM on Serine). (C) MS/MS/MS of DGQVINTS*QHDDL showing a neutral loss event on serine where the sequence ladder matching glutamine (Q) to serine (S) has a mass difference of -18 Da which corresponds to the loss of phosphoric acid from the phosphorylated residue to produce a di-dehydroamino acid.

Figure 1

Figure 3-1. Processing workflow for MS2 and MS3 data using Proteome Discoverer 2.1.

Fig. 2a

#1	Seq.	y"
1	D	
2	G	1721.77253
3	Q	1664.75107
4	V	1536.69249
5	I	1437.62408
6	N	1324.54001
7	E	1210.49708
8	T	1081.45449
9	S	980.40681
10	Q	893.37479
11	H	765.31621
12	H	628.25730
13	D	491.19838
14	D	376.17144
15	L	261.14450
16	E	148.06043

Fig. 2b

#1	Seq.	y"
1	D	
2	G	1801.73886
3	Q	1744.71740
4	V	1616.65882
5	I	1517.59041
6	N	1404.50634
7	E	1290.46342
8	T	1161.42082
9	S-Phospho	1060.37314
10	Q	893.37479
11	H	765.31621
12	H	628.25730
13	D	491.19838
14	D	376.17144
15	L	261.14450
16	E	148.06043

Fig. 2c

A	B	C
#1	Seq.	y"
1	D	
2	G	1703.76197
3	Q	1646.74050
4	V	1518.68193
5	I	1419.61351
6	N	1306.52945
7	E	1192.48652
8	T	1063.44393
9	S-Dehydrated	962.39625
10	Q	893.37479
11	H	765.31621
12	H	628.25730
13	D	491.19838
14	D	376.17144
15	L	261.14450
16	E	148.06043

Figure 3-2 MS/MS of DGQVINTSQHDDLE showing sequence ladder matching glutamine (Q) to serine (S) and has a mass difference of 87 Da confirming unmodified serine. (B) MS/MS of DGQVINTS*QHDDL showing a sequence ladder matching glutamine (Q) to serine (S) and has a mass difference of 167 Da (addition of phosphate group (HPO3) confirming phosphorylated serine (+80 ΔM on Serine). (C) MS/MS/MS of DGQVINTS*QHDDL showing a neutral loss event on serine where the sequence ladder matching glutamine (Q) to serine (S) has a mass difference of -18 Da which corresponds to the loss of phosphoric acid from the phosphorylated residue to produce a di-dehydroamino acid.

128

3.6   References

Alpert, A.J., Hudecz, O. and Mechtler, K. 2015. Anion-exchange chromatography of phosphopeptides: Weak anion exchange versus strong anion exchange and anion-exchange chromatography versus electrostatic repulsion-hydrophilic interaction chromatography. *Analytical Chemistry,* 87(9), pp.4704-4711.

Altamirano, C., Paredes, C., Cairo, J.J. and Godia, F. 2000. Improvement of CHO cell culture medium formulation: Simultaneous substitution of glucose and glutamine. *Biotechnology Progress,* 16(1), pp.69-75.

Baycin-Hizal, D., Tabb, D.L., Chaerkady, R., Chen, L., Lewis, N.E., Nagarajan, H., Sarkaria, V., Kumar, A., Wolozny, D., Colao, J., Jacobson, E., Tian, Y., O'Meally, R.N., Krag, S.S., Cole, R.N., Palsson, B.O., Zhang, H. and Betenbaugh, M. 2012. Proteomic analysis of chinese hamster ovary cells. *Journal of Proteome Research,* 11(11), pp.5265-5276.

Brinkrolf, K., Rupp, O., Laux, H., Kollin, F., Ernst, W., Linke, B., Kofler, R., Romand, S., Hesse, F., Budach, W.E., Galosy, S., Muller, D., Noll, T., Wienberg, J., Jostock, T., Leonard, M., Grillari, J., Tauch, A., Goesmann, A., Helk, B., Mott, J.E., Puhler, A. and Borth, N. 2013. Chinese hamster genome sequenced from sorted chromosomes. *Nature Biotechnology,* 31(8), pp.694-695.

Choudhary, C. and Mann, M. 2010. Decoding signalling networks by mass spectrometry-based proteomics. *Nature Reviews.Molecular Cell Biology,* 11(6), pp.427-439.

Cohen, P. 2001. The role of protein phosphorylation in human health and disease. the sir hans krebs medal lecture. *European Journal of Biochemistry,* 268(19), pp.5001-5010.

Dephoure, N., Gould, K.L., Gygi, S.P. and Kellogg, D.R. 2013. Mapping and analysis of phosphorylation sites: A quick guide for cell biologists. *Molecular Biology of the Cell,* 24(5), pp.535-542.

Farrell, A., McLoughlin, N., Milne, J.J., Marison, I.W. and Bones, J. 2014. Application of multi-omics techniques for bioprocess design and optimization in chinese hamster ovary cells. *Journal of Proteome Research,* 13(7), pp.3144-3159.

Farrell, A.S., Allen-Petersen, B., Daniel, C.J., Wang, X., Wang, Z., Rodriguez, S., Impey, S., Oddo, J., Vitek, M.P., Lopez, C., Christensen, D.J., Sheppard, B. and Sears, R.C. 2014. Targeting inhibitors of the tumor suppressor PP2A for the treatment of pancreatic cancer. *Molecular Cancer Research : MCR,* 12(6), pp.924-939.

Guo, Y., Xiao, P., Lei, S., Deng, F., Xiao, G.G., Liu, Y., Chen, X., Li, L., Wu, S., Chen, Y., Jiang, H., Tan, L., Xie, J., Zhu, X., Liang, S. and Deng, H. 2008. How is mRNA expression predictive for protein expression? A correlation study on human circulating monocytes. *Acta Biochimica Et Biophysica Sinica,* 40(5), pp.426-436.

Gutierrez, J.M. and Lewis, N.E. 2015. Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. *Biotechnology Journal,* 10(7), pp.939-949.

Hopkins, B.D., Hodakoski, C., Barrows, D., Mense, S.M. and Parsons, R.E. 2014. PTEN function: The long and the short of it. *Trends in Biochemical Sciences,* 39(4), pp.183-190.

Kaas, C.S., Kristensen, C., Betenbaugh, M.J. and Andersen, M.R. 2015. Sequencing the CHO DXB11 genome reveals regional variations in genomic stability and haploidy. *BMC Genomics,* 16pp.160-015-1391-x.

Kildegaard, H.F., Baycin-Hizal, D., Lewis, N.E. and Betenbaugh, M.J. 2013. The emerging CHO systems biology era: Harnessing the 'omics revolution for biotechnology. *Current Opinion in Biotechnology,* 24(6), pp.1102-1107.

Lewis, N.E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., Bordbar, A., Roth, A.M., Rosenbloom, J., Bian, C., Xie, M., Chen, W., Li, N., Baycin-Hizal, D., Latif, H., Forster, J., Betenbaugh, M.J., Famili, I., Xu, X., Wang, J. and Palsson, B.O. 2013. Genomic landscapes of chinese hamster ovary cell lines as revealed by the cricetulus griseus draft genome. *Nature Biotechnology,* 31(8), pp.759-765.

Liu, S., Zhang, C., Campbell, J.L., Zhang, H., Yeung, K.K., Han, V.K. and Lajoie, G.A. 2005. Formation of phosphopeptide-metal ion complexes in liquid chromatography/electrospray mass spectrometry and their influence on phosphopeptide detection. *Rapid Communications in Mass Spectrometry : RCM,* 19(19), pp.2747-2756.

Martini, M., De Santis, M.C., Braccini, L., Gulluni, F. and Hirsch, E. 2014. PI3K/AKT signaling pathway and cancer: An updated review. *Annals of Medicine,* 46(6), pp.372-383.

Meleady, P., Hoffrogge, R., Henry, M., Rupp, O., Bort, J.H., Clarke, C., Brinkrolf, K., Kelly, S., Muller, B., Doolan, P., Hackl, M., Beckmann, T.F., Noll, T., Grillari, J., Barron, N., Puhler, A., Clynes, M. and Borth, N. 2012. Utilization and evaluation of CHO-specific sequence databases for mass spectrometry based proteomics. *Biotechnology and Bioengineering,* 109(6), pp.1386-1394.

Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., Brunak, S. and Mann, M. 2010. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Science Signaling,* 3(104), pp.ra3.

Prentice, H.L., Ehrenfels, B.N. and Sisk, W.P. 2007. Improving performance of mammalian cells in fed-batch processes through "bioreactor evolution". *Biotechnology Progress,* 23(2), pp.458-464.

Roitinger, E., Hofer, M., Kocher, T., Pichler, P., Novatchkova, M., Yang, J., Schlogelhofer, P. and Mechtler, K. 2015. Quantitative phosphoproteomics of the ataxia telangiectasia-mutated (ATM) and ataxia telangiectasia-mutated and rad3-related (ATR) dependent DNA damage response in arabidopsis thaliana. *Molecular & Cellular Proteomics : MCP,* 14(3), pp.556-571.

Simon, G.M. and Cravatt, B.F. 2008. Challenges for the 'chemical-systems' biologist. *Nature Chemical Biology,* 4(11), pp.639-642.

Spivak, M., Weston, J., Bottou, L., Kall, L. and Noble, W.S. 2009. Improvements to the percolator algorithm for peptide identification from shotgun proteomics data sets. *Journal of Proteome Research,* 8(7), pp.3737-3745.

Taus, T., Kocher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C. and Mechtler, K. 2011. Universal and confident phosphorylation site localization using phosphoRS. *Journal of Proteome Research,* 10(12), pp.5354-5362.

Thingholm, T.E., Jensen, O.N., Robinson, P.J. and Larsen, M.R. 2008. SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Molecular & Cellular Proteomics : MCP,* 7(4), pp.661-671.

Walsh, G. 2014. Biopharmaceutical benchmarks 2014. *Nature Biotechnology,* 32(10), pp.992-1000.

Whitmarsh, A.J. and Davis, R.J. 2000. Regulation of transcription factor function by phosphorylation. *Cellular and Molecular Life Sciences : CMLS,* 57(8-9), pp.1172-1183.

Wurm, F.M. 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotechnology,* 22(11), pp.1393-1398.

Xu, X., Nagarajan, H., Lewis, N.E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., Andersen, M.R., Neff, N., Passarelli, B., Koh, W., Fan, H.C., Wang, J., Gui, Y., Lee, K.H., Betenbaugh, M.J., Quake, S.R., Famili, I., Palsson, B.O. and Wang, J. 2011. The genomic sequence of the chinese hamster ovary (CHO)-K1 cell line. *Nature Biotechnology,* 29(8), pp.735-741.

Novel panel of protein biomarkers to predict response to bortezomib-containing induction regimens in multiple myeloma patients

Authors: Kay Reen Ting, <u>Michael Henry,</u> Justine Meiller, Annemarie Larkin, Martin Clynes, Paula Meleady, Despina Bazou, Paul Dowling and Peter O'Gorman

Within this work, Michael Henry contributed to the methodology of the two step serum depletion strategy. Michael specifically performed the in-solution protein digests and LC-MS/MS analysis on all the protein samples. Michael Henry implemented the differential expression analysis tools using label free software Progensis QI and searched the statistically significant protein features using MS search algorithms. Michael Henry prepared the protein differential data results. In the writing of the publication prepared the proteomic sections in the materials and methods and results section.

Abstract

Multiple myeloma (MM) is a complex heterogeneous disease. Various risk stratification models have been recommended including cytogenetic and FISH analysis to identify high-risk patients who may benefit from novel treatments, but such facilities are not widely available. The International Scoring System (ISS) using beta-2-microglobulin and albumin remains a widely used prognostic scoring system in many clinical practices; however, it is not useful in predicting response to treatment in MM. The aim of this study is to identify clinically useful biomarkers to predict response to treatment containing bortezomib.

Methods: 17 MM patient serum samples (9 responders/8 non-responders) were used for the discovery phase (label-free mass spectrometry) and an additional 20 MM patient serum samples were used for the ELISA-based validation phase (14 responders/6 non-responders).

Results: CLU and ANG mean levels were higher in the responders group, while Complement C1q had lower concentrations. The combination of all standard biomarkers (albumin, beta-2-microglobulin (ß2M), para protein and kappa/lambda (K/L) ratio) had an AUC value of 0.71 with 65% correct classification, while an overall combination of new candidate protein biomarkers with standard biomarkers had an AUC value of 0.89 with 85.3% correct classification.

Conclusions: A combination of new and standard biomarkers consisting of CLU, ANG, C1Q, albumin, ß2M, paraprotein and K/L ratio may have potential as a novel panel of biomarkers to predict MM response to treatment containing bortezomib.

General significance: Use of this biomarker panel could facilitate a more personalised therapy approach and to minimize unnecessary side effects from ineffective drugs.

4.1 Introduction

Multiple Myeloma (MM) is a plasma cell disorder characterised by bone marrow infiltration with clonal plasm a cells, which secrete monoclonal immunoglobulin detectable in serum and/or urine. The development of novel targeted therapies has markedly improved the response rate and survival outcome, but MM remains incurable (N. Kumar et al. 2008). Bortezomib, which was the first proteasome inhibitor anticancer drug, is one of the many novel agents being used. It has numerous functionalities such as inducing apoptosis and growth arrest in the myeloma cell cycle, altering the bone marrow micro-environment, inhibiting nuclear factor kappa B and reversing chemoresistance in myeloma cells (Field-Smith, Morgan and Davies 2006, Mitsiades et al. 2003). It also demonstrates an inhibitory effect on angiogenesis and DNA repair but also has no permanent impact on normal hematopoietic stem cells.

Following the APEX and SUMMIT trials, bortezomib monotherapy may overcome relapsed and refractory MM patients with poorer prognosis (Jagannath et al. 2007, Richardson et al. 2006). Bortezomib is now established as the backbone agent for combined induction therapy regimens (Cavo et al. 2010, Richardson et al. 2010, Jagannath et al. 2005).

Diagnostic criteria and frontline standards guidelines have been shown to also play an important role in deciding the outcome of MM disease and individualising treatment (Fonseca et al. 2009, Fonseca et al. 2004). Unfortunately, only a minority of centers has such extensive access to molecular and genetic sequencing studies and imaging facilities that will provide a better stratification of patients according to their disease burden (Inamoto et al. 2009).

While novel agents have improved treatment outcomes, identification of biomarkers that will facilitate clinicians in determining which treatment is optimum for high-risk patients following initial diagnosis is a crucial area in the clinical management of MM. Quantitative measurement of plasma/serum and urinary protein concentrations plays a significant role in the monitoring of patients with MM. In 2005, an international consortium of researchers used serum β2M and serum albumin to create the International Staging System (ISS), which enabled clinicians to stage patients and predict their long-term prognosis, a process that has been refined with the addition of fluorescent in situ hybridization (FISH) data (Avet-Loiseau et al. 2013, Greipp et al. 2005). The serum free light chain assay also plays an important role in prognosis, with the International Myeloma

Working Group identifying several uses for serum free light chain analysis in MM (Dispenzieri et al. 2009, van Rhee et al. 2007, Durie et al. 2006).

In view of the limited and restricted access to facilities such as advanced imaging and genetic testing, identifying the most specific and reliable prognostic biomarker remains a great challenge. We have focused specifically on identifying a panel of protein biomarkers that could be used to predict response following combined agent use in newly diagnosed MM patients. The ability to combine routine laboratory measurements, for example β2M, with newly discovered candidate biomarkers for predicting response to bortezomib therapy was an important consideration. In this investigation, all patients were naïve to bortezomib therapy at the time of sampling in order to facilitate the identification of predictive biomarker of treatment response. Some MM patients have de novo resistance to bortezomib therapy, with the identification of novel biomarkers of early resistance critical for their clinical management (Fall et al. 2014).

The availability of predictive biomarkers would be useful in avoiding ineffective treatments, and allow for the administration of alternative regimens which are continuing to be approved for the treatment of MM. Ineffective treatments can also be associated with undesired side effects, such as peripheral neuropathy and thrombocytopenia, which are important to avoid particularly if no benefit is associated with the specific treatment.

4.2   Material and methods

4.2.1   Patients' selection and sample collection

Thirty seven newly diagnosed MM patients from 2011 to 2013 were selected for this study as shown in Table 4-1. They were naïve to bortezomib therapy. These patients were stratified to either responders or non-responders group according to the International Myeloma Working Group (IMWG) uniform response criteria for MM (Durie et al. 2006). Responders were considered Complete Response (CR), Very Good Partial Response (VGPR) and Partial Response (PR) with non-responders considered Progressive Disease (PD) and Stable Disease (SD) as determined using the International Myeloma Working Group Uniform Response Criteria.

136

Seventeen patient samples were used in the discovery phase and all samples were used in the validation phase.

Serum samples were obtained at diagnosis or prior to commencing therapy. The participating subjects gave written informed consent in accordance with the Declaration of Helsinki that was approved by local ethics committees. These samples were collected according to standard phlebotomy procedures. 10 ml of blood was collected into additive free (serum) blood tubes and was allowed to clot for 30 min to 1 h at room temperature. The serum was denuded by pipette from the clot and aliquoted into a clean tube. The tubes were centrifuged at 400 ×g for30 min at 4 °C. Serum was aliquoted in cryovial tubes, labelled and stored at −80 °C until time of analysis. The time from sample procurement to storage at −80 °C was < 3 h.

4.2.2   Serum protein sample fractionation and enzymatic digestion

The Proteome Purity™ 12 Human Serum Protein Immunodepletion Resin from R & D Systems, United Kingdom was selected to remove high abundance protein. It depletes the 12 most abundant proteins (alpha 1- Acid Glycoprotein, alpha 1-Antitrypsin, alpha 2-Macroglobulin,Albumin, Apolipoprotein A-I, Apolipoprotein A-II, Fibrinogen, Haptoglobin, IgA, IgG, IgM, Transferrin) from the serum as described by the manufacturer's protocol (Millioni et al. 2011).

100 µL of the Blue Nanotrap particles (RB4VSA) from Ceres Nano was added to each prepared immunodepleted sample that contained Tris-HCl buffer. These particles were resuspended into the prepared samples and allowed to incubate for 30 min at room temperature. The suspension was centrifuged at 16,800 ×g (Hettich Mikro 200R, United Kingdom) for 10 min at room temperature and the supernatant was carefully removed from each sample and transferred to individually labelled clean micro-centrifuge tubes. The pelleted particles were resuspended in 500 µL of LC-MS grade water. Once more, the particles were centrifuged at 16,800 ×g for 10 min at room temperature. The supernatant was removed and discarded. A total of 3 washes with the Ultra High Purity Water were performed leaving a pellet particle to be resuspended in an eluent.

100 µL of freshly made protein elution buffer (70% acetonitrile (ACN) from Sigma-Aldrich/10% ammonium hydroxide from Sigma-Aldrich) was added to the particles before vortexing and then

sonicating using the waterbath method (Precision ultrasonic cleaning, DP201–00, Ultrawave, UK) for 2 min. These samples were centrifuged (Hettich Mikro 200R, United Kingdom) at 16,800 ×g for 7 min at room temperature and the eluted proteins obtained were carefully removed and transferred to a clean microcentrifuge tube. The protein elution step was repeated for a total of 200 μL. The combined eluent, collected in the microcentrifuge tubes, was vacuum dried at room temperature. Samples were then digested with trypsin overnight according to standard procedures (Dowling, Hayes et al. 2014).

4.2.3   Nano HPLC and mass spectrometry analysis

The nano LC-MS/MS analysis of responders versus non-responders patient samples was carried out using an Ultimate 3000 nanoLC system (Dionex) coupled to an LTQ XL Orbitrap mass spectrometer (Thermo Fisher Scientific, Dublin, Ireland) in the Proteomics Facility of the National Institute for Cellular Biotechnology, Dublin City University. Peptide mixtures (5 μl volume) were loaded onto a C18 trap column (C18 PepMap, 300 μm id × 5 mm, 5 μm particle size, 100 Å pore size; Dionex). Desalting was achieved at a flow rate of 25 μl/min in 0.1%TFA for 3 min. The trap column was switched on-line with an analytical PepMap C18 column (75 μm id × 500 mm, 3 μm particle and 100 Å pore size; Dionex). Peptides generated from the digestion were eluted with the following binary gradients: solvent A (2% ACN and 0.1%formic acid in LC-MS grade water) and 0–25% solvent B (80% ACN and 0.08% formic acid in LC-MS grade water) for 160 min and 25–50% solvent B for a further 20 min. The column flow rate was set to 350 nl/min. Data was acquired with Xcalibur software, version 2.0.7 (Thermo Fisher Scientific). The MS apparatus was operated in data-dependent mode. Survey MS scans were acquired in the Orbitrap in the 400–1200 m/z range with the resolution set to a value of 30,000 at m/z 400 and lock mass set to 445.120025 m/z. CID fragmentation was carried out in the linear ion trap with the three most intense ions per scan. Within 40 s, a dynamic exclusion window was applied. Normalised collision energy of 35%, an isolation window of 2 m/z and one microscan were used to collect suitable tandem mass spectra.

4.2.4 Quantitative protein profiling by label-free LC-MS/MS analysis

Processing of the raw data generated from LC-MS/MS analysis was carried out with Progenesis QI for Proteomics label-free LC-MS software (version 3.1; Non-Linear Dynamics, Newcastle upon Tyne, UK). Data alignment was based on the LC retention time of each sample, allowing for any drift in retention time given and adjusted retention time for all runs in the analysis. A reference run was established with the sample run that yielded most features (i.e. peptide ions). The retention times of all of the other runs were aligned to this reference run and peak intensities were then normalised. Prior to exporting the MS/MS output files to Proteome Discoverer 1.4 (Thermo Fisher Scientific) for protein identification, a number of criteria were employed to filter the data including: (i) peptide features with ANOVA < 0.05 between experimental groups, (ii) mass peaks (features) with charge states from +2, +3, and (iii) greater than one isotope per peptide. The exported MS/MS file from Progenesis label-free LC-MS software was searched using a dual algorithm search with both MASCOT and SEQUEST in ProteomeDiscoverer 1.4 against the human UniSwiss Prot database (taxonomy:Homo sapiens). The following search parameters were used for protein identification: (i) MS/MS mass tolerance set at 0.6 Da, (ii) peptide mass tolerance set to 20 ppm, (iii) carbamidomethylation set as a fixed modification, (iv) up to two missed cleavages were allowed and (v) methionine oxidation set as a variable modification. The following criteria were applied to assign positively identified proteins: (i) an ANOVA score between experimental groups of ≤ 0.05, (ii) proteins with ≥ 1 peptides matched and (iii) a MASCOT score > 40.

4.2.5 Enzyme-linked immunosorbent assays (ELISA)

ELISA kits were used to verify the different levels for selected proteins identified from the discovery phase using label free mass spectrometry analysis. Four potential biomarkers were chosen and were validated using raw unfractionated serum samples from the same cohort of patients. Four commercially available kits for these four proteins; angiogenin (ANG) [Abcam, UK - ab99970], clusterin (CLU)[R & D system, UK - DCLU00], C-C Motif Chemokine 18 (CCL18) [Abcam, UK - ab100620], and Complement C1q [Abcam, UK -ab170246] were used. Duplicate and triplicate serum samples were used during this analysis. Each of these ELISA assays was performed according to their individual manufacturer's protocol and guidelines. The

concentration of each protein in the serum samples was measured by comparing the optical density (OD) using a microplate reader (Bio-Tek and Luminex). Standard curves were calculated for each ELISA kit.

4.2.6 Statistical analysis

Proteins differentially expressed in responder and non-responder groups underwent statistical analyses. Multivariate logistic regression (LR) and receiver operating characteristic (ROC) curve analysis were selected to interpret the various protein combinations. The ROC plots were obtained by plotting all sensitivity values (true positive fraction) on the y-axis against their equivalent (1-specificity) values (false positive fraction) for all available thresholds on the x-axis (MedCalc for Windows 8.1.1.0, Medcalc Software, Mariakerke, Belgium). The probability of correctly predicting a given model was calculated from the ROC curve by determining the area under curve (AUC)(R. Kumar and Indrayan 2011, Zou, O'Malley and Mauri 2007). In our study, we considered AUC values ranging from $0.5 \rightarrow 0.7$ as poor, $0.7 \rightarrow 0.8$ as average, $0.8 \rightarrow 0.9$ as good and > 0.9 as outstanding. Proteins and any combination of proteins providing an AUC value >0.8 was deemed to be effective for the discrimination of responders from non-responders (Hosmer et al. 1997).

4.3 Results

4.3.1 Sample group

The study group consists of 14 females and 23 males; mean age of patients was 66.5 ± 10.5 years (range 44–87 years). As per International Myeloma Working Group (IMWG) uniform response criteria, 14 patients were identified as non-responders and 23 patients were responders to induction therapy containing bortezomib (Figure 4-1 and Table 4-1). Age was closely matched for both groups with responders mean age of 65 ± 9.5 years (range 44– 80 years) and non-responders mean age of 68.7 ± 11.9 years (range 46–87 years). As ISS classification remains important in providing prognostic outcome of MM disease, patients were also classified to each respective stages; I (n = 11 cases), II (n = 17 cases) and III (n = 9 cases).

4.3.2 Proteomic profiling and identification

16 proteins showing differential expression with significant *p*-values and fold changes were identified using Progenesis QI for Proteomics. Four proteins (ANG, CLU, CCL18 & C1Q) were selected for further validation based on the availability of commercial ELISA kits to evaluate the abundance of these proteins in the validation cohort Table 4-2. ANG, CLU, CCL18 & C1Q also had the following criteria: p-value ≤0.05, protein fold change ≥1.5 and proteins with ≥2 peptides matched.

4.3.3 ELISA data and ROC curve analysis

Standard markers such as albumin, ß2M, paraprotein and K/L ratio that are routinely used in clinical practice were measured and compared with the new candidate biomarkers. CLU, ANG, C1Q and CCL18levels were compared between the responder and non-responder patients Table 4-3. A similar analytical approach was applied to this cohort of patient as described in our previous study with thalidomide-based therapy samples (Rajpal et al. 2011).

CLU showed higher levels in the responders compared to non-responders with a mean of 291.123 ± 243.13 ng/mL and181.760 ± 208.096 ng/mL respectively; corresponding with the

discovery phase pattern observed by LC-MS analysis. The ROC curve from this ELISA data was found to have an AUC of 0.651 (p = 0.129). ANG levels were also significantly higher in the responders compared to non-responders group with a mean of 14.383 ± 0.436 ng/mL and13.839 ± 0.635 ng/mL respectively. ROC curve analysis for ANG had an AUC value of 0.748 (p = 0.012). No significant difference in the mean and AUC for CCL18 protein was observed.

C1Q was found to have higher levels in the non-responders com-pared to responders group that was also observed in the discovery phase. Mean levels for non-responders were 249.170 ± 190.604 µg/L and 149.675 ± 138.883 µg/L in the responders group. An AUC of 0.722 with a p-value of 0.032 was calculated from the ROC curve

Albumin and K/L ratio mean abundance levels were higher in the responders group as shown in Table 4-3. ß2M and paraprotein mean levels were higher in the non-responders group. Paraprotein was the only protein to have a significant p-value (0.027) and an associated AUC value of 0.719.

In clinical settings, patients would generally have several different markers measured to assess their disease burden and progression. To identify a predictive panel of protein biomarkers, the standard biomarkers consisting of albumin, ß2M, paraprotein and K/L ratio were assessed initially. A combination of proteins that are used alone in the ISS (albumin, ß2M) had an AUC of 0.66 with 67.6% correct classification. By combining albumin and ß2M with paraprotein and K/L ratio, the AUC increased to 0.708 with 64.9% correct classification. Based upon this analysis, further combinations of proteins were performed to identify a better predictive model.

Four new combinations of biomarkers from this study were found to be successful in predicting response to bortezomib therapy Table 4-4.These combinations consist of either all four new candidate biomarkers (ANG, CLU, C1Q and CCL18) with an AUC = 0.850 and 76.5% correct classification, or three new candidate biomarkers without CCL18,AUC = 0.802 and 76.5% correct classification. The other two combinations include all four standard biomarkers that are routinely used in most clinical settings with either 3 new candidate biomarkers (ANG, CLU and C1Q) with an AUC = 0.890 and 85.3% correct classification or 4 new candidate biomarkers including CCL18, AUC = 0.905 and 82.4% correct classification.
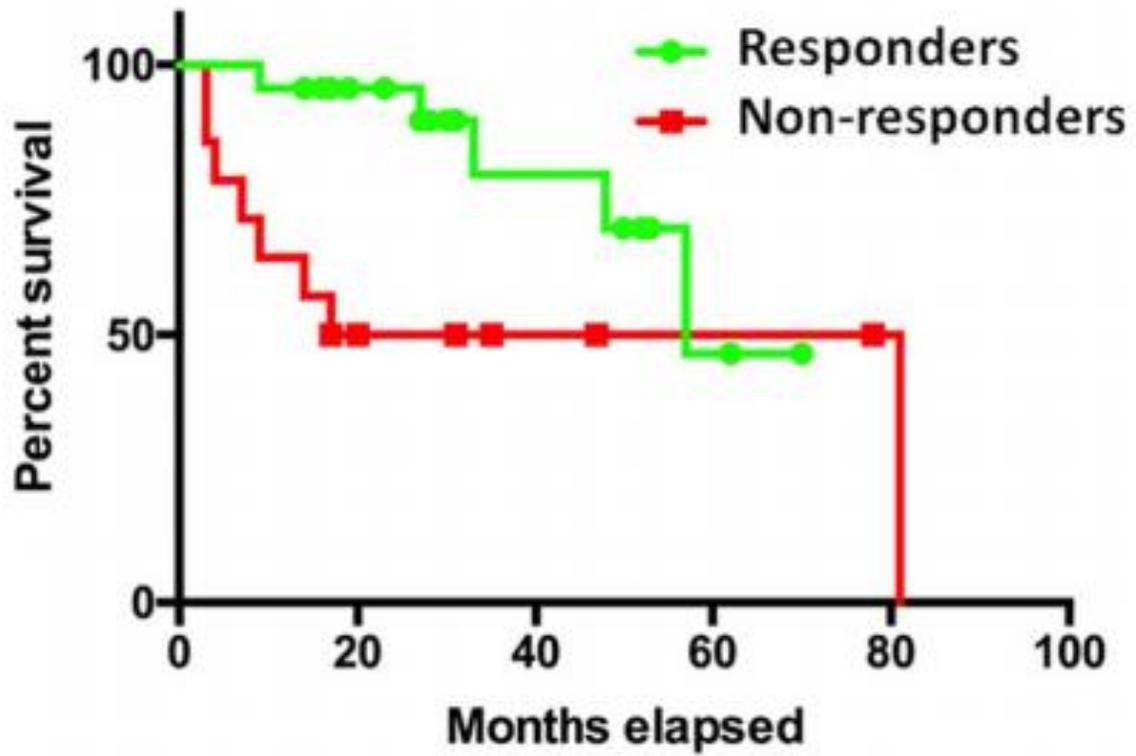
Figure 4-1. Survival data. Overall survival for responders and non-responders patients following induction therapy containing bortezomib

Table 4-1. Patient clinical details. Patient details used in the study with 23 responders and 14 non-responders

Code	Gender	Age	ISS	Immunofixation	*Predicted response*	*Bortezomib*	IMIDs	Overall survival (mo)
1	Female	65	3	IgG Lambda	*Responder*	Y		17.00
2	Male	54	1	Kappa Light Chain	*Responder*	Y	Y	27.00
3	Female	61	1	IgG Kappa	*Responder*	Y		57.00
4	Female	60	2	Kappa Light Chain	*Responder*	Y	Y	19.00
5	Male	51	1	Lambda Light Chain	*Responder*	Y	Y	28.00
6	Male	70	2	IgG Kappa	*Responder*	Y	Y	28.00
7	Male	69	1	IgA Kappa	*Responder*	Y	Y	50.00
8	Male	56	3	Kappa Light Chain	*Responder*	Y	Y	17.00
9	Female	75	2	IgA Lambda	*Responder*	Y		53.00
10	Male	58	2	IgA Kappa	*Responder*	Y	Y	27.00
11	Female	78	3	Lambda Light Chain	*Responder*	Y	Y	31.00
12	Female	64	2	IgA Lambda	*Responder*	Y		31.00
13	Male	76	2	Kappa Light Chain	*Responder*	Y	Y	62.00
14	Female	64	2	IgA Kappa	*Responder*	Y		70.00
15	Male	80	3	IgG Kappa	*Responder*	Y	Y	30.00
16	Male	67	2	IgG Kappa	*Responder*	Y		48.00
17	Female	54	1	IgG Kappa	*Responder*	Y		16.00
18	Male	64	2	IgG Kappa	*Responder*	Y	Y	50.00

19	Male	44	1	IgG Kappa	*Responder*	Y	Y	52.00
20	Male	64	2	IgG Lambda	*Responder*	Y	Y	23.00
21	Male	68	2	IgG Lambda	*Responder*	Y	Y	33.00
22	Male	77	3	IgA Kappa	*Responder*	Y	Y	9.00
23	Male	78	3	IgA Kappa	*Responder*	Y	Y	14.00
24	Male	66	3	IgG Kappa	*Non-responder*	Y	Y	17.00
25	Female	55	1	IgG lambda	*Non-responder*	Y	Y	3.00
26	Female	75	2	IgG Lambda	*Non-responder*	Y	Y	9.00
27	Male	83	2	IgG Lambda	*Non-responder*	Y		7.00
28	Female	58	2	IgA Kappa	*Non-responder*	Y		81.00
29	Male	76	2	IgG Kappa	*Non-responder*	Y		3.00
30	Male	65	2	IgA Kappa	*Non-responder*	Y		4.00
31	Male	87	2	IgA Kappa	*Non-responder*	Y	Y	47.00
32	Male	76	3	IgA Lambda	*Non-responder*	Y		14.00
33	Female	83	2	IgA Lambda	*Non-responder*	Y	Y	35.00
34	Male	67	2	IgG Kappa	*Non-responder*	Y		17.00
35	Male	67	3	IgG Lambda	*Non-responder*	Y		31.00
36	Female	45	1	IgG Kappa	*Non-responder*	Y	Y	20.00
37	Female	58	1	IgG Lambda	*Non-responder*	Y		78.00

Table 4-2. Label-free mass spectrometry data. List of statistically significant discovered proteins using LC-MS analysis. Proteins (in bold) were selected for further validation using ELISA.

Accession	Peptide count	Confidence score	Anova (_p_)	Max fold change	Highest mean condition	Lowest mean condition	Protein Description
P10909	**3**	**139.15**	**0.003**	**1.87**	**Responders**	**Non-responders**	**Clusterin**
P01024	3	100.75	0.01	1.85	Responders	Non-responders	Complement C3
P08603	10	369.28	0.02	2.75	Responders	Non-responders	Complement factor H
P01009	3	71.37	0.02	2.09	Responders	Non-responders	Alpha-1-antitrypsin
P55774	**2**	**106.68**	**0.02**	**2.26**	**Responders**	**Non-responders**	**C-C motif chemokine 18**
P10720	2	103.88	0.02	2.53	Responders	Non-responders	Platelet factor 4 variant
P02790	2	118.53	0.03	1.58	Non-responders	Responders	Hemopexin
Q14624	3	125.97	0.03	3.62	Responders	Non-responders	Inter-alpha-trypsin inhibitor heavy chain H4
P03950	**3**	**92.77**	**0.03**	**1.86**	**Responders**	**Non-responders**	**Angiogenin**
P02647	1	49.62	0.03	1.56	Responders	Non-responders	Apolipoprotein A
Q03591	6	208.05	0.03	1.93	Responders	Non-responders	Complement factor H-related protein 1
P00751	1	68.05	0.04	1.73	Responders	Non-responders	Complement factor B
P01031	1	71.55	0.04	4.15	Responders	Non-responders	Complement C5
P00747	1	46.69	0.04	2.13	Responders	Non-responders	Plasminogen
P02745	**2**	**64.53**	**0.05**	**2.96**	**Non-responders**	**Responders**	**Complement C1q subcomponent subunit A**
Q9BXR6	1	63.61	0.05	2.32	Responders	Non-responders	Complement factor H-related protein 5

Table 4-3. ELISA data. Mean, SD, Area under the curve (AUC) and p-value for each of the new and standard proteins found in the groups of patients compared.

Proteins		n	Mean	Std. Deviation	AUC	*p*-value
Candidate biomarkers						
Clusterin (ng/mL)	Responders	23	291.12	243.13	0.651	0.129
	Non-responders	14	181.76	208.1		
Complement C1q (µg/mL)	Responders	21	149.68	138.88	0.722	0.032
	Non-responders	13	249.17	190.6		
Angiogenin (ng/mL)	Responders	23	14.38	0.44	0.748	0.012
	Non-responders	14	13.84	0.64		
C-C motif ligand 18 (ng/mL)	Responders	23	7	0.4	0.549	0.633
	Non-responders	13	6.96	0.41		
Standard biomarkers						
Albumin (g/L)	Responders	23	33.04	7.11	0.672	0.082
	Non-responders	14	28.21	7.68		
Beta 2 Microglobulin (mg/L)	Responders	23	3.45	2.24	0.565	0.511
	Non-responders	14	4.39	3.16		
Paraprotein (g/L)	Responders	23	16.13	19.12	0.719	0.027
	Non-responders	14	29.07	18.34		
Kappa/Lambda ratio	Responders	23	176.33	425.43	0.512	0.908
	Non-responders	14	102.09	257.69		

Table 4-4. Logistic regression analysis data. List of different protein combinations used to establish the best model that can be used as a predictive panel for response to induction therapy containing bortezomib regime.

Protein combinations	AUC	Hosmer & Lemeshow test (*p*-value)	Correct classification (%)
International Scoring System: albumin and ß2M	0.66	0.702	67.6
Standard biomarkers: albumin, ß2M, paraprotein and K/L Ratio	0.708	0.75	64.9
Candidate biomarkers: Angiogenin, Clusterin, Complement C1q	0.802	0.76	76.5
Candidate biomarkers: Angiogenin, Clusterin, Complement C1q, CCL18	0.85	0.636	76.5
4 standard and 3 candidate biomarkers (Angiogenin, Clusterin and Complement C1q)	0.89	0.495	85.3
4 standard and 3 candidate biomarkers (Angiogenin, Clusterin and Complement C1q, CCL18)	0.905	0.238	82.4

4.4   Discussion

Numerous clinical trials have shown a dramatic improvement in the clinical outcome for MM patients over the last decade, but all patients will still relapse following treatment. Over the last decade, the US Food and Drug Administration (FDA) have approved the use of bortezomib as a frontline treatment for MM. Despite the developments of novel biological agents such as first and second generation proteasome inhibitors (bortezomib and carfilzomib), many MM patients will develop drug resistance and in some cases have de novo resistance to these novel agents (Fall et al. 2014, N. Kumar et al. 2008, Markovina et al. 2008). Furthermore, bortezomib has a range of adverse reactions such as thrombocytopenia, neutropenia, anaemia, infections, gastrointestinal discomfort, musculoskeletal pain and commonly peripheral neuropathy that may affect a patient's subsequent treatment (Argyriou et al. 2014, Kouroukis et al. 2014). The ability to avoid side effects from ineffective treatments is a crucial area where biomarkers can play a significant role.

Biomarkers have been effective in individualising treatments in various malignancies including breast, lung and nasopharyngeal cancers (Cagle, Raparia and Portier 2016, Calhoun and Collins 2015, Kang, Kiess and Chung 2015). In MM, there are still limited published studies on biomarkers that could be used to predict treatment (e.g. bortezomib) response. Recently, Cereblon (CRBN) has been found to be a target for immunomodulatory agents (IMiDs) and its lack of expression shows a high correlation with resistance to these agents (Broyl et al. 2013, Egan et al. 2013). Employing CRBN as a biomarker in MM, accurately predicts response rate and survival in patients treated with IMiDs.

Albumin, ß2M, paraprotein and kappa/lambda ratio are the standard biomarkers currently used in the clinical setting. Despite that these panels of blood tests are not specific to predict response to any form of therapy (Greipp et al. 2005). In this study, a number of proteins were found to be significantly changed in abundance between responders and non-responders to treatment containing bortezomib. C1Q is a hexamer of three unique protein subunits (A, B, and C) and plays a key role in apoptotic cell, immunological and inflammatory processes. It is a 460 kDa protein composed of 18 polypeptide chains of (6 A-chains, 6 B-chains and 6 C-chains) (Kishore and Reid 2000). Its role has been thought to be

involved with advanced glycation end (AGEs) products through its binding properties in human serum (Chikazawa et al. 2016), and studies have shown that these proteins have the ability to induce precipitation of soluble gamma globulin complexes (Eggleton, Reid and Tenner 1998). C1Q receptors are directed to the heavy polypeptide chains of IgG and IgM (Sontheimer, Racila and Racila 2005). It has been suggested that some myeloma IgG proteins undergo unusual glycosylation processes during disease progression (Aurer et al. 2007). These glycosylation studies together with our current findings suggest that there is an association between the increased levels of C1Q in the non-responders group.

ANG, a 14 kDa angiogenic ribonuclease, is an actin binding protein on the surface of endothelial cells, promoting cell invasion and migration (Gao and Xu 2008). Various studies have shown that immunomodulatory drugs such as thalidomide have an inhibitory effect on angiogenesis in cancer, thus reinforcing the importance of ANG's role in MM disease progression (Terpos et al. 2008, Li et al. 2003, Thomas et al. 2003). A study has recently shown that myeloma patients who were treated with bortezomib and had responded well, displayed a significant decrease in microvessel density, which suggests that bortezomib may have an anti-angiogenic effect (Wang et al. 2015).

CLU, a ubiquitous extracellular protein is expressed in a range of diseases that arise from protein misfolding and deposition of highly structured protein (Falgarone and Chiocchia 2009). CLU exists in several isoforms; secretory and nuclear CLU. Its exact mechanism of action remains unknown but it is thought to play a role in promoting cancer cell survival through the activation of Akt and NFκB pathways (Zoubeidi and Gleave 2012, Zoubeidi et al. 2010, Trougakos and Gonos 2002, Moll et al. 1998). Studies have shown that bortezomib plays an important role with transcription nuclear factor kB (NFκB) by blocking the degradation of its inhibitor, IkB (Chen et al. 2011).

CCL18, also known as macrophage inflammatory protein 4, is a chemotactic cytokine that has no known receptor. Yuan and co-workers recently found using survival and multivariate analysis that CCL18 was an independent favourable prognostic factor in patients with colorectal cancer. Similarly, in this investigation, CCL18 levels were found to be significantly increased in the responder group (Yuan et al. 2013).

By combining these new candidate biomarkers (ANG, CLU, C1Q) with standard biomarkers (Albumin, ß2M, Paraprotein and K/L ratio), we have developed a novel panel of biomarkers to predict response to bortezomib treatment in MM. These ELISA kits are widely available and inexpensive in comparison to other prognostic methods. This novel biomarker panel may assist clinicians in choosing a more personalised and efficacious treatment, whilst minimising unnecessary side effects. Further prospective multi-centre randomised studies will be useful to determine the efficacy of this panel of biomarkers. MM is not one disease and therefore for any potential biomarker to be useful in monitoring or predicting response to treatment, it will most likely be as part of a panel of disease relevant biomolecules with direct connection to the phenotypic machinery associated with MM. Most biomarkers current used in MM management are either diagnostic or prognostic. This study has demonstrated that by combining biomarkers that are currently used in the management of MM, with additional candidate protein biomarkers, it is possible to develop a sensitive test that predict response to bortezomib treatment, an established component of anti-myeloma therapy.

In this study, a combination of fraction methods was employed to reduce the dynamic range of proteins under investigation and facilitate the detection/quantification of more low abundant molecules by high abundant protein removal/depletion. However, high abundant proteins sequestering low abundant cargo proteins can help to increase their abundance and protect them from elimination by the body. A drawback of compressing the dynamic range by protein removal/depletion is the potential loss of information on the abundance levels for these proteins under different conditions (response to treatment) and the valuable proteins that can be found attached to these molecules (Dowling, Palmerini et al. 2014).The ISS predicts survival of newly diagnosed MM patients by using a combination of two routine biomarkers, ß2M and albumin, and uses this information to separate patients into three stages with a distinct prognosis. However, this approach is not specific to any individual treatment. Therefore, the focus of this investigation was to combine routine biomarkers, such as those used by the ISS, with novel biomarkers discovered using proteomics analysis of MM patient samples. Ultimately the aim is to develop a predictive model using easily quantifiable and accessible platforms, directly associated with specific therapeutic options, such as bortezomib. Such theranostics are intended to give clinicians the means to implement a treatment plan customised for each

patient. Prospective studies will be planned to validate the proposed risk stratification panel in this study and will also take into account the revised International Staging System (R-ISS), to improve risk stratification by evaluating the presence of chromosomal abnormalities detected by iFISH (t(4;14), t(14;16) and del17p) and elevated serum lactate dehydrogenase in addition to ISS.

4.5   References

Argyriou, A.A., Cavaletti, G., Bruna, J., Kyritsis, A.P. and Kalofonos, H.P. 2014. Bortezomib-induced peripheral neurotoxicity: An update. *Archives of Toxicology,* 88(9), pp.1669-1679.

Aurer, I., Lauc, G., Dumic, J., Rendic, D., Matisic, D., Milos, M., Heffer-Lauc, M., Flogel, M. and Labar, B. 2007. Aberrant glycosylation of igg heavy chain in multiple myeloma. *Collegium Antropologicum,* 31(1), pp.247-251.

Avet-Loiseau, H., Durie, B.G., Cavo, M., Attal, M., Gutierrez, N., Haessler, J., Goldschmidt, H., Hajek, R., Lee, J.H., Sezer, O., Barlogie, B., Crowley, J., Fonseca, R., Testoni, N., Ross, F., Rajkumar, S.V., Sonneveld, P., Lahuerta, J., Moreau, P., Morgan, G. and International Myeloma Working Group. 2013. Combining fluorescent in situ hybridization data with ISS staging improves risk assessment in myeloma: An international myeloma working group collaborative project. *Leukemia,* 27(3), pp.711-717.

Broyl, A., Kuiper, R., van Duin, M., van der Holt, B., el Jarari, L., Bertsch, U., Zweegman, S., Buijs, A., Hose, D., Lokhorst, H.M., Goldschmidt, H., Sonneveld, P., Dutch-Belgian HOVON group and German GMMG Group. 2013. High cereblon expression is associated with better survival in patients with newly diagnosed multiple myeloma treated with thalidomide maintenance. *Blood,* 121(4), pp.624-627.

Cagle, P.T., Raparia, K. and Portier, B.P. 2016. Emerging biomarkers in personalized therapy of lung cancer. *Advances in Experimental Medicine and Biology,* 890pp.25-36.

Calhoun, B.C. and Collins, L.C. 2015. Predictive markers in breast cancer: An update on ER and HER2 testing and reporting. *Seminars in Diagnostic Pathology,* 32(5), pp.362-369.

Cavo, M., Tacchetti, P., Patriarca, F., Petrucci, M.T., Pantani, L., Galli, M., Di Raimondo, F., Crippa, C., Zamagni, E., Palumbo, A., Offidani, M., Corradini, P., Narni, F., Spadano, A., Pescosta, N., Deliliers, G.L., Ledda, A., Cellini, C., Caravita, T., Tosi, P., Baccarani, M. and GIMEMA Italian Myeloma Network. 2010. Bortezomib with thalidomide plus dexamethasone compared with thalidomide plus dexamethasone as induction therapy before, and consolidation therapy after, double autologous stem-cell transplantation in newly diagnosed multiple myeloma: A randomised phase 3 study. *Lancet (London, England),* 376(9758), pp.2075-2085.

Chen, D., Frezza, M., Schmitt, S., Kanwar, J. and Dou, Q.P. 2011. Bortezomib as the first proteasome inhibitor anticancer drug: Current status and future perspectives. *Current Cancer Drug Targets,* 11(3), pp.239-253.

Chikazawa, M., Shibata, T., Hatasa, Y., Hirose, S., Otaki, N., Nakashima, F., Ito, M., Machida, S., Maruyama, S. and Uchida, K. 2016. Identification of C1q as a binding protein for advanced glycation end products. *Biochemistry,* 55(3), pp.435-446.

Dispenzieri, A., Kyle, R., Merlini, G., Miguel, J.S., Ludwig, H., Hajek, R., Palumbo, A., Jagannath, S., Blade, J., Lonial, S., Dimopoulos, M., Comenzo, R., Einsele, H., Barlogie, B., Anderson, K., Gertz, M., Harousseau, J.L., Attal, M., Tosi, P., Sonneveld, P., Boccadoro, M., Morgan, G., Richardson, P., Sezer, O., Mateos, M.V., Cavo, M., Joshua, D., Turesson, I., Chen, W., Shimizu, K., Powles, R., Rajkumar, S.V., Durie, B.G. and International Myeloma Working Group. 2009. International myeloma working group guidelines for serum-free light chain analysis in multiple myeloma and related disorders. *Leukemia,* 23(2), pp.215-224.

Dowling, P., Hayes, C., Ting, K.R., Hameed, A., Meiller, J., Mitsiades, C., Anderson, K.C., Clynes, M., Clarke, C., Richardson, P. and O'Gorman, P. 2014. Identification of proteins found to be significantly altered when comparing the serum proteome from multiple myeloma patients with varying degrees of bone disease. *BMC Genomics,* 15pp.904-2164-15-904.

Dowling, P., Palmerini, V., Henry, M., Meleady, P., Lynch, V., Ballot, J., Gullo, G., Crown, J., Moriarty, M. and Clynes, M. 2014. Transferrin-bound proteins as potential biomarkers for advanced breast cancer patients. *BBA Clinical,* 2pp.24-30.

Durie, B.G., Harousseau, J.L., Miguel, J.S., Blade, J., Barlogie, B., Anderson, K., Gertz, M., Dimopoulos, M., Westin, J., Sonneveld, P., Ludwig, H., Gahrton, G., Beksac, M., Crowley, J., Belch, A., Boccadoro, M., Cavo, M., Turesson, I., Joshua, D., Vesole, D., Kyle, R., Alexanian, R., Tricot, G., Attal, M., Merlini, G., Powles, R., Richardson, P., Shimizu, K., Tosi, P., Morgan, G., Rajkumar, S.V. and International Myeloma Working Group. 2006. International uniform response criteria for multiple myeloma. *Leukemia,* 20(9), pp.1467-1473.

Egan, J.B., Kortuem, K.M., Kurdoglu, A., Izatt, T., Aldrich, J., Reiman, R., Phillips, L., Baker, A., Shi, C.X., Schmidt, J., Liang, W.S., Craig, D.W., Carpten, J.D. and Stewart, A.K. 2013. Extramedullary myeloma whole genome sequencing reveals novel mutations in cereblon, proteasome subunit G2 and the glucocorticoid receptor in multi drug resistant disease. *British Journal of Haematology,* 161(5), pp.748-751.

Eggleton, P., Reid, K.B. and Tenner, A.J. 1998. C1q--how many functions? how many receptors? *Trends in Cell Biology,* 8(11), pp.428-431.

Falgarone, G. and Chiocchia, G. 2009. Chapter 8: Clusterin: A multifacet protein at the crossroad of inflammation and autoimmunity. *Advances in Cancer Research,* 104pp.139-170.

Fall, D.J., Stessman, H., Patel, S.S., Sachs, Z., Van Ness, B.G., Baughn, L.B. and Linden, M.A. 2014. Utilization of translational bioinformatics to identify novel biomarkers of bortezomib resistance in multiple myeloma. *Journal of Cancer,* 5(9), pp.720-727.

Field-Smith, A., Morgan, G.J. and Davies, F.E. 2006. Bortezomib (velcadetrade mark) in the treatment of multiple myeloma. *Therapeutics and Clinical Risk Management,* 2(3), pp.271-279.

Fonseca, R., Barlogie, B., Bataille, R., Bastard, C., Bergsagel, P.L., Chesi, M., Davies, F.E., Drach, J., Greipp, P.R., Kirsch, I.R., Kuehl, W.M., Hernandez, J.M., Minvielle, S., Pilarski, L.M., Shaughnessy, J.D.,Jr, Stewart, A.K. and Avet-Loiseau, H. 2004. Genetics and cytogenetics of multiple myeloma: A workshop report. *Cancer Research,* 64(4), pp.1546-1558.

Fonseca, R., Bergsagel, P.L., Drach, J., Shaughnessy, J., Gutierrez, N., Stewart, A.K., Morgan, G., Van Ness, B., Chesi, M., Minvielle, S., Neri, A., Barlogie, B., Kuehl, W.M., Liebisch, P., Davies, F., Chen-Kiang, S., Durie, B.G., Carrasco, R., Sezer, O., Reiman, T., Pilarski, L., Avet-Loiseau, H. and International Myeloma Working Group. 2009. International myeloma working group molecular classification of multiple myeloma: Spotlight review. *Leukemia,* 23(12), pp.2210-2221.

Gao, X. and Xu, Z. 2008. Mechanisms of action of angiogenin. *Acta Biochimica Et Biophysica Sinica,* 40(7), pp.619-624.

Greipp, P.R., San Miguel, J., Durie, B.G., Crowley, J.J., Barlogie, B., Blade, J., Boccadoro, M., Child, J.A., Avet-Loiseau, H., Kyle, R.A., Lahuerta, J.J., Ludwig, H., Morgan, G., Powles, R., Shimizu, K., Shustik, C., Sonneveld, P., Tosi, P., Turesson, I. and Westin, J. 2005. International staging system for multiple myeloma. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology,* 23(15), pp.3412-3420.

Hosmer, D.W., Hosmer, T., Le Cessie, S. and Lemeshow, S. 1997. A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in Medicine,* 16(9), pp.965-980.

Inamoto, Y., Kurahashi, S., Imahashi, N., Fukushima, N., Adachi, T., Kinoshita, T., Tsushita, K., Miyamura, K., Naoe, T. and Sugiura, I. 2009. Combinations of cytogenetics and international scoring system can predict poor prognosis in multiple myeloma after high-dose chemotherapy and autologous stem cell transplantation. *American Journal of Hematology,* 84(5), pp.283-286.

Jagannath, S., Durie, B.G., Wolf, J., Camacho, E., Irwin, D., Lutzky, J., McKinley, M., Gabayan, E., Mazumder, A., Schenkein, D. and Crowley, J. 2005. Bortezomib therapy alone and in combination with dexamethasone for previously untreated symptomatic multiple myeloma. *British Journal of Haematology,* 129(6), pp.776-783.

Jagannath, S., Richardson, P.G., Sonneveld, P., Schuster, M.W., Irwin, D., Stadtmauer, E.A., Facon, T., Harousseau, J.L., Cowan, J.M. and Anderson, K.C. 2007. Bortezomib appears to overcome the poor prognosis conferred by chromosome 13 deletion in phase 2 and 3 trials. *Leukemia,* 21(1), pp.151-157.

Kang, H., Kiess, A. and Chung, C.H. 2015. Emerging biomarkers in head and neck cancer in the era of genomics. *Nature Reviews.Clinical Oncology,* 12(1), pp.11-26.

Kishore, U. and Reid, K.B. 2000. C1q: Structure, function, and receptors. *Immunopharmacology,* 49(1-2), pp.159-170.

Kouroukis, T.C., Baldassarre, F.G., Haynes, A.E., Imrie, K., Reece, D.E. and Cheung, M.C. 2014. Bortezomib in multiple myeloma: Systematic review and clinical considerations. *Current Oncology (Toronto, Ont.),* 21(4), pp.e573-603.

Kumar, R. and Indrayan, A. 2011. Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatrics,* 48(4), pp.277-287.

Kumar, S. and Rajkumar, S.V. 2008. Many facets of bortezomib resistance/susceptibility. *Blood,* 112(6), pp.2177-2178.

Kumar, S.K., Rajkumar, S.V., Dispenzieri, A., Lacy, M.Q., Hayman, S.R., Buadi, F.K., Zeldenrust, S.R., Dingli, D., Russell, S.J., Lust, J.A., Greipp, P.R., Kyle, R.A. and Gertz, M.A. 2008. Improved survival in multiple myeloma and the impact of novel therapies. *Blood,* 111(5), pp.2516-2520.

Li, X., Liu, X., Wang, J., Wang, Z., Jiang, W., Reed, E., Zhang, Y., Liu, Y. and Li, Q.Q. 2003. Effects of thalidomide on the expression of angiogenesis growth factors in human A549 lung adenocarcinoma cells. *International Journal of Molecular Medicine,* 11(6), pp.785-790.

Markovina, S., Callander, N.S., O'Connor, S.L., Kim, J., Werndli, J.E., Raschko, M., Leith, C.P., Kahl, B.S., Kim, K. and Miyamoto, S. 2008. Bortezomib-resistant nuclear factor-kappaB activity in multiple myeloma cells. *Molecular Cancer Research : MCR,* 6(8), pp.1356-1364.

