Quantitative Analysis of the Plasma Proteome in Pre-eclampsia

A thesis submitted to the University of Manchester for the degree of PhD in the Faculty of Medical and Human Sciences

2012

Christal Fisher

School of Medicine
Table of Contents

List of Figures ......................................................................................................................... 10
List of Tables ........................................................................................................................... 15
List of Abbreviations ................................................................................................................. 17

Abstract ................................................................................................................................... 19

1 Introduction ............................................................................................................................ 21
  1.1 Pre-eclampsia ....................................................................................................................... 22
     1.1.1 Clinical Syndrome ........................................................................................................ 22
     1.1.2 Definition of Pre-eclampsia ........................................................................................ 23
            1.1.2.1 The ISSHP Guidelines for Diagnosis of Pre-eclampsia ....................................... 23
     1.1.3 Risk Factors for Pre-eclampsia .................................................................................. 24
            1.1.3.1 Health after Pre-eclampsia ............................................................................... 24
     1.1.4 Pathology ..................................................................................................................... 25
            1.1.4.1 Trophoblast Invasion ....................................................................................... 25
            1.1.4.2 Circulating Factors ....................................................................................... 25
            1.1.4.3 Early and Late Onset Pre-eclampsia ............................................................... 26
     1.1.5 Diagnosis ..................................................................................................................... 27
     1.1.6 Prevention, Management and Treatment ...................................................................... 27
     1.1.7 Screening Tests ........................................................................................................... 28
            1.1.7.1 Evaluation of Screening Tests ........................................................................... 31
            1.1.7.2 Screening Test Study Design ............................................................................ 31
            1.1.7.3 The SCOPE Study ......................................................................................... 32
            1.1.7.4 Candidate Predictive Factors ......................................................................... 34
            1.1.7.4.1 Angiogenic Growth Factors ......................................................................... 34
            1.1.7.4.2 The Coagulation System ............................................................................. 35
            1.1.7.4.3 Inflammation .............................................................................................. 36
            1.1.7.4.4 Placental Secretions ..................................................................................... 37
     1.1.8 Development of Personalised Clinical Risk Profiles ..................................................... 38
            1.1.8.1 A Personalised Clinical Risk Profile in Use - Ovarian Cancer Risk of Malignancy
                     Index ..................................................................................................................... 39
     1.1.9 Project Rationale .......................................................................................................... 39
  1.2 Proteomics .......................................................................................................................... 40
    1.2.1 Biomarker Discovery ................................................................................................... 40
1.2.2 Proteomic Methods for Biomarker Discovery ......................................................... 42
  1.2.2.1 Hypothesis Generating Proteomic Methods .................................................. 42
    1.2.2.1.1 Two Dimensional Gel Electrophoresis (2D-Gel) ..................................... 42
    1.2.2.1.2 Global Mass Spectrometry .................................................................. 43
  1.2.2.2 Hypothesis Driven Proteomics Methods ...................................................... 43
    1.2.2.2.1 Antibody Methods ........................................................................ 44
    1.2.2.2.2 Targeted Mass Spectrometry ......................................................... 44
  1.2.3 Plasma Proteomics ......................................................................................... 44
    1.2.3.1 Plasma Complexity .............................................................................. 45
  1.2.4 Application of Proteomics to the Study of Pre-eclampsia ............................. 47
    1.2.4.1 Two Dimensional Gel Electrophoresis and Pre-eclampsia ...................... 47
    1.2.4.2 Mass Spectrometry and Pre-eclampsia ............................................. 48
  1.2.5 Project Rationale – Method .......................................................................... 49

1.3 Mass Spectrometry .......................................................................................... 49
  1.3.1 Components of a Mass Spectrometer ............................................................ 49
    1.3.1.1 Ionisation Source ................................................................................ 50
      1.3.1.1.1 Electrospray Ionisation .................................................................. 50
      1.3.1.1.2 Matrix-Assisted Laser Desorption Ionisation ............................... 50
    1.3.1.2 Mass Analyser ....................................................................................... 52
      1.3.1.2.1 Time-of-Flight Mass Analyser ..................................................... 52
      1.3.1.2.2 Quadrupole Mass Analyser .......................................................... 52
    1.3.1.2.3 Ion Trap Mass Analyser .................................................................. 53
    1.3.1.3 Data Output .......................................................................................... 55
  1.3.2 Application of Mass Spectrometry to Proteomics ........................................... 56
    1.3.2.1 Peptide Mass Fingerprinting .............................................................. 56
    1.3.2.2 Tandem Mass Spectrometry .............................................................. 57
      1.3.2.2.1 MASCOT: An Example of Search Engine Software ..................... 58
    1.3.2.3 Sample Processing for Tandem Mass Spectrometry .............................. 60
      1.3.2.3.1 Immunodepletion of High Abundance Proteins ............................ 60
      1.3.2.3.2 High Performance Liquid Chromatography ................................. 60
    1.3.2.4 Multiple Reaction Monitoring ............................................................. 61
  1.3.3 Quantitative Analysis using Mass Spectrometry .............................................. 63
    1.3.3.1 Stable Isotope Labelling ....................................................................... 63
      1.3.3.1.1 Isobaric Tags for Relative and Absolute Quantification .................. 64
      1.3.3.1.2 Quantitative Analysis by Multiple Reaction Monitoring ............... 65

2 Aims ....................................................................................................................... 66
3 Methods................................................................................................................................. 68

3.1 Patient Samples .................................................................................................................. 68
  3.1.1 Plasma Sample Collection ......................................................................................... 68
  3.1.2 Early Pregnancy Plasma Samples from The SCOPE Study ........................................ 68
  3.1.3 Plasma Samples used for Method Development ........................................................ 70

3.2 Identification of Candidate Markers by iTRAQ .................................................................. 70
  3.2.1 Plasma Pooling ........................................................................................................... 70
  3.2.2 Sample Preparation: IgY 14-SuperMix Immunodepletion Workflow ....................... 70
    3.2.2.1 IgY 14-SuperMix Immunodepletion Workflow Overview .................................. 71
      3.2.2.1.1 IgY-SuperMix Immunodepletion ................................................................. 71
      3.2.2.1.2 Ultrafiltration .............................................................................................. 73
      3.2.2.1.3 Analysis of IgY-SuperMix Processed Samples ............................................. 73
        3.2.2.1.3.1 Coomassie Protein Assay ............................................................... 73
        3.2.2.1.3.2 In-Gel Digest ................................................................. 73
        3.2.2.1.3.3 Zip Tip Cleanup ............................................................................... 74
      3.2.2.1.4 Protein Digestion and iTRAQ Labelling ..................................................... 74
      3.2.2.1.5 High pH Reverse Phase Fractionation (3 µm Fortis C18 Column) ............... 74
  3.2.3 Sample Preparation: MARS 14 Immunodepletion System Workflow .................... 75
    3.2.3.1 MARS 14 Immunodepletion System Overview ................................................ 75
      3.2.3.1.1 MARS 14 Immunodepletion ................................................................. 76
      3.2.3.1.2 Macroporous Reversed-Phase C18 Column .............................................. 77
      3.2.3.1.3 Analysis of MARS 14 Processed Samples .................................................. 78
        3.2.3.1.3.1 Protein Assay ................................................................................... 78
        3.2.3.1.3.2 Protein 230 Assay for Qualitative Assessment .................................... 78
      3.2.3.1.4 Digestion and iTRAQ Labelling ............................................................... 79
      3.2.3.1.5 High pH Reverse Phase Fractionation (3µm Extend-C18 Column) .......... 80
  3.2.4 Mass Spectrometry ....................................................................................................... 81
    3.2.4.1 Analysis of Sample Preparation .......................................................................... 81
      3.2.4.1.1 Ultraflex MALDI TOF-TOF ...................................................................... 81
      3.2.4.1.2 Esquire HCT Ion Trap .............................................................................. 81
    3.2.4.2 Analysis of iTRAQ Labelled Samples .................................................................. 82
      3.2.4.2.1 QStar XL ESI qTOF .............................................................................. 82
      3.2.4.2.2 5800 MALDI TOF-TOF ......................................................................... 83
  3.2.5 Data Analysis ................................................................................................................. 84
    3.2.5.1 MASCOT ............................................................................................................. 84
    3.2.5.2 ProteinPilot (version 3.0) .................................................................................. 84
3.3 Validation of Candidate Markers ................................................................. 85

3.3.1 Multiple Reaction Monitoring ............................................................... 85

3.3.1.1 Preparation of Samples for Multiple Reaction Monitoring ................. 85

3.3.1.1.1 Sample Spiking and Preparation .................................................. 85

3.3.1.1.2 Digestion .................................................................................. 85

3.3.1.1.3 Supelco Analytical Discovery C18 Column .................................. 85

3.3.1.2 Optimisation of Sample Preparation .................................................. 86

3.3.1.2.1 Molecular Weight Spin Columns ............................................... 86

3.3.1.2.2 Sepherose Columns .................................................................. 87

3.3.1.2.3 Supelco Analytical Discovery C18 Column for Quantification .......... 87

3.3.1.2.4 6530 qTOF for Sample Assessment .......................................... 88

3.3.1.3 Multiple Reaction Monitoring Assay ................................................ 88

3.3.1.4 Data Analysis .................................................................................... 88

3.3.2 Validation Using Antibodies .................................................................. 89

3.3.2.1 SDS-PAGE and Western Blotting .................................................... 89

3.3.2.1.1 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis ........ 89

3.3.2.1.2 Coomassie Staining .................................................................. 89

3.3.2.1.3 Western Blot ............................................................................. 89

3.3.2.2 ELISA ............................................................................................ 90

4 Development and Optimisation of an iTRAQ Workflow .................................. 91

4.1 High Abundance Protein Depletion: Comparison of Immunodepletion Methods for Identification of Changed Proteins using iTRAQ ..................................................... 91

4.1.1 Introduction ............................................................................................ 91

4.1.2 Aim ........................................................................................................ 91

4.1.3 Method ................................................................................................... 92

4.1.3.1 The Genway Seppro® IgY-SuperMix Immunodepletion System .......... 92

4.1.3.2 The Agilent Human 14 Multiple Affinity Removal LC Column ............... 92

4.1.3.3 Application to SCOPE Samples ..................................................... 93

4.1.3.4 Comparison of the IgY 14-SuperMix System and the MARS 14 Column .... 93

4.1.4 Results ................................................................................................... 96

4.1.4.1 Optimisation of the IgY 14-SuperMix System .................................... 96

4.1.4.2 Application of the IgY 14-SuperMix System to SCOPE Samples ............. 98

4.1.4.3 Optimisation of the MARS 14 Column ............................................. 99

4.1.4.3.1 Investigation of Injection Volume and Flow-through Peak Area .......... 99

4.1.4.3.2 Application of Plasma from Non Pregnant and Pregnant Women .... 101

4.1.4.4 Optimisation of Assessment of the MARS 14 Column ....................... 102

4.1.4.4.1 Assessment of Depletion using the Flow-through Peak Area .......... 102
4.1.4.4.2 Qualitative Assessment of Immunodepleted Proteins .............................................104
4.1.4.4.3 Optimisation of Injection Volume for iTRAQ labelling ........................................107
4.1.4.5 Application of the MARS 14 Column to SCOPE Samples ......................................109
4.1.4.6 Comparison of Immunodepletion of SCOPE Samples by the IgY 14-SuperMix
System and the MARS 14 Column .....................................................................................112
4.1.4.6.1 Proteome Coverage and Dynamic Range .........................................................112
4.1.4.6.2 Efficiency of High Abundance Protein Removal ...............................................115
4.1.4.6.3 Reproducibility .................................................................................................116
4.1.4.7 Results Summary ..................................................................................................118
4.1.5 Discussion ...............................................................................................................119
4.1.5.1 The Genway Seppro® IgY-SuperMix Immunodepletion System ..............................119
4.1.5.2 The Agilent Human 14 Multiple Affinity Removal LC Column ..............................120
4.1.5.3 Application of Plasma from Non Pregnant and Pregnant Women .........................121
4.1.5.4 Comparison of Immunodepletion of SCOPE Samples by the IgY 14-SuperMix
System and the MARS 14 Column .....................................................................................122
4.1.6 Conclusion ...............................................................................................................125
4.2 Comparison of Performance of an ESI-qTOF and MALDI-TOF-TOF in Analysis of iTRAQ
Labelled Plasma ..............................................................................................................126
4.2.1 Introduction .............................................................................................................126
4.2.2 Aim .........................................................................................................................126
4.2.3 Method ....................................................................................................................127
4.2.3.1 Sample Preparation ............................................................................................127
4.2.3.2 Data Analysis .....................................................................................................127
4.2.3.3 Comparative Analysis .......................................................................................129
4.2.4 Results ....................................................................................................................131
4.2.4.1 Practical Considerations .....................................................................................131
4.2.4.2 Peptide Identifications .....................................................................................131
4.2.4.3 Peptide Characterisation ....................................................................................133
4.2.4.3.1 Lysine and Arginine Distribution ................................................................136
4.2.4.4 Protein Identifications .......................................................................................137
4.2.4.5 Analysis of Relative Quantification Data ................................................................140
4.2.4.5.1 Comparison of Relative Quantification between Methods ..............................140
4.2.4.5.2 Analysis of Repeated Measurement of Relative Quantification .........................142
4.2.4.6 Results Summary ................................................................................................142
4.2.5 Discussion ...............................................................................................................145
4.2.5.1 Practical Considerations .....................................................................................145
4.2.5.2 Peptide and Protein Identifications .....................................................................145
4.2.5.3 Peptide Characterisation .....................................................................................148
4.2.5.4 Analysis of Relative Quantification Data ..........................................................150
4.2.6 Conclusion ............................................................................................................152

4.3 Evaluation of Two Optimised Workflows for Relative Quantification of iTRAQ Labelled Plasma .............................................................................................................153
4.3.1 Introduction ..........................................................................................................153
4.3.2 Aim .......................................................................................................................153
4.3.3 Methods ...............................................................................................................154
4.3.3.1 Preparation of Samples for Relative Quantification by iTRAQ ....................154
4.3.3.2 Comparison of Protein Identification and Relative Quantification Data .......154
4.3.4 Results ..................................................................................................................156
4.3.4.1 Preparation of Samples for Relative Quantification by iTRAQ ....................156
4.3.4.2 Comparison of Protein Identification and Relative Quantification Data .......158
4.3.4.2.1 Assessment of Digestion and Labelling with iTRAQ 8-plex reagent ........158
4.3.4.2.2 High Confidence Protein Identifications ...................................................160
4.3.4.2.3 Assessment of iTRAQ Labelling ..............................................................162
4.3.4.2.4 Assessment of iTRAQ Relative Quantification ........................................163
4.3.4.3 Summary of Samples for Analysis .................................................................165
4.3.5 Discussion ..........................................................................................................166
4.3.5.1 Preparation of Samples for Relative Quantification by iTRAQ ....................166
4.3.5.2 Protein Identifications and Relative Quantification by iTRAQ ....................167
4.3.6 Conclusion ..........................................................................................................171

5 Identification and Validation of Changed Plasma Proteins in Early Pregnancy in Women Who Develop Pre-eclampsia ...............................................................................................172
5.1 Identification of Changed Plasma Proteins at 15 Weeks Gestation in Women Who Develop Pre-eclampsia Using iTRAQ ..................................................................................172
5.1.1 Introduction ......................................................................................................172
5.1.2 Aim ....................................................................................................................172
5.1.3 Method ..............................................................................................................173
5.1.3.1 Samples for Relative Quantification ............................................................173
5.1.3.2 Identification of Proteins Changed in Abundance .......................................173
5.1.4 Results ..............................................................................................................176
5.1.4.1 Data Analysis Stage One - Proteins with Reproducible Quantification .......176
5.1.4.2 Data Analysis Stage Two - Proteins Changed in Abundance .......................179
5.1.4.3 Proteins Changed in Abundance between Pre-eclampsia and Control ........183
5.1.4.3.1 Proteins Changed in Abundance in Early Onset Pre-eclampsia ...............183
5.1.4.3.2 Proteins Changed in Abundance in Late Onset Pre-eclampsia .................190
5.1.4.3.3 Proteins Changed in Abundance in Late and Early Onset Pre-eclampsia..194
5.3.3.3 Western Blots ........................................................................................................ 233
5.3.3.4 ELISA .................................................................................................................. 233
5.3.4 Results ................................................................................................................... 235
  5.3.4.1 Optimisation of Western Blotting ................................................................. 235
  5.3.4.1.1 Assessment of Reproducibility of Gel Loading ....................................... 235
  5.3.4.1.2 Antibody Specificity - PSG9 ................................................................. 235
  5.3.4.2 Validation by Western Blot ............................................................................ 238
  5.3.4.3 Validation by ELISA ...................................................................................... 238
  5.3.4.4 Summary of Western Blot and ELISA Validation of iTRAQ Results ........ 241
5.3.5 Discussion ............................................................................................................. 242
  5.3.5.1 Validation using Antibodies ............................................................................ 242
  5.3.5.1.1 Western Blotting ..................................................................................... 244
  5.3.5.1.2 ELISA ..................................................................................................... 246
  5.3.5.2 Validation of iTRAQ Discovery Phase ....................................................... 247
5.3.6 Conclusion ............................................................................................................. 249

5.4 Identification and Validation of Changed Plasma Proteins in Early Pregnancy in Women
   Who Develop Pre-eclampsia: Summary ................................................................. 250

6 Discussion ............................................................................................................... 254
  6.1 Future Work .......................................................................................................... 260
  6.2 Summary ............................................................................................................... 260

References .................................................................................................................. 261

Appendix : Glossary of ProteinPilot (version 3.0) terms ............................................. 281

Word Count: 88285
List of Figures

Figure 1.1.1 The basis for the development of the maternal syndrome of pre-eclampsia is established long before the clinical presentation of the condition (black).................................33

Figure 1.2.1 The dynamic range of plasma spans a $10^{12}$ fold concentration range and exceeds the dynamic range of most analytical methods.........................................................46

Figure 1.3.1 Components of a Mass Spectrometer. A mass spectrometer consists of a sample inlet, an ion source, a mass analyser and an ion detector.................................49

Figure 1.3.2 Gas phase ions are produced by the ion source..................................................51

Figure 1.3.3 The mass analyser separates ions by $m/z$ ratio enabling detection.......................54

Figure 1.3.4 A mass spectrum is produced by processing software plotting the $m/z$ against the intensity of the signal detected at each $m/z$. .........................................................55

Figure 1.3.5 Tandem mass spectrometry can be used to identify proteins from complex mixtures.......................................................................................................................59

Figure 1.3.6 Multiple reaction monitoring is a highly sensitive targeted method for identification of proteins using a triple quadrupole mass spectrometer........62

Figure 1.3.7 Isobaric Tags for Relative and Absolute Quantification (iTRAQ) consist of a reporter group, a balance group and an amine specific peptide reactive group. 66

Figure 3.2.1 Agilent Protein 200 Plus ladder (Agilent, UK) in both electropherogram and gel view were used to assess the quality of each assay performed........................................79

Figure 4.1.1 Pooled plasma samples were immunodepleted using the Genway Seppro* IgY 14-SuperMix Liquid Chromatography Column System or the Agilent Multiple Affinity Removal Liquid Chromatography Column (Human 14), labelled with iTRAQ reagent and analysed by LC-MS/MS using a 5800 MALDI–TOF-TOF. ............................95

Figure 4.1.2 Assessment of replicate plasma samples depleted using the IgY 14-SuperMix System by SDS-PAGE with Coomassie Staining showed depletion to be variable between replicates..................................................................................................................97

Figure 4.1.3 Immunodepletion of pooled plasma samples acquired from the SCOPE study using the IgY 14-SuperMix System was variable.................................................................98

Figure 4.1.4 Depletion of plasma using the MARS 14 System produces a characteristic peak pattern and the area under the flow-through peak is proportional to the volume of plasma applied to the column.................................................................100

Figure 4.1.5 Peak area for the flow-through fraction is smaller when depleting plasma from a non pregnant woman than from a pregnant woman using the MARS 14 System.................................................................101

Figure 4.1.6 Plasma incompletely depleted using a compromised MARS 14 Column could be differentiated from plasma completely depleted using a working MARS 14 Column using the peak area of the flow-through fraction. .................................103
Figure 4.1.7  Assessment of depleted plasma by the Protein 230 Assay is reproducible in both gel and electropherogram mode and application of whole plasma, bound fraction containing high abundance proteins and albumin produces a characteristic band and peak at approximately 60kDa.  

Figure 4.1.8  Plasma incompletely depleted using a compromised MARS 14 Column could be differentiated from plasma completely depleted using an alternative MARS 14 Column using the peak at 60kDa following Protein 230 Assay.  

Figure 4.1.9  Qualitative assessment of immunodepletion of pooled SCOPE samples showed depletion to be reproducible and complete.  

Figure 4.1.10  Peak area for the flow-through fraction of immunodepleted SCOPE samples were reproducible and indicated complete depletion.  

Figure 4.1.11  More lower abundance proteins were identified from LC-MS/MS analysis of sample depleted using the IgY 14-SuperMix System than the MARS 14 System.  

Figure 4.1.12  More proteins were identified at lower concentration in sample immunodepleted using the IgY 14-SuperMix System than the MARS 14 system.  

Figure 4.1.13  Immunodepletion by the IgY 14-SuperMix System is less reproducible and specific than immunodepletion by the MARS 14 Column.  

Figure 4.2.1  Samples were prepared by labelling with 8-plex iTRAQ reagent and LC-MS/MS analysis was carried out using a QStar XL qTOF and 5800 MALDI-TOF-TOF.  

Figure 4.2.2  The QSTAR identified more spectra and made more peptide spectral matches than the 5800 but identified fewer unique peptides.  

Figure 4.2.3  Peptides identified by the QSTAR only, had a larger average molecular weight and lower m/z ratio than those identified by the 5800 only.  

Figure 4.2.4  Peptides identified by the QSTAR only had an average lower isoelectric point and higher relative hydrophobicity than those identified by the 5800 only.  

Figure 4.2.5  The QSTAR identified a higher proportion of peptides with lysine at the carboxyl terminus and 5800 a higher proportion of peptides with arginine at the carboxyl terminus. Plasma labelled using iTRAQ 8-plex reagent was analysed using both QSTAR and 5800 LC-MS/MS systems.  

Figure 4.2.6  The QSTAR identified a similar number of proteins at high confidence, less with 3+ peptides and a similar number of unique proteins when compared to the 5800.  

Figure 4.2.7  The number of peptides per protein and UPS were both lower for proteins identified by the QSTAR compared to proteins identified by the 5800.  

Figure 4.2.8  Relative quantification ratios calculated for peptides (>95% confidence) used for quantification and proteins (high confidence) identified by both the QSTAR and 5800 were measured equally by both methods.
Figure 4.2.9 The Bland Altman Method of Agreement showed ratios obtained using the 5800 were greater than those from the QSTAR at the extremes of the range of the ratios.

Figure 4.3.1 Pooled plasma samples acquired from the SCOPE Study were immunodepleted, labelled with iTRAQ 8-plex reagent and analysed by LC-MS/MS.

Figure 4.3.2 High pH reverse phase chromatography was optimised using a cytochrome C digest and used to separate depleted pooled plasma samples labelled using iTRAQ 8-plex reagent.

Figure 4.3.3 A total of 502 proteins were identified with high confidence, with 113 proteins identified in all three datasets.

Figure 4.3.4 Assessment of processing of duplicate samples confirmed early onset pre-eclampsia and control samples were reproducibly processed in the IgY 14-SuperMix datasets and superpool, early and late onset pre-eclampsia samples were reproducibly processed in the MARS 14-5800 dataset.

Figure 5.1.1 Stage one of the data analysis method - identification of proteins with reproducible quantification between duplicate samples.

Figure 5.1.2 Stage two of the data analysis method identified proteins which changed in abundance between pre-eclampsia and control samples greater than between the two control samples.

Figure 5.1.3 Transformed relative quantification ratios for proteins identified in duplicate samples were plotted and considered reproducibly quantified when between -0.35 and 0.35.

Figure 5.1.4 Proteins where the difference between early onset pre-eclampsia and control samples fell outside the range of biological variation were considered changed.

Figure 5.1.5 Changes in plasma proteins in women who develop pre-eclampsia were identified using a two stage analysis.

Figure 5.1.6 A total of 113 proteins were identified as changed between early onset pre-eclampsia and control samples in all data.

Figure 5.1.7 Proteins where the difference between late onset pre-eclampsia and the control sample iTRAQ relative quantification ratios was outside the range of biological variation were considered changed.

Figure 5.1.8 A total of 24 proteins were identified as changed in abundance in both late and early onset pre-eclampsia in the MARS 14-5800 dataset.

Figure 5.1.9 Proteins changed in abundance between early onset pre-eclampsia and control for validation.

Figure 5.1.10 Filamin, Pregnancy-specific beta-1-glycoprotein 9, Talin and Vinculin were all increased in early onset pre-eclampsia compared to controls, Fibulin was decreased.
Enolase peptides were used to calculate a ‘normalisation factor’ to correct for variation in sample processing and LC-MS/MS analysis.

Non human proteins were spiked into plasma and quantified by MRM. Proteins were spiked into plasma alongside enolase.

The CVs of peak areas for enolase peptides used for normalisation were between 9 and 21%.

The CV for peak areas of peptides of spiked proteins normalised against peak areas for enolase were calculated.

Normalised peak areas for peptides assayed by MRM were compared to the concentration of spiked protein in each sample set.

The CV of spiked proteins decreased as the concentration of protein spiked increased.

Peptides of proteins of interest were selected and retention times and a minimum of three transitions optimised.

Quantification of enolase peptides was more variable in early onset pre-eclampsia and replicate samples than in control samples.

The CV of the peak area for peptides assayed by MRM in replicated samples increased on normalisation against summed peak areas of spiked enolase peptides.

Peak areas for peptides of PSG9 and platelet basic protein indicated an increase in abundance in early onset pre-eclampsia compared to control.

The MRM sample preparation workflow was optimised to improve reproducibility of samples processing.

Samples prepared using a MWCO were digested more efficiently, assayed more consistently and resulted in more protein identifications than the sample prepared using the mRP column.

The CV of the peak area for peptides assayed by MRM in 80 individual assays was consistent and fell below 10% on normalisation.

NAP-2 was used as a surrogate for platelet basic protein. Antibodies for platelet basic protein were not available.

Ponceau S staining of fibulin blots shows loading of plasma is consistent. Plasma proteins were separated by SDS-PAGE gel and target proteins identified by western blotting as previously described.

Specificity of the PSG9 antibody was questionable as signal was also detected on analysis of plasma from healthy non-pregnant individuals.

Plasma levels of platelet basic protein were increased in early onset pre-eclampsia samples compared to control samples by western blotting.
Figure 5.3.5  Plasma levels of platelet basic protein were increased in early onset pre-eclampsia samples compared to control samples using commercially available ELISA assays.

Figure 5.4.1  Summary of platelet basic protein abundance in early onset pre-eclampsia and control samples calculated by iTRAQ, relative MRM, western blot and ELISA. ....252
## List of Tables

| Table 1.1 | Screening Tests for Pre-eclampsia. Summary of proposed clinical, biophysical and biochemical tests for the prediction of pre-eclampsia (Adapted from Conde-Agudelo et al. 2004). | 30 |
| Table 1.2 | The biomarker discovery pipeline. Proteins of interest are identified during discovery phase which quantifies large numbers of potential biomarkers in a small number of samples. | 41 |
| Table 3.1 | Maternal characteristics and pregnancy outcome for women contributing plasma to the SCOPE study. | 69 |
| Table 3.2 | Composition of Buffers for Immunodepletion of plasma samples using the Seppro IgY-SuperMix Liquid Chromatography Column system. | 72 |
| Table 3.3 | Immunodepletion of plasma samples using the Seppro IgY-SuperMix Liquid Chromatography Column system. | 72 |
| Table 3.4 | Immunodepletion of plasma samples using the MARS 14 column. Pooled plasma samples as previously described were depleted of high abundance proteins. | 76 |
| Table 3.5 | Concentration and desalting of samples using a Macroporous Reversed-Phase C18 Column. | 77 |
| Table 3.6 | A non linear gradient was applied to a 3µm Extend-C18 Column to separate peptides by Reverse Phase Fractionation. | 80 |
| Table 3.7 | Sample preparation was assessed using an Esquire HCT Ion Trap. | 81 |
| Table 3.8 | Analysis of iTRAQ labelled samples using an Applied Biosystems QStar XL qTOF. | 82 |
| Table 3.9 | A non-linear gradient was applied to an Acclaim PepMap 100 C18 column to separate peptides prior to spotting with HCCA matrix and analysis using an Applied Biosystems 5800 MALDI TOF-TOF. | 83 |
| Table 3.10 | Desalting of samples using a Sulcelco Analytical Discovery C18 Column. | 86 |
| Table 3.11 | Quantification of peptide content using a Sulcelco Analytical Discovery C18 Column. | 87 |
| Table 4.1 | Immunodepletion of 25 µl of plasma from a pregnant woman by the MARS 14 System is ideal for iTRAQ labelling. | 108 |
| Table 4.2 | More of the proteins which should have been depleted were identified in plasma depleted using the IgY 14-SuperMix System than the MARS 14 System. | 115 |
| Table 4.3 | Summary of a performance comparison between the IgY 14-SuperMix System and MARS 14 System in the depletion of human plasma. | 118 |
| Table 4.4 | Summary of performance comparison between an ESI-qTOF and MALDI-TOF-TOF in the Analysis of an 8-plex iTRAQ Labelled Plasma Sample Set. | 144 |
Table 4.3.1  The rate of missed cleavage and iTRAQ labelling of lysine residues indicated digestion efficiency and labeling by iTRAQ reagent was similar for both IgY 14-SuperMix and MARS 14 sample preparation methods. ........................................159

Table 4.3.2  The bias correction factor calculated by ProteinPilot for the 119:113 (control: superpool) iTRAQ relative quantification ratio for the MARS 14-5800 dataset was low indicating the ratio was unreliable and unfit for use in further data analysis. ........................................................................................................162

Table 4.3.3  Samples available for analysis following immunodepletion, iTRAQ labelling, LC-MS/MS analysis and data processing are shown in colour, samples unavailable for analysis due to sample processing issues are shown in grey. ......................................................165

Table 5.1.1  Relative quantification ratios were compared for proteins in duplicate samples to identify proteins with reproducible quantification. .................................................................177

Table 5.1.2  Control samples were used to define the range of biological variation to identify proteins changed in abundance. ....................................................................................179

Table 5.1.3  A total of 113 proteins were identified as changed between early onset pre-eclampsia and control samples. ........................................................................................................185

Table 5.1.4  A total of 32 proteins were identified as changed between late onset pre-eclampsia and the control sample.................................................................192

Table 5.2.1  Yeast enolase was spiked into each sample at 10 µg/m of plasma. Four enolase peptides were quantified in each sample.........................................................205

Table 5.3.1  Summary of conditions used for SDS-PAGE gels and antibodies for Western Blots. Plasma was diluted 1:10 in ddH₂O and then 1:1 in Lamelli buffer... .......................234

Table 5.3.2  Summary of changes in plasma proteins identified in early onset pre-eclampsia samples by iTRAQ and validated by western blot and ELISA.................................241

Table 5.4.1  Summary of relative quantification ratios calculated for expression of platelet basic protein in early onset pre-eclampsia and control samples by iTRAQ, relative MRM, western blot and ELISA.................................................................253
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-Lac</td>
<td>β-Lactoglobulin</td>
</tr>
<tr>
<td>2D-Gel</td>
<td>Two Dimensional Gel Electrophoresis</td>
</tr>
<tr>
<td>S800</td>
<td>S800 MALDI–TOF-TOF</td>
</tr>
<tr>
<td>ACN</td>
<td>Acetonitrile</td>
</tr>
<tr>
<td>ADH</td>
<td>Alcohol Dehydrogenase</td>
</tr>
<tr>
<td>AmBic</td>
<td>Ammonium bicarbonate</td>
</tr>
<tr>
<td>Apo</td>
<td>Apolipoprotein</td>
</tr>
<tr>
<td>ASSHP</td>
<td>Australasian Society for the Study of Hypertension in Pregnancy</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index</td>
</tr>
<tr>
<td>CPEP</td>
<td>The Calcium for Pre-eclampsia Prevention Trial</td>
</tr>
<tr>
<td>CRF</td>
<td>Corticotrophin-releasing factor</td>
</tr>
<tr>
<td>CRF-BP</td>
<td>CRF-binding protein</td>
</tr>
<tr>
<td>CRP</td>
<td>C Reactive protein</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of Variation</td>
</tr>
<tr>
<td>DIGE</td>
<td>Difference Gel Electrophoresis</td>
</tr>
<tr>
<td>EIC</td>
<td>Extracted Ion Chromatogram</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-Linked Immunosorbent Assay</td>
</tr>
<tr>
<td>FA</td>
<td>Formic Acid</td>
</tr>
<tr>
<td>FDR</td>
<td>False Discovery Rate</td>
</tr>
<tr>
<td>Flt-1</td>
<td>fms-like Tyrosine Kinase 1</td>
</tr>
<tr>
<td>HAP</td>
<td>High Abundance Proteins</td>
</tr>
<tr>
<td>hCG</td>
<td>human Chorionic Gonadotropin</td>
</tr>
<tr>
<td>HDL</td>
<td>High Density Lipoprotein</td>
</tr>
<tr>
<td>HELLP</td>
<td>Haemolysis, Elevated Liver enzymes and Low Platelets</td>
</tr>
<tr>
<td>HPLC</td>
<td>High-performance liquid chromatography</td>
</tr>
<tr>
<td>Ig</td>
<td>Immunoglobulin</td>
</tr>
<tr>
<td>IgY 14-SuperMix System</td>
<td>The Seppro® IgY 14-SuperMix Liquid Chromatography Column System</td>
</tr>
<tr>
<td>Immunodepletion</td>
<td>Immunoaffinity-based protein subtraction chromatography</td>
</tr>
<tr>
<td>ISSHP</td>
<td>The International Society for the Study of Hypertension in Pregnancy</td>
</tr>
<tr>
<td>iTRAQ</td>
<td>Isobaric Tags for Relative and Absolute Quantification</td>
</tr>
<tr>
<td>IUGR</td>
<td>Intrauterine growth restriction</td>
</tr>
<tr>
<td>LDL</td>
<td>Low Density Lipoprotein</td>
</tr>
<tr>
<td>MARS 14</td>
<td>Multiple Affinity Removal LC Column-Human 14</td>
</tr>
<tr>
<td>MDH</td>
<td>Malate Dehydrogenase</td>
</tr>
<tr>
<td>MIAPE</td>
<td>Minimum Information About A Proteomics Experiment</td>
</tr>
<tr>
<td>MRM</td>
<td>Multiple Reaction monitoring</td>
</tr>
</tbody>
</table>
MS  Mass spectrometry
MWCO  Molecular Weight Cut Off
NAP2  Neutrophil-Activating Protein-2
NHBPPEP  National High Blood Pressure Education Program
NICE  National Institute for Health and Clinical Excellence
PAI  Plasminogen Activator Inhibitor
PAPP-A  Pregnancy-Associated Plasma Protein A
PARIS  The Perinatal Antiplatelet Review of International Studies Collaboration
PIGF  Placental Growth Factor
PSG9  Pregnancy-specific beta-1-glycoprotein 9
QQQ  Triple Quadrupole
QSTAR  QStar XL qTOF
QTOF  Quadrupole-Time-of-flight
RMI  Risk of Malignancy Index
RP  Reverse phase
SCOPE  SCreening fOr Pregnancy Endpoints
SD  Standard Deviation
SDS  Sodium Dodecyl Sulphate
SDS-PAGE  Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis
SELDI  Surfaced Enhanced Laser Desorption and Ionisation
sEng  Soluble Endoglin
SERPINA  Serine-type endopeptidase inhibitor activity
sFlt-1  Soluble Flt-1
SGA  Small for Gestational Age
SICAPA  Stable Isotope Standards and Capture by Anti-Peptide Antibodies
SILAC  Stable Isotope Labelling in Culture
SPE  Solid Phase Extraction
TCEP  Tris (2-carboxyethyl) Phosphine
TEAB  Triethylammonium Bicarbonate
TFA  Trifluoroacetic Acid
TGF  Transforming growth factor
tPA  Tissue Plasminogen Activator
UPS  Unused Protein Score
UV  UltraViolet
VEGF  Vascular Endothelial Growth Factor
VLDL  Very Low Density Lipoprotein
ABSTRACT OF THESIS submitted by Christal Fisher for the degree of PhD and entitled Quantitative Analysis of the Plasma Proteome in Pre-eclampsia. Submitted February 2012.

There is currently no clinically useful screening test available to identify nulliparous women at high risk of developing pre-eclampsia. This study aimed to identify novel biomarkers using hypothesis generating proteomic methods applied to plasma samples obtained prior to clinical diagnosis of pre-eclampsia.

Plasma samples taken at 15 weeks gestation from women who subsequently developed late pre-eclampsia (>34 weeks), early pre-eclampsia (<34 weeks) and two distinct groups of women with uncomplicated pregnancies (each n=12) were pooled. Pooled plasma was immunodepleted, labelled using iTRAQ-8plex reagent and separated into fractions using high pH reverse phase chromatography. Fractions were analysed by LC-MS/MS and data interrogated using ProteinPilot 3.0.

The merits of two immunodepletion systems were compared; the Seppro® IgY 14-SuperMix LC column system removes up to 100 highly abundant plasma proteins and the Multiple Affinity Removal LC column depletes 14 highly abundant plasma proteins. Removal of more high abundance proteins allowed identification of more, potentially interesting, low abundance proteins, but was less reproducible than removing fewer proteins. Two methods of LC-MS/MS analysis were assessed; the QStar XL qTOF and 5800 MALDI–TOF-TOF. The protein identifications and the quantification data acquired by each method was comparable and complementary and increased the total number of proteins identified.

A total of 502 proteins were identified. A stringent two stage analysis was developed to identify candidate proteins which changed in abundance in plasma from women who later developed pre-eclampsia compared to women with uncomplicated pregnancies. Analysis identified a total of 113 proteins which were both reproducibly quantified and changed by more than the expected range of biological variation. Six candidate proteins changed in abundance in the plasma taken from women who subsequently developed early pre-eclampsia were selected for further validation. A high throughput, low cost, method of multiple reaction monitoring which allows relative quantitation without the use of costly isotopically labelled peptides was developed to validate candidate proteins. Candidate proteins were also assessed by western blot and ELISA. Only one candidate protein; platelet basic protein, was validated by all three methods and demonstrated similar increases in the abundance.

This investigation suggests that measurement of platelet basic protein at 15 weeks gestation is a novel candidate predictive marker for pre-eclampsia. Validation of platelet basic protein in a large, independent, sample set is required to confirm changes in protein expression and to evaluate potential, alongside other factors, to identify nulliparous women at high risk of developing pre-eclampsia later in pregnancy.
Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/
Acknowledgments

This work has in no way been a solo effort and there are so many people to whom I owe an enormous debt of gratitude. Firstly I would like to thank my supervisors Dr Jenny Myers and Dr Melissa Westwood for their steadfast support, guidance and encouragement throughout this work. I would also like to thank Professor Philip Baker for giving me this opportunity and for setting me on my way. I would also like to thank all those I have worked with at the Maternal and Fetal Health Research Group for their help and support.

Without Dr Richard Blankley much of this work would not exist. His unfailing help and support has spanned several guises over the past few years and has been invaluable. I am incredibly grateful to Richard, not least for a wicked sense of humour and a steady supply of biscuits.

I would also like to express my gratitude to Dr Richard Unwin for his generosity with both his time and extensive expertise, from a first introduction to mass spectrometry to viva advice and a healthy dose of common sense along the way. I must also thank the group at CADET for letting me ‘hang around’ and putting up with more than one writing tantrum, but most of all Dr Katherine Hollywood for keeping me on the straight and narrow.

I would also like to show my appreciation to my advisor Professor Tony Whetton and in particular Dr Michael Walker and Dr Andrew Williamson for the generous use of the QSTAR and 5800. I must also thank Tony and his research group for allowing me to participate in lab meetings and allowing me to present my work. The opportunity to be involved with a proteomics group, and the guidance and feedback provided has been instrumental in the development of this project.

I must also thank those from The SCOPE Consortium for the provision of the samples which formed the basis of this work. A particular mention must go to the 48 women who provided these samples who have been my constant companions over the past 4 year but whom I know so little about.
1 Introduction

1.1 Pre-eclampsia

Pre-eclampsia is a multi-systemic disorder of human pregnancy associated with significant maternal and perinatal morbidity and mortality. Characterised by hypertension and proteinuria, the pathophysiology of the condition is poorly understood. Worldwide, 2-8% of pregnancies are affected by pre-eclampsia (Duley 2009); 3-5% in developed countries (Hernández-Díaz S et al. 2009). In the UK pre-eclampsia is a leading cause of maternal death and stillbirth (CMACE 2011a; CMACE 2011b) and as delivery of the infant is the only effective intervention, a significant number of premature births. Pre-eclampsia and associated prematurity has implications for the health of the child (Saigal and Doyle 2008; Wu, Nohr et al. 2009). Maternal health in later life is also at risk with pre-eclampsia linked to an increased incidence of cardiovascular disease and associated complications (Irgens et al. 2001; Lampinen et al. 2006).

1.1.1 Clinical Syndrome

Pre-eclampsia is a multi-systemic disorder thought to originate following abnormal placentation. The development of the compromised placenta leads to a cascade of events in multiple maternal systems. The effects are manifold and result in a wide range of maternal signs and symptoms. Pre-eclampsia is characterised by proteinuria and hypertension, however other less reliable signs including oedema, rapid weight gain and a small for gestational infant may also be present (Davey and Macgillivray 1985; Milne et al. 2005). Symptoms may include epigastric pain and vomiting, frontal headache, visual disturbance, and reduced fetal movement (Davey and Macgillivray 1985; Milne et al. 2005). These signs and symptoms may be the first indication of the developing condition.

The maternal syndrome is thought to represent ‘end organ damage’; the later stages in the cascade of events give rise to the outward signals of the developing condition. Complications arising from pre-eclampsia include acute renal failure, pulmonary oedema, liver failure, HELLP (haemolysis, elevated liver enzymes and low platelets) syndrome and development of eclampsia (maternal seizures associated with the signs of pre-eclampsia)(Sibai et al. 2005). Neonatal complications include fetal growth restriction and neurologic injury as a result of hypoxia. Complications may also arise as a result of preterm delivery, with the magnitude of the risk dependent on gestational age at delivery (Sibai 2006).
1.1.2 Definition of Pre-eclampsia

Although characteristic of pre-eclampsia, the presence of hypertension and proteinuria during pregnancy may also indicate a number of underlying pregnancy complications. In order to define the hypertensive disorders of pregnancy Davey and Macgillivray developed diagnostic criteria using hypertension and proteinuria as defining characteristics (Davey and Macgillivray 1985). Classification of the disorders differentiates between pre-existing chronic hypertension and/or renal disease and hypertension and/or proteinuria developed during pregnancy in a previously healthy woman. Based on this classification, several groups have developed guidelines for diagnosis of pre-eclampsia including the National High Blood Pressure Education Program (NHBPEP), Australasian Society for the Study of Hypertension in Pregnancy (ASSHP) and The International Society for the Study of Hypertension in Pregnancy (ISSHP) (Brown et al. 2001). The differing definitions of pre-eclampsia reflect the multi-systemic nature and difficulty in identifying the condition.

1.1.2.1 The ISSHP Guidelines for Diagnosis of Pre-eclampsia

The ISSHP guidelines (Brown et al. 2001) for diagnosis of pre-eclampsia used in this work are summarised as follows;

- Blood pressure of 140/90 mm Hg occurring after the 20th week of pregnancy in a previously normotensive woman.

- Proteinuria – excretion of more than 300mg per day.

- Resolution by 6 weeks postpartum.
1.1.3 Risk Factors for Pre-eclampsia

There are multiple factors which indicate a greater risk of developing pre-eclampsia including; age, parity, multiple pregnancy, body mass index, duration of sexual relationship, previous pre-eclampsia or a family history of pre-eclampsia or coronary heart disease (Duckitt and Harrington 2005; Kho et al. 2009; North et al. 2010). Pre-existing medical conditions including hypertension or renal disease, insulin dependent diabetes, chronic autoimmune disease and antiphospholipid syndrome are all associated with an increased risk of developing pre-eclampsia. Nulliparous women are particularly at risk as a previous unaffected pregnancy has been shown to be protective in future pregnancies (Bhattacharya et al. 2009). Conversely, the biggest indicator of risk in developing pre-eclampsia is a previous pregnancy complicated by pre-eclampsia (Bhattacharya et al. 2009). While parity is protective of pre-eclampsia in future pregnancies, risk can be ‘reset’ where there is a long inter-pregnancy interval, change of partner (Trogstad et al. 2001; Skjærvén et al. 2002) or increase in body mass index (Getahun et al. 2007). Nulliparous women represent the largest group at risk of developing pre-eclampsia, however obstetric history which may indicate risk status, is not applicable. Accurately identifying otherwise healthy nulliparous women who will develop pre-eclampsia during pregnancy is not possible at present.

1.1.3.1 Health after Pre-eclampsia

The health of both mother and child is also at risk following a pregnancy complicated by pre-eclampsia. Impaired vasodilatation and increased levels of C reactive protein were both observed in women several years after a pregnancy complicated by pre-eclampsia (Lampinen et al. 2006; Hubel et al. 2008) indicating a predisposition for cardiovascular disease. An increased incidence of maternal cardiovascular disease (ischemic heart disease; Smith et al. 2001; hypertension; Bellamy et al. 2007), end stage renal disease (Vikse et al. 2008) and diabetes (Callaway et al. 2007; Libby et al. 2007) have been observed.

It has also been suggested that children born of a pregnancy complicated by pre-eclampsia may also be at greater risk of cardiovascular disease (Oglaend et al. 2009) and diabetes (Libby et al. 2007). Health risks in later life for the child are more a factor of premature birth and low birth weight associated with pre-eclampsia (Hansen et al. 2009). Being born early and/or at low weight both carry inherent risks for the child’s health in later life (Hack et al. 2002). There is evidence however, that children born at term of pregnancies complicated by pre-eclampsia may also be at risk of health complications. Wu et al. demonstrated a child born at term of pregnancy complicated by pre-eclampsia is at increased risk of being hospitalised for a variety of conditions during the first 25 years of life (Wu et al. 2009).
1.1.4 Pathology

Pre-eclampsia only occurs during pregnancy. The condition is a response to the presence of the placenta as it can also occur in molar pregnancies where there is a placenta but no fetus (Chun et al. 1964). Pre-eclampsia resolves postpartum (as stated in the diagnostic criteria) however the syndrome will continue if not all of the placenta is removed (Piering et al. 1993).

The exact mechanism by which the condition develops has yet to be established. Several causes have been suggested including excessive maternal inflammatory response to pregnancy (Redman et al. 1999; Sargent et al. 2006), maternal genetic susceptibility (Moses et al. 2000), hypoxia resulting from breathing disorders (Jerath et al. 2009) and an extension of an inflammatory response to periodontal disease (Kunnen et al. 2010). There are a variety of hypothesis regarding the aetiology of pre-eclampsia. The most widely accepted theory states that deficient trophoblast invasion leads to reduced placental perfusion and ultimately maternal vascular endothelial dysfunction (Redman 1991).

1.1.4.1 Trophoblast Invasion

During the first 20 weeks of gestation, the maternal uterine circulation is dramatically altered. Extravillous trophoblast cells invade and remodel the maternal myometrial spiral arteries. The arteries are initially plugged, limiting blood flow to the placenta, and then transformed into slow flowing flaccid vessels. Invasion and remodelling is incomplete in pregnancies complicated by pre-eclampsia (Naicker et al. 2003). Abnormal placental perfusion is thought to be the first stage in the development of the maternal syndrome of pre-eclampsia.

1.1.4.2 Circulating Factors

The link between abnormal placentation and the maternal syndrome is not yet fully understood. The maternal syndrome is thought to occur as a result of activation of the maternal endothelium. Unlike other types of hypertension in pregnancy, renal biopsies have shown that glomerular endotheliosis is present in up to 80% of women with pre-eclampsia as a result of activation of the endothelium and resolves following delivery (Spargo et al. 1959; Redman et al. 1976). The accepted hypothesis, suggested by Roberts et al. is that the compromised poorly perfused placenta releases a factor into the maternal circulation which activates the endothelium (Roberts et al. 1989).
Investigation of the endothelial response in pre-eclampsia has been carried out using small vessel wire myography. An abnormal endothelial cell function was demonstrated in small arteries dissected from the subcutaneous fat layer of women with pre-eclampsia compared to healthy pregnant and non-pregnant women (McCarthy et al. 1993; Knock and Poston 1996). The presence of a ‘circulating factor’ was confirmed by application of plasma from a woman with pre-eclampsia to arteries from a healthy pregnant woman; endothelium-dependent relaxation was impaired (Ashworth et al. 1998; Hayman et al. 2000). The effect was lost on removal of the vascular endothelium demonstrating the factor affects only the endothelium and not the smooth muscle (Hayman et al. 2000). Several studies have attempted to characterise the ‘factor’ using myography. Hayman et al. concluded the factor was a high molecular weight protein; the effect was reduced on heat treatment, lost following protease digestion and only appeared in fractions of plasma containing proteins above 100kDa (Hayman et al. 2001). Myers et al. demonstrated the factor is present prior to diagnosis of the maternal syndrome of pre-eclampsia by retrospective application of plasma samples acquired from women at 22-26 weeks gestation who subsequently developed pre-eclampsia (Myers et al. 2005). The presence of multiple synergistic factors was also demonstrated by Myers et al; plasma from women who developed pre-eclampsia divided into less complex fractions had no effect on endothelial dependent relaxation, however when fractions were re-combined the effect was restored (Myers et al. 2007). This study also demonstrated that the factor is not concurrently transported with the high abundance plasma proteins including albumin, which were removed by targeted depletion prior to application to the myometrial arteries.

1.1.4.3 Early and Late Onset Pre-eclampsia

The gestational age at which pre-eclampsia develops is not taken into consideration in the definition of the condition; however it is an important variable in the maternal and perinatal outcome. The risk of maternal mortality is 20 times higher where pre-eclampsia develops between gestational ages 20-28 weeks (MacKay et al. 2001). Development of pre-eclampsia at an early gestation is also associated with fetal growth restriction and low birth weight (Xiong et al. 2002), and is more likely to result in a premature birth (Sibai 2006). To reflect these differences, pre-eclampsia is further subdivided into two conditions based on gestational age at presentation. Presentation of the maternal syndrome of pre-eclampsia requiring delivery before 34 weeks of gestation is defined as early onset pre-eclampsia, whereas presentation after 34 weeks is defined as late onset pre-eclampsia (von Dadelszen et al. 2003).
There is an argument that early and late onset pre-eclampsia represent two different conditions; there is evidence that gene expression (Sitas et al. 2009), morphology (Egbor et al. 2006), and protein expression (von Dadelszen et al. 2002) of the placenta differ in early and late onset pre-eclampsia. Von Dadelszen et al. propose that early onset pre-eclampsia may be of more use to differentiate changes between pre-eclamptic and unaffected pregnancies (von Dadelszen et al. 2003). There is also a suggestion, however, that the two conditions represent different ends of the same spectrum and that complication of early onset pre-eclampsia with other conditions results in a more severe syndrome (Huppertz 2008). Whether different conditions or a different presentation of the same condition, the maternal and perinatal outcome of pre-eclampsia differ depending on the gestational age at presentation of pre-eclampsia. The conditions are divided based on survival rates for the neonate; survival rates approach 95% at 34 weeks gestation (Field et al. 1991).

1.1.5 Diagnosis

A diagnosis of pre-eclampsia is currently made on observation of newly arising hypertension and proteinuria during pregnancy. Guidelines for the diagnosis and management of pre-eclampsia have been issued by several societies based on these signs, including the ISSHP (see 1.1.2.1) (Helewa et al. 1997; Brown et al. 2000; NHBPEP 2000; Brown et al. 2001; Milne et al. 2005). Quantification of blood pressure as an indication of hypertension and proteinuria is error prone (Waugh et al. 2004) and extensive recommendations to reduce error are included within the guidelines for diagnosis (Milne et al. 2005). Although diagnosis following these guidelines will identify most cases of pre-eclampsia, presentation of the syndrome is variable. Eclampsia and HELLP syndrome are both severe complications of pre-eclampsia, however some women who develop these conditions present without hypertension or proteinuria (Douglas and Redman 1994; Sibai 2004). Diagnosis of pre-eclampsia is difficult as it relies on measurement of error prone signs which vary depending on the guidelines used.

1.1.6 Prevention, Management and Treatment

As pre-eclampsia originates from the presence of the placenta, the only effective treatment is delivery. Delivery of a pregnancy complicated by pre-eclampsia represents tertiary prevention as the aim is to prevent further complications. The National Institute for Health and Clinical Excellence (NICE) issued clinical guidelines for the management of hypertensive disorders during pregnancy and recommend delivery for women who develop pre-eclampsia after 37 weeks gestation (NICE clinical guideline 107 2010). Delivery is also recommended for women who develop pre-eclampsia after 34 weeks, once blood pressure is controlled and a course of corticosteroids can be administered. Where pre-eclampsia presents before 34 weeks gestation (early onset pre-eclampsia) NICE recommends management of the condition with careful monitoring until maternal compromise and delivery is unavoidable. Antihypertensives may be used to control blood pressure, multiple treatment
options are available and applicable to differing presentations of the syndrome as summarised by Magee and von Dadelszen, and McCoy and Baldwin (Magee and von Dadelszen 2009; McCoy and Baldwin 2009). Magnesium sulphate is the ‘drug of choice’ to prevent eclampsia; maternal seizures associated with the signs of pre-eclampsia, and can halve the risk of pre-eclampsia developing into eclampsia (Altman et al. 2002).

Primary prevention of pre-eclampsia (complete prevention of the syndrome) is not possible as the cause of the condition is still unknown. Secondary prevention aims to prevent progression of pre-eclampsia to the clinically recognisable syndrome. Proposed interventions have focused on the pathophysiology of the disease process; successful intervention however is reliant on identification of women at greatest risk of developing pre-eclampsia. The efficacy of calcium supplementation, low dose aspirin and fish oil supplementation were reviewed by Dekker and Sibai (Dekker and Sibai 2001) who concluded that the treatments may have some modest effect but were of limited use in the general population.

More recently, supplementation with the antioxidant vitamins C and E was investigated by several studies on the basis that oxidative stress was a major causative factor in the development of pre-eclampsia (Poston et al. 2004; Rumbold et al. 2006; McCoy and Baldwin 2009; McCance et al. 2010). The outcome of a total of 10 studies was summarised by The Cochrane Collaboration and concluded that antioxidant supplementation did not reduce the risk of developing pre-eclampsia or other serious complications in pregnancy (Rumbold et al. 2008). The use of antiplatelet agents, in particular low dose aspirin, was investigated by metaanalysis by The PARIS (Perinatal Antiplatelet Review of International Studies) Collaboration (individual patient data for 32,217 women) (Askie et al. 2007) and The Cochrane Collaboration (review of 42 trials) (Duley et al. 2007). Both studies concluded there was a moderate reduction in the relative risk of pre-eclampsia, however counselled that further investigation of which women would be most likely to benefit, when treatment should be started and at what dose would be required before widespread use was recommended (Askie et al. 2007; Duley et al. 2007). In conclusion, there is currently no effective intervention for the prevention of pre-eclampsia.

1.1.7 Screening Tests

Management of pre-eclampsia requires identification of those at greatest risk. The basis for the development of the maternal syndrome of pre-eclampsia is established long before the clinical presentation of the condition (Figure 1.1.1). Blood flow to the placenta is established between 10 and 12 weeks gestation and so identification of pregnancies at risk of pre-eclampsia may be possible by this stage in development. There is evidence to suggest that differential placentation may be established as early as 7 weeks gestation (Salomon et al. 2003). Based on this knowledge there is potential for the development of a screening test to identify the women at greatest risk of developing pre-eclampsia.
Current tests proposed to predict the onset of pre-eclampsia measure a range of clinical, biophysical and biochemical markers (Table 1.1.1). Doppler ultrasound indicates the amount of trophoblast invasion by measuring the resistance in blood flow of the uterine arteries (Campbell et al. 1983). It was noted in pregnancies complicated by hypertension there was an altered waveform indicating an increase in vascular resistance. Doppler analysis has been carried out on women at high and low risk of developing early and late onset pre-eclampsia at a range of gestational time points (Audibert et al. 2005; Yu et al. 2005; Ghi et al. 2009; Groom et al. 2009; Sciscione and Hayes 2009). It was concluded, however that the technique has limited diagnostic accuracy in predicting onset of pre-eclampsia (Chien et al. 2000).

Early identification of women at risk of developing pre-eclampsia would be beneficial to antenatal management; antenatal care could be tailored to the relative risk associated with the pregnancy. The number of late antenatal appointments which are primarily in place to monitor for signs of pre-eclampsia, could be reduced for those at least risk and increased for those at most risk (Figure 1.1.1). The best estimate of the additional average cost associated a case of pre-eclampsia is £9,000 (Meads et al. 2008). With current knowledge of prediction of pre-eclampsia, economic modelling suggests a ‘no test-treat all’ strategy to be the most effective (Meads et al. 2008). Suggested interventions including rest and calcium supplementation are relatively low cost. The modest cost of application to all pregnant women is therefore outweighed by the small numbers who benefit from the treatment and require less antenatal care during pregnancy. The application of this treatment strategy however is not considered viable until further evidence of benefit is available. The ability to identify women at greater risk would provide a platform for the development of a more specific and beneficial intervention. Development of a screening test for early pregnancy to identify women at greater risk of developing pre-eclampsia is therefore of the utmost importance.

The World Health Organization Systematic Review of Screening Tests for Pre-eclampsia states that the ideal predictive test for pre-eclampsia should be ‘simple, innocuous, rapid, inexpensive, reproducible, and non-invasive as well as easy to perform early in pregnancy’ (Conde-Agudelo et al. 2004). The review compared proposed screening tests for predicting pre-eclampsia (tests listed in Table 1.1.1). Tests were categorised by those which measured dysfunction of the placental perfusion and vascular resistance dysfunction, feto-placental endocrinology, the renal system, and endothelial and oxidative stress. The conclusion of this and several more recent reviews (Meads et al. 2008; Briceno-Perez et al. 2009; Cnossen et al. 2009) is that there is currently no clinically useful screening test to predict the onset of pre-eclampsia.
Table 1.1.1  **Screening Tests for Pre-eclampsia.** Summary of proposed clinical, biophysical and biochemical tests for the prediction of pre-eclampsia (Adapted from Conde-Agudelo et al. 2004).