Millioni, R., Tolin, S., Puricelli, L., Sbrignadello, S., Fadini, G.P., Tessari, P. and Arrigoni, G. 2011. High abundance proteins depletion vs low abundance proteins enrichment: Comparison of methods to reduce the plasma proteome complexity. *PloS One,* 6(5), pp.e19603.

Mitsiades, N., Mitsiades, C.S., Richardson, P.G., Poulaki, V., Tai, Y.T., Chauhan, D., Fanourakis, G., Gu, X., Bailey, C., Joseph, M., Libermann, T.A., Schlossman, R., Munshi, N.C., Hideshima, T. and Anderson, K.C. 2003. The proteasome inhibitor PS-341 potentiates sensitivity of multiple myeloma cells to conventional chemotherapeutic agents: Therapeutic applications. *Blood,* 101(6), pp.2377-2380.

Moll, S., Menoud, P.A., French, L., Sappino, A.P., Pastore, Y., Schifferli, J.A. and Izui, S. 1998. Tubular up-regulation of clusterin mRNA in murine lupus-like nephritis. *The American Journal of Pathology,* 152(4), pp.953-962.

Rajpal, R., Dowling, P., Meiller, J., Clarke, C., Murphy, W.G., O'Connor, R., Kell, M., Mitsiades, C., Richardson, P., Anderson, K.C., Clynes, M. and O'Gorman, P. 2011. A novel panel of protein biomarkers for predicting response to thalidomide-based therapy in newly diagnosed multiple myeloma patients. *Proteomics,* 11(8), pp.1391-1402.

Richardson, P.G., Barlogie, B., Berenson, J., Singhal, S., Jagannath, S., Irwin, D.H., Rajkumar, S.V., Srkalovic, G., Alsina, M. and Anderson, K.C. 2006. Extended follow-up of a phase II trial in relapsed, refractory multiple myeloma:: Final time-to-event results from the SUMMIT trial. *Cancer,* 106(6), pp.1316-1319.

Richardson, P.G., Weller, E., Lonial, S., Jakubowiak, A.J., Jagannath, S., Raje, N.S., Avigan, D.E., Xie, W., Ghobrial, I.M., Schlossman, R.L., Mazumder, A., Munshi, N.C., Vesole, D.H., Joyce, R., Kaufman, J.L., Doss, D., Warren, D.L., Lunde, L.E., Kaster, S., Delaney, C., Hideshima, T., Mitsiades, C.S., Knight, R., Esseltine, D.L. and Anderson, K.C. 2010. Lenalidomide, bortezomib, and dexamethasone combination therapy in patients with newly diagnosed multiple myeloma. *Blood,* 116(5), pp.679-686.

Sontheimer, R.D., Racila, E. and Racila, D.M. 2005. C1q: Its functions within the innate and adaptive immune responses and its role in lupus autoimmunity. *The Journal of Investigative Dermatology,* 125(1), pp.14-23.

Sontheimer, R.D., Racila, E. and Racila, D.M. 2005. C1q: Its functions within the innate and adaptive immune responses and its role in lupus autoimmunity. *The Journal of Investigative Dermatology,* 125(1), pp.14-23.

Terpos, E., Kastritis, E., Roussou, M., Heath, D., Christoulas, D., Anagnostopoulos, N., Eleftherakis-Papaiakovou, E., Tsionos, K., Croucher, P. and Dimopoulos, M.A. 2008. The combination of bortezomib, melphalan, dexamethasone and intermittent thalidomide is an effective regimen for relapsed/refractory myeloma and is associated with improvement of abnormal bone metabolism and angiogenesis. *Leukemia,* 22(12), pp.2247-2256.

Thomas, D.A., Estey, E., Giles, F.J., Faderl, S., Cortes, J., Keating, M., O'Brien, S., Albitar, M. and Kantarjian, H. 2003. Single agent thalidomide in patients with relapsed or refractory acute myeloid leukaemia. *British Journal of Haematology,* 123(3), pp.436-441.

Trougakos, I.P. and Gonos, E.S. 2002. Clusterin/apolipoprotein J in human aging and cancer. *The International Journal of Biochemistry & Cell Biology,* 34(11), pp.1430-1448.

van Rhee, F., Bolejack, V., Hollmig, K., Pineda-Roman, M., Anaissie, E., Epstein, J., Shaughnessy, J.D.,Jr, Zangari, M., Tricot, G., Mohiuddin, A., Alsayed, Y., Woods, G., Crowley, J. and Barlogie, B. 2007. High serum-free light chain levels and their rapid reduction in response to therapy define an aggressive multiple myeloma subtype with poor prognosis. *Blood,* 110(3), pp.827-832.

Wang, L., Du, F., Zhang, H.M., Zhang, W.J. and Wang, H.X. 2015. Changes in circulating endothelial progenitor cells predict responses of multiple myeloma patients to treatment with bortezomib and dexamethasone. *Brazilian Journal of Medical and Biological Research = Revista Brasileira De Pesquisas Medicas e Biologicas,* 48(8), pp.736-742.

Yuan, R., Chen, Y., He, X., Wu, X., Ke, J., Zou, Y., Cai, Z., Zeng, Y., Wang, L., Wang, J., Fan, X., Wu, X. and Lan, P. 2013. CCL18 as an independent favorable prognostic biomarker in patients with colorectal cancer. *The Journal of Surgical Research,* 183(1), pp.163-169.

Zou, K.H., O'Malley, A.J. and Mauri, L. 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation,* 115(5), pp.654-657.

Zoubeidi, A., Ettinger, S., Beraldi, E., Hadaschik, B., Zardan, A., Klomp, L.W., Nelson, C.C., Rennie, P.S. and Gleave, M.E. 2010. Clusterin facilitates COMMD1 and I-kappaB degradation to enhance NF-kappaB activity in prostate cancer cells. *Molecular Cancer Research : MCR,* 8(1), pp.119-130.

Zoubeidi, A. and Gleave, M. 2012. Small heat shock proteins in cancer therapy and prognosis. *The International Journal of Biochemistry & Cell Biology,* 44(10), pp.1646-1656.

Metabolomic and proteomic analysis of breast cancer patient samples suggests that glutamate and 12-HETE in combination with CA15-3 may be useful biomarkers reflecting tumour burden

Authors: Paul Dowling, <u>Michael Henry</u>, Paula Meleady, Colin Clarke, Kathy Gately, Kenneth O'Byrne, Elizabeth Connolly, Vincent Lynch, Jo Ballot, Giuseppe Gullo, John Crown, Michael Moriarty, Martin Clynes

Within this work, Michael Henry contributed to the methodology of the serum proteome equalisation strategy. Michael Henry specifically performed the in-solution protein digests and LC-MS/MS analysis on all the protein samples. Michael Henry implemented the differential expression analysis tools using label free software Progenesis QI and searched the statistically significant protein features using MS search algorithms. Michael Henry prepared the protein differential data results. In the writing of the publication prepared the proteomic sections in the materials and methods and results section.

Abstract

Evaluation of protein and metabolite expression patterns in blood using mass spectrometry and high-throughput antibody-based screening platforms has potential for the discovery of new biomarkers for managing breast cancer patient treatment. Previously identified blood-based breast cancer biomarkers, including cancer antigen 15.3 (CA15-3) are useful in combination with imaging (computed tomography scans, magnetic resonance imaging, X-rays) and physical examination for monitoring tumour burden in advanced breast cancer patients. However, these biomarkers suffer from insufficient levels of accuracy and with new therapies available for the treatment of breast cancer, there is an urgent need for reliable, non-invasive biomarkers that measure tumour burden with high sensitivity and specificity so as to provide early warning of the need to switch to an alternative treatment. The aim of this study was to identify a biomarker signature of tumour burden using cancer and non-cancer (healthy controls/non-malignant breast disease) patient samples.

Results demonstrate that combinations of three candidate biomarkers from Glutamate, 12-Hydroxyeicosatetraenoic acid, Beta-hydroxybutyrate, Factor V and Matrix metalloproteinase-1 with CA15-3, an established biomarker for breast cancer, were found to mirror tumour burden, with AUC values ranging from 0.71 to 0.98 when comparing non-malignant breast disease to the different stages of breast cancer.

Further validation of these biomarker panels could potentially facilitate the management of breast cancer patients, especially to assess changes in tumour burden in combination with imaging and physical examination.

5.1 Introduction

Breast cancer is the most common cancer and the leading cause of cancer-related death in women worldwide. Declines in breast cancer mortality have been reported in several developed countries including northern America, Australia, and several European regions, (Soerjomataram et al. 2012), however the overall incidence rate continues to increase (Benson and Jatoi 2012). Metastatic breast cancer (MBC) is not a curable condition but treatment can prolong life, delay the progression of the cancer, relieve cancer related symptoms, and improve quality of life (Senkus, Cardoso and Pagani 2014). MBC is treatable by means of serial administration of endocrine, cytotoxic, or biologic therapies. Sensitive reproducible biomarkers of response/resistance are urgently needed to avoid unnecessary toxic side effects of ineffective therapies and to signal the need to switch to a different treatment regime. Treatment response can be assessed with the use of serial imaging, but logistics, cost and radiation exposure considerations limit the applicability of this approach.

The most commonly used serum biomarkers in patients with breast cancer is the soluble form of Mucin 1 (MUC1/ CA15-3) (Duffy, Evoy and McDermott 2010), primarily used for monitoring therapy outcomes and disease progression in MBC patients (Duffy 2006). MUC1 is a transmembrane mucin which can become overexpressed in epithelial cancer cells (Brayman, Thathiah and Carson 2004). CA15-3 is an FDA approved assay that detects shed or soluble forms of MUC1 protein (Hayes et al. 1985). A CA15-3 test which utilises an antibody pair, DF3 and 115-D8 to detect the soluble forms of MUC-1 protein is the most widely used serum marker in patients for monitoring MBC. CA 27.29 also detects MUC-1 through the monoclonal antibody BR 27.29. CA15-3 is clinically useful in some patients with MBC for monitoring response to therapy (Kurebayashi et al. 2004, Tondini et al. 1988), however ASCO guidelines state that CA15-3 should not be used alone but can be used in conjunction with diagnostic imaging, history and physical examination (Graham et al. 2014, Duffy 2013, Harris et al. 2007).

Carcinoembryonic antigen (CEA), circulating cytokeratins, for example tissue polypeptide antigen, tissue polypeptide specific antigen and cytokeratin 19 fragment (CYFRA 21-1), and the proteolytically cleaved ectodomain of the human epidermal growth factor receptor 2 (s-HER2) are also useful in the management of breast cancer patients (Moreno-Aspitia et al. 2013, Nakata et al. 2000, Mughal et al. 1983, Lokich, Zamcheck and Lowenstein 1978, Haagensen Jr

et al. 1978). Ideally, a serum tumour biomarker should be able to detect disease early, predict response or the emergence of resistance to specific therapeutic strategies and for monitoring patient's post-treatment (Mirabelli and Incoronato 2013). Only serum tests for MUC1 (CA27.29 or CA15-3) and CEA are FDA approved for use in breast cancer, however, these tests lack the high levels of overall accuracy needed and are most effective when used in combination with imaging techniques currently used routinely by oncology departments.

The aim of this study was to determine the serum levels for a number of candidate biomarkers, identified using different discovery platforms and analysed in a larger independent sample set with different cancer stages, non-malignant breast disease and the healthy subjects. Moreover, we defined the receiver operating characteristics (ROC) curve for all tested parameters and performed logistic regression (LR) analysis on combinations of the most promising candidate biomarkers together with the FDA approved CA15-3.

The data presented identifies a panel of protein and metabolite biomolecules that correlate significantly with increasing tumour burden.

5.2   Materials and methods

5.2.1   Patient information and sample collection

Samples were collected through a collaborative project involving participating sites (St Vincent's University Hospital, Dublin 4, Ireland and St. James's Hospital, Dublin 8, Ireland), coordinated through ICORG (the All Ireland Co-operative Oncology Research Group, www. icorg.ie) and by Asterand in the USA (Detroit, MI, USA). The samples were collected according to standard phlebotomy procedures from consented patients before the commencement of treatment. Ethical consent was granted from each of the respective Hospital Ethics Committees. A total of 10 ml of blood was collected into additive free (serum) blood tubes and was allowed to clot for 30 min to 1 h at room temperature. The serum was denuded by pipette from the clot and poured into a clean tube. The tubes were centrifuged at 1000xg for 30 min at 4°C. Serum was aliquoted in the cryovial tubes, labelled and stored at -80°C until time of

analysis. The time from sample procurement to storage at -80°C was less than 3 h. Each serum sample underwent not more than three freeze/thaw cycles prior to analysis. Table 5-1 provides a summary of the total number of patients analysed and a breakdown of their clinicopathologic parameters.

5.2.2   Metabolomics analysis

Serum specimens (250 µl) were shipped on dry-ice to Metabolon Inc., North Carolina, USA, where the metabolomics profiling was performed. Metabolon incorporates three independent complementary analysis platforms to maximise the number of small molecules and metabolites that the combined systems can identify and measure. Two independent ultra-high performance liquid chromatography/ tandem mass spectrometry (UHPLC/MS/MS2) injections (one optimised for basic compounds and the other for acidic compounds) and one GC/MS injection per sample are performed. Firstly, small molecules were extracted from serum specimens using methanol to allow precipitation of proteins. The extract supernatant was then split into four equal aliquots; two for UHPLC/MS, one for GC/MS and one reserve aliquot. Aliquots were then dried overnight to remove solvent.

For the UHPLC methods, one aliquot was reconstituted in 50 µl 0.1 % formic acid and the other in 50 µl 6.5 mM ammonium bicarbonate pH 8.0. For GC/MS analysis, aliquots were derivatized using equal parts N,O-bistrimethylsilyltrifluoroacetamide and a solvent mixture of acetonitrile:dichloromethane:cyclohexane (5:4:1) with 5 % triethylamine at 60°C for 1 h. All reconstitution solvents contained instrument internal standards used to monitor instrument performance. UHPLC/MS was carried out using a Waters Acquity UHPLC coupled to an LTQ mass spectrometer equipped with an electrospray ionization source. Two independent

UHPLC/MS injections were performed on each sample. The acidic injections were monitored for positive ions and the basic injections were monitored for negative ions. The derivatized samples for GC/MS were analysed on a Thermo-Finnigan Trace DSQ fast-scanning single-quadrupole MS. The resulting MS/MS2 data was then searched against Metabolon's reference standard library. This library was generated from 1500 standards and contains the retention time/index, mass to charge (m/z), and MS/MS spectral data for all molecules in the library,

including their associated adducts, in-source fragments, and multimers. The library allows identification of experimentally detected metabolites based on a multi-parameter match basis. All identifications and quantifications were subjected to QC to verify the quality of the identification and peak integration.

5.2.3 ProteoMiner™ fractionation

Serum protein equalisation was performed using Proteo-Miner™ enrichment kit (Bio-Rad Laboratories, Hercules, CA) according to the manufacturer procedure. In summary, the storage solution was first washed out from the spin column containing 100 µl of peptide beads with deionised water. Thereafter, the column was washed with the 10 mM NaH_2PO_4, 150 mM NaCl, pH 7.4 solution provided with the kit. When the spin column was ready for sample binding, 1 ml of centrifuged serum sample was added to the column and equilibrated at room temperature for 2 h. The unbound proteins were removed with the wash buffer and the captured proteins were eluted by 3 x 100 µl of 8 M urea containing 2 % CHAPS dissolved in 5 % acetic acid.

Following vortexing, sonication and centrifugation, the protein concentration of control and stage IV breast cancer patient samples was determined. Volumes of protein suspensions were equalised using label-free solubilisation buffer and then reduced for 30 min with 10 mM DTT and alkylated for 20 min in the dark with 25 mM iodoacetamide in 50 mM ammonium bicarbonate. The proteolytic digestion of proteins was carried out in 2 steps. Firstly, digestion was performed with sequencing grade Lys-C at a ratio of 1:100 (protease/protein) for 4 h at 37°C, followed by diluted with four times the initial sample volume in 50 mM ammonium bicarbonate. Secondly, further digestion was based on incubation with sequencing grade trypsin at a ratio of 1:25 (protease/protein) overnight at 37°C. The protease-treated serum protein suspensions were diluted 3:1 (v/v) with 2 % trifluoroacetic acid in 20 % acetonitrile. To ensure an even suspension of peptide populations from control and stage IV breast cancer patient samples, the samples were briefly vortexed and sonicated.

5.2.4 Label-free LC–MS/MS analysis

The nano LC–MS/MS analysis of control versus stage IV breast cancer patient samples was carried out with the help of an Ultimate 3000 nanoLC system (Dionex UK, Camberley, United Kingdom) coupled to a an LTQ Orbitrap XL mass spectrometrer (Thermo Fisher Scientific, Dublin, Ireland) in the Proteomics Facility of the National Institute for Cellular Biotechnology, Dublin City University. The optimised methodology has been as previously described in detail) (Meleady, Hoffrogge et al. 2012a, Meleady et al. 2012). Peptide mixtures (5µl volume) were loaded onto a C18 trap column (C18 Pep-Map, 300 µm x id 50 mm, 5 µm particle size, 100 Å pore size; Dionex). Desalting was achieved at a flow rate of 25 µl/min in 0.1 % TFA containing 2% ACN for 3 min. The trap column was switched on-line with an analytical PepMap C18 column (75 µm id x 500 mm, 3 µm particle and 100 Å pore size; Dionex). Peptides generated from muscle proteins were eluted with the following binary gradients: solvent A (2 % ACN and 0.1 % formic acid in LC–MS grade water) and 0–25 % solvent B (80 % ACN and 0.08 % formic acid in LC–MS grade water) for 240 min and 25–50 % solvent B for a further 60 min. The column flow rate was set to 350 nl/min. Data was acquired with Xcalibur software, version 2.0.7 (Thermo Fisher Scientific). The MS apparatus was operated in data-dependent mode and externally calibrated. Survey MS scans were acquired in the Orbitrap in the 375–2000 m/z range with the resolution set to a value of 30,000 at m/z 400 and lock mass set to 445.120025. CID fragmentation was carried out in the linear ion trap with the three most intense ions per scan. Within 60 s, a dynamic exclusion window was applied. A normalised collision energy of 35 %, an isolation window of 2 m/z and one microscan were used to collect suitable tandem mass spectra.

5.2.5 Quantitative profiling by label-free LC–MS/MS analysis

Processing of the raw data generated from LC–MS/MS analysis was carried out with Progenesis label-free LC–MS software (version 3.1; Non-Linear Dynamics, Newcastle upon Tyne, UK). Data alignment was based on the LC retention time of each sample, allowing for any drift in retention time given and adjusted retention time for all runs in the analysis (Meleady, Hoffrogge et al. 2012b, Meleady, Gallagher et al. 2012). A reference run was established with the sample run that yielded most features (i.e. peptide ions). The retention times of all of the

other runs were aligned to this reference run and peak intensities were then normalised. Prior to exporting the MS/MS output files to MASCOT (www.matrixscience.com) for protein identification, a number of criteria were employed to filter the data including: (i) peptide features with ANOVA\0.05 between experimental groups, (ii) mass peaks (features) with charge states from +2, +3, and (iii) greater than one isotope per peptide. A MASCOT generic file was generated from all exported MS/MS spectra from Progenesis software. The MASCOT generic file was used for peptide identification with MASCOT (version 2.2) and searched against the UniProtKB-Swiss-Prot database (downloaded in January 2013) with 16,638 proteins (taxonomy: Homo sapiens). The following search parameters were used for protein identification: (i) MS/MS mass tolerance set at 0.5 Da, (ii) peptide mass tolerance set to 20 ppm, (iii) carbamidomethylation set as a fixed modification, (iv) up to two missed cleavages were allowed and (v) methionine oxidation set as a variable modification. For further consideration and re-importation back into Progenesis LC–MS software for further analysis, only peptides with ion scores of 40 and above were chosen. Importantly, the following criteria were applied to assign a serum associated proteins as properly identified: (i) an ANOVA score between experimental groups of ≥0.05, (ii) proteins with >2 peptides matched and (iii) a MASCOT score >40.

5.2.6   Serum biomarker screening and colorimetric assays

Serum samples were screened using different protein panel selections, including MILLIPLEX MAP Human Circulating Cancer Biomarker Magnetic Bead Panel, MILLIPLEX MAP Human MMP Magnetic Bead Panel 2 and MILLIPLEX MAP Human Angiogenesis/Growth Factor Magnetic Bead Panel (Merck-Millipore UK: HCCBP1MAG-58K, HMMP2MAG-55K & HAGP1MAG- 12K). Multiplex assays were processed on a MAGPIX Plate Reader (Luminex, Austin, TX, US), and analysed with 3.1 xPONENT software. Protein expression levels were expressed as median fluorescent intensities generated by analyzing 50–100 magnetic beads for each analyte in each sample. The concentrations of analytes were quantitated from median fluorescence intensities using standard curves generated by Analyst 5.1 five-parameter curve fitting to the series of known concentrations for each analyte. All samples were performed in duplicate. Significantly

elevated markers were reanalysed by MILLIPLEX MAP Magnetic Bead Panels now tailored for the markers of interest in a larger independent cohort of samples.

13-HODE, 12-hydroxyeicosatetraenoic acid (12-HETE) and Serotonin levels were analysed using 13(S)-HODE EIA kits, 12(S)-HETE EIA kits and Serotonin ELISA kits respectively which are colorimetric competitive enzyme immunoassay with absorbances read at 405 nm (Enzo Life Sciences (UK) LTD: ADI-900-108, ADI-900-050 & ADI- 900-175). ELISA kits to Fibronectin (competitive enzyme immunoassay technique) and Factor V (sandwich enzyme immunoassay technique) were performed to quantify abundance level for these label-free LC–MS identified proteins (Universal Biologicals (Cambridge) Ltd: EF1045- 1 & EF1005-1). Assays for Glutamate, and Beta- Hydroxybutyrate were performed according to the manufacturers guidelines and quantified by colorimetric (spectrophotometry at k = 450 nm) method (Abcam: ab83389 & ab83390). All samples were performed in duplicate.

5.2.7   Statistical analysis

Student t-test was used to identify statistically significant changes in abundance levels for specific proteins between control, non-malignant breast disease and stages I–IV breast cancer patient serum samples (discovery phase— Table 5-2, Table 5-3, Table 5-4). Non-parametric Kruskal–Wallis tests were used for a global assessment of possible differences in the distributions of the candidate biomarker measurements across the different benign and disease categories (validation phase). Following the statistically significant global test, a pairwise Wilcoxon Rank Sum test was performed to determine differences between groups for each candidate biomarker with a Benjamini Hochberg correction used to adjust the p values for multiple testing. ROC curve analysis was performed as it is a useful tool in assessment of biomarker accuracy (Dowling et al. 2012). The ROC plots were obtained by plotting all sensitivity values (true positive fraction) on the y-axis against their equivalent (1- specificity) values (false positive fraction) for all available thresholds on the x-axis (MedCalc, version 13-0-0-0 64-bit, Medcalc Software, Mariakerke, Belgium). The area under the curve (AUC) was calculated to provide a summary of overall classifier effectiveness. In our study, we consider AUC values ranging from 0.5-0.7 as poor, 0.7-0.8 as average, 0.8-0.9 as good and >0.9 as

outstanding. For multivariate analysis of biomarker combinations, LR analysis of the serum biomarker levels in these patients groups was performed.

5.3   Results

5.3.1   Metabolomic profiling

Samples were split into four aliquots for untargeted metabolic profiling and randomised for analysis. Samples were characterised using three independent platforms: ultrahigh-performance liquid chromatography/tandem mass spectrometry (UHPLC-MS/MS) in the negative ion mode, UHPLC-MS/MS in the positive ion mode, and gas chromatography-mass spectrometry (GC-MS). Of the molecules that were analysed, a number of them were found to be statistically significant when comparing normal controls to advanced breast cancer patient serum samples (Table 5-2). The vast majority of these molecules were found to have an increased abundance levels in the advanced patient samples with only Serotonin (5-HT) found to be higher in the normal control group. 13-HODE, 12-HETE, Glutamate, Beta-hydroxybutyrate and Serotonin were selected for further analysis based on the availability of biochemical or antibody-based assays for further analysis in an independent cohort of patient samples.

5.3.2   Proteomic profiling

Prior to label-free mass spectrometry analysis, serum samples from normal control and advanced breast cancer patients were fractionated using ProteoMiner[TM] technology. The dynamic range of serum protein concentrations comprises at least nine orders of magnitude. Without fractionation, high abundant serum protein such as albumin, Haptoglobin and immunoglobulins would mask the vast majority of proteins making their detection and identification impossible. Many groups have previously reported on using ProteoMiner[TM] technology as a sample preparation tool to compression of the dynamic range in complex protein mixtures such as serum (Milan et al. 2012, Monari et al. 2011). ProteoMiner[TM]

technology contains millions of randomly synthesised hexapeptide ligands that are equally represented and when a complex sample such as serum is processed, high-abundance proteins quickly saturate their binding sites, while low abundant proteins do not saturate the corresponding hexapeptide ligands, therefore allowing for quantitative analysis in subsequent experiments.

ProteoMiner™ fractionated samples were analysed by label-free mass spectrometry analysis. The resultant data was transferred to Progenesis LC–MS software to compare normal controls to advanced breast cancer patient serum samples. Following Progenesis LC–MS analysis peptide features with ANOVA <0.05 and +2/+3 charge states were subjected to MASCOT database searching. The resultant MASCOT mgf files were then resubmitted to the Progenesis software to yield a list of identified proteins (Table 5-3). Similarly, to the metabolomics data, the majority of proteins were found to be increased in abundance levels in the advanced breast cancer patient group compared to normal controls. We included proteins that were only identified by a single peptide (the so-called ''one-hit wonders) as the complexity of the serum sample even after ProteoMiner™ fractionation was very high and resulted in many single hit identifications. We selected Fibronectin and Coagulation Factor V for further ELISA-based analysis as these were strong candidates in terms of *p*-values and MASCOT scores and some of the other proteins we had previously investigated (Dowling et al. 2012).

5.3.3   Analyte screen

Single ELISAs have been the mainstay of immunoassay technology for nearly half a century, however, the development of multiplexed assays has proved to be a significant platform in screening clinical specimens for panels of known analytes producing data in a matter of hours in large numbers of samples (Bigbee et al. 2012). In this investigation, our approach was to use the Luminex xMAP technology platform to screen serum samples from advanced breast cancer patients compared to normal controls using assays for circulating cancer biomarkers, matrix metalloproteinases (MMPs) and angiogenesis/growth factors. Luminex xPONENT_ software was used to analyse the data with a number of analytes found to be statistically significant. Prolactin, Cancer Antigen 15-3 (CA15-3), Cancer Antigen 125 (CA125), Cancer Antigen 19-9 (CA19-9),Matrixmetalloproteinase-1 (MMP-1),Matrix metalloproteinase-9 (MMP-9), Epidermal

growth factor and vascular endothelial growth factor (VEGF) were all found to have increased levels in advanced breast cancer compared to the normal control group (Table 5-4). A number of these proteins were selected for further analysis in a larger independent cohort of patient samples based on their fold-change, p-value and established usage in breast cancer management.

5.3.4   Evaluation of candidate markers in an independent sample set

Single ELISAs, multiplex and biochemical assays were used to analyse a range of metabolites and proteins in an independent sample set comprised of 34 normal control serum samples, 12 non-malignant breast disease serum samples and samples from breast cancer patients stage I (n = 29), stage II (n = 30), stage III (n = 20) and stage IV (n = 16 or n = 50). Candidate biomarkers were selected for analysis in this cohort of patient samples based on the availability of suitable assays and also the amount of sample needed. For CA15-3, Glutamate and 12-HETE, the number of samples in the stage V breast cancer group was 50 due to patient sample availability at the time of performing these assays. Table 5-5 summarises the data including mean ± standard deviation, median, range and n-value data for a number of selected metabolites/proteins analysed in the independent sample set. The non-malignant breast disease patient serum samples were from individuals diagnosed with fibroadenoma of the breast, fibrocystic mastopathy of the breast, fibrocystic change with fibroadenoma of the breast, duct papillomas of the breast or atypical ductal hyperplasia of the breast.

Comparison *p*-values and AUC-values between the different groups for 13-HODE, 12-HETE, Fibronectin, Glutamate, Beta Hydroxybutyrate, Factor V, CA15-3, CA125, VEGF, Serotonin and MMP-1 are included in the supplementary data (https://link.springer.com/article/10.1007%2Fs11306-014-0723-1). From this data, 12-HETE, Fibronectin, Glutamate, Factor V and CA15-3 were consistently found to be the most significant between the different comparisons.

Kurskal-Wallis tests were used to evaluate the significance of 12HETE, 13HODE, Fibronectin, Glutamate, Beta Hydroxybutyrate, Factor V, CA15-3, CA125, VEGF, MMP- 1 and Serotonin across all groups (Table 5-6). 12HETEFibronectin, Glutamate, Beta Hydroxybutyrate, Factor V,

CA15-3 and MMP-1 were found to be significant. A pair wise Wilcoxon Rank Sum test was performed to determine differences between groups for each candidate biomarker with a Benjamini Hochberg correction used to adjust the *p*-values for multiple testing (see supplementary data (https://link.springer.com/article/10.1007%2Fs11306-014-0723-1) for individual comparison tables and box & whisker plots).

5.3.5 Multivariate analysis of biomarker panel using logistic regression

Table 5-7 summarises LR analysis for combinations of 3 candidate biomarkers (3 was found to be the optimum number as additional biomarkers did not improve the AUC values). CA15-3 was always included as this is the most significant breast cancer biomarkers currently used in the clinic, with 12-HETE, Glutamate, Beta Hydroxybutyrate, Factor V and MMP-1 also selected as these were found to be significant using Kurskal–Wallis tests. Fibronectin was not selected for LR analysis as there was discordance between the discovery label-free analysis and subsequent measurement in the independent sample set. Fibronectin was found to be 2.77-fold increase in abundance levels in serum samples from advanced breast cancer patient group compared to normal controls. However, analysis by ELISA in the larger sample set showed a significant decrease in abundances levels, most noticeably in the advanced breast cancer disease group.

A number of biomarker combinations were found to be significant, especially when focusing on advanced disease compared to normal health controls and non-malignant breast disease patient samples. An example of one of the most significant LR AUC-values was for the combination of CA15-3/12-HETE/Glutamate in control compared to stage IV and non-malignant breast disease compared to stage IV. Figure 4-1 graphically represents the LR AUC-values for the combination of CA15-3/12-HETE/Glutamate and CA15-3 alone in the comparison of control to non-malignant breast disease, stage I, stage II, stage III and stage IV breast cancer patient serum specimens respectively (A) and LR AUC values for the combination of CA15-3/12-HETE/Glutamate and CA15-3 alone in the comparison of non-malignant breast disease to stage I, stage II, stage III and stage IV breast cancer patient serum specimens respectively (B).

Discovery set	*n*	Validation set	*n*
Normal healthy control		**Normal healthy control**	
No. cases	16	No. cases	34
Age (average ± standard deviation)	55 ± 10	Age (average ± standard deviation)	54 ± 8
Breast cancer		**Non-malignant breast disease**	
No. cases	26	No. cases	12
Age (average ± standard deviation)	55 ± 13	Age (average ± standard deviation)	43 ± 20
Tumour stage		**Type**	
IV	26	Fibroadenoma of the breast	6
ER status (positive/negative)	19/6 (1*)	Fibrocystic mastopathy of the breast	1
Type		Fibrocystic change with fibroadenoma of the breast	2
Invasive lobular carcinoma	6	Duct papillomas of the breast	1
Invasive ductal carcinoma	20	Atypical ductal hyperplasia of the breast	2
		Breast cancer	
		No. cases	129
		Age (average ± standard deviation)	54 ± 2
		Tumour stage	
		I	29
		II	30
		III	20
		IV	50
		ER status (positive/negative)	101/22(6*)
		HER-2 status (positive/negative)	21/101 (7*)
		Type	
		Invasive lobular carcinoma	20
		Invasive ductal carcinoma	109

Table 5-1. Patient data. Clinicopathologic features of the cancer samples used for proteomics and metabolomics analyses. The discovery sample set was used to generate the data displayed in Tables 5-2, 5-3 and 5-4. The validation sample set was used to evaluate 13-HODE, 12-HETE, Fibronectin, Glutamate, Beta Hydroxybuterate, Factor V, CA15-3, CA125, VEGF, Serotonin and MMP-1 in a range of different sample groups (normal healthy controls, non-malignant breast disease and breast cancer (stages I–IV) a not determined.

HMDB	Biochemical name	Platform	Fold change	Highest mean condition	Lowest mean condition	*p*-value
HMDB00172	Isoleucine	LC/MS pos	1.5	Breast Cancer	Control	≤0.001
HMDB03374	Ornithine	GC/MS	2.7	Breast Cancer	Control	≤0.001
n/a	Cyclo(phe-phe)	GC/MS	2.1	Breast Cancer	Control	≤0.001
HMDB00673	Linoleate	LC/MS neg	2.2	Breast Cancer	Control	≤0.001
HMDB00806	Myristate (14:0)	LC/MS neg	2.2	Breast Cancer	Control	≤0.001
HMDB02259	Margarate (17:0)	LC MS neg	2.4	Breast Cancer	Control	≤0.001
n/a	10-heptadecenoate (17:ln7)	LC MS neg	3.0	Breast Cancer	Control	≤0.001
HMDB00827	stearate(18:0)	LC MS neg	1.7	Breast Cancer	Control	≤0.001
n/a	10-nonadecenoate (19:ln9)	LC MS neg	3.1	Breast Cancer	Control	≤0.001
HMDB06111	12-HETE	LC/MS neg	37.8	Breast Cancer	Control	≤0.001
HMDB01388	Linolenate [alpha or gamma; (18:3n3 or 6)]	LC/MS neg	2.7	Breast Cancer	Control	≤0.01
HMDB00772	Nonadecanoate (19:0)	LC/MS neg	2.0	Breast Cancer	Control	≤0.01
HMDB00207	Oleate (18:ln9)	GC/MS	2.9	Breast Cancer	Control	≤0.01
HMDB02231	Eicosenoate (20:ln9 or 11)	LC/MS neg	3.0	Breast Cancer	Control	≤0.01
HMDB02000	Myristoleate (14:ln5)	LC/MS neg	3.0	Breast Cancer	Control	≤0.01
HMDB00162	Proline	LC/MS pos	1.7	Breast Cancer	Control	≤0.01
n/a	13-HODE + 9-HODE	LC/MS neg	3.7	Breast Cancer	Control	≤0.01
HMDB00191	Aspartate	GC/MS	2.0	Breast Cancer	Control	≤0.01
HMDB03339	Glutamate	LC/MS pos	2.3	Breast Cancer	Control	≤0.01
HMDB03736	3-methyl-2-oxovalerate	LC/MS neg	1.6	Breast Cancer	Control	≤0.05
HMDB00357	3-hiydroxybutyrate	GC/MS	2.9	Breast Cancer	Control	≤0.05
n/a	Dihomo-linoleate (20:2n6)	LC/MS neg	2.1	Breast Cancer	Control	≤0.05
HMDB00765	Mannitol	GC/MS	3.7	Breast Cancer	Control	≤0.05
HMDB00259	Serotonin (5HT)	LC/MS pos	1.8	Control	Breast Cancer	≤0.05
HMDB00267	5-Oxoproline	LC/MS pos	1.7	Breast Cancer	Control	≤0.05
HMDB00672	Hexadecanedioate	LC/MS neg	2.1	Breast Cancer	Control	≤0.05
HMDB00169	Mannose	GC/MS	1.9	Breast Cancer	Control	≤0.05
HMDB00706	Aspartylphenylalanine	LC/MS pos	2.4	Breast Cancer	Control	≤0.05

Table 5-2. Metabolomics screen. List of metabolites found to be differentially expressed at statistically significant levels when comparing control (n = 8) to stage IV breast cancer (n = 16) patient serum samples. Included in the table is the Human Metabolome Database (HMDB) entry, biochemical name, platform used to analyse the biochemicals, fold-change, highest/lowest mean change and p value. (n/a = not available).

Gene symbol	Protein Identification	Platform	Peptides used for quantification	Confidence score	Fold change	Highest mean condition	Lowest mean Condition	*p*-value
H4	Histone H4	LC/MS	2	90.8	111.1	Breast cancer	Control	≤0.001
ACTB	Actin, cytoplasmic 1	LC/MS	1	111.6	13.4	Breast cancer	Control	≤0.001
VTDB	Vitamin D-binding protein	LC/MS	1	55.4	3.5	Breast cancer	Control	≤0.001
1433F	14-3-3 protein eta	LC/MS	1	52.6	9.9	Breast cancer	Control	≤0.001
BID	Biotinidase	LC/MS	1	64.2	3.8	Breast cancer	Control	≤0.001
CIS	Complement Cl	LC/MS	5	382.1	1.8	Breast cancer	Control	≤0.001
FMC	Fibronectin	LC/MS	36	3030.0	2.8	Breast cancer	Control	≤0.001
APOA1	Apolipoprotein A-1	LC/MS	1	54.4	5.7	Breast cancer	Control	≤0.001
FIB A	Fibrinogen alpha chain	LC/MS	2	209.6	2.4	Breast cancer	Control	≤0.001
CBPB2	Carboxypeptidase B2	LC/MS	2	130.9	3.7	Breast cancer	Control	≤0.001
ITIH3	Inter-alpha-trypsin inhibitor heavy chain H3	LC/MS	5	385.8	4.0	Breast cancer	Control	≤0.001
ALEU	Serum albumin	LC/MS	2	290.3	3.3	Breast cancer	Control	≤0.001
FA9	Coagulation factor IX	LC/MS	1	56.1	3.9	Breast cancer	Control	≤0.01
HBB	Hemoglobin subunit beta	LC/MS	3	157.2	6.7	Breast cancer	Control	≤0.01
CLUS	Clusterin	LC/MS	2	121.9	1.8	Breast cancer	Control	≤0.01
AACT	Alpha-1-antichymotrypsin	LC/MS	4	287.1	4.0	Breast cancer	Control	≤0.01
LG3BP	Galectin-3-binding protein	LC/MS	5	400.6	2.3	Breast cancer	Control	≤0.01
FBLN1	Fibulin-1	LC/MS	2	203.8	1.9	Breast cancer	Control	≤0.01
IGHM	Ig mu chain C region	LC/MS	1	46.0	5.5	Control	Breast cancer	≤0.01
FA5	Coagulation factor V	LC/MS	3	266.5	1.8	Breast cancer	Breast cancer	≤0.01
CERU	Ceruloplasmin	LC/MS	1	41.8	4.2	Control	Breast cancer	≤0.01
ACTC	Actin, alpha cardiac muscle 1	LC/MS	1	116.4	5.6	Breast cancer	Control	≤0.01
IGHA1	Ig alpha-1 chain C region	LC/MS	1	54.9	5.5	Breast cancer	Control	≤0.01
A1AG1	Alpha-1-acid glycoprotein 1	LC/MS	3	173.3 5	3.8	Breast cancer	Control	≤0.01
A1AT	Alpha-1-antitrypsin	LC/MS	9	756.5	2.5	Breast cancer	Control	≤0.01
APOB	Apolipoprotein B-100	LC/MS	2	108.2	1.7	Breast cancer	Control	≤0.01
HEMO	Hemopexin	LC/MS	1	56.8	2.7	Breast cancer	Control	≤0.01
CRP	C-reactive protein	LC/MS	5	274.0	10.4	Breast cancer	Control	≤0.01
IC1	Plasma protease CI inhibitor	LC/MS	1	67.0	2.6	Breast cancer	Control	≤0.01
PLF4	Platelet factor 4	LC/MS	1	40.7	2.4	Control	Breast cancer	≤0.01

APOA2	Apolipoprotein A-II	LC/MS	1	44.0	82.3	Control	Breast cancer	≤0.01
HPT	Haptoglobin	LC/MS	3	168.9	2.4	Breast cancer	Control	≤0.01
C1R	Complement C1r subcomponent	LC/MS	4	250.9	1.8	Breast cancer	Control	≤0.01
FIBG	Fibrinogen gamma chain	LC/MS	1	40.3	1.5	Breast cancer	Control	≤0.01
HBA	Hemoglobin subunit alpha	LC/MS	2	220.5	7.9	Breast cancer	Control	≤0.01
IGHG4	Ig gamma-4 chain C region	LC/MS	1	40.9	3.1	Breast cancer	Control	≤0.01
A2AP	Alpha-2-antiplasmin	LC/MS	2	96.5	1.7	Breast cancer	Control	≤0.01
VTNC	Vitronectin	LC/MS	1	74.9	1.5	Breast cancer	Control	≤0.05
HABP2	Hyaluronan-binding protein 2	LC/MS	1	41.5	1.5	Breast cancer	Control	≤0.05
C03	Complement C3	LC/MS	4	199.0	1.5	Control	Breast cancer	≤0.05
THRB	Prothrombin	LC/MS	2	89.7	1.3	Breast cancer	Control	≤0.05
C04A	Complement C4-A	LC/MS	10	737.1	1.5	Breast cancer	Control	≤0.05
APOA4	Apolipoprotein A-IV	LC/MS	2	99.6	3.9	Control	Breast cancer	≤0.05

Table 5-3. Label-free proteomics screen List of proteins found to be differentially expressed at statistically significant levels when comparing control (n = 16) to stage IV breast cancer (n = 26) patient serum samples. Included in the table is the gene symbol, protein identification, platform used to analyse the proteins, foldchange, highest/lowest mean change and *p*-value.

Gene symbol	Protein Identification	Platform	Fold change	Highest mean condition	Lowest mean condition	p-value
PLR	Prolactin	LuminexxMAP	1.8	Breast cancer	Control	≤0.01
CA15-3	Cancer Antigen 15-3	LuminexxMAP	3.6	Breast cancer	Control	≤0.001
CA125	Cancer Antigen 125	LuminexxMAP	2.0	Breast cancer	Control	≤0.05
CA19-9	Cancer Antigen 19-9	LuminexxMAP	3.9	Breast cancer	Control	≤0.05
MMP-1	Matrix metalloproteinase-1	LuminexxMAP	1.9	Breast cancer	Control	≤0.001
MMP-9	Matrix metalloproteinase-9	LuminexxMAP	1.9	Breast cancer	Control	≤0.001
EGF	Epidermal growth factor	LuminexxMAP	3.2	Breast cancer	Control	≤0.001
VEGF	Vascular endothelial growth factor	LuminexxMAP	4.1	Breast cancer	Control	≤0.01

Table 5-4. Multiplex proteomics screen List of proteins found to be differentially expressed at statistically significant levels when comparing control (n = 16) to stage IV breast cancer (n = 26) patient serum samples. Included in the table is the gene symbol, protein identification, platform used to analyse the proteins, foldchange, highest/lowest mean change and *p*-value.

Biomarker	Control				Non-malignant breast disease				Stage I			
	Mean ± SD	Median	Range	*n*	Mean ± SD	Median	Range	*n*	Mean ± SD	Median	Range	*n*
13 HODE (ng/ml)	92 ± 132	34.0	15-374	34	42 ± 12	39.0	29-73	12	45 ± 22	41	22-133	29
12 HETE (ng/ml)	101.7 ± 216.5	9.8	1.3-1055.4	34	573.8 ± 587.6	326.5	26.6-1624.4	12	175.0 ± 317.4	37.3	3.0-1365.8	29
Fibronectin (lg/ml)	245 ± 127	241.0	59-634	34	357 ± 92	354.0	173-529	12	241 ± 120	258	26-427	29
Glutamate (nmoL/ml)	14.7 ± 4.3	14.0	11.7-28.6	34	15.5 ± 3.5	16.3	13.0-18.6	12	15.32± 4.8	14.9	12.0-30.9	29
Beta Hydroxybutyrate (nmol/ml)	14.5 ± 3.2	13.5	10.8-22.5	34	16.1 ± 3.7	15.2	12.6-24.0	12	17.2 ± 4.6	16.1	11.5-29.7	29
Factor V (lg/ml)	7.6 ± 3.4	6.6	3.9-20.7	34	12.0 ± 5.3	10.3	5.6-19.1	12	11.6 ± 4.3	10.8	5.9-25.1	29
CA15-3 (U/ml)	3.4 ± 2.6	2.8	0.5-12.4	34	3.0 ± 1.6	3.0	1.0 -5.7	12	8.0 ± 16.9	4.6	0.7-93.8	29
CA125 (U/ml)	8.6 ± 5.0	7.2	3.8-31.2	34	9.2 ± 5.4	7.9	3.1-21.5	12	68.3 ± 272.4	10.4	1.8-1480.3	29
VEGF (pg/ml)	13.6 ± 5.9	13.7	4.4-26.5	34	16.0 ± 7.2	16.3	2.5-31.1	12	24.2 ± 20.7	16.1	2.5-92.0	29
Serotonin (ng/ml)	56.5 ± 32.9	50.9	13.3-158.2	34	49.4 ± 20.8	46.1	15.9-103.3	12	41.6 ± 32.3	28.1	15.4-116.3	29
MMP-1 (pg/ml)	3419 ± 4209	2269	122-22701	34	4055 ± 4334	3044	360-13519	12	2406 ± 2789	1116	122-7448	29

Biomarker	Stage II				Stage III				Stage IV			
	Mean ± SD	Median	Range	*n*	Mean ± SD	Median	Range	*n*	Mean ± SD	Median	Range	*n*
13 HODE (ng/ml)	41 ± 18	36.0	20-89	30	50 ± 21	43.0	22-100	20	63 ± 38	54	19-180	16
12 HETE (ng/ml)	300.1 ± 506.4	65.1	7.1-2201.4	30	208.8 ± 274.2	67.0	2.4-585.2	20	1507.0 ± 543.3	1523	507.6-2318	50
Fibronectin (lg/ml)	242 ± 129	209.0	59-589	30	217 ± 120	222.0	37-516	20	155 ± 74	140	57-296	16
Glutamate (nmoL/ml)	13.1 ± 2.5	13.0	11.2-18.97	30	15.6 ± 7.0	13.8	11.7-40.4	20	22.1 ± 5.8	23	16.8-31.5	50
Beta Hydroxybutyrate (nmol/ml)	16.1 ± 4.7	14.2	12.4-33.7	30	19.7± 11.2	17.4	12.3-64.5	20	16.2 ± 4.8	14.4	11.5-26.2	16
Factor V (lg/ml)	12.0 ± 8.9	9.1	3.8-49.0	30	10.9 ± 3.5	10.3	6.2-18.1	20	14.7 ± 6.9	12.1	8.3-100	16
CA15-3 (U/ml)	6.1 ± 13.8	2.5	0.8-77.7	30	4.6 ± 3.7	3.8	0.6-14.4	20	26.0 ± 37.1	12.2	3.5-893.6	50
CA125 (U/ml)	30.1 ± 107.5	8.2	2.1-592.5	30	10.3 ± 12.4	6.4	2.3-57.2	20	128.3 ± 296.6	10.8	4.3-53.2	16
VEGF (pg/ml)	22.1 ± 15.1	16.9	6.0-71.0	30	16.2 ± 8.4	14.9	4.3-34.7	20	14.4 ± 11.2	11.9	8.0-141.5	16
Serotonin (ng/ml)	41.2 ± 24.0	33.7	14.8-87.1	30	394.2 ± 19.2	34.8	5.2-81.5	20	39.8 ± 26.2	36	19-76632	16
MMP-1 (pg/ml)	2240 ± 1937	1597	122-5324	30	2465 ± 1524	2301	122-4737	20	12749 ± 22443	4372	122-7448	16

Table 5-5. Validation results. Mean ± SD, median, range and *n* value data for a number of metabolites/proteins analysed in an independent sample set of control, non-malignant breast disease, stage I, stage II, stage III and stage IV breast cancer patient serum specimens.

Biomarker	Kmskal-Walbs chi squared	df	p-value	Significant
12 HETE	89	5	1.3E-17	Yes
13 HODE	8.2	5	0.15	No
Fibronectin	21	5	0.00074	Yes
Glutamate	15	5	0.012	Yes
Beta Hydroxybuterate	12	5	0.031	Yes
Factor V	38	5	4.6E-07	Yes
CA15-3	77	5	3.5E-15	Yes
CA125	11	5	0.058	No
VEGF	11	5	0.051	No
MMP-1	22	5	0.00047	Yes
Serotonin	9.8	5	0.082	No

Table 5-6. Kurskal–Wallis test The Kurskal–Wallis test was used to evaluate the significance of 12 HETE, 13 HODE, Fibronectin, Glutamate, Beta Hydroxybuterate, Factor V, CA15-3, CA125, VEGF, MMP-1 and Serotonin across all groups.

Biomarker combinations	Control v benign breast disease	Control v stage I	Control v stage II	Control v stage III	Control v stage IV	Benign breast disease v stage I	Benign breast disease v stage II	Benign breast disease stage III	Benign breast disease stage IV
CA15-3 + 2 HETE+glutamate	0.868 (0.0007)	0.671 (0.0238)	0.716 (0.0053)	0.684 (0.1592)	0.997 (<0.0001)	0.773 (0.0283)	0.758 (0.0375)	0.713 (0.1054)	0.977 (<0.0001)
CA15-3 +12 HETE+Beta-Hydroxybuterale	0.86 (0.0009)	0.71 (0.0077)	0.711 (0.0227)	0.791 (0.0038)	0.996 (<0.0001)	0.784 (0.0271)	0.692 (0.3664)	0.729 (0.0972)	0.932 (0.0001)
CA15-3 +12 HETE+Factor V	0.885 (\0.0001)	0.845 (0.0001)	0.798 (0.0011)	0.819 (0.0018)	0.998 (<0.0001)	0.799 (0.0159)	0.728 (0.2805)	0.754 (0.0407)	0.958 (<0.0001)
CA15-3 + 12 HETE+MMP1	0.887 (0.0008)	0.71 (0.0851)	0.671 (0.5520)	0.704 (0.2118)	0.997 (<0.0001)	0.675 (0.2917)	0.667 (0.2084)	0.712 (0.2024)	0.978 (<0.0001)
CA15-3 + glutamate+Beta-Hydroxybuterate	0.657 (0.1610)	0.71 (0.0077)	0.707 (0.0111)	0.778 (0.0093)	0.928 (<0.0001)	0.704 (0.1539)	0.758 (0.0423)	0.683 (0.2309)	0.906(0.0001)
CA15-3 + glutamate+Factor V	0.777 (0.0042)	0.845 (0.0001)	0.774 (0.0032)	0.8 (0.0103)	0.976 (<0.0001)	0.716 (0.1284)	0.75 (0.0352)	0.671(0.3541)	0.906 (0.0002)
CA15-3 + glutamate+MMPl	0.581 (0.7790)	0.645 (0.1886)	0.577 (0.7779)	0.683 (0.1334)	0.966 (<0.0001)	0.708 (0.3526)	0.7 (0.3587)	0.727 (0.2561)	0.965 (<0.0001)
CA15-3 + Beta-Hydroxybuterate+Factor V	0.777 (0.0042)	0.853 (\0.0001)	0.794 (0.0087)	0.819 (0.0008)	0.958 (<0.0001)	0.71 (0.1314)	0.633 (0.5492)	0.704 (0.1606)	0.849 (0.0021)
CA15-3 + Beta-Hydroxybuterate + MMPl	0.675 (0.1021)	0.797 (0.0090)	0.748 (0.0992)	0.727 (0.0310)	0.823 (0.0001)	0.717 (0.3382)	0.642 (0.5138)	0.682 (0.2494)	0.807 (0.0029)
CA15-3 + Factor V+MMP1	0.772 (0.0098)	0.868 (0.0062)	0.665 (0.6026)	0.812 (0.0405)	0.946 (<0.0001)	0.667 (0.3212)	0.758 (0.1102)	0.705 (0.2771)	0.833 (0.0019)

Table 5-7. Multivariate analysis LR analysis was performed on combinations of 3 candidate biomarkers found to be significant in the Kurskal–Wallis test, with the AUC and p value in brackets shown. CA15-3 was always included, with 12-HETE, Glutamate, Beta Hydroxybuterate, Factor V and MMP-1 also selected. Fibronectin was not included even-though this was significant as the results from the discovery and validation phases were opposing.
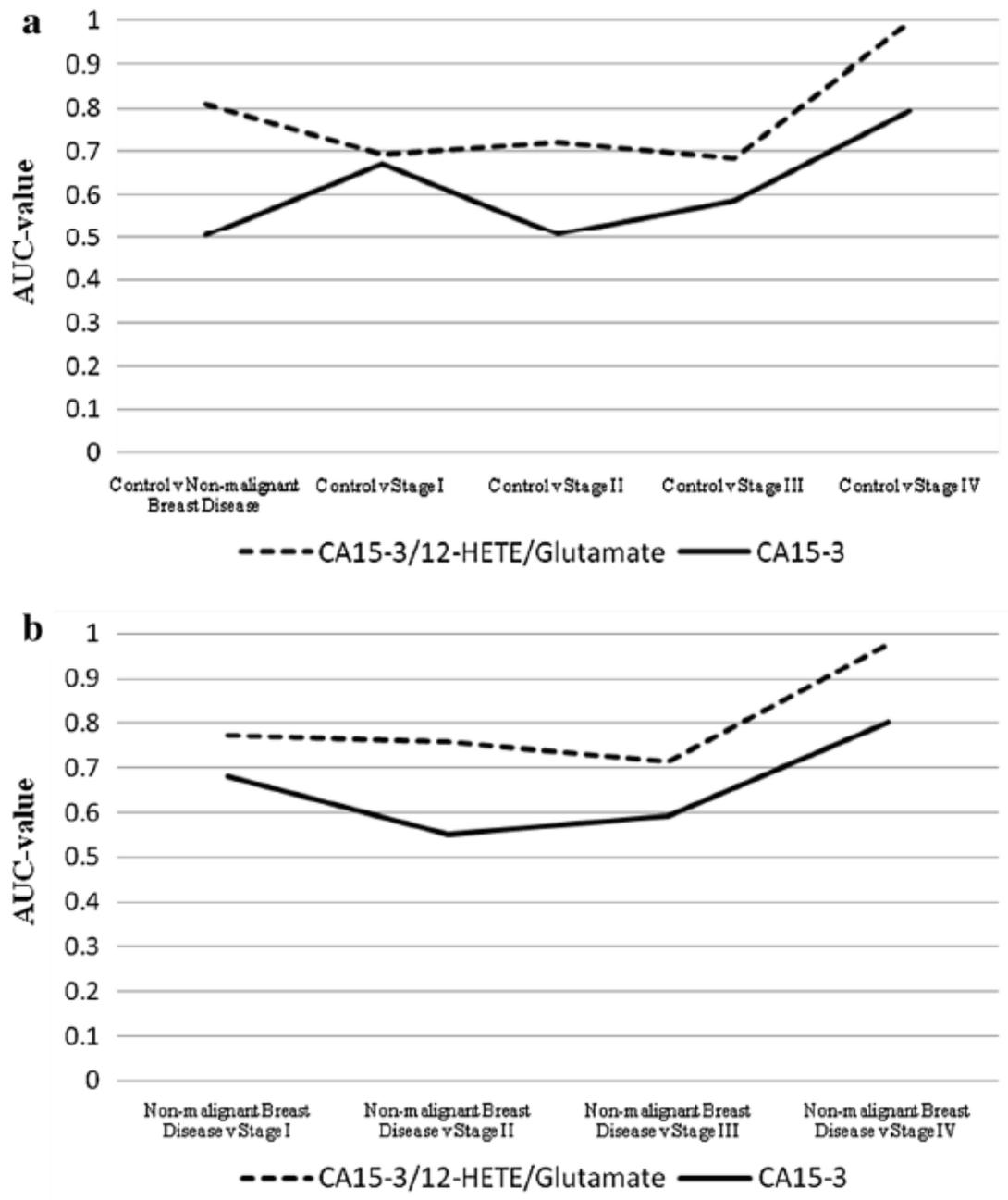
Figure 5-1. Logistic regression AUC-values. A Comparison of logistic regression AUC-values for the combination of CA15-3/ 12-HETE/Glutamate and CA15- 3 alone for control compared to non-malignant breast disease, stage I, stage II, stage III and stage IV breast cancer patient serum specimens respectively. b Comparison of logistic regression AUC-values for the combination of CA15-3/12- HETE/Glutamate and CA15-3 alone for non-malignant breast disease compared to stage I, stage II, stage III and stage IV breast cancer patient serum specimens respectively.

5.4 Discussion

5.4.1 Established tumour biomarkers

Circulating tumour markers are soluble molecules in blood, usually glycoproteins, detected by a single or pair of monoclonal antibodies. They are released into the blood by tumour cells or by other cells within the tumour microenvironment. Established markers are widely examined for detection of malignancies, for assessing outcome or predicting recurrence and for monitoring the response to anti-cancer therapies. Various studies have demonstrated that the diagnostic sensitivity of the CA15-3 test is about 10–15 % (stage I), 20–25 % (stage II) and 30–45 % (stage III) (Seregni et al. 2004). The accuracy of established tumour markers improves in advanced disease with levels of CA15-3 found to be significant in patients with distant metastases (Molina et al. 2005). Molina et al. found that combinations of CA15-3 and other clinically relevant biomarkers was a useful tool in the prognostic evaluation of patients with primary breast cancer (Molina, Auge, Farrus et al. 2010). This finding was also expanded by including serum HER-2 (only in those patients with tissue overexpression), with similar results (Molina, Auge, Escudero et al. 2010).

5.4.2 Clinical relevance

Clinically relevant and accurate biomarkers, reflecting tumour burden, present a minimally invasive clinical tests that clinicians can use to help further stratify patients along with other relevant tests. By assessing for example, information ascertained about the metastatic nature of the disease based on protein measurements, clinicians potentially can reduce over-treatment of localised tumours and avoid under-treatment of aggressive tumours.

The mainstay of early breast cancer detection is X-ray mammography. Researchers have not yet reached the necessary levels of sensitivity and specificity to routinely detect early stage breast cancer using single or multiple protein biomarkers, as found in this study with AUC-values failing to reach above 0.8 when comparing benign breast disease to stage I cancer. However, the X-ray mammogram, with its associated exposure to small levels of X-ray radiation, does have lower sensitivity in women with dense breasts and is unable to differentiate between morphologically similar benign or malignant breast lesions (Drukteinis et al. 2013). Therefore, protein biomarkers may complement existing scanning techniques in these situations.

5.4.3 Biomarker combinations in tumour burden measurement

Our results confirm a clear correlation between CA15-3 and disease progression, with stage IV breast cancer patient samples showing the highest levels of CA15-3 (AUC = 0.8 normal control v stage IV & AUC = 0.8 nonmalignant breast disease v stage IV). An example of three candidate biomarkers used in combination is CA15-3/ Glutamate/12-HETE, which was found to accurately reflect the degree of tumour burden when compared to both normal healthy control and non-malignant breast disease groups. The most significant AUC values was found when comparing normal and non-malignant breast disease groups to advanced breast cancer (AUC = 0.99 & 0.97 respectively). By the addition of both glutamate and 12-HETE measurements to CA15-3, there is a significant improvement in the AUC values compared to CA15-3 alone. As previously described, CA15-3 is an established biomarkers in breast cancer management with strict guidelines in place as to where these markers add value in patient care and was found to be an integral part of the three biomolecules panel for measuring tumour burden in this evaluation. Glutamate is a non-essential amino acid and an excitatory neurotransmitter in the central nervous system.

Glutamine, an abundant extracellular nutrient, is steadily converted into glutamate by phosphate-dependent glutaminase (GLS), an enzyme found within the inner mitochondrial membrane. The altered cellular metabolism displayed by malignant cells was first acknowledged most notably in Otto Warburg's description of a switch in glucose metabolism from oxidative phosphorylation to glycolysis, known as the Warburg effect (Koppenol, Bounds and Dang 2011). To help maintain a functioning citric acid cycle compromised due to the Warburg effect, cancer cells often rely on elevated glutamine metabolism (Erickson and Cerione 2010). As a result, GLS has been shown to be overexpressed in both solid tumours and cell line models of cancer (Katt et al. 2012, Maity, Chakraborty and Bhattacharya 2000). Interestingly, targeting GLS activity has been proposed as a potential therapeutic strategy against human malignancies (Zhao, Butler and Tan 2013).

Our findings support the major set of changes in cancer metabolism resulting in a shift to increased rates of glutamine metabolism, through the accelerated hydrolysis of glutamine to glutamate, as catalyzed by mitochondrial GLS activity. Using a standard glutamate biochemical assay for evaluating abundance, we found increased levels of glutamate in breast cancer patient serum samples compared to normal control and non-malignant breast disease. The most noticeable increase in abundance for glutamate was in the MBC group, perhaps reflecting the greatest increase in glutamine metabolism. Other groups have reported on elevated glutamate levels and associated signalling pathways in malignancies and

correlating this with a more aggressive phenotype (Speyer et al. 2014, Koochekpour et al. 2012, Ripka et al. 2010, Namkoong et al. 2007).

We also found increased levels of 12-HETE when comparing control groups (normal and non-malignant breast disease) to breast cancer patient samples. Similar to glutamate, the most significant elevation in abundance for this molecule was found in the MBC cohort. 12-HETE is an eicosanoid formed by the action of a 12-lipoxygenase (12-LO) enzyme on arachidonic acid. The evidence to date suggests a significant role for 12-HETE and 12-lipoxygenase, the enzyme responsible for the production of 12-HETE in many aspects of tumour biology, including cell growth, angiogenesis, metastatic potential and immunomodulation (Fürstenberger et al. 2006). Recent reports on 12-HETE involvement in tumour biology have found it to induce pro-metastatic protein expression patterns and show NF-jB-dependent up-regulation of the mesenchymal marker protein S100A4 and of transcriptional repressor ZEB1 secreted by MCF-7 spheroids (Vonach et al. 2011). Previously, Lui and co-workers described how 12-HETE increased the motility and invasion of AT2.1 cells (low metastatic rat prostate tumour cells) through activation of protein kinase C (PKC), a process that was inhibited by a selective PKC inhibitor, calphostin C (Liu et al. 1994). 12-lipoxygenase has been reported to be significantly increased in both breast cancer serum and tissue patient samples compared to healthy controls (Singh et al. 2011, Jiang, Douglas-Jones and Mansel 2003), with this increase found to be significantly decreased after chemotherapy (Singh et al. 2011).

Of the other interesting candidate biomarkers that were found to have significant AUC-values when used in combination with other proteins/metabolites, Factor V & MMP-1 have previous been found elevated in the blood proteome from cancer patients, especially in the advanced stages, reflecting the increasing role of the coagulation cascades in some cases, together with the function proteases play in tumour progression (Klee et al. 2012, Boström et al. 2011). Beta-hydroxybutyrate, one of three sources of ketone bodies, has previously been reported to be significantly increased in breast cancer, participating in regulating histone acetylation and gene expression in tumour development (Asiago et al. 2010). Reports on combining traditional biomarkers for breast cancer with other proteins have established this approach as very significant with impressive data to support their findings. Lawicik and co-workers demonstrated the usefulness of macrophage-colony stimulating factor (M-CSF) and VEGF in the diagnosis of breast cancer, but showed that only M-CSF was found to discriminate between cancer and non-carcinoma lesions, especially in combination with CA15-3 (Ławicki, Będkowska and Szmitkowski 2013). A recent publication by Park and co-workers describes how measurement of human cytosolic

thioredoxin in combination with established breast cancer biomarkers, including CA15-3, improved the overall accuracy for early detection of breast cancer (Park, Cha and Kim 2014).

5.5   Concluding remarks

Using a combination of different metabolomics and proteomics discovery platforms, we identified a number of significant biomolecules when comparing normal controls serum samples to a MBC cohort. Subsequently, we analysed a number of these significant molecules in a larger independent cohort of patient samples, comprising of normal controls, non-malignant breast disease and different stages of breast cancer. Established markers used in breast cancer patient care are an important resource in the clinical setting and our approach was to try to improve the overall accuracy of measurements of CA15-3 when combined with the most significant molecules coming out of this investigation.

In this investigation, advanced stage breast cancer serum samples were analysed in the discovery phase analysis to increase the chance of identifying abnormal proteins/metabolites due to the associated high tumour burden. However, it is likely that many advanced stage biomarkers do not represent the early stage phenotype, due to the specific processes and mechanisms activated throughout the course of tumour development. Nevertheless, many of the molecules selected for further analysis in the validation cohort did correlate with tumour burden, suggests this is a worthwhile strategy for initial biomarker discovery investigations.

It was established that glutamate, 12-HETE, Betahydroxybutyrate, Factor V and MMP-1 were the most significant and further analysis using combinations of these candidate biomarkers with CA15-3 resulted in noticeably improved AUC values compared to CA15-3 alone, especially when comparing advanced breast cancer groups to normal controls and non-malignant breast disease. The example expanded on in this discussion, shows how a combination of CA15-3/Glutamate/12-HETE is more accurate than CA15-3 alone, and indeed the same is true of other combinations including Beta-hydroxybutyrate, Factor V and MMP-1, especially in the advanced setting. Clinical applications of biomarkers that reflect tumour burden include screening in asymptomatic individuals, confirming a suspected diagnosis and assisting in tumour classification/staging. Many of the candidate biomarkers discussed in this study accurate correlate with tumour burden, especially in advanced stages where these biomarkers could potentially provide clinicians with additional confirmation on staging.

5.6   References

Asiago, V.M., Alvarado, L.Z., Shanaiah, N., Gowda, G.A., Owusu-Sarfo, K., Ballas, R.A. and Raftery, D. 2010. Early detection of recurrent breast cancer using metabolite profiling. *Cancer Research,* 70(21), pp.8309-8318.

Benson, J.R. and Jatoi, I. 2012. The global breast cancer burden. *Future Oncology,* 8(6), pp.697-702.

Bigbee, W.L., Gopalakrishnan, V., Weissfeld, J.L., Wilson, D.O., Dacic, S., Lokshin, A.E. and Siegfried, J.M. 2012. A multiplexed serum biomarker immunoassay panel discriminates clinical lung cancer patients from high-risk individuals found to be cancer-free by CT screening. *Journal of Thoracic Oncology,* 7(4), pp.698-708.

Boström, P., Söderström, M., Vahlberg, T., Söderström, K., Roberts, P.J., Carpén, O. and Hirsimäki, P. 2011. MMP-1 expression has an independent prognostic value in breast cancer. *BMC Cancer,* 11(1), pp.348.

Brayman, M., Thathiah, A. and Carson, D.D. 2004. MUC1: A multifunctional cell surface component of reproductive tissue epithelia. *Reproductive Biology and Endocrinology,* 2(1), pp.4.

Dowling, P., Clarke, C., Hennessy, K., Torralbo-Lopez, B., Ballot, J., Crown, J., Kiernan, I., O'byrne, K.J., Kennedy, M.J. and Lynch, V. 2012. Analysis of acute-phase proteins, AHSG, C3, CLI, HP and SAA, reveals distinctive expression patterns associated with breast, colorectal and lung cancer. *International Journal of Cancer,* 131(4), pp.911-923.

Drukteinis, J.S., Mooney, B.P., Flowers, C.I. and Gatenby, R.A. 2013. Beyond mammography: New frontiers in breast cancer screening. *The American Journal of Medicine,* 126(6), pp.472-479.

Duffy, M.J., Evoy, D. and McDermott, E.W. 2010. CA 15-3: Uses and limitation as a biomarker for breast cancer. *Clinica Chimica Acta,* 411(23-24), pp.1869-1874.

Duffy, M.J. 2013. Tumor markers in clinical practice: A review focusing on common solid cancers. *Medical Principles and Practice: International Journal of the Kuwait University, Health Science Centre,* 22(1), pp.4-11.

Duffy, M.J. 2006. Serum tumor markers in breast cancer: Are they of clinical value? *Clinical Chemistry,* 52(3), pp.345-351.

Erickson, J.W. and Cerione, R.A. 2010. Glutaminase: A hot spot for regulation of cancer cell metabolism? *Oncotarget,* 1(8), pp.734-740.

Fürstenberger, G., Krieg, P., Müller-Decker, K. and Habenicht, A. 2006. What are cyclooxygenases and lipoxygenases doing in the driver's seat of carcinogenesis? *International Journal of Cancer,* 119(10), pp.2247-2254.

Graham, L.J., Shupe, M.P., Schneble, E.J., Flynt, F.L., Clemenshaw, M.N., Kirkpatrick, A.D., Gallagher, C., Nissan, A., Henry, L., Stojadinovic, A., Peoples, G.E. and Shumway, N.M. 2014. Current approaches and challenges in monitoring treatment responses in breast cancer. *Journal of Cancer,* 5(1), pp.58-68.

Haagensen Jr, D.E., Kister, S.J., Vandevoorde, J.P., Gates, J.B., Smart, E.K., Hansen, H.J. and Wells Jr, S.A. 1978. Evaluation of carcinoembryonic antigen as a plasma monitor for human breast carcinoma. *Cancer,* 42(S3), pp.1512-1519.

Harris, L., Fritsche, H., Mennel, R., Norton, L., Ravdin, P., Taube, S., Somerfield, M.R., Hayes, D.F. and Bast Jr, R.C. 2007. American society of clinical oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *Journal of Clinical Oncology,* 25(33), pp.5287-5312.

Jiang, W.G., Douglas-Jones, A. and Mansel, R.E. 2003. Levels of expression of lipoxygenases and cyclooxygenase-2 in human breast cancer. *Prostaglandins, Leukotrienes and Essential Fatty Acids,* 69(4), pp.275-281.

Katt, W.P., Ramachandran, S., Erickson, J.W. and Cerione, R.A. 2012. Dibenzophenanthridines as inhibitors of glutaminase C and cancer cell proliferation. *Molecular Cancer Therapeutics,* 11(6), pp.1269-1278.

Koochekpour, S., Majumdar, S., Azabdaftari, G., Attwood, K., Scioneaux, R., Subramani, D., Manhardt, C., Lorusso, G.D., Willard, S.S., Thompson, H., Shourideh, M., Rezaei, K., Sartor, O., Mohler, J.L. and Vessella, R.L. 2012. Serum glutamate levels correlate with gleason score and glutamate blockade decreases proliferation, migration, and invasion and induces apoptosis in prostate cancer cells. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research,* 18(21), pp.5888-5901.

Ławicki, S., Będkowska, G.E. and Szmitkowski, M. 2013. VEGF, M-CSF and CA 15-3 as a new tumor marker panel in breast malignancies: A multivariate analysis with ROC curve. *Growth Factors,* 31(3), pp.98-105.

Liu, B., Maher, R.J., Hannun, Y.A., Porter, A.T. and Honn, K.V. 1994. 12 (S)-HETE enhancement of prostate tumor cell invasion: Selective role of PKCα. *JNCI: Journal of the National Cancer Institute,* 86(15), pp.1145-1151.

Lokich, J.J., Zamcheck, N. and Lowenstein, M.W. 1978. Sequential carcinoembryonic antigen levels in the therapy of metastatic breast cancer: A predictor and monitor of response and relapse. *Annals of Internal Medicine,* 89(6), pp.902-906.

Maity, P., Chakraborty, S. and Bhattacharya, P. 2000. A general survey of glutamine level in different tissues of murine solid tumor bearing mice before and after therapy with purified glutaminase. *Journal of Experimental & Clinical Cancer Research : CR,* 19(2), pp.161-164.

McLarty, K., Cornelissen, B., Scollard, D.A., Done, S.J., Chun, K. and Reilly, R.M. 2009. Associations between the uptake of 111In-DTPA-trastuzumab, HER2 density and response to trastuzumab (herceptin) in athymic mice bearing subcutaneous human tumour xenografts. *European Journal of Nuclear Medicine and Molecular Imaging,* 36(1), pp.81-93.

Meleady, P., Gallagher, M., Clarke, C., Henry, M., Sanchez, N., Barron, N. and Clynes, M. 2012. Impact of miR-7 over-expression on the proteome of chinese hamster ovary cells. *Journal of Biotechnology,* 160(3-4), pp.251-262.

Meleady, P., Hoffrogge, R., Henry, M., Rupp, O., Bort, J.H., Clarke, C., Brinkrolf, K., Kelly, S., Muller, B., Doolan, P., Hackl, M., Beckmann, T.F., Noll, T., Grillari, J., Barron, N., Puhler, A., Clynes, M. and Borth, N. 2012. Utilization and evaluation of CHO-specific sequence databases for mass spectrometry based proteomics. *Biotechnology and Bioengineering,* 109(6), pp.1386-1394.

Milan, E., Lazzari, C., Anand, S., Floriani, I., Torri, V., Sorlini, C., Gregorc, V. and Bachi, A. 2012. SAA1 is over-expressed in plasma of non small cell lung cancer patients with poor outcome after treatment with epidermal growth factor receptor tyrosine-kinase inhibitors. *Journal of Proteomics,* 76 Spec No.pp.91-101.

Mirabelli, P. and Incoronato, M. 2013. Usefulness of traditional serum biomarkers for management of breast cancer patients. *BioMed Research International,* 2013pp.685641.

Molina, R., Auge, J.M., Escudero, J.M., Filella, X., Zanon, G., Pahisa, J., Farrus, B., Munoz, M. and Velasco, M. 2010. Evaluation of tumor markers (HER-2/neu oncoprotein, CEA, and CA 15.3) in patients with locoregional breast cancer: Prognostic value. *Tumour Biology : The Journal of the International Society for Oncodevelopmental Biology and Medicine,* 31(3), pp.171-180.

Molina, R., Auge, J.M., Farrus, B., Zanon, G., Pahisa, J., Munoz, M., Torne, A., Filella, X., Escudero, J.M., Fernandez, P. and Velasco, M. 2010. Prospective evaluation of carcinoembryonic antigen (CEA) and carbohydrate antigen 15.3 (CA 15.3) in patients with primary locoregional breast cancer. *Clinical Chemistry,* 56(7), pp.1148-1157.

Molina, R., Barak, V., van Dalen, A., Duffy, M.J., Einarsson, R., Gion, M., Goike, H., Lamerz, R., Nap, M., Soletormos, G. and Stieber, P. 2005. Tumor markers in breast cancer- european group on tumor markers recommendations. *Tumour Biology : The Journal of the International Society for Oncodevelopmental Biology and Medicine,* 26(6), pp.281-293.

Monari, E., Casali, C., Cuoghi, A., Nesci, J., Bellei, E., Bergamini, S., Fantoni, L.I., Natali, P., Morandi, U. and Tomasi, A. 2011. Enriched sera protein profiling for detection of non-small cell lung cancer biomarkers. *Proteome Science,* 9(1), pp.55-5956-9-55.

Moreno-Aspitia, A., Hillman, D.W., Dyar, S.H., Tenner, K.S., Gralow, J., Kaufman, P.A., Davidson, N.E., Lafky, J.M., Reinholz, M.M., Lingle, W.L., Kutteh, L.A., Carney, W.P., Dueck, A.C. and Perez, E.A. 2013. Soluble human epidermal growth factor receptor 2 (HER2) levels in patients with HER2-positive breast cancer receiving chemotherapy with or without trastuzumab: Results from north central cancer treatment group adjuvant trial N9831. *Cancer,* 119(15), pp.2675-2682.

Mughal, A.W., Hortobagyi, G.N., Fritsche, H.A., Buzdar, A.U., Yap, H.Y. and Blumenschein, G.R. 1983. Serial plasma carcinoembryonic antigen measurements during treatment of metastatic breast cancer. *Jama,* 249(14), pp.1881-1886.

Nakata, B., Ogawa, Y., Ishikawa, T., Ikeda, K., Kato, Y., Nishino, H. and Hirakawa, K. 2000. Serum CYFRA 21-1 is one of the most reliable tumor markers for breast carcinoma. *Cancer,* 89(6), pp.1285-1290.

Namkoong, J., Shin, S.S., Lee, H.J., Marin, Y.E., Wall, B.A., Goydos, J.S. and Chen, S. 2007. Metabotropic glutamate receptor 1 and glutamate signaling in human melanoma. *Cancer Research,* 67(5), pp.2298-2305.

Park, B.J., Cha, M.K. and Kim, I.H. 2014. Thioredoxin 1 as a serum marker for breast cancer and its use in combination with CEA or CA15-3 for improving the sensitivity of breast cancer diagnoses. *BMC Research Notes,* 7pp.7-0500-7-7.

Ripka, S., Riedel, J., Neesse, A., Griesmann, H., Buchholz, M., Ellenrieder, V., Moeller, F., Barth, P., Gress, T.M. and Michl, P. 2010. Glutamate receptor GRIA3--target of CUX1 and mediator of tumor progression in pancreatic cancer. *Neoplasia (New York, N.Y.),* 12(8), pp.659-667.

Senkus, E., Cardoso, F. and Pagani, O. 2014. Time for more optimism in metastatic breast cancer? *Cancer Treatment Reviews,* 40(2), pp.220-228.

Seregni, E., Coli, A., Mazzucca, N. and Italian Group RIA-IRMA Test, Italian Association of Nuclear Medicine. 2004. Circulating tumour markers in breast cancer. *European Journal of Nuclear Medicine and Molecular Imaging,* 31 Suppl 1pp.S15-22.

Singh, A.K., Kant, S., Parshad, R., Banerjee, N. and Dey, S. 2011. Evaluation of human LOX-12 as a serum marker for breast cancer. *Biochemical and Biophysical Research Communications,* 414(2), pp.304-308.

Soerjomataram, I., Lortet-Tieulent, J., Parkin, D.M., Ferlay, J., Mathers, C., Forman, D. and Bray, F. 2012. Global burden of cancer in 2008: A systematic analysis of disability-adjusted life-years in 12 world regions. *Lancet (London, England),* 380(9856), pp.1840-1850.

Speyer, C.L., Hachem, A.H., Assi, A.A., Johnson, J.S., DeVries, J.A. and Gorski, D.H. 2014. Metabotropic glutamate receptor-1 as a novel target for the antiangiogenic treatment of breast cancer. *PloS One,* 9(3), pp.e88830.

Teh, J.L., Shah, R., La Cava, S., Dolfi, S.C., Mehta, M.S., Kongara, S., Price, S., Ganesan, S., Reuhl, K.R., Hirshfield, K.M., Karantza, V. and Chen, S. 2015. Metabotropic glutamate receptor 1 disrupts mammary acinar architecture and initiates malignant transformation of mammary epithelial cells. *Breast Cancer Research and Treatment,* 151(1), pp.57-73.

Tondini, C., Hayes, D.F., Gelman, R., Henderson, I.C. and Kufe, D.W. 1988. Comparison of CA15-3 and carcinoembryonic antigen in monitoring the clinical course of patients with metastatic breast cancer. *Cancer Research,* 48(14), pp.4107-4112.

Viola, K., Kopf, S., Rarova, L., Jarukamjorn, K., Kretschy, N., Teichmann, M., Vonach, C., Atanasov, A.G., Giessrigl, B., Huttary, N., Raab, I., Krieger, S., Strnad, M., de Martin, R., Saiko, P., Szekeres, T., Knasmuller, S., Dirsch, V.M., Jager, W., Grusch, M., Dolznig, H., Mikulits, W. and Krupitza, G. 2013. Xanthohumol attenuates tumour cell-mediated breaching of the lymphendothelial barrier and prevents intravasation and metastasis. *Archives of Toxicology,* 87(7), pp.1301-1312.

Vonach, C., Viola, K., Giessrigl, B., Huttary, N., Raab, I., Kalt, R., Krieger, S., Vo, T.P., Madlener, S., Bauer, S., Marian, B., Hammerle, M., Kretschy, N., Teichmann, M., Hantusch, B., Stary, S., Unger, C., Seelinger, M., Eger, A., Mader, R., Jager, W., Schmidt, W., Grusch, M., Dolznig, H., Mikulits, W. and Krupitza, G. 2011. NF-kappaB mediates the 12(S)-HETE-induced endothelial to mesenchymal transition of lymphendothelial cells during the intravasation of breast carcinoma cells. *British Journal of Cancer,* 105(2), pp.263-271.

Zhang, C., Yuan, X.R., Li, H.Y., Zhao, Z.J., Liao, Y.W., Wang, X.Y., Su, J., Sang, S.S. and Liu, Q. 2015. Anti-cancer effect of metabotropic glutamate receptor 1 inhibition in human glioma U87 cells: Involvement of PI3K/Akt/mTOR pathway. *Cellular Physiology and Biochemistry : International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology,* 35(2), pp.419-432.

Zhao, Y., Butler, E.B. and Tan, M. 2013. Targeting cellular metabolism to improve cancer therapeutics. *Cell Death & Disease,* 4pp.e532.

7B7: a novel antibody directed against the Ku70/Ku80 heterodimer blocks invasion in pancreatic and lung cancer cells

Authors: Dermot O'Sullivan*, Michael Henry*, Helena Joyce, Naomi Walsh, Edel Mc Auley, Paul Dowling, Niall Swan, Michael Moriarty, Paul Barnham, Martin Clynes, Annemarie Larkin

* contributed equally to this study.

Within this work, Michael Henry specifically performed the methodology and verification of the immunoprecipitation technique used. Michael Henry performed experiments followed by SDS-PAGE protein separation and gel staining experiments. Michael Henry conducted the in-gel protein digestions and was responsible for the LC-MS/MS analysis. Michael Henry implemented the protein identification analysis tools. In the writing of the publication prepared the proteomic sections in the materials and methods and results section.

Abstract

Development of more effective therapeutic strategies for cancers of high unmet need requires the continued discovery of disease-specific protein targets for therapeutic antibody targeting. In order to identify novel proteins associated with cancer cell invasion/metastasis, we present here an alternative to antibody targeting of cell surface proteins with an established role in invasion; our functional antibody screening approach involves the isolation and selection of MAbs that are primarily screened for their ability to inhibit tumour invasion. A clonal population of the Mia PaCa-2, a pancreatic ductal adenocarcinoma (PDAC) cell line, which displays a highly invasive phenotype, was used to generate MAbs with the objective of identifying membrane targets directly involved in cancer invasion. Selected MAb 7B7 can significantly reduce invasion in a dose-responsive manner in Mia PaCa-2 clone 3 and DLKP-M squamous lung carcinoma cells. Using immunoprecipitation and liquid chromatography tandem mass spectrometry (LC-MS-MS) analysis, the target antigen of anti-invasive antibody, 7B7, was determined to be the heterodimeric Ku antigen, Ku70/80, a core protein composed of the Ku70 and Ku80 subunits which is involved in non-homologous end-joining (NHEJ) DNA repair. RNA interference-mediated knockdown of Ku70 and Ku80 resulted in a marked decrease in the invasive capacity of Mia PaCa-2 clone 3 and DLKP-M cells, indicating that Ku70/Ku80 is functionally involved in pancreatic and lung cancer invasion. Immunohistochemical analysis demonstrated Ku70/Ku80 immunoreactivity in 37 PDAC tumours, indicating that this heterodimer is highly expressed in this aggressive cancer type. This study demonstrates that a functional MAb screening approach coupled with immunoprecipitation/proteomic analyses can be successfully applied to identify functional anti invasive MAbs and potential novel targets for therapeutic antibody targeting.

Keywords: Monoclonal antibody. Cancer metastasis. Cancer invasion. Ku70/Ku80. Target discovery

6.1 Introduction

Metastasis remains the major driver of mortality in patients with cancer, and, in contrast to primary cancer, metastasis is a major therapeutic challenge; currently, there are no effective treatments for the vast majority of metastatic cancers. Only a very small number of therapeutic targets have been investigated/evaluated clinically for their ability to limit metastasis—none of these are in clinical use (Weber 2013, Eckhardt et al. 2012). Identification of targets with a role in cancer invasion/metastasis may have the potential to be developed as novel treatment strategies to treat (or delay) metastasis. Monoclonal-antibody-based treatment of cancer has, in the last two decades, become established as one of the most successful therapeutic strategies for treating patients with haematological and solid tumours (Scott, Wolchok and Old 2012). Antibody cancer therapeutics are highly specific and potent drugs, which exert their therapeutic effect through modulation of key signalling pathways to reduce cell growth or induce apoptosis and/or triggering of antibody-unique effect or functions. There are now 14 antibody-based products approved to treat cancer patients in the European Union (EU) and the USA (Lambert 2013, Chames et al. 2009). Therapeutic cancer antibodies are, however, focused on a relatively narrow set of targets; one third of all antibodies (approved and in development) are specific for just five targets (Reichert 2014, Lambert 2013). Currently, the key challenge in new antibody cancer therapeutics is the continued discovery of novel targets that are associated with hallmarks of cancer pathophysiology such as metastatic potential and drug sensitivity that are suitable for antibody-based targeting. The ideal tumour target antigen will be easily accessible from the tumour neovasculature and shows abundant tumour expression with limited normal tissue expression.

First generation antibodies such as trastuzumab and rituximab show significant clinical benefits and have revolutionised treatment of human epidermal growth factor receptor 2 (HER-2) positive breast cancer and CD-20 positive lymphomas. However, conventional unmodified antibodies have limitations such as low tumour to blood ratio due to long serum half-life, limited tissue penetration and specificity for a single epitope, and they can sometimes be limited in efficacy (Chames et al. 2009). There has been a drive to improve established targeted therapeutic antibodies using glycoengineering approaches to increase potency (e.g. development of non-fucosylated antibodies) and the use of antibody fragments for better tissue penetration. Development of bispecific antibodies (bsABs) (e.g. catumaxomab) and trifunctional antibodies is a big growth area; these antibodies can bind to two different tumour antigens or to a tumour antigen and an antigen in the

tumour stroma/microenvironment such as an immune killer cell (Sliwkowski and Mellman 2013, Scott et al. 2012). Antibody-drug conjugates (ADCs) such as the recently approved trastuzumab-DM1 aim to overcome the limitations of non-specific cytotoxics and ineffective MAb therapy, through the use of a highly potent cytotoxic agent (payload) which is directed to the tumour by appending it to an antibody or antibody fragment. ADCs currently account for around 15% of the clinical stage anticancer-antibody-based pipeline outnumbering bispecific antibodies and antibody fragments (Mullard 2013). Currently, the key challenge for ADC cancer therapeutics is the ongoing identification of tractable targets (Sievers and Senter 2013).

Large numbers of potential cancer-associated antigens have been identified in a number of different cancer types using various high-throughput "omics" platforms (Roti and Stegmaier 2012). However, for the most part, these approaches did not distinguish between targets that are directly involved in the cancer pathology from those that are merely bystanders with no functional contribution to tumourogenesis. All of these candidate targets, obviously then, require extensive characterisation and validation to define their relevant functional role, if any, in cancer. It is imperative that a greater emphasis is placed upon targets that are functionally relevant to cancer development and progression. The significant contribution of phenotypic screening is well recognised in first in-class small molecule drug discovery (Swinney and Anthony 2011). The most commonly used approach to therapeutic antibody discovery has been to generate antibodies that are specifically directed at pre-defined, well-characterised cell surface proteins or newly described proteins that are less well characterised, with respect to cancer. Early antibody candidates are generally identified using traditional hybridoma technology developed by Köhler and Milstein or phage display platforms which are adapted easily for high-throughput antibody generation (Köhler and Milstein 1975). Therapeutic antibody discovery is now also starting to focus on MAb targets that are functionally relevant to the cancer disease process, with "function first" discovery or "phenotypic antibody screening" approaches being described for identification of candidate MAbs that are functionally involved in cancer tumourogenesis (Frendéus 2013, Rust et al. 2013, Kinch et al. 2009).

Our laboratory previously isolated isogenic populations of the PDAC cell line, Mia PaCa-2 clone 3, which displays high levels of invasion *in vitro*, and Mia PaCa-2 clone 8, which is poorly invasive *in vitro* (Walsh et al. 2008). The use of isogenic cell lines with different phenotypes with respect to metastasis is a very useful model to aid in the identification of potential cancer targets. In this study, we have used a functional MAb screening approach (based on ability to inhibit tumour invasion *in vitro*) to generate functional anti-invasive MAbs using the Mia PaCa-2 clone 3 cell line as an

immunogen, in order to identify potential targets with a direct role in cancer invasion. Immunoprecipitation followed by liquid chromatography tandem mass spectrometry (LC-MS-MS) analyses was used for target identification of the selected anti-invasive antibody.

6.2 Materials and Methods

6.2.1 Cell lines

The triple-negative breast carcinoma (TNBC), MDA-MB- 231, HER-2 positive breast, SKBR-3, large cell lung NCIH1229 and the HCT-116 colon carcinoma cell lines were obtained from the American Tissue Culture Collection (ATCC), Rockville, MD, USA. The SNB-19 glioma cell line was obtained from the German Collection of Microorganisms and Cell Cultures. The Mia PaCa-2 clone 3 pancreatic cancer cell line and the DLKP-M squamous lung carcinoma cell line were established in our own laboratory (Walsh et al. 2008, McBride et al. 1998).

6.2.2 Hybridoma generation

The immunogen chosen for the Mab generation was Mia PaCa-2 clone 3, an invasive clonal variant of the pancreatic cell line Mia PaCa-2 (Walsh et al. 2008). Prior to immunisations, a non-enzymatic cell dissociation buffer was used to remove cells from flasks (2×10^6 cells (50 µl volume)) following which cells were washed and re-suspended in phosphate-buffered saline (PBS) plus an equal volume of lipopeptide adjuvant, Pam3Cys-SKKKK (EMC micro-collections, Germany; L2000). Each mouse received a primary immunisation administered by an intraperitoneal injection, followed by four booster immunisations administered at three weekly intervals. Three days prior to cell fusion, mice received one final immunisation. Cell fusion of the immunised spleen B-cells with SP/2/O-Ag myeloma cells was carried out using a modification of the fusion protocol outlined by Kohler and Milstein (Köhler and Milstein 1975). Fused cells were plated out in 48-well plates and incubated undisturbed for 10–12 days in a humidified chamber at 37°C/5%CO2. Immunisations were carried out at the Royal College of Surgeons (RCSI) Biomedical Research Facility, Beaumont Hospital, Dublin 9, Ireland; all immunisation protocols were licensed by the Department of Health and Children, Ireland.

6.2.3 Hybridoma generation

The immunogen chosen for the Mab generation was Mia PaCa-2 clone 3, an invasive clonal variant of the pancreatic cell line Mia PaCa-2 (Walsh et al. 2008). Prior to immunisations, a non-enzymatic cell dissociation buffer was used to remove cells from flasks (2×106 cells (50 µl volume)) following which cells were washed and re-suspended in phosphate-buffered saline (PBS) plus an equal volume of lipopeptide adjuvant, Pam3Cys-SKKKK (EMC micro-collections, Germany; L2000). Each mouse received a primary immunisation administered by an intraperitoneal injection, followed by four booster immunisations administered at three weekly intervals. Three days prior to cell fusion, mice received one final immunisation. Cell fusion of the immunised spleen B-cells with SP/2/O-Ag myeloma cells was carried out using a modification of the fusion protocol outlined by Kohler and Milstein (Köhler and Milstein 1975). Fused cells were plated out in 48-well plates and incubated undisturbed for 10–12 days in a humidified chamber at 37°C/5%CO2. Immunisations were carried out at the Royal College of Surgeons (RCSI) Biomedical Research Facility, Beaumont Hospital, Dublin 9, Ireland; all immunisation protocols were licensed by the Department of Health and Children, Ireland.

6.2.4 Screening of hybridomas

Hybridomas were allowed to form large colonies and grow for at least 10 days undisturbed before supernatant was removed to screen for specific antibody production. Hybridoma supernatants were screened directly for their ability to inhibit cell invasion in Mia PaCa-2 clone 3 cells and for their ability to show membrane immunoreactivity on unfixed Mia PaCa-2 clone 3 cells.

6.2.5 Immunofluorescence studies

Cells were suspended to a concentration of 2×10^{6} cells/ml, added to 10-well 7-mm microscope slides (Thermo Fisher Scientific), and incubated overnight at 37°C. Excess supernatant was removed, and the microscope slides were rinsed gently with PBS. Hybridoma supernatants were added to each well (PBS was added to one well as a negative control). For fixed cell immunofluorescence studies, cells were fixed in ice cold acetone for 2 min, followed by air-drying for 15 min prior to application of antibody 7B7. All slides were incubated overnight at 4°C. Primary antibodies were removed by gently washing three times in PBS. A volume of 30 µl of secondary antibody, fluorescein isothiocyanate (FITC)-linked rabbit anti-mouse Ig (Dako, F0261), diluted 1/40, was added to each well, followed by a 30-min incubation at room temperature. Secondary antibody was removed by gently washing three times in PBS. Each slide was

mounted using VECTASHIELD mounting medium (Vector, H-100), cover slipped and viewed using a Nikon phase contrast microscope fitted with an FITC filter. 96-well invasion assays. This screening assay was adapted from the standard Boyden chamber transwell assay described previously (Walsh et al. 2009), which is a modification of the method of Albini et al. (Albini 1998). Briefly, using Millipore 96-well invasion plates (Millipore, MAMIC8S10), 50 µl of Matrigel (1 mg/ml) (BD Biosciences, 354234) was dispensed into each insert; the plate was incubated overnight at 4°C, followed by the removal of any excess Matrigel. Cells were harvested at a concentration of 2.5×10^4 cells/ml in media, and 50 µl was added to each insert, along with an equal volume of antibody or control hybridoma medium (no antibody). A volume of 150 µl media was added to each insert well. Cells were incubated at 37 °C for 48 h. Following this, the inserts were washed with PBS to remove any cells, while the outside of the insert was stained with 0.25% crystal violet. Stained inserts were viewed using a Nikon inverted microscope and scored as having either decreased levels of invasion or no effect on invasion relative to control hybridoma medium inserts (no hybridoma supernatant, representing 100 % invasion). Those supernatants showing a decrease in invasion were re-screened using this same method.

6.2.6 24-well invasion assays

For 24-well invasion assays, Costar 24-well plates (Costar, 3524), containing 8.0 µm pore-sized inserts (BD Biosciences, 353097), were used. A volume of 100 µl of Matrigel was dispensed into each insert, a cell suspension of 1×10^6 cells/ml was used, and 100 µl of antibody or control hybridoma medium (no antibody) was added to each insert; 500 µl media were added to each insert well. Cells were incubated at 37°C for either 24 or 48 h, depending on the cell line used. The total number of invading cells was determined by counting the number of cells per field in 10 random fields at ×200 magnification. The average number of cells per field was then multiplied by a factor of 140 as described previously (Walsh et al. 2009).

6.2.7 Motility assays

Motility assays were carried out in an identical manner to 24- well invasion assays, as described above, with the exception that the inserts were not coated in Matrigel.

6.2.8 Proliferation assays

Cells were harvested at a concentration of 2×10^4 cells/ml in media. Volumes of 100 µl/well of these cell suspensions were added to a 96-well plate (Costar, 3596) using a multichannel pipette. Plates were gently agitated to ensure an even dispersion of cells over a given plate. Cells were incubated at 37°C in 5 % CO_2 overnight. Following this, 100 µl of antibody was added to each well. Control wells had 100 µl hybridoma media added to the cell suspension. Plates were gently agitated, as described above, and incubated at 37°C in 5 % CO2 for 6/7 days, until the control wells have reached 80–90 % confluency. Assessment of cell survival in the presence of the antibodies relative to control (no MAb—hybridoma control medium only) was determined by acid phosphatase assay as previously described (Martin and Clynes 1991).

6.2.9 Adhesion assays

Adhesion assays were performed as described previously (Walsh et al. 2009). A volume of 250 µl of extracellular matrix (ECM) proteins was placed into each well of 24-well plates and incubated overnight at 4°C. Excess solution was removed from the wells, followed by washing two times with sterile PBS. To reduce non-specific binding, 0.5 ml of sterile 0.1 % BSA/PBS solution was dispensed into each well. The plates were incubated at 37°C for 20 min and then rinsed twice again with sterile PBS. Cells were harvested and re-suspended in media at a concentration of 2.5×10^4 cells/ml. Aliquots of 1 ml of cells were plated onto 24-well plates, in triplicate, and incubated for 60 min. The medium was then removed from the wells and rinsed gently with sterile PBS. The cell number attached was assessed using the acid phosphatase assay.

6.2.10 Zymography

DLKP-M cells were treated with MAb 7B7 (50 µg/ml) or control hybridoma (no MAb) for 24 h and subsequently transferred into serum-free medium for the collection of conditioned medium for zymographic analysis (over 72 h). Conditioned medium was concentrated 20× using an Amicon Ultra Centrifugal Filter device (5 k MW cut-off) and loaded onto Novex gelatin (Invitrogen EC6175) and casein (Invitrogen EC6405) gels using a 2× non-reducing sample buffer (Novex), renatured in zymogram renaturing buffer and developed in zymogram developing buffer (Invitrogen). Gels were stained with Brilliant blue G Colloidal Coomassie (Sigma, B2025); zones of gelatinolytic/casein activity were detected as clear bands against a blue background.

6.2.11 Isotype analysis

Isotyping of MAb 7B7 was carried out using the Isostrip Monoclonal Antibody Isotyping kit (Roche Diagnostics Gmbh, 1493027). Prior to immunoprecipitation, MAb 7B7 was purified using the Pierce NAb Spin Purification Kit (Pierce, 20530) and dialyzed using a Pierce Slide-A-Lyzer MINI Dialysis Unit (Pierce, 69576). Cells were grown to 80–90 % confluency, removed from flasks, pelleted and washed three times in PBS. Whole cell extracts were prepared by re-suspending cells in 1.5 ml of radio-immunoprecipitation assay (RIPA) lysis buffer (Sigma, R0278), sonicating for 5 min and centrifuging at 14,000 rpm for 10 min at 4°C. The 7B7 antigen(s) were immunoprecipitated from these samples using the Direct IP Kit from Pierce (45335), as per the manufacturer's instructions. Briefly, protein-L purified MAb 7B7 was covalently immobilised on aldehyde activated agarose beads overnight at 4°C. The antigen (whole cell extracts of Mia PaCa-2 clone 3 or DLKP-M cells) was then added to immobilised antibody gel mixture together with binding buffer on a rocking platform overnight at 4°C. The antigens were eluted from the beaded agarose into the elution buffer (IgM Elution Buffer, Pierce) using microcentrifuge spin cups. Fractions were neutralised through the addition of Tris-HCl, pH 8.8. Eluted fractions were prepared for gel electrophoresis using NuPAGE LDS sample buffer (Invitrogen, NP007). Following gel electrophoresis, gels were stained with Brilliant Blue Colloidal Coomassie to visualise precipitated proteins. For validation of mass spectrometry identifications, the gels were probed with relevant antibodies (Ku70 and Ku80 specific antibodies, Abcam, UK) for Western blot analysis.

6.2.12 Protein identification using LC-MS-MS

LC-MS/MS was performed on an UltiMate 3000 Nano-LC system (Dionex), interfaced to an LTQ Orbitrap XL (Thermo Fisher Scientific). Preparation of samples was based on the method described by Shevehenko et al. (Shevchenko et al. 2006). Briefly, protein bands were excised from the Coomassie-stained gel and were destained. Samples were then dehydrated and digested, and peptides were extracted with acetonitrile/0.1% formic acid. A 5 µl peptide mixture was loaded onto a C18 trap column (C18 PepMap, 300 µm id × 5 mm, 5 µm particle size, 100 Å pore size; Dionex) for desalting before separation on an analytical PepMap C18 column (75 µm id × 250 mm, 3 µm particle and 100 Å pore size; Dionex). Peptides were separated over a 60 minute period at a flow rate of 350 nl/min into the Orbitrap XL. The instrument was operated in data-dependent mode. Survey MS scans were acquired in the Orbitrap in the 400–1500 m/z range a resolution of 30,000 at m/z 400 and CID fragmentation was carried out in the linear ion trap with the ten most intense ions per scan.

6.2.13 Bioinformatic profiling of identified proteins

Protein database searches were performed using TurboSEQUEST software (Bioworks Browser version 3.3.1) (Thermo Fisher Scientific) using the human subset from the SWISS-PROT database. The following search parameters were used for protein identification: (i) MS/MS mass tolerance set at 0.6 Da, (ii) peptide mass tolerance set to 20 ppm, (iii) carbamidomethylation set as a fixed modification, (iv) methionine oxidation set as a variable modification, (v) up to two missed cleavages were allowed and (vi) for charge state 1, xcorr>1.5; for charge state 2, xcorr>2.5; and for charge state 3, xcorr>3.5 were accepted.

6.2.14 RNA interference-mediated knockdown of Ku70 and Ku80

Two pre-designed small interfering RNAs (siRNAs, Ambion) were chosen that target Ku70 and Ku80, respectively (Ku70—Ambion s14952 and s14953; Ku80—Ambion s5455 and s52594), and transiently transfected into cells. Each set of siRNA transfections carried out was accompanied with a control (non-transfected) and a scrambled (SCR) transfection (Ambion, 4390843). Solutions of siRNA at a final concentration of 30 nm were prepared in Opti-MEM (Invitrogen). The siRNA experiments were set up using 2-μl NeoFX (Ambion, AM4511), to transfect 30 nm siRNA at a cell density of 3× 105 ml in a six-well plate. Transfection medium was removed after 24 h and replaced with fresh growth medium. A kinesin specific siRNA (Ambion) was used as a positive control in all knockdown experiments to assess the efficiency of transfections. The transfected cells were collected at 72 and 96 h for immunoblotting and assayed for changes in invasion capacity at 24 or 48 h using the *in vitro* invasion assay, as described above.

6.2.15 Immunoblotting

Whole cell lysates were extracted using RIPA buffer (Sigma) containing protease inhibitor cocktail (Roche) and centrifuged for 15 min at 16,000g at 4°C. Protein concentrations were determined using the BCA protein assay kit (Pierce) according to the manufacturer's instructions. Concentrations of 15-30μg of protein (whole cell extracts or immunoprecipitated proteins) were separated using 4–12 % gradient gels (Invitrogen) under reducing conditions. Proteins were transferred to PVDF membranes (Roche Diagnostic GmbH). The membranes were blocked for 2 h at room temperature with Starting Block^{TM} Blocking Buffer (Pierce). The membranes were probed with MAb 7B7 supernatants, anti-Ku70 and anti-Ku80 antibodies (Abcam, UK), β-actin or α-tubulin (T5168, A5441, Sigma). The membranes were washed

three times for 5 min with 1× Tris-buffered saline (TBS)/Tween-20 (0.5 %v/v) and incubated with an anti-mouse HRP-conjugated secondary antibody (Dako) for 1 h at room temperature and washing step repeated. Detection was performed with ECL reagents and Amersham Hyperfilm™ chemiluminescence film (GE Healthcare).

6.2.16 Immunohistochemistry

A commercially available pancreatic cancer tissue microarray (TMA) (PA1001, US BioMax), comprising of duplicate tissue cores from 40 pancreatic cancers (including 37 pancreatic ductal adenocarcinoma (PDAC) tumours) and 6 adjacent non-cancer tissues, was stained for Ku70/Ku80 expression using an anti-Ku70/Ku80 polyclonal antibody suitable for immunohistochemistry (IHC) (Abcam, UK). Staining was performed using an automated IHC staining apparatus (Autostainer, Dako) according to the manufacturer's guidelines. Briefly, following de-paraffinisation and rehydration, the slides were subjected to an antigen retrieval step consisting of 20-min incubation in pH 6.0 buffer (Target Retrieval, Dako). The slides were counterstained with haematoxylin, dehydrated in graded alcohols and glass-mounted (DPX, Sigma). Negative control (primary antibody omitted, replaced by TBS) and positive control (tonsil) slides were processed at the same time. TMA cores were scored according to the intensity of the Ku70/Ku80 immunoreactivity observed (weak, moderate, and strong).

6.2.17 Quantification of MAb 7B7

MAb 7B7 was quantified using the mouse IgM ELISA quantification from Bethyl Laboratories, USA (E90-101).

6.2.18 Statistical analysis

All data are presented as mean ± standard deviation (SD). Analysis of the difference of comparisons, scrambled control siRNA versus siRNA-treated mean invasion and motility counts, adherence absorbance, and percentage survival calculated was performed using the Student's t test (two-tailed, unpaired with equal variance). A p-value of $\leq$0.05, $\leq$0.01, and $\leq$0.005 was deemed statistically significant. Statistical analysis was performed using Microsoft Excel. All experiments were repeated a minimum of three times, with the exception of initial hybridoma invasion screening assays and IHC analysis of Ku70/Ku80.

6.3   Results

6.3.1   Isolation of anti-invasive MAbs

Since trypsin may lead to the loss of cell surface antigens, a non-enzymatic cell disassociation buffer (Sigma, C5914) was used to remove cells from flasks prior to preparation of immunogen and for all screening assays for antibody characterisation studies. Following fusion, 335 hybridoma supernatants (77.5 % fusion efficiency) were screened using a 96-well Boyden chamber invasion assay platform for their ability to block invasion in Mia PaCa-2 clone 3 cells. Selected hybridoma supernatants were re-screened twice (duplicate samples) using the 96-well invasion Boyden chamber assay method. Seven supernatants, 8B3, 1E5, 7C3, 7B7, 9D6, 5B1 and 6E2, were selected on the basis of a decrease in invasion observed in relation to invasion observed on controls (hybridoma control medium). These seven hybridomas were then expanded for further characterisation. Supernatants (triplicate samples) were re-screened by both 96-well and 24-well invasion assays. Hybridomas 7B7 and 6E2 were selected based on their ability to inhibit invasion. In parallel, hybridoma supernatants were screened for membrane fluorescence immunoreactivity using an FITC-labelled secondary antibody, on unfixed Mia PaCa-2 clone 3 cells. Twenty-two supernatants (from a total of 335 screened) showed strong immunoreactivity on Mia Paca-2 clone 3 cells. Hybridomas 8F4, 3E2, 5B5, 1E5, 6D6 and 7B7 showed strong immunoreactivity; 8F4 and 7B7 showed more intense staining in Mia PaCa-2 clone 3 cells compared with the poorly invasive Mia PaCa- 2 clone 8 cells. Selected antibodies 7B7 and 6E2 were cloned by limiting dilution and re-screened (duplicate samples) using both the 24-well and the 96-well invasion assay methods. The 7B7 clone G52 was selected for further study as it showed the strongest anti-invasive effects in Mia PaCa-2 clone 3 cells and showed strong cell surface expression on these cells (Figure 6-1 Figure 6-2) (note that 7B7 G52 is referred to, throughout this manuscript as 7B7, its parental name).

6.3.2   Isotyping of MAb 7B7

Using a commercial isotyping kit, the isotype of antibody 7B7 was determined to be an IgM (data not shown).

6.3.3 MAb 7B7 impedes the invasive potential of aggressive cancer cells *in vitro*

In order to ascertain if antibody 7B7 could also impede invasion in other aggressive cancer cell types, we also investigated its anti-invasive effects in glioma, lung, colon and breast cancer cell lines. Antibody 7B7 blocked the ability of DLKP-M squamous lung carcinoma cells (mean inhibition level of 32.2 %) (Figure 6-2) and H1229 large cell lung carcinoma cells (mean inhibition level of 30.6 %) to invade through Matrigel (Figure 6-5). The anti-invasive effects of 7B7 in the other aggressive cancer types were not as marked: SNB-19 glioma, mean inhibition level of 27 %; HCT-116 colon adenocarcinoma, mean inhibition level of 24.2 %; SKBR-3 HER-2 positive breast, mean inhibition level of 25.6 %; and TNBC MDA-MB-231, mean inhibition level of 23 % in the presence of identical concentrations of 7B7 (50 µg/ml) (Figure 6-5

Figure 6-6) (mean inhibition, figures from a minimum of three independent experiments). To further characterise the 7B7 functional blocking antibody, its ability to inhibit invasion in a dose-responsive manner was determined in Mia PaCa2-clone 3 cells and DLKP-M squamous carcinoma cells (Figure 6-1). Antibody 7B7 significantly inhibited invasion in a dose-responsive manner in both cell lines; represented histograms are presented in figure 6-1. Furthermore, 7B7 was also found to inhibit motility in a dose-responsive manner in Mia PaCa-2 clone 3 cells (see supplementary data https://link.springer.com/article/10.1007%2Fs13277-014-1857-5); effects on motility in other aggressive cell types were not determined, so it cannot be concluded if antibody 7B7 is capable of blocking cancer cell motility.

6.3.4 MAb 7B7 recognises a membrane associated target antigen

Immunofluorescence staining of highly invasive Mia PaCa-2 clone 3 cells and drug resistant Mitox6p cells with 7B7 revealed strong membrane reactivity on both of these cell lines (Figure 6-2); weaker immunoreactivity was observed on poorly invasive Mia PaCa-2 clone 8 cells and DLKP-SQ, the parental drug-sensitive clone of DLKP-SQ Mitox-BCRP 6P. These results suggested that the 7B7 antigen may possibly be associated with an aggressive cancer phenotype, i.e. highly invasive/drug sensitive cell. Taken together with 7B7s ability to impede the invasive potential of various cancer cell types, the observed 7B7 immunoreactivity strongly suggests that this antibody binds to its native putative target antigen on the cell surface in these invasive cell lines.

6.3.5 Identification of the 7B7 target antigen

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) separation followed western blotting analysis with 7B7 revealed no potential reactive antigen band(s), suggesting that this antibody is recognising a conformational and not a liner epitope. We previously successfully used an immunoprecipitation approach coupled with proteomic analysis to identify the target antigen of antibody 5C3 (Larkin et al. 2005). As we had determined the isotype of antibody 7B7 to be an IgM, we used a modification of our previous approach utilising a commercially sourced immunoprecipitation kit with involved direct immobilisation of purified antibody to aldehyde-activated agarose beads. So, in addition to eliminating immunoprecipitation contaminating antibody, protein A/G is not involved, thus its suitability for IgM immobilisation. Cell extracts from Mia PaCa-2 clone 3 cells and DLKP-M cells were immunoprecipitated with MAb 7B7. SDS-PAGE and MS-compatible Coomassie Blue staining revealed two reactive bands of approximately 80 and 70 KDa, in immunoprecipitates from both cell lines which were not present on control mouse IgM immunoprecipitates. A representative gel is shown in Figure 6-3 showing reactive bands in immunoprecipitated Mia PaCa-2 clone 3 extracts. These two bands were excised and subjected to in-gel digestion followed by liquid chromatography mass spectrometry. They were identified as the ATP dependent DNA helicase 2 subunit 1 (Ku70) (molecular weight 69.79 KDa) and the ATP dependent DNA helicase 2 subunit 2 (Ku80) (molecular weight 82.65 KDa) (Table 6-1). Immunoprecipates were identified with commercial antibodies directed against Ku70 and Ku80, confirming these identifications (see supplementary data https://link.springer.com/article/10.1007%2Fs13277-014-1857-5).

6.3.6 RNA interference-mediated knockdown of Ku70 and Ku80 suppresses the invasive capacity of Mia PaCa-2 clone 3 and DLKP-M cells

As we had demonstrated that incubation with MAb 7B7 can significantly inhibit tumour invasion in a dose-responsive manner in the Mia PaCa-2 clone 3 and DLKP-M cell lines, we then determined the effects of siRNA-mediated knockdown on proliferation and invasion to establish if Ku70/Ku80 is functionally involved and thus has therapeutic relevance in these two aggressive cancer types. Two independent siRNAs were used to successfully knockdown both subunits of the Ku heterodimer in the two cell lines. Down regulation of both Ku70 and Ku80 significantly reduced the invasive capacity of both cell lines to invade compared with cells transfected with scrambled control siRNA as indicated by a marked decrease in the number of cells invading through matrigel (Figure 6-4 Figure 6-5). Cells

transfected with Ku70 and Ku80 siRNA grew at a similar rate to scrambled control transfected cells indicating that down regulation of Ku70/Ku80 has no effect on proliferation in these cell lines, indicating that the anti-invasive effects of antibody 7B7 are not secondary to the effects on cell growth (see supplementary data https://link.springer.com/article/10.1007%2Fs13277-014-1857-5). These siRNA knockdown experiments demonstrate that Ku70 and Ku80 are required for invasion of these aggressive PDAC and squamous lung carcinoma cell lines and thus are functionally involved in pancreatic and lung cancer invasion *in vitro*.

6.3.7 MAb 7B7 impedes the adhesion potential of Mia PaCa-2 clone 3 and DLKP-M cells

Adhesion of a tumourigenic cell to the basement membrane is a crucial stage in the invasion process. Thus, the effect of MAb 7B7 on the adhesive potential of Mia PaCa-2 clone 3 cells was assessed. In the presence of MAb 7B7, a small decrease in the adhesion of Mia PaCa-2 clone 3 cells to matrigel and DLKP-M cells to fibronectin was observed (see supplementary data https://link.springer.com/article/10.1007%2Fs13277-014-1857-5).

6.3.8 MAb inhibits MMP activity in DLKPM cells

Following treatment of squamous lung DLKP cells with antibody 7B7, effects on MMP activity were determined. As demonstrated in Figure 6.6, zymographic analysis of conditioned medium (collected over 72 h) from DLKP-M cells treated for 24 h with antibody 7B7 revealed reductions in pro and active MMP-10 and pro-MMP-2 compared with conditioned medium from cells treated with hybridoma control medium only. These results indicated that 7B7 may be eliciting its anti-invasive effects through reduction of MMP-2 and MMP-10 activities in DLKP-M cells.