<table>
<thead>
<tr>
<th>Renal Dysfunction</th>
<th>Fetoplacental Unit Endocrinology Dysfunction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>•</strong> Serum uric acid</td>
<td><strong>•</strong> Human chorionic gonadotropin</td>
</tr>
<tr>
<td><strong>•</strong> Microalbuminuria</td>
<td><strong>•</strong> Alpha fetoprotein</td>
</tr>
<tr>
<td><strong>•</strong> Urinary calcium excretion</td>
<td><strong>•</strong> Estriol</td>
</tr>
<tr>
<td><strong>•</strong> Urinary kallikrein</td>
<td><strong>•</strong> Inhibin A</td>
</tr>
<tr>
<td><strong>•</strong> Microtransferrinuria</td>
<td><strong>•</strong> Pregnancy-associated plasma protein A</td>
</tr>
<tr>
<td><strong>•</strong> N-acetyl-β-glucosaminidase</td>
<td><strong>•</strong> Activin A</td>
</tr>
<tr>
<td></td>
<td><strong>•</strong> Corticotrophin release hormone</td>
</tr>
<tr>
<td><strong>Endothelial and Oxidant Stress Dysfunction</strong></td>
<td><strong>Placental Perfusion and Vascular Resistance Dysfunction</strong></td>
</tr>
<tr>
<td><strong>•</strong> Platelet count</td>
<td><strong>•</strong> Mean blood pressure in second trimester</td>
</tr>
<tr>
<td><strong>•</strong> Fibronectin</td>
<td><strong>•</strong> Roll over test</td>
</tr>
<tr>
<td><strong>•</strong> Platelet activation and endothelial cell adhesion molecules</td>
<td><strong>•</strong> Isometric exercise test</td>
</tr>
<tr>
<td><strong>•</strong> Endothelin</td>
<td><strong>•</strong> Intravenous infusion of angiotensin II</td>
</tr>
<tr>
<td><strong>•</strong> Prostacyclin</td>
<td><strong>•</strong> Platelet angiotensin II binding</td>
</tr>
<tr>
<td><strong>•</strong> Thromboxane</td>
<td><strong>•</strong> Platelet calcium response to arginine vasopressin</td>
</tr>
<tr>
<td><strong>•</strong> Cytokines</td>
<td><strong>•</strong> Renin</td>
</tr>
<tr>
<td><strong>•</strong> Homocysteine</td>
<td><strong>•</strong> 24-hour ambulatory blood pressure monitoring</td>
</tr>
<tr>
<td><strong>•</strong> Isoprostanes</td>
<td><strong>•</strong> Doppler ultrasound</td>
</tr>
<tr>
<td><strong>•</strong> Serum lipids</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Insulin resistance</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Antiphospholipid antibodies</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Plasminogen activator inhibitor</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Placental growth factor</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Leptin</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Hematocrit</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Total proteins</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Antithrombin III</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Magnesium</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Calcium</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Ferritin</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Transferrin</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Haptoglobin</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Atrial natriuretic peptide</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> B2-microglobulin</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Genetic markers</td>
<td></td>
</tr>
</tbody>
</table>
1.1.7.1 Evaluation of Screening Tests

Screening tests can be evaluated by several criteria, including sensitivity; the proportion of true positives that are correctly identified by the test, and specificity; the proportion of true negatives that are correctly identified (Bland and Altman 1994a). The probability that the test will give the correct diagnosis can be calculated by the positive and negative predictive values. The positive predictive value is the proportion of patients with positive test results who are correctly diagnosed, and conversely the negative predictive value is the proportion of patients with negative test results who are correctly diagnosed (Bland and Altman 1994b). In the case of pre-eclampsia, a high false positive rate is preferable to a false negative rate (Cnossen et al. 2009). Women who do not develop pre-eclampsia are unlikely to be negatively affected by increased antenatal observation and potential secondary intervention. The consequences are much more severe for women predicted of being at low risk of developing pre-eclampsia who are not monitored.

1.1.7.2 Screening Test Study Design

Studies involving patient sampling have many inherent problems, making cross study comparisons difficult, and relevance and reliability of findings questionable. Selection bias can be introduced by lack of randomisation of samples and subjects. Randomised control trials are considered the ‘gold standard’ of medical research (Greenhalgh 1997), however, in the case of developing screening tests and identifying predictive markers, cross sectional surveys and case control studies are more applicable. Cross sectional studies involve investigation of a representative sample of subjects where data is collected at a particular time point and may also involve retrospective or prospective information, i.e. the subsequent development of pre-eclampsia. Case control studies involve comparison of a subject of interest to a matched control subject of similar circumstance except for the condition of interest, i.e. with and without pre-eclampsia at a certain time of gestation.

Due to the nature of pre-eclampsia, presentation differs between subjects and various definitions of the disease exist. The guidelines issued by the ISSHP are most commonly used for diagnosis, however different studies may use differing definitions, introducing exclusion bias and also making comparison between studies impossible. The pathology of pre-eclampsia is established long before the clinical presentation; therefore studies are often carried out retrospectively, when the outcome of the pregnancy is known. In order to avoid detection bias, samples should be processed blind, i.e. without knowledge of the origins of the sample. Inconsistency may also be introduced during collection and storage of the sample; plasma and serum are commonly collected, however are not comparable. Standard operating procedures incorporating method of collection, storage and number of freeze thaw cycles, vary widely between studies and departments. Differences in detection and methodology also make comparison of studies difficult, as does the lack of standardised statistical evaluation and assessment. Numerous studies have been carried out into
methods of prediction of pre-eclampsia; however, results are often inconsistent and contradictory. As discussed previously, several systematic reviews applied appropriate meta-analysis (Conde-Agudelo et al. 2004; Meads et al. 2008) to these studies and concluded that there is currently no clinically useful predictive test for pre-eclampsia.

1.1.7.3 The SCOPE Study

The SCreening fOr Pregnancy Endpoints (SCOPE) Study was established as a platform from which research into identification, development, testing and validation of molecular biomarkers for prediction of pregnancy complications could take place (Australian and New Zealand Clinical Trials Registry ACTRN12607000551493; The SCOPE Pregnancy Research Study 2004). SCOPE recruited healthy nulliparous women with singleton pregnancies. Women were recruited into the study between 14 and 16 weeks of pregnancy. At this point blood and urine samples were collected and an interview conducted.

Women were seen for a second time at 20 weeks; blood and urine tests were repeated and a second interview carried out (metadata data collected is described by North et al. 2010). Women were scanned to measure fetal growth, blood flow to the placenta and cervical length, a vaginal swab was also collected. Complications arising later in pregnancy were documented and assessment of the child’s health and a sample of cord blood and DNA were taken on delivery. Samples collected as part of the SCOPE study were well curated; standard operating procedures and stringent control and tracking mechanisms were put in place to ensure samples were of the highest quality.

The SCOPE study was designed to fit with existing antenatal appointments; the timing of routine antenatal appointments (purple) and SCOPE study visits (green) through pregnancy are shown in Figure 1.1.1. The study recruited a total of 5,690 women from New Zealand, Australia, Ireland, the UK and the USA. Plasma samples collected at 15 weeks gestation were obtained for investigation. Changes in plasma proteins have been documented at 15 weeks gestation. At the time the study was designed, this gestation also coincided with blood tests made routinely during antenatal care. Although women no longer make a routine antenatal visit at 15 weeks, most women have booked with their midwife by this stage in the pregnancy and use of a prognostic biomarker could easily be incorporated into existing antenatal care.
Figure 1.1.1  The basis for the development of the maternal syndrome of pre-eclampsia is established long before the clinical presentation of the condition (black). There are documented changes in plasma protein expression and uterine blood flow in pregnancies complicated by pre-eclampsia prior to clinical diagnosis (red). There are monthly antenatal appointments during late pregnancy which are primarily in place to monitor for signs of pre-eclampsia (purple). Presentation of pre-eclampsia requiring delivery prior to 34 weeks gestation is termed early onset pre-eclampsia while presentation after 34 weeks is defined as late onset pre-eclampsia. The SCOPE study was designed to fit with existing antenatal appointments (green).
1.1.7.4 Candidate Predictive Factors

Multiple plasma proteins have been investigated as candidate markers predictive of pre-eclampsia. Although several have shown promise, a clinically applicable marker or combination of markers has yet to be identified for nulliparous women. Potential markers have been identified based on pathophysiological observations of pre-eclampsia, including placental dysfunction, endothelial activation, inflammatory response, and activation of the coagulation system. Proteins in amniotic fluid (Vascotto et al. 2007) and urine (Buhimschi et al. 2008) have also been investigated for predictive value, however, clinically useful markers have yet to be discovered in either.

1.1.7.4.1 Angiogenic Growth Factors

Angiogenesis and regulation of trophoblast invasion are regulated by angiogenic growth factors including vascular endothelial growth factor (VEGF) and Placental Growth Factor (PIGF) (Desai et al. 1999; Crocker et al. 2001; Tjwa et al. 2003). VEGF is required for maintenance of healthy arteries and binds to membrane bound fms-like tyrosine kinase 1(Flt-1). PIGF is homologous to VEGF and required for early placental development. Soluble Flt-1 (sFlt-1) is an alternative splice isoform of membrane bound Flt-1 and a VEGF and PIGF antagonist.

A cross sectional study applying the ISSHP guidelines to define pre-eclampsia and enzyme-linked immunosorbent assays (ELISA) found a higher median concentration of sFlt in plasma from pregnancies complicated by pre-eclampsia (Chaiworapongsa et al. 2004). Concentration of sFlt was found to be higher in early onset pre-eclampsia (<34 weeks) and the level of increase to correspond directly to the severity of the clinical presentation (Chaiworapongsa et al. 2004). An increase in sFlt-1 and concurrent decrease in PIGF was also shown by Sunderji et al. in time of disease plasma samples (Sunderji et al. 2009). This data however, was collected across a range of gestational ages and further investigation would be required to determine the clinical applicability at one gestational time point. The soluble form of Flt has a higher affinity for VEGF and PIGF than the membrane bound form, and so there is concurrent decrease in VEGF and PIGF levels (Maynard et al. 2003). Animal models have also shown that an increase in sFlt-1 levels produces a pre-eclampsia like phenotype and therefore high circulating levels of sFlt-1 and low levels of VEGF and PIGF are indicated in the pathogenesis of pre-eclampsia (Maynard et al. 2003). A nested case control study carried out within the CPEP Trial measured sFlt, VEGF and PIGF monthly, showed increased levels of sFlt and reduced PIGF prior to clinical onset of pre-eclampsia (Levine et al. 2004a). The concentration of VEGF was also measured but showed little difference between control and pre-eclampsia samples with the exception of measurements at 21 to 32 weeks (if 5 weeks prior to presentation), and at 37 to 41 weeks, where concentrations were lower in pre-eclampsia samples. Therefore, measurement of sFlt and VEGF 5 weeks prior to clinical presentation was suggested to be predictive of pre-
eclampsia. Conflicting evidence does exist however where PLGF and sFlt-1 were measured in early pregnancy (Ong et al. 2001; Thadhani et al. 2004). Mijal et al. also showed that concentrations of sFlt-1 and PIGF varied in pregnancy dependent on lifestyle, ethnicity, parity, BMI and other factors, further complicating the clinical applicability of these markers (Mijal et al. 2010).

Endoglin is a component of the transforming growth factor (TGF)-β receptor complex, expressed at high concentration on cell membranes of vascular endothelium and syncytiotrophoblast (Cheifetz et al. 1992; Gougos et al. 1992). Soluble endoglin (sEng) is released into the maternal circulation by up-regulation of placental endoglin in pre-eclampsia (Venkatesha et al. 2006). Association of sEng with endothelial nitric oxide synthase in the plasma membrane of the endothelium results in dysfunction (Venkatesha et al. 2006). A nested case control study within the CPEP Trial carried out blinded analysis of sEng and sFlt levels by ELISA (Levine et al. 2006). Multivariate analysis indicated that an increase in sEng and the ratio of sFlt:PIGF was associated with pre-eclampsia and that measurement of both was more strongly predictive than either alone. Increased levels of sEng and sFlt-1 however, were both observed in normotensive pregnancies complicated by IUGR and SGA suggesting the markers are not specific to pre-eclampsia alone (Wathen et al. 2006; Stepan et al. 2007; Erez et al. 2008; Jeyabalan et al. 2008).

1.1.7.4.2 The Coagulation System

In normal pregnancy, the coagulation system is in a heightened state of activation (Halligan et al. 1994). The effects are vastly exaggerated in pre-eclampsia due to the associated dysfunction of the vascular endothelium which has a central role in the coagulation cascade. Symptoms include endothelial injury, increased platelet activation and enhanced clotting (Whigham et al. 1978; Saleh et al. 1992). Thrombin cleaves the fibrinopeptides A and B from fibrinogen to form fibrin. Fibrin deposition contributes to the multisystemic organ failure found in pre-eclampsia although changes in levels of fibrinogen and fibrinopeptides correlated with onset of clinical symptoms rather than preceded it (Borok et al. 1984). Under normal circumstances, fibrin production is tightly regulated by the action of tissue plasminogen activator (tPA) and urokinase on inactive plasminogen, forming active plasmin which degrades fibrin. Longitudinal cross sectional observations have shown that in normal pregnancy levels of plasminogen activator inhibitor (PAI)-1 and plasminogen activator inhibitor -2 increase progressively (Halligan et al. 1994). A prospective longitudinal case control study showed that an increased ratio of PAI-1:PAI-2 could be used to distinguish pre-eclampsia from low risk subjects at 24 weeks, prior to clinical onset of pre-eclampsia (Chappell et al. 2002). When combined with other markers (Leptin, placental growth factor, and uric acid) the ratio of PAI-1/PAI-2 may be predictive as early as 20 weeks but only in high risk cases (Chappell et al. 2002).
1.1.7.4.3 Inflammation

Maternal systemic inflammation is a feature of the third trimester of normal pregnancy; in pre-eclampsia this inflammatory response occurs earlier and in an exaggerated state (Redman and Sargent 2003). This lack of distinction between normal and diseased states is thought to contribute to the difficulty in diagnosis of the condition.

C-Reactive protein (CRP) is a positive acute phase reactant; circulating levels rise quickly and steeply in response to inflammatory stimuli (Morley and Kushner 1982). CRP acts as a scavenger removing dangerous molecules and has been shown to be at a raised concentration in high risk women with pre-eclampsia (Teran et al. 2001). A prospective nested case control study found CRP increased 2-fold in the first trimester of pregnancy in women who subsequently developed pre-eclampsia (Wolf et al. 2001). High circulating CRP levels were found to be associated with risk of developing pre-eclampsia, however, when corrected for body mass index (BMI) this association was lost (Wolf et al. 2001). Increased CRP levels have previously been associated with obesity (Visser et al. 1999) and coronary heart disease (Rifai and Ridker 2002); both of which are risk factors for pre-eclampsia, therefore, pre-pregnancy levels of CRP must be accounted for in order to correct for levels in pregnancy and any further changes found in pregnancies complicated by pre-eclampsia.

Reduction in high-density lipoprotein cholesterol and altered cholesterol transport are also associated with systemic inflammation and pre-eclampsia. These changes promote atherosclerosis and may explain the link between pre-eclampsia and subsequent increased risk of cardiovascular disease (Rodie et al. 2004). A prospective case control study of high risk women showed increased serum lipid concentrations in those who developed pre-eclampsia, however, this increase was also found in those who delivered SGA infants (Chappell et al. 2002). Lipid profiles (cholesterol, triglycerides, very low density lipoprotein (VLDL), low density lipoprotein (LDL) and high density lipoprotein (HDL)) were increased at 15-20 weeks gestation in women who subsequently developed pre-eclampsia (Baker et al. 2009). This trend continued through pregnancy until approximately 30 weeks (Uzun et al. 2005) but at term triglycerides, total cholesterol and LDL cholesterol were all significantly increased in pre-eclamptic woman while levels of HDL cholesterol were decreased (Kocyigit et al. 2004). Differential levels of Apo E and Apo C-11 were also indicated at term by Atkinson et al. in pre-eclamptic pregnancies compared to normotensive controls (Atkinson et al. 2009). Specificity of measuring levels of lipoproteins and cholesterol is questionable; there is evidence that an altered lipid profile both before and after pregnancy may be associated with a number of pregnancy complications rather than pre-eclampsia alone (Wiznitzer et al. 2009).

Systemic inflammation and oxidative stress are intrinsically linked; an inflammatory response leads to oxidative stress and conversely the products of oxidative stress cause inflammation. The generation of free radicals by oxidative stress leads to oxidative damage...
of proteins, lipids, and DNA (Burton and Jauniaux 2004). Various markers of lipid peroxidation have been studied in relation to pre-eclampsia including homocysteine (D’Anna et al. 2004), the prostaglandins (Fitzgerald et al. 1987) and malondialdehyde (Hubel et al. 1996), although no conclusive evidence of a predictive value could be found.

1.1.7.4.4 Placental Secretions

As pre-eclampsia is dependent on the presence of the placenta and is thought to originate due to altered placental function, it is reasonable to assume that levels of placental derived proteins would also be altered. Upward of 50 proteins are secreted by the placenta with several of these proteins investigated as potential predictive markers of pre-eclampsia.

Placental protein 13 (PP-13) is expressed by the syncytiotrophoblast and levels are thought to reflect the functional status (Huppertz et al. 2008). Mean PP-13 levels in pregnancies complicated by pre-eclampsia, IUGR and preterm delivery and measured by custom ELISA were lower during the first trimester and high in second and third trimester than in healthy controls (Burger et al. 2004). Significantly lower levels of PP-13 were measured in the first trimester of pregnancies complicated by pre-eclampsia, IUGR and preterm delivery by ELISA (Chafetz et al. 2007). For pre-eclampsia alone however, specificity of prediction was 90% and sensitivity 79% with a receiver operating characteristic analysis area under the curve of 0.91 (Chafetz et al. 2007). Nicolaides et al. combined first trimester PP-13 serum concentrations with measurement of pulsatility index of blood flow in the uterine arteries to predict pregnancies which would subsequently be affected by pre-eclampsia (Nicolaides et al. 2006). The detection rate was 90% with a false-positive rate of 6% compared to 12% and 31%, when each method was used alone (Nicolaides et al. 2006).

A peak in human Chorionic Gonadotropin (hCG) concentration in the maternal circulation is linked to trophoblast invasion in the 12th week of gestation; unlike other placental hormones, levels then fall and remain low (Bischof and Irminger-Finger 2005). Retrospective studies using data obtained while screening for trisomy 21 at 15 to 18 weeks, found maternal serum hCG (multiples of the median) to be significantly raised in samples from pregnancies complicated by pre-eclampsia (Muller et al. 1996). An association between high serum hCG and pre-eclampsia has been shown by other studies, and when combined with uterine artery Doppler ultrasound the positive predictive value increased. Sensitivity however, was found to be too low to be of useful predictive value (Audibert et al. 2005).

The dimeric glycoprotein hormones inhibin A and activin A are members of the TGF-β family and primarily secreted by the placenta. In normal pregnancy inhibin A and activin A levels increase from early pregnancy and peak at delivery (Silver et al. 1999). A prospective case control study measuring inhibin A by ELISA found significantly higher concentrations in maternal serum in those who developed pre-eclampsia as early as 7-13 weeks gestation.
A nested case-control study measured inhibin A and activin A in early pregnancy and demonstrated concentration of both were higher in early onset pre-eclampsia samples than control samples at each time point however sensitivity and specificity were low (Muttukrishna et al. 2000). The predictive values of inhibin A and activin A were greatly increased when combined with uterine artery ultrasound (Spencer et al. 2006). The association of activin A with pre-eclampsia was lost when serum levels were measured by immunoassay in women with underlying health complications (chronic hypertension, renal disease or previous pre-eclampsia) at high risk of developing the condition (Blackburn et al. 2003).

During pregnancy, fetal cells, including erythroblasts and cell-free fetal DNA (cff-DNA), pass across the placenta into the maternal circulation. Cell-free fetal DNA has been shown to be at increased levels prior to clinical onset of pre-eclampsia where the fetus was male (Lo et al. 1999; Zhong et al. 2002; Levine et al. 2004b). This increase was thought to correspond to an increased release of necrotic and apoptotic cff-DNA containing syncytiotrophoblast fragments by the hypoxic placenta. Any potential predictive relevance of cff-DNA to pre-eclampsia is extremely limited as current methods are complex, costly (Baumann et al. 2007), and only applicable where the fetus is male (Levine et al. 2004b).

**1.1.8 Development of Personalised Clinical Risk Profiles**

As previous studies have demonstrated it is unlikely that there is any one plasma protein with sufficient predictive value to identify women at risk of pre-eclampsia. The specificity and sensitivity of several tests was improved when multiple proteins were considered (Chappell et al. 2002; Levine et al. 2006) and on inclusion of measurements made by Doppler Ultrasound (Audibert et al. 2005; Spencer et al. 2006). Currently women at risk of developing pre-eclampsia in pregnancy are identified using a limited set of clinical risk factors (see 1.1.3) (Milne et al. 2005).

North et al. expanded this and considered an extensive list of demographic and clinical risk factors in a prospective multicentre cohort study of nulliparous woman at 15 weeks of pregnancy (North et al. 2010) The aim of the investigation was to provide a personalised clinical risk profile for nulliparous women. It was concluded that the predictive ability of clinical phenotype was ‘modest’ and that addition of biochemical markers would greatly improve predictive value (North et al. 2010). It has also been demonstrated that levels of some protein markers are dependent on demographic and socioeconomic factors, including age, body mass index, smoking status and ethnicity (Mijal et al. 2010). In conclusion, the most powerful predictive tool for pre-eclampsia is likely to incorporate clinical risk factors and demographic data with biophysical and biochemical measurements to determine an individualised risk factor with could be used to determine a personalised antenatal care pathway.
1.1.8.1 A Personalised Clinical Risk Profile in Use - Ovarian Cancer Risk of Malignancy Index

Ovarian cancer has a high mortality rate in developed countries, particularly where a late diagnosis is made (Oriel et al. 1999). Differentiation between benign and malignant conditions is difficult, particularly where resources are limited. The Risk of Malignancy Index (RMI) is used as an indicator for women who would benefit most from efficient and effective allocation of resources (van den Akker et al. 2010). The RMI was developed following analysis of the several clinical, biochemical and biophysical markers thought to be predictive of ovarian cancer. It was concluded that alone these markers did not have sufficient predictive value, however, combining three factors: ultrasound scan results, menopausal status and serum CA 125 measurement, effectively discriminated between cancerous and benign lesions with a sensitivity of 85% and specificity of 97% (van den Akker et al.; Jacobs et al. 1990; Tingulstad et al. 1996; Tingulstad et al. 1999; Yamamoto et al. 2009). The RMI has been adopted into clinical practice to stratify women according to risk; NICE recommends women with a high score on the RMI are referred immediately to a specialist multidisciplinary team (NICE clinical guideline 122 2011).

1.1.9 Project Rationale

As a result of abnormal placentation it is suggested that a ‘circulating factor’ is released into the maternal circulation. This factor causes a cascade of reactions which ultimately activate the maternal endothelium and has been shown to be active in the maternal circulation prior to clinical presentation. Changes in expression of several circulating plasma proteins have been identified prior to clinical presentation; evidence suggests expression of some proteins is altered as early as 7 weeks gestation (Salomon et al. 2003). Changes in expression of these plasma proteins are ideal targets for prediction of subsequent pre-eclampsia. Targeted methods have yet to identify proteins with sufficient predictive value in a nulliparous population; a global and unbiased strategy able to identify novel proteins of predictive value may be of more use.
1.2 Proteomics

The term ‘proteome’ is used to define the entire complement of expressed proteins and so ‘proteomics’ the field studying the proteome (Wilkins et al. 1996). Anderson and Anderson defined the goal of proteomics as the ‘comprehensive, quantitative, description of protein expression and its changes’ (Anderson and Anderson 1998). The proteome is dynamic; the expression and degradation, post-translational modification and interaction of proteins are altered in response to environmental factors. Study of the genome can indicate the potential of an organism to a particular outcome but it cannot identify the onset of that outcome, for example the risk of developing diabetes and actual development of the condition. The proteome of an organism reflects the environmental factors which also contribute to the outcome, for example, the lifestyle and dietary factors which may result in development of diabetes in one person with genes indicating risk but not in another. Analysis of the proteome is therefore a useful tool for monitoring disease progression, particularly in the case of pre-eclampsia where the hypothetical ‘circulating factor’ has been shown to have a protein component.

1.2.1 Biomarker Discovery

Protein indicators or ‘biomarkers’ can be used to identify the risk, presence or stage of development of a disease. Despite investigation of multiple candidates a clinically useful prognostic biomarker for pre-eclampsia has yet to be identified.

Identification of clinically applicable biomarkers for any condition follows a defined pathway or pipeline (Pepe et al. 2001; Rifai et al. 2006). The biomarker pipeline can be divided into two stages; firstly discovery, where biomarkers of interest are identified and secondly validation, where observations made in the discovery phase are tested (Table 1.2.1). The discovery phase investigation is commonly applied to a large number of potential biomarkers in a small number of samples or a model system. Methods used for discovery phase investigation have low throughput but must be able to quantify multiple components of each sample. A large proportion of the candidate biomarkers identified during discovery phase are likely to be false positives, often these are identified as an artefact of the discovery method or model system used. A small number of the candidate biomarkers identified in the discovery phase are then validated. Validation of biomarkers requires multiple stages. Increasing numbers of clinically relevant samples are used first to confirm discovery observations, secondly to assess specificity and sensitivity of the biomarker, and finally to develop a clinically applicable assay. Samples must be drawn from the population to which the biomarker will be applied and the methods used must be sensitive, specific and have high-throughput.
Table 1.2.1  **The biomarker discovery pipeline.** Proteins of interest are identified during discovery phase which quantifies large numbers of potential biomarkers in a small number of samples. A small number of potential biomarkers identified during the discovery phase are subsequently validated in increasing numbers of samples (Adapted from Rifai et al. 2006).

<table>
<thead>
<tr>
<th>Phase</th>
<th>Aims</th>
<th>Samples</th>
<th>Number of Biomarkers</th>
<th>Number of Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discovery</strong></td>
<td>Identify promising candidate markers</td>
<td>Proximal Fluids, Human plasma, Animal model plasma</td>
<td>1,000s</td>
<td>10s</td>
</tr>
<tr>
<td><strong>Qualification</strong></td>
<td>Development of test demonstrating differential expression</td>
<td>Human plasma</td>
<td>30-100</td>
<td>10s</td>
</tr>
<tr>
<td><strong>Verification</strong></td>
<td>Retrospective longitudinal studies evaluating specificity and sensitivity</td>
<td>Population derived human plasma</td>
<td>10s</td>
<td>100s</td>
</tr>
<tr>
<td><strong>Testing and Validation</strong></td>
<td>Trial, assay optimisation and impact analysis</td>
<td>Population derived human plasma</td>
<td>4-10</td>
<td>1,000s</td>
</tr>
</tbody>
</table>
1.2.2 Proteomic Methods for Biomarker Discovery

As previously discussed, the proteome is potentially a source of biomarkers for many conditions including pre-eclampsia. Application of proteomic methods to the biomarker pipeline has lead to the identification of multiple candidate biomarkers (Labugger et al. 2003; de Vera et al. 2006; Simpson et al. 2009).

1.2.2.1 Hypothesis Generating Proteomic Methods

The aim of hypothesis generating proteomic methods is to quantify the proteome of a system. These methods are global, unbiased and un-directed. Hypothesis generating proteomic methods are most applicable to the discovery phase of the biomarker pipeline as all proteins in the sample are considered.

1.2.2.1.1 Two Dimensional Gel Electrophoresis (2D-Gel)

Proteins are separated first by isoelectric point and secondly by molecular weight, producing a map of protein spots which can be visualised. To prepare samples for separation, proteins are first denatured, reduced and solubilised. Sample preparation can introduce variation into the analysis and care must be taken to remove or inactivate interfering compounds such as proteases and salts. Solubilisation of proteins also limits the proteins which can be analysed using this method; membrane and other highly hydrophobic proteins require specialised buffers to prevent precipitation (Rabilloud 1998). The sample is then applied to first dimension isoelectric focusing. The development of immobilised pH gradients (Bjellqvist et al. 1982) has improved the reproducibility of the technique (Corbett et al. 1994) and the increased specificity and pH range which can be targeted (Gorg et al. 1985). In the second dimension proteins are separated by molecular weight by Sodium Dodecyl Sulphate-PolyAcrylamide Gel Electrophoresis (SDS-PAGE). The technique is again limited by a lack of ability to identify proteins of extreme molecular weight, targeting of these proteins requires additional gels (Schägger and von Jagow 1987).

Separated proteins are visualised by staining; anionic dyes, metal cations, fluorescent compounds and radioactive isotopes can be used and must be chosen to suit the requirements of the analysis carried out. Comparison of separated proteins is carried out by overlaying images of the gels. Difference Gel Electrophoresis (DIGE) was developed to improve quantitative comparison between gels (Ünlü et al. 1997). Two protein samples labelled with fluorescent dyes are run on one gel and the fluorescence image compared post run, increasing throughput and reducing gel to gel variability. Although analysis of the protein spot pattern can identify changes between samples, the identity of the changed protein is unknown. In order to determine the identity of the changed proteins, the spot must be excised from the gel and the protein identified by mass spectrometry. This requires precise excision of the differential protein spot to prevent contamination with other proteins and may be further complicated where one spot contains multiple proteins.
2D Gel electrophoresis is an ideal technique for resolving larger number of proteins in complex samples, particularly proteins which have undergone post translational modification (Mann and Jensen 2003). The technique however, is limited by a high level of variability both in sample preparation and between gels (Corbett et al. 1994). Molloy et al. estimated the technical variation from the process as 20-30% coefficient of variation (CV), dependent on the type of sample used (Molloy et al. 2003). The technique also lacks the ability to identify all classes of proteins particularly those at the extremes of pH and molecular weight and at low abundance in complex samples (Gygi et al. 2000).

1.2.2.1.2 Global Mass Spectrometry

Mass spectrometry measures the mass to charge ratio ($m/z$) of gas phase ions with high sensitivity. The technique has been adapted for use with proteins and has become a powerful method for high throughput analysis in the field of proteomics. Using stable isotope labelling, simultaneous identification and quantification of proteins in complex samples is possible by mass spectrometry (Aebersold and Mann 2003; Lane 2005; Hoffmann and Stroobant 2007; Simpson et al. 2009).

Schmidt et al. compared the use of stable isotope labelling with mass spectrometry and 2D gel electrophoresis in identification and quantitation of the Mycobacterium tuberculosis proteome (Schmidt et al. 2004). More proteins were identified and quantified using mass spectrometry and while 27 proteins were identified by both techniques, the expression ratios were not comparable. Choe et al. investigated the proteome of Escherichia coli (E.coli) in response to rhsA induction using an alternative stable isotope label mass spectrometry technique and 2D gel electrophoresis (Choe et al. 2005). Quantitation was more reproducible using mass spectrometry: 95% of the expression ratios for quantified proteins had a CV less than 53% compared to 81% using 2D- gel electrophoresis. Both investigations concluded that different populations of proteins were identified by each method. Due to the ability for multiplex, high-throughput, reliable and sensitive analysis it was concluded that quantitation using mass spectrometry was of most use in the identification of changes in plasma proteins in pre-eclampsia.

1.2.2.2 Hypothesis Driven Proteomics Methods

Hypothesis driven methods are targeted methods. Knowledge derived from hypothesis generating methods, pathophysiology, literature and other sources are used to define protein(s) for quantification. The validation phase of the biomarker pipeline relies upon hypothesis driven proteomics methods (Rifai et al. 2006).
1.2.2.2.1 Antibody Methods

Antibodies can be raised to almost any protein provided pure protein is available. To produce polyclonal antibodies, protein is inoculated into a suitable host mammal and the sera containing the antibody collected and purified. As a result of the in vivo generation of these antibodies, the epitope may vary between antibodies. Monoclonal antibodies are produced in vitro by hybridoma cells and as a result, are specific to one common epitope (Kohler and Milstein 1975). Antibodies are highly specific and can be utilised in a variety of techniques including western blotting and ELISA. Both western blotting and ELISA have been used for initial validation of hypothesis generating proteomics (Atkinson et al. 2009; Blankley et al. 2009; Blumenstein et al. 2009a; Blumenstein et al. 2009b). Ultimately, the goal of any biomarker discovery pipeline is the development of an ELISA. Due to the high level of specificity and ease of use, clinical laboratories routinely use ELISAs for diagnostic tests (Lequin 2005).

1.2.2.2.2 Targeted Mass Spectrometry

Mass spectrometry can also be used for targeted proteomic investigations. Increasingly, multiple reaction monitoring is being used for validation of candidate biomarkers as discussed by Kitteringham et al. (Kitteringham et al. 2009). Prior knowledge of the protein of interest is required, including an elution time and information on how the protein behaves on analysis by mass spectrometry. This information can be acquired hypothetically by software, from databases collating previous observations of the protein or from in-house application of global mass spectrometry.

1.2.3 Plasma Proteomics

Blood is an ideal candidate for proteomic analysis; it is easily obtained and perfuses all tissues. Proteomic methods have been applied to other bodily fluids including synovial (Baillet et al. 2009), cerebrospinal (Norwitz et al. 2005) and amniotic fluid (Vascotto et al. 2007), however these are difficult to obtain. Urine is more accessible but highly variable and complicated to process (Buhimschi et al. 2008). Blood products offer a unique picture of the whole body (Katz et al. 2005) and are highly applicable to clinical tests. Blood products are routinely collected and tested by clinical laboratories for a multitude of conditions.

Serum is the fractionated product of clotted blood and plasma the product of blood collected in the presence of an anticoagulant. While serum is less complex and so easier to study it is less stable (Boyanton and Blick 2002) than plasma which is more complex but more stable as the presence of the anticoagulant prevents the action of proteases (Evans et al. 2001). Measurement of some proteins can be affected by the blood product it is measured in; VEGF is actively secreted by platelets and therefore found at higher levels in serum than plasma (Benoy et al. 2002; Spence et al. 2002). Increasingly plasma is being used for proteomic analysis where a stable, uniform sample is required (Anderson and
Anderson 2002). The plasma proteome however, is highly variable from sample to sample and person to person. Proteins in the plasma reflect internal and external environmental factors and so represent a snapshot of a system at any one point in time. Factors affecting the proteome include pregnancy (Lockitch 1997), gender (Ogata et al. 2007), age (Buckley and Dorsey 1970), season (Letellier and Desjarlais 1982) and time of day (Macy et al. 1993).

1.2.3.1 Plasma Complexity

Plasma is the most complex human derived proteome containing an estimated 10,000,000 protein forms (Anderson and Anderson 2002). Putnam defined a true plasma protein as one which functions in the circulation, however, this definition excludes many proteins which are also found in plasma (Putnam 1975). Anderson and Anderson expanded this definition to reflect the complexity of plasma and classify the entire protein content (Anderson and Anderson 2002). They defined the following groups; proteins secreted by solid tissues that act in plasma, immunoglobulins, long distance receptor ligands, local receptor ligands, temporary passengers, tissue leakage products, aberrant secretions, and foreign proteins.

Protein content is on average between 60-80mg per ml of plasma (Jiang et al. 2004; Hortin and Sviridov 2009). Some of the highly abundant classical plasma proteins occur at milli-molar concentration, while lower abundance proteins including interleukins, aberrant secretions and tissue leakage products can occur at femto-molar concentrations (Hortin and Sviridov 2009). The dynamic range of plasma therefore spans a $10^{12}$ fold concentration range and exceeds the dynamic range of most analytical methods (Figure 1.2.1)(Hortin et al. 2006). It is argued that all proteins are of interest as biomarkers and diagnostic tests exist for proteins throughout the dynamic range. The majority of diagnostic tests, however, target lower abundance proteins (Figure 1.2.1). This presents a problem for proteomic analysis. Approximately 22 high abundance proteins account for 99% of the plasma proteome by mass; the transport protein albumin, for 50% alone (Tirumalai et al. 2003). The remaining 1% of low abundance proteins consists of several classes of physiologically important proteins including cytokines, chemokines, and peptide hormones, which are of primary interest for biomarker discovery (Tirumalai et al. 2003).

Various methods have been applied to plasma to reduce the dynamic range and remove highly abundant proteins which overwhelm any attempt to analyse lower abundance proteins.
The dynamic range of plasma spans a 10^12 fold concentration range and exceeds the dynamic range of most analytical methods. Although all proteins are of interest as biomarkers, the majority of diagnostic tests target lower abundance proteins. Plasma proteins were defined by concentration by Anderson and Anderson within the concentration range of the Classical Plasma Proteins (1x10^6 to 1x10^10 pg/mL), Tissue Leakage Products (1x10^3 to 1x10^6 pg/mL) and Interleukins (1x10^3 to 1x10^6 pg/mL). (Adapted from Anderson and Anderson 2002).
1.2.4 Application of Proteomics to the Study of Pre-eclampsia

Several studies have applied hypothesis generating proteomic methods to plasma to identify proteins with changed expression in pre-eclampsia. The established 2D gel and DIGE method have been applied to early pregnancy and term, plasma and serum samples (Atkinson et al. 2009; Blumenstein et al. 2009a; Blumenstein et al. 2009b). Mass spectrometry methods are less well established and several studies have demonstrated suitability for application to pre-eclampsia (Blankley et al. 2009; Auer et al. 2010).

1.2.4.1 Two Dimensional Gel Electrophoresis and Pre-eclampsia

Atkinson et al. carried out a case control study using serum and plasma samples from women who developed pre-eclampsia near term (Atkinson et al. 2009). Samples were depleted of 6 high abundance proteins; serum was investigated by 2D electrophoresis and plasma by DIGE which is more technically reliable. In serum, 6 spots were identified which could differentiate case from control samples and in plasma, 4 spots were associated with pre-eclampsia. Apo E was identified as up-regulated in serum as a basic isoform and down-regulated as an acidic isoform in pre-eclampsia. This was validated by western blot, however the protein was identified in only 11 of 12 women with pre-eclampsia and three of 12 healthy pregnant women. Blumenstein et al. also used DIGE to compare plasma taken at 20 weeks gestation from women who developed pre-eclampsia with and without a small for gestational age infant to uncomplicated pregnancies (Blumenstein et al. 2009a). Plasma was depleted of 6 highly abundant proteins. A total of 49 spots were identified by multivariate analysis as differentiating between all three groups. Western blot analysis confirmed that two proteins: fibrinogen and alpha-1-antichymotrypsin, were differentially regulated prior to development of pre-eclampsia.

A second longitudinal study by Blumenstein et al. again applied DIGE to plasma sampled from women with pre-eclampsia and healthy controls (Blumenstein et al. 2009b). Plasma was sampled at 20 weeks gestation prior to a diagnosis of pre-eclampsia and at 33-36 weeks after subsequent diagnosis of the condition. Women who did not develop pre-eclampsia were used as healthy controls. Eight protein spots were identified as significantly different between case and control. Vitronectin and high molecular weight kininogen were identified as the primary proteins in those eight spots by MS. Changes in vitronectin were validated at 20 weeks gestation by western blot. High molecular weight kininogen could only be validated in samples taken from women with pre-eclampsia and a small for gestational age infant. All three studies demonstrated the difficulty in identifying single proteins from spots on 2D gels. The proteins identified were at relatively high concentration and difficult to identify even when samples were depleted of high abundance proteins. Several of the proteins identified were specific to pre-eclampsia with a small for gestational age infant
which may limit use as a clinical marker. The ability of 2D gels to identify isoforms of proteins is clearly demonstrated, however translation into a clinical test would be difficult using antibody based methods.

1.2.4.2 Mass Spectrometry and Pre-eclampsia

Several studies have applied hypothesis generating mass spectrometry based methods to the identification of plasma proteins that are changed in pre-eclampsia. Myers et al. used the selective mass spectrometry method Surfaced Enhanced Laser Desorption and Ionisation (SELDI) to carry out a pilot study (Myers et al. 2004). SELDI allows the selective binding of peptides to a target prior to MS analysis. Plasma from women was taken at 26 weeks gestation prior to diagnosis of pre-eclampsia and compared to matched controls. Five proteins were identified as significantly up-regulated and the method proved potentially useful for a larger sample set. The method identifies mass values of candidate proteins, however the identity of the protein cannot be determined and requires further investigation. Due to a lack of sensitivity and protein identification data which is essential for investigation of candidate markers, the SELDI method was determined to be of limited use (Tang et al. 2004).

Blankley et al. used iTRAQ, a labelled MS method capable of quantifying and identifying multiple proteins, in a proof of principle investigation. The method was applied to pooled plasma samples depleted of 12 highly abundant proteins. Plasma was acquired at time of disease from women who developed pre-eclampsia and controls matched for gestational age. Multiple proteins were identified as changed in expression between case and control samples, including several proteins previously demonstrated to be differentially expressed. Proteins identified as changed included PAPP-A and endoglin which were also validated by immunoassay. This initial investigation demonstrated the ability of the method to identify proteins changed in abundance. Advantages of the iTRAQ method include the ability to quantify and identify proteins simultaneously. The iTRAQ method was also applied to term plasma depleted of high abundance protein and pooled by phenotype by Auer et al. (Auer et al. 2010). Plasma was acquired from healthy controls and women who developed pre-eclampsia, pre-eclampsia with intrauterine growth restriction and intrauterine growth restriction alone. Several proteins were identified as changed in expression in pre-eclampsia alone including SERPINA3 and CRP. Changes in SERPINA3 expression were confirmed by western blot of each of the samples contributing to the pooled samples used for the iTRAQ investigation.
1.2.5 Project Rationale – Method

The ability of hypothesis generating quantitative mass spectrometry to identify proteins changed in expression in pregnancies complicated by pre-eclampsia has been established. The iTRAQ method has been applied to late pregnancy samples (Blankley et al. 2009; Auer et al. 2010), but has yet to be applied to early pregnancy plasma samples for the identification of prognostic biomarkers for pre-eclampsia.

1.3 Mass Spectrometry

Mass spectrometry (MS) is a powerful and versatile technique for chemical characterisation. The technique relies on the generation of gas phase ions; the ratio of the mass of the ion to the charge can then be measured and used to deduce the chemical structure of the compound.

1.3.1 Components of a Mass Spectrometer

An MS instrument consists of a sample inlet, an ion source, a mass analyser, and an ion detector. The sample is applied to the instrument via the sample inlet (Figure 1.3.1). Gas phase ions are produced by the ionisation source and directed toward the mass analyser. The mass analyser separates ions by the mass to charge ratio ($m/z$), the intensity of the ions at each $m/z$ is then quantified by the ion detector. The result is a mass spectrum which plots the intensity of the ions at each $m/z$ and can be used to determine the molecular weight of the ions measured. Calculating the molecular weight of the ion from the mass spectrum allows structural information about the sample to be determined. The instrument must operate under vacuum to prevent interruption to the movement and measurement of the ions.

Figure 1.3.1 Components of a Mass Spectrometer. A mass spectrometer consists of a sample inlet, an ion source, a mass analyser and an ion detector.
1.3.1.1 Ionisation Source

The role of the ion source is to produce gas phase ions as the sample must be
vaporised and ionised to be measured by MS. Charged gaseous particles can be
electrostatically directed and quantified within the vacuum environment of the mass
spectrometer. Proteins and polypeptides are liable to excessive fragmentation and thermal
decomposition if exposed to the high energy ionisation methods used to ionise more stable
chemical compounds. Low energy ‘soft’ ionisation techniques were therefore developed to
ionise proteins and polypeptides (Tanaka et al. 1988; Griffiths et al. 2001). The two most
common ‘soft’ ionisation techniques used in proteomics are electrospray ionisation (ESI) and
matrix-assisted laser desorption ionisation (MALDI).

1.3.1.1.1 Electrospray Ionisation

ESI involves transfer of ions in solution into gaseous phase ions at atmospheric
pressure by application of charge (Yamashita and Fenn 1984a; Yamashita and Fenn
1984b). The exact mechanism by which the peptides become vaporised and ionised is not
yet completely understood (Gaskell 1997). In summary three stages are involved; droplet
formation by application of the sample to a capillary in a strong electric field, droplet
shrinkage by gas flow or heat, and ion formation by explosion of the droplet at the point
where the electrostatic repulsion exceeds the surface tension (Figure 1.3.2-A). Multiple
charged ions are commonly produced due to the presence of multiple ionisable sites; a
denatured protein typically has one charge per 1000Da (Lane 2005). ESI is typically used in
combination with HPLC; chromatography is directly coupled to the MS instrument via the
electrospray capillary (Liu et al. 2002). MS analysis is carried out dynamically on-line as the
sample elutes from the chromatography column.

1.3.1.1.2 Matrix-Assisted Laser Desorption Ionisation

MALDI involves sublimation and ionisation of the analyte from a dry crystalline
matrix using laser pulses (Figure 1.3.2-B)(Karas and Hillenkamp 1988). The sample is co-
crystallised with an excess of light absorbing matrix material. Weak UV-absorbing organic
acids such as 4-hydroxy-alpha-cyanocinnamic acid (HCCA) or dihydroxybenzoic acid are
commonly used as matrix. The matrix is excited by pulses of light from a laser; a nitrogen
laser at 337nm is typically used. The matrix is rapidly heated causing sublimation of matrix
crystals; almost all of the light energy is absorbed by the chromophore in the matrix. The
sample enters gas phase alongside the matrix. A small amount of energy is transferred from
the matrix to the sample allowing ionisation without fragmentation. The exact mechanism by
which ions are produced is not yet fully understood. The most accepted hypothesis suggests
gas phase proton transfer occurs during expansion of the matrix into gas phase producing
singly charged ions (Hoffmann and Stroobant 2007). Ions are then accelerated toward the
mass analyser by application of an electric field.
SELDI is a variation of the MALDI technique allowing selection of proteins from crude mixtures prior to ionisation. Proteins are applied to a solid phase surface designed to select for proteins according to chemical or biochemical properties (Tang et al. 2004). Unbound proteins and other unwanted substance (for example salts, detergents, and lipids) are removed by washing with an appropriate buffer. Bound proteins are then co-crystallised with matrix and ionised utilising the MALDI technique.

**Figure 1.3.2** Gas phase ions are produced by the ion source. A- Electrospray ionisation involves transfer of ions in solution into gaseous phase ions at atmospheric pressure by application of charge. B- Matrix-Assisted Laser Desorption Ionisation involves sublimation and ionisation of the analyte from a dry crystalline matrix using laser pulses.
1.3.1.2 Mass Analyser

The mass analyser separates ions by \( m/z \) ratio enabling detection. The mass analyser is, in the large part, responsible for the performance of the mass spectrometer and the quality of the data output. The most commonly used mass analysers in proteomics are time-of-flight (TOF), quadrupole (q or Q) and ion traps.

1.3.1.2.1 Time-of-Flight Mass Analyser

Time-of-flight mass analysers determine \( m/z \) by measuring the time ions take to travel a set distance through a field free region. The kinetic energy of an ion can be calculated using the accelerating potential and the mass and charge of the ion. As the same accelerating potential is applied to all ions the kinetic energy of an ion is a function of the mass and charge. Ions of the same charge but different mass travel at different velocities and so reach the detector at different times (Figure 1.3.3-A).

Mass resolution can be poor due to spatial and temporal differences in formation of the ions, and variations in flight times of ions with similar \( m/z \) ratios. Poor mass resolution can be overcome by use of delayed extraction, where ions are held in a field free region until accelerated by the application of a voltage pulse. Variations in flight times can also be avoided by use of reflectrons which deflect high energy ions. The length of the flight path is increased proportionally to energy of the ion and high energy ions arrive at the detector at the same time as lower energy ions of the same \( m/z \). TOF mass analysers are commonly coupled to MALDI ionisation sources. The used of delayed extraction is ideally applied to the pulsed formation of ions at each laser shot when using MALDI.

1.3.1.2.2 Quadrupole Mass Analyser

Quadrupole mass analysers separate ions using direct current and alternate current or radio frequency. Fields are established across pairs of parallel rods of opposite charge arranged around a central axis (Figure 1.3.3–B). Ions are introduced to the quadrupole continuously and filtered by \( m/z \). The positive rods allow ions above a selected \( m/z \) to pass through the central axis while the negative rods allow ions below a selected \( m/z \) to pass. The arrangement of the rods in a quadrupole overlaps the mass filter regions of the positive and negative rods. This region only allows ions between the selected \( m/z \) values to continue to the detector. Ions with \( m/z \) outside this range hit the rods and do not pass through quadrupole. Altering the current and radio frequency applied to the rods alters the range of the mass filter regions and so the resolution of separation that can be achieved. Scanning of the quadrupole mass analyser can be carried out by changing the amplitude of current and radio frequency at a constant ratio allowing a range of \( m/z \) ratios to be observed. Quadrupole mass analysers are commonly coupled to ESI sources. Ions are produced continuously by ESI and the quadrupole is able to filter \( m/z \) of the generated ions on a continuous basis.
1.3.1.2.3 Ion Trap Mass Analyser

An ion trap mass analyser collects a continuous stream of ions in a three dimensional electric field (Figure 1.3.3-C). Ions are trapped until the space charge limit; the number of ions which can be held in the field without distorting it is reached. The m/z ratio of the ions is determined by application of a radio frequency voltage which causes ions to oscillate. The frequency of oscillation is a function of the m/z value of the ion. To scan a range of m/z ratios, the radio frequency voltage is increased. The frequency of oscillation of the ions increases until the ion becomes destabilised and is ejected from the trap. As the oscillating frequency is a function of the m/z of an ion, ions with different m/z are ejected at different voltages and times. Ion trap mass analysers are robust and increasingly be used for proteomics, however the mass accuracy is relatively low (Hager 2002).
Figure 1.3.3  The mass analyser separates ions by $m/z$ ratio enabling detection.

A- Time of flight mass analysers determine $m/z$ by measuring the time ions take to travel a set distance. The use of delayed extraction, where ions are held in a field free region until accelerated and reflectrons which deflect high energy ions of the same $m/z$, can be applied to improve mass resolution. B- Quadrupole mass analysers separate ions using direct current and alternate current or radio frequency. Quadrupole mass analysers are commonly coupled to ESI sources. Ions are produced continuously by ESI and the quadrupole is able to filter $m/z$ of the generated ions on a continuous basis. C- Ion trap mass analyser collect a continuous stream of ions in a three dimensional electric field. Radio frequency voltage is applied causing ions to oscillate. The frequency of oscillation is a function of the $m/z$ value of the ion.
1.3.1.3 Data Output

Once ions generated by the ion source have been filtered and separated by \( m/z \) by the mass analyser the ions are directed to a detector. The detector measures the numbers of ions at each \( m/z \). This information is converted to a signal and a mass spectrum produced by processing software (Figure 1.3.4). The relative intensity of ions detected at each \( m/z \) is plotted by the mass spectrum. This information can be used to determine the molecular weight of ions generated. Algorithms are used to process this data and infer structural information about the sample applied to the mass spectrometer.

Figure 1.3.4  A mass spectrum is produced by processing software plotting the \( m/z \) against the intensity of the signal detected at each \( m/z \).
1.3.2 Application of Mass Spectrometry to Proteomics

Analysis of multiple whole proteins by MS is complex: proteins are soluble under different conditions, sensitivity of the MS is limited and the vast number of modifications and isoforms of any one protein make identifying that protein based on molecular weight alone difficult. As a result proteins are commonly digested into peptides for MS analysis. Peptides have more homologous physiochemical properties than proteins and so are soluble in similar conditions. Identification of a protein (and isoforms and modifications) is more straightforward using the molecular weight of component peptides or the amino acid sequence. The identity of the peptides and so protein can be deduced from data acquired by MS. This information is most efficiently obtained for peptides of approximately 20 amino acid residues (Steen and Mann 2004).

The protease trypsin is most commonly used to digest proteins into peptides (Olsen et al. 2004). Trypsin is a stable and highly specific protease which cleaves proteins on the carboxy-terminus side of arginine and lysine residues. Cleavage products of trypsin are therefore highly predictable and can be used to identify the originating protein. Accurate predictions of the sequences of tryptic peptides can be made from the genome of an organism. Databases containing predicted sequences of tryptic peptides based on the genome and previous proteomic experiments have been populated and can be used to identify the peptides observed on LC-MS/MS analysis.

1.3.2.1 Peptide Mass Fingerprinting

Peptide mass fingerprinting is the simplest method of identifying an unknown protein by MS. The method utilises the molecular weight of tryptic peptides and is only applicable to single or very simple protein mixes. The method is reliant on the purity of the sample, the number of fragments obtained and the accuracy in measurement of the mass of those fragments (Clauser et al. 1999). Proteins are digested using trypsin, and the molecular weight of the peptides determined using MS. MS data obtained can then be compared to database entries to identify the protein observed by MS. Several commercially available programs including MASCOT (MatrixScience Ltd 2007), Sequest (Eng et al. 1994) and Protein Prospector (Baker and Clauser) utilise this data to infer the identity of proteins from MS data.
1.3.2.2 Tandem Mass Spectrometry

Tandem MS (MS/MS) utilises the ability of a mass analyser to select peptides of a certain m/z and the ability of a linear arrangement of mass analyser to further refine this selection. ‘Precursor’ ions of a determined m/z are selected by the first mass analyser (Figure 1.3.5-A). The precursor ions are directed into a collision chamber and further induced to fragment by collision with an inert gas (Nitrogen, Argon or Helium). This process is known as Collision Induced Dissociation (CID) and results in ‘product’ ions. The product ions are directed to a second mass analyser and further refined by m/z prior to detection. Common combinations of mass analysers include TOF and TOF, often in combination with MALDI as the ionisation source (MALDI-TOF-TOF), ESI a quadrupole and TOF (qTOF) and three quadrupole mass analysers (QQQ) where the central quadrupole analyser is used as a collision chamber.

Collision induced fragmentation of ionised precursor peptides selected by the first mass analyser occurs along the peptide backbone producing a series of product ions. Product ions are directed to the second mass analyser and the m/z of the product ions determined. The nature of the precursor ion is dependent on the ionisation method used: ESI produces ions with multiple charge while MALDI results in singly charged ions. The charge state of the precursor ion determines the fragmentation pattern which results in product ion peptides. Multiple protons on the charged precursor ion peptide cause sequential fragmentation along the peptide backbone resulting in a majority of N and C-terminus containing peptides while the single proton of MALDI induced ions results in less regular fragmentation (Cramer and Corless 2001). Biemann determined nomenclature for identifying these product ions: product ions containing the N-terminus were determined a, b and c, and the C-terminus x, y and z (Figure 1.3.5-B) (Biemann 1988). The difference between the mass of b and y ions corresponds to the mass of the amino acid lost and can be used to determine the sequence of the peptide (Figure 1.3.5-C).

The amino acid sequence of the tryptic peptides present in the sample can be determined. This information is applied to search engine software, for example MASCOT, which utilise protein sequence databases to determine the identity of the proteins in the sample analysed. The confidence of the protein identification returned by the software is dependent on the resolution of the instrument and settings used for the database search. The ability of the MS instrument to accurately measure mass must be reflected in the stringency used in matching observed and predicted protein sequences. The presence of incompletely or unexpectedly digested proteins and post translational modifications (arising naturally or as a result of sample processing) must also be taken into account when assessing the confidence in protein identification.
1.3.2.2.1 MASCOT: An Example of Search Engine Software

MASCOT (Matrix Science, London, UK) is a commonly used search engine which uses MS data to identify proteins from primary sequence databases by several methods. MASCOT uses the MOlecular Weight SEarch (MOWSE) algorithm to search a database of calculated molecular weights derived from a set of specific enzyme or reagent cleavage rules to analyse peptide-mass fingerprint data (Pappin et al. 1993). By specifying the enzyme used to digest the protein, fidelity of the enzyme and approximate mass range of the peptides, matching mass values can be calculated on a probabilistic basis (MatrixScienceLtd 2007). The probability of a match is reported as a score (calculated by \(-10^{10\times \log_{10}(P)}\)) where scores greater than 67 are significant \((p=0.05)\) (Pappin et al. 1993). MS/MS data can also be analysed in a similar manner, peak list and intensity data pairs are searched. This MS/MS ion search is most commonly used to analyse LC-MS/MS data where matches to multiple peptides from a single protein indicate a high level of confidence in a correct result (MatrixScienceLtd 2007).

The most powerful search available via MASCOT is the Sequence Query where one or more peptide molecular masses are combined with sequence, composition and fragment ion data (MatrixScienceLtd 2007). A ‘sequence tag’, the combination of a small amount of sequence data and molecular weight information, is used to search a database which can be up to 1 million-fold more discriminating than using partial sequence information alone and is able to identify peptides in the presence of unknown post-translational modifications and amino acid substitutions (Mann and Wilm 2002). A further error tolerant search can also be carried out, where unmatched spectra due to, for example; enzyme non-specificity, underestimated mass, unknown chemical or post-translational modifications, or inaccurate determination of precursor charge, can be carried out (Creasy and Cottrell 2002). The search is carried out without determining enzyme specificity and sequentially tests a comprehensive list of chemical and post-translational modifications in order to avoid loss of discrimination (Creasy and Cottrell 2002).
Figure 1.3.5  Tandem mass spectrometry can be used to identify proteins from complex mixtures. A- Precursor ions of a determined m/z are selected by the first mass analyser. B- Collision induced fragmentation of ionised precursor peptides selected by the first mass analyser occurs along the peptide backbone producing a series of product ions. Product ions containing the N-terminus are determined a, b and c, and the C-terminus x, y and z (Biemann 1988). C- Product ions are directed to the second mass analyser and the m/z of the product ions determined. The difference between the mass of b and y ions corresponds to the mass of the amino acid lost and can be used to determine the sequence of the peptide.
1.3.2.3 Sample Processing for Tandem Mass Spectrometry

Tandem MS can be applied to tryptic digests of complex protein mixtures including plasma. As previously discussed the dynamic range of plasma exceeds the sensitivity of MS. In order to identify proteins through a wider dynamic range plasma undergoes sequential processing prior to analysis. A typical sample processing ‘workflow’ involves removal of high abundance proteins or enrichment for proteins of interest, protein digestion using trypsin, and further peptide fractionation using HPLC or peptide enrichment and orthogonal separation of peptides directly prior to MS/MS analysis (LC-MS/MS) (Qian et al. 2006).

1.3.2.3.1 Immunodepletion of High Abundance Proteins

Immunoadfinity-based protein subtraction chromatography (immunodepletion) removes highly abundant proteins from plasma to reduce the dynamic range. Plasma is applied to immobilised affinity purified polyclonal antibodies against a number of high abundance proteins. Various commercial products are available which target the top 12-14 high abundance proteins, for example the Seppro® IgY 14 -SuperMix Liquid Chromatography Column system (Genway, San Diego, USA) and the Agilent Multiple Affinity Removal LC Column (Human 14; Agilent, UK). Immunodepletion is high throughput and highly specific and so commonly used in proteomic studies (Echan et al. 2005; Zolotarjova et al. 2005; Whiteaker et al. 2007).

1.3.2.3.2 High Performance Liquid Chromatography

High-performance liquid chromatography (HPLC) fractionation is commonly applied to reduce the complexity of samples (Gilar et al. 2005). HPLC involves the application of sample to a stationary support phase engineered to form reversible interactions with amino acid side chains, which can subsequently be modified by application of a mobile phase. HPLC can be applied in multiple dimensions, producing simplified fractions separated in each dimension by a different property.