6.3.9 IHC analysis of Ku70/Ku80 expression in pancreatic cancer

TMA cores were scored according to the intensity of the Ku70/Ku80 immunoreactivity observed (weak, moderate, and strong) (Table 6-2). Duplicate cores were stained in all cases; there was complete concordance between duplicates. Nuclear Ku70/Ku80 immunoreactivity was observed in all cases; some cytoplasmic localised staining was also observed. In addition, cytoplasmic staining of stromal cells was observed in a number of tumours. Ku70/Ku80 immunoreactivity was present in 40/40 tumours (37 PDAC, 2 acinar, and 1 squamous). All of the high-grade tumours studied exhibited either moderate or strong Ku70/Ku80 immunoreactivity, suggesting a possible trend of increased expression of Ku70/Ku80 with more aggressive PDAC tumours. Weak Ku70/Ku80 expression was also observed in 2/5 normal

pancreatic tissues; 3/5 tissues showed intermediate immunoreactivity. Representative

photomicrographs are shown in

Figure **6-7**. We were not able to establish if membrane Ku70/Ku80 expression is present in PDAC, as this

commercial antibody has not been demonstrated to be specific for Ku70/Ku80 membrane expression.
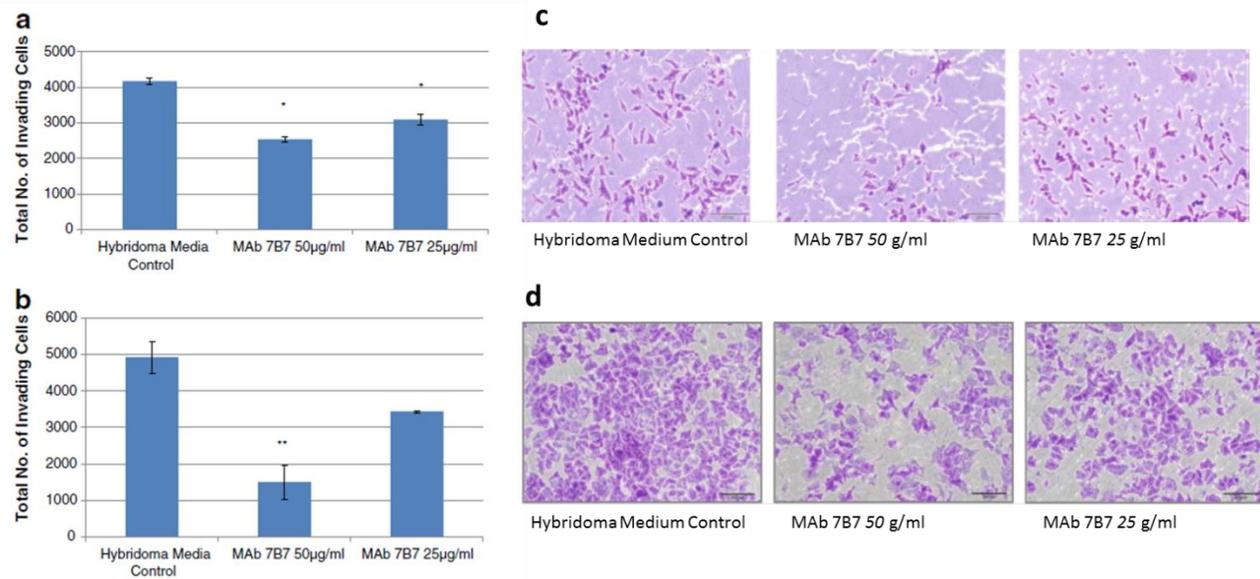
Figure 6-1. MAb 7B7 blocks cancer invasion in PDAC Mia PaCa-2 clone 3 cells and DLKP-M squamous lung cells. a The invasive potential of Mia PaCa-2 clone 3 cells is significantly reduced in the presence of 50 µg/ml of antibody 7B7; invasion is also reduced, to a lesser extent, in the presence of 25 µg/ml of antibody, indicating that MAB 7B7 can significantly reduce the invasive potential of this cell line in a dose responsive manner. b The invasive potential of DLKP-M cells is significantly reduced in the presence of 50 µg/ml of antibody 7B7; invasion is also reduced, to a lesser extent, in the presence of 25 µg/ml of antibody, indicating that MAB 7B7 can significantly reduce the invasive potential of this cell line in a dose-responsive manner. In both cases, the total number of cells invading through Matrigel is shown. Data plotted represent the mean ± standard deviation from a representative experiment carried out in duplicate. Statistics: *p≤0.05, **p≤0.01, and ***p≤0.005 compared with hybridoma medium control (no MAb). Student's t test (two-tailed with equal variance, unpaired). c Representative photomicrographs of Mia PaCa-2 clone cells invaded through Matrigel-coated Boyden invasion chambers (8-µm pore size) (stained with 0.25 % crystal violet (w/v) in the presence of two different concentrations of antibody (original magnification, ×100; scale bar= 200 µm). d Representative photomicrographs of DLKPM cells invaded through Matrigel-coated Boyden invasion chambers (8-µm pore size) (stained with 0.25 % crystal violet (w/v) in the presence of two different concentrations of antibody (original magnification, ×100; scale bar= 200 µm). All experiments were performed a minimum of three times, and representative results are presented.
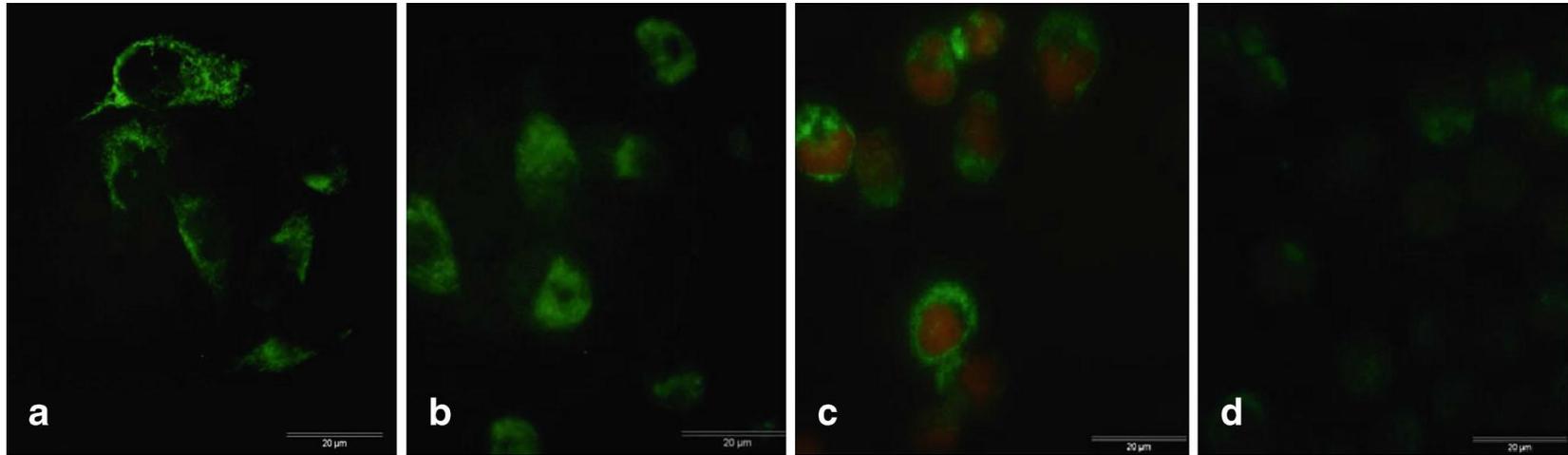
Figure 6-2. Immunofluoresence cell surface staining of antibody 7B7. MiaPaCa-2 clone 3 (a) and Mia PaCa-2 clone 8 (b) **pancreatic ductal adenocarcinoma cells and DLKP Mitox-BCRP 6P, drug-resistant lung squamous carcinoma cells (H. Joyce, unpublished)** (c) **with their sensitive parent cells, DLKPSQ** (d) **and stained with antibody 7B7. Membrane 7B7 immunoreactivity is observed in invasive Mia PaCa-2 clone 3 and drug-resistant DLKP Mitox-BCRP 6P cells, Mia PaCa-2 clone 8 cells which are poorly invasive** *in vitro* **show more cytoplasmic-like 7B7 immunoreactivity only, DLKPSQ cells show very weak 7B7 immunoreactivity. c, d Counter stained with propidium iodide. Original magnification of all photomicrographs, ×400, scale bar=20 μm).**

206
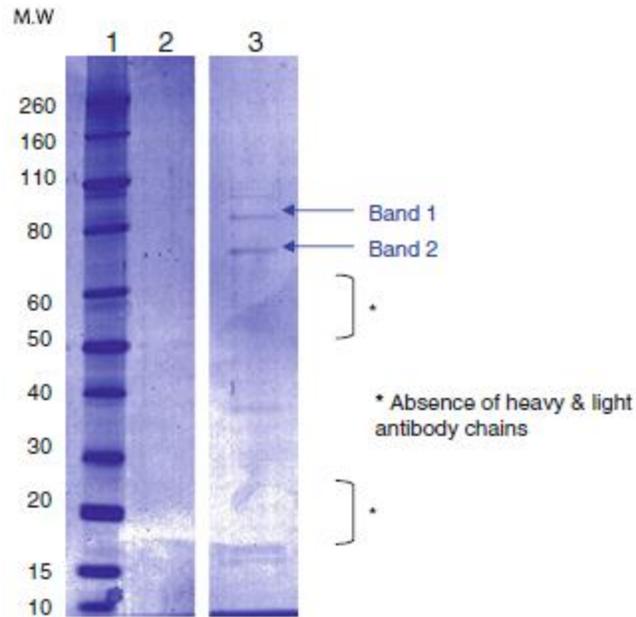
Figure 6-3. Identification of MAb 7B7 target antigens by immunoprecipitation. Cross-linked immunoprecipitation of cell lysates with 7B7MAb and control mouse IgM separated by SDS-PAGE and stained with 0.25 % Coomassie Blue. Lane 1 molecular weight markers, lane 2 Mia PaCa-2 clone 3 immunoprecipitated with mouse IgM, lane 3 Mia PaCa-2 clone 3 immunoprecipitated with MAb 7B7. Two bands are observed; at approximately 80 kDa, 70 kDa were excised and run through an Orbitrap XL mass spectrometer for identification. No antibody heavy and light chains are present. Non-binding agarose beads were also used as a control (data not shown). Experiment was performed three times, and a representative scanned gel is shown. DLKP-M cells were also immunoprecipitated with antibody 7B7 and resulted in the same 70 and 80-kDa reactive bands (data not shown).
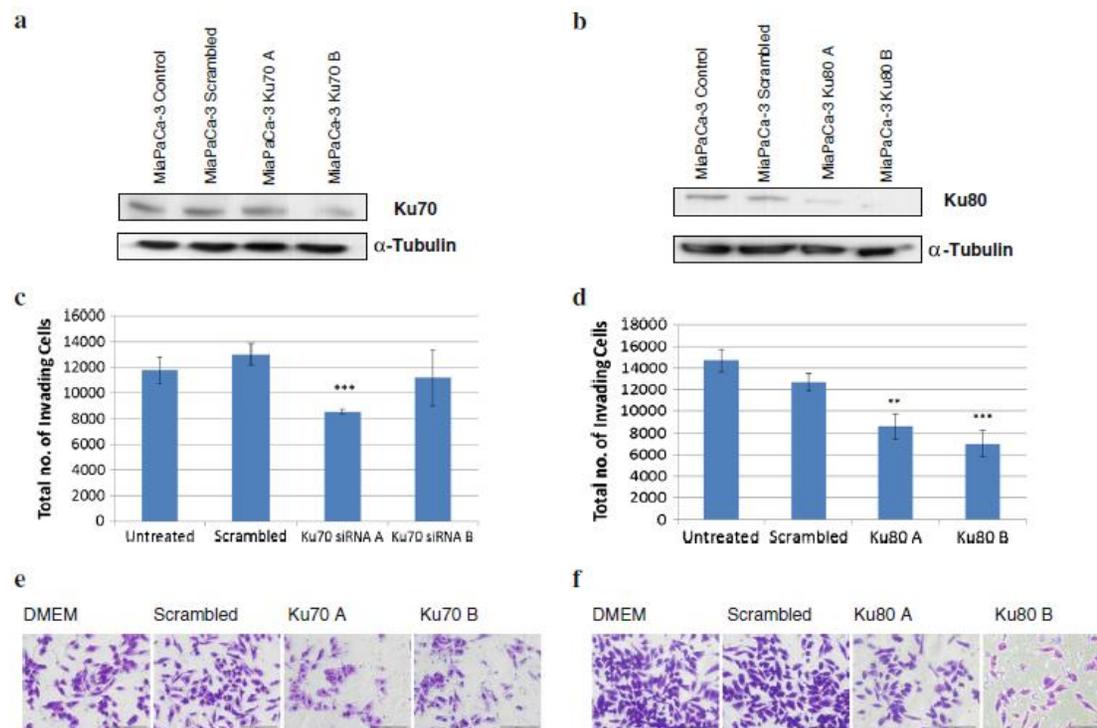
Figure 6-4. The siRNA-mediated interference knockdown of both Ku70 and Ku80 reduces invasive capacity of Mia PaCa-2 clone 3 cells. (a), (b) **Immunoblots showing effective knockdown of Ku70** (a) **and Ku80** (b**) at 48-h post-transfection in Mia PaCa-2 clone 3 cells transfected with two independent siRNAs targeting Ku70 and Ku80, respectively, relative to scrambled control siRNA-transfected cells. α-Tubulin served as a loading control.** (c), (d) **Representative histograms showing reduced invasive capacity of Mia PaCa-2 clone 3 cells in a Boyden chamber assay following transfection with Ku70- and Ku80-targeted siRNAs. The total number of cells invading through Matrigel is shown. Data plotted represent the mean ± standard deviation of duplicate transwell inserts from a representative experiment. Statistics: *p≤0.05, **p≤0.01, and ***p≤0.005 compared with scrambled siRNA control. Student's t test (two-tailed with equal variance, unpaired).** (e), (f) **Representative photomicrographs showing invasion status of Mia PaCa-2 clone 3 cells following transfection with siRNAs targeting Ku70** (e) **and Ku80** (f**), respectively. A decrease in invasion can be observed following siRNA transfections of both Ku70 and Ku80 compared with scrambled control insert. Original magnification, ×100; scale bar=200 μm. All experiments were performed in a minimum of three times, and representative results are presented.**
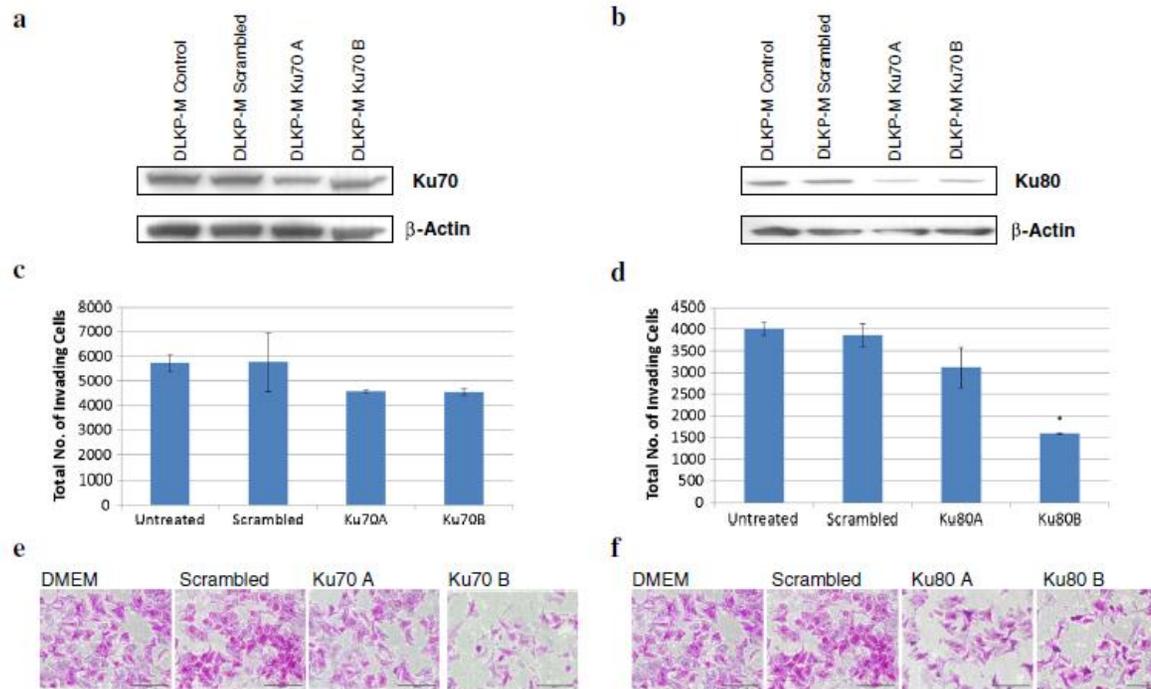
Figure 6-5. siRNA-mediated interference knockdown of both Ku70 and Ku80 reduces invasive capacity of DLKP-M cells. (a), (b) **Immunoblots showing effective knockdown of Ku70 (a) and Ku80 (b) at 48-h post-transfection in DLKP-M cells transfected with two independent siRNAs targeting Ku70 and Ku80, respectively, relative to scramble control siRNA-transfected cells. α-Tubulin served as a loading control. (c), (d) Representative histograms showing reduced invasive capacity of DLKP-M cells in a Boyden chamber assay following transfection with Ku70 and Ku80 siRNAs. The total number of cells invading through matrigel is shown. Data plotted represent the mean ± standard deviation of duplicate transwell inserts from a representative experiment. Statistics: *p≤0.05, **p≤0.01, and ***p≤0.005 compared with scrambled siRNA control. Student's t test (two tailed with equal variance, unpaired). e, f Representative photomicrographs showing invasion status of DLKP-M cells following transfection siRNAs targeting Ku70 (e) and Ku80 (f), respectively. A decrease in invasion can be observed following siRNA transfection of both Ku70 and Ku80 compared with scrambled control insert. Original magnification, ×100; scale bar=200 μm). All experiments were performed in a minimum of three times, and representative results are presented.**

Protein name	Gene symbol	Protein Accession	MW (kDa)
ATP-dependent DNA helicase 2 subunit 1	XRCC6	gi|169145199|	69.79
ATP-dependent DNA helicase 2 subunit 2	XRCC5	gi|125731|	82.65

Table 6-1. The 7B7 target antigens identified by immunoprecipitation/LCMS- MS. These proteomic identifications were validated by probing Mia PaCa-2 clone 3 and DLKP-M immunoprecipitates with commercial antibodies specific for both the Ku70 and Ku80 subunits of the Ku heterodimer, confirming these proteomic identifications.

Grade 3	5/12 strong
	7/12 moderate
Grade 2	3/12 strong
	5/12 moderate
	4/12 weak
Grade 1	3/13 strong
	7/13 moderate
	3/13 weak

Table 6-2. Ku70/Ku80 immunoreactivity in grades 1, 2 and 3 PDAC tumours.
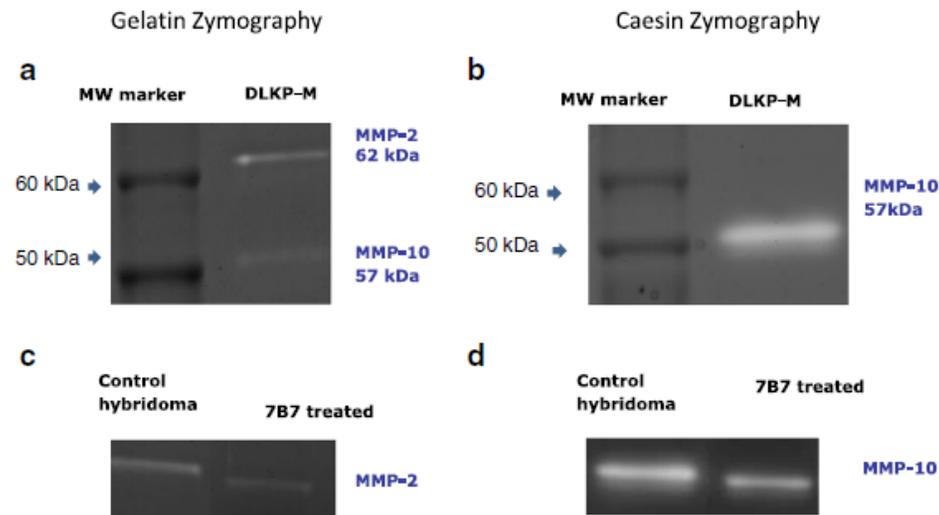
Figure 6-6. Antibody 7B7 reduces MMP activity in squamous lung carcinoma DLKP-M cells. MMP activity in DLKP-M cells. a Gelatin zymography analysis of conditioned medium collected over 72 h from DLKP-M cells showing presence of both pro and active (lower band) MMP-2 in DLKP-M cells. b Caesin zymography analysis of conditioned medium collected over 72 h from DLKP-M cells showing the presence of both pro and active (lower band) MMP-10 in DLKP-M cells. MMP activity in DLKP-M cells treated with antibody 7B7. c Gelatin zymography analysis of conditioned medium collected over 72 h following treatment of DLKP-M cells with 50 μg of MAb 7B7. A reduction in MMP-2 activity in DLKP-M antibody-treated cells is observed compared with control hybridoma medium only (no antibody). D Caesin zymography analysis of conditioned medium collected over 72 h following treatment with antibody 7B7. A marked reduction inMMP-10 activity (pro and active) is observed in DLKP-M cells treated with antibody 7B7 compared with cells treated with control hybridoma medium only (no antibody). Three independent experiments were performed in all cases, and representative zymograms are shown.
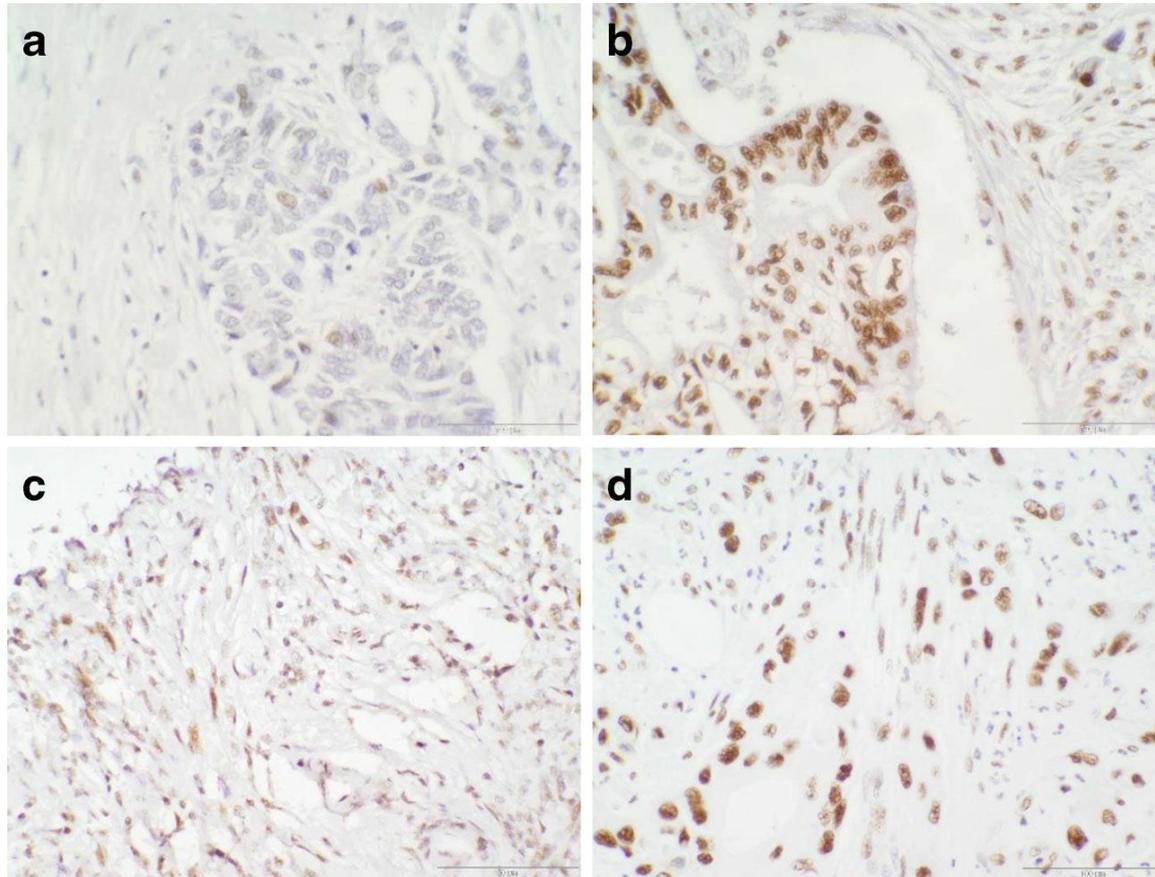
Figure 6-7. Immunohistochemical analysis of Ku70/Ku80 in (PDAC). Ku70/Ku80 expression was studied in 37 PDAC tumours and 5 normal pancreatic tissues using a commercial pancreatic cancer TMA (US Biomax). Duplicate cores from each patient were stained, and Ku70/Ku80 immunoreactivity was scored according to the staining intensity observed (weak, moderate, and strong. Representative photomicrographs are shown, (a) a low-grade PDAC tumour showing weak Ku70/Ku80 immunoreactivity, while strong Ku70/Ku80-positive staining is observed in grade 2 (b) and grade 3 tumours (d). Stromal Ku70/Ku80 immunoreactivity can be observed in (d) and in a second high grade PDAC (c). Original magnification of photomicrographs of a, b and d is×400 (scale bar=100 µm; c×200 (scale bar=50 µm).

6.4   Discussion

Despite the considerable success of MAbs in cancer therapy, there is still an unmet need for novel antibody therapeutics in oncology. Metastasis remains a huge therapeutic challenge, and a number of aggressive cancers currently lack well defined therapeutic targets. The key challenge for antibody discovery for these unmet cancers is the continued identification of novel targets that are suitable for therapeutic antibody or ADC targeting. Therapeutic antibody target discovery can broadly be classified into two main approaches: antibody targeting of proteins with established, validated roles in cancer and antibody development without prior knowledge of the target which is based on functional screening.

Here, we report the isolation and characterisation of an anti-invasive blocking antibody which was identified using such a "functional screening" or function first hybridoma screening approach. Such phenotypic-based antibody screening approaches are increasingly been described for antibody discovery (Beck et al. 2010). Our strategy involved the generation and selection of functional anti-invasive MAbs directed against the Mia PaCa-2 clone 3 cell line, with the overall objective of identifying cell surface protein targets with a direct role in cancer invasion that may be tractable to antibody/ADC therapy. This isogenic cell line which was previously isolated in this laboratory displays the characteristics of an aggressive cancer, with decreased adhesion facilitating increased invasion and motility coupled with an ability to survive in anchorage-independent conditions (Walsh et al. 2009). Furthermore, by immunising with cancer cells expressing functionally important antigens in their native cell surface configuration, it was more likely that we would isolate highly tumour-selective antibodies. Following fusion, hybridomas were screened based on their ability to inhibit invasion in their target cell line, Mia PaCa-2 clone 3. In parallel, hybridoma supernatants were also screened by immunofluorescence staining on unfixed Mia PaCa-2 clone 3 cells. Cell surface receptors are key targets for antibody-based therapeutics and account for over two thirds of current drug targets. The selected antibody 7B7 was chosen for further study, based on its ability to significantly impede invasion in Mia PaCa-2 clone 3

cells and its strong membrane immunoreactivity in Mia PaCa-2 clone 3 cells with reduced staining in poorly invasive clone 8 cells. DLKP SQ Mitox6p cells, a drug-resistant clone also exhibited strong 7B7 immunoreactivity compared with its drug-sensitive parent (H. Joyce, unpublished). So, it appeared that the 7B7 target antigen may possibly be more highly expressed in cell lines that are representative of a more aggressive phenotype, i.e. highly invasive and/or chemo-resistant.

We further demonstrated that antibody 7B7 could reduce cancer invasion in a panel of cell lines representing other aggressive tumour types (lung, glioma, colon, and breast). Its anti-invasive effects, however, were not as strong as that observed in Mia PaCa-2 clone 3. The observed variation in the extent of inhibition may reflect varying expression of the 7B7 target antigen in these cell lines, or it may suggest that our antibody is more selective for Mia PaCa-2 clone 3 PDAC cells. As the most striking anti-invasive effects were observed in pancreatic and lung cell lines, we then demonstrated that MAb 7B7 can significantly reduce invasion in a dose dependent manner in both Mia PaCa-2 clone 3 and DLKP M cells. Tumour invasion involves both the formation and the loss of adhesive contacts of tumour cells with the extracellular matrix. A small reduction in adhesion of Mia PaCa-2 clone 3 cells to ECM protein, fibronectin, and in adhesion of DLKPM cells to Matrigel in the presence of MAb 7B7 was also observed; this inhibitory effect on adhesion may be contributing to the significant decreases in the invasive potential observed in these aggressive cancer cells lines with MAb 7B7.

In addition to identifying a functional MAb with invasion-blocking activity, our study also aimed to reveal its target antigen. No signals could be detected when western blotting was performed with antibody 7B7, suggesting that 7B7 is recognising a conformational and not a linear epitope. The anti-invasive effects of antibody 7B7 and cell surface immunoreactivity observed in invasive cancer cells indicate that this antibody is recognising a native target antigen on the membrane of these cells.

Using an immunoprecipitation cross-lined method, mass spectrometry analysis using LC-MS-MS identified the immunoprecipitated 70 and 80 kDa bands as the ATP dependent DNA helicase 2 subunit 1 (Ku70) and the 80 kDa ATP-dependent DNA helicase 2 subunit 2 (Ku80) Table 6-1. The antibody immunoprecipitates were validated with commercial

antibodies directed against the Ku70 and Ku80 subunits, confirming the proteomic identifications (see supplementary data https://link.springer.com/article/10.1007%2Fs13277-014-1857-5). Ku is a heterodimeric protein composed of two subunits, Ku70 and Ku80; it was originally identified as an auto-antigen recognised by the sera of patients with autoimmune disease (Tuteja et al. 1994). It is normally found in the nucleus where it plays a key role in non-homologous end joining (NHEJ), a major pathway in DSB repairs in humans (Muller et al. 2005). Ku70/Ku80 is a key mediator in NHEJ and has been implicated in the radiosensitivity of a number of cancer types including the bladder, lung, breast, glioma, prostate, head and neck and cervix (Groselj, Kerr and Kiltie 2013, Bouchaert et al. 2012, Ma et al. 2012, Hassan et al. 2011, Petera et al. 2011, Jia et al. 2010, Söderlund Leifler et al. 2010, Cosaceanu et al. 2007). Ku has also been reported to be involved in cisplatin chemoresistance of some aggressive cancers such as lung adenocarcinoma, ovarian and leukaemic cells (Groselj et al. 2013, Bouchaert et al. 2012, Ma et al. 2012, Hassan et al. 2011, Petera et al. 2011).

Functional antibody-blocking experiments with MAb 7B7 directed against Ku70/Ku80 had demonstrated that incubation of Mia PaCa-2 clone 3 and DLKP-M cells with this anti-invasive antibody strongly suppressed invasion in a dose dependent manner in both cell lines. We then demonstrated that, for the first time, siRNA duplexes specific for both Ku70 and Ku80 could substantially reduce the invasive capacity of PDAC (Mia PaCa-2 clone 3) and lung squamous cancer cells (DLKP-M). These RNA interference-mediated knockdown results further strengthen our antibody-blocking results, indicating the functional involvement of Ku70/Ku80 in pancreatic and lung cancer invasion. The Ku heterodimer has been found to be expressed on the cell surface of various cancer cell lines (e.g. neuroblastoma, breast, cervix, and multiple myeloma) (Muller et al. 2005). Ku80 has been found to associate with latent and active MMP-9 on the cell surface of highly invasive haemapoietic cells (Monferran et al. 2004). MAb 7B7 showed strong membrane reactivity on invasive Mia PaCa-2 clone 3 cells and on drug-resistant Mitox6p cells. Together with the anti-invasive effects of MAb 7B7, these results strongly suggest that our antibody shows specificity for membrane-associated Ku70/Ku80. A number of antibodies have been developed against Ku70/Ku80; only one of these, the INCA-X

antibody, has been demonstrated to show membrane-associatedKu70/Ku80 expression in human cancer tissue, glioma (Persson et al. 2010). This group also demonstrated, using their antibody, that Ku70/Ku80 is internalised into pancreatic carcinoma cells, suggesting that it has potential therapeutic value (Fransson and Borrebaeck 2006). To our best knowledge, the involvement of Ku70/Ku80 in invasion of this aggressive cancer type or in lung cancer has not been studied to date. Studies on the role of Ku70/Ku80 in cancer invasion/ progression are limited. Inhibition of Ku80 with either TRK mediated breast cancer invasion cancer (Lagadec et al. 2010).

To investigate how 7B7 may be exerting its anti-invasive effects, we then determined if antibody 7B7 could alter MMP activity. Treatment of DLKP-M cells with antibody 7B7 resulted in a reduction in both MMP-2 and MMP-10 activities. Ku70/Ku80 has been shown to interact with MMP-9 in highly invasive cancer cells; to our knowledge, interactions with other MMPs have not been reported. So, antibody 7B7 may possibly be eliciting its anti-invasive effects by reducing MMP activity, at least in the cell line studied here.

It is likely that tumour progression is associated with enhanced NHEJ, and up regulation of Ku70/Ku80 has been reported in a number of aggressive tumours such as lung adenocarcinoma, breast and gastric cancers (Alshareeda et al. 2013, Li et al. 2013, Ma et al. 2012). Conversely, it has been suggested that deficient expression or lack of Ku70/Ku80 may result in genomic instability leading to tumourogenesis (Ferenczi et al. 2010). Earlier phases of melanoma progression are associated with loss of expression of Ku70/Ku80, and the binding activity of Ku70/Ku80 was reported to be strongly reduced at the advanced tumour stage in breast and bladder tumours (Pucci et al. 2009, Korabiowska et al. 2002). There are limited investigations focused on DNA repair proteins in pancreatic cancer. Ku70 has been implicated in the radiosensitivity of pancreatic carcinoma cells, has been shown to block Bax-mediated apoptosis and has been shown to be a secondary target for MDM2 inhibitor (Azmi et al. 2010). PDAC is highly refractory to systemic therapies and is characterised by a high propensity for local invasion and distant metastasis. There is an urgent need to develop new therapeutic strategies for this neoplasm. To ascertain if altered expression of Ku70/Ku80 is involved in tumourogenesis in PDAC, we determined, using IHC, Ku70/Ku80 expression in 37

PDAC tumours and 6 normal pancreatic tissues. Our characterisation studies indicated that the 7B7 epitope was destroyed upon formalin fixation as we did not observe any immunoreactivity on archival tissues; therefore, a commercial Ku70/Ku80 antibody was used for IHC studies. The Ku protein was expressed in all tumours studied, with a trend towards possible elevated expression of Ku70/Ku80 in higher grade, i.e. more aggressive tumours. Ku70/Ku80 immunoreactivity was also observed in stromal cells in a number of tumours. We did observe Ku70/Ku80 expression in some of the normal pancreatic tissues stained, albeit weaker immunoreactivity than that observed in the majority of the tumours. Determination of Ku70/Ku80 immunoreactivity (and tumour membrane specificity, if this can be demonstrated) in a larger PDAC patient cohort with follow-up will clarify the role of Ku70/Ku80 in PDAC progression. In terms of its application as a potential antibody target for PDAC (or other aggressive neoplasms), immunohistochemical expression levels in an extended panel of normal tissues (including haemopoietic cells) will need to be carried out.

The preliminary IHC results presented here, however, demonstrate, for the first time, that Ku70/Ku80 is widely expressed in PDAC. Taken together with invasion blocking and siRNA-mediated knockdown experiments, our study provides further evidence that targeting Ku70/Ku80 with antibodies has therapeutic potential in pancreatic cancer. Our observed anti-invasive effects in other cancer cells, while not as striking as Mia PaCa-2 clone 3 and DLKP-M, suggest that Ku70/Ku80 may be involved in the invasion of other aggressive cancer types; further, RNA-mediated interference knockdown studies would need to be carried out to confirm any potential functional role for Ku70/Ku80 in these cancer types.

In summary, we have identified a cancer cell surface protein representing a potential therapeutic target through characterisation of an anti-invasive MAb generated using a functional hybridoma screening approach. Selection of MAbs based on their ability to inhibit cancer invasion and immunoprecipitation-based proteomic identification of putative target antigens has the potential to identify novel candidate membrane proteins for therapeutic antibody targeting or ADC targeting. Furthermore, our approach may also complement pre-defined target discovery approaches by assigning cancer-invasive-associated functions to known targets.

6.5   References

Albini, A. 1998. Tumor and endothelial cell invasion of basement membranes. *Pathology & Oncology Research,* 4(3), pp.230-241.

Alshareeda, A.T., Negm, O.H., Albarakati, N., Green, A.R., Nolan, C., Sultana, R., Madhusudan, S., Tighe, P., Ellis, I.O. and Rakha, E.A. 2013. Clinicopathological significance of KU70/KU80, a key DNA damage repair protein in breast cancer. *Breast Cancer Research and Treatment,* 139(2), pp.301-310.

Azmi, A., Philip, P., Aboukameel, A., Wang, Z., Banerjee, S., Zafar, S., Goustin, A., Almhanna, K., Yang, D. and Sarkar, F. 2010. Reactivation of p53 by novel MDM2 inhibitors: Implications for pancreatic cancer therapy. *Current Cancer Drug Targets,* 10(3), pp.319-331.

Beck, A., Wurch, T., Bailly, C. and Corvaia, N. 2010. Strategies and challenges for the next generation of therapeutic antibodies. *Nature Reviews Immunology,* 10(5), pp.345.

Bouchaert, P., Guerif, S., Debiais, C., Irani, J. and Fromont, G. 2012. DNA-PKcs expression predicts response to radiotherapy in prostate cancer. *International Journal of Radiation Oncology* Biology* Physics,* 84(5), pp.1179-1185.

Chames, P., Van Regenmortel, M., Weiss, E. and Baty, D. 2009. Therapeutic antibodies: Successes, limitations and hopes for the future. *British Journal of Pharmacology,* 157(2), pp.220-233.

Cosaceanu, D., Budiu, R., Carapancea, M., Castro, J., Lewensohn, R. and Dricu, A. 2007. Ionizing radiation activates IGF-1R triggering a cytoprotective signaling by interfering with ku-DNA binding and by modulating Ku86 expression via a p38 kinase-dependent mechanism. *Oncogene,* 26(17), pp.2423.

Eckhardt, B.L., Francis, P.A., Parker, B.S. and Anderson, R.L. 2012. Strategies for the discovery and development of therapies for metastatic breast cancer. *Nature Reviews Drug Discovery,* 11(6), pp.479.

Ferenczi, K., Ohtola, J., Aubert, P., Kessler, M., Sugiyama, H., Somani, A., Gilliam, A., Chen, J., Yeh, I. and Matsuyama, S. 2010. Malignant T cells in cutaneous T-cell lymphoma lesions contain decreased levels of the antiapoptotic protein Ku70. *British Journal of Dermatology,* 163(3), pp.564-571.

Fransson, J. and Borrebaeck, C.A. 2006. The nuclear DNA repair protein Ku70/80 is a tumor-associated antigen displaying rapid receptor mediated endocytosis. *International Journal of Cancer,* 119(10), pp.2492-2496.

Frendéus, B. 2013. Function-first antibody discovery: Embracing the unpredictable biology of antibodies. *Oncoimmunology,* 2(8), pp.e25047.

Groselj, B., Kerr, M. and Kiltie, A.E. 2013. Radiosensitisation of bladder cancer cells by panobinostat is modulated by Ku80 expression. *Radiotherapy and Oncology,* 108(3), pp.429-433.

Hassan, M.K., Watari, H., Christenson, L., Bettuzzi, S. and Sakuragi, N. 2011. Intracellular clusterin negatively regulates ovarian chemoresistance: Compromised expression sensitizes ovarian cancer cells to paclitaxel. *Tumor Biology,* 32(5), pp.1031.

Jia, Q., Li, Y., Xu, D., Li, Z., Zhang, Z., Zhang, Y., Liu, D., Liu, X., Pu, P. and Kang, C. 2010. Radiosensitivity of glioma to gamma knife treatment enhanced *in vitro* and in vivo by RNA interfering Ku70 that is mediated by a recombinant adenovirus. *Journal of Neurosurgery,* 113(Special Supplement), pp.228-235.

Kinch, M.S., Kohli, M., Goldblatt, M. and Li, W. 2009. Function-first approaches to improve target identification in cancer. *Future Oncology,* 5(5), pp.617-623.

Köhler, G. and Milstein, C. 1975. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature,* 256(5517), pp.495.

Korabiowska, M., Tscherny, M., Stachura, J., Berger, H., Cordon-Cardo, C. and Brinck, U. 2002. Differential expression of DNA nonhomologous end-joining proteins Ku70 and Ku80 in melanoma progression. *Modern Pathology,* 15(4), pp.426.

Lagadec, C., Romon, R., Tastet, C., Meignan, S., Com, E., Page, A., Bidaux, G., Hondermarck, H. and Le Bourhis, X. 2010. Ku86 is important for TrkA overexpression-induced breast cancer cell invasion. *PROTEOMICS–Clinical Applications,* 4(6-7), pp.580-590.

Lambert, J.M. 2013. Drug-conjugated antibodies for the treatment of cancer. *British Journal of Clinical Pharmacology,* 76(2), pp.248-262.

Larkin, A., Moran, E., Kennedy, S.M. and Clynes, M. 2005. Monoclonal antibody 5C3 raised against formalin fixed paraffin-embedded invasive breast tumour tissue: Characterisation of its reactive antigen via immunoprecipitation and internal sequencing. *Journal of Immunological Methods,* 303(1-2), pp.53-65.

Li, W., Xie, C., Yang, Z., Chen, J. and Lu, N.H. 2013. Abnormal DNA-PKcs and ku 70/80 expression may promote malignant pathological processes in gastric carcinoma. *World Journal of Gastroenterology,* 19(40), pp.6894-6901.

Ma, Q., Li, P., Xu, M., Yin, J., Su, Z., Li, W. and Zhang, J. 2012. Ku80 is highly expressed in lung adenocarcinoma and promotes cisplatin resistance. *Journal of Experimental & Clinical Cancer Research,* 31(1), pp.99.

Martin, A. and Clynes, M. 1991. Acid phosphatase: Endpoint for *in vitro* toxicity tests. *In Vitro Cellular & Developmental Biology-Animal,* 27(3), pp.183-184.

McBride, S., Meleady, P., Baird, A., Dinsdale, D. and Clynes, M. 1998. Human lung carcinoma cell line DLKP contains 3 distinct subpopulations with different growth and attachment properties. *Tumor Biology,* 19(2), pp.88-103.

Monferran, S., Paupert, J., Dauvillier, S., Salles, B. and Muller, C. 2004. The membrane form of the DNA repair protein ku interacts at the cell surface with metalloproteinase 9. *The EMBO Journal,* 23(19), pp.3758-3768.

Mullard, A. 2013. *Maturing antibody–drug Conjugate Pipeline Hits 30,*

Muller, C., Paupert, J., Monferran, S. and Salles, B. 2005. The double life of the ku protein: Facing the DNA breaks and the extracellular environment. *Cell Cycle,* 4(3), pp.438-441.

Persson, O., Salford, L.G., Fransson, J., Widegren, B., Borrebaeck, C.A. and Holmqvist, B. 2010. Distribution, cellular localization, and therapeutic potential of the tumor-associated antigen Ku70/80 in glioblastoma multiforme. *Journal of Neuro-Oncology,* 97(2), pp.207-215.

Petera, J., Sirak, I., Beranek, M., Vosmik, M., Drastikova, M., Paulikova, S. and Soumarova, R. 2011. Molecular predictive factors of outcome of radiotherapy in cervical cancer. *Neoplasma,* 58(6), pp.469-475.

Pucci, S., Mazzarelli, P., Sesti, F., Boothman, D.A. and Spagnoli, L.G. 2009. Interleukin-6 affects cell death escaping mechanisms acting on bax-Ku70-clusterin interactions in human colon cancer progression. *Cell Cycle,* 8(3), pp.473-481.

Reichert, J.M. 2014. Antibodies to watch in 2014. *IN:* Antibodies to watch in 2014. *Mabs.* Taylor & Francis, pp.5-14.

Roti, G. and Stegmaier, K. 2012. Genetic and proteomic approaches to identify cancer drug targets. *British Journal of Cancer,* 106(2), pp.254.

Rust, S., Guillard, S., Sachsenmeier, K., Hay, C., Davidson, M., Karlsson, A., Karlsson, R., Brand, E., Lowne, D. and Elvin, J. 2013. Combining phenotypic and proteomic approaches to identify membrane targets in a 'triple negative'breast cancer cell type. *Molecular Cancer,* 12(1), pp.1.

Scott, A.M., Allison, J.P. and Wolchok, J.D. 2012. Monoclonal antibodies in cancer therapy. *Cancer Immunity Archive,* 12(1), pp.14.

Scott, A.M., Wolchok, J.D. and Old, L.J. 2012. Antibody therapy of cancer. *Nature Reviews Cancer,* 12(4), pp.278.

Shevchenko, A., Tomas, H., Havli, J., Olsen, J.V. and Mann, M. 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols,* 1(6), pp.2856.

Sievers, E.L. and Senter, P.D. 2013. Antibody-drug conjugates in cancer therapy. *Annual Review of Medicine,* 64

Sliwkowski, M.X. and Mellman, I. 2013. Antibody therapeutics in cancer. *Science (New York, N.Y.),* 341(6151), pp.1192-1198.

Söderlund Leifler, K., Queseth, S., Fornander, T. and Askmalm, M.S. 2010. Low expression of Ku70/80, but high expression of DNA-PKcs, predict good response to radiotherapy in early breast cancer. *International Journal of Oncology,* 37(6), pp.1547-1554.

Swinney, D.C. and Anthony, J. 2011. How were new medicines discovered? *Nature Reviews Drug Discovery,* 10(7), pp.507.

Tuteja, N., Tuteja, R., Ochem, A., Taneja, P., Huang, N., Simoncsits, A., Susic, S., Rahman, K., Marusic, L. and Chen, J. 1994. Human DNA helicase II: A novel DNA unwinding enzyme identified as the ku autoantigen. *The EMBO Journal,* 13(20), pp.4991-5001.

Walsh, N., Clynes, M., Crown, J. and O'Donovan, N. 2009. Alterations in integrin expression modulates invasion of pancreatic cancer cells. *Journal of Experimental & Clinical Cancer Research,* 28(1), pp.140.

Walsh, N., Dowling, P., O'Donovan, N., Henry, M., Meleady, P. and Clynes, M. 2008. Aldehyde dehydrogenase 1A1 and gelsolin identified as novel invasion-modulating factors in conditioned medium of pancreatic cancer cells. *Journal of Proteomics,* 71(5), pp.561-571.

Weber, G.F. 2013. Why does cancer therapy lack effective anti-metastasis drugs? *Cancer Letters,* 328(2), pp.207-211.

Differential Phosphoproteomic Analysis of Recombinant Chinese Hamster Ovary Cells Following Temperature Shift

Authors: <u>Michael Henry</u>, Martin Power, Prashant Kaushik, Orla Coleman, Martin Clynes, Paula Meleady

Within this work, Michael Henry specifically performed the methodology and verification of the phosphopeptide enrichment analytical methods used. Michael Henry implemented the differential expression analysis tools on the Phosphopeptide enriched samples using label free software Progenesis QI and searched the statistically significant protein features using MS search algorithms and site-specific phosphorylation algorithms. Michael Henry prepared presentation of the published work, specifically the visualisation and data presentation.

Abstract

Phosphorylation is one of the most important post-translational modifications, playing a crucial role in regulating many cellular processes, including transcription, cytoskeletal rearrangement, cell proliferation, differentiation, apoptosis, and signal transduction. However, to date, little work has been carried out on the phosphoproteome in CHO cells. In this study we have carried out a large-scale differential phosphoproteomic analysis of recombinant CHO cells following a reduction of culture temperature (temperature shift). The reduction of culture temperature during the exponential phase of growth is commonly employed by the biopharmaceutical industry to increase product yield; however, the molecular mechanisms of temperature shift in CHO cells remain poorly understood. We have identified 700 differentially expressed phosphopeptides using quantitative label-free LC-MS/MS phosphoproteomic analysis in conjunction with IMAC and TiO_2 phosphopeptide enrichment strategies, following a reduction in temperature from 37 to 31 °C. Functional assessment of the phosphoproteomic data using gene ontology analysis showed a significant enrichment of biological processes related to growth (e.g., cell cycle, cell division), ribosomal biogenesis, and cytoskeleton organization, and molecular functions related to RNA binding, transcription factor activity, and protein serine/threonine kinase activity. Differential phosphorylation of two proteins, ATF2 and NDRG1, was confirmed by Western blotting. This data suggests the importance of including the post-translational layer of regulation, such as phosphorylation, in CHO "omics" studies. This study also has the potential to identify phosphoprotein targets that could be modified using cell line engineering approaches to improve the efficiency of recombinant protein production.

Key Words: Chinese hamster ovary, phosphoproteomics, site-specific phosphorylation, biopharmaceuticals, temperature shift, recombinant protein production.

7.1   Introduction

The Chinese hamster ovary (CHO) cell line is still the animal cell expression system of choice for production of recombinant therapeutic proteins due to their ability to produce grams/L of high quality protein, their robust nature, their track record in industry, and their safety record (Walsh 2014, Wurm 2004). As a result, there is significant scientific and commercial interest in research that could improve the productivity of CHO cells as this dominance is likely to continue. Improving the efficiency of production of these biologics will be critical in controlling costs to healthcare systems as more of these drugs come to market.

There has been considerable success in developing high-producing CHO cell culture processes using approaches such as optimisation of media formulation, improvements in expression vector design, and also improvements in the design of bioreactors (Hacker, De Jesus and Wurm 2009). The next generation of improvements are being made via genetic engineering of the host (CHO) cell itself to increase or decrease the expression of endogenous genes depending on the desired outcome, in order to improve the efficiency of the production of therapeutic protein product (Fischer, Handrick and Otte 2015). In order to enhance the production capabilities and efficiency of the host cell line, an increased understanding of cellular physiology of CHO cells is of critical importance.

There have been a number of proteomic studies in CHO to try to understand phenotypes relating to growth and productivity by our group and others (Sommeregger et al. 2016, Orellana et al. 2015, Meleady et al. 2011, Doolan et al. 2010, Carlage et al. 2009, Meleady et al. 2008a). However, the vast majority of 'omic based studies in Chinese hamster ovary (CHO) cells to date has missed out on the post-translational level of regulation, which is surprising given their pivotal role in regulating cellular events such as growth and apoptosis (Farrell et al. 2014). Genomic and transcriptomic approaches are blind to post-translational modifications (PTMs) such as phosphorylation, making proteomics the only way to study these on a global scale (Angel et al. 2012).

Reversible protein phosphorylation occurring on serine, threonine, and tyrosine residues is one of the most important and most studied post-translational modifications, and plays a

crucial role in the regulation of many cellular processes including transcription, cytoskeletal rearrangement, cell proliferation, differentiation, apoptosis and signal transduction pathways (Yan and He 2008), hence are likely to be central to understanding and controlling bioprocess-relevant phenotypes. PTMs determine protein function by altering activity, cellular location, turnover, and interaction with other proteins (Choudhary and Mann 2010). Protein modifications are involved in most signaling events driving communication from the cell membrane to the nucleus in response to external stimuli. Because PTMs are a central mechanism for signal transduction and regulation, knowledge of the protein targets and modified amino acid sites sheds light on important aspects of signaling and regulation.

Despite this, to date little work has been carried out on the phosphoproteome in CHO cells. Phosphoproteomics, when mapped to metabolic and genetic networks, have the potential to improve our understanding in controlling key regulatory events in the cell (Gutierrez and Lewis 2015). One of the earliest profiling studies in CHO cells found alterations in the levels of tyrosine phosphorylation on an 80 kDa and 180 kDa protein following temperature shift from 37°C to 31°C; this provided the first evidence of CHO cells actively responding at the post-translational level to changes in culture temperature (Kaufmann et al. 1999). The reduction of culture temperature during the exponential phase of growth is commonly employed by biopharma to increase product yield (Becerra et al. 2012). There has been some work looking at protein translation and phosphorylation events following reduction in temperature. A decrease in global protein synthesis has been attributed to the phosphorylation of the initiation factor eIF2α (Underhill et al. 2005). More recently exposure of CHO-K1 cells to mild hypothermia was found to activate the ATR (ataxia telangiectasia mutated- and Rad3-related kinase) –p53 – p21 signaling pathway, where it was found that p53 is phosphorylated at Ser15 through ATR activation; at the same time ATR activates p38 MAPK, resulting in the phosphorylation of p53 at Ser33 and Ser46, that subsequently enhances Ser15 phosphorylation. P53 subsequently activates transcription of the downstream target p21, which leads to cell cycle arrest (Roobol et al. 2011).

In recent years high-throughput and quantitative phosphoproteomics based on mass spectrometry has emerged as a powerful tool to study signal transduction pathways and cellular responses (Macek, Mann and Olsen 2009, Jensen et al. 2006, Cohen 2001). The

complexity and importance of the cellular processes that rely upon phosphorylation for such signaling events becomes apparent when it is realised that approximately 30% of all eukaryotic proteins can become phosphorylated (Olsen et al. 2010, Cohen 2001), though some of these phosphorylation events may be physiologically irrelevant and may not be required for cell signaling (Dephoure et al. 2013).

In this 'proof-of-concept' study we have carried out a phosphoproteomic analysis of CHO-SEAP cells subjected to a culture temperature shift from 37°C to 31°C. Phosphopeptide enrichment of protein digests in combination with quantitative label-free LC-MS/MS proteomic analysis was used to compare differential expression of phosphorylated proteins following temperature shift. Phosphopeptide enrichment using immobilized metal ion (Fe3+ or Ga3+) affinity chromatography (IMAC) and metal oxide affinity chromatography (MOAC) using TiO_2 or ZrO_2 are widely used in phosphoproteomic studies (Angel et al. 2012) and are often used in parallel to increase phosphoproteome coverage (von Stechow, Francavilla and Olsen 2015). Their use in the phosphoproteomic analysis of CHO cells has not been reported to date making this study an important step towards a better understanding of CHO cell physiology at a systems biology level.

7.2 Materials and Methods

7.2.1 Cell Culture

CHO K1 cells stably transfected with an expression vector encoding human secreted alkaline (SEAP) were previously established in our laboratory.[26] Eight shake flasks of CHO SEAP cells seeded at $2x10^5$ cells/mL were set up in a 50 mL volume containing CHO-S-SFMII (Thermo Fisher Scientific) serum-free media in a batch culture at 37°C and 5% CO_2 in a Kuhner Climo-Shaker ISF1-X orbital shaker (Kuhner) at 170 rpm. After 48 hours, four flasks were selected for temperature-shift and were placed in a 31°C orbital shaker for the remainder of the culture period while the remaining 4 flasks were maintained at 37°C. After a further 36 hrs incubation CHO SEAP cells were harvested from both temperature-shifted and non-temperature-shifted cultures. 36 hours was selected because this is more than sufficient time for consistent phosphorylation events in direct response to culture temperature reduction to take place and

227

allow any changes in phosphorylation that may be as a result of shock to subside. Based on other growth profiling experiments, 36 hrs post temperature shift is also still within the exponential phase of growth for this cell line. Cellular growth and viability was assayed using a Guava Benchtop Cytometer after staining with Viacount^{TM} (Millipore) while SEAP activity was measured using the phosphatase assay as described previously (Barron et al. 2011, Lipscomb et al. 2005). SEAP was reported in terms of activity (unit/min/mL) rather than protein quantity. Normalised activity referred to the measured activity divided by the cell number per unit volume. All cell lines were free of *Mycoplasma* contamination, as tested monthly with the indirect Hoechst staining method.

7.2.2   Cell Lysate Sample Preparation

Cells were pelleted by centrifugation, washed with 50 mM ice cold HEPES and then snap frozen and stored at -80˚C until required. Cells were lysed in a buffer containing 7 M Urea, 2 M Thiourea, 4% CHAPS, 30 mM Tris, pH 8.5, 1x HALT protease inhibitors (Thermo Fisher Scientific), 1X HALT phosphatase inhibitors (Thermo Fisher Scientific) and incubated on ice for 20 minutes with occasional vortexing. Sample lysates were centrifuged at 14000 xg for 15 minutes at 4˚C and the supernatant was transferred to a new microcentrifuge tube and assayed for protein concentration using the Quick Start^{TM} Bradford Protein Assay (BioRad). The protein sample was stored in aliquots at -80˚C.

7.2.3   In-Solution Protein Digestion and Phosphopeptide Enrichment

Protein from the cell lysates was cleaned up using the ReadyPrep 2D Cleanup Kit (BioRad) according to manufacturer's instructions. The cleaned protein pellets were resuspended in a buffer containing 6 M Urea, 2 M Thiourea, 30 mM Tris, pH 8.5 and assayed for protein concentration using the QuickStart^{TM} Bradford Protein Assay. Five hundred microgrammes of cleaned up protein was diluted with 50 mM Ammonium Bicarbonate so that the final urea concentration was less than 1 M. The protein was reduced with freshly prepared dithiothreitol to a final concentration of 5 mM for 20 minutes at 56˚C followed by alkylation with freshly prepared iodoacetamide to a final concentration of 15 mM and incubated in the dark for 20

minutes at room temperature. Digestion of protein samples was carried out using Endoproteinase Lys-C (Promega) at a ratio of 100:1, protein/enzyme w/w at 37°C for 3 hours followed by trypsin digestion (sequence grade, Promega) at a ratio of 50:1 protein/enzyme w/v at 37°C overnight. Trifluoroacetic acid (TFA) was added to a final concentration of 0.5% to each sample to inactivate the trypsin. Peptide concentration was determined using the Pierce™ Quantitative Fluorometric Peptide Assay (Thermo Fisher Scientific). 20 µg of digested peptide were retained from each sample as non-enriched whole cell lysate samples and were purified with C18 spin columns (Pierce, Thermo Fisher Scientific) as per manufacturer's instructions. Eluted peptides were gently dried with a vacuum evaporator (SpeedVac). The remaining peptides for phosphopeptide enrichment were then purified with graphite spin columns (Pierce, Thermo Fisher Scientific) as per manufacturer's instructions and eluted peptides were gently dried with a vacuum evaporator (SpeedVac). To ensure reproducibility of the graphite spin columns we used the same Lot no. for the complete experiment. The average peptide loss was 31% ± 2% for each sample that was subjected to graphite spin column clean-up). Dried peptides were resuspended in the relevant resuspension buffer depending on the phosphopeptide enrichment used as per manufacturers' instructions. TiO$_2$ phosphopeptide enrichment and Fe-NTA (IMAC) spin column kits were used in the analysis (both from Pierce, Thermo Fisher Scientific). Eluted phosphopeptides from both techniques were once again individually cleaned up with graphite spin columns and gently dried with a SpeedVac.

7.2.4   LC-MS/MS analysis

20 µg of dried phosphopeptide and non-phosphopeptide enriched samples were re-solubilised in 25 µL of LC-MS grade water with 0.1% TFA and 2% Acetonitrile (ACN). Nano LC–MS/MS analysis was carried out using a Dionex Ultimate 3000 RSLCnano system (Thermo Fisher Scientific) coupled to a hybrid linear ion trap/Orbitrap mass spectrometer (LTQ Orbitrap XL; Thermo Fisher Scientific). SilicaTip™ Standard Coating Tubing OD/ID 360/20µm Tip, ID 10 µm, Length 5cm (New Objective) were used as emitter tips for nano electrospray.

A 5 µL injection of sample was picked up using the autosampler and loaded onto a C18 trap column (C18 PepMap, 300 µm ID × 5 mm, 5 µm particle size, 100 Å pore size; Thermo Fisher Scientific). The sample was desalted for 3 minutes using a flow rate of 25 µL/min in 0.1% TFA

containing 2% ACN. The trap column was then switched online with the analytical column (PepMap C18, 75 µm ID × 250 mm, 3 µm particle and 100 Å pore size; (Thermo Fisher Scientific)) using a column oven at 35°C and peptides were eluted with the following binary gradients of: Mobile Phase Buffer A and Mobile Phase Buffer B: 0–25% solvent B in 240 minutes and 25–50% solvent B in a further 60 minutes, where solvent A consisted of 2% acetonitrile (ACN) and 0.1% formic acid in water and solvent B consisted of 80% ACN and 0.08% formic acid in water. The column flow rate was set to 300 nL/min. Data was acquired with Xcalibur software, version 2.0.7 (Thermo Fisher Scientific).

The LTQ Orbitrap XL was operated in data-dependent and externally calibrated. Survey MS scans were acquired in the Orbitrap in the 400–1800 m/z range with the resolution set to a value of 30,000 at m/z 400. Up to three of the most intense ions (1+, 2+ and 3+) per scan were CID fragmented in the linear ion trap. Dynamic exclusion was enabled with a repeat count of 1, repeat duration of 30 seconds, exclusion list size of 500 and exclusion duration of 40 seconds. The minimum signal was set to 2000. All tandem mass spectra were collected using a normalised collision energy of 35%, an isolation window of 2 m/z with an activation time of 30 milliseconds.

7.2.5 Quantitative Label-free LC-MS/MS Data Analysis

The raw MS data files obtained were processed using Progenesis QI for Proteomics software (version 2.0; Non-Linear Dynamics, a Waters company, Newcastle upon Tyne, UK). Peptide LC retention times from all MS data files were aligned to an assigned reference run (a raw file chosen by the software as the most suitable reference for all runs). The samples from the two experimental groups were then set up within the Progenesis software for differential analysis. Peptide features were filtered using the following parameters; (1) peptide features with ANOVA ≤ 0.05 between experimental groups, (2) mass peaks with charge states from +1 to +3 and greater than one isotope per peptide. For the phosphoproteomic analysis a mascot generic file (mgf) was then generated from all exported MS/MS spectra which was used for peptide and protein identification via Proteome Discoverer 2.1 using Sequest HT (Thermo Fisher Scientific) and Percolator (Kall et al. 2007) followed by phospho*RS* 3.1 (Taus et al. 2011) against the NCBI *Cricetulus griseus* database CRIGR containing 44,065 sequences (fasta file

downloaded November 2015). The following search parameters were used for protein identification: (1) peptide mass tolerance set to 20 ppm, (2) MS/MS mass tolerance set to 0.6 Da, (3) up to two missed cleavages were allowed, (4) carbamidomethylation of cysteine set as a fixed modification and (5) methionine oxidation, and serine, threonine and tyrosine phosphorylation were set as variable modifications. Only high confident peptide identifications with an FDR ≤ 0.01 (identified using a SEQUEST HT workflow coupled with Percolator validation in Proteome Discoverer 2.1 (Thermo Fisher Scientific)) were imported into Progenesis QI software for further analysis if the phosphorylation site in the peptide was localised through phospho*RS* in Proteome Discoverer 2.1 after applying a site probability cut off score of 75% or greater for S, T or Y amino acids (Alpert, Hudecz and Mechtler 2015, Roitinger et al. 2015). Once successful phosphopeptide identifications were imported into Progenesis, statistical criteria of ANOVA p-value less than 0.05 and a minimum fold-change of relative peptide abundance between the two experimental groups of ≥ 1.5-fold was set. Phosphopeptides which passed these criteria were then deemed to be differentially expressed phosphopeptides between the two experimental groups. Data visualisation was achieved using volcano plots and heatmaps which were generated using ggplot package in R to show (1) the distribution of all identified phosphopeptides based on statistics and fold-change and (2) the relative expression levels of the differentially expressed phosphopeptides across all of the samples, respectively. For the non-enriched whole cell lysate samples, a statistical criteria of ANOVA p-value less than 0.05, a minimum of two peptides matched to a protein and a ≥ 1.5-fold change in relative abundance between the two experimental conditions was used as the criteria for identification as a differentially expressed protein.

7.2.6 Gene Ontology (GO) Analysis

Identified differentially expressed proteins were assigned to mouse or human official Gene Symbol identifiers. These gene symbols were then imported into DAVID (https://david.ncifcrf.gov) for functional pathway analysis using GO. In order to refine the number of pathways identified, an adjusted p-value (Benjamini Hochberg) of ≤0.05 was used.

7.2.7 Western Blot Analysis

Samples for western blot analysis were prepared in 2X Laemmli sample buffer (Sigma-Aldrich), heated at 95 °C for 5 min and then cooled on ice. Protein samples were then loaded onto 4-12% NuPAGE Bis-Tris gels (Invitrogen™, Life Technologies). Electrophoretic transfer, blocking and development of Western blots were carried out as described previously (Meleady et al. 2008b). The following primary antibodies were used; polyclonal rabbit anti-ATF2 (dilution 1:1000, v/v; Abcam, Cambridge, UK); polyclonal rabbit anti-phospho ATF2 (Thr69/Thr51) (dilution 1:1000, v/v; Abcam); polyclonal rabbit anti-NDRG1 (dilution 1:1000, v/v; Cell Signalling Technology, Danvers, MA, USA); polyclonal rabbit anti-phospho NDRG1 (Ser330) (dilution 1:1000, v/v; Cell Signalling Technology); monoclonal mouse anti-GAPDH (clone 6C5, dilution 1:10.000, v/v; Abcam); polyclonal goat anti-β-actin (dilution 1:1000, v/v, Abcam). Bound antibodies were detected using horseradish peroxidase conjugated secondary antibodies (anti-mouse, dilution 1:2000, v/v; anti-rabbit, dilution 1:2000 v/v; anti-goat, 1:2000, v/v (all from Dako, Glostrup, Denmark)). Bands were quantified by densitometry using Image Quant software version 5.0 (Amersham Biosciences). Statistical analysis was performed using Students t-test.

7.3 Results

7.3.1 Differential Phosphoproteomic Analysis of CHO–SEAP Cells Following Temperature Shift

Eight biological replicates of CHO-SEAP cells were cultured for 48 hours, the culture temperature was then reduced to 31°C in one set of the four cultures and the second set of four cultures were maintained at 37°C (i.e. four biological replicate cultures for each of the two experimental conditions). Cells were harvested from both sets of cultures 36 hours post temperature shift. 36 hours was selected because this is more than sufficient time for consistent phosphorylation events in direct response to culture temperature reduction to take

place and allow any changes in phosphorylation that may be as a result of reduced temperature shock to subside. Based on previous growth profiling experiments (results not shown), 36 hours post temperature shift was also still within the exponential phase of growth for this cell line. This time-point was chosen for cell harvesting as cell signalling events as a result of temperature shift may become difficult to decipher from other signalling events later in the culture, e.g., reduced growth rate due to cells entering stationary phase. Figure 7-1 displays the growth and productivity data for CHO-SEAP cells showing the reduced growth of the cells after 36 hours post-temperature shift compared to the 37°C cultures with an associated increase in productivity at 31°C.

Five hundred micrograms of whole cell lysate sample was digested and enriched for phosphopeptides using TiO_2 and Fe-NTA spin columns and then prepared for LC-MS/MS analysis; see Figure 7-2 for an outline of the experimental workflow. Five hour reverse phase gradients were used to separate the peptides from each enrichment method used. Raw MS data was interrogated using Progenesis QI for proteomics label-free LC-MS software to identify differentially expressed phosphopeptides between the 37°C and the 31°C cultures. Significant increased or decreased expression of phosphopeptides between samples was determined by using a fold-change cut-off of +/- 1.5-fold, and a p-value > 0.05 between the two culture sample sets. The phosphorylation site in the peptide was localised through phosphoRS 3.1 in Proteome Discoverer 2.1 after applying a site probability cut off score of ≥ 75%.[29]

Unsupervised Pearson clustering of the differentially expressed phosphopeptides shows clear separation of the sample sets into distinct sample groups relating to the two temperature cultures based on phosphopeptide abundance difference as visualised by heatmaps for both enrichment methods (see Figure 7-3). From this analysis we identified 384 differentially expressed phosphopeptides (182 up and 202 down at 31°C) from the TiO_2 enriched samples, and 504 differentially expressed phosphopeptides from the Fe-NTA enriched samples (184 up and 320 down at 31°C). These phosphopeptides corresponded to 319 differentially expressed proteins from the TiO_2 samples (150 up and 169 down at 31°C) and 402 differentially expressed proteins from the Fe-NTA enriched samples (143 up and 259 down at 31°C), due to in some cases more than one differentially expressed phosphopeptides were identified from the same protein. See supplementary tables S-1 and S-2

(https://pubs.acs.org/doi/10.1021/acs.jproteome.6b00868#notes-2) for the full lists of

differentially expressed phosphopeptides from the TiO_2 and Fe-NTA enriched samples). The total number of phosphopeptides identified was 2,466 from the TiO2 enriched samples and 3,351 from the Fe-NTA enriched samples. From these total numbers of phosphopeptides identified, 15.6% from the TiO_2 enrichment and 15% from the Fe-NTA enrichment were found to change in relative abundance following a reduction in culture temperature. Volcano plots for the two enrichment methods show the distribution of differentially expressed phosphopeptides based on p-value on fold-change between the two experimental conditions (see Figure 7-4). Only those phosphopeptides which pass the required criteria of a minimum fold-change of 1.5 and a p-value of ≤0.05 (i.e. the peptides both above and to the right and left of the threshold lines) were deemed differentially expressed between the two culture conditions.

From the results a number of differentially expressed phosphopeptides were found to be in common between the two enrichment methods; 44 phosphopeptides were in common with decreased expression at 31°C and 50 in common showing increased expression at 31°C. Tables 7-1 and 7-2 show the overlap of common phosphopeptides between the two enrichment methods and differentially expressed in the same direction. A closer look at this analysis revealed that 106 phosphosites were identified in common between the two enrichment methods – 46 from phosphopeptides with decreased expression and 60 from phosphopeptides with increased expression at 31°C compared to 37°C (some of these phosphopeptides had more than one phosphosite identified by mass spectrometry). Examples of common phosphopeptides from proteins identified between the two enrichment methods include Eukaryotic translation initiation factor 4 gamma 1 (EIF4G), Cyclic AMP-dependent transcription factor ATF-2 (ATF2) and Microtubule-associated protein 1B (MAP1B) all showing decreased expression at 31°C compared to 37°C (see Table 7-1). Protein NDRG1 (NDRG1), Ras and Rab interactor 2 (RIN2) and Serine/threonine-protein kinase 10 (STK10) all show increased expression at 31°C compared to 37°C (see Table 7-2). Phosphopeptides relating to ATF2 and NDRG1 are highlighted in the Volcano plots as shown in Figure 7-4 to emphasise that both surpassed the criteria for being differentially expressed. It has also been found that >85% of the phosphorylation sites identified on the CHO proteins are homologous to sites identified in human, rat or mouse and shown in Tables 7-1 and 7-2 (as determined using PhosphoSitePlus ® and Uniprot). The remaining unmatched sites of phosphorylation are potentially novel CHO-specific phosphosites.

7.3.2 Overlap of phosphoproteomic analysis with non-enriched whole cell lysate proteomic analysis of CHO-SEAP cells following temperature shift.

Quantitative label-free LC-MS/MS proteomic analysis of the non-phosphopeptide enriched whole cell lysate samples resulted in the differential expression of 149 proteins with 96 showing increased expression and 53 with decreased expression following a temperature reduction to 31°C (see supplementary Tables S-5 and S-6 https://pubs.acs.org/doi/10.1021/acs.jproteome.6b00868#notes-2). Overlapping the differential protein expression list from the non-enriched analysis with the phosphopeptide enrichment lists (from both TiO_2 and Fe-NTA enrichment) yielded 6 proteins with peptides in common (See supplementary Table S-8 https://pubs.acs.org/doi/10.1021/acs.jproteome.6b00868#notes-2). It should be noted that site-specific phosphorylation was not detected on these common peptides in the non-enriched whole cell lysate proteomic analysis. A further 16 proteins overlapped the phosphopeptide enrichment protein list that did not share a common peptide (see supplementary Table S-9 https://pubs.acs.org/doi/10.1021/acs.jproteome.6b00868#notes-2).

One of the differentially expressed proteins found to share common peptides between the non-enriched and phosphopeptide-enriched samples was vimentin. From the non-enriched whole cell lysate analysis 26 peptides were matched to vimentin that contributed to its differential expression and identification (see supplementary Table S-7 (https://pubs.acs.org/doi/10.1021/acs.jproteome.6b00868#notes-2). Four of the differentially expressed phosphopeptides were in common with these 26 peptides (see supplementary Table S-8 https://pubs.acs.org/doi/10.1021/acs.jproteome.6b00868#notes-2). Interestingly one of the differentially expressed phosphopeptides from vimentin (TYS*LGSALRPSTSR) showed a 14 fold increase in expression following a reduction in temperature compared to the 1.7 fold increase in expression of the non-phosphorylated version of the peptide identified from the non-enriched whole cell lysate analysis. The overall increase in vimentin expression following temperature shift from the whole cell lysate proteomic study was 17 fold (see supplementary Table S-5 https://pubs.acs.org/doi/10.1021/acs.jproteome.6b00868#notes-2). Vimentin is a major intermediate filament protein and phosphorylation is known to regulate its activity (Snider and Omary 2014a). Expression of vimentin has previously been shown to be increased

following temperature shift by our group and others (Snider and Omary 2014b, N. Kumar et al. 2008).Further work would need to be carried out to assess the extent of altered phosphorylation of these phosphopeptides compared to their non-phosphorylated form to determine a potential role of phosphorylation in regulating vimentin activity following temperature shift.

Phosphorylation is a low stoichiometry modification; as a result phosphopeptides (and phosphoproteins) need to be enriched from complex mixtures that have high backgrounds of non-phosphorylated proteins (Riley and Coon 2016a), which may explain the relatively low overlap between the non-enriched whole cell lysate proteomic and the phosphoproteomic data from our study. It could also be due to the low abundant nature of many of the phosphoproteins identified that were undetectable in the non-enriched whole cell lysate proteomic analysis.

7.3.3 Western Blot Analysis Confirms Differential Phosphorylation of NDRG1 and ATF2 Following Temperature Shift

Western blotting was carried out to validate the results found by quantitative label-free phosphoproteomic analysis. Similar approaches have been previously used to determine differential phosphopeptide expression using quantitative label-free phosphoproteomic approaches followed by altered phosphorylation confirmation by methods such as western blotting (Zhang et al. 2016). This would suggest if there were altered phosphorylation levels as a result of temperature shift rather than just an increase or decrease in levels of the phosphoprotein itself (Chen et al. 2012).

Primary phospho-specific antibody selection against a target protein was based on the ability to detect peptides with homologous amino acid sequences to those identified in this study in species such as human, mouse and rat, as few CHO-specific antibodies are available commercially. The second requirement was that the antibody was capable of detecting the correct site of phosphorylation on the peptide of interest, i.e. the specific phosphorylation site was conserved between species. Based on these stipulations, two antibodies were sourced for the purpose of this validation, against phosphorylated and total NDRG1 and ATF2 proteins. See supplementary Figures S-1 and S-2

(https://pubs.acs.org/doi/10.1021/acs.jproteome.6b00868#notes-2) for sequence alignment for human NDRG1 and ATF2 and the CHO-specific forms of NDRG1 and ATF2 derived from the CHO protein sequence database (Meleady et al. 2012). The phosphorylation sites highlighted are specifically detected by the phospho-specific antibodies chosen to validate the proteomic study.

Following quantitative label-free LC-MS/MS analysis of TiO_2 phosphopeptide enriched temperature-shifted CHO SEAP whole cell lysates, the monophosphorylated phosphopeptide TASGSSVTSLEGPR from NDRG1 was found to be have increased expression at 31˚C compared to 37˚C (Table 7-1) and detected using both enrichment strategies (see Figure 7-5A) for Progenesis label-free LC-MS/MS software analysis output showing increased expression of this phosphopeptide at 31°C). Western blot validation of the phosphorylated form of NDRG1 clearly shows an increase in the phosphorylated form of the protein at 31˚C compared to 37˚C (Figure 7-5 B and C). Although total NDRG1 was detected in the control cell line (human breast cancer cells, SKBR3) it was not detected in the CHO samples (Figure 7-5B); this is possibly as a result of the antibody for total NDRG1 not being able to detect the CHO form of the protein or that the levels of NDRG1 are undetectable in these CHO-SEAP cells. It should be noted that a search of the RAW MS files from LC-MS/MS analysis of CHO whole cell lysate failed to identify the NDRG1 protein in the non-enriched samples (not shown); this indicates that NDRG1 is likely to be a low abundance protein in CHO-SEAP cells and lends further weight to the possibility that it could not be detected by the antibody due to its low abundant nature.

Decreased expression of the phosphopeptide NDSVIVADQTPTPTR from ATF2 at 31˚C was shown using both enrichment methods; the Progenesis output from the label-free LC-MS proteomic data is shown in Figure 7-6A. Figure 7-6B shows a western blot confirming the down-regulation of the phosphopeptide NDSVIVADQTPTPTR from the protein ATF2 at 31˚C. There was no change to the expression of total ATF2 detected in the cell lysates by western blot analysis which suggests decreased phosphorylation of ATF2 following a reduction in culture temperature, as shown by densitometry analysis of the western blot data (Figure 7-6C).

Because of lack of commercially available phospho-specific antibodies derived from CHO cells against the CHO-specific sites of phosphorylation identified in this study, future work would use targeted phosphoproteomic approaches such as Selection Reaction Monitoring (SRM) (Narumi et al. 2012) or Parallel Reaction Monitoring (PRM) (Abelin et al. 2016) to quantify

changes in phosphorylation of other phosphopeptides identified as differentially expressed following a reduction in culture temperature.

7.3.4 Gene Ontology (GO) Analysis

Two individual differentially expressed lists were submitted to DAVID for gene ontology analysis. One list was from proteins with phosphopeptides which showed increased expression at 31˚C from both enrichment strategies and a second was from proteins that showed increased expression of phosphopeptides at 37˚C from both enrichment strategies. GO biological processes which were found to be over-represented in the up-regulated at 31˚C differentially expressed list (Table 7-3) included ribosomal biogenesis and RNA processing while molecular functions which were over-represented included RNA binding and transcription factor regulation. GO biological processes over-represented in the down-regulated at 31˚C list (Table 7-4) included many linked to growth including chromosome organization, mitosis, cell cycle, kinase activity and signal transduction while molecular functions associated with transcription factor activity, cytoskeleton binding and serine/threonine kinase activity were over-represented.

(A)

(B)
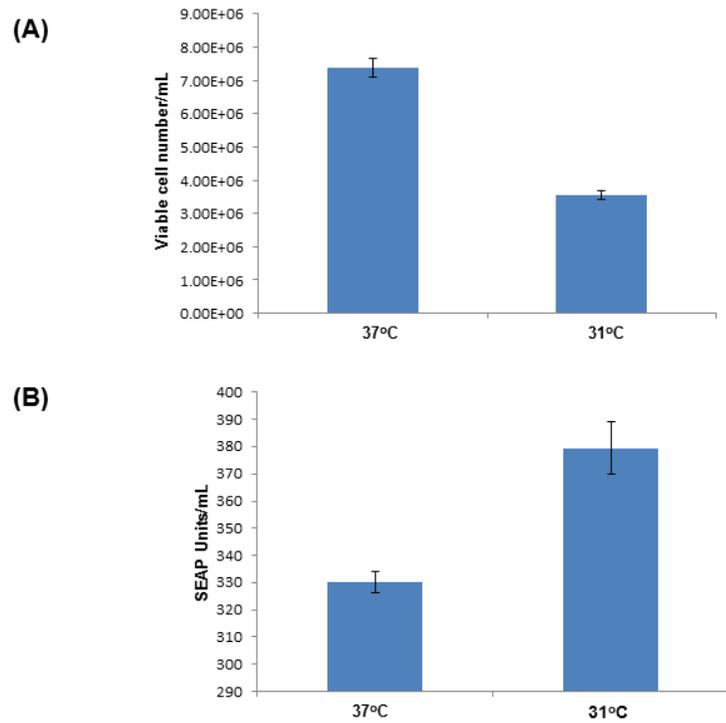
Figure 1

Figure 7-1. Growth and productivity of CHO-SEAP cells at 37°C and 31°C 36 hours post temperature shift. (A) **Viable cell numbers/mL measured using a Guava Viacount assay. (B) Secreted recombinant SEAP protein was determined by measuring its activity using a kinetic absorbance assay. Error bars represent standard deviation from data obtained from four biological replicate cultures at each temperature (n=4).**

239

CHO-SEAP Whole Cell Lysate
37°C v 31°C cultures

Protein Digest

Phosphopeptide
Enrichment
(TiO$_2$)

Phosphopeptide
Enrichment
(IMAC - Iron)

Quantitative label
free LC-MS/MS

Quantitative label
free LC-MS/MS
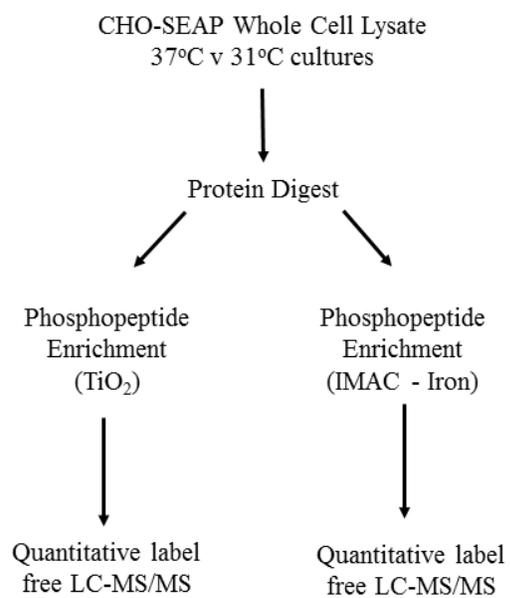
Figure 2

Figure 7-2. Outline of experimental workflow for differential phosphoproteomic analysis using the two phosphopeptide enrichment strategies.

(A)

Differential Analysis Fe Enrichment- Pearson clustering

(B)

Differential Analysis TiO2 Enrichment- Pearson clustering
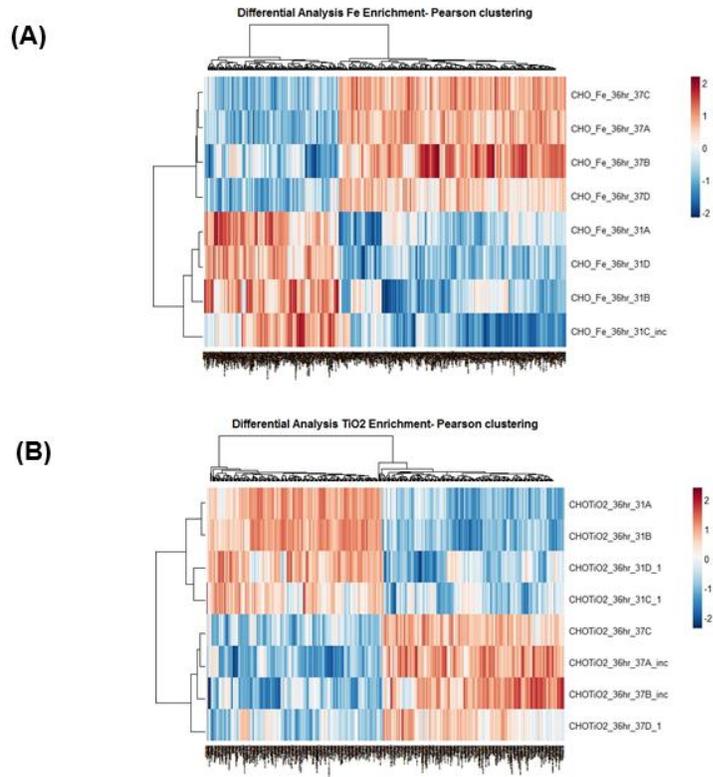
Figure 3

Figure 7-3. Unsupervised Pearson clustering shows that the expression of the differentially expressed phosphopeptides identified from the experimental samples separate into two distinct sample groups from the 37^oC (37A-37D) and 31^oC (31A-31D) cultures grown in biological quadruplicate. (A) Fe-NTA (IMAC) enriched phosphopeptides and (B) TiO$_2$ enriched phosphopeptides.

(A)   (B)
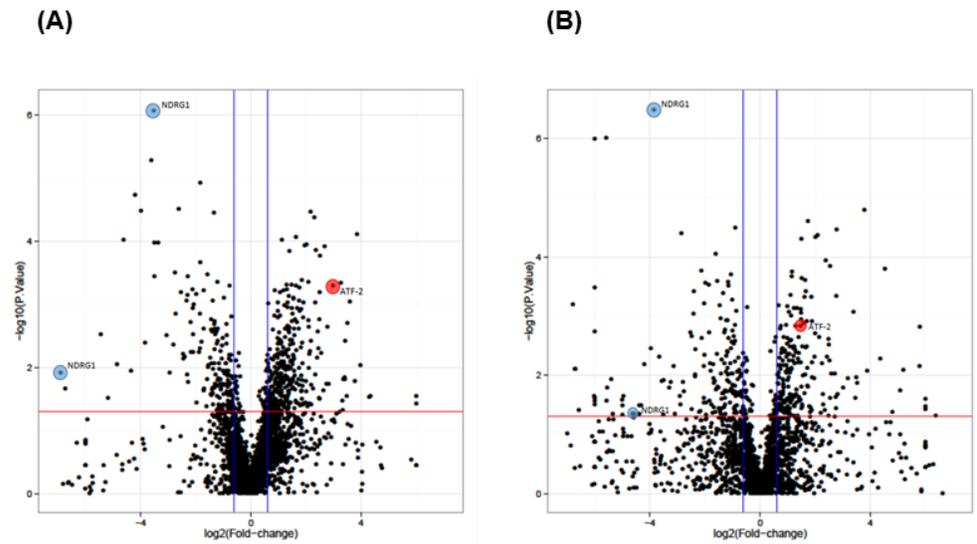
Figure 4

Figure 7-4. Volcano plot representations of phosphopeptide abundance differences caused by temperature shift in CHO-SEAP cells. Volcano plots of all identified phosphopeptides showing their distribution according to Anova p-value (-log10 P-value) and fold change (log2 fold-change) between the 37°C and 31°C samples from (A) Fe-NTA (IMAC) and (B) TiO$_2$ enrichment. Phosphopeptides above the horizontal line are deemed statistically significant (p-value <0.05) and those to the right and left of the vertical lines indicate relative fold changes ≥ 1.5. Phosphopeptides relating to NDRG1 (blue) and ATF2 (red) are highlighted.
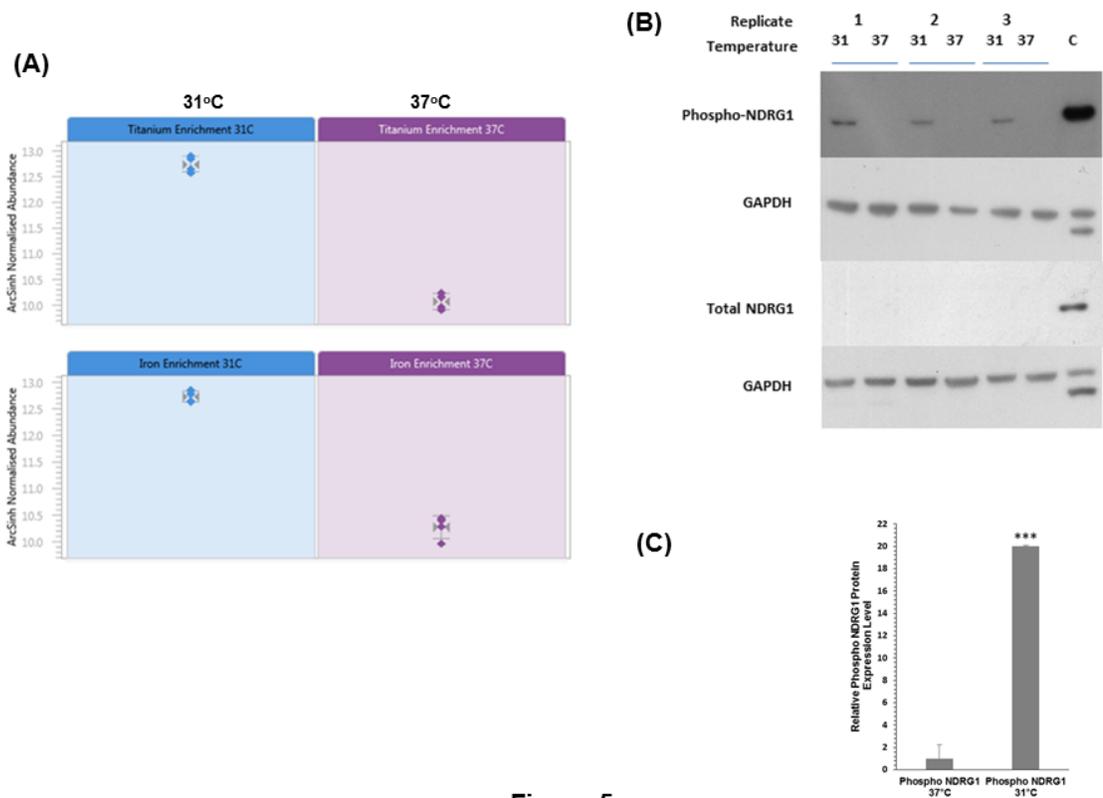
Figure 5

Figure 7-5. (A) **Progenesis QI for Proteomics software output showing increased expression of the phosphopeptide TASGSSVTSLEGPR at 31°C from NDRG1. The vertical axis represents normalised abundance volumes (log).** (B) **Western blot confirming increased expression of phospho-NDRG1 in 3 replicate CHO-SEAP samples at 31°C. Total NDRG1 was not detected in all CHO-SEAP samples. GAPDH was used as a loading control. C – human breast cancer cell lysate from SKBR3 cell line.** (C) **Phospho-NDRG1 protein expression levels at 31°C was densitometrically quantified relative to its expression at 37°C and confirms increased phosphorylation of NDRG1 at 31°C. The data represents the three replicates for each sample at 37°C and 31°C and are expressed as mean values ± standard deviation. (*** = p≤0.001). (No densitometry analysis is shown for total NDRG1 due to the lack of signal on the immunoblot).**

243
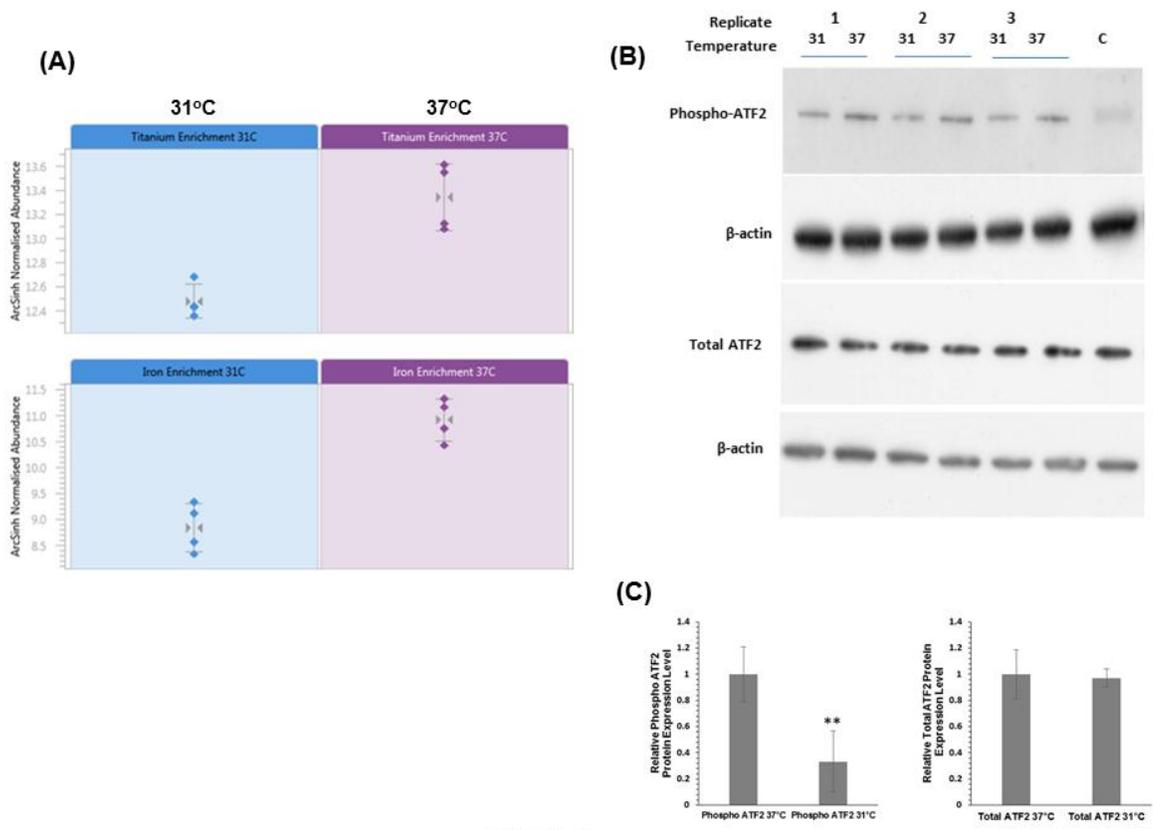
Figure 6

Figure 7-6. (A) **Progenesis QI for Proteomics software output showing a reduction in expression of the phosphopeptide NDSVIVADQTPTPTR at 31°C from ATF2. The vertical axis represents normalised abundance volumes (log).** (B) **Western blot confirming reduced expression of phospho-ATF2 in 3 replicate CHO-SEAP samples at 31°C. Total ATF2 showed no change in expression at both temperatures. β-actin was used as a loading control. C – human breast cancer cell lysate from SKBR3 cell line.** (C) **Phospho-ATF (left panel) and total ATF2 (right panel) protein expression levels at 31°C were densitometrically quantified relative to their expression at 37°C, and confirms decreased phosphorylation of ATF2 at 31°C. The data represents the three replicates for each sample at 37°C and 31°C and are expressed as mean values ± standard deviation. (** = p≤0.01).**

Accession	Description	Sequence	Titanium Enrichment		Fe-NTA (IMAC) Enrichment	
			PhosphoRS: Best Site Probabilities	Fold change	PhosphoRS: Best Site Probabilities	Fold change
625200413	60S ribosomal export protein NMD3 isoform X1	DSTVPVESDTDDEGAPR	S8(Phospho): 100; T10(Phospho): 100	2.01	T10(Phospho): 99.39	2.41
625265154	alpha-2-macroglobulin receptor-associated protein isoform X3	DTQMVHSNSLNEDTQDELGDPR	S9(Phospho): 100	1.65	S9(Phospho): 98.18	2.61
354503849	AN1-type zinc finger protein 5	MSPMGTASGSNSPTSDSASVQR	S2(Phospho): 99.14	2.45	S2(Phospho): 100	2.49
625191482	ashwin isoform X1	KSPSGPVKSPPLSPVGTTPVK	S9(Phospho): 99.7; T17(Phospho): 78.36	3.07	S9(Phospho): 100	2.02
625212872	AT-rich interactive domain-containing protein 1A isoform X1	TLLDPGRFSK	S9(Phospho): 100	2.60	S9(Phospho): 100	2.09
625240398	autophagy-related protein 16-1 isoform X1	SVSSIPVPQDNVDTHPASGK	S1(Phospho): 97	2.29	S1(Phospho): 97.95	2.73
625187501	bcl-2-associated transcription factor 1-like	AEGEPQEESPLKSK	S9(Phospho): 100	5.09	S9(Phospho): 99.99	2.29
625273094	brain-specific angiogenesis inhibitor 1-associated protein 2 isoform X2	NSYATTENKTLPR	T10(Phospho): 100	4.05	T10(Phospho): 100	3.48
625240603	CCAAT/enhancer-binding protein zeta	NEDEDSIEDVDDKEFEK	S6(Phospho): 100	2.11	S6(Phospho): 100	1.77

625192987	DNA ligase 1 isoform X1	KAAQVLSSEGEEEDEAPSTPK	S7(Phospho): 100; S8(Phospho): 100	1.52	S7(Phospho): 100; S8(Phospho): 100; T19(Phospho): 91.26	1.81
625192987	DNA ligase 1 isoform X1	AAQVLSSEGEEEDEAPSTPK	S6(Phospho): 100; S7(Phospho): 100	1.58	S6(Phospho): 100; S7(Phospho): 100	1.75
354494097	DNA replication licensing factor MCM3 isoform X1	KKPSEDESDLEDEEEKSQEDPEQK	S4(Phospho): 100; S8(Phospho): 100	2.13	S4(Phospho): 100; S8(Phospho): 100	1.85
625181241	ELL-associated factor 1	TSPLKDNPSPEPQLDDIKR	T1(Phospho): 92.16; S9(Phospho): 100	1.51	S9(Phospho): 100	1.94
625279516	eukaryotic translation initiation factor 3 subunit B isoform X2	AAQPEEETATSPAVSPPPQPAER	S11(Phospho): 99.5; S15(Phospho): 100	1.78	S11(Phospho): 97.67; S15(Phospho): 99.98	1.60
625200598	eukaryotic translation initiation factor 5B isoform X1	VEIFSGSEDDDDSNKLSK	S5(Phospho): 100; S7(Phospho): 100	3.13	S7(Phospho): 99.94	2.13
346716294	gap junction alpha-1 protein	SDPYHATTGPLSPSK	S12(Phospho): 92.5	1.92	S12(Phospho): 90.22	2.31
625205384	GAS2-like protein 1 isoform X1	RYSGDSDSSASSAQSGPPGAR	Y2(Phospho): 91.95	11.32	Y2(Phospho): 99.29	1.87
625207697	high mobility group protein HMG-I/HMG-Y isoform X3	KLEKEEEEGISQESSEEEQ	S14(Phospho): 100; S15(Phospho): 100	2.36	S14(Phospho): 100; S15(Phospho): 100	2.52
625192744	kinectin isoform X1	EIQNGAIHESDSENVLR	S10(Phospho): 98.35	2.53	S10(Phospho): 99.99	1.68
625287056	lamin isoform X2	LRLSPSPTSQR	S4(Phospho): 99.84; S6(Phospho): 86.85	2.00	S4(Phospho): 97.08; S6(Phospho): 97.51	1.73

625259556	leucine-rich repeat flightless-interacting protein 1 isoform X2	SVNEGAEQSLER	S1(Phospho): 100	2.34	S1(Phospho): 100	1.78
625234358	MKI67 FHA domain-interacting nucleolar phosphoprotein	QPVSTVKEEGQNTPTPAR	T13(Phospho): 98.26	2.43	T13(Phospho): 99.27	2.04
625213904	neuroblast differentiation-associated protein AHNAK isoform X1	GGVTGSPEASISGSKGDLK	S6(Phospho): 98.05; S10(Phospho): 96.21	1.95	S6(Phospho): 97.98; S10(Phospho): 99.71; S12(Phospho): 87.94	27.27
625234518	neutral amino acid transporter B(0) isoform X2	TKTPSSEPELIQVK	S6(Phospho): 92.03	23.79	T3(Phospho): 92.28; S6(Phospho): 99.25	2.75
625227303	nuclear fragile X mental retardation-interacting protein 2-like	SLSSDEATNPISR	S3(Phospho): 98.41	1.87	S3(Phospho): 99.09	1.79
625195520	nuclear ubiquitous casein and cyclin-dependent kinase substrate 1 isoform X1	TSASSPLEKSGDEGSEDEAPSGED	S10(Phospho): 100; S15(Phospho): 100	4.87	S10(Phospho): 100; S15(Phospho): 100	2.63
625274316	phosphoserine aminotransferase isoform X2	ASLYNAVTTEDVEK	S2(Phospho): 99.95	137.00	S2(Phospho): 99.12	4.35
625281518	protein NDRG1 isoform X2, partial	SRTASGSSVTSLEGPR	T3(Phospho): 99.31; S5(Phospho): 99.93	24.87	T3(Phospho): 92.96; S5(Phospho): 99.99	120.25
625281518	protein NDRG1 isoform X2, partial	TASGSSVTSLEGPR	S3(Phospho): 99.19	14.45	S3(Phospho): 99.23	11.50
625213093	protein phosphatase 1 regulatory subunit 12A isoform X3	RSTQGVTLTDLQEAEK	T3(Phospho): 93.79	21.72	T3(Phospho): 93.95	13.33

625188491	pumilio homolog 1 isoform X1	RDSLTGSSDLYK	S3(Phospho): 100	5.00	S3(Phospho): 99.96	1.57
625183895	ras and Rab interactor 2 isoform X1	SPEPEPEPEPELGTAGHAGGAPPR	S1(Phospho): 85.21	2.23	S1(Phospho): 100	1.81
625183895	ras and Rab interactor 2 isoform X1	QASFLEAEGSAK	S3(Phospho): 100	2.27	S3(Phospho): 100	2.95
625239472	rho GTPase-activating protein 1 isoform X1	SSSPEPVTHLK	S3(Phospho): 99.39	3.42	S3(Phospho): 92.19	2.79
354494802	ribonucleoside-diphosphate reductase subunit M2 isoform X1	LSLADKENTPPSLSGTR	T9(Phospho): 98.21	6.99	T9(Phospho): 100	2.71
625272657	serine/threonine-protein kinase 10 isoform X2	ANQSRPNSVALETLGGEK	S4(Phospho): 100; S8(Phospho): 100	2.47	S4(Phospho): 100; S8(Phospho): 100	3.45
354481628	serine/threonine-protein kinase DCLK1 isoform X2	SGKSPSPSPTSPGSLRK	S4(Phospho): 98.39; S11(Phospho): 98.59	3.13	S4(Phospho): 75.21	3.42
625282303	serine/threonine-protein kinase SMG1 isoform X3	NLATSADTPPSTIPGTGK	T8(Phospho): 84.47	1.54	S5(Phospho): 78.86; T8(Phospho): 78.86	2.19
625181054	serine/threonine-protein phosphatase 4 regulatory subunit 2 isoform X1	AHSDSPASEPEVSSLSPVK	S3(Phospho): 82.67; S16(Phospho): 97.67	2.33	S3(Phospho): 88.77; S16(Phospho): 88.77	4.04
625240893	small integral membrane protein 13	ELVGDTGSQEGDHEPSGSETEEDPSASPHR	S16(Phospho): 99.41; S18(Phospho): 98.07	2.09	S16(Phospho): 99.12; S18(Phospho): 96.95	3.70
625285767	supervillin isoform X7	EMEKSFDEHNVPK	S5(Phospho): 100	2.10	S5(Phospho): 100	1.77
354477403	thyroid hormone receptor-associated protein 3	ERSPALKSPLQSVVVR	S3(Phospho): 100; S8(Phospho): 99.94	2674.42	S3(Phospho): 100; S8(Phospho): 99.6	3.03

354477403	thyroid hormone receptor-associated protein 3	SPALKSPLQSVVVR	S6(Phospho): 99.3	95.78	S6(Phospho): 99.86	5.49
625290034	transcription intermediary factor 1-beta isoform X2	SRSGEGEVSGLMR	S3(Phospho): 99.48	1.75	S3(Phospho): 94.28	1.78
625265776	uncharacterized protein LOC100772531	TRTDGDSDAEDPGAPLEAVGAGR	S7(Phospho): 98.45	1.60	S7(Phospho): 99.91	1.84
625262091	uncharacterized protein LOC103161432	SRDATPPVSPINMEDQER	S9(Phospho): 100	5.31	T5(Phospho): 99.19; S9(Phospho): 100	4.31
354482483	vimentin	ETNLESLPLVDTHSK	S6(Phospho): 99.95	2.03	S6(Phospho): 99.97	3.29
354482483	vimentin	DGQVINETSQHHDDLE	S9(Phospho): 94.36	4.04	S9(Phospho): 93.11	2.55
354468693	zinc finger SWIM domain-containing protein 8 isoform X1	HTGMASIDSSAPETTSDSSPTLSR	S19(Phospho): 99.58	3.86	S19(Phospho): 90.04	2.51
625262958	zinc transporter SLC39A7 isoform X2	EKPSSEEEEKETGVLR	S4(Phospho): 100; S5(Phospho): 100	4.91	S4(Phospho): 100; S5(Phospho): 100	2.75

Table 7-1. Differentially expressed phosphopeptides with increased expression at 31°C compared to 37°C that were identified in common between the two enrichment methods. PhosphoRS: Best Site Probabilities shows the amino acid that is phosphorylated and % probability score that this site is phosphorylated.

			Titanium Enrichment		Fe-NTA (IMAC) Enrichment	
Accession	Description	Sequence	PhosphoRS: Best Site Probabilities	Fold change	PhosphoRS: Best Site Probabilities	Fold change
625219250	AMP deaminase 2 isoform X1	TDSDSDLQLYK	S3(Phospho): 99.49	2.03	S3(Phospho): 99.51	1.76
625281556	arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 1 isoform X1	TLSDPPSPLPHGPPNK	S3(Phospho): 91.86; S7(Phospho): 100	2.55	S3(Phospho): 98.68; S7(Phospho): 100	2.68
625188804	ataxin-2 isoform X1	TNSPSISPSVLSNSEHK	S3(Phospho): 99.7; S7(Phospho): 99.83	2.11	S3(Phospho): 99.72	3.02
625257315	Bloom syndrome protein isoform X3	DLDDSDKEKDILNTSK	S5(Phospho): 100	2.49	S5(Phospho): 100	2.56
625189773	C2 domain-containing protein 5 isoform X1	SQSESSDEVTELDLSHGK	S3(Phospho): 90.52	4.29	S3(Phospho): 98.06	2.98
625289462	calcium-regulated heat stable protein 1	DRSPSPLRGNVVPSPLPTR	S3(Phospho): 100; S5(Phospho): 100; S14(Phospho): 99.91	1.98	S3(Phospho): 100; S5(Phospho): 100	1.99
625188785	CAP-Gly domain-containing linker protein 1 isoform X1	TASESISNLSEAGSVK	S10(Phospho): 99.47; S14(Phospho): 100	3.06	S5(Phospho): 95.19; S14(Phospho): 99.57	1.90
625186195	catalase isoform X1	SLQKPDVLTTGGGNPIGDK	S1(Phospho): 100	3.15	S1(Phospho): 100	2.94

625235353	cdc42 effector protein 2	LHLESPQPSPK	S5(Phospho): 100	2.53	S5(Phospho): 99.33	1.85
625246587	cdc42-interacting protein 4 isoform X5	VPSDSSLGTPDCRPELR	S3(Phospho): 97.34	4.98	S3(Phospho): 98.25	2.68
625240816	C-Jun-amino-terminal kinase-interacting protein 4 isoform X6	SASQSSLDKLDQELK	S3(Phospho): 83.77	1.56	S3(Phospho): 85.2	3.62
625188428	cyclic AMP-dependent transcription factor ATF-2 isoform X1	NDSVIVADQTPTPTR	T10(Phospho): 99.99; T12(Phospho): 92.62	2.42	T10(Phospho): 99.96	7.85
625210034	cyclic AMP-dependent transcription factor ATF-7	TDSVIIADQTPTPTR	T10(Phospho): 98.85; T14(Phospho): 90.31	10.66	T10(Phospho): 99.99; T12(Phospho): 92.12	2.70
625227028	DENN domain-containing protein 4C isoform X1	VPSGLFDTNNR	S3(Phospho): 100	3.05	S3(Phospho): 100	2.54
625258329	DENN domain-containing protein 5B isoform X2	RLSITSLTGK	S3(Phospho): 98.31	37.11	S3(Phospho): 87.65	2.86
625271453	epidermal growth factor receptor substrate 15-like 1 isoform X7	STPSHGSVSSLNSTGSLSPK	S4(Phospho): 88.79; S18(Phospho): 97.97	18.91	S18(Phospho): 99.6	5.11
625226222	eukaryotic elongation factor 2 kinase isoform X1	KPESDEDSLGSYGR	S4(Phospho): 100; S8(Phospho): 98.43	1.80	S8(Phospho): 99.95	1.51

625269033	eukaryotic translation initiation factor 4 gamma 3 isoform X3	RSPGPVQTAIIAPK	S2(Phospho): 99.75	9.56	S2(Phospho): 99.92	2.67
625269033	eukaryotic translation initiation factor 4 gamma 3 isoform X3	SPGPVQTAIIAPK	S1(Phospho): 100	2.78	S1(Phospho): 100	1.93
625233296	FERM, RhoGEF and pleckstrin domain-containing protein 1 isoform X3	VSALESGPHQSPALSK	S11(Phospho): 99.99	5.70	S11(Phospho): 99.96	2.17
625202041	girdin isoform X1	SSSQENLLDEVMK	S3(Phospho): 88.21	3.10	S3(Phospho): 99.45	10.20
625248717	golgin subfamily A member 4 isoform X2	SEASPPQSGDTQTFAQK	S4(Phospho): 98.91	2.60	S4(Phospho): 99.75	2.71
354468872	interferon regulatory factor 2-binding protein 2, partial	LEEPPELNRQSPNPR	S11(Phospho): 100	2.06	S11(Phospho): 100	2.69
354468872	interferon regulatory factor 2-binding protein 2, partial	AQPAHRSPADSLSGASGASELGADGAGK	S7(Phospho): 98.44	1.82	S7(Phospho): 99.85	2.13
625288986	la-related protein 4B	SPSPAHLPEDPK	S3(Phospho): 99.29	1.74	S3(Phospho): 99.98	1.85
625261027	meiosis-specific nuclear structural protein 1 isoform X2	SLHVQVLNK	S1(Phospho): 100	4.58	S1(Phospho): 100	2.02
625260808	microtubule-associated protein 1B isoform X2	EVPSKEEQSPVKAEVAEK	S9(Phospho): 99.33	2.23	S9(Phospho): 99.52	3.36

625260808	microtubule-associated protein 1B isoform X2	AQSLEGEKLSPKSDISPLTPR	S10(Phospho): 99.99; S13(Phospho): 99.95; T19(Phospho): 99.7	3.05	S10(Phospho): 95.68; S13(Phospho): 98.81	2.27
625260808	microtubule-associated protein 1B isoform X2	SLMSSPEDLTK	S4(Phospho): 99.99; S5(Phospho): 99.99	3.31	S4(Phospho): 99.95; S5(Phospho): 99.77	5.54
625260808	microtubule-associated protein 1B isoform X2	AQSLEGEKLSPK	S10(Phospho): 100	2.65	S10(Phospho): 100	2.54
625260808	microtubule-associated protein 1B isoform X2	SPSLSPSPPSPIEK	S10(Phospho): 99.87	2.21	S10(Phospho): 99.99	1.74
625265796	oxysterol-binding protein-related protein 11 isoform X2	VSESEGKLEGLATAVTPNK	S4(Phospho): 90.88	1.57	S4(Phospho): 91.25	1.52
625189056	oxysterol-binding protein-related protein 3 isoform X1	ALVHQLSNESR	S7(Phospho): 99.57	8.04	S7(Phospho): 100	5.84
625264630	pleckstrin homology domain-containing family A member 5 isoform X3	RAESVKELDPEHR	S4(Phospho): 100	6.05	S4(Phospho): 100	3.01
625283168	poly(A)-specific ribonuclease PARN isoform X2	VSSPCSHTNSFTATGMTGK	S3(Phospho): 92.79	5.81	S3(Phospho): 89.53	1.97
625287066	polyamine-modulated factor 1 isoform X1	GPEGSSPESVPVDTTVSR	S6(Phospho): 78.66	2.05	S6(Phospho): 98.2	1.96

354494786	protein chibby homolog 1 isoform X1	SASLSNLHSLDR	S3(Phospho): 99.2	2.14	S3(Phospho): 89.27	1.80
625288290	protein LSM12 homolog	TETPPPLASLNVSK	T3(Phospho): 99.97	2.82	T3(Phospho): 99.97	2.18
354478888	pygopus homolog 2 isoform X1	GGGTPDANSLAPPGK	T4(Phospho): 99.69	2.33	T4(Phospho): 100	2.45
625212310	tensin-3 isoform X2	LDPVEGPGRSPGQQGDESIAGR	S10(Phospho): 100	27.89	S10(Phospho): 100	3.72
625187665	tumor suppressor p53-binding protein 1 isoform X1	SEGDGENTQIEDTEPLSPVTNSK	S1(Phospho): 97.36; S17(Phospho): 98.9	4.04	S17(Phospho): 97.69	1.77
625184675	uncharacterized protein C1orf198 homolog	SSSLDALGPSR	S3(Phospho): 99.91	5.18	S3(Phospho): 99.97	4.90
625261639	WD repeat and HMG-box DNA-binding protein 1 isoform X2	SHILGDDENSVDAPMLK	S10(Phospho): 100	2.15	S10(Phospho): 100	1.81
625233942	zinc finger protein 609 isoform X2	AEEGKSPFRESSGDGIK	S6(Phospho): 99.98	13.70	S6(Phospho): 99.98	2.53

Table 7-2. Differentially expressed phosphopeptides with decreased expression at 31°C compared to 37°C that were identified in common between the two enrichment methods. PhosphoRS: Best Site Probabilities shows the amino acid that is phosphorylated and % probability score that this site is phosphorylated.

7.4   Discussion

There is much research ongoing to characterise CHO cell factories through the application of 'omics-based approaches' (e.g. genomics, transcriptomics, proteomics, glycomics, metabolomics, fluxomics) that will enable a fundamental understanding of CHO cell physiology. This would result in a better knowledge and understanding of recombinant protein production which will be instrumental in metabolic and cell engineering interventions to enhance CHO cell lines (Kildegaard et al. 2013, Datta, Linhardt and Sharfstein 2013). However, to date little work has been carried out on the phosphoproteomic analysis of recombinant CHO cells. In this 'proof-of-concept' study we have used a quantitative phosphoproteomic approach and have identified 700 phosphopeptides that were found to be differentially expressed following culture temperature shift, using a combined TiO_2 and Fe-NTA phosphopeptide (IMAC) enrichment strategy in conjunction with quantitative label-free LC-MS/MS proteomic analysis. This suggests additional complexity in the regulation of biological and molecular processes in recombinant CHO cells in bioprocess-relevant conditions such as temperature shift that has not been extensively studied to date.