Reverse phase (RP) HPLC separates peptides according to reversible hydrophobic interactions between the amino acid side chains of the peptide and the hydrophobic alkyl chains in the stationary phase (Simpson 2003; Steen and Mann 2004). Peptides are loaded on to the stationary phase under favourable binding conditions and eluted into fractions according to hydrophobicity by modification of the mobile phase (Simpson 2003). RP-HPLC carried out at high pH followed by RP-HPLC at low pH has been shown to give greatest separation (Gilar et al. 2005).
1.3.2.4 Multiple Reaction Monitoring

Multiple reaction monitoring (MRM) is a highly sensitive targeted method for identification of proteins. Pre-existing knowledge of the $m/z$ of the product ions of the proteins to be investigated is required. MRM commonly utilises triple quadrupole (QQQ) mass spectrometers with online chromatography. Peptides are injected onto an LC column and ionised using ESI. Ionised peptides of a predetermined $m/z$ are selected by the first quadrupole (Figure 1.3.6-A). Selected peptides are directed to the second quadrupole which acts as a collision chamber to further fragment the peptide into product ions. The combination of precursor and product ion $m/z$ determines the protein identified and is known as a ‘transition’ (Figure 1.3.6-B). For confident identification of a protein, a minimum of two transitions are commonly used (Sherman et al. 2009a). Multiple transitions can be observed in a short space of time meaning multiple proteins can be identified during one MS analysis. The product ions are directed toward the third quadrupole which selects product ions by $m/z$ and directs the ions toward the detector. Data output appears as a plot of time of elution of the selected peptide during online chromatographic separation against intensity of ion detection (Figure 1.3.6-C) and the area under the peak is calculated (Figure 1.3.6-D).

Selection of optimal transitions must take into account multiple factors: the peptide must not contain missed enzyme cleavage sites, sites of potential post transitional modifications (unless of interest), the product and precursor ions must fall within the mass range of the MS instrument (7-30 amino acids is ideal) and the transitions should uniquely identify the peptide of interest. In order to maximise sensitivity, the MS instrument should be optimised for each transition. The selection and optimisation of transitions can be aided by software which uses sophisticated algorithms including MRMPilot (Applied Biosystems), MassHunter Optimiser (Agilent) and Skyline (MacLean et al. 2010). Input of MS data and a degree of manual selection and refinement is still required to identify optimal transitions.
Figure 1.3.6. Multiple reaction monitoring is a highly sensitive targeted method for identification of proteins using a triple quadrupole mass spectrometer. 

A - Precursor ions of a pre-determined m/z are selected by the first quadrupole. B - Precursor ions are fragmented in the second quadrupole. C - Product ions are monitored at the elution time for the precursor ion by the third quadrupole producing a plot of signal intensity against time. D - A minimum of two transitions are commonly used for each product ion. The area under the peak for the measurement of each transition is summed for each product ion and compared between samples.
1.3.3 Quantitative Analysis using Mass Spectrometry

Mass spectrometry is not inherently quantitative; information relating to protein abundance is lost during LC-MS/MS analysis. A sample produced by digestion of a single protein analysed by MS should contain the constituent peptides at an equimolar ratio, however this relationship is not reflected in the mass spectrum. The heights of the peaks plotted on the mass spectrum represent the intensity with which the ions are detected. The size of the peaks does not reflect the quantity of the peptide in the sample due to differences in solubility, protease activity, ionisation efficiency and technical drift of the MS instrument. This effect also prohibits the comparison of spectra both within and between LC-MS/MS experiments. A variety of methods utilising stable isotopes have therefore been developed to quantify proteins by MS.

1.3.3.1 Stable Isotope Labelling

Quantitative mass spectrometry methods involve chemical modification of either proteins or peptides with stable isotopes prior to sample processing. The resulting labelled peptides are identical to the unlabelled peptides in all but mass. Labelled and unlabelled peptides therefore can be combined and behave identically during sample processing and on application to the mass spectrometer. The addition of the stable isotope means peptides have different m/z and so appear as discrete peaks in the mass spectrum. Comparison of the labelled and unlabelled peaks in the mass spectrum provides an estimate of the relative abundance between the peptides in the labelled and unlabelled sample.

Stable isotope labels can be introduced to samples at both the protein and peptide level. Where samples are obtained from culture, stable isotope labelling in culture (SILAC) can be used (Ong et al. 2002; Parker et al. 2009). A heavy amino acid (for example; d₃ leucine, ¹³C₆ arginine or ¹³C₆ lysine) is introduced to the sample in media during culturing. The sample must be cultured with the heavy amino acid until isotopic equilibrium is reached. This method allows up to three samples to be compared but cannot be applied to primary material and is reliant on protein turnover for uptake and incorporation of the heavy amino acid. As previously discussed, proteins are digested using trypsin prior to MS analysis and this process requires H₂O. By digesting proteins in heavy water (H₂¹⁸O), ¹⁸O can be incorporated into all newly digested peptides allowing comparison of two samples from any source (Yao et al. 2001; Reynolds et al. 2002). Modification of the N terminus lysine residue of a peptide to homoarginine using methylisourea can also be used to incorporate a heavy isotope label (Brancia et al. 2004). These methods rely on the efficiency of incorporation of the label which introduces a relatively small mass difference. Observation of the change in mass on labelling may be obscured by the naturally occurring range in isotopes in a sample and further confused by the presence of unlabelled peptides. Isotope coded affinity tags (ICAT) chemically modify peptides by binding to cysteine residues (DeSouza et al. 2005; Wu et al. 2006; Bantscheff et al. 2007). The tags contain cleavable biotin groups which can be
used for affinity purification ensuring only labelled peptides are quantified. The mass
difference of the tags is generated with multiple deuterium or \(^{13}\)C\(_6\), ensuring a larger and
more determinable shift in mass. This method is also limited to the comparison of two
samples and requires the presence of cysteine which is a relatively low abundance amino
acid.

**1.3.3.1.1 Isobaric Tags for Relative and Absolute Quantification**

Development and use of the commercially available 'Isobaric Tags for Relative and
Absolute Quantification' (iTRAQ) was first reported by Ross et al. for the analysis of the
Saccharomyces cerevisiae proteome (Ross et al. 2004). The labels bind to the primary
amino group of tryptic peptides and the \(\varepsilon\)-amino group of lysine ensuring almost all peptides
are labelled. iTRAQ is unique; each tag is isobaric and so labelled peptides behave
identically during sample preparation and fractionation. Each tag consists of a reporter
group, a balance group and an amine specific peptide reactive group and so has equal mass
and chemical structure. The tags are designed, so that during MS/MS analysis of the
peptide, the tag fragments (Figure 1.3.7.-A) The reporter group is released on CID producing
reporter ions of specific \(m/z\) which can be used for relative quantification (Thompson et al.
2003). Multiplexing of samples is possible with kits available with 4 or 8 isobaric labels.
Samples are labelled following digestion and subsequently pooled and processed as one
which reduces the variation introduced by processing. Labelled peptides from each sample
behave identically until fragmented in the collision cell, where the reporter ions are released.
Reporter ions appear in the 'quieter' low \(m/z\) region of the spectra (\(m/z\) 114-117 for 4 plex,
\(m/z\) 113-119 and 121 for 8 plex kits) where peaks for few other fragmented peptides appear
(Figure 1.3.7.-B). The relative intensity of the spectra of the reporter ions can be used to
determine the relative abundance of each peptide in the original sample.

Wu et al. carried out a comparative study of DIGE, ICAT and iTRAQ using a
standard six protein mix spiked into cell lysates (Wu et al. 2006). They concluded that all
three techniques returned quantitative data of 'reasonable accuracy'. The DIGE method was
complicated by the identification of multiple proteins in differential spots compromising
accuracy in quantification. Quantification sensitivity was determined by the number of
peptides detected for each protein; DIGE was found to be least sensitive and similar to ICAT
while iTRAQ was found to be most sensitive. It was noted however, that the iTRAQ method
is prone to errors in precursor ion isolation which can be avoided by manual inspection of
the data. DeSouza et al. used ICAT and iTRAQ to identify changes in protein expression in
dermatial tissue in women with and without endometrial cancer (DeSouza et al. 2005).
Nine potential markers were identified using both techniques. It was suggested that the
methods were complementary, however the use of ICAT resulted in identification of more
low abundance proteins due to the selective nature of labelling and iTRAQ in more high
abundance proteins. Due to the ability to analyse multiple samples and obtain data on
protein identification and quantitation during one analysis, iTRAQ has been used extensively in the study of a range of diseases including pre-eclampsia (Blankley et al. 2009; Auer et al. 2010; Chen et al. 2011).

1.3.3.1.2 Quantitative Analysis by Multiple Reaction Monitoring

MRM can also be used for quantitation using an appropriate standard. For each peptide of interest a synthetic stable isotope labelled peptide is required. The synthetic peptide is spiked into the sample at known concentration, the labelled and endogenous peptide elute at the same time during LC separation. The relative intensity of the spectra for the endogenous and labelled peptide can be used to determine the abundance of the exogenous peptide in the original sample. Minimal preparation is required to measure through a dynamic range of 4.5 orders of magnitude, which can be increased on coupling to enriching preparative methods (Anderson et al. 2004). Data obtained by MRM is reproducible; Anderson and Hunter measured each of 53 high and medium abundance plasma proteins over 10 analyses and recorded a maximum coefficient of variation of 22% (Anderson and Hunter 2006).
Isobaric Tags for Relative and Absolute Quantification (iTRAQ) consist of a reporter group, a balance group and an amine specific peptide reactive group.

A - The tags bind to the amino group of tryptic peptides. Each tag is isobaric and so isobaric peptides behave identically during sample preparation and fractionation. Labelled precursor ions from each sample are selected together during MS. B - The reporter group of the tag is released on CID. Reporter ions appear in the low m/z region of the spectra where peaks for few other fragmented peptides appear.
2 Aims

There is currently no clinically useful screening test available to identify nulliparous women at high risk of developing pre-eclampsia prior to clinical presentation. There are changes in expression of plasma proteins in early pregnancy before a diagnosis of pre-eclampsia can be made. Targeted investigations of changed proteins to identify predictive markers of pre-eclampsia have so far been unsuccessful. A global, unbiased, proteomics strategy to identify proteins of predictive value is therefore proposed. Several studies have demonstrated the ability of quantitative mass spectrometry to identify changes in plasma proteins in pre-eclampsia in late pregnancy however the technique has yet to be applied to the analysis of early pregnancy plasma. This investigation aims to identify proteins that are changed in abundance at 15 weeks gestation and have predictive value prior to onset of pre-eclampsia using iTRAQ.

The aims of this investigation are:

- Development and optimisation of a workflow for analysis of early pregnancy plasma samples using iTRAQ.

- Identification and validation of changes in plasma proteins in early pregnancy in women who develop pre-eclampsia compared to healthy controls.
3 Methods

The application and context of the methods used is discussed in the following chapters. Detailed protocols for methods used are reported below. HPLC-grade water and Acetonitrile were supplied by BDH (Leicestershire, UK). All other chemicals were analytical grade and supplied by Sigma Aldrich unless otherwise stated (Dorset, UK).

3.1 Patient Samples

3.1.1 Plasma Sample Collection

Venepuncture was performed according to a standardised protocol; plasma samples were collected into BD EDTA-Vacutainer® tubes (BD, USA), centrifuged at 1500 x g at 4°C for 10 minutes and stored at -80°C. Samples were processed within 3 hours of collection. Plasma samples were defrosted on ice and maintained below 4°C during experiments wherever possible.

3.1.2 Early Pregnancy Plasma Samples from The SCOPE Study

As previously described, early pregnancy plasma samples were acquired via the SCOPE Study (Australian and New Zealand Clinical Trials Registry ACTRN12607000551493; The SCOPE Pregnancy Research Study 2004). Healthy, nulliparous women, with a singleton pregnancy were recruited and tracked throughout pregnancy. Full ethical approval was obtained and all patients gave written informed consent. Venepuncture was performed at 15±1 week’s gestation.

The definition of pre-eclampsia used in the research carried out was taken from the guidelines for diagnosis issued by The ISSHP (see 1.1.2.1). Plasma samples were obtained from women who subsequently developed late onset pre-eclampsia (> 34 weeks; n=12) and early onset pre-eclampsia (<34 weeks; n=12). Control samples matched to each group of case sample were also obtained from women with uncomplicated pregnancies (control group 1 and control group 2, each n=12). Maternal and neonatal characteristics are detailed in Table 3.1.1.
Table 3.1.1  Maternal characteristics and pregnancy outcome for women contributing plasma to the SCOPE study. Values are presented as median and range. Significant differences between the pre-eclampsia and matched control group were assessed using Mann-Whitney U and are indicated by *.

<table>
<thead>
<tr>
<th>Study Characteristics</th>
<th>Late onset pre-eclampsia (n=12)</th>
<th>Late Control (n=12)</th>
<th>Early onset pre-eclampsia (n=12)</th>
<th>Early Control (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>26 (20-31)</td>
<td>28 (15-40)</td>
<td>32 (24-40)</td>
<td>24 (14-34)</td>
</tr>
<tr>
<td>BMI at booking (kg/m²)</td>
<td>34 (20-49)</td>
<td>25 (17-34)</td>
<td>28 (21-36)</td>
<td>33 (21-45)</td>
</tr>
<tr>
<td>Caucasian</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Systolic blood pressure at 15 weeks (mm Hg)</td>
<td>110 (86-134)</td>
<td>110 (90-130)</td>
<td>120* (100-140)</td>
<td>105 (90-120)</td>
</tr>
<tr>
<td>Diastolic blood pressure at 15 weeks (mm Hg)</td>
<td>64 (52-76)</td>
<td>62 (50-74)</td>
<td>81* (70-92)</td>
<td>64 (58-70)</td>
</tr>
<tr>
<td>Maximum Systolic blood pressure (mm Hg)</td>
<td>155* (140-170)</td>
<td>121 (110-132)</td>
<td>170* (140-200)</td>
<td>115 (100-130)</td>
</tr>
<tr>
<td>Maximum Diastolic blood pressure (mm Hg)</td>
<td>105* (90-120)</td>
<td>72 (110-132)</td>
<td>107* (90-125)</td>
<td>74 (60-88)</td>
</tr>
<tr>
<td>Gestation at diagnosis (weeks+days)</td>
<td>36+5 (32+4-40+5)</td>
<td>28+7 (24+1-33+5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gestation at delivery (weeks+days)</td>
<td>37+6* (35+0-40+5)</td>
<td>39+6 (38+2-41+3)</td>
<td>31+2* (28+4-33+6)</td>
<td>40+4 (39+1-41+6)</td>
</tr>
<tr>
<td>Birthweight (g)</td>
<td>2725 (1610-3840)</td>
<td>3830 (3080-4580)</td>
<td>1820* (980-2660)</td>
<td>3783 (3190-4375)</td>
</tr>
<tr>
<td>Birthweight Centile</td>
<td>47 (0-94)</td>
<td>64 (32-64)</td>
<td>47* (0.3-94)</td>
<td>59 (27-91)</td>
</tr>
</tbody>
</table>
3.1.3 Plasma Samples used for Method Development

Venous blood samples were obtained from healthy women with singleton pregnancies booked for antenatal care and delivery at the Ninewells Hospital, Dundee as part of a previous study (Myers 2005). Approval was given by the Dundee and Manchester Local Research Ethics committees and written informed consent was obtained from patients for all samples. Plasma samples were obtained at 22-26 weeks gestation and were prepared by laboratory staff at the University of Dundee. Samples were transported to Manchester on dry ice. Non-pregnant samples were obtained with informed consent from male and female healthy volunteers. Plasma samples were used during development of methods and as a reference and were not used for identification of prognostic biomarkers.

3.2 Identification of Candidate Markers by iTRAQ

The mass spectrometry method iTRAQ was applied to plasma sampled in early pregnancy from women who developed pre-eclampsia compared to women who had uncomplicated pregnancies.

3.2.1 Plasma Pooling

Plasma samples acquired from the SCOPE Study (see 3.1.2) were pooled. Four phenotypic pooled plasma samples were produced; late pre-eclampsia, early pre-eclampsia, control group 1 and control group 2. Each sample group contained 12 individual plasma samples, 25 µl of each of these samples was used to produce a pooled sample for each group. A further reference pool containing all samples was also produced and referred to as the ‘superpool’. One aliquot from each of the four phenotypic pools was used to produce the superpool.

3.2.2 Sample Preparation: IgY 14-SuperMix Immunodepletion Workflow

Pooled plasma samples were depleted of high abundance proteins using the IgY 14-SuperMix immunodepletion system and prepared for analysis as described below.
3.2.2.1 IgY 14-SuperMix Immunodepletion Workflow Overview

Pooled plasma acquired from the SCOPE study (see 3.1.2) was immunodepleted using the IgY-SuperMix Immunodepletion system. Two superpool, late and early onset pre-eclampsia pooled samples were immunodepleted; one pooled plasma sample of each control group was immunodepleted. For each pooled plasma sample two injections of 150 µl were made to the column; for each sample a total of 300 µl of plasma was immunodepleted. The eluent from each 150 µl injection were combined, concentrated and exchanged into a suitable buffer by ultrafiltration. Protein concentration was quantified by protein assay and completeness of depletion assessed by SDS-PAGE (see 3.3.2.1.1) with Coomassie stain (see 3.3.2.1.2). Protein (80 µg) was digested and iTRAQ labelled and the efficiency of these reactions assessed using the Esquire HCT Ion trap. Combined iTRAQ labelled samples were separated into multiple fractions by high pH RP-HPLC, fractions were collected and applied to the QSTAR XL ESI qTOF and 5800-MALDI TOF-TOF for analysis.

3.2.2.1.1 IgY-SuperMix Immunodepletion

Plasma was subjected to depletion of high abundance proteins using the Seppro® IgY 14-SuperMix Liquid Chromatography Column system (Genway, San Diego, USA). The system consists of two columns; the IgY 14 column and the SuperMix column, which are run sequentially. The IgY 14 (12.7 x 79mm) column removes fourteen of the most abundant plasma proteins (albumin, immunoglobulin (Ig) G, alpha-1-antitrypsin, IgA, IgM, transferrin, haptoglobin, alpha-1-acid glycoprotein, alpha-2 macroglobulin, HDL (Apo A-I & A-II), fibrinogen, complement C3 and LDL (mainly Apo B)) and the SuperMix Column (6.4 x 63mm) removes approximately 75 undefined moderately abundant plasma proteins. Manufacturer’s instructions were followed using a 2795 Separation Module (Waters, Milford, MA) and a 783A Programmable Absorbance Detector (Applied Biosystems, Warrington, UK) to measure UV absorbance of eluting proteins at 280nm.

The liquid chromatography system was prepared by wet priming with buffers (Table 3.2.1) and following addition of the columns, rinsed with dilution buffer (Table 3.2.1) to achieve a stable UV reading. Plasma (100-200µl) was centrifuged at 4°C for 10 minutes at 10,000rpm, diluted to 1mL using dilution buffer and passed through 0.22µm Spin Filters (Agilent, UK) at 10,000rpm for 1 minute. Plasma was injected using a 200 µl injection loop and Program A (Table 3.2.2) applied. Flow-through was collected into pre-cooled tubes between 16 and 26 minutes, stored on ice for the duration of the experiment and at -20 °C until further use. The IgY 14 column was removed and Program B (Table 3.2.2) applied to strip and re-equlibrate the SuperMix column (eluant was collected between 8 and 17 minutes). The SuperMix column was replaced with the IgY 14 column and Program C (Table 3.2.2) applied to strip and re-equlibrate the IgY 14 column.
Table 3.2.1  Composition of Buffers for Immunodepletion of plasma samples using the Seppro® IgY-SuperMix Liquid Chromatography Column system.

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Buffer Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dilution</td>
<td>10 mM Tris-HCl, 0.15M NaCl, pH 7.4</td>
</tr>
<tr>
<td>Stripping</td>
<td>0.1 M Glycine, pH 2.5</td>
</tr>
<tr>
<td>Neutralization</td>
<td>0.1 M Tris-HCl, pH 8.0</td>
</tr>
</tbody>
</table>

Table 3.2.2  Immunodepletion of plasma samples using the Seppro® IgY-SuperMix Liquid Chromatography Column system. Pooled plasma samples as previously described were depleted of high abundance proteins. Manufacturer's instructions were followed using buffers detailed in Table 3.2.1. Proteins bind to the columns during Program A, collected eluant is depleted plasma. Program B strips bound proteins and re-equilibrates the SuperMix Column, Program C strips and re-equilibrates the IgY14 Column.

Program A- Column Loading

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Time (min)</th>
<th>Buffer</th>
<th>Flow Rate (mL/min)</th>
<th>Max Pressure (psi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Injection</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wash</td>
<td>15</td>
<td>Dilution</td>
<td>0.5</td>
<td>200</td>
</tr>
<tr>
<td>Wash</td>
<td>25</td>
<td>Dilution</td>
<td>1.0</td>
<td>200</td>
</tr>
</tbody>
</table>

Program B- SuperMix Wash

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Time (min)</th>
<th>Buffer</th>
<th>Flow Rate (mL/min)</th>
<th>Max Pressure (psi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wash</td>
<td>5</td>
<td>Dilution</td>
<td>2.0</td>
<td>200</td>
</tr>
<tr>
<td>Elution</td>
<td>14</td>
<td>Stripping</td>
<td>2.0</td>
<td>200</td>
</tr>
<tr>
<td>Neutralisation</td>
<td>21</td>
<td>Neutralisation</td>
<td>1.0</td>
<td>200</td>
</tr>
<tr>
<td>Re-equilibrium</td>
<td>28</td>
<td>Dilution</td>
<td>1.0</td>
<td>200</td>
</tr>
</tbody>
</table>

Program C- IgY 14 Wash

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Time (min)</th>
<th>Buffer</th>
<th>Flow Rate (mL/min)</th>
<th>Max Pressure (psi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wash</td>
<td>4</td>
<td>Dilution</td>
<td>2.0</td>
<td>200</td>
</tr>
<tr>
<td>Elution</td>
<td>25</td>
<td>Stripping</td>
<td>2.0</td>
<td>200</td>
</tr>
<tr>
<td>Neutralisation</td>
<td>35</td>
<td>Neutralisation</td>
<td>2.0</td>
<td>200</td>
</tr>
<tr>
<td>Re-equilibrium</td>
<td>45</td>
<td>Dilution</td>
<td>2.0</td>
<td>200</td>
</tr>
</tbody>
</table>
3.2.2.1.2 Ultrafiltration

Depleted samples were concentrated using 5kDa molecular weight cut-off (MWCO) spin columns (Agilent, UK). Samples were applied to the column and spun at 4,000 rpm, at 4 ºC, until 300 µl remained. Columns were rinsed twice by application of 10mLs of 0.5M triethylammonium bicarbonate (TEAB) and spun for a further 30 minutes. Approximately 300 µl of concentrated sample were recovered.

3.2.2.1.3 Analysis of IgY-SuperMix Processed Samples

The completeness of immunodepletion and protein concentration was assessed for each sample.

3.2.2.1.3.1 Coomassie Protein Assay

The concentration of protein in each sample were determined using Coomassie Protein Assay Reagent (Thermo Fisher Scientific, UK). Standards (0-1mg/mL) were prepared using bovine serum albumin (BSA) dissolved in TEAB. Coomassie Protein Assay Reagent (1.5mL) and sample or standard (25 µl) were mixed and incubated for 10 minutes at room temperature. Measurements were made in triplicate at 595 nm using a Jenway 6300 Spectrophometer calibrated using TEAB. Concentration of samples was calculated by plotting a standard curve of concentration against blank-adjusted optical density measurement of the standards.

3.2.2.1.3.2 In-Gel Digest

Following ultrafiltration, completeness of depletion was assessed by SDS-PAGE (see 3.3.2.1.1) with Coomassie stain (see 3.3.2.1.2). Bands of interest were excised from a destained SDS-PAGE gel and rinsed twice by agitation in water for 10 minutes. Bands were de-stained by incubation with agitation in 500µl 100mM ammonium bicarbonate (AmBic) solution, followed by the addition of 500µl acetonitrile (ACN) and a further 10 minute incubation with agitation; this was repeated 8 times or until the gel band was completely destained. De-stain solutions were removed and the band incubated in reduction solution (4mM DL-Dithiothreitol (DTT) in 50mM AmBic) for 10 minutes at 60ºC. Once cooled to room temperature and following addition of 175µl 200mM iodoacetamide (IAM), bands were incubated for 30 minutes in darkness. The reduction reaction was quenched by addition of 4µl 100mM DTT and incubation for 5 minutes. Reduction reagents were removed and gel bands incubated with agitation in 500µl 100mM AmBic for 10 minutes. Gel bands were incubated in 500µl 100mM AmBic, followed by addition of 500µl 100% ACN and incubation with agitation for a further 10 minutes. The gel bands were dehydrated by removal of all solution and incubation in 100µl 100% ACN with agitation for 15 minutes, followed by air drying until white and opaque.
Proteins were digested by addition of 0.6µg trypsin in 40mM AmBic and incubated for
15 minutes. After removal of residual trypsin solution, bands were incubated for 5 minutes in
a volume of 40mM AmBic in 5% ACN which covered the bands by a minimum of 2mm. Once
suitably re-hydrated, bands were incubated at 37ºC with agitation for 16-18hours. Following
addition of an equal volume of ACN and further 20 minutes incubation, supernatant was
collected and dried using a SpeedVac (Thermo Savant, UK) to 50% volume to remove the
ACN and stored at -20ºC until further use.

3.2.2.1.3.3 Zip Tip Cleanup

ZipTip® 10µl Pipette tips (Millipore, MA, USA) were equilibrated by washing with
equilibration solution (0.1% TFA). Peptides were bound by up to 10 cycles of aspiration and
dispensing of sample and washed with washing solution (0.1% trifluoroacetic acid (TFA)).
Peptides were eluted by washing in the minimum volume of elution solution (50% ACN in
0.1% TFA).

3.2.2.1.4 Protein Digestion and iTRAQ Labelling

The manufacturer’s protocol was followed to label peptides using 8 plex iTRAQ®
Reagents (Isobaric Tags for Relative and Absolute Quantification; Applied Biosystems,
Warrington, UK). Following ultrafiltration, samples were dried using a SpeedVac (Thermo
Savant, UK). Dried plasma samples (80 µg protein) were resuspended in 20 µl dissolution
buffer (0.5M TEAB, pH 8.5) followed by addition of 1µl denaturant (2% SDS) to disrupt
protein-protein interactions, 2µl reducing reagent (50mM Tris (2-carboxyethyl) phosphine
(TCEP)) to reduce disulphide bonds and incubation at 60 ºC for 1 hour. Cysteine blocking
reagent (1µl- 200mM methylmethane thiosulphonate in isopropanol (MMTS) was added and
incubated at room temperature for 10 minutes. Proteins were digested with trypsin (1:20 w/w
enzyme:substrate) for 12-16hrs at 37ºC and dried to a total volume less than 20µl. iTRAQ
reagents resuspended in 50-70µl isopropanol were transferred to the appropriate samples
and incubated at room temperature for 1 hour. On completion, labelled samples were
combined and dried (SpeedVac, Thermo Savant, UK).

3.2.2.1.5 High pH Reverse Phase Fractionation (3 µm Fortis C18 Column)

Offline high pH RP fractionation using a 3 µm Fortis C18 column (150 x 3.0mm)
(Fortis Technologies Ltd, Neston, UK) run on an Ultimate 3000 HPLC system (Dionex, CA,
USA) was used to separate labelled peptides. Dried and labelled samples were
resuspended in Buffer A (2% ACN, 0.1% ammonia solution). With the column maintained at
65 ºC and a flow rate of 425 µl/min, peptides were separated using a linear gradient (0-50%)
of solvent B (90% ACN, 0.1% ammonia solution) over 30 minutes. Fractions were collected
and dried.
3.2.3 Sample Preparation: MARS 14 Immunodepletion System Workflow

Pooled plasma samples were depleted of high abundance proteins using the MARS 14 immunodepletion system and prepared for analysis as described below.

3.2.3.1 MARS 14 Immunodepletion System Workflow Overview

Pooled plasma (see 3.1.2) was immunodepleted, in duplicate for case and superpool, and singly for each control pool. Depleted samples were concentrated using the mRP C18 column, dried and resuspended for digestion and iTRAQ labelling. The efficiency of depletion was assessed prior to digestion using the chromatographic trace from immunodepletion and the Protein 230 assay. Each iTRAQ label was applied to the processed product of 25µl of plasma. Combined iTRAQ labelled samples were separated into multiple fractions by high pH RP, fractions were collected and applied to the 5800-MALDI TOF-TOF for analysis.
3.2.3.1.1 MARS 14 Immunodepletion

Plasma was depleted of high abundance plasma proteins using the Multiple Affinity Removal LC Column- Human 14 (MARS 14; Agilent, UK). The MARS 14 column (4.6 x 100mm) removes up to 94% of total protein mass by removing the top 14 abundant plasma proteins (albumin, alpha-1-acid glycoprotein, alpha-2-macroglobulin, alpha-1-antitrypsin, Apo Al, Apo All, complement C3, fibrinogen, haptoglobin, IgA, IgG, IgM, transferrin and transthyretin). Manufacturer’s instructions and buffers were used with an Agilent 1200 Series HPLC system controlled using ChemStation (both Agilent, UK). UV absorbance was detected at 280nm.

The HPLC system was prepared by purging with Buffer A and B (Agilent, UK), purging the column with Buffer A and completion of a method blank (Table 3.2.3). Plasma was centrifuged at 4 ºC for 10 minutes at 10,000 rpm, diluted 4 fold in Buffer A and passed through a 0.22µm spin filter (Agilent, UK) at 10,000g for 1 minute to remove particulate matter. Plasma was injected at a flow rate of 500µl/min with switching from main-pass to bypass after 6 minutes and the recommended program applied (Table 3.2.3). The depleted flow-through fraction was collected between 9 and 16 minutes and the bound high abundance protein fraction between 22 and 22.5 minutes (see 4.1.4.3.1, Figure 4.1.4 p100 for trace). Collected fractions were maintained at 4 ºC by the automated sample collector.

Table 3.2.3 Immunodepletion of plasma samples using the MARS 14 column.

Pooled plasma samples as previously described were depleted of high abundance proteins. Manufacturer’s buffers were used with the recommended program to remove the top 14 high abundance proteins from plasma samples. Depleted plasma was collected between 9 and 16 minutes.

<table>
<thead>
<tr>
<th>Time (minutes)</th>
<th>% Buffer B</th>
<th>Flow (mL/min)</th>
<th>Maximum Pressure (bar)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.125</td>
<td>60</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>1</td>
<td>60</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>1</td>
<td>60</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
<td>1</td>
<td>60</td>
</tr>
<tr>
<td>38</td>
<td>0</td>
<td>1</td>
<td>60</td>
</tr>
</tbody>
</table>
3.2.3.1.2 Macroporous Reversed-Phase C18 Column

Depleted plasma was concentrated and desalted using a Macroporous Reversed-Phase C18 Column (mRP-C18, 4.6mm x 50mm, Agilent, UK). Manufacturer's instructions were used with an Agilent 1200 Series HPLC system (Agilent, UK). UV absorbance was detected at 280nm. The HPLC system was prepared by wet priming with appropriate buffers without, and with, the column. Depleted plasma was applied to the column maintained at 60°C and the recommended program applied (Table 3.2.4, main-pass to bypass after 2.75 minutes, Buffer A; 0.1% TFA, Buffer B; Acetonitrile, 0.1% TFA). The bound protein fraction was collected between 7 and 9 minutes and maintained at 4°C by the automated sample collector. Samples were dried using a vacuum drier (Eppendorf, UK) and stored at -20°C until further use or resuspended in 47.5µl TEAB (pH 8.5) and 1.5µl 2% SDS.

Table 3.2.4 Concentration and desalting of samples using a Macroporous Reversed-Phase C18 Column. Pooled plasma samples were depleted of high abundance proteins and concentrated and desalted using a Macroporous Reversed-Phase C18 Column. Manufacturer's recommended program was used with Buffer A; 0.1% TFA and Buffer B; Acetonitrile, 0.1% TFA, eluant was collected between 7 and 9 minutes.

<table>
<thead>
<tr>
<th>Time (minutes)</th>
<th>% Buffer B</th>
<th>Flow Rate (mL/min)</th>
<th>Maximum Pressure (bar)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
<td>0.750</td>
<td>200</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>0.750</td>
<td>200</td>
</tr>
<tr>
<td>5</td>
<td>70</td>
<td>0.750</td>
<td>200</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>0.750</td>
<td>200</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>0.750</td>
<td>200</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0.750</td>
<td>200</td>
</tr>
</tbody>
</table>
3.2.3.1.3 Analysis of MARS 14 Processed Samples

The completeness of immunodepletion and protein concentration was assessed for each depleted plasma sample.

3.2.3.1.3.1 Protein Assay

Protein concentration was assessed using Bio-Rad protein assay (BioRad, Hemel Hempstead, UK). To 1mL of a 1 in 5 dilution of Bio-Rad Protein Assay reagent, 35 µl of a 1 in 10 dilution of sample were added, mixed and incubated at room temperature for 10 minutes. Measurements were made in triplicate at 595nm (Jenway 6300 Spectrophometer) and concentration of protein calculated against standards (0-1mg/mL) prepared using BSA in appropriate buffer. Concentration of samples was calculated by plotting a standard curve of concentration against blank-adjusted optical density measurement of the standards.

3.2.3.1.3.2 Protein 230 Assay for Qualitative Assessment

Qualitative protein assessment of immunodepleted plasma was carried out using the Protein 230 Assay performed on a 2100 Bioanalyser microfluidics-based lab-on-a-chip platform (Agilent, UK). Manufacturer's protocols were followed; in brief samples were prepared by addition of Denaturing Solution, incubation at 100°C for 10 minutes and dilution by addition of ddH2O. Chips were prepared by injection of Gel-Dye mix and subsequently loaded with sample, Agilent Protein 200 Plus ladder Destaining Solution, and Gel-Dye mix (Agilent, UK). The chip was inserted into the 2100 Bioanalyser and analysed using 2100 Expert Software (Agilent, UK). The quality of each assay was assessed using the ladder electropherogram (Figure 3.2.1). Samples were analysed for reproducibility of immunodepletion by comparison in both gel and electropherogram view and completeness of immunodepletion by assessment of characteristic peaks using the electropherogram.
Figure 3.2.1 Agilent Protein 200 Plus ladder (Agilent, UK) in both electropherogram and gel view were used to assess the quality of each assay performed. Assay standardisation markers at 4.5 and 240kDa and standard proteins between 7 and 150kDa were measured as shown by the peaks in the electropherogram view. The peaks can also be viewed as bands on a simulated SDS-PAGE gel (see 3.3.2.1.1) stained with Coomassie Blue (see 3.3.2.1.2) for ease of analysis.

3.2.3.1.4 Digestion and iTRAQ Labelling

Depleted plasma samples were digested and peptides labelled according to the manufacturer’s protocol for plasma using 8plex iTRAQ® Reagents (Applied Biosystems, Warrington, UK). Samples to be labelled were derived from injection of equal volumes of plasma onto the MARS-14 LC column (see 3.2.3.1.1) and contained approximately 80µg of protein. Dried samples were resuspended in 47.5µl Dissolution buffer (1M TEAB, pH 8.5) and 1.5µl Denaturant (2% SDS) to disrupt protein-protein interactions. From each sample, 10µl was removed in order to assess completeness of abundant protein depletion (see 3.2.3.1.3.2). Proteins were then reduced by addition of 1.5µl Reducing Reagent (50mM (TCEP)) and incubation at 60 ºC for 1 hour. Cysteine residues were blocked by addition of 1.5ul 84mM IAM and incubation in the dark at room temperature for 30 minutes. Proteins were digested with proteomics grade trypsin (1:20 w/w enzyme:substrate, Sigma Aldrich, Dorset, UK) for 12-16hrs at 37ºC, dried to a total volume less than 5µl and resuspended with addition of 30ul Dissolution buffer. iTRAQ reagents, resuspended in 80µl isopropanol, were transferred to the plasma samples and incubated at room temperature for 2 hours; the reaction was quenched by addition of 115 µl ddH₂O. Samples were combined and dried (Eppendorf, UK).
3.2.3.1.5 High pH Reverse Phase Fractionation (3µm Extend-C18 Column)

To simplify the labelled peptide mixture and remove unbound iTRAQ reagents, peptides were fractionated and washed offline using a 3µm Extend-C18 column (4.6 x 100 mm; Agilent, UK) run on an ICS-3000 HPLC system (Dionex, CA, USA). Dried and labelled samples were resuspended in Buffer A (0.1% ammonia hydroxide), injected onto the column and washed for 1 hour at a flow rate of 0.4 mL/min with 0.5% Buffer B (Acetonitrile, 0.1% ammonia hydroxide, pH 10). Peptides were eluted using a non-linear gradient (Table 3.2.5) at a flow rate of 0.4mL/min. Fractions were collected every 15 seconds for 30 minutes, dried and stored at -20º.

Table 3.2.5  A non linear gradient was applied to a 3µm Extend-C18 Column to separate peptides by Reverse Phase Fractionation. Dried and labelled samples were resuspended in Buffer A (0.1% ammonia hydroxide) applied to the column and eluted in Buffer B (Acetonitrile, 0.1% ammonia hydroxide, pH 10); fractions were collected every 15 seconds.

<table>
<thead>
<tr>
<th>Time (minutes)</th>
<th>% Buffer B</th>
<th>Flow (mL/min)</th>
<th>Maximum Pressure (bar)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5</td>
<td>0.40</td>
<td>100</td>
</tr>
<tr>
<td>30</td>
<td>50</td>
<td>0.40</td>
<td>100</td>
</tr>
<tr>
<td>34</td>
<td>75</td>
<td>0.40</td>
<td>100</td>
</tr>
<tr>
<td>40</td>
<td>75</td>
<td>0.40</td>
<td>100</td>
</tr>
<tr>
<td>41</td>
<td>0.5</td>
<td>0.40</td>
<td>100</td>
</tr>
<tr>
<td>71</td>
<td>0.5</td>
<td>0.40</td>
<td>100</td>
</tr>
<tr>
<td>72</td>
<td>0.5</td>
<td>0.40</td>
<td>100</td>
</tr>
</tbody>
</table>
3.2.4 Mass Spectrometry

Methods for LC-MS/MS analysis are described as follows.

3.2.4.1 Analysis of Sample Preparation

Samples were assessed at several stages of each workflow using the methods and LC-MS/MS platforms detailed below.

3.2.4.1.1 Ultraflex MALDI TOF-TOF

Analysis using a MALDI-TOF TOF Mass Spectrometer (Bruker Daltonics, MA, USA). Samples were prepared for analysis using 1:1 HCCA as matrix (2mg/mL, 70% MeCN, 0.1% TFA) and calibrated using a standard peptide mix, (Bruker Calibration Mix 2, Bruker Daltonics, MA, USA). Laser power and tuning parameters were optimised for each sample plate. Standard Bruker methods were used for MS (RPC_Proteomics) and MS/MS (Lift); 4000 shots were acquired and summed together. Data were analysed using MASCOT (Matrix Science, London, UK; see 3.2.5.1).

3.2.4.1.2 Esquire HCT Ion Trap

Online Reverse Phase fractionation was carried out using a PepMap™ 100 (C18) Nanocolumn (Dionex, CA, USA) run on an Ultimate 3000 HPLC system (Dionex, CA, USA) coupled to an Esquire HCT Ion Trap (Bruker Daltonics, MA, USA). Peptides were separated using a linear gradient of standard LC buffers over 45 minutes. Automated selection of peaks was carried out and three peptides per MS survey scan were selected by data dependent acquisition for MS/MS using the conditions shown below (Table 3.2.6). Data were analysed using MASCOT (Matrix Science, London, UK; see 3.2.5.1).

Table 3.2.6 Sample preparation was assessed using an Esquire HCT Ion Trap.

Peptides were separated by High pH Reverse Phase and analysed by tandem mass spectrometry using automated selection of peaks to analyse three peptides per survey scan.

<table>
<thead>
<tr>
<th>Esquire HCT Ion Trap</th>
<th>MS</th>
<th>MS/MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass Acquisition Time</td>
<td>200ms</td>
<td>200ms</td>
</tr>
<tr>
<td>Maximum Trap Fill</td>
<td>200,000 ions</td>
<td>200,000 ions</td>
</tr>
<tr>
<td>Acquisition Mass Range</td>
<td>300-2000 m/z</td>
<td>200-2000 m/z</td>
</tr>
<tr>
<td>Number of Precursor Ions pre Survey Scan</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Charge State Selection</td>
<td>&gt;+1</td>
<td></td>
</tr>
</tbody>
</table>
3.2.4.2 Analysis of iTRAQ Labelled Samples

Depleted plasma samples were labelled with iTRAQ reagent and analysed using the LC-MS/MS platforms detailed below.

3.2.4.2.1 QStar XL ESI qTOF

Online Reverse Phase fractionation was performed using a PepMap™ 100 Column (C18, Dionex, CA, USA) run on a UltiMate pump (LC Packings, Amsterdam, The Netherlands), and coupled to a QStar XL qTOF (Applied Biosystems, Warrington, UK). Sample was loaded at 30ul/min in Buffer A (2% ACN, 0.1% Formic Acid (FA)) and eluted using a linear gradient (0-70% over 100 minutes) of Buffer B (80% ACN, 0.1% Formic Acid) at a flow rate of 0.25 µl/min. An independent data acquisition protocol was applied to acquire data; an MS scan was taken and the two highest abundance multiply charged (2 to 4) ions above a 10 count threshold with m/z between 480 and 2000 were selected for fragmentation and subsequently dynamically excluded for 1 minute (for conditions see Table 3.2.7). Data acquired was processed using ProteinPilot version 3.0 (see 3.2.5.2).

Table 3.2.7 Analysis of iTRAQ labelled samples using an Applied Biosystems QStar XL qTOF. Data were acquired by an independent data acquisition protocol. The two highest abundance multiply charged ions above a 10 count threshold with an m/z between 480 and 2000 were selected for fragmentation and subsequently dynamically excluded for 1 minute by the MS survey scan.

<table>
<thead>
<tr>
<th>QStar XL qTOF</th>
<th>MS</th>
<th>MS/MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accumulation Time (seconds)</td>
<td>0.999950</td>
<td>1.500051</td>
</tr>
<tr>
<td>Masses (Da)</td>
<td>400-1200</td>
<td>50-1600</td>
</tr>
<tr>
<td>Duration (min)</td>
<td>119.994</td>
<td>119.995</td>
</tr>
<tr>
<td>Cycles</td>
<td>1800</td>
<td>1800</td>
</tr>
<tr>
<td>Cycle Time</td>
<td>3.9998</td>
<td>3.9998</td>
</tr>
<tr>
<td>Switch Criteria</td>
<td>480-1200</td>
<td></td>
</tr>
<tr>
<td>Charge State</td>
<td>2-4</td>
<td></td>
</tr>
<tr>
<td>Mass Tolerance</td>
<td>0.1Da</td>
<td></td>
</tr>
</tbody>
</table>
3.2.4.2.2 5800 MALDI TOF-TOF

Analysis was performed using a 5800 MALDI TOF-TOF Mass Spectrometer (Applied Biosystems, Warrington, UK). Sample was resuspended in Buffer A (2% Acetonitrile, 0.1% TFA) and separated using an Acclaim PepMap 100 C18 column (3 µm, 1mm x150 mm; Dionex, CA, USA) with 5 µm Acclaim PepMap µ-Precolumn (300 µm x 5mm; Dionex, CA, USA) with a gradient of 80% ACN, 0.1% TFA at a flow rate of 0.8mL/min using a Ultimate 3000 HPLC system ((Table 3.2.8: Dionex, CA, USA). For each LC run, 330 spots were collected and mixed online with HCCA as matrix (2mg/mL, 70% MeCN, 0.1% TFA) by a Probot Micro Fraction Collector (LC Packings, Amsterdam, The Netherlands).

Each sample plate was manually calibrated using Applied Biosystems standard calibrant mix (Applied Biosystems, Warrington, UK), and laser power and tuning parameters optimised. MS scans of 2000 laser shots were taken at fixed laser intensity and the most abundant 27 precursors per fraction selected for MS/MS at the highest intensity. The acquisition method used excluded ions with signal to noise less than 20 and filtered out identical peaks detected in adjacent spots, selecting only the strongest precursors. In order to achieve the maximum signal intensity for low abundance peptides, the precursor ions with the weakest signal to noise ratio were acquired first. MS/MS spectra were acquired using a 2kv method with CID and a relative precursor mass window resolution of 300, 4000 shots were acquired at a fixed laser intensity. Data acquired was processed using ProteinPilot version 3.0 (see 3.2.5.2).

Table 3.2.8  A non-linear gradient was applied to an Acclaim PepMap 100 C18 column to separate peptides prior to spotting with HCCA matrix and analysis using an Applied Biosystems 5800 MALDI TOF-TOF. Dried pre-fractionated samples were resuspended in Buffer A (2% Acetonitrile, 0.1% TFA) applied to the column and eluted in 80% ACN, 0.1% TFA. 330 spots were collected for each applied fraction and mixed online 1:1 with HCCA matrix.

<table>
<thead>
<tr>
<th>Time (minutes)</th>
<th>% Buffer B</th>
<th>Flow (mL/min)</th>
<th>Maximum Pressure (bar)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>0.80</td>
<td>70</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>0.80</td>
<td>70</td>
</tr>
<tr>
<td>75</td>
<td>15</td>
<td>0.80</td>
<td>70</td>
</tr>
<tr>
<td>150</td>
<td>75</td>
<td>0.80</td>
<td>70</td>
</tr>
<tr>
<td>158</td>
<td>100</td>
<td>0.80</td>
<td>70</td>
</tr>
<tr>
<td>168</td>
<td>100</td>
<td>0.80</td>
<td>70</td>
</tr>
<tr>
<td>169</td>
<td>5</td>
<td>0.80</td>
<td>70</td>
</tr>
</tbody>
</table>
3.2.5 Data Analysis

Search algorithms used for analysis of LC-MS/MS data are described below.

3.2.5.1 MASCOT

For analysis of sample preparation using LC-MS/MS, database searches were performed using the internet based program MASCOT (see 1.3.2.2.1 for more information, Matrix Science, London, UK) Search parameters were as follows;

- Database: SwissProt
- Taxonomy: Homo Sapiens
- Enzyme: Trypsin
- Fixed modifications: Methythiolation of cysteine residues where MMTS used and carbamidomethyl if using IAM
- Variable Modification: oxidation of methionine
- MS Peptide Mass Tolerance: 30-50ppm
- MS/MS Mass Tolerance: 0.1-0.4 Da
- Max Missed Cleavages: 1

3.2.5.2 ProteinPilot (version 3.0)

Data were interrogated using ProteinPilot version 3.0 (Applied Biosystems, Warrington, UK) with trypsin as the digest agent and default settings for iTRAQ 8-plex labelling. Peptide identifications were made using the Paragon algorithm (Shilov et al. 2007) searching against the ipi.HUMAN database (v3.59.). A 95% confidence interval cut off was used for significant peptides, and single peptide identifications were included in grouped protein identifications. False discovery rate was calculated by searching all peptide data against a concatenated database containing both forward and reversed protein sequences. An average relative quantification ratio was calculated from the unique iTRAQ reporter ions associated with the peptides which identify each protein. The software applies a normalisation factor to each iTRAQ reporter ion ratio to compensate for any differences in the amount of starting protein material prior to labelling. To be included in the calculation of the ratio, peptides must be unique to that protein, identified with greater than 1% confidence and have a summed signal-to-noise ratio for peak pairs greater than 9. Relative ratios are calculated against a user defined denominator (A glossary of terms used by ProteinPilot are detailed in Appendix).
3.3 Validation of Candidate Markers

3.3.1 Multiple Reaction Monitoring

3.3.1.1 Preparation of Samples for Multiple Reaction Monitoring

Plasma samples were depleted of high abundance proteins using the MARS 14 immunodepletion system and prepared for analysis by MRM as described below.

3.3.1.1.1 Sample Spiking and Preparation

Plasma samples (see 3.1.2) were prepared for MARS 14 Immunodepletion. Plasma samples (30µl) were centrifuged at 4°C for 10 minutes at 10,000g and diluted 4 fold in buffer A spiked with yeast enolase (10 µg/mL of plasma). Samples were depleted as described in MARS 14 Immunodepletion (3.2.3.1.1). Depleted plasma was concentrated and desalted using an mRP column (see 0) and dried using a vacuum drier (Eppendorf, UK). Dried samples were resuspended in 50 µl TEAB:TFE (2,2,2-trifluoro-ethanol) and incubated at 90 °C for 5 minutes.

3.3.1.1.2 Digestion

Proteins were reduced by addition of 10µl 25mM DTT and incubated at 60°C for 30 minutes. Cysteine residues were blocked by addition of 10µl 100mM IAM and incubation at room temperature in the dark for an hour. The reaction was quenched by addition of 12 µl 62.5mM DTT and further incubation at room temperature for 30 minutes. Proteins were digested using trypsin (1:50 w/w enzyme to substrate) in 0.5M TEAB for 12-16hrs at 37°C.

3.3.1.1.3 Supelco Analytical Discovery C18 Column

Peptides were washed and desalted offline using a Sulpeolco Analytical Discovery C18 Column (300 x 2.1mm, 5 µm; Sigma Aldrich, Dorset, UK) with an Agilent 1200 Series HPLC system (Agilent, UK). Digested samples were diluted with 500 µl buffer B (Acetonitrile, 0.1% TFA) and half of the sample was applied to the column. The HPLC system was prepared by wet priming with appropriate buffers without and with the column. Sample was applied to the column (main-pass to bypass after 2.75 minutes) maintained at 35°C in buffer A (0.1% TFA) and washed and eluted in buffer B using the recommended program (Table 3.3.1). The bound protein fraction was collected between 7.5 and 19 minutes and maintained at 4°C by the automated sample collector.
Table 3.3.1  Desalting of samples using a Sulpelco Analytical Discovery C18 Column. Following digestion, samples were desalted and washed using a Sulpelco Analytical Discovery C18 Column. Manufacturer’s recommended program was used with Buffer A; 0.1% TFA and Buffer B; Acetonitrile, 0.1% TFA, eluant was collected between 7.5 and 10 minutes.

<table>
<thead>
<tr>
<th>Time (minutes)</th>
<th>% Buffer B</th>
<th>Flow Rate (mL/min)</th>
<th>Maximum Pressure (bar)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>0.75</td>
<td>150</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.75</td>
<td>150</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>0.40</td>
<td>150</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>0.40</td>
<td>150</td>
</tr>
<tr>
<td>8</td>
<td>50</td>
<td>0.40</td>
<td>150</td>
</tr>
<tr>
<td>8.5</td>
<td>100</td>
<td>0.80</td>
<td>150</td>
</tr>
<tr>
<td>10.5</td>
<td>100</td>
<td>0.80</td>
<td>150</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>0.75</td>
<td>150</td>
</tr>
</tbody>
</table>

3.3.1.2 Optimisation of Sample Preparation

The sample processing method was subsequently optimised using the methods detailed below.

3.3.1.2.1 Molecular Weight Spin Columns

Following enolase spiking and immunodepletion using the MARS 14 system, sample was concentrated and exchanged into suitable buffer for digestion using Amicon® Ultra 0.5mL Centrifugal Filters (MWCO-10kDA, Millipore, MA, USA). Samples were applied to the MWCO and spun at 15,000g at 4 °C for 15 minutes and washed a minimum of 4 times with 0.5M TEAB. To remove the concentrated sample, tubes were inverted in a new collection tube and spun at 1,000g at 4 °C for 2 minutes. Samples were adjusted to an equal volume by addition of further 0.5M TEAB. In preparation for digestion TFE was added to a final volume of 5%.
3.3.1.2.2 Sepharose Columns

Following digestion, samples were desalted using SamplIQ Polymer SCX 1mL individual columns (Agilent, UK). Columns were prepared by passing through 1mL of MeOH 0.1% FA and 2mLs of 2% FA. Samples were made to 1mL with 2% FA and applied to the column. Sample was washed with 1mL of 2% FA, 3 mLs of 0.1% FA and 3mLs of MeOH and eluted in a total of 500µl of 3% ammonium hydroxide in MeOH.

3.3.1.2.3 Supelco Analytical Discovery C18 Column for Quantification

To quantify peptide content, samples were applied to a Sulpelco Analytical Discovery C18 Column (300 x 2.1mm, 5 µm; Sigma Aldrich, Dorset, UK) using an Agilent 1200 Series HPLC system (Agilent, UK). The HPLC system was prepared by wet priming with appropriate buffers without and with the column. Sample and appropriate standards were applied to the column (main-pass to bypass after 2.75 minutes) maintained at 35ºC in Buffer A (0.1% TFA) and washed and eluted in Buffer B (Acetonitrile, 0.1% TFA) using the recommended program (Table 3.3.2). Protein content was compared using the area under the elution peak measured at 214nm.

Table 3.3.2 Quantification of peptide content using a Sulpelco Analytical Discovery C18 Column. Sample protein content was quantified using a Sulpelco Analytical Discovery C18 Column. Manufacturer’s recommended program was used with Buffer A; 0.1% TFA and Buffer B; Acetonitrile, 0.1% TFA, protein content was compared using the area under the elution peak measured at 214nm.

<table>
<thead>
<tr>
<th>Time (minutes)</th>
<th>% Buffer B</th>
<th>Flow Rate (mL/min)</th>
<th>Maximum Pressure (bar)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>0.750</td>
<td>150</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>0.750</td>
<td>150</td>
</tr>
<tr>
<td>4.05</td>
<td>50</td>
<td>0.750</td>
<td>150</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>0.750</td>
<td>150</td>
</tr>
<tr>
<td>5.5</td>
<td>100</td>
<td>0.750</td>
<td>150</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>0.750</td>
<td>150</td>
</tr>
<tr>
<td>7.3</td>
<td>5</td>
<td>0.750</td>
<td>150</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>0.750</td>
<td>150</td>
</tr>
</tbody>
</table>
3.3.1.2.4 6530 qTOF for Sample Assessment

The 6530 qTOF was used for qualitative analysis of samples. Prepared samples were dried and re-suspended in 4% ACN, 0.1% FA with sonication. Online reverse phase separation was performed using a Poroshell 120 Column (250x2.1mm; Agilent, UK) run on a 1290 ultra HPLC pump (Agilent, UK). Sample was loaded in buffer A (0.1% FA) and separated over 45 minutes using a linear gradient of 5-35% buffer B (methanol, 0.1% FA). A flow rate of 200µl/min was used with the column maintained at 35°C. A standard data dependent acquisition method was used for analysis of sample using the 6530 qTOF (Agilent, UK). Data acquired was searched using MASCOT (see 3.2.5.1).

3.3.1.3 Multiple Reaction Monitoring Assay

Prepared samples were dried and resuspended in 100µl of 4% ACN, 0.1% FA with sonication. Online reverse phase separation was performed using a Poroshell 120 Column (250x2.1mm; Agilent, UK) run on a 1290 uHPLC pump (Agilent, UK). Sample was loaded in buffer A (0.1% formic acid) and separated over 45 minutes using a linear gradient of 5-35% buffer B (methanol, 0.1% formic acid). A flow rate of 200µl/min was used with the column maintained at 35°C.

The 6460 QQQ uses a JetStream electrospray source; the following parameters were applied- gas temperature 300°C, gas flow 130 l/min, nebuliser 40 psi, sheath gas temperature 300°C, sheath gas flow 10 l/min. Peptide retention times and optimised collision energies were supplied to MassHunter (B03.01; Agilent, UK) to establish a dynamic MRM scheduling method. The method was based on input parameters of 0.12 min delta retention times and 800 ms cycle times. The minimum and maximum dwell times established by the method were 100 and 350 ms respectively.

3.3.1.4 Data Analysis

Data were analysed using Skyline (MacLean et al. 2010). The total area under the peak for all transitions was summed to give a total peak area for each peptide. Data were exported to Microsoft® Excel for analysis. Statistical analysis was carried out using Graphpad Prism (version 4.01; San Diego, California).
3.3.2 Validation Using Antibodies

3.3.2.1 SDS-PAGE and Western Blotting

3.3.2.1.1 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

Plasma samples were reduced and denatured by dilution 1:10 in dH₂O, addition of an equal volume of Lamelli buffer (0.125M Tris (pH 6.8), 5% SDS, 50% Glycerol, 25% β-mercaptoethanol, 0.05% w/v bromophenol blue) and incubation at 90 °C for 20 minutes. Samples were loaded onto a 1.5mm stacking gel and appropriate percentage resolving gel with Amersham Full-Range Rainbow Molecular Weight Markers (12-225 kDa, GE Healthcare, UK). Electrophoresis was carried out in Running buffer (25mM Tris-base, 191mM Glycine, 3.4mM SDS, pH8.0) using a Mini Protean III Electrophoresis System (BioRad, Hemel Hempstead, UK).

3.3.2.1.2 Coomassie Staining

To visualise separated proteins, gels were stained in Coomassie stain (brilliant blue R-250 10% acetic acid 45% methanol) and rinsed in de-stain (10% acetic acid 45% methanol) with each step carried out with agitation.

3.3.2.1.3 Western Blot

Following electrophoresis, the gel containing separated proteins, Amersham Hybond ECL Nitrocellulose Membrane (GE Healthcare, UK), blotting papers and fibre pads were equilibrated for 5 minutes in Transfer buffer (25mM tris-base, 190mM Glycine, 0.35mM SDS, in 20 % Methanol). Proteins were transferred to the membrane at 100V for 1-2 hours using a Mini Protean III Protein Blotting System (BioRad, Hemel Hempstead, UK). Nitrocellulose membranes were blocked in 5% milk in tris buffered saline-Tween 20 (TBS-T; 2.42g tris-base, 8g sodium chloride, 3.8mL 1M HCL, (pH7.6) 0.5mL Tween 20) to prevent non specific reactions. The membrane was incubated in an appropriate primary antibody (abcam® Cambridge, UK) in 0.5% milk in TBS-T at 4 ºC overnight. The membrane was then washed 6 times for 10 minutes with TBS-T and secondary antibody conjugated with horseradish peroxidase (polyclonal goat anti-mouse IgG and polyclonal goat anti-rabbit IgG, both Dako, Denmark) diluted 1/5000 in 0.5% milk or BSA in TBS-T applied for 1 hour. The membrane was washed again in TBS-T; all incubations were carried out with agitation. Membranes were developed using the Pierce ECL Western Blotting Substrate (Pierce, Rockford, IL, USA), visualised using a Molecular Imager Chemidoc XRS+ with Image Lab™ 2.0 Software (BioRad, Hemel Hempstead, UK). Proteins bands were compared to molecular weight markers for sizing. After imaging membranes were incubated in Ponceau S Stain (0.1% (w/v) Ponceau S in 5% (v/v) acetic acid) for 5 minutes with agitation and re-imaged to assess loading.
3.3.2.2 ELISA

Measurement of candidate proteins was carried out using commercially available ELISA kits following the manufacturer's protocol. Absorbance was measured using a SpectraMax M5 Plate Reader (Molecular Devices, CA, USA). Data were exported to Graphpad Prism version 4.01 (San Diego, California) for statistical analysis.
4 Development and Optimisation of an iTRAQ Workflow

4.1 High Abundance Protein Depletion: Comparison of Immunodepletion Methods for Identification of Changed Proteins using iTRAQ

4.1.1 Introduction

The dynamic range of human plasma spans 10 orders of magnitude with just 22 high abundance proteins accounting for 99% of the protein content; no one detection method can accurately quantify proteins across this range (Anderson and Anderson 2002). Removal of these high abundance proteins is therefore an essential step in analysis of the plasma proteome. Removal of high abundance proteins allows analysis of more proteins present at lower concentrations. These proteins are most likely to be reflective of health status as they enter the plasma as a result of leakage from different tissues or to serve as signalling molecules (Anderson and Anderson 2002).