GO analysis of the data using DAVID identified a large number of functional classes altered as a result of temperature shift suggesting the complexity of the molecular and cellular processes that underpin this phenotype in recombinant CHO cells. Many of the enriched functional classes identified are in common with previous studies on recombinant CHO cells subjected to temperature shift including growth, transcription, mRNA translation, and cytoskeletal reorganization (Kantardjieff et al. 2010a, Masterton et al. 2010, Yee, Gerdtzen and Hu 2009a, Roobol et al. 2009, Baik et al. 2006), but the additional complexity surrounding the regulation of these processes by phosphorylation, though well studied in other biological systems and diseases, has not been explored in great detail to date in recombinant CHO cells.

From GO analysis an enrichment of molecular and biological processes linked to growth, in particular cell division, cell cycle, and protein kinase activity, showed decreased expression following temperature shift which is as expected from the observed reduction in growth following temperature shift. A reduction in culture temperature has been shown to arrest cells in G1 phase of the cell cycle with an associated increase in productivity of recombinant protein (Becerra et al. 2012). Phosphopeptides from proteins involved in cell cycle regulation include

CDK14, CDK17, NEK1, KIF23, ANAPC1, and MELK all showed decreased expression following temperature shift. Anaphase-promoting complex subunit 1 (ANAPC1) (and also known as APC1) is a component of the anaphase promoting complex/cyclosome (APC/C) that controls progression through mitosis and the G1 phase of the cell cycle, and phosphorylation of ANAPC1 is required for its activity (Zhang et al. 2016). Maternal embryonic leucine zipper kinase (MELK) is a serine/threonine-protein kinase involved in cell cycle regulation with increased expression during mitosis and is then degraded when cells enter into G1 phase of the cell cycle (Badouel et al. 2010). It has been shown to be overexpressed in various cancers including melanoma and breast cancer resulting in its identification as a potential therapeutic target (Jiang and Zhang 2013). Interestingly is has been shown to be critical for mitotic progression of cancer cells where it directly phosphorylates eIF4B regulating protein synthesis during mitosis (Wang et al. 2016). We also identified a phosphopeptide from eIF4B with reduced expression following temperature shift. Phosphopeptides from a number of proteins which play key roles in MAPK signaling showed decreased expression following temperature shift including MAP3K7, MAP4K4 and PAK1 (P21-activated kinase-1). MAP3K7 (also known as TAK1) is a key regulator of both MAPK and NFkB signaling pathways (Roobol et al. 2009). Recently shRNA silencing of MAP4K4 reduced proliferation and induced G1 cell cycle arrest of gastric cancer cells (Y. F. Liu et al. 2016). PAK1 mediates activation of downstream MAP kinases and is also involved in reorganisation of the actin cytoskeleton (R. Kumar et al. 2017). The phosphoproteomic data therefore suggests a potential role for phosphorylation in regulating control mechanisms during cell cycle arrest of recombinant CHO cells as a result of temperature shift.

From our phosphoproteomic study one of the most significantly changed proteins which showed increased levels of the phosphopeptide TAS*GSSVTSLEGPR at 31˚C compared to 37˚C and the same phosphopeptide detected using both TiO_2 and Fe-NTA enrichment methodologies was identified as Protein NDRG1 (N-Myc down-regulated gene 1). Western blot analysis also shows increased levels of phosphorylation of this specific phosphopeptide at the reduced culture temperature. NDRG1 is highly conserved across a number of multicellular organisms, typically expressed in response to cell stress and is known as a tumour metastasis suppressor in several different types of cancer including breast, colon, pancreatic and prostate (Ghalayini et al. 2013, Bae et al. 2013, McCaig et al. 2011). Reduced expression of NDRG1 has been shown to promote proliferation of prostate cancer cells (Li et al. 2015). Phosphorylation

of NDRG1 also appears to be involved in cell cycle arrest mechanisms (Ghalayini et al. 2013, McCaig et al. 2011). In this study, we found NDRG1 to have increased phosphorylation upon temperature shift. Although cell cycle arrest in temperature shifted CHO cells has been attributed to the activation of the ATR-p53-p21 pathway (Roobol et al. 2011), cellular stress has been shown to result in the expression of NDRG1 in a p53 dependent manner (Roobol et al. 2011, Stein et al. 2004). This would suggest that NDRG1 may be involved in the cell cycle arrest of temperature-shifted CHO cells; however further work would be required to determine the precise mechanisms by which this would occur and if phosphorylation is functionally required for this.

Gene Ontology analysis of all differentially expressed phosphopeptides identified a significant number of proteins involved in the regulation of transcription. Phosphorylation can impact target gene expression by affecting the stability, location and structure of transcription factors (Filtz, Vogel and Leid 2014). Phosphorylation may also regulate the status of other PTMs such as methylation on a protein which can influence the transcription of target genes (Kowenz-Leutz et al. 2010). Phosphopeptides identified from both ATF2 and ATF7, which are both highly conserved members of the AP-1 transcription factor family (Bae et al. 2013), were found to show decreased expression at 31˚C compared to 37˚C. There is also a suggestion from western blot analysis that there is a reduction in phosphorylated levels of ATF2 following temperature shift. While phosphorylation of ATF2 can result in its degradation as well as induction of activity the observed decrease in phosphorylation of Thr-69 following temperature shift (see Figure 7-6 6) would appear to suggest that reduced culture temperature conditions results in a reduction of the active form of ATF2 (Fuchs, Tappin and Ronai 2000). ATF2 is a known target of SAPK/JNK and p38 MAP kinase signalling pathways and has been found to be directly phosphorylated by p38 MAPK and SAPK (Ouwens et al. 2002). Transcription intermediary factor 1-beta (TRIM28) is a transcriptional core-repressor protein involved in a wide variety of functions such as cell growth, apoptosis and DNA repair (Iyengar and Farnham 2011) and was identified as up-regulated at 31˚C using both enrichment strategies. It has been demonstrated using shRNAi that TRIM28 depletion led to increased cell proliferation and overexpression of TRIM28 led to decreased cell proliferation in breast cancer and lung cancer cell lines (Chen et al. 2012). Recently, phosphorylation of TRIM28 on Ser473 by (MAPKAP) kinase 2 (MK2) has been shown to lead to a reduction in its activity (King 2013a).

From our phosphoproteomic study we have also identified a number of differentially expressed phosphopeptides from proteins involved in mRNA translation, with certain phosphopeptides showing increased expression and others showing decreased expression following a reduction in culture temperature. From other studies it has been shown that there is a general reduction in translation following temperature shift but there can also be increased expression of specific target genes including product mRNAs (Becerra et al. 2012). Biological processes related to ribosome biogenesis and ribonucleoprotein complex assembly were enriched following functional analysis of the phosphopeptides showing increased expression at 31$^{\text{o}}$C. Phosphorylation is known to play a role in the regulation of RNA-binding proteins and ribonucleoprotein complex assembly (Thapar 2015). In our study we identified two phosphopeptides using both enrichment strategies from the translation initiation factor EIF4G3 with both showing decreased expression following temperature shift. EIF4G becomes phosphorylated during mitosis inhibiting cap-dependent translation through a reduction in its interaction with eIF4E (Pyronnet, Dostie and Sonenberg 2001). EIF4G3, along with eIF4G1 (phosphopeptide identified as down-regulated also at 31˚C from TiO$_2$ enrichment), form the two functional homologs of eIF4G, part of the cap binding protein complex of eIF4F (Ivanov et al. 2011). In addition to phosphorylation cleavage of EIF4G resulting in loss of function is also known to cause inhibition of cap-dependent translation (Marcet-Palacios et al. 2011). These reports show the pivotal role that EIF4G plays in initiation of protein translation. Altered levels of phosphopeptides from other translation initiation factors were also identified in the study including EIF5B, EIF3A, EIF3B, EIF4EBP1, EIF3G and EIF4B showing the potential importance of phosphorylation in regulating protein synthesis in CHO cells following temperature shift. Recently it has been shown that EEF2K, a threonine kinase that regulates protein synthesis by controlling the rate of peptide chain elongation, regulates the cold stress response by slowing translation elongation through phosphorylation of eEF2 (Knight et al. 2015). Increased phosphorylation of eEF2K was also shown 4 hours and 24 hours following a reduction in temperature from 37$^{\text{o}}$C to 32$^{\text{o}}$C (King 2013b). However, our results show decreased levels of two phosphopeptides to eEF2K following a reduction in culture temperature. Interestingly from closer analysis of the western blots and immunoprecipitation data shown by Knight et al (Knight et al. 2015) no obvious change in Ser500 phosphorylation levels were seen 4 hours after cooling but increased phosphorylation was seen 24 hours after cooling compared to the 37$^{\text{o}}$C culture. Expression levels of eEF2K phosphorylated at Ser398 were increased 4 hours post

temperature reduction but appeared to be back at basal levels by 24 hours (Knight et al. 2015). We did not identify phosphopeptides with these equivalent sites of phosphorylation as being differentially expressed in our study; this could be due to the time point of 36 hours post temperature shift that was used in our study compared to the 4 and 24 hour time points chosen in the other study (Knight et al. 2015). The four different sites of phosphorylation identified in our study were different to those assessed by Knight et al. (i.e. Ser398 and Ser500) (Knight et al. 2015). Three out of the four CHO-specific sites that we identified were conserved between human and CHO and were equivalent to Ser392, Ser470 and Ser474 from human EEF2K as determined using PhosphositePlus (http://www.phosphosite.org/). EEF2K is known to be phosphorylated at multiple different sites and is also prone to auto-phosphorylation demonstrating the complexity of its regulation, and which is still poorly understood (Kenney et al. 2014, Pyr Dit Ruys et al. 2012). It is known, however, that the activity of eEF2K is regulated by calcium ions and calmodulin (Tavares et al. 2017, Kenney et al. 2014). Cytosolic calcium concentrations were shown to be increased more than 10 fold after 15 and 30 minutes of rapid cooling compared with CHO cells maintained at 37°C, providing direct evidence of a potential role of calcium signaling during cooling, and the source of the calcium was determined to be from the ER (Knight et al. 2015). It would be of interest in future phosphoproteomic studies to sample at multiple time-points following a reduction in temperature to capture phosphoproteomic changes of eEF2K over time, to assess altered phosphorylation of the various different phosphosites on eEF2K and to determine which ones are of functional relevance. As phosphorylation events can be rapid, transient, static and reversible (Riley and Coon 2016b) which can affect protein functionality and subsequent signaling cascades and molecular processes that are regulated by phosphorylation, this would suggest that the phosphoproteome would change depending on whether a rapid or slower rate of cooling was used to reduce culture temperature. It would therefore be of interest to carry out a time course study during a slow versus a rapid cooling of culture temperature (which may be of relevance to the working volume of a culture where, for example, larger media volumes would cool slower than smaller media volumes) to determine variations in the phosphoproteome under different cooling conditions.

Functional analysis using GO identified a number of phosphopeptides from proteins involved in regulating cellular shape were identified as showing altered expression in CHO-SEAP cells following temperature shift. Biological processes and molecular functions related to actin

filament organization, cytoskeletal protein binding and actin binding were found to be over-represented from GO analysis of the proteins down-regulated following temperature shift. For example girdin, showing decreased expression at 31˚C and was identified using both enrichment strategies, is a key modulator of the Akt-mTOR signaling pathway and associates with actin and tubulin in the cytoskeleton to help regulate cell shape (C. Liu et al. 2012, Enomoto, Ping and Takahashi 2006). Phosphorylation of Girdin by AKT plays an important role in cytoskeletal reorganisation by regulating cell migration and polarisation (Yamamura et al. 2015). shRNA knockdown of girdin has been found to inhibit cell growth of subcutaneously transplanted HeLa-derived tumors in nude mice (Mao et al. 2012) and phosphorylation of girdin appears to play a role in prostate cancer cell migration and invasion (Tomiyama et al. 2015). We have also identified 4 phosphopeptides found to be down-regulated following temperature shift from the protein Microtubule-associated protein 1B (MAP1B). Phosphorylation of MAP1B is known to play a role in cytoskeletal changes involved in neurite extension (Villarroel-Campos and Gonzalez-Billault 2014). It is known to be phosphorylated by GSK3b, CDKs such as CDK5 and MAPKs such as JNK1 and KNK2 regulating its activity (Villarroel-Campos and Gonzalez-Billault 2014). Previous proteomic and transcriptomic studies on CHO cells subjected to temperature shift to increase productivity have shown an enrichment of molecular processes related to cytoskeletal reorganization (Becerra et al. 2012, Kantardjieff et al. 2010b, Yee, Gerdtzen and Hu 2009b). Our study is suggesting additional potential complexity to the regulation of these processes in CHO cells.

Phosphoproteins themselves and the enzymes that regulate phosphorylation, such as kinases and phosphatases, have the potential to be engineering targets in CHO cells. There have been a small number of studies where signaling pathways, in which phosphorylation plays a key role, have been manipulated to improve growth and productivity of CHO cells. For example, an engineering strategy has also been employed to overexpress mTOR (mammalian target of rapamycin) which is a serine/threonine kinase, in CHO cells, and was found to improve growth, viability, cell size and antibody production (Dreesen and Fussenegger 2011). The mammalian target of rapamycin complex 1 (mTORC1) pathway was recently chemically inhibited using Rapamycin leading to increased growth and productivity in CHO cells (Dadehbeigi and Dickson 2015). Recently Amgen reported on the use of a small molecule inhibitor that directly targets cyclin dependent kinases (CDK) 4/6 in order to control cell growth, and found improved productivity and product quality of a Mab produced from CHO cells (Du et al. 2015).

7.5   Conclusion

Combining all levels of regulation through systems biology models will unveil the underlying complexity inherent in CHO cell biology and will ultimately enhance and accelerate CHO productive capabilities in the coming decades. The identification of 700 differentially expressed phosphopeptides following a reduction in bioprocess culture temperature highlights the importance of including phosphorylation/phosphoproteomic studies in systems biology analyses of CHO cell factories, and which have largely been neglected in CHO 'omics studies to date. The identification of signalling events involved in the regulation of transcription raises some interesting questions regarding the control of transcription in temperature-shifted cells; are such phosphorylation events potentially involved in the increased transcription of the recombinant molecule, attenuating the translation of endogenous cellular proteins or in some way leading to enhanced transcription of target genes? Furthering our understanding of these signalling mechanisms could provide cell engineering targets that would enable the preferential translation of specific gene targets.

Further work will also be required to assess if the site-specific phosphorylation sites identified in the CHO cell proteins from this study have a functional role and also to assess if these phosphoproteins show altered levels of phosphorylation rather than just altered levels of total phosphoprotein itself. The CRISPR/Cas9 genome editing technique which is becoming more widely used for CHO cell engineering (Tomiyama et al. 2015) is an ideal tool to study the phosphoproteome as it is also amenable to targeted site-specific mutagenesis as well as gain/loss of function studies.

7.6   References

Abelin, J.G., Patel, J., Lu, X., Feeney, C.M., Fagbami, L., Creech, A.L., Hu, R., Lam, D., Davison, D., Pino, L., Qiao, J.W., Kuhn, E., Officer, A., Li, J., Abbatiello, S., Subramanian, A., Sidman, R., Snyder, E., Carr, S.A. and Jaffe, J.D. 2016. Reduced-representation phosphosignatures measured by quantitative targeted MS capture cellular states and enable large-scale comparison of drug-induced phenotypes. *Molecular & Cellular Proteomics : MCP,* 15(5), pp.1622-1641.

Angel, T.E., Aryal, U.K., Hengel, S.M., Baker, E.S., Kelly, R.T., Robinson, E.W. and Smith, R.D. 2012. Mass spectrometry-based proteomics: Existing capabilities and future directions. *Chemical Society Reviews,* 41(10), pp.3912-3928.

Badouel, C., Chartrain, I., Blot, J. and Tassan, J.P. 2010. Maternal embryonic leucine zipper kinase is stabilized in mitosis by phosphorylation and is partially degraded upon mitotic exit. *Experimental Cell Research,* 316(13), pp.2166-2173.

Bae, D.H., Jansson, P.J., Huang, M.L., Kovacevic, Z., Kalinowski, D., Lee, C.S., Sahni, S. and Richardson, D.R. 2013. The role of NDRG1 in the pathology and potential treatment of human cancers. *Journal of Clinical Pathology,* 66(11), pp.911-917.

Baik, J.Y., Lee, M.S., An, S.R., Yoon, S.K., Joo, E.J., Kim, Y.H., Park, H.W. and Lee, G.M. 2006. Initial transcriptome and proteome analyses of low culture temperature-induced expression in CHO cells producing erythropoietin. *Biotechnology and Bioengineering,* 93(2), pp.361-371.

Barron, N., Kumar, N., Sanchez, N., Doolan, P., Clarke, C., Meleady, P., O'Sullivan, F. and Clynes, M. 2011. Engineering CHO cell growth and recombinant protein productivity by overexpression of miR-7. *Journal of Biotechnology,* 151(2), pp.204-211.

Carlage, T., Hincapie, M., Zang, L., Lyubarskaya, Y., Madden, H., Mhatre, R. and Hancock, W.S. 2009. Proteomic profiling of a high-producing chinese hamster ovary cell culture. *Analytical Chemistry,* 81(17), pp.7357-7362.

Chen, L., Chen, D.T., Kurtyka, C., Rawal, B., Fulp, W.J., Haura, E.B. and Cress, W.D. 2012. Tripartite motif containing 28 (Trim28) can regulate cell proliferation by bridging HDAC1/E2F interactions. *The Journal of Biological Chemistry,* 287(48), pp.40106-40118.

Choudhary, C. and Mann, M. 2010. Decoding signalling networks by mass spectrometry-based proteomics. *Nature Reviews.Molecular Cell Biology,* 11(6), pp.427-439.

Cohen, P. 2001. The role of protein phosphorylation in human health and disease. the sir hans krebs medal lecture. *European Journal of Biochemistry,* 268(19), pp.5001-5010.

Dadehbeigi, N. and Dickson, A.J. 2015. Chemical manipulation of the mTORC1 pathway in industrially relevant CHOK1 cells enhances production of therapeutic proteins. *Biotechnology Journal,* 10(7), pp.1041-1050.

Datta, P., Linhardt, R.J. and Sharfstein, S.T. 2013. An 'omics approach towards CHO cell engineering. *Biotechnology and Bioengineering,* 110(5), pp.1255-1271.

Dephoure, N., Gould, K.L., Gygi, S.P. and Kellogg, D.R. 2013. Mapping and analysis of phosphorylation sites: A quick guide for cell biologists. *Molecular Biology of the Cell,* 24(5), pp.535-542.

Diring, J., Camuzeaux, B., Donzeau, M., Vigneron, M., Rosa-Calatrava, M., Kedinger, C. and Chatton, B. 2011. A cytoplasmic negative regulator isoform of ATF7 impairs ATF7 and ATF2 phosphorylation and transcriptional activity. *PloS One,* 6(8), pp.e23351.

Doolan, P., Meleady, P., Barron, N., Henry, M., Gallagher, R., Gammell, P., Melville, M., Sinacore, M., McCarthy, K., Leonard, M., Charlebois, T. and Clynes, M. 2010. Microarray and proteomics expression profiling identifies several candidates, including the valosin-containing protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines. *Biotechnology and Bioengineering,* 106(1), pp.42-56.

Dreesen, I.A. and Fussenegger, M. 2011. Ectopic expression of human mTOR increases viability, robustness, cell size, proliferation, and antibody production of chinese hamster ovary cells. *Biotechnology and Bioengineering,* 108(4), pp.853-866.

Du, Z., Treiber, D., McCarter, J.D., Fomina-Yadlin, D., Saleem, R.A., McCoy, R.E., Zhang, Y., Tharmalingam, T., Leith, M., Follstad, B.D., Dell, B., Grisim, B., Zupke, C., Heath, C., Morris, A.E. and Reddy, P. 2015. Use of a small molecule cell cycle inhibitor to control cell growth and improve specific productivity and product quality of recombinant proteins in CHO cell cultures. *Biotechnology and Bioengineering,* 112(1), pp.141-155.

Enomoto, A., Ping, J. and Takahashi, M. 2006. Girdin, a novel actin-binding protein, and its family of proteins possess versatile functions in the akt and wnt signaling pathways. *Annals of the New York Academy of Sciences,* 1086pp.169-184.

Farrell, A., McLoughlin, N., Milne, J.J., Marison, I.W. and Bones, J. 2014. Application of multi-omics techniques for bioprocess design and optimization in chinese hamster ovary cells. *Journal of Proteome Research,* 13(7), pp.3144-3159.

Filtz, T.M., Vogel, W.K. and Leid, M. 2014. Regulation of transcription factor activity by interconnected post-translational modifications. *Trends in Pharmacological Sciences,* 35(2), pp.76-85.

Fischer, S., Handrick, R. and Otte, K. 2015. The art of CHO cell engineering: A comprehensive retrospect and future perspectives. *Biotechnology Advances,* 33(8), pp.1878-1896.

Fuchs, S.Y., Tappin, I. and Ronai, Z. 2000. Stability of the ATF2 transcription factor is regulated by phosphorylation and dephosphorylation. *The Journal of Biological Chemistry,* 275(17), pp.12560-12564.

Ghalayini, M.K., Dong, Q., Richardson, D.R. and Assinder, S.J. 2013. Proteolytic cleavage and truncation of NDRG1 in human prostate cancer cells, but not normal prostate epithelial cells. *Bioscience Reports,* 33(3), pp.10.1042/BSR20130042.

Gutierrez, J.M. and Lewis, N.E. 2015. Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. *Biotechnology Journal,* 10(7), pp.939-949.

Hacker, D.L., De Jesus, M. and Wurm, F.M. 2009. 25 years of recombinant proteins from reactor-grown cells - where do we go from here? *Biotechnology Advances,* 27(6), pp.1023-1027.

Ivanov, I.P., Firth, A.E., Michel, A.M., Atkins, J.F. and Baranov, P.V. 2011. Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Research,* 39(10), pp.4220-4234.

Iyengar, S. and Farnham, P.J. 2011. KAP1 protein: An enigmatic master regulator of the genome. *The Journal of Biological Chemistry,* 286(30), pp.26267-26276.

Jensen, L.J., Jensen, T.S., de Lichtenberg, U., Brunak, S. and Bork, P. 2006. Co-evolution of transcriptional and post-translational cell-cycle regulation. *Nature,* 443(7111), pp.594-597.

Jiang, P. and Zhang, D. 2013. Maternal embryonic leucine zipper kinase (MELK): A novel regulator in cell cycle control, embryonic development, and cancer. *International Journal of Molecular Sciences,* 14(11), pp.21551-21560.

Jiang, P. and Zhang, D. 2013. Maternal embryonic leucine zipper kinase (MELK): A novel regulator in cell cycle control, embryonic development, and cancer. *International Journal of Molecular Sciences,* 14(11), pp.21551-21560.

Kall, L., Canterbury, J.D., Weston, J., Noble, W.S. and MacCoss, M.J. 2007. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods,* 4(11), pp.923-925.

Kantardjieff, A., Jacob, N.M., Yee, J.C., Epstein, E., Kok, Y.J., Philp, R., Betenbaugh, M. and Hu, W.S. 2010. Transcriptome and proteome analysis of chinese hamster ovary cells under low temperature and butyrate treatment. *Journal of Biotechnology,* 145(2), pp.143-159.

Kaufmann, H., Mazur, X., Fussenegger, M. and Bailey, J.E. 1999. Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells. *Biotechnology and Bioengineering,* 63(5), pp.573-582.

Kenney, J.W., Moore, C.E., Wang, X. and Proud, C.G. 2014. Eukaryotic elongation factor 2 kinase, an unusual enzyme with multiple roles. *Advances in Biological Regulation,* 55pp.15-27.

Kildegaard, H.F., Baycin-Hizal, D., Lewis, N.E. and Betenbaugh, M.J. 2013. The emerging CHO systems biology era: Harnessing the 'omics revolution for biotechnology. *Current Opinion in Biotechnology,* 24(6), pp.1102-1107.

King, C.A. 2013. Kaposi's sarcoma-associated herpesvirus kaposin B induces unique monophosphorylation of STAT3 at serine 727 and MK2-mediated inactivation of the STAT3 transcriptional repressor TRIM28. *Journal of Virology,* 87(15), pp.8779-8791.

Knight, J.R., Bastide, A., Roobol, A., Roobol, J., Jackson, T.J., Utami, W., Barrett, D.A., Smales, C.M. and Willis, A.E. 2015. Eukaryotic elongation factor 2 kinase regulates the cold stress response by slowing translation elongation. *The Biochemical Journal,* 465(2), pp.227-238.

Kowenz-Leutz, E., Pless, O., Dittmar, G., Knoblich, M. and Leutz, A. 2010. Crosstalk between C/EBPbeta phosphorylation, arginine methylation, and SWI/SNF/Mediator implies an indexing transcription factor code. *The EMBO Journal,* 29(6), pp.1105-1115.

Kumar, R., Sanawar, R., Li, X. and Li, F. 2017. Structure, biochemistry, and biology of PAK kinases. *Gene,* 605pp.20-31.

Kumar, N., Gammell, P., Meleady, P., Henry, M. and Clynes, M. 2008. Differential protein expression following low temperature culture of suspension CHO-K1 cells. *BMC Biotechnology,* 8pp.42-6750-8-42.

Lee, J.S., Grav, L.M., Lewis, N.E. and Faustrup Kildegaard, H. 2015. CRISPR/Cas9-mediated genome engineering of CHO cell factories: Application and perspectives. *Biotechnology Journal,* 10(7), pp.979-994.

Li, Y., Pan, P., Qiao, P. and Liu, R. 2015. Downregulation of N-myc downstream regulated gene 1 caused by the methylation of CpG islands of NDRG1 promoter promotes proliferation and invasion of prostate cancer cells. *International Journal of Oncology,* 47(3), pp.1001-1008.

Lipscomb, M.L., Palomares, L.A., Hernandez, V., Ramirez, O.T. and Kompala, D.S. 2005. Effect of production method and gene amplification on the glycosylation pattern of a secreted reporter protein in CHO cells. *Biotechnology Progress,* 21(1), pp.40-49.

Liu, Y.F., Qu, G.Q., Lu, Y.M., Kong, W.M., Liu, Y., Chen, W.X. and Liao, X.H. 2016. Silencing of MAP4K4 by short hairpin RNA suppresses proliferation, induces G1 cell cycle arrest and induces apoptosis in gastric cancer cells. *Molecular Medicine Reports,* 13(1), pp.41-48.

Macek, B., Mann, M. and Olsen, J.V. 2009. Global and site-specific quantitative phosphoproteomics: Principles and applications. *Annual Review of Pharmacology and Toxicology,* 49pp.199-221.

Mao, J.Z., Jiang, P., Cui, S.P., Ren, Y.L., Zhao, J., Yin, X.H., Enomoto, A., Liu, H.J., Hou, L., Takahashi, M. and Zhang, B. 2012. Girdin locates in centrosome and midbody and plays an important role in cell division. *Cancer Science,* 103(10), pp.1780-1787.

Marcet-Palacios, M., Duggan, B.L., Shostak, I., Barry, M., Geskes, T., Wilkins, J.A., Yanagiya, A., Sonenberg, N. and Bleackley, R.C. 2011. Granzyme B inhibits vaccinia virus production through proteolytic cleavage of eukaryotic initiation factor 4 gamma 3. *PLoS Pathogens,* 7(12), pp.e1002447.

Masterton, R.J., Roobol, A., Al-Fageeh, M.B., Carden, M.J. and Smales, C.M. 2010. Post-translational events of a model reporter protein proceed with higher fidelity and accuracy upon mild hypothermic culturing of chinese hamster ovary cells. *Biotechnology and Bioengineering,* 105(1), pp.215-220.

McCaig, C., Potter, L., Abramczyk, O. and Murray, J.T. 2011. Phosphorylation of NDRG1 is temporally and spatially controlled during the cell cycle. *Biochemical and Biophysical Research Communications,* 411(2), pp.227-234.

Meleady, P., Doolan, P., Henry, M., Barron, N., Keenan, J., O'Sullivan, F., Clarke, C., Gammell, P., Melville, M.W., Leonard, M. and Clynes, M. 2011. Sustained productivity in recombinant chinese hamster ovary (CHO) cell lines: Proteome analysis of the molecular basis for a process-related phenotype. *BMC Biotechnology,* 11pp.78-6750-11-78.

Meleady, P., Henry, M., Gammell, P., Doolan, P., Sinacore, M., Melville, M., Francullo, L., Leonard, M., Charlebois, T. and Clynes, M. 2008. Proteomic profiling of CHO cells with enhanced rhBMP-2 productivity following co-expression of PACEsol. *Proteomics,* 8(13), pp.2611-2624.

Narumi, R., Murakami, T., Kuga, T., Adachi, J., Shiromizu, T., Muraoka, S., Kume, H., Kodera, Y., Matsumoto, M., Nakayama, K., Miyamoto, Y., Ishitobi, M., Inaji, H., Kato, K. and Tomonaga, T. 2012. A strategy for large-scale phosphoproteomics and SRM-based validation of human breast cancer tissue samples. *Journal of Proteome Research,* 11(11), pp.5311-5322.

Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., Brunak, S. and Mann, M. 2010. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Science Signaling,* 3(104), pp.ra3.

Orellana, C.A., Marcellin, E., Schulz, B.L., Nouwens, A.S., Gray, P.P. and Nielsen, L.K. 2015. High-antibody-producing chinese hamster ovary cells up-regulate intracellular protein transport and glutathione synthesis. *Journal of Proteome Research,* 14(2), pp.609-618.

Ouwens, D.M., de Ruiter, N.D., van der Zon, G.C., Carter, A.P., Schouten, J., van der Burgt, C., Kooistra, K., Bos, J.L., Maassen, J.A. and van Dam, H. 2002. Growth factors can activate ATF2 via a two-step mechanism: Phosphorylation of Thr71 through the ras-MEK-ERK pathway and of Thr69 through RalGDS-src-p38. *The EMBO Journal,* 21(14), pp.3782-3793.

Pyr Dit Ruys, S., Wang, X., Smith, E.M., Herinckx, G., Hussain, N., Rider, M.H., Vertommen, D. and Proud, C.G. 2012. Identification of autophosphorylation sites in eukaryotic elongation factor-2 kinase. *The Biochemical Journal,* 442(3), pp.681-692.

Pyronnet, S., Dostie, J. and Sonenberg, N. 2001. Suppression of cap-dependent translation in mitosis. *Genes & Development,* 15(16), pp.2083-2093.

Riley, N.M. and Coon, J.J. 2016. Phosphoproteomics in the age of rapid and deep proteome profiling. *Analytical Chemistry,* 88(1), pp.74-94.

Roh, Y.S., Song, J. and Seki, E. 2014. TAK1 regulates hepatic cell survival and carcinogenesis. *Journal of Gastroenterology,* 49(2), pp.185-194.

Roitinger, E., Hofer, M., Kocher, T., Pichler, P., Novatchkova, M., Yang, J., Schlogelhofer, P. and Mechtler, K. 2015. Quantitative phosphoproteomics of the ataxia telangiectasia-mutated (ATM) and ataxia telangiectasia-mutated and rad3-related (ATR) dependent DNA damage response in arabidopsis thaliana. *Molecular & Cellular Proteomics : MCP,* 14(3), pp.556-571.

Roobol, A., Carden, M.J., Newsam, R.J. and Smales, C.M. 2009. Biochemical insights into the mechanisms central to the response of mammalian cells to cold stress and subsequent rewarming. *The FEBS Journal,* 276(1), pp.286-302.

Roobol, A., Roobol, J., Carden, M.J., Bastide, A., Willis, A.E., Dunn, W.B., Goodacre, R. and Smales, C.M. 2011. ATR (ataxia telangiectasia mutated- and Rad3-related kinase) is activated by mild hypothermia in mammalian cells and subsequently activates p53. *The Biochemical Journal,* 435(2), pp.499-508.

Snider, N.T. and Omary, M.B. 2014. Post-translational modifications of intermediate filament proteins: Mechanisms and functions. *Nature Reviews.Molecular Cell Biology,* 15(3), pp.163-177.

Sommeregger, W., Mayrhofer, P., Steinfellner, W., Reinhart, D., Henry, M., Clynes, M., Meleady, P. and Kunert, R. 2016. Proteomic differences in recombinant CHO cells producing two similar antibody fragments. *Biotechnology and Bioengineering,* 113(9), pp.1902-1912.

Stein, S., Thomas, E.K., Herzog, B., Westfall, M.D., Rocheleau, J.V., Jackson, R.S.,2nd, Wang, M. and Liang, P. 2004. NDRG1 is necessary for p53-dependent apoptosis. *The Journal of Biological Chemistry,* 279(47), pp.48930-48940.

Taus, T., Kocher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C. and Mechtler, K. 2011. Universal and confident phosphorylation site localization using phosphoRS. *Journal of Proteome Research,* 10(12), pp.5354-5362.

Tavares, C.D., Giles, D.H., Stancu, G., Chitjian, C.A., Ferguson, S.B., Wellmann, R.M., Kaoud, T.S., Ghose, R. and Dalby, K.N. 2017. Signal integration at elongation factor 2 kinase: THE ROLES OF CALCIUM, CALMODULIN, AND SER-500 PHOSPHORYLATION. *The Journal of Biological Chemistry,* 292(5), pp.2032-2045.

Thapar, R. 2015. Structural basis for regulation of RNA-binding proteins by phosphorylation. *ACS Chemical Biology,* 10(3), pp.652-666.

Tomiyama, L., Sezaki, T., Matsuo, M., Ueda, K. and Kioka, N. 2015. Loss of Dlg5 expression promotes the migration and invasion of prostate cancer cells via girdin phosphorylation. *Oncogene,* 34(9), pp.1141-1149.

Underhill, M.F., Birch, J.R., Smales, C.M. and Naylor, L.H. 2005. eIF2alpha phosphorylation, stress perception, and the shutdown of global protein synthesis in cultured CHO cells. *Biotechnology and Bioengineering,* 89(7), pp.805-814.

Villarroel-Campos, D. and Gonzalez-Billault, C. 2014. The MAP1B case: An old MAP that is new again. *Developmental Neurobiology,* 74(10), pp.953-971.

von Stechow, L., Francavilla, C. and Olsen, J.V. 2015. Recent findings and technological advances in phosphoproteomics for cells and tissues. *Expert Review of Proteomics,* 12(5), pp.469-487.

Walsh, G. 2014. Biopharmaceutical benchmarks 2014. *Nature Biotechnology,* 32(10), pp.992-1000.

Wang, Y., Begley, M., Li, Q., Huang, H.T., Lako, A., Eck, M.J., Gray, N.S., Mitchison, T.J., Cantley, L.C. and Zhao, J.J. 2016. Mitotic MELK-eIF4B signaling controls protein synthesis and tumor cell survival. *Proceedings of the National Academy of Sciences of the United States of America,* 113(35), pp.9810-9815.

Wang, Y., Hu, L., Ji, P., Teng, F., Tian, W., Liu, Y., Cogdell, D., Liu, J., Sood, A.K., Broaddus, R., Xue, F. and Zhang, W. 2016. MIIP remodels Rac1-mediated cytoskeleton structure in suppression of endometrial cancer metastasis. *Journal of Hematology & Oncology,* 9(1), pp.112-016-0342-6.

Wurm, F.M. 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotechnology,* 22(11), pp.1393-1398.

Yamamura, Y., Asai, N., Enomoto, A., Kato, T., Mii, S., Kondo, Y., Ushida, K., Niimi, K., Tsunoda, N., Nagino, M., Ichihara, S., Furukawa, K., Maeda, K., Murohara, T. and Takahashi, M. 2015. Akt-girdin signaling in cancer-associated fibroblasts contributes to tumor progression. *Cancer Research,* 75(5), pp.813-823.

Yan, G.R. and He, Q.Y. 2008. Functional proteomics to identify critical proteins in signal transduction pathways. *Amino Acids,* 35(2), pp.267-274.

Yang, Z., Yang, F., Zhang, Y., Wang, X., Shi, J., Wei, H., Sun, F. and Yu, Y. 2018. Girdin protein: A potential metastasis predictor associated with prognosis in lung cancer. *Experimental and Therapeutic Medicine,* 15(3), pp.2837-2843.

Yee, J.C., Gerdtzen, Z.P. and Hu, W.S. 2009. Comparative transcriptome analysis to unveil genes affecting recombinant protein productivity in mammalian cells. *Biotechnology and Bioengineering,* 102(1), pp.246-263.

Zhang, S., Chang, L., Alfieri, C., Zhang, Z., Yang, J., Maslen, S., Skehel, M. and Barford, D. 2016. Molecular mechanism of APC/C activation by mitotic phosphorylation. *Nature,* 533(7602), pp.260-264.

The expression pattern of the phosphoproteome is significantly changed during the growth phases of recombinant CHO cell culture.

Authors: Prashant Kaushik, Michael Henry, Martin Clynes, Paula Meleady

Within this work, Michael Henry specifically performed the methodology and verification of the LC-MS/MS analysis on the Orbitrap Fusion for Phosphopeptide or whole cell lysate label free differential expression analysis. Michael Henry implemented the differential expression analysis tools from both sets of samples using label free software Progenesis QI and searched the statistically significant protein features using MS search algorithms and site specific phosphorylation algorithms. Michael Henry prepared presentation of the published work, specifically the visualisation and data presentation.

Abstract

Post-translational modification of proteins by reversible phosphorylation plays a pivotal role in regulating key cellular processes including transcription, translation, cell proliferation, differentiation, apoptosis, and signal transduction. Despite the importance of the phosphorylation level of regulation, little work has been carried out on the phosphoproteomic characterisation of Chinese hamster ovary (CHO) cells in bioprocess-relevant conditions. Growth control strategies are often used to prolong culture duration and increase specific productivity; however, the cellular mechanisms and regulatory pathways underlying growth strategies are poorly understood in CHO cells. Phosphorylation changes are dynamic and will respond to changes in culture conditions; this may reflect the status of the cells with respect to growth and viability of the culture. In this study, we have used a phosphopeptide enrichment strategy in conjunction with LC-MS/MS to carry out a large-scale differential phosphoproteomic analysis of IgG producing CHO DP12 cells at various phases of growth in serum-free suspension batch culture to characterise dynamic changes to the phosphoproteome with changing culture conditions. In total over the various growth phases, we have identified 3,777 differentially expressed unique phosphopeptides from 1,415 differentially expressed unique phosphoproteins. Analysis of the whole cell lysate without phosphopeptide enrichment over the various growth phases revealed the differential expression of 834 unique proteins, with an overlap of 188 proteins between the proteomic and phosphoproteomic analyses. The inclusion of phosphoproteomic data significantly improves proteome coverage but also gives insights into the post-translational level of regulation during cellular growth of recombinant CHO cells.

8.1   Introduction

Chinese hamster ovary (CHO) cells are the expression system of choice for the industrial production of recombinant protein therapeutics due to their ability to grow to high densities in suspension culture, their robust nature, their track record in industry, their ability to produce human-like post-translational modifications (e.g. glycosylation) and their capability of producing high quality protein therapeutics (Walsh 2014, Omasa, Onitsuka and Kim 2010, Wurm 2004). As a result, there is significant scientific and commercial interest in research to significantly improve the efficiency of biotherapeutic production from CHO cells (Fischer, Handrick and Otte 2015). In recent years, the protein production capabilities of recombinant CHO cells have significantly improved (Bandaranayake and Almo 2014), routinely producing high-quality protein in grams/L concentrations (Huang et al. 2010).   Most of these improvements in productivity and product quality can be attributed to bioprocess optimisation approaches such as in media optimisation, feed development, bioreactor design, etc., and often these have been based on iterative empirical approaches which can be time-consuming and costly (Stolfa et al. 2018, Hacker, De Jesus and Wurm 2009).   In order to further improve the production capabilities and efficiency of CHO cells, an increased understanding of CHO cell biology is of the utmost importance. The next generation of improvements is expected to be made via genetic engineering of the host (CHO) cell itself to improve the efficiency of the production of biotherapeutic protein products (Fischer et al. 2015). Drug discovery pipelines of biopharmaceutical companies are producing ever-more complex molecules which are putting increasing pressure on cell line development groups to generate enough material to meet requirements for clinical trials. The recent sequencing of CHO genomes has opened up new possibilities for the application of systems biology to understand and manipulate CHO cell biology (Kaas et al. 2015, Lewis et al. 2013, Brinkrolf et al. 2013, Xu et al. 2011) and will enable a fundamental understanding of CHO cell physiology. This improved understanding of recombinant protein production will be instrumental in metabolic and cell engineering interventions to enhance CHO cell line (Datta, Linhardt and Sharfstein 2013, Kildegaard et al. 2013).

The vast majority of 'omic based studies in Chinese hamster ovary (CHO) cells carried out to date have continued to miss out on the post-translational level of regulation. Phosphorylation events are crucial to the regulation of growth, cell cycle arrest, apoptosis, transcription, signal transduction, etc. (Yan and He 2008), and hence are likely to be central to understanding and controlling bioprocess-relevant phenotypes. Post-translational modifications, either by proteolytic cleavage or by covalent modification of amino acids, can bring about changes in protein characteristics and functionality (Choudhary and Horvath 1996). Furthermore, genomic and transcriptomic studies are likely to reveal an incomplete picture of regulation if post-translational regulation of the final effector molecules, i.e., through a post-translational modification on the regulatory protein, is ignored (V. M. Tan, Cheng and Drake 2016). Hence, in order to gain a complete insight into the biology of CHO cells, a global differential study of post-translational modifications such as reversible phosphorylation of proteins is of great significance (Henry et al. 2017, Gutierrez and Lewis 2015).

Post-translational modification of proteins by reversible phosphorylation on serine, threonine and tyrosine residues is a major switch mechanism of protein activity(Humphrey, James and Mann 2015). The importance of phosphorylation regulation is demonstrated in that up to 2% of the eukaryotic genome codes for proteins from the kinase gene family (Caenepeel et al. 2004). The human genome contains approximately 568 protein kinases and 156 protein phosphatases to regulate phosphorylation events and play a crucial role in the control of biological processes such as proliferation and apoptosis (Ardito et al. 2017). Furthermore, up to 30% of all cellular proteins are estimated to be phosphorylated on at least one residue at any given time, though not all may be functionally relevant (Cohen 2000).

Mass spectrometry has now become the principal technique for global and targeted analysis of phosphoproteins and phosphorylation events (Vyse, Desmond and Huang 2017, Savitski et al. 2010). Despite the widespread phosphorylation of proteins, a functionally phosphorylated isoform of a particular regulatory protein is significantly lower than the total amount of the protein present in the cell, therefore the large-scale analysis of phosphoproteins/phosphopeptides is generally achieved by enrichment strategies such as

immobilised metal ion ($Fe3^+$ or $Ga3^+$) affinity chromatography (IMAC) and metal oxide affinity chromatography (MOAC) using TiO_2 or ZrO_2 (Angel et al. 2012).

It is expected that proteomic and phosphoproteomic patterns will change during the different growth phases of a batch culture and these changes may give an indication of the cellular state regarding viability, growth, and productivity similar to mRNA and miRNA profiling studies of CHO cells grown in batch suspension culture(Bort et al. 2012). In this study, we have applied a well-established technique for phosphopeptide enrichment from CHO DP12 whole cell lysate samples. Immobilized metal affinity chromatography (IMAC; Fe-NTA) (Wang et al. 2015) was used to enrich phosphopeptides from CHO DP12 cells prior to analysis by LC-MS/MS at different phases of cellular growth in serum-free suspension batch culture, i.e. during the exponential phase, stationary phase and decline/death phase of culture. The data generated emphasises the dynamic nature of the changing proteome and phosphoproteome during the different growth phases over time in culture. The data also clearly demonstrates the importance of including multiple time points in experiments when comparing CHO cellular phenotypes or culture conditions (Bort et al. 2012).

8.2 Materials and Methods

8.2.1 Assessment of cell number and culture viability.

IgG-producing CHO DP12 cells (purchased from the ATCC) were seeded at 3×10^5 cells/mL in a 50 mL working volume containing CHO-S-SFM II serum free media (Thermo Fisher Scientific) in suspension batch culture at 37 ^{o}C and 5 % CO_2 in a Kuhner Climico-Shaker ISF1-X orbital shaker at 170 rpm. A total of 12 flasks were seeded, and three replicate flasks were harvested for sample collection at each time point, i.e., at 48 hours, 72 hours, 96 hours and 120 hours after initial seeding, to capture early exponential, exponential, stationary and death phase of culture based on the growth profiling experiments. Cellular growth and culture viability were assayed using a Guava Benchtop Cytometer after staining with Viacount (Millipore). All cells were tested for *Mycoplasma* using the Hoeschst indirect culture test carried out in-house and were found to be free of contamination.

8.2.2 Cell lysate preparation

Cells were pelleted by centrifugation, washed with cold PBS and were stored at -80°C until required. Cells were resuspended in 8 M urea, 50 mM Tris, 75 mM NaCl (pH 8.2) buffer supplemented with 1X Halt protease inhibitors (Thermo Fisher Scientific), 1X Halt phosphatase inhibitors (Thermo Fisher Scientific), and 0.1% ProteaseMAX surfactant detergent (Promega). Cells were sonicated at a medium setting for 30 seconds for a total of three pulses or until a homogeneous, free-flowing mixture of cell lysate was observed with at least a 5 minute rest between each sonication pulse. The cell lysate was centrifuged at 14,000 x*g* for 10 minutes at 4 °C and the supernatant was transferred to a clean microcentrifuge tube. Protein quantitation was carried out using the Quick Start Bradford Protein Assay (BioRad). The remaining supernatant was stored at -80 °C.

8.2.3 In-solution protein digestion and phosphopeptide enrichment

2 mg of each cell lysate protein preparation was transferred to a new centrifuge tube, and all samples were equalised to the same volume using the same resuspension buffer as described above. A fresh stock of 0.5 M dithiothreitol (DTT) was prepared, and an appropriate volume of DTT was added to achieve a final concentration of 5 mM. Samples were reduced for 25 minutes at 56 °C. Before alkylation, samples were allowed to cool to room temperature and an appropriate volume of freshly prepared 0.5 M iodoacetamide was added to a final concentration of 14 mM. Samples were alkylated for 30 minutes at room temperature in the dark. Protein samples were diluted at a ratio of 1:5 in 25 mM Tris-HCl, pH 8.2, to reduce the concentration of urea to less than 1.6 M. Samples were digested with sequence grade modified trypsin (Thermo Fisher Scientific) at a ratio of 1:100 enzyme:substrate for 4 hours at 37 °C. Fresh trypsin at the same ratio was then added to each sample for an overnight digestion at 37 °C. After overnight digestion, TFA was added to a final concentration of 0.4% to inactivate the trypsin. Digested samples were desalted and concentrated using a Sep-Pak C18 Vac cartridge (Waters) with negative pressure. The C18 cartridge was washed and conditioned by adding 9 mL of acetonitrile (ACN) followed by 3 mL of 50% ACN and 0.5% acetic acid. The cartridge was then equilibrated with 9 mL of 0.1% TFA, and the samples were loaded in 0.4% TFA. Loaded samples were desalted with 9 mL 0.1% TFA. TFA was removed with 1 mL 0.5% acetic acid. Desalted peptides were eluted with 6 mL of 50% ACN, 0.5% acetic acid. The eluates from each sample were snap frozen with liquid nitrogen, lyophilized by freeze drying to dry powder and stored at -20 °C. 10 µg of digested protein was set aside for non-enriched sample analysis. 2 mg of tryptically digested protein samples were phosphopeptide-enriched using Fe-NTA (IMAC) spin columns (Pierce, Thermo Fisher Scientific) as per manufacturer's instructions.

Non-enriched peptide and phosphopeptide sample concentrations were determined using a Nanodrop One (Laptech International, UK).

8.2.4   LC-MS/MS analysis

LC-MS/MS separations were performed using an UltiMate 3000 nanoRSLC system (Thermo Scientific) coupled in-line with an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific).

For non-enriched samples, 1.3 µL of cell lysate samples (equivalent to 1 µg of digested peptides) were loaded onto the trapping column (PepMap100, C18, 300 µm × 5 mm) (Thermo Scientific) for 3 minutes at a flow rate of 25 µL/min with 2% (v/v) ACN, 0.1% (v/v) TFA. Peptides were resolved on an analytical column (Easy-Spray C18 75 µm × 250 mm, 2 µm bead diameter column) (Thermo Scientific) using a gradient of 98% A (0.1% (v/v) formic acid (FA)): 2% B (80% (v/v) ACN, 0.08% (v/v) FA) to 35% B over 120 min at a flow rate of 300 nL/min.

MS(/MS) data were acquired on the Orbitrap Fusion Tribrid as follows for the non-enriched whole cell lysate samples: all MS1 spectra were acquired over m/z 400–1500 in the Orbitrap (120 K resolution at 200 m/z), automatic gain control (AGC) was set to accumulate $4{\times}10^5$ ions with a maximum injection time of 100 ms. Data-dependent tandem MS analysis was performed using a top-speed approach (cycle time of 3 s), and the normalised collision energy was optimised at 35% for Collision Induced Dissociation (CID). MS2 spectra were acquired with a fixed first m/z of 100. The intensity threshold for fragmentation was set to 5000 and included charge states 2+ to 7+. A dynamic exclusion of 50 s was applied with a mass tolerance of 10 ppm.

For the phosphopeptide-enriched samples, 2.5 µL of sample (equivalent to 1 µg of digested phosphopeptides) were loaded onto the trapping column (PepMap100, C18, 300 µm × 5 mm), using injection pick-up, for 3 min at a flow rate of 25 µL/min with 2% (v/ v) ACN, 0.1% (v/v) TFA and then resolved on an analytical column (Easy-Spray C18 75 µm × 250 mm, 2 µm bead diameter column) using a gradient of 98% A (0.1% (v/v) formic acid (FA)): 2% B (80% (v/v) ACN, 0.0.08% (v/v) FA) to 35% B over 120 min at a flow rate of 300 nL/min. For the phosphopeptide-enriched samples MS(/MS) data were acquired on the Orbitrap Fusion Tribrid as follows: all MS1 spectra were acquired over m/z 400–1400 in the Orbitrap (120 K resolution at 200 m/z), automatic gain control (AGC) was set to accumulate 4×10^5 ions, with a maximum injection time of 50 ms. Data-dependent tandem MS analysis was performed using a top-speed approach (cycle time of 3 s) and the normalised collision energy CID was optimised at 35%.

MS2 spectra were acquired with a fixed first m/z of 100. The intensity threshold for fragmentation was set to 10,000 and included charge states 2+ to 6+. A dynamic exclusion of 50 s was applied with a mass tolerance of 10 ppm. Multistage activation was enabled for neutral-loss triggered fragmentation for all precursor ions exhibiting neutral loss of mass 97.9763 or 80 Da with a mass tolerance of 0.5 m/z, where the neutral loss ion was one of the top 10 most intense MS2 ions. AGC was set to 20,000 with a maximum injection time set at 90 ms.

8.2.5 Quantitative Label-free LC-MS/MS Data Analysis

Differential quantitative label-free LC-MS/MS data analysis was carried by interrogating the resultant LC-MS/MS files using the software package Progenesis QI for Proteomics (NonLinear Dynamics, Waters) essentially as previously described (Henry et al. 2017). Raw MS data files from each of the 12 samples were imported into the Progenesis QI software interface, and automatic reference run alignment was carried out to account for variability between runs. Upon alignment of all runs, identified features were filtered based on ANOVA p-value < 0.05 between experimental groups. For proteomic and phosphoproteomic analysis, a mascot generic file (mgf) was generated from all exported MS/MS spectra and analysed using Proteome Discoverer v.2.1 (Thermo Fisher Scientific) in conjunction with SEQUEST. Peak lists were searched against an NCBI-Chinese hamster ovary (*Cricetulus griseus*) protein database (fasta file downloaded 29th November 2017 containing 24,906 sequences). Parameters were set as follows: MS1 tolerance of 10 ppm; MS2 mass tolerance of 0.6 Da for ion trap detection; enzyme specificity was set as trypsin with two missed cleavages allowed; carbamidomethylation of cysteine was set as a fixed modification; and phosphorylation of serine, threonine, and tyrosine (for phosphopeptide analysis) and oxidation of methionine were set as variable modifications. For phosphosite identification, the Phospho*RS* algorithm (Taus et al. 2011) was run through ProteomeDiscover 2.1 using diagnostic fragment ions and analyzer-specific fragment ion tolerances, as described above. Data were filtered to a 1% false discovery rate (FDR) on PSMs using automatic decoy searching in SEQUEST and by applying a phosphosite probability score of 75% or greater for S, T or Y amino acids in Phospho*RS* (Taus et al. 2011). Once peptides and phosphopeptides were successfully identified, a statistical criteria

of ANOVA p-value ≤ 0.05 and a relative minimum fold change of ±1.5 between the experimental groups were set. Only peptides and phosphopeptides which passed these criteria were then deemed to be differentially expressed between the relevant experimental groups being analysed.

8.2.6   Gene Ontology (GO) Analysis

Identified differentially expressed proteins and phosphoproteins were assigned to mouse or human official Gene Symbol identifiers. These gene symbols were then imported into DAVID (https://david.ncifcrf.gov), ToppGene Suit (https://toppgene.cchmc.org) and Enrichr (http://amp.pharm.mssm.edu/Enrichr/) online gene ontology platforms for functional pathway analysis. To refine the number of pathways identified, an adjusted p-value (Benjamini-Hochberg) of ≤0.05 between experimental groups was used. For each phosphoprotein of interest, the phosphosite number was assigned from the corresponding human protein sequence match. The functional information of such phosphosites was obtained from phosphosite.org (www.phosphosite.org), an online resource for providing information and tools for the study of protein post-translational modifications. For phosphopeptides with no complementary human sequence match, the phosphosite number was assigned through CHO-K1 protein database (www.chogenome.org).

8.3 Results

8.3.1 Proteomic and phosphoproteomic analysis of the growth of CHO cells in batch culture

In this comparative proteomic and phosphoproteomic study of the IgG producing CHO DP12 cells, we aimed to identify differentially expressed proteins and phosphoproteins as the cells transition through the exponential, stationary and decline phases of batch growth in suspension-cultured CHO cells. The cells were harvested on Day 2 and then after every 24 hours until Day 5, and are referred to as conditions Day 2, Day 3, Day 4 and Day 5. These time points were selected based on our understanding of the behaviour of CHO DP12 cells in serum-free media in suspension batch culture. During this short culture duration, selected time points capture the culture transition from early to late exponential phase (Day 2 to Day 3), through to the stationary phase (Day 3 to Day 4) and ultimately to the decline phase of the cell culture (Day 4 to Day 5). Figure 8-1A shows the growth and culture viability of the CHO DP12 cells used for this study.

For each condition or day, 2 mg of protein was tryptically digested. 10 µg from each digested sample was analysed as the whole cell lysate non-enriched sample and the remainder was used for phosphopeptide enrichment. Each sample was analysed by LC-MS/MS. Raw MS files generated from the Orbitrap Fusion Tribrid mass spectrometer were interrogated using Progenesis QI for Proteomics in order to identify statistically significant differentially expressed proteins and phosphoproteins over the culture duration. From this LC-MS data, we identified approximately 4,900 proteins in total from the 2 hour LC-MS runs from the non-enriched whole cell lysate samples and approximately 7,200 phosphopeptides in total from the phosphopeptide-enriched fractions.

To best understand the progressive changes in the proteome and the phosphoproteome of the CHO DP12 cells over time in batch suspension culture, LC-MS/MS data files from the three replicate samples from two consecutive day samples were compared with each other to obtain an account of differentially expressed peptides and phosphopeptides that may have changed in abundance during the transition from each day of culture to the next, i.e., Day 2 vs. Day 3, Day 3 vs. Day 4, and Day 4 vs. Day 5. Significantly increased or decreased expression of

phosphopeptides between samples was determined by using a fold-change cut-off of ±1.5 fold and a p-value ≤0.05 between experimental groups (Henry et al. 2017). The phosphorylation site in the peptide was localised through phospho*RS* 3.1 in ProteomeDiscoverer 2.1 after applying a phosphosite probability cut-off score of ≥75 (Taus et al. 2011). Peptides from the non-enriched whole cell lysate samples with fold changes ±1.5 and a p-value ≤0.05 between experimental groups were also deemed to be differentially expressed. Supplementary Table 1 provides an overview of the numbers of differentially expressed peptides and phosphopeptides and their corresponding proteins over time in culture (https://onlinelibrary.wiley.com/doi/full/10.1002/biot.201700221). Unsupervised Euclidian clustering analysis of the differentially expressed phosphopeptides and non-enriched peptides show a clear clustering of each sample set into the different days of the batch suspension culture, as shown in Figure 8-1B, i.e. Day 2, Day 3, Day 4, and Day 5 of the cell culture. There is some variation between the samples from each time point; however, the Unsupervised Euclidian clustering of the phosphoproteomic and total proteomic data clustered the three biological replicates from each time point into distinct sample groups. This relationship can be confirmed from a Principal Component Analysis (PCA) plot of the phosphoproteomic and proteomic data, as shown in supplementary data (https://onlinelibrary.wiley.com/doi/full/10.1002/biot.201700221).