Multiple methods utilising the properties of recombinant proteins, peptides and chemical ligands are available for the separation and enrichment of plasma. The ideal depletion system needs to have high specificity and accuracy, low cross reactivity for non targeted proteins, good recovery of bound proteins, high capacity, simple operation and be easily automated (Fang and Zhang 2008). Immunodepletion, the immunoaffinity subtraction of high abundance proteins using polyclonal antibodies, is commonly used in proteomic studies due to the high level of specificity afforded by the antibodies and high throughput (Echan et al. 2005; Zolotarjova et al. 2005; Whiteaker et al. 2007). This study aimed to investigate the best available immunodepletion systems. Initial investigation used the Seppro® IgY 14-SuperMix Liquid Chromatography Column System, which removes 14 high abundance proteins and approximately 75 undefined moderately abundant proteins. The Agilent Multiple Affinity Removal LC Column (Human 14) and associated workflow subsequently became available and depletion by both systems was compared.

4.1.2 Aim

- To compare the immunodepletion of human plasma by the Genway Seppro® IgY 14-SuperMix Liquid Chromatography Column System and the Agilent Multiple Affinity Removal LC Column (Human 14) and the applicability of each system to a hypothesis generating investigation using iTRAQ.
4.1.3 Method

4.1.3.1 The Genway Seppro® IgY-SuperMix Immunodepletion System

The Seppro® IgY 14-SuperMix Liquid Chromatography Column System (hereafter referred to as the IgY 14-SuperMix System) consists of the IgY 14 column which removes fourteen of the most abundant plasma proteins (albumin, IgG, alpha-1-antitrypsin, IgA, IgM, transferrin, haptoglobin, alpha-1-acid glycoprotein, alpha-2 macroglobulin, HDL (apo A-I & A-II), fibrinogen, complement C3 and LDL (mainly apo B)) coupled to the SuperMix column which removes approximately 75 undefined moderately abundant plasma proteins. Both columns utilise avian IgY antibodies, with the antibodies for the SuperMix column collected from the yolk of eggs produced by chickens inoculated with human plasma depleted by the IgY 14 column (Huang et al. 2005). The manufacturer’s instructions were followed to deplete plasma (see 3.2.2).

To optimise the IgY 14-SuperMix System, varying volumes of plasma from women who were and were not pregnant (see 3.1.3) were applied to the columns. Samples were buffer exchanged by ultrafiltration and assessed by SDS-PAGE (10% acrylamide) with Coomassie Staining (see 3.3.2.1.1.). To determine the identity of protein in the bands of interest in the SDS-PAGE gels, an in-gel digest and Zip Tip peptide cleanup was carried out; proteins were identified by MS analysis and MASCOT search (see 3.2.4.1 and 3.2.5.1).

4.1.3.2 The Agilent Human 14 Multiple Affinity Removal LC Column

Plasma was depleted of high abundance plasma proteins using the Multiple Affinity Removal LC Column-Human 14 (hereafter referred to as the MARS 14 Column). The MARS 14 Column removes up to 94% of total protein mass by removing the top 14 abundant plasma proteins (albumin, alpha-1-acid glycoprotein, alpha-2-macroglobulin, alpha-1-antitrypsin, apo AI, apo AII, complement C3, fibrinogen, haptoglobin, IgA, IgG, IgM, transferrin and transthyretin). The manufacturer’s instructions and buffers were used to deplete plasma (see 3.2.3).

To optimise the MARS 14 Column, varying volumes of plasma from women who were and were not pregnant (see 3.1.3) were applied to the column and the peak area of the flow-through fraction, as reported by ChemStation, assessed. Data were exported to Microsoft® Excel for further analysis; the peak area for the flow-through fraction was compared between injection volumes and between plasma from non-pregnant and pregnant women. Statistical tests were carried out using Graphpad Prism version 4.01 (San Diego, California). Following depletion, samples were concentrated and desalted using mRP-C18 column dried and resuspended. Qualitative protein assessment of immunodepleted plasma was carried out using 10µl of resuspended sample. Sample was applied to the 2100 Bioanalyzer microfluidics-based lab-on-a-chip platform and the Protein 230 Assay performed.
following the manufacturer's protocol. Quantitative assessment of protein concentration was carried out by Bio-Rad Protein Assay (BioRad, Hemel Hempstead, UK). For protein identification, resuspended samples were digested using trypsin, analysed by LC-MS/MS and proteins identified by database search using MASCOT (see 3.2.4.1 and 3.2.5.1).

4.1.3.3 Application to SCOPE Samples

Plasma samples taken at 15 weeks gestation were obtained from the SCOPE Study from women who developed late and early onset pre-eclampsia (each n=12) later in pregnancy and two distinct sets of healthy controls (each n=12) (see 3.1.2). Plasma samples were pooled according to phenotype and an additional reference superpool consisting of equal amounts of each of the case and control plasma samples produced.

A proportion of the pooled plasma was first immunodepleted using the IgY 14-SuperMix System and subsequently, when available, using the MARS 14 System. Superpool and case samples were immunodepleted in duplicate and control samples once, resulting in a total of 8 depleted plasma samples (Figure 4.1.1). Depleted plasma was processed appropriately, digested and iTRAQ labelled (see 3.2). Immunodepletion carried out by the IgY 14-SuperMix System was assessed prior to digestion by SDS-PAGE and post digestion by LC-MS analysis with proteins identified using a MASCOT search. Protein concentration was calculated by Coomassie Protein Assay (ThermoScientific, USA) and the volume of sample to be labelled adjusted appropriately. Immunodepletion by the MARS 14 Column was assessed qualitatively using the Protein 230 Assay and the peak area for the flow-through fraction.

4.1.3.4 Comparison of the IgY 14-SuperMix System and the MARS 14 Column

Following depletion and labelling using iTRAQ reagent, plasma samples acquired from the SCOPE study were analysed by LC-MS/MS using a 5800 MALDI–TOF-TOF. Data from each LC-MS/MS analysis was interrogated using ProteinPilot 3.0 (see 3.2.5.2) searching against the Human protein database from the IPI (version 3.59). The relative quantification ratios for iTRAQ reporter ions were calculated by ProteinPilot against the reference superpool labelled with iTRAQ reporter ion 113 and data were exported to Microsoft® Excel. Data were filtered to produce a list of proteins considered identified with high confidence; the FDR (see 3.2.5.2) for the data set must be less than 5% and the UPS greater than 1.3. The proteins identified were assessed and filtered where necessary to identify the total number proteins identified with high confidence (see 4.3.4.2.2). Comparative analysis was carried out using this high confidence dataset for both depletion methods.
**Dynamic Range** – The total number of proteins identified from plasma processed by both immunodepletion methods and the unique proteins identified following depletion by either the IgY 14-SuperMix System or MARS 14 column were counted. The depth of the plasma proteome sampled was assessed using the normal plasma concentration for proteins identified. Where available in the literature, normal reference ranges for plasma proteins identified following depletion by each method were noted and plotted. (Sources for plasma protein concentration reference ranges: Macko et al. 2002; Tsoutsou et al. 2004; Gong et al. 2006; Polanski and Anderson 2007; Ernoult et al. 2008; Bandow 2010). The number of proteins identified from sample processed by each immunodepletion method falling within the concentration ranges of the classical plasma proteins (1x10^{-10}-1x10^{6} pg/mL), tissue leakage products (1x10^{6}-1x10^{3} pg/mL), and Interleukins (1x10^{3}-1 pg/mL) as defined by Anderson and Anderson (Anderson and Anderson 2002) were also calculated.

**Efficiency of High Abundance Protein Removal** – To assess the efficiency of immunodepletion by both the IgY 14-SuperMix System and MARS 14 Column, proteins identified from samples processed using each method were searched for proteins which should have been depleted.

**Reproducibility** - Reproducibility of sample processing was assessed using the iTRAQ relative quantification ratios for plasma samples depleted twice and forming technical replicates (Figure 4.1.1). The ratio of iTRAQ relative quantification ratios for each replicate pair of processed plasma samples was calculated and transformed (Log_{2} (replicate iTRAQ ratio/replicate iTRAQ ratio)). The range of transformed comparative iTRAQ relative quantification ratios and percentage frequency between 0.05 intervals were calculated and plotted.
Figure 4.1.1  Pooled plasma samples were immunodepleted using the Genway Seppro® IgY 14-SuperMix Liquid Chromatography Column System or the Agilent Multiple Affinity Removal Liquid Chromatography Column (Human 14), labelled with iTRAQ reagent and analysed by LC-MS/MS using a 5800 MALDI–TOF-TOF. Plasma samples were obtained from women who developed late and early onset pre-eclampsia (each n=12) and two distinct sets of healthy controls (each n=12). Plasma samples were pooled according to phenotype and an additional reference superpool consisting of equal amounts of each of case and control plasma produced. Plasma was immunodepleted using manufacturer’s instructions with either the Genway Seppro® IgY 14-SuperMix Liquid Chromatography Column System or the Agilent Multiple Affinity Removal LC Column (Human 14). Following tryptic digestion, pooled samples were labelled with iTRAQ 8-plex reagent and separated using high pH reverse phase chromatography. LC-MS/MS analysis was carried out using a 5800 MALDI–TOF-TOF, data were interrogated using ProteinPilot 3.0 searching against the Human protein database from the IPI (version 3.59) and exported to Microsoft® Excel for comparative analysis.
4.1.4 Results

4.1.4.1 Optimisation of the IgY 14-SuperMix System

The performance of the IgY 14-SuperMix System was investigated. The manufacturer's instructions recommend application of up to 200 µl of plasma to the columns. In order to assess the reproducibility of removal of high abundance proteins, replicate injections of varying volumes of plasma from a pregnant woman were made to the column. The depleted plasma was processed accordingly and assessed by SDS-PAGE with Coomassie Staining. An initial injection volume of 200µl was tested; the SDS-PAGE gel indicated differential depletion between replicates as shown by the discrepancy in protein band patterns (Figure 4.1.2; A-lanes 1-3, B-lanes 1-3, C-lanes 1 and 2). The band at approximately 66kDa also indicated incomplete depletion of albumin which was confirmed by LC-MS/MS analysis.

Application of 100µl of plasma to the IgY 14-SuperMix System (Figure 4.1.2; C-lane 3) showed some improvement in depletion compared to application of 200µl (Figure 4.1.2; C-lanes 1 and 2) as indicated by lower density staining on the SDS-PAGE gel. At both injection volumes the presence of plasma proteins which should be removed by depletion were confirmed by LC-MS/MS, including alpha-2- macroglobulin, alpha-1- antitrypsin and fibrinogen. An injection volume of 100 µl was applied to both the IgY 14-SuperMix System (Figure 4.1.2 C-lane 3) and the IgY 14 Column alone (Figure 4.1.2; C-lane 4). Use of the SuperMix Column in addition to the IgY 14 Column shows an improvement in depletion as indicated by the decreased amount of protein on the SDS-PAGE gel. From injections of plasma from a pregnant woman onto the IgY 14-SuperMix System it was concluded that the injection of less than the recommended 200 µl of plasma to the column was preferable but also that depletion was variable across all 9 injections.

Changes to the plasma proteome occur during pregnancy, a plasma sample from a woman who was not pregnant was also applied the IgY 14-SuperMix System for comparison. It was noted that depleted plasma from a non-pregnant woman, when compared to an equal volume of depleted plasma from a pregnant woman, contained far fewer proteins (Figure 4.1.2; B-lane 4, C- lane 5).
Figure 4.1.2  Assessment of replicate plasma samples depleted using the IgY 14-SuperMix System by SDS-PAGE with Coomassie Staining showed depletion to be variable between replicates. Plasma from women who were and were not pregnant was applied to the IgY 14-SuperMix System in the volumes indicated. Depleted plasma was concentrated by ultrafiltration, separated by SDS-PAGE and visualised using Coomassie stain. A- An injection volume of 200 µl of plasma from a pregnant woman showed differential depletion between replicates, the band at approximately 66kDa was confirmed to be albumin by LC-MS/MS analysis. B- A second replicate analysis of 200 µl of plasma from a pregnant woman showed variance in depletion as indicated by the differing protein band pattern. Depleted plasma from a woman who was not pregnant contained fewer proteins (lane 4). C- An injection volume of 100 µl of pregnant plasma (lane 3) showed some improvement in depletion compared to 200 µl (lanes 1 and 2). The same volume (100 µl) of plasma depleted using the IgY 14 column alone (lane 4) contains far more protein than plasma depleted by the IgY14-SuperMix System (lane 3). Injection of the same volume of plasma from a woman who was not pregnant (lane 5) to IgY 14-SuperMix System indicated the depleted non-pregnant plasma contained fewer proteins than the same volume of depleted pregnant plasma (lane 3).
4.1.4.2 Application of the IgY 14-SuperMix System to SCOPE Samples

Superpool and case samples were depleted twice to assess technical variability within the workflow, while control samples were depleted once (Figure 4.1.1). A total of 300 µl of pooled plasma was depleted in two injections of 150 µl to provide enough protein for iTRAQ labelling. Depleted plasma from both 150µl injections were combined into one during buffer exchange. Protein concentration was quantified by protein assay and the immunodepletion assessed by SDS-PAGE gel (Figure 4.1.3). An appropriate volume of protein was digested and labelled using an 8-plex iTRAQ reagent kit. Immunodepletion was assessed by LC-MS and indicated that the bands seen at approximately 65kDa in superpool 1 and late onset pre-eclampsia case 1 samples in the SDS-PAGE gel (Figure 4.1.3) were albumin. Alpha-1-antitrypsin and alpha-2-macroglobulin were also shown to be present, signifying incomplete immunodepletion of both samples.

Figure 4.1.3  Immunodepletion of pooled plasma samples acquired from the SCOPE study using the IgY 14-SuperMix System was variable. Pooled plasma samples obtained via the SCOPE Study from women who developed late and early onset pre-eclampsia (n=12), two distinct sets of healthy controls (each n=12) and a reference superpool consisting of an equal proportion of each pool were immunodepleted using the IgY 14-SuperMix System. For each sample two injections of 150µl were made to the column and combined during subsequent buffer exchange. Samples were assessed by SDS-PAGE with Coomassie staining and revealed differential depletion in both superpool 1 and late onset pre-eclampsia 1 samples; the presence of proteins which should have been removed during depletion was confirmed by LC-MS/MS analysis.
4.1.4.3 Optimisation of the MARS 14 Column

4.1.4.3.1 Investigation of Injection Volume and Flow-through Peak Area

The MARS 14 Column and associated workflow subsequently became available; the performance of the MARS 14 column and workflow was investigated and optimised. The manufacturer’s instructions recommend application of up to 40 µl of plasma to the column. Injections of varying volumes of plasma on to the column were made to assess immunodepletion. Running of the MARS 14 column was automated using an Agilent 1200 Series HPLC system. The HPLC system includes a UV detector which measures at 280nm, the resulting chromatogram was used to assess injections of plasma onto the column. The example chromatogram provided by the manufacturer shows two distinct peaks; the first wider and shallower peak represents the flow-through fraction which is collected, the second taller and narrower peak represents the bound high abundance proteins which are eluted later in the gradient (Figure 4.1.4-A). Application of plasma to the column produced a similar chromatogram and the area under the peak for both flow-through and a bound fraction was calculated by the Agilent ChemStation software (Figure 4.1.4-B).

Assessment of the flow-through peak area showed the peak area increased with the volume of plasma applied to the column (Figure 4.1.4-B). A total of 51 injections of 20 to 45 µl of plasma from women who were and were not pregnant were made to the column. Plotting peak area against injection volume showed a linear relationship (Figure 4.1.4-C). Reproducibility of the peak area was also assessed within and between runs. The CV for the 8 replicate injections of 30 µl of plasma plotted in Figure 4.1.4-C was 12%, while the injections at 25 µl (n=34) were made over several runs and had a CV of 17%, indicating reproducibility of peak area both within and between sets of depletions.
Figure 4.1.4  Depletion of plasma using the MARS 14 System produces a characteristic peak pattern and the area under the flow-through peak is proportional to the volume of plasma applied to the column. Plasma was immunodepleted using a MARS 14 column run using an Agilent 1200 Series HPLC system (Agilent, UK). Chromatograms of UV absorbance at 280nm were recorded by ChemStation software (Agilent, UK) and the area under each peak calculated. A- The manufacturer’s example chromatogram shows two distinct peaks labelled flow-through and bound fraction. B- Application of plasma to the column produced a similar chromatogram with flow-through and bound fraction peaks, and showed the peak area increased with injection of increasing volume (20-45µl) of plasma. C- Area under the flow-through peak was found to be proportional to the volume of plasma injected onto the column as indicated by the linear relationship on plotting injection volume against peak area.
4.1.4.3.2 Application of Plasma from Non Pregnant and Pregnant Women

As previously shown during investigation of the IgY 14-SuperMix System, depleted plasma from women who are pregnant is different to women who are not pregnant. Application of plasma from pregnant women to the MARS 14 column was therefore investigated. Injections of varying volumes (25-40 µl) of plasma from a woman who was not pregnant and a woman who was pregnant were applied to the MARS 14 column and the peak area for the flow-through fraction assessed. The UV trace for injections of plasma from the non pregnant and pregnant women at each injection volume were assessed and indicated that the peak area for plasma from pregnant women was larger at all but the 25 µl injection volume, this was confirmed by plotting injection volume and peak area for each paired set of injections (Figure 4.1.5). This effect was noted throughout the investigation, however only data from a direct comparison of injection volume using pregnant and non pregnant plasma samples is shown.

![Graph](chart.png)

Figure 4.1.5  Peak area for the flow-through fraction is smaller when depleting plasma from a non pregnant woman than from a pregnant woman using the MARS 14 System. When using the MARS 14 system to deplete plasma, peak areas calculated from chromatograms of replicate injections of increasing volume of plasma from a woman who was not pregnant were smaller than peak areas for plasma from a woman who was pregnant. On plotting peak area against injection volume, peak area was smaller for non pregnant plasma at all injection volumes but 25 µl and for plasma from both women peak area increased with injection volume.
4.1.4.4 Optimisation of Assessment of the MARS 14 Column-

Assessment of the fidelity of immunodepletion is recommended to determine the quality of depleted plasma samples. Immunodepletion using the IgY 14-SuperMix System was assessed using SDS-PAGE with Coomassie staining. Alternative methods for assessment of immunodepletion using the MARS 14 System were investigated.

4.1.4.4.1 Assessment of Depletion using the Flow-through Peak Area

As shown previously, the peak area for the flow-through fraction is both reproducible and proportionate to the volume of sample applied when plasma is immunodepleted using the MARS 14 Column. The peak area for the flow-through fraction was therefore also investigated as a measure of quality of depletion. The peak area for replicate samples incompletely immunodepleted by a compromised MARS 14 column was compared to samples thought to be completely immunodepleted using an alternative MARS 14 column (Figure 4.1.6-A).

The median peak area for incomplete immunodepletions was 204,300 mAUs compared to 129,700 mAUs for complete immunodepletions and was found to be significantly different using Mann Whitney U (p<0.05). Peak areas were also plotted, and mean±2SD for the peak areas of complete immunodepletions calculated; peak areas for incomplete immunodepletions fell outside mean±2SD for complete injections (Figure 4.1.6-B). Although demonstrated on a limited sample set, it is possible to differentiate complete from incomplete depletions using peak area. The peak area for the flow-through fraction of sample immunodepleted using the MARS 14 column is therefore a useful indication of the quality of the depletion.
Figure 4.1.6  Plasma incompletely depleted using a compromised MARS 14 Column could be differentiated from plasma completely depleted using a working MARS 14 Column using the peak area of the flow-through fraction. Plasma was immunodepleted using a MARS 14 System known to be faulty and a working system and the peak area for the flow-through fraction assessed. A- Chromatograms of UV absorbance at 280nm recorded by ChemStation software appeared larger for incomplete immunodepletions than for complete depletions. B- Peak areas for incomplete and complete depletions were plotted and mean±2SD calculated (dashed line). Peak areas for incomplete depletions fell outside mean±2SD (red); peak areas for complete depletions (blue) fell within mean±2SD.
4.1.4.4.2 Qualitative Assessment of Immunodepleted Proteins

Completeness of immunodepletion was assessed using the Protein 230 Assay performed on a 2100 Bioanalyser microfluidics-based lab-on-a-chip platform. The lab-on-a-chip Protein 230 Assay is designed for sizing and quantitation of protein samples between 14 and 230kDa; samples are processed, applied to a microfluidics chip and the absorbance over time as the sample migrates across the chip measured. The migration and absorbance are presented as size (kDa) and absorbance (FU) in an electropherogram, the information is also represented as an ‘SDS-PAGE gel’ by the 2100 Expert Software (Agilent, UK).

Immunodepleted plasma was processed and assessed using the Protein 230 Assay. The quality of each assay was assessed using the electropherogram for the size ladder. Reproducibility of immunodepletion was assessed by comparison of samples in gel mode and electropherogram view; gel bands between replicate injections were reproducible (Figure 4.1.7; A- lanes 2-7) with electropherogram traces similar but less reproducible (Figure 4.1.7; B- traces 2-7). Both gel bands and electropherogram traces correlated weakly to the volume of plasma applied. It was noted that application of whole plasma and bound fraction containing high abundance proteins (HAP) produced a characteristic peak at approximately 60kDa in gel and electropherogram mode (Figure 4.1.7; A-lane 1, 8 and 9, B-traces 1, 8 and 9). Application of human or bovine serum albumin to the assay also produced a peak at approximately this size in gel and electropherogram mode (Figure 4.1.7; A- lane 10, B- trace 10) suggesting this peak in whole plasma and the bound fraction is albumin. Completeness of immunodepletion was assessed by the presence or absence of the characteristic signal at 60 kDa in gel and electropherogram mode.

To test this method, replicate samples known to be incompletely immunodepleted by a compromised MARS 14 column were compared to samples thought to be completely immunodepleted using an alternative MARS 14 column. The band was visible in both gel mode (Figure 4.1.8; A-lanes 1-4 ‘Bad’ incomplete immunodepletion, lanes 5-8 complete immunodepletion) and electropherogram mode in incompletely depleted plasma (Figure 4.1.8; B- traces 1-4 ‘Bad’ incomplete immunodepletion, traces 5-8 complete immunodepletion). The peak at approximately 60kDa was found to differentiate sample from complete and incomplete immunodepletion and was therefore used to assess the completeness of depletion in subsequently processed samples.
Assessment of depleted plasma by the Protein 230 Assay is reproducible in both gel and electropherogram mode and application of whole plasma, bound fraction containing high abundance proteins and albumin produces a characteristic band and peak at approximately 60kDa. Plasma immunodepleted using the MARS 14 System was assessed using the Protein 230 Assay performed on a 2100 Bioanalyser microfluidics-based lab-on-a-chip platform (Agilent, UK). A- When viewed in gel mode gel band patterns were reproducible and weakly correlated to the volume of plasma immunodepleted. A characteristic band was present in sample containing albumin including whole plasma, bound fraction containing high abundance proteins (MARS 14 HAP) and human serum albumin (lanes 1 and 8-10 respectively). B- On viewing the results of the assay in electropherogram mode, traces were similar but less reproducible and correlated weakly to injection volume. The characteristic peak thought to indicate the presence of albumin could be seen at approximately 60 kDa (trace 1, 8 and 10).
Figure 4.1.8  Plasma incompletely depleted using a compromised MARS 14 Column could be differentiated from plasma completely depleted using an alternative MARS 14 Column using the peak at 60kDA following Protein 230 Assay. Plasma was immunodepleted using a MARS 14 System known to be faulty and a working system and was assessed using the Protein 230 Assay performed on a 2100 Bioanalyser microfluidics-based lab-on-a-chip platform (Agilent, UK). A- Viewed in gel mode, the characteristic peak at 60kDa (27 seconds) thought to be albumin (both lane 10) could be seen in the samples labelled 'MARS 14 Bad' which were incompletely depleted but not in the samples labelled 'MARS 14 25 µl' which were completely depleted. B- In electropherogram mode the characteristic peak could be seen in the incompletely depleted samples labelled ‘MARS 14 25 µl BAD’ and bovine serum albumin sample (trace 10) but not in completely depleted samples labelled ‘MARS 14 25µl’.
4.1.4.4.3 Optimisation of Injection Volume for iTRAQ labelling

Following immunodepletion using the IgY 14-SuperMix System, the protein concentration of depleted plasma was calculated and the volume of the sample adjusted so 80 µg of protein were iTRAQ labelled. As immunodepletion by the MARS 14 System was found to be highly reproducible, a set volume of each sample to be iTRAQ labelled was immunodepleted and processed with no adjustment for amount of protein present required.

In order to determine the ideal injection volume for samples to be labelled with iTRAQ reagent, the number of proteins which could be identified and amount of protein in resuspended sample following injections of varying volumes were investigated. Plasma from a pregnant woman was injected onto the MARS 14 column (20-45 µl, see Figure 4.1.7 for Quality Control), processed and assessed by Bio-Rad Protein Assay and LC-MS/MS. The average number of proteins identified by a MASCOT search of three replicate LC-MS analyses, number of proteins identified of the 14 which should be removed, and actual and theoretical amount of protein following depletion were compared for each injection volume (Table 4.1.1). Theoretical protein return following immunodepletion was calculated assuming the concentration of proteins in plasma was 65mg/mL (Jiang et al. 2004), that the MARS 14 System removed 94% of plasma proteins and that no protein is lost in subsequent sample processing.

An injection volume of 40 µl was found to return the largest number of protein identifications; however a small increase in injection volume to 45 µl resulted in the identification of three proteins which should have been depleted. Each injection is to be processed and labelled with iTRAQ reagent; the labelling capacity for one iTRAQ label is 100µg of protein and so an injection of greater than 30 µl would exceed the capacity of the iTRAQ label. The theoretical amount of protein present in depleted plasma was similar to the actual amount calculated by protein assay (with the exception of the 35 µl injection) indicating that the MARS 14 System depletes the high abundance proteins efficiently and there is limited loss of protein in downstream sample processing. An injection volume of 25 µl was therefore found to balance the number of protein identifications with the capacity of both the column, indicated by the absence of proteins which should be removed and the iTRAQ label, and was therefore used for all plasma to be iTRAQ labelled.
Table 4.1.1  Immunodepletion of 25 µl of plasma from a pregnant woman by the MARS 14 System is ideal for iTRAQ labelling. Plasma from a pregnant woman was injected onto the MARS 14 column (20-45µl), processed and assessed by Bio-Rad Protein Assay and LC-MS/MS. The average number of proteins identified by a MASCOT search of three replicate LC-MS analyses, number of proteins identified of the 14 which should be removed, and actual and theoretical amount of protein following depletion was compared for each injection volume. Theoretical protein return following immunodepletion was calculated assuming the concentration of proteins in plasma was 65mg/mL (Jiang et al. 2004) and the MARS 14 System removed 94% of plasma proteins and no protein is lost in subsequent sample processing. An injection volume of 25 µl was found to balance the number of protein identifications with the capacity of both the column, indicated by the absence of proteins which should be removed and the iTRAQ label and was therefore used for all plasma to be iTRAQ labelled.

<table>
<thead>
<tr>
<th>Volume of plasma injected onto MARS 14 Column (µl)</th>
<th>Average number of proteins identified</th>
<th>Number of proteins identified by which should be depleted</th>
<th>Theoretical amount of protein in depleted sample (µg)</th>
<th>Actual amount of protein in depleted sample (µg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>20</td>
<td>0</td>
<td>78</td>
<td>79</td>
</tr>
<tr>
<td>25</td>
<td>17</td>
<td>0</td>
<td>98</td>
<td>91</td>
</tr>
<tr>
<td>30</td>
<td>10</td>
<td>0</td>
<td>117</td>
<td>117</td>
</tr>
<tr>
<td>35</td>
<td>21</td>
<td>0</td>
<td>137</td>
<td>86</td>
</tr>
<tr>
<td>40</td>
<td>22</td>
<td>0</td>
<td>156</td>
<td>144</td>
</tr>
<tr>
<td>45</td>
<td>17</td>
<td>3</td>
<td>176</td>
<td>123</td>
</tr>
</tbody>
</table>
4.1.4.5 Application of the MARS 14 Column to SCOPE Samples

Pooled plasma samples acquired from the SCOPE study were immunodepleted in duplicate for superpool and case samples to assess technical variability in the workflow and singly for each control sample (Figure 4.1.1). An injection volume of 25 µl of plasma was used. Qualitative assessment of the immunodepletions showed gels band to be reproducible between injections (Figure 4.1.9-A), and no peak at 60kDA indicating complete immunodepletion (Figure 4.1.9-B).

Peak areas for the flow-through fraction were also assessed (Figure 4.1.10). Peak area was recorded for a total of 33 injections of the pooled plasma samples acquired from the SCOPE study. Area of the flow-through peak was recorded as an indication of quality and completeness of depletion. The mean peak area and standard deviation were calculated and the peak areas and mean±2SD (dashed line) plotted. All immunodepleted pooled samples used here for iTRAQ labelling (in red) fell within mean±2SD indicating complete and reproducible immunodepletion. The peak area for only one of the remaining samples fell outside this limit indicating a problem with the depletion. The peak areas for the for the other 25 samples fell within this limit suggesting immunodepletion using the MARS 14 system is highly reproducible.
Figure 4.1.9  Qualitative assessment of immunodepletion of pooled SCOPE samples showed depletion to be reproducible and complete. Pooled plasma samples acquired from the SCOPE study and taken at 15 weeks gestation from women who did and did not develop pre-eclampsia were immunodepleted (injection volume 25 µl) in duplicate for reference and case samples and assessed using the Protein 230 assay and measurement of the area under the flow though peak. A- The protein banding pattern appeared reproducible between injections with no band at 60kDa indicating complete depletion. B- The characteristic albumin peak (grey and green traces) was not present in processed SCOPE samples indicating complete depletion.
Figure 4.1.10 Peak area for the flow-through fraction of immunodepleted SCOPE samples were reproducible and indicated complete depletion. Peak area was recorded for immunodepletion of a total of 33 injections of the pooled plasma samples acquired from the SCOPE study and plotted. Mean±2SD (dashed line) was also plotted and used as an indication complete and reproducible depletion. All 8 injections used for iTRAQ labelling in this investigation fell within this range indicating complete depletion of high abundant proteins. Of the remaining 25 samples (blue crosses), the peak area for only one fell outside this range indicating a high level of reproducibility between depletions.
4.1.4.6 Comparison of Immunodepletion of SCOPE Samples by the IgY 14-SuperMix System and the MARS 14 Column

The proteins identified following LC-MS/MS analysis of immunodepleted iTRAQ labelled plasma samples from the SCOPE study were compared. The dynamic range, specificity and reproducibility of each immunodepletion method were compared in order to determine any benefit to the data set produced by application of either method.

4.1.4.6.1 Proteome Coverage and Dynamic Range

Total and unique protein identifications made from analysis of samples immunodepleted using both methods were compared to assess proteome coverage. A total of 392 proteins were identified on combining the data sets. Of those proteins 203 (52%) were from analysis of samples depleted using the IgY 14-SuperMix System only and 61 (16%) from samples depleted using the MARS 14 system (Figure 4.1.11-A). A total of 128 proteins (33% of the total data set) were identified in samples processed by both methods; 38% of the total proteins identified from IgY 14-SuperMix System depleted samples and 68% of total proteins from MARS 14 depleted samples.

The depth of the plasma proteome sampled following immunodepletion by both methods was assessed. Normal concentration ranges for proteins identified were noted from the literature. Concentration ranges were found for a total of 69 proteins identified in sample depleted using the IgY 14-SuperMix System; this accounted for 21% of the total proteins identified in sample processed by this method. Of proteins identified following depletion using the MARS 14 system concentration ranges were found for 46 proteins; 24% of the total proteins identified by this method. For both depletion methods, proteins were identified in the range of the classical plasma proteins ($1 \times 10^5$-$1 \times 10^6$ pg/mL), tissue leakage products ($1 \times 10^5$-$1 \times 10^3$ pg/mL), and interleukins ($1 \times 10^3$-$1 \times 10^4$ pg/mL) as defined by Anderson and Anderson (Anderson and Anderson 2002) (Figure 4.1.11). The distribution of these proteins within each classification was calculated for each depletion method (Figure 4.1.10-B). A larger proportion of tissue leakage products and interleukins were identified in sample depleted using the IgY 14-SuperMix system than the MARS 14 system, for which a larger proportion of high abundance proteins were identified.
Figure 4.1.11  More lower abundance proteins were identified from LC-MS/MS analysis of sample depleted using the IgY 14-SuperMix System than the MARS 14 System. As previously described pooled plasma samples were depleted using both IgY 14-SuperMix System and MARS 14 Systems, iTRAQ labelled and analysed by LC-MS/MS. A - Total and unique protein identification following analysis of sample immunodepleted using each method were assessed, 391 proteins were identified in total. B - A concentration range for 69 proteins identified from IgY 14-SuperMix System depleted plasma and 46 from MARS 14 depleted plasma could be found in the literature. The proportion of proteins falling within the concentration range of the classical plasma proteins (1x10^6 to 1x10^10 pg/mL), tissue leakage products (1x10^3 to 1x10^6 pg/mL) and interleukins (up to 1x10^3 pg/mL) as defined by Anderson and Anderson (Anderson and Anderson 2002) were calculated.
More proteins were identified at lower concentration in sample immunodepleted using the IgY 14-Supermix System than the MARS 14 System. Concentration ranges for proteins identified following LC-MS/MS analysis of plasma depleted using the IgY 14-Supermix System and the MARS 14 System were noted from the literature and plotted. A concentration range for a similar proportion of proteins from sample depleted using each method was found (21% and 24% respectively). Proteins were identified within the concentration range of the Classical Plasma Proteins (1x10^6 to 1x10^9 pg/mL), Tissue leakage products (1x10^3 to 1x10^6 pg/mL), and Interleukins (up to 1x10^3 pg/mL) as defined by Anderson and Anderson (2002). However, more proteins were identified at a lower concentration range from plasma immunodepleted using the IgY 14-Supermix System than the MARS 14 System.
4.1.4.6.2 Efficiency of High Abundance Protein Removal

The efficiency of high abundance protein depletion by each method was compared. Protein identifications made following LC-MS/MS analysis of plasma immunodepleted using the IgY 14-SuperMix System and MARS 14 Systems were searched for named proteins which should have been depleted. Of the proteins to be depleted, 13 of the 14 were common to both methods, with the addition of LDL (ApoB) for the IgY 14 column and Transthyretin for the MARS 14 column. In samples processed using the IgY 14-SuperMix System 9 proteins which should have been depleted were identified, 5 proteins were incompletely depleted using the MARS 14 System, although for both this included 2 isoforms of Fibrinogen (Table 4.1.2). Depletion of high abundance proteins was less efficient using the IgY 14-SuperMix System than the MARS 14 column.

<table>
<thead>
<tr>
<th>Proteins Depleted by the IgY 14 Column</th>
<th>Protein Identifications</th>
<th>Proteins depleted by the MARS 14 Column</th>
<th>Protein Identifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antitrypsin</td>
<td></td>
<td>Antitrypsin</td>
<td></td>
</tr>
<tr>
<td>Albumin</td>
<td>1</td>
<td>Albumin</td>
<td></td>
</tr>
<tr>
<td>Alpha-1-acid glycoprotein</td>
<td>1</td>
<td>Alpha-1-acid glycoprotein</td>
<td></td>
</tr>
<tr>
<td>Alpha-2-macroglobulin</td>
<td></td>
<td>Alpha-2-macroglobulin</td>
<td>1</td>
</tr>
<tr>
<td>Apolipoprotein Al</td>
<td>1</td>
<td>Apolipoprotein Al</td>
<td>1</td>
</tr>
<tr>
<td>Apolipoprotein All</td>
<td>1</td>
<td>Apolipoprotein All</td>
<td>1</td>
</tr>
<tr>
<td>Complement C3</td>
<td>1</td>
<td>Complement C3</td>
<td></td>
</tr>
<tr>
<td>Fibrinogen</td>
<td>2</td>
<td>Fibrinogen</td>
<td>2</td>
</tr>
<tr>
<td>Haptoglobin</td>
<td>1</td>
<td>Haptoglobin</td>
<td></td>
</tr>
<tr>
<td>IgA</td>
<td></td>
<td>IgA</td>
<td></td>
</tr>
<tr>
<td>IgG</td>
<td></td>
<td>IgG</td>
<td></td>
</tr>
<tr>
<td>IgM</td>
<td></td>
<td>IgM</td>
<td></td>
</tr>
<tr>
<td>Transferrin</td>
<td></td>
<td>Transferrin</td>
<td></td>
</tr>
<tr>
<td>LDL (ApoB)</td>
<td>1</td>
<td>LDL (ApoB)</td>
<td></td>
</tr>
<tr>
<td>Transthyretin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>9</strong></td>
<td><strong>5</strong></td>
<td></td>
</tr>
</tbody>
</table>
4.1.4.6.3 Reproducibility

To assess the reproducibility and specificity of sample processing using both methods, the iTRAQ relative quantification ratios for technical replicates (superpool and case) were compared. The iTRAQ relative quantification ratios indicate the abundance of protein in a sample; for technical replicates, the relative quantification ratio for each protein identification should be the same in each replicate if the depletion is reproducible. Relative quantification ratios were compared between replicates, transformed (Log$_2$ (replicate iTRAQ ratio/replicate iTRAQ ratio)), and the percentage frequency (0.05 intervals) of the total protein identifications for each method calculated and plotted (Figure 4.1.13).

If the depletion method is specific and reproducible, transformed comparisons for replicate ratios should equal zero indicating the iTRAQ relative quantification ratio and so protein abundance is the same for each protein in each replicate sample. For all technical replicates, plots of the comparative iTRAQ relative quantification ratios showed a wider and shallower peak for the IgY 14-SuperMix System than the MARS 14 System, for which all three peaks were closely centred around zero. Transformed comparative iTRAQ relative quantification ratios for all replicate samples processed by the IgY 14-SuperMix System had a minimum of -1.96 and maximum value of 5.27, compared to -1.05 and 0.87 respectively for the MARS 14 processed sample replicates.

The wider range of comparative iTRAQ relative quantification ratios for the IgY 14-SuperMix System indicates that proteins identified and quantified in samples processed in duplicate by this system were at differing abundance. This suggests immunodepletion by the IgY 14-SuperMix System to be less reproducible and specific than the MARS 14 Column.
Figure 4.1.13  Immunodepletion by the IgY 14-SuperMix System is less reproducible and specific than immunodepletion by the MARS 14 Column. Pooled plasma samples were depleted using both IgY 14-SuperMix System and MARS 14 Systems, iTRAQ labelled and analysed by LC-MS/MS. Reference superpool and case samples were immunodepleted in duplicate. Relative quantification ratios were compared between replicates, transformed (Log₂(replicate iTRAQ ratio/replicate iTRAQ ratio)) and the percentage frequency (0.05 intervals) of the total protein identifications for each method calculated and plotted. The range of total transformed comparative iTRAQ relative quantification ratios was larger for samples immunodepleted using the IgY 14-SuperMix System than the MARS 14 System (7.24 and 1.12 respectively).
### 4.1.4.7 Results Summary

The aim of this analysis was to compare the performance of the IgY 14-SuperMix System and MARS 14 System in the removal of highly abundant proteins from human plasma. Each system was optimised and proteome penetration, efficiency, specificity and reproducibility compared as summarised in Table 4.1.3.

**Table 4.1.3 Summary of a performance comparison between the IgY 14-SuperMix System and MARS 14 System in the depletion of human plasma.** Pooled plasma samples obtained from women who developed early and late onset pre-eclampsia (each n=12) and 2 distinct sets of healthy controls (each n=12) were immunodepleted using each system. Depleted plasma was labelled using iTRAQ 8-plex reagent and LC-MS/MS analysis carried out. Data were assessed for specificity, reproducibility and depth of proteome penetration, the practical aspects of each system were also considered.

<table>
<thead>
<tr>
<th>Immunodepletion Method</th>
<th>IgY 14-SuperMix System</th>
<th>MARS 14 System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommended maximum application of plasma (µl)</td>
<td>200</td>
<td>40</td>
</tr>
<tr>
<td>Method of assessment of Immunodepletion</td>
<td>SDS-PAGE-Coomassie Stain and LC-MS/MS</td>
<td>Protein 230 Assay and Flow-through Peak Area</td>
</tr>
<tr>
<td>Optimised injection volume used to process SCOPE samples (µl)</td>
<td>150</td>
<td>25</td>
</tr>
<tr>
<td>Protein identification (&gt;95% confidence)</td>
<td>331</td>
<td>189</td>
</tr>
<tr>
<td>Proteins identified by one method (391 total)</td>
<td>203 (52%)</td>
<td>61 (16%)</td>
</tr>
<tr>
<td>Proteins with a concentration range in the literature</td>
<td>69</td>
<td>46</td>
</tr>
<tr>
<td>Classical Plasma proteins identified (% of above)</td>
<td>65%</td>
<td>78%</td>
</tr>
<tr>
<td>Tissue Leakage products identified</td>
<td>31%</td>
<td>20%</td>
</tr>
<tr>
<td>Interleukin Range identified</td>
<td>4%</td>
<td>2%</td>
</tr>
<tr>
<td>Number of proteins identified which should be depleted</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Reproducibility- Range of transformed comparative iTRAQ relative quantification ratios</td>
<td>7.24</td>
<td>1.12</td>
</tr>
</tbody>
</table>
4.1.5 Discussion

Immunodepletion is an essential step in the exploration of the plasma proteome; it allows analysis of proteins through a wider concentration range. Use of the IgY 14-SuperMix System and MARS 14 Column were optimised. The application of both systems to a quantitative hypothesis generating investigation using iTRAQ assessed. The performance of both systems was compared including ease of operation and sample processing, efficiency and reproducibility of depletion.

4.1.5.1 The Genway Seppro® IgY-SuperMix Immunodepletion System

Optimisation of the IgY 14-SuperMix System showed immunodepletion to be variable. Differences between replicate injections were visible on Coomassie stained SDS-PAGE gels and the presence of high abundance proteins which should have been removed by the column were confirmed by LC-MS/MS analysis.

The manufacturer recommends application of up to 200 µl of plasma to the column, however optimisation found that application of less plasma resulted in better depletion. To produce sufficient plasma for iTRAQ labelling two injections of 150 µl of each sample were required. Operation of the column using the 2795 Separation Module required significant manual input; the system uses two columns each must be stripped and re-equilibrated individually, sample was injected manually using an injection needle and depleted plasma was also collected manually using readings from the UV detector as an indication of eluting protein. This may contribute to variation in depletion of the samples and was time consuming, more so as two injections were required per sample. The volume of flow-through collected was relatively large (up to 10mL), requiring downstream processing to exchange into a suitable buffer and workable volume of sample. The completeness of depletion was assessed by SDS-PAGE stained with Coomassie which required large volumes of sample (up to 120 µl) and is relatively time consuming. Use of the IgY 14-SuperMix System with the method reported here was labour intensive and irreproducible. In other hands however, the system and its predecessors which removed fewer proteins, were reported to be reproducible (Liu et al. 2006; Qian et al. 2008; Bandow 2010; Polaskova et al. 2010).

Removal of more proteins is thought to result in the ability to identify more proteins in depleted plasma. As shown here, use of the IgY 14 and SuperMix columns decreases protein content of flow-through; up to an 83% increase in protein identifications by LC-MS/MS has been reported on addition of the SuperMix Column to the IgY 14 Column (Qian et al. 2008). The SuperMix column removes approximately 75 undefined moderately abundant proteins and is produced by inoculating chickens with IgY 14 depleted human plasma; antibodies raised as a result of the challenge are isolated from eggs, conjugated with support medium and packaged into a column (Gassmann et al. 1990; Huang et al. 2005). The amount and specificity of the SuperMix Column antibodies are therefore
dependent on the immune response of the host and vary between columns. Reproducibility of depletion by IgY 14-SuperMix System between columns therefore cannot be guaranteed. The method of production of the SuperMix Column may also contribute to variation in depletion using one column. The binding capacity of the antibodies in the SuperMix Column may be exceeded resulting in differential depletion from sample to sample; this is of particular concern for samples destined for quantitative investigation as the changes in protein levels measured will not reflect the natural differences (Qian et al. 2008). Reproducibility of depletion between samples by the SuperMix Column is questionable, and may contribute to the variability in depletion seen here between replicate samples.

Pooled plasma samples obtained via the SCOPE Study were immunodepleted using the IgY 14-SuperMix System and subsequently labelled using iTRAQ 8-plex reagent. Two injections of 150 µl were made to the columns which were combined during post-depletion buffer exchange. Reproducibility of depletion between samples was shown to be poor; analysis of the SDS-PAGE gel revealed differential depletion of the samples designated superpool 1 and late onset pre-eclampsia 1, with the presence of high abundance proteins which should be depleted confirmed by LC-MS/MS. Although these samples were processed and labelled with iTRAQ reagent, analysis of the LC-MS/MS data for these samples and the duplicate late onset pre-eclampsia 2 sample was not possible due to the differential immunodepletion. Usable data were acquired for only four of a potential seven iTRAQ relative quantification ratios.

4.1.5.2 The Agilent Human 14 Multiple Affinity Removal LC Column

As demonstrated by others, immunodepletion using the MARS 14 System was reproducible (Polaskova et al. 2010; Tu et al. 2010; Smith et al. 2011). Little difference between replicate injections was apparent on analysis of the chromatographic trace and area under the flow-through peak.

Samples were applied to the MARS 14 Column using an Agilent 1200 Series HPLC system which is highly automated, requiring limited manual input and returning less than 1 mL of depleted plasma. Buffer exchange was still required, although the need for downstream processing was greatly reduced compared to sample processing following IgY 14-SuperMix System depletion. The manufacturer recommends application of up to 40 µl of plasma to the column; however optimisation found 25 µl to be preferable and was subsequently used for sample processing. The area under the flow-through peak and the Protein 230 Assay were used as a measure of quality control to assess the completeness of depletion. The peak area for the flow-through peak is reported with the chromatographic trace produced by the HPLC system as part of the depletion process. Although a separate step in the sample processing, the Protein 230 Assay requires no further sample processing, uses only 5 µl of sample and returns results relatively quickly. Reproducibility has been demonstrated over upward of 200 depletions (Brand et al. 2006; Dekker et al. 2007) and in
the course of this investigation only 1 of 33 injections of pooled plasma on to the MARS 14 System failed to meet quality control standards put in place to assess depletion.

The pooled plasma samples acquired via the SCOPE Study were immunodepleted using the MARS 14 System and subsequently labelled using iTRAQ-8 plex reagents. A total of 25µl of plasma from each sample to be labelled was applied to the column; approximately 8 µg of protein was assessed using the Protein 230 Assay and the remainder of each sample was iTRAQ labelled (approximately 90 µg). All samples processed passed quality control measures and data for each iTRAQ label acquired following LC-MS/MS analysis was used. Usable data were acquired for all 7 iTRAQ relative quantification ratios compared to just 4 for IgY 14-SuperMix System depleted sample.

4.1.5.3 Application of Plasma from Non Pregnant and Pregnant Women

It was noted that depletion using both systems differed when plasma from pregnant women was applied to the columns compared to plasma from women who were not pregnant. Investigation using the IgY 14-SuperMix System showed depleted plasma from women who were not pregnant contained less protein than women who were. Application of plasma from women who were not pregnant to the MARS 14 system resulted in a significantly smaller flow-through peak area than for women who were pregnant. The composition of the proteome is highly variable and differs between the sexes (Ogata et al. 2007) it can also be affected by a variety of factors including time of day (Buckley and Dorsey 1970; Macy et al. 1993), age (Buckley and Dorsey 1970), season (Letellier and Desjarlais 1982) and sleep (Linkowski et al. 1998). During pregnancy multiple physiological changes including changes in blood pressure, heart rate and cardiac output occur. Plasma volume increases by approximately 40% (Hays et al. 1985) and changes in the plasma proteome occur, including levels of plasma proteins depleted by both IgY 14-SuperMix System and MARS 14 Systems. Of the proteins depleted by the columns, albumin (Lockitch 1997), IgG, IgA, IgM (Yasuhara et al. 1992; Larsson et al. 2008) and alpha-1-acid glycoprotein (Burtis et al. 2001; Larsson et al. 2008) levels decrease throughout pregnancy, whereas fibrinogen (Lockitch 1997), alpha-1-antitrypsin (Burtis et al. 2001; Larsson et al. 2008) and apo Al and AII (Belo et al. 2002) increase. These changes may affect the performance of the columns. Several studies have applied immunodepletion methods to plasma from pregnant women with and without pre-eclampsia, however no comment was made on the application of these methods and the changes in plasma proteins during pregnancy (Atkinson et al. 2009; Blankley et al. 2009; Blumenstein et al. 2009a; Blumenstein et al. 2009b; Auer et al. 2010). To date there is no documented study of either depletion system with plasma from pregnant women. For both systems injection volumes below the recommended capacity were used to accommodate the changes in protein abundance which occurs during pregnancy.
4.1.5.4 Comparison of Immunodepletion of SCOPE Samples by the IgY 14-SuperMix System and the MARS 14 Column

The protein identifications and relative quantification ratios acquired following depletion of plasma samples acquired from the SCOPE study using both the IgY 14-SuperMix System and MARS 14 System and analysis by LC-MS/MS were assessed. Proteome coverage, specificity and reproducibility of depletion were compared for each data set to determine any advantage or disadvantage to the relative quantification data acquired using either depletion method. Removal of high abundance proteins should improve ‘accessibility’ of the proteome. Where previously the proteome was dominated by a relatively small number of proteins at very high concentration, removal of these proteins by immunodepletion allows access to proteins at lower concentration.

As shown here removal of more proteins is thought to lead to more protein identifications through a wider concentration range (Qian et al. 2008; Bandow 2010; Polaskova et al. 2010). More protein identifications were made following LC-MS/MS analysis of samples depleted using the IgY 14-SuperMix System, which removed approximately 100 high and moderately abundant proteins. Far fewer proteins were identified in samples depleted using the MARS 14 system which removes only the top 14 abundant proteins. Approximately a third of the total number of proteins identified were common to plasma depleted using either system however, this accounted for a smaller proportion of the proteins identified in IgY 14-SuperMix System depleted sample than the MARS 14 depleted sample.

Where possible, concentration ranges for proteins identified were noted from the literature in order to assess the depth of the proteome accessible following depletion. A concentration could be found for a similar proportion of proteins identified following depletion with each method. The samples assessed here were taken at 15 weeks gestation from women who did and did not develop pre-eclampsia while reference concentration ranges taken from the literature were for healthy non-pregnant individuals. Although established that there is some change in the concentration of plasma proteins during pregnancy and illness, ‘normal’ concentration ranges still reflect the depth of proteome which can be sampled following each depletion method. For both methods, classical plasma proteins, tissue leakage products and interleukins were identified down to the 100pg/mL concentration range. A larger proportion of the proteins assigned concentrations were within the tissue leakage and interleukin range for samples depleted using the IgY 14-SuperMix System compared to the MARS 14 System.

Due to the method of operation, LC-MS/MS analysis is biased toward the identification of higher abundance proteins (Qian et al. 2008; Tu et al. 2010). Filtering of data for high confidence protein identifications also introduces a bias toward highly abundant proteins. Proteins with low numbers of peptides are removed from the dataset; low
abundance proteins are likely to be identified by fewer peptides and therefore may be removed from the dataset by filtering for high confidence identifications. Sample processed using the IgY 14-SuperMix System required removal of proteins with less than 2 peptides while data from sample immunodepleted using the MARS 14 column had proteins with less than 3 peptides removed. This may bias the data acquired following IgY 14-SuperMix System depletion toward lower abundance proteins. Approximately the same amount of protein (80-90 µg) from each sample was labelled with each iTRAQ reagent resulting in a similar amount of protein analysed by LC-MS/MS and so the larger number of proteins identified following IgY 14-SuperMix System cannot be accounted for by analysis of a larger amount of protein. While the depletion and subsequent processing of the IgY 14-SuperMix System was a well established method, both depletion and downstream processing using the MARS 14 system required optimisation. The smaller number of proteins identified on analysis of MARS 14 depleted samples was initially thought to be a result of the workflow requiring further development (optimisation of buffer exchange and desalting and chromatography). Although this may contribute, the number of proteins identified here are similar to other studies which have used the MARS 14 System for depletion of human plasma with analysis by both 2D-Gel and LC-MS/MS (Tonack et al. 2009; Auer et al. 2010; Tu et al. 2010). Data from this investigation therefore suggests that removal of more highly and moderately abundant proteins improves the number of proteins and depth of the proteome that can be sampled.

If depletion of more highly abundant proteins results in more protein identifications, it follows that the use of the ProteoPrep® Kit (Sigma Aldrich, UK), the only other depletion method to remove more defined proteins than those used here, should result in more protein identifications. The ProteoPrep® Kit removes 20 highly abundant proteins, more than the IgY 14 and MARS 14 Columns. While efficiency and specificity of depletion were shown to be good, protein identification by LC-MS/MS and 2D-Gel analysis was not significantly improved compared to samples processed using the IgY 12 and MARS 6 and 14 depletion systems (removing 12, 6 and 14 highest abundant proteins respectively) (Roche et al. 2009; Polaskova et al. 2010; Smith et al. 2011). The lack of additional identifications is thought to be due to the method of depletion; the ProteoPrep® Kit used was in spin filter format. Depletion by spin filter is less effective, labour intensive, plasma capacity is limited and multiple depletions have to be carried out to produce enough depleted plasma for analysis (Echan et al. 2005; Polaskova et al. 2010). The same practical problems were also demonstrated using an earlier version of the IgY 14-SuperMix System; the IgY 12 spin column (Corrigan et al. 2011).

Efficiency of depletion by each method was also assessed; protein identifications made following LC-MS/MS analysis of samples depleted using the IgY 14-SuperMix System and MARS 14 system were searched for the 14 named proteins which should be depleted by each system. Specificity of the SuperMix column could not be assessed as the proteins
which are depleted are undefined. Of the 14 named proteins depleted by the IgY 14 and MARS 14 column, 13 of the 14 were depleted by both methods (with the addition of LDL Apo B for the IgY 14 Column and Transthyretin for the MARS 14 Column). More of these proteins which should have been depleted were found in samples processed by the IgY 14-SuperMix System than the MARS 14 System (although both included two isoforms of Fibrinogen). As previously discussed concentrations of many proteins change during pregnancy, including increases in antitrypsin, fibrinogen and apo AI and AII, the concentration of these proteins may therefore outstrip binding capacity of the columns. Although several of the proteins which increase during pregnancy were identified in sample processed by both methods, several proteins which decrease during pregnancy were also identified suggesting that the presence of proteins which should be depleted is not related to pregnancy.

A lack of efficiency of depletion for both systems has been documented elsewhere using non-pregnant plasma. The IgY 14-SuperMix System removes approximately 99% of the protein content of a sample (Bandow 2010), however incomplete depletion has been reported in several other studies (Liu et al. 2006; Qian et al. 2008; Polaskova et al. 2010) with fibrinogen particularly inefficiently removed (Bandow 2010). The MARS 14 System claims to remove approximately 94% of protein from whole plasma, fibrinogen is also reported to be ineffectively removed (Tu et al. 2010) and incomplete depletion of other proteins has also been reported (Polaskova et al. 2010; Tu et al. 2010). Efficiency of depletion using the MARS 6 column was shown to decrease over time, particularly of the immunoglobulins, however the column retained upwards of 98% efficiency of removal (Brand et al. 2006).

Conversely, several studies have reported multiple non-targeted proteins present on analysis of the fraction which contains the depleted proteins (Brand et al. 2006; Gong et al. 2006; Liu et al. 2006; Polaskova et al. 2010; Tu et al. 2010; Corrigan et al. 2011). Proteins may be co-depleted if associated with targeted proteins like albumin (Gong et al. 2006), or have immunoaffinity for column antibodies as a result of sequence homology (Brand et al. 2006; Liu et al. 2006; Tu et al. 2010). Non-targeted proteins may also be retained by the column antibodies; the IgY 14-SuperMix System uses avian IgY antibodies which have higher avidity and demonstrate less non-specific binding to human proteins via the Fc region (Stuart et al. 1988; Larsson et al. 1991; Huang et al. 2005; Corrigan et al. 2011). The MARS 14 Column uses IgG antibodies which have greater affinity for human proteins via the Fc region which may result in non-specific depletion (Larsson et al. 1991; Tu et al. 2010). Several studies have shown non-specific depletion to be reproducible between samples for both depletion methods and therefore total depletion of targeted and non-targeted proteins should be the same between samples where the depletion itself is reproducible (Liu et al. 2006; Qian et al. 2008; Tu et al. 2010). Although not assessed here, it is reasonable to assume that proteins not specifically targeted by the columns were removed by both depletion methods during this investigation and that those proteins may be of interest. The
The overall aim of this investigation was to identify plasma proteins that are differently regulated in women who develop pre-eclampsia during pregnancy in comparison to those who do not develop pre-eclampsia. Plasma from pregnant women who developed pre-eclampsia was shown to retain vasoactive properties after removal of the top six abundant proteins (albumin, IgG, antitrypsin, IgA, transferrin, and haptoglobin) suggesting that the circulating factors responsible for pre-eclampsia are not concurrently removed on immunodepletion of high abundance proteins (Myers et al. 2007) and that proteins of interest are still present in plasma post depletion.

Reproducibility of immunodepletion was assessed across a large number of proteins using the iTRAQ relative quantification ratios for case and reference superpool samples processed in duplicate as technical replicates. Relative quantification ratios were compared for all proteins identified in sample processed using both depletion methods between the paired samples. Comparative relative quantification ratios were shown to vary greatly for proteins identified from sample processed by the IgY 14-SuperMix System 14-SuperMix System indicating depletion to be highly variable from sample to sample. Depletion was shown be particularly variable between superpool and late onset pre-eclampsia samples which had already been qualitatively demonstrated to be differentially depleted. Comparative ratios for proteins identified from samples processed by the MARS 14 System showed less variance indicating greater reproducibility of depletion across a wide range of proteins and between samples.

4.1.6 Conclusion

This aim of this study was to assess the efficiency, reproducibility and effect on the data acquired from LC-MS/MS analysis of iTRAQ labelled human plasma immunodepleted using either the IgY 14-SuperMix or MARS 14 System. This investigation suggests that where proteome penetration is the goal i.e. increasing numbers of protein identifications and the depth of dynamic range sampled, the IgY 14-SuperMix System is of most use. The IgY 14-SuperMix System removes up to 100 high and moderately abundant proteins and depletion of more proteins results in more protein identifications across a greater concentration range. Reproducibility however, was shown to be an issue with the method used here. Where efficiency, specificity and reproducibility of immunodepletion are required the MARS 14 System is ideal, although number of protein identifications and depth of dynamic range may have to be compromised. All of these qualities were required for this investigation. As a result, data acquired from sample immunodepleted using both the IgY 14-SuperMix and MARS 14 System were considered for the subsequent quantitative hypothesis generating study.
4.2 Comparison of Performance of an ESI-qTOF and MALDI-TOF-TOF in Analysis of iTRAQ Labelled Plasma

4.2.1 Introduction

There is evidence to suggest that the application of multiple LC-MS/MS systems to the analysis of one sample produces complementary data, particularly when those systems utilise different ionisation methods. Several studies have demonstrated this including: a comparison of a Global ESI qTOF (Micromass Ltd, Manchester, UK) and 4700 MALDI-TOF-TOF (Applied Biosystems, Warrington, UK) in the analysis of E.coli DNA binding proteins (Stapels and Barofsky 2004); an LTQ Orbitrap (Thermo Fisher Scientific, UK) and 5800 MALDI TOF-TOF (Applied Biosystems, Warrington, UK) using E.coli cell lysate (Stapels and Barofsky 2004; Seymour et al.); and an analysis of the yeast proteome on both the Linear Trap Quadrupole and a QSTAR XL (Applied Biosystems, Warrington, UK)(Elias et al. 2005).