In this analysis, 1867 phosphopeptides were found to be differentially expressed between Day 2 and Day 3 of culture as the cells transition from early through to late exponential phase of growth, with 923 phosphopeptides showing increased expression and 944 phosphopeptides showing a decrease in expression at Day 3 of culture compared to Day 2. In the comparison between Day 3 and Day 4, 1096 phosphopeptides were differentially expressed, with 370 phosphopeptides showing increased expression at Day 4 and 726 phosphopeptides showing decreased expression at Day 4 of culture as the cells transition from the exponential to the stationary phase. Similarly, differential analysis of Day 4 vs Day 5 of the culture identified 1521 differentially expressed phosphopeptides with 407 phosphopeptides showing increased expression at Day 5 and 1114 phosphopeptides showing decreased expression at Day 5 as the cells transition into the decline/death phase of growth. A summary of the total number of differentially expressed phosphopeptides and peptides and their corresponding proteins is shown in Figure 8-2A. Figure 8-2B shows the overlap between the phosphoproteomic and proteomic differentially expressed proteins for each comparative analysis. The full lists of

differentially expressed phosphopeptides and their corresponding phosphoproteins are shown in supplementary data (https://onlinelibrary.wiley.com/doi/full/10.1002/biot.201700221). The full lists of differentially expressed non-enriched peptides and the corresponding non-enriched proteins from the whole cell lysate (WCL) samples are shown in supplementary data (https://onlinelibrary.wiley.com/doi/full/10.1002/biot.201700221).

A number of phosphopeptides were differentially expressed between multiple comparisons; for example, 379 phosphopeptides were common between the Day 2 vs. Day 3 and Day 3 vs. Day 4 comparisons, and 428 phosphopeptides were commonly identified between and Day 3 vs. Day 4 and Day 4 vs. Day 5 comparative analyses. Interestingly, 111 phosphopeptides corresponding to 97 phosphoproteins were common between all comparison sets. Such phosphopeptides represent the dynamic nature of the phosphoproteome and give an account of potential acute changes in phosphorylation in response to cell growth. A change diagram of these constitutively differentially phosphopeptides is shown in Figure 8-3, showing the phosphopeptides that generally increase in expression over time in culture as Cluster 1 and phosphopeptides that generally decrease over time in culture as Cluster 2. A list of phosphopeptides that were observed to be differentially expressed at all phases of growth is included in supplementary Table 3 (https://onlinelibrary.wiley.com/doi/full/10.1002/biot.201700221) and includes reference to whether the phosphopeptides belong to Cluster 1 or Cluster 2.

8.3.2 Overlap of proteomic and phosphoproteomic datasets

A significant limitation in phosphoproteomic studies is the ambiguity of total peptide stoichiometry concerning its phosphorylated isoform. As a result one of the major challenges in phosphoproteomics is that abundance differences of a phosphopeptide could arise from (1) differential phosphorylation or (2) the differential overall expression of the total phosphoprotein. A comparison between the relative abundance pattern of a phosphopeptide and the non-phosphorylated version of the same peptide, if identified, can provide evidence for differential phosphorylation/dephosphorylation rather than just a change in abundance of a phosphopeptide and hence the phosphoprotein itself. With rapidly improving capabilities of high resolution and highly sensitive mass spectrometers such as the Orbitrap Fusion^{TM} Tribrid^{TM}

mass spectrometer used in this study, a good overlap between phosphopeptide identification and total proteome identification is expected. Supplementary Table 4 (https://onlinelibrary.wiley.com/doi/full/10.1002/biot.201700221) shows the list of all differentially expressed phosphopeptides (i.e. 532 in total) for which the corresponding peptide was identified in the non-enriched whole cell lysate, but not differentially expressed, suggesting altered phosphorylation of these proteins over the growth phases in batch culture. From this list of 532 phosphopeptides, three examples are shown in Figure 8-4 to demonstrate differential expression of a phosphopeptide with no change in the corresponding peptide from the total protein. For example the phosphopeptide S*GEGEVSGLMR, with identified phosphosite corresponding to Ser473 of human Transcription intermediary factor 1-beta (TRIM28), was found to be differentially expressed at all phases of growth, whereas the non-phosphorylated version of the peptide did not change over time in culture (see Figure 8-4A). The phosphopeptide abundance varies significantly across the growth curve with relative abundance firstly decreasing 1.9 fold from day 2 to day 3 and then gradually rising from day 3 until the end of the culture duration. However, no significant fold changes were observed from the analysis of the non-phosphorylated version of the peptide over time in culture. It has previously been shown that depletion of TRIM28 using shRNAi leads to increased cellular proliferation (Chen et al. 2012) and phosphorylation of TRIM28 on Ser473 by MAP kinase-activated protein kinase 2 (MK2) has been shown to reduce the activity of TRIM28 (King 2013). These previous studies tie in with the phosphoproteomic data shown here with the initial decrease in phosphorylation during the exponential phase of growth (day 2 to day 3) and then an increase in phosphorylation as the growth rate decreases and eventually the culture entering the decline phase by day 5. Similarly, the phosphopeptide QAGPS*PEAELR of protein Pleckstrin-like domain-containing family G member 6 (PLEKHG6), which is a guanine nucleotide exchange factor that can activate RhoG and RhoA and plays a role in cytokinesis (Samson et al. 2010, Wu et al. 2006), was identified to be differentially expressed at all phases of growth, initially showing a decrease in expression by 1.8 fold between day 2 and day 3, and then increasing in expression 1.8 fold from day 3 to day 4 and continuing to increase (1.7 fold) from day 4 to day 5. However, the abundance of the non-phosphorylated peptide QAGPSPEAELR identified from the non-enriched whole cell lysate did not change significantly over the culture duration (see Figure 8-4B). Similarly, phosphopeptide VSHYIINSSGPRPPVPPSPAQPPPGVSPS*R (Ser85) identified from Proto-oncogene c-CRK (also known as Adaptor molecule CRK), which

belongs to a family of adaptor proteins that act as major convergence points of tyrosine kinase signaling pathways(Braiman and Isakov 2015), shows a 2 fold decrease from day 2 to day 3, a 1.8 fold decrease from day 3 to day 4, and a 1.9 fold decrease from day 4 to day 5 of the CHO cell culture, whereas the non-phosphorylated version of the peptide shows no significant change in abundance over time in culture (see Figure 8-4C).

8.3.3   Functional classification of significantly differentially expressed proteins and phosphoproteins

All differentially expressed proteins were assigned human official gene identifiers for functional classification. Functional annotation of these genes was performed using DAVID (https://david.ncifcrf.gov) and Toppgene suite for functional analysis (https://toppgene.cchmc.org). For refined identification of statistically significant enrichment terms, an adjusted p-value (Benjamini Hochberg) of ≤0.05 was used. The most significant of the GO: Biological Process terms identified from the phosphoproteomic study were related to cell cycle, RNA processing, regulation of organelle organization, cytoskeleton organization, chromatin organization, histone modifications, etc. The molecular function of these proteins indicated that the differentially expressed phosphoproteins were of importance in, for example, transcription factor and transcription cofactor activity, transcription factor binding, GTPase regulator activity and kinase binding. (See supplementary Table 5 for the top 20 most-enriched GO: Biological Processes and GO: Molecular Functions from the phosphoproteomic and the non-enriched whole cell lysate proteomic analyses (https://onlinelibrary.wiley.com/doi/full/10.1002/biot.201700221)). We also identified the differential expression of 94 kinases along the growth curve, as shown in supplementary Table 6 (https://onlinelibrary.wiley.com/doi/full/10.1002/biot.201700221), which is a remarkable number given that the human kinome contains close to 518 kinases (Duong-Ly and Peterson 2013).

From KEGG pathway analysis we observed the enrichment of a number of pathways during the various phases of growth in batch culture from analysis of the lists of differentially expressed phosphoproteins. For example, there was a significant enrichment of growth-related signalling pathways such as ErB signalling, insulin signalling and cancer-related pathways. Table 8-1 lists

the KEGG pathways that were identified to be active at various phases of cell growth. Two of the pathways that were found to be highly enriched in the exponential and stationary phases were the mTOR and autophagy pathways. The mTOR pathway has been manipulated previously in CHO cells to improve production efficiencies (Josse et al. 2016, Dadehbeigi and Dickson 2015a, Dreesen and Fussenegger 2011a), and recently autophagy in CHO cells have gained attention as an opportunity for controlled cell death in bioprocess (Baek et al. 2016). In our study, we identified 18 phosphoproteins from the mTOR pathway, including the components of both mTORC1 and mTORC2 complexes. We also observed the differential expression of 17 autophagy-related phosphoproteins, predominantly at the exponential phase of cell culture; Supplementary Table 7 provides the complete list of differentially expressed phosphopeptides and associated phosphosites identified from the mTOR and autophagy pathways (https://onlinelibrary.wiley.com/doi/full/10.1002/biot.201700221). Supplementary Figure 2 shows KEGG pathways highlighting the differentially expressed phosphoproteins from the mTOR and autophagy pathways identified in this study (https://onlinelibrary.wiley.com/doi/full/10.1002/biot.201700221). Three of the phosphosites identified from WIPI2 and ATG2A from the autophagy pathway are potentially CHO-specific as no equivalent phosphosites were found in human sequences.

KEGG pathway analysis of the non-enriched whole cell lysate differentially expressed proteins showed a different set of enriched pathways and included an enrichment of proteins related to metabolism such as the Citrate (TCA) cycle, carbon metabolism, amino acid synthesis and degradation (see Table 8-1). This again demonstrates the extra depth of information achieved when combining both the proteomic and phosphoproteomic datasets.
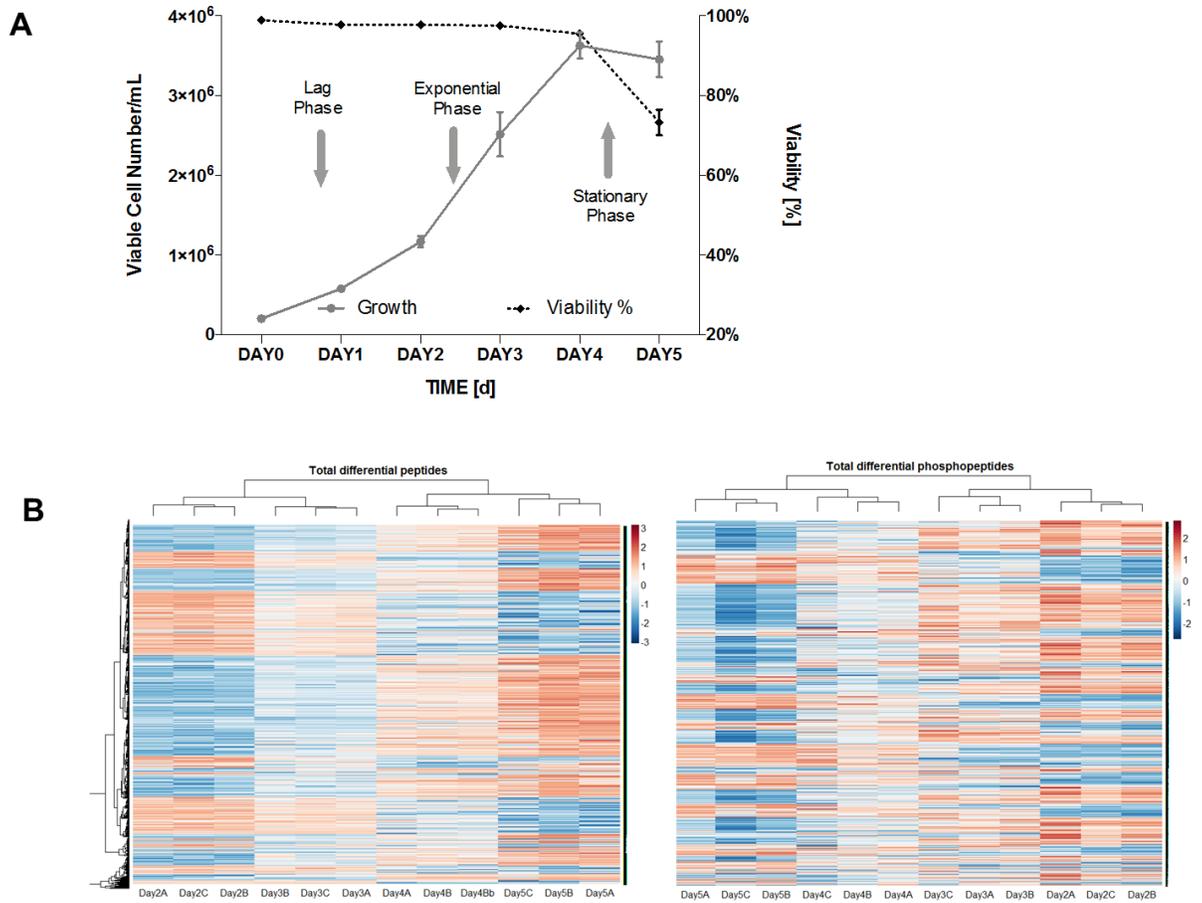
Figure 8-1. Growth curve of CHO DP12 cell line and heat maps of differentially expressed peptides and phosphopeptides over time in culture. (A) Cell number and culture viability (%) over time in serum-free suspension batch culture. Samples for proteomic and phosphoproteomic analysis were captured on day 2 (early exponential), day 3 (exponential), day 4 (stationary) and day 5 (early death) phases of growth. **(B)** Unsupervised Euclidian clustering shows that the expression of the differentially expressed non-enriched peptides (left) and phosphopeptides (right) identified from the experimental samples separate into four distinct sample groups i.e. Day 2, Day 3, Day 4, and Day 5 of the cell culture. Three replicate samples (labelled A-C) are shown for each time point.

A

Comparison	Phosphopeptides	Phosphoproteins	Non-enriched peptides	Non-enriched proteins
DAY 2 vs DAY3 (Early to late exponential)	1867 923 ↑ Day3 944 ↓ Day3	835 440 ↑ Day3 395 ↓ Day3	1332 418 ↑ Day3 914 ↓ Day3	311 219 ↑ Day3 92 ↓ Day3
DAY3 vs DAY4 (Exponential to Stationary)	1096 370 ↑ Day4 726 ↓ Day4	585 213 ↑ Day4 372 ↓ Day4	1873 1075 ↑ Day4 798 ↓ Day4	432 214 ↑ Day4 218 ↓ Day4
DAY4 vs DAY5 (Stationary to Decline)	1521 407 ↑ Day5 1114 ↓ Day5	754 214 ↑ Day5 540 ↓ Day5	1465 934 ↑ Day5 531 ↓ Day5	367 209 ↑ Day5 158 ↓ Day5

B
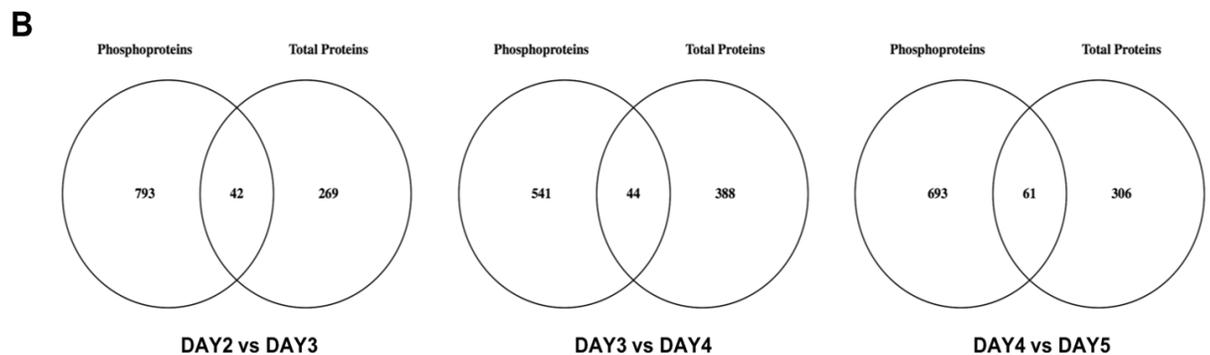
Figure 8-2. Overview of the total number of differentially expressed phosphopeptides, and peptides for each comparative analysis. (A) Summary of differentially expressed peptides, phosphopeptides and their corresponding proteins, and their direction of fold change. (B) Venn diagrams showing the overlap between the proteomic and phosphoproteomic protein identifications between the three comparative analyses. Peptides and phosphopeptides and their corresponding proteins were deemed differentially expressed if the fold change between two conditions was ±1.5 with an ANOVA p-value ≤ 0.05 between experimental groups.
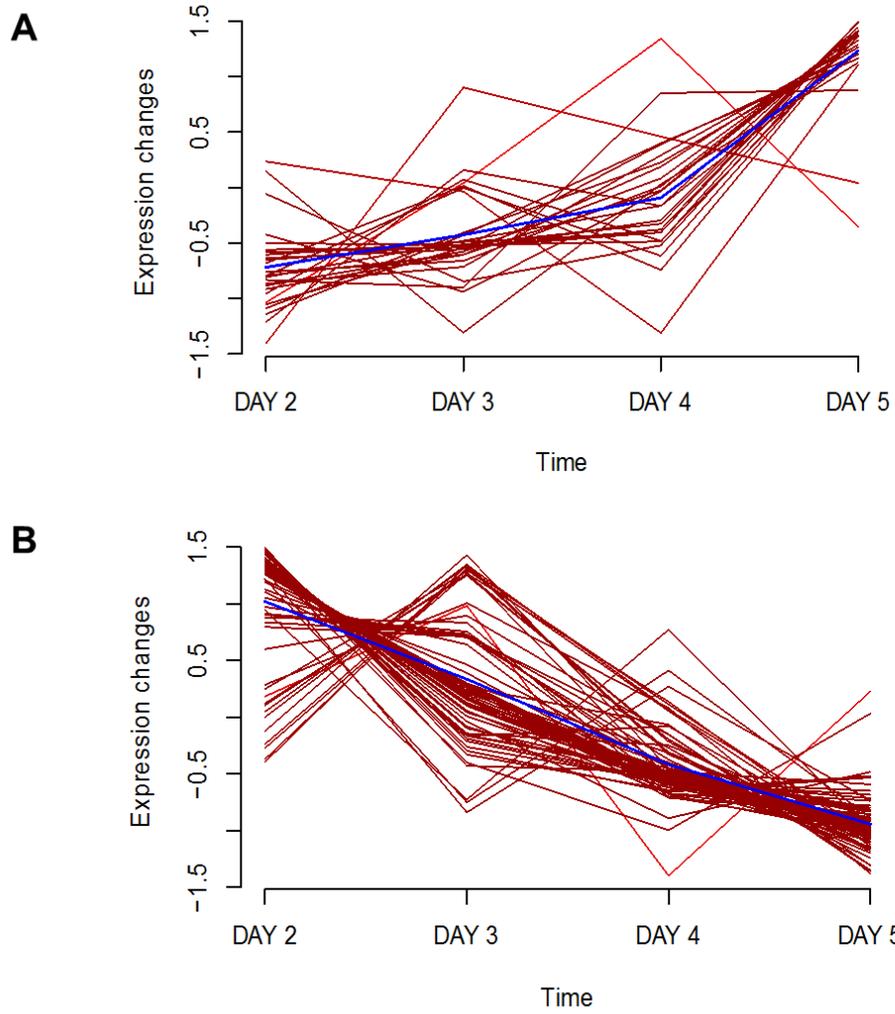
Figure 8-3. A change diagram of phosphopeptides shown to be differentially expressed at all growth phases. Unsupervised soft clustering of phosphopeptides revealed two different cluster structures, (A) Cluster 1 and (B) Cluster 2. Log10 transformed phosphopeptide abundance values on the y-axis are plotted against the sample collection time points on the x-axis.
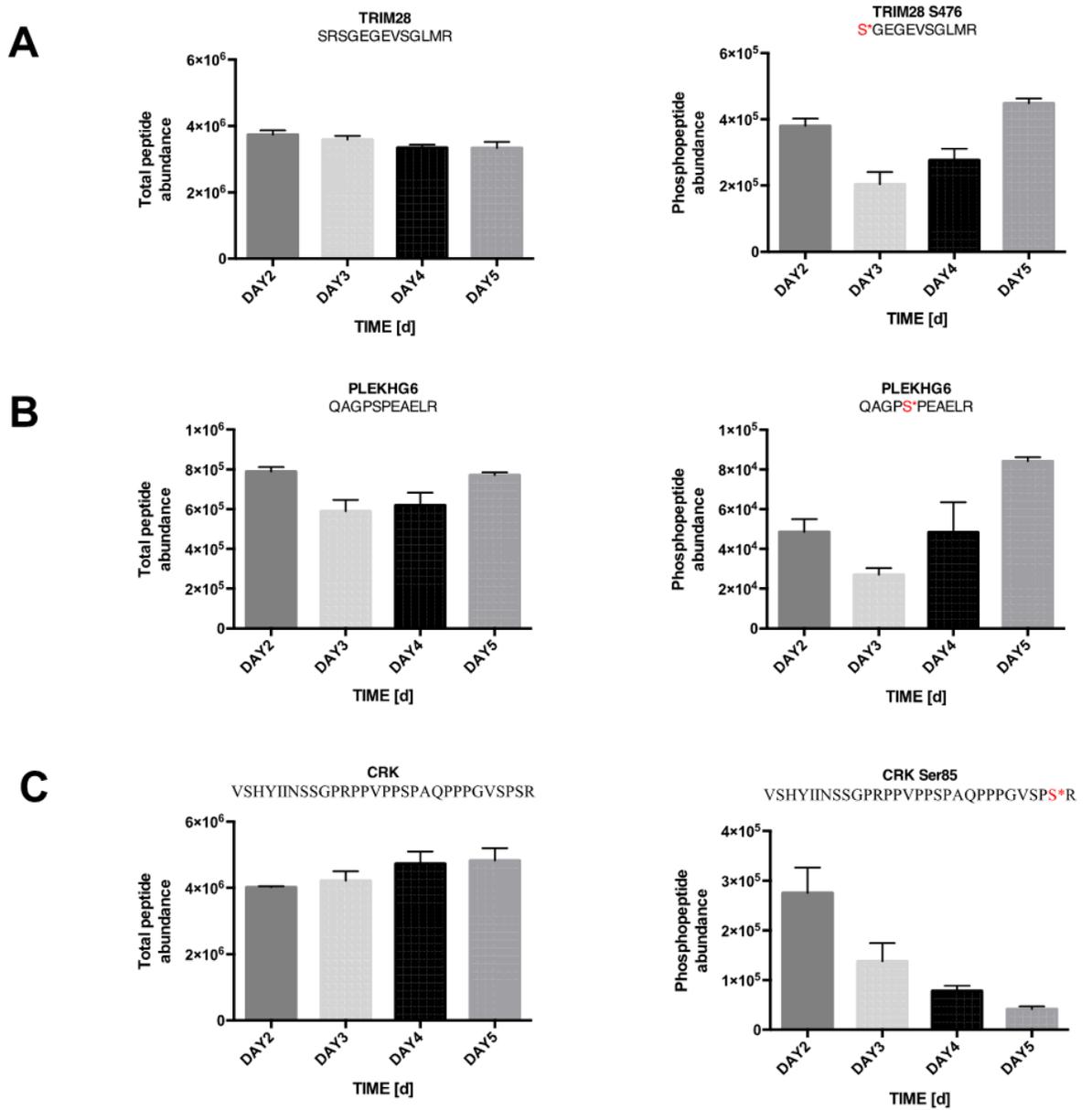
Figure 8-4. Altered phosphorylation of phosphoproteins during batch suspension culture. Expression patterns of peptides and phosphopeptides from (A) TRIM28, (B) PLEKHG6, and (C) CRK over time in culture. Left panel shows the change in expression of the non-phosphorylated peptide identified from the non-enriched whole cell lysate over time in culture. Right panel show changes in expression of the phosphorylated version of the same peptide over time in culture. X-axis, days of cell culture; Y-axis, peptide/phosphopeptide abundance from MS data.

288

8.4   Discussion

The data in this study shows significant changes in the proteome and phosphoproteome during growth of CHO cells in suspension batch culture. The entire study identified 3,777 differentially expressed CHO phosphopeptides corresponding to 1,415 differentially expressed phosphoproteins and 834 differentially expressed proteins over the culture duration. We also identified 94 differentially expressed kinases in the study. The regulation of phosphorylation by kinases and phosphatases allows cells to respond to internal and external signaling events on the fly (Caenepeel et al. 2004).

From the gene ontology analysis of the phosphoproteomic datasets we identified a significant enrichment of proteins related to transcription factor activity; this enrichment was not observed in the whole cell lysate proteomic data. Phosphorylation is known to target gene transcription by affecting the stability, location and structure of transcription factors (Vogel et al. 2014). Interestingly gene ontology analysis of the whole cell lysate samples showed an enrichment of molecular functions primarily associated with metabolic pathways such as the TCA cycle, carbon metabolism, amino acid metabolism, etc. as the culture progresses over time. An improved knowledge of such metabolic pathways in CHO cells which can lead to the production and accumulation of growth-inhibiting and toxic metabolites will lead to improved growth and productivity of CHO cells (Pereira, Kildegaard and Andersen 2018).

One of the pathways that were found to be highly enriched in this study was the mTOR pathway. This pathway is of high interest in CHO bioprocessing where a number of studies have involved manipulation of the mTOR pathway to improve CHO cell growth and productivity (Josse et al. 2016, Dadehbeigi and Dickson 2015b, Dreesen and Fussenegger 2011b). In this study, 18 phosphoproteins from the mTOR pathway were identified, including components of both mTORC1 and mTORC2 complexes as outlined in supplementary Table 7 (https://onlinelibrary.wiley.com/doi/full/10.1002/biot.201700221). We identified the peptide TDS*YSAGQSVEILDGVELGEPAHK with a phosphorylated serine corresponding to Ser2248 (human) of the mTOR protein. The phosphorylation of Ser2248 on mTOR by p70S6K has been shown to be responsible for nuclear translocation on the mTORC1 complex (Leal et al. 2013). In this study we observed that total mTOR protein levels followed an upward trend from day 2 until day 5 (fold change >7) of culture, however the expression of phosphopeptide

TDS*YSAGQSVEILDGVELGEPAHK (Ser2248) from mTOR shows only a 1.5 fold increase from Day 2 to Day 3 of the cell culture and thereafter following a downward trend up until day 5.

We also identified differentially expressed phosphopeptides related to two crucial downstream substrates 4E-BP1 and ULK1 of the mTORC1 complex. 4E-BP1 is a well-known substrate of the mTORC1 complex (Hara et al. 1997) and the phosphorylation status of 4E-BP1 has been used as a marker of mTOR activity (Qin, Jiang and Zhang 2016). We identified three functionally relevant (human) phosphosites i.e. Thr41, Ser65 and Thr70 on protein 4E-BP1. Activated mTORC1 hyperphosphorylates 4E-BP1 on multiple sites including Ser65 and Thr70 which lead to a release of eIF4E from 4E-BP1/eIF4E complex and initiation of cap-dependent translation (Gingras et al. 2001). We observed a 2 fold upregulation of a phosphopeptide (Thr70) from day 2 to day 3 of culture and a 4.5 fold decrease in the expression of a phosphopeptide containing the equivalent of Ser65 from day 3 to day 4 of culture. While the potential increase in phosphorylation of Thr70 indicates increased mTOR activity and a resultant increase in cap-dependent translation, the dephosphorylation of Ser65 from day 3 to day 4 indicates a potential decrease in mTOR activity. These observations potentially show that the highest mTOR activity is registered on day 3 of the batch culture, which correlates with the phosphopeptide containing Ser2248 on mTOR being the most abundant at day 3. Increased phosphorylation of Ser2248 has been shown to be observed upon treatment of adult cardiac muscle cells with mTOR agonists such as endothelin-1 (ET-1), phenylephrine (PE), TPA, or insulin (Moschella et al. 2007), and mTOR activity has been shown to be negatively regulated by the dephosphorylation of Ser2248 triggered by overexpression in cultured adipocytes (Guntur et al. 2010). The increased activity of mTOR at day 3 of cell culture is also indicated by the higher relative abundance of the phosphopeptide ILDTSSLTQSAPAS*PTNK (Ser863) at day 3 as Ser863 on RAPTOR has been shown to be required for positive regulation of mTORC1 activity (Foster et al. 2010). RAPTOR functions as a scaffold for mTORC1 complex substrates (Foster et al. 2010). Like RAPTOR, PRAS40 and other members of the mTORC1 complex can also regulate mTOR activity (Yuan and Guan 2016). The phosphorylation of Thr246 of PRAS40 by PI3K has been shown to induce dissociation of PRAS40 from mTOR thus inhibiting mTOR activity (Wiza, Nascimento and Ouwens 2012). We found that the phosphopeptide LNT*SDFQK containing phosphosite Thr246 showed a 5.3 fold decrease from day 2 to day 5 of culture. Similar changes in abundance patterns were observed for phosphopeptide SSDEENGPPS*SPDLDR (Ser211) and SSDEENGPPSS*PDLDR (Ser212) of PRAS40. As a central

pathway for cell proliferation and protein synthesis, the mTOR pathway has been manipulated in CHO cells to understand and improve bioprocess attributes. The phosphoproteomic data shows a much deeper insight into the complexity of the mTOR pathway and the potential importance of post-translational regulation by phosphorylation of the mTOR pathway in recombinant CHO cells that would not be observed in proteomic studies.

Pathway analysis of phosphoproteomic data also revealed a significant enrichment of autophagy-related proteins with differential expression between day 2 and day 4 of the batch suspension culture. Recently autophagy has gained importance in bioprocess as an opportunity to genetically and chemically control cell survival and productivity (Baek et al. 2016, Nasseri et al. 2014, Y. J. Kim et al. 2013, Han et al. 2011). It is therefore important to have a detailed understanding of the autophagic pathway in order to select relevant genes and chemical targets to manipulate autophagy in CHO cells to enhance the production of biotherapeutics (Y. J. Kim et al. 2013). To increase culture longevity and increase protein production, numerous attempts have been made to delay apoptosis of CHO cells (J. G. Tan et al. 2015, Druz et al. 2013), however little is understood regarding the regulation of autophagy in CHO cells in bioprocess-relevant conditions. In our study we have observed the differential expression of 17 autophagy-related phosphoproteins, predominantly at the exponential phase of cell culture. For example, the phosphorylation status of the ULK1–ATG13–FIP200–ATG101 is crucial for autophagy initiation (Ganley et al. 2009). In this study we identified differentially expressed phosphopeptides from ULK1 and ATG13. Studies have shown ULK1 and ATG13 as direct substrates of the mTORC1 complex and the inhibition of mTORC1 by rapamycin of starvation conditions leads to dephosphorylation of these proteins (J. Kim et al. 2011, Kamada et al. 2010). We found a large 50-fold decrease in abundance of the phosphopeptide AS*PHDVLETIFVR (Ser361) between day 3 and day 4 of the culture. This protein was not found to be differentially expressed in the non-enriched proteomic analysis. For the onset of autophagy, ATG13 is phosphorylated by ULK1 (Alers et al. 2012). Here, we identified 6 phosphopeptides from ULK1 protein from which phosphopeptides containing the identified phosphosites Ser556 on LHS*APNLSDFHVVRPK and Thr670 on NRT*LPDLSEAGPFQGQQLGSGLRPAEDTR show differential expression between day 2 and day 3 of culture with a net decrease in phosphopeptide abundance of 34 and 11 fold respectively. The phosphopeptide SGSTS*PLGFAR containing Ser469 showed a 31-fold decrease from day 3 to day 4 of culture. Three other phosphopeptides containing the phosphosites Ser714

(AAFGTQASDS*GSTDSLQEKPMEIAPSAGFGGTLHPGAR), Thr452 (IEQNLQSPT*QHQTAR), and Ser477 (AS*PSPPSHTDGAMLAR) were found to show decreased expression from day 2 to day 5 of culture.

Several other phosphoproteins with potential implication in recombinant protein productivity and product quality were identified to be differentially expressed between the growth phases. For example, we found four differentially expressed phosphopeptides from the intermediate filament protein, vimentin, during culture progression. Previous proteomic studies have shown that increased expression of vimentin is associated with high specific productivity (Carlage et al. 2009, Kumar et al. 2008, Baik et al. 2006). We identified four phosphopeptides with phosphosites Ser55 (SLYSS*SPGGAYVTR), Ser56 (SLYSSS*PGGAYVTR), Ser459 (DGQVINETS*QHHDDLE), Ser412 (LLEGEESRIS*LPLPNFSSLNLR) and Ser325 (QVQS*LTCEVDALK) of vimentin to show decreased expression over the culture duration. We also found differential expression of three phosphopeptides from signal recognition particle subunit (SRP72), an essential protein for targeting secreted protein to the rough endoplasmic reticulum (Becker et al. 2017). SUMOylation, the covalent attachment of SUMO (small ubiquitin-like modifier) of a protein product, has shown to increase product quality by improving folding, solubility, and stability of the recombinant protein (Peroutka et al. 2008). We identified differential expression of a number of phosphoproteins associated with the SUMOylation pathway including SENP1 and NUP153. In high producer cells, the secretory pathway remains a bottleneck (Le Fourn et al. 2014) . A recent study showed a dramatic increase in recombinant protein productivity by knockdown of two secretory pathway proteins, i.e. Ceramide Synthase 2 (CERS2) and Tbc1D20 (Le Fourn et al. 2014). We identified a 2.7 fold increase in expression of a phosphopeptide expression from CERS2 from day 2 to day 3 of culture. We also identified differentially expressed phosphopeptides from the heat shock proteins, Hspa5 and Hap90b1; the expression of these two proteins has been previously linked with increased protein productivity (Dorner, Wasley and Kaufman 1989).

As the cellular mechanisms controlling CHO cell growth and productivity are intertwined, cyclin-dependent kinases (CDKs) have been shown to be prime targets of interest to control CHO cell growth and improve productivity (Kumar, Gammell and Clynes 2007). In our study we identified differential expression of eight CDKs including Cdk6, Cdk5rap2, Cdk2, Cdk17, Cdk16, Cdk14 Cdk12, Cdk1 and Cdk6. Interestingly a previous study has demonstrated that

manipulation of signalling pathways using small molecule inhibitors that can directly target CDK 4/6 to arrest cell growth after a few days of culture has been found to improve productivity and quality of a Mab (Du et al. 2015a). This suggests the potential to manipulate expression of such target phosphoproteins identified to be differentially expressed during the growth phases in our study, either through genetic engineering or chemical inhibitor approaches, to control cell growth and potentially improve productivity.

8.5   Conclusion

The data in this study clearly demonstrates that studying the proteome and phosphoproteome adds additional dimensions to the understanding of recombinant CHO cells in bioprocess conditions. These phosphoproteins and the kinases and phosphatases that are involved in their regulation have the potential to be engineering targets in CHO cells. Kinases and phosphatases also have the potential to be chemically manipulated, for example to control cell growth and increase Mab productivity (Du et al. 2015b). From a perspective of future CHO cell signaling research, the identified kinases are of particular interest as to fully understand the signaling dynamics of CHO cells grown in bioprocess-relevant conditions. It is equally important to understand the kinase-substrate relationship to pinpoint the responsible kinases for observed differences in phosphosite abundance and provide the opportunity to manipulate CHO cell behaviour by regulation of endogenous kinases.

Future work will involve quantification of phosphorylation changes using mass spectrometry based approaches such as Parallel Reaction Monitoring (Abelin et al. 2016) due to the lack of availability of CHO-specific anti-phospho antibodies. Genome editing techniques such as CRISPR/Cas9 are also ideal tools to study the phosphoproteome in further detail and will allow the assignment of functionality to specific phosphosites of interest through targeted site-specific mutagenesis approaches (Lee et al. 2015).

Acknowledgements

8.6   References

Abelin, J.G., Patel, J., Lu, X., Feeney, C.M., Fagbami, L., Creech, A.L., Hu, R., Lam, D., Davison, D., Pino, L., Qiao, J.W., Kuhn, E., Officer, A., Li, J., Abbatiello, S., Subramanian, A., Sidman, R., Snyder, E., Carr, S.A. and Jaffe, J.D. 2016. Reduced-representation phosphosignatures measured by quantitative targeted MS capture cellular states and enable large-scale comparison of drug-induced phenotypes. *Molecular & Cellular Proteomics: MCP,* 15(5), pp.1622-1641.

Alers, S., Loffler, A.S., Wesselborg, S. and Stork, B. 2012. Role of AMPK-mTOR-Ulk1/2 in the regulation of autophagy: Cross talk, shortcuts, and feedbacks. *Molecular and Cellular Biology,* 32(1), pp.2-11.

Angel, T.E., Aryal, U.K., Hengel, S.M., Baker, E.S., Kelly, R.T., Robinson, E.W. and Smith, R.D. 2012. Mass spectrometry-based proteomics: Existing capabilities and future directions. *Chemical Society Reviews,* 41(10), pp.3912-3928.

Ardito, F., Giuliani, M., Perrone, D., Troiano, G. and Lo Muzio, L. 2017. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy. *International Journal of Molecular Medicine,* 40(2), pp.271-280.

Baek, E., Kim, C.L., Kim, M.G., Lee, J.S. and Lee, G.M. 2016. Chemical inhibition of autophagy: Examining its potential to increase the specific productivity of recombinant CHO cell lines. *Biotechnology and Bioengineering,* 113(9), pp.1953-1961.

Baik, J.Y., Lee, M.S., An, S.R., Yoon, S.K., Joo, E.J., Kim, Y.H., Park, H.W. and Lee, G.M. 2006. Initial transcriptome and proteome analyses of low culture temperature-induced expression in CHO cells producing erythropoietin. *Biotechnology and Bioengineering,* 93(2), pp.361-371.

Becker, M.M., Lapouge, K., Segnitz, B., Wild, K. and Sinning, I. 2017. Structures of human SRP72 complexes provide insights into SRP RNA remodeling and ribosome interaction. *Nucleic Acids Research,* 45(1), pp.470-481.

Bort, J.A.H., Hackl, M., Höflmayer, H., Jadhav, V., Harreither, E., Kumar, N., Ernst, W., Grillari, J. and Borth, N. 2012. Dynamic mRNA and miRNA profiling of CHO-K1 suspension cell cultures. *Biotechnology Journal,* 7(4), pp.500-515.

Braiman, A. and Isakov, N. 2015. The role of crk adaptor proteins in T-cell adhesion and migration. *Frontiers in Immunology,* 6pp.509.

Caenepeel, S., Charydczak, G., Sudarsanam, S., Hunter, T. and Manning, G. 2004. The mouse kinome: Discovery and comparative genomics of all mouse protein kinases. *Proceedings of the National Academy of Sciences of the United States of America,* 101(32), pp.11707-11712.

Carlage, T., Hincapie, M., Zang, L., Lyubarskaya, Y., Madden, H., Mhatre, R. and Hancock, W.S. 2009. Proteomic profiling of a high-producing chinese hamster ovary cell culture. *Analytical Chemistry,* 81(17), pp.7357-7362.

Chen, L., Chen, D.T., Kurtyka, C., Rawal, B., Fulp, W.J., Haura, E.B. and Cress, W.D. 2012. Tripartite motif containing 28 (Trim28) can regulate cell proliferation by bridging HDAC1/E2F interactions. *The Journal of Biological Chemistry,* 287(48), pp.40106-40118.

Choudhary, C. and Mann, M. 2010. Decoding signalling networks by mass spectrometry-based proteomics. *Nature Reviews.Molecular Cell Biology,* 11(6), pp.427-439.

Cohen, P. 2000. The regulation of protein function by multisite phosphorylation–a 25 year update. *Trends in Biochemical Sciences,* 25(12), pp.596-601.

Cohen, P. 2001. The role of protein phosphorylation in human health and disease. the sir hans krebs medal lecture. *European Journal of Biochemistry,* 268(19), pp.5001-5010.

Dadehbeigi, N. and Dickson, A.J. 2015. Chemical manipulation of the mTORC1 pathway in industrially relevant CHOK1 cells enhances production of therapeutic proteins. *Biotechnology Journal,* 10(7), pp.1041-1050.

Datta, P., Linhardt, R.J. and Sharfstein, S.T. 2013. An 'omics approach towards CHO cell engineering. *Biotechnology and Bioengineering,* 110(5), pp.1255-1271.

Dorner, A.J., Wasley, L.C. and Kaufman, R.J. 1989. Increased synthesis of secreted proteins induces expression of glucose-regulated proteins in butyrate-treated chinese hamster ovary cells. *The Journal of Biological Chemistry,* 264(34), pp.20602-20607.

Dreesen, I.A. and Fussenegger, M. 2011. Ectopic expression of human mTOR increases viability, robustness, cell size, proliferation, and antibody production of chinese hamster ovary cells. *Biotechnology and Bioengineering,* 108(4), pp.853-866.

Druz, A., Son, Y.J., Betenbaugh, M. and Shiloach, J. 2013. Stable inhibition of mmu-miR-466h-5p improves apoptosis resistance and protein production in CHO cells. *Metabolic Engineering,* 16pp.87-94.

Du, Z., Treiber, D., McCarter, J.D., Fomina-Yadlin, D., Saleem, R.A., McCoy, R.E., Zhang, Y., Tharmalingam, T., Leith, M., Follstad, B.D., Dell, B., Grisim, B., Zupke, C., Heath, C., Morris, A.E. and Reddy, P. 2015. Use of a small molecule cell cycle inhibitor to control cell growth and improve specific productivity and product quality of recombinant proteins in CHO cell cultures. *Biotechnology and Bioengineering,* 112(1), pp.141-155.

Duong-Ly, K.C. and Peterson, J.R. 2013. The human kinome and kinase inhibition. *Current Protocols in Pharmacology,* 60(1), pp.2.9. 1-2.9. 14.

Foster, K.G., Acosta-Jaquez, H.A., Romeo, Y., Ekim, B., Soliman, G.A., Carriere, A., Roux, P.P., Ballif, B.A. and Fingar, D.C. 2010. Regulation of mTOR complex 1 (mTORC1) by raptor Ser863 and multisite phosphorylation. *The Journal of Biological Chemistry,* 285(1), pp.80-94.

Ganley, I.G., Lam du, H., Wang, J., Ding, X., Chen, S. and Jiang, X. 2009. ULK1.ATG13.FIP200 complex mediates mTOR signaling and is essential for autophagy. *The Journal of Biological Chemistry,* 284(18), pp.12297-12305.

Gingras, A.C., Raught, B., Gygi, S.P., Niedzwiecka, A., Miron, M., Burley, S.K., Polakiewicz, R.D., Wyslouch-Cieszynska, A., Aebersold, R. and Sonenberg, N. 2001. Hierarchical phosphorylation of the translation inhibitor 4E-BP1. *Genes & Development,* 15(21), pp.2852-2864.

Gutierrez, J.M. and Lewis, N.E. 2015. Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. *Biotechnology Journal,* 10(7), pp.939-949.

Hacker, D.L., De Jesus, M. and Wurm, F.M. 2009. 25 years of recombinant proteins from reactor-grown cells - where do we go from here? *Biotechnology Advances,* 27(6), pp.1023-1027.

Han, Y.K., Ha, T.K., Lee, S.J., Lee, J.S. and Lee, G.M. 2011. Autophagy and apoptosis of recombinant chinese hamster ovary cells during fed-batch culture: Effect of nutrient supplementation. *Biotechnology and Bioengineering,* 108(9), pp.2182-2192.

Hara, K., Yonezawa, K., Kozlowski, M.T., Sugimoto, T., Andrabi, K., Weng, Q.P., Kasuga, M., Nishimoto, I. and Avruch, J. 1997. Regulation of eIF-4E BP1 phosphorylation by mTOR. *The Journal of Biological Chemistry,* 272(42), pp.26457-26463.

Huang, Y., Hu, W., Rustandi, E., Chang, K., Yusuf-Makagiansar, H. and Ryll, T. 2010. Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment. *Biotechnology Progress,* 26(5), pp.1400-1410.

Humphrey, S.J., James, D.E. and Mann, M. 2015. Protein phosphorylation: A major switch mechanism for metabolic regulation. *Trends in Endocrinology & Metabolism,* 26(12), pp.676-687.

Josse, L., Xie, J., Proud, C.G. and Smales, C.M. 2016. mTORC1 signalling and eIF4E/4E-BP1 translation initiation factor stoichiometry influence recombinant protein productivity from GS-CHOK1 cells. *The Biochemical Journal,* 473(24), pp.4651-4664.

Kamada, Y., Yoshino, K., Kondo, C., Kawamata, T., Oshiro, N., Yonezawa, K. and Ohsumi, Y. 2010. Tor directly controls the Atg1 kinase complex to regulate autophagy. *Molecular and Cellular Biology,* 30(4), pp.1049-1058.

Kildegaard, H.F., Baycin-Hizal, D., Lewis, N.E. and Betenbaugh, M.J. 2013. The emerging CHO systems biology era: Harnessing the 'omics revolution for biotechnology. *Current Opinion in Biotechnology,* 24(6), pp.1102-1107.

Kim, J., Kundu, M., Viollet, B. and Guan, K.L. 2011. AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1. *Nature Cell Biology,* 13(2), pp.132-141.

Kim, Y.J., Baek, E., Lee, J.S. and Lee, G.M. 2013. Autophagy and its implication in chinese hamster ovary cell culture. *Biotechnology Letters,* 35(11), pp.1753-1763.

King, C.A. 2013. Kaposi's sarcoma-associated herpesvirus kaposin B induces unique monophosphorylation of STAT3 at serine 727 and MK2-mediated inactivation of the STAT3 transcriptional repressor TRIM28. *Journal of Virology,* 87(15), pp.8779-8791.

Kumar, N., Gammell, P. and Clynes, M. 2007. Proliferation control strategies to improve productivity and survival during CHO based production culture : A summary of recent methods employed and the effects of proliferation control in product secreting CHO cell lines. *Cytotechnology,* 53(1-3), pp.33-46.

Kumar, N., Gammell, P., Meleady, P., Henry, M. and Clynes, M. 2008. Differential protein expression following low temperature culture of suspension CHO-K1 cells. *BMC Biotechnology,* 8pp.42-6750-8-42.

Leal, P., Garcia, P., Sandoval, A., Buchegger, K., Weber, H., Tapia, O. and Roa, J.C. 2013. AKT/mTOR substrate P70S6K is frequently phosphorylated in gallbladder cancer tissue and cell lines. *OncoTargets and Therapy,* 6pp.1373-1384.

Lee, J.S., Grav, L.M., Lewis, N.E. and Faustrup Kildegaard, H. 2015. CRISPR/Cas9-mediated genome engineering of CHO cell factories: Application and perspectives. *Biotechnology Journal,* 10(7), pp.979-994.

Meleady, P., Hoffrogge, R., Henry, M., Rupp, O., Bort, J.H., Clarke, C., Brinkrolf, K., Kelly, S., Muller, B., Doolan, P., Hackl, M., Beckmann, T.F., Noll, T., Grillari, J., Barron, N., Puhler, A., Clynes, M. and Borth, N. 2012. Utilization and evaluation of CHO-specific sequence databases for mass spectrometry based proteomics. *Biotechnology and Bioengineering,* 109(6), pp.1386-1394.

Moschella, P.C., Rao, V.U., McDermott, P.J. and Kuppuswamy, D. 2007. Regulation of mTOR and S6K1 activation by the nPKC isoforms, PKCepsilon and PKCdelta, in adult cardiac muscle cells. *Journal of Molecular and Cellular Cardiology,* 43(6), pp.754-766.

Nasseri, S.S., Ghaffari, N., Braasch, K., Jardon, M.A., Butler, M., Kennard, M., Gopaluni, B. and Piret, J.M. 2014. *Increased CHO cell fed-batch monoclonal antibody production using the autophagy inhibitor 3-MA or gradually increasing osmolality.*

Omasa, T., Onitsuka, M. and Kim, W. 2010. Cell engineering and cultivation of chinese hamster ovary (CHO) cells. *Current Pharmaceutical Biotechnology,* 11(3), pp.233-240.

Pereira, S., Kildegaard, H.F. and Andersen, M.R. 2018. Impact of CHO metabolism on cell growth and protein production: An overview of toxic and inhibiting metabolites and nutrients. *Biotechnology Journal,* 13(3), pp.1700499.

Peroutka, R.J., Elshourbagy, N., Piech, T. and Butt, T.R. 2008. Enhanced protein expression in mammalian cells using engineered SUMO fusions: Secreted phospholipase A2. *Protein Science : A Publication of the Protein Society,* 17(9), pp.1586-1595.

Puente, C., Hendrickson, R.C. and Jiang, X. 2016. Nutrient-regulated phosphorylation of ATG13 inhibits starvation-induced autophagy. *The Journal of Biological Chemistry,* 291(11), pp.6026-6035.

Qin, X., Jiang, B. and Zhang, Y. 2016. 4E-BP1, a multifactor regulated multifunctional protein. *Cell Cycle,* 15(6), pp.781-786.

Samson, T., Welch, C., Monaghan-Benson, E., Hahn, K.M. and Burridge, K. 2010. Endogenous RhoG is rapidly activated after epidermal growth factor stimulation through multiple guanine-nucleotide exchange factors. *Molecular Biology of the Cell,* 21(9), pp.1629-1642.

Savitski, M.M., Scholten, A., Sweetman, G., Mathieson, T. and Bantscheff, M. 2010. Evaluation of data analysis strategies for improved mass spectrometry-based phosphoproteomics. *Analytical Chemistry,* 82(23), pp.9843-9849.

Stein, S., Thomas, E.K., Herzog, B., Westfall, M.D., Rocheleau, J.V., Jackson, R.S.,2nd, Wang, M. and Liang, P. 2004. NDRG1 is necessary for p53-dependent apoptosis. *The Journal of Biological Chemistry,* 279(47), pp.48930-48940.

Stolfa, G., Smonskey, M.T., Boniface, R., Hachmann, A., Gulde, P., Joshi, A.D., Pierce, A.P., Jacobia, S.J. and Campbell, A. 2018. CHO-Omics review: The impact of current and emerging technologies on chinese hamster ovary based bioproduction. *Biotechnology Journal,* 13(3), pp.1700227.

Tan, J.G., Lee, Y.Y., Wang, T., Yap, M.G., Tan, T.W. and Ng, S.K. 2015. Heat shock protein 27 overexpression in CHO cells modulates apoptosis pathways and delays activation of caspases to improve recombinant monoclonal antibody titre in fed-batch bioreactors. *Biotechnology Journal,* 10(5), pp.790-800.

Taus, T., Kocher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C. and Mechtler, K. 2011. Universal and confident phosphorylation site localization using phosphoRS. *Journal of Proteome Research,* 10(12), pp.5354-5362.

Vogel, W.K., Gafken, P.R., Leid, M. and Filtz, T.M. 2014. Kinetic analysis of BCL11B multisite phosphorylation–dephosphorylation and coupled sumoylation in primary thymocytes by multiple reaction monitoring mass spectroscopy. *Journal of Proteome Research,* 13(12), pp.5860-5868.

Vyse, S., Desmond, H. and Huang, P.H. 2017. Advances in mass spectrometry based strategies to study receptor tyrosine kinases. *Iucrj,* 4(2), pp.119-130.

Walsh, G. 2014. Biopharmaceutical benchmarks 2014. *Nature Biotechnology,* 32(10), pp.992-1000.

Wang, Z., Lv, N., Bi, W., Zhang, J. and Ni, J. 2015. Development of the affinity materials for phosphorylated proteins/peptides enrichment in phosphoproteomics analysis. *ACS Applied Materials & Interfaces,* 7(16), pp.8377-8392.

Wiza, C., Nascimento, E.B. and Ouwens, D.M. 2012. Role of PRAS40 in akt and mTOR signaling in health and disease. *American Journal of Physiology.Endocrinology and Metabolism,* 302(12), pp.E1453-60.

Wu, D., Asiedu, M., Adelstein, R.S. and Wei, Q. 2006. A novel guanine nucleotide exchange factor MyoGEF is required for cytokinesis. *Cell Cycle (Georgetown, Tex.),* 5(11), pp.1234-1239.

Wurm, F.M. 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotechnology,* 22(11), pp.1393-1398.

Yan, G.R. and He, Q.Y. 2008. Functional proteomics to identify critical proteins in signal transduction pathways. *Amino Acids,* 35(2), pp.267-274.

Yuan, H.X. and Guan, K.L. 2016. Structural insights of mTOR complex 1. *Cell Research,* 26(3), pp.267-268.

Clonal Variation in Productivity and Proteolytic Clipping of an Fc-Fusion Protein in CHO Cells: Proteomic Analysis Suggests a Role for Defective Protein Folding and the UPR.

Authors: <u>Michael Henry</u>, Clair Gallagher, Ronan M. Kelly, Christopher C. Frye, Matthew D. Osborne, Ciaran P Brady, Niall Barron, Martin Clynes, Paula Meleady.

Within this work, Michael Henry specifically performed the methodology and verification of the protein digestions for label free differential analysis on the selected cell clones. Michael Henry implemented the differential expression analysis tools on the peptide samples using label free software Progensis QI and searched the statistically significant protein features using MS search algorithms. Michael Henry prepared presentation of the published work, specifically writing the initial draft.

Abstract

Product degradation, such as clipping, is a common quality issue in the production of Fc-fusion proteins from Chinese hamster ovary (CHO) cells. Degradation of proteins is mainly due to action of either intracellular or extracellular host cell proteases. This study was carried out to understand more fundamentally the intracellular events that may play a role in determining why cell lines from the same cell line development project can vary with regards to the extent of Fc-fusion protein clipping. The cell lines that displayed the highest levels of clipping also produced less product than the cell lines with a lower level of clipping. In this study we applied differential quantitative label-free LC-MS/MS proteomic analysis to group clonally-derived cell lines (CDCLs) based on the level of clipping of the Fc-fusion protein. The analysis was carried out over two times points in culture and clones were designated as either having 'high' or 'low' clipping phenotypes. We have identified 200 differentially expressed proteins using quantitative label-free LC-MS/MS analysis between the two experimental groups. Functional assessment of the resultant proteomic data using Gene Ontology analysis showed a significant enrichment of biological processes and molecular functions related to protein folding, response to unfolded protein and protein translation. The levels of several proteases were also increased. This study identified protein targets that could be modified using cell line engineering approaches to improve the quality of recombinant Fc-fusion protein production in the biopharmaceutical industry.

Key Words: Chinese hamster ovary (CHO) cells, Fc-fusion protein, productivity, label-free quantitative proteomics, biopharmaceuticals.

9.1   Introduction

The Chinese hamster ovary (CHO) cell line is the principal animal cell expression system for the production of recombinant therapeutic monoclonal antibody proteins (Walsh, 2014). Fc-fusion proteins are rapidly emerging as alternatives in drug development because of their longer serum half-lives (Czajkowsky et al., 2012; Konterman, 2016; Strohl, 2017). Biologically active proteins have very short half-lives due to fast renal clearance (Maack et al. 1979). Antibodies have been very successfully used as biotherapeutics to date because the constant region or Fc domain prolongs the serum half-life due to the pH dependent binding of the neonatal Fc receptor (FcRn), thus avoiding endosomal degradation along with their larger size which limits renal degradation (Lencer and Blumberg 2005).

During recombinant protein production, certain CHO host cell lines, when grown under certain process conditions, display undesirable product proteolysis or 'clipping' (Chirino et al. 2004). This results in the generation of product fragments, which can impact negatively on the yield of intact, high quality, functional protein with associated potential immunogenic risks (Eon-Duval et al. 2012). Several causes have been identified for this phenomenon, often related to a reduction of disulfide bonds or aberrant glycosylation that allows access by proteases in the culture media to cleavage sites in the protein product (Kao et al. 2010). However, a fundamental understanding of the molecular events, both intracellularly and extracellularly, that can lead to product degradation, such as 'clipping', is lacking.

In this study, six recombinant CHO clonally-derived cell lines (CDCLs) from a single transfection and selected bulk culture exhibiting varying levels of clipping of the Fc-fusion molecule were compared over distinct time points from a fed-batch shake flask experiment.   It has been shown previously through protease inhibitor studies that proteolysis is involved in the extracellular clipping of the Fc-fusion protein. The cell lines that displayed a higher level of clipping also produced less final product when compared to the cells with a lower level of clipping. The current study was carried out to determine the intracellular events that play a role in impacting the range of Fc-fusion protein clipping. The six cell lines were grouped into two sets of three cell lines that were compared experimentally using quantitative label-free LC-MS/MS proteomic analysis; one group of

cell lines displayed a 'low' clipping phenotype while the second group displayed a 'high' clipping phenotype. Comparative proteomic profiling of the cell lines was carried out on day 6 and day 10 cell pellet samples. Gene Ontology analysis of the differentially expressed proteins indicated an over-representation of proteins associated with protein folding, the Unfolded Protein Response (UPR) and protein translation.