Quantitative data derived from the analysis of iTRAQ labelled samples analysed by multiple systems have also been compared. Shirran and Botting compared measurement of an iTRAQ labelled spiked protein mix using the QSTAR XL and 4800 MALDI-TOF-TOF (both Applied Biosystems, Warrington, UK)(Shirran and Botting 2010) while Kuzyk et al. used the next generation QSTAR Elite and 4800 MALDI-TOF-TOF (both Applied Biosystems) to analyse various ratios of iTRAQ labelled E.coli tryptic digests (Kuzyk et al. 2009). However no comparison of the available LC-MS/MS systems; a QStar XL qTOF and 5800 MALDI–TOF-TOF (both Applied Biosystems, Warrington, UK), has been documented to date.

4.2.2 Aim

- To evaluate the complementary nature of data acquired in the analysis of an 8 plex iTRAQ labelled plasma sample set using both QStar XL qTOF and 5800 MALDI–TOF-TOF LC-MS/MS systems and the applicability of such data to a hypothesis generating investigation.
4.2.3 Method

4.2.3.1 Sample Preparation

Samples were prepared for relative quantification by iTRAQ using the IgY 14-SuperMix Immunodepletion system (see 3.2) and analysed by LC-MS/MS using a QStar XL qTOF and 5800 MALDI–TOF-TOF (hereafter referred to as the QSTAR and 5800 respectively; see 3.2.4.2).

The workflow is outlined in Figure 4.2.1; in brief pooled plasma samples from women who developed early onset pre-eclampsia (n=12) were immunodepleted in duplicate, concentrated, digested and labelled with iTRAQ reagent 117 and 118. One immunodepletion was also carried out on pooled plasma from two distinct sets of healthy controls (each n=12), these control samples were labelled with iTRAQ reporters 119 and 121. A reference superpool consisting of equal amounts of each of the case and control plasma pools was also applied to the workflow and labelled with iTRAQ reporter ion 113 to serve as a reference point for calculation relative quantification ratios.

4.2.3.2 Data Analysis

Data from each LC-MS/MS analysis were interrogated using ProteinPilot 3.0 (see 3.2.5.2) searching against the Human protein database from the IPI (version 3.59) and exported to Microsoft® Excel for comparative analysis.
Figure 4.2.1  Samples were prepared by labelling with 8-plex iTRAQ reagent and LC-MS/MS analysis was carried out using a QStar XL qTOF and 5800 MALDI-TOF-TOF. Pooled plasma samples from women who developed pre-eclampsia (n=12) and 2 distinct sets of healthy controls (each n=12) were depleted of abundant proteins using the IgY 14-SuperMix system. Following tryptic digestion, duplicate pooled samples were labelled with iTRAQ 8-plex reagent and fractionated using high pH reverse phase chromatography. LC-MS/MS analysis was carried out using a QStar XL qTOF and 5800 MALDI–TOF-TOF. The data were interrogated using ProteinPilot 3.0 searching against the Human protein database from the IPI (version 3.59) and exported to Microsoft® Excel for comparative analysis.
4.2.3.3 Comparative Analysis

Data from each LC-MS/MS analysis was compared across several variables, including running time, effectiveness of peptide and protein identification, peptide properties, accuracy and precision of quantification and comprehensiveness. Comparisons were carried out as follows:

Practical Considerations – The total time required for LC-MS/MS analysis of the samples using both the QSTAR and 5800 was noted and compared. The running time and manual input required to operate each LC-MS/MS system was also assessed.

Peptide Identifications - The number of spectra analysed as calculated by ProteinPilot 3.0, number of spectra resulting in a > 95% confidence peptide identification from exported ProteinPilot data (known as a ‘peptide spectral match’) and the number of unique peptides (unique amino acid sequences identified by only one method) calculated from the number of peptides identified at > 95% confidence were compared. The number of peptides identified (> 95% confidence) by both the QSTAR only, the 5800 only and both methods were also compared.

Peptide Characterisation - The MW (as estimated by ProteinPilot) of unique peptide precursors identified by each LC-MS/MS method and with at least one iTRAQ label were compared. The percentage frequency of peptides with MW between 100kDA intervals was calculated. Significant differences between the methods were assessed using Mann-Whitney U (Graphpad Prism version 4.01; San Diego, California).

The pl of unique peptides identified by each method were calculated using the freely available pl Calculator developed by Gauci et al. (https://bioinformatics.chem.uu.nl/supplementary/gauci_proteomics/, Gauci et al. 2008). Unique peptides identified by each method were applied to the calculator, pl for each peptide recorded and the percentage frequency of peptides with pl between each 0.5 increment calculated. Data were assessed for significance using Mann-Whitney U (Graphpad Prism version 4.01; San Diego, California). The hydrophobicity of unique peptides identified by the QSTAR and the 5800 were calculated using the online Sequence Specific Retention Calculator (http://hs2.proteome.ca/SSRCalc/SSRCalcX.html) developed and described by Krokhin et al. (Krokhin et al. 2004). The percentage frequency of peptides falling between 0.5 increments were calculated for each method and the medians assessed for significant differences using Mann-Whitney U (Graphpad Prism version 4.01; San Diego, California).

Lysine and Arginine Distribution- The frequency of tryptic peptides identified by each LC-MS/MS method with lysine or arginine at the carboxyl terminus was calculated. Significant difference in the frequencies for each method was assessed using Fisher’s exact test (Graphpad Prism version 4.01; San Diego, California).
**Protein Identification** - For proteins to be considered identified with 'high confidence', the FDR for the data set must be less than 5% and the UPS greater than 1.3 (see 3.2.5.2). The total proteins identified were assessed and filtered where necessary to identify the total number proteins identified at high confidence (see 4.3.4.2.2). The number of proteins identified by more than 3 peptides of these high confidence proteins was then calculated for both LC-MS/MS methods. The number of proteins (high confidence) identified by each method or both methods were also calculated. The number of peptides and the UPS of the proteins identified with high confidence were assessed. Data were assumed to be non-parametric and data acquired by each method were analysed for significance using Mann-Whitney U (Graphpad Prism version 4.01; San Diego, California).

**Analysis of Relative Quantification Data** - The relative quantification ratios for iTRAQ reporter ions 117, 118, 119 and 121 were calculated by ProteinPilot against the reference superpool labelled with iTRAQ reporter ion 113. The relative quantification ratios for proteins identified with high confidence were assessed and the proportion of proteins with complete quantification data calculated. The efficiency of iTRAQ labelling was also determined by assessing the number of iTRAQ labels attached at potential labelling sites.

**Comparison of Relative Quantification between Methods** - The variation in measurement of one iTRAQ label using two LC-MS/MS methods was assessed. The relative quantification ratios calculated by ProteinPilot for iTRAQ labels 117, 118, 119 and 121 for peptides (>95% confidence) used for quantification and proteins (high confidence) identified by both the QSTAR and 5800 were considered. Relative quantification ratios were calculated against the reference superpool labelled with iTRAQ reporter ion 113. Relative quantification ratios for each peptide and protein identified by both methods were transformed (Log₂) and any difference in the values assessed. The relationship between the measurement of each iTRAQ label was visualised using a scatter plot, with the transformed relative quantification ratio calculated from the QSTAR data on the horizontal axis and the 5800 on the vertical. The perpendicular bisectors across the positive and negative quadrants were also plotted.

**Analysis of Repeated Measurement of Relative Quantification Data** - The variation was also assessed using the Bland-Altman Method of Agreement (Bland and Altman 1986). The method was applied to relative quantification ratios calculated by ProteinPilot against the reference superpool (label 113) for iTRAQ labels 117, 118, 119 and 121 for peptides (>95% confidence) used for quantification and proteins (high confidence) identified by both LS-MS/MS methods. Relative quantification ratios were transformed (Log₂) and the average of the ratios for each peptide and protein identified by both the QSTAR and 5800 calculated and plotted against the difference between those ratios. The mean and standard deviation for protein ratios were also calculated and indicated on the plot (dashed line) and the trend line for each protein data set reported (solid line).
4.2.4 Results

4.2.4.1 Practical Considerations

The practical considerations associated with each method; time taken to run the samples and manual input, were compared to identify any operational benefit to either machine. The total time required for sample analysis using the QSTAR, including LC separation, was 90 hours with limited manual input. The 5800 required a total of 125 hours; 105 hours chromatography and spotting time plus 70 hours MS time. The 5800 also required more manual input due to the offline nature of the chromatography with an estimated 4 hours of input to complete the sample run.

4.2.4.2 Peptide Identifications

The number of spectra and peptides identified, and the efficiency of peptide identification were compared for each method. Approximately 1.5 times more MS/MS spectra were acquired following QSTAR analysis; however efficiency of peptide identification was comparable with 48% of QSTAR spectra and 49% of 5800 spectra contributing to a peptide spectral match. Although the QSTAR resulted in more peptide spectral matches, the 5800 identified a greater number of unique peptides i.e. a greater number of different amino acid sequences, indicating a larger amount of repeated peptide identifications in the QSTAR data (Figure 4.2.2-A). A total of 3,716 unique peptides (> 95% confidence) were identified across both methods, 29% were only identified by the QSTAR, a larger proportion (45%) was identified by the 5800 only, and 25% were identified by both methods (Figure 4.2.2-B).
Figure 4.2.2  The QSTAR identified more spectra and made more peptide spectral matches than the 5800 but identified fewer unique peptides. As previously described pooled plasma samples were depleted of high abundance proteins, iTRAQ labelled and analysed by LC-MS/MS using both the QSTAR and 5800. A- The number of spectra analysed, number of peptides spectral matches and the number of unique peptides identified by each method were calculated. B-The number and proportion of peptides identified by each method and those identified by both methods were assessed, 3716 unique peptides were identified in total.
4.2.4.3 Peptide Characterisation

The MW, $m/z$ ratios, pl and relative hydrophobicity of the peptides identified by the QSTAR only and the 5800 only were compared to develop a profile of the peptides identified by each LC-MS/MS method.

The peptides identified by only the QSTAR had an average larger MW than those identified by the 5800 only, with medians of 1883 and 1636 kDa respectively (Figure 4.2.3-A). The $m/z$ ratios for peptides identified by the QSTAR ranged from 480 to 1191, with a median value of 654 and charge states between 2+ and 6+. For peptides identified by the 5800, the $m/z$ ratios fell between 897 and 3894, with a median value of 1637; all had a charge state 1+. These differences were both shown to be statistically significant using Mann-Whitney U ($p<0.0001$) (Figure 4.2.3-B). The peptides identified by the QSTAR only had a lower pl (median-5.83) and higher relative hydrophobicity (median-30.8) compared to peptides identified by the 5800 only (median; pl-7.99, relative hydrophobicity- 19.1, Figure 4.2.4). These differences were both found to be statistically significant using Mann-Whitney U ($p<0.0001$ for both).
Peptides identified by the QSTAR only, had a larger average molecular weight and lower m/z ratio than those identified by the 5800 only. As previously described, plasma labelled using iTRAQ 8-plex reagent was analysed using both QSTAR and 5800 LC-MS/MS systems. The relative percentage frequencies of unique peptides identified by one method only were compared. A-The relative percentage frequency of unique peptide precursors with MW between 100kDa intervals with at least one iTRAQ label was evaluated. The MW range of peptides identified by the QSTAR only was greater and the median value (1883kDa) larger than those identified by the 5800 only which had a median value of 1636kDa. B- The m/z ratios of peptides identified by the QSTAR only (median; 654) were on average lower and had a smaller range than those identified by the 5800 alone (median: 1637). Relative percentage frequencies for each LC-MS/MS method were compared between intervals of 100. These differences were all found to be statistically significant using Mann-Whitney U (all p<0.0001).
Peptides identified by the QSTAR only had an average lower isoelectric point and higher relative hydrophobicity than those identified by the 5800 only. As previously described, plasma labelled using iTRAQ 8-plex reagent was analysed using both QSTAR and 5800 LC-MS/MS systems. The characteristic of the peptides identified by one method only were compared. A- The pI for peptides identified by one method only was calculated using the freely available pI calculator developed by Gauci et al. (Gauci et al. 2008). The median and relative percentage frequencies of pI for each method were calculated and showed peptides identified by the QSTAR to have lower average pI (median 5.83) compared to the 5800 (median 7.99). B- The median value and percentage frequency of peptides falling between 0.5 increments of relative hydrophobicity as calculated using the Sequence Specific Retention Calculator described by Krokhin et al. (Krokhin et al. 2004) were reported for each method. Peptides identified by the QSTAR alone had higher relative hydrophobicity (median- 30.8) than peptides identified by the 5800 (median 19.1). These differences were both found to be statistically significant using when using Mann-Whitney U (all p<0.0001).
4.2.4.3.1 Lysine and Arginine Distribution

It has been noted in the literature that ESI favours the ionisation of peptides with lysine at the carboxyl terminus and MALDI peptides with arginine at the carboxyl terminus (Krause et al. 1999; Brancia et al. 2000; Stapels and Barofsky 2004; Seymour et al.). Consequently the data acquired here was assessed for this tendency and found to support previously published studies. On analysis of the carboxyl terminus of the unique tryptic peptides, the QSTAR identified a higher proportion of peptides ending in lysine and the 5800 a higher proportion of peptides ending in arginine, this was found to be statistically significant using Fisher's exact test (p<0.001) (Figure 4.2.5).

Figure 4.2.5  The QSTAR identified a higher proportion of peptides with lysine at the carboxyl terminus and 5800 a higher proportion of peptides with in arginine at the carboxyl terminus. Plasma labelled using iTRAQ 8-plex reagent was analysed using both QSTAR and 5800 LC-MS/MS systems. The percentages of unique tryptic peptides with lysine or arginine at the carboxyl terminus of the total number of unique tryptic peptides were compared. The QSTAR was found to identify a higher proportion of tryptic peptides ending in lysine and the 5800 peptides ending in arginine, both were found to be statistically significant using Fisher’s exact test (p<0.001, as indicated by *).
4.2.4.4 Protein Identifications

Several features of the protein data were compared in order to determine any difference or advantage in data acquired by the QSTAR opposed to the 5800. To assess the depth of proteome penetration, total and unique protein identifications for each method were compared. Coverage and confidence in individual protein identifications was assessed by comparing the number of peptides per protein, peptide length and UPS.

The QSTAR identified 319 proteins at high confidence, 56% of those were identified by 3+ peptides and 39% of the total proteins were identified by that method only. A total of 331 proteins at high confidence were identified by the 5800, 75% of those by 3+ peptides, 41% of the total proteins were identified only by the 5800 (Figure 4.2.6-A). On combining the data sets, a total of 455 proteins were identified with high confidence, 28% of these proteins were identified by the QSTAR only, 30% by the 5800 only and 42% were identified with high confidence by both methods (Figure 4.2.6-B).

Comparison of the number of peptides per protein identified by the QSTAR and the 5800, revealed the proteins identified by the 5800 had an average greater number of peptides per protein with a median of 3 and 5, and maximum of 191 and 313, respectively (Figure 4.2.7-A). The UPS of the proteins identified by each method were also compared; proteins identified by the QSTAR had a median UPS of 5 compared to a median of 7 for the 5800 (Figure 4.2.7-B). These difference between both the number of peptides and UPS were found to be statistically significant using Mann-Whitney U (p<0.0001 for both). The average number of amino acids in the unique peptides identified by each method was also calculated and indicated that peptides identified by the QSTAR consisted on average of 14 residues. Peptides identified by the 5800 were shorter and consisted on average of 11 amino acid residues.
Figure 4.2.6  The QSTAR identified a similar number of proteins at high confidence, less with 3+ peptides and a similar number of unique proteins when compared to the 5800. As previously described, plasma labelled using iTRAQ 8-plex reagent was analysed using both QSTAR and 5800 LC-MS/MS systems. A- The total number of proteins identified with high confidence, total number of proteins identified (high confidence) by 3+ peptides and the numbers of unique peptides identified by each method were compared. B- The number and proportion of proteins identified (high confidence) by each method and those identified by both methods were assessed, 455 proteins were identified in total.
Figure 4.2.7  The number of peptides per protein and UPS were both lower for proteins identified by the QSTAR compared to proteins identified by the 5800. As previously described, pooled plasma labelled using iTRAQ 8-plex reagent was analysed using both QSTAR and 5800 LC-MS/MS systems. A- The number of peptides per protein identified by the QSTAR and 5800 were compared. B- The UPS is calculated by ProteinPilot during the analysis of the MS data and is a measure of confidence in protein identification. These differences were both found to be statistically significant using when using Mann-Whitney U (p<0.0001 for both, as indicated by *).
4.2.4.5 Analysis of Relative Quantification Data

Relative quantification ratios calculated by ProteinPilot from the iTRAQ reporter ions associated with each pooled sample were assessed to determine any influence of the LC-MS/MS analysis method used on the quantification data acquired. Relative quantification ratios were calculated for iTRAQ reporter ions 117, 118, 119 and 121 against the reference superpool (labelled with iTRAQ reporter ion 113) following analysis using both the QSTAR and 5800 LC-MS/MS systems. As expected, a similar proportion of proteins identified (high confidence) by each method were found to have incomplete relative quantification data; 4% and 6% of proteins identified by the QSTAR and 5800 respectively.

4.2.4.5.1 Comparison of Relative Quantification between Methods

The relative quantification ratios calculated from iTRAQ reporter ions representing replicates (117, 118, 119 and 121) were also compared across methods in order to assess any effect of the LC-MS/MS method on the resulting quantification ratio. Each LC-MS/MS method must be able to measure the relative quantification ratios equally in order to be applicable. The transformed (Log₂) relative quantification ratios for peptides (>95% confidence and used for quantification) and proteins (high confidence) identified by both the QSTAR and 5800 were plotted against each other in order to assess variation (Figure 4.2.8).

The relative quantification ratios for both peptides and proteins identified by both QSTAR and 5800 followed a linear relationship when plotted against each other (data not shown for 118, 119 and 121). Ratios at the protein level showed less variation than at the peptide level as indicated by the reduction in scatter. This indicates that both the peptide and protein ratios were measured equally by the QSTAR and 5800. It was noted however, that for each iTRAQ label (peptide and protein) more points fell toward the axis representing the 5800 suggesting a bias toward these measurements and further analysis of this relationship was undertaken.
Figure 4.2.8 Relative quantification ratios calculated for peptides (>95% confidence) used for quantification and proteins (high confidence) identified by both the QSTAR and 5800 were measured equally by both methods. As previously described, pooled plasma labelled using iTRAQ 8-plex reagent was analysed using both QSTAR and 5800 LC-MS/MS systems. Transformed (Log₂) relative quantification ratios (calculated against reference superpool labelled with reporter ion 113) for iTRAQ label 117 (data not shown for 118,119 and 121) of peptides (>95% confidence) used for quantification and proteins (high confidence) were plotted. Plots were used to assess variation between methods. The dashed line indicates a 1:1 relationship, peptide data is in blue, protein in red.
4.2.4.5.2 Analysis of Repeated Measurement of Relative Quantification

In order to further assess the relative quantification ratios calculated from iTRAQ reporter ion peak areas measured by the QSTAR and 5800, the Bland-Altman Method of Agreement was applied (Bland and Altman 1986). The relative quantification ratios calculated from iTRAQ reporter ions 117,118,119 and 121 against the reference superpool labelled with iTRAQ reporter ion 113 for peptides (>95% confidence) used for quantification and proteins (high confidence) identified by both the QSTAR and 5800 were used (Figure 4.2.9). According to the Bland-Altman method, if repeated measurements taken by different methods are in agreement, the plot should show a trend toward the horizontal.

The mean and standard deviation for protein ratios were calculated. The large proportion of both peptide and protein data points falling within this range indicates good reproducibility of measurement of the ratios by both methods. However, the trend-line for the protein data were skewed from the horizontal. This indicates a consistent bias by one method, in this case the 5800, to measure a more positive result when the ratio is greater than one (i.e. indicating an increase in that protein compared to the denominator) and a more negative result when the ratio is less than one, (i.e. indicating an decrease in that protein compared to control the denominator) and so the difference in ratios between methods is larger at the extremes of the range of ratio measurements.

4.2.4.6 Results Summary

The aim of this analysis was to compare the performance of the QStar and 5800 LC-MS/MS system in the analysis of a single 8-plex iTRAQ labelled plasma sample set. Several variables were compared including running time, peptide and protein identifications and accuracy of quantification, the results are summarised in Table 4.2.1.
The Bland Altman Method of Agreement showed ratios obtained using the 5800 were greater than those from the QSTAR at the extremes of the range of the ratios. The Bland-Altman Method of Agreement is used to assess the level of agreement in repeated measurements made using two different systems. Relative quantification ratios calculated for pooled plasma samples labelled with iTRAQ 8plex reagents (labels 117, 118, 119 and 121, with ratios calculated against 113) and analysed using a QSTAR and 5800 LC-MS/MS system were compared. The average of the transformed (Log2) relative quantification ratios for iTRAQ labels 117, 118, 119 and 121 of peptides (>95% confidence) used for quantification and proteins (high confidence) identified by both the QSTAR and 5800 were calculated and plotted against the difference between those ratios (Data not shown for 118, 119 and 121). The line of best fit is indicated by the continuous line and one standard deviation from the mean by the dashed line both calculated from protein data.
Table 4.2.1  Summary of performance comparison between an ESI-qTOF and MALDI-TOF-TOF in the Analysis of an 8-plex iTRAQ Labelled Plasma Sample Set.

Plasma samples were pooled, immunodepleted and labelled using iTRAQ 8-plex reagent. LC-MS/MS analysis was carried out on both a QSTAR and 5800 and the results compared. Data were assessed for dissimilarity in running time, effectiveness of peptide and protein identification, and accuracy and precision of quantification.

<table>
<thead>
<tr>
<th>LC-MS/MS System</th>
<th>QStar XL qTOF</th>
<th>5800 MALDI TOF-TOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running time</td>
<td>90 hours</td>
<td>125 hours</td>
</tr>
<tr>
<td>Number of spectra analysed</td>
<td>52987</td>
<td>39665</td>
</tr>
<tr>
<td>Peptide spectral matches (% of spectra)</td>
<td>25554 (48%)</td>
<td>19532 (49%)</td>
</tr>
<tr>
<td>Unique peptide sequences (% of spectra)</td>
<td>2030 (8%)</td>
<td>2620 (13%)</td>
</tr>
<tr>
<td>Peptide sequences identified by each method (3716 total)</td>
<td>1096 (30%)</td>
<td>1686 (45%)</td>
</tr>
<tr>
<td>Unique tryptic peptides with Lysine at the C-terminus (Fisher’s exact test, p&lt;0.001)</td>
<td>1032 (53%)</td>
<td>1107 (43%)</td>
</tr>
<tr>
<td>Unique tryptic peptides with Arginine at the C-terminus (Fisher’s exact test, p&lt;0.001)</td>
<td>922 (47%)</td>
<td>1445 (57%)</td>
</tr>
<tr>
<td>Tryptic peptide sequences identified by each method</td>
<td>1954</td>
<td>2552</td>
</tr>
<tr>
<td>MW of unique peptides with iTRAQ label(s) (median, Mann-Whitney U, p&lt;0.0001)</td>
<td>1883</td>
<td>1636</td>
</tr>
<tr>
<td>pI of unique peptides (Median, Mann-Whitney U, p&lt;0.0001)</td>
<td>5.83</td>
<td>7.99</td>
</tr>
<tr>
<td>m/z ratio of peptides (Median, Mann-Whitney U, p&lt;0.0001)</td>
<td>654</td>
<td>1637</td>
</tr>
<tr>
<td>Charge State</td>
<td>1+</td>
<td>2+ to 6+</td>
</tr>
<tr>
<td>Relative hydrophobicity of unique peptides (median, Mann-Whitney U, p&lt;0.0001)</td>
<td>30.7</td>
<td>19.1</td>
</tr>
<tr>
<td>Protein identifications (high confidence)</td>
<td>319</td>
<td>331</td>
</tr>
<tr>
<td>Proteins with 3+ peptides (% of total high confidence identifications)</td>
<td>180 (56%)</td>
<td>248 (75%)</td>
</tr>
<tr>
<td>Proteins identified by one method (455 total)</td>
<td>126 (28%)</td>
<td>136 (30%)</td>
</tr>
<tr>
<td>Peptides per protein (median, Mann-Whitney U, p&lt;0.0001)</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Average amino acids per peptide</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>Unused Protein Score (median, Mann-Whitney U, p&lt;0.0001)</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Percentage of proteins with incomplete quantification data</td>
<td>4%</td>
<td>6%</td>
</tr>
</tbody>
</table>
4.2.5 Discussion

It has been suggested that analysis of one sample by multiple LC-MS/MS methods is of benefit to the dataset produced, particularly when the systems utilised have different ionisation methods. This study aimed to compare data produced in analysis of one iTRAQ labelled plasma sample set by both a QStar XL qTOF and 5800 MALDI–TOF-TOF LC-MS/MS. Data acquired was assessed in order to determine any difference in running time, effectiveness of peptide and protein identification, accuracy and precision of quantification.

4.2.5.1 Practical Considerations

LC-MS/MS analysis using the QSTAR required less time and input than the 5800. The extra time taken for sample analysis using the 5800 however was considered insignificant unless time allowed for analysis is a limiting factor. In this investigation 30 fractions from the previous High pH Reverse Phase separation were analysed by the QSTAR compared to 35 for the 5800. The analysis of these extra fractions may account for the longer run time required by the 5800 however the average time to analyse one fraction is still longer for the 5800 than the QSTAR. Although requiring more total time, an advantage of the static nature of ionisation and offline chromatography utilised by the 5800 is the production of a sample set that may be stored and re-analysed. Evidence suggests samples can be stored on a target plate at room temperature for up to 5 months with little effect on protein identification or iTRAQ relative quantification ratios (Kuzyk et al. 2009). This is also of benefit if there are mechanical problems during the run allowing reanalysis of the plate whereas the dynamic online nature of ESI used by the QSTAR means samples are lost when mechanical faults occur. Although not utilised here, the ability to store and re-analyse samples prepared for MALDI analysis is a unique advantage allowing for the possibility of repeat analysis after further optimisation of data collection or validation of a prior analysis.

4.2.5.2 Peptide and Protein Identifications

Features of the peptide and protein data were compared in order to determine any advantage in the data acquired by the QSTAR or the 5800 and the effect on the complete data set. Repeated analysis of a sample set will increase the number of peptides identified (Stapels and Barofsky 2004; Song et al. 2008), repeated analysis using complementary methods also increases the total number of peptides identified (Elias et al. 2005; Molle et al. 2009; Seymour et al. 2009). In this study, peptide identification was increased by 45% by using both the QSTAR and 5800. An increase of up to 50% has been documented using complementary methods compared to 25% per additional run when repeating analysis using the same method (Stapels and Barofsky 2004). This indicates a clear advantage to the number of peptides identified by repeating analysis using complementary methods.
The numbers of unique peptides (i.e. different amino acid sequences) identified by each method were assessed on the assumption that this would contribute to the identification of more proteins. The increased number of peptide identifications made by the repeated analysis was not evenly distributed between methods; in accordance with previous studies, more unique peptides were identified by the 5800 (Stapels and Barofsky 2004; Seymour et al. 2009). Identification of more unique peptides using the 5800 was expected due to the method of operation. The static nature of MALDI ionisation (see 1.3.1.1.2) allows a survey scan of precursor ions to be carried out prior to selected re-analysis of the sample and selection of targets at the highest intensities. In this LC-MS/MS, analysis the 5800 selected the most abundant 27 precursors per fraction for MS/MS analysis; identical peaks in adjacent spots were filtered to select the strongest precursor and precursor ions with the weakest signal to noise ratio were acquired first to maximise signal intensity for low abundance peptides (see 3.2.4.2.2). Filtering of precursor peaks is particularly effective when using MALDI as it results primarily, in singly charged precursor ions; as a result each precursor has only one \( m/z \) ratio and can be effectively excluded from subsequent analysis.

The QSTAR uses ESI (see 1.3.1.1.1) a dynamic method where analysis of precursor ions is dependent on precursor intensity. The choice of peptide for analysis is carried out in real time, reducing the opportunity to refine the choice of ion. During ESI analysis, limitations on the intensity of a precursor ion and application of dynamic exclusion, where repeated analysis of the same precursor ion is prevented over a short period of time, help to ensure precursor ions are selected at the optimal elution point. In this experiment, the QSTAR selected the two highest abundance ions satisfying charge, count and \( m/z \) criteria for fragmentation and dynamic exclusion was applied for 1 minute (see 3.2.4.2.1). This prevents repeated analysis through the remainder of the chromatographic peak. This does not exclude analysis of the ion for the remainder of the run resulting in re-analysis of the same product ions multiple times and the identification of fewer unique peptides. Filtering of precursors is less effective here as ESI results in a range of multiply charged precursor ions and as a result one precursor ion can have multiple \( m/z \) ratios. While one \( m/z \) can be excluded from re-analysis, alternative charge states at different \( m/z \) ratios are not recognised as the same precursor and may be selected for analysis again, contributing to redundancy in the identification of unique peptide sequences.

Although the method of ionisation is thought to be a major contributor to the nature and number of the peptides identified by each LC-MS/MS system, there are also other differences between the systems which must be taken into account. The QSTAR couples ESI to a quadrupole and TOF mass analyser while the 5800 uses a TOF-TOF combination. Several studies have utilised split flow LC, coupled simultaneously to an ESI source and MALDI spotter, with analysis carried out by a hybrid MS/MS system which can accommodate both ionisation methods (Bodnar et al. 2003; Molle et al. 2009). While Bodnar et al. found similar results by changing only the ionisation method, the properties and
proportions of the peptides and proteins identified by Molle et al. differed from previous studies using separate LC-MS/MS systems. Molle et al. suggested that coupling of the ionisation method to its idealised analysis system accentuates the differences between the ionisation methods, however it may also be concluded that the benefit to any system is in optimisation of that system to utilise the strengths of the individual components.

Repeat analysis of a sample set using complementary methods also increases the number of proteins identified; this investigation showed approximately 30% more proteins were identified by additional analysis with either the QSTAR or 5800, with 42% of identifications made by both methods. Although previously documented that repeated analysis using complementary methods increases the number of protein identifications, the proportions of the increase were found to differ from that shown here (Bodnar et al. 2003; Stapels and Barofsky 2004; Molle et al. 2009; Seymour et al. 2009).

Previous studies using E.coli lysates and bovine subcellular fractions showed a larger proportion of total proteins identified by both methods and that more peptides and proteins were identified by repeated analysis using a MALDI based LC-MS/MS platform (Bodnar et al. 2003; Stapels and Barofsky 2004; Molle et al. 2009; Seymour et al. 2009). In comparison, in this investigation, the 5800 identified more unique peptides; however this did not contribute to more total or unique protein identifications. More fractions (5) from the RP-HPLC separation were analysed by the 5800, however as only 4% more proteins were identified by the 5800 than the QSTAR, these fractions must be of relatively low protein content. In comparison to the refined protein mixes used in the studies discussed, the human plasma samples used in this study are highly complex with a far greater dynamic range. Proteomic analysis cannot identify all the proteins within the plasma proteome during one or repeated analyses. In protein mixes with a more limited dynamic range, a MALDI platform may be able to penetrate the proteome and the additional peptides identified contribute to new protein identifications. In a sample with a larger dynamic range, although a MALDI LC-MS/MS system can identify more unique peptides; these still contribute to the same proteins as identified by an ESI platform. The range of detection is flooded by peptides from a limited number of proteins occurring within the dynamic range of detection. This suggest that when characterising total protein content additional analysis using a MALDI based platform may be of more use. When sampling a proportion of a proteome however, additional analysis using an ESI or MALDI LC-MS/MS system is equally beneficial. In this case, data from both the QSTAR and 5800 is equally applicable.

If the greater number of unique peptides identified by the 5800 does not increase the number of protein identifications, is the benefit seen elsewhere; does a MALDI based platform identify a greater number of peptides per protein and how might this affect the quantification data? The median number of peptides per protein was higher for proteins identified by the 5800 suggesting the larger number of unique peptides identified by the 5800 contributes to an increase in the number of peptides identified per protein. This did not
increase the average sequence coverage which was similar for both methods. Peptides identified by the 5800 were, on average, shorter than those identified by the QSTAR which may account for the lack of increase in sequence coverage, however this may be of benefit to the confidence of the identifications. The UPS, as calculated by Protein Pilot (see Appendix), is a measure of confidence in protein identification and so a comparison of the difference in the UPS gives an indication of any difference in confidence of protein identification between methods. This data showed proteins identified by the 5800 to have a larger UPS in agreement with Seymour et al. who surmised that the greater confidence in identification was the result of the larger number of peptides identified per protein (Seymour et al.).

4.2.5.3 Peptide Characterisation

Characteristics including MW, pl and relative hydrophobicity, of the peptides identified by the QSTAR and the 5800 were compared to determine any differences in the profile of the peptides identified by each LC-MS/MS system. In agreement with previous studies, the QSTAR identified peptides with a larger average molecular weight and over a wider mass range than the 5800 (Stapels and Barofsky 2004). The range of m/z ratios for identified precursors was smaller for the QSTAR compared to the ratios of precursors identified by the 5800. While the QSTAR identified precursors with charge states ranging between 2+ and 6+, the 5800 identified only singly charged precursor ions. As a function of the multiply charged precursors produced by ESI ionisation the m/z ratio for the QSTAR is compressed allowing identification of precursors through a much wider mass range than the 5800 for which singly charged ions predominate. The range of m/z ratios identified by the 5800 therefore directly reflects the MW range of the tryptic peptides present in the sample and so larger peptides falling outside the range of detection cannot be identified.

Peptides with at least one iTRAQ label were considered in this analysis; however the MW of the peptide increases proportionally to the number of iTRAQ labels (Ernoult et al. 2008) and there is also evidence to suggest that the iTRAQ label also increases the charge state of the ion generated (Thingholm et al. 2010). This may limit the number of iTRAQ labelled peptides which can be identified by the 5800 as both MW and charge state fall outside detectable limits. While the multiple charge states generated by the QSTAR are a hindrance to the identification of unique peptides, it allows identification of peptides over a larger natural mass range. This also enables identification peptides which may have both mass and charge increased due to the presence of multiple iTRAQ labels.

As shown in other studies, the QSTAR using ESI ionisation identified more acidic peptides while the 5800 using MALDI ionisation identified more basic peptides (Stapels and Barofsky 2004; Molle et al. 2009). The presence of basic amino acids is thought to promote proton transfer between the matrix and peptide during MALDI ionisation resulting in more efficient fragmentation (Olumee et al. 1995; Zhu et al. 1995) which is enhanced by the
addition of an iTRAQ label (Ross et al. 2004; Aggarwal et al. 2006; Zieske 2006). The tendency for LC-MS/MS systems utilising ESI ionisation to identify more acidic peptides has been previously documented (Zhen, Xu et al. 2004; Aggarwal, Choe et al. 2006) and is thought to be due to enhanced fragmentation (Qin and Chait 1995), however the molecular basis for this bias has yet to be discovered.

This study showed a preference for QSTAR to identify more hydrophobic peptides than the 5800, which has also been documented in other studies (Stapels and Barofsky 2004; Molle et al. 2009). It has been suggested that during formation of the ESI droplet, hydrophobic amino acid containing peptides tend toward the surface and so are preferentially ionised (Cech and Enke 2000). This tendency was exaggerated here, due to the use of formic acid as an ion pairing agent during RP-HPLC separation by the QSTAR LC-MS/MS system. While RP-HPLC was carried out using the same column for both systems, TFA was used as an ion pairing agent for the 5800 LC-MS/MS system. TFA provides superior resolution and recovery of peptides but cannot be used with an ESI ionisation source as it prevents formation of the droplet required for ionisation of peptides (Apffel et al. 1995; Gustavsson et al. 2001; García 2005). Formic acid binds peptides to the column less strongly than TFA and so more hydrophobic peptides which would have remained bound to the column using TFA are eluted. However, a tendency to identification of hydrophobic peptides has been shown where one RP-HPLC separation has subsequently been split and applied to LC-MS/MS systems utilising both ESI and MALDI and is therefore also a function of the LC-MS/MS system used as well as the ion pairing agent (Molle et al. 2009).

The data in this study also demonstrated the previously reported tendency for an ESI platform to favour ionisation of peptides with lysine at the carboxyl terminus and a MALDI based system for peptides terminating in arginine (Krause et al. 1999; Brancia et al. 2000; Stapels and Barofsky 2004; Seymou et al.). Although shown in other studies (Stapels and Barofsky 2004), the tendency of ESI for lysine terminating peptides is not well understood and it has been suggested that ESI has no affinity for peptides ending in either residue (Ernoult et al. 2008). It may be that the larger proportion of lysine terminating peptides identified is simply a reflection of the larger proportion of peptides with lysine at the carboxyl terminus in any tryptic digest (Zhen et al. 2004; Wysocki et al. 2005).

The preference of MALDI for the identification of peptides with arginine at the carboxyl terminus is thought to be due to the stabilising arginine side chain (Bonetto et al. 1997; Brancia et al. 2000). The bias is more pronounced during MS mode where stability is an advantage, however the side chain offers little benefit during MS/MS investigation and so this preference is less distinct (Zhen et al. 2004), even more so on addition of an iTRAQ label (Ross et al. 2004; Ernoult et al. 2008). In order to avoid this tendency it would be possible to use an alternative protease, for example GluC which cleaves at the carboxyl side of glutamate, or AspN which cleaves on the amino side of aspartate, instead of trypsin which
results in arginine and lysine terminating peptides. Trypsin is the industry standard and the majority of databases required for identification of proteins from their peptide components are based on tryptic digests. There is also evidence to suggest that the LC-MS/MS system used may contribute toward identification of a biased peptide population at an individual amino acid level which would not be prevented by using an alternative protease (Stapels and Barofsky 2004; Zhen et al. 2004).

This study supports the previously reported tendency for ESI and MALDI based LC-MS/MS based systems to preferentially identify certain subsets of peptides. Although these tendencies may affect simpler protein mixes, in the analysis of highly complex and heterogeneous samples such as human plasma where there are many more peptides than can be identified, the tendency is not detrimental. Identification of a differential subset of peptides by the QSTAR and 5800 may actually increase coverage of the proteome, again suggesting that the use of complementary systems is highly beneficial.

4.2.5.4 Analysis of Relative Quantification Data

The relative quantification ratios calculated from iTRAQ reporter ions were compared between methods. Both methods showed similar efficiency of iTRAQ labelling and numbers of proteins with incomplete relative quantification data indicating both methods are equally able to identify the iTRAQ tag. Each LC-MS/MS method must also be able to measure the relative quantification ratios equally in order to be applicable. The relative ratios calculated from iTRAQ reporter ions measured by each LC-MS/MS method were also compared. Initial investigation indicated that both methods produced comparable iTRAQ relative ratios for proteins and to an expected lesser extent, peptides, identified by both methods. Protein iTRAQ relative ratios are produced from multiple relative ratios calculated at the peptide level and combined in a weighted manner into a single relative ratio for the protein those peptides are assigned to, therefore removing some of the variance seen at the peptide level. Further investigation indicated that there was a consistent trend for the relative ratio acquired following 5800 analysis to be larger than that acquired using the QSTAR. This was confirmed using the Bland-Altman Method of Agreement which showed relative ratios calculated from the 5800 acquired data to be larger than that acquired using the QSTAR with the effect increasing toward the extremes of the measured relative ratios.

Differences in observed and expected iTRAQ relative quantification ratios have been previously documented. This is thought to be due to poor selection of precursor ions in complex mixtures resulting in contamination of reporter ion peaks by non-reporter ions causing compression, and ultimately, an underestimation of the expected ratio (Pierce et al. 2008; Kuzyk et al. 2009; Ow et al. 2009; Shirran and Botting 2010). The level of underestimation increases with increasing relative ratio and the choice of processing software used to analyse the MS data also contributes (Shirran and Botting 2010). The data from this investigation showed ratios calculated from the QSTAR acquired data to be
consistently lower, as shown in the literature the underestimation was more pronounced
when using an ESI based platform where the ratio is calculated from an average of multiple
measurements which may include less accurate ratios measured when the peptide was not
optimal (Shirran and Botting 2010). Ratios calculated from data acquired using a MALDI LC-
MS/MS system are thought to be more accurate as the increased ionisation efficiency of
MALDI results in one optimised measurement of the reporter ion for each peptide (Shirran
and Botting 2010). This may be an advantage of the larger numbers of peptides identified
per protein when using the 5800. Accuracy of ratio measurement using both systems can be
improved by optimising LC separation prior to MS/MS analysis to reduce the complexity of
the sample (Ow et al. 2011) however this must be balanced with the practicalities of
analysing multiple fractions. Contamination of the iTRAQ reporter ions by isotopes of other
reporter ions may also contribute to the underestimation of relative ratios, however, this can
be successfully countered by application of isotope correction factors provided by the
manufacturer with the iTRAQ reagents (Ow et al. 2009).

All of the evidence suggests that underestimation of the iTRAQ ratios within an
experiment is systematic; Karp et al. demonstrated a linear relationship between the
observed and expected ratios which became more pronounced with larger fold changes and
that this relationship was independent of signal strength (Karp et al. 2010). As a result, they
suggested that if proteins of a known ratio were spiked into a complex sample, a correction
factor could be calculated and applied in order to overcome the underestimation in the
observed fold changes (Karp et al.). The experiment here contains no proteins at a known
ratio and therefore, it is not possible to confirm if the difference in ratios is due to an
underestimation of the fold change however, there is no reason to suspect that this does not
apply here. The lack of proteins at a known ratio also means that the linearity of the
underestimation and so a correction factor cannot be calculated, nor the contribution of
processing the data using ProteinPilot. Again, there is also no reason that the relationship
would not be linear as previously demonstrated and so the effect applied equally throughout
the dataset. This effect is not unique to iTRAQ; a similar underestimation of relative ratios
has been shown using several other quantitative mass spectrometry methods including
SILAC (Ong et al. 2002; Parker et al. 2009) and label free spectral counting (Old et al.
2005).

In this case, the iTRAQ method was applied to a hypothesis generating study, the aim
being to identify proteins for which there is a significant difference in relative abundance
compared to the remainder of the data and the exact magnitude of the fold change is
unimportant. The iTRAQ method, despite the documented lack of accuracy in relative
quantification ratios when analysed using the QSTAR and 5800 is therefore still valid when
applied to hypothesis generating investigations.
4.2.6 Conclusion

This investigation has shown that while there is overlap between the datasets, there are also differences in the peptide and protein populations identified by the QSTAR and 5800 LC-MS/MS systems. No one analysis method is able to identify all the proteins in a complex sample such as human plasma and the divergence of the protein populations identified by the two different LC-MS/MS systems expands the range of the data which can be collated. This study has also shown that there is some discrepancy in the measurement of iTRAQ relative quantification ratios by each system, however application of both systems to hypothesis generating studies utilising relative quantification is by no means compromised, although an awareness of this issue during downstream processing may be necessary.

This study has therefore shown that the QSTAR and 5800 LC-MS/MS systems produce complementary data and are both equally applicable to the analysis of iTRAQ labelled human plasma samples. As a result of this analysis, data acquired using both QSTAR and 5800 LC-MS/MS systems were considered for the subsequent hypothesis generating investigation.
4.3 Evaluation of Two Optimised Workflows for Relative Quantification of iTRAQ Labelled Plasma

4.3.1 Introduction

The quantitative MS method iTRAQ has been applied to plasma to study a range of diseases including pre-eclampsia (Blankley et al. 2009; Auer et al. 2010). Advantages of the method include the ability to analyse multiple samples and obtain data on protein identification and quantitation during a single analysis. There is however, no established method for the preparation of plasma for iTRAQ labelling. Several protocols have been published (Ernoult et al. 2008; Song et al. 2008; Tonack et al. 2009; Burkhart et al. 2011) however sample processing remains highly dependent on the equipment and resources available.

During the course of this investigation two workflows were developed and optimised. The first workflow was developed to label plasma samples which had been immunodepleted using the IgY 14-SuperMix System with iTRAQ reagent. The MARS 14 immunodepletion system was subsequently acquired. A second workflow was developed to process plasma depleted by the MARS 14 system for labelling with iTRAQ reagent. Two LC-MS/MS systems; the QSTAR and 5800 were available for analysis of iTRAQ labelled samples. A direct comparison of the advantages and disadvantages of the immunodepletion systems and LC-MS/MS methods has been carried out (see 4.1 and 4.2). Each immunodepletion method and LC-MS/MS analysis lead to the identification of a different population of proteins; the reliability and reproducibility of quantitative data acquired in each dataset is of particular interest as it is used to identify candidate biomarkers. A comparison of all three datasets was carried out in order to establish if each provided data of equal value for future analysis.

4.3.2 Aim

- To compare and evaluate optimised workflows for labelling of plasma proteins with iTRAQ reagent and the applicability of data subsequently acquired to a hypothesis generating investigation.
4.3.3 Methods

4.3.3.1 Preparation of Samples for Relative Quantification by iTRAQ

Plasma samples were obtained from women who subsequently developed late onset pre-eclampsia (n=12), early onset pre-eclampsia (n=12) and two distinct sets of healthy control (each n=12) matched for gestational age only. Plasma samples were pooled according to phenotype and an additional reference superpool consisting of equal amounts of each plasma sample produced. The pooled plasma samples were then prepared for relative quantification by iTRAQ using two parallel workflows as outlined in Figure 4.3.1 (see 3.2). In brief, pooled plasma samples were immunodepleted using the IgY 14-SuperMix System (Figure 4.3.1-blue workflow, see 3.2.2). Depleted plasma was concentrated and exchanged into a suitable buffer, digested using trypsin and labelled using iTRAQ 8-plex reagent. Pooled plasma for superpool, late and early onset pre-eclampsia samples were processed in duplicate forming technical replicates (labelled with iTRAQ reagent 115 and 116, and 117 and 118 respectively). Each control group was processed once and labelled with iTRAQ reagent 119 and 121. Labelled samples were pooled and separated into multiple fractions using high pH reverse phase chromatography. Fractionated samples were analysed by LC-MS/MS using both the QSTAR and 5800 resulting in two datasets referred to as IgY 14-SuperMix-QSTAR and IgY 14-SuperMix-5800. Sample processing was subsequently repeated using the MARS 14 workflow. Pooled plasma samples were processed using the MARS 14 immunodepletion column, concentrated, digested and labelled with iTRAQ reagent as outlined above (Figure 4.3.1-red workflow, see 3.2.3). Labelled samples were pooled and separated into multiple fractions using high pH reverse phase. Fractionated samples were analysed using the 5800 MALDI--TOF-TOF resulting in the dataset referred to as MARS 14-5800.

4.3.3.2 Comparison of Protein Identification and Relative Quantification Data

Data from each LC-MS/MS analysis were interrogated using ProteinPilot 3.0 (see 3.2.5.2), searching against the Human protein database from the IPI (version 3.59). The relative quantification ratios for iTRAQ reporter ions were calculated by ProteinPilot, the reference superpool labelled with iTRAQ reporter ion 113 was chosen as the denominator. Data were exported to Microsoft® Excel. Comparison of the datasets was carried out across several criteria.
Figure 4.3.1  Pooled plasma samples acquired from the SCOPE Study were immunodepleted, labelled with iTRAQ 8-plex reagent and analysed by LC-MS/MS.
Pooled plasma samples from women who developed late and early onset pre-eclampsia (each n=12) and two distinct sets of healthy controls (each n=12) were pooled according to phenotype and a reference superpool consisting of equal amounts of each case and control plasma produced. Pooled plasma was immunodepleted using either the IgY 14-SuperMix System or MARS 14 column, digested and labelled using iTRAQ 8-plex reagent. Labelled samples were pooled and separated into multiple fractions by high pH RP. Sample processed using the IgY 14-SuperMix System Workflow was analysed by LC-MS/MS using the QStar XL qTOF and 5800 MALDI–TOF-TOF producing the ‘IgY 14-SuperMix-QSTAR’ and ‘IgY 14-SuperMix-5800’ datasets. Sample processed using the MARS 14 Workflow was analysed by LC-MS/MS using the 5800 MALDI–TOF-TOF producing the ‘MARS 14-5800’ dataset.
4.3.4 Results

Pooled plasma samples were prepared for relative quantification by iTRAQ using two parallel workflows. The efficiency of the workflow and quality of the data produced was assessed.

4.3.4.1 Preparation of Samples for Relative Quantification by iTRAQ

As previously discussed, immunodepletion using the IgY 14-SuperMix System was less reproducible than the MARS 14 system. As a result one superpool and the late onset pre-eclampsia samples were not reproducibly processed in the IgY 14-SuperMix datasets; these samples were not considered suitable for further analysis. Following immunodepletion using the IgY 14-SuperMix System, protein concentration for the remaining samples was calculated using the Bradford assay. Protein concentration in depleted samples ranged from 180 to 366 µg/ml; 80 µg of protein was digested and labelled with iTRAQ 8-plex reagent. Digestion and labelling were optimised as part of a prior investigation (Blankley et al. 2009). Digestion and labelled was assessed by LC-MS/MS and analysed using MASCOT. High confidence protein identifications were assessed and found to have one or no missed cleavages and complete iTRAQ modification.

An injection of 25 µl of plasma was used for samples immunodepleted using the MARS 14 column and the entire processed product of each injection labelled with iTRAQ 8-plex reagent (approximately 90 µg). Digestion and labelling was carried out according to the iTRAQ protocol and was assessed briefly and qualitatively by MS/MS analysis using the 5800. Digestion and labelling was found to be sufficient to proceed.

Labelled samples immunodepleted with the IgY 14-SuperMix system were combined and separated into multiple fractions using a 3 µm Fortis C18 high pH reverse phase column. Chromatography was first optimised using trypsin digested cytochrome C; elution peaks were sharp and narrow, each peptide eluted over approximately 30 seconds, and the UV trace returned to baseline between each peak indicating good separation and column performance (Figure 4.3.2-A). Fractions were therefore collected at 30 second intervals between 9 and 36.5 minutes of the gradient resulting in 55 fractions. The fractions containing the majority of the peptides were analysed by LC-MS/MS (Figure 4.3.2-B).

Labelled samples immunodepleted using the MARS 14 column were combined and separated into multiple fractions using a 3µm Extend-C18 high pH reverse phase column. Chromatography was also optimised with trypsin digested cytochrome C. Fractions were collected every 15 seconds for a total of 30 minutes, resulting in a total of 120 fractions. The fractions containing the most peptides were analysed by LC-MS/MS.
Figure 4.3.2  High pH reverse phase chromatography was optimised using a cytochrome C digest and used to separate depleted pooled plasma samples labelled using iTRAQ 8-plex reagent. A- High pH reverse phase using a 3 µm Fortis C18 high column was optimised using a cytochrome C digest. Peak widths for separation of the single protein digest were approximately 30 seconds wide. B- Pooled plasma from women who did and did not develop pre-eclampsia during pregnancy was immunodepleted using the IgY 14-SuperMix System, digested and labelled using iTRAQ 8-plex reagent. Labelled samples were combined and separated into multiple fractions by high pH RP. Fractions were collected at 30 second intervals between 9 and 36.5 minutes, fractions containing the majority of the peptide content collected between 11 and 26 minutes were subsequently analysed by LC-MS/MS.
4.3.4.2 Comparison of Protein Identification and Relative Quantification Data

Protein identifications and relative quantification ratios as calculated by ProteinPilot against the superpool sample labelled with iTRAQ reporter ion 113 were exported to Microsoft® Excel for further analysis. Quality of the data were assessed using various measures.

4.3.4.2.1 Assessment of Digestion and Labelling with iTRAQ 8-plex reagent

The number of unexpected missed cleavages, lysine residues and labelled lysine residues were calculated for each dataset acquired following LC-MS/MS analysis (Table 4.3.1). The rate of missed cleavage indicated trypsin digestion was of equal efficiency in each sample preparation method. Labelling efficiency was also similar for each method at 100% for the MARS 14-5800 dataset and slightly less for the IgY 14-SuperMix-QSTAR and IgY 14-SuperMix-5800 datasets. In summary, sample preparation across all three datasets was comparable.
The rate of missed cleavage and iTRAQ labelling of lysine residues indicated digestion efficiency and labeling by iTRAQ reagent was similar for both IgY 14-SuperMix and MARS 14 sample preparation methods. Data were acquired by LC-MS/MS analysis of pooled plasma samples from women who did and did not develop pre-eclampsia during pregnancy labelled with iTRAQ 8-plex reagent. To indicate the efficiency of trypsin digestion the rate of missed cleavage was calculated from the number unexpected missed cleavage sites and total number of peptides. The efficiency of labelling by iTRAQ reagent was estimated from the number of lysine residues with an iTRAQ label.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>IgY 14-SuperMix-QSTAR</th>
<th>IgY 14-SuperMix-5800</th>
<th>MARS 14-5800</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unexpected Missed Cleavage Sites</td>
<td>782</td>
<td>794</td>
<td>1301</td>
</tr>
<tr>
<td>Number of Peptides</td>
<td>25554</td>
<td>19532</td>
<td>48353</td>
</tr>
<tr>
<td>Rate of Missed Cleavage</td>
<td>3.06%</td>
<td>4.06%</td>
<td>2.69%</td>
</tr>
<tr>
<td>Number of iTRAQ Labelled Lysine Residues</td>
<td>7803</td>
<td>5503</td>
<td>11820</td>
</tr>
<tr>
<td>Number of Lysine Residues</td>
<td>7819</td>
<td>5525</td>
<td>11820</td>
</tr>
<tr>
<td>iTRAQ Labelling Efficiency (%)</td>
<td>99.8%</td>
<td>99.6%</td>
<td>100%</td>
</tr>
</tbody>
</table>
4.3.4.2.2 High Confidence Protein Identifications

The total number of high confidence protein identifications made following analysis of each dataset was compared. For the IgY 14-SuperMix-QSTAR dataset, proteins with high confidence were reported by ProteinPilot and the FDR for the dataset was 0.61%. Datasets where analysis was carried out by the 5800 required filtering to produce a list of proteins considered identified with high confidence. Proteins with a UPS less than 1.3 were considered identified with low confidence and were removed from the list. The FDR for the dataset must also be below 5% (Elias and Gygi 2007), in order to achieve this proteins with the least number of peptides were removed from the dataset. For the IgY 14-SuperMix-5800 dataset proteins with a UPS less than 1.3 and with fewer than 2 peptides were removed and for the MARS 14-5800 dataset proteins with a UPS less than 1.3 and with fewer than 3 peptides were removed. The FDR for the filtered IgY 14-SuperMix-5800 dataset was 0.63% and 1.63% for the filtered MARS 14-5800 dataset.

A total of 502 high confidence protein identifications were made; 319 proteins were identified in the IgY 14-SuperMix-QSTAR dataset, 331 by the IgY 14-SuperMix-5800 dataset and 189 in the MARS 14-5800 dataset (Figure 4.3). A total of 113 proteins were identified in all 3 datasets; 23% of the total protein identifications (35%, 34% and 60% of each dataset respectively). Of these high confidence identifications, 307 (99%) of the proteins identified in the IgY 14-SuperMix-QSTAR dataset had complete quantification data, 309 (93%) of the IgY 14-SuperMix-5800 dataset and 164 (86%) in the MARS 14 dataset.
A total of 502 proteins were identified with high confidence, with 113 proteins identified in all three datasets. Total protein identifications made by ProteinPilot interrogation of data acquired following LC-MS/MS analysis of iTRAQ labelled immunodepleted pooled plasma samples were filtered to remove low confidence identifications. A total of 502 proteins considered to be identified with high confidence remained following filtering, with 113 of those proteins identified in IgY 14-SuperMix-QSTAR, IgY 14-SuperMix-5800, and MARS 14-5800 datasets.
4.3.4.2.3 Assessment of iTRAQ Labelling

ProteinPilot analysis was carried out using ‘bias correction’ which corrects for uneven mixing of labelled sample. A bias correction algorithm is applied on the assumption that most proteins do not change in expression and that labelled samples are combined in equal ratios. A bias correction factor is calculated for each iTRAQ relative quantification ratio and applied. Bias correction factors as reported by ProteinPilot were compared for each LC-MS/MS analysis (Table 4.3.2). The bias correction factor for the 119:113 (control: superpool) relative quantification ratio for the MARS 14-5800 dataset was extremely low indicating either unequal combination of sample or a problem with the 119 iTRAQ reagent. This sample was not considered suitable for further analysis.

Table 4.3.2 The bias correction factor calculated by ProteinPilot for the 119:113 (control: superpool) iTRAQ relative quantification ratio for the MARS 14-5800 dataset was low indicating the ratio was unreliable and unfit for use in further data analysis. Pooled plasma samples were immunodepleted using the IgY 14-SuperMix System and MARS 14 column, labelled with iTRAQ 8-plex reagent and analysed by LC-MS/MS using a QStar XL qTOF and 5800 MALDI–TOF-TOF. Data from each LC-MS/MS analysis were interrogated using ProteinPilot 3.0 which calculates a bias correction factor for each relative quantification ratio to correct for uneven mixing of labelled sample.

<table>
<thead>
<tr>
<th>Ratio</th>
<th>Sample</th>
<th>IgY 14-SuperMix-QSTAR</th>
<th>IgY 14-SuperMix-5800</th>
<th>MARS 14-5800</th>
</tr>
</thead>
<tbody>
<tr>
<td>114:113</td>
<td>Superpool</td>
<td></td>
<td></td>
<td>0.98</td>
</tr>
<tr>
<td>115:113</td>
<td>Late onset pre-eclampsia</td>
<td></td>
<td></td>
<td>1.02</td>
</tr>
<tr>
<td>116:113</td>
<td>Late onset pre-eclampsia</td>
<td></td>
<td></td>
<td>1.34</td>
</tr>
<tr>
<td>117:113</td>
<td>Early onset pre-eclampsia</td>
<td>1.29</td>
<td>1.17</td>
<td>1.31</td>
</tr>
<tr>
<td>118:113</td>
<td>Early onset pre-eclampsia</td>
<td>1.23</td>
<td>1.08</td>
<td>1.31</td>
</tr>
<tr>
<td>119:113</td>
<td>Control</td>
<td>1.57</td>
<td>1.30</td>
<td>0.15</td>
</tr>
<tr>
<td>121:114</td>
<td>Control</td>
<td>1.35</td>
<td>1.15</td>
<td>0.90</td>
</tr>
</tbody>
</table>
### 4.3.4.2.4 Assessment of iTRAQ Relative Quantification

A final assessment of the cumulative reproducibility of each workflow was made. The iTRAQ relative quantification ratios for duplicate samples (superpool, late and early onset pre-eclampsia and control) were compared. The relative quantification ratios should be the same for both samples if sample processing is reproducible. Technical replicates (superpool, late and early onset pre-eclampsia) where sample is processed in duplicate and therefore only reflects variance introduced by the workflow, are expected to be highly reproducible. Control samples were also compared although the samples do not represent a replicate. The two controls groups consist of women with healthy pregnancies, comparing the two groups reflects not only the variance introduced by the workflow but also the differences between the two control populations. Although not a true replicate, comparison of the two control groups still represents a useful indication of the suitability of the samples for further analysis.

Relative quantification ratios were compared between duplicate samples, transformed (\(\log_2(\text{duplicate ratio/duplicate ratio})\)) and the percentage frequency (0.05 intervals) of the total protein identifications for each dataset calculated and plotted (Figure 4.3.4). A reproducible replicate should show a clear peak around zero indicating the iTRAQ relative quantification ratio and so protein abundance is the same for each protein in each replicate sample.

The superpool and late onset pre-eclampsia replicates for the MARS 14-5800 data produced a sharp centred peak indicating highly reproducible data. Early onset pre-eclampsia samples produced a centred peak in all three datasets, however the peak was sharper in the MARS 14-5800 dataset indicating reproducibility may be greater in that workflow. As expected the peak produced by analysis of the two control sample groups in the IgY 14-SuperMix datasets was wider than that for technical replicates. The peak was centred at approximately zero, indicating reproducible quantification across the two sample groups.
Figure 4.3.4  Assessment of processing of duplicate samples confirmed early onset pre-eclampsia and control samples were reproducibly processed in the IgY 14-SuperMix datasets and superpool, early and late onset pre-eclampsia samples were reproducibly processed in the MARS 14-5800 dataset. Pooled plasma samples were depleted using both IgY 14-SuperMix System and MARS 14 Systems, iTRAQ labelled and analysed by LC-MS/MS. Reference superpool and case samples were immunodepleted in duplicate and each control sample singly. Relative quantification ratios were compared between replicates, transformed (Log₂ (replicate ratio/replicate ratio)) and the percentage frequency (0.05 intervals) of the total protein identifications for each method calculated and plotted.
4.3.4.3 Summary of Samples for Analysis

In conclusion, several samples were deemed unsuitable for further analysis. As discussed previously, one superpool sample and both late onset pre-eclampsia samples in the IgY 14-SuperMix QSTAR and 5800 datasets were not reproducibly immunodepleted. These samples were assessed in the context of all three datasets and were not considered suitable for further analysis. The control pool labelled with reagent 119 in the MARS 14-5800 dataset was also found to be unsuitable for analysis due to an abnormal bias correction factor. In all datasets the superpool (iTRAQ reagent 113) and early onset pre-eclampsia samples were available for analysis. In the IgY 14-SuperMix QSTAR and 5800 datasets both control samples were processed reproducibly and therefore suitable for analysis, while only one control sample was available for the MARS 14-5800 dataset. A summary of samples remaining for analysis is shown in Table 4.3.3.

| Table 4.3.3 | Samples available for analysis following immunodepletion, iTRAQ labelling, LC-MS/MS analysis and data processing are shown in colour, samples unavailable for analysis due to sample processing issues are shown in grey. Pooled plasma samples acquired from the SCOPE Study were immunodepleted using the IgY 14-SuperMix System and MARS 14 column, labelled with iTRAQ 8-plex reagent and analysed by LC-MS/MS using a QStar and 5800. Pooled plasma samples from women who developed late and early onset pre-eclampsia (each n=12) and two distinct sets of healthy controls (each n=12) were pooled according to phenotype and a reference superpool consisting of equal amounts of each case and control plasma produced. Samples in the IgY 14-SuperMix-QSTAR and IgY 14-SuperMix-5800 datasets were not analysed due to irreproducible immunodepletion, one sample in the MARS-14-5800 dataset due to a problem with the iTRAQ reagent; samples indicated in grey. |
|---|---|---|---|---|
| | Technical Replicate | Technical Replicate | Technical Replicate | Control Samples |
| iTRAQ Label | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 121 |
| IgY 14-SuperMix - QSTAR | Super pool 2 | Super pool 1 | Late PE Case Pool 1 | Late PE Case Pool 2 | Early PE Case Pool 1 | Early PE Case Pool 2 | Control 1 Pool | Control 2 Pool |
| IgY-14-SuperMix - 5800 | Super pool 2 | Super pool 1 | Late PE Case Pool 3 | Late PE Case Pool 2 | Early PE Case Pool 1 | Early PE Case Pool 2 | Control 1 Pool | Control 2 Pool |
| MARS 14 – 5800 | Super pool 1 | Super pool 2 | Late PE Case Pool 1 | Late PE Case Pool 2 | Early PE Case Pool 1 | Early PE Case Pool 2 | Control 1 Pool | Control 2 Pool |
4.3.5 Discussion

Two optimised workflows were developed to process plasma samples for analysis using iTRAQ. The first workflow was developed to process samples which had been immunodepleted using the IgY 14-SuperMix System. A second workflow was developed when the MARS 14 Immunodepletion system subsequently became available. Pooled plasma samples were processed using both workflows. Samples processed using the IgY 14-SuperMix workflow were analysed using the QSTAR and 5800 and samples processed using the MARS 14 workflow were assessed using the 5800, producing a total of three datasets. Evaluation and comparison of all three datasets is essential prior to further analysis in order to determine if all three datasets provide data of equal quality.

4.3.5.1 Preparation of Samples for Relative Quantification by iTRAQ

As previously discussed, immunodepletion is an essential step in the analysis of the plasma proteome as it allows access to lower abundance proteins that are most likely to be reflective of health status. Two different immunodepletion methods were used in this study; the IgY 14-SuperMix System and the MARS 14 Column. Investigation of the different immunodepletion methods revealed advantages and disadvantages to each system and that each system results in identification of a different protein population (see 4.1). Use of the IgY 14-SuperMix System was not always reproducible and as a result, three of the pooled samples immunodepleted using this method were unfit for further analysis. The MARS 14 system was shown to be extremely reproducible and all pooled samples processed using this method were suitable for further analysis.

Approximately the same amount of protein was digested and labelled for both the IgY 14-SuperMix System and the MARS 14 column. Although equal volumes of plasma were applied to the IgY 14-SuperMix System, protein concentration was calculated post-depletion and was found to vary widely across all pooled samples. Adjustment of the sample volume was therefore required to ensure equal amounts of protein were digested and labelled with iTRAQ reagent. Investigation of the MARS 14 column demonstrated protein content post depletion to be reproducible between depletions (see 4.1). A set volume of plasma was therefore immunodepleted for each pooled sample and labelled with iTRAQ reagent with no further adjustment made for protein concentration. For relative quantification ratios to be reproducible and representative, it is essential that the iTRAQ reagents are applied to the same amount of sample; however how that same amount of sample is measured can be interpreted differently. Here, both protein concentration and sample volume were used. Measurement of the protein concentration using a protein assay introduces a further source of variance into the workflow, with up to 30% variance in the BCA assay reported (Burkhart et al. 2011), however it is essential where the immunodepletion method is not reproducible. Application of the iTRAQ reagent to a set volume of plasma would therefore seem to be a more reliable method but risks introducing variance through the inherent variation in protein.
concentration between healthy people. For example, serum albumin concentration in plasma can vary widely from person to person, and may be a result of biology, hydration state or disease status (Larsson et al. 2008). Without any normalisation it may be impossible to tell which of these factors is responsible for any changes in protein abundance observed. Even with normalisation of protein concentration between samples, observed changes in protein abundance may also be a result of differing natural ranges of proteins rather than disease state. Where the immunodepletion method is shown to be reproducible, measurement of sample amount by volume of plasma is also valid and is ultimately a more useful measure. This reflects the situation in a clinical laboratory setting where normalisation for protein concentration prior to sample assay is not possible.

The digested and labelled mixture was then separated into multiple fractions for analysis by high pH reverse phase chromatography, which also removes unbound iTRAQ labels. Separation was optimised using a simple one protein cytochrome C digest and elution times applied to the separation of the more complex plasma digest. Simplified fractions were then applied to the LC-MS/MS for further orthogonal separation and analysis. As previously discussed, LC-MS/MS analysis was carried out by both a QSTAR and 5800 which use ESI and MALDI ionisation respectively. The LC conditions used with both systems were optimised to achieve full separation of peptides by liquid chromatography. Ionisation and collision energy were optimised to produce good spectra for both quantification and protein identification as demonstrated in previous successful experimental runs (Unwin et al. 2005; Pierce et al. 2008). Investigation as part of this study of both the QSTAR and 5800 systems in the analysis of iTRAQ labelled human plasma samples demonstrated that both systems were equally applicable and produced complementary data (see 4.1.1). No one analysis method is able to identify all proteins in a plasma sample and the use of the two complementary LC-MS/MS systems increases the range of proteins which can be identified and quantified.

4.3.5.2 Protein Identifications and Relative Quantification by iTRAQ

Assessment of digestion and labelling using iTRAQ reagent, both during sample processing and from the LC-MS/MS data, indicated both processes were efficient. The rate of missed cleavage and labelling efficiency were similar across all datasets despite the different processing methods used. Although several optimised iTRAQ workflows have been published, trypsin digestion and to a lesser extent labelling with iTRAQ reagent, will never be completely efficient (Tonack et al. 2009; Burkhart et al. 2011). Non tryptic peptides may be present in the digested sample as a result of plasma proteases; non tryptic digestion was reduced here by maintaining samples at 4º C at all possible times while processing. Protease inhibitors may also be used, however inhibitors are more applicable to serum samples where plasma proteases are more active as a result of the clotting cascade and were not used with these samples (Yi et al. 2008). Labelling efficiency of the iTRAQ reagent
is calculated by assessing the number labelled lysine residues. Although the iTRAQ reagent also reacts with the free amine group at the N-terminus of peptides, labelling is less efficient and may be blocked by modifications or subsequently lost or cleaved giving a less accurate measure of labelling efficiency. As previously discussed (see 4.1.1) it has been suggested that ESI based LC-MS/MS platforms such as the QSTAR have a bias toward the identification of lysine terminating peptides; this may be a factor of the larger number of lysine peptides in any tryptic digest. In comparison, MALDI based LC-MS/MS systems such as the 5800, show a bias toward identification of arginine terminating peptides, these instrument biases may account for the larger number of lysine residues identified in the IgY 14-SuperMix-QSTAR sample set compared to the IgY 14-SuperMix-5800 for which only the LC-MS/MS analysis step of the workflow differed. Although not completely efficient, digestion and labelling efficiency were comparable across all three datasets.

The choice of search algorithm used can also affect the quality of the protein identification and quantitation data obtained (Kapp et al. 2005). ProteinPilot was chosen for this study. ProteinPilot uses the Paragon™ Algorithm to identify peptides from tandem mass spectrometry data. As opposed to other methods of peptide identification, which rely on user entered data on modifications and either estimated sequence data or precursor mass to search databases, The Paragon™ Algorithm models probabilities of modifications, substitutions and cleavage events in conjunction with the Paragon™ Algorithm Sequence Temperature Values tag method (Shilov et al. 2007). This method identifies multiple sequence tags which can be used to reduce the ‘search space’; areas of the database are considered ‘hot’ if strongly implicated by the tag set for that spectrum and are considered with less common post translational modifications. Regions of the database which are considered ‘cold’ i.e. not implicated by the tags are only considered with more common modifications (Shilov et al. 2007). Searching of a larger search space can be carried out without losing sensitivity; this is not possible using other search algorithms (Kapp et al. 2005). While this may result in loss of peptides with low frequency or poor fragmentation, a balance must be maintained between speed, discrimination and accuracy (Shilov et al. 2007).

ProteinPilot also uses the Pro Group™ Algorithm, which recognises that the same data cannot be used multiple times to justify detection of multiple proteins (Biosystems. March 2009). The algorithm bundles identified proteins into groups of related proteins which are compared according to the spectra they explain rather than the identified peptides they contain. Protein groups contain ‘winner’ proteins which are the primary identified protein and for which there is the greatest amount of evidence not explained by better proteins and ‘competitor’ proteins for which there is less evidence. Relative quantification ratios are then calculated by ProteinPilot using only unique peptides (altered expression of splice forms and isoforms of variant proteins can therefore be calculated) and ensuring that reporter ions do not contribute to relative quantification ratios for multiple proteins. The relative quantification
ratios of the unique peptides are then combined to produce a weighted average ratio for the ‘winning’ protein group to which the peptide contributes. A total of 502 proteins were identified across all three datasets, approximately 300 by each IgY 14-SuperMix system dataset and less than 200 for the MARS 14-5800 datasets, as previously discussed the differences in numbers may be a result of the depth of the proteome sampled following each immunodepletion method. Matching of proteins across the datasets may be hampered by inconsistent processing and identification by ProteinPilot, while one ‘winner’ protein may be identified in one dataset from one body of evidence, a different ‘winner’ may be identified from the same evidence in another dataset. Ideally all three datasets should be processed as one, combining the whole body of evidence into one set of protein identifications and quantification data. This was not possible using ProteinPilot (v3.0) as the program requires information on the LC-MS/MS instrument used for analysis to be entered. Knowledge of the instrument used for analysis allows the program to define parameters for data processing, including charge state and fragmentation patterns. Combined analysis of datasets produced by different LC-MS/MS instruments is not possible.

The data acquired by each LC-MS/MS method was reported differently by ProteinPilot. Only high confidence protein identifications were reported for sample analysed by the QSTAR. All protein identifications were reported by ProteinPilot following processing of data acquired using the 5800. As a result, filtering was required to produce a dataset containing only high confidence protein identifications. This filtering method may result in the removal of low abundance proteins identified by only one or two peptides but still be of interest to the investigation; however to guarantee the integrity of the dataset and confidence in the protein identifications this filtering is necessary. In order to try to overcome the experimentally generated differences in protein abundance between samples, ProteinPilot calculates a ‘bias correction factor’ which corrects for uneven labelling. Based on the assumption that most proteins do not change in abundance between samples, a correction factor is calculated and applied to each iTRAQ relative quantification ratio. In this investigation bias correction factors indicated samples were of similar amounts and measurement of sample by protein concentration and volume were equally valid. The bias correction factor is also an indication of the behaviour of the iTRAQ reagent as it corrects for both uneven mixing and labelling of sample. The bias correction factor for the control sample relative quantification ratio (119:113) in the MARS 14-5800 dataset was abnormally low. This suggests a problem with the iTRAQ reagent rather than the amount of sample as correction factors for the remaining quantification ratios for sample produced using the same workflow fell within the expected range. It has been anecdotally noted that the iTRAQ reagent 119 is less reliable than other labels, quantification data for sample labelled with this reagent is often lost (Personal communication; Unwin R.D.). As a result the relative quantification ratio for the control sample labelled with iTRAQ reagent 119 in the MARS 14 5800 dataset was considered unreliable and unfit for further analysis.
Analysis of samples using iTRAQ allows multiple samples to be compared at once. Samples are immunodepleted, digested and labelled individually, once labelled, downstream processing is carried out on the whole of the labelled sample reducing variation in sample processing. Due to the isobaric nature of the labels, peptides from differently labelled samples have the same m/z and so elute and enter the collision chamber together. This increases the signal strength and so sensitivity of LC-MS/MS analysis. On fragmentation of the amino acid backbone, the iTRAQ reporter ion is also fragmented allowing sequencing of the peptide and relative quantification. The LC-MS/MS platform must be optimised for the iTRAQ reagent kit used, while both 4-plex and 8-plex labels increase the charge state peptides, the effect is more pronounced with the 8-plex iTRAQ kit (Thingholm et al. 2010). Collision energy must also be optimised to achieve efficient fragmentation of both the amino acid backbone but also the iTRAQ reporter ion and a compromise must be made to achieve both. Although the iTRAQ reporter ions appear in the ‘quiet’ region of the spectra where few other ions will affect the quantification, as previously discussed, reporter ions may become contaminated in complex mixtures such as plasma resulting in an underestimation of the relative quantification ratio (See 4.1.1 Pierce et al. 2008; Kuzyk et al. 2009; Ow et al. 2009; Shirran and Botting 2010). The magnitude of the underestimation increases with increasing relative ratio and may be affected by the LC-MS/MS system and data processing software used (Shirran and Botting 2010). Underestimation can be reduced by optimisation of the chromatography to produce less complex fractions for analysis. The need for more, but still not completely accurate relative quantification ratios must be balanced with the practicalities and expense of increasing the numbers of fractions for analysis (Ow et al. 2011).

Other methods of relative quantification are available; however these also suffer from the same disadvantages as iTRAQ without the benefits of multiplex analysis and accurate quantitation in MS/MS mode. The iCAT method developed by Gygi et al. only allows comparison of two samples and labels only cysteine containing peptides which vastly reduces the number of proteins which can be quantified (Gygi et al. 1999). A similar underestimation of relative ratios has been shown using several other quantitative mass spectrometry methods including SILAC, which is not applicable to clinical samples (Ong et al. 2002; Parker et al. 2009) and label free spectral counting (Old et al. 2005). While label free methods allow comparison of an infinite number of samples, a highly reproducible sample processing method is required to avoid introduction of variance between samples, data analysis is extremely complex and error prone and depending on the method used, may not include the identity of changed proteins (Neilson et al. 2011).

A final assessment of the iTRAQ relative quantitation ratios was made to ensure sample processing, LC-MS/MS analysis, data processing and filtering and the inbuilt issues discussed previously did not affect the ratios calculated for each sample. The use of technical replicates was carefully considered in design of this hypothesis generating investigation; replicates allow assessment of the variation introduced by each workflow.
Superpool and late and early onset pre-eclampsia samples were processed in duplicate forming technical replicates. Variation between the ratios of the two processed samples is therefore a result of the workflow and analysis. Two groups of control samples from women with healthy pregnancies were also processed. Although the two control groups do not represent a true replicate, the two groups are useful as an indication of the range of biological variation within a control population. Subsequent data analysis to identify proteins which were changed in abundance in pre-eclampsia samples compared to healthy controls relied upon the reproducibility of these ‘duplicate’ samples. Assessment of the reproducibility of these samples is therefore required prior to further data analysis.

The superpool and late onset pre-eclampsia ratios for the MARS 14-5800 dataset appeared reproducible. Data indicated early onset pre-eclampsia samples were well processed and reproducible. As expected the data for the control samples for the IgY 14-SuperMix datasets was less defined but still indicated a comparable level of reproducibility. As such, despite the biological differences between the two sets of samples, the two control groups were considered ‘paired’ for further analysis.

4.3.6 Conclusion

In conclusion, the two workflows developed produced comparable data. Sample processing using the IgY 14-SuperMix workflow however, was less reproducible. A total of seven samples were unusable, six of which were a direct result of the IgY 14-SuperMix immunodepletion. While removing more high abundance proteins with the IgY 14-SuperMix system may be an advantage as it allows identification of lower abundance proteins, the lack of reproducibility is a distinct disadvantage. Investigations using iTRAQ require highly reproducibly processed samples to calculate relative quantification ratios, in this case for the identification of proteins changed in abundance in pre-eclampsia. The remaining samples were found to be reproducibly processed; including early onset pre-eclampsia and control samples in the IgY 14-SuperMix datasets and superpool, late and early onset pre-eclampsia and one control sample in the MARS 14-5800 dataset. Comparison across all three datasets indicated data from each was of equal value and subsequent analysis should consider each dataset with equal weight.
5 Identification and Validation of Changed Plasma Proteins in Early Pregnancy in Women Who Develop Pre-eclampsia

5.1 Identification of Changed Plasma Proteins at 15 Weeks Gestation in Women Who Develop Pre-eclampsia Using iTRAQ

5.1.1 Introduction

The ability to identify women at risk of developing pre-eclampsia prior to clinical presentation would be advantageous. It has been established that there are proteins which cause endothelial dysfunction present in the plasma of women prior to a diagnosis of pre-eclampsia (Myers et al. 2005). Multiple plasma proteins have been investigated as early pregnancy predictive markers of pre-eclampsia, however, a predictive marker with sufficient sensitivity and specificity for clinical use in a nulliparous low risk population has yet to be identified.

Hypothesis generating proteomic methods are unbiased and have been applied to the identification of predictive markers for pre-eclampsia. Several studies have demonstrated the ability of quantitative mass spectrometry to identify changes in plasma proteins in pre-eclampsia in late pregnancy. Investigations of plasma sampled at time of disease (Blankley et al. 2009) and at term (Auer et al. 2010) has been carried out using iTRAQ. Both studies identified and subsequently, validated proteins which changed in abundance in the plasma of women who developed pre-eclampsia compared to those who did not, suggesting the iTRAQ method is applicable to the identification of candidate proteins predictive of pre-eclampsia. To date, however, iTRAQ has not been applied to plasma samples taken from pre-symptomatic women. Early onset pre-eclampsia samples were obtained via the SCOPE Study. The study recruited healthy nulliparous women with singleton pregnancies and collected plasma samples between 14 and 16 weeks of pregnancy.

At the time the study was designed, this gestation coincided with the timing of blood tests performed during routine antenatal care. Although women no longer make a routine antenatal visit at 15 weeks, a test for a prognostic biomarker identified at this gestation could easily be incorporated into existing antenatal care.

5.1.2 Aim

- To utilise iTRAQ to identify plasma proteins that change in abundance at 15 weeks gestation in women who subsequently develop pre-eclampsia compared to women with uncomplicated pregnancies.
5.1.3 Method

5.1.3.1 Samples for Relative Quantification

Pooled plasma samples collected at 15 weeks gestation as part of the SCOPE study were processed and analysed as previously described (see 4.3.3.1). Plasma samples were obtained from women who subsequently developed late onset pre-eclampsia (n=12), early pre-eclampsia (n=12) and two distinct sets of healthy controls (each n=12) matched for gestational age only. An aliquot of all four pooled samples was also combined to produce a ‘superpool’ sample for use as a reference. In brief, pooled samples were immunodepleted using the IgY 14-SuperMix and MARS 14 systems and labelled using iTRAQ 8-plex reagent. Labelled samples were analysed using both the QSTAR and 5800 resulting in three datasets referred to as the ‘IgY 14-SuperMix-QSTAR’, ‘IgY 14-SuperMix-5800’ and the ‘MARS 14-5800’. Data from each LC-MS/MS analysis was interrogated using ProteinPilot 3.0 searching against the Human protein database from the IPI (version 3.59). The relative quantification ratios for iTRAQ reporter ions were calculated by ProteinPilot against the reference superpool labelled with iTRAQ reporter ion 113 and data were exported to Microsoft® Excel. Each dataset was filtered to produce a list of proteins considered to be identified with high confidence. As previously discussed, the data acquired for several samples of each dataset were unsuitable for further analysis (see 4.2.1).

5.1.3.2 Identification of Proteins Changed in Abundance

A two stage analysis was carried out to identify proteins which are changed in abundance between pre-eclampsia and control samples (outlined in 5.1.4.1 and 5.1.4.2). Data were first filtered to identify proteins which were quantified reproducibly between samples processed in duplicate. Duplicate samples were defined as technical replicates (superpool, late onset pre-eclampsia and early onset pre-eclampsia) and control samples (control pool 1 and 2). Although not truly equivalent, the two control pools were considered duplicates for this stage of the analysis. Reproducibly quantified proteins were then assessed to identify proteins which changed in abundance between pre-eclampsia and control samples. Development of this analysis method is described as follows (see 5.1.4.1 and 5.1.4.2). The method was applied to early onset pre-eclampsia samples across all three datasets and late onset pre-eclampsia samples in the MARS 14-5800 dataset.
Table 5.1.1 Stage one of the data analysis method - identification of proteins with reproducible quantification between duplicate samples.

The iTRAQ relative quantification ratio for one duplicate was divided by the iTRAQ relative quantification ratio for the second duplicate sample and the value transformed \((\log_2)\). The mean and 2SD of the transformed values for each set of duplicate samples was calculated. Based on these values for technical replicates, limits for reproducible quantification were defined. Proteins where the transformed ratio for duplicate iTRAQ relative quantification ratios fell between -0.35 and 0.35 were considered reproducibly quantified. These limits were also applied to the duplicated control samples.

<table>
<thead>
<tr>
<th>Stage One: Reproducible Quantification</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.35&gt;Log2(Duplicate iTRAQ Ratio/Duplicate iTRAQ Ratio)&lt;0.35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage One: Reproducible Quantification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculate average of duplicate ratios</td>
</tr>
</tbody>
</table>

| Figure 5.1.1 Stage one of the data analysis method - identification of proteins with reproducible quantification between duplicate samples. The iTRAQ relative quantification ratio for one duplicate was divided by the iTRAQ relative quantification ratio for the second duplicate sample and the value transformed \((\log_2)\). The mean and 2SD of the transformed values for each set of duplicate samples was calculated. Based on these values for technical replicates, limits for reproducible quantification were defined. Proteins where the transformed ratio for duplicate iTRAQ relative quantification ratios fell between -0.35 and 0.35 were considered reproducibly quantified. These limits were also applied to the duplicated control samples. |
Figure 5.1.2 Stage two of the data analysis method identified proteins which changed in abundance between pre-eclampsia and control samples greater than between the two control samples. The iTRAQ relative quantification ratios calculated by ProteinPilot ratios were transformed (Log2). The difference in the value between the two control samples was used to define the range of biological variance of plasma proteins in normal uncomplicated pregnancies. An average of the values for the duplicated case and control samples was calculated. The difference between the average values for case and control samples was calculated and a protein defined as changed if this value fell outside the defined range of biological variation.

<table>
<thead>
<tr>
<th>iTraq Ratio</th>
<th>Average</th>
<th>Average</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgY 14-SuperMix - QSTAR</td>
<td></td>
<td>Early PE Case</td>
<td>Control</td>
</tr>
<tr>
<td>IgY-14-SuperMix - S800</td>
<td></td>
<td>Early PE Case</td>
<td>Control</td>
</tr>
<tr>
<td>MARS 14 – S800</td>
<td>Late PE Case</td>
<td>Early PE Case</td>
<td></td>
</tr>
</tbody>
</table>

**Stage Two:**
**Proteins Changed in Abundance**

-0.3 < (average pre-eclampsia – average control) > 0.3

**Candidate Biomarkers**
5.1.4 Results

A two stage analysis method was developed using iTRAQ relative quantification ratios for the proteins which were identified in the IgY 14-SuperMix QSTAR, IgY 14-SuperMix-5800 and MARS 14-5800 datasets (Figure 4.3.3). These proteins were indentified in multiple datasets; identification and quantification is therefore of higher confidence than for proteins identified in just one dataset. These proteins were thought to be most useful in developing a stringent method of analysis. A total of 113 proteins were identified in all three datasets, 99 of which had complete quantification data.

5.1.4.1 Data Analysis Stage One - Proteins with Reproducible Quantification

To identify reproducibly quantified proteins, iTRAQ relative quantification ratios were compared between duplicate samples within each dataset. For each of the 99 proteins, the iTRAQ relative quantification ratio was divided by the iTRAQ relative quantification ratio for the duplicate sample and the value transformed (Log₂). If the relative quantification of the protein is similar between duplicate samples, the transformed value should approximate zero. Duplicated samples in each dataset were assessed previously (see 4.2.1); early onset pre-eclampsia and control duplicate samples were considered for the IgY 14-SuperMix QSTAR and 5800 datasets. Late and early onset pre-eclampsia duplicate samples were considered for the MARS 14-5800 dataset; the superpool relative quantification ratio was also considered a duplicate sample in this dataset.

The mean and 2SD of the transformed values for each set of duplicate samples were calculated (Figure 5.1.1); mean values all approximated to zero as expected (Table 5.1.1). Based on the SD values for technical replicates, limits for reproducible quantification were defined. Proteins where the transformed ratio for duplicate iTRAQ relative quantification ratios fell between -0.35 and 0.35 were considered reproducibly quantified and all other proteins were removed. These limits were also applied to the duplicated control samples.

The 99 proteins used to define this range were plotted in alphabetical order in Figure 5.1.3. The transformed iTRAQ relative quantification ratios for each of the duplicate samples were plotted to demonstrate reproducible quantification. Proteins which are quantified reproducibly in each duplicate sample show fall along the diagonal line (solid line) indicating a 1:1 relationship. The limits for reproducible quantification are indicated by the dashed line (-0.35 and 0.35). Proteins must be reproducibly quantified for all duplicated samples in all datasets. A total of 57 protein identifications were considered reproducibly quantified, the remainder were excluded.
Table 5.1.1  Relative quantification ratios were compared for proteins in duplicate samples to identify proteins with reproducible quantification. For each of the 99 proteins identified in the IgY 14-SuperMix QSTAR, IgY 14-SuperMix 5800 and MARS 14-5800 datasets, the iTRAQ relative quantification ratio was divided by the iTRAQ relative quantification ratio for the duplicate sample and the value transformed (Log₂). The mean and 2SD of the transformed ratios were calculated. Based on these values for technical replicates, limits for reproducible quantification were defined. These limits were also applied to the values for the control duplicate samples. Transformed paired observations falling between -0.35 and 0.35 were considered reproducibly quantified.

<table>
<thead>
<tr>
<th>Paired Samples</th>
<th>Dataset</th>
<th>-2SD</th>
<th>Mean</th>
<th>+2SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superpool</td>
<td>MARS 14-5800</td>
<td>-0.27</td>
<td>0.03</td>
<td>0.33</td>
</tr>
<tr>
<td>Late onset pre-eclampsia</td>
<td>MARS 14-5800</td>
<td>-0.44</td>
<td>-0.03</td>
<td>0.38</td>
</tr>
<tr>
<td>Early onset pre-eclampsia</td>
<td>IgY 14-SuperMix-QSTAR</td>
<td>-0.38</td>
<td>-0.02</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>IgY 14-SuperMix-5800</td>
<td>-0.53</td>
<td>0.00</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>MARS 14-5800</td>
<td>-0.26</td>
<td>-0.01</td>
<td>0.24</td>
</tr>
<tr>
<td>Limits of Reproducible Quantification</td>
<td></td>
<td>-0.35</td>
<td></td>
<td>0.35</td>
</tr>
<tr>
<td>Control</td>
<td>IgY 14-SuperMix-QSTAR</td>
<td>-0.59</td>
<td>-0.07</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>IgY 14-SuperMix-5800</td>
<td>-0.62</td>
<td>-0.08</td>
<td>0.47</td>
</tr>
</tbody>
</table>
Figure 5.1.3  Transformed relative quantification ratios for proteins identified in duplicate samples were plotted and considered reproducibly quantified when between -0.35 and 0.35. A total of 99 proteins were identified in all three datasets with complete quantification data. Early onset pre-eclampsia and control duplicate samples were considered for the IgY 14-SuperMix QSTAR and 5800 datasets. Late and early onset pre-eclampsia duplicate samples were considered for the MARS 14-5800 dataset. The superpool relative quantification ratio was also considered a duplicate in this dataset; the iTRAQ relative quantification ratio was transformed. The limits of reproducible quantification (-0.35 and 0.35) were defined using technical replicates and are indicated by the dotted line. A total of 57 protein identifications were considered reproducibly quantified in duplicate samples in all three datasets.
5.1.4.2 Data Analysis Stage Two - Proteins Changed in Abundance

The relative quantification ratios calculated by ProteinPilot were then assessed to identify proteins that changed in abundance in pre-eclampsia samples compared to control samples.

The control samples were used to define the range of biological variation of plasma proteins in normal uncomplicated pregnancies. The iTRAQ relative quantification ratios calculated by ProteinPilot were transformed (Log2). The difference between the transformed iTRAQ relative quantification ratios for each of the control samples was calculated. This value was calculated for the 57 proteins with reproducible quantification. As one control ratio was unavailable in the MARS 14-5800 dataset, this range could only be calculated for the IgY 14-SuperMix-QSTAR and IgY 14-SuperMix-5800 datasets. The mean and 2SD for the difference between transformed iTRAQ relative quantification ratios for control samples was calculated (Table 5.1.2). These values were used to define the range of biological variation, which was set at -0.3 to 0.3.

Table 5.1.2 Control samples were used to define the range of biological variation to identify proteins changed in abundance. Proteins identified in both IgY 14-SuperMix datasets with reproducible quantification were used to define the range of biological variation (57 proteins). The difference between transformed (Log2) relative quantification ratios for control samples was calculated. The mean and 2SD was calculated and used to define an inclusive biological range.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>-2SD</th>
<th>Mean</th>
<th>+2SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteins Reproducibly Quantified</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgY 14-SuperMix-QSTAR</td>
<td>-0.33</td>
<td>-0.08</td>
<td>0.16</td>
</tr>
<tr>
<td>IgY 14-SuperMix-5800</td>
<td>-0.32</td>
<td>-0.03</td>
<td>0.26</td>
</tr>
<tr>
<td>Biological Range</td>
<td>-0.3</td>
<td></td>
<td>0.3</td>
</tr>
</tbody>
</table>
The difference between pre-eclampsia and control iTRAQ relative quantification ratios for the 57 reproducibly quantified proteins were then calculated to identify changes which fell outside this range. There are two pre-eclampsia and two control iTRAQ relative quantification ratios for each protein identification. The consistency of quantification was assessed in stage one of the method (See 5.1.4.1) and so it was concluded that an average of these two relative quantification ratios would be representative. An average value for the transformed relative quantification ratios for pre-eclampsia and control samples was taken. The difference between the average pre-eclampsia and average control ratios was then calculated. Where this difference fell outside the range of biological variation (-0.3 to 0.3) the protein was considered changed in abundance in pre-eclampsia.

The 57 reproducibly quantified proteins are plotted (in alphabetical order) in Figure 5.1.4; the range of biological variation (-0.3 to 0.3) is indicated by the dotted lines. As previous analysis suggests the datasets are comparable (See 4.2.1), the range of biological variance calculated for the IgY 14-SuperMix datasets was also applied to the MARS 14-5800 dataset. Proteins with differences between early onset pre-eclampsia and control samples falling outside this range were considered changed in early onset pre-eclampsia. A total of 26 proteins were changed outside the range of biological variation.

Relative quantification ratios for three example proteins are plotted in Figure 5.1.5. An average value for case and control relative quantification ratios has been calculated and plotted. Haptoglobin (Figure 5.1.5-A) was not reproducibly quantified and was excluded at this stage of the analysis. Tenascin-X and talin-1 (Figure 5.1.5- B and C) fell within the limits of reproducible quantification. The difference between the average values for Tenascin-X is small and does not fall outside the range of biological variation (Figure 5.1.5-B). The difference between the average case and control values for talin-1 is greater than 0.3 and so this protein is considered up-regulated in the case sample compared to the control sample (Figure 5.1.5-C).
Proteins where the difference between early onset pre-eclampsia and control samples fell outside the range of biological variation were considered changed. A two stage analysis method was applied to the 99 proteins with complete quantification data identified in all three datasets. A total of 57 of those proteins had reproducible quantification in all three datasets. The difference between the average transformed iTRAQ relative quantification ratios for early onset pre-eclampsia and controls was calculated and plotted (in alphabetical order) for those proteins.
Changes in plasma proteins in women who develop pre-eclampsia were identified using a two stage analysis. Data were acquired by LC-MS/MS analysis of pooled plasma samples from women who did and did not develop pre-eclampsia during pregnancy labelled with iTRAQ 8-plex reagent. Data were filtered to identify high confidence protein identifications. Relative quantification ratios were transformed (Log2) and plotted with confidence intervals calculated by ProteinPilot. A- Proteins with inconsistent quantification between duplicate samples were excluded from further analysis. B- An average value for case and control transformed relative quantification ratios were calculated for proteins with consistent replicates. Where the difference between these average values fell within the range of biological variance (-0.3 to 0.3) proteins were not considered changed. C- Proteins with a difference in average transformed relative quantification ratios outside the range of biological variance were considered changed. The difference between the transformed relative quantification ratios is greater than 0.3 and so this protein is up-regulated.
5.1.4.3 Proteins Changed in Abundance between Pre-eclampsia and Control

The two stage analysis method detailed above was applied to all data acquired. Data for early onset pre-eclampsia samples was available in all three datasets; data for late onset pre-eclampsia samples was only available in the MARS 14-5800 dataset.

5.1.4.3.1 Proteins Changed in Abundance in Early Onset Pre-eclampsia

Having demonstrated the ability of the analysis method to identify changed proteins, the analysis method was applied to the 502 proteins identified across all three datasets (Figure 4.3.3). A total of 113 proteins were identified as changed between early onset pre-eclampsia and control samples (Table 5.1.3). The proteins were distributed between the datasets as shown in Figure 5.1.6.
A total of 113 proteins were identified as changed between early onset pre-eclampsia and control samples in all data. Pooled plasma samples were immunodepleted and iTRAQ labelled. Data acquired following LC-MS/MS analysis was interrogated by ProteinPilot and filtered to remove low confidence identifications. A total of 502 proteins were considered identified with high confidence. These proteins were further filtered to identify those proteins which were processed reproducibly and measured reproducibly by LC-MS/MS. These proteins were then assessed for changes between early onset pre-eclampsia and control samples which fell outside the range of biological variation. A total of 43 proteins were identified as changed between early onset pre-eclampsia and control samples in the IgY 14-SuperMix-QSTAR dataset, 49 in the IgY 14-SuperMix-5800 dataset and 45 in the MARS 14-5800 dataset (shown in bracket. The number of proteins identified as changed between early onset pre-eclampsia and control samples and identified in more than one dataset are indicated.
A total of 113 proteins were identified as changed between early onset pre-eclampsia and control samples. The magnitude of the change and number of peptides for each proteins identified are reported. Proteins ‘identified’ in a dataset but not reproducibly quantified are indicated.

<table>
<thead>
<tr>
<th>Name</th>
<th>Magnitude of Fold Change</th>
<th>Peptides (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IgY 14-SuperMix-QSTAR</td>
<td>IgY 14-SuperMix-5800</td>
</tr>
<tr>
<td>A2M Alpha-2-macroglobulin</td>
<td>-0.292</td>
<td>-0.471</td>
</tr>
<tr>
<td>ACTB Beta actin variant (Fragment)</td>
<td>0.646</td>
<td>0.75</td>
</tr>
<tr>
<td>AFM Afamin</td>
<td>0.344</td>
<td>Identified</td>
</tr>
<tr>
<td>AGT Angiotensinogen</td>
<td>-0.126</td>
<td>-0.245</td>
</tr>
<tr>
<td>AHSG 29 kDa protein</td>
<td>0.67</td>
<td>0.75</td>
</tr>
<tr>
<td>APOA1 Apolipoprotein A-I</td>
<td>Identified</td>
<td>Identified</td>
</tr>
<tr>
<td>APOC4 Apolipoprotein C-IV</td>
<td>Identified</td>
<td>Identified</td>
</tr>
<tr>
<td>APOH Beta-2-glycoprotein 1</td>
<td>0.4</td>
<td>0.765</td>
</tr>
<tr>
<td>ATRN Isoform 1 of Attractin</td>
<td>Identified</td>
<td>-0.307</td>
</tr>
<tr>
<td>BCHE Cholinesterase precursor</td>
<td>0.067</td>
<td>0.055</td>
</tr>
<tr>
<td>BIN2 Isoform 2 of Bridging integrator 2</td>
<td>Identified</td>
<td>0.765</td>
</tr>
<tr>
<td>BTD cDNA FLJ51892, highly similar to Biotinidase</td>
<td>Identified</td>
<td>0.311</td>
</tr>
<tr>
<td>C1QA 23 kDa protein</td>
<td>-0.309</td>
<td>0.389</td>
</tr>
<tr>
<td>C1QA Complement C1q subcomponent subunit A</td>
<td>-0.389</td>
<td>-0.314</td>
</tr>
<tr>
<td>C1QB Complement component 1, q subcomponent, B chain</td>
<td>-0.545</td>
<td></td>
</tr>
<tr>
<td>C1QB complement component 1, q subcomponent, B chain precursor</td>
<td>-0.421</td>
<td>Identified</td>
</tr>
<tr>
<td>C1QC Complement C1q subcomponent subunit C</td>
<td>-0.379</td>
<td>-0.625</td>
</tr>
<tr>
<td>C1RL 48 kDa protein</td>
<td>-0.483</td>
<td>-0.252</td>
</tr>
<tr>
<td>Protein Name</td>
<td>Identified</td>
<td>Identified</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>C3 Complement C3 (Fragment)</td>
<td>Identified</td>
<td>Identified</td>
</tr>
<tr>
<td>C4B Complement C4-B</td>
<td>Identified</td>
<td>Identified</td>
</tr>
<tr>
<td>C8B Complement component C8 beta chain</td>
<td>-0.151</td>
<td>-0.311</td>
</tr>
<tr>
<td>C8G Complement component C8 gamma chain</td>
<td>Identified</td>
<td>Identified</td>
</tr>
<tr>
<td>CA1 Carbonic anhydrase 1</td>
<td>-0.586</td>
<td>-0.62</td>
</tr>
<tr>
<td>CAP1 Isoform 2 of Adenylyl cyclase-associated protein 1</td>
<td>0.951</td>
<td></td>
</tr>
<tr>
<td>CD44 Isoform 5 of CD44 antigen</td>
<td>-0.355</td>
<td></td>
</tr>
<tr>
<td>CD5L CD5 antigen-like</td>
<td>Identified</td>
<td>Identified</td>
</tr>
<tr>
<td>cDNA FLJ51034, highly similar to Vitamin K-dependent protein C</td>
<td>Identified</td>
<td>Identified</td>
</tr>
<tr>
<td>cDNA FLJ53558, highly similar to Protein disulfide-isomerase A3</td>
<td>0.576</td>
<td></td>
</tr>
<tr>
<td>cDNA FLJ53670, highly similar to EGF-containing fibulin-like extracellular matrix protein 1</td>
<td>0.353</td>
<td></td>
</tr>
<tr>
<td>cDNA FLJ55673, highly similar to Complement factor B</td>
<td>Identified</td>
<td>Identified</td>
</tr>
<tr>
<td>cDNA FLJ60461, highly similar to Peroxiredoxin-2</td>
<td>-0.33</td>
<td></td>
</tr>
<tr>
<td>CFD Complement factor D preproprotein</td>
<td>0.108</td>
<td>0.452</td>
</tr>
<tr>
<td>CGA Glycoprotein hormones alpha chain</td>
<td>0.3</td>
<td>0.559</td>
</tr>
<tr>
<td>CGB1 Isoform 1 of Choriogonadotropin subunit beta variant 1</td>
<td>0.647</td>
<td></td>
</tr>
<tr>
<td>CGB2 Isoform 1 of Choriogonadotropin subunit beta variant 2</td>
<td>0.397</td>
<td></td>
</tr>
<tr>
<td>CLIC1 Chloride intracellular channel protein 1</td>
<td>0.498</td>
<td></td>
</tr>
<tr>
<td>CLU clusterin isoform 1</td>
<td>0.072</td>
<td>0.113</td>
</tr>
<tr>
<td>COL6A1 Collagen alpha-1(VI) chain</td>
<td>0.657</td>
<td>0.111</td>
</tr>
<tr>
<td>CPN1 Carboxypeptidase N catalytic chain</td>
<td>0.057</td>
<td>-0.163</td>
</tr>
<tr>
<td>CRP Isoform 1 of C-reactive protein</td>
<td>Identified</td>
<td>Identified</td>
</tr>
<tr>
<td>CST3 Cystatin-C</td>
<td>Identified</td>
<td>Identified</td>
</tr>
<tr>
<td>CTBS Di-N-acetylchitobiase</td>
<td>-0.571</td>
<td></td>
</tr>
<tr>
<td>Gene Name</td>
<td>Identified</td>
<td>phenotypic value</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>------------</td>
<td>------------------</td>
</tr>
<tr>
<td>ENO1 Isoform alpha-enolase of Alpha-enolase</td>
<td>Identified</td>
<td>0.685</td>
</tr>
<tr>
<td>ENPP2 Isoform 3 of Ectonucleotide pyrophosphatase/phosphodiesterase family member 2</td>
<td>0.075</td>
<td>0.038</td>
</tr>
<tr>
<td>F10 Coagulation factor X</td>
<td>-0.025</td>
<td>-0.038</td>
</tr>
<tr>
<td>F13A1 coagulation factor XIII A1 subunit precursor</td>
<td>-0.128</td>
<td>-0.031</td>
</tr>
<tr>
<td>F13B Coagulation factor XIII B chain</td>
<td>-0.413</td>
<td>-0.645</td>
</tr>
<tr>
<td>F7 Factor VII active site mutant immunoconjugate</td>
<td>-0.013</td>
<td>0.57</td>
</tr>
<tr>
<td>F9 Coagulation factor IX</td>
<td>-0.056</td>
<td>-0.191</td>
</tr>
<tr>
<td>FAM20C family with sequence similarity 20, member C</td>
<td>-0.304</td>
<td>-0.336</td>
</tr>
<tr>
<td>FCGBP IgGFc-binding protein</td>
<td>-0.205</td>
<td>-0.404</td>
</tr>
<tr>
<td>FCGR3B Protein</td>
<td>0.332</td>
<td>Identified</td>
</tr>
<tr>
<td>FGG 50 kDa protein</td>
<td>Identified</td>
<td>0.624</td>
</tr>
<tr>
<td>FLNA Filamin A, alpha</td>
<td>Identified</td>
<td>0.685</td>
</tr>
<tr>
<td>GAPDH Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>Identified</td>
<td>0.65</td>
</tr>
<tr>
<td>GC vitamin D-binding protein precursor</td>
<td>Identified</td>
<td>-0.372</td>
</tr>
<tr>
<td>GSTP1 Glutathione S-transferase P</td>
<td>0.37</td>
<td>-0.356</td>
</tr>
<tr>
<td>H6PD GDH/6PGL endoplasmic bifunctional protein</td>
<td>-0.301</td>
<td>0.237</td>
</tr>
<tr>
<td>HABP2 Hyaluronan-binding protein 2</td>
<td>0.301</td>
<td>0.237</td>
</tr>
<tr>
<td>HGFAC Hepatocyte growth factor activator</td>
<td>-0.247</td>
<td>-0.302</td>
</tr>
<tr>
<td>HRG Histidine-rich glycoprotein</td>
<td>Identified</td>
<td>0.581</td>
</tr>
<tr>
<td>HSPA5 HSPA5 protein</td>
<td>0.055</td>
<td>0.081</td>
</tr>
<tr>
<td>IGF1 Insulin-like growth factor IB</td>
<td>0.558</td>
<td>0.241</td>
</tr>
<tr>
<td>IGFBP2 Insulin-like growth factor-binding protein 2</td>
<td>-0.256</td>
<td>-0.331</td>
</tr>
<tr>
<td>ITIH3 Isoform 2 of Inter-alpha-trypsin inhibitor heavy chain H3</td>
<td>0.312</td>
<td>Identified</td>
</tr>
<tr>
<td>ITIH4 ITIH4 protein</td>
<td>Identified</td>
<td>0.429</td>
</tr>
<tr>
<td>Gene Name</td>
<td>Mean</td>
<td>Score</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>KLKB1 Plasma kallikrein</td>
<td>-0.078</td>
<td>0.577</td>
</tr>
<tr>
<td>KNG1 44 kDa protein</td>
<td>0.389</td>
<td>103</td>
</tr>
<tr>
<td>KNG1 Isoform HMW of Kininogen-1</td>
<td>Identified</td>
<td>0.35</td>
</tr>
<tr>
<td>LOC100126583 Putative uncharacterized protein DKFZp686K04218 (Fragment)</td>
<td>-0.583</td>
<td>21</td>
</tr>
<tr>
<td>LPA Apolipoprotein(a)</td>
<td>-0.517</td>
<td>-0.459</td>
</tr>
<tr>
<td>LRG1 Leucine-rich alpha-2-glycoprotein</td>
<td>-0.251</td>
<td>-0.378</td>
</tr>
<tr>
<td>MBL2 Mannose-binding protein C</td>
<td>-0.391</td>
<td>Identified</td>
</tr>
<tr>
<td>NOTUM Protein notum homolog</td>
<td>-0.395</td>
<td>-0.458</td>
</tr>
<tr>
<td>PAEP Glycodelin</td>
<td>0.352</td>
<td>1</td>
</tr>
<tr>
<td>PGLYRP2 Isoform 1 of N-acetylmuramoyl-L-alanine amidase</td>
<td>Identified</td>
<td>-0.583</td>
</tr>
<tr>
<td>PKM2 Isoform M2 of Pyruvate kinase isozymes M1/M2</td>
<td>0.417</td>
<td>2</td>
</tr>
<tr>
<td>PLEK Pleckstrin</td>
<td>0.626</td>
<td>0.641</td>
</tr>
<tr>
<td>PLG Plasminogen</td>
<td>Identified</td>
<td>0.387</td>
</tr>
<tr>
<td><strong>PPBP Platelet basic protein</strong></td>
<td>0.563</td>
<td>0.376</td>
</tr>
<tr>
<td>PRSS2 Protease serine 2 isoform B</td>
<td>Identified</td>
<td>-0.739</td>
</tr>
<tr>
<td>PSG2 Pregnancy-specific beta-1-glycoprotein 2</td>
<td>0.42</td>
<td>7</td>
</tr>
<tr>
<td>PSG3 cDNA PLJ099068 4.0, clone PLACE100526, highly similar to Pregnancy-specific beta-1 glycoprotein 5</td>
<td>0.314</td>
<td>0.677</td>
</tr>
<tr>
<td>PSG5 Similar to Pregnancy-specific beta 1 glycoprotein</td>
<td>0.483</td>
<td>8</td>
</tr>
<tr>
<td><strong>PSG9 Pregnancy-specific beta-1-glycoprotein 9</strong></td>
<td>0.431</td>
<td>0.583</td>
</tr>
<tr>
<td>PTPRD 215 kDa protein</td>
<td>-0.392</td>
<td>3</td>
</tr>
<tr>
<td>PZP Isoform 1 of Pregnancy zone protein</td>
<td>Identified</td>
<td>0.409</td>
</tr>
<tr>
<td>RBP4 Retinol-binding protein 4</td>
<td>0.804</td>
<td>Identified</td>
</tr>
<tr>
<td>SELL Selectin L (Lymphocyte adhesion molecule 1), isoform CRA_b</td>
<td>-0.269</td>
<td>-0.424</td>
</tr>
<tr>
<td>SEPP1 selenoprotein P isoform 2</td>
<td>0.209</td>
<td>0.314</td>
</tr>
<tr>
<td>Protein Name</td>
<td>Description</td>
<td>Rank 1</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>--------</td>
</tr>
<tr>
<td>SERPINA10</td>
<td>Protein Z-dependent protease inhibitor</td>
<td>-0.33</td>
</tr>
<tr>
<td>SERPINA3</td>
<td>Isoform 1 of Alpha-1-antichymotrypsin</td>
<td>-0.361</td>
</tr>
<tr>
<td>SERPINA4</td>
<td>Kallistatin</td>
<td>-0.182</td>
</tr>
<tr>
<td>SERPINA5</td>
<td>Plasma serine protease inhibitor</td>
<td>Identified</td>
</tr>
<tr>
<td>SERPINA6</td>
<td>Corticosteroid-binding globulin</td>
<td>-0.216</td>
</tr>
<tr>
<td>SERPINA7</td>
<td>Thyroxine-binding globulin</td>
<td>0.019</td>
</tr>
<tr>
<td>SERPIN D1</td>
<td>Serpin peptidase inhibitor, clade D (Heparin cofactor), member 1</td>
<td>0.033</td>
</tr>
<tr>
<td>SH3BGL3</td>
<td>Putative uncharacterized protein</td>
<td>0.01</td>
</tr>
<tr>
<td>SOD3</td>
<td>Extracellular superoxide dismutase [Cu-Zn]</td>
<td>0.418</td>
</tr>
<tr>
<td>SPARC</td>
<td>SPARC</td>
<td></td>
</tr>
<tr>
<td>SPP2</td>
<td>Secreted phosphoprotein 24</td>
<td>-0.499</td>
</tr>
<tr>
<td>TAGLN2</td>
<td>24 kDa protein</td>
<td>0.524</td>
</tr>
<tr>
<td>THBS1</td>
<td>Thrombospondin-1</td>
<td>0.46</td>
</tr>
<tr>
<td>TLN1</td>
<td>Talin-1</td>
<td>0.505</td>
</tr>
<tr>
<td>TMSB4X</td>
<td>TMSB4X protein (Fragment)</td>
<td>Identified</td>
</tr>
<tr>
<td>TPI1</td>
<td>triosephosphate isomerase 1 isoform 2</td>
<td>0.384</td>
</tr>
<tr>
<td>TPM4</td>
<td>Isoform 2 of Tropomyosin alpha-4 chain</td>
<td>0.557</td>
</tr>
<tr>
<td>V(k)3</td>
<td>sequence of NG9 gene from fetal liver DNA</td>
<td></td>
</tr>
<tr>
<td>VASP</td>
<td>42 kDa protein</td>
<td></td>
</tr>
<tr>
<td>VCL</td>
<td>Isoform 2 of Vinculin</td>
<td>0.713</td>
</tr>
<tr>
<td>YWHAZ</td>
<td>14-3-3 protein zeta/delta</td>
<td>0.411</td>
</tr>
<tr>
<td>ZYX</td>
<td>67 kDa protein</td>
<td>0.486</td>
</tr>
<tr>
<td>ZYX</td>
<td>Zyxin</td>
<td></td>
</tr>
</tbody>
</table>
5.1.4.3.2 Proteins Changed in Abundance in Late Onset Pre-eclampsia

The two stage analysis method was also applied to proteins identified in late onset pre-eclampsia samples in the MARS 14-5800 dataset. Late onset pre-eclampsia samples were only available in this dataset due to incomplete immunodepletion using the IgY 14-SuperMix system. A total of 133 of 164 proteins with complete quantification data were quantified reproducibly. Only one control sample was available for analysis in the MARS 14-5800 dataset. The average late onset pre-eclampsia transformed iTRAQ relative quantification ratios were compared to the single transformed relative quantification ratio for the control sample. The range of biological variation defined in section 5.1.4.2 (calculated using control samples in the IgY 14-SuperMix datasets) was also applied here. The difference between the average transformed relative quantification ratios for duplicate late onset pre-eclampsia samples and the transformed relative quantification ratio control was calculated. The differences for the 133 proteins considered reproducibly quantified are plotted in alphabetical order in Figure 5.1.7; the range of biological variation (-0.3 to 0.3) is indicated by the dotted lines. Proteins with differences between late onset pre-eclampsia and control samples falling outside this range are considered changed in late onset pre-eclampsia. A total of 32 proteins were identified as changed in late onset pre-eclampsia (Table 5.1.4).
Proteins where the difference between late onset pre-eclampsia and the control sample iTRAQ relative quantification ratios was outside the range of biological variation were considered changed. The two stage analysis method was applied to the 164 proteins with complete quantification data identified in the MARS 14-5800 dataset. A total of 133 of those proteins had reproducible quantification. The difference between the average transformed relative quantification ratios for late and early onset pre-eclampsia and the transformed relative quantification ratio for the control sample was calculated. The differences were plotted in alphabetical order for those proteins. The difference between transformed relative quantification ratios for control samples in the IgY 14-SuperMix-QSTAR and 5800 datasets was used to define the biological range (dashed line) of protein expression in uncomplicated pregnancies.
Table 5.1.4  
A total of 32 proteins were identified as changed between late onset pre-eclampsia and the control sample. The magnitude of the change and number of peptides for each protein identified are reported.

<table>
<thead>
<tr>
<th>Name</th>
<th>Change in Abundance</th>
<th>Peptides (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGT Angiotensinogen</td>
<td>-0.346</td>
<td>47</td>
</tr>
<tr>
<td>ALB Putative uncharacterized protein ALB</td>
<td>0.347</td>
<td>39</td>
</tr>
<tr>
<td>APOA1 Apolipoprotein A-I</td>
<td>-0.305</td>
<td>12</td>
</tr>
<tr>
<td>BCHE Cholinesterase precursor</td>
<td>-0.328</td>
<td>5</td>
</tr>
<tr>
<td>BTD cDNA FLJ51892, highly similar to Biotinidase</td>
<td>0.335</td>
<td>11</td>
</tr>
<tr>
<td>C3 Complement C3 (Fragment)</td>
<td>0.466</td>
<td>46</td>
</tr>
<tr>
<td>C4B Complement C4-B</td>
<td>-0.979</td>
<td>349</td>
</tr>
<tr>
<td>C4BPA C4b-binding protein alpha chain</td>
<td>-0.397</td>
<td>17</td>
</tr>
<tr>
<td>C8G Complement component C8 gamma chain</td>
<td>-0.447</td>
<td>17</td>
</tr>
<tr>
<td>CA1 Carbonic anhydrase 1</td>
<td>-0.384</td>
<td>3</td>
</tr>
<tr>
<td>cDNA FLJ51034, highly similar to Vitamin K-dependent protein C</td>
<td>-0.465</td>
<td>6</td>
</tr>
<tr>
<td>CP 20 kDa protein</td>
<td>0.366</td>
<td>51</td>
</tr>
<tr>
<td>CPN1 Carboxypeptidase N catalytic chain</td>
<td>0.659</td>
<td>11</td>
</tr>
<tr>
<td>CST3 Cystatin-C</td>
<td>-0.635</td>
<td>4</td>
</tr>
<tr>
<td>ECM1 Isoform 4 of Extracellular matrix protein 1</td>
<td>-0.333</td>
<td>22</td>
</tr>
<tr>
<td>Gene Name</td>
<td>Value</td>
<td>Rank</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>ENPP2 Isoform 3 of Ectonucleotide pyrophosphatase/phosphodiesterase family member 2</td>
<td>0.316</td>
<td>12</td>
</tr>
<tr>
<td>F10 Coagulation factor X</td>
<td>-0.358</td>
<td>9</td>
</tr>
<tr>
<td>F9 Coagulation factor IX</td>
<td>0.556</td>
<td>7</td>
</tr>
<tr>
<td>FN1 Isoform 9 of Fibronectin</td>
<td>0.379</td>
<td>77</td>
</tr>
<tr>
<td>HRG Histidine-rich glycoprotein</td>
<td>0.447</td>
<td>45</td>
</tr>
<tr>
<td>HSPA5 HSPA5 protein</td>
<td>0.904</td>
<td>5</td>
</tr>
<tr>
<td>PGLYRP2 Isoform 1 of N-acetylmuramoyl-L-alanine amidase</td>
<td>-0.505</td>
<td>27</td>
</tr>
<tr>
<td>PLEK Pleckstrin</td>
<td>0.581</td>
<td>3</td>
</tr>
<tr>
<td>PLG Plasminogen</td>
<td>0.332</td>
<td>104</td>
</tr>
<tr>
<td>PON1 Serum paraoxonase/arylesterase 1</td>
<td>-0.334</td>
<td>7</td>
</tr>
<tr>
<td>PPBP Platelet basic protein</td>
<td>0.708</td>
<td>4</td>
</tr>
<tr>
<td>PRSS2 Protease serine 2 isoform B</td>
<td>-0.700</td>
<td>3</td>
</tr>
<tr>
<td>PSG4 cDNA FLJ90668 fis, clone PLACE1005426, highly similar to Pregnancy-specific beta-1-glycoprotein 5</td>
<td>0.477</td>
<td>9</td>
</tr>
<tr>
<td>PZP Isoform 1 of Pregnancy zone protein</td>
<td>0.497</td>
<td>73</td>
</tr>
<tr>
<td>SERPINA3 cDNA FLJ35730 fis, clone TESTI2003131, highly similar to ALPHA-1-ANTICYTOMOTRYPSIN</td>
<td>-0.349</td>
<td>64</td>
</tr>
<tr>
<td>SERPINA6 Corticosteroid-binding globulin</td>
<td>-0.386</td>
<td>9</td>
</tr>
<tr>
<td>SERPING1 Plasma protease C1 inhibitor</td>
<td>-0.302</td>
<td>30</td>
</tr>
</tbody>
</table>
5.1.4.3.3 Proteins Changed in Abundance in Late and Early Onset Pre-eclampsia

The data for proteins changed in late and early onset pre-eclampsia in the MARS 14-5800 dataset was compared. A total of 45 proteins were identified as changed outside the range of biological variation between early onset pre-eclampsia and the control sample in the MARS 14-5800 dataset. Fewer proteins (32) were identified as changed in abundance between late onset pre-eclampsia and the control sample. A total of 24 proteins were changed in abundance in both late and early onset pre-eclampsia at 15 weeks gestation (Figure 5.1.8).