9.2   Materials and Methods

9.2.1   Fed-batch cultivation of CHO cell lines

Chosen CDCLs were seeded in E250 mL shake flasks containing 100 mL of Lilly propriety production medium without antibiotics or methionine sulfoximine (MSX) at 0.75 x 10^6 cells/mL. All chosen cell lines were cultured in duplicate at 150 rpm, 6% CO_2 and 35 °C for 14 days in a Kuhner Shaker ISF1-X. Cell density and culture viability were determined using an automated Vicell^{TM} XR cell viability analyzer (Beckman Coulter, Brea, CA).

9.2.2   Protein A HPLC

Protein concentration (titer) was measured via analytical protein A affinity chromatography using an Applied Biosystems POROS A 20 µm, 2.1 x 30 mm column (Applied Biosystems, Foster City, CA) on an Agilent 1100 high performance liquid chromatography (HPLC) with UV detection (Agilent Technologies, Santa Clara, CA).

9.2.3   Protein A capture

Cell culture supernatant was purified in Whatman 24 well filter plates (GE Healthcare, Uppsala, Sweden) using GE Healthcare MabSelect Protein A resin, eluting retained product using 50 mM acetic acid. Protein A captured material was used for various analytical assays as described below.

9.2.4 Reversed-phase HPLC for determination of % clipping of intact Fc-fusion protein

Separation of ProA purified Fc-fusion samples were analysed for percentage clipping using an Acquity UPLC BEH 300 C4, 1.7 μm, 2.1x150 mm column system (Waters, Milford, MA) on an Acquity UPLC system (Waters, Milford, MA) equipped with UV detection. 20 μL of sample was injected onto the column in each analysis. Column temperature was maintained at 70°C. The column was eluted across a buffer gradient with 30 - 35% acetonitrile (ACN) at a flow rate of 0.2 mL/min, detected by UV absorbance at 214 nm wavelength. Instrument control and data analysis was completed by Empower 3 software (Waters).

9.2.5 SDS PAGE for analysis of Fc-fusion protein clipping

Protein content from conditioned medium was determined using a Bradford protein assay (BioRad). 1 μg of conditioned medium was mixed with an equal volume of SDS Laemmli buffer (Sigma Aldrich). The samples were heated for 1 min at 100°C and loaded immediately onto a NuPAGE 4-12% Bis-Tris gel (Invitrogen). The protein samples were separated at a constant voltage (200 V) for 45 mins. The gel was then stained using Brilliant Blue G Colloidal concentrate (Sigma Aldrich), as per manufacturer's instructions.

9.2.6 Sample preparation for proteomic analysis

Protein extraction and digestion of whole cell lysates was carried out essentially as previously described (Sommeregger et al. 2016). 200 μg of total protein from each cell line was precipitated using a 2D-Clean-Up Kit (GE Healthcare), as per manufacturer's instructions. The resulting protein pellets were resuspended in 6 M urea, 2 M thiourea, 0.1 M Tris-HCL pH 8.0 and quantified again using Bradford protein assay. 20 μg of each sample was transferred to a fresh tube and was diluted with 50 mM ammonium bicarbonate to a final volume of 50 μL. Each protein sample was reduced with 0.5 M DTT for 20 minutes at 56 °C and alkylated with 0.55 M iodoacetamide for 20 minutes at room temperate in the dark. 0.5 μL of a 1% Protease Max Surfactant Trypsin Enhancer (Promega) solution was added to each sample. 0.5 μg of sequence grade trypsin (Promega) was added to each sample (1:20 trypsin:protein ratio) and

the samples were digested overnight at 37 °C. Digestion was halted by adding 10 µL of a 10 % Trifluroacetic acid solution. The peptide samples were desalted using C18 Spin Columns (Pierce, Thermo Fisher Scientific) and washed using 2 % Acetonitrile (ACN). Eluted peptides were dried using a SpeedVac and stored at -20 °C until they were ready for LC-MS/MS analysis.

9.2.7   LC-MS/MS analysis

Dried peptide samples were re-solubilised in 25 µL of LC-MS grade water with 0.1% formic acid (FA) and 2% ACN. Nano LC–MS/MS analysis was carried out using an Ultimate 3000 RSLCnano system (Thermo Fisher Scientific) coupled to a hybrid linear ion trap/Orbitrap mass spectrometer (LTQ Orbitrap XL; Thermo Fisher Scientific). SilicaTip™ Standard Coating Tubing OD/ID 360/20µm Tip, ID 10 µm, Length 5cm (New Objective) were used as emitter tips for nano electrospray.

A 5 µL injection was picked up using the Ultimate 3000 nanoLC system autosampler and loaded onto a C18 trap column (C18 PepMap, 300 µm ID × 5 mm, 5 µm particle size, 100 Å pore size; Thermo Fisher Scientific). The sample was desalted for 3 min using a flow rate of 25 µL/min in 0.1% TFA containing 2% acetonitrile The trap column was then switched online with the analytical column (PepMap C18, 75 µm ID × 250 mm, 3 µm particle and 100 Å pore size; (Thermo Fisher Scientific)) using a column oven at 40 °C and peptides were eluted with the following binary gradients of: Mobile Phase Buffer A and Mobile phase buffer B: 0–25% solvent B in 160 min and 25–50% solvent B in a further 20 min, where solvent A consisted of 2% ACN and 0.1% FA in water and solvent B consisted of 80% ACN and 0.08% FA in water. Column flow rate was set at 300 nL/min. Data were acquired with Xcalibur software, version 2.0.7 (Thermo Fisher Scientific).

The LTQ Orbitrap XL was operated in data-dependent mode and externally calibrated. Survey MS scans were acquired in the Orbitrap in the 400–1800 m/z range with the resolution set to a value of 30,000 at m/z 400. Up to three of the most intense ions (1+, 2+ and 3+) per scan were CID fragmented in the linear ion trap. Dynamic exclusion was enabled with a repeat count of 1, repeat duration of 30 seconds, exclusion list size of 500 and exclusion duration of 40 seconds.

The minimum signal was set to 500. All tandem mass spectra were collected using a normalised collision energy of 32%, an isolation window of 2 m/z with an activation time of 30.

9.2.8 Quantitative label-free LC-MS/MS data analysis

The raw MS data files obtained were processed using Progenesis QI for Proteomics software (version 2.0; Non-Linear Dynamics, a Waters company, Newcastle upon Tyne, UK). Peptide LC retention times from all MS data files were aligned to an assigned reference run (the run with the most peptides features). Peptide features were filtered using the following parameters; (1) peptide features with ANOVA ≤ 0.05 between experimental groups, (2) mass peaks with charge states from +1 to +3 and greater than one isotope per peptide. An MGF file format was then generated from all exported MS/MS spectra which was used for peptide identification via Proteome Discoverer 2.1 using Sequest HT (Thermo Scientific) and MASCOT (www.matrixscience.com) using Percolator against NCBI database CRIGR (downloaded November 2016 containing 24,904 sequences) and the Fc-fusion protein sequence. The following search parameters were used for protein identification: (1) peptide mass tolerance set to 20 ppm, (2) MS/MS mass tolerance set to 0.6 Da, (3) up to two missed cleavages were allowed, (4) carbamidomethylation set as a fixed modification and (5) methionine oxidation as a variable modification. For re-importation back into Progenesis LC–MS software for further analysis, only high confidence peptides (FDR<1%) with XCorr scores >1.5 for singly charged ions, >2.0 for doubly charged ions and >2.5 for triply charged ions (from SEQUEST) and MASCOT scores of >40 were considered. Statistical criteria of ANOVA p-value less than 0.05 and fold change greater/less than 1.2 fold (Sommeregger et al., 2016) was then applied to the differentially expressed peptides between the two experimental groups (i.e. Day 6 high clipping vs low clipping, Day 10 high clipping vs low clipping). Day 14 was not compared due to the large variation in viability at this stage of culture between the 6 clones (see Figure 9-2).

9.2.9 Gene Ontology (GO) analysis

Identified differentially expressed proteins were assigned official Gene Symbol identifiers by entering the Protein GI Accession information into DAVID (https://david.ncifcrf.gov). These

gene symbols were then imported into DAVID for functional pathway analysis using GO. In order to refine the number of pathways identified, an adjusted p-value (Benjamini Hochberg) of ≤0.05 was used. The STRING database (http://string-db.org; version 10.0) (Szklarczyk et al. 2010) of known and predicted protein interactions that include direct physical and indirect functional protein associations was also used to analyse the data.

9.3   Results

9.3.1   Phenotypic assessment of CHO cell lines with a 'High' and 'Low' clipping phenotype

For characterisation of the Fc-fusion molecule clipping, size exclusion studies followed by mass spectrometry characterization identified three separate cleavage sites within the IgG4 Fc-fusion molecule. Cleavage at all three sites in the partner moiety of the molecule, which is a member of the endocrine subfamily, occurred after a charged aliphatic amino acid. Bioassay analysis showed that cleavage of the fusion partner peptide affects the potency of the molecule. There is no impact on Protein A binding due to the truncation of the partner moiety of the molecule only, as the IgG4 CH2 domain remains intact. Figure 9-1A shows an overview of structure of Fc-fusion molecule showing the cleavage of the molecule within the fusion partner sequence occurring at three individual sites (S). Proteolytic cleavage is thought to occur in a sequential order from the C-terminal end of the partner moiety of the molecule (S1, S2, S3). Figure 9-1B shows an SDS-PAGE gel of conditioned medium from two cell lines producing the Fc-fusion molecule clearly shows product clipping when compared to purified full-length protein.

Six cell lines from the a single transfection project were chosen for this study based on their consistent behaviour with regards to the amount of intact Fc-fusion protein produced over time in culture. Three of these cell lines displayed a "high" clipping phenotype and three clones a "low" clipping phenotype. The percentage of unclipped Fc-fusion protein was measured on Day 6, Day 10 and Day 14 using HPLC-MS and the difference between the two cell experimental

groups at each time point is statistically significant as shown in Figure 9-2A. Product titre measured at days 8, 10 and 14 is shown in Figure 9-2B and shows that the cell lines with the 'high' clipping phenotype consistently produced less product than the 'low' clipping cell lines at each day of culture. Viable cell density and culture viability for the cell lines with the 'high' clipping compared to the 'low' clipping phenotype at each time point at day 6, 10 and 14 are shown in Figure 9-2C and Figure 9-2D. By day 14 the culture viability had dropped to 60% or less. There was no correlation between percentage viability and percentage unclipped Fc-fusion protein as the culture viability of the two cell line groups remains similar at each time point.

9.3.2   Proteomic analysis of cell lines with 'high' and 'low' clipping phenotypes

To assess what is happening at the intracellular level in these clipping phenotypes a proteomic analysis was carried out on the two experimental groups of cell lines at two time points, on Day 6 and Day 10. Day 14 was not analysed due to the lower culture viability at this time (<60%) (Figure 9-2D). Tryptically digested proteins from whole cell lysate samples were subjected to LC-MS/MS analysis. Label-free LC-MS/MS quantitative analysis of the cell lines with the "high" clipping and the "low" clipping phenotypes was carried out using Progenesis QI for Proteomics differential analysis software. During this analysis the overall percentage score for peptide/feature alignment from each LC-MS run was of high quality at greater than 89% which compares favourably with other studies (Theron et al. 2014). The software calculates the ion intensity for each aligned peptide/feature to determine the relative abundances of proteins between the cell lines with the "high" and "low" clipping phenotypes. The MS/MS spectrum of the statistically significant quantified peptides was searched using a combination of the algorithms SEQUEST and MASCOT against a CRIGR fasta database so the peptide identification and abundance were obtained and compared between the two experimental groups.

Data visualisation using heat maps was generated using ggplot in R to build a Pearson correlation matrix for the differential data for the individual time points and showed good separation of the cell lines into the two experimental groups at each time point (see Figure 9-3). From the results of the label-free relative quantitative analysis carried out comparing the two experimental groups it was found that 81 proteins had a statistically significant level of

differential expression on day 6. 42 proteins showed increased expression and 39 showed decreased expression in the cell lines with the "high" clipping versus 'low' clipping phenotype at day 6 (see supplementary Table 1 (https://ars.els-cdn.com/content/image/1-s2.0-S016816561830172X-mmc2.xlsx). 145 proteins showed a significant level of differential expression on day 10, where 75 proteins were overexpressed and 70 showed decreased expression in the cell lines with the "high" clipping clones compared to the cell lines with the 'low' clipping phenotype (see supplementary Table 2 https://ars.els-cdn.com/content/image/1-s2.0-S016816561830172X-mmc2.xlsx). All proteins displaying a statistically significant level of differential expression from the day 6 and day 10 individual proteomic comparisons were overlapped, and 23 proteins were identified as following a similar expression profile on both days (see supplementary table https://ars.els-cdn.com/content/image/1-s2.0-S016816561830172X-mmc2.xlsx). Two examples of differentially expressed proteins, i.e. Wolframin (WFS1) and intracellular adhesion molecule 1 (ICAM-1) were identified using this analysis and are outlined in Figure 9-4.

From the proteomic data it is also important to note that we observed a relative increase in intracellular levels of Fc-fusion protein in the cell lines with the 'low' clipping/higher productivity phenotype compared to the cell lines with the 'high' clipping/low productivity phenotype (2.2 fold increase at Day 6 with 2 peptides matched to Fc-fusion protein and 1.7 fold increase at Day 10 with 3 peptides matched to the Fc-fusion protein) by interrogating the LC-MS/MS data files against the full sequence of the Fc-fusion molecule.

9.3.3   Pathway Analysis of Differentially Expressed Proteins

The two lists of differentially expressed proteins from day 6 and day 10 comparisons were submitted to DAVID for gene ontology (GO) analysis. GO biological processes which were found to be over-represented included protein folding and translation while molecular functions which were over-represented included unfolded protein binding (see Table 9-2). Proteins that were identified that were related to the enriched biological process of protein folding include HSP90AB1, PDIA4, PDIA5, HSPA5, ERP29. Proteins related to mRNA translation with decreased expression in the cell lines with the 'high' clipping/reduced productivity phenotype include a number of ribosomal proteins including Rpl15 (1.38 down at day 10), Rpl4 (1.26 fold down at

310

day 10), Rpl8 (1.25 fold down at day 10), RPL18 (1.22 fold down at day 10), RPS19 (1.21 fold down at day 10), Rpl3 (1.2 fold down at day 10), RPL35 (725 fold down at day 6) and RPL23A (1.39 fold down at day 6). Proteins related to the enriched molecular function of the UPR include HSP90AB1, Hsp90b1, HSPA5, WFS1 and UGGT1.

Data mining was also carried out using the STRING database (Szklarczyk et al. 2011). Pathway analysis on Day 10 showed network interactions linked to protein folding and response to unfolded protein significantly overrepresented (Figure 9-4A and Figure 9-4B). Table 9-3 outlines the list of proteins identified with links to protein folding and the unfolded protein response as identified and annotated using STRING analysis.

			Day 6		Day 10	
Accession no.	Gene Name	Protein Description	Peptides matched	Fold Change	Peptides matched	Fold Change
Increased expression in cell lines with 'High Clipping' phenotype						
EGV91615.1	WFS1	Wolframin	1	3.22	1	2.78
EGW13109.1	ITGB1	Integrin beta-1	2	2.38	3	2.33
AAD00079.1	GLG1	latent TGF-beta complexed protein	1	1.92	2	1.60
AAG30280.1	Icam1	intracellular adhesion molecule 1	9	3.69	8	3.64
EGW06206.1	Creld1	Cysteine-rich with EGF-like domain protein 1	1	2.28	1	2.82
EGW07262.1	Ckap4	Cytoskeleton-associated protein 4	1	1.65	3	1.55
EGW01344.1	HSP90AB1	Heat shock protein HSP 90-beta	1	1.59	1	1.42
EGV93182.1	Pdia4	Protein disulfide-isomerase A4	1	1.36	1	2.40
EGW00520.1	Ppid	40 kDa peptidyl-prolyl cis-trans isomerase	1	1.31	1	1.81

EGV98481.1	Fkbp4	FK506-binding protein 4	2	1.25	3	1.48

Decreased expression in cell lines with 'High Clipping' phenotype

EGV97184.1	SORD	Sorbitol dehydrogenase	3	3.00	2	1.24
EGV93786.1	VPS26A	Vacuolar protein sorting-associated protein 26A	1	1.83	1	1.30
EGV93120.1	DLD	Dihydrolipoyl dehydrogenase, mitochondrial	2	1.78	1	1.25
EGV99884.1	NEDD4	E3 ubiquitin-protein ligase NEDD4	1	1.63	1	1.38
EGW12490.1	ATP5B	ATP synthase subunit beta, mitochondrial	1	1.57	1	1.31
XP_016825423.1	ATXN2L	PREDICTED: ataxin-2-like protein isoform X13	1	1.56	1	1.52
EGW12175.1	BLMH	Bleomycin hydrolase	1	1.39	1	1.39
EGW00228.1	LOC100770709	Heterogeneous nuclear ribonucleoprotein G	1	1.27	1	1.37
EGV99768.1	Hspa2	Heat shock-related 70 kDa protein 2	1	2.50	2	1.63

EGW10632.1	Coro1c	Coronin-1C	1	1.76	2	1.37
EGW02055.1	Vim	Vimentin	14	1.48	4	1.47
EGW02330.1	Hsd17b10	3-hydroxyacyl-CoA dehydrogenase type-2	1	1.45	1	1.22
EGW00358.1	Mdh1	Malate dehydrogenase, cytoplasmic	2	1.22	2	1.26

Table 9-1. Significantly differentially expressed proteins that overlap from Day 6 and Day 10 proteomic comparisons.

GO ID	GO Term	P Value	BH Adj.
Biological Process			
GO:0006457	protein folding	2.05E-13	2.79E-10
GO:0006414	translational elongation	2.50E-08	1.70E-05
GO:0006412	translation	7.66E-06	0.003459
GO:0016052	carbohydrate catabolic process	4.06E-05	0.013678
GO:0044275	cellular carbohydrate catabolic process	5.97E-05	0.016077
GO:0006091	generation of precursor metabolites and energy	7.77E-05	0.017424
GO:0051789	response to protein stimulus	2.55E-04	0.048183
Molecular Function			
GO:0051082	unfolded protein binding	1.64E-10	6.73E-08
GO:0051287	NAD or NADH binding	1.78E-06	3.65E-04
GO:0003723	RNA binding	3.16E-05	0.004308
GO:0003735	structural constituent of ribosome	5.05E-05	0.005164
GO:0000166	nucleotide binding	8.18E-05	0.006685
GO:0003755	peptidyl-prolyl cis-trans isomerase activity	9.24E-05	0.006295
GO:0016859	cis-trans isomerase activity	1.20E-04	0.006982
GO:0005198	structural molecule activity	1.33E-04	0.00681
GO:0048037	cofactor binding	3.05E-04	0.013811

Table 9-2. Overview of the most significantly enriched differentially expressed pathways from the comparison of 'high' and 'low' clipping phenotypes at Day 6 and Day 10. Enrichment was considered significant upon observation of a p-value ≤ 0.05 and a Benjamini Hochberg (BH) adjusted p-value ≤ 0.05.

Accession no.	Gene symbol	Protein Description	Network	Network	Unique peptides	Confidence score	Anova	Fold change
Increased expression in cell lines with 'High Clipping' phenotype								
EGW09512.1	Aars	Alanyl-tRNA synthetase, cytoplasmic	Protein folding	Response to unfolded Protein	2	101.25	0.0220	1.50
EGW01989.1	Dnajb11	DnaJ-like subfamily B member 11	Protein folding	Response to unfolded Protein	2	93.06	0.0229	1.41
EGW01344.1	HSP90AB1	Heat shock protein HSP 90-beta	Protein folding	Response to unfolded Protein	1	43.21	0.0494	1.42
EGW09701.1	Hsp90b1	Endoplasmin	-	Response to unfolded Protein	11	560.01	0.0179	1.36
NP_001233668.1	Hspa5	78 kDa glucose-regulated protein precursor	Protein folding	Response to unfolded Protein	6	300.38	0.0095	1.51
AAB00689.1	Hyou1	170 kDa glucose regulated protein	-	Response to unfolded Protein	9	495.85	0.0143	1.35
EGV96414.1	Pdia5	Protein disulfide-isomerase A5	Protein folding	Response to unfolded Protein	1	49.95	0.0466	1.48
EGW02226.1	Serpinh1	Serpin H1	-	Response to unfolded Protein	3	223.28	0.0005	1.51

EGV91615.1	Wfs1	Wolframin	Protein folding	Response to unfolded Protein	1	36.58	0.0017	2.78
EGV97474.1	TRAP1	Glycosyltransferase 25 family member 1	Protein folding	-	1	37.71	0.0228	2.37
EGV93182.1	Pdia4	Protein disulfide-isomerase A4	Protein folding	-	1	35.99	0.0437	2.40
EGV91638.1	Fkbp9	FK506-binding protein 9	Protein folding	-	2	88.97	0.0030	2.51
EGW08023.1	Fkbp10	FK506-binding protein 10	Protein folding	-	2	126.13	0.0027	2.42
EGV98481.1	Fkbp4	FK506-binding protein 4	Protein folding	-	3	164.86	0.0041	1.48
EGV92756.1	Ganab	Neutral alpha-glucosidase AB	Protein folding	-	6	315.65	0.0091	1.68
EGW00369.1	Cct4	T-complex protein 1 subunit delta	Protein folding	-	1	51.81	0.0067	1.21
EGW07173.1	Ppib	Peptidyl-prolyl cis-trans isomerase B	Protein folding	-	1	33.01	0.0290	1.27
EGV92871.1	Dnajc10	DnaJ-like subfamily C member 10	Protein folding	-	2	98.39	0.0343	1.60
EGW07426.1	Erp29	Endoplasmic reticulum protein ERp29	Protein folding	-	1	42.76	0.0452	1.28

EGW00520.1	Ppid	40 kDa peptidyl-prolyl cis-trans isomerase	Protein folding	-	1	23.09	0.0195	1.81
EGV96168.1	Prkcsh	Glucosidase 2 subunit beta	Protein folding	-	4	236.53	0.0065	1.39
EGV93626.1	Trap1	Heat shock protein 75 kDa, mitochondrial	Protein folding	-	1	13.42	0.0320	1.63
EGW02487.1	Txndc5	Thioredoxin domain-containing protein 5	Protein folding	-	1	46.73	0.0362	1.66
EGW06464.1	Uggt1	UDP-glucose:glycoprotein glucosyltransferase 1	Protein folding	Response to unfolded Protein	2	76.38	0.0090	1.51
Decreased expression in cell lines with 'High Clipping' phenotype								
EGV99768.1	Hspa2	Heat shock-related 70 kDa protein 2	Protein folding	Response to unfolded Protein	2	53.3	0.0271	1.63
EGW11739.1	Sec31a	Protein transport protein Sec31A	-	Response to unfolded Protein	1	33.24	0.0010	1.21
EGW05640.1	SRPRB	Signal recognition particle receptor subunit beta	-	Response to unfolded Protein	2	113.44	0.0361	1.51

Table 9-3. Significantly differentially expressed proteins associated with the functional interaction networks of 'Protein Folding' and 'Response to Unfolded Protein' as identified from quantitative label-free LC-MS/MS proteomic analysis on Day 10.
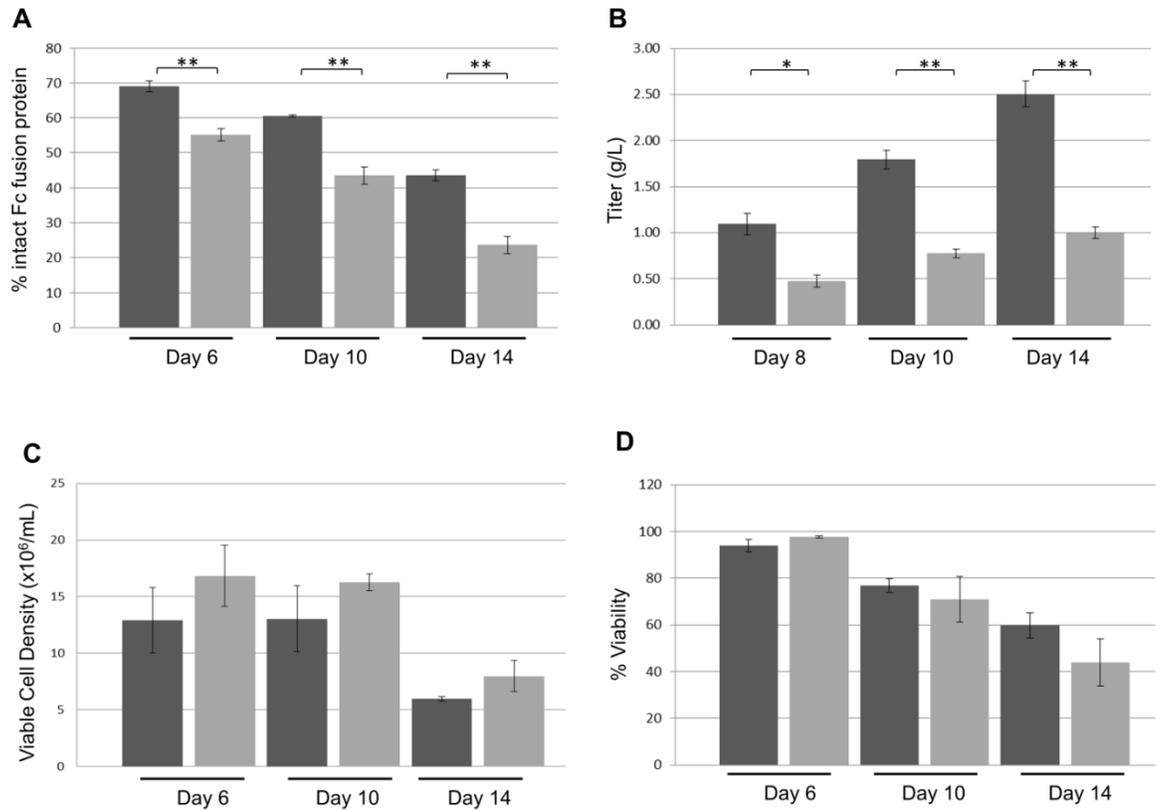
Figure 9-1. (A) **Mean % unclipped Fc-fusion protein (intact protein) for cell lines designated with a "high" and "low" clipping phenotype at the three time points (Day 6, 10 and 14).** (B) **Mean titre of the Fc-fusion protein in g/L for the cell lines with the "high" and "low" clipping phenotypes at three time points (Day 8, 10 and 14) (Day 6 data was not available).** (C) **Comparison of viable cell density and** (D) **% viability for the individual cell lines at Day 6, Day 10 and Day 14. Error bars represent the standard error calculated from the standard deviation of the mean. (*<0.05, **<0.005). Black bars – cell lines with 'low' clipping phenotype; Grey bars – cell lines with 'high' clipping phenotype.**
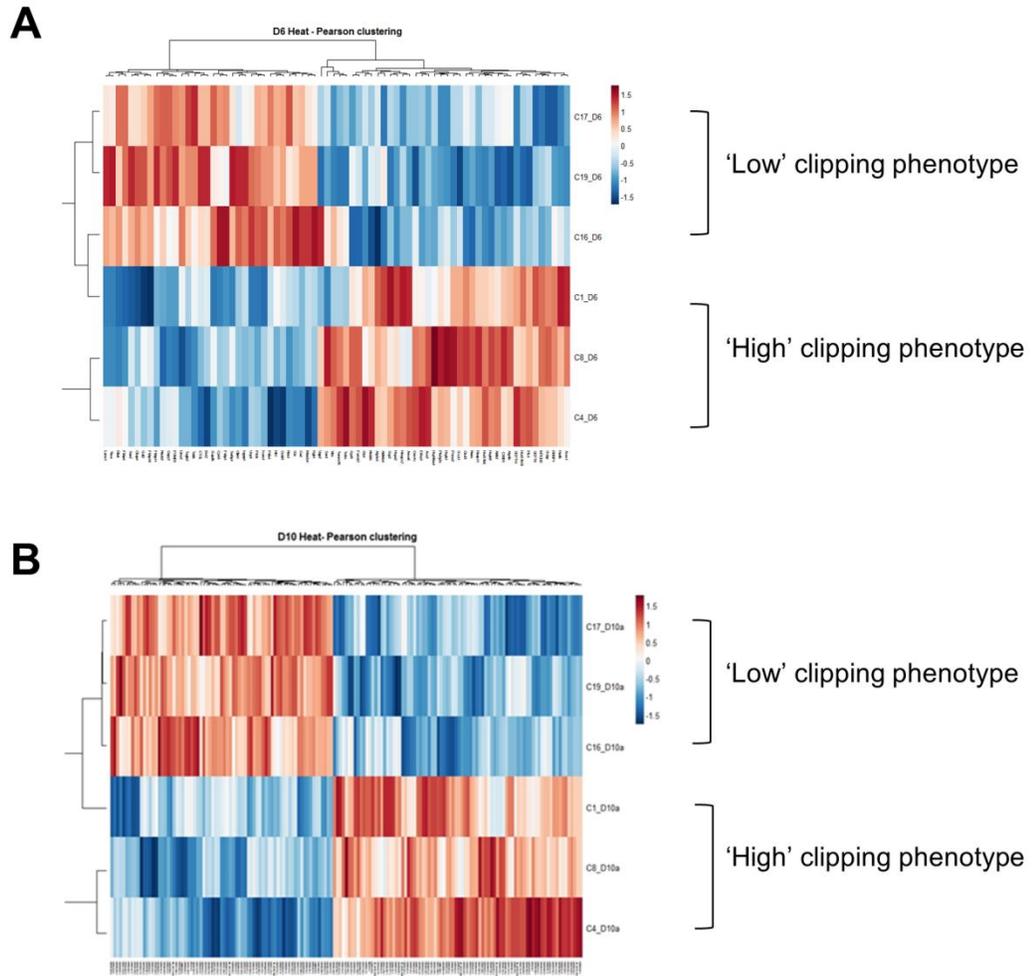
Figure 9-2. Data visualisation using a heatmap generated using ggplot package in R to show the distribution of all identified proteins based on statistical analysis and fold-change across all of the samples used in the quantitative label-free LC-MS/MS proteomic analysis. Cell lines with 'high' and 'low' clipping phenotypes compared at (A) **Day 6 and** (B) **Day 10.**
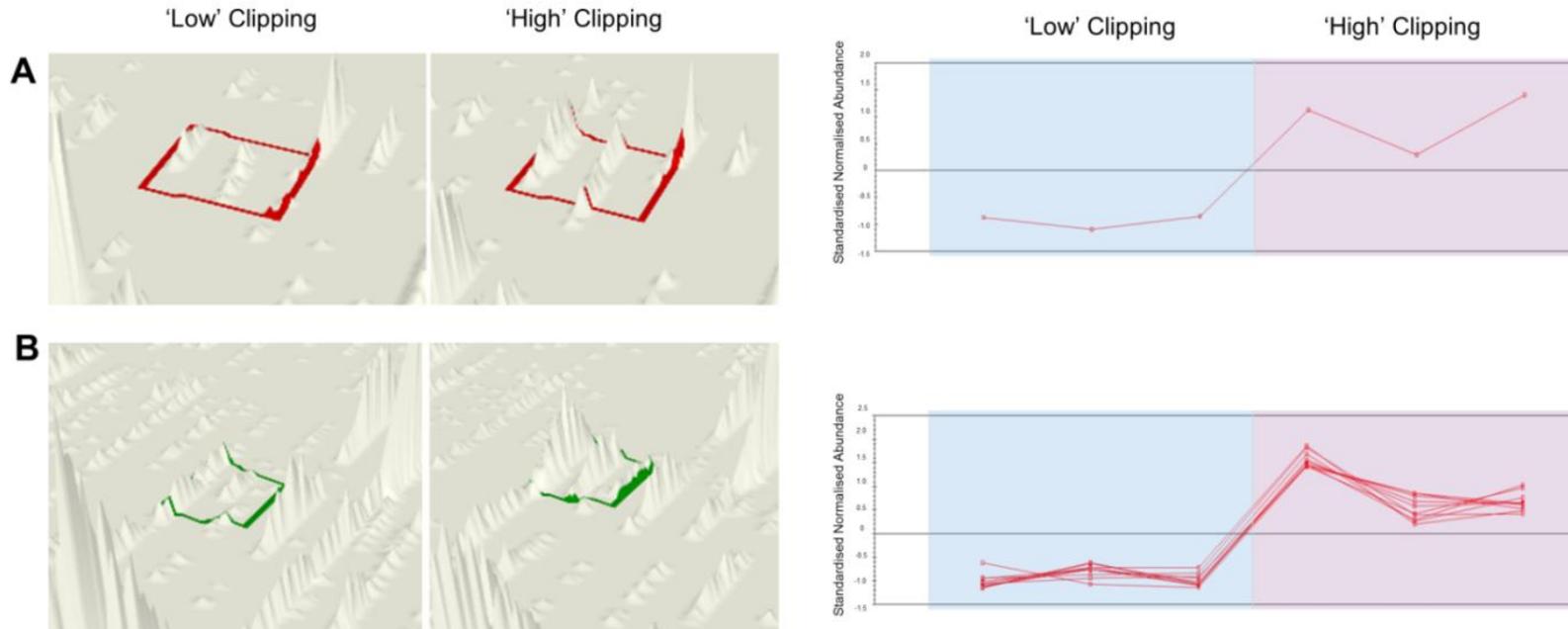
Figure 9-3. Relative quantitative label-free LC-MS/MS analysis using Progenesis QI for Proteomics software showing the abundance of a peptide from (A) **Wolframin (WFS1) and** (B) **ICAM-1 demonstrating increased expression in the cell lines with the 'High' clipping phenotype on Day 10 compared to the cell lines with the 'Low' clipping phenotype. Left panel - peptide view showing a representation of the altered abundance of one peptide from WFS1 and ICAM-1. Right panel - protein view showing the average normalised abundance volumes of the individual peptides identified from WFS1 and ICAM-1. The horizontal axis represents the three individual cell lines from the two experimental groups ('high' clipping and 'low' clipping'). The vertical axis represents normalised abundance volumes (log).**
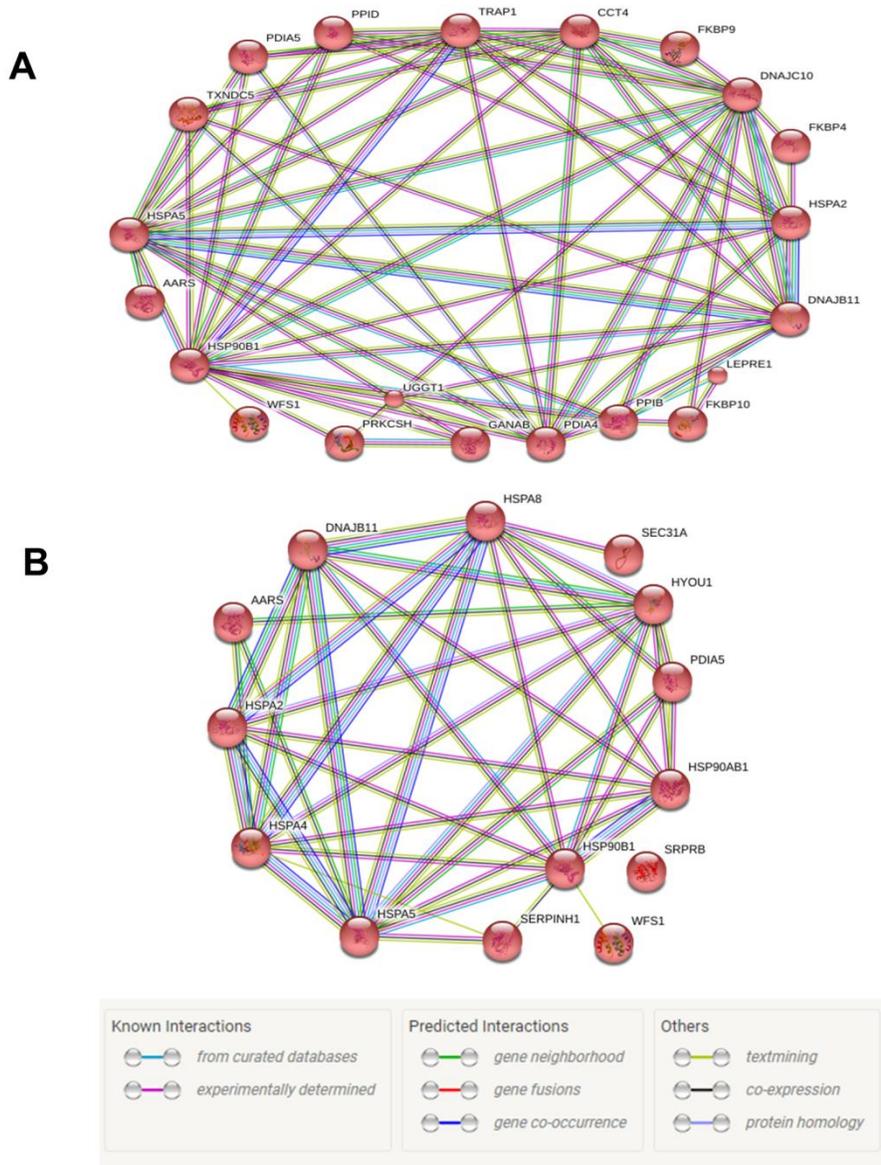
Figure 9-4. Functional interactions network analysis using STRING shows an enrichment of protein representing (A) 'protein folding' and (B) 'response to unfolded protein' following analysis of the significantly differentially expressed proteins on Day 10 between the cell lines with the "high" and "low" clipping phenotypes. The full list of proteins and direction of fold change is outlined in Table 9-3.

9.4 Discussion

The purpose of this proteomic study was to shed light on the intracellular mechanisms involved in both proteolysis and productivity in CHO cell lines derived from a single transfection project that express an Fc-fusion protein with varying levels of product quality. Clipping of Fc-fusion proteins is a major problem in the biopharmaceutical industry leading to significant differences in intact protein and product titre between different cell lines (Eon-Duval et al. 2012). In this study we compared three cell lines with a 'high' clipping phenotype with three cell lines which displayed a 'low' clipping phenotype; the 'clipping' phenotype was based on the level of clipping of the Fc-fusion protein. The cell lines with the 'high' clipping phenotype also produced less product over the course of the fed-batch study. We identified over 200 differentially expressed proteins between the two samples groups, with pathway analysis suggesting an enrichment of pathways linked to induction of the UPR and protein folding. The proteomic data suggests that the primary defect in the cell lines with the 'high' clipping and reduced productivity phenotype could be potentially due to reduced efficiency in protein folding, leaving the Fc-fusion protein vulnerable to proteolysis at a variety of sites which would ordinarily be more efficiently protected by correct folding. This defect in folding may lead to the induction of the UPR which is turn may induce the expression and/or activation of several proteases that may lead to external clipping of the product. It is also a possibility that the induction of the UPR could be a cellular response to the clipping itself and not due to unfolded material that then gets clipped.

Protein secretion starts with translocation of protein into the endoplasmic reticulum (ER) where equilibrium is maintained by a coordinated regulation of the following pathways; endoplasmic reticulum associated degradation (ERAD), unfolded protein response (UPR) and endoplasmic reticulum associated folding (ERAF) (Prashad and Mehra, 2015). In this proteomic study a number of proteins related to protein folding and the UPR showed altered expression in CDCLs displaying a 'high' clipping/reduced productivity phenotype. The UPR comprises three signalling branches arising from the transmembrane transducers, inositol-requiring enzyme 1 (IRE1), activated transcription factor 6 (ATF6) and protein kinase RNA-activated ER kinase (PERK) (Ron and Walter, 2007). If there is an accumulation of unfolded or mis-folded protein in the ER, the UPR pathway attempts to overcome ER stress by attenuating mRNA translation and

up-regulating the capacity of the cells to fold proteins by increasing the transcriptional level of chaperones (Schroder, 2006; Hussain et al., 2014; Wang and Kaufman, 2014). Induction of the UPR can also result in the up-regulation of degradation enzymes to remove mis-folded proteins to maintain homeostasis (Hussain et al. 2014). In this study, there is evidence for an attenuation of mRNA translation with decreased expression of a number of ribosomal proteins in the cell lines with the 'high' clipping/reduced productivity phenotype including Rpl15, Rpl4, Rpl8, RPL18, RPS19, Rpl3, RPL35 and RPL23A. Although the reduction in certain ribosomal and translation-related proteins might suggest a reduction in global protein synthesis in the low productivity/'high' clipping clones, this is unlikely since there is no reduction in growth rate in these clones.

The correct folding of a protein is essential for protein function. Incorrect post-translational modifications, disassembly of oligomeric complexes, cellular stress, mutation and off-pathway folding can cause a protein to mis-fold (Cyr et al. 2002). Chaperones and folding enzymes reside in the ER lumen and participate in all stages of protein folding. The main chaperone families of the ER are Hsp70s, Hsp40s, Hsp90, Peptidyl-prolyl isomerases, calnexin & calreticulin and Thiol-disulphide oxidoreductases (Sitia and Braakman, 2003). To deal with mis-folded proteins in the cell, molecular chaperones can refold misfolded protein or degrade misfolded protein (Houck et al. 2012). The two most abundant ER proteins are the chaperones, glucose-regulated protein 78 (also known as GRP78, HSPA5 or BiP) and glucose-regulated protein 94 (also known as GRP94 or HSP90B1) (Schröder, 2008; Prashad and Mehra, 2015). Both HSPA5 and HSP90B1 showed increased expression in the 'high' clipping/reduced productivity cell lines at Day 10. Under non-stressed conditions in the ER, the three transmembrane sensors that mediate the UPR, PERK, ATF6 and IRE1, are inactive and are bound to HSPA5 (Hussain et al. 2014). When a protein misfolds in the ER, the misfolded protein sequesters HSP5A and there is a reduction in the level of free HSP5A, activating signalling pathway to induce its transcription along with other UPR protein chaperones (Ma and Hendershot, 2004). In this study, we have shown that HSP5A is over expressed by 1.5 fold in the "high" clipping cell lines at day 10, suggesting possible activation of the UPR. The chaperone HSP90B1, which was 1.4 fold increased in cell lines with the 'high' clipping phenotype, is known to bind to partially folded proteins and acts sequentially after HSPA5 (Schröder, 2008). In addition to these chaperones, a number of members of the protein disulphide isomerase (PDI) family of proteins were identified as differentially expressed in this

study; PDIA4 was overexpressed at both time points in the "high" clipping cell lines (1.4 fold at day 6 and 2.4 fold by Day 10), while PDIA5 was overexpressed by 1.5 fold at day 10 in the "high" clipping cell lines. Disulphide bond formation is catalysed by PDIs to ensure protein structural stability and promote assembly of multi-protein complexes (Wilkinson and Gilbert, 2004).

Several other UPR-related proteins were found to be differentially expressed in this study including Wolframin (WFS1) and UDP-glucose glycoprotein glucosyltransferase 1 (UGGT1). WFS1 protein is located in the endoplasmic reticulum and shows increased expression in the cell lines with the 'high' clipping phenotype at day 6 (3.2 fold increased) and day 10 (2.8 fold increased). The WFS1 protein is thought to regulate the amount of calcium in cells. WFS1 has been shown to be a downstream target of IRE1 and PERK signalling, induced transcriptionally and translationally in response to ER stress, and when suppressed, causes high levels of ER stress in the β-cell (Fonseca and Fukuma et al. 2005). WFS1 also regulates ATF6 activity, targeting ATF6 to the E3 ubiquitin ligase HRD1, which results in its ubiquitination and proteasomal degradation (Fonseca et al. 2005). UGGT1 was found to be increased by 1.5 fold on Day 10 in the cell lines with the 'high' clipping phenotype. The function of UGGT1 is to recognise glycoproteins with minor folding defects which are partially degraded and are not recognised by HSP5A (Prashad and Mehra, 2015). It will reglucosylate single N-glycans near the misfolded protein acting as a quality control for protein folding in the ER. Reglucosylated proteins are recognised by calreticulin for recycling to the ER and refolding or degradation (Arnold et al. 2000). Further work will involve the characterisation of the induction of classical UPR markers such as XBP1 in the high clipping/low productivity cell lines that were not observed to be differentially expressed in the proteomic data, possibly due to low abundance levels of such proteins not detectable by mass spectrometry.

We have also identified a number of proteases that show increased expression in the cell lines with the 'high' clipping phenotype. This could be due to the increased UPR activity in these cell lines as the UPR is known to upregulate degradation pathways and proteases (Wang and Kaufman, 2014). Proteases are responsible for the removal of unfolded, damaged or unwanted proteins and contribute to the clearance of damaged proteins by degrading them into short peptides in an attempt to prevent accumulation of protein aggregates (Gur et al. 2013). In this study, we have identified Cathepsin D and Cathepsin L1 with increased expression in the 'high'

clipping cell lines compared to the 'low' clipping cell lines at Day 10. The synergistic induction of the pro-cathepsin secretion is predominately regulated by activation of the inositol-requiring enzyme 1α (IRE1α) axis of the UPR pathway (Yan et al. 2016). The UPR also induces expression of cathepsin D in A549 lung cells (Nguyen and Uhal, 2016). Cathepsin D has previously been found to cause proteolytic degradation of Fc-fusion proteins produced from recombinant CHO cells (Robert et al. 2009; Lim et al. 2017). We also identified Scpep1 protein, which showed increased expression in the cell lines with the 'high' clipping phenotype. Scpep1 is a lysosomal serine carboxypeptidase and is a member of the α/β hydrolase family of proteins that use a Ser-Asp-His catalytic triad to cleave the carboxyterminal peptide bonds on their protein substrates (Kourist et al. 2010). This increased expression of a number of proteases in the cell lines with the 'high' clipping phenotype could potentially be responsible for the external clipping of the Fc-fusion protein product.

9.5   Conclusions

The aim of this study was to shed light on the intracellular mechanisms that render an Fc-fusion protein produced by CHO cells prone to external clipping and a reduction in intact product titre. Analysis of the proteomic data generated suggests that the cell lines with the 'high' clipping/reduced productivity phenotype may have a reduced efficiency in global protein folding leaving the Fc-fusion protein vulnerable to proteolysis at a variety of sites that would ordinarily be more efficiently protected by folding. Our data does not identify a specific cause for the hypothesised poor protein folding in the 'high' clipping/low productivity cell lines. It may be a response to some unusual conformation of the "unnatural" fusion protein, and this response may vary on a random clonal basis. This defect in folding may lead to the induction of the UPR which is turn may induce the expression and/or activation of several proteases that may lead to external clipping of the product. Future work will involve the use of specific protease inhibitors and cell line engineering approaches to improve the quality of recombinant Fc-fusion protein production in these recombinant cell lines.

9.6   References

Arnold, S. M., Fessler, L. I., Fessler, J. H., Kaufman, R. J. 2000. Two homologues encoding human UDP-glucose: Glycoprotein glucosyltransferase differ in mRNA expression and enzymatic activity. Biochemistry 39(9), 2149-2163.

Chirino, A. J., Mire-Sluis, A. 2004. Characterizing biological products and assessing comparability following manufacturing changes. Nat. Biotechnol. 22(11), 1383. doi:10.1038/nbt1030

Cyr, D. M., Höhfeld, J., Patterson, C. 2002. Protein quality control: U-box-containing E3 ubiquitin ligases join the fold. Trends Biochem Sci. 27(7), 368-375.

Czajkowsky, D. M., Hu, J., Shao, Z., Pleass, R. J. 2012. Fc-fusion proteins: New developments and future perspectives. EMBO Mol. Med. 4(10), 1015-1028. doi: 10.1002/emmm.201201379

Eon-Duval, A., Broly, H., Gleixner, R. 2012. Quality attributes of recombinant therapeutic proteins: An assessment of impact on safety and efficacy as part of a quality by design development approach. Biotechnol. Prog. 28(3), 608-622. doi: 10.1002/btpr.1548.

Fonseca, S. G., Fukuma, M., Lipson, K. L., Nguyen, L. X., Allen, J. R., Oka, Y., Urano, F. 2005. WFS1 is a novel component of the unfolded protein response and maintains homeostasis of the endoplasmic reticulum in pancreatic beta-cells. J. Biol. Chem. 280(47), 39609-39615. doi: 10.1074/jbc.M507426200

Gur, E., Ottofueling, R., Dougan, D. A. 2013. Machines of destruction–AAA proteases and the adaptors that control them. In: Dougan D.A. (ed) Regulated proteolysis in microorganisms. Springer, Subcell. Biochem. 66, pp3-33.

Houck, S. A., Singh, S., Cyr, D. M. 2012. Cellular responses to misfolded proteins and protein aggregates. Methods Mol. Biol. 832: 455–461. doi: 10.1007/978-1-61779-474-2_32

Hussain, H., Maldonado-Agurto, R., Dickson, A. J. 2014. The endoplasmic reticulum and unfolded protein response in the control of mammalian recombinant protein production. Biotechnol. Lett. 36(8), 1581-1593. doi: 10.1007/s10529-014-1537-y.

Kao, Y., Hewitt, D. P., Trexler-Schmidt, M., Laird, M. W. 2010. Mechanism of antibody reduction in cell culture production processes. Biotechnol. Bioeng. 107(4), 622-632. doi: 10.1002/bit.22848.

Kontermann, R. E. 2016. Half-life extended biotherapeutics. Expert Opin. Biol. Ther. 16(7), 903-915. doi: 10.1517/14712598.2016.1165661

Kourist, R., Jochens, H., Bartsch, S., Kuipers, R., Padhi, S. K., Gall, M., Böttcher, D., Joosten H. J, Bornscheuer, U. T. 2010. The α/β-hydrolase fold 3DM database (ABHDB) as a tool for protein engineering. Chembiochem 11(12), 1635-1643. doi: 10.1002/cbic.201000213

Lencer, W. I., Blumberg, R. S. 2005. A passionate kiss, then run: Exocytosis and recycling of IgG by FcRn. Trends Cell Biol. 15(1), 5-9. doi: 10.1016/j.tcb.2004.11.004

Lim, A., Doyle, B. L., Kelly, G. M., Reed-Bogan, A. M., Breen, L. H., Shamlou, P. A., Lambooy, P. K. 2017. Characterization of a cathepsin D protease from CHO Cell-Free medium and mitigation of its impact on the stability of a recombinant therapeutic protein. Biotechnol. Prog. Jul 20. doi: 10.1002/btpr.2530. [Epub ahead of print].

Liu, P., Cheng, H., Roberts, T. M., Zhao, J. J. 2009. Targeting the phosphoinositide 3-kinase pathway in cancer. Nat. Rev. Drug Discov. 8(8), 627-644. doi: 10.1038/nrd2926

Ma, Y., Hendershot, L. M. 2004. The role of the unfolded protein response in tumour development: Friend or foe? Nat. Rev. Cancer 4(12), 966. doi: 10.1038/nrc1505

Maack, T., Johnson, V., Kau, S. T., Figueiredo, J., Sigulem, D. 1979. Renal filtration, transport, and metabolism of low-molecular-weight proteins: A review. Kidney Int. 16(3), 251-270.

Nguyen, H., Uhal, B. D. 2016. The unfolded protein response controls ER stress-induced apoptosis of lung epithelial cells through angiotensin generation. Am. J. Physiol. Lung Cell. Mol. Physiol. 311(5), L846-L854. doi: 10.1152/ajplung.00449.2015

Prashad, K., Mehra, S. 2015. Dynamics of unfolded protein response in recombinant CHO cells. Cytotechnology 67(2), 237-254. doi: 10.1007/s10616-013-9678-8.

Robert, F., Bierau, H., Rossi, M., Agugiaro, D., Soranzo, T., Broly, H., Mitchell-Logean, C. 2009. Degradation of an Fc-fusion recombinant protein by host cell proteases: Identification of a CHO cathepsin D protease. Biotechnol. Bioeng. 104(6), 1132-41. doi: 10.1002/bit.22494

Ron, D., Walter, P. 2007. Signal integration in the endoplasmic reticulum unfolded protein response. Nat. Rev. Mol. Cell Biol. 8(7), 519. doi: 10.1038/nrm2199

Schröder, M. 2008. Endoplasmic reticulum stress responses. Cell. Mol. Life Sci. 65(6), 862-894. doi: 10.1007/s00018-007-7383-5

Schröder, M. 2006. The unfolded protein response. Mol. Biotechnol. 34(2), 279-290. doi: 10.1385/MB:34:2:279

Schröder, M., Kaufman, R. J. 2005. The mammalian unfolded protein response. Annu. Rev. Biochem. 74, 739-789. doi: 10.1146/annurev.biochem.73.011303.074134

Sitia, R., Braakman, I. 2003. Quality control in the endoplasmic reticulum protein factory. Nature 426(6968), 891. doi: 10.1038/nature02262

Sommeregger, W., Mayrhofer, P., Steinfellner, W., Reinhart, D., Henry, M., Clynes, M., Meleady, P., Kunert, R. 2016. Proteomic differences in recombinant CHO cells producing two similar antibody fragments. Biotechnol. Bioeng. 113(9), 1902-12. doi: 10.1002/bit.25957

Strohl, W. R. 2017. Current progress in innovative engineered antibodies. Protein Cell, 1-35. doi: 10.1007/s13238-017-0457-8

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., Jensen, L.J., von Mering, C. 2011. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. Jan;39(Database issue), D561-8. doi: 10.1093/nar/gkq973

Theron, L., Gueugneau, M., Coudy, C., Viala, D., Bijlsma, A., Butler-Browne, G., Maier, A., Béchet, D., Chambon, C. 2014. Label-free quantitative protein profiling of vastus lateralis muscle during human aging. Mol. Cell. Proteomics 13(1), 283-294. doi: 10.1074/mcp.M113.032698

Walsh, G. 2014. Biopharmaceutical benchmarks 2014. Nat. Biotechnol. 32(10), 992-1000. doi: 10.1038/nbt.3040

Wang, M., Kaufman, R. J. 2014. The impact of the endoplasmic reticulum protein-folding environment on cancer development. Nat. Rev. Cancer 14(9), 581. doi: 10.1038/nrc3800

Wilkinson, B., Gilbert, H. F. 2004. Protein disulfide isomerase. Biochim. Biophys. Acta (BBA)-Proteins and Proteomics 1699(1), 35-44. doi: 10.1016/j.bbapap.2004.02.017

Yan, D., Wang, H., Bowman, R. L., Joyce, J. A. 2016. STAT3 and STAT6 signaling pathways synergize to promote cathepsin secretion from macrophages via IRE1α activation. Cell Rep. 16(11), 2914-2927. doi: 10.1016/j.celrep.2016.08.035

Overall conclusions

Over the last decade, the proteomic analytic hardware equipment involved in identifying proteins from biological samples has reached a level of sophistication where researchers need and want to identify thousands of confident identifications; however, the bioinformatic interpretation and data processing is struggling to keep up and find a better biological understanding from the data generated.

Chinese hamster ovary cells are the principal workhorse for the production of therapeutic recombinant proteins. The ultimate goal is improved production of these therapeutic proteins and this can be achieved by a greater understanding of mechanisms of protein expression. The proteomic analysis of CHO cells in our laboratory routinely results in +5,000 protein identifications generated from +30,000 peptides from a single injection strategy. The integration of proteome and phosphoproteome data improves the overall proteome coverage and gives better insights into the post-translational level of regulation of recombinant CHO cells.

I believe with the advent of big 'omic data continuously being generated, that possibly the biological question initially in mind should drive the application rather than just generating huge amounts of biological data, and also thought should be put into how to integrate the data generated from each new study with existing sources of data and studies. For example, combining datasets from intracellular CHO cell studies looking at protein degradation and protein translation with datasets from CHO host cell protein analysis studies to understand the networks associated with protein expression, product quality and production titres.

Integrating data, however, is difficult especially ifthey are from different information systems, data sources and/or databases so I believe a major focus should be to develop methods to combine these data outputs into a single information system to allow deep data analysis. This software system would need to access to raw data files from multiple outputs, be easy to use and most importantly "open-source" to fully allow for large-scale integration of data.

Although, discovering potential biomarkers in proteomic laboratories is quite common, very few to date have made it into clinical practice. Biomarkers need to be clinically robust in order for them to make it into the clinic. One of the goals of biomarker proteomic research is to identify biomarkers that would stratify patients to the most effective treatment (especially with the advent of personalised medicine). The challenges include setting up a good clinical study, analysing samples by mass spectrometry using the highest possible quality control to produce data that is robust. Proteomic labs have experts in sample preparation, LC-MS/MS and bioinformatic analysis to discover biomarkers using instruments that are expensive, fast, sensitive and capable of measuring very low-abundance proteins in the presence of high-abundance proteins. If the goal to replace antibody-based assays in the hospital with mass spectrometry platforms then the pipeline needs simple/cheap/rapid sample preparation methods, cheap and easy to maintain robust LC-MS instruments that are sensitive enough to identify biomarker panels. I believe this is possible over the next 10 years.