**Figure 5.1.8** A total of 24 proteins were identified as changed in abundance in both late and early onset pre-eclampsia in the MARS 14-5800 dataset. A total of 189 proteins were considered identified with high confidence in the MARS 14-5800 dataset. A two stage method of analysis was applied to the proteins to identify proteins which were reproducibly quantified and changed between pre-eclampsia and a control sample outside the range of biological variation. A total of 45 proteins were identified as changed in abundance between early onset pre-eclampsia samples and the control sample and 32 between late onset pre-eclampsia and the control sample. Of those proteins, 24 were identified as changing in both late and early onset pre-eclampsia. Proteins in red were up-regulated, proteins in blue, down-regulated.
5.1.4.4 Validation of Proteins Changed in Abundance

Further validation of the changes in protein abundance identified here is required. Validation of all proteins is not practical and so a smaller number of proteins were chosen for further investigation. Proteins selected for validation must be identified in at least two datasets and by 3 or more peptides in each dataset. Proteins involved in the complement cascade or immune response and those which should have been immunodepleted (or with homology to proteins which should have been immunodepleted) were also excluded. A total of five proteins were selected for validation using these criteria including: filamin, platelet basic protein, pregnancy-specific beta-1-glycoprotein 9 (PSG9), talin and vinculin. Fibulin was also included for further investigation as it was identified in two datasets with 7 and 16 peptides and fell just outside the range of biological variation (-0.246 and -0.290). Proteins were distributed across the datasets as shown in Figure 5.1.9. All proteins were increased in early onset pre-eclampsia compared to the control, with the exception of Fibulin, which decreased (Figure 5.1.10).

None of the proteins identified as changed in abundance in late onset pre-eclampsia were selected for validation as the data were not available at the time of initial validation. Platelet basic protein, however, was subsequently identified as increased in late onset pre-eclampsia compared to controls.
Proteins changed in abundance between early onset pre-eclampsia and control for validation. Data were acquired by LC-MS/MS analysis of pooled plasma samples from women who developed early onset pre-eclampsia and healthy controls labelled with iTRAQ 8-plex reagent. Data were filtered to identify high confidence protein identifications with reproducible quantification data which were changed in abundance between early onset pre-eclampsia and control samples. Proteins of interest were identified in all three datasets. The area of each segment represents the total number of proteins identified in each dataset as noted in brackets; the second figure is the number of proteins identified as changed between early onset pre-eclampsia and control samples in each dataset.
Figure 5.1.10 Filamin, Pregnancy-specific beta-1-glycoprotein 9, Talin and Vinculin were all increased in early onset pre-eclampsia compared to controls, Fibulin was decreased. Transformed (Log2) relative quantification ratios for late and early onset pre-eclampsia and control samples calculated against the reference superpool (labelled with iTRAQ reporter ion 113) for proteins of interest. Relative quantification ratios were plotted for each dataset available. The size of the data points are proportional to the number of peptides identified for each protein in each dataset, which is also noted in the legend.
5.1.5 Discussion

Although many studies have investigated changes in plasma proteins prior to diagnosis of pre-eclampsia, a clinically useful biomarker(s) for a low risk nulliparous population has yet to be identified. The aim of this study was to identify novel candidate biomarkers using iTRAQ.

5.1.5.1 Plasma Samples

Plasma samples used for this investigation were collected at 15 weeks gestation using a standardised procedure as part of the SCOPE Study. Standardised sample collection is essential to minimise pre-analytic variance and ensure samples are fit for purpose. The conditions under which the samples are collected can greatly impact upon the investigative work carried out and therefore the experimental conclusions drawn from that investigation. Standardisation of sample collection and primary processing is vitally important for proteomic investigations due to the variable nature of the proteome. The proteome sampled at the time of collection can be influenced by multiple factors including the posture of the patient, application of the tourniquet (Statland et al. 1974), the type of sample collection tube used, the length of time taken to process the sample (Yi et al. 2008; Randall et al. 2010) and storage conditions (Insenser et al. 2010; Jackson and Banks 2010). Samples collected as part of the SCOPE study were well curated; standard operating procedures and stringent control and tracking mechanisms were put in place to ensure samples were of the highest quality (1.1.7.3).

A total of 12 plasma samples were acquired for each case and control group. The number of samples required in investigative quantitative studies is more usually dictated by the power required for the study. For a quantitative study to be successful it must be sufficiently powered to produce a significant result (Kapp et al. 2005). Power is, however, not applicable to pooled samples as used in this investigation and so the number of samples included was arbitrarily set at 12. This number was thought to incorporate sufficient biological variability for an initial hypothesis generating investigation while still being practical for sample processing and data analysis.

Plasma was pooled according to phenotype into late onset pre-eclampsia, early onset pre-eclampsia and two control groups. A further reference superpool containing equal amounts of all samples was also produced and used as the denominator against which all iTRAQ ratios were calculated. The reference superpool was required as a homologous representative sample which should contain all proteins from all samples. Had a control sample been used as the denominator, a protein appearing in a case sample but not in a control sample would not have been quantified. Pooling of samples is a controversial issue; while common in microarray investigations, (Kendziorski et al. 2005), opinion is divided on the use and validity of pooled samples in the proteomics community (Karp and Lilley 2007;
Song et al. 2008). Pooled samples suffer from biological averaging; one contributing sample with a particularly high or low concentration of a protein will skew measurement of that protein in the pool. As a result, quantification is biased to this sample and the value may not be representative of the outlying sample or the sample population on the whole. Dilution of low abundance proteins past the point of observation is also a risk (Song et al. 2008). Conversely, pooling of samples has been suggested to be advantageous in experiments where small amounts or numbers of samples are available, the number of replicates is limited by time or cost, biological variance is high, or where the aim of the investigations is to characterise a population rather than an individual (Karp and Lilley 2007). To analyse multiple samples on an individual basis using iTRAQ would require multiple experiments; pooling samples allows direct comparison of sample populations without loss of data integrity though multiple normalisations (DeSouza et al. 2007). As previously discussed this study aimed to characterise protein expression at a global level. Pooled samples take into account the range of biological variability. Pooling of samples is therefore ideal for ‘first pass’ exploratory studies constrained by practicalities of time and cost such as this. Samples can subsequently be analysed on an individual basis for proteins of interest in future verification steps.

5.1.5.2 Data Analysis

There is no standardisation of experimental design, methodology or data analysis within proteomics (Taylor et al. 2007; Tonack et al. 2009; Burkhart et al. 2011). Experimental design is often dictated by sample type, availability and the resources accessible. The data analysis method can also greatly affect the data output; different analysis software uses different search algorithms (Kapp et al. 2005) and can be applied to multiple databases which ultimately result in identification of differing proteins (see 4.1 and 4.1.1). Different analysis software packages employ different algorithms for analysis of iTRAQ reporter ions and calculation of relative quantification ratios (Kapp et al. 2005; Hill et al. 2008; Nesvizhskii 2010; Rodríguez-Suárez et al. 2010). The analysis software used, however, is often dictated by the manufacturer of the LC-MS/MS and the data output format favoured by that vendor (Rodríguez-Suárez et al. 2010). Efforts are being made to standardise reporting of proteomics experiments with development of the Minimum Information About A Proteomics Experiment (MIAPE) guidelines for collection, integration, storage and dissemination of proteomics data (Taylor et al. 2007). Data analysis however remains highly individualised and must be tailored to each experiment.

To identify proteins changed in abundance in this investigation, a two-step filtering method was developed. Firstly, reproducibly quantified proteins were identified and secondly, the difference in abundance between case and control samples was calculated. In order to be confident in a change in protein abundance, the relative quantification ratios must be reproducible between duplicate samples, indicating that the protein has been
reproducibly processed, analysed and quantified. Relative quantification ratios were compared between duplicated samples; a reproducibly processed and quantified protein should have a replicate ratio of, or near, zero when transformed. Proteins with a transformed ratio for duplicate samples falling outside the range calculated using the mean and standard deviation of all technical replicates, were considered to be irreproducible and were excluded. The limits for reproducible quantification were calculated using transformed ratios for technical replicates and applied to all duplicated samples including control samples. As expected the values for the two control samples were larger than for technical replicates as the values also incorporate the range of biological variation as well as the variation introduced during sample processing. The range of protein abundance in normal healthy women must also be considered when identifying a biomarker as highly variable proteins will be of limited use in a clinical setting. The limits of reproducible quantification were therefore set to be exclusive rather than inclusive, particularly for control samples. For the IgY 14-SuperMix processed samples, two duplicate samples were available (early onset pre-eclampsia); for the MARS 14-5800 dataset three duplicate sample were available for analysis (superpool, late onset pre-eclampsia and early onset pre-eclampsia). Reproducibility of proteins identified in the MARS 14-5800 dataset could be considered to be of greater confidence than those identified by the IgY 14-SuperMix datasets.

Relative quantification ratios were also collated for proteins identified in multiple datasets and several proteins were demonstrated to be reproducibly quantified in all duplicated samples in all datasets. In order to calculate the change in protein abundance between case and control samples an average of the transformed relative quantification ratios for the two duplicate samples was taken. The difference between the average values for the case and control sample was then calculated. As replicates had already been established to be reproducible, the average value should be representative of the relative ratios of both replicates. To define the range of biological variation, the difference between transformed relative ratios for control samples was first calculated for proteins identified in all three datasets. Proteins which changed outside this range should therefore be due to disease state rather than biological variance. Due to the missing control relative quantitative ratio in the MARS 14-5800 dataset the biological range could only be calculated for the IgY 14-SuperMix datasets. As previous examination showed a high degree of similarity throughout the datasets, application of this range to the MARS 14-5800 dataset was justified.

The relative quantification ratios calculated here may not represent the actual change in abundance of the proteins identified in the women sampled. As previously discussed (see 4.1.1), there is a documented discrepancy between the observed and expected iTRAQ relative quantification ratios which increases with relative quantification ratio and is dependent on the LC-MS/MS platform used (Karp et al. 2010). Equal numbers of proteins were identified as changed from the IgY 14-SuperMix QSTAR and 5800 datasets and there was no observable discrepancy between the biological range for each dataset. This
suggests the influence of this discrepancy is negligible in this case. The relative quantification ratios were also calculated against the reference superpool which represents protein abundance across all sample groups rather one phenotype. This compresses the apparent change in protein abundance due to the biological averaging of the superpool, however it avoids having to integrate and compare multiple case to control ratios calculated against two non-equivalent control groups.

This data processing method is based on the statistical identification of outlying values and is therefore in danger of simply identifying those proteins with outlying relative quantification ratios rather than those which change biologically. The method of identifying changed proteins was carefully designed to incorporate and minimise variance introduced by sample processing and the range of biological variation between sample pools. When changes in abundance for proteins were plotted, proteins identified as changed were shown to fall outside the range of biological variation. Filtering was designed to be inclusive rather than exclusive, ensuring the identification of all potential candidate biomarkers and allowing further manual assessment where required.

5.1.5.3 Proteins Changed in Abundance in Early Onset Pre-eclampsia

Proteins identified as changed in abundance included Choriogonadotropin; CGB1 Isoform 1 of Choriogonadotropin subunit beta variant 1 and variant 2 in the IgY 14-SuperMix-5800 and QSTAR datasets respectively. This protein has been shown to be increased at 15 weeks gestation in women who subsequently developed pre-eclampsia (Muller et al. 1996; Audibert et al. 2005) and therefore supports the validity and reliability of the investigation. The data analysis method used identified 113 proteins which changed in abundance between early onset pre-eclampsia and control samples across all three datasets. The IgY 14-SuperMix System depleted approximately 100 high and medium abundance proteins resulting in the identification of lower abundance proteins in comparison to the MARS 14 column which removes only the top 14 high abundance proteins. Low abundance proteins are thought to be more reflective of health status and so changed proteins identified by the IgY 14-SuperMix datasets may be more clinically relevant than proteins identified in the MARS 14-5800 dataset. Late onset pre-eclampsia samples could only be analysed in one dataset. More proteins were identified as changed in early onset pre-eclampsia than in late onset pre-eclampsia when compared to the control sample. The majority of proteins identified as changed in late onset pre-eclampsia were also identified as changed in early onset pre-eclampsia. This suggests it may be possible to differentiate between women who will develop early and late onset pre-eclampsia at 15 weeks of pregnancy. This difference in the number of proteins affected may reflect the severity and rapid progression of the early form of pre-eclampsia. None of the proteins differing in abundance in late onset pre-eclampsia were validated; this dataset however is of great interest for future work.
Proteins for validation were selected from the larger list using several criteria. Proteins must pass limits set in place for reproducible quantification between duplicate samples and the change between case and control samples must fall outside the range of biological variation defined. As previously discussed this ensures proteins can be reproducibly quantified and change outside the normal range of biological variation. Proteins must be identified as changed in abundance in two or more datasets; multiple observations of a protein changing in abundance increase the confidence in that change being real. The protein must also be identified by 3 or more peptides in each dataset. Datasets produced by the 5800 required filtering to produce a list of proteins considered identified with high confidence. For the MARS 14-5800 dataset proteins with a UPS less than 1.3 and with fewer than 3 peptides were removed. A lower limit of 3 peptides was therefore set for proteins for validation to reflect the necessity of filtering across in datasets. An exception was made in the case of fibulin which was identified in two datasets and fell just outside the range of biological variation. No other reproducibly quantified proteins were identified within this range and so a single exception was made. While stringent data analysis methods were developed and applied, the dataset was also investigated manually at all stages to ensure the limits applied were reflective of the dataset on the whole. Proteins involved in the complement cascade or immune response were not considered as changes in these proteins are unlikely to be unique to pre-eclampsia alone. Proteins which should have been immunodepleted or with homology to those which were immunodepleted were also not considered as the change in abundance may be an artefact of the depletion rather than disease state. With the inclusion of fibulin, six proteins remained for validation.

5.1.6 Conclusion

The aim of this study was to use iTRAQ to identify plasma proteins which change in abundance at 15 weeks gestation in women who subsequently develop pre-eclampsia. The experiment was designed with technical replicates allowing assessment of sample processing variation and with duplicated control samples allowing comparison of the range of biological variation. A data analysis method was developed to identify proteins which were reproducibly quantified and changed in abundance outside the range of biological variation. This investigation has identified several proteins which potentially change in abundance at 15 weeks gestation in pooled plasma of women who subsequently developed both early and late onset pre-eclampsia compared to women who did not develop pre-eclampsia. A shortlist of proteins which change in early onset pre-eclampsia were selected for validation. Proteins identified as changed require validation in each of the individual plasma samples contributing to the pooled samples used in this analysis. Validation is required to confirm changes in protein expression and ultimately, to determine clinical usefulness.
5.2 Relative Quantification by Multiple Reaction Monitoring: Method Development and Validation of Candidate Proteins

5.2.1 Introduction

A large number of proteins were identified as changed in early onset pre-eclampsia samples when compared to control samples using iTRAQ. Validation of these changes is required to confirm the observations made during the discovery phase of the investigation. A large proportion of the proteins identified as changed in abundance during discovery are likely to be false positives as a result of the method and samples used for analysis. During the discovery phase of this investigation iTRAQ was applied to pooled plasma samples and relative quantification calculated against a homologous reference sample. As a result the changes in protein abundance observed may not be representative of the actual change in the samples. Validation of the protein change in each sample contributing to the pool is therefore required in order to confirm the changes in the protein identified are real.

Multiple Reaction Monitoring (MRM; see 1.3.2.4 and 1.3.3.1.2) is a highly selective targeted mass spectrometry method. MRM is increasingly being used to validate candidate markers identified during discovery phase (McKay et al. 2007; DeSouza et al. 2009). Where discovery phase investigations are also carried out using mass spectrometry the technique is particularly useful. MS data for peptides of candidate proteins identified during discovery phase can be applied directly to the MRM assay. Current quantitative MRM methods require the generation of costly stable isotope-labelled peptide standards. A stable isotope-labelled standard peptide is required for each peptide quantified (Anderson et al. 2004; Anderson and Hunter 2006). Validation of a large number of proteins identified as changed during discovery phase is not a practical option due to the prohibitive cost of stable isotope-labelled peptide standards. The MRM method used here was designed to quantify large numbers of changed proteins at low cost prior to application of more rigorous and costly methods of validation to a refined number of candidates. The aim in developing this method was to provide an interim step between the discovery phase and validation.

5.2.2 Aims

- To develop and test a high throughput label free MRM method for relative quantification.
- To validate changes in abundance of proteins identified in the plasma of women who developed early onset pre-eclampsia using iTRAQ.
5.2.3 Method

5.2.3.1 Relative Quantification – Enolase Normalisation

A non-human reference standard was used to assess and correct for variation in the MRM workflow allowing for relative quantitation of endogenous peptides. Plasma samples were spiked with a non-human protein standard - yeast enolase - prior to sample processing and analysis (see 3.3.1.1). Transitions from four enolase peptides with limited homology to endogenous human enolase were selected and retention times optimised for each transition (Table 5.2.1). The peak areas for the transitions were summed for each enolase peptide. These values were then used to calculate a ‘normalisation factor’ to correct for variation in sample processing and LC-MS/MS analysis. The peak areas for all four yeast peptides were summed and the sample with the largest peak area for the yeast enolase peptides was identified. The ratio of the summed enolase peak area for each sample to the largest summed enolase peak area was calculated. This value was used as a normalisation factor and was applied to peak areas for each peptide assayed to produce a normalised peak area (Figure 5.2.1).
Yeast enolase was spiked into each sample at 10 µg/m of plasma. Four enolase peptides were quantified in each sample. Three transitions were assessed for each peptide at the retention time stated.

<table>
<thead>
<tr>
<th>Peptide Sequence</th>
<th>Precursor m/z</th>
<th>Retention Time (min)</th>
<th>Fragment Ion</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVNDVIAPAFVK</td>
<td>643.8588</td>
<td>30.32</td>
<td>y10</td>
<td>1073.599</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>y7</td>
<td>745.4607</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>632.3766</td>
</tr>
<tr>
<td>AVDDFLISLDGTANK</td>
<td>789.9041</td>
<td>37.37</td>
<td>y9</td>
<td>918.4891</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>y8</td>
<td>805.405</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>605.2889</td>
</tr>
<tr>
<td>WLTGPQLADLYHSLMK</td>
<td>624.9941</td>
<td>43.65</td>
<td>y14</td>
<td>787.4058</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>y13</td>
<td>736.8819</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>y12</td>
<td>708.3712</td>
</tr>
<tr>
<td>AADALLLK</td>
<td>407.7553</td>
<td>24.93</td>
<td>y7</td>
<td>743.4662</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>672.4291</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>y5</td>
<td>557.4021</td>
</tr>
</tbody>
</table>
Enolase peptides were used to calculate a ‘normalisation factor’ to correct for variation in sample processing and LC-MS/MS analysis. Relative quantification of Apo E in five replicate samples was used to demonstrate the effect of normalisation. Enolase was spiked into each sample at 10 µg/ml. Four peptides were assessed and the peak area for all four peptides summed for each sample. A normalisation factor was calculated using the ratio of the summed peak area for enolase peptides to the sample with the largest enolase summed peak. The normalisation factor was applied to peak areas for each peptide proteins assayed to produce a normalised peak area for relative quantification.

![Enolase Peptides](image)

<table>
<thead>
<tr>
<th>Enolase Peptide</th>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 3</th>
<th>Sample 4</th>
<th>Sample 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AADALLUK</td>
<td>24507</td>
<td>21468</td>
<td>25196</td>
<td>25779</td>
<td>28218</td>
</tr>
<tr>
<td>WLTGPOLADLVHSLMK</td>
<td>17708</td>
<td>17387</td>
<td>19690</td>
<td>19295</td>
<td>19546</td>
</tr>
<tr>
<td>NVNDVARAPFK</td>
<td>29750</td>
<td>23188</td>
<td>27800</td>
<td>27876</td>
<td>27853</td>
</tr>
<tr>
<td>AvDCLDSLQGKAN</td>
<td>8946</td>
<td>7778</td>
<td>8682</td>
<td>8696</td>
<td>9032</td>
</tr>
<tr>
<td>Summed Peak Area</td>
<td>76560</td>
<td>69821</td>
<td>81446</td>
<td>81956</td>
<td>82559</td>
</tr>
<tr>
<td>Normalisation Factor</td>
<td>1.083</td>
<td>1.184</td>
<td>1.015</td>
<td>1.010</td>
<td>1.000</td>
</tr>
</tbody>
</table>

**Figure 5.2.1** Enolase peptides were used to calculate a ‘normalisation factor’ to correct for variation in sample processing and LC-MS/MS analysis. Relative quantification of Apo E in five replicate samples was used to demonstrate the effect of normalisation. Enolase was spiked into each sample at 10 µg/ml. Four peptides were assessed and the peak area for all four peptides summed for each sample. A normalisation factor was calculated using the ratio of the summed peak area for enolase peptides to the sample with the largest enolase summed peak. The normalisation factor was applied to peak areas for each peptide proteins assayed to produce a normalised peak area for relative quantification.
5.2.4 Results

5.2.4.1 Proof of Principle

To test the relative MRM method developed, non-human proteins were spiked into plasma alongside enolase and quantified. A sample of plasma from a female who was not pregnant was divided into three homologous aliquots and spiked with non-human proteins at varying concentrations (Figure 5.2.2). Each aliquot was further divided into 5 replicate spiked plasma samples however only 4 replicates of sample C were available for analysis. Spiked plasma samples were immunodepleted, exchanged into a suitable buffer, digested and desalted (see 3.3.1.1).

5.2.4.1.1 Transition Optimisation

Non-human proteins were reduced, alkylated, digested (see 3.3.1.1.2) and assessed by LC-MS/MS using a 6530 qTOF (see 3.3.1.3). Data were searched using MASCOT (see 3.2.5.1). For each protein, the three peptides with the greatest score and no missed cleavages were selected. The peptide data were inspected to determine the most abundant charge state and y-ions and the sequences imported into Skyline (MacLean et al. 2010). Each protein digest was analysed using the 6460 QQQ (see 3.3.1.2.4) to determine retention times and collision energies were optimised using Peptide Optimizer (Agilent, UK). A minimum of three optimal transitions were selected for each peptide.
Non human proteins were spiked into plasma and quantified by MRM.

Proteins were spiked into plasma alongside enolase. Proteins were spiked at a range of concentrations ratios to three homologous plasma aliquots. Spiked plasma samples were immunodepleted and concentrated digested using trypsin and desalted. Peptides were assayed by MRM.

<table>
<thead>
<tr>
<th>Proteins Spiked</th>
<th>Protein Concentration (ng/mL plasma)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Species</td>
</tr>
<tr>
<td>Yeast</td>
<td>Enolase</td>
</tr>
<tr>
<td>E.coli</td>
<td>Malate Dehydrogenase (MDH)</td>
</tr>
<tr>
<td>Gallus gallus</td>
<td>Lysozyme</td>
</tr>
<tr>
<td>Equus</td>
<td>Cytochrome C</td>
</tr>
<tr>
<td>Equus</td>
<td>Myoglobin</td>
</tr>
<tr>
<td>Bovine</td>
<td>B-Lactoglobulin (β-Lac)</td>
</tr>
<tr>
<td>Yeast</td>
<td>Alcohol Dehydrogenase (ADH)</td>
</tr>
<tr>
<td>Bovine</td>
<td>Catalase</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Protein</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeast</td>
<td>Enolase</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td>E.coli</td>
<td>Malate Dehydrogenase (MDH)</td>
<td>10,000</td>
<td>15,000</td>
<td>30,000</td>
</tr>
<tr>
<td>Gallus gallus</td>
<td>Lysozyme</td>
<td>2000</td>
<td>1500</td>
<td>1000</td>
</tr>
<tr>
<td>Equus</td>
<td>Cytochrome C</td>
<td>200</td>
<td>400</td>
<td>600</td>
</tr>
<tr>
<td>Equus</td>
<td>Myoglobin</td>
<td>1,500</td>
<td>1,200</td>
<td>1,000</td>
</tr>
<tr>
<td>Bovine</td>
<td>B-Lactoglobulin (β-Lac)</td>
<td>200</td>
<td>300</td>
<td>400</td>
</tr>
<tr>
<td>Yeast</td>
<td>Alcohol Dehydrogenase (ADH)</td>
<td>50</td>
<td>75</td>
<td>100</td>
</tr>
<tr>
<td>Bovine</td>
<td>Catalase</td>
<td>30</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>
5.2.4.1.2 Enolase Normalisation

Enolase was spiked into all plasma samples at a concentration of 10µg/mL and used for normalisation. Four enolase peptides were quantified, each with three transitions. The summed peak areas for the transitions of the enolase peptides were reported by Skyline. The CV was calculated across all samples for the enolase peptides quantified. The CV for peak areas measured in each replicate of the three sample sets were between 9 and 21% (Figure 5.2.3).

Figure 5.2.3 The CVs of peak areas for enolase peptides used for normalisation were between 9 and 21%. Non-human proteins at varying concentrations were spiked into three homologous plasma samples and quantified by MRM. Enolase was spiked into plasma at a concentration of 10µg/mL. The CV for peak areas of each enolase peptide measured across replicates of each plasma sample (A and B n=5, C n=4) were all above 9%.
5.2.4.1.3 Quantification of Spiked Proteins

Peak areas for peptides assayed were normalised against the four summed peak areas for the enolase peptides. The CV for the summed peak areas were calculated before and after normalisation (Figure 5.2.4). The CV for peptides assayed generally decreased on normalisation. The CV of peak areas for ADH and catalase increased on normalisation and were above 60% for at least one set of spiked sample replicates (not presented).

The peak areas for transitions of each peptide were summed and normalised for peptides assayed in each replicate sample and plotted. The significance of differences between peak areas was tested using a Kruskal-Wallis test (indicated by * in Figure 5.2.5). The concentrations of protein spiked into each sample were also plotted (Figure 5.2.5). The relative quantification of peak area for each peptide was compared to the concentrations at which the proteins were spiked. Three peptides derived from MDH were assayed; all three showed a relative increase in peak area which reflected the ratio of spiked protein. The peak area for the two peptides of myoglobin assayed also showed a similar pattern to the concentration at which the protein was spiked. Two peptides were assayed for lysozyme and cytochrome C. For both, the peak area for one peptide reflected the concentration of spiked protein, while one showed no relationship to the spiked protein concentration. The lysozyme peptide which was not quantified accurately had a CV above 20% for all three sets of spiked samples. The cytochrome C peptide which was not accurately quantified had a relatively low CV (<20%) in all three sets of spiked samples. The CV for the inaccurately quantified peptide was lower than the correctly quantified peptides. The CV for the incorrectly quantified peptide increased on normalisation rather than decreased. Peptides assayed from β-Lac, ADH and catalase showed no correlation to the concentration ratio at which the proteins were spiked (not shown).
The CV for peak areas of peptides of spiked proteins normalised against peak areas for enolase were calculated. Non-human proteins at varying concentrations were spiked into three homologous samples of plasma from a female who was not pregnant; each sample was further split into 5 replicates. Spiked proteins were quantified by MRM. The summed peak areas for enolase peptides were used to calculate a correction factor for each sample. The correction factor was applied to peak areas for each peptide assayed. The CV was calculated for peak areas of each peptide assayed across replicates of each sample (A and B n=5, C n=4) and were improved by normalisation.
Normalised peak areas for peptides assayed by MRM were compared to the concentration of spiked protein. Significant differences in the concentration of spiked protein reflected the concentration of MSH, lysozyme, cytochrome C and proteins were spiked into plasma alongside MDH and enolase. Proteins were spiked at a range of concentrations across three sample sets; A, B and C. At least one peptide from each protein in each sample set was compared to the concentration of spiked proteins assessed by MRM. Proteins were spiked into plasma alongside

Figure 5.2: Normalised peak areas

Concentration (ng/ml) vs Normalised Peak Area of Assayed Peptides
The CV of the normalised peak areas for the peptides assayed for each protein was calculated. The CV was plotted alongside the concentration at which each protein was spiked (Figure 5.2.6). A decrease in CV correlated with an increase in concentration of spiked protein. The CV fell below 20%, an indication of reproducible measurement, at a concentration range of approximately 200-400 ng/mL. This suggests the assay is reproducible to a minimum concentration range of 200 ng/mL. This proof of principle investigation of spiked non-human proteins demonstrated the ability of the MRM method to identify significant changes in abundance of peptides in the range of 200 ng/mL to 30 µg/mL. Although several proteins were accurately quantified, this investigation did suggest some care must be taken when normalising against enolase and that normalisation may be of limited use when applied to this workflow.

Figure 5.2.6 The CV of spiked proteins decreased as the concentration of protein spiked increased. Non-human proteins at varying concentrations were spiked into three homologous samples of plasma from a female who was not pregnant; each sample was further split into 5 replicates. Yeast enolase was spiked at 10 µg/mL into each sample. Peptides of each protein were assayed by MRM and the peak area normalised against that of the enolase peptides. The CV for each protein in each set of replicates was calculated and plotted against the concentration of the protein. The CV fell below 20% at a concentration of approximately 200-400 g/mL suggesting this assay is reproducible to this concentration range. The CV and concentration for sample A are shown.
5.2.4.2 Validation by MRM- Application to SCOPE Samples

The relative MRM method was applied to individual pre-eclampsia and matched control samples (both n=12) acquired from the SCOPE Study (see 3.1.2). A replicate sample was also processed alongside case and control samples to assess the variability introduced by the workflow. The replicate sample was produced from a pool of plasma from a non-pregnant female divided into homologous aliquots (n=5; see 3.1.3). Samples were spiked with enolase and applied to the workflow as outlined in 3.3.1.1. Replicate samples were run at intervals between early onset pre-eclampsia and control samples during the MRM assay (see 3.3.1.3).

5.2.4.2.1 Transition Optimisation for Plasma Proteins

Proteins of interest were identified from iTRAQ experiments both as part of this investigation and others (Table 5.7, Blankley et al. 2009; Blankley et al. 2010). Tryptic peptides of proteins of interest were selected from Peptide Atlas (Farrah et al. 2011) or from iTRAQ data. Peptides of 8-20 amino acids with the highest number of observations in the HumanPlasma spectral library (Farrah et al. 2011) and no missed cleavages were selected. A synthetic version of selected peptides from each protein (SpikeTides, JPT Peptide Technologies, Germany) were analysed on the 6460 QQQ (see 3.3.1.3) to determine retention times; collision energies were optimised using Peptide Optimizer (Agilent, UK). A minimum of two optimal transitions were selected for each peptide.
Peptides of proteins of interest were selected and retention times and a minimum of three transitions optimised.

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Peptide Sequence</th>
<th>Precursor m/z</th>
<th>Retention Time (min)</th>
<th>Fragment Ion</th>
<th>Product m/z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apo C2</td>
<td>TYLPADVDEK</td>
<td>518.271</td>
<td>21.35</td>
<td>y7</td>
<td>771.425</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>658.341</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>z7</td>
<td>378.207</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y8</td>
<td>957.431</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y7</td>
<td>870.399</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>783.367</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y9</td>
<td>953.494</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y8</td>
<td>854.425</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y7</td>
<td>753.378</td>
</tr>
<tr>
<td>Apo C3</td>
<td>ESLSSYWESAK</td>
<td>643.799</td>
<td>28.21</td>
<td>y7</td>
<td>792.400</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>721.363</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y5</td>
<td>592.320</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y8</td>
<td>898.474</td>
</tr>
<tr>
<td>Apo E</td>
<td>LQAEAFQAR</td>
<td>517.275</td>
<td>18.59</td>
<td>y7</td>
<td>872.437</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>721.363</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y5</td>
<td>592.320</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y8</td>
<td>898.474</td>
</tr>
<tr>
<td>Clusterin</td>
<td>AATVGSAGQPLQER</td>
<td>749.405</td>
<td>22.93</td>
<td>y7</td>
<td>785.452</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>686.383</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y8</td>
<td>783.367</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y7</td>
<td>753.378</td>
</tr>
<tr>
<td>hCG</td>
<td>VLQGVLAPLQCVCNNR</td>
<td>642.693</td>
<td>38.73</td>
<td>y8</td>
<td>1035.504</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y5</td>
<td>711.324</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>518.526</td>
</tr>
<tr>
<td>Pigment</td>
<td>DTDTGALLFIGK</td>
<td>625.835</td>
<td>34.85</td>
<td>y10</td>
<td>1034.588</td>
</tr>
<tr>
<td>epitheliun-</td>
<td></td>
<td></td>
<td></td>
<td>y9</td>
<td>919.561</td>
</tr>
<tr>
<td>derived factor</td>
<td>LQSLFDSPDFS overshoot predicting a peptide retention time</td>
<td>692.343</td>
<td>32.49</td>
<td>y10</td>
<td>1142.536</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y8</td>
<td>818.513</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y7</td>
<td>761.492</td>
</tr>
<tr>
<td>Platelet</td>
<td>GTHCNQVEVIATLK</td>
<td>523.940</td>
<td>23.89</td>
<td>y7</td>
<td>773.477</td>
</tr>
<tr>
<td>Basic Protein</td>
<td>NIQSLEVIGK</td>
<td>550.819</td>
<td>26.28</td>
<td>y6</td>
<td>644.434</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y5</td>
<td>545.366</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y7</td>
<td>773.477</td>
</tr>
<tr>
<td>PSG-1,3,4,6,7,9</td>
<td>SVPVTLLNLLPK</td>
<td>598.848</td>
<td>33.01</td>
<td>y8</td>
<td>994.630</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>697.577</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y7</td>
<td>798.526</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>697.461</td>
</tr>
<tr>
<td>PSG-1,7</td>
<td>SVPVTLLNVYLPDLP</td>
<td>585.981</td>
<td>36.52</td>
<td>y7</td>
<td>817.420</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>654.357</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y5</td>
<td>597.335</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y7</td>
<td>777.487</td>
</tr>
<tr>
<td>PSG-1,8</td>
<td>TLFLLGVTK</td>
<td>496.313</td>
<td>36.76</td>
<td>y6</td>
<td>630.418</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y5</td>
<td>517.334</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y7</td>
<td>811.471</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>664.403</td>
</tr>
<tr>
<td>PSG-3,4,11</td>
<td>TLFLFGVTK</td>
<td>513.305</td>
<td>38.72</td>
<td>y5</td>
<td>551.319</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y8</td>
<td>930.504</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y7</td>
<td>817.420</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>597.335</td>
</tr>
<tr>
<td>PSG-3,8</td>
<td>SYPVTLLNVYLPDLP</td>
<td>601.993</td>
<td>38.17</td>
<td>y7</td>
<td>817.420</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>597.335</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y8</td>
<td>930.504</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y7</td>
<td>817.420</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>597.335</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y5</td>
<td>551.319</td>
</tr>
<tr>
<td>PSG-4,5,7</td>
<td>LFIPQITTK</td>
<td>530.824</td>
<td>29.51</td>
<td>y7</td>
<td>800.488</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>687.404</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y5</td>
<td>590.351</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y7</td>
<td>827.466</td>
</tr>
<tr>
<td>PSG-6,9</td>
<td>LQSLFDSPDFS overshoot predicting a peptide retention time</td>
<td>521.303</td>
<td>34.79</td>
<td>y6</td>
<td>664.403</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y5</td>
<td>551.319</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>590.351</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y7</td>
<td>827.466</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>597.335</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y5</td>
<td>551.319</td>
</tr>
<tr>
<td>PSG-9</td>
<td>SNPVILNVYLPDLP</td>
<td>589.665</td>
<td>41.25</td>
<td>y6</td>
<td>654.357</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y5</td>
<td>597.335</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y6</td>
<td>727.446</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y4</td>
<td>614.362</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y5</td>
<td>517.309</td>
</tr>
<tr>
<td>PSG-9</td>
<td>LFIPQITTK</td>
<td>494.303</td>
<td>32.08</td>
<td>y4</td>
<td>614.362</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y5</td>
<td>517.309</td>
</tr>
</tbody>
</table>
5.2.4.2.2 Quantification of Plasma Proteins

The CV for the summed peak areas of enolase peptides was calculated for early onset pre-eclampsia, control and replicate samples (Figure 5.2.8-A). The CV for enolase peptides spiked into early onset pre-eclampsia and replicate samples approximated 20%, and was greater than control samples (5-10%). Peak areas for samples in each group were plotted to identify outlying values (Figure 5.2.8-B); quantification of one early onset pre-eclampsia sample was identified as very low. Removal of this sample from the dataset improved the CV of the enolase peptides to a small extent. Quantification of the enolase peptides within replicate samples was variable as indicated by the CV and no one sample could be identified as an outlier.

Peak areas for peptides assayed were normalised against summed peak areas for enolase peptides. The CV for each peptide was calculated for replicate samples before and after normalisation (Figure 5.2.9). Normalisation increased the CV for all peptides indicating an increase in variance on normalisation as opposed to the intended decrease.
Quantification of enolase peptides was more variable in early onset pre-eclampsia and replicate samples than in control samples. Enolase was spiked into plasma samples at 10µg/ml, samples were applied to the workflow and enolase peptides quantified by MRM. A-The CV for peak areas of enolase peptides spiked into early onset pre-eclampsia and replicate samples were larger than control samples. B- Peak areas for each enolase peptide were plotted for samples in each group. One outlying sample was identified in the early onset pre-eclampsia sample group (indicated by *); removal of this sample decreased the CV for the sample group to approximately 10% for all peptides. No one outlying sample could be identified in the replicate sample group.
Peptides of proteins of interest were assessed by MRM. Peptide areas for peptides assayed were normalised against the peak areas of the peptide of enolase which was spiked into plasma at 10µg/ml. Peptide areas for peptides assayed were normalised against summed peak areas of spiked enolase peptides. Replicate samples of plasma spiked with enolase were processed alongside early onset pre-eclampsia and control samples of plasma spiked with enolase. Peptide areas for peptides assayed by MRM in replicated samples increased on normalisation against summed peak areas of spiked enolase peptides. The CV of the peak area for peptides assayed by MRM in replicated samples increased on normalisation against summed peak areas of spiked enolase peptides.
5.2.4.2.3 Quantification of Candidate Proteins

As demonstrated in the proof of principle experiment an increase in the CV on normalisation indicates a decrease in the quality of the quantification data. Summed peak areas for candidate proteins of particular interest identified during the discovery phase (5.1) were therefore investigated without enolase normalisation. The peak areas for peptides of platelet basic protein and PSG9 quantified were plotted (Figure 5.2.10). The CV in replicate samples for both PSG9 peptides was 5%, for the peptides of platelet basic protein the CVs were 5% and 10% indicating peptide quantification was reproducible. Each peptide showed an increase in abundance in early onset pre-eclampsia samples compared to control samples. The relative increase in abundance for each peptide was calculated from the average peak area for early onset pre-eclampsia and control samples. The relative ratio for both PSG9 peptides was 1.31. The relative ratios for peptides of platelet basic protein were 1.50 and 1.52. Differences between early onset pre-eclampsia samples and control samples were not found to be significant using the Mann Whitney U test.

Figure 5.2.10  Peak areas for peptides of PSG9 and platelet basic protein indicated an increase in abundance in early onset pre-eclampsia compared to control. The relative MRM method was applied to early onset pre-eclampsia, matched controls and 5 replicate samples. A minimum of two transitions were used to quantify two peptides for each protein. Peak areas for all transitions were summed and plotted.
5.2.4.3 Workflow Optimisation

The CVs for spiked enolase peptides were variable and showed distinct differences between sample groups. As a result normalisation using the summed peak areas for enolase peptides increased, rather than decreased, the variation in quantitation data as intended. This suggests that variation was introduced by the sample processing method used. The distinct differences between the CVs for each sample group are a result of the samples being processed in batches and indicate sample processing is not reproducible over time (Figure 5.2.3). To improve the accuracy and reproducibility of sample processing the workflow was optimised.

Each stage of the MRM workflow was investigated with the aim of improving the reproducibility and accuracy of the method (see 3.3.1.2). Sample concentration following immunodepletion (mRP column vs. MWCO spin columns), desalting following digestion (C18 vs. Solid Phase Extraction (SPE) columns) and re-suspension prior to LC–MS analysis were investigated. The alternative sample preparation methods were compared by using the Sulpelco Analytical Discovery C18 column for peptide quantitation and the 6530 qTOF.

5.2.4.3.1 Comparison of Concentration Methods

Following immunodepletion using the MARS 14 column, samples were concentrated and desalted using an mRP column. Samples were then dried and re-suspended into a buffer suitable for digestion. Drying and re-suspending samples was thought to be a highly variable stage in the workflow. Alternative methods of concentration and buffer exchange were investigated.

Replicate plasma samples (n=16) were immunodepleted using the MARS 14 system (Figure 5.2.11). Eight of those depleted samples were applied to the mRP column, dried and re-suspended. The remaining eight samples were exchanged into TEAB using MWCO spin columns. The samples were digested and four of the samples processed using each method were applied to solid phase extraction (SPE) columns for desalting. All samples were dried and re-suspended into appropriate buffer for MRM analysis. Samples prepared using the mRP columns were noticeably more difficult to re-suspend. The amount of peptide in each sample was quantified using the C18 column (Figure 5.2.12-A). More peptide was recovered from samples prepared using the MWCO spin columns than the mRP column. The amount of peptides in samples desalted using the SPE column was lower but less variable than samples which were not desalted. Samples were also assessed using the qTOF; the CV was calculated for the peak area of the extracted ion chromatogram (EIC) for 10 peptides (Figure 5.2.12-B). The EIC is a plot of the signal intensity measured for a selected m/z plotted against chromatographic retention time. Data acquired from the qTOF was searched using MASCOT and the number of proteins identified in each sample compared (Figure 5.2.12-B). Peptide recovery and protein identification was greater using the MWCO spin
columns and more reproducible following desalting using the SPE column. As a result MWCO spin columns were used for concentration and exchange of immunodepleted sample into a suitable buffer prior to digestion. No concentration or desalting was applied post digestion.

Figure 5.2.11 The MRM sample preparation workflow was optimised to improve reproducibility of samples processing. Concentration and desalting are thought to introduce variance and were investigated. Replicate plasma samples were applied to the MARS 14 system, half were concentrated using the mRP column (n=8), half using MWCO spin columns (n=8). All samples were digested, and half off those concentrated using each method were desalted using the SPE columns (n=4), half had no further desalting (n=4). Samples were dried and resuspended in appropriate buffer. Analysis was carried out using a quantitative applied to the C18 column and the 6530QTOF.
Figure 5.2.12  Samples prepared using a MWCO were digested more efficiently, assayed more consistently and resulted in more protein identifications than the sample prepared using the mRP column. Replicated samples were immunodepleted. Eight depleted samples were applied to the mRP column, dried, re-suspended and digested; four of those samples were then applied to the SPE column. Eight depleted samples were applied to MWCO, and exchanged into TEAB buffer, digested and four of those samples applied to SPE columns. A- The amount of peptide in each sample was quantified using the C18 column. B- Samples were assessed using the TOF. The average CV for the peak areas of the extracted ion chromatogram (EIC) of 10 peptides was calculated. The data were also searched using MASCOT and the average number of peptides identified in each sample set compared.
### 5.2.4.3.2 LC-MS/MS and MRM Assay Reproducibility

To investigate the reproducibility of the LC-MS/MS system and the MRM assay, the remainder of the early onset pre-eclampsia, control samples and replicates discussed prepared as described previously (see 5.2.4.2) were applied to the LC-MS/MS. The samples were pooled and divided into 20 aliquots. The aliquoted samples were dried and re-suspended to simulate the final stages of the sample preparation workflow. Each aliquot was assayed 4 times by MRM. The CV was calculated for peak areas of peptides quantified before and after enolase normalisation (Figure 5.2.13). Before normalisation the CV for all peptides fell below 15%, after normalisation the CV fell below 10%. This indicates peak areas were measured reproducibly across all 4 assays of all 20 aliquoted samples; a total of 80 individual assays. This demonstrates the final steps of the workflow including drying, re-suspension and assay of the sample is reproducible.

### 5.2.4.3.3 Summary of Optimised Workflow

A more efficient and reproducible workflow was established as follows: samples were spiked with enolase, immunodepleted using the MARS 14 system, washed and exchanged into TEAB using MWCO spin columns. Samples were reduced, alkylated and digested as described previously; the same method of digestion was used for all plasma samples depleted using the MARS 14 column and this method was shown to be efficient on analysis of iTRAQ data (Table 4.3.1). No further optimisation of this stage in the sample processing was required. Dried and re-suspended sample was exchanged into a buffer suitable for LC-MS/MS analysis. Samples were randomised for processing to avoid the ‘batch effect’ observed on analysis of the samples from the SCOPE study. The application of this workflow to the relative MRM method is currently under investigation.
The CV of the peak area for peptides assayed by MRM in 80 individual assays was consistent and fell below 10% on normalisation.

Early onset pre-eclampsia, control samples and replicates were prepared for assay by MRM as described above. A proportion of the samples were applied to the LC-MS/MS for analysis. The remainder of the samples were used to investigate the reproducibility of the LC-MS/MS system and the MRM assay.

Each aliquot was assayed 4 times for peptides with optimised transitions.
5.2.5 Discussion

The aim of this investigation was to further develop and test a high throughput, low cost MRM method for relative quantification of proteins and to apply the method to the validation of candidate biomarkers. The method is designed for application to large numbers of candidate biomarkers as an interim validation step. More sensitive and stringent methods of validation could then be applied to filtered lists of candidate proteins.

5.2.5.1 Multiple Reaction Monitoring

Using mass spectrometry for both identification and validation of candidate biomarkers is a natural and straightforward extension of the biomarker discovery pipeline. The workflow for both discovery and validation are similar; sample preparation revolves around key stages including removal of highly abundant proteins and tryptic digestion of proteins into peptides. As a result, peptide data for candidate proteins acquired during the discovery phase investigation can be used to optimise transitions for the subsequent MRM assay. Several studies have used MRM to investigate proteins indicated to be dysregulated in disease (Keshishian et al. 2007; McKay et al. 2007; Ahn et al. 2009; Fortin et al. 2009) including several which used MS methods for the discovery phase (DeSouza et al. 2009; Lopez et al. 2010).

As previously discussed, mass spectrometry is not inherently quantitative and the data obtained does not directly reflect the quantity of the peptide in the sample. The data obtained from an MS spectrum indicates the intensity at which the peptide was measured which can be affected by the sample processing method used, ionisation efficiency of the peptide and technical drift of the MS instrument. Stable isotope labelled peptides are therefore applied in order to infer quantitative data from MRM analysis. For each peptide of interest, a synthetic isotopically labelled peptide is spiked into the sample at a known concentration. Endogenous and spiked peptides elute together but appear at a different m/z on MRM analysis. The concentration of endogenous peptide can then be calculated by comparison to the spiked labelled peptide (Keshishian et al. 2007). A matched synthetic isotopically labelled peptide is required for each peptide of interest. Accurate quantitation therefore relies on the quality and behaviour of the synthetic peptide. Quantitation also relies on the assumption that the amount of synthetic peptide observed on analysis is equal to the amount spiked with no loss during sample processing. Quantitation using stable isotopes is sensitive and reproducible; proteins with concentrations in the ng/ml level have been quantified and CVs below 20% reported (Anderson and Hunter 2006; Keshishian et al. 2007; Addona et al. 2009). The ease of multiplex analysis is a major advantage of the MRM method; however investigation of multiple proteins using this method may be constrained by the cost of producing multiple synthetic peptides.
The discovery phase of the current investigation identified a large number of proteins that were changed in abundance between pooled case and control plasma samples in a very small number of observations. A large proportion of those proteins are likely to be false positive observations. As previously established, validation of those protein changes is required. Application of validation methods such as stable isotope labelled MRM or antibody methods to all proteins identified are prohibited by both cost and time.

Several methods for low cost quantification have been developed which could be applied to the biomarker pipeline. Silva et al. developed a method of absolute quantification by application of LC-MS/MS\textsuperscript{E}; a Waters label free quantification method (Waters, Milford, MA), to samples spiked with a reference protein (Silva et al. 2006). The method, however, is reliant upon specific instruments and software, and cannot be used to target specific proteins. Ishihama et al. also developed a method for global relative quantification by relating the number of observed peptides to the number of potentially observed peptides (Ishihama et al. 2005). Again the method is global and cannot be used to target proteins of interest. Of immediate application to the biomarker discovery pipeline, Lopez et al. developed a method using receiver operating characteristic curve algorithms to narrow down lists of candidate proteins identified by label free LC-MS/MS (Lopez et al. 2010). The algorithm uses sensitivity and specificity to identify proteins which differentiated case from control samples. Proteins of most ‘usefulness’ were then assessed and validated by stable isotope labelled MRM (Lopez et al. 2010). However, assessment of diagnostic usefulness of un-validated protein changes in a limited sample set (case; n=24, control n=21), seems premature. An interim method to reliably reduce the number of changed proteins requiring more stringent and costly validation would therefore be of great use to the biomarker discovery pipeline.

5.2.5.2 Proof of Principle

The method described here uses enolase, which is cheap and readily available as a reference standard, reducing the cost of the assay. In this small scale proof of principle investigation, non-human proteins were spiked at known concentration into plasma alongside enolase. Peptides of the exogenous proteins were assayed and relative quantification of each peptide compared. The CV of replicate samples was used as a measure of variance in the assay. The CV for enolase peptides in each set of samples indicated a good level of reproducibility. Enolase peptides were used to normalise peak areas; for most of the endogenous peptides the CV decreased on normalisation indicating a reduction in spread of the data. The quantitative ability of the assay was investigated by comparing spiked protein concentrations against the relative ratios of normalised peak areas. The assay was able to identify significant differences between peptides spiked at different concentrations. Although the CV for peptide TGQAPGFTYTDANK of cytochrome C is larger than for peptide EDLIAYLK the relative quantification for peptide
TGQAPGFTYTDANK was significantly different between samples and reflects the concentration at which the protein was spiked. The relative quantification for peptide EDLIAYLK does not reflect the spiked concentrations even though the lower CV may indicate ‘better’ data. While the CV of a dataset may be a useful indicator of the variation within a sample set it may not give an accurate representation of quality of quantification. This suggests careful selection and optimisation of peptides using cheaply produced synthetic peptides is essential as contradictory data such as this is less easily resolved when measuring endogenous peptides at unknown concentrations.

Peptides for quantification must be robust and behave reproducibly during both sample processing and LC-MS/MS analysis in order to be representative of the protein. Variation in quantification of a peptide can be introduced by sample processing including loss during handling, poor digestion or naturally occurring or artificially induced unexpected post-translational modifications (Addona et al. 2009). Peptides must elute from the LC column in a highly reproducible manner in order to be quantified within the specified retention time and produce a good strength of signal during MS/MS analysis. Peptides with a low signal strength may be obscured by background signal, particularly in complex samples such as plasma and also, where used, from heavy labelled peptides (Addona et al. 2009). Care must also be taken when selecting peptides and transitions representative of the protein for quantification. While a peptide may have a unique amino acid sequence, the m/z on ionisation may be homologous to other peptides leading to redundancy in the assay and false positive identifications (Sherman et al. 2009a). Using the LC retention time may differentiate peptides with homologous ions, however this approach requires the retention time to be extremely well defined (Sherman et al. 2009a). Digestions of the spiked non-human proteins were used here to determine retention times for proteins quantified, however this is more difficult for endogenous proteins. Synthetic peptides are relatively cheap to obtain and were subsequently used to determine retention times for endogenous proteins assayed in plasma samples. A minimum of two transitions were selected for each peptide. Sherman et al. demonstrated that greater than 96% of human proteins with a sequence unique peptide could be quantified using two unique ion signatures, i.e. transitions that could confer peptide and protein identity (Sherman et al. 2009b). Three transitions are more commonly used (Addona et al. 2009), with multiple peptides per protein to overcome redundancy of sequence and m/z.

As part of this investigation a preliminary comparison of the average CV for each protein against the spiked concentration was carried out and indicates the assay has comparable sensitivity to that reported for stable isotope labelled MRM. This comparison, however, suggests that the quantitation of proteins that are present below 400ng/ml of plasma may be unreliable using this assay. Using stable isotope labelled methods, Addona et al. reported identification of peptides from spiked proteins in the range of 0.70-3.34 µg/ml in whole plasma, however CVs of up to 45% were reported (Addona et al. 2009). Fortin et al.
quantified prostate specific antigen in albumin depleted plasma at the ng/mL level with comparable performance to an ELISA, however relatively large volumes (100 µl) of plasma were required for the MRM assay. Keshishian et al. quantified spiked proteins in the 1-10ng/mL range with CVs from 3-15% in immunodepleted fractionated plasma (Keshishian et al. 2007). The Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA) method developed by Anderson and Anderson (Anderson et al. 2004) uses antibodies to enrich for peptides of interest prior to MRM analysis. Peptide enrichment increases the sensitivity of the assay to the ng/ml range; Ahn et al. identified proteins at approximately 0.8ng/ml in 1.7ul of serum using SISCAPA (Ahn et al. 2009). Whiteaker et al. achieved detection limits in the ng/ml range in 10µl of plasma with a median CV of 12.6%; sensitivity was increased to the low pg/ml range by using larger volumes (1ml) of plasma (Whiteaker et al. 2009). SISCAPA, however, requires synthetic isotopically labelled peptides as well as the generation of peptide antibodies making it both costly and time consuming. Further investigation is required to assess the linearity of response and true lower limit of detection and quantification for the assay described here. Use of larger volumes of plasma and further fractionation of the sample may improve upon these limits.

5.2.5.3 Validation by MRM - Application to SCOPE Samples

The MRM method was applied to early onset pre-eclampsia and matched control samples. The aim of the assay was to validate the increase in abundance of the proteins identified by iTRAQ in the discovery phase of this investigation. Relative quantification of peptides from both PSG9 and platelet basic protein indicated an increase in abundance of these proteins in early onset pre-eclampsia samples compared to control samples. Two peptides from each protein were assayed and showed highly consistent increases in abundance for each protein.

Validation by MRM is highly sensitive and specific; isoforms and known post translational modifications can be identified and unique peptides targeted. PSG9 is a member of a family of highly homologous proteins (Bayhan et al. 2005) which presents a problem for validation using other less specific methods (see 5.2.1). Peptides from unique regions of the protein can be selected for assay by MRM offering a significant advantage. Relative quantification by MRM for both PSG9 and platelet basic protein supported data acquired by iTRAQ during the discovery phase investigation. Validation by MRM used information on peptides acquired on LC-MS/MS analysis iTRAQ labelled plasma. Transitions were identified for peptides previously observed by LC-MS/MS analysis. Peptides assigned to proteins of interest identified during discovery phase have already proved reproducible during sample processing and LC-MS/MS analysis. Peptides previously identified can then be applied to the MRM providing a shortcut in the search for suitable representative peptides. Following MRM assay, relative quantification is carried out at the peptide level and an average value for the change in abundance produced, as such quantification by MRM is
an almost direct repetition and so validation of the iTRAQ method. Validation of the change in abundance is still required using orthogonal methods, particularly for this data where no normalisation was applied to confirm reproducibility of the MRM method and changes may be an artefact of that method.

The CV for enolase peptides was consistently greater for early onset pre-eclampsia and replicate samples than control samples. The differences between the CV in the different sample sets may be a result of the samples being processed in batches. Replicate samples were interspaced between early onset pre-eclampsia and control samples during the MRM assay, however were processed in two batches with both early onset pre-eclampsia and control samples. Separate processing may account for the discrepancy in CV for enolase peptides in each sample set. It is imperative that all samples are spiked with enolase from a single mix of sample buffer and processed in a random order. While stages of the workflow such as immunodepletion must be carried out sequentially due to the nature of the technique, performance must be monitored throughout. Where possible other stages of the workflow, for example, digestion and buffer exchange, must be carried out on all samples as one to avoid batch to batch variance. For this experiment the CV for peak area of assayed peptides increased on enolase normalisation indicating an increase in the spread of the data. This suggests normalisation against enolase peptides is only of use where measurement of those peptides is highly consistent and reproducible.

5.2.5.4 Workflow Optimisation

The workflow used to prepare samples was investigated and optimised to improve accuracy and reproducibility of the MRM method.

The aim was to reduce the batch processing effect observed. Preparation of a larger number of samples may be necessary in batches and a more reproducible sample processing method would reduce the variation introduced by this. As previously discussed, immunodepletion is a sequential process which may introduce variation into the workflow; performance of the immunodepletion column was monitored throughout sample processing to identify any change in column performance. Earlier in the course of this work, immunodepletion using the MARS 14 column was investigated and shown to be reproducible (see 4.1); no further investigation of this stage in the workflow was necessary. Buffer exchange following immunodepletion and concentration of samples was investigated. Samples were concentrated using an mRP column, then dried and resuspended in the buffer used for digestion. Drying and re-suspending samples is extremely variable and reliant on the solubility and behaviour of the dried plasma proteins. This stage in the workflow was replaced by MWCO spin filters which allow concentration and buffer exchange without the need to dry and re-suspend samples. Concentration and desalting following digestion were also investigated but found to be of negligible benefit to the workflow. The exchange of samples into a mass spectrometry compatible buffer and the MRM assay were
also investigated. Buffer exchange and MRM analysis of a homologous sample set showed a high level of reproducibility. A robust and reproducible LC-MS/MS method is essential to the assay; the LC-MS/MS system used here was found to be highly reproducible.

Application of the method developed to samples acquired from the SCOPE study was not possible during the course of this investigation due to time restrictions. This optimised workflow is currently under investigation and the MRM method developed shows promise as a first pass validation method for iTRAQ data.

5.2.6 Conclusion

The first aim of this investigation was to develop and test a high throughput low cost MRM method for relative quantification. The method was design for validation of large numbers of candidate biomarkers generated during discovery investigations without use of stable isotope labelled synthetic peptides. A small scale proof of principle investigation was carried out which demonstrated the ability of the assay to reproducibly quantify significant differences in concentration between exogenous spiked peptides. The second aim of this investigation was to validate changes in abundance of proteins identified in the plasma of women who developed early onset pre-eclampsia using iTRAQ. The MRM method was applied to early onset pre-eclampsia and matched control samples used for the discovery phase of this investigation. In accordance with the iTRAQ data; relative quantitation of peptides of PSG9 and platelet basic proteins showed a highly consistent increase in plasma levels in early onset pre-eclampsia samples. The results of the MRM assay suggest both PSG9 and platelet basic proteins are worth prioritising for further validation.

The workflow was investigated to improve reproducibility of sample processing. Modifications to the workflow suggested in the course of this research are currently being investigated to optimise the method. This method shows great potential as an interim method for refining lists of candidate biomarkers for more thorough validation using more specific and costly methods.
5.3 Validation of Candidate Proteins by Western Blot and ELISA

5.3.1 Introduction

The discovery phase of a biomarker investigation quantifies large numbers of proteins in a small number of samples using a hypothesis generating method. As a result, candidate biomarkers identified may be artefacts of the samples or method used. Immunodepletion can leave candidate biomarkers at artificial concentrations levels in the depleted samples. Relative quantification is averaged at the peptide level and leaves candidate markers vulnerable to small numbers of abnormally quantified peptides. Proteins identified may also be easily identified by LC-MS/MS but not by other methods which ultimately leaves the candidate marker of limited use in a clinical setting. Validation of the candidate proteins identified is therefore necessary.

Antibody methods, including western blotting and ELISA, are relatively quick, easy and specific and so ideally suited to targeted studies. Antibody methods are commonly used to validate changes in protein abundance identified in discovery phase investigations. Application of antibody methods to validation of proteomic investigations of pre-eclampsia has been demonstrated. Atkinson et al. used 2D western blotting to validate proteins changed in abundance identified using DIGE analysis (Atkinson et al. 2009). Blumenstein et al. validated changes in six proteins identified by DIGE using western blot (Blumenstein et al. 2009a) and two more proteins were identified as changed using DIGE and validated using western blot in a later study (Blumenstein et al. 2009b). Blankley et al. used ELISA assays to validate changes identified in several proteins using iTRAQ (Blankley et al. 2009). Western blots and ELISA were also used by Auer et al. to validate changes in protein abundance identified by iTRAQ, immunohistochemistry was applied to further investigate the related biology implicated by the changes identified (Auer et al. 2010).

5.3.2 Aim

- To use antibody methods to validate changes in abundance of plasma proteins identified at 15 weeks gestation in women who subsequently developed pre-eclampsia compared to women with uncomplicated pregnancies.
5.3.3 Methods

5.3.3.1 Plasma Samples

Plasma samples collected at 15 weeks gestation were acquired from women who subsequently developed early onset pre-eclampsia and from women with uncomplicated pregnancies for use as matched controls (both n=12). These samples were previously investigated as two phenotypic pools using iTRAQ (see 3.1.2); for validation each of the twelve samples was investigated. Plasma samples were assessed by western blot and ELISA.

5.3.3.2 NAP-2 as a Platelet Basic Protein Surrogate

Platelet basic protein was identified using iTRAQ; peptides assigned to platelet basic protein by ProteinPilot and used for iTRAQ relative quantitation are shown in Figure 5.3.1. Antibodies to platelet basic protein are not commercially available and so neutrophil-activating protein-2 (NAP-2) was used as a surrogate. NAP-2 is a cleavage product of platelet basic protein (amino acid residues 59-128; highlighted in yellow in Figure 5.3.1). Antibodies for western blotting and an ELISA kit were available for NAP-2.

MSLRLDTPS CNSARPLHAL QVLLLLSLLL TALASSTKGQ TKRNLAKGKE
ESLDSDLAYE LRCMCIKTTS GIHPKNIQSL EVIGKTHCN QVEVIATLKD
GRKICLDPDA PRIKKIVQKK LAGDESAD

IgY 14-SuperMix—QSTR peptides
IgY 14-SuperMix-5800 peptides
MARS 14- 5800 peptides
MRM Quantified Peptides

NAP-2

Figure 5.3.1 NAP-2 was used as a surrogate for platelet basic protein. Antibodies for platelet basic protein were not available. The amino acid sequence for platelet basic protein is annotated with peptides assigned to platelet basic protein and used for iTRAQ relative quantitation for each dataset and quantified by MRM. NAP-2 is a 70 amino acid (highlighted in yellow) cleavage product of platelet basic protein. Antibodies and an ELISA kit for NAP-2 were readily available and so were used as an indicator of platelet basic protein abundance.
5.3.3.3 Western Blots

Individual plasma samples were diluted 1:10 in ddH₂O and applied in equal volumes to an SDS-PAGE gel as described in 3.3.2.1.1. Proteins were identified by western blotting (see 3.3.2.1.3) using conditions detailed in Table 5.3.1. All 12 plasma samples were assessed for each protein of interest; four early onset pre-eclampsia samples and four control samples were loaded alongside a reference sample on one gel and three gels were run at the same time for each protein. The reference sample used was plasma acquired at 26 weeks gestation from a woman with an uncomplicated pregnancy (3.1.3). Relative quantification was carried out against the reference sample loaded on each gel using Image Lab™ 2.0 Software (BioRad, Hemel Hempstead, UK). Gel bands quantified were defined manually. Data were exported to Graphpad Prism version 4.01 (San Diego, California) for statistical analysis.

5.3.3.4 ELISA

Platelet basic protein (NAP-2) and vinculin were measured using commercially available ELISA kits; manufacturer’s assay protocols were followed (NAP-2; RayBiotech Inc. GA, USA. Vinculin; Biotang Inc. MA, USA). The NAP-2 ELISA kit had a minimum detectable limit of 8.5pg/mL, an average recovery of 94% in plasma and the CV for intra-assay reproducibility was less than 10%. The plasma samples were diluted 1/3000 in provided buffer. The range of detection for the vinculin ELISA kit was 3-100ng/mL and plasma was diluted 1/200 using appropriate buffer. Data were exported to Graphpad Prism version 4.01 (San Diego, California) for statistical analysis.
Table 5.3.1  Summary of conditions used for SDS-PAGE gels and antibodies for Western Blots. Plasma was diluted 1:10 in ddH$_2$O and then 1:1 in Lamelli buffer. An appropriate volume of diluted plasma was loaded onto a gel of the optimal percentage acrylamide for the molecular weight of the protein targeted. Proteins were identified by western blotting using primary antibodies from abcam® (Cambridge, UK).

<table>
<thead>
<tr>
<th>Protein</th>
<th>SDS-PAGE Gel (% Acrylamide)</th>
<th>Volume of Plasma (µl)</th>
<th>Molecular Weight (kDa)</th>
<th>abcam® Code</th>
<th>Primary Antibody Dilution</th>
<th>Primary Antibody</th>
<th>Western Blot Immunogen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibulin</td>
<td>10</td>
<td>1</td>
<td>77</td>
<td>ab67625</td>
<td>1/1000 (0.5% Milk)</td>
<td>Mouse polyclonal</td>
<td>Mouse polyclonal to Fibulin 1</td>
</tr>
<tr>
<td>Filamin</td>
<td>5</td>
<td>1</td>
<td>250</td>
<td>ab3261</td>
<td>Not established</td>
<td>Mouse monoclonal</td>
<td>Platelet filamin</td>
</tr>
<tr>
<td>Platelet Basic Protein (NAP-2)</td>
<td>15</td>
<td>2</td>
<td>14</td>
<td>ab9554</td>
<td>1/500 (1% BSA)</td>
<td>Rabbit polyclonal</td>
<td>Highly pure (&gt;98%) recombinant hNAP-2</td>
</tr>
<tr>
<td>Pregnancy-specific beta-1-glycoprotein 9</td>
<td>10</td>
<td>2</td>
<td>58</td>
<td>ab64425</td>
<td>1/200 (1% BSA)</td>
<td>Rabbit polyclonal</td>
<td>A region within synthetic peptide: VYSNASLQLONVTDLGTYTLHIKRGDET REEIRHFTFTLYLETPKPY, corresponding to amino acids 101-150 of Human PSG9</td>
</tr>
<tr>
<td>Talin</td>
<td>5</td>
<td>1</td>
<td>270</td>
<td>ab57758</td>
<td>Not established</td>
<td>Mouse monoclonal</td>
<td>SALSVVQNLKDLOEVEKAAARDGKLKPLP GETMEKCTQDLGNSGTAVSSAIAQGLGEVA QGNENYAGIARDVAGGLRSLAQAARGVA ALTSDPAVQA, corresponding to amino acids 1052-1150 of Human Talin.</td>
</tr>
<tr>
<td>Vinculin</td>
<td>7.5</td>
<td>1</td>
<td>130</td>
<td>ab18058</td>
<td>1/1500 (0.5% milk)</td>
<td>Mouse monoclonal</td>
<td>Full length native protein (semi-purified) (Human).</td>
</tr>
</tbody>
</table>
5.3.4 Results

5.3.4.1 Optimisation of Western Blotting

Optimal antibody conditions and dilutions were investigated for each candidate protein primary antibody. The percentage acrylamide gel for ideal separation, blocking buffer, primary and secondary antibody dilution, and time and temperature for incubation were investigated and optimised. Protein was detected at the correct molecular weight for four of the candidates investigated. Talin and Filamin could not be identified; most probably due to the size of the proteins.

5.3.4.1.1 Assessment of Reproducibility of Gel Loading

Ponceau staining was carried out after blotting and imaging of each gel and assessed; an example of the staining is shown in Figure 5.3.2. Relative quantification of protein bands was carried out on images of Ponceau stained membranes and indicated that plasma loading was consistent (data not shown).

5.3.4.1.2 Antibody Specificity - PSG9

During optimisation, specificity of the PSG9 antibody was found to be questionable. PSG9 should not be present in plasma from non-pregnant healthy individuals as it is primarily produced by the placenta; however signal was detected by western blot of plasma from pregnant and non-pregnant subjects (Figure 5.3.3). A Blast search of the PSG9 antibody epitope was carried out (Altschul et al. 1997). The search revealed the epitope to which the antibody was raised had a high level of homology with PSG11. The epitope also showed some homology with several other PSG proteins, but not to proteins of other families. Ponceau staining indicated a larger volume of protein was present at the same molecular weight as the signal detected and may account for the condensed signal pattern. Blots were also investigated using primary antibody only and secondary antibody only and gave no signal (data not shown). It was assumed the signal was a result of a non-specific reaction with an unknown protein.
Figure 5.3.2  Ponceau S staining of fibulin blots shows loading of plasma is consistent. Plasma proteins were separated by SDS-PAGE gel and target proteins identified by western blotting as previously described. After imaging of target protein signals, membranes were incubated in Ponceau S stain and re-imaged. Staining was assessed by eye and where possible relative quantification carried out on prominent protein bands.
Specificity of the PSG9 antibody was questionable as signal was also detected on analysis of plasma from healthy non-pregnant individuals. Plasma proteins were separated by SDS-PAGE gel and PSG9 identified by western blot. Four plasma samples from pregnant women and three plasma samples from non-pregnant individuals (two female, one male) were used. A pooled sample consisting of plasma from each of the individual samples was also run for both pregnant and non-pregnant samples. PSG9 is not thought to be present in the plasma of healthy individuals who are not pregnant as it is primarily produced by the placenta. A large volume of protein was present at the same molecular weight as the signal identified. The origin of the signal in non-pregnant plasma is unknown.
5.3.4.2 Validation by Western Blot

Plasma levels of fibulin, platelet basic protein and vinculin were all quantified by western blot (Figure 5.3.4). Quantification data were acquired for 12 early onset pre-eclampsia and control samples for each protein with the exception of platelet basic protein. One blot for platelet basic protein did not develop and only 8 early onset pre-eclampsia and control samples could be quantified. The difference between early onset pre-eclampsia and control plasma levels was found to be statistically significant using the Mann Whitney U test for platelet basic protein only (median; 0.0713 and 0.0253 respectively, p=0.0379).

5.3.4.3 Validation by ELISA

Imunoassays were also performed for two proteins; platelet basic protein and vinculin. For vinculin the assay was carried out on all 12 early onset pre-eclampsia and control samples. Each sample and standard was assessed in triplicate and the average CV for the assay was 56.3%. Due to the large value for the inter-assay CV no conclusions could made from this assay (data not shown).

Assays were carried out on 10 of the 12 early onset pre-eclampsia and control plasma samples for platelet basic protein. Each sample and standard was assessed in triplicate to determine the technical variability of the assay; the average CV for samples and standards was 16.2%. Plasma levels of platelet basic protein were significantly increased in early onset pre-eclampsia samples compared to control samples (Mann Whitney U test, median 1377 and 801 ng/ml respectively, p=0.0355; Figure 5.3.5).
Plasma levels of platelet basic protein were increased in early onset pre-eclampsia samples compared to control samples by western blotting. Plasma samples were applied to an SDS-PAGE gel of appropriate percentage acrylamide and proteins of interest identified by western blotting using antibodies detailed previously. For fibulin and vinculin; 12 early onset pre-eclampsia and 12 control samples were applied to 3 SDS-PAGE gels and quantification carried out. For platelet basic protein, only 8 pre-eclampsia and 8 control samples could be quantified. The difference between early onset pre-eclampsia and control plasma levels was found to be statistically significant using the Mann Whitney U test for platelet basic protein only (p=0.0379, as indicated by *).
Figure 5.3.5  Plasma levels of platelet basic protein were increased in early onset pre-eclampsia samples compared to control samples using commercially available ELISA assays. Early onset pre-eclampsia and control plasma samples (both n=10) were assessed using manufacturer’s protocols provided with the ELISA kit. Samples and standards were measured in triplicate to assess technical variability. The average CV for the platelet basic protein assay was 16.2%. Plasma levels of platelet basic protein were significantly increased in early onset pre-eclampsia samples compared to control samples (Mann Whitney U test, median 1377 and 801 ng/ml respectively, p=0.0355).
Summary of Western Blot and ELISA Validation of iTRAQ Results

Changes in plasma levels of six proteins were identified in pooled plasma samples using iTRAQ from women who developed early onset pre-eclampsia compared to women with uncomplicated pregnancies. These proteins were assessed in each plasma sample contributing to the pool by western blot and ELISA as summarised in Table 5.3.2. Changes in abundance of platelet basic protein measured by both western blot and ELISA were found to be significant using Mann Whitney U (both p<0.05, indicated by * in Table 5.3.2).

Table 5.3.2 Summary of changes in plasma proteins identified in early onset pre-eclampsia samples by iTRAQ and validated by western blot and ELISA. Changes in protein abundance were validated for platelet basic protein by both western blot and ELISA. Changes found to be significant using Mann Whitney U are indicated by *.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Plasma levels in early onset pre-eclampsia samples compared to controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>iTRAQ</td>
</tr>
<tr>
<td>Fibulin</td>
<td>Decreased</td>
</tr>
<tr>
<td>Filamin</td>
<td>Increased</td>
</tr>
<tr>
<td>Platelet Basic Protein (NAP-2)</td>
<td>Increased</td>
</tr>
<tr>
<td>Pregnancy-specific beta -1-glycoprotein 9</td>
<td>Increased</td>
</tr>
<tr>
<td>Talin</td>
<td>Increased</td>
</tr>
<tr>
<td>Vinculin</td>
<td>Increased</td>
</tr>
</tbody>
</table>
5.3.5 Discussion

Proteins changed in abundance which were identified during discovery phase using iTRAQ, require validation using alternative methods. Several proteins were identified as changed and a small number were chosen for validation.

5.3.5.1 Validation using Antibodies

Antibodies were purchased for the six candidate proteins and conditions optimised for western blotting. Platelet basic protein was identified using iTRAQ, however, antibodies to platelet basic protein were not available and NAP-2 was used as a surrogate. NAP-2 is a cleavage product of platelet basic protein consisting of amino acid residues 59-128. Peptides were identified and quantified outside this region by iTRAQ indicating that the protein identified in the discovery phase investigation was platelet basic protein rather than the cleavage product. Platelet basic protein is stored in the alpha-granules of platelets in an inactive form. The inhibitory N terminus prevents activity and is cleaved by cell surface cathepsin G producing a number of truncated active forms including NAP-2 (Brandt et al. 1991; Ehlert et al. 1998). Quantification of NAP-2 as a surrogate for platelet basic protein is therefore reliant on the activity of cathepsin G, the activity of NAP-2 and the rate of degradation of the truncated protein. The plasma levels of NAP-2 may not be representative of the plasma levels of platelet basic protein.

A polyclonal antibody raised against highly pure recombinant human NAP-2 was obtained from abcam® for the western blot. Polyclonal antibodies recognise multiple epitopes which can result in more tolerant and robust, but less specific, identification of the antigen. In this case, the lack of specificity may be an advantage; if the epitope recognised is common to multiple truncated forms of platelet basic protein (as suggested by the double bands present on the blot) this may give a better indication of the abundance of full length platelet basic protein as opposed to the single NAP2 cleavage product. The antibody used by the ELISA kit is not specified; however none of the cleavage products of platelet basic protein are included in the list of proteins to which it does not cross react. It can therefore be assumed that the ELISA kit also identifies multiple cleavage products of platelet basic protein.

Signal was identified at the correct molecular weight for four proteins; however the identity of the protein resulting in that signal is dependent on the specificity of the antibody. Further investigation of the PSG9 antibody using plasma from women who had uncomplicated pregnancies, healthy males and females who were not pregnant also resulted in signal from the antibody. PSG9 is primarily produced by the placenta (Grudzinskas et al. 1977) and is expressed by differentiated trophoblasts as early as 18 days gestation (Grudzinskas et al. 1977) with expression higher in earlier pregnancy than at term (Zhou et al. 1997). Although the function of the PSG family is unknown they have been
shown to induce the proangiongenic factors TGF-β1 and VEGF in the placenta (see 1.1.7.4.1)(Wu et al. 2008) and low levels of PSGs are associated with complicated pregnancies (Towler et al. 1977; Karg et al. 1981; Arnold et al. 1999; Silver et al. 1999). Deregulated expression of PSG9 is also associated with cancer cells (Salahshor et al. 2005). All non-pregnant plasma samples investigated were obtained from healthy individuals and so expression of PSG9 was not expected. There is a high level of homology between the eleven members of the PSG family (Bayhan et al. 2005) which have a characteristic N terminus homologous to the immunoglobulin variable region (Bayhan et al. 2005). The antibody epitope may not be specific to PSG9, resulting in cross reaction with other PSG proteins. The presence of PSG proteins in plasma of individuals who were not pregnant, however, is still not expected. PSG9 has a predicted molecular weight of 58kDa, which is similar to the molecular weight of serum albumin (67kDa). Serum albumin is a highly abundant plasma protein; Ponceau S staining of membranes indicated a large volume of protein at this molecular weight range. The compressed protein band identified using the PSG9 antibody may be a result of the large volume of protein in this region of the gel and may also account for the cross-reaction of the antibody. There are no published accounts of use of this antibody for western blotting.

Multiple methods of confirming identification of a protein identified by western blot have been developed. The simplest method uses a positive control run alongside samples of interest. The molecular weight of protein identified in the positive control and the sample can then be compared. A positive control may either be recombinant protein or a sample in which the protein of interest is known to be highly expressed. Recombinant protein was not available for PSG9. The hepatocellular carcinoma cell line, HepG2, was recommended as a positive control for the antibody. Placental lysate was also unsuccessfully tested as a positive control. In comparison to whole plasma, which is highly complex, tissue and cell lysates are relatively simple and so may not be a representative positive control in this case, particularly where expression of the protein of interest is of a similar molecular weight to albumin which is present at lower concentration in either lysate. A peptide blocking assay may also be carried out to assess specificity of the antibody. The antibody is incubated with an excess of peptide that corresponds to the epitope recognised by the antibody and so the antibody becomes ‘blocked’. Blocked and un-blocked antibodies are then applied to western blots side by side. If the antibody is specific to the epitope of the protein of interest no signal should be present using the blocked antibody. This technique again requires peptide to be available and so was not applicable to several of the proteins investigated here.

A method utilising MS for identification of protein transferred onto nitrocellulose has also been developed (Luque-Garcia et al. 2008) and was unsuccessfully trialled during this investigation. A western blot was carried out and the location of signal noted, the nitrocellulose was stripped of antibodies and the section indicated to contain proteins of interest excised. Nitrocellulose was processed according to the Blotting and Removal of
Nitrocellulose method described by Luque et al. (Luque-Garcia et al. 2008). Proteins were digested while bound to nitrocellulose stripped of antibodies and peptides extracted. Peptides were analysed using LC-MS/MS and proteins identified by database searching. The method was applied to western blots of several proteins of interest. The method was unable to identify the protein of interest; however several high abundance proteins were identified. With optimisation, this method may be a useful tool for confirmation of proteins bands identified by western blot in the absence of a suitable positive control or blocking peptide.

5.3.5.1.1 Western Blotting

Western blotting is a relatively simple method for the identification and quantification of proteins. The method, however, requires proteins to be soluble which makes identification of extremely hydrophobic proteins difficult (Rabilloud 1998). Proteins at extreme molecular weights are also difficult to identify and quantify by western blot. Obtaining sufficient separation of proteins in the polyacrylamide gel and subsequent transfer to the nitrocellulose membrane of large proteins (greater than 200kDa) is difficult. Two candidate proteins, talin and filamin, could not be identified by western blot; thought to be due to the large size of the proteins making separation and transfer to the nitrocellulose membrane difficult. Even in the low percentage acrylamide gels used here, high molecular weight proteins move slowly through the pores of the polyacrylamide gel despite the high charge state conferred by the size of the protein. Separation and transfer requires more time than smaller proteins and the polyacrylamide gel and nitrocellulose must be maintained at low temperature to prevent damage. One published study has used the filamin antibody for western blotting, however this was applied to cell cytoskeletal fractions which are far less complex than whole plasma (Jackson et al. 2008); no reports of use of the talin antibody could be found.

The blots used for quantification were subsequently stained with Ponceau S to assess equal loading of plasma onto the polyacrylamide gel and the complete and equal transfer of proteins to the nitrocellulose membrane. Equal volumes of plasma samples were used across all three of the gels used to quantify one protein. As previously discussed, standardising samples by volume has several inherent disadvantages (see 5.1.5.1) but is the most relevant to plasma samples which will ultimately be assessed in a clinical environment. Ponceau S staining is not considered to be completely quantitative; however this method was the most suitable loading control available. Actin and tubulin are commonly used as loading controls for western blots of cell lysates where the quantity of protein loaded can be standardised. As the levels of plasma proteins can vary widely between individuals, evaluation of loading is not possible using a single protein in plasma. Ponceau S staining indicated that loading of plasma was approximately equal across all gels.
Each early onset pre-eclampsia and control sample was quantified only once during this investigation; reproducibility of the method could therefore not be assessed. Early onset pre-eclampsia and control samples were also processed without blinding or randomisation which may introduce user and batch bias (Cairns 2011). Three gels were run to quantify all 12 samples for each protein. All three gels were processed together to avoid variation in the processing conditions but this may also introduce bias. Exposure of the blot using imaging software was not regulated which resulted in overexposure of several blots. The use of a digital photoimaging system for acquisition of the blot image, however, reduces the error introduced by the need to digitise traditionally used X-ray films for analysis (Gassmann et al. 2009). Quantification was also carried out manually against an unrelated reference sample which can also introduce user error and bias in the selection of protein bands to quantify. Standardisation of densitometry and reporting of the results is not yet common practice and is highly dependent on the facilities available to each investigation (Gassmann et al. 2009). While the same method was applied to all blots here, this method is highly user dependent, and as a result vulnerable to user error and bias.

A more robust method of quantifying proteins using western blot was also briefly investigated but not reported as due to time pressure this method was not applied to proteins of interest. Samples were blinded, randomised and assessed in triplicate to avoid user bias on analysis. A pooled reference sample consisting of equal amounts of all plasma samples was applied to the gel in incremental volumes. A pooled reference sample is more appropriate than the external reference used for the blots quantified here. Where previously, the averaging of samples on pooling is a disadvantage, here it is advantageous, as the reference sample is representative of the samples investigated and should fall well within the range of relative quantification. The reference sample was also used to assess the linearity of the signal response as proof that the blot is applicable for quantitation. A standardised method for image acquisition was optimised and subsequently applied to all blots to ensure images were consistently exposed allowing blot to blot comparison. Quantification however, was still carried out by manually defining the bands for comparison and therefore this aspect of the method is still user dependent. This aim of this optimised method was to minimise the introduction of user bias and error which was prevalent in the method previously used. This ensures a greater confidence in changes in protein abundance identified using this method.
5.3.5.1.2 ELISA

The ELISA is the 'gold standard' of clinical protein quantitation; the assay is robust, reproducible and routinely used. An ELISA is simple to run, high throughput and unlike western blots, suffers from less environmental day to day variation. ELISA is therefore ideally applied to routine diagnostic analysis in a clinical setting. Unlike western blotting an ELISA can be used for direct absolute quantification by use of synthetic standards run alongside samples, which also serve as an indication of the reproducibility and linearity of the assay. An ELISA is also carried out without immobilisation of proteins; while high abundance proteins are still present there is no competition with lower abundance proteins of similar molecular weight. Larger volumes of plasma may also be used improving detection of low abundance proteins. The specificity of a capture-based ELISA, as used here, is similar to that of a western blot, as the protein of interest is identified by one antibody. Specificity can be greatly improved by used of a 'sandwich' ELISA, which uses two antibodies identifying two epitopes of the protein of interest for both capture and identification. While a western blot is easier to optimise and useful for an initial investigation of a protein, an ELISA is more difficult to optimise but more specific and more useful in the long term where investigation of a large number of samples is required.

On the basis of the western blot data and availability of commercial kits, NAP2 and vinculin were also assessed by ELISA. Commercially available kits were used for both and the reproducibility for both samples and controls assessed. The CV for triplicate analysis of NAP2 was higher than stated by the kit but still within reasonable limits (<20%). No CV was reported for the vinculin kit; the CV for triplicate analysis of sample and controls was unacceptably high and no conclusion could be made from the analysis. The lack of reproducibility may be due to irreproducible sample preparation and application of the assay, a lack of protein in the sample or homogeneity in the replicate samples, insufficient binding of protein to the well surface or the antibody lacking in specificity. As previous assays run under the same conditions were reproducible and replicate samples produced as one aliquot it is more likely the lack of reproducibility is a function of the kit used. An ELISA kit for vinculin was purchased on the strength of the western blot data. Signal was identified in all samples although did not indicate a significant change in protein concentration between case and control samples. The ELISA kit, however, does not use the same antibody and therefore may not recognise the same ‘vinculin’ protein. It is possible the antibody for either the western blot or ELISA or even both is not specific. Several published studies have used the abcam® antibody for western blotting, although only in cell lysate and not plasma (Hahn et al. 2009; Raya et al. 2009; Ceccaldi et al. 2011) while there is no published use of the ELISA kit. A competition assay can be carried out to assess specificity in a similar manner to the peptide blocking assay used with a western blot; however, this was not used here due to a lack of specific antigen.
5.3.5.2 Validation of iTRAQ Discovery Phase

As previously discussed there are several weaknesses to the iTRAQ method used for the discovery phase of this investigation, including pooling of samples, averaging of ratios calculated at the peptide level and identification of a high level of false positive protein changes (see 5.1.5). MRM was used to investigate the same sample set on an individual basis; however, more robust methods of quantification are still required for proteins indicated to be of particular interest by the MRM (see 5.1.1). Changes in protein abundance identified in the discovery phase require validating using alternative methods to ensure confidence in the change. Antibody methods are relatively quick and easy methods for validation, however as previously discussed, require careful choice of antibodies and may not be applicable to all proteins. Antibody methods have been used to validate changes in protein levels in pre-eclampsia identified by iTRAQ and other proteomic methods.

Atkinson et al. identified an up-regulation in the basic isoform of Apo E and a down-regulation of the acidic isoform by DIGE analysis of depleted serum (Atkinson et al. 2009). This was confirmed by 2D western blotting of the same 12 depleted serum samples; the basic isoform of Apo E could only be identified in serum from 11 women with pre-eclampsia and only 3 healthy controls. These methods are very similar and were carried out on the same sample set making confirmation of the initial result unsurprising. When Apo E and Apo C II were measured by multiplexed immunoassay in 18 alternative, undefined, serum samples from women with pre-eclampsia and 18 healthy controls, no significant difference was identified. Confirmation of the change in both proteins could not be made using an unrelated method and different samples, suggesting this change was an artefact of the discovery method, or the sample set and may not be applicable in the wider population.

Blumenstein et al. investigated six proteins identified as changing between immunodepleted plasma taken at 20 weeks gestation from women who subsequently developed pre-eclampsia (n=27), pre-eclampsia with SGA (n=12) and healthy controls (n=57) by DIGE 2D gel electrophoresis (Blumenstein et al. 2009a). Western blot was used to confirm these changes in the same plasma samples without immunodepletion; significant differences could be identified for only two of the six proteins. Blumenstein et al. also identified differential processing of vintronectin and high molecular weight kininogen by DIGE analysis of depleted plasma samples from pregnant women with and without pre-eclampsia (both 20 weeks and 33-36 weeks gestation; n=6) (Blumenstein et al. 2009b). Western blots were carried out on the same plasma samples without immunodepletion and also on whole plasma samples from women who developed pre-eclampsia with SGA later in pregnancy and healthy controls (both 20 weeks gestation n=4). Western blots confirmed the changes identified by DIGE but not for the pre-eclampsia with SGA sample set. This contradiction may be a result of the two samples sets used.
Blankley et al. used immunoassays to confirm changes in three proteins identified using iTRAQ in pooled immunodepleted plasma taken at term from women who developed pre-eclampsia compared with healthy controls (both n=23) (Blankley et al. 2009). Immunoassays were carried out on the subsection of the same plasma samples used for the original observation without immunodepletion. Changes could only be confirmed for one protein; unacceptable levels of variance were obtained using the commercial ELISA kits for the remaining two proteins. Auer et al. also used iTRAQ; the method was applied to pooled plasma samples obtained from women who developed pre-eclampsia (n=7), IUGR (n=7), pre-eclampsia with IUGR (n=7) and healthy controls (n=6) at the time of diagnosis of disease (Auer et al. 2010). SERPINA 3 was identified as differentially regulated and confirmed by western blot; both demonstrated an increase in expression in case samples compared to healthy controls. The same individual immunodepleted plasma samples were used for discovery and validation and samples were normalised for protein concentration. The investigations discussed above used equal volumes of plasma for validation. Normalisation for protein volume in plasma is subject to the large variance in protein concentration between individuals. This variation is present even in immunodepleted plasma and may skew any change in abundance measured in concentration normalised samples. The initial iTRAQ investigation of term plasma also identified an increase in CRP in samples from pregnancies complicated by pre-eclampsia only (Auer et al. 2010). This was confirmed by automated ELISA in another set of plasma samples; a significant increase in CRP was identified in plasma samples from women who developed pre-eclampsia (n=7) compared to healthy controls (n=10).

The investigations discussed above encountered similar difficulties to the validation carried out here; proteins could not be identified and/or quantified, changes in abundance identified during discovery phase were no longer significant when a different method or sample set was used and commercially available kits for quantification lacked sufficient reproducibility. The change in abundance identified during the discovery phase may be an artefact of the sample processing or method used and so a different method should be used for validation to increase confidence in the change. Where pooled samples are used for discovery, each sample should be investigated individually for validation; the change in abundance identified may be a result of one sample contributing to the pool with abnormal concentration levels. Where immunodepleted plasma is used for discovery, whole plasma should be used for validation. Should the change in protein abundance be an artefact of the immunodepletion method used, validation using immunodepleted plasma will only confirm this artefact. This does, however, result in some difficulty in identifying low abundance proteins as demonstrated here by the difficulty in identifying PSG9, which has a similar molecular weight to albumin.
Six proteins of interest were identified as changed in abundance by iTRAQ between pooled plasma acquired from women who subsequently developed early onset pre-eclampsia and pooled plasma from women with uncomplicated pregnancies. Both western blot and ELISA analysis of the same plasma samples on an individual basis confirmed the changes in platelet basic protein identified by iTRAQ. Platelet basic protein, using antibodies to NAP2 as a surrogate, was shown to increase in early onset pre-eclampsia plasma samples compared to control samples. Although five other proteins were investigated, changes in the abundance for these proteins could not be validated. Changes in the abundance of platelet basic protein were validated in individual whole plasma samples using a different method to the discovery phase. Orthogonal validation increases the confidence in the protein change being real, and is further increased by validation using a different sample set. Ultimately validation using antibody methods is most useful as these methods are commonly used in clinical laboratories. The end goal of any biomarker discovery investigation is therefore production of an ELISA which can be easily introduced into routine diagnostic analysis.

5.3.6 Conclusion

The aim of this investigation was to validate changes in plasma levels of several proteins between samples taken from women who developed early onset pre-eclampsia and women with uncomplicated pregnancies. Six proteins of interest were identified as changed in abundance by iTRAQ in pooled plasma samples. Antibodies were acquired for all six proteins; however talin and filamin could not be identified and specificity of the PSG9 antibody was questionable. Western blots were carried out on individual plasma samples for three proteins: fibulin, platelet basic protein, and vinculin. A significant change was identified for only one protein; platelet basic protein was identified as increased in early onset pre-eclampsia plasma samples compared to control samples. Commercial ELISA kits were also acquired for two proteins; platelet basic protein and vinculin. The platelet basic protein ELISA was highly reproducible and indicated an increase in platelet basic protein in early onset pre-eclampsia samples compared to control samples. Reproducibility of the vinculin ELISA kit was not sufficient for conclusions to be drawn from the assay.

In conclusion, platelet basic protein was identified as increased in abundance in plasma from women who developed early onset pre-eclampsia compared to healthy controls by both western blot and ELISA, using NAP2 as a surrogate. This investigation validated changes identified during the discovery phase using iTRAQ. Validation of this protein as a predictive marker for pre-eclampsia, however, requires significant further investigation using larger numbers of samples.
5.4 Identification and Validation of Changed Plasma Proteins in Early Pregnancy in Women Who Develop Pre-eclampsia:

Summary

Plasma collected at 15 weeks gestation was acquired from the SCOPE study. Plasma samples (each n=12) were obtained from women who subsequently developed late onset pre-eclampsia, early onset pre-eclampsia and two distinct groups of women with healthy pregnancies matched for gestational age only. A proportion of all four pooled samples were also combined to produce a ‘superpool’ sample for use as a reference. Two superpool and pre-eclampsia pooled samples were processed and iTRAQ labelled. Each control pool was processed once resulting in a total of 8 labelled samples. Two parallel workflows were used to process pooled plasma samples and resulted in a total of three distinct sets of data. Plasma samples were depleted of abundant proteins using the IgY14-SuperMix System and MARS 14 immunodepletion systems. Depleted plasma was labelled with iTRAQ reagent and analysed using by LC-MS/MS using a QTAR and 5800. Data were interrogated using ProteinPilot 3.0 searching against the Human protein database from the IPI (version 3.59). The relative quantification ratios for iTRAQ reporter ions were calculated by ProteinPilot against the homologous reference superpool.

Protein identifications were filtered to produce a list of proteins considered to be identified with high confidence; a total of 502 high confidence proteins were identified. Proteins changed in abundance between early onset pre-eclampsia and control samples were identified using a two stage method of analysis. This method was developed using proteins identified in all three datasets and required proteins to be consistently quantified between paired observations and changed in abundance outside the range of protein expression in healthy controls. A total of 32 proteins were identified as changed between late onset pre-eclampsia and control. A total of 113 proteins were identified as changed between early onset pre-eclampsia and control samples. Only a small number of proteins identified as changed in early onset pre-eclampsia were selected for validation for practical reasons.

Proteins selected for validation were chosen by several criteria; proteins were identified in at least two datasets by three or more peptides and were not immunodepleted. Validation was carried out on each plasma sample individually. Proteins were validated by multiple reaction monitoring; a label free method was developed for relative quantification. Proteins were also validated by western blot and ELISA. One protein was identified as increased in early onset pre-eclampsia samples compared to control samples in all methods. Platelet basic protein was identified and found to be changed in abundance in all three iTRAQ datasets (Figure 5.4.1). Validation by western blot, ELISA and relative quantification by MRM also identified platelet basic protein as changed in abundance in early onset pre-eclampsia samples (Figure 5.4.1). When plasma samples were assessed individually by
validation methods, platelet basic protein was consistently identified at higher concentration in a small number of samples compared to the remainder of the sample group. Three samples numbered 135, 1098 and 1297 were quantified at a higher concentration than the remaining samples in the group by all methods. Sample number 2706 was quantified at a higher concentration by MRM and western blot but not by ELISA. Conversely sample number 2328 which was quantified at high abundance by ELISA was quantified at lower concentration by MRM and western blot. This result may be an artefact of the ELISA method. The ratio of platelet basic protein in early onset pre-eclampsia samples compared to control samples was calculated for all methods of analysis (Table 5.4.1). The ratio was calculated from an average value for platelet basic protein in early onset pre-eclampsia and control samples for validation methods. Ratios of change were similar for all method of analysis indicating consistent quantitation across all methods.
Summary of platelet basic protein abundance in early onset pre-eclampsia and control samples calculated by iTRAQ, relative MRM, western blot and ELISA. Discovery phase investigation using iTRAQ identified platelet basic protein at a higher abundance in pooled plasma samples from women who developed early onset pre-eclampsia compared to women who had uncomplicated pregnancies. This change in abundance was confirmed by investigating the plasma samples contributing to the pool individually using MRM, western blot and ELISA. When investigated individually the same samples; sample numbers 135, 1098, 1297, 2706 and 2328, were identified at a higher abundance than the remaining samples.
Table 5.4.1  Summary of relative quantification ratios calculated for expression of platelet basic protein in early onset pre-eclampsia and control samples by iTRAQ, relative MRM, western blot and ELISA. Relative ratios for iTRAQ are calculated for pooled early onset pre-eclampsia plasma samples compared to pooled control samples. The same samples contributing to the pools were subsequently assessed by MRM, western blot and ELISA. Relative ratios for MRM and antibody methods are calculated from an average value for all samples assessed.

<table>
<thead>
<tr>
<th>Method</th>
<th>Disease to Control Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ITRAQ</strong></td>
<td></td>
</tr>
<tr>
<td>IgY 14-SuperMix-QSTAR</td>
<td></td>
</tr>
<tr>
<td>117:119</td>
<td>1.36</td>
</tr>
<tr>
<td>118:119</td>
<td>1.55</td>
</tr>
<tr>
<td>117:121</td>
<td>1.41</td>
</tr>
<tr>
<td>118:121</td>
<td>1.61</td>
</tr>
<tr>
<td>IgY 14-SuperMix-5800</td>
<td></td>
</tr>
<tr>
<td>117:121</td>
<td>1.39</td>
</tr>
<tr>
<td>118:121</td>
<td>1.35</td>
</tr>
<tr>
<td>117:119</td>
<td>1.27</td>
</tr>
<tr>
<td>118:119</td>
<td>1.23</td>
</tr>
<tr>
<td>MARS 14-5800</td>
<td></td>
</tr>
<tr>
<td>117:121</td>
<td>1.39</td>
</tr>
<tr>
<td>118:121</td>
<td>1.53</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>1.41</td>
</tr>
<tr>
<td><strong>MRM</strong></td>
<td></td>
</tr>
<tr>
<td>GTHCNQVEVIATLK</td>
<td>(n=12) 1.50</td>
</tr>
<tr>
<td>NIQSLEVIGK</td>
<td>(n=12) 1.52</td>
</tr>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
</tr>
<tr>
<td>Western Blot</td>
<td>(n=8) 2.84</td>
</tr>
<tr>
<td>ELISA</td>
<td>(n=10) 1.91</td>
</tr>
</tbody>
</table>
6 Discussion

It is well established that there are ‘circulating factors’ which cause endothelial dysfunction present in the plasma of women who develop pre-eclampsia (Myers et al. 2005). These factors consist of synergistic proteins and are present in plasma prior to development of the maternal syndrome (Myers et al. 2007). Multiple proteins have been identified as changed in abundance in the plasma of women who develop pre-eclampsia compared to women who have uncomplicated pregnancies (see 1.1.7.4)(Chappell et al. 2002; Tjoa et al. 2004; Carty et al. 2008). Identification of proteins changed in abundance in early pregnancy with sufficient predictive value to identify women who will develop pre-eclampsia later in pregnancy would be beneficial. Combining measurement of changed proteins with other risk factors may allow identification of women at high risk of developing pre-eclampsia prior to a clinical diagnosis (North et al. 2010). As yet there is no screening test which is applicable to an otherwise healthy, nulliparous population. Nulliparous women are particularly at risk of developing pre-eclampsia as a previous uncomplicated pregnancy has been demonstrated to be protective (Bhattacharya et al. 2009). Previous obstetric history, the most significant obstetric risk factor which informs the provision of antenatal care, is also absent in nulliparous women. Identification of nulliparous women at most risk would allow stratification of antenatal care, targeting women who would most benefit from extra monitoring throughout pregnancy. Identification of proteins with predictive value may also inform development of potential therapies to prevent and treat pre-eclampsia. This study aimed to identify novel predictive markers for pre-eclampsia using hypothesis generating proteomic methods applied to plasma samples obtained prior to clinical diagnosis.

Samples were collected at 15 weeks gestation as part of the SCOPE study; at the time the SCOPE study was designed this gestation coincided with a routine antenatal visit and optional blood test for Down’s syndrome screening. Since the inception of the study the timing of this blood test has changed and it is now offered earlier in pregnancy. There is now, no routine antenatal visit or blood collection at 15 weeks gestation however this time point still remains relevant. A predictive test for pre-eclampsia developed at 15 weeks gestation would also allow sufficient time for monitoring and potential intervention. Most women have booked for antenatal care by this stage in the pregnancy and a screening test could easily be incorporated into routine antenatal care. Changes in expression of plasma proteins have been documented as early as 7 weeks of gestation and therefore, there is potential to identify proteins with predictive value at this gestation (Salomon et al. 2003).

Plasma samples were acquired from women who subsequently developed late and early onset pre-eclampsia. The SCOPE study aimed to recruit women at low risk of developing pregnancy complications; women at high risk of developing pre-eclampsia or other pregnancy complications (small for gestational age or spontaneous preterm birth) during pregnancy were excluded from the study (North et al. 2010). Each group of pre-
Eclampsia samples were paired with a set of healthy control samples matched for gestational age at the time of sample collection. The control samples used were taken from healthy women with uncomplicated pregnancies. Samples used as controls were not matched for factors such as age, BMI, or family history of pregnancy complications (although the data was collected), all of which are risk factors for pre-eclampsia (Duckitt and Harrington 2005). This study aimed to identify global changes in protein expression which could potentially be used to predict pre-eclampsia in low risk women. Close matching of control samples reduces the range of biological variability represented by the control group and so limits the population to which a predictive test would be applicable.

Hypothesis driven studies have so far been unsuccessful in identifying plasma proteins with predictive value in nulliparous women at risk of pre-eclampsia later in pregnancy. Hypothesis generating proteomic methods are unbiased and able to identify changes in the expression of multiple proteins simultaneously. These methods are therefore ideally suited to the search for early pregnancy plasma markers predictive of pre-eclampsia. Several plasma proteins have been identified by 2D gel electrophoresis as changed in abundance in women who subsequently developed pre-eclampsia when compared to uncomplicated pregnancies but validation of the changes identified proved inconclusive (Atkinson et al. 2009; Blumenstein et al. 2009a; Blumenstein et al. 2009b). The 2D gel method is highly variable (Corbett et al. 1994; Molloy et al. 2003), and limited in the number, level of abundance and classes of proteins which can be identified (Gygi et al. 2000).

Changes in abundance of plasma proteins in the later stages of pregnancies complicated by pre-eclampsia have been identified using iTRAQ and validated using alternative methods (Blankley et al. 2009; Auer et al. 2010). Despite documented issues (Pierce et al. 2008; Kuzyk et al. 2009; Ow et al. 2009; Shirran and Botting), iTRAQ allows identification of proteins and relative quantification of protein abundance between multiple samples. Relative quantification using iTRAQ is therefore, the most useful and relevant method for this hypothesis generating study. The method was applied to pooled samples allowing comparative phenotypic profiling. Although pooling of samples is controversial in proteomics (Karp and Lilley 2007; Song et al. 2008), it is commonly used by other disciplines (Kendzierski et al. 2005). Applied to investigations such as this, where the aim is to characterise a population, the sample pooling is particularly useful as it allows analysis of a large number of samples within a manageable workload. The disadvantage, however, is that pooled samples suffer from biological averaging and do not represent the actual concentration range of the samples contributing to the pool. Sample pooling also restricts the statistics which can be applied to the investigation; the number of samples required to detect a change in protein abundance would usually be defined by a power calculation. Power, however, cannot be applied to pooled samples and so a total of 12 samples were acquired for each phenotypic group. This number was thought to incorporate sufficient
biological variability for an initial hypothesis generating investigation while still being practical for sample processing and data analysis.

Two workflows were developed to process and label samples with iTRAQ 8plex reagent. There is no standardised method for sample processing for iTRAQ quantification in proteomics. Two immunodepletion methods were used here; each method has advantages and disadvantages. The sample processing method used can affect the data acquired and so must be carefully chosen for the qualities required for the work in hand. The IgY 14-SuperMix system removes more high and medium abundance proteins, which lead to identification of more low abundance proteins which are potentially of greater interest. This use of the IgY 14-SuperMix system however, lead to loss of an unacceptably high number of samples. The MARS 14 immunodepletion system proved to be the most reproducible method for depleting plasma and so was used to process samples during the validation phase. Two LC-MS/MS methods were also used for analysis of prepared samples, resulting in three distinct datasets. Although the choice of LC-MS/MS is more likely to be constrained by instrument availability, the system used can also affect the data acquired. Comparison of data from each LC-MS/MS system confirmed the established underestimation of relative ratios by iTRAQ (Pierce et al. 2008; Ow et al. 2009); this issue is not unique to this method and has been demonstrated in several other quantitative proteomics methods (Ong et al. 2002; Old et al. 2005; Parker et al. 2009). Multiple analyses of one sample set using iTRAQ are not commonly carried out and so the three datasets were carefully compared to ensure data were comparable. The iTRAQ method is designed for unbiased identification and quantification of multiple proteins from one analysis of a small number of samples and multiple analyses is therefore not deemed necessary. It is clear from this investigation however, that multiple analysis of a sample set is beneficial as it increases the number of proteins which can be identified and also the confidence in quantification.

There is also no standardisation in analysis of proteomics data and so a method to identify proteins changed in abundance between samples was developed. The method developed relied upon careful experimental design, allowing limits for reproducibility of sample quantification and the range of biological variation of the control samples to be defined. Where a protein was identified as changed in abundance in more than one dataset, there was greater confidence in the change than proteins identified as changed in only one dataset. Despite the rigorous method developed, approximately a fifth of the total proteins identified were found to change in abundance suggesting the analysis method was inclusive rather than exclusive. This investigation used iTRAQ to identify plasma proteins changed in abundance at 15 weeks of gestation between women who subsequently develop early onset pre-eclampsia and women with healthy uncomplicated pregnancies. As the relative quantification ratios calculated using iTRAQ may not represent the natural changes in protein abundance due to biological averaging and underestimation of iTRAQ ratios, rigorous validation of the data obtained is crucial.
A small number of proteins identified as changed were selected for validation based on the confidence of the identification and quantification. This shortlist is inherently biased; however, the aim was to identify the most promising candidates for further investigation. Validating large numbers of candidate proteins identified by discovery investigations such as this, has been identified as a bottleneck in the search for biomarkers using proteomics (Keshishian et al. 2007; Paulovich et al. 2008; Addona et al. 2009; Whiteaker et al. 2009). There is currently no accessible method which can efficiently quantify hundreds of proteins. Until this bottleneck can be addressed the high throughput advantage to iTRAQ is somewhat lost, as the large number of proteins identified as changed cannot be validated. Proteins of interest must be selected using arbitrarily defined criteria from a longer list which undermines the hypothesis generating nature of the investigation.

A method for relative quantification by MRM was developed here to address this need. The method was designed for relative quantification of large numbers of candidate proteins prior to application of more stringent and robust methods of absolute quantification to the most promising candidates. This method was applied to selected candidate proteins identified here; the increase in abundance of platelet basic protein and PSG9 identified by iTRAQ in plasma samples taken from women who developed early onset pre-eclampsia compared to women with uncomplicated pregnancies was confirmed. Candidate proteins were also validated by western blot and ELISA. The ELISA assay is routinely used by clinical diagnostic laboratories and therefore the most relevant method for validation of potential biomarkers. PSG9 could not be identified using antibodies as there is a high level of homology between the proteins of the PSG family. Until a reproducible and high throughput method, accessible to clinical diagnostic facilities, for identifying PSG9 can be developed, it is of limited use as a biomarker. Antibodies to NAP2 were used as a surrogate to quantify platelet basic protein. It was confirmed that platelet basic protein was increased in abundance in the plasma of women who subsequently developed early onset pre-eclampsia compared to women with uncomplicated pregnancies by both western blot and ELISA.

As previously discussed, several hypothesis generating studies have investigated changes in abundance of plasma proteins in pre-eclampsia using 2D gels (Atkinson et al. 2009; Blumenstein et al. 2009a; Blumenstein et al. 2009b) and MS methods (Blankley et al. 2009; Auer et al. 2010). Each investigation carried out a single discovery phase analysis and identified a subset of proteins changed in abundance for validation. Very few of those proteins were successfully validated on application of an alternative method. The end result for each study was the identification of one, at most two validated proteins. In this investigation, attempts were made to validate six proteins using multiple methods. Only one protein could be validated by more than one method of validation. Although similar to the studies discussed, this is a small return from the discovery phase investigation which identified over a hundred proteins changed in abundance from a total of over five hundred
protein identifications. Investigation of the remainder of the proteins identified as changed in abundance may reveal other candidate biomarkers, however was not possible or practical within the timeframe of this investigation. This highlights the established weakness in the biomarker pipeline when moving between discovery and validation phase investigation.

Validation using multiple methods adds to the confidence in the changes in protein abundance within this sample set. Relative quantification of platelet basic protein was similar across all methods when plasma samples from women who later developed pre-eclampsia were compared to plasma from women with uncomplicated pregnancies. It has been established that the iTRAQ method underestimates relative quantification ratios; calculation of a smaller relative ratio between early onset pre-eclampsia and control samples using this method is therefore expected. As previously discussed, there were difficulties in quantifying platelet basic protein by western blot. While the method indicates differential expression between early onset pre-eclampsia and control samples, the ratio calculated may not be a true reflection of the change in expression between the groups. The ratio of platelet basic protein between early onset pre-eclampsia and control samples calculated by MRM and ELISA is likely to be the most accurate. This value, however, represents a change in expression in a very small sample group which may not be reflective of the general population.

A subgroup of the plasma samples from women who developed early onset pre-eclampsia had a higher concentration of platelet basic protein than the remainder of the samples in the group. While iTRAQ was applied to plasma samples pooled by phenotypic group, validation methods were applied to each individual sample contributing to that pool. The data acquired here by MRM, western blot and ELISA indicates the relative ratio calculated in pooled samples may be skewed by a higher concentration of platelet basic protein in a small number of samples. This highlights the need to validate in individual samples where discovery phase is carried out using pooled samples. This also suggests that the change in abundance in platelet basic protein may be exaggerated in a small subset of women with a particular presentation of pre-eclampsia. It may be that this change occurs as a result of a particular aspect of the pathophysiology of that presentation of pre-eclampsia. Further investigation of platelet basic protein may indicate a mechanistic role in the development of pre-eclampsia. Other studies which identified differentially regulated proteins thought to be predictive of pre-eclampsia found the association was lost when results were corrected for other factors, including BMI (Wolf et al. 2001). A large amount of metadata was collected on each SCOPE participant; over the course of the study several thorough interviews were carried out with participants. Data collected included medical and reproductive history, lifestyle questionnaires and psychological and behavioural assessments (North et al. 2010). An integrated retrospective analysis of potential predictors of pre-eclampsia, including platelet basic protein alongside this metadata is therefore
possible. This may identify sub-populations of women for whom platelet basic protein may be more strongly predictive of risk.

Platelet basic protein belongs to the CXC chemokine family and is primarily located in the cytoplasm and expressed by hematopoietic cells in the bone marrow and trophoblast cells in the placenta (Soleymanlou et al. 2005; Vasarhelyi et al. 2006; Uhlen et al. 2010). Involvement of platelet basic protein in the pathophysiology of pre-eclampsia is highly possible; the coagulation system is in a heightened state of activation (see 1.1.7.4.2) which includes increased platelet activity (Inglis et al. 1982; Janes et al. 1995). Platelet basic protein has been shown to induce transmigration of neutrophils through vascular endothelium and therefore may contribute to the vascular endothelial dysfunction which is characteristic of pre-eclampsia (Schenk et al. 2002). Platelet basic protein is stored in the alpha-granules of platelets and secreted in an inactive form on platelet activation (Brandt et al. 1991; Schenk et al. 2002). It can be further proteolytically processed into 10 active truncated forms including low affinity platelet factor 4, β-thromboglobulin (Holt et al. 1986) and NAP-2 (of varying lengths) (Ehlert et al. 1998). β-thromboglobulin is measured as an indicator of platelet activation (Kaplan and Owen 1981) and has been identified as increased in both blood and urine of women with pregnancies complicated by pre-eclampsia compared to uncomplicated pregnancies (Douglas et al. 1982; Inglis et al. 1982; Leiberman et al. 1985; Leiberman et al. 1988; Hayashi et al. 2002). NAP-2 binds to polymorphonuclear neutrophil granulocytes inducing degranulation and chemotaxis (Walz et al. 1989; Ehlert et al. 2000). Active NAP-2 promotes leukocyte and endothelial cell activation resulting in an inflammatory response (Smith et al. 2006) which is well documented as a feature of pre-eclampsia (see 1.1.7.4.3).

Platelet basic protein has also been identified in other proteomic studies of pre-eclampsia and pregnancy complications using iTRAQ. Platelet basic protein was identified as decreased in the media of term placental explants cultured under hypoxic conditions to simulate the chronic hypoxic insult thought to occur in pre-eclampsia, compared to explants cultured under normoxic conditions to simulate normal pregnancy (Blankley et al. 2010). The decrease in abundance of platelet basic protein identified in placental explant media by Blankley et al. contradicts the increase in abundance identified in early pregnancy plasma in this investigation. Gene expression of the pro-platelet basic protein gene, which encodes platelet basic protein, was also identified as increased in expression by microarray in near term placental samples taken from pregnancies complicated by pre-eclampsia compared to uncomplicated pregnancies (Soleymanlou et al. 2005). The discrepancies measurement of abundance may occur for many reasons; the artificial nature of the explant model does not incorporate the maternal contribution to the expression of platelet basic protein and may not reflect the in vivo expression of the protein.

Platelet basic protein was also identified using iTRAQ, as decreased in abundance in plasma sampled between 11 and 13 weeks gestation of women carrying a Down
syndrome fetus compared to women with normal euploid pregnancies (Kolla et al. 2009). Proteins used to diagnose Down’s syndrome, including PAPP-A and hCG, have also been investigated in early pregnancy for predictive value in women who subsequently developed pre-eclampsia (Muller et al. 1996; Saller and Canick 2008). Both PAPP-A and hCG were unable to discriminate between pregnancies complicated by pre-eclampsia, and other pregnancies complications such as SGA and IUGR (Muller et al. 1996; Ong et al. 2000; Smith et al. 2002; Yaron et al. 2002; Audibert et al. 2005). Platelet basic protein was identified as decreased in abundance in pregnancies with a Down syndrome fetus (Kolla et al. 2009) compared to the increase in abundance in pregnancies complicated by pre-eclampsia identified here. This dichotomy suggests measurement of platelet basic protein may be of use in defining future pregnancy complications. Further investigation however, will be required to determine if platelet basic protein has sufficient predictive value to differentiate between pregnancies complicated by pre-eclampsia and other conditions such as SGA, IUGR and preterm birth.

6.1 Future Work
As a result of this study, there are several areas being considered for future investigation;

- Development of the method for low cost high throughput relative quantification by MRM for application to future biomarker discovery studies.
- Investigation and validation of other plasma proteins identified as changed at 15 weeks gestation prior to development of late and early onset pre-eclampsia.
- Validation of an increase in platelet basic protein at 15 weeks gestation in the plasma of women who developed pre-eclampsia compared to women with uncomplicated pregnancies by ELISA in a larger sample set.
- Evaluation of clinical metadata and pregnancy outcome for the subset of women identified with an increased plasma concentration of platelet basic protein at 15 weeks gestation.
- Investigation and evaluation of the application of platelet basic protein to clinical risk prediction for pre-eclampsia alongside other biochemical markers and clinical risk factors using principle component analysis.

6.2 Summary
This investigation suggests that measurement of platelet basic protein at 15 weeks gestation is a novel candidate predictive marker for pre-eclampsia. Further work is required to determine the clinical significance of this protein in a larger and more diverse sample group. If successful, platelet basic protein has the potential to be developed into an early pregnancy screening test, alongside other clinical, biophysical and biochemical markers, to identify healthy nulliparous women at risk of developing pre-eclampsia.
References


Baker, P. R. and K. R. Clauser. "http://prospector.ucsf.edu." Reference only


Proteomics 11(16): 3415-3419.


Molecular & Cellular Proteomics 6(7): 1170-1182.


Analytical Chemistry 81(9): 3462-3470.


BMJ 309: 1395-1400.


Seminars in Perinatology 33(3): 130-137.


Myers, J. (2005). Circulating Factors in Pre-eclampsia. Faculty of Medical and Human Sciences, School of Medicine, Manchester, University of Manchester. PhD: 321.


gonadotropin, activin A and inhibin A at 22 + 0 to 24 + 6 weeks’ gestation."


Appendix: Glossary of ProteinPilot (version 3.0) terms.

(Adapted from ProteinPilot (version 3.0) Software Help; Applied Biosystems, Warrington, UK).

Bias Correction
- Normalisation factor applied to each iTRAQ reporter ion

Compensation applied during calculation of quantification ratios to account for unequal mixing during combination of the labelled sample; based on the assumption that the majority of proteins do not change in expression.

Denominator
- iTRAQ label used to calculate relative quantification ratios

Defined by the user as the iTRAQ label representing a control or reference sample. Denominators were iTRAQ label 113 representing either superpool reference sample.

Error Factor
- A measure of the error in the average relative quantification ratio and indication of data quality

Statistic used to calculate the non symmetrical error in the relative quantification ratio derived from the 95% confidence error. The error factor decreases as the number of ratios used to calculate the average ratio increases and as the consistency of ratio measurement increases. The error factor is used to calculate the confidence intervals of the ratios. Perfect data has an error factor of 1.

False Discovery Rate (FDR)
- Indication of the rate of incorrectly identified proteins

False discovery rate was calculated by searching all peptide data against a concatenated database containing both forward and reversed protein sequences. The FDR is expressed as a percentage of reverse identifications made of the total identifications (Elias and Gygi 2007).

Percentage Coverage
- Amino acid sequence coverage

The percentage of matching amino acids from identified peptides, divided by the total number of amino acids in the sequence. The confidence in the amino acid is stated in brackets and is 0 when no figure is stated, i.e. % Coverage (50) uses peptides with a confidence greater or equal to 50%.
ProtScore Unit

- User friendly expression of the protein confidence.

The ProtScore is calculated by transformation of the percent confidence of the protein identification i.e. if a protein is identified by 5 peptides each with a confidence of 99% the confidence in the protein is 99.99999999% and the ProtScore is 10.

Relative Quantification Ratio

- A weighted average ratio for the protein identified measured against a chosen denominator.

An average weighted ratio calculated by ProteinPilot from the measurement of iTRAQ labels during MS/MS mode. The average ratio includes correction for experimental bias and uses peptides which are unique to that protein, identified with greater than 1% confidence and with a summed signal-to-noise ratio for peak pairs greater than 9.

Total Protein Score

- A measure of the total evidence for a detected protein.

The total protscore value calculated from the confidence values of the peptides detected for the protein.

Unused Protein Score (UPS)

- Indication of confidence in an identification of a protein.

The Unused Protein Score is a measure assigned by ProteinPilot summing the score of the peptides which are unique to that protein. It also takes into account where the protein was ranked in the list of detected proteins, the number of spectra the protein explains and the number of spectra which could be assigned to other proteins. It is reported in ProtScore Units. A protein with one 95% confidence peptide must have a UPS of 1.3 or above.

Winner Protein

- The protein believed to be present in the sample.

Protein groups contain ‘winner’ proteins which are the primary identified protein and for which there is the greatest amount of evidence not explained by better proteins and ‘competitor’ proteins for which there is less evidence